Dissertation

Submitted to the

Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

Felix Geist (Dipl. Molekularmediziner)

born in Lüneburg, Germany

Oral examination: 4th June 2019

Identification of predictors of patient survival with TSPAN8 as a mediator of tumor aggressiveness in clear cell renal cell carcinoma

Referees: Prof. Dr. Andreas Trumpp

Dr. Michael Milsom

Abstract

With the help of a patient-derived clear cell renal cell carcinoma (ccRCC) model system previously established in our laboratory, which recapitulates the heterogeneity of the originating tumor, we were able to study ccRCC on a functional level[1,2]. In five rounds and in four biological replicates of an *in vivo* selection, we transplanted lung metastases of orthotopically transplanted tumor cells into the renal capsules of NOD scid gamma (NSG) mice. The tumor was enriched for cells with increased growth and higher metastatic potential compared to the initial heterogeneous population. Comparative gene expression analysis revealed candidate genes associated with enhanced malignant growth and metastasis. Least absolute shrinkage and selection operator (LASSO) regression identified a gene signature that can robustly predict cancer specific patient survival. The prognostic power of our signature was additionally verified in independent patient cohorts suggesting that this approach leverages efficient stratification of patients into distinctive risk groups. Intra- and intertumor heterogeneity remains a clinical challenge as estimated survival rates could vary substantially when comparing different tumor regions.

Tetraspanin-8 (*TSPAN8*) was identified as one of the hallmark genes in the generated ccRCC signature and is known to alter cellular signaling. Therefore, we hypothesized that TSPAN8 contributes to tumor aggressiveness and thus to growth and metastasis of ccRCC. In fact, in knockdown and overexpression xenografts experiments, we could confirm an essential role for tumor aggressiveness *in vivo* suggesting that TSPAN8 is an attractive target for treatment of clear cell renal cell carcinoma.

Zusammenfassung

Ein zuvor in unserem Labor etabliertes patientenbasiertes Modellsystem des klarzelligen Nierenzellkarzinoms (ccRCC), welches die Heterogenität des Ursprungtumors widerspiegelt, ermöglichte es uns, ccRCC auf funktioneller Ebene zu untersuchen[1,2]. In vier biologischen Replikaten und in jeweils fünf *in vivo* Selektionsrunden wurden Lungenmetastasen orthotopisch transplantierter Tumorzellen in die Nierenkapseln von NOD scid gamma (NSG)-Mäusen retransplantiert. Dabei wurde der Tumor, verglichen mit der anfänglich heterogenen Population, für Zellen mit erhöhtem Wachstums- und Metastasierungspotenzial angereichert. Durch eine vergleichende Genexpressionsanalyse wurden Kandidatengene identifiziert, die mit bösartigen Tumorwachstum und Metastasierung assoziiert werden. Wir konnten mit Hilfe einer LASSO-Regression (Least Absolute Shrinkage and Selection Operators) eine Gensignatur erstellen, die das krebsspezifische Patientenüberleben zuverlässig vorhersagen kann. Die prognostische Aussagekraft unserer Signatur wurde zusätzlich in unabhängigen Patientenkohorten überprüft, was darauf hindeutet, dass ihr Einsatz eine effiziente Stratifizierung der Patienten in unterschiedliche Risikogruppen ermöglicht. Die Tumorheterogenität zwischen den Patienten und innerhalb des Tumors selbst bleibt jedoch eine klinische Herausforderung, da die geschätzten Überlebensraten beim Vergleich verschiedener Tumorregionen stark variieren können.

Tetraspanin-8 (*TSPAN8*) ist eines der wichtigsten Gene in der ccRCC-Signatur und dafür bekannt, die zellulären Signalwege zu modulieren. Daher haben wir angenommen, dass TSPAN8 zur Tumoraggressivität und damit zum Wachstum und Metastasierung von ccRCC beiträgt. Tatsächlich konnten wir durch den Knockdown und die Überexpression des Gens im Xenograft Modell dessen wesentlichen Einfluss auf die Aggressivität des Tumors *in vivo* bestätigen. Dies deutet darauf hin, dass TSPAN8 ein attraktives Ziel für die Behandlung von klarzelligem Nierenkrebs ist.

Declarations

The work presented in this dissertation was performed from September 2014 until March 2019 at the German Cancer Research Center (DKFZ, Heidelberg) and the Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM, Heidelberg) under the supervision of Prof. Dr. Andreas Trumpp and Dr. Martin Sprick.

Declarations according to § 8 (3) c), d) and h) of the Doctoral Degree Regulations:

a) I hereby declare that I have written the submitted dissertation myself and in this process have used no other sources or materials than those explicitly indicated.
b) I hereby declare that I have not applied to be examined at any other institution, nor have I used the dissertation in this way or any other form at any other institution as an examination paper nor submitted it to any other faculty as dissertation.
c) I hereby consent to the verification of the dissertation by means of electronic data processing programs against standing scientific standards.

Contents

1 Introduction

1.1 Kidney Cancer – Epidemiology and Subtypes

Kidney cancer is the 14th most common cancer worldwide[3]. Approximately 400 000 new cases of kidney cancer were registered worldwide, an age-standardized risk of 4.4 per 100 000 inhabitants, with a mortality of approximately 175 000 every year. In countries with a high human development index (developed countries), incidence rates are higher (**Figure 1a, b**). The highest age-standardized incidence rate has been reported in the Eastern European countries Belarus, Latvia and Lithuania with 16.8, 15.2 and 14.8 cases per 100 000 inhabitants, respectively[3]. Two-third of the patients are male and average age of diagnosis is 64

(**Figure 2a**). The epidemiological data in Germany is comparable to other Western countries, with an incidence of 16.5 per 100 000 males and 7.8 per 100 000 females. Therefore, kidney cancer is the 8th and 10th most common cancer type in Germany, respectively[4]. In Germany, the relative 5-year survival rate that represents the cancer-specific survival (CSS) is 77 % and relative 10-year survival rate 71 % (**see section 1.3.1**).

Kidney cancer incidence has been reported to rise in the future (**Figure 2b**)[5,6]. This rise has been linked to an aging population[7], an increased prevalence of the main risk factors, such as obesity[8], smoking[9] and hypertension[10]. However, a detection bias has been discussed, which arises by improvements in screening methods and increasing awareness of kidney cancer[11].

1.1.1 Renal Cell Carcinoma

Approximately 90-95 % of kidney cancers are renal cell carcinomas (RCC)[12]. The classical clinical symptoms of RCC are a combined symptom triad of flank pain, hematuria and flank mass. These symptoms are present in only 10 % of the patients and only in advanced tumors. Other symptoms of renal cell carcinomas are anemia, weight loss, fever and hypertension. Many patients present themselves additionally with paraneoplastic syndromes that may mislead a correct diagnosis, such as hypercalcemia, erythrocytosis, Stauffer syndrome, amyloidosis and anemia. The complex clinical presentation of renal cell carcinoma explains late diagnosis, when 15-25 % of the patients have already developed local or distant metastases (**Figure 2c**)[4,13,14].

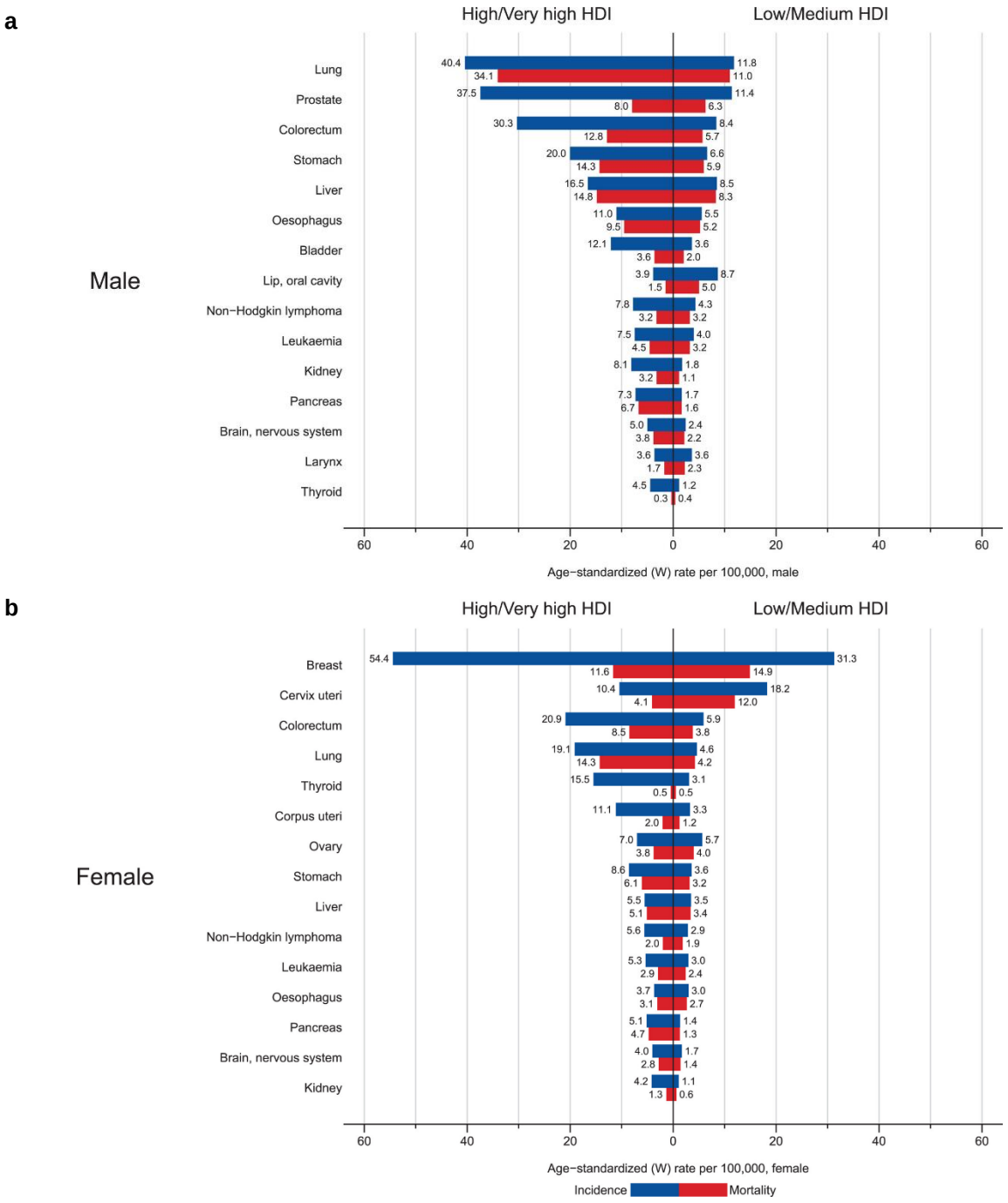


Figure 1 –Cancer Statistics

GLOBOCAN 2018 age-standardized cancer incidence and mortality rates between **(a)** men and **(b)** woman in regions of high/very-high human development index (HDI) and regions of low/medium HDI regions in 2018. In the 2018 edition of the Global Cancer Statistics, Bray et al. present the worldwide estimates of incidence and mortality for 36 types of cancer in a bar chart of descending order of the overall age-standardized rate[3]. 400 000 new cases of kidney cancer were registered worldwide with a mortality of approximately 175 000 every year. Countries with high human development index (developed countries) have higher incidence rates than countries with a low human development index.

Figure reprinted with kind permission of John Wiley and Sons (License Number 4494191362627).

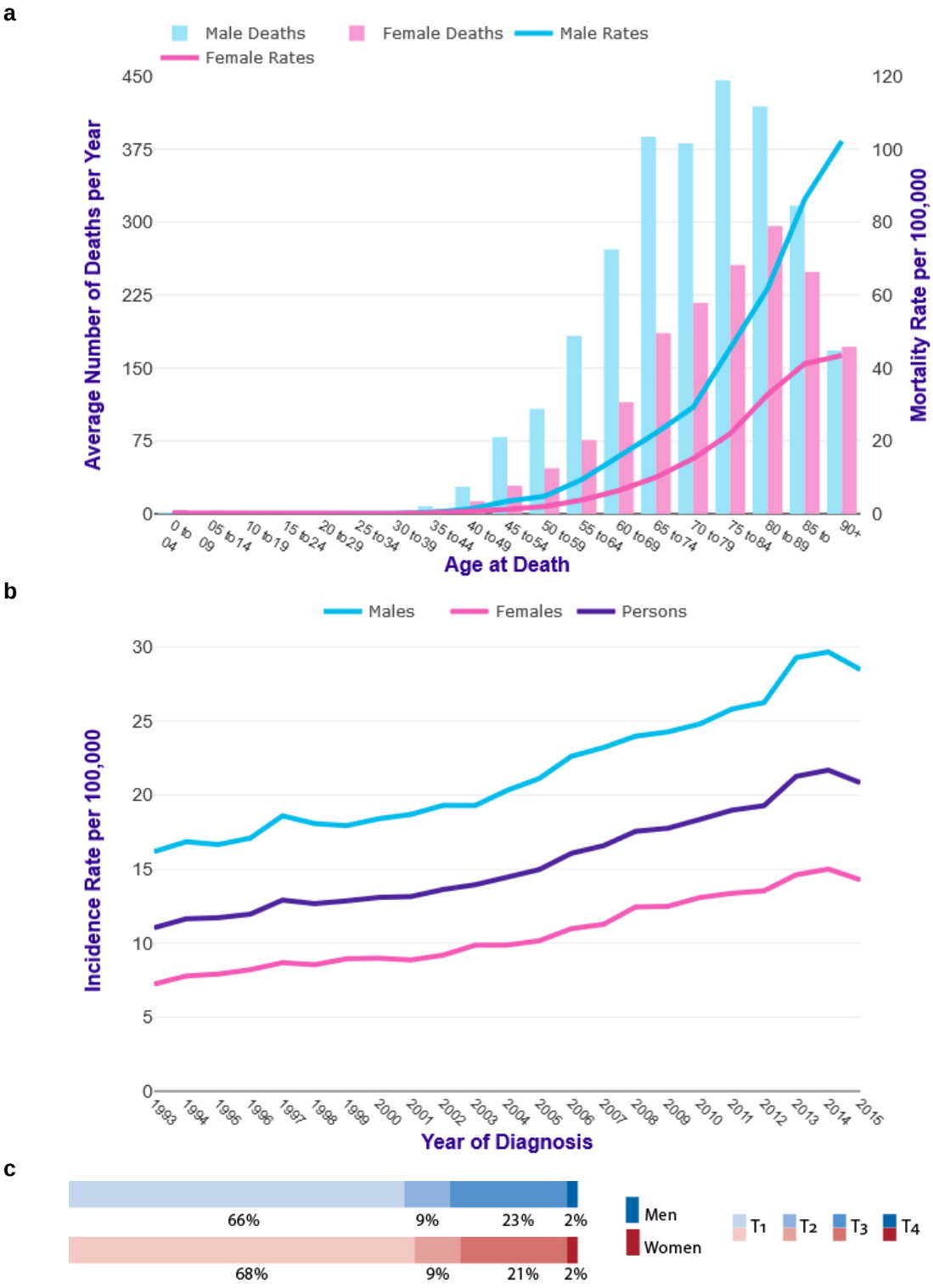


Figure 2 – Kidney Cancer Statistics

(a) Cancer Research UK reported age-specific mortality rate and average number of death by kidney cancer per year, UK, 2014-2016. Rates of kidney cancers in the UK are comparable with high HDI countries. This plot illustrates that kidney cancer is an age and gender related tumor. Credit: Cancer Research UK.

(b) Cancer Research UK reported increasing incidence rates of kidney cancer between 1990 and 2015. Credit: Cancer Research UK.

(c) Distribution of T-Stages of RCC at first diagnosis by gender, Germany 2013-2014.

Adapted from "Krebs in Deutschland", 2017, with kind permission of Zentrum für Krebsregisterdaten, Robert Koch-Institut[4]

Correct diagnosis is accomplished by ultrasonography, computed tomography scan or magnetic resonance imaging (MRI) and provides necessary information for surgical resections. Preoperative biopsy may validate diagnosis but is not used in German general practice[15,16].

1.1.2 Histologic Subtypes of RCC

RCC is subtyped into five major distinct histological subtypes. The most common subtype is the clear cell type with 60-70 % of the renal cell carcinomas. Other types are papillary (10-15 %), chromophobe (3-5%), oncocytomas (3 %) and tumors originating from the collecting duct (1 %). Additionally, some rare subtypes of RCC have been identified: Cystic-solid (1 %), Medullary (< 1 %), Xp11 translocations, "Mucinous tubular and Spindle cell", and "Associated with neuroblastoma"[17,18].

The RCCs of the different histological phenotypes differ in growth, aggressiveness, metastatic patterns, cell of origin and cytogenetics among others (**Table 1**).

Table 1 – Pathologic classification of the most common types of renal cell carcinoma[17,19]

Type	Features	Growth Pattern	Cell of Origin	Common Cytogenetics
Clear cell	Most common, majority sporadically	Acinar or sarcomatoid	Proximal tubule	3p-, 5q+, 8p-, 9p-, 14q-
Papillary	Bilateral and multifocal	Papillary or sarcomatoid	Proximal tubule	+7, +17, -Y, 12+, 16+, 20+
Chromophobe	Indolent course	Solid, tubular, or sarcomatoid	Cortical collecting duct	Hypodiploid
Oncocytic	Rarely metastasize	Tumor nests	Cortical collecting duct	Undetermined
Collecting duct	Very aggressive	Papillary or sarcomatoid	Medullary collecting duct	Undetermined

1.1.3 Clear Cell Renal Cell Carcinoma

Clear cell Renal Cell Carcinoma (ccRCC), the most common histologic variant of RCC, displays the worst prognosis in comparison to other histologic subtypes[20]. The most prominent feature of this subtype is the vast accumulation of lipids and glycogen in the cytoplasm of the epithelial cells. These molecules are washed out during the fixation process for the histological examination and therefore the cells appear "clear" (**Figure 3a**)[21]. ccRCCs may also contain various amounts of cells with granular eosinophilic cytoplasm (**Figure 3b**)[22].

ccRCCs grow in acinar, nested compact-alveolar and/or microcystic patterns, often in a combination in one single tumor. Typically, the epithelial cells are surrounded by a branched

network of connective tissue and blood vessels separated by a well-defined basal lamina (**Figure 3a,c**)[22].

Additional cells of the microenvironment are tumor-infiltrating immune cells, mainly T cells (51 %), tumor-associated macrophages (31 %), natural killer cells (9 %) and B cells (4 %)[23]. Also granulocytes are found in the ccRCC tumor microenvironment at low levels.

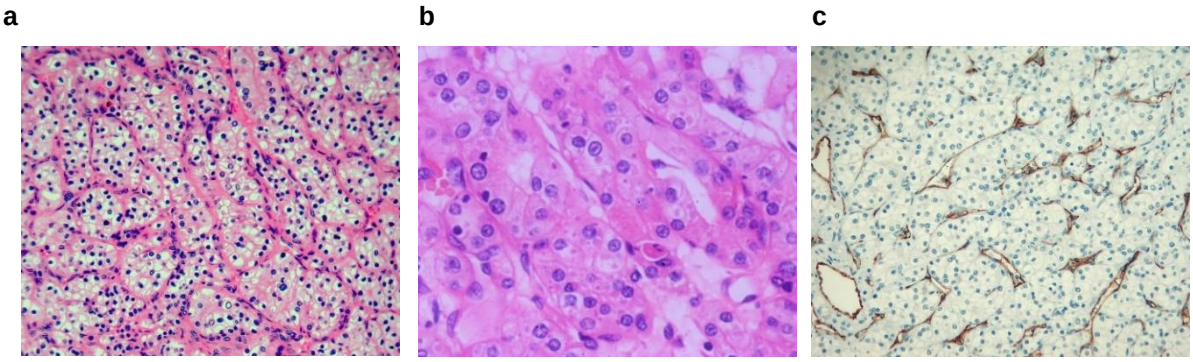


Figure 3 – Typical Histological Appearance of Clear Cell Renal Cell Carcinoma

(a) Histology of a ccRCC, fixated and stained with hematoxylin and eosin. The carcinoma is characterized by a clear, name-giving cytoplasm of the epithelial cells. The tumor is highly infiltrated by a branched network of vasculature.
(b) Variant of ccRCC with eosinophilic granules
(c) The extensive network of vasculature can be observed by staining for CD31
Image reproduced with permission from Medscape Drugs & Diseases (https://emedicine.medscape.com/), Pathology of Clear Cell Renal Cell Carcinoma, 2016, available at: https://emedicine.medscape.com/article/1612043-overview

Macroscopically, ccRCCs are generally solid and solitary tumors. In early stages they are often well surrounded by a capsule or a so called pseudocapsule comprised of compressed normal renal parenchyma and fibrous tissue[24]. Intracellular accumulation of lipids gives ccRCCs a yellow-golden color, which may be intersected by brown hemorrhages and grey fibrotic areas or necrotic areas[22]. Tumor sizes can range from microscopic lesions < 0.3 cm to maximal diameters large as 30 cm [25]. In average, patients present themselves with a tumor size of 6-7 cm at diagnosis. Modern imaging technologies and a higher awareness allow earlier tumor detection and therefore tumors are diagnosed at smaller sizes[20].

1.2 Pathophysiology of ccRCC

1.2.1 The role of VHL and HIF in ccRCC

ccRCC arises from proximal convoluted tubular epithelium of the nephron[26,27]. The majority of ccRCCs are sporadic cancers, but approximately 2-4 % of ccRCCs are caused by hereditary syndromes such as the rare Von Hippel-Lindau (VHL) syndrome and tuberous sclerosis[26]. In about 90 % of the tumors, the VHL pathway is affected: Mutations (53-37 %) of the tumor suppressor gene *VHL*, hypermethylation of its promoter (5-10 %), mutations in genes of the

VHL complex or loss of the short arm of chromosome 3, on which the *VHL* gene is located at 3p25 have been reported[28-31].

The gene product of *VHL* is part of the E3 ubiquitin ligase complex VCB-CR that targets hydroxylated proteins for proteasomal degradation[32]. Under normoxic conditions, prolyl hydroxylases (PHD) enzymes hydroxylate proline residues of the hypoxia-inducible transcription factors HIF-1α and HIF-2α, which are subsequently recognized by pVHL of the VCB-CR complex to start the ubiquitination and degradation process (**Figure 4a**)[33-35]. Deprivation of oxygen suppresses PHD dependent hydroxylation of HIFα resulting in HIFα stabilization and accumulation. Together with HIF-1β, HIF-α forms a heterodimer, which is shuttled into the nucleus, binds hypoxia response elements (HREs) and induces the transcription of more than 500 hypoxic target genes (**Figure 4b**)[36-38]. Thereby, pathways of glycolysis, lipogenesis, cell-cycle and angiogenesis are activated and apoptosis is inhibited[39-46].

Defective VCB-CR complexes lead to normoxic accumulation of HIF-α, irrespectively of its proline residue hydroxylation status. As a consequence, HIF-α-HIF-1β complexes become stabilized resulting in non-physiological HREs activation (**Figure 4c**)[35,47]. Although not expressed in normal tubular kidney epithelium, the expression and stabilization of the isoform HIF-2α plays an essential role as an early event in the tumorigenesis of ccRCCs[48-51]. In this aspect, ccRCC differs from other cancer types in which both subunits of HIF-α are associated with tumor progression[52]. In ccRCC, it has been noted that HIF-1α could even act as a tumor suppressor. HIF-1α loss of function mutations have been identified and additionally, knockdown and mutations in HIF-1α promote tumor growth in mouse models[53-55]. On the contrary, overexpression of HIF-2α in mouse ccRCC-xenograft models increases tumor formation[48,56], whereas its inhibition stops tumor progression *in vivo*[57]. Both isoforms share common DNA binding regions like the ubiquitous HIF target genes *CA9* and *GLUT1*. One of the target genes that is exclusively regulated by the heterodimeric HIF-2α-HIF-1β complex is *Cyclin D1*, a cell cycle promoting protein, which is expressed in early ccRCC lesions[48]. Moreover, c-Myc activity is solely enhanced by tumors only expressing HIF-2α[58,59]. HIF-2α stabilizes c-Myc/Max complexes post-transcriptionally due to direct binding, while HIF-1α represses c-Myc transcription factor activity by disrupting c-Myc/Max and c-Myc/Sp1 complexes. Hence, c-Myc promotes tumor progression by enhanced transcription of cell cycle activators such as cyclin D2, E2F1 and inhibition of cell cycle repressors such as cyclin-dependent kinase inhibitor 1 (p21) and cyclin-dependent kinase inhibitor 1B (p27)[60]. Another gene that is upregulated by HIF-2α is the embryonic stem cell factor *POU5F1* (Oct-4)[61].

The tumor-suppressor gene *TP53*, a hallmark gene of many tumor entities, is mutated in only 2.6 % of ccRCCs[62,63]. However, HIF-2α inhibition has been shown to promote p53 pathway activity[64]. Nevertheless, the crosstalk between HIF and TP53 is complex and its role in tumor progression controversially discussed[65].

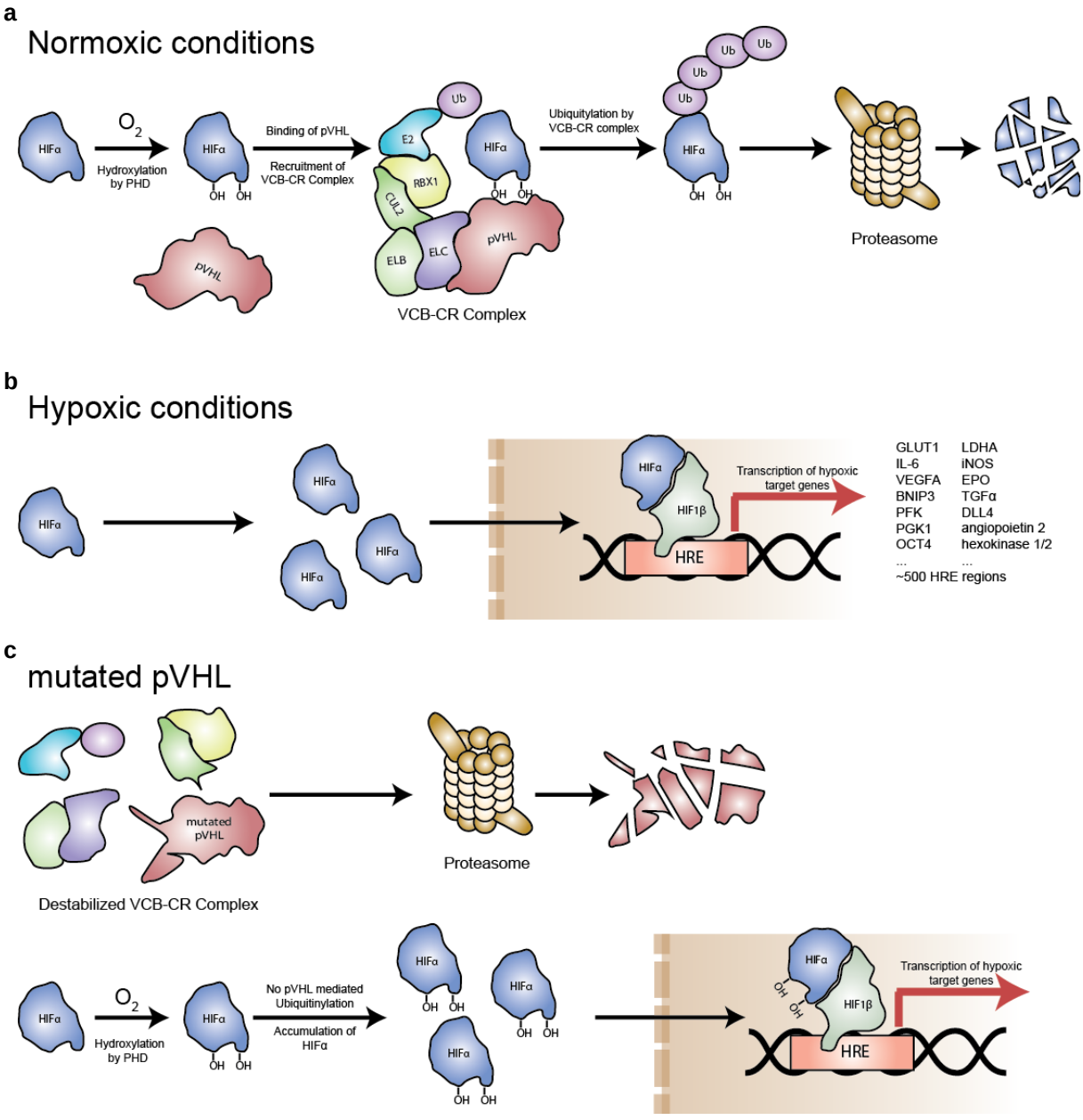


Figure 4 – pVHL and HIF in physiological conditions and disease

(a) In normoxic conditions, all isoforms of hypoxia inducible factor alpha (HIF-α) are oxygen dependent hydroxylated by prolyl hydroxylases (PHD). Subsequently, HIF-α is bound by pVHL and the E3 ubiquitin ligase complex VCB-CR is recruited. The complex ubiquitinates HIF-α and targets it for ubiquitin-mediated proteolysis.

(b) In hypoxic conditions, HIF-α is not degraded. It accumulates, forms a heterodimer with HIF1-β and translocates to the nucleus where it binds to hypoxia response elements (HREs) and regulates the transcription of more than 500 genes.

(c) Mutated pVHL is unable to bind hydroxylated HIF-α, which thereby accumulates and non-physiologically regulates HREs.

1.2.2 Genomic evolution of ccRCC

In many ccRCC patients, loss of chromosomal arm 3p, which harbors *VHL*, is a hallmark event that occurs early in childhood to adolescence (**Figure 5**)[62,66,67]. Often, a chromothriptic rearrangement between 3p and 5q is followed by 3p loss and 5q gain. This unbalanced translocation can be found in up to 60 % of all ccRCC patients[67-70]. Following Knudson's two-hit hypothesis for tumor suppressor genes[71], the subsequent mutation or epigenetic inactivation of the second *VHL* gene on the other allele of 3p is often found on the trunk of the phylogenetic tree that describes the ccRCC tumor evolution[72,73].

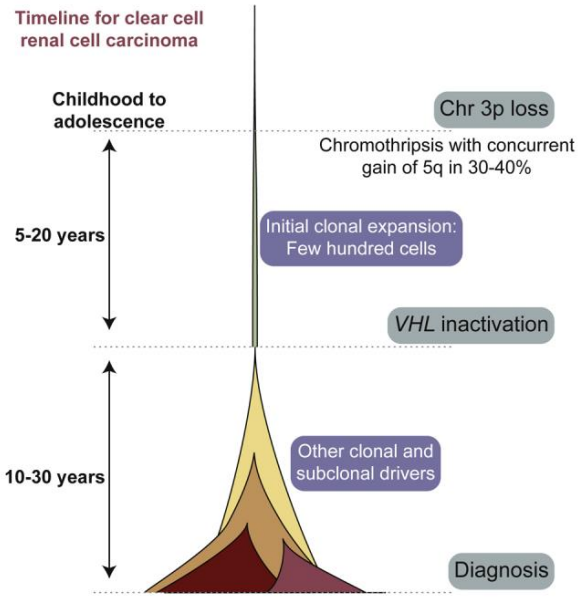


Figure 5 – Chronologic sequence of genomic events in ccRCC development

Already in childhood and adolescence, loss of the short arm of chromosome 3 by a chromotrypsis event is often the initial event of ccRCC development. Loss of 3p leads to a loss of heterozygosity for many tumor suppressors. Years later, inactivation of the second *VHL* allele is found clonal in most tumors, followed by subclonal mutations. *From Mitchell et al.*[67]

The chromosomal deletion of 3p leads to a loss of heterozygosity (LOH) of many genes. Next to *VHL*, 3p harbors genes like Polybromo-1 (*PBRM1*), BRCA1-associated protein (*BAP1*), and Su(var), Enhancer of zeste, Trithorax-domain containing 2 (*SETD2*), which are mutated in ~50 %, ~15 % and ~15 % of the patients, respectively (**Figure 6a**)[74-77]. While 3p loss and inactivation of *VHL* are ubiquitously found in ccRCC, mutations in *PBRM1*, *BAP1* and *SETD2* are often found in spatial separated subclones. They are subclonal driver mutations of branches of a phylogenetic tree describing the tumor evolution[73,78-82]. The branched evolution of the tumor leads to a vast intratumor heterogeneity (ITH) with a median of 4 subclones per tumor that complicates stratification of patients for targeted treatment according to a single biopsy (**Figure 7**). In fact, the number of subclones correlates with tumor stage and grade but only with a tumor of a size up to ~10 cm [28].

Turajilic *et al.* identified seven major evolutionary subtypes of ccRCC subclones by multiregion sequencing[28]. The first subtype, often found in early stage tumors, is solely driven by VHL inactivation and has a stable genome (**Figure 6b-1**). The second subtype has an

additional *BAP1* mutation driving tumor progression to high grade tumors with increased genome instability (**Figure 6b-2**).

Instead of *BAP1*, *PBRM1* drives progression of three other distinct evolutionary subtypes with high subclonal heterogeneity. Subsequent mutation of *SETD2*, often mutated in parallel evolution, results in slow tumor progression with high intratumor heterogeneity (> 10 clones per tumor) (**Figure 6b-3**). Evolutionary subtypes that show PI3K/AKT/mTOR pathway activation (**Figure 6b-4**) or subclonal somatic copy number aberrations following *PBRM1* mutation are often low grade and have a good prognosis (**Figure 6b-5**). Common subclonal chromosomal events are 14q loss (with *HIF1A*), 9p loss (with *CDKN2A*) and 4q loss (with *CXXC4*) (**Figure 6c**).

In the subtypes described so far, *BAP1* and *SETD2* or *PBRM1* mutations are mutually exclusive[28,77]. However, such mutations are found together in advanced tumors of the last evolutionary subtype defined by Turajulic *et al.* This "multiple clonal driver" subtype with *VHL*, *BAP1*, *PBRM1*, *SETD2* or *PTEN* clonal mutations shows low intratumor heterogeneity but high chromosomal complexity (**Figure 6b-6**).

The last evolutionary subtype is a highly proliferative *VHL* wild-type tumor with high genome instability (**Figure 6b-7**). This rare subtype (10-18 %) is characterized by a hyperdiploid karyotype with common amplifications of chromosome 2, 3q, 7, 8q, 12, 16, 20, 21 and loss of 9p[83]. It often lacks the classic 3p21-25 deletion and shows rapid tumor progression as well as sarcomatoid differentiation[84].

The different evolutionary subtypes were grouped by Mitchell *et al.* according to their evolution into: "Punctuated Evolution" with rapid progression ("multiple clonal driver", "BAP1 driven" and "VHL wildtype"), "Branched Evolution" with attenuated progression (PBRM1 driven tumors) and "Linear Evolution" that may represent early tumor stages (**Figure 6b**)[67].

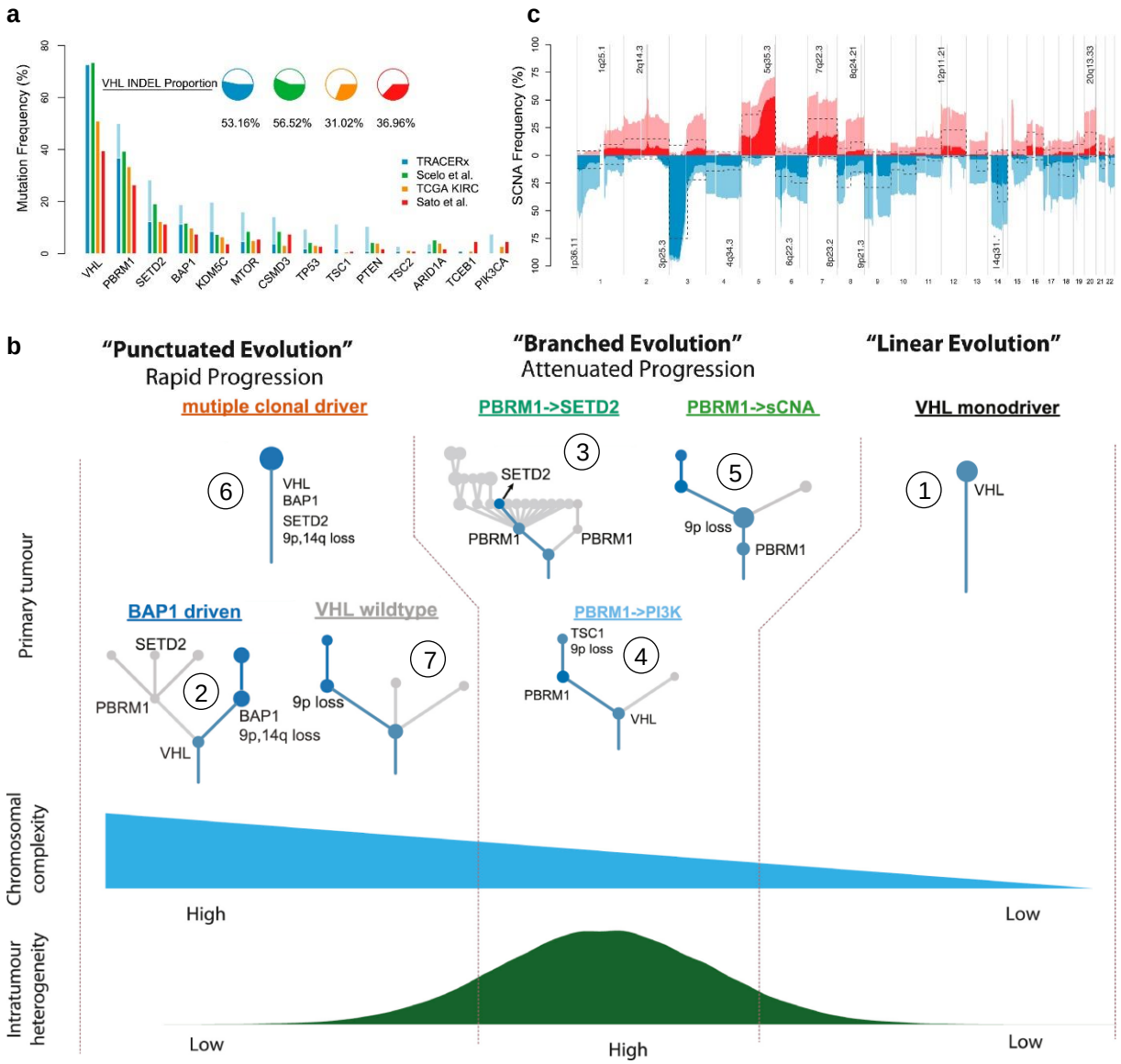


Figure 6 – Somatic drivers of ccRCC

(a) Mutation frequency of somatic drivers of ccRCC in the TRACERx Renal cohort, TCGA-KIRC, Sato *et al.*[31] and Scelo *et al.*[85] ccRCC studies. Dark shaded columns indicates clonal and light shade subclonal mutations in the TRACERx study. *Adapted from Turajilic et al.*[28]

(b) Evolutionary subtypes of primary tumors of ccRCC subclones, grouped according to chromosomal complexity. Exemplary chronology sequence of mutations and somatic copy number alterations (SCNA) is annotated from the bottom to the top. Genomic alterations and driver mutations found in one exemplary tumor subclone are colored in blue for each evolutionary subtype. BAP1-driven tumors for example harbor clonal *VHL* inactivation, following subclonal *BAP1* mutation, 9p and 14q loss. ccRCC subclones that contain the sequence of VHL→PBRM1 mutations tend to have higher intratumor heterogeneity. The length of the path that describes a tumor subclone is meaningless. *Adapted from Turajilic et al.*[86]

(c) Frequencies of somatic copy number alterations (SCNAs) in the TRACERx renal cohort. Red marks clonal copy number gains and blue marks clonal copy number losses. Light shades are subclonal changes in copy numbers. The dotted line indicates SCNAs in the TCGA KIRC dataset. *Adapted from Turajilic et al.*[28]

PBRM-1, the second most common mutated gene in ccRCC, encodes for the nucleosome remodeling protein BAF180. It is part of the chromatin-remodeling complex Switch/Sucrose nonfermentable (SWI/SNF) and associated with cell cycle control[87,88]. Unchecked cell cycle control can also result from mutations in the tumor suppressor gene *BAP-1* that together with the host cell factor-1 (HCF-1) inhibits cell proliferation[75].

The chromatin remodeler SETD2 has a dual function. On the one hand, SETD2 is important for nucleosome stabilization, RNA polymerase II-mediated transcriptional elongation and splicing by trimethylation of H3 histones on lysine 36[89,90]. On the other hand, the loss of SETD2 mediated microtubule methylation during mitosis results in mitotic defects with increased genomic instability[91].

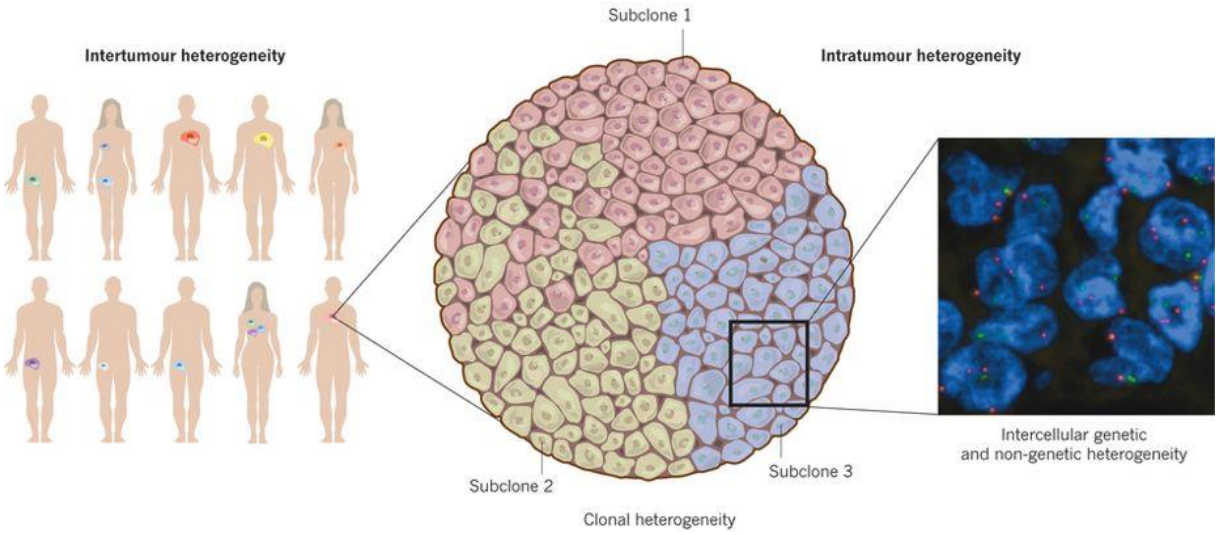


Figure 7 – Schematic overview of intertumor and intratumor heterogeneity

ccRCCs show vast genetic and phenotypic heterogeneity between patients (intertumor heterogeneity). Although the majority of patients (~90 %) harbor the trunk mutation *VHL*, subsequent tumor development varies greatly between them. Tumor evolution of patients is never identical, even though they are all classified histologically as ccRCC. The tumors themselves may consist of several subclones (intratumor heterogeneity), of which cells of the subclones may be mixed (subclone 1 and 2) or separated spatially from other subclones (subclone 3). *Figure reprinted with kind permission of Springer Nature (License Number: 4514240325627)*

1.2.3 mTOR signaling in ccRCC

In many ccRCCs, activation of the PI3K/AKT/mTOR pathway plays a central role in enhancing cell growth and metabolism (**Figure 8**). Phosphatidylinositol-3-kinase (PI3K or p110-α) is phosphorylated by receptor tyrosine kinases (RTKs), such as epidermal growth factor (EGF) receptor, Platelet-derived growth factor (PDGF) receptor or vascular endothelial growth factor (VEGF) receptor family members after extracellular stimulation[92]. PI3K phosphorylates phosphatidylinositol-4,5-bisphosphate (PIP2) in the plasma membrane to phosphatidylinositol-3,4,5-triphosphate (PIP3) recruiting protein kinase B (PKB/Akt) to the membrane. After successful recruitment, PKB/Akt is activated by phosphorylation by

mammalian target of rapamycin complex 2 (mTORC2 or PDK2) and subsequently by phosphoinositide dependent kinase 1 (PDPK1)[93-95]. This process is antagonized by dephosphorylation of PIP3 by phosphatase and tensin homolog (PTEN). PKB/Akt is a serine/threonine protein kinase that promotes proliferation and cell growth but inhibits apoptosis via various substrates. Through PKB dependent phosophorylation of tuberous sclerosis complex 2 (TSC2), the GTPase function of the heterodimer TSC1/TSC2 is inactivated. Subsequently, the Ras homolog enriched in brain (RHEB)-GTP complex is not hydrolyzed to RHEB-GDP and can activate the mechanistic target of rapamycin complex 1 (mTROC1)[96]. This in turn activates many processes required for cell growth and proliferation, such as increased lipid and protein synthesis[97].

HIF molecules are mTOR dependently transcribed and translated, promoting tumorigenesis of ccRCC[98,99]. As a negative feedback loop, they usually regulate the expression of *regulated in development and DNA damage responses 1* (*REDD1*) by a HRE. REDD1 activates TSC2 and inhibits mTORC1[99,100]. Nevertheless, *TSC1* and *TSC2* are downregulated in *VHL* and *PBRM1* deficient kidney tumors in a JAK/STAT3 dependent manner and therefore, the negative feedback loop is inactive in ccRCCs[101]. Moreover, mTORC1 signaling in ccRCC is indirectly activated due to a HIF-2α specific upregulation of the amino acid carriers *SLC7A5* and *SLC43A1*, leading to an increased uptake of branched chain amino acids[102].

Glycolysis in ccRCC is increased by elevated import of glucose into the cell via HIF dependent upregulation of *GLUT1* and HIF-2α dependent c-Myc activation[103,104]. Increased glycolysis can activate mTORC1 via pyruvate kinase muscle isoform 2 (PKM2) dependent phosphorylation of the mTORC1 inhibitor proline-rich AKT substrate 1 (PRAS40)[105]. PRAS40 is also inactivated by PKB signaling[99].

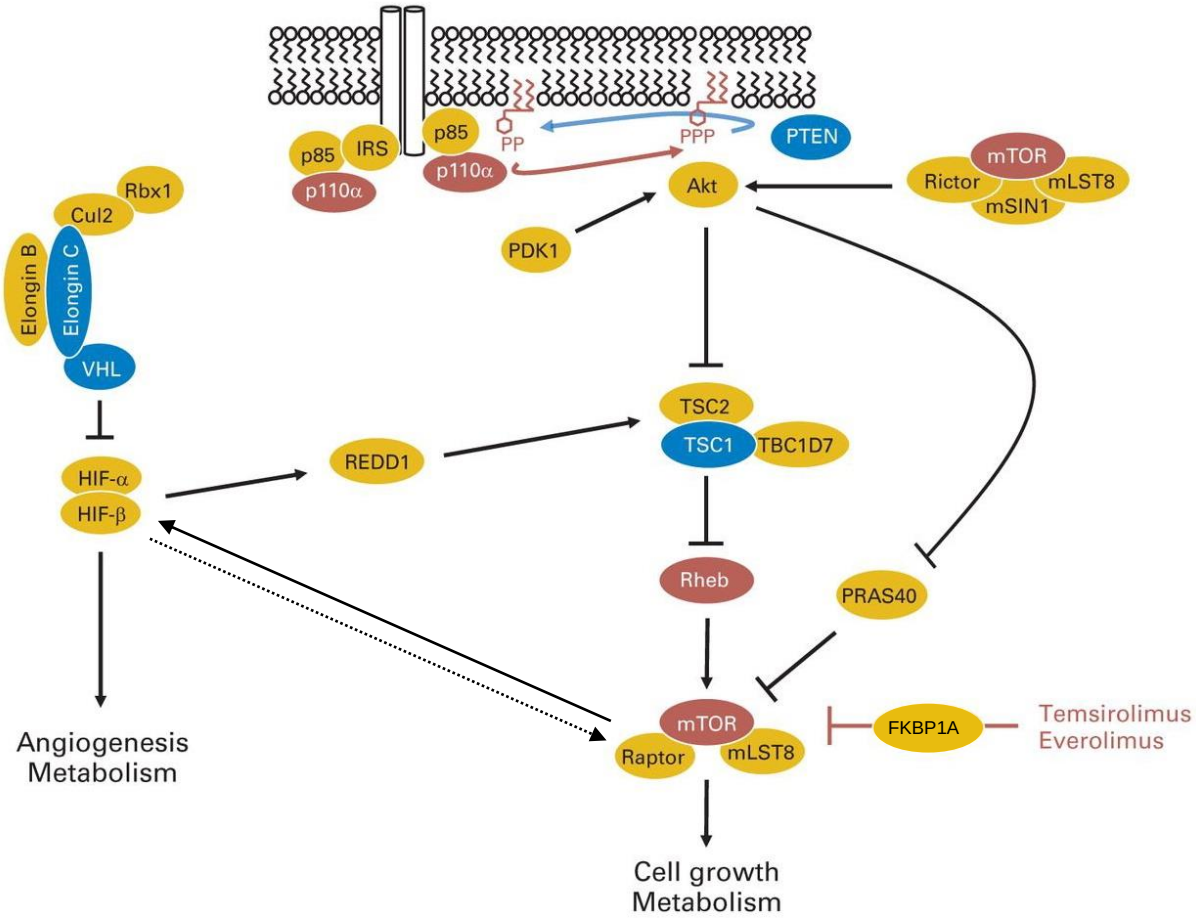


Figure 8 – Interplay of PI3K/AKT/mTOR and VHL signaling in ccRCC

Activation of the PI3K/AKT/mTOR pathway plays a central role in the development of ccRCC. Extracellular stimuli activate receptor tyrosine kinases (RTKs) that activate phosphatidylinositol-3-kinase (PI3K) by phosphorylation. PI3K in turn phosphorylates phosphatidylinositol-4,5-bisphosphate (PIP2) to phosphatidylinositol-3,4,5-triphosphate (PIP3), which recruits protein kinase B (PKB/Akt) to the membrane where it gets activated by mechanistic target of rapamycin complex 2 (mTORC2/PDK2) and phosphoinositide dependent kinase 1 (PDPK1/PDK1). Phosphatase and tensin homolog (PTEN) dephosphorylates PIP3 again to PIP2 and inhibits activation of PKB. Activated PKB induces proliferation, cell growth and inhibits apoptosis. It activates the mTOR pathway by inhibiting tuberous sclerosis complex 2 (TSC2). Thereby, the Ras homolog enriched in brain (Rheb) is not inactivated anymore and can activate mTORC1. In a hypoxic environment, mTORC1 is regulated by a negative feedback loop via upregulation of *HIF* and *regulated in development and DNA damage responses 1* (*REDD1*). This negative control slows down tumorigenesis of *VHL* mutated patients. Using small molecule inhibitors to chemically inhibit FK506 Binding Protein 1A (FKBP1A/FKBP12), mTORC1 signaling can be specifically blocked. Brown-colored oncogenes are commonly activated in ccRCC; blue-labeled tumor-suppressor genes are commonly inactivated in ccRCC. Dotted line indicates indirect activation of mTORC1 by HIF-mediated signaling pathways.
Adapted from Kucejova et al.[99]

1.2.4 Angiogenesis in ccRCC

Angiogenesis is the process of blood vessel formation. In physiological settings angiogenesis is tightly controlled and only active during embryogenesis, wound healing processes and menstruation[106]. New vessel formation can be induced by hypoxia and HIF signaling, via secretion of mitogens like the vascular endothelial growth factor (VEGF) and the platelet-derived growth factor (PDGF). VEGF binds to VEGF receptors on endothelial cells and induces proliferation and tube formation of the forming capillaries[107,108]. PDGF stimulates recruitment of vascular smooth muscle cells and pericytes to cover the newly formed vessels[109-111].

In 2000, Hanahan and Weinberg postulated sustained angiogenesis as one of the hallmarks of cancer[112]. As growing tumors need constant oxygen and nutrient supply, they initiate *de novo* angiogenesis by activating pro-angiogenic activators, such as VEGF, which is secreted from the tumor, and by suppressing the expression angiogenesis inhibitors like thrombospondin-1 (*TSP-1*)[113].

Harboring a hyperactive HIF signaling, *VHL* mutant ccRCC exhibit increased VEGF and PDGF secretion resulting in hyper-vascularization of the tumor (**Figure 3c**)[114]. This is further enhanced by an activation of the PI3K/Akt/mTOR signaling pathway, which leads to a downregulation of the anti-angiogenic factor TSP-1[115]. In late tumor development, hyperactivated angiogenesis signaling induces an undifferentiated microvasculature with bad prognosis[116].

1.2.5 Metastasis of ccRCC

Metastasis is a multi-step process by which a localized tumor disseminates into distant parts of the body[92]. First, the primary tumor invades the local tissue and breaches through the basement membrane. Next, single tumor cells or groups of cells intravasate into blood or lymphatic vessels and get transported through the body. Eventually, the tumor cells get trapped in microvessels and may extravasate into the surrounding tissue. These micrometastases can stay dormant for a long period of time or proliferate at one point to form macrometastases.

Metastasis of ccRCC follows the path of least resistance, most frequent along renal veins and the renal sinus, as migrating tumor cells do not have to cross connective tissue[117]. Lymphatic metastasis is less frequent in renal cell carcinoma than hematogenous[118] explaining the distribution of distant metastatic sites: Lung (45 %), bone (30 %), lymph node (22 %), liver (20 %), adrenal (9 %), brain (8 %) and retroperitoneum (7 %)[119]. With only 16.6 % relative

survival probability, metastasis deteriorates the relatively good prognosis of kidney cancer (**see section 1.3.1**).

Turajlic *et al.* analyzed the metastatic potential of ccRCCs[86]. They found that tumors of the "punctuated evolution" type (tumors with low ITH and elevated SCNAs) were more prone to metastasize rapidly, whereas tumors of the "branched evolution" type (tumors with high ITH but low SCNAs) showed an attenuated progression (**Figure 9a**). The authors identified driver events of many metastasis seeding subclones, mainly of the "punctuated evolution" subtype, in copy-number losses of either chromosome 9p or 14q (**Figure 9b**). Furthermore, 9p loss was pre-determining the prognosis of patients' survival. This chromosome encodes for the cell cycle inhibitor *CDKN2A* and 14q hosts the ccRCC tumor suppressor *HIF1A* gene among others. Tumors of the "linear evolution" did not metastasize in their analysis. However, more research needs to be done to identify genes and pathways associated with metastases (**see chapter 1.2.1**).

Early dissemination of metastases has been observed in breast cancer[120-122]. This process, by which already early lesions seed cells that form subsequent metastases has in general not been observed in ccRCCs[86]. Usually, metastatic clones develop from late subclones of the primary tumor or from already seeded metastases (**Figure 9b**). However, Turajlic *et al.* identified patients with late pancreatic metastases that appeared in median 15 years after the first presentation. The clones that seeded these metastases most likely developed already from primitive ancestral clones and lacked the metastatic 9p loss. Therefore, the pancreas may represent a more permissive metastatic niche for dormant ccRCC cells than other organs[123].

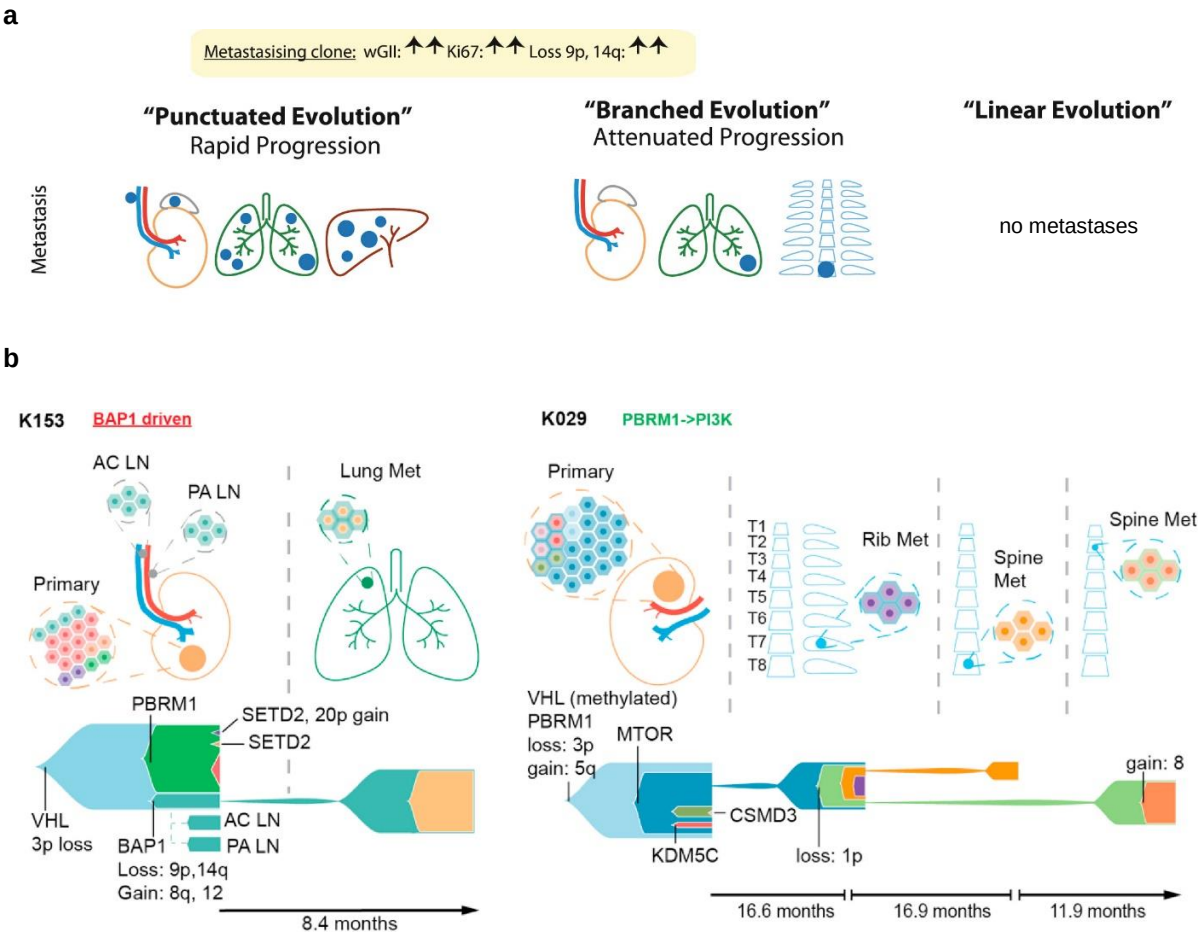


Figure 9 – The evolution of ccRCC metastases

(a) Schematic of the evolutionary subtypes and their potential to metastasize. The "punctuated evolution" subtype depicts rapid progression with many metastases at various locations and the "branched evolution" subtype shows slow progression with only few, singular metastases. The "linear evolution" subtype metastasizes rarely.

(b) Summary of two selected cases of the metastatic route of ccRCC. The fishplots illustrate disease evolution by annotating driver events. For example, the BAP1 driven route of metastasis of patient K153 (punctuated evolution) shows that a subclonal BAP1 mutation together with a loss of 9p and 14q and a gain of 8q and 12 led to the growth of a distinct subclone. This subclone had already metastasized into two lymphnodes (aorto-caval (AC) and para-aortic (PA)) when the patient underwent surgery. 8.4 month later, this subclone was found in the lung. The patient K029 with a PBRM1 → P13K evolutionary subtype (branched evolution) showed a slower progressing metastasis development. At 6the time of the first surgery, no metastases where found. After a period of 16.6 month, the patient had a metastasis in the ribs, which originated from an mTOR driven subclone of the primary tumor. Twice more metastasis where excised, both from the spine and both of the metastases originated from the primary metastasis.

Figures adapted from Turajlic et al.[86]

In many tumors, hypoxia is a key microenvironmental factor that drives migration and metastasis[124]. HIF activation regulates many steps of the metastatic cascade. For example, it induces the epithelial to mesenchymal transition (EMT) pathway[125,126].

EMT is essential in early embryogenesis for processes such as gastrulation and neural crest formation, but also for early steps of tumor cell dissemination[127,128]. Thereby, epithelial cells lose contact to their neighboring cells and the basal lamina, mainly due to a loss of E-cadherin-mediated adherens junctions. Subsequent loss of cellular polarity and gain of a mesenchymal phenotype allows the tumor cells to migrate and invade into surrounding tissue. Important transcriptional regulators of this process are Twist Family BHLH Transcription Factor 1 (TWIST1), Snail Family Transcriptional Repressor 1 (SNAI1 or SNAIL) and Zinc finger E-box-binding homeobox 1/2 (ZEB1/2).

Other processes next to EMT of the metastatic cascade are also mediated via hypoxia signaling. Local invasion of migrating tumor cells, both at the primary tumor and metastatic site, is supported by modifying the extracellular matrix. Key enzymes of this process are HIF-1α dependent secretion of the extracellular lysyl oxidase (LOX)[129], matrix-metalloproteinase-1 and -2 (MMP1/2), urokinase-type plasminogen-activator receptor (uPAR) and cathepsin D (CTSD)[130].

Hypoxia induced overexpression of VEGF not only stimulates angiogenesis, but also increases vascular permeability and interstitial fluid pressure[131]. Thereby, tumor cells can intra- and extravasate more easily. It has been shown that organ specific homing of metastatic tumor cells can be explained by chemokine-receptor interactions. Hypoxia dependent upregulation of the chemokine receptor CXCR4 allow the tumor cells to home to organs that express the ligand stromal cell-derived factor-1alpha (SDF-1α), such as the lung and the bones[132].

In summary, many steps of the multi-step process of metastases are regulated by HIF and the hypoxia pathway which is omnipresent in *VHL* deficient ccRCCs. Nevertheless, the period until the first metastasis appears in ccRCC is relatively long in comparison to other tumors and except for pancreatic metastases, additional mutations and genomic events have to occur before the tumor metastasizes to distant organs. Especially the loss of chromosome 9p and *CDKN2A* seems to play a pivotal role in the initiation of metastasis in ccRCC.

1.3 Management of clear cell Renal Cell Carcinoma

1.3.1 Prognostic stages, grades and signatures

In most classification systems, assessing patients' survival no distinction is made between different kidney cancers and histological subtypes. In the following, the most common systems in clinical use are summarized.

1.3.1.1 Staging of kidney cancer:

The American Joint Committee on Cancer (AJCC) makes use of the TNM classification to stage kidney cancer into four stages. The system differentiates primary tumor size and extent (T), lymph node infiltration (N) and distant metastasis (M). The 8th edition of the TNM classification uses the system shown in **Table 2**[133,134].

Table 2 – 8th edition of the AJCC TNM classification system for kidney cancer

T	**TX**	Tumor size of the primary tumor cannot be evaluated
	T0	No evidence of primary tumor
	T1a	Primary Tumor diameter < 4cm
	T1b	Primary Tumor diameter > 4cm but ≤ 7cm
	T2a	Primary Tumor diameter > 7cm but ≤ 10cm
	T2b	Primary Tumor diameter > 10cm, limited to the kidney
	T3a	Primary tumor extends into the renal vein or its segmental branches, or invades the pelvicalyceal system, or invades perirenal and/or renal sinus fat but not beyond Gerota's fascia
	T3b	Primary tumor extends into vena cava below the diaphragm
	T3c	Primary tumor grossly extends into vena cava above diaphragm or invades wall of vena cava
	T4	Tumor invading beyond Gerota's fascia
N	**NX**	Lymph node involvement cannot be evaluated
	N0	No regional lymph node metastases
	N1	Single or multiple regional lymph nodes involved (regional lymph nodes: hilar, caval, aortic)
M	**M0**	No distant metastases
	M1	Distant metastasis, including noncontiguous adrenal involvement

With the help of the TNM classification, the kidney tumor can be assigned to a prognostic cancer stage (**Table 3**), which, in most cases, determines the treatment decision and is one of the best prognostic factors for treatment success.

Table 3 – AJCC Prognostic Groups, stage specific 5-year relative survival probability kidney cancer 1998-2016, Munich Cancer Registry[135]

Stage	TNM	Alternative TNM	5-year Relative Survival Probability
I	T1, N0, M0		97.6 %
II	T2, N0, M0		88.6 %
III	T1-2, N1, M0	T3a-c, N0-1, M0	71.9 %
IV	T4	Any T, any N, M1	16.6 %

1.3.1.2 Fuhrman nuclear grading system

ccRCCs are graded based on the microscopic findings of nuclear morphology[136] (**Table 4**). Nevertheless, interobserver reproducibility of the Fuhrman grading is low. Therefore, researchers tried to simplify the system by combining Fuhrman grade 1 and 2 as well as grade 3 and 4 into a 2-tiered grading system[137] or by combining grade 1 and 2 with unchanged grade 3 and 4 into a 3-tiered system[138]. The simplified systems have the same accuracy relatively to the conventional grading system[139-141].

Table 4 – Fuhrman nuclear grading system, grade specific 5-year Cancer-Specific Survival Probability assessed by Gudbiartsson[25]

Grade	Nuclear characteristics	5-year Cancer-Specific Survival Probability
1	Nuclei < 10µm, finely round, granular chromatin, small nucleoli	87.3 %
2	Nuclei < 15 µm, finely granular chromatin, small nucleoli	70.5 %
3	Nuclei < 20 µm, oval, coarsely granular chromatin, prominent nucleoli	45.9 %
4	Pleomorphic nuclei, open chromatin, single/multiple macronucleoli	14.9 %

1.3.1.3 SSIGN Score:

To predict outcome of ccRCCs after radical nephrectomy, Frank *et al.* proposed a scoring system termed SSIGN score based on TNM stage, tumor size, nuclear grade and histological tumor necrosis (**Table 5**)[142]. Zigeuner *et al.* validated the SSIGN score on an external, single center cohort and suggest to apply the system routinely[143].

Table 5 – Estimated Cancer-Specific Survival rates according to SSIGN Scores[142]

	% estimated cancer-specific survival rates				
SSIGN Score	**Year 1**	**Year 3**	**Year 5**	**Year 7**	**Year 10**
0 - 2	99.7	98.7	97.3	96	93.6
3 - 4	98.1	93.4	89.8	84	77.9
5 - 6	92.9	83.8	74.1	65	57.3
7 - 9	76.5	46.9	38.6	29	25.9
≥ 10	43.3	21.9	19.2	19.2	19.2

1.3.2 Prognostic survival factors (MSKCC Score)

In addition to tumor stage and grade information, Motzer *et al.* categorized metastatic renal cell carcinoma patients according to five clinical risk factors[144]. Patients with limited levels of self-sufficiency (assessed by a so called Karnofsky performance status), increased serum lactate dehydrogenase (LDH) level, low hemoglobin levels, increased serum calcium levels and patients already nephrectomized displayed reduced survival expectations (**Table 6**). The risk assessment of the MSKCC Score is used to give patients with advanced tumors the optimal first line treatments taking adverse effects into account[16].

Table 6 – Relative survival rates of patients with metastasized ccRCC according to risk stratification by Motzer *et al.*[144]

	% survival rates		
Number of risk factors	**Year 1**	**Year 2**	**Year 3**
0	71	45	31
1-2	42	17	7
3-5	12	3	0

An interesting finding regarding an additional prognostic survival factor was observed by Albiges *et al.* They showed that obese patients (BMI >25) with metastatic RCC and targeted treatment showed improved survival[145].

1.4 Treatment of ccRCC

The gold standard treatment of localized renal cell carcinoma is surgical resection with potential curative outcome[16,146]. Depending on tumor size, stage, location and patient performance, either radical or partial nephrectomy (also called nephron-sparing) surgery is preferred.

Early stage Ia/b tumors are preferably laparoscopically or robotically resected while keeping most of the surrounding kidney intact[147,148]. Stage II and stage III tumors are generally removed by radical nephrectomy[16,146]. An adjuvant therapy is not recommended when imaging gave unsuspicious findings[149-151].

Advanced ccRCCs (stage IV, relapsed or recurrent disease) cannot be completely surgically removed. Therefore, patients have to undergo systemic treatment.

ccRCC tumors are highly resistant to chemotherapy with cytotoxic or cytostatic agents, showing only 5-10 % response rates and no improved overall survival[152-154]. Only isolated cases with sarcomatoid differentiated tumors show limited benefit of chemotherapeutics. The resistance to chemotherapy has been correlated to a high HIF-2α dependent expression of the ABC-transporter Multidrug resistance protein 1 (MDR1/p-glycoprotein)[155-158]. Another

resistance mechanism may be a PTEN dependent inhibition of p53-mediated apoptosis in ccRCC[159].

Besides chemotherapy, ccRCCs are also highly resistant to radiotherapy[160,161], which has been associated with HIF-2α mediated resistance to DNA damage and interferon signaling via STAT1[162-165]. As operations in the brain are complicated and risky, brain metastases are nevertheless treated with high doses of radiation to overcome potential resistance mechanisms[16,146]. Furthermore, radiotherapy of metastases is a palliative option with possible symptom reduction[166-168].

The guidelines for patients with advanced ccRCCs recommend a cytoreductive nephrectomy to reduce primary tumor mass as much as possible prior systemic treatment with additional resection of metastases[16,146,169]. However, a new trial result suggests that perioperative risks may outweigh the benefit of this intervention and recommend targeted therapy alone[170].

1.4.1 First generation of systemic treatment in advanced RCC: Immunotherapy

As chemo- and radiotherapy have been proven to give unsatisfactory results in the management of advanced ccRCC, patients were historically treated with interferon-alpha (IFNα) and high dose interleukin-2 (IL-2). This treatment option was considered, as some tumors were anecdotally found to regress in the absence of systemic treatment, implying that the immune system might have recognized them[171,172]. However, this immunotherapy has been found to be effective in only 15-25 % of the patients and some of them experienced strong adverse effects[173,174].

1.4.2 Second generation of systemic treatment in ccRCC: Tyrosine Kinase Inhibitors and mTOR inhibitors

A new era for the treatment of renal cell carcinoma emerged from the study of the role of *VHL* in ccRCC. Sorafenib was the first broad acting tyrosine kinase inhibitor (TKI) that blocks receptors of the pro-angiogenic factors VEGF and PDGF (**Figure 11**). This therapy doubled progression free survival of patients from 2.8 to 5.5 months and improved overall survival from 14.3 to 17.8 months[175]. Sorafenib was approved as first line treatment of advanced renal cell carcinoma by the FDA and the EMA in 2005 and 2006, respectively[176,177].

In the following years, more and more TKIs were approved by the regulatory authorities and especially sunitinib and pazopanib have emerged as the first choice in first line treatment of advanced renal cell carcinoma. Their efficacy is similar but pazopanib may show a better

safety profile[178]. Other approved TKIs are axitinib, cabozantinib, lenvatinib and tivozanib, all with slightly different tyrosine kinase binding capabilities.

Bevacuzimab is a monoclonal antibody that directly targets VEGF-A. It has been approved in the EU and 2009 in the US as combination therapy together with IFN-α2a as first line treatment option. However, an improvement over the traditional IFN monotherapy was only notable in the extended progression free survival time, whereas overall survival remained unchanged[179].

Another overstimulated signaling pathway in ccRCC is the PI3K/Akt/mTOR pathway. Blockage of the mTORC1 complex using temsirolimus and everolimus has been shown to improve patients' progression free survival for both agents from 1.9 months to 5.5 months[180,181]. Especially temsirolimus has been shown to be effective in high risk patients, whereas single-agent everolimus has been relegated back in the line.

The current approved treatment options for advanced RCC according to the German kidney cancer guidelines (2/2017) are summarized in **Figure 10**.

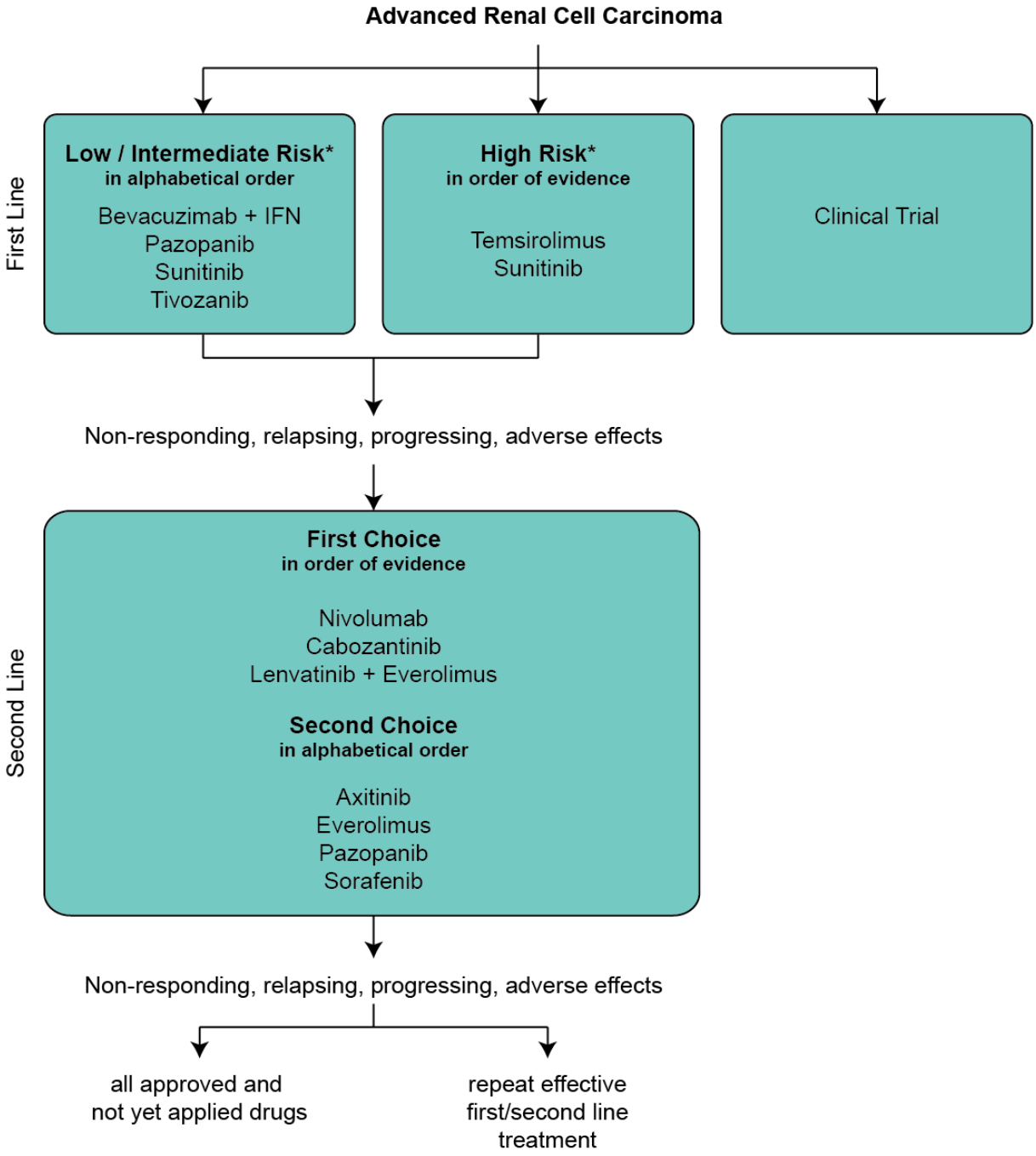


Figure 10 – Sequence of treatment approaches for the advanced renal cell carcinoma according to the German kidney cancer guidelines 2/2017 [16]

* Risk assessment according to the MSKCC Score (see **chapter 1.3.2**)

1.4.3 New treatment options on the horizon: Immune checkpoint inhibitors, combinational treatments and vaccines

Blockage of tyrosine kinases and the mTOR pathway has been shown to significantly improve survival of metastatic renal cell carcinoma patients. Nevertheless, clinical guidelines in the USA and Europe suggest to enroll patients into ongoing clinical trials instead of using current treatment options in order to identify more effective therapy options[16,146].

In 1992 PD-1 was identified as an inhibitory surface molecule on T cells[182] as binding of the ligand PD-L1 or PD-L2 leads to inhibition of T cell receptor (TCR) signaling[183]. A variety of tumors have been found to express PD-L1 on their surface[184], thereby repressing the anticancer T cell response. In renal cell carcinoma, PD-L1 is higher expressed on advanced tumors and associated with reduced survival[185].

Another checkpoint inhibitor under investigation is CTLA-4. CTLA-4 is expressed on T cells and T regulatory cells as response to TCR activation[186]. It binds CD80 and CD86 on antigen presenting cells with higher affinity than CD28, which acts as second signal for T cell activation, thereby reducing the signal amplitude of T cell priming[187,188]. Furthermore, CTLA-4 plays an essential role in the development of memory function and tolerance to self-antigens. There are further targets for immunotherapy in clinical development, such as CD25, OX40 [189].

The first approved monoclonal antibody as second line treatment for advanced renal cell carcinoma was the immune checkpoint inhibitor nivolumab, targeting PD-1[190]. It outcompeted the mTOR inhibitor everolimus in the Checkmate 025 phase III study, even though only in high risk MSKCC patients. In these patients, it could nearly double the median overall survival from 7.9 (everolimus) to 15.3 months (nivolumab)[191].

Already included in the most recent ESMO guidelines as first line treatment recommendation for intermediate and poor risk metastatic ccRCC, the immune checkpoint inhibitors are not yet recommended in the German treatment guidelines[192]. In the following paragraphs, current or recently finished clinical trials of substances with various cellular targets are summarized that may increase treatment options in the future.

The Checkmate 214 clinical phase III trial was the first first-line treatment trial with immune checkpoint inhibitors to show prolonged overall survival in advanced ccRCC[193]. Additionally, it was the first co-treatment trial in ccRCC with two different immune checkpoint inhibitors: Nivolumab (mAB αPD-1) together with ipilimumab (mAB αCTLA4) against a monotherapy with sunitinib (TKI). Especially high risk patients showed improved response rates and

overall survival (18-months OS 75 % vs 60 %). Still, adverse effects led to discontinuation of treatments in 22 % of the patients.

The IMmotion151 clinical phase III trial combined two compounds with different mode of action: The immune checkpoint inhibitor atezolizumab (mAB αPD-L1) with the anti-angiogenic bevacuzimab (mAB αVEGF-A). This combination of drugs was also compared to sunitinib in advanced ccRCCs as first line treatment option and progression free survival was significantly improved from 7.7 (sunitinib alone) to 11.2 months (atezolizumab + bevacuzimab)[194].

A combination of avelumab (mAB αPD-L1) and axitinib (TKI) in the treatment of naïve metastatic RCCs was tested in the JAVELIN Renal 101 trial[195]. This trial may have set a new standard in the management of advanced RCC: Compared to sunitinib, this combination nearly doubled the patients' progression free survival in all risk groups (13.8 vs 8.4 months), regardless the PD-L1 status. Interestingly, the safety of this combination showed fewer adverse events than treating patients with sunitinib alone.

A similar combination of drug targets is currently being tested in the KEYNOTE-426 trial with pembrolizumab (mAB αPD-1) and axitinib (TKI). Also in this trial, interim results suggested that PD-L1 expression on tumor cells had no effect on the efficacy of immune checkpoint inhibitor treatment, as progression free and overall survival did not significantly differ so far[196].

Currently, more clinical trials are ongoing, testing the efficacy of other monoclonal PD-1 and PD-L1 antibodies in the advanced renal cell carcinoma setting alone or in combination therapy[197-199].

Another treatment approach has been opened by testing vaccines in the metastatic renal cell carcinoma setting. However, IMA901, a multipeptide cancer vaccine as co-treatment to sunitinib in a first line setting tested in the IMPRINT trial did not improve patients overall survival[200].

The ADAPT trial was maybe the most personalized immunotherapy that has been tested in the renal cell carcinoma context so far. The trial tested, whether sunitinib in combination with patient derived tumor neoantigens primed dendritic cells would improve patients' survival. These cells (called AGS-003) were then given back to the patient. This trial failed, as it did not improve patients' survival[201,202].

At the time of this thesis, 166 interventional clinical trials were running for the metastatic RCC alone, with many new therapy approaches in early phases of clinical development[203]. In phase III studies are currently only immune checkpoint inhibitors as mono- and combinational-therapies in various combinations with already approved drugs to treat RCC. However, in phase I and II studies a broad variety of other compounds is being tested. One example is the specific inhibition of HIF-2α with the compound PT2977[204,205] (**see chapter 1.2.1**).

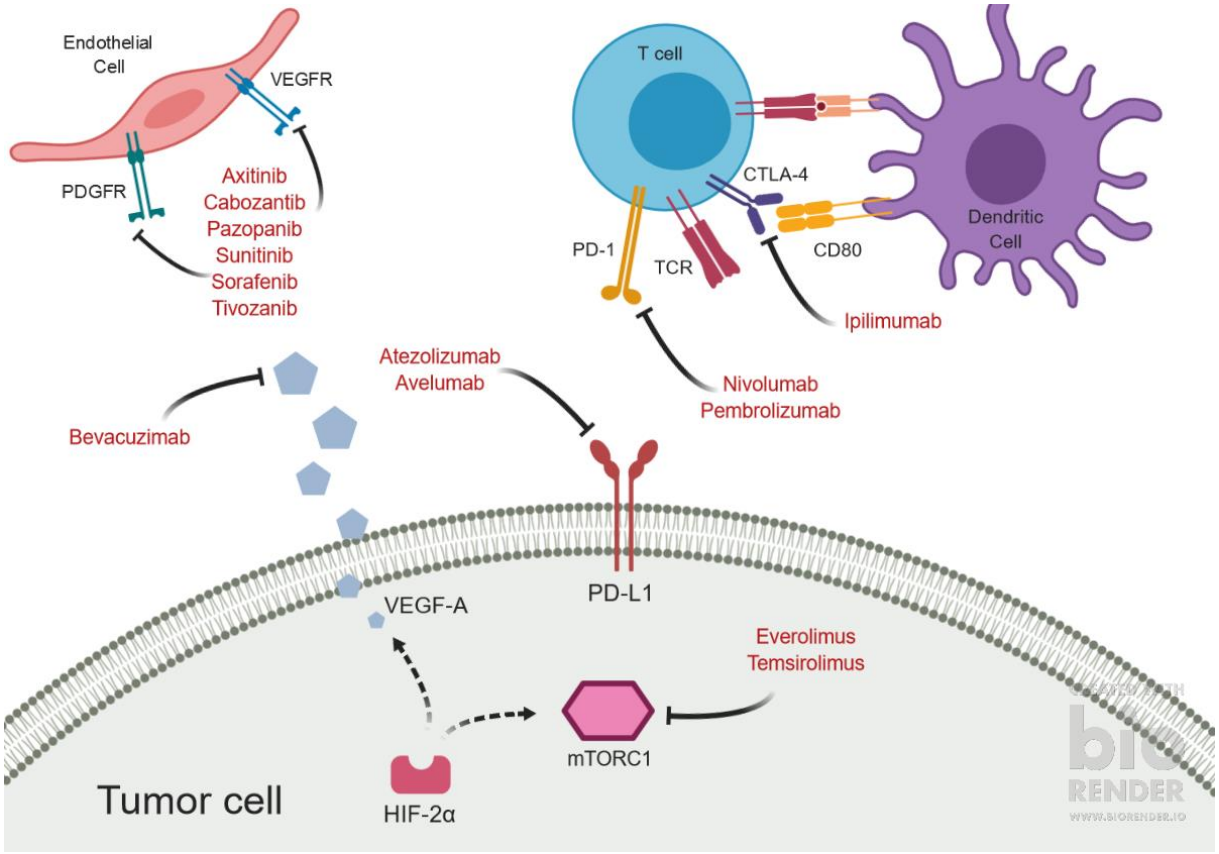


Figure 11 – Mechanism of therapies of advanced RCC (approved or successful clinical phase III studies)

1.5 Model Systems of ccRCC

In order to study various aspects of ccRCC in the laboratory, different model systems have been established including conventional cell lines, primary patient derived cell lines, genetic engineered mouse models and patient derived xenografts.

1.5.1 Conventional cell lines

Supplementation of cell culture medium with fetal calf serum (FCS) allowed the generation of hundreds of cancer cell lines. With the help of these homogenous cell lines, fundamental insights into cellular and molecular biology were gained as they are fast and easy to manipulate. Conventional cell lines allowed for drug screenings in large scale before testing them *in vivo*, complying with the main principle of animal research: Replacement, Reduction and Refinement (3Rs)[206].

However, the cell culture setting oversimplifies the complex tissue architecture of the human system. In general, cells are cultured in a two-dimensional monolayer of adherent cells without mechanic and biochemical interactions of normal tissue environments, which are essential to maintain cellular differentiation fates[207]. Some cell lines have already been cultured for decades, like the commonly used HeLa cell line, which has been established in 1951 and passaged since then[208]. This long-term passage of cells with FCS leads to an acquisition of mutations with alterations in gene expression and pathways over time[209,210]. Media supplemented with FCS is not a fully defined medium and concentrations of growth factors vary from batch to batch[211]. Therefore, it is not surprising that long-term passage lead to a high clonal and genetic variability of the cell lines between different laboratories and studies[212]. Re-injection of cancer cell lines into immunodeficient mice led to tumors that do not resemble the primary tumor[213-216]. In summary, usage of conventional cell lines can complicate the reproducibility and significance of preclinical studies[217-219].

Brodaczewska *et al.* discuss the usage of cell lines as model systems for ccRCC[220]. Many of the established cell lines are described to be tumorigenic in immunodeficient mice, some even metastatic, whereas others fail to implant[221]. Commonly used ccRCC cell lines are 786-O (*VHL* mutant), UM-RC-2 (*VHL* mutant) and the metastatic Caki-1 cell line (*VHL* wild type). Often also papillary RCC cell lines, such as ACHN, A-498 or Caki-2, are used as model system to examine the clear cell phenotype. Various more cell line collections are available, such as one from the Memorian Sloan Kettering Cancer Center and the National Cancer Institute in Bethesda. Nevertheless, gene expression patterns and CNAs of most of the conventional cell lines differ often from primary patient derived tumors[222]. Additionally, only

few conventional ccRCC cell lines are able to recapitulate the distinct clear cell phenotype *in vivo*.

1.5.2 Genetically engineered mouse models

Genetically engineered mouse models (GEMMs) are an excellent tool to study tumor-microenvironment interactions in an immune competent setting (**Figure 12**). Tissue specific overexpression of tumor oncogenes, dominant negative tumor-suppressor genes, the knockout of tumor-suppressor genes, the insertion of targeted mutations or complex combinations of all those methods can induce *de novo* tumorigenesis[223]. With the help of inducible promoters, this process can be timely adjusted and controlled[224].

However, the usage of GEMMs is limited as long as the complex process of tumorigenesis, which differs among the various cancer types, is poorly understood. Increased length of telomeres in mice hamper the development of an instable genome, a common feature of human cancers[225]. The mayor drawback of GEMMs is that a simplified tumorigenesis model in the mouse is used to draw conclusions for the human counterpart, bearing the risk of unreliable results.

Thus, GEMMs are required to study the effect of distinct genes on the development of tumors but cannot reproduce the complex genetic human landscape of a cancer.

Knockout of the ccRCC hallmark gene *Vhl* in mice is embryonic lethal[226]. Heterozygous *VHL* patients are predisposed to develop ccRCCs but mice with heterozygous *Vhl*, even with additional treatment of the renal carcinogen streptozotocin, do not develop renal cancers[227]. An organ specific conditional inativation of *Vhl* in renal epithelial cells has no tumorigenic effect either but shows renal cyst formation[228]. The first mouse model that led to carcinomas in the kidney had a constitutively overexpression of a mutant active HIF-1α in the proximal tubules of the kidney[229].

In 2017, the combined homozygous knockout of *Vhl*, *Rb1* and *Trp53* (encoding for p53) in renal epithelial cells led to the formation of advanced ccRCCs[230]. However, both Rb1 and p53 are only altered in 0.8 % and 3 % of human ccRCC, respectively (cBioportal, TCGA, PanCancer Atlas), which limits the use of this model.

One month later, another clear cell renal cell carcinoma model in mice was presented by Gu *et al.*[231]. They followed the pathogenesis of ccRCC and conditionally deleted *Vhl* and either *Bap1* (the driver of the punctuated evolution type) or *Pbrm1* (the driver of the branched evolution type). *Bap1* deletion gave rise to a more aggressive, high grade tumor and deletion of *Pbrm1* gave rise to a slow progressing tumor. An additional heterozygous inactivation of

the mTORC1 inhibitor *Tsc1* in the *Vhl* and *Pbrm1* deficient kidney cells increased the tumor aggressiveness. At the time of this thesis, this mouse model resembles the development of ccRCC best. An interesting finding of this mouse model were pre-neoplastic lesions found in the parietal epithelial cells of the Bowman capsule (the lining of the glomerulus, the functional unit of the kidney), which challenges the current view that proximal tubule cells are the cell of origin of ccRCC.

1.5.3 Xenograft and Syngeneic ccRCC Models

A xenograft is a heterologous transplant of cells, tissue or organs, for example human material into a mouse. Xenotransplantation has to be performed with immunocompromised mouse models, as the foreign material would be rejected by the hosts' immune system (Host-versus-Graft reaction). The ability of the foreign material to engraft into the new host depends on the functionality of the host immune system. Athymic nude mice, lacking the thymus and thus functional T cells, were the first mice to receive xenotransplants[232]. Engraftment efficiency was low and therefore a new mouse model was established: The non-obese diabetic mouse strain with severe combined immunodeficiency (NOD/SCID)[233,234]. These mice are unable to rearrange T cell receptor and immunoglobulins, which leads to a combined B and T cell deficiency in those mice. A derivative of this mouse strain is the NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl} (NSG) mouse strain[235]. The nonfunctional interleukin-2 receptor γ-chain impairs B and T cell, natural killer cell and lymph node development, but improved engraftment efficiency[234].

Xenograft mouse models allow the study of cell lines and primary patient tumor material *in vivo*. Samples can be transplanted ectopically, for example subcutaneously or orthotopic at the native site of the originating tumor (**Figure 12**). The microenvironment of the orthotopic site in the mouse resembles tumor-originating tissue in matrix composition, nutrient availability and growth factors. Such a surrounding is impossible to emulate in a cell culture setting[236]. Regarding RCCs, the tumor material can be injected orthotopically into the renal subcapsule[237,238]. This method is technically very challenging and may lead to increased animal morbidity. Also tumor take and growth rates may vary in between experiments.

A unique feature of mouse models in cancer research is their ability to study metastasis. While subcutaneous transplantation of tumor material shows very low metastatic capacity[238,239], orthotopic injection has been reported to be the best method to study metastasis. With the help of this method, researchers are able to study metastases development in lung, bone and peritoneal organs (**Figure 12**), which resembles the human route of metastases in renal cancer (**see chapter 1.2.5**).

In addition to conventional cell lines used for orthotopic transplantation[237], various xenograft models of ccRCC were generated from primary patient material with varying engraftment potentials[240-244]. These mouse xenograft *in vivo* models are increasingly used in preclinical studies to test the efficacy of new drugs[245].

A disadvantage of xenograft models is the mouse strain itself as these mice lack a functional immune system, which plays a crucial role in tumor and metastasis control. New treatment strategies targeting the immune system of a patient, cannot be tested in a xenograft transplantation setting and therefore, syngeneic mouse models are used.

Syngeneic mouse models are mouse tumor allografts meaning that they have the same genetic background as the mouse strain in which the tumor is transplanted (**Figure 12**). For renal cell carcinomas, the syngeneic mouse model Renca is available[246,247]. The renal tumor arose spontaneously in a BALB/c mouse and a cell line was established from this tumor. When orthotopically re-transplanted into the renal capsule of BALB/c mice, the tumor develops and forms metastasis. However, the tumor is not mutated in *Vhl* and its mouse origin makes it complicated to compare to the human system.

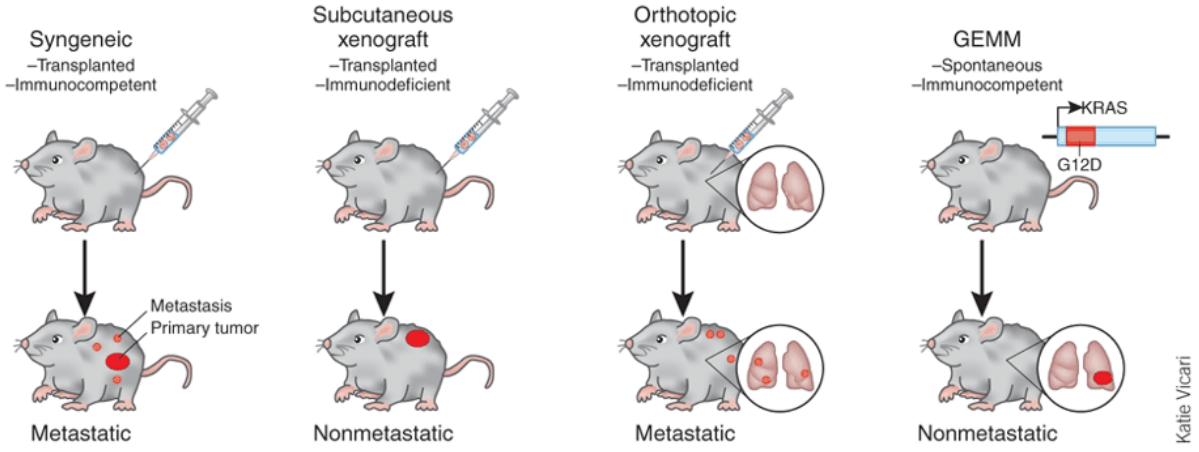


Figure 12 – *In vivo* tumor models of ccRCC

Several different approaches have been developed to study tumor formation *in vivo*. In the syngeneic mouse model a mouse tumor is transplanted as allograft. In this mode, mice have a full functional immune system and orthotopic transplantation often leads to metastases. However, the number of tumor models is limited and the comparison to the human system is hampered.
In subcutaneous xenograft experiments, primary patient tumor material is injected into immunodeficient mice subcutaneously. This is an easy handling to study primary tumor formation. To study the physiological tumor formation, tumor cells are injected orthotopically.
Genetic engineered mouse models (GEMM) with mutations in oncogenes and/or tumor suppressor genes, often organ specific, have been generated. Thereby, spontaneous tumors arise and the influence of single genes on the tumor formation process can be studied. Often, GEMMs oversimplify the tumor development and no metastases develop.
In the figure by Katie Vicari the different tumor mouse models at the example of lung cancer are shown.
With kind permission of Springer Nature, License Number: 4511280297518

1.5.4 Advanced primary serum-free cultured cell lines and xenografts

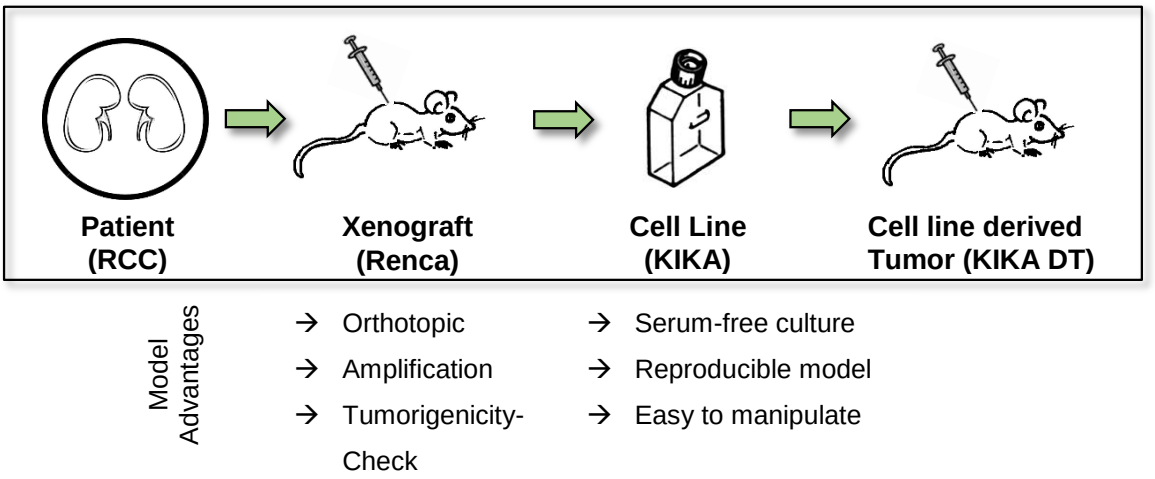
In our laboratory, novel patient-derived *in vivo* and *in vitro* ccRCC models were developed[1,248] (**Figure 13a**). In short, primary tumor material was mechanically and enzymatically dissociated, and orthotopically transplanted into immunodeficient NSG mice. Engraftment efficiency of primary patient material correlated with tumor size and tumor stage. Engraftment time and metastatic capacity showed great intertumor heterogeneity.

Subsequently, tumor-initiating serum free cell culture models, termed KIKA, were isolated. With our chemically defined cell culture medium (adapted from Vermeulen *et al.*[249]), disadvantages of the conventional cell lines were avoided[216] (**see chapter 1.5.1**). To preserve tumorigenicity of the cell lines, various cell culture conditions were tested. Except for one line, 3D spheroid conditions showed best tumor initiating capacity.

The xenografts and KIKA cell line derived tumors were able to recapitulate both the intra- and inter heterogeneity of the original ccRCC (**Figure 13b**)[1]. The histological hallmark phenotype of ccRCCs, the clear cells, were retained and additional renal cell carcinoma and epithelial markers such as AE1/3, CD10, Vimentin, CK18 and Kl1 stained positively[1]. Human origin of the material was verified by positive staining of the proliferation marker hKi67 and negative staining for the mouse marker H2kD.

With our approach to establish novel ccRCC models, we are able to amplify limited starting material of primary patient material, check for tumorigenicity and use it as a tool to study the tumor both *in vitro* and *in vivo*.

a



b

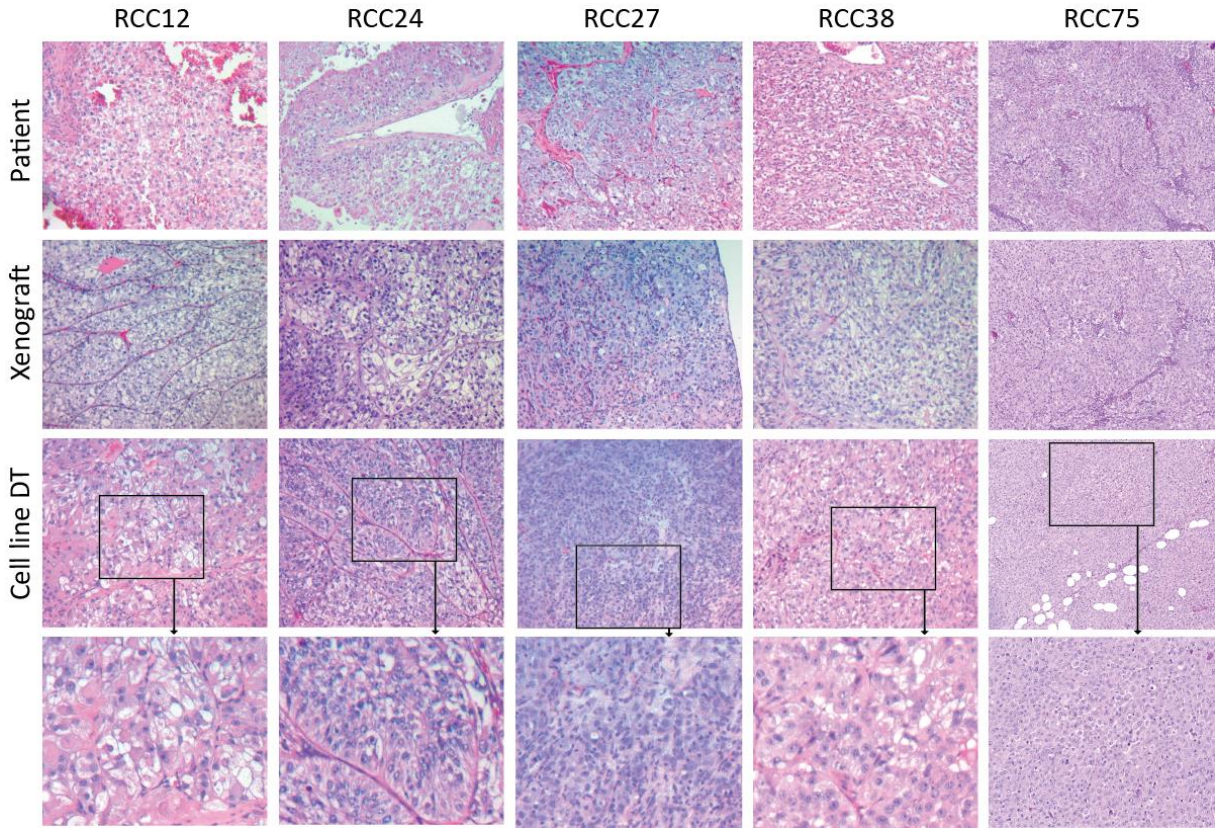


Figure 13 – Novel patient derived ccRCC xenograft and serum free cell culture model

(a) Schema describes the establishment of our ccRCC models. Adapted from Watermeier[1]
(b) HE-staining of the primary patient tumor material, the respective xenograft and cell line derived tumor models used in this study. DT = derived tumor

1.6 Survival analysis

Survival analysis is a method to analyze time dependent outcomes and is mostly used to estimate the survival time of a cohort[250]. However, the outcomes can be any event, from the death of a person, marriage to the birth of a child. The time from a well-defined starting point until an event occurs can be measured in any appropriate timescale, for example days, years, generations or atom oscillations.

In medical research, studies that analyze survival data are in general prospective cohort studies following patients over a period of time. In these studies, it is common that the outcome of an observation is not known for every event. These events are so called censored events. There are several types of censoring, but in survival analysis, we have usually right censored events. Meaning that either an event occurs or the event has not occurred and is censored at the end of the observation period (**Figure 14a**).

With the help of the survival function $S(t)$, it is possible to calculate the probability to survive longer than time *t*, as long as the time is $t \geq 0$. Let *T* be the non-negative variable that stands for the time until an event occurs:

$$S(t) = P(\{T > t\}) \tag{1}$$

With the help of the hazard function $h(t)$, one can calculate the hazard rate: The risk that an event occurs in the next period of time, in case the event has not occurred until then. In the context of survival, the function describes the risk of dying at a specific time *t*, in case, the subject survived that long:

$$h(t) = \lim_{\Delta t \to 0} \left(\frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \right) \tag{2}$$

The complexity of this analysis is that common regression methods cannot handle censored data. Therefore, Edward L. Kaplan and Paul Meier formulated their product-limit estimator in 1958[251].

Let *i* be the individuals of the cohort, and events (not censored observations) be ordered along the time $t_{(1)} < t_{(2)} < \cdots < t_{(i)}$. Let d_i be the events that occur at time $t_{(i)}$ and n_i be the total number of observations in risk right before $t_{(i)}$:

$$\hat{S}(t) = \prod_{i:t_{(i)}<t} \left(1 - \frac{d_i}{n_i}\right) \tag{3}$$

With this formula, survival step functions can be drawn with the survival probability as dependency of the time (**Figure 14b**). To survive a specific timespan, $t_0 \rightarrow t_{(i)}$ is the product of all survival probability intervals up until $t_{(i)}$.

In order to compare two survival curves, several tests have been developed. The most commonly used test is the nonparametric log rank test (or Mantel-Cox test)[252]. It tells whether two survival curves significantly differ between each other, which indicates that a particular grouping variable might be prognostically relevant[253].

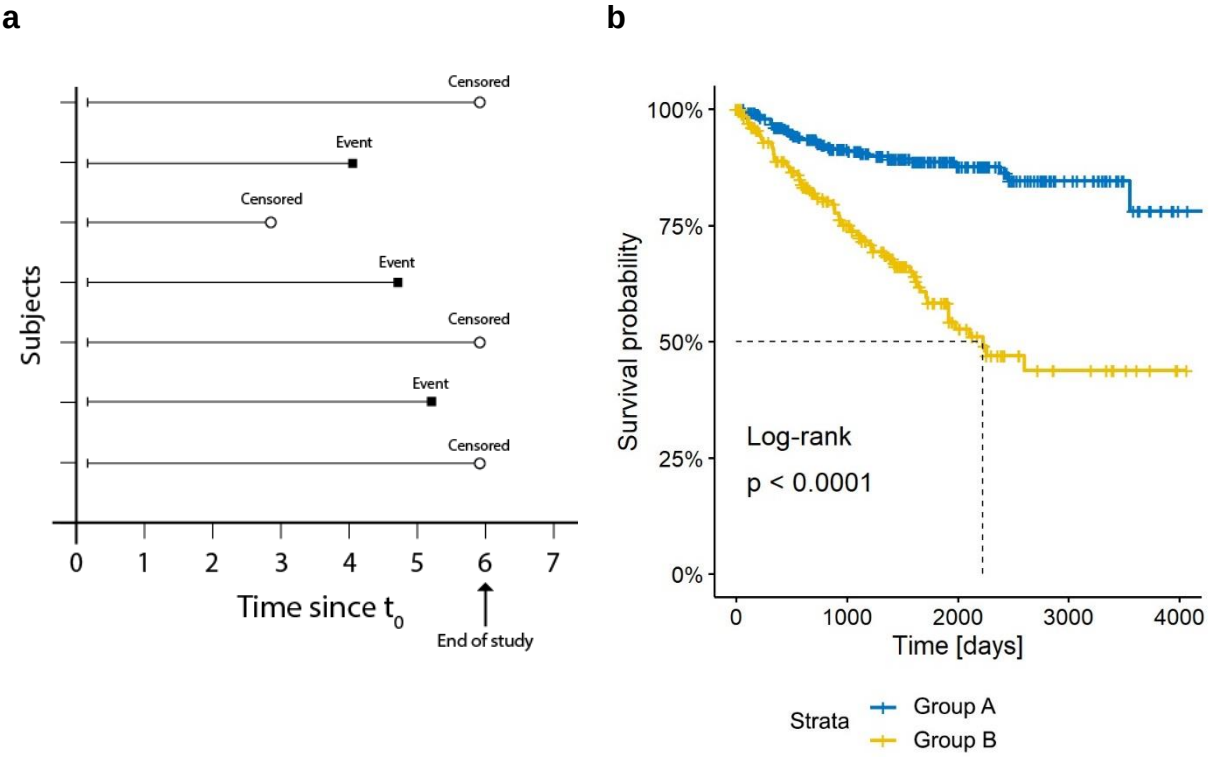


Figure 14 – Survival Analysis

(a) Exemplary Time-to-Event illustration. The study starts at a defined time point and monitors the occurrence of a specific event. If the outcome of the event is unknown, either because the study ends or a subjects exits the study early, the event is censored.
(b) Exemplary Kaplan-Meier estimator. Group A has a significant better survival than Group B. At the end of the study, the probability that an event occurs for Group A is approximately 80 %, whereas for Group B, it is around 40 %. Dotted line indicates median survival of Group B meaning that after around 2200 days, the probability that an event occurs until the next time interval is 50 %. Censored events are represented by a vertical line.

1.6.1 The Cox proportional hazards model

A Cox proportional hazard model (Cox regression, Cox PH model) can be calculated to obtain an estimation of factors influencing survival[254]. Factors can be, for example the administration of a drug, the gender or the age of a participant. They are termed explanatory variables, risk factors or covariates and have a specific weight on the hazard rate that is assumed to be independent from each other[255]. The proportional hazard assumption of this model states that

each unit difference in any of the covariates scales the hazard proportionally at any given time.

Let $X_i = \{x_{i1}, x_{i2}, \dots, x_{ip}\}$ be a vector of the realized values of the covariates p for the subjects i and let β be the corresponding coefficients of the covariates p. $h_0(t)$ represents the baseline hazard when all covariates are $X_i = 0$. According to the Cox proportional hazard model, the hazard rate can be calculated as followed:

$$h(t|X_i) = h_0(t)exp\big(\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}\big) \tag{4}$$

$$h(t|X_i) = h_0(t)exp(X_i \cdot \beta)$$

The baseline hazard is often complicated to evaluate as covariates, such as blood pressure, cannot be set to 0 in physiological settings. Therefore, the Cox proportional hazard model uses a semi-parametric approach to avoid estimation of the baseline hazard. By declaring the hazards for two subjects proportionally, a ratio of hazards can be formulated:

$$HR = \frac{h(t|X_1)}{h(t|X_2)} = \frac{h_0(t)exp(X_1 \cdot \beta)}{h_0(t)exp(X_2 \cdot \beta)} = exp\big(\beta \cdot (X_1 - X_2)\big) \tag{5}$$

From this semi-parametric approach, we can estimate the β of a specific covariate k, by holding all other covariates constant and increasing k by only one unit:

$$HR = \frac{h\big(t\big|(x_1, \cdots, x_k + 1, \cdots, x_p)_1\big)}{h\big(t\big|(x_1, \cdots, x_k, \cdots, x_p)_2\big)} = exp(\beta_k) \tag{6}$$

With all other covariates remain unchanged and only the covariate x_k is increased by one unit, β_k corresponds to the log hazard ratio. In practical use in a medical setting and patient survival as outcome, this formulation describes the changed risk of a patient to die when a covariate increases by one unit. For assuming that the covariate x is smoking, then $x_{Smoking} = 1$ represents that a patient smokes and $x_{Smoking} = 0$ that a patient is a non-smoker. If the patient now is a smoker, the risk to die changes by $exp\big(\beta_{Smoking}\big)$ in comparison to the non-smoker. If the hazard ratio is above one, the risk to die for a patient with the covariate x is increased, if it is below one, the risk is reduced. Hence, the hazard ratio always compares two groups of subjects that differ in covariates.

There are various methods to estimate and validate multiple β coefficients of a Cox PH model[256-258]. In short, the methods try to fit the observed data to an ideal set of coefficients by maximizing the partial likelihood. To assess the validity of the regression and its parameters, tests like the Wald's test, the Likelihood ratio test and the Score (log rank) test are commonly

used[259]. In order to estimate, how good the calculated model predicts the data, the concordance index (C-index, "Harrel's C") can be calculated[260,261].

1.6.2 Parameter Shrinkage – The Lasso Regression

By profiling all expressed genes of a patient cohort, the resulting data is high-dimensional (number of patients multiplied by the number of genes). In most settings, the number of patients *n* is much smaller than the number of genes *p* that are profiled ($n \ll p$). Estimating a Cox regression model that includes all possible gene combinations would not only over parameterize the model and difficult to interpret, but also be a computational challenge[262]. With every additional covariate included, the possible number of parameter combinations in the regression model growths exponentially. This effect has been termed "the curse of dimensionality". Even if all covariates were estimated, only a small number of gene combinations would most likely have a major influence on the outcome.

A reduced number of predictors for assessing the model can lower the prediction error. Next to preselection of a set of genes by an educated guess (for example additional experiments that indicate the importance of a specific set of genes), stepwise regression and penalized regression methods have been developed to shrink the number of predictors[262]. All these methods profit from the bias-variance tradeoff:

Models with a high variance tend to over-fit the noisy training data. Models with a high bias oversimplify the training data and thus the model (**Figure 15**).

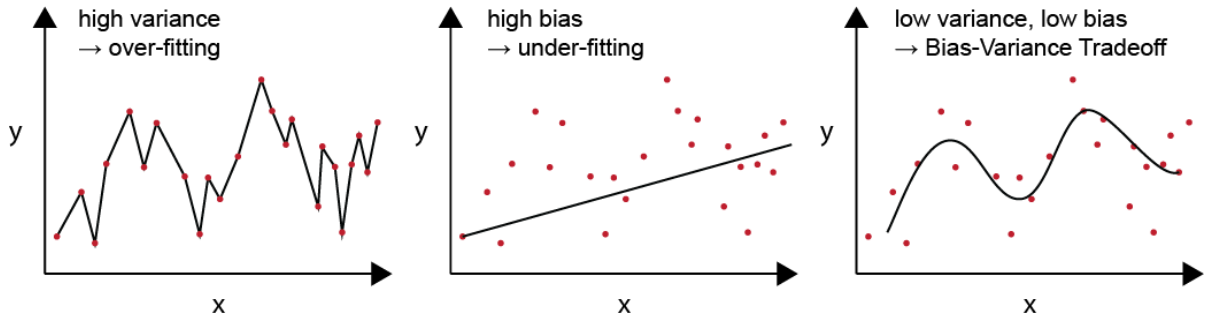


Figure 15 – Bias-Variance Tradeoff

Many algorithms profit from the bias-variance tradeoff. Models with high variance have a high sensitivity to the noisy training data, but they tend to over-fit the data. Models with high bias are under-fitting, as they oversimplify the underlying data and cannot predict the true data structure. Models with low variance are less complex and models with low bias are more complex. The bias-variance tradeoff finds an optimal balance between over- and under-fitting and minimizes the error of the model[263].

The penalized regression models introduce a factor that penalizes less realistic coefficients to get a simpler model. A so called regularization parameter λ helps to control the bias-variance tradeoff. When the parameter λ has been chosen too small, the training data will be over-fit and a λ that is too large is often oversimplifying the model by introducing a high bias.

The least absolute shrinkage and selection operator (LASSO) regression method is one penalized regression method that shrinks the values of the coefficients and at the same time selects the numbers of predictors[264]. In the context of Cox's proportional hazards models, the method has been adapted to fit censored survival data (Regularized Cox Regression)[265]. Thereby, many of the covariates are due to the nature of the used penalty term *L-1* norm minimized exactly to zero. The LASSO regression therefore allows a shrinkage of covariates in an unbiased manner[266]. The regularization parameter λ controls for the number of non-zero covariates in the model and can be selected by cross-validation.

2 Aim of the Dissertation

Clear cell renal cell carcinoma (ccRCC) is the most common subtype of renal cell carcinoma and the one with the worst prognosis[20]. With complex clinical presentation, diagnosis often occurs by accident and at late stages, when the tumor has already developed metastases[14]. Even with progress in novel checkpoint inhibitor therapies, prognosis of patients with advanced ccRCCs is poor with a median overall survival of 15 months[191].

After surgical resection of both metastatic and non-metastatic patients, it is essential to assess the risk of relapse in order to ensure adequate follow-up[16]. In the clinical routine, the patients' risk is most widely assessed by using the AJCC TNM classification or the SSIGN Score[133,142]. Molecular markers and complex signatures failed to improve the prognostic value of these classifications so far[267].

Therefore, the objective of this study is to identify a robust and simple gene signature allowing for a more precise prediction of patients' risk.

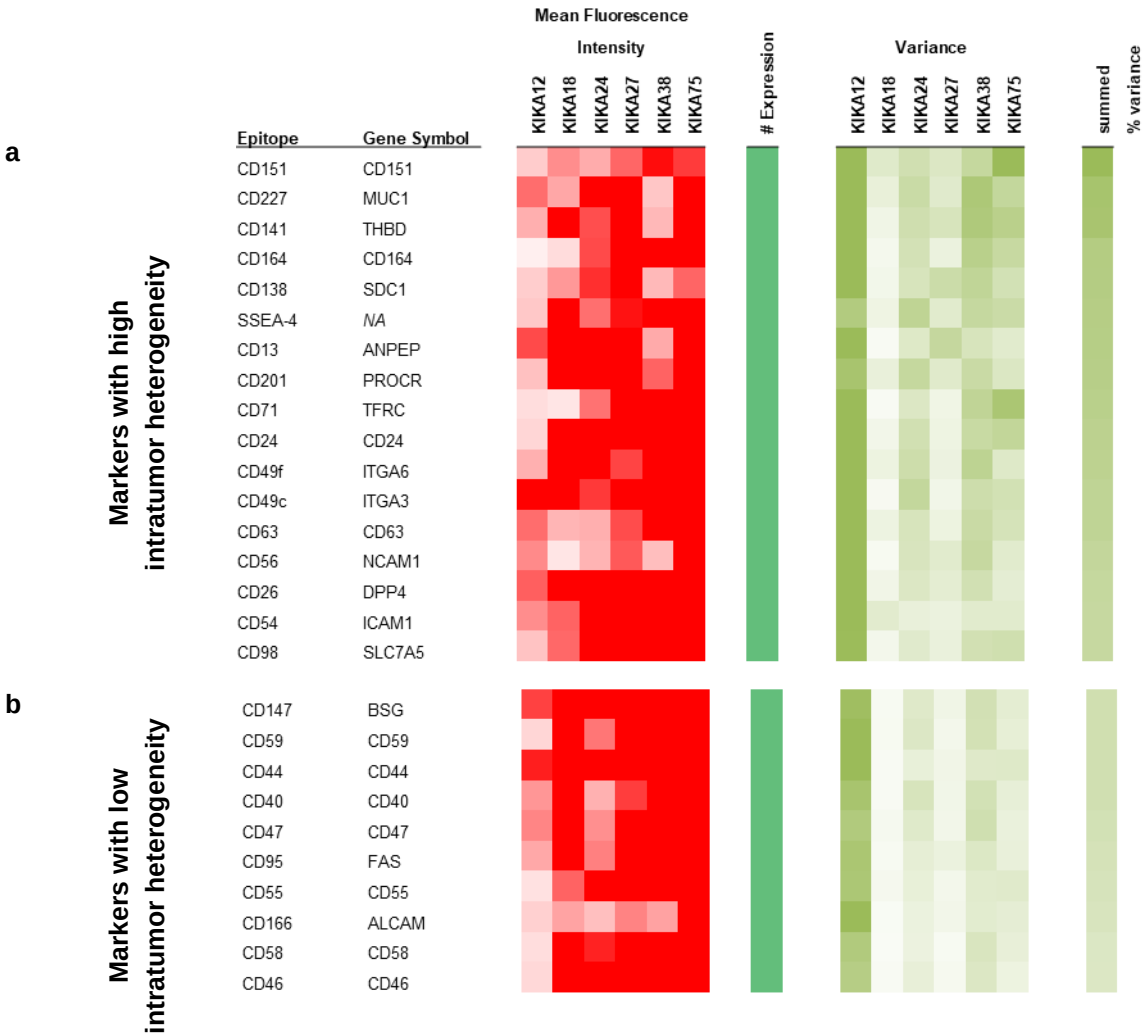
On the molecular level, ccRCCs show significant inter- and intra-tumor heterogeneity[73]. Therefore, the use of a prognostic signature that assesses the risk of ccRCC patients was analyzed with regards to tumor heterogeneity.

The identified signature comprises genes, which might play a role in mediating tumor aggressiveness and metastasis. Making use of *in vitro* and *in vivo* experiments, we aim to elucidate their role in cancer progression to identify possible targets for the treatment of clear cell renal cell carcinoma.

3 Results

3.1 KIKA ccRCC models are immunophenotypically heterogeneous

In a first step, we characterized six of the KIKA cell culture ccRCC models (see **section 1.5.4**) by SNP array to evaluate CNAs (**Supplemental Figure 1**) and by screening 242 cell surface markers by flow cytometer analysis. Selected results are shown in **Figure 16**. We identified surface markers, which are expressed in all cell lines but show high intratumor heterogeneity (**Figure 16a**) and markers expressed in all cell lines that show low intratumor heterogeneity (**Figure 16b**). Moreover, we found markers that are heterogeneously expressed between the cell lines (**Figure 16c**) and markers that are not expressed on the KIKA cell culture ccRCC models (**Figure 16d**). The complete list of cell surface markers is found in **Supplemental Figure 2**.



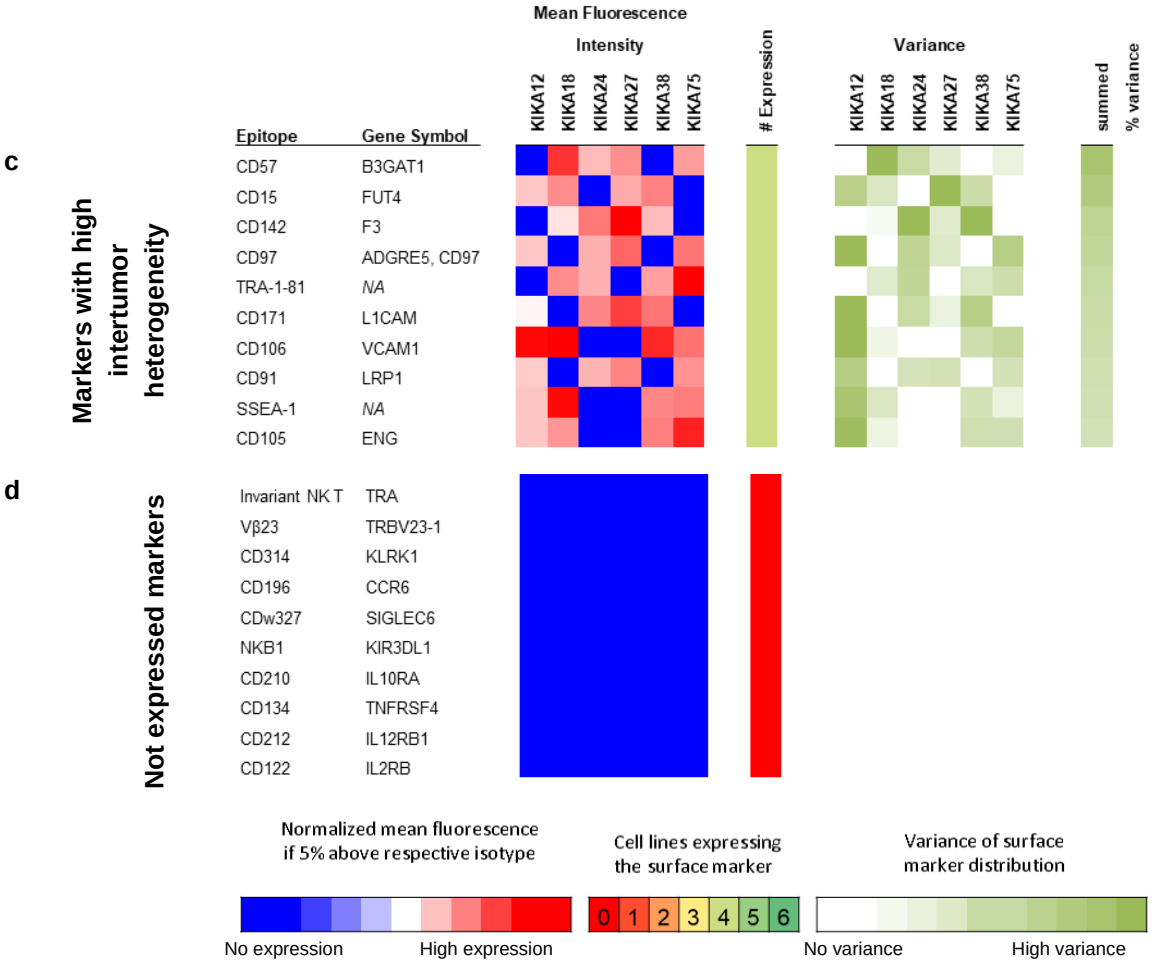
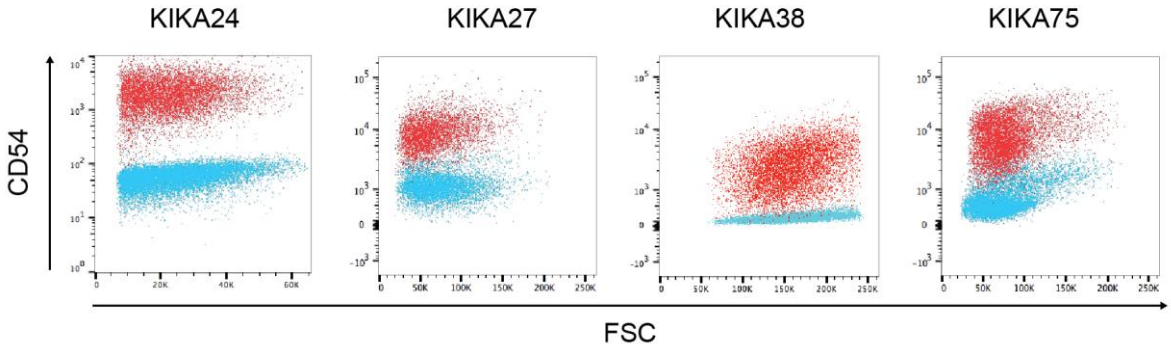


Figure 16 – Excerpt of 242 analyzed cell surface markers of six different KIKA ccRCC cell culture models

Cell surface markers measured by flow cytometry were sorted according to the number of cell lines expressing a surface marker (5 % above respective isotype) and according to the standardized robust coefficient of variation (CV). Shown are the top surface markers with **(a)** high intratumor heterogeneity, **(b)** low intratumor heterogeneity or **(c)** intertumor heterogeneity. **(d)** shows exemplary markers that are not expressed on the KIKA cell lines (no mean fluorescent signal above isotype). **Supplemental Figure 2** lists all cell surface markers of the immunophenotypisation.

Exemplary, we validated one of the surface markers, ICAM1/CD54, in the KIKA cell lines by flow cytometry and *in vivo* by immunohistochemistry on four KIKA cell line derived tumors (**Figure 17**). We find ICAM1 expressed in all tested cell lines in cell culture, with high expression in KIKA24 cells and a heterogeneous expression in KIKA27, KIKA38 and KIKA75 cells. Xenotransplanted KIKA cells form tumors that show intratumor heterogeneous staining for ICAM1. In KIKA75, and KIKA38 derived tumors, a majority of tumor cells membranes stain positively for ICAM1. In KIKA27 and especially KIKA24 derived tumors, only a fraction of tumor cells stain positively for ICAM1, suggesting a high intratumor heterogeneity of ICAM1 expression. **Supplemental Figure 3** shows the gating scheme for CD54 FACS staining.

a



b

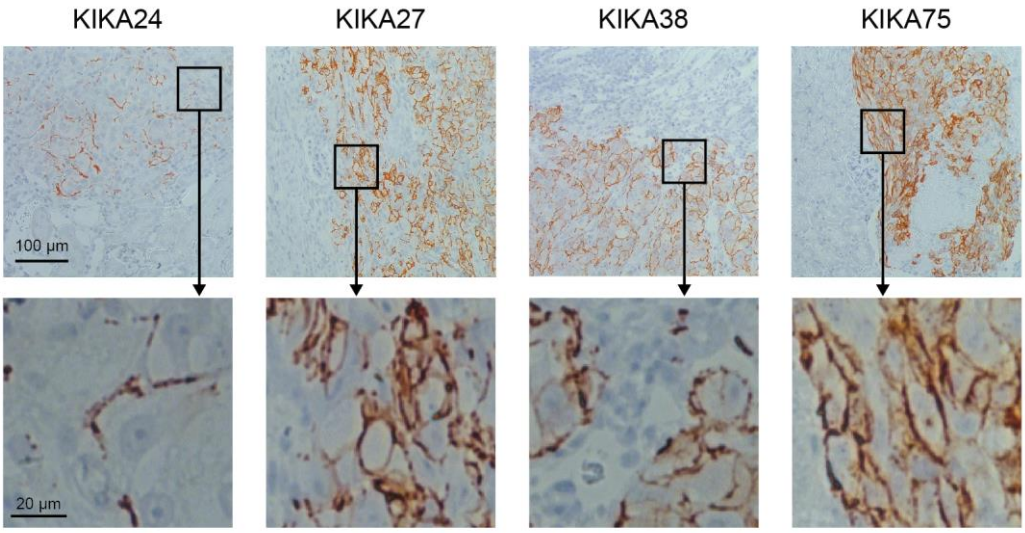


Figure 17 – ICAM1/CD54 is expressed *in vitro* and *in vivo* KIKA ccRCC models

(a) CD54 expression in flow cytometry. CD54 is expressed in all cell lines with varying intensity. Light blue indicates isotype staining.
(b) Expression of CD54 in xenograft tumors. Staining for CD54 reveals heterogeneous distribution with some cells of high CD54 expression on the surface and others with no CD54 expression.

3.2 KIKA cell lines are resistant to mTOR and tyrosine kinase inhibitors

KIKA cell lines were tested for their drug resistance against the two most commonly used small molecule inhibitors (see **section 1.4.2**) targeting the mTOR (everolimus) and receptor tyrosine kinases (sunitinib). Additionally, we tested the EGF receptor inhibitor erlotinib. The titration with sunitinib, everolimus and erlotinib revealed high resistance of all tested cell lines against all tested drugs (Figure 18).

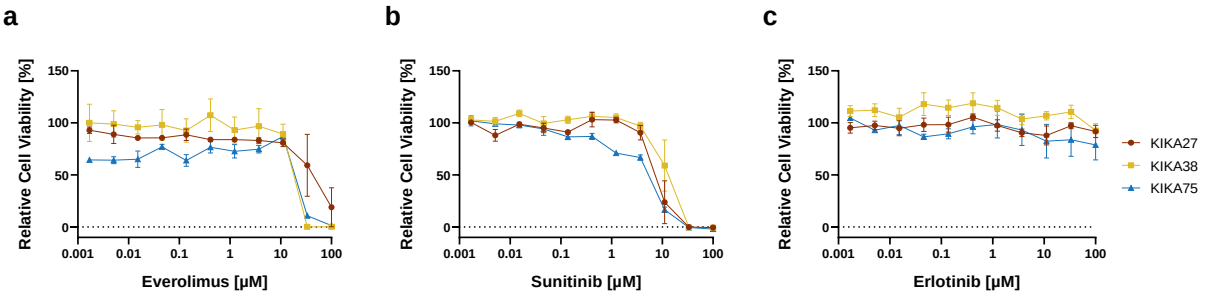


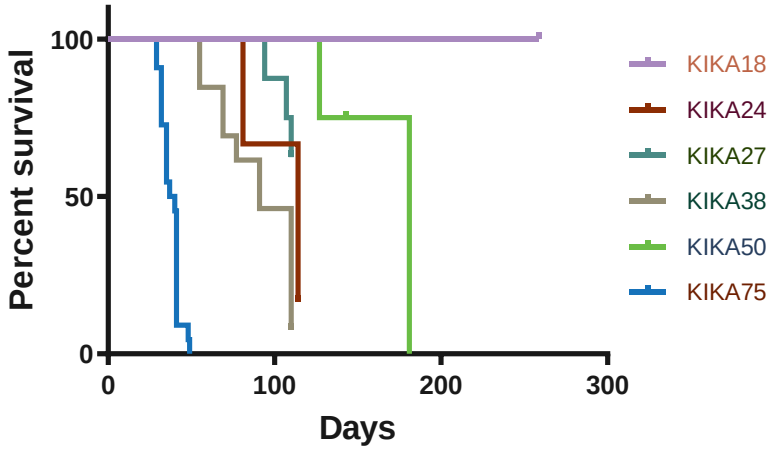
Figure 18 – KIKA cell lines are resistant to mTOR and tyrosine kinase inhibition

KIKA cell lines were treated 48 h (KIKA27) or 72h (KIKA38 & KIKA75) with increasing concentrations of **(a)** the mTOR inhibitor Everolimus, **(b)** the receptor tyrosine kinase inhibitor Sunitinib or the EGF receptor inhibitor Erlotinib. Cell growth was measured by CellTiter Blue metabolism and measured by fluorescence in a plate reader. Error bars depict mean ± s.e.m.

3.3 Xenograft ccRCC models are heterogeneous in survival and metastasis formation

By injecting six different primary patient xenograft models orthotopically into NSG mice, we observed different survival proportions (**Figure 19a**) and capacities to form lung metastases (**Figure 19b**), indicating an intertumor heterogeneity. Especially the KIKA75 model, with a median survival of 38.5 days, develops highly likely lung metastasis. Therefore, we decided to use this model to identify genes that have an impact on the metastatic progression. We further examined genomic architecture of this cell line by chromosome painting and whole exome sequencing, which revealed a highly instable genome with many rearrangements (**Supplemental Figure 4**).

a



b

Xenografts models	% formation of lung metastasis	Patient TNM-Status	Histological grading
KIKA12	-	T3a N1 M0	G3
KIKA18	0	T3a Nx M1	G2
KIKA24	33.3	T3b N0 M1	G3
KIKA27	25	T3a Nx M1	G3
KIKA38	30.8	T4 N0 M0	G3
KIKA50	0	T3a N0 M1	G3
KIKA75	85.7	T3a Nx Mx	G4

Figure 19 – Intertumor-heterogeneity in tumor development, aggressiveness and potential to form lung metastases in primary patient derived xenograft models

(a) Survival estimates of the different primary patient derived ccRCC xenografts. 100000 KIKA cells of the different tumor models were injected into the renal capsule of NSG mice. When the tumor size reached its stopping criteria, mice were euthanized. KIKA75 has the fastest tumor progression (median survival = 38.5 days).

(b) Overview of ccRCC models, with clinical parameters of the primary patient. Formation of lung metastases was assessed by eye. KIKA75 has the highest capacity to form metastases in NOD scid-gamma (NSG) mice (85.7 %).

3.4 Generation of an *in vivo* selection model to identify mediators of tumor aggressiveness

In five rounds and in three biological replicates of an in vivo selection, we transplanted the metastases of orthotopically transplanted KIKA75 cells into the renal capsules of NSG mice **(Figure 20)**. Primarily, we wanted to select for cells with a higher capacity to metastasize, but we are well aware that at the same time cells enriched, selected or adapted for a more aggressive phenotype (**Figure 21**). Survival of mice in the last passage was reduced by 22 %, primary tumor size was 50 % increased and more and bigger lung metastases were observable.

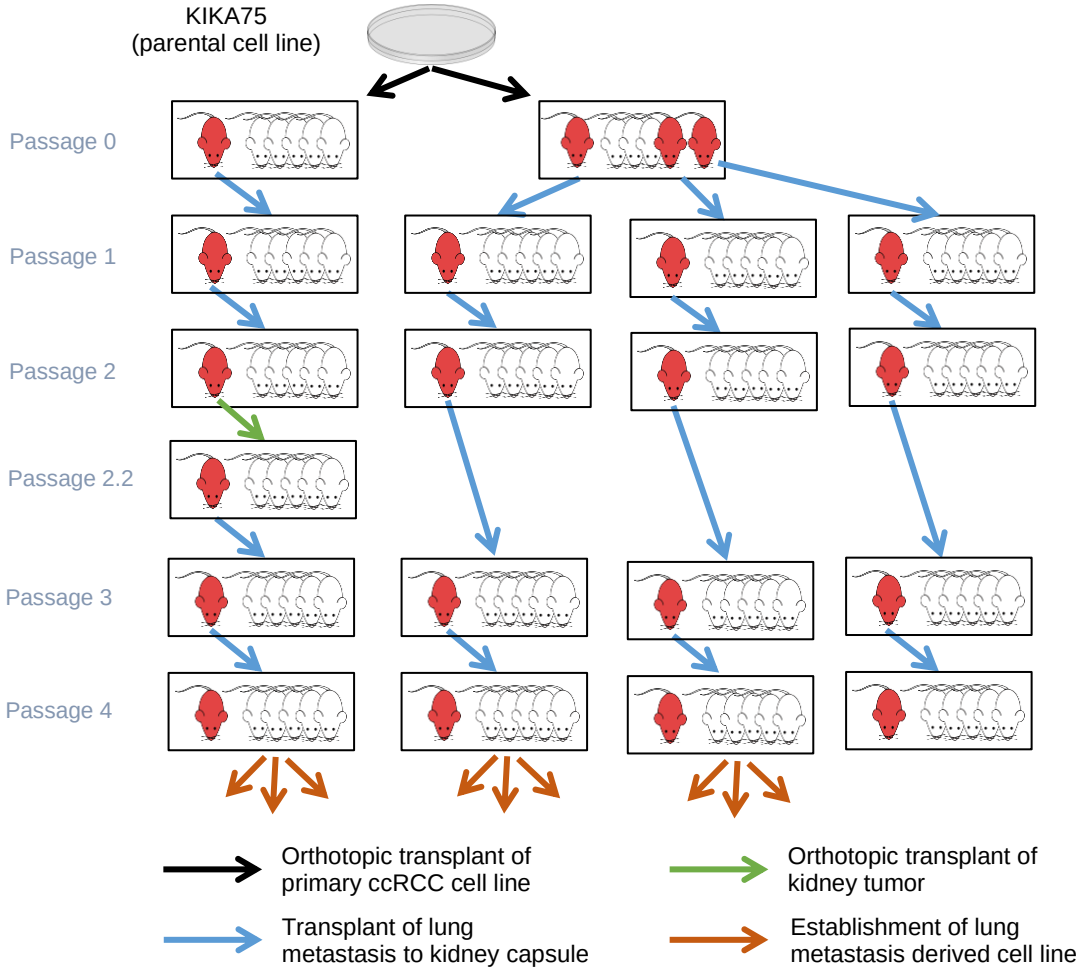


Figure 20 – Schematic overview of the *in vivo* selection model.

The primary ccRCC cell line KIKA75 was transplanted orthotopically into the kidney capsule of NSG mice. When the mice had to be euthanized according to the previously defined stopping criteria, the lungs of one of the mice per biological replicate were digested and reinjected into the renal capsule of mice of the next round. In one round of the first biological replicate, no metastases developed and cells of the digested primary tumor were orthotopically transplanted into the next round of mice. After the last round, cell lines were established from the first three biological replicates from the lung metastases of three or four mice per replicate group. At every step of the first three replicates of the selection process tumor material was collected and expression profiled. For the last replicate gene expression data was generated from only the passage 0 and passage 4 samples.

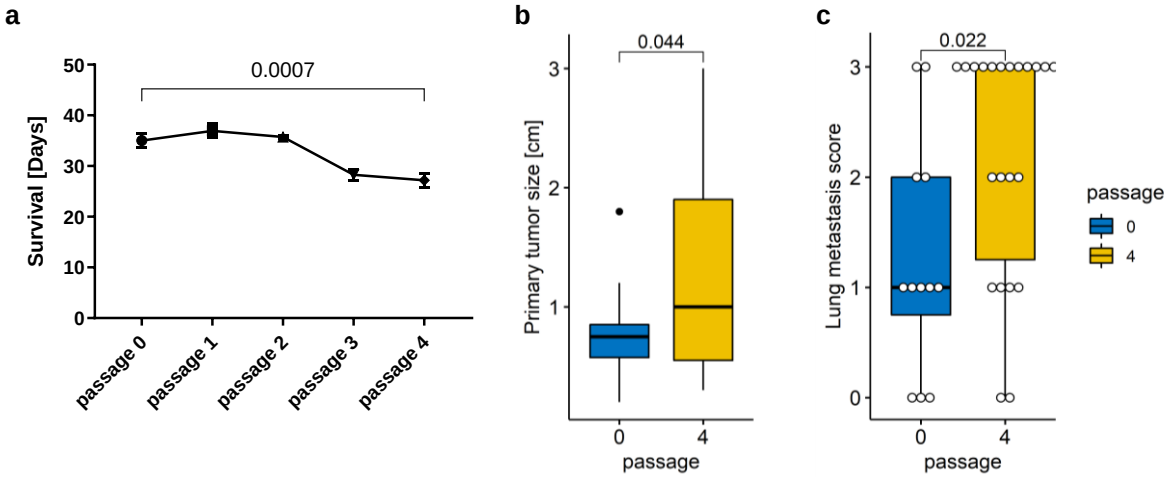


Figure 21 – The *in vivo* selection model selected for more aggressive and metastatic cells

(a) Mice from the first passage had to be euthanized after 35.0 ± 1.4 days and mice from the last passage after 27.2 ± 1.3 days (*P* = 0.0007) when any of the stop criteria defined by the GV-SOLAS was reached. Error bars represent mean ± s.e.m. **(b)** Primary tumor size at time of euthanization of the first passage was 0.8 ± 0.4 cm (n = 12) and primary tumors size from the last passage was 1.2 ± 0.8 cm (n = 22; p = 0.044). **(c)** Scoring the lung metastases by eye and histology revealed an increase from 1.3 ± 0.3 (n = 12) in passage 0 to 2.2 ± 0.2 (n = 22; *P* value = 0.0218) in passage 4. *P* value was calculated using a two-sided Student's *t*-test. Error bars depict mean ± 1.5*IQR.

3.4.1 Passage 4 derived lung metastasis cell lines exhibit an increased carrying capacity

To explain the *in vivo* differences in tumor formation between the originating KIKA75 cell line derived tumors and the tumors of the last passage of the *in vivo* selection, the growth of cells isolated from the lungs of the last passage and from the originating cell line was measured. Subsequently, growth rate and carrying capacity were estimated by logistic regression **(7)**, fitting the results of metabolic turnover as an estimator for cell number (**Figure 22**).

Logistic regression formula to calculate growth rates

$$N_t = \frac{K}{1 + \left(\frac{K - N_0}{N_0}\right) e^{-rt}} \tag{7}$$

N_t: Population size at time t

K: Maximum possible population size, or carrying capacity

r: Growth rate

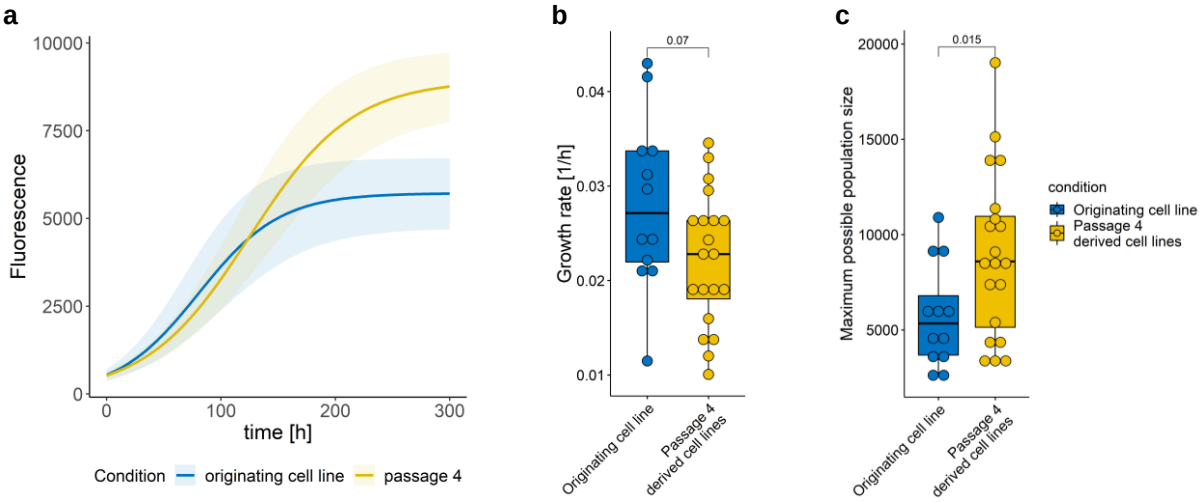


Figure 22 – Cell lines derived from lung metastases of passage 4 had no significant difference in growth rate but maximum possible population size

(a) Growth curves estimated from the originating cell lines and passage 4 lung metastases derived cell lines. Cell growth of three biological replicates of each condition with 6 technical replicates each was measured by CellTiter Blue metabolism. Growth curves were fitted using the *growthcurver* package in R. **(b)** The growth rate of cells before and after the selection revealed no significant difference in tumor cell growth rate (passage 0: r = 0.028 ± 0.009 h^{-1}, n = 12; passage 4: r = 0.022 ± 0.007 h^{-1}, n = 20; *P* = 0.089). **(c)** The maximum possible population size of the passage 4 derived cells is about 1.6 times larger than of the originating cell line (*P* = 0.015, two sided Student's *t*-test). Error bars depict mean ± 1.5*IQR.

3.5 Expression and genomic analyses of the *in vivo* selection

To elucidate the differences in growth potential and metastases formation, we analyzed samples of the *in vivo* selection with the help of gene expression profiles, DNA methylation analyses and exome sequencing.

3.5.1 Exome Sequencing reveals that the *in vivo* selection did not select for subclones

By comparing the sequenced exome of the primary tumors derived from the originating cell line and the primary tumors of the last passage, changes in clonal distribution can be characterized. For this analysis, Gregor Warsow identified all functional somatic single nucleotide variants (SNVs) and copy number variations (CNVs) of the bulk kidney tumor samples. Mouse cell contamination was reduced by alignment to a mouse-human hybrid genome. Between 84.9 % and 93.5 % of the reads were human origin. Using the heuristic Marcov Chain Monte Carlo method[268,269], the mutations were allocated to individual mutation clusters, which are distributed along the branches of a phylogenetic tree. The tree describes the relatedness between the subclones. Each subclone carries a set of mutations which is collected when traveling from the root of the phylogenetic tree to the subclones respective leaf. The aim of the Marcov Chain Monte Carlo method is to achieve a global optimum distributing mutations along the tree. Subsequently, to each tumor sample the proportions of the different subclones were assigned. However, with this method not all subclones

necessarily possess all mutations, which are found along the path from root to leaf. Five different subclones with a characteristic mutational pattern were identified (**Figure 23**). Although clone 3 and 4 show subtle variance, the general picture indicates stable clonal heterogeneity. Therefore, the changes in aggressiveness between the passages may not be explained by clonal outgrowth but by differential expression and epigenetic regulations.

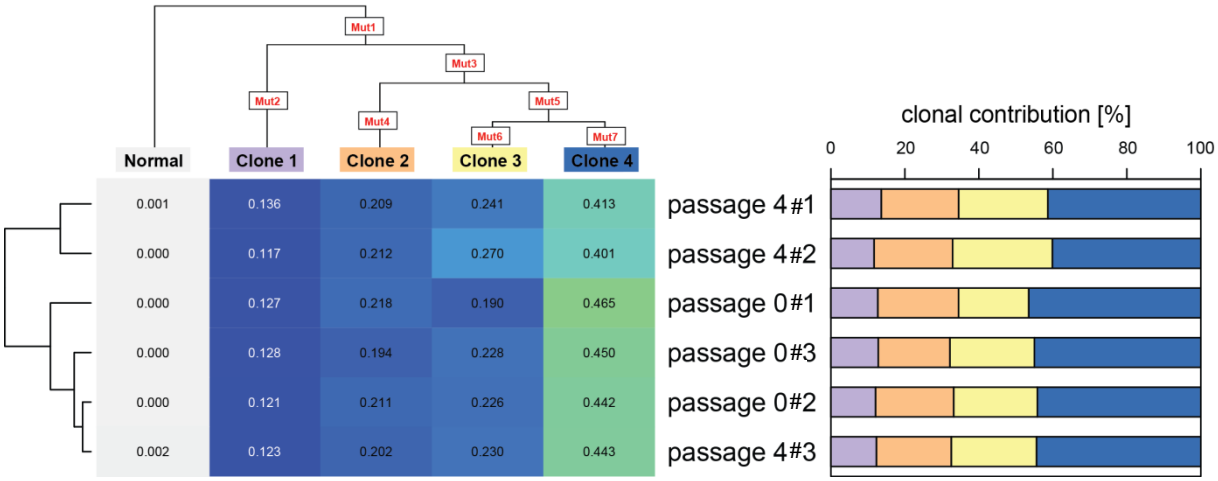


Figure 23 – No predominant subclone emerges during the *in vivo* selection

Using a heuristic Marcov Chain Monte Carlo method, mutations called by exome sequencing were distributed along the branches of a phylogenetic tree. Each subclone is characterized by a set of mutations (Mut). Clone 3 and 4 show subtle variance but the general picture indicates stable clonal heterogeneity. Posterior probability 32.3 %. Data analyzed by Gregor Warsow.

3.5.2 Primary Kidney Samples were used for Gene Expression Analyses

We analyzed the gene expression profile of the kidney tumors, the lungs containing the metastases of each step, as well as the isolated cell lines derived from the lung metastases of the last passage and the originating cell lines by gene expression microarray analysis. Following the *limma* pipeline[270,271], arrays were corrected for background, normalized and log2-transformed. To increase the sensitivity of microarrays and to reduce the multiple testing problem, probes of the microarrays were filtered for duplicates and variance (interquartile range of variance < 0.5). Quality control was performed and thereby one outlier was identified and removed (**Supplemental Figure 5**). Gene annotation was subsequently updated to the recent ensemble version (Ensembl Release 94, October 2018).

We decided not to compare the isolated cell lines to the originating cell lines alone, as potential gene expression effects induced by the microenvironment of the tumors might be affected by the isolation, digestion and culturing procedure conducted to obtain tumor derived cell lines.

Data of the bulk lung with the metastases were analyzed with the caveat of false positive results which might have risen from the sampling process in which small metastases from the lungs could have resulted in a contamination with normal mouse lung tissue. This mouse tissue could be false positively recognized by the human microarray and is less pronounced in the bigger primary tumors, where only little adjacent normal tissue was sampled (**Figure 24**). By comparing the expression profile of bulk kidney and lung tumors of the first biological replicate to gene expression data of isolated cell lines generated from the lung metastases of the last passage, we could show that expression data from the bulk tumor correlates better to the cell lines ($r^2 = 0.78 \pm 0.01$) than the lung metastases to the cell lines ($r^2 = 0.69 \pm 0.06$; $p = 0.05$).

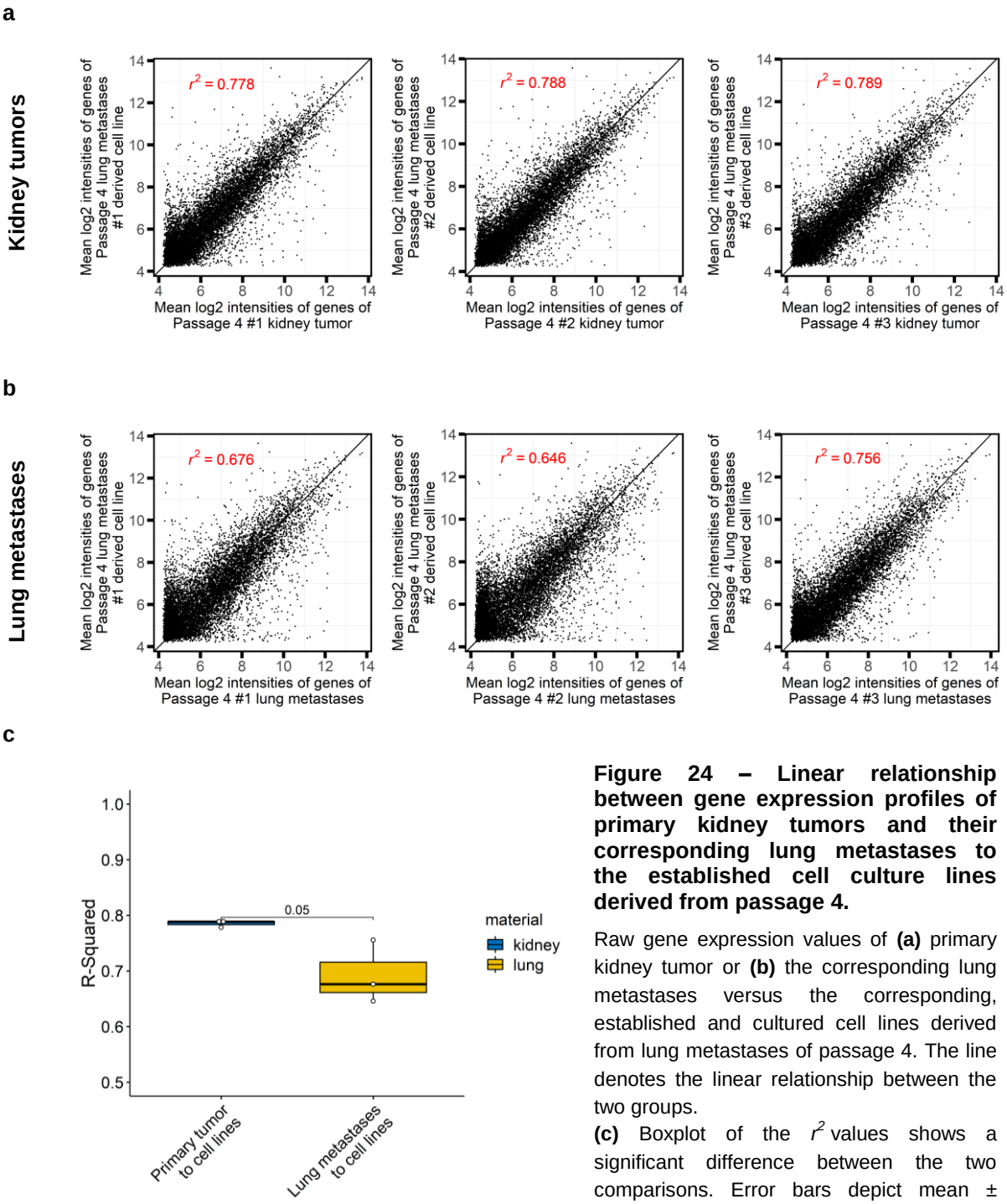


Figure 24 – Linear relationship between gene expression profiles of primary kidney tumors and their corresponding lung metastases to the established cell culture lines derived from passage 4.

Raw gene expression values of **(a)** primary kidney tumor or **(b)** the corresponding lung metastases versus the corresponding, established and cultured cell lines derived from lung metastases of passage 4. The line denotes the linear relationship between the two groups.

(c) Boxplot of the r^2 values shows a significant difference between the two comparisons. Error bars depict mean ± 1.5*IQR.

3.5.3 Principle component analysis of the gene expression profiles clusters samples according to tissue type

The degree of differences within the cell lines derived from the lung metastases, the lung metastases themselves and the primary tumors can also be observed by performing a principle component analysis (**Figure 25a**). The samples cluster according to their tissue of origin in the first two principle components (cell line, kidney tumor, lung metastases) and biological replicates. Nevertheless, only 34.9 % of the observed variance can be explained by the first two principal components. For that reason, the samples had to be subsetted, in order to observe significant results in a differential gene expression analysis. By annotating the bead chips, we were able to exclude a batch arising from the usage of different microarrays (**Figure 25b**).

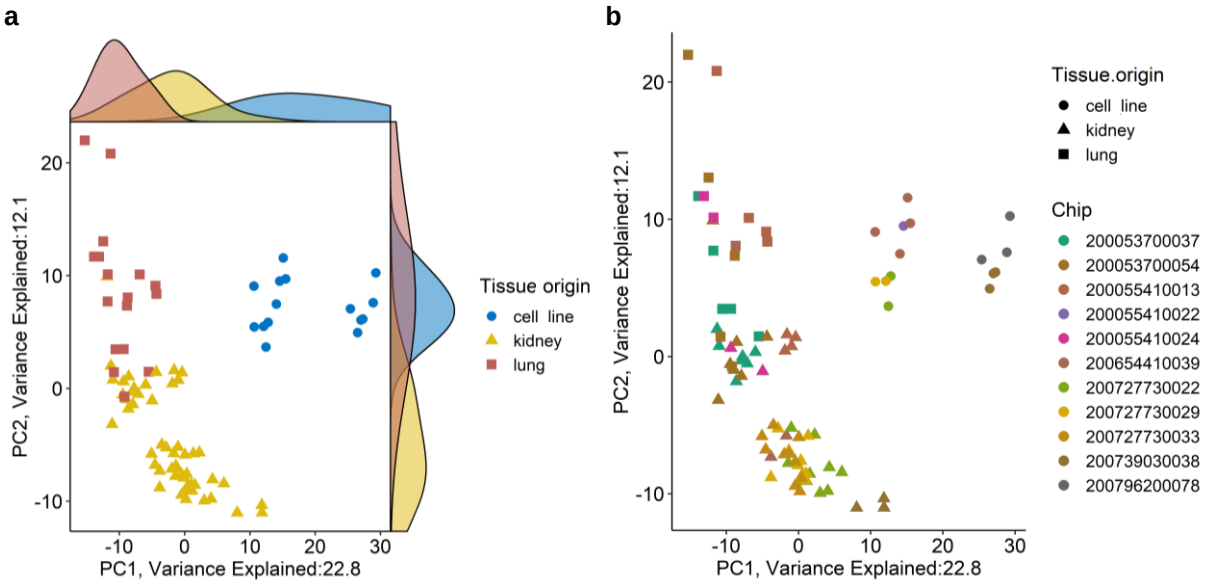


Figure 25 – Principal Component Analysis separates samples into clusters of tissue origin and show no batch effect

The top 500 variant standardized gene expression data of all used Illumina HT12-v4 expression arrays were analyzed in a principal component analysis and samples labeled according to **(a)** tissue origin and **(b)** microarray bead chips. Samples cluster in the first two principal components foremost according to tissue origin and show no extensive batch effect.

3.5.4 Principal component analysis of each biological replicate clusters samples according to the *in vivo* passage

By subsetting the primary tumor derived gene expression profiles biological replicates and *in vivo* passages cluster together in a principal component analysis, indicating a common trajectory of variance explained by the principal components (**Figure 26**).

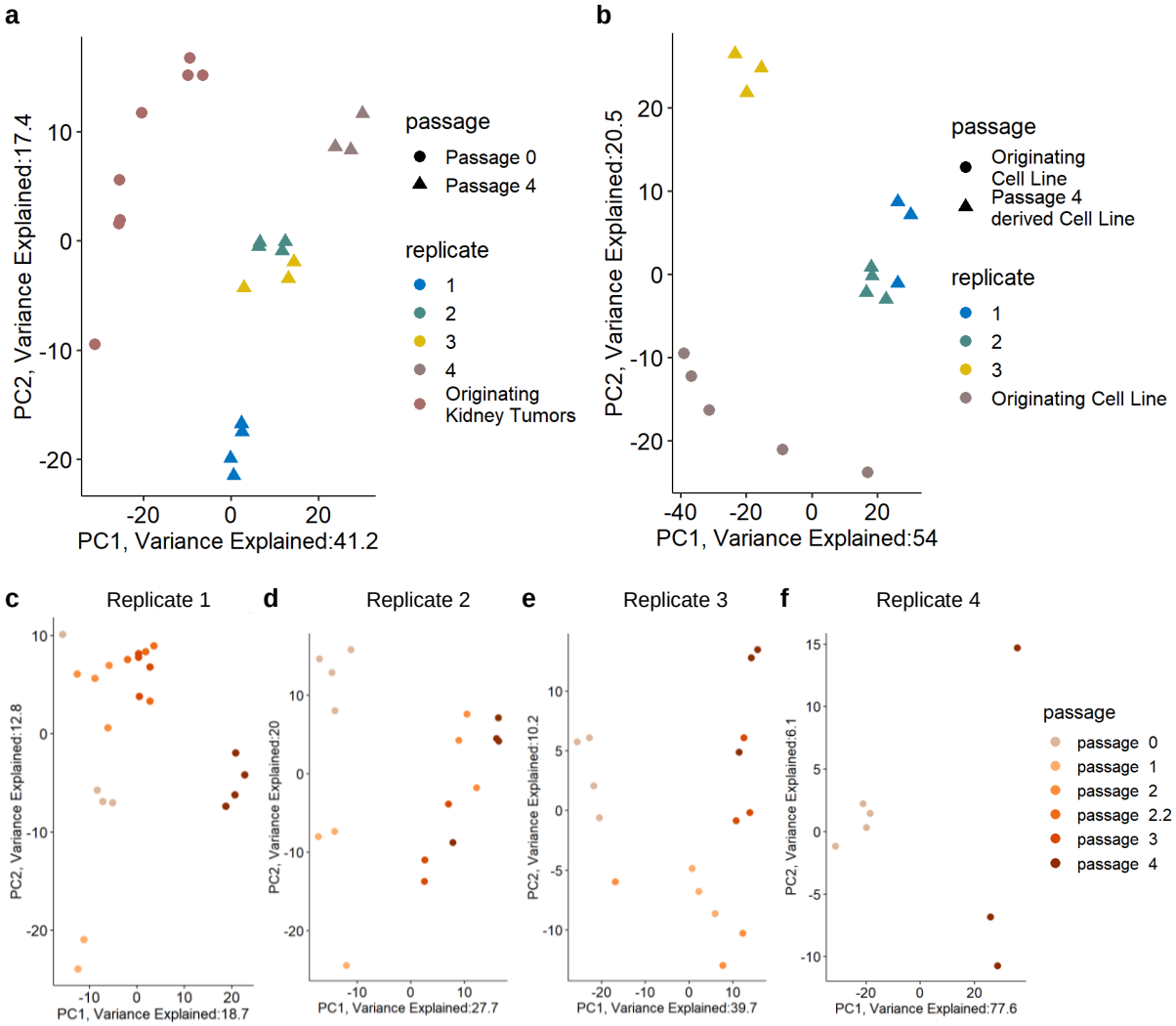


Figure 26 – The principal component analysis visualizes a common trajectory of variance between the passages of the *in vivo* selection

(a) Principal Component Analysis of gene expression profiles of the top 500 variant genes between the kidney samples. The first principal component describes most of the variance between the samples (41.2 %) and samples cluster by their passage of the in vivo selection. The primary tumor material that was not passaged yet clusters to the left and primary tumor material from the last passage clusters to the right. The second principle component explains variance derived from the different biological replicates (17.4 %).
(b) Principal Component Analysis of gene expression profiles of the top 500 variant genes between the analyzed cell lines. The first principal component explains 54 % of the variance between the samples and mostly describes the difference between the originating cell line and cell lines derived from the lung metastases of the last passage of replicate 3, 1 and 2. Differences between the originating cell line and passage 4 are mostly explained by the second principal component (20.5 % variance explained).
(c-f) Subset of each biological replicate of the different passages and kidney tumors. Variance between the first and the last passage can be explained by the first principal component.

3.5.5 DNA Methylation analysis recapitulates the variance of the gene expression profiles

To elucidate whether the samples not only vary in gene expression profiles but also on a genomic level, DNA methylation profiles of the first and the last passage of genomic DNA of the bulk kidney tumors were analyzed using an Illumina Infinium MethylationEpic BeadChip. Primary kidney tumors derived directly from orthotropic transplanted originating cell lines show a homogenous DNA methylation profile (**Figure 27a**). Variances in DNA methylation between the samples of passage 4 is more pronounced (**Figure 27b**). Nevertheless, no general trend and differences in global methylation on CpG Island (CGI) relation is observable. The hierarchical clustering of the most variant methylated CGIs by manhatten distance metric groups biological replicates together (**Figure 27c**). Replicates are also grouped together in a principal component analysis of the variant methylated probes and the primary tumors of the first passage cluster apart from tumors of the last passage (**Figure 27d**).

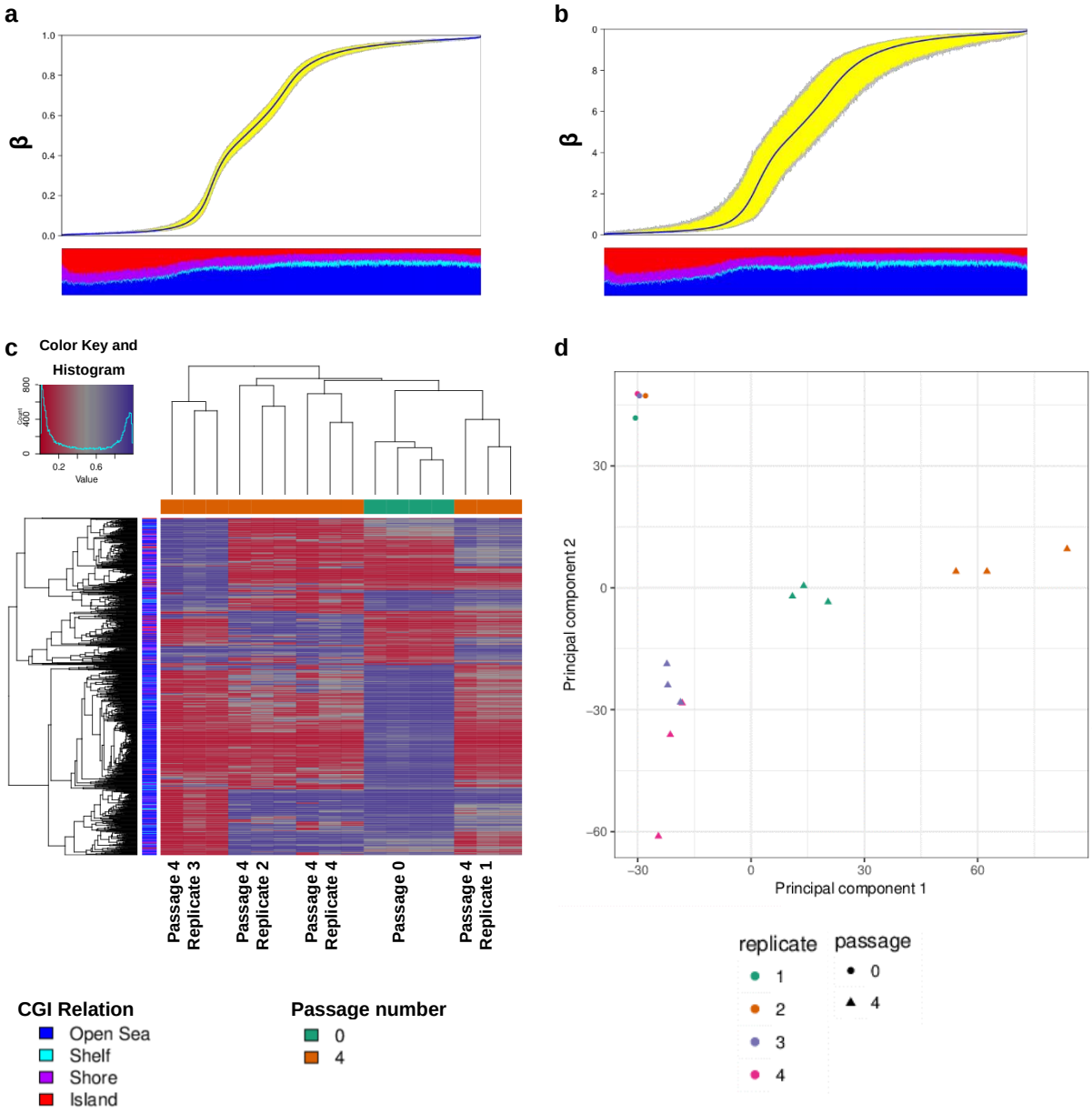


Figure 27 – DNA Methylation Analysis

(a) DNA methylation variance of genomic DNA of bulk kidney tumors is low, **(b)** whereas the DNA methylation in the kidney tumors of the last passage shows higher variance.
(c) Visualisation of the top 1000 variable loci in a heatmap; dissimilarity metric is manhattan distance and agglomeration strategy by average linking. Samples are colored based on the *in vivo* passage and sites are colored based on CGI relation.
(d) Principal component analysis of the first two principal components. Each biological replicate and each *in vivo* passage clusters together as a distinct group.
Plots were generated using the *RnBeads* package.

3.5.6 Differential expressed genes are associated with cell migration, extracellular binding and extracellular matrix

To obtain a list of significantly differentially expressed genes, linear models were fitted by generalized least squares probe-wise following the *limma* differential gene expression pipeline, biological replicates group-mean parametrized and gene-wise ranked using the empirical Bayes method[270,271].

In order to identify genes that mediate the increased aggressiveness and metastatic outgrowth, we compared the differentially expressed genes between both the bulk tumors of passage 0 and passage 4 and between the originating cell line and passage 4 lung metastases derived cell lines. With the knowledge that the human specific bead array might false positively identify genes of the mouse microenvironment, as stated in **section 3.5.2**, we decided to solely rely on differential expressed genes of the primary tumors that are also significantly differentially expressed in the cell culture setting (adj. *P* value < 0.05). To reduce the possibility of false negative exclusion of primary tumor specific genes we did not apply any fold-change cutoff on the cultured cells.

Of the 3356 significantly differentially expressed genes in the cell culture setting, 183 genes were at least two-fold differentially expressed between primary tumors of passage 4 versus passage 0 (**Figure 28**). 140 genes were only differentially expressed *in vivo* and might therefore tumor microenvironment associated (**Supplemental Figure 6**).

The 183 differentially expressed genes are highlighted in a volcano plot in **Figure 28b** and the 16 genes that are at least 8-fold differentially expressed are additionally labeled. The top upregulated genes are Annexin A10 (*ANXA10*), Tetraspanin 8 (*TSPAN8*), Biglycan (*BGN*), C-X-C Motif Chemokine Ligand 5 (*CXCL5*) and Interleukin 13 Receptor Subunit Alpha 2 (*IL13RA2*). The top down-regulated genes are Somatomedin B And Thrombospondin Type 1 Domain Containing (*SBSPON*), Secreted Phosphoprotein 1 (*SPP1*), Plasminogen Activator, Urokinase (*PLAU*), SLIT And NTRK Like Family Member 4 (*SLITRK4*) and Collagen Type VIII Alpha 1 Chain (*COL8A1*). The biological replicates, except of one sample, cluster together when the differentially expressed genes are plotted in a heat map (**Figure 28c)**.

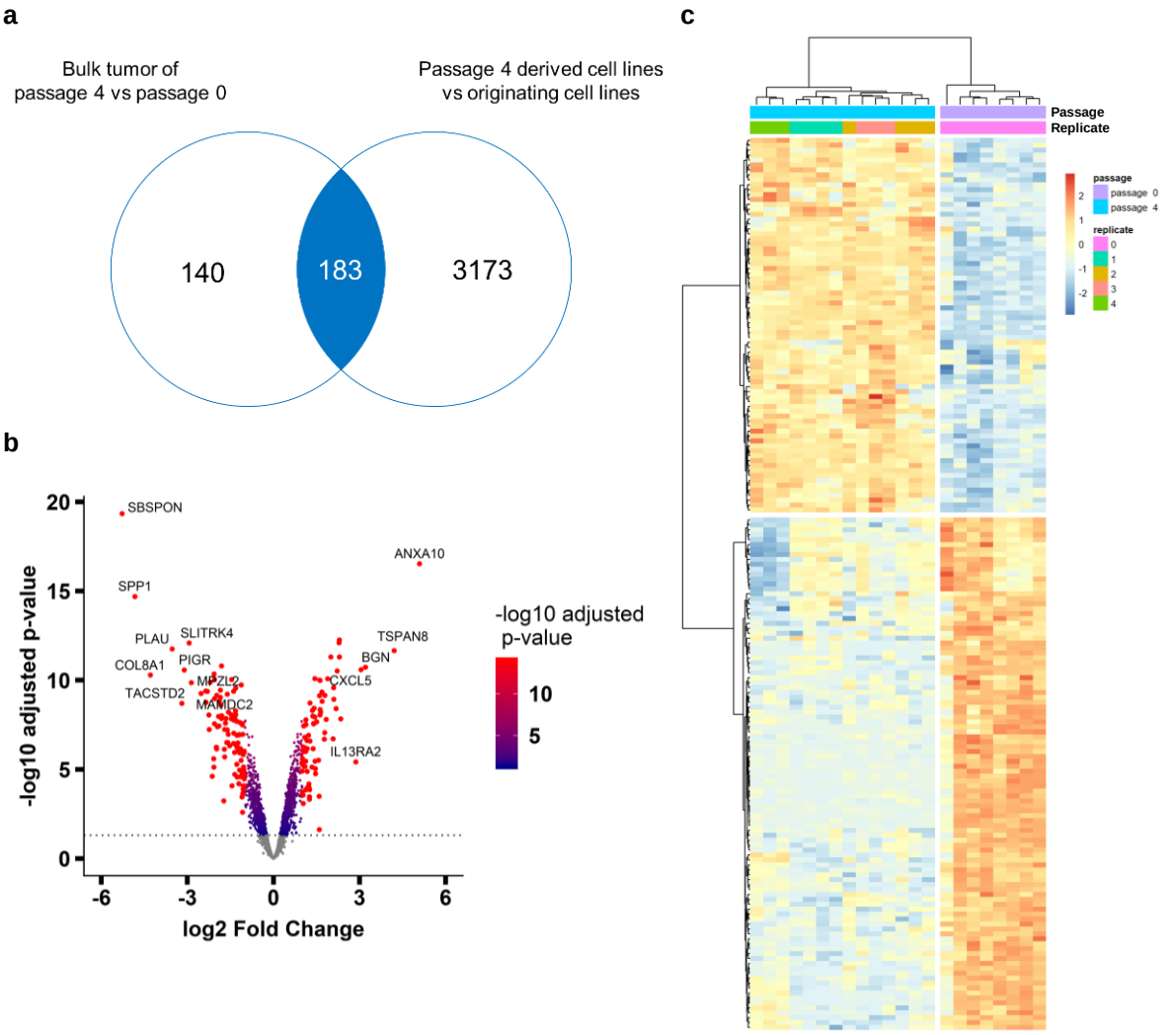


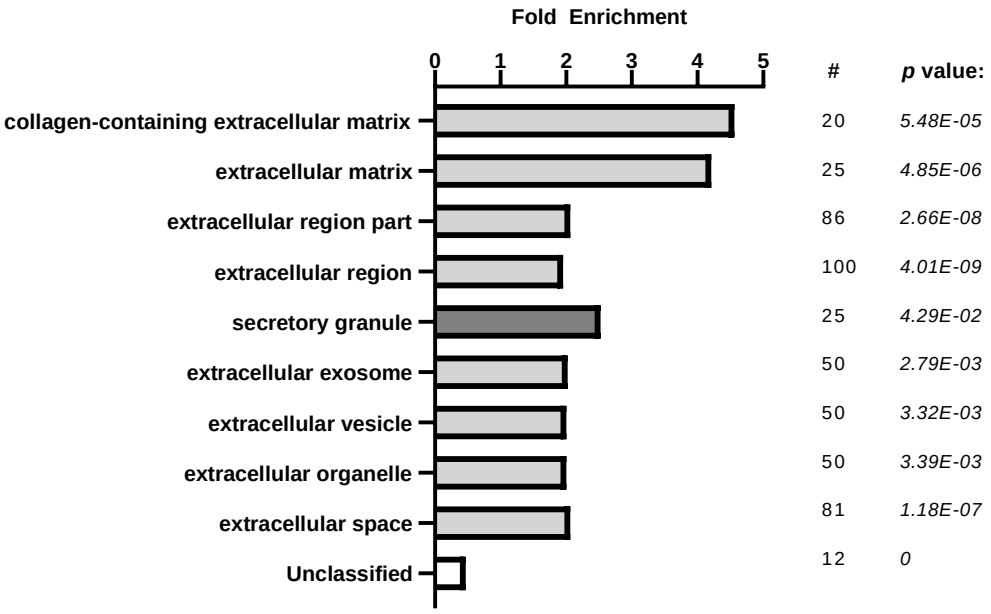
Figure 28 – Differential Gene Expression analysis

(a) VENN Diagram of all differentially expressed genes with an adjusted *P* value < 0.05 (BH corrected) between the originating and passage 4 derived cell lines, and at least two-fold differentially expressed genes of the bulk kidney tumors of passage 4 and passage 0.
(b) Volcano plot of all differentially expressed genes of the bulk tumor, which are also differentially expressed in the cell culture setting. Highlighted in red are all 183 genes that are at least two-fold differentially expressed. Annotated are all genes that are at least 8-fold differentially expressed.
(c) Heat map of the 183 differentially expressed genes. Biological replicates cluster, except of one sample, together (Manhattan distance, ward.D linkage).

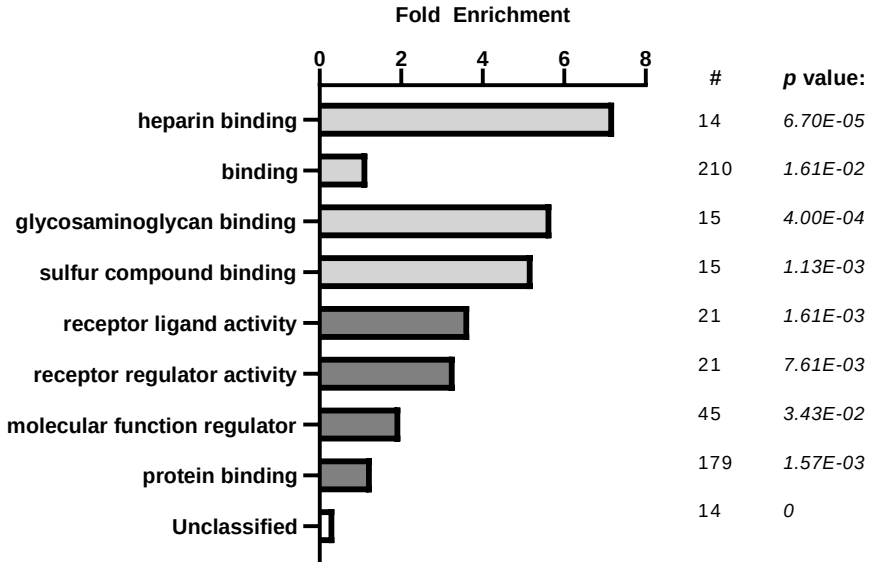
3.5.7 The in vivo selection selected for extracellular associated genes

The cellular component of the genes is enriched for either extracellular or extracellular associated genes (**Figure 29a**). In addition, the molecular function of these genes is associated with extracellular binding and communication (**Figure 29b**) and their biological function with circulation, cell migration, structure regulation and response to stimuli (**Figure 29c**).

a



b



c GO biological process

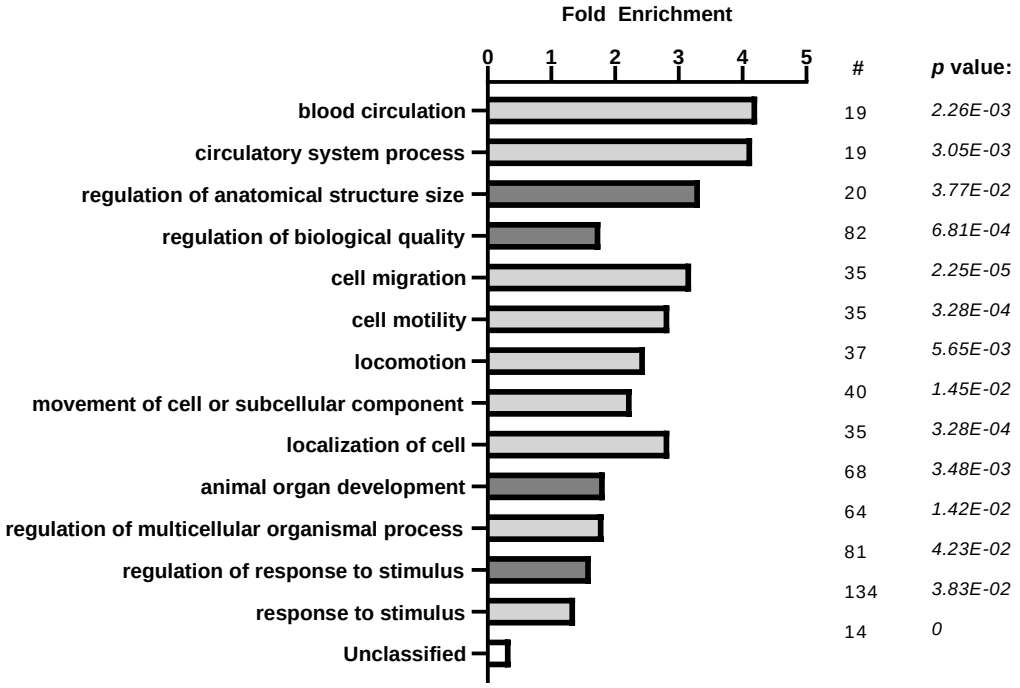


Figure 29 – PANTHER overrepresentation test for the gene ontology of the differentially expressed genes

(a) The differentially expressed genes are enriched in the GO cellular component for either extracellular or extracellular associated genes. **(b)** The GO molecular functions are enriched for genes that have a role in extracellular binding and communication. **(c)** The most enriched GO biological processes are associated with circulation, cell migration, structure regulation and response to stimuli.
Terms of the same adjacent biological items are colored identically.
PANTHER Overrepresentation Test (Released 2018-11-13). GO Ontology database released 2018-12-01.
Fisher's exact test, Bonferroni correction for multiple testing.

3.6 Selection of clinical relevant genes from the *in vivo* selection

The genes we identified with our *in vivo* selection experiment correlate with extracellular communication and migration, indicating that they might play a role in metastatic outgrowth. Therefore, the question arose whether these genes might be clinically relevant and would correlate with metastatic outgrowth and tumor aggressiveness.

We made use of the publically available TCGA-KIRC patient dataset. Next to RNASeq gene expression data, methylation, mutation profiles, survival data and clinical characteristics have been collated. Median follow-up was 39 months (**Supplemental Figure 7**).

In order to identify clinical relevant processes essential for primary tumor outgrow, a hallmark of stage III tumors, we correlated the differentially expressed genes of the *in vivo* selection to processes that happen between stage I and stage III of ccRCCs. Stage III tumors have by definition not metastasized to distant organs yet but invaded local tissues.

3.6.1 Identification of clinical covariates that are predictive for cancer specific survival

It is important to include clinical characteristics that are prognostic for cancer specific survival into a gene signature model, as they might correlate with the expression of various genes. Therefore, univariate analysis (univariate Cox regression) was performed on all patients of stage I to III with all clinical variables of the dataset that were annotated in at least 90 % of the patients and had > 3 patients per group (see **Supplemental Figure 7**). The hazard function *h(t,x)* and from this function derived hazard ratios (HR) for the univariate analysis are defined in **equation (8) and (9)**. To identify the optimal cut point for stratifying patients into risk groups for age and tumor size the maximally selected rank statistics for the outcome of cancer specific survival probability was calculated. Thereby, the optimal cutoff for the age was identified at 51 years ≡ -0.791 standardized age and tumor size at 7.7 cm.

Univariate Cox PH model

$$h(t|\, covariate) = h_0(t) \times e^{\beta \times charasteristic} \tag{8}$$

Hazard Ratio (HR)

$$HR_{\frac{covariate}{reference}} = \frac{h(t|covariate)}{h(t|reference)} \tag{9}$$

$h_0(t)$: Baseline hazard

β: Corresponding coefficient of the covariate

Of the 17 tested biomarkers for the univariate regression, eight were significantly associated with cancer specific survival of stage I to III patients ($P \leq 0.05$) (**Figure 30**). Stratification of patients according to the TNM staging system, tumor grade and the SSIGN Score show a significant survival benefit. Interestingly, primary tumors located in the right kidney show favorable survival (HR = 0.38; $P = 0.003$) and patient age at diagnosis had no significant association with cancer specific survival (HR = 1.3; $P = 0.078$).

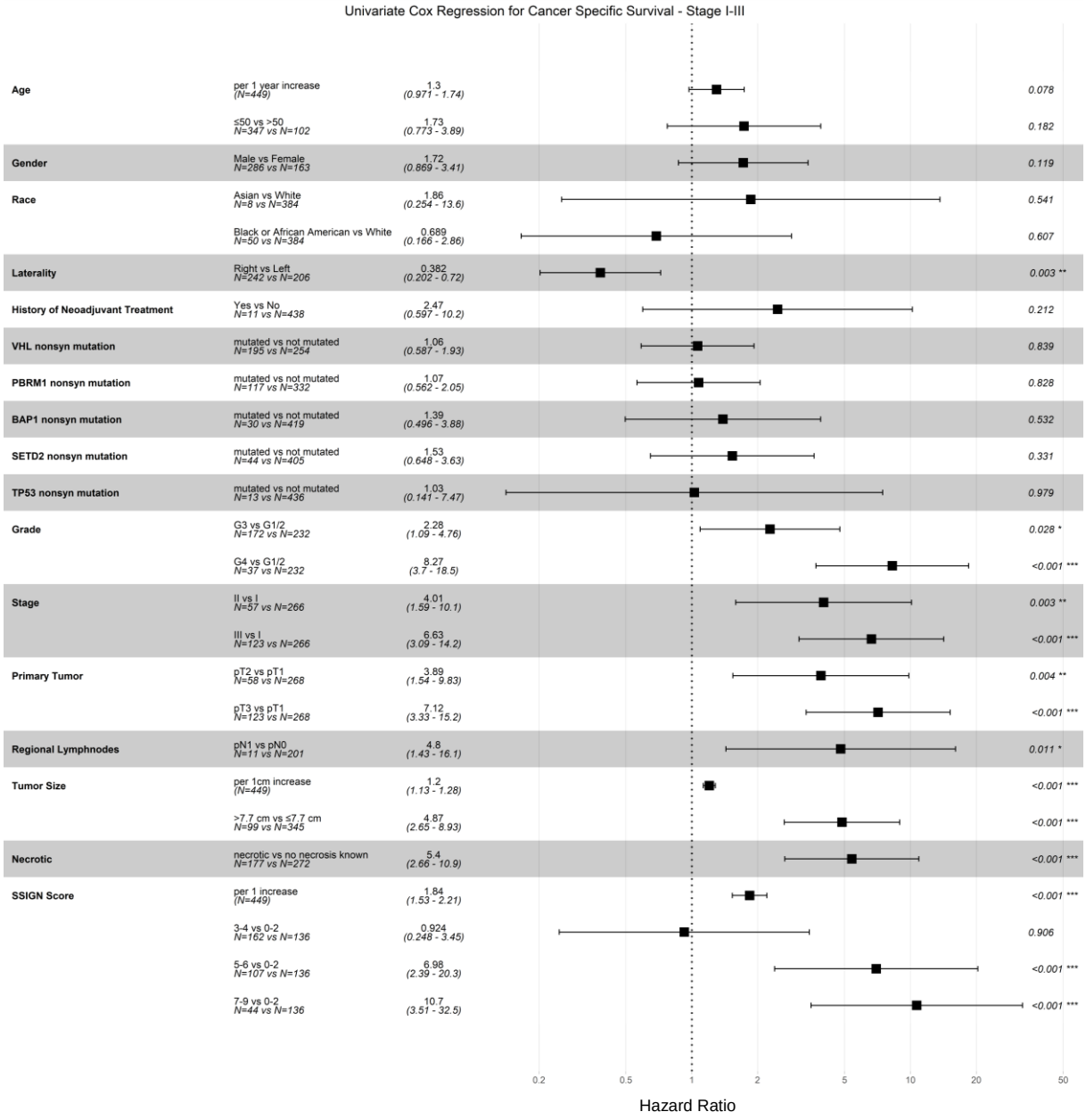


Figure 30 – Forest plot: Univariate Cox Regression for Cancer Specific Survival (Stage I-III)

Univariate regression was calculated for the TCGA-KIRC dataset separately for each comparison group using 8. The impact size of each covariate is estimated by their associated beta coefficients and hazard ratios. Age as a continuous variable was standardized to exclude outlier effects. Grade 1 and 2 patients were combined as reference group for patient grade. Clinical parameters were analyzed when data was present for > 90 % of the patients. Groups with < 3 patients per group were excluded. nonsyn = nonsynonymous * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Error bars represent the 95 %-CI.

Nevertheless, standardized patient age correlates with overall survival of patients (HR = 1.41; $P < 0.001$) and especially with survival of a non-cancer caused death (HR = 2.10; $P < 0.001$) (**Figure 31a-c)**. While stratification into two age groups could significantly separate patients "overall survival" (**Figure 31d;** log rank $P = 0.00047$) and "not cancer specific survival" probability (death not by cancer; **Figure 31e;** log rank $P = 0.0035$), the estimators were not significantly different for "cancer specific survival" (CSS) (**Figure 31f)**.

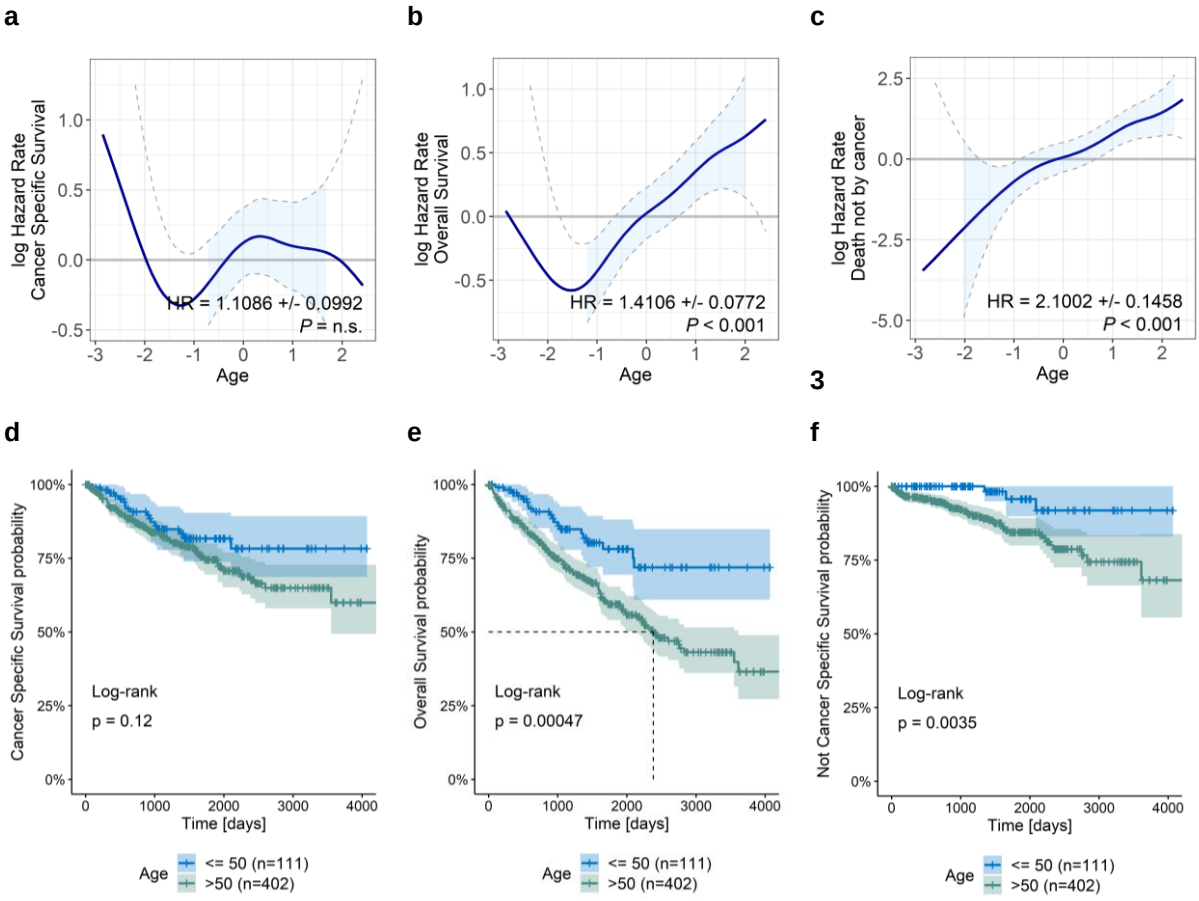


Figure 31 – Age is not a significant predictor for cancer specific survival

(a) The log hazard rate of cancer specific survival for all tumor stages shows no significant association with the standardized age of patients at diagnosis. **(b)** Age correlates significantly with overall survival (HR = 1.41 ± 0.08) and especially with **(c)** death not related to cancer (HR = 2.10 ± 0.15).
Kaplan-Meier Estimator of **(d)** cancer specific survival, **(e)** overall survival and **(f)** not cancer specific survival for patients stratified by age 51. This stratification cannot separate cancer specific survival probability into significant different estimators.
The optimal cut point for stratification by age was calculated using the maximally selected rank statistics. Filled areas represent the 95 %-confidence intervals. *P* value calculated by log rank test.

All clinical and histopathological characteristics for tumor stage, histopathological primary tumor size (T-stage), the presence of regional (N-stage) and distant metastases (M-stage), tumor grade, tumor size, presence of metastases and the SSIGN Score could significantly predict patients' cancer specific survival of all tumor stages (**Figure 32**).

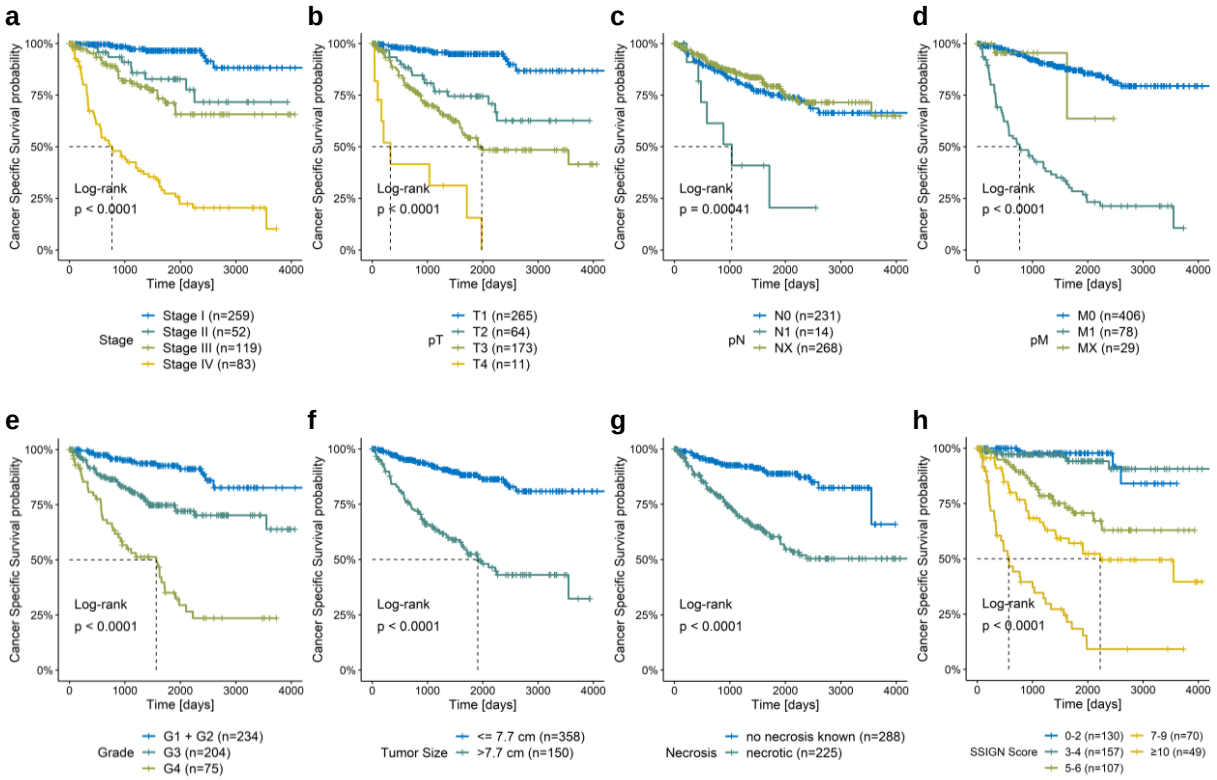


Figure 32 – Kaplan-Meier estimators for clinical/pathologic characteristics and cancer specific survival

ccRCC patients can be stratified significantly into significantly distinct risk groups by **(a)** tumor stage, **(b)** primary tumor size, the presence of **(c)** regional and **(d)** distant metastases, by **(e)** tumor grade, **(f)** tumor size, **(g)** presence of necrosis and **(h)** by the SSIGN-Score. *P* values calculated by log rank test.

3.6.2 Lasso Regression variable shrinkage

To obtain a model able to identify groups of patients with low survival probability, especially in early stage cancers which reflect the model of the *in vivo* selection, least absolute shrinkage and selection operator (LASSO) regression in the context of Cox's proportional hazards model (Cox regression) was performed on a total of 444 stage I-III patients (early stages) of the ccRCC TCGA Dataset (**Figure 33**). Two left-censored patients and three patients discrepant for stage annotation were excluded from the analysis.

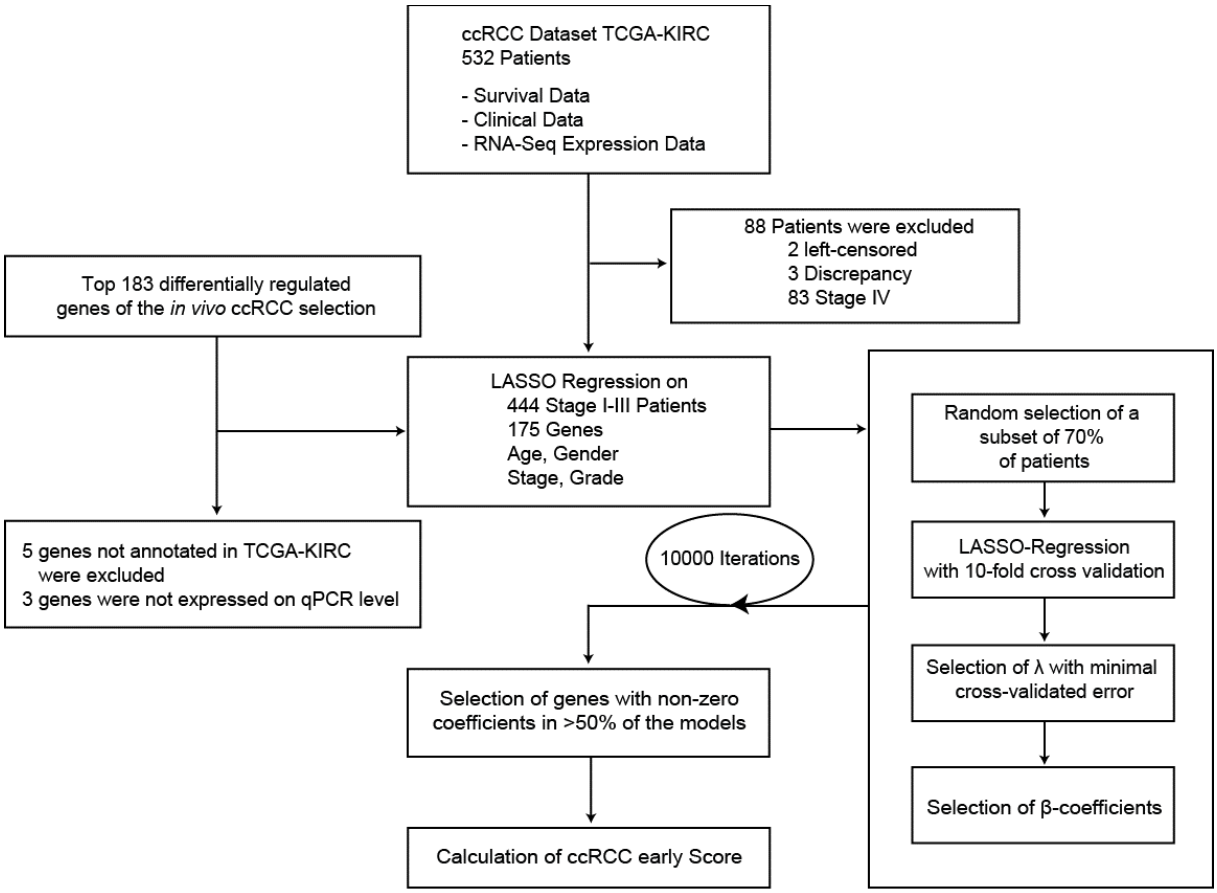


Figure 33 – Schema describing the method used to obtain the ccRCC early score

RNASeq Gene expression data of 178 of the 183 highly differentially expressed genes identified in the *in vivo* selection was annotated in the TCGA dataset. Three of the genes could not be verified by qRT-PCR (**Supplemental Figure 8**). Standardized gene expression data together with clinical parameters that had a *P* value below the threshold of 0.2 in the univariate Cox regression (**see Figure 30**) were used as covariates to estimate a Cox proportional hazard model[272]. The Mayo SSIGN score, tumor laterality, tumor size in cm and necrosis status were predictive in univariate analysis for patients' survival. These variables were not included as covariates into the regression, as these clinical characteristics were missing from the Sato *et al.* test dataset (**see Supplemental Figure 7 & Figure 48**).

Hazard rates of the tumor stages (**Figure 34a**) and primary tumor classification (**Figure 34b**) are not distributed linearly in both cancer specific survival and overall survival, whereas tumor grade stratifies patients approximately linearly (**Figure 34c**). This can also be noted in the Kaplan-Meier estimators of these clinical characteristics. Stage II and III alone for example cannot separate the survival curves significantly (**Figure 32a**; $p = 0.255$) and estimators of primary tumor size pT2 and pT3 alone are marginally significantly different (**Figure 32b**; $p = 0.043$).

The clinical characteristics stage, pT stage and grade do not correlate linearly with the relative hazard. Therefore, instead of replacing these ordinal clinical characteristics with an ascending numeric sequence as covariate in the regression, a binary covariate model has been chosen: Each level of the clinical characteristics was given its own variable separately.

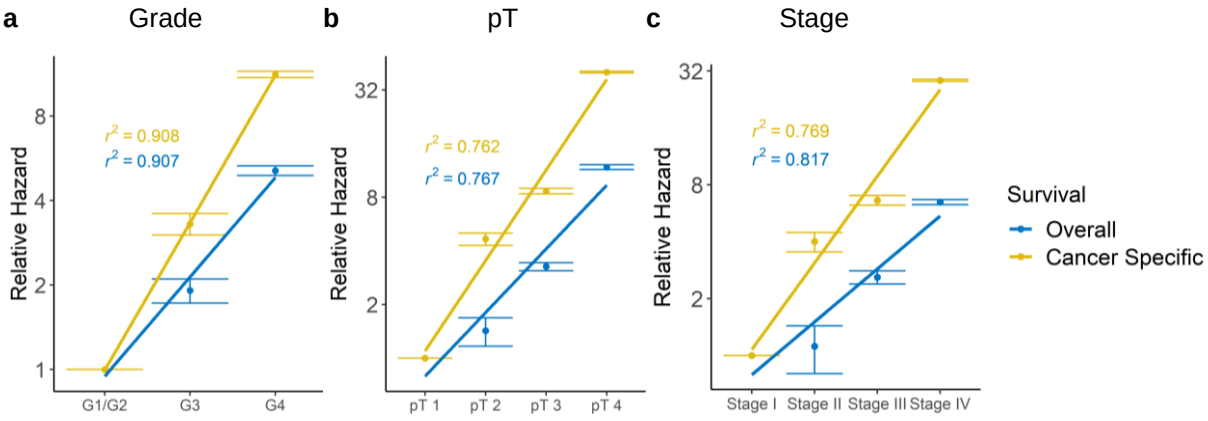


Figure 34 – Correlation of stage, grade and primary tumor size with the relative log hazard of overall and cancer specific survival

(a) Tumor grade correlates approximately linearly with the relative log hazard in both overall and cancer specific survival. The coefficient of determination R^2 of the linear correlation of **(b)** primary tumor T-stage **(c)** and tumor stage was less pronounced.
Linear regressions for each group are drawn. Error bars depict HR ± standard error.

Nodal status in 268 and distant metastases status of 29 patients were not assigned in the clinical annotation of the dataset. Hence, only the AJCC staging system was included, as it is based on the TNM staging system and clinical information of regional and distant metastases are included in the annotation.

In summary, the clinical parameters stage and grade, together with the standardized gene expression data of 175 genes were assessed as covariates for the regression model. As negative controls we included additionally standardized age and gender.

When calculating a model with many regression coefficients, the *L-1* norm penalized estimation by the LASSO regression method can be used to perform variable shrinkage and selection[264]. Some regression coefficients for genes and clinical parameters (covariates of the

regression) are thereby due to the nature of the used *L-1* norm minimized exactly to zero. Thus, the LASSO regression allows a gene selection in an unbiased manner to score patients by gene expression values and predict patient survival[266]. The number of variables in the model is controlled by the amount of regularization using a tuning parameter λ. The optimal λ-value for each regression was calculated by 10-fold cross-validation.

To avoid the risk of overfitting the model to the dataset, a bootstrapped approach was developed (**see Figure 33**). 70 % of the patients of the TCGA-KIRC dataset were selected randomly as a training dataset on which the LASSO-Regression is applied by using the *coxnet* function of the *glmnet* package in R[265,273]. This was repeated for 10000 iterations on the DKFZ ODCF PBS computer cluster. Subsequently, only factors with non-zero regression coefficients in more than 50 % of the LASSO-models were selected to calculate the beta values of the Stage I-III model (**Figure 35**)[274].

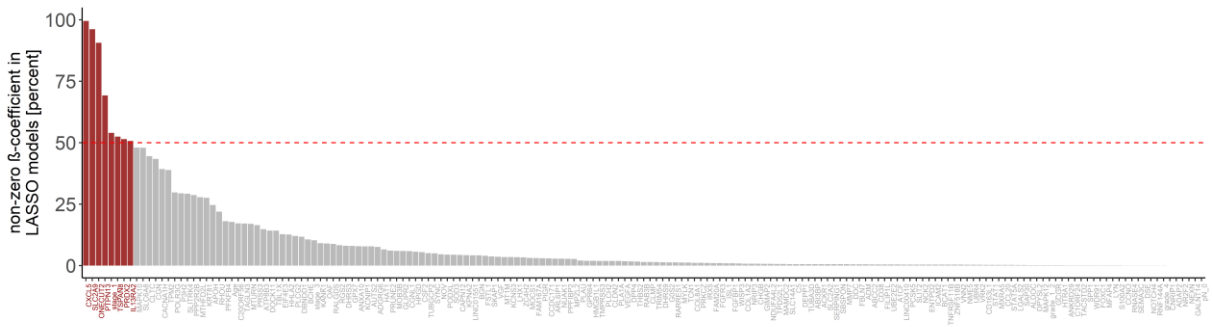


Figure 35 – Percentage of non-zero β-coefficients of all covariates of the LASSO regression

The genes *CXCL5*, *SLC2A9*, *ONECUT2*, *PTPN3*, *TSPAN8*, *PRDX2* and *IL13RA2* of the *in vivo* selection and the clinical covariate "stage I" are the only covariates that were selected by the LASSO regression into more than 50 % of the models.

We termed the model of the weighted sum of expression and patient stage I "**ccRCC early Score**":

$$\begin{aligned} \textit{ccRCC early Score} = {} & 0.084 \times \textbf{\textit{IL13RA2}} - 0.342 \times \textbf{\textit{SLC2A9}} - 0.132 \times \textbf{\textit{PTPN13}} + \\ & 0.327 \times \textbf{\textit{CXCL5}} + 0.010 \times \textbf{\textit{PRDX2}} + 0.115 \times \textbf{\textit{TSPAN8}} + \\ & 0.172 \times \textbf{\textit{ONECUT2}} - 1.398 \times \textbf{stage 1} \end{aligned} \qquad (10)$$

As the list of differentially expressed genes used by the LASSO regression was initially derived from microarray data with the chance of false positive results, we validated the expression of the seven selected genes by quantitative real-time PCR (qRT-PCR) using human specific TaqMan probes (**Figure 36**).

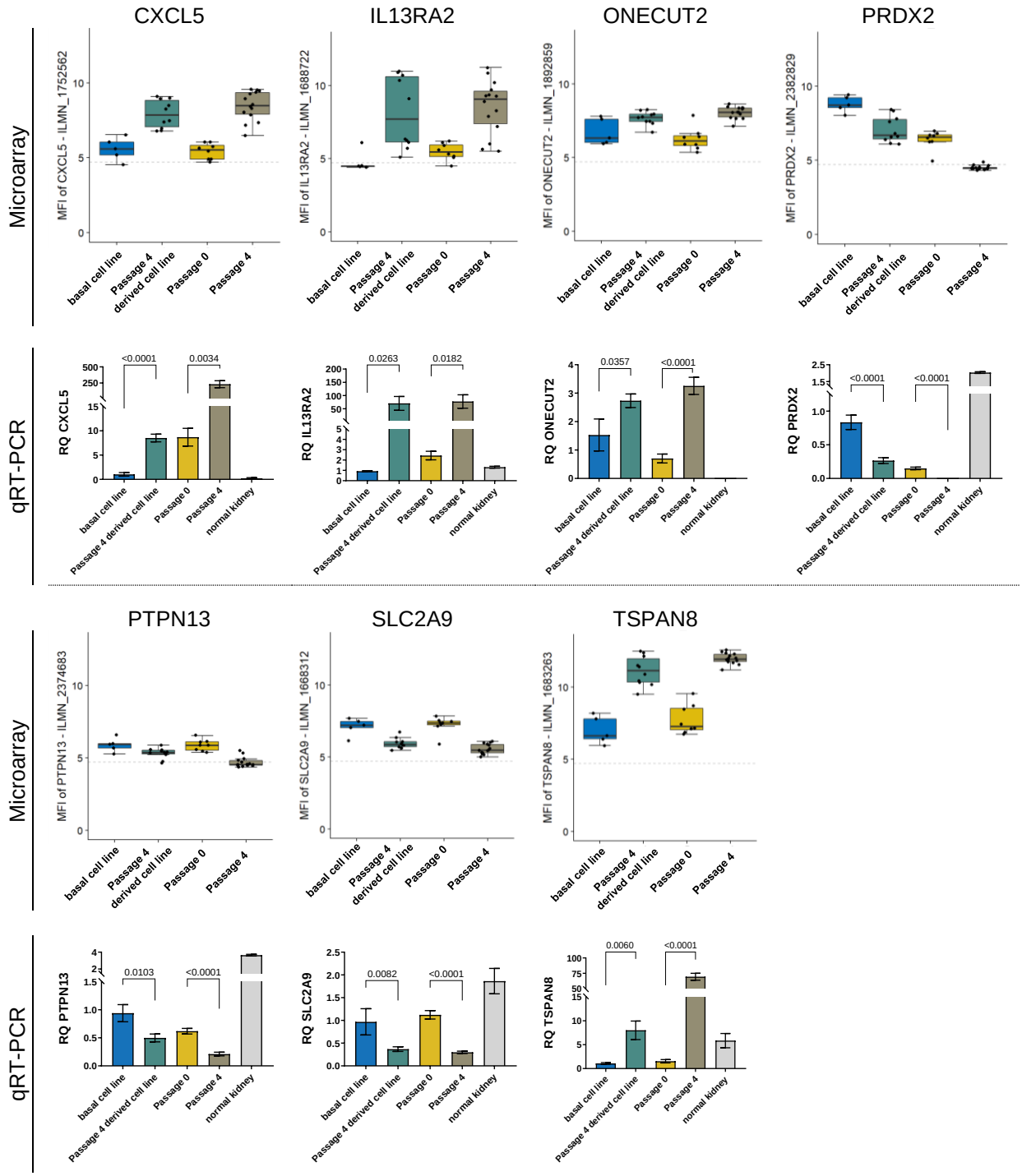


Figure 36 – Validation of ccRCC early score genes by quantitative TaqMan real-time PCR

The expression of ccRCC early score genes is consistent between microarray and qRT-PCR.
Microarray: Boxplot, error bars depict median ± 1.5*IQR
qRT-PCR: Barplot, *P* values calculated using an unpaired, two-sided Student's *t*-test. Error bars depict mean ± s.e.m.

To validate whether the ccRCC early Score can recapitulate the *in vivo* experiment, the score was calculated on the standardized gene expression microarray data of the *in vivo* selection (**Figure 37**). We saw a high correlation between the ccRCC early score and the passage number of the *in vivo* expression data. This reveals us that the ccRCC early score derived from the differentially expressed genes of our experimental setup can predict the *in vivo* passage. The ccRCC early score was obtained with the help of the RNAseq expression data of the TCGA-KIRC dataset and gene expression data of the *in vivo* selection experiment was assessed by a microarray. This gives us a hint that the ccRCC early score may be applicable across platforms.

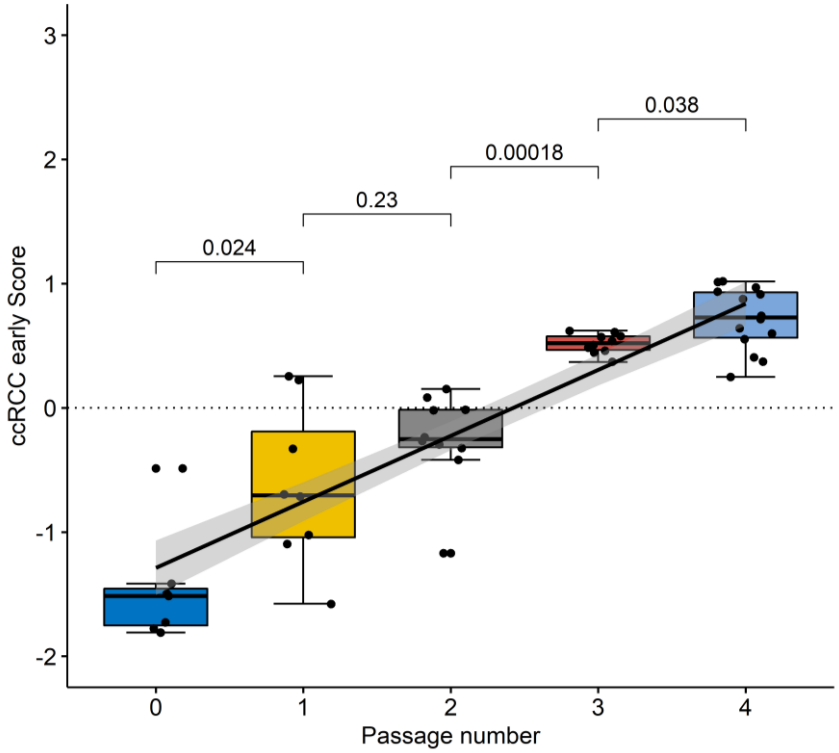


Figure 37 – Calculation of ccRCC early scores from gene expression data of primary tumors of the *in vivo* selection model

To validate the *in vivo* selection, ccRCC early scores were calculated from the microarray gene expression data (KIKA75 was derived from a stage III tumor). Kidney xenograft tumors of the originating tumor have a ccRCC early Score of -1.55 ± 0.49 and kidney xenografts of the last passage have a ccRCC early Score of 0.71 ± 0.25. Between the score and the passages is a linear relationship with an adjusted R^2 of 0.794.
The black line represents the linear fit between the different passages and grey areas the 95 % confidence intervals of the regression. *P* value calculated using a two-sided Student's *t*-test between adjacent passages. Error bars depict mean ± 1.5*IQR.

To assess the ability of each of the ccRCC early Score genes to predict the clinical outcome of patients' survival, the univariate hazard ratios of stage I to stage III patients of each gene alone were plotted against the log fold-changes of the same genes in the differential expression between the last passage and the originating xenografts of the *in vivo* selection (**Figure 38**). All genes with a log fold-change > 1 have an increased risk of cancer specific death and all genes with a log fold-change < 1 have an improved survival probability. Therefore, the *in vivo* selection model selected for genes that correlate to the survival probability of ccRCC patients.

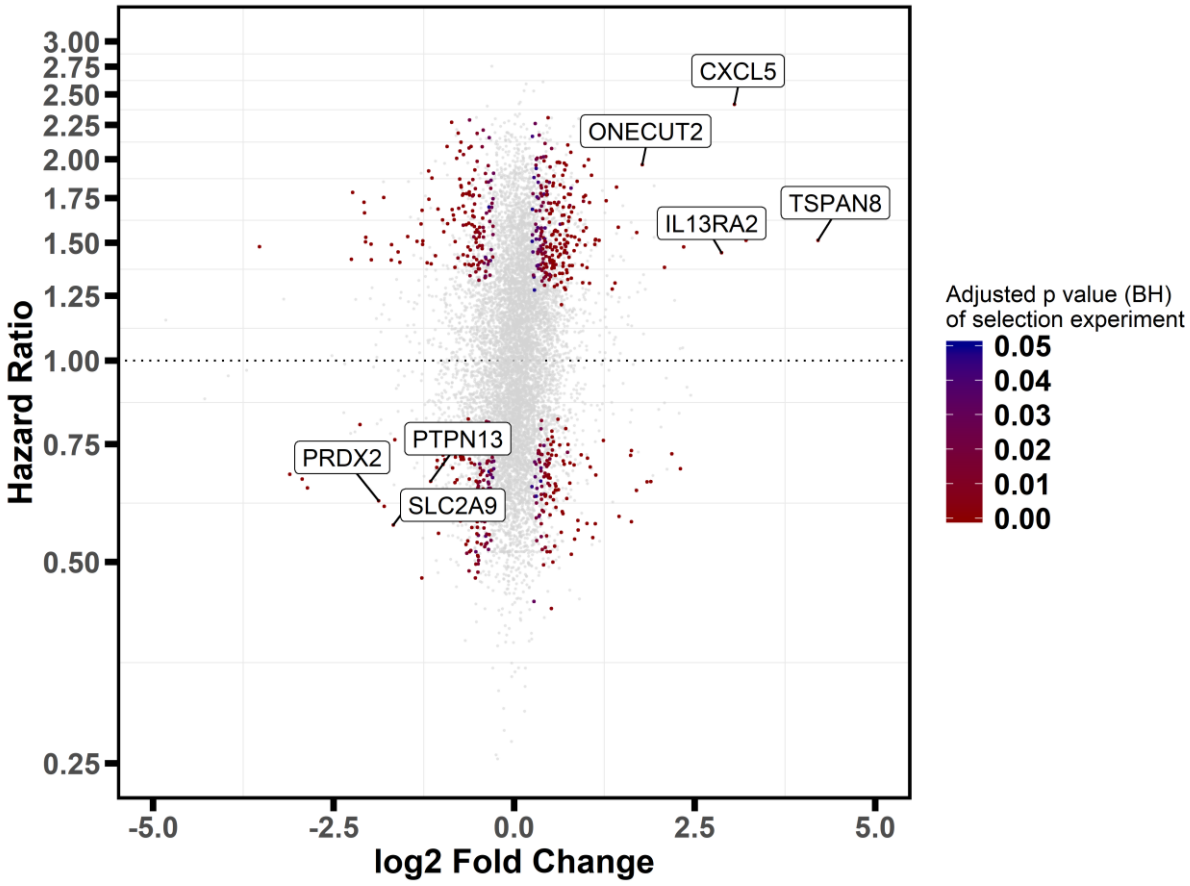


Figure 38 – Univariate hazard ratio vs. log2 fold change of the *in vivo* selection

Scatter plot of the differential expressed genes between parental tumors and passage 4 as well as the corresponding univariate hazard ratios of all genes annotated in the TCGA-KIRC dataset of patients of stage I-III patients. Labeled are the genes that were selected by the LASSO regression.
Genes with a significant univariate hazard ratio which are significantly differentially expressed between passage 4 and passage 0 (FDR < 0.05) and which are additionally significantly differentially expressed between passage 4 lung metastases derived cell lines and the originating cell line are labeled according to their adjusted *P* value of the differential expression. Otherwise they are labeled light grey.

3.6.3 The ccRCC early score is the a significant predictor of stage I-III patient survival in a multivariate analysis

To assess the efficiency of the variable shrinkage of the LASSO regression, a multivariate analysis (Cox multiple regression) was performed using the ccRCC early score and all clinical variables that were used as covariates for the regression (**Equation 11**). The results are presented as a Forest plot in (**Figure 39**).

Cox multiple PH model

$$h(t,x) = h_0(t) \times \exp(\beta_1 \times clinical\ attribute\ 1 + \beta_2 \times clinical\ attribute\ 2 + \cdots + \beta_n \times clinical\ attribute\ n) \quad (11)$$

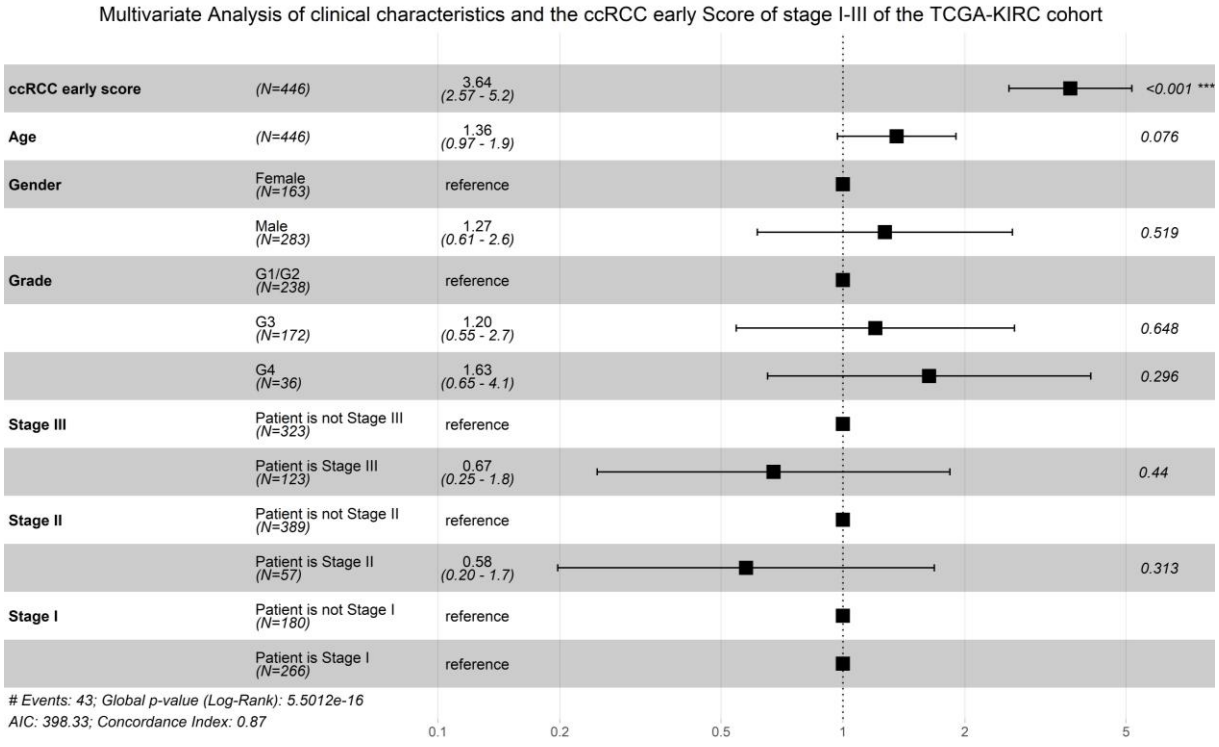


Figure 39 – Forest Plot of the multivariate analysis of clinical phenotypes and the ccRCC Score of stage I-III patients of the TCGA-KIRC dataset

Clinical parameters used in the LASSO regression were included in the multivariate analysis (Cox multiple regression as in **(11)**. Only the ccRCC early score shows a significant *P* value. 84 patients of stage IV or those with missing information in any of the covariates where excluded from the analysis. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Error bars represent the 95 %-CI.

3.6.4 Obtaining a general applicable ccRCC score

When patients of all stages in the ccRCC dataset were scored and tested in a multivariate analysis, only stage IV stays together with the ccRCC early score as significant predictor for patients' survival (**Figure 40**).

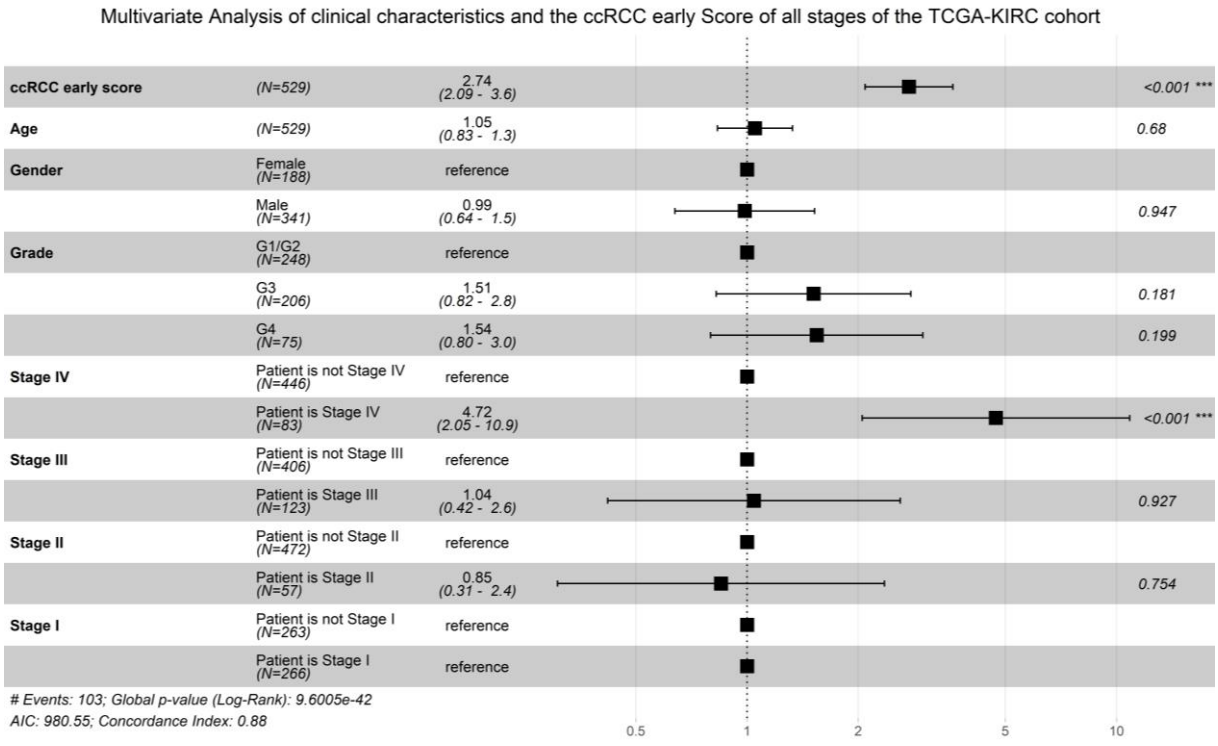


Figure 40 – Forest plot of the multivariate analysis of clinical parameters and the ccRCC Score in all patients of the TCGA-KIRC dataset

Only the ccRCC Score and the clinical parameter "Stage IV" showed significant *P* values. 3 patients with missing information in any of the covariates where excluded from the analysis. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Error bars represent the 95 %-CI.

To enhance the predictive power of the ccRCC early score for all patients over all stages, a multiple Cox regression was performed by including the ccRCC early score together with stage IV annotation (**Figure 41 and Equation 12**).

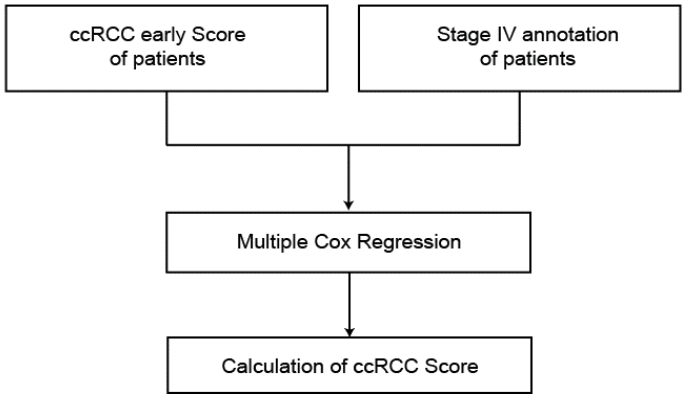


Figure 41 – Schema describing the method to obtain a generally applicable ccRCC score

Obtaining the generally applicable ccRCC Score model

$$ccRCC\ Score = \beta_1 \times \text{ccRCC early Score} + \beta_2 \times \text{stage IV} \quad (12)$$

$$= 1.085 \times \text{ccRCC early Score} + 1.621 \times \text{stage IV}$$

$$= 1.085 \times (0.084 \times IL13RA2 - 0.342 \times SLC2A9 - 0.132 \times PTPN13 + 0.327 \times CXCL5 - 0.010 \times PRDX2 + 0.115 \times TSPAN8 + 0.172 \times ONECUT2 - 1.398 \times \text{stage I}) + 1.621 \times \text{stage IV}$$

$$\mathbf{= 0.091 \times IL13RA2 - 0.371 \times SLC2A9 - 0.143 \times PTPN13 + 0.355 \times CXCL5 - 0.011 \times PRDX2 + 0.125 \times TSPAN8 + 0.187 \times ONECUT2 - 1.516 \times \text{stage I} + 1.621 \times \text{stage IV}}$$

3.7 The ccRCC Score robustly predicts cancer specific survival of ccRCC patients

With the help of the new ccRCC Score we calculated an individual ccRCC score for every patient in the TCGA-KIRC dataset. To assess the effect of the continuous predictor on survival, we plotted the Cox proportional log hazard smoothed by the ccRCC Scores of the patients[275] (**Figure 42**).

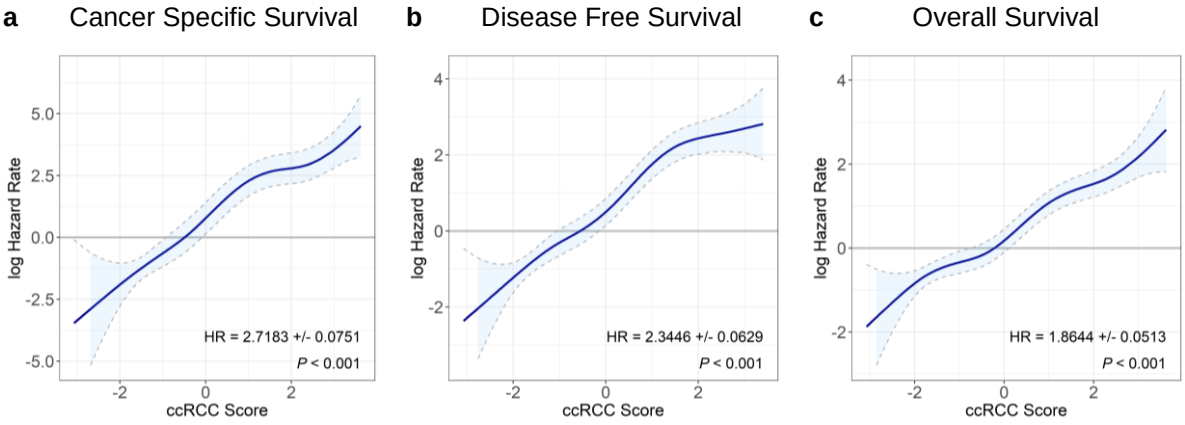


Figure 42 – Cox proportional log hazard smoothed by ccRCC Scores of the patients.

The log hazard ratio of **(a)** cancer specific survival, **(b)** disease free survival and **(c)** overall survival with the ccRCC Score of a patient show highly significant ($p < 0.001$) linear relationships.

The higher the ccRCC Score of a patient, the higher is the individual hazard risk for both cancer specific survival (CSS) and disease free survival (DFS).

For example, the estimated hazard of a patient with a ccRCC Score of 2 has a log hazard rate of 2.82, another patient with a ccRCC Score of -2 has a log hazard rate of -1.91. The estimated hazard for the patient with the ccRCC Score of 2 is:

$$estimated\ hazard = \frac{e^{2.82}}{e^{-1.91}}$$

$$\approx 113.30$$

Therefore the patient with a ccRCC Score of 2 has an about 113 times increased cancer mortality risk in comparison to the patient with a ccRCC Score of -2.

The observation of a correlation between ccRCC Score and log hazard ratio holds also true when patient survival was analyzed for each stage separately (**Figure 43a-d**).

The ccRCC Score was trained on patients of tumor stage I-III of the TCGA-KIRC dataset. Hence, the ccRCC Score could also predict survival in stage IV patients, which is an additional validation of the score and its method (**Figure 43d**).

The ccRCC Score in stage I ranges from approximately -3 to 1, in stage II from -1 to 2, stage III from -1 to 3 and in stage IV from 0 to 4. Therefore, the question arose whether the distribution of ccRCC Scores according to stage was significantly different (**Figure 43e**). Testing correlation between stage and score using Kendall's rank correlation method shows a significant correlation of R^2 = 0.71 (P < 0.0001). Testing for differences between the mean showed significant differences between all of the stages, except between stage II and stage III patients.

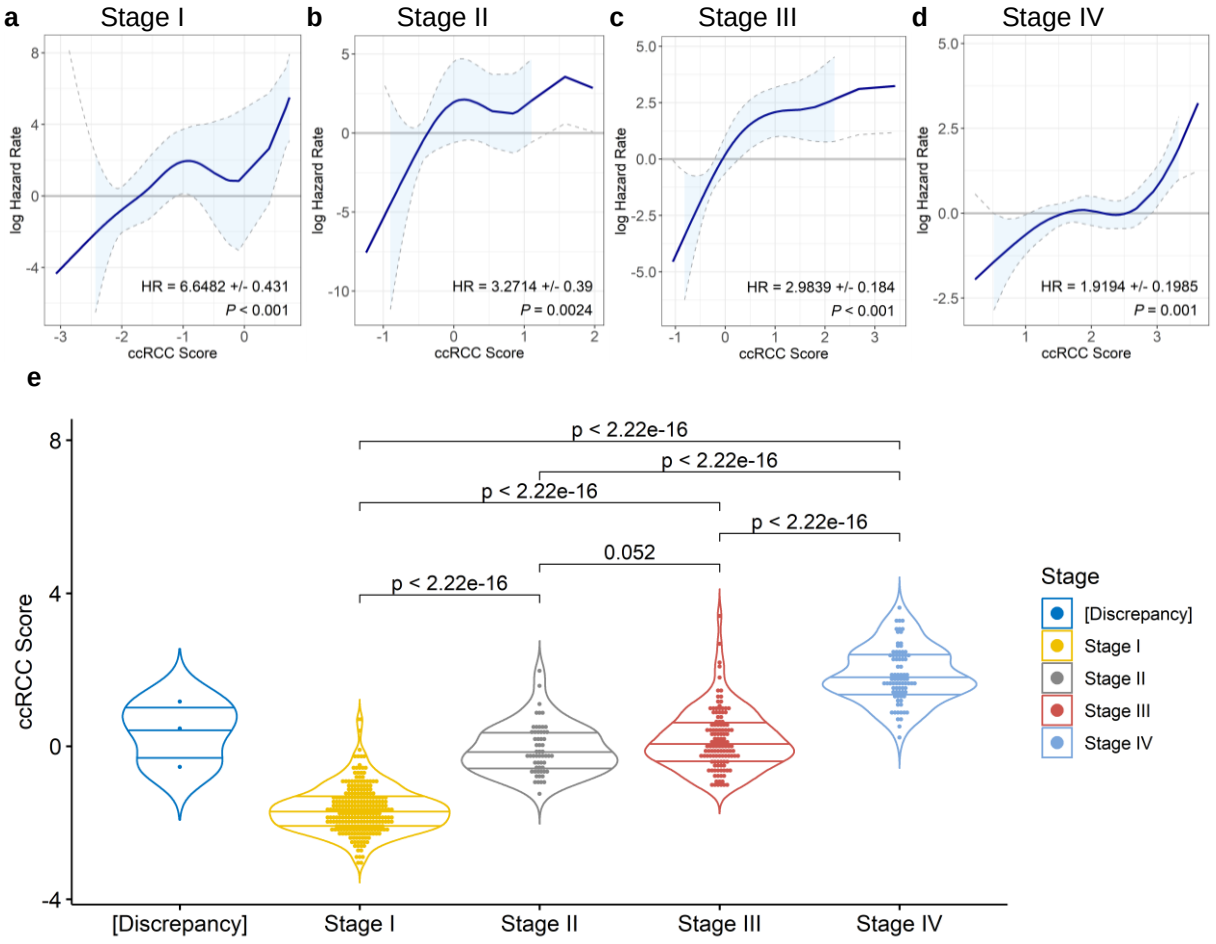


Figure 43 – Cox proportional log hazard smoothed by ccRCC Scores of patients stratified according to tumor stage.

(a-d) Patients with low tumor stage have generally lower ccRCC Scores and patients with high tumor stages have higher ccRCC Scores.

(e) ccRCC Score correlates with tumor stage (Kendall R^2 = 0.71, P < 0.0001). Mean ccRCC Score of Stage I patients (-1.580 ± 0.663) is significantly different from Stage II (-0.104 ± 0.737), Stage III (0.218 ± 0.857) and Stage IV (1.900 ± 0.781), whereas the ccRCC Score of patients with stage III and stage IV tumors does not significantly differ. Horizontal lines in the violet plot indicate 0.25, 0.50 and 0.75 quantiles of the samples. *P* values are calculated using the two-samples Wilcoxon test.

Since ccRCC Score and tumor stage correlate with each other, we were wondering whether the score remained a significant predictor for patients' cancer specific survival in a multivariate analysis with all patients and with all complete annotated clinical covariates.

Only necrosis status proved to be another significant and thus independent predictor of cancer specific survival. (**Figure 44**). That means when comparing to non-necrotic patients, patients with a histopathologic presence of necrotic areas in ccRCC specimens have an increased risk of a cancer specific death by 3.59 and with every rise of this ccRCC Score by 1 a 2.34 times increased risk of cancer specific death.

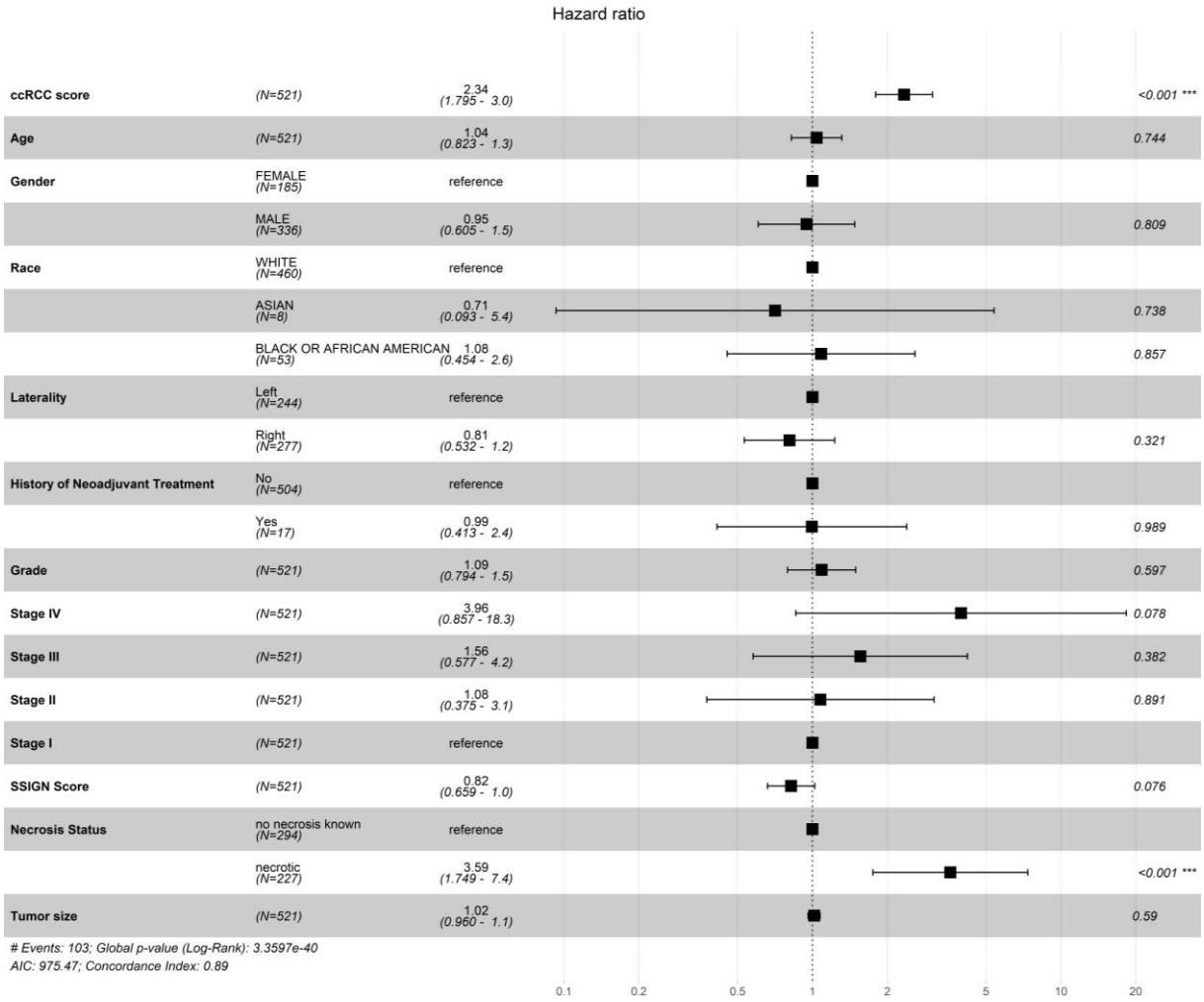


Figure 44 – Forest Plot of the multivariate analysis of clinical phenotypes and the ccRCC Score in the TCGA-KIRC dataset

All clinical parameters with > 90 % complete data were included in the multivariate analysis (Cox multiple regression as in **11**). Only the ccRCC Score and the clinical parameter "Necrosis Status" showed significant *P* values. 11 patients with missing information in any of the covariates where excluded from the analysis. * $P < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Error bars represent the 95 %-CI.

Survival estimators are generally displayed by the method of Kaplan-Meier, which was achieved by dividing the patients into either two or three groups (**Figure 45**). Two groups were defined by separating patients by the median ccRCC Score. For three patient groups the lower and upper quantile of the ccRCC Score were classified as low risk and high risk, whereas the 50 % of patients in between were classified as intermediate risk group.

The stratification of patients into both, two risk and three risk groups, significantly separates cancer specific survival probabilities. The high risk group of the binary stratification has only 73 months median CSS probability, whereas at the end of the observation period after more than 12 years the survival probability of the low risk patients was still 90.9 % (**Figure 45a**). By stratifying patients into three risk groups, patients of the high risk group have a median survival of 34 months and intermediate risk patients have still an 86.2 % cancer specific survival probability after 12 years follow up. In the low risk patient group only one patient dies during the observation period and the cancer specific survival probability is still 94.4 % (**Figure 45b**).

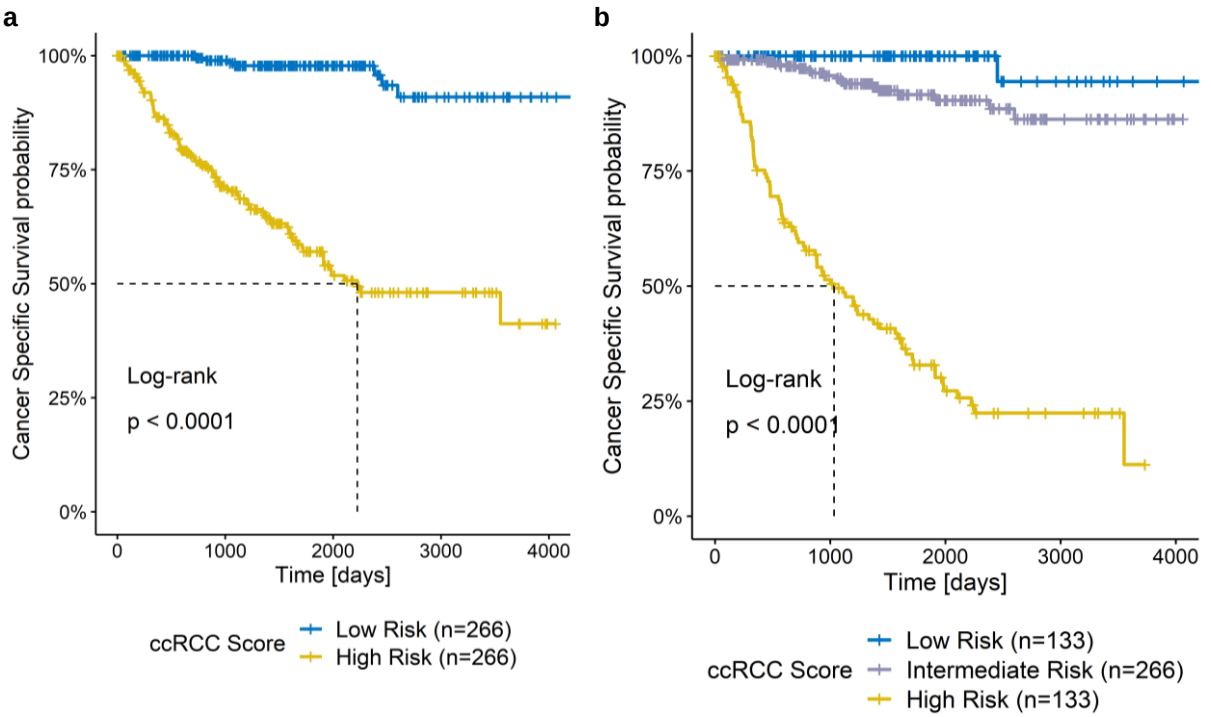


Figure 45 – The ccRCC Score efficiently stratifies patients into risk groups

(a) Kaplan-Meier estimator for cancer specific survival (CSS) of the TCGA-KIRC dataset for patients classified by median ccRCC Score (-0.751). Patients in the high risk group (n=266) have a median CSS of 2225 days (~73 month), whereas patients in the low risk group (n=266) have a CSS probability of 90.9 % after 12 years of survival. **(b)** For an ordinal ccRCC Score stratification, patients were grouped into three groups. The high risk group (n=133; ccRCC Score > 0.449) has a median survival of 1033 days (~34 month), the intermediate (n=266) and low risk group (n=133; ccRCC Score < -1.700) have at the end of the follow up 86.2 % and 94.4 % CSS probability respectively. Dotted lines indicate median survival. Stratifications separate CSS probabilities significantly different (log rank test).

In a next step, patients were stratified according to both tumor stage and the ccRCC Score-derived three risk group classification. The ccRCC Scores used to allocate all patients of the dataset into risk groups in **Figure 45** were also applied to allocate patients into risk groups that were stratified according to tumor stage (**Figure 46a-d**). For each tumor stage the assigned risk groups can estimate cancer specific survival probabilities. Nevertheless, the number of patients per risk group is dependent on tumor stage (**Figure 46e**). In stage I only two patients were classified as high risk and 133 patients as low risk. In contrast, in stage II no patient was classified as low risk, while 11 of the 57 stage II patients in the dataset were at high risk with a median survival of 2254 days (73 months). No patient of stage III and IV was classified as low risk. The high risk group of stage III patients has a median survival of 1229 days and of stage IV patients only 721 days. Still, the intermediate risk group showed significantly improved cancer specific survival probability.

By grouping patients with specific ranges of ccRCC scores, the 5-year cancer specific survival rates can be calculated (**Figure 46f)**. Patients with low ccRCC Scores have a very good prognosis whereas high scores indicate worse 5-year cancer specific survivals.

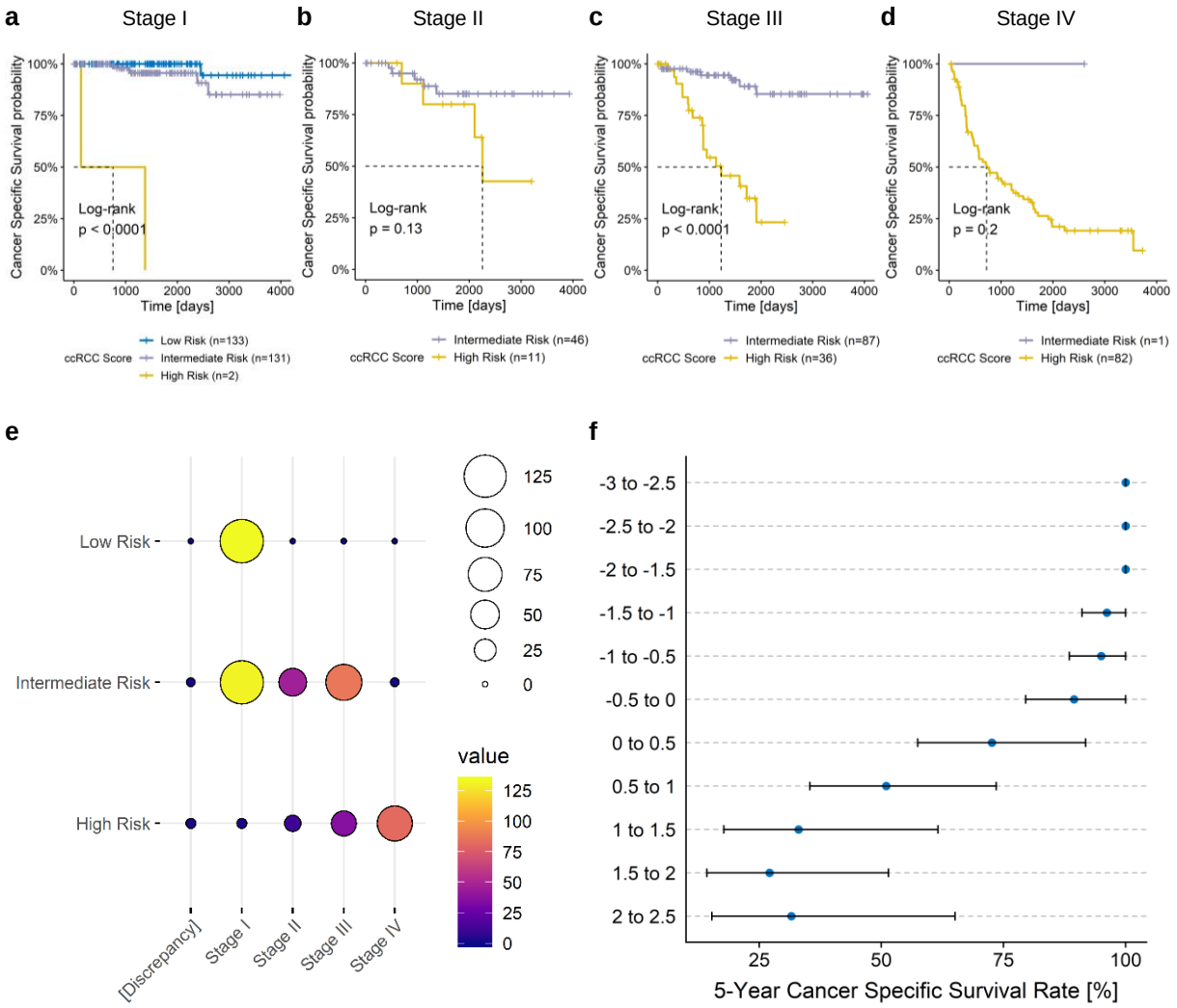


Figure 46 – Cancer specific survival probability of ccRCC patients stratified according to ccRCC Score and patient stage.

(a-d) Kaplan-Meier estimators for cancer specific survival with stratification of patients according to tumor stage and three risk groups using the same stratification boundaries for the ccRCC Score as applied for all patients Dotted lines indicate median survival. (e) Balloon-Plot displaying the number of patients per tumor stage and ccRCC risk group. **(f)** 5-year cancer specific survival of patients in ccRCC Score ranges. Patients were grouped according to a ccRCC Score range of -3 to 2.5 in 0.5 intervals. 5-year cancer specific survival rates with their respective 95 %-CI were calculated for each group.

3.8 The ccRCC Score has been validated in an independent patient cohort

The ccRCC score was validated using the independent RNA Sequencing dataset published by Sato *et al.*[31]. Median follow-up of these 100 patients containing validation cohort was 50 months (**Supplemental Figure 7**). RNASeq data was TNM normalized, voom transformed and standardized according to the TCGA-KIRC dataset to allow comparison and the ccRCC Score for each patient calculated (**Supplemental Figure 9**).

Also in the validation dataset the log hazard ratio significantly correlates with the ccRCC Score (**Figure 47**). The 95 % confidence interval is larger than the interval of the TCGA-KIRC dataset due to lower patient number.

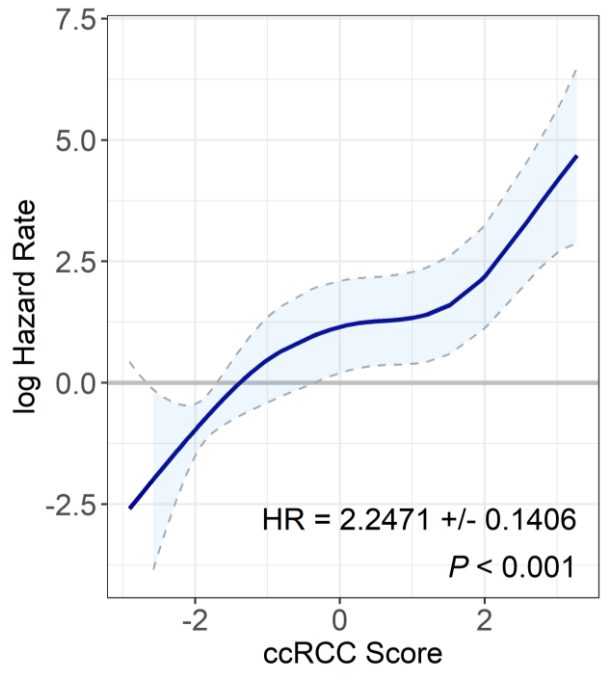


Figure 47 – Validation of the ccRCC Score in the independent ccRCC patient cohort of Sato *et al.*

Cox proportional log hazard smoothed by ccRCC Scores shows a significant correlation with a hazard ratio of 2.2471 ± 0.1406 ($P < 0.001$)

In the Sato *et al.* cohort, the ccRCC Score is the only significant prognostic factor in a multivariate analysis with annotated clinical covariates (**Figure 48**).

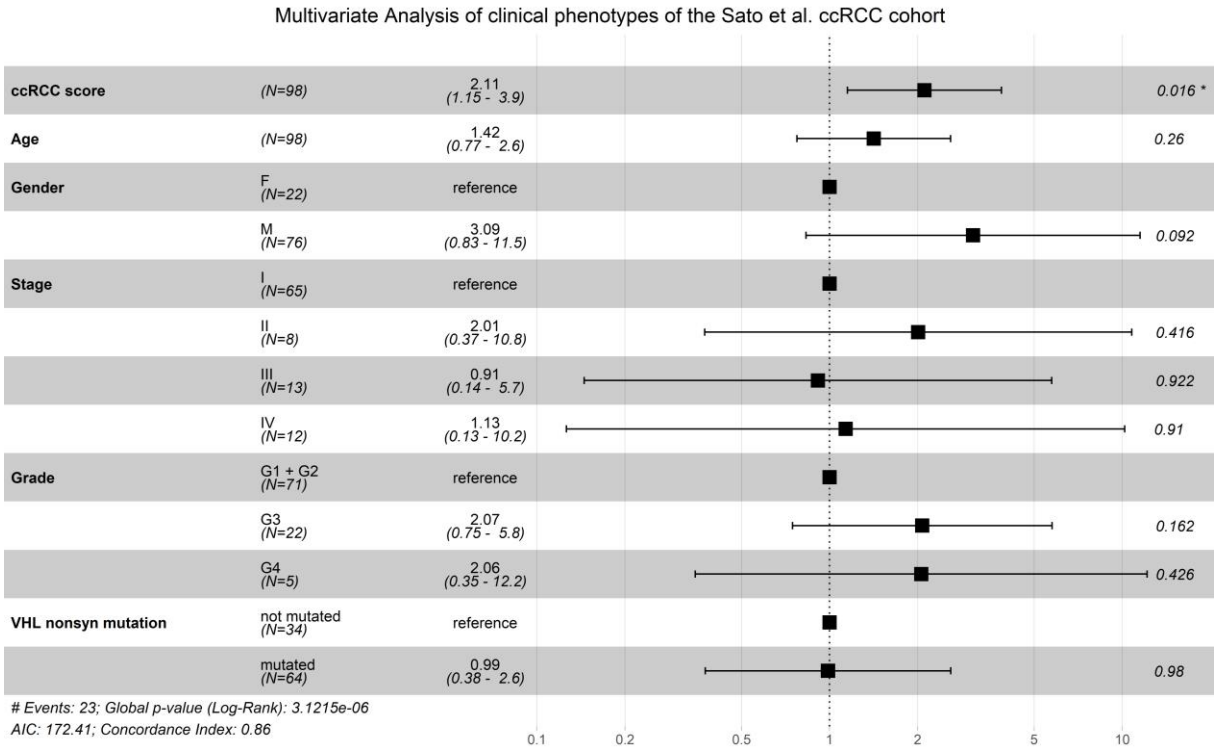


Figure 48 – Forest Plot of the multivariate analysis of clinical phenotypes and the ccRCC Score in the Sato *et al.* ccRCC dataset

For the multivariate analysis (Cox multiple regression as in **equation 11**) all clinical parameters with > 90 % complete data were included. Only the ccRCC Score exhibits significant *P* values of the regression. There was no patient with missing information in any of the covariates that had to be excluded from the analysis. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Error bars represent the 95 %-CI.

Subsequently, patients were divided into risk groups using the same ccRCC Scores that allocated patients into risk groups of the TCGA-KIRC dataset. In the validation dataset of Sato *et al.*, the ccRCC Score is highly predictive for patient survival in both binary and ordinal stratifications (log rank $P < 0.0001$). Median survival of the high risk group (n = 28) is 44 months regardless the stage in the binary model. The survival probability for patients in the low risk group (n = 72) that have survived for 145 months of follow-up is still 84.1 % (**Figure 49a**).

In an ordinal stratification approach, high risk patients (n = 18) have a median survival of only 33 months, whereas the intermediate risk group (n = 41) has a survival probability of 64.7 % at the end of the follow-up period. The low risk group (n = 41) has a survival probability of 92.2 % (**Figure 49b**).

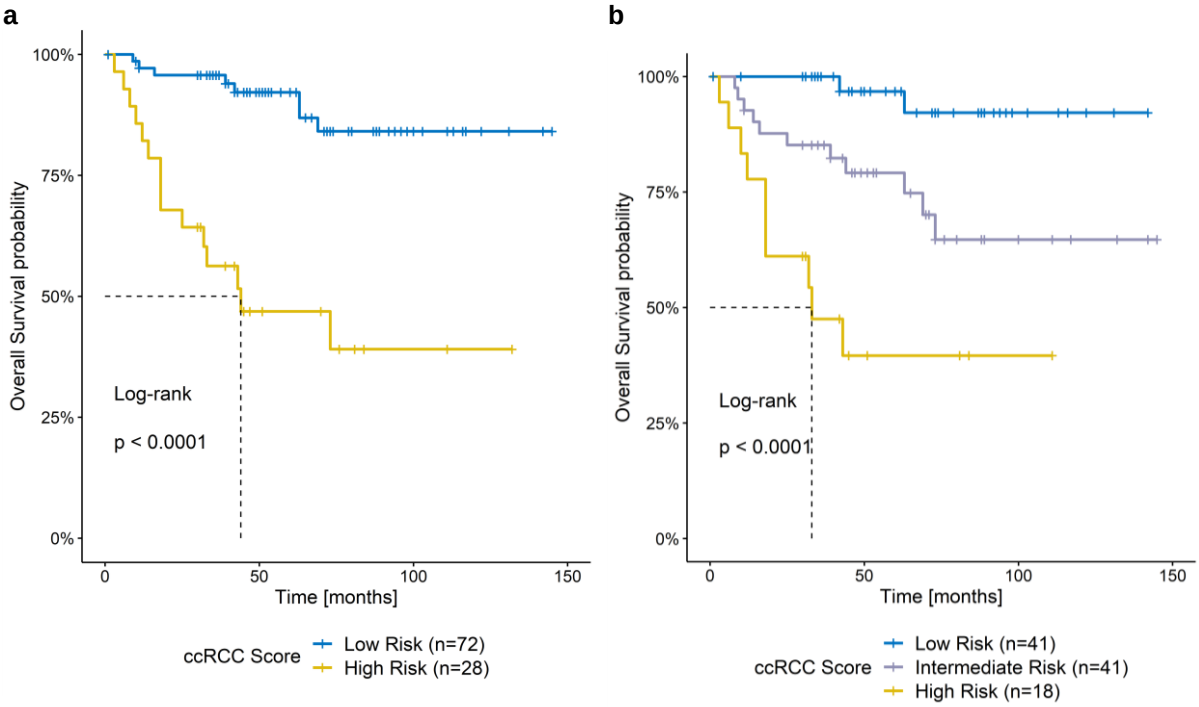


Figure 49 – Kaplan-Meier estimators for overall survival probability of ccRCC of the Sato *et al.* cohort stratified according to the ccRCC Score.

(a) The same ccRCC Score (-0.759) as in the TCGA-KIRC dataset was applied to divide patients into high risk (n = 28) and low risk (n = 72) groups. Median survival of the high risk group was 44 months and the low risk group had an overall survival of 84.1 % after the follow up period. **(b)** Stratification into three risk groups according the TCGA-KIRC cohort with ccRCC Scores of -1.700 and 0.449 resulted in 18 patients in the high risk group with a median survival of 33 months and 41 patients with intermediate risk and 64.7 % survival probability at the end of the observation period. The 41 low risk patients had 92.2 % survival probability after 145 months. Dotted lines indicate median survival. Stratifications separate patient survival probabilities significantly different (log rank test).

3.9 The ccRCC Score has superior predictive power

To assess the power of the ccRCC Score to predict cancer specific survival (CSS), it was compared in the two independent cohorts to the predictive power of other published signatures for the ccRCC that are based on gene expression data.

In 2014, a systematic approach by Gulati *et al.*[267] identified six prognostic signatures that are based on gene expression analysis. They evaluated the signatures by comparing their prognostic power to predict CSS of the TCGA-KIRC dataset. At the timepoint of the study, only data from 350 patients of the dataset were published. A multiregion biopsy data set of 10 ccRCCs was additionally analyzed to assess intratumor heterogeneity. However, a true validation on an independent large patient cohort was not performed.

Expression data of 21 of the 23 genes of the Cluster A vs. C and 41 of the 48 genes of the Cluster B vs. A/C gene signature of Beleut *et al.*[276], 144 of the 157 genes of the gene signature of Boström *et al.*[277], 34 of the 34 genes of the ClearCode34 gene signature from Brooks *et al.*[278], which was derived from the ccA/ccB signature from Brannon *et al.*[279], 28 of the 35 genes of the gene signature of Kosari *et al.*[280], 40 of 44 genes of the gene signature of Lane *et al.*[281] and 224 of the 259 genes of the gene signature of Zhao *et al.*[282] were annotated in the TCGA KIRC and Sato RNA-seq data sets.

Interestingly, in none of the tested gene signatures any gene of the ccRCC Score is included.

At the time of this thesis, the TCGA KIRC dataset comprises of 532 patients. These patients and the validation cohort of 100 patients of the Sato *et al.* dataset were assigned to their respective subgroups by unsupervised NMF consensus clustering as proposed by Gulati *et al.*[267] (**Supplemental Figure 14 & Supplemental Figure 15**). Exemplary the clustering of patients is presented for the ClearCode34 signature in both datasets in **Figure 50**.

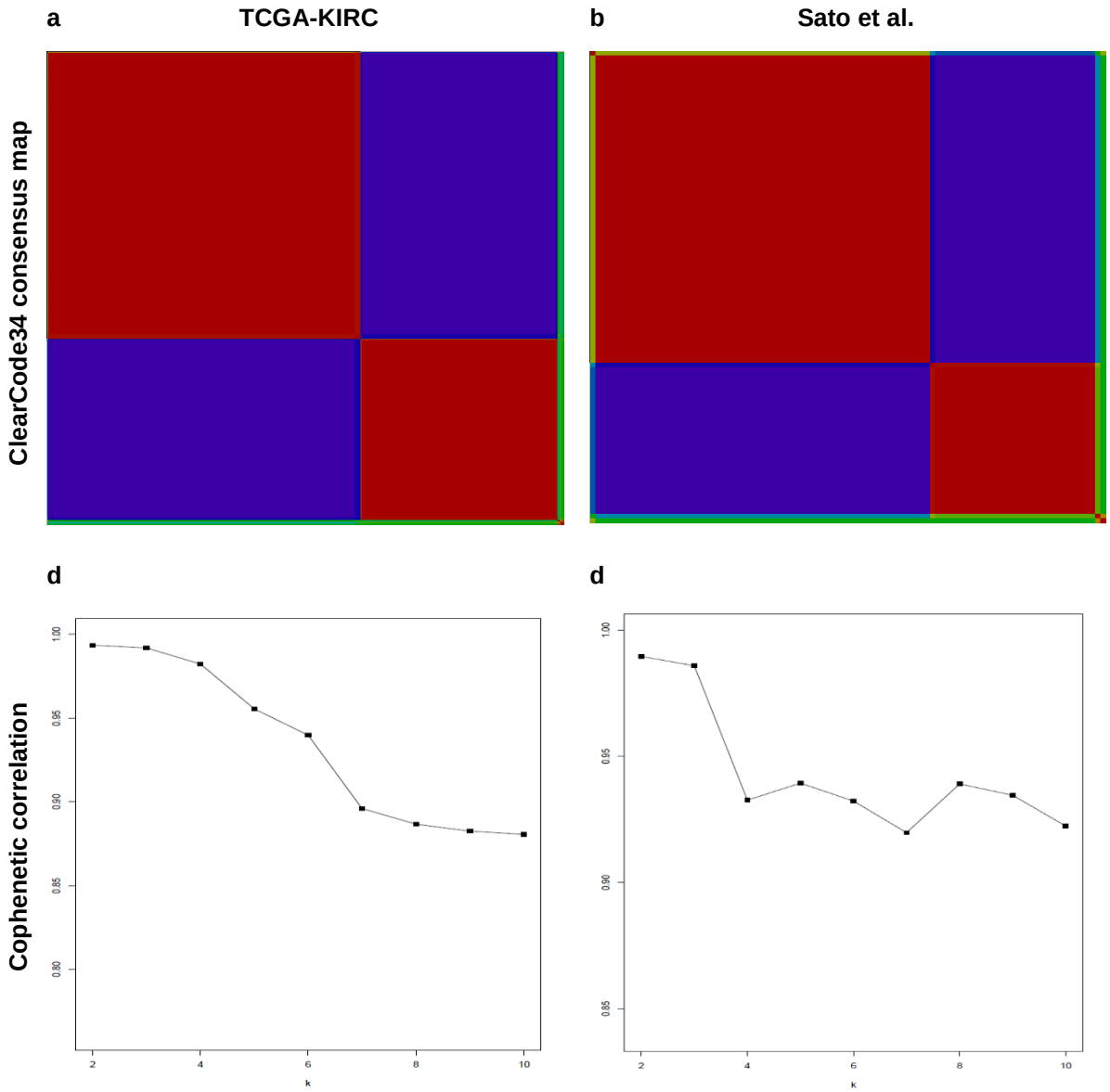


Figure 50 – NMF Consensus Clustering for TCGA-KIRC and Sato *et al.* patients according to the ClearCode34 signature

Ordered consensus NMF clustering maps for k = 2 for **(a)** TCGA-KIRC and **(b)** Sato *et al.* patients.
Cophenetic correlation plots generated for the **(c)** TCGA-KIRC and **(d)** Sato *et al.* datasets were used to identify the best cluster representation by NMF Consensus clustering. In both datasets, two clusters represent the best separation of patients according to the gene expression data of the tested genes, as published for the ClearCode34 signature.
All NMF consensus plots and reports were generated using the cloud.genepattern.org plattform by the Broad Institute[283].

To evaluate whether the ccRCC score correlates with any of the other signatures or with ordinal and continuous clinical parameters, Kendall rank correlation coefficients were calculated (tau test)[284] (**Figure 51 & Figure 52**).

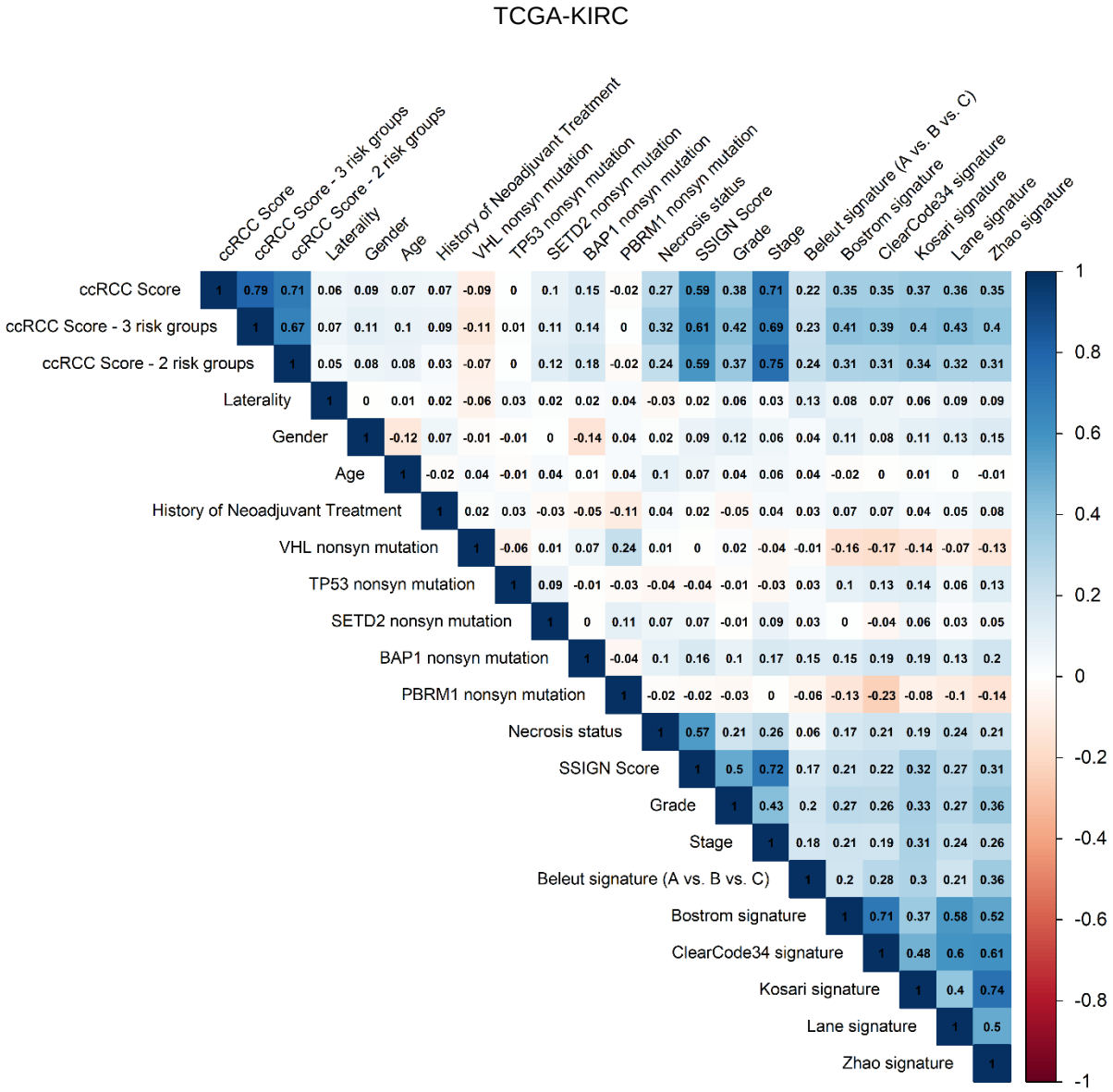


Figure 51 – Kendall rank correlation coefficients for clinical variables, ccRCC Score and published signatures in the TCGA-KIRC ccRCC dataset.

With the help of the tau test, statistical dependencies of ordered covariates such as tumor stage and grade could be estimated. Two and three risk groups of the ccRCC Score and the linear ccRCC Score were compared to both clinical variables and other published gene expression signatures for ccRCC. Tumor stage shows a high correlation and Fuhrman grade a low correlation with the score in the TCGA-KIRC patient datasets. Other signatures correlated only with low correlation to the ccRCC Score, whereas for example Kosari's and Zhao's correlate highly with a correlation coefficient of 0.74.

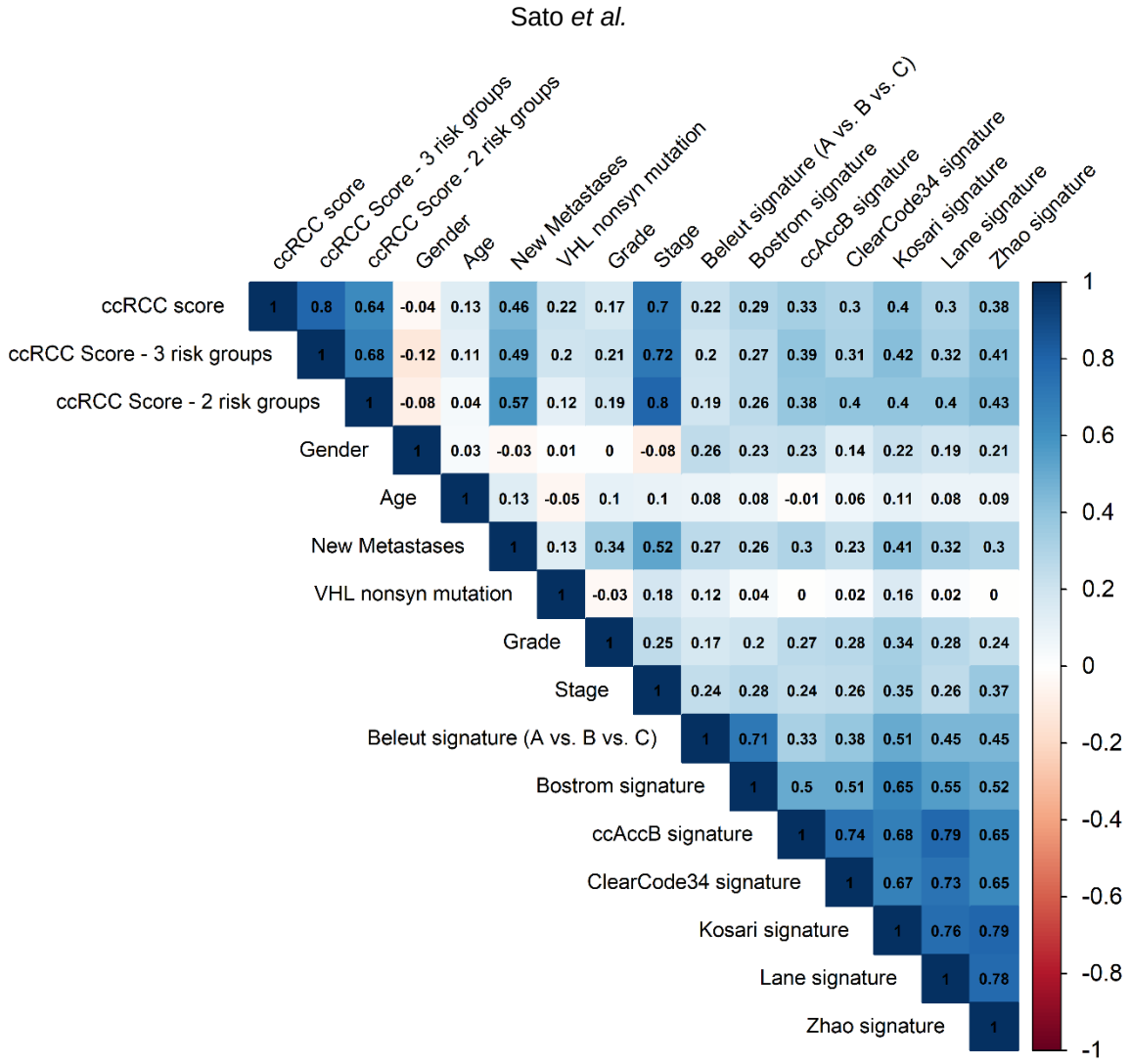


Figure 52 – Kendall rank correlation coefficients for clinical variables, ccRCC Score and published signatures in the Sato *et al.* ccRCC dataset.

As in **Figure 51**, tumor stage shows a high correlation and Fuhrman grade a low correlation with the ccRCC score in both patient datasets. The other published gene signatures correlate only with low correlation to the ccRCC Score, whereas for example Kosari's and Zhao's correlate highly with a correlation coefficient of 0.78. In general, the published signatures correlate much stronger with each other in the Sato *et al.* dataset.

As expected, the ccRCC Score correlates with tumor stage in both datasets, as this clinical attribute was included as covariate in the model. Fuhrman grade has a slight correlation of 0.38 in the TCGA-KIRC dataset (**Figure 51**) but only a correlation of 0.17 in the Sato *et al.* dataset (**Figure 52**). Tumor laterality, gender, history of neoadjuvant treatment and VHL mutations do not correlate with the ccRCC Score. To the other signatures only minor correlations of 0.22 to 0.43 in TCGA-KIRC and 0.19 to 0.43, depending on risk stratification, are recognizable. The other published signatures however correlate highly with each other, for example the signature of Zhao and Kosari with a correlation of 0.74 and 0.79. Interestingly, the annotation of new metastases in the Sato *et al.* dataset shows quite some association to the ccRCC Score.

Kaplan-Meier estimators of the published signatures depict that all analyzed signatures can significantly discriminate cancer specific survival in the TCGA-KIRC patient cohort, with $P < 0.001$ (log rank), confirming their results in the most recent TCGA-KIRC cohort (**Figure 54)**.

Patients in Beleut's cluster A had significantly better prognosis than cluster B and C, whereas between cluster B and C no significant survival differences could be identified. TGF-β score high classified patients of the Boström signature show worse survival. The ccB cluster of the ClearCode34 signature (the simplified ccA/ccB signature) has lower survival probabilities. CSS of the aggressive subgroup, defined by Kosari, was worse than the non-aggressive subgroup. CSS of the aggressive subgroup, defined by Lane, was worse than the indolent subgroup of patients. Patients with a high TGFβ score had worse survival than patients with a low score. Patients in Zhao's poor prognosis cluster 2 had worse CSS than cluster 1, supporting the results of Gulati *et al.*[267].

Testing the signatures in the dataset of Sato *et al.* shows that the six published signatures can also significantly stratify ccRCC patients into prognostic groups **(Figure 54**).

The predicted hazard ratios of the more aggressive subgroup in comparison to the benign group of the various gene signatures are summarized for both the TCGA-KIRC dataset as well as the Sato *et al.* patient cohort in (**Table 7**). For completeness, I additionally stated the hazard ratios for these datasets that were calculated by Gulati *et al.* on a previous version of the TCGA-KIRC dataset with only 350 patients[267]. Unfortunately, the authors of this study did not publish their patient classifications for comparison.

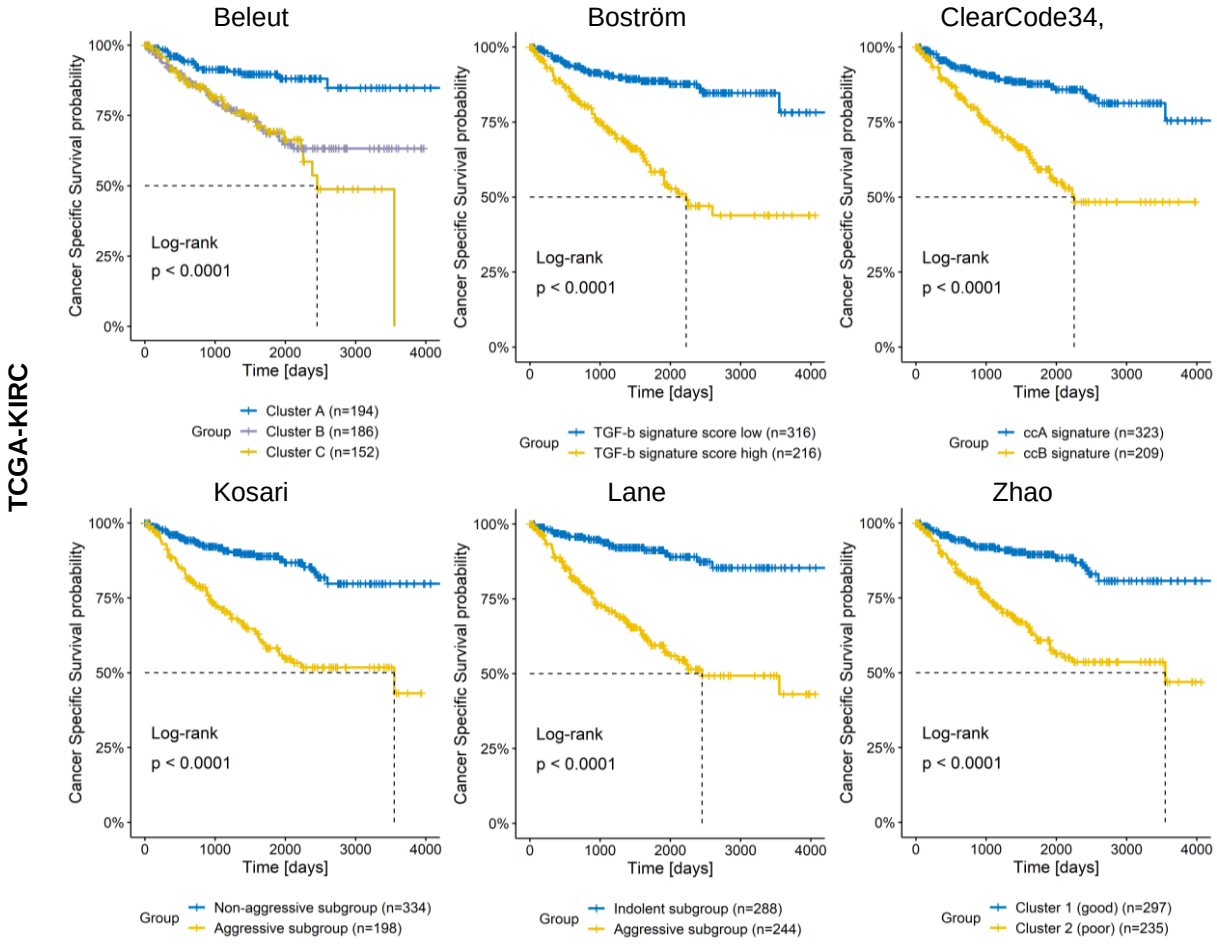


Figure 53 – Published gene expression signatures show different power in predicting ccRCC cancer specific survival in the TCGA-KIRC cohort

Kaplan-Meier estimators of cancer specific survival for ccRCC patient groups of the TCGA-KIRC dataset identified by published gene expression signatures. All signatures stratify patients into significantly different groups with increased and decreased survival probabilities (logrank test).
Dotted lines indicate median survival time.

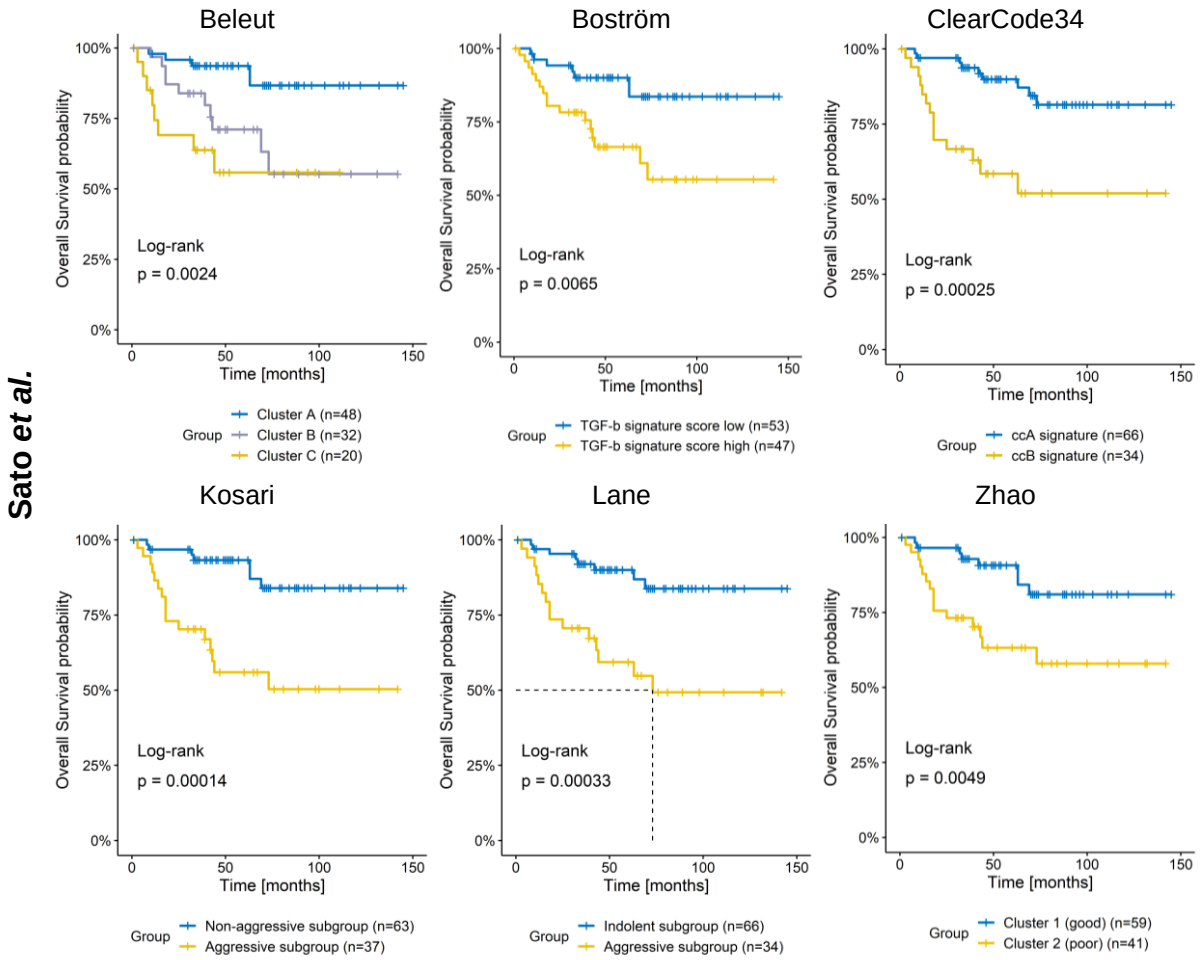


Figure 54 – Published gene expression signatures show different power in predicting ccRCC overall survival in the Sato *et al.* cohort

Kaplan-Meier estimators of overall survival for ccRCC patient groups of the Sato *et al.* dataset identified by published gene expression signatures. All signatures stratify patients into significantly different groups with increased and decreased survival probabilities (logrank test).
Dotted lines indicate median survival time.

Table 7 – Univariate hazard ratios of ccRCC patients classified by published signatures and the ccRCC Score

Study	HR [TCGA-KIRC] (95 % CI)	No. Of cases (n = 532) (%)	*P* value (log rank test)	HR [TCGA - Gulati *et al.*] (95 % CI)	No. Of cases (n = 350 (%)	*P* value	HR [Sato *et al.*] (95 % CI)	No. Of cases (n = 100) (%)	*P* value (log rank test)
Beleut signature									
Cluster A	1.00 (Ref)	194 (36)		1.00 (Ref)	127 (36)		1.00 (Ref)	48 (48)	
Cluster B	2.87 (1.67-4.94)	186 (35)		2.27 (1.31-3.96)	175 (50)		3.71 (1.27-10.90)	32 (32)	
Cluster C	3.23 (1.84-5.65)	152 (29)	< 0.001	2.30 (1.13-4.66)	48 (14)	0.005	5.92 (1.93-18.23)	20 (20)	0.003
Boström signature:									
TGF-b signature low	1.00 (Ref)	316 (59)		1.00 (Ref)	175 (50)		1.00 (Ref)	53 (53)	
TGF-b signature high	3.97 (2.61-6.03)	216 (41)	< 0.001	1.98 (1.23-3.16)	175 (50)	0.003	3.22 (1.32-7.86)	47 (47)	0.007
ClearCode34 signature:									
ccA	1.00 (Ref)	323 (61)		1.00 (Ref)	240 (69)		1.00 (Ref)	66 (66)	
ccB	3.3 (2.2-5.0)	209 (39)	< 0.001	4.90 (3.09-7.76)	110 (31)	< 0.001	4.30 (1.85-10.02)	34 (34)	< 0.001
Kosari signature:									
Nonaggressive subgroup	1.00 (Ref)	334 (63)		1.00 (Ref)	242 (69)		1.00 (Ref)	63 (63)	
Aggressive subgroup	3.66 (2.43-5.51)	198 (37)	< 0.001	2.85 (1.84-4.43)	108 (31)	< 0.001	4.80 (1.97-11.70)	37 (37)	< 0.001
Lane signature:									
Indolent	1.00 (Ref)	288 (54)		1.00 (Ref)	219 (63)		1.00 (Ref)	66 (66)	
Aggressive subgroup	4.96 (3.12-7.88)	244 (46)	< 0.001	4.21 (2.62-6.77)	131 (37)	< 0.001	4.25 (1.80-10.03)	34 (34)	< 0.001
Zhao signature:									
Cluster 1 (good)	1.00 (Ref)	297 (56)		1.00 (Ref)	269 (77)		1.00 (Ref)	59 (59)	
Cluster 2 (poor)	3.61 (2.35-5.54)	235 (44)	< 0.001	5.26 (3.37-8.22)	81 (23)	< 0.001	3.21 (1.36-7.59)	41 (41)	0.005
ccRCC Score (3 groups):									
low risk	1.00 (Ref)	133 (25)					1.00 (Ref)	41 (41)	
intermediate risk	8.98 (1.20-67.30)	266 (50)					6.06 (1.34-27.33)	41 (41)	
High risk	113.97 (15.87-818.60)	133 (25)	< 0.001				18.21 (3.96-83.75)	18 (18)	< 0.001
ccRCC Score (2 groups):									
low risk	1.00 (Ref)	266 (50)					1.00 (Ref)	72 (72)	
high risk	16.51 (7.66-35.56)	266 (50)	< 0.001				6.55 (2.77-15.53)	28 (28)	< 0.001
ccRCC Score (linear):	2.72 (2.35-3.15)	532 (100)	< 0.001				2.25 (1.71-2.96)	100 (100)	< 0.001

The *C* index, also called concordance index or "Harrel's *C*", is used to validate the predictive ability of censored data such as in a Cox Regression model[260]. It was calculated for both published signatures and the ccRCC Score (**Figure 55a, b**). In both datasets, the ccRCC Score had the highest *C* index. In contrast to the continuous ccRCC Score which goes along with a high predictive power (Harrel's *C* Index of 0.88), stratification of patients lowers the predictive power (Harrel's *C* Index of ≤ 0.84). It is notable that none of the published signatures could reach the predictive power of the classical clinical annotation stage.

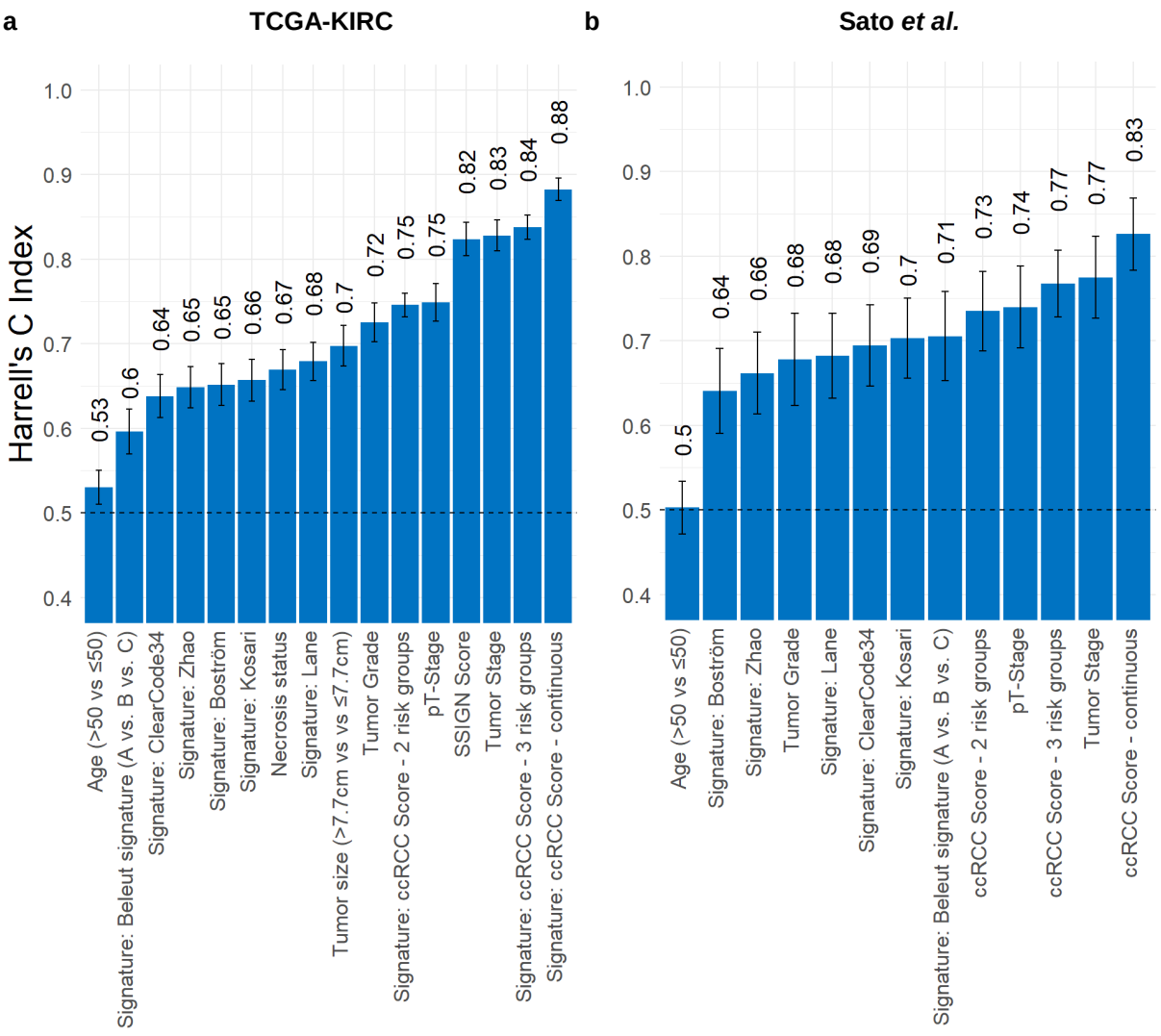


Figure 55 – Comparative analysis of the predictive power of clinical phenotypes, published gene signatures and the ccRCC Score

The Harrel's C Index was calculated in order to compare the fits of the gene signatures Cox regression models. In both datasets, the ccRCC Score had superior survival predictive power. All published signatures based on gene expression data have worse predictive power than most of the clinical characteristics such as tumor grade or SSIGN Score. Error bars depict C-Index ± standard error.

Another common metric to compare survival predictions is the likelihood ratio test[285]. When comparing the scores, the continuous ccRCC Score has superior predictive power above the other variables (**Figure 56**).

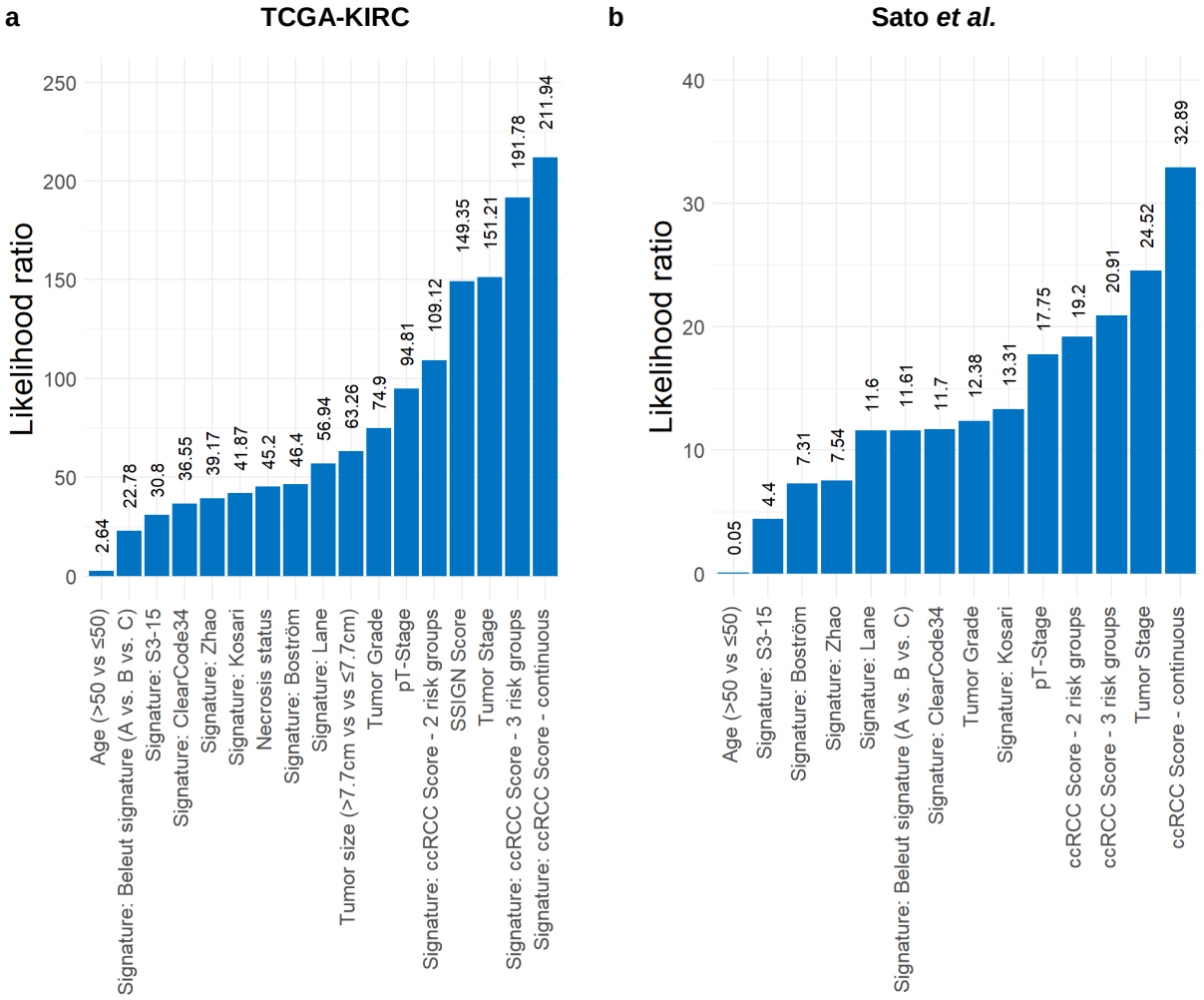


Figure 56 – Likelihood ratios of clinical phenotypes, published gene signatures and the ccRCC Score to predict patients survival

The Likelihood ratio was calculated in order to compare the predictive power of the different gene signatures and clinical parameters. In both datasets, the ccRCC Score had superior survival predictive power.

3.10 The Heidelberg ccRCC mini cohort

We collected follow up data of those patients, which were used to establish ccRCC tumor models (**Supplemental Figure 10**). Microarray gene expression profiling was performed by Teresa Rigo-Watermeier[1] and the data imported and normalized as described before (**see section 3.5.2**). Subsequently, gene expression data of the ccRCC score genes was standardized using the scaling factors used for the TCGA-KIRC data standardization (**see Supplemental Figure 9**).

Our follow up was incomplete to estimate cancer specific survival, yet sufficient to estimate progression free survival. We calculated the ccRCC score for each patient, classified them according to their score into "low", "intermediate" and "high" risk groups (**Supplemental Figure 11**) and calculated Kaplan-Meier estimators (**Figure 57a**). No patient was classified as low risk, four patients were of intermediate and nine patients of high risk. In addition, in this small cohort, the ccRCC score significantly separated patients' progression free survival into two distinct groups.

One could argue that the survival difference of the small cohort was due to the presence of metastasis at the time point of tumor resection. However, when calculating Kaplan-Meier estimators for either metastatic or non-metastatic patients, the separation did not show a significant difference between the two groups (**Figure 57b**). Two early relapsing patients with non-metastatic tumors were classified by the ccRCC score as high risk patients. Therefore, the score had improved the risk classification of our patient cohort.

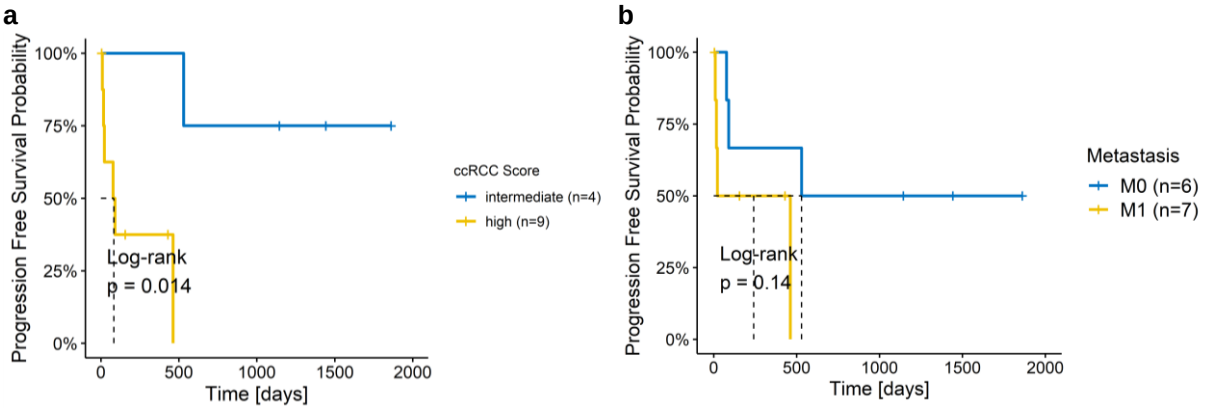


Figure 57 – Progression free survival of the Heidelberg ccRCC mini cohort

(a) Stratification of patients of the Heidelberg ccRCC cohort into three risk groups according the TCGA-KIRC cohort with ccRCC Scores of -1.700 and 0.449 resulted in 9 patients with high risk and 4 patients with intermediate risk. Stratifications separate progression free survival probabilities significantly different (logrank test). (b) Stratification of the patients according to their metastasis status did not significantly separate progression free survival probabilities.

3.11 Distinct subclonal regions have higher ccRCC scores

Gerlinger *et al.* described the vast intratumor heterogeneity of ccRCC by multiregion sequencing (**see section 1.2.2**)[72,73]. In their studies of 2012 and 2014, most regions of ten analyzed patients were additionally profiled for gene expression using microarrays. This allowed us to calculate the ccRCC score for most of the identified subclones. We stratified the different regions of the tumors into the 3-tiered ccRCC Score risk categories, presented in **section 3.7**. Gulati *et al.* applied this approach exemplarily for the ccAccB signature[267]. We likewise visually delineated the subclonal regions with identical ccRCC Score categories on the phylogenetic trees describing tumor development (**Figure 58**). In eight of ten patients, all regions of a tumor were classified into a single risk category. Two patients (EV001 and EV006) had tumor regions of both intermediate and high risk (**Supplemental Figure 12**).

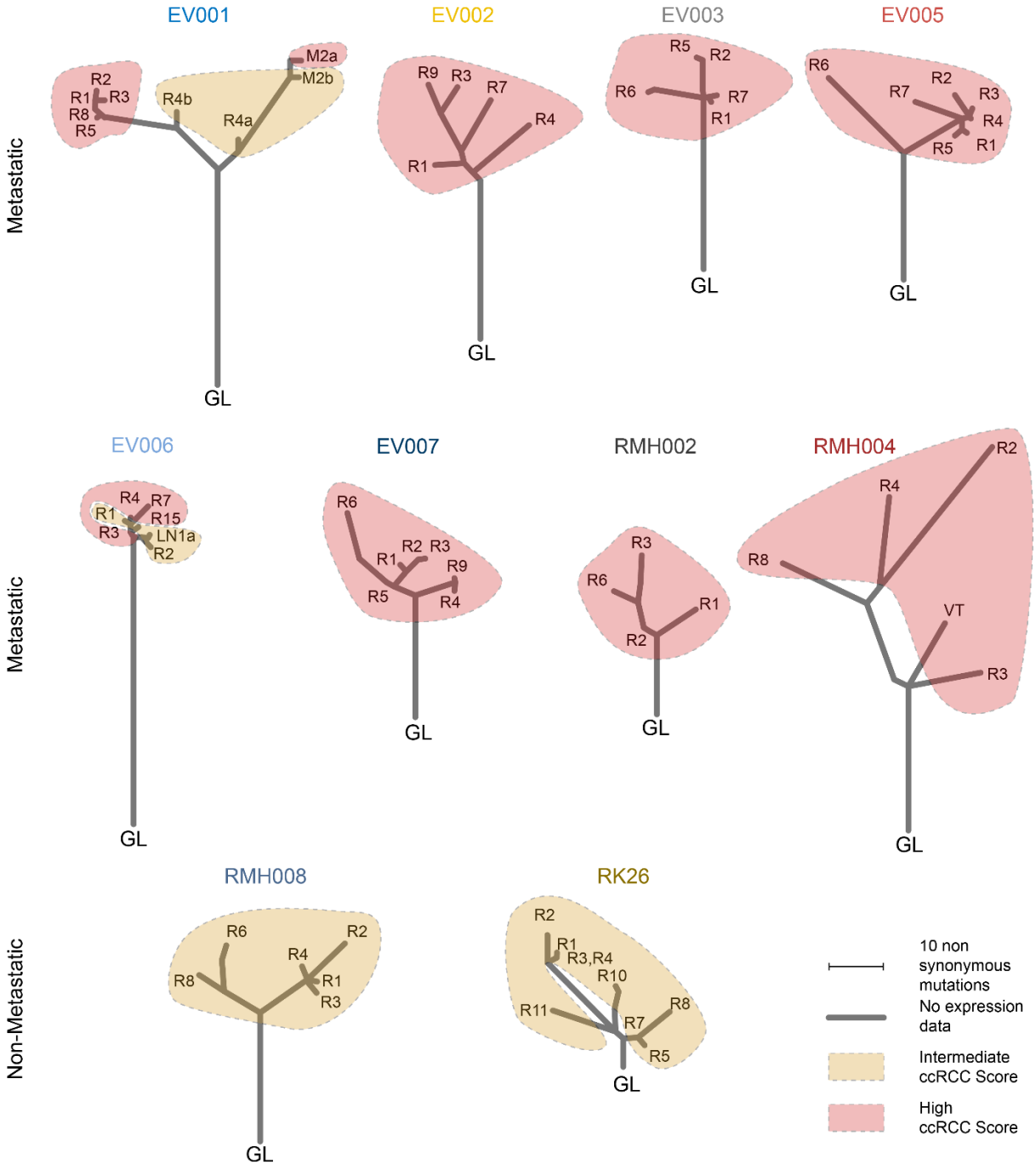


Figure 58 – Intratumor heterogeneity of ccRCC risk scores categories

Regions within ccRCC tumors were colored according to their ccRCC Score into intermediate risk (yellow; ccRCC Score < 0.449) or high risk (red; ccRCC Score > 0.449) categories. Metastatic patients had stage IV tumors and non-metastatic patients had stage II tumors. GL = Germline. R = primary tumor region, M = metastatic region, VT = venous thrombus. Clones without gene expression data described by original publication were left out.

Subsequently, the phylogenetic trees were recolored according to the ccRCC Score to illustrate changes of the ccRCC Score along tumor evolution (**Figure 59**).

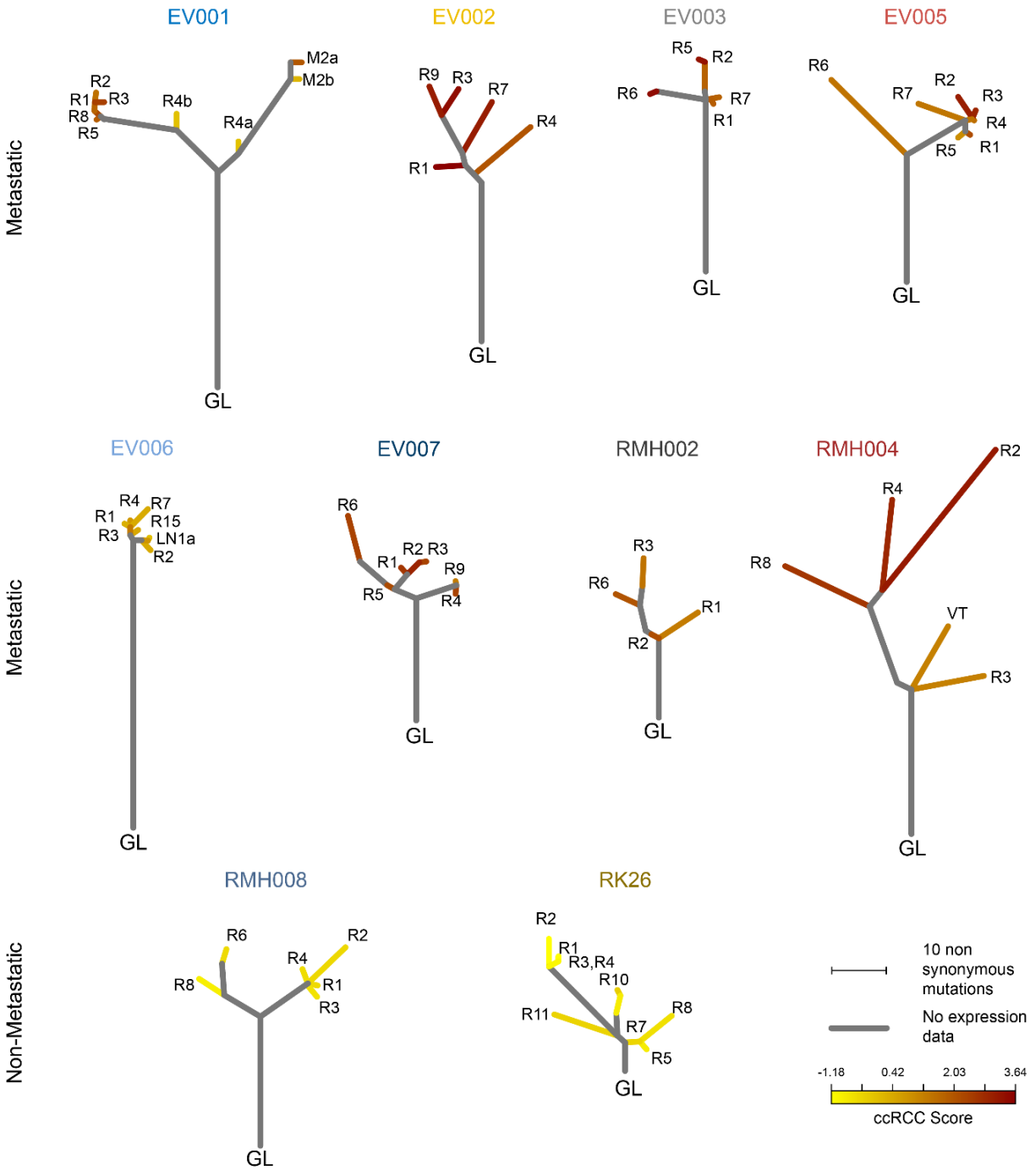


Figure 59 – Intratumor heterogeneity of ccRCC Scores

The phylogenetic trees that describe the tumor evolution of each patient were recolored according to their ccRCC Scores. Metastatic patients had stage IV tumors and non-metastatic patients had stage II tumors. GL = Germline. R = primary tumor region, M = metastatic region, VT = venous thrombus. Subclones without gene expression data described by original publication were left out.

This analysis revealed a vast intratumor heterogeneity of the ccRCC scores. As expected, the non-metastatic patients showed lower maximal subclonal ccRCC scores (RMH008 = -0.43, RK26 = -0.26) accompanied by a lower range of differences in scores between the subclones (RMH008 = 0.43, RK26 = 0.92) (**Table 8**). Metastatic patients had higher maximum ccRCC scores (up to 3.64 in patient EV002) as well as larger differences between the subclonal ccRCC scores (up to 2.38 in EV001 and 2.12 in EV005). This means that the predicted 5-year survival rate varies substantially when comparing the subclonal regions of a single tumor. For example, if only region R4a of the patient EV001 would have been sampled for gene expression analysis, a 5-year survival probability of 82.3 % would have been estimated by calculating the ccRCC score. However, the phylogenetic more distant subclone R3 has a predicted 5-year survival probability of only 12.2 %.

Table 8 – Overview of the multiregion analysis of ten ccRCC patients

Patient	ccRCC Score range	Max. ccRCC Score difference	Predicted 5-year survival (min > max ccRCC score in percent)
EV001	0.01 → 2.389	2.379	82.3 → 12.2
EV002	2.012 → 3.642	1.630	23.6 → 0.1
EV003	1.763 → 3.441	1.678	32.4 → 0.2
EV005	0.943 → 3.061	2.118	60.9 → 1.6
EV006	0.066 → 1.368	1.302	81.4 → 46.9
EV007	1.288 → 2.919	1.631	49.6 → 2.8
RK26	-1.184 → -0.264	0.920	94.3 → 86.2
RMH002	1.299 → 2.146	0.847	49.3 → 19.2
RMH004	1.204 → 3.135	1.931	52.5 → 1.2
RMH008	-0.856 → -0.426	0.430	92.1 → 88.1

We wondered whether the ccRCC Score showed a linear relationship to the number of non-synonymous mutations. Therefore, the two parameters were plotted against each other and the corresponding linear models were calculated (**Figure 60**). We found that for six patients an increasing mutation number corresponded to an increasing ccRCC score. One patient (EV002) showed no such correlation and three patients had a negative correlation (RMH002, EV006 & RK26).

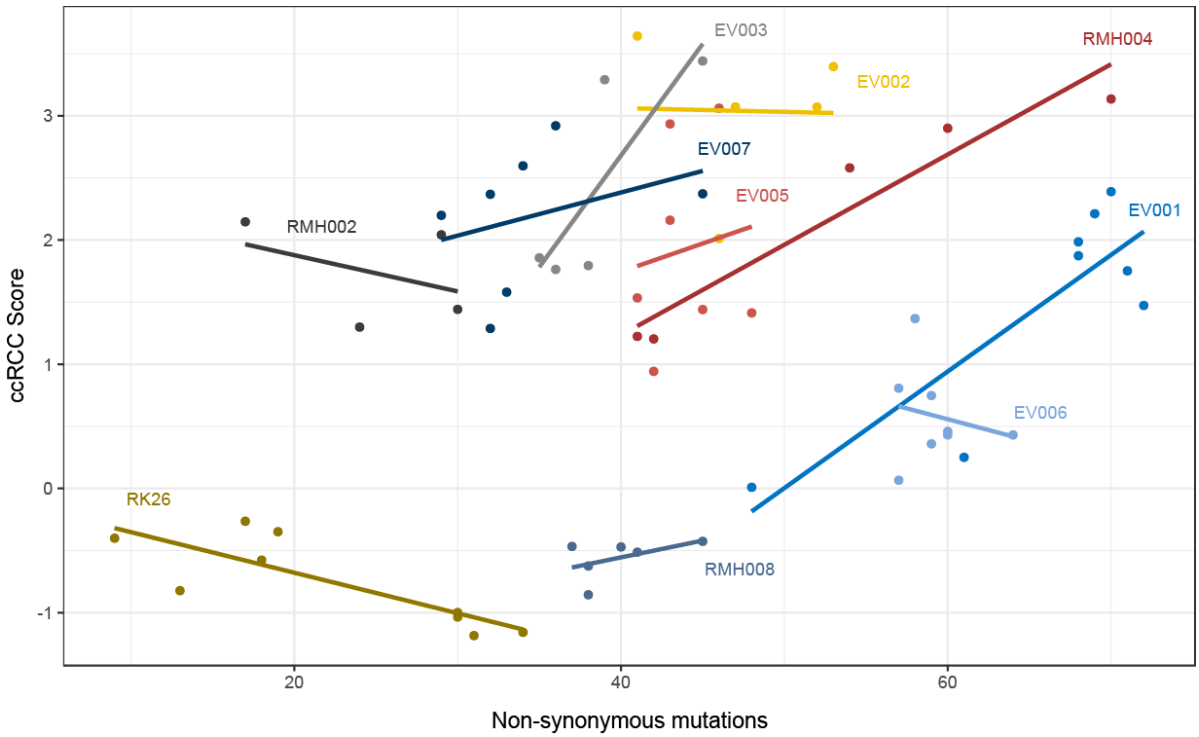


Figure 60 – Relationship between the number of non-synonymous mutations (evolutionary distance) originating from the germline and the ccRCC Score

For 5 patients, a positive linear relationship between the number of non-synonymous mutations and the ccRCC Score was identified, whereas for 2 patients the relationship was negative.

To identify a possible correlation between ccRCC Score and mutational load, we plotted the number of non-synonymous mutations of the bulk tumors of patients of the TCGA dataset against their ccRCC Score (**Figure 61**). We could not identify a significant correlation between the number of mutations and the ccRCC Score.

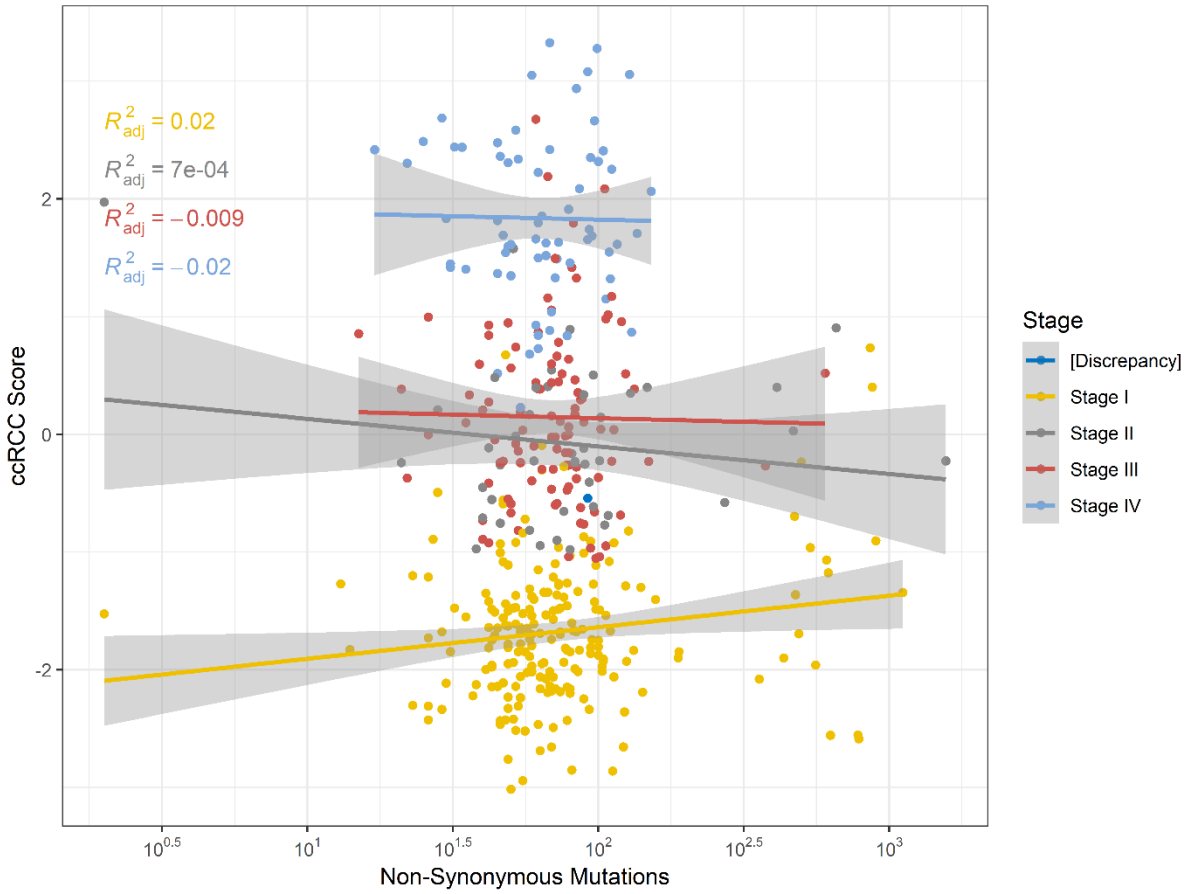


Figure 61 – Dotplot of non-synonymous mutations of the TCGA patient cohort and their corresponding ccRCC Score, subsetted according to patient stage

The number of non-synonymous mutations does not correlate with the ccRCC Score. The number of non-synonymous mutations was assessed by summing up all annotated mutations of cBioportal for each patient.

3.12 The ccRCC-Score gives added value above the classical TNM-stage classification

To estimate the value the ccRCC-Score adds to patients' risk stratification over stage classification, we tested whether patients within a specific stage classification could be identified. For this purpose, we applied an individual risk stratification cutpoint for patients within a tumor stage according to maximally selected rank statistics[286] and draw Kaplan-Meier estimators (**Figure 62**). Hereby, we could define patients with very good prognosis, especially for those of stage I to stage III. 10-year survival probability of the low risk group was between 94.8 % (stage III) and 100 % (Stage II). This shows the benefits of ccRCC Score risk stratifications within a stage group over the classical clinical stage classification.

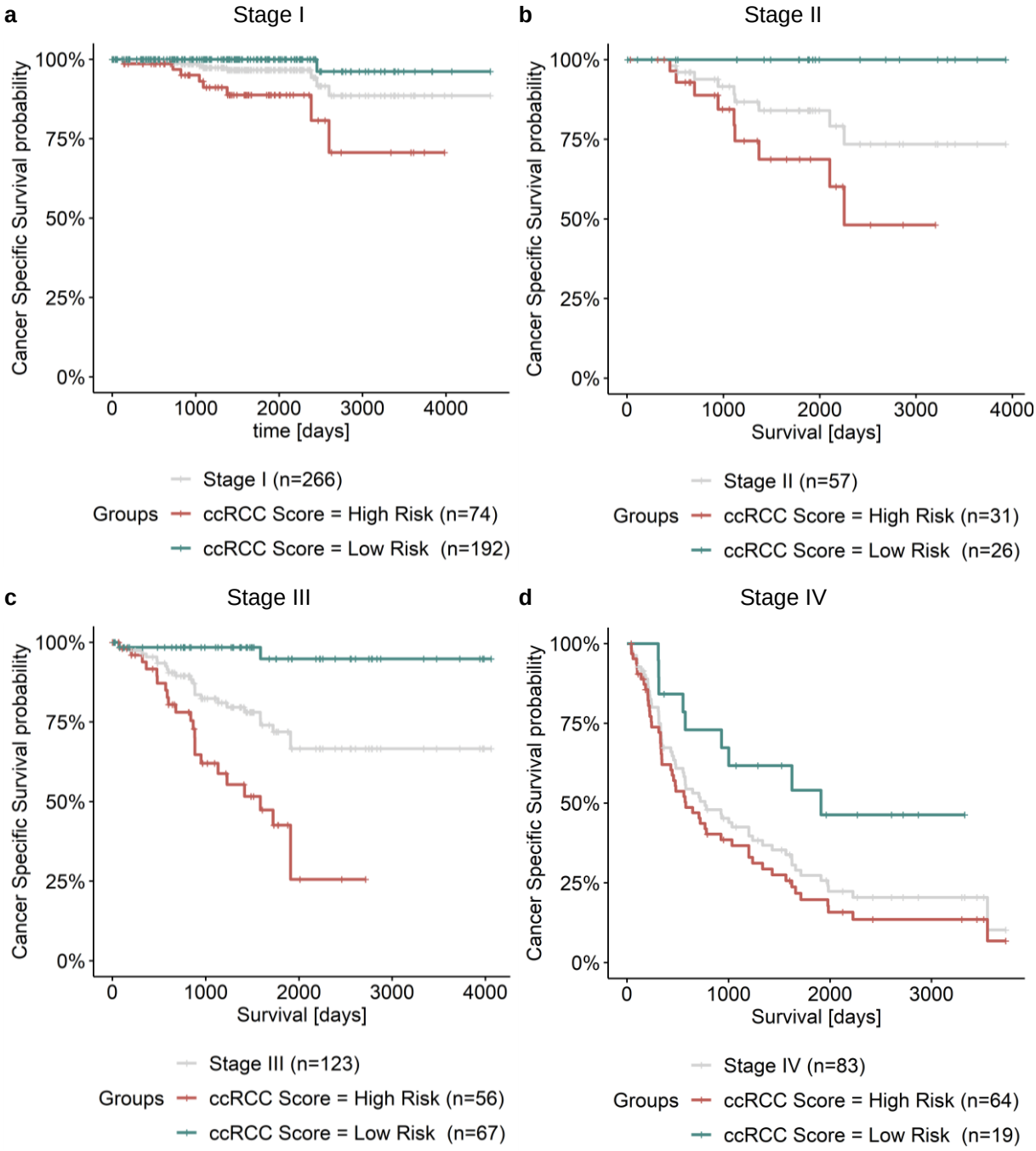


Figure 62 – Risk Stratification by the ccRCC Score gives added value to the stage classification of patients by identifying patients with good prognosis within a stage group

Overlay of survival curves estimating survival for patients either according to the stage alone (grey) and for patients within a stage group stratified by the ccRCC Score (High risk: Red curve; Low Risk: Dark green curve). Patients were stratified using following ccRCC scores calculated by maximally selected rank statistics. Stage I: -1.376, Stage II: -0.237, Stage III: 0.104, Stage IV: 1.367

3.13 Identification of potential drug targets for stage I patients

We were interested in identifying potential drug targets by making use of the ccRCC Score derived risk stratification within the group of stage I patients (**Figure 62a**). Therefore, we analyzed RNASeq gene expression data of high risk and low risk patients and calculated differentially expressed genes between the groups (**Figure 63**). We identified 547 genes that were at least two-fold upregulated and had an adjusted *P* value below 0.05.

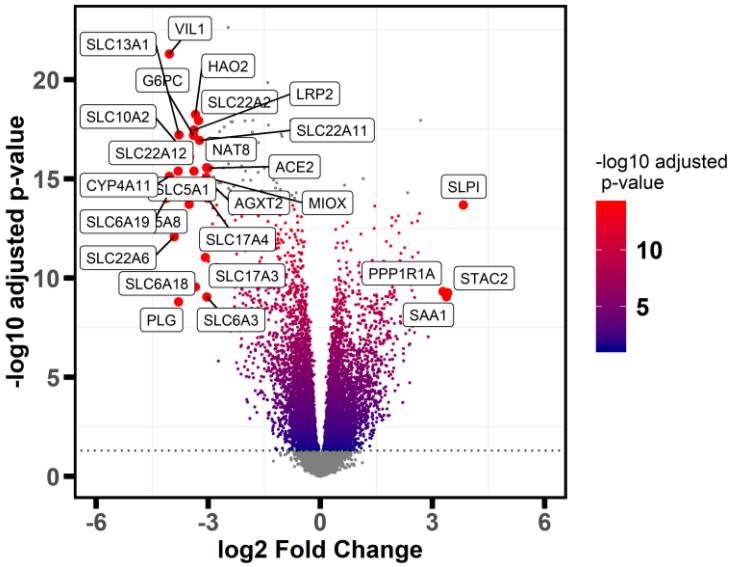


Figure 63 – Volcano plot of differentially expressed genes between high risk and low risk stage I patients

All genes that are at least 8-fold differentially expressed are labeled and highlighted in red.

We made use of the DGIdb drug interaction database to identify druggable targets in the high risk stage I patient cohort[287]. We entered all genes that were at least two-fold differentially expressed and made use of the following preset filters: "FDA approved", "antineoplastic" and "immunotherapies". Additionally we searched only for the following interaction types "inhibitor", "antagonist", "antibody", "blocker", "channel blocker", "cleavage", "desensitize the target", "gating inhibitor", "incorporation into and destabilization", "inhibitor, competitive", "inhibitory allosteric modulator", "inhibitory immune response", "inverse agonist", "negative modulator", "neutralizer", "partial antagonist", "reducer", "suppressor" and "vaccine".

The drug interaction database recognized 546 of the 547 genes and identified nine potential drugs for six potential druggable targets, which could be tested in high risk stage I patients (**Table 9**).

Table 9 – DGIdb Drug Interaction Database results for high risk stage I patients sorted by differentially expression

Gene	Drug	Interaction types	log FC	adj.P.Val
GPR87	CYCLOPHOSPHAMIDE	antagonist	1.9119	6.6E-10
ERBB4	IBRUTINIB	inhibitor	1.4718	1.6E-04
B4GALNT1	DINUTUXIMAB	antibody	1.3220	7.7E-05
IL6	SILTUXIMAB	antagonist|antibody|inhibitor	1.2444	2.6E-04
TUBB3	TRASTUZUMAB EMTANSINE	inhibitor	1.2209	1.6E-10
TUBB3	PACLITAXEL	inhibitor	1.2209	1.6E-10
TUBB3	CABAZITAXEL	inhibitor	1.2209	1.6E-10
TUBB3	BRENTUXIMAB VEDOTIN	inhibitor	1.2209	1.6E-10
KCNH2	AMSACRINE	inhibitor	1.1036	2.9E-05

3.14 *TSPAN8* expression increases throughout the *in vivo* selection

Among the ccRCC signature genes, Tetraspanin-8 (*TSPAN8*) was one of the most differentially expressed genes of the *in vivo* selection (18.37-fold, adj. *P* Value < 0.001). We validated these results using Taqman qRT-PCR (**Figure 64**) and observed a strong induction of this gene along the passages (67.83 ± 7.062-fold).

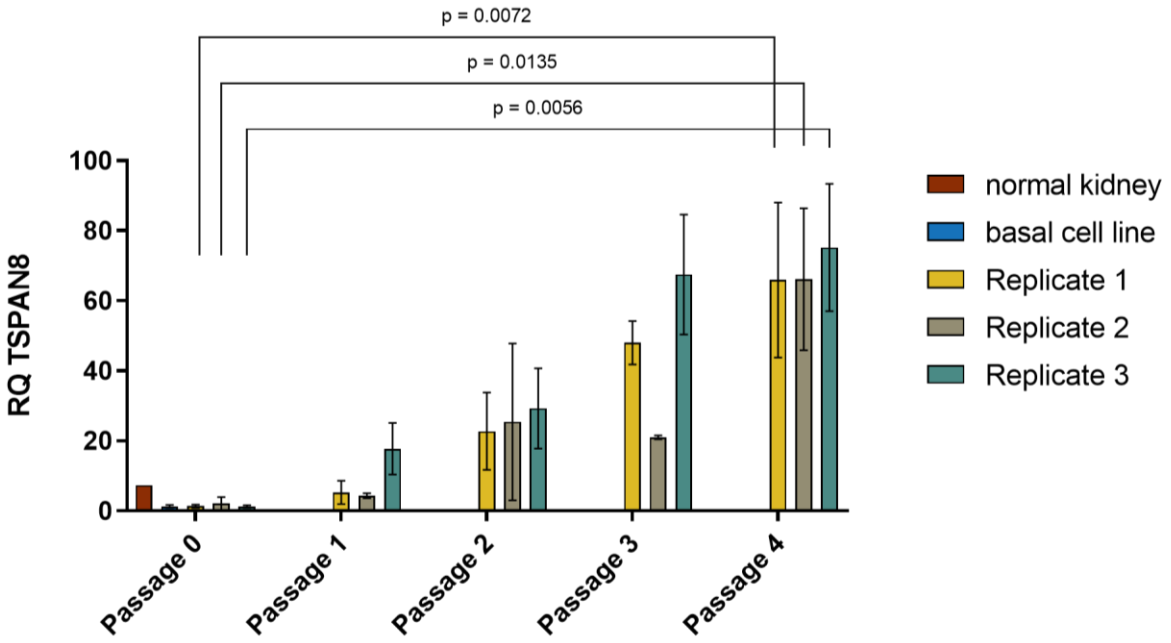


Figure 64 – Relative mRNA expression of *TSPAN8* in different passages of the *in vivo* selection

Expression of *TSPAN8* RNA levels in the primary tumors of the different passages was assessed by Taqman-PCR and normalized to *TSPAN8* expression of the basal KIKA75 cell line. Normal kidney expression of *TSPAN8* was included for comparison. *TSPAN8* is significantly and increasingly upregulated along the passages of the *in vivo* selection. 2-4 biological replicates in 3 technical replicates were analyzed for each passage and replicate. *P* value was calculated using a two-sided Student's *t*-test. Error bars depict mean ± s.e.m.

In addition to transcriptional levels, TSPAN8 protein levels were validated immunohistochemically using an antibody generated by Ailane *et al.*[288]. In passage 0, only few cell clusters positive for TSPAN8 staining were detected (**Figure 65**). The normal proximal tubule cells of the kidney showed weak TSPAN8 expression at their luminal boarder.

In most tumor regions of the last *in vivo* passage, the cellular membrane of tumor and metastases cells were strongly enriched for TSPAN8. Other tumor regions showed less TSPAN8 enrichment. Still we could not identify any specific co-localization of TSPAN8 expressing cells with vessels, extracellular matrix, tumor boarder or necrotic tumor regions.

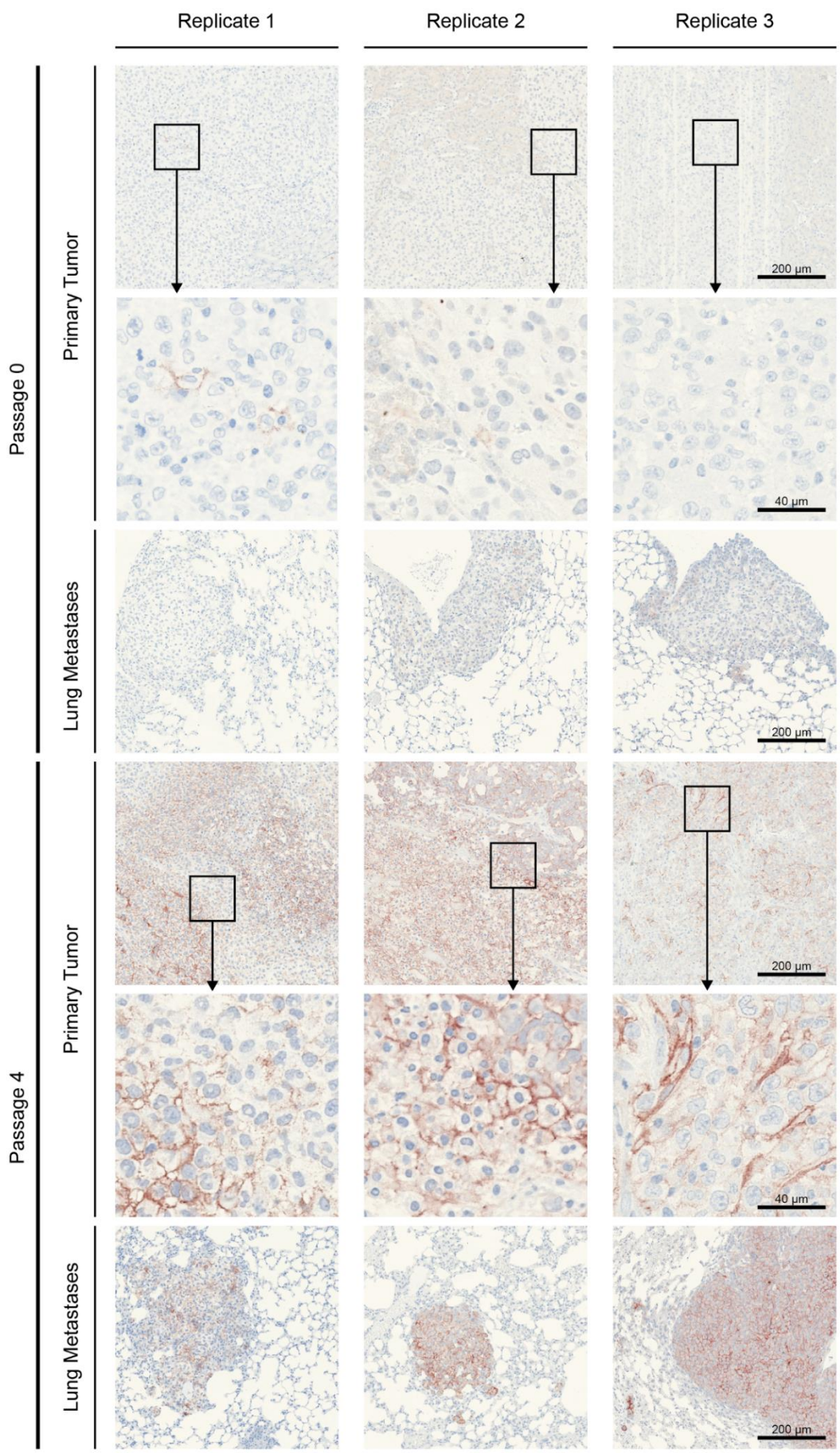

Replicate 1
Replicate 2
Replicate 3
Passage 0
Primary Tumor
Lung Metastases
Passage 4
Primary Tumor
Lung Metastases
200 µm
40 µm
200 µm
200 µm
40 µm
200 µm

Figure 65 – Immunohistochemistry of TSPAN8 of primary tumor and corresponding metastases of the first and the last passage of the in vivo selection

Paraffin embedded tumor material was stained for TSPAN8 (TS29.2). In the primary tumor and the lung metastases of the first passage only little and dispersed staining for TSPAN8 is detectable. The apical boarder of the proximal kidney epithelium stained weakly positive for TSPAN8.
The cell membrane of passage 4 tumor and metastases cells show high TSPAN8 expression in most regions, whereas in other regions no or only little TSPAN8 staining is detected. Staining of TSPAN8 did not co-locate with tumor boarder, vessels, extracellular matrix or necrotic tumor regions.
Enlarged are regions with heterogeneous TSPAN8 expression. Staining was performed by Ornella Kossi.

The established cell lines from the lung metastases of the last passage (**see section 3.4.1**) continued to express *TSPAN8* in their first passages, whereas higher passaged cells showed declining expression of *TSPAN8* (**Figure 66**). This indicates that the expression of *TSPAN8* was either induced by the tumor microenvironment or repressed caused by unknown cell culture conditions.

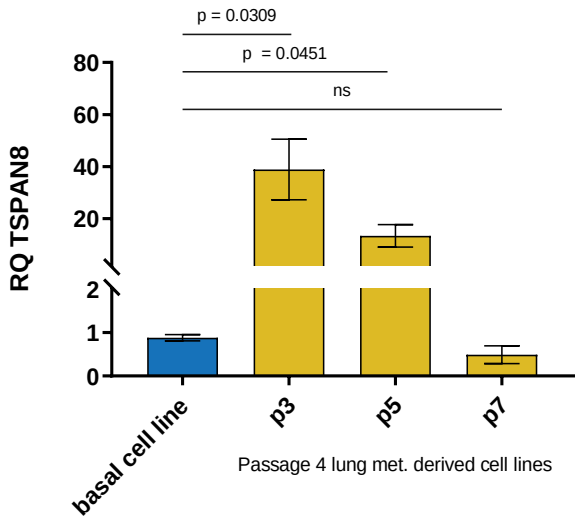


Figure 66 – *TSPAN8* expression in passage 4 lung metastases derived cell lines in comparison to the basal cell line

The expression of *TSPAN8* in comparison to the basal cell line KIKA75 is high in the first passages after generating the cell lines from passage 4 lung metastases. However, the higher the passages of the cell lines, the lower the expression of *TSPAN8*, eventually returning to the basal cell level after seven *in vitro* passages.
Three cell lines per *in vivo* selection replicate were generated and three replicates each for each *in vitro* passage were analyzed. *P* value was calculated using a two-sided Student's *t*-test. Error bars depict mean ± s.e.m.

3.15 *TSPAN8* is heterogeneously expressed between ccRCC patients

We tested *TSPAN8* expression in our primary patient derived ccRCC cell lines and compared it to the expression of *TSPAN8* in normal kidney tissue (**Figure 67a**). The cell lines KIKA12, KIKA27 and KIKA75 had only little *TSPAN8* mRNA expression (0.012, 0.013 and 0.089-fold respectively), whereas the cell lines KIKA24 and KIKA38 had levels close to the normal kidney control (0.721 and 0.479-fold). We tested the expression of TSPAN8 by flow cytometry and found similar results on protein level (**Figure 67b**). To test whether this inter patient heterogeneity was observable in a larger patient set, we analyzed *TSPAN8* expression level of the TCGA KIRC dataset and compared it to annotated normal kidney samples (**Figure 67c**). Interestingly, *TSPAN8* mRNA expression in the tumor material was significantly lower compared to normal kidney tissue (log2 normalized RSEM counts:

Tumor = 3.48 ± 2.91; Normal = 9.84 ± 1.29). Additionally, the intertumor heterogeneity of *TSPAN8* expression was much higher in tumor than normal tissue.

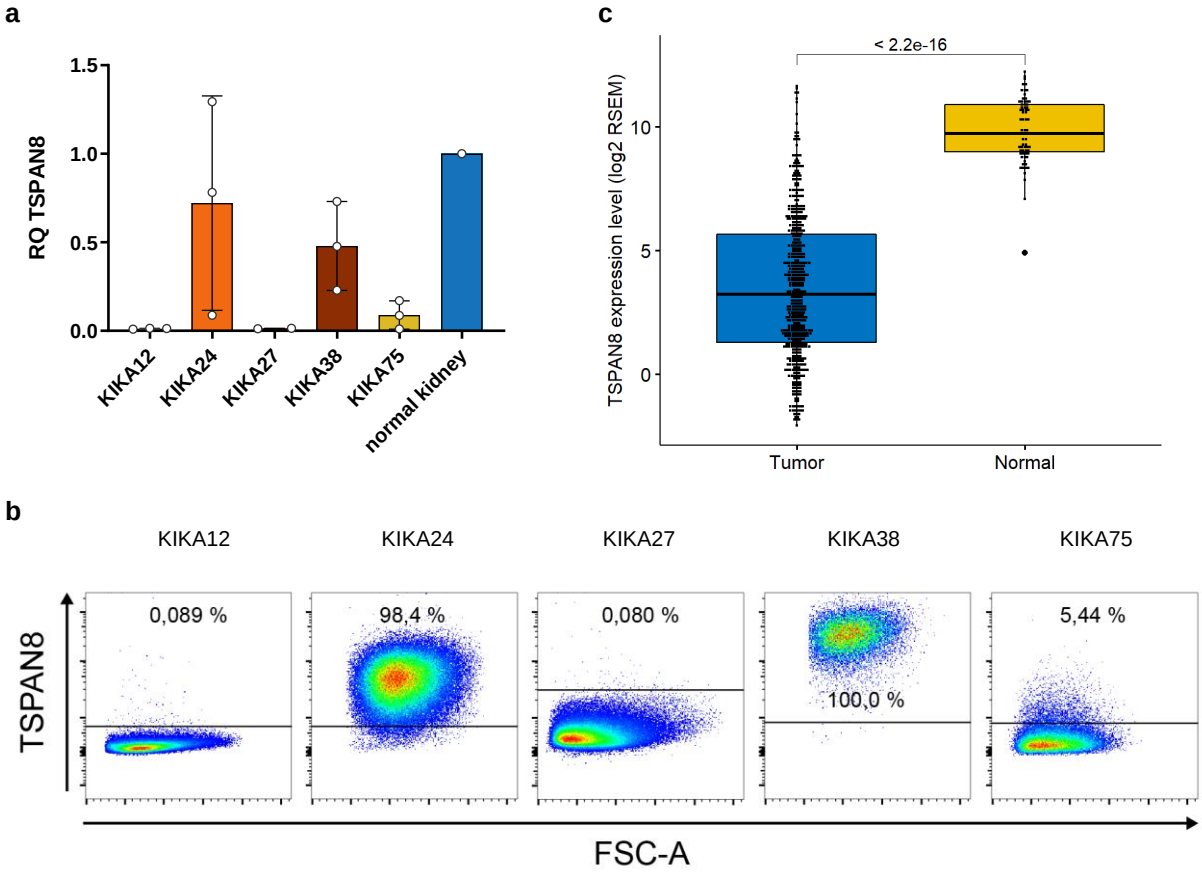


Figure 67 – Relative Expression of *TSPAN8* in the different primary patient derived cell lines in comparison to normal kidney expression

(a) The cell lines KIKA12, KIKA27 and KIKA75 express TSPAN8 mRNA only slightly in comparison to normal kidney, whereas KIKA24 and KIKA38 express levels of TSPAN8 comparable to normal kidney tissue (qRT-PCR). Error bars depict mean ± 95 %-CI.
(b) Protein levels of TSPAN8 correspond to the mRNA level results. In KIKA12, KIKA27 and KIKA75, only a subset of cells stained positive for TSPAN8. The line indicates the isotype control. Error bars depict mean ± 1.5*IQR.
(c) RNASeq expression of TSPAN8 in the TCGA KIRC patient dataset. TSPAN8 is significantly lower expressed in bulk tumor tissue in comparison to normal kidney tissue. Nevertheless, a vast intertumor heterogeneity in the expression of TSPAN8 is discernible. *P* value was calculated using a two-sided Student's *t*-test.

3.16 Expression levels of *TSPAN8* predict patient survival of late ccRCC tumor stages

We were wondering whether the different expression levels of *TSPAN8* also have an influence on patients' cancer specific survival probability. Therefore, we stratified patients according to maximally selected rank statistics[286] into *TSPAN8* high and low expressing tumors and calculated their survival probabilities (**Figure 68**).

Although showing a trend, *TSPAN8* could not significantly separate patients' cancer specific survival in the TCGA-KIRC dataset, according to the classical levels of significance. However, in the Sato *et al.* validation cohort, *TSPAN8* expression profiles significantly separated patients' overall survival.

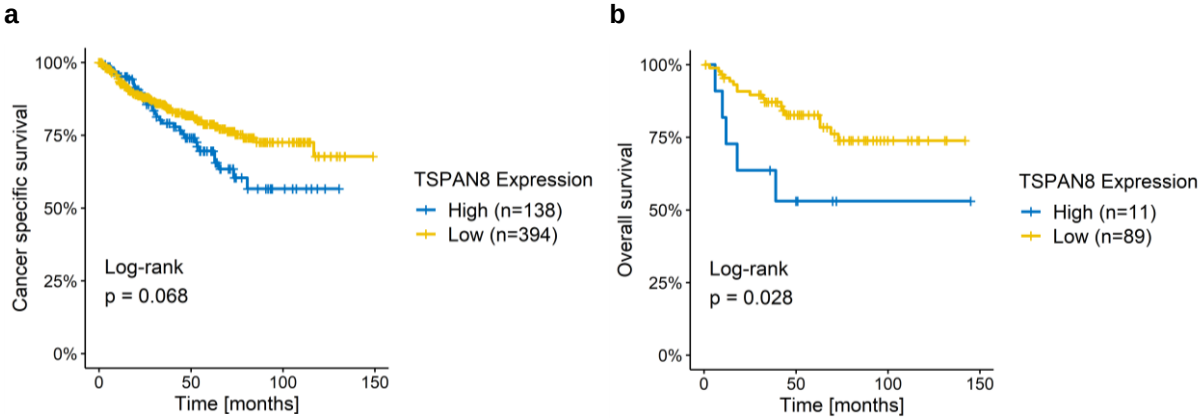


Figure 68 – Kaplan Meier estimators of patients' survival, stratified by the expression of *TSPAN8*

While stratification of *TSPAN8* expression profiles of the tumors does not separate patient survival probabilities in the **(a)** TCGA-KIRC dataset, survival of the **(b)** Sato *et al.* cohort could be significantly separated (log rank test).

Hence, we were asking whether *TSPAN8* only influences the survival probability of patients at a specific tumor stage. To answer this question, patients were additionally stratified according to tumor stage (**Figure 69**). Patients were grouped into either stage I & II or stage III & IV patients, as the Sato *et al.* dataset consists only of 13 stage III and 12 stage IV patients (see **Supplemental Figure 7**).

Patients of the early stages I and II showed no difference in survival in both datasets. However, TCGA-KIRC dataset of 206 patients of advanced stage III and IV showed significantly different survival probabilities. The cancer specific survival of patients harboring *TSPAN8* low expressing tumors was enhanced. In the Sato *et al.* stage III & IV dataset, including only 25 patients, survival of patients was not significantly different according to the classic significance level ($p = 0.073$), although the same trend as for the TCGA-KIRC dataset was discernable.

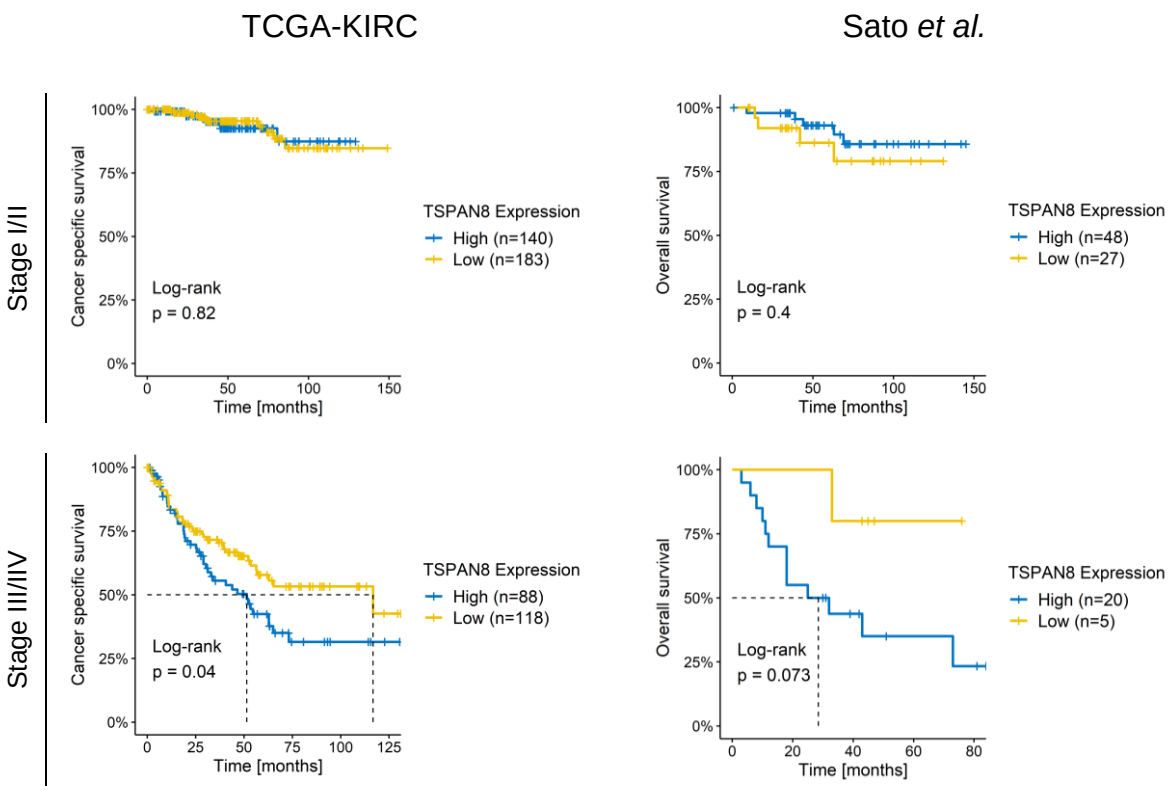


Figure 69 – Survival stratified according to *TSPAN8* expression and tumor stage

Patients of both the TCGA KIRC and Sato *et al.* datasets were stratified according to *TSPAN8* expression and tumor stage. Patients of stage I and II, and patients of stage III and IV were combined, respectively. The TCGA-KIRC dataset patients of Stage III/IV show significantly different survival probabilities when stratified for *TSPAN8* expression profiles, revealing enhanced survival in *TSPAN8* low patients. For patients of the Sato *et al.* cohort, survival probabilities are not significantly different, yet the same trend for enhanced overall survival of patients with *TSPAN8* low expressing tumors can be seen.
For each group, the optimal *TSPAN8* expression cutoff to divide patients was calculated separately by maximally selected rank statistics with a minimal proportion of 20 % per group. *P* values calculated by log rank test.

This difference in survival probabilities according to *TSPAN8* expression motivated us to check for the distribution of TSPAN8 expression according to tumor stage (**Figure 70**). However, no correlation of *TSPAN8* expression and tumor stage was found.

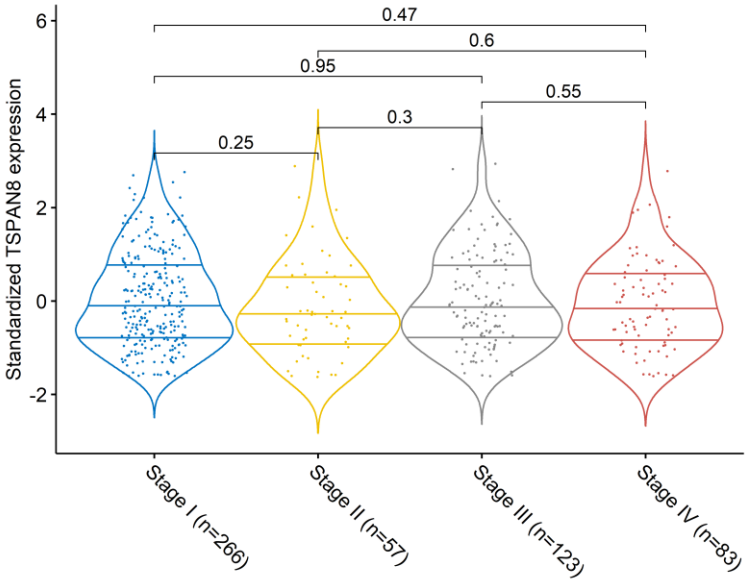


Figure 70 – Violin plots of *TSPAN8* expression and ccRCC tumor stages

Standardized *TSPAN8* expression levels of TCGA-KIRC patients were plotted against their respective tumor stage. There is no discernible difference between *TSPAN8* and tumor stage. *P* value was calculated using a two-sided Student's *t*-test.

3.17 Generation of *TSPAN8* knockdown and overexpression KIKA models

The tetraspanin transmembrane protein family is known to alter cellular signaling properties[289,290]. Therefore, we hypothesized that TSPAN8 contributes to tumor growth, metastasis and thus to the malignancy of this disease, which had been previously shown in other cancer entities[291,292].

As depicted earlier (**Figure 67b**), two of our KIKA cell line models show high *TSPAN8* expression levels (KIKA24 and KIKA38), whereas in the KIKA27 and KIKA75 cell lines, only a minority of the cells express *TSPAN8*. We transduced the cell lines to stably express a luciferase transgene (Renilla-luciferin 2-monooxygenase), which allowed us to monitor their growth *in vivo* after xenotransplantation.

Cell lines were transduced with lentiviral vectors expressing inducible short hairpin RNAs (shRNAs) to silence *TSPAN8* expression via RNA interference (RNAi)[293]. We aimed for a transduction efficiency of approximately 30 % to ensure a single vector integration per cell. Subsequently, transduced cells were selected with puromycin, as the cells proved to be highly susceptible for cell sorting. Thereby, three different cell lines were established for each KIKA model: Two different shRNAs (shTSPAN8(1) and shTSPAN8(2)) and, as a control, a non-silencing shRNA was used that shows minimal homology to the human genome. Efficiency of the knockdown was evaluated by qRT-PCR (**Figure 71a**) and FACS (**Figure 72**).

We saw a significant knockdown of *TSPAN8* on mRNA levels in all generated cell lines. The knockdown efficiency varied between the cell lines and shRNAs. The first shRNA led to a decreased *TSPAN8* mRNA levels of 50, 13, 21 and 35 percent in KIKA24, KIKA27, KIKA38 and KIKA75, respectively in comparison to the not induced cell lines. The second shRNA was slightly more efficient and led to decreased mRNA levels of 17, 13, 10 and 35 percent.

To validate the knockdown on protein level, the generated cell lines were stained with a TSPAN8 specific antibody[288] and analyzed on a flow cytometer (**Figure 72 & Supplemental Figure 13**). Thereby, we could show that the levels of TSPAN8 were also reduced on protein level. Nevertheless, the cell lines expressing high basal TSPAN8 levels showed a reduction in TSPAN8 staining intensity, yet still detectable above isotype staining. Interestingly, the knockdown of *TSPAN8* with the first shRNA in the KIKA24 cell line lead only to an mRNA reduction of 50 %, but showed a strong reduction on protein level.

Next, we established inducible *TSPAN8* overexpressing cell lines. Especially the cell lines with low basal *TSPAN8* expression showed a strong leakiness of the vector on mRNA level (**Figure 71b)**, which could not be seen on protein levels (**Figure 72**). After induction with doxycycline, *TSPAN8* showed a strong induction on both mRNA and protein levels in all used cell lines.

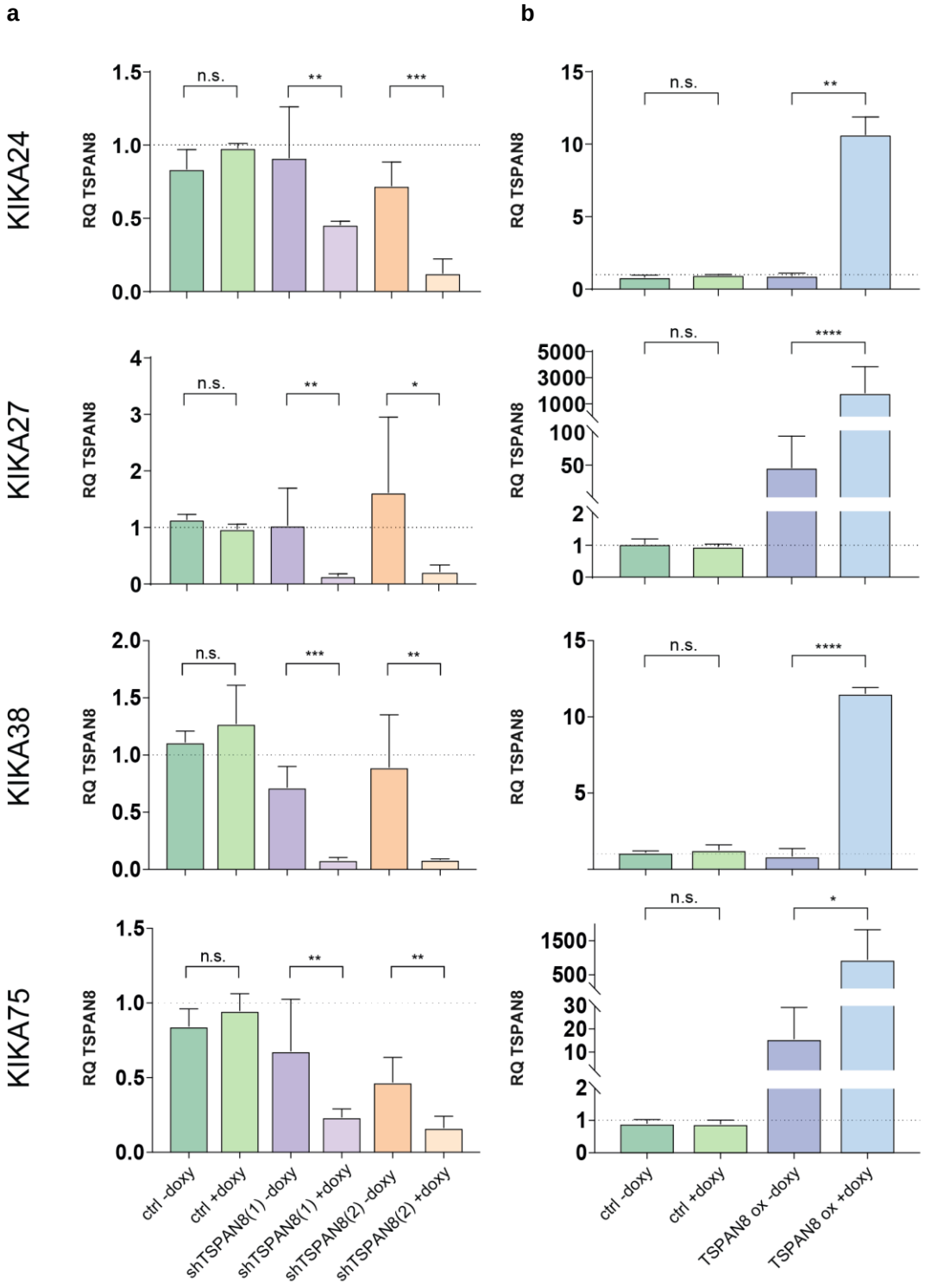


Figure 71 – qRT-PCR of *TSPAN8* knockdown and overexpression in different KIKA models

(a) *TSPAN8* mRNA levels are significantly reduced after doxycycline induction in both tested shRNA vectors in all tested KIKA models. **(b)** Overexpression of *TSPAN8* shows leakiness in the cell lines with low *TSPAN8* levels. The levels of *TSPAN8* are significantly enriched after inducing the overexpression with doxycycline. *P* value was calculated using a two-sided Student's *t*-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns = not significant, ox = overexpression. Error bars depict mean ± s.e.m.

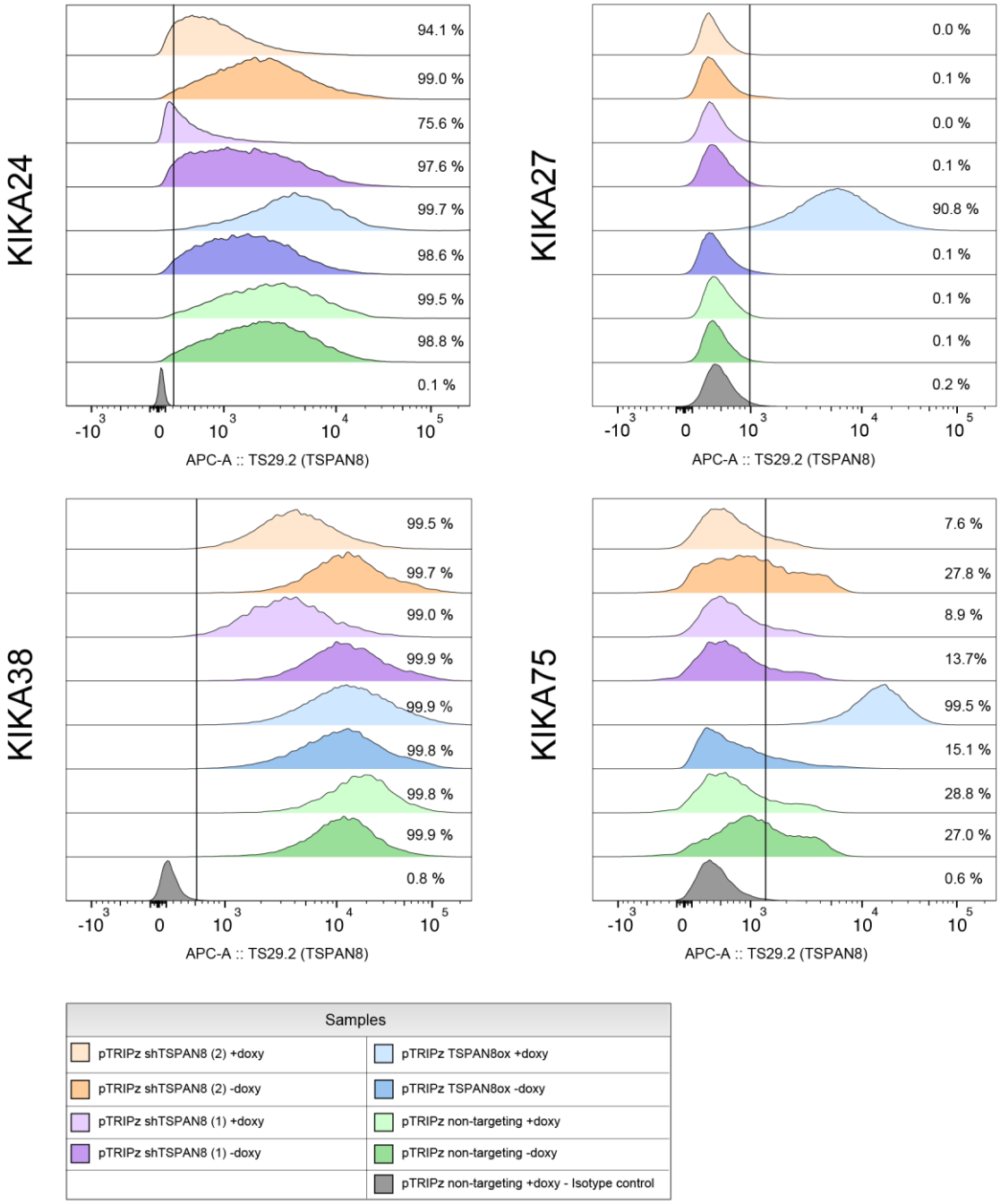


Figure 72 – Protein levels of TSPAN8 in TSPAN8 knockdown and overexpressing KIKA models measured by flow cytometry

Overexpression of TSPAN8 is highly efficient in both of the TSPAN8 low expressing cell lines and is discernible in the KIKA24 cell line. *TSPAN8* knockdown shows a strong reduction of TSPAN8 in KIKA24, KIKA38 and KIKA75. ox = overexpression

3.18 The knockdown of TSPAN8 delays tumor growth

We pretreated NSG mice with doxycycline for seven days, orthotopically injected the generated TSPAN8 knockdown KIKA cell lines and finally measured tumor growth over time as shown in **Figure 73**.

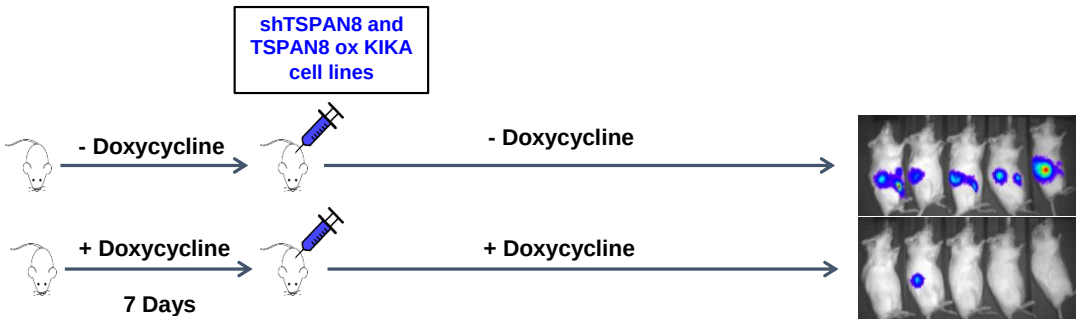


Figure 73 – Schema of *in vivo TSPAN8* knockdown experiments

Mice of the knockdown group were pretreated with doxycycline for 7 days before orthotopic injection of doxycycline pretreated KIKA cells carrying a luciferase transgene and either a shRNA targeting TSPAN8 or a non-targeting scramble shRNA. Tumor growth was monitored with IVIS 200 imager system at least once a week.

In all tested models, doxycycline alone had no significant effect on tumor growth (**Figure 74a, Figure 75a, Figure 76a, Figure 77a**). The effect of TSPAN8 knockdown varied between the tumor models and the used shRNAs. In the KIKA24 model, the shTSPAN8(1) showed only a 50 % reduction of *TSPAN8* on mRNA level but was highly efficient in the reduction of TSPAN8 on protein level (**Figure 71a, Figure 72**). The induction of this shRNA in the *in vivo* tumor growth experiment showed a significant reduction of tumor growth (**Figure 74b**). shTSPAN8(2) showed no effect on tumor growth (**Figure 74c**).

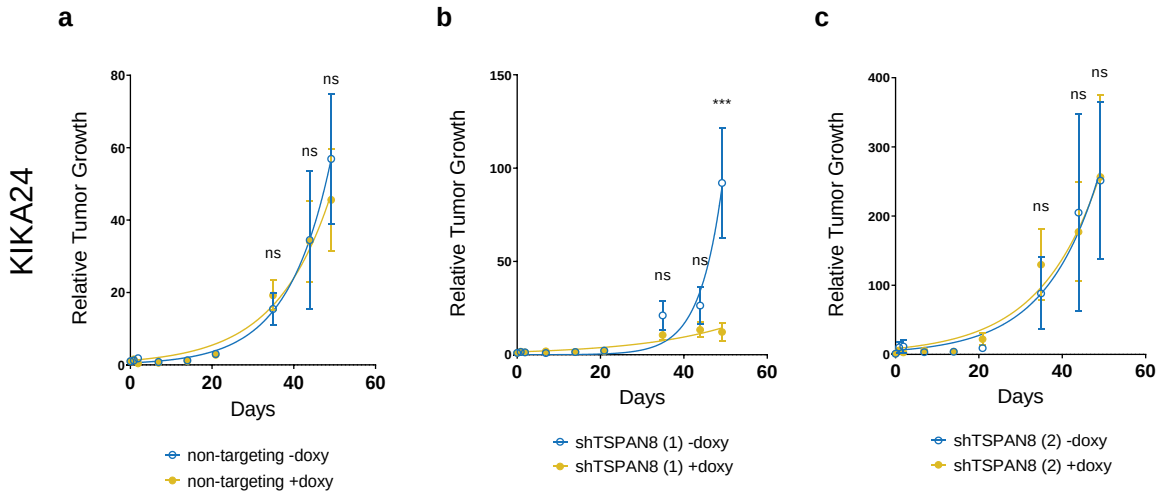


Figure 74 – Orthotopic tumor growth of KIKA24 in dependence of TSPAN8

(a) Doxycycline treatment has no effect on tumor growth. **(b)** Knockdown of TSPAN8 with shTSPAN8(1) significantly affects tumor growth **(c)** Induction of shTSPAN8(2) has no effect on tumor growth.
Tumor signal was normalized to day 0 and followed up for 49 days. *P* value was calculated for the last three time points using a two-sided Student's *t*-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns = not significant. Error bars depict mean ± s.e.m.

KIKA27 is the cell line with the lowest basal *TSPAN8* levels in our study (**Figure 67a,b**). Knockdown of *TSPAN8* in this line reduced the mRNA levels even further. As protein levels *in vitro* were already low, knockdown efficiency on protein level could not be determined. Strikingly, *in vivo*, knockdown of *TSPAN8* with both shRNAs had a strong effect on tumor growth as it stopped tumor growth in shTSPAN8(1) cell line derived tumors (**Figure 75b**) and delayed tumor growth substantially in shTSPAN8(2) tumors (**Figure 75c**).

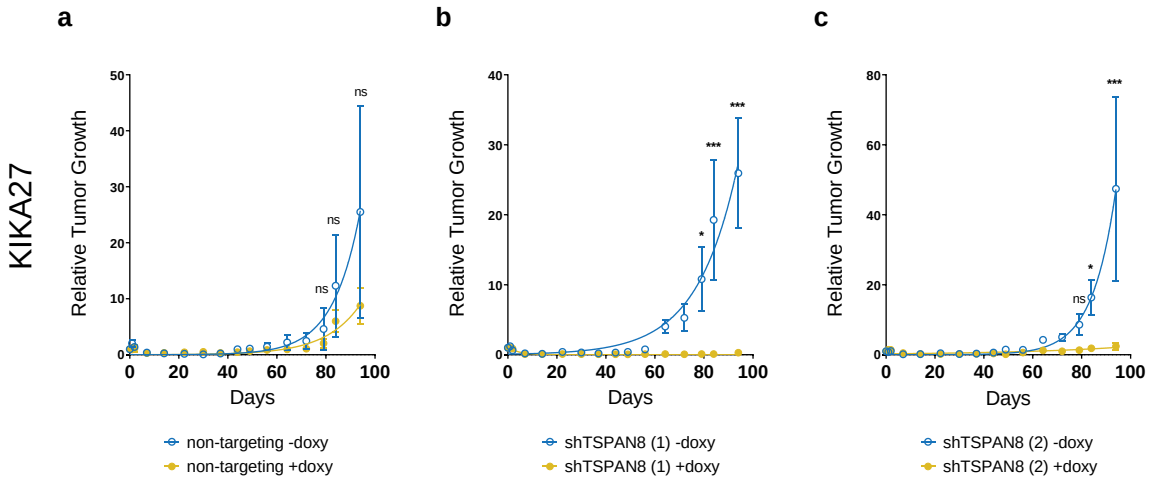


Figure 75 – Orthotopic tumor growth of KIKA27 in dependence of TSPAN8

(a) Doxycycline treatment has no effect on tumor growth. **(b,c)** Knockdown of TSPAN8 with both shRNAs affects tumor growth significantly.
Tumor signal was normalized to day 0 and followed for 94 days. *P* value was calculated for the last three time points using a two-sided Student's *t*-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns = not significant. Error bars depict mean ± s.e.m.

KIKA38 show high basal TSPAN8 levels (**Figure 67a,b**) and knockdown was highly efficient on mRNA levels (*TSPAN8* levels of 21 % and 10 %, **Figure 71a**). Nevertheless, with reduced TSPAN8 protein levels, the majority of cells still stained positively for TSPAN8 (**Figure 72**).

The reduction of TSPAN8 levels in the KIKA38 derived tumors by doxycycline induction *in vivo* had no significant influence on tumor growth (**Figure 76b,c**).

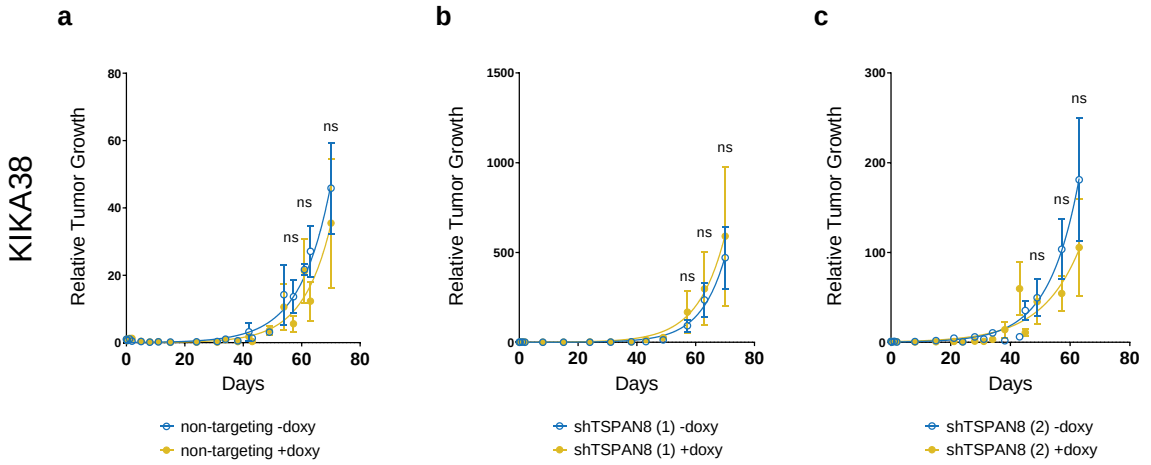


Figure 76 – Orthotopic tumor growth of KIKA38 in dependence of TSPAN8

(a) Doxycycline treatment has no effect on tumor growth. **(b,c)** Knockdown of TSPAN8 with both shRNAs has no effect on tumor growth.
Tumor signal was normalized to day 0 and followed for 63 days. *P* value was calculated for the last three time points using a two-sided Student's *t*-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns = not significant. Error bars depict mean ± s.e.m.

The KIKA75 line contains of cells with heterogeneous TSPAN8 expression. Whereas the majority of cells show no TSPAN8 signal, 5-30 % of the cells have varying levels of TSPAN8 (**Figure 67b, Figure 72**). Knockdown of *TSPAN8* on mRNA levels to 35 % of the normal level led to a reduction on protein level, yet incomplete (**Figure 71a, Figure 72)**. The knockdown delayed cell line derived tumor growth with both shRNAs (**Figure 77b,c**).

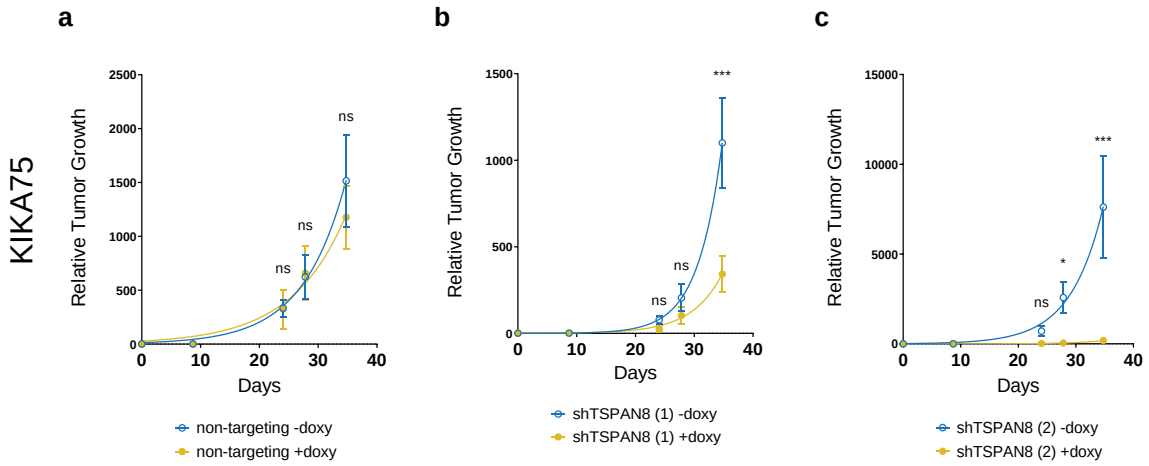


Figure 77 – Orthotopic tumor growth of KIKA75 in dependence of TSPAN8

(a) Doxycycline treatment has no effect on tumor growth. **(b)** Knockdown of TSPAN8 with shTSPAN8(1) significantly delays tumor growth. **(c)** Induction of shTSPAN8(2) has a strong effect on tumor growth. Tumor signal was normalized to day 0 and followed for 35 days. *P* value was calculated for the last three time points using a two-sided Student's *t*-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns = not significant. Error bars depict mean ± s.e.m.

3.19 Overexpression of TSPAN8 has no effect on tumor growth

We wondered whether high levels of TSPAN8 alone are sufficient enough to promote tumor growth. Therefore, we tested *in vivo* tumor formation of the KIKA models with low TSPAN8 levels with an inducible TSPAN8 overexpression vector and the TSPAN8 high KIKA38 model in comparison.

Overexpression of TSPAN8 had no significant effect on tumor growth in KIKA27 and KIKA38 derived tumors and surprisingly reduced tumor growth of KIKA75 derived tumors (**Figure 78**). Thus, the *in vivo* TSPAN8 overexpression did not show the opposite effect of the TSPAN8 knockdown models.

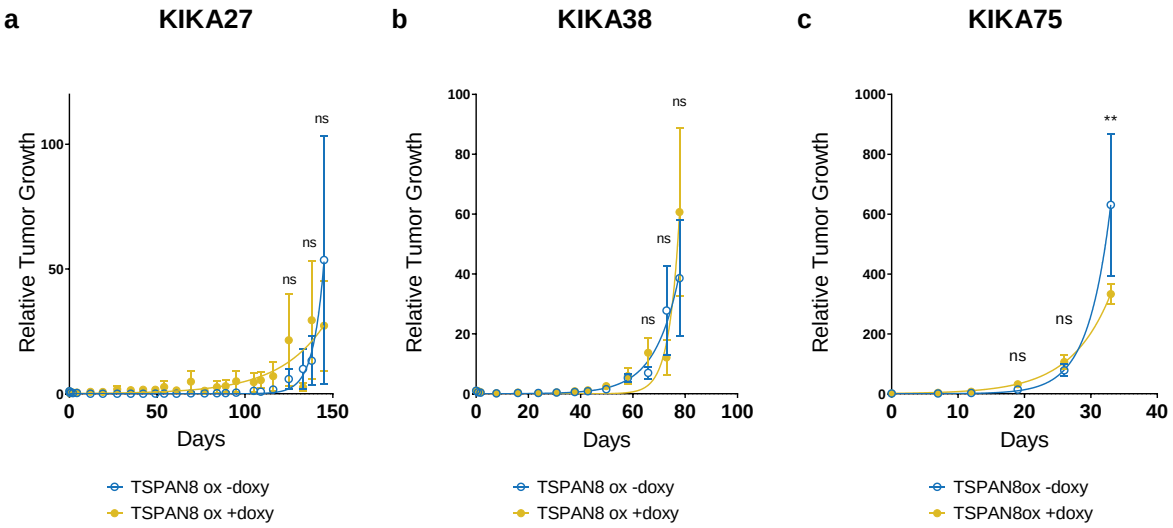


Figure 78 – Orthotopic tumor growth of KIKA models with TSPAN8 overexpression

(a,b) TSPAN8 overexpression shows no significant effect on tumor growth in KIKA27 and KIKA38 derived tumors. **(c)** The overexpression of TSPAN8 in KIKA75 derived tumors slows down tumor growth significantly at day 33. *P* value was calculated for the last three time points using a two-sided Student's *t*-test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns = not significant, ox = overexpression. Error bars depict mean ± s.e.m.

3.20 Knockdown of TSPAN8 stops tumor growth on already established tumors

As shown earlier, knockdown of TSPAN8 in TSPAN8 low tumor models inhibited tumor formation (see **Figure 75** and **Figure 77**). Next, we wondered whether knockdown of TSPAN8 may also have an effect on already established tumors (**Figure 79**). For this purpose, we let KIKA27 derived tumors grow for 56 days, randomized the mice according to tumor size and induced the TSPAN8 knockdown in one of the groups (**Figure 80**).

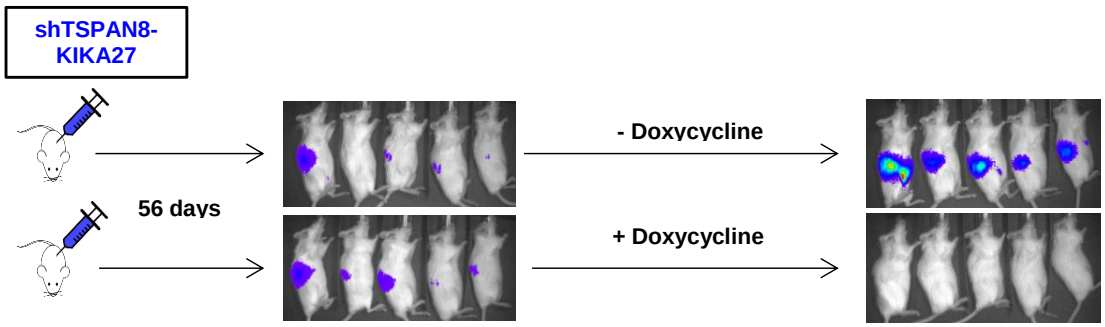


Figure 79 – Schema of *in vivo* TSPAN8 knockdown on established tumors

NSG mice were orthotopically injected with KIKA cells carrying a luciferase transgene and one of two different shRNA targeting TSPAN8. Tumor growth was monitored with IVIS 200 imager system at least once a week. After 56 or 90 days, mice were randomized into two groups of similar tumor size. We subsequently induced the knockdown in one of the groups by giving doxycycline containing drinking water until the end of the experiment and measured tumor growth over time.

Both shTSPAN8 KIKA27 derived tumors showed normal tumor development after orthotopic cell injection. A strong decrease in tumor signal right after injection was observed, which recovered over time and reached approximate starting values after 56 days of tumor growth (**Figure 80a+c**). When TSPAN8 knockdown was induced in one of the randomized groups, tumor growth was halted (**Figure 80b**) or delayed (**Figure 80d**), whereas the non-induced group showed exponential tumor growth. This result indicates that TSPAN8 plays a general role in tumor growth of our ccRCC models.

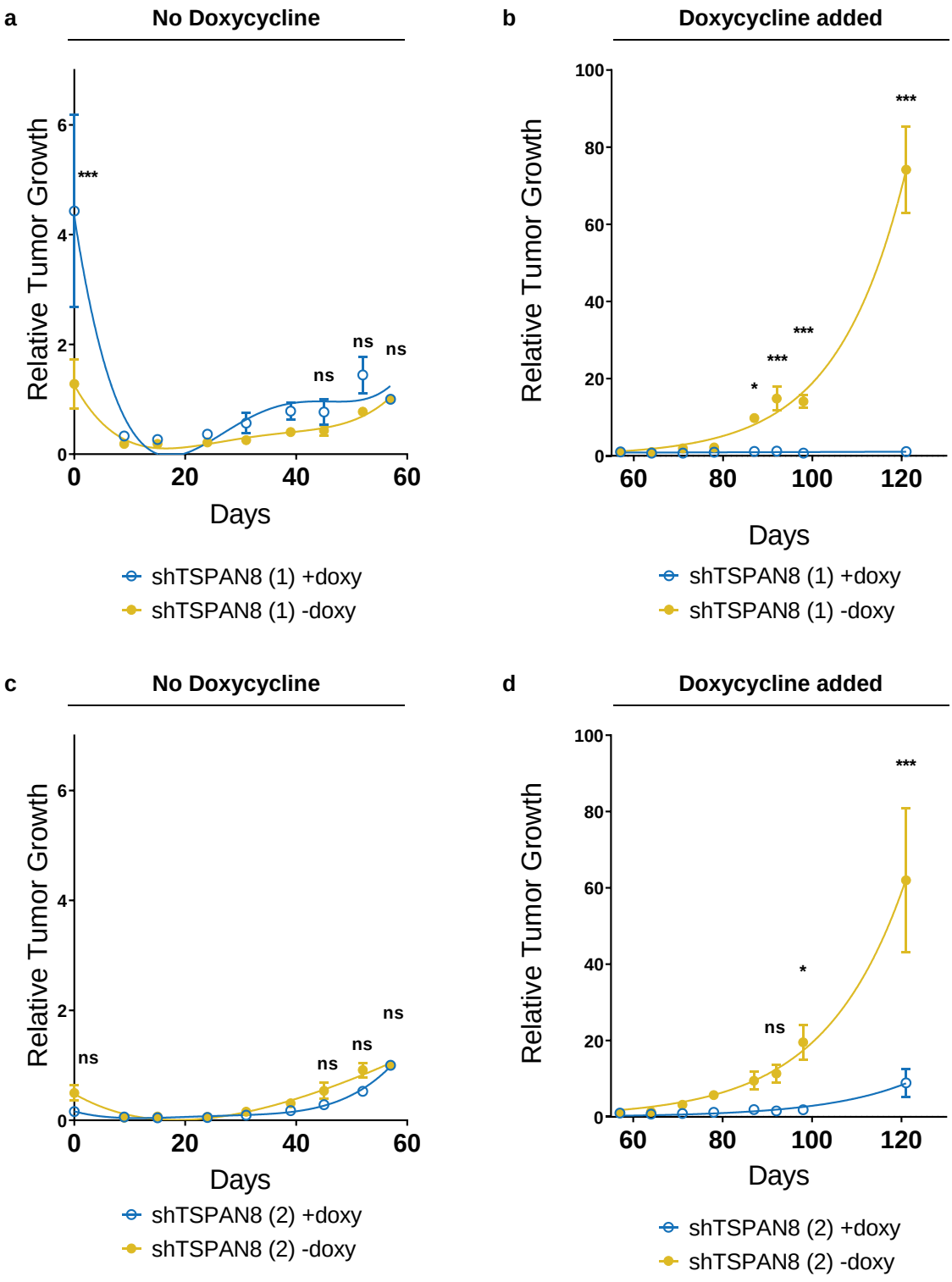


Figure 80 – Knockdown of TSPAN8 delays tumor growth of established KIKA27 tumors

After 56 days of tumor growth, we randomized mice of either shTSPAN8(1) or shTSPAN8(2) into two groups of equal tumor size **(a,c)**. One of each groups was given doxycycline containing drinking water and tumor growth was followed over time. Tumors with the TSPAN8 knockdown **(b)** halt in growth or **(d)** show delayed growth.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns = not significant. Error bars depict mean ± s.e.m.

3.21 TSPAN8 overexpression has no influence on gene expression *in vitro*

We wondered whether TSPAN8 plays a potential role in tumor progression. Therefore, we tested if differential expression of *TSPAN8* in the KIKA cell lines had an influence on cellular expression profiles *in vitro*.

For this purpose, RNA of overexpressed *TSPAN8* in the KIKA27 and KIKA75 cell lines that have low basal *TSPAN8* expression levels were analyzed by comparative gene expression profiling. The microarray analysis shows increased *TSPAN8* signal intensity (**Figure 81a,b**), which has already been validated by qRT-PCR and FACS (see **section 3.17**).

From this gene expression profiling data we performed differential expression analysis by treatment-contrast parametrization (**Figure 81c,d**). Only a few genes were differentially expressed, among them *TSPAN8*. In the KIKA27 cell line, the differential expression significance level of *TSPAN8* did not reach an adjusted *P* value below 0.05 (adj. *P* value = 0.0578), possibly due to only two analyzed samples per group. This indicates that overexpression of TSPAN8 does not greatly change cellular gene expression patterns. Among the upregulated genes in the KIKA27 cell line was *IL13RA2*, one gene of the ccRCC signature.

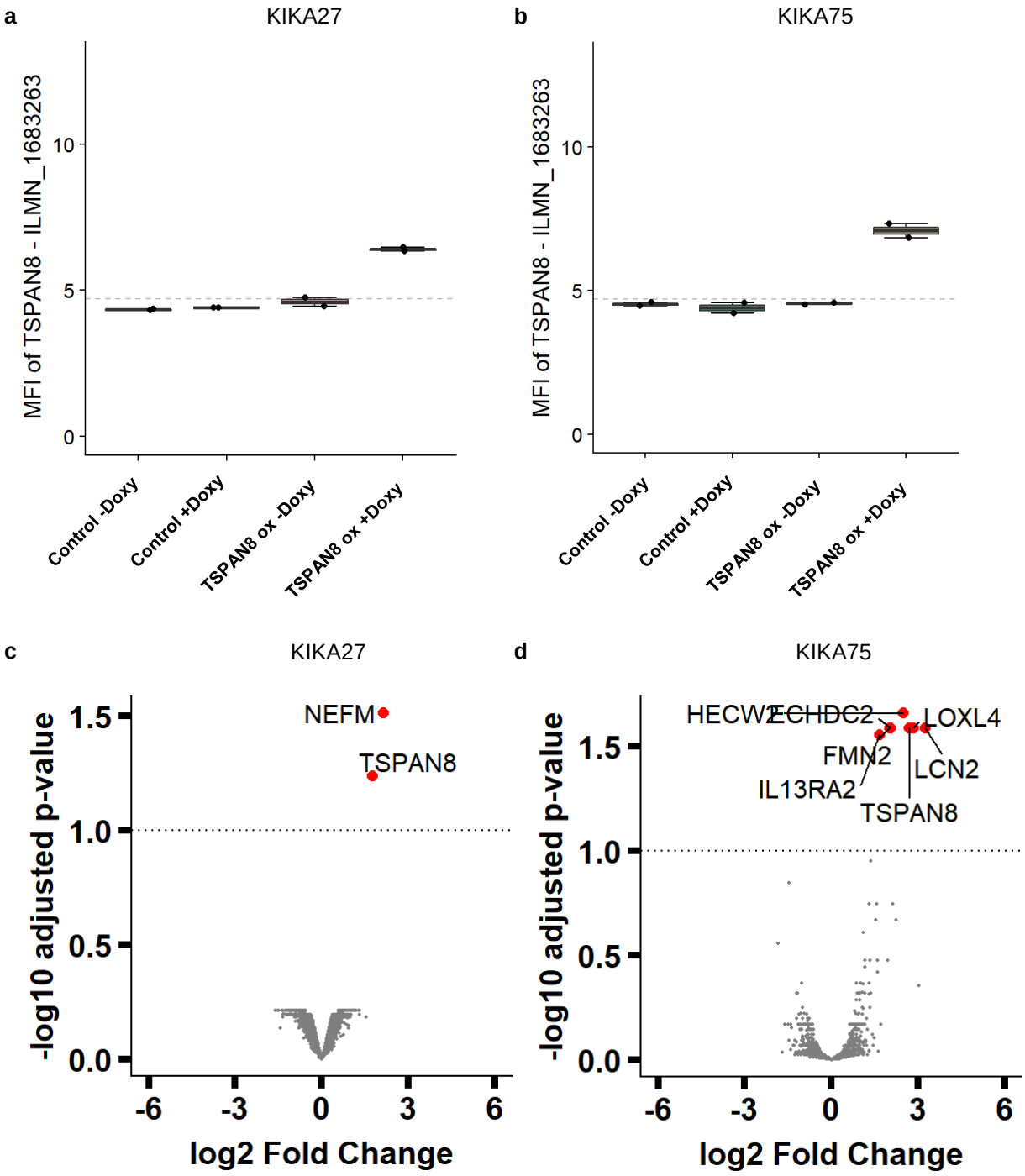


Figure 81 – Overexpression of TSPAN8 has only a minor impact on gene expression

Microarray analysis shows that in both, **(a)** KIKA27 and **(b)** KIKA75 cell lines, mean *TSPAN8* signal intensity is raised above background when the TSPAN8 overexpression cell lines are induced with doxycycline (compare **section 3.17**). Dotted line indicates background signal intensity. Error bars depict mean ± 1.5*IQR. Volcano Plot of differentially expressed genes of the TSPAN8 overexpressing cell lines **(c)** KIKA27 or **(d)** KIKA75. Genes with an adj. *P* value > 0.1 and logFC > 1.5 are annotated. Dotted line indicates an adjusted *P* value = 0.1

3.22 TSPAN8 knockdown has no influence on gene expression *in vitro*

Gene expression profiling was performed from RNA of KIKA38 *TSPAN8* knockdown cells, which showed high endogenous *TSPAN8* expression. As already validated by qRT-PCR and FACS analysis (see **section 3.17**), *TSPAN8* signal intensity was also reduced also in the microarray analysis (**Figure 82a**).

We performed differential expression analysis of the paired samples (**Figure 82b**). Only a few genes were significantly differentially regulated, among them *TSPAN8* (logFC = -2.90).

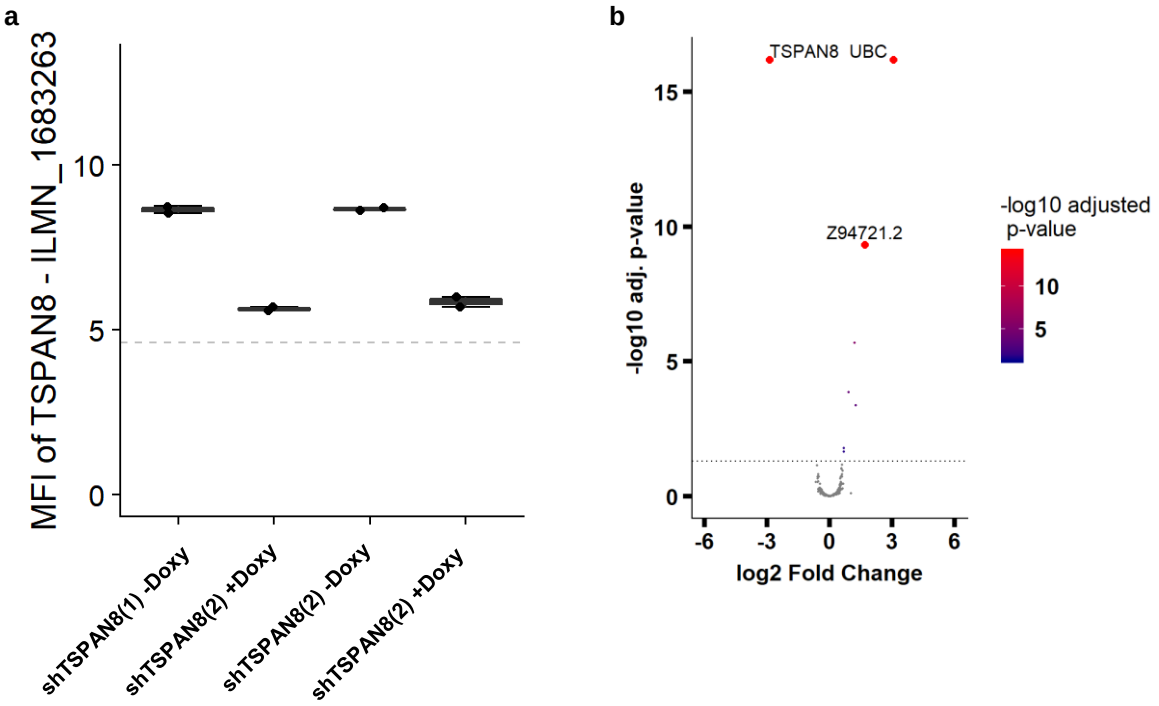


Figure 82 – Knockdown of TSPAN8 in the KIKA38 cell line has a minor impact on gene expression

(a) In both *TSPAN8* knockdown cell lines, mean *TSPAN8* signal intensity of the microarrays is reduced after doxycycline induction (compare **section 3.17**). Dotted line indicates background signal intensity.
(b) Volcano Plot of differentially expressed genes of the KIKA38 TSPAN8 knockdown cell line. Labeled are genes with a logFC > 1.5 and an adj. *P* value > 0.05. Dotted line indicates an adjusted *P* value = 0.05, genes with a smaller *P* value are colored according to their significance level. Error bars depict mean ± 1.5*IQR.

3.23 Changes in TSPAN8 level have no influence on growth rates *in vitro*

We wondered whether the inhibition of tumor growth after *TSPAN8* knockdown was a cell intrinsic effect. Therefore, cell growth of KIKA27 and KIKA75 cell lines was analyzed, which showed strongest tumor growth inhibition *in vitro*. After pretreating the cells with doxycycline for three days, cell growth was followed for 12 days without replacing the media and analyzed by CellTiter Blue consumption as a measurement for cell viability. Growth curves were fitted by logistic regression and mean growth rates were estimated for *TSPAN8* knockdown as well as *TSPAN8* overexpression cell lines.

The results indicate that there is no significant difference in cellular growth rates between the induced knockdown of *TSPAN8* in both KIKA27 (**Figure 83a**) and KIKA75 cell lines (**Figure 83b**). Similarly, overexpression of *TSPAN8* had no significant effect on cell growth.

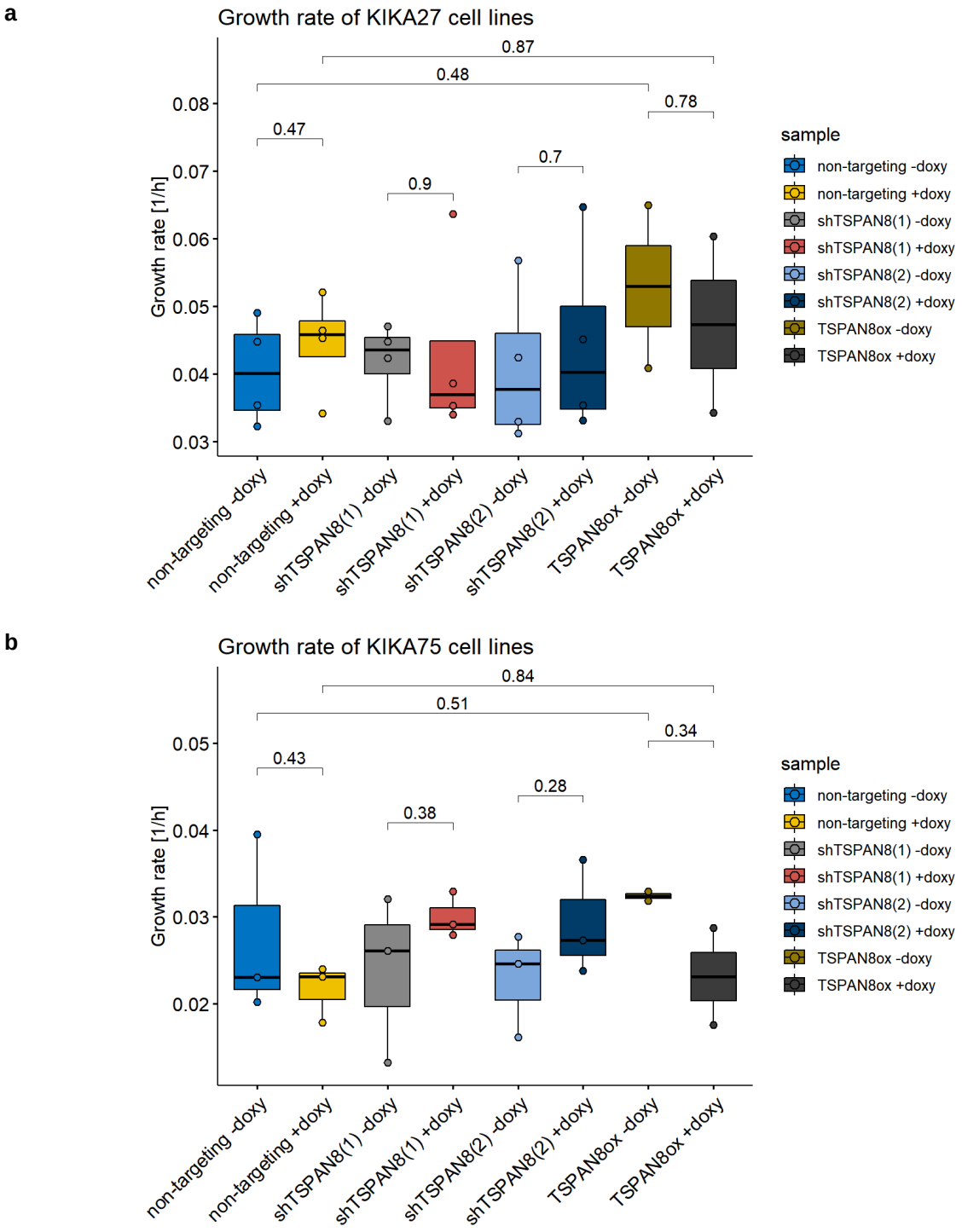


Figure 83 – Growth rates of KIKA27 cell lines estimated by logistic regression

shTSPAN8 cell lines were seeded in four (**(a)** KIKA27) or three (**(b)** KIKA75) and TSPAN8 overexpression cell lines in two biological replicates. For each replicate, 500 cells per time point were seeded in three technical replicates and cell growth was assessed by CellTiter Blue consumption as a measurement for cell viability. The growth rate was estimated by logistic regression of the growth curves. No significant difference in growth rates between the different experimental conditions was observed. *P* value was calculated using a two-sided Student's *t*-test. Error bars depict mean ± 1.5*IQR.

3.24 *TSPAN8* expression is not dependent on *TP53*

Agaësse *et al.* demonstrated that p53 regulates the expression of *TSPAN8* to prevent tumor invasiveness in melanoma[294]. To test this hypothesis in renal cancer, we analyzed *TSPAN8* expression according to the mutational status of p53 (**Figure 84a**). Of the 532 patients of the TCGA KIRC cohort, eight patients had a truncating p53 mutation (1.5 %) and seven patients a p53 missense mutation (1.3 %). We could not observe a significant correlation of mutational status of *TP53* and *TSPAN8* expression levels. Additionally, *TSPAN8* mRNA levels in the TCGA-KIRC patient cohort did not correlate with either *TP53* mRNA (**Figure 84b**) or protein expression levels (**Figure 84c**).

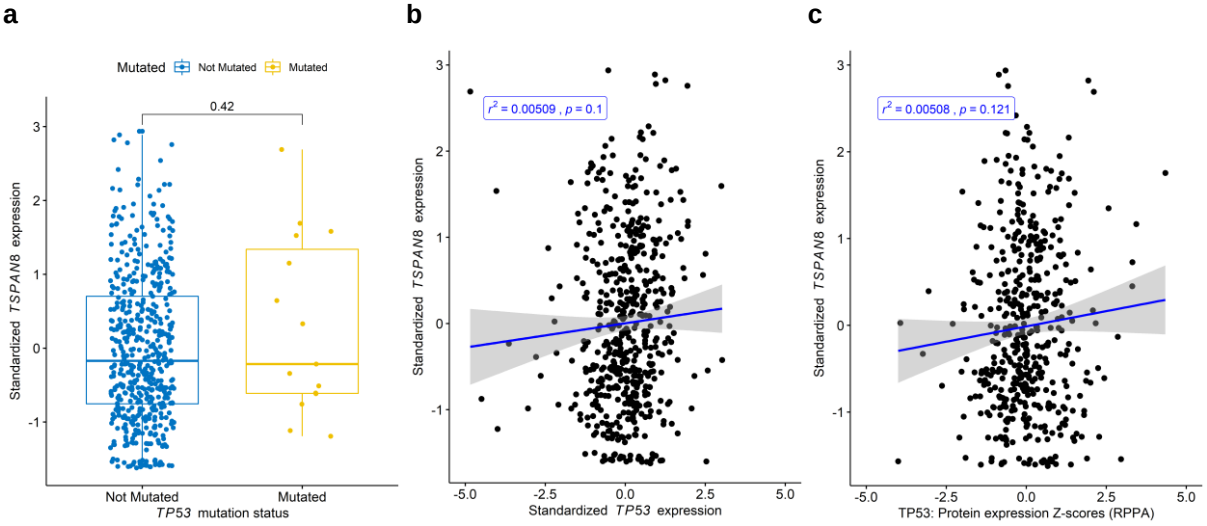


Figure 84 – p53 mutation or expression status does not correlate with *TSPAN8* gene or protein expression

(a) Mutational status of p53 in patients of the TCGA-KIRC cohort was plotted against standardized TSPAN8 expression levels. *P* value was calculated using a two-sided Student's *t*-test.
(b) Scatter plot of standardized TP53 mRNA expression levels with standardized TSPAN8 expression levels. No correlation is observable.
(c) Scatter plot of z-scored TP53 protein expression levels with standardized TSPAN8 expression levels. No correlation is observable.
The linear model, R-squared values and F-statistics were calculated using the *stats* R-package.

3.25 FOXA2 target genes are enriched in the *in vivo* selection

It has been shown that FOXA2 drives tumor progression in ovarian and uterine carcinomas[295,296] and that high *FOXA2* expression in ccRCC has dismal patient outcomes[297]. We found an enrichment for FOXA2 (also called HNF3B) target genes in the list of differentially regulated genes of the *in vivo* selection, indicating that increased aggressiveness of the selected tumors might be regulated by FOXA2 (**Figure 85**).

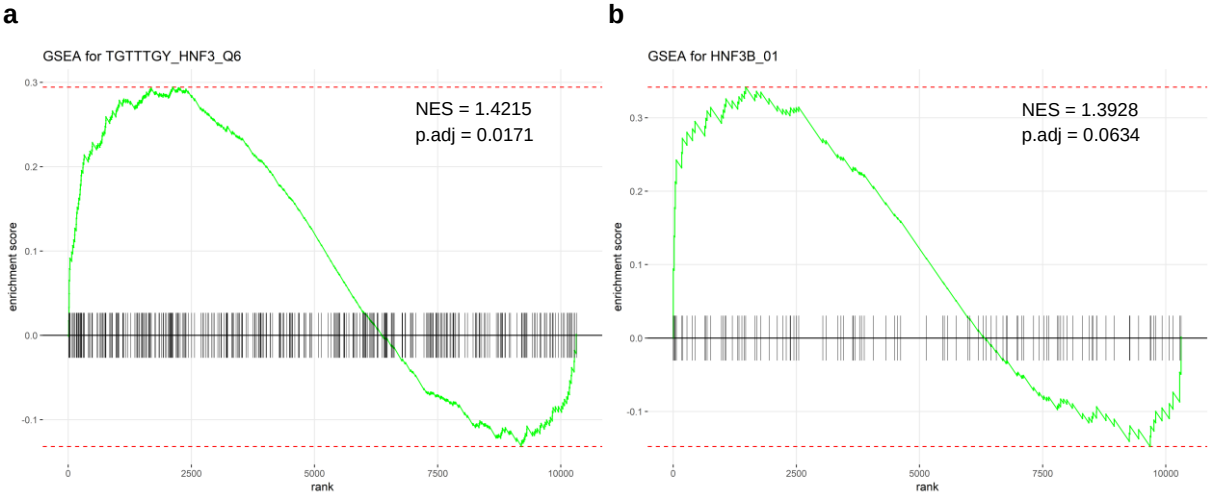


Figure 85 – Gene Set Enrichment Analysis of FOXA2 target genes

(a) Enrichment for genes containing the FOXA2 consensus sequence in their promotor
(b) Enrichment for genes harboring the transcription factor binding site around 4kb of their transcription start sites (v7.4 TRANSFAC).

3.26 Knockdown of *FOXA2* delays tumor growth by regulating *TSPAN8* expression

FOXA2 was, relatively to normal kidney, highly upregulated in the KIKA models as shown in **Figure 86a**. Moreover, *FOXA2* expression was found to be expressed in the TCGA-KIRC patient cohort (**Figure 86b**). However, no difference in *FOXA2* expression between tumor and normal tissue is discernible. Nevertheless, stage I patients have significantly lower *FOXA2* expression levels, than late stage patients. This is also reflected in the cancer specific survival probabilities showing that patients with low *FOXA2* expression levels have significantly better survival probabilities than patients with high *FOXA2* levels (**Figure 86c**).

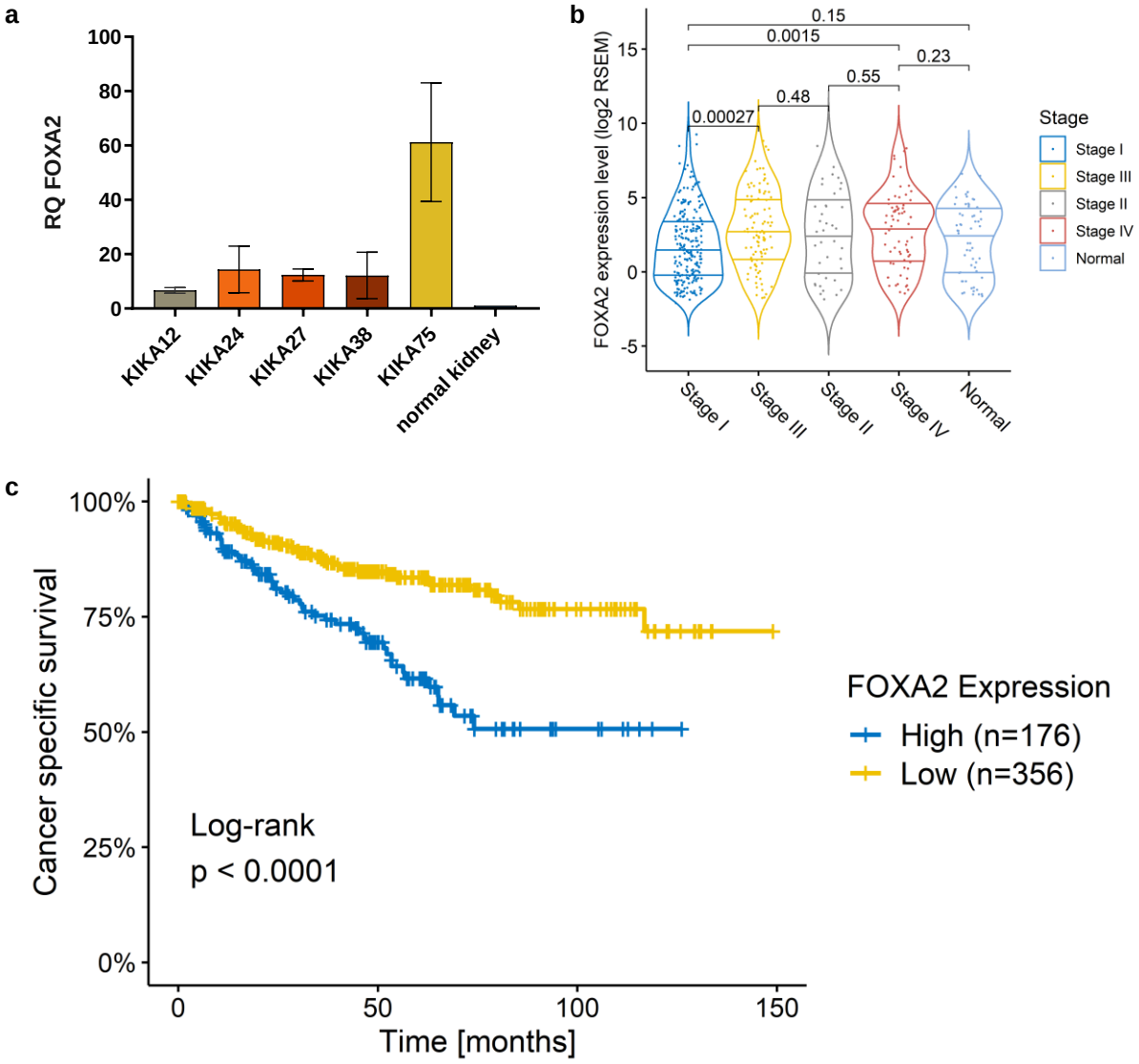


Figure 86 – *FOXA2* is highly expressed in KIKA cell lines and aggressive ccRCCs

(a) FOXA2 expression levels of KIKA models were determined by qRT-PCR and normalized to normal kidney. Error bars depict mean ± 95 %-CI**.**

(b) RNASeq log2 RSEM expression values of *FOXA2* in the TCGA KIRC patient dataset over the stages. *FOXA2* expression shows no difference between bulk tumor tissue and normal kidney tissue, but FOXA2 is significantly higher expressed between stage I and the other stages. *P* value was calculated using a two-sided Student's *t*-test. Error bars depict mean ± 1.5*IQR.

(c) Kaplan Meier estimators of patients' survival, stratified by the expression of *FOXA2*. Stratification of patients according to *FOXA2* expression levels separates patient survival probabilities significantly (log rank test).

We established *FOXA2* knockdown in the KIKA27 cell line by transducing two shRNAs simultaneously. Thereby, we achieved a knockdown efficiency of 77 % (**Figure 86a**). In a next step, a *FOXA2* overexpressing cell line was generated. Even though *FOXA2* has relatively high basal expression levels, the vector showed strong leakiness (**Figure 87b**). By inducing the vector, we observed a relatively to the not induced cell line 105-fold overexpression of *FOXA2*. However, the cells started to detach from the surface and became apoptotic (data not shown). Surprisingly, *TSPAN8* expression was dependent on *FOXA2* expression: Knockdown of *FOXA2* led to a reduction of *TSPAN8* levels (60.3 %) and overexpression to an increase of *TSPAN8* levels (6.6-fold). The increased basal levels of *FOXA2* due to leakiness of the vector had no influence on *TSPAN8* levels.

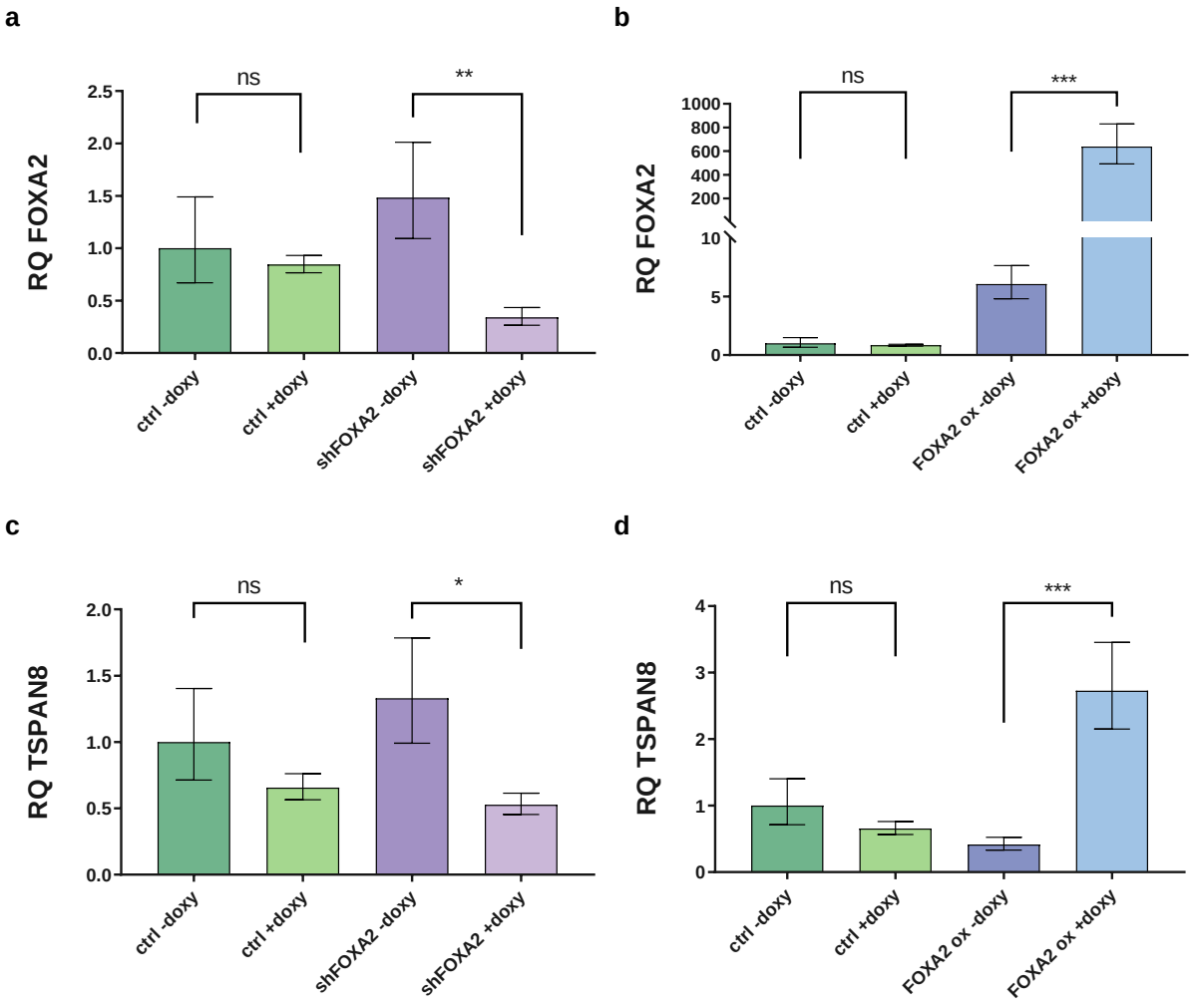


Figure 87 – Expression levels of *FOXA2* in the KIKA27 cell line correlate with *TSPAN8* expression levels

(a) Knockdown with a combination of two shRNAs and **(b)** overexpression efficiency of *FOXA2* was estimated by qRT-PCR. Expression levels of *TSPAN8* significantly change upon FOXA2 **(c)** knockdown and **(d)** overexpression. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns = not significant. Error bars depict mean ± 95 %-CI

The question arose whether FOXA2 regulates the expression of *TSPAN8* directly. We analyzed the chromosomal region of *TSPAN8* and corresponding Chip-Seq data, which the ENCODE project has collected for the HepG2 cell line and liver tissue (**Figure 88a**). Indeed, FOXA2 binds the transcription start site of *TSPAN8* in these tissues. With the help of the Methylation Plotter[298], the beta values of the CpGs covered by the Epic methylation array were plotted for the genomic *TSPAN8* region[299,300]. Interestingly, these FOXA2 binding sites were also differentially methylated between the first and the last passage of the *in vivo* selection (**Figure 88b**).

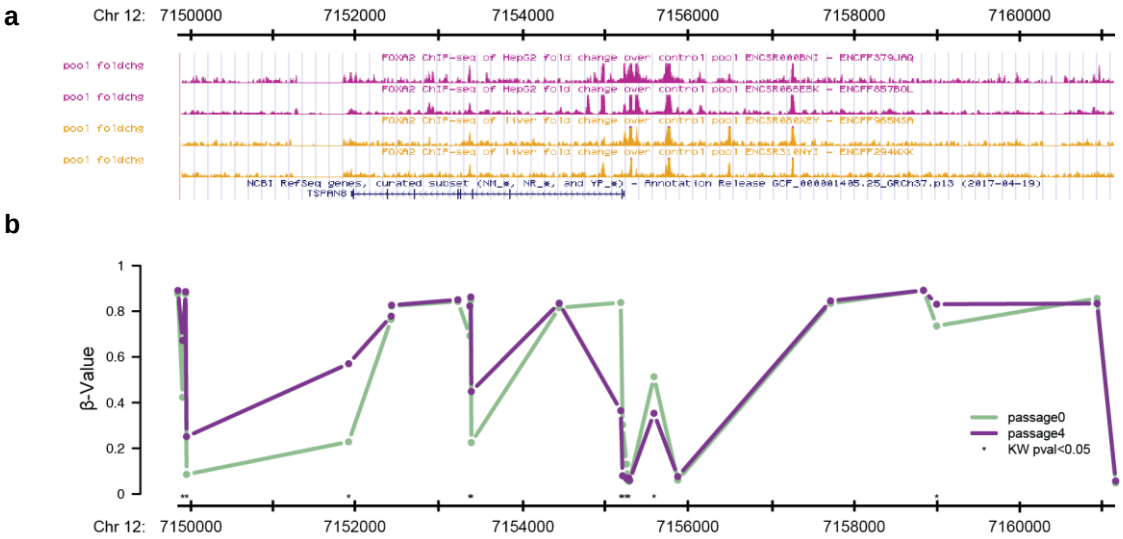


Figure 88 – The genomic Region of TSPAN8 is differentially methylated

(a) FOXA2 Chip-Seq binding sites in HepG2 and liver for the chromosomal region of TSPAN8 (Chr12:71498410-71610830). We accessed the Encode Chip-Seq datasets (ENCFF379JAQ, ENCFF857BOL, ENCFF965MSA, ENCFF294WXK) via the UCSC Genome Browser (https://www.encodeproject.org/)

(b) The β-Values of the CpGs for the chromosomal region of TSPAN8 covered by the epic methylation array were plotted by the Methylation Plotter. Methylation means for each group and for each CpGs of the genomic TSPAN8 region (Chromosome 12q21.1) are shown in a methylation profile plot. Significant differentially methylated positions (calculated with the Kruskal-Wallis test) are indicated by asterisks.

To investigate the effect of FOXA2 knockdown *in vivo*, we xenotransplanted the *FOXA2* knockdown KIKA27 cell line and followed tumor growth. Doxycycline induced *FOXA2* knockdown in KIKA27 derived tumors delayed tumor growth significantly in comparison to normal *FOXA2* levels (**Figure 89**).

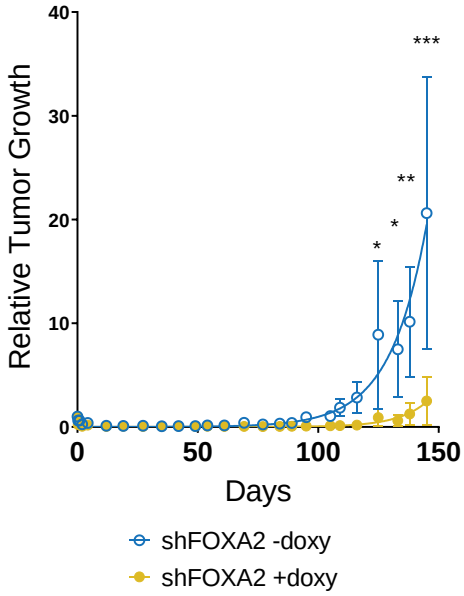


Figure 89 – In vivo KIKA27 derived tumor growth with FOXA2 knockdown

Knockdown of FOXA2 significantly delays tumor growth. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns = not significant. Error bars depict mean ± s.e.m.

4 Discussion

4.1 The ccRCC Score signature

We developed a LASSO regression based ccRCC signature that surpasses the survival predictions of all other published gene expression based signatures for ccRCC (**Figure 55 & Figure 56**) and is capable of identifying high risk patients in early stage tumors (**Figure 62**).

The advantage of our signature over other published signatures is the inclusion of clinical parameters into the regression (**Equation (12)**). The clinical stage or the SSIGN-score alone showed better prediction power than all signatures published that are based on gene expression data (**Figure 55 & Figure 56**). Including clinical variables such as stage into a regression model improves the predictive power. Unfortunately, we could not reproduce the signatures presented by Wang *et al.*[301] and Buttner *et al.*[302,303]. Neither did their studies report the classification of each patient in the TCGA-KIRC cohort, nor could we reproduce the classification according to their publications. Therefore, we were unable to assess the power of our ccRCC score in comparison to their patient classification method.

We hypothesize that our *in* vivo model selects for genes that promote a more aggressive phenotype with enhanced capacity to metastasize (**Figure 20**). As we were especially interested in clinical relevant processes, mediating tumor outgrow, we correlated differentially expressed genes of this selection with processes that occur between stage I and stage III of ccRCCs.

We included all clinical covariates into the LASSO regression that were annotated in both the TCGA-KIRC training and the Sato. *et al.* test cohort, and which had at least a univariate *P* value for the hazard ratio below the threshold of 0.2 (**Figure 30**). As tumor stage did not correlate linearly with the cancer specific relative hazard, we estimated its hazard using binary predictors (**Figure 34c**). The in the TCGA-KIRC and Sato *et al.* dataset reported four-tiered tumor grade is outdated, therefore, we used the WHO-ISUP grading system, which combines the old grade I and II patients into a single group[304]. Interestingly, age is not a significant predictor for cancer specific survival, but for non-cancer related survival (**Figure 31a,c**). This is contrary to the fact that ccRCC is an age related disease (**Figure 2a**). Nonetheless, most studies only present overall survival, not taking into account that age itself is a life limiting factor. Moreover, we found that laterality of the tumor is predictive for patients' survival (**Figure 30**). Chandrasekar *et al.* analyzed a large patient cohort of more than 50,000 RCC patients[13] and could not observe a significant correlation of tumor laterality and hazard ratio, yet he saw a low but significant correlation with age. This argues that our dataset was not

large enough to estimate linear predictors with minimal influence on the cancer specific survival. However, variables with little influence on the survival were not selected by the parameter shrinkage of the LASSO regression.

The final ccRCC model that had been selected by the LASSO regression on stage I to stage III patients (**Figure 33**) was extended to stage IV patients (**Figure 41**). It comprises of seven genes and the clinical covariates stage I and stage IV as binary classifiers. *CXCL5, IL13RA2, ONECUT2* and *TSPAN8* have a positive weight in the ccRCC Score, whereas expression of *PRDX2, PTPN13, SLC2A9* reduce the score. In the following, the genes their role in cancer progression will be shortly introduced.

The C-X-C Motif Chemokine Ligand 5 (CXCL5) is a chemokine that mediates chemotaxis of neutrophils[305]. Several roles of neutrophils in tumor development have been proposed, among them the production of matrix metalloproteases enabling tumor cell migration and invasion[306,307]. Recently, the role of neutrophils in the progression of cancer has been reported to lie additionally in the escort of circulating tumor cells facilitating thereby metastasis[308].

Interleukin 13 receptor alpha 2 (IL13RA2) has been associated with acquired sunitinib resistance in ccRCC by inhibiting apoptosis[309]. IL13RA2 binds IL-13 with very high affinity but has no signaling domain[310]. Thereby, it was suggested that it might act as a decoy receptor[311]. It has been reported that NKT cells act with the help of IL-13 as tumor suppressors by inducing cell differentiation[312].

Studies reporting about the role of the transcription factor One cut homeobox 2 (ONECUT2) in cancer progression are limited, but it has been shown to drive neuroendocrine tumors[313-315]. In those tumors, it induces the expression of a survival program and supports metastases[316,317].

Tetraspanin-8 (TSPAN8) is a member of the transmembrane 4 superfamily and has been associated with tumor progression and metastasis[292,318-321]. TSPAN8 will be further discussed in **section 4.2**.

The tumor repressor Peroxiredoxin 2 (PRDX2) reduces the reactive oxygen species (ROS) alkyl hydroperoxides and hydrogen peroxide[322]. High PRDX2-levels protect the cell from high ROS levels and from ROS-induced mutations like DNA double strand breaks, which impair the genomic stability[323,324].

The Protein Tyrosine Phosphatase, Non-Receptor Type 13 (PTPN13) dephosphorylates Fas receptor, IκB, PTEN, Ephrin B and has been shown to regulate Rho signaling pathways[325-327]. Thereby, it can act as a tumor suppressor by inactivating the Src tyrosine kinase[328].

Additionally, it has been shown to promote apoptosis via insulin receptor substrate 1 dephosphorylation and thereby inactivating insulin-like growth factor signaling via the Akt pathway[329].

The Solute Carrier Family 2 Member 9 (SLC2A9) is a glucose/fructose/uric acid transporter, which is physiological expressed in the proximal tubules of the kidney[330-332]. Functioning as an antioxidant, uric acid can reduce oxidative stress when its intracellular levels are elevated upon SLC2A9 expression[333]. Therefore, SLC2A9 has been identified as a tumor suppressor in hepatocellular and prostate carcinoma[334,335].

The described functions of *CXCL5, IL13RA2, ONECUT2, TSPAN8* as pro-tumorigenic genes and *PRDX2, PTPN13, SLC2A9* as tumor suppressors can be correlated to the survival of ccRCC patients. On the one hand, the LASSO regression on stage I-III selected for pro-tumorigenic genes with a high hazard ratio. These genes got upregulated during the *in vivo* selection (**Figure 38**). On the other hand, it selected for genes with a hazard ratio below one, which got downregulated in the *in vivo* selection. Hence, using the LASSO regression approach, clinically relevant genes were selected that might be new and interesting clear cell renal cell carcinoma targets.

Tumor necrosis is a covariate of the SSIGN Score (see **section 1.3.1.3**) and was the only significant clinical parameter in a multivariate analysis with the ccRCC score (**Figure 44**). Unfortunately, the patients of the Sato *et al.* cohort had no tumor necrosis annotation. Including this variable into the final ccRCC score model might even improve the predictive power.

We calculated the ccRCC Scores for all patients of the TCGA-KIRC and thereby estimated cancer specific survival (**Figure 42a**). As expected from a training data set, the log hazard rate strongly correlated with the ccRCC Score. Additionally, not only cancer specific survival but also disease free and overall survival could be estimated from the ccRCC Score (**Figure 42b,c**). We validated the training cohort with an independent ccRCC RNASeq cohort, published by Sato and colleagues in 2013[31]. We observed the same linear correlation between ccRCC Score and survival (**Figure 47**), with the ccRCC Score as the only significant predictor for patients' survival in a multivariate analysis (**Figure 48**).

For everyday clinical situations, it is easier to classify patients into risk categories. Therefore, patients of the TCGA-KIRC cohort were stratified either by the median into two risk groups (ccRCC Score = -0.751) or by the lower (ccRCC Score = -1.700) and upper quartile (ccRCC Score = 0.449), resulting in three risk groups (**Figure 45**). We used the same ccRCC Scores in

order to classify patients into risk groups, independent of the cohort and source. Stratification into risk groups distributed patients of both the TCGA-KIRC and the Sato *et al.* dataset into very distinct groups (**Figure 45 & Figure 49**). This was also true for the estimation of progression free survival of the mini cohort from Heidelberg including 13 patients of which the gene expression profile was analyzed (**Figure 57**).

Stratification of patients into risk groups reduces the predictive power of the analysis[336]. We obtain higher C-indices and likelihood ratios when applying the continuous ccRCC Score than the three-tier or even two-tier risk stratification (**Figure 56**). A categorization into smaller intervals allowed us for more precise estimations of the 5-year survival (**Figure 46f**).

Intratumor heterogeneity is a clinical challenge in ccRCC. Gerlinger and his team were the first to show that within a single ccRCC tumor of a patient, multiple distinct different subclones coexist[72,73]. By sequencing, they were able to generate phylogenetic trees that represent the tumor evolution with driver mutations in the trunk of the tree and subclonal genomic events distinguishing the branches of the tree. The group classified 63 different regions of the ten patients with the help of the ccA/ccB signature[267,279]. Gulati *et al.* observed extensive intratumor heterogeneity when applying the ccA/ccB risk categorization. Only two tumors (EV005 and RMH008) had homogeneous risk scores for all tumor regions. In comparison, tumor regional risk classification by the ccRCC Score classified all but two patients (EV001 and EV006) into a homogeneous risk group (**Figure 58**). This is an advantage over the ccAccB signature as a single biopsy might already help to stratify patients into risk categories.

Nevertheless, when analyzing the subclonal regions with a continuous ccRCC Score, vast intratumor heterogeneous differences are observable (**Figure 59**), with differences in ccRCC Scores of up to 2.379 (EV001). Estimation of predicted 5-year survival probabilities for each region using the ccRCC Score (see above) revealed vast differences. The most benign tumor region R4a of patient EV001 has a good prognosis of 82.3 % 5-year survival probability. The region R3 however, which is appeared later in tumor evolution, indicates only a potential 5-year survival probability of 12.2 %. This demonstrates that risk stratification overly simplifies the complex intratumor heterogeneous structure of the tumor and a continuous risk predictor might map the system better. Therefore, in order to provide a well-founded approximation of patient risk using the ccRCC score, several tumor regions must be analyzed.

We were interested in identifying a relationship between evolutionary distance and tumor aggressiveness. Therefore, we plotted the number of non-synonymous mutations as a

measurement of evolution against the ccRCC Score (**Figure 60**). In six patients, increased mutational burden was associated with an increased ccRCC Score. One patient had no such correlation and three patients were invers correlated, among them the non-metastatic stage II patients.

Since the relationship between evolutionary distance of a tumor region with its aggressiveness could not be answered conclusively by the ten patients of the multiregional analysis, we plotted the number of non-synonymous mutations against the ccRCC Score of patients of the TCGA-KIRC cohort (**Figure 61**). We could not identify any correlation between ccRCC Score as measurement of tumor aggressiveness and evolutionary distance.

As stated before, we show that the ccRCC Score has superior predictive power over all ccRCC signatures that are based on gene expression data (**Figure 55 & Figure 56**). Nevertheless, as the risk classification by stage or SSIGN score is already good in predicting cancer specific survival we wondered, whether the ccRCC Score could give additional information over these classical clinical parameters. Hence, we calculated optimal cutoffs for the ccRCC Score within each stage category and stratified patients into high and low risk groups (**Figure 62**). Surprisingly, low risk patients within stage I to III had very good cancer specific survival probabilities between 94.8 and 100 %. The identification of a subpopulation of patients with such a good prognosis would help to determine the optimal need for follow-up after nephrectomy. So far, in Germany, clinical follow up is performed according to the stage and grade of the tumor, dividing patients into three risk groups[16]. The main difference between the risk groups is the recommended frequency of computerized tomography (CT) or magnetic resonance imaging (MRI) scans. Those examinations are not only expensive, but in the case of CT scans the radiation is also burdening the patient. A reduced amount of examinations would be beneficial for low risk patients, especially patients with stage III tumors, as they would have been grouped into the high risk group with up to eight scans within two years after nephrectomy. For low risk patients, the guidelines recommend only three to four scans in the same period. Patients of the high risk group within a stage category defined by the ccRCC Score could be the potential target group and benefit from an adjuvant therapy.

Patients of the low risk score groups of stage I to III show very good prognosis. We were wondering whether we might identify differentially expressed genes in the high risk group of patients that might be druggable. Hence, we made use of the DGIdb drug interaction database to search for already approved drugs that interact with genes upregulated in the high risk group of stage I patients (**Figure 63**). Thereby, we identified six upregulated genes that

interact with nine different drugs (**Table 9**), among them is the receptor tyrosine kinase Erb-B2 Receptor Tyrosine Kinase 4 (ERBB4 or HER4). Many patients benefit from targeting RTKs by multi-tyrosine kinase inhibitors such as sunitinib and sorafenib. Therefore, it could be of interest to investigate signaling through ERBB4 in ccRCC and whether ccRCC patients would benefit from targeting ERBB4 as well. Another possible target might be Beta-1,4-N-Acetyl-Galactosaminyltransferase 1 (B4GALNT1) catalyzing the biosynthesis of the disialoganglioside GD2, which was found to be a potential target to treat in neuroblastoma with the monoclonal antibody dinutuximab[337]. Hence, identifying upregulated genes in high risk stage I ccRCC patients as potential druggable targets represent another application by which patient stratification via the ccRCC Score could help to improve the understanding and therapy of ccRCC.

4.2 *TSPAN8* is an attractive surface marker to target ccRCC

Tetraspanin 8 (TSPAN8; also TM4SF3 or CO-029) is encoded by the *TSPAN8* gene and belongs to the transmembrane 4 superfamily (tetraspanin family)[292]. With short intracellular N- and C-termini, tetraspanins span the membrane four times **(Figure 90**). Of importance is the second extracellular loop, which is larger than the others. This loop consists of a constant region and of a variable region which bind other tetraspanins and other transmembrane proteins respectively, to form tetraspanin-enriched membrane microdomains (TEMs)[338,339]. TEMs are stabilized by cholesterol and gangliosides but differ structurally from lipid rafts[340].

TEMs are highly enriched for integrins (mostly α3β1, α4β1 and α6β1)[341], G-protein-coupled receptors[342], growth factor receptors[343], peptidases, immunoglobulin superfamily members[344,345] and other transmembrane proteins like EPCAM and CD44[340]. Thereby, TEMs can act as platforms for a variety of signaling pathways. The tetraspanin scaffold determines the function of the microdomaine based on the combination of binding partners rather than by tetraspanins directly[346].

In these TEMs, tetraspanins regulate cell adhesion, spreading, migration and via exosomes cell-cell communication. Exosomes are small vesicles (30-120nm) that are released from cells and fuse or bind selectively with target cells[347,348]. Inside the lipid bilayer of the exosomes enzymes, mRNA and microRNA can be transferred to the target cells and modulate their function[292,349,350]. Exosomes are highly enriched for tetraspanins but whether they have an influence on the exosome composition is unknown[351,352].

Expression of the tetraspanins CD82 and CD9 has been associated with tumor suppression[353], whereas CD151 and TSPAN8 seem to support tumor progression and metastasis[292,318-321].

TSPAN8 is significantly upregulated in many cancers, such as gastrointestinal[354-357], liver[358,359], ovary[360], pancreatic[361], prostate[362], esophagus cancer[363], glioma[364] and melanoma[365,366]. As it is the case for most tetraspanins, the pro-invasive and pro-metastatic role of TSPAN8 is mainly regulated via scaffolding of binding partners such as E-cadherin[356], claudin 7, EPCAM, α6β4 integrin, CD44v6 and EWIF[345,355,359]. It has been reported that TSPAN8 might modulate the tumor microenvironment with the help of exosomes, which might promote angiogenesis of endothelial cells in cooperation with CD151[321,354,367,368]. In addition, exosomes have been linked to matrix remodeling by activating matrix metalloproteases[367,369], which support tumor growth and invasion.

.

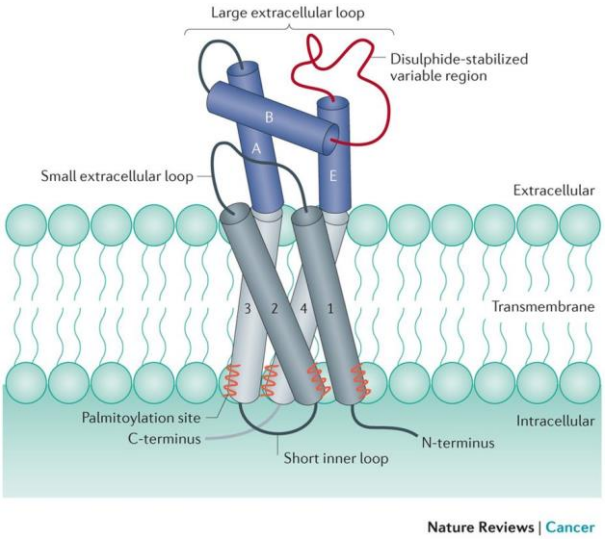


Figure 90 – Structure of Tetraspanins

Tetraspanins have short cytosolic C- and N-termini and cross the membrane four times with hydrophobic transmembrane domains. They are connected by loops of which the large extracellular loop is of importance, as it binds other tetraspanins with its constant region (labeled with A, B and E) to form a so called tetraspanin-web. With its variable region, tetraspanins bind other partners. Together they form tetraspanin-enriched microdomains (TEMs), which may act as signaling hubs.
With kind permission of Springer Nature, License Number 4518130548095

A proteomic analysis in colorectal cancer cell lines suggested that TSPAN8 together with the tetraspanin CD9 facilitates clustering of epidermal growth factor receptors (EGFRs) and CD44[370]. We find CD44 as one of the most abundant expressed surface molecules on the surface of all KIKA cell lines (**Figure 16b**). The hyaluronic acid binding surface molecule CD44 has been shown to activate EGFRs and thereby promotes cell survival and remodeling of the tumor microenvironment[371]. In ccRCC, the expression of CD44 has been correlated with highly aggressive tumor progression[372]. We observed EGFR expression on all KIKA cell lines, but with less pronounced intensity (**Supplemental Figure 2**). In renal carcinoma cell line experiments, the activation of EGFR signaling has been proposed as a potential resistance mechanism for Sunitinib treatments[373]. Nevertheless, treatment with the EGFR inhibitor Erlotinib alone did not show any effect on cell viability in our patient derived ccRCC cell lines (**Figure 18c**).

TSPAN8 is the most differentially regulated gene in the *in vivo* selection of the seven hallmark genes of the identified ccRCC signature (**Figure 28 & Figure 38**). Looking at

TSPAN8 expression it became increasingly upregulated along the passages (**Figure 64**). With the help of IHC, we observed intratumor heterogeneity of TSPAN8 positive cells (**Figure 65**). In the primary tumors, which derived from xenotransplantation of the originating KIKA75 cell line only few TSPAN8 positive cells were found, mostly clustering together. The tumors of the last in vivo selection stained highly positive for membranous TSPAN8. But still, we could observe areas with TSPAN8 positive cells and areas with no or only weak TSPAN8 expression. We did not observe any relationship of TSPAN8 positive cells to tumor boarder, vessels or extracellular matrix.

Also within our different KIKA cell lines, TSPAN8 was heterogeneously expressed (**Figure 67a,b**). This heterogeneous expression of TSPAN8 was also observable in the TCGA-KIRC patient cohort (**Figure 67c**).

Stratification of patients by the expression of TSPAN8 shows a significant overall survival benefit for patients with low *TSPAN8* expression in the Sato *et al.* patient cohort. Although, this survival benefit is not significant for cancer specific survival in the TCGA-KIRC patient cohort, the trend is pointing towards the same direction (**Figure 68**). In early stage patients of both ccRCC cohorts, *TSPAN8* expression levels have no prognostic value for patients' survival (**Figure 69**). In patients of the TCGA-KIRC cohort with stage III and IV, however, high TSPAN8 expression correlates with worse prognosis. For the Sato *et al.* cohort which includes only 25 stage III and IV patients, stratification into risk groups using TSPAN8 expression shows the same trend for patient prognosis, but is not significant. (**Figure 69**).

To investigate the role of TSPAN8 in the disease progression of ccRCC, we used two cell lines with high basal TSPAN8 expression (KIKA24 & KIKA38), a cell line with low (KIKA75) and a cell line with no basal TSPAN8 expression (KIKA27) on protein level (**Figure 67b**). For each of them, we generated two TSPAN8 knockdown and one TSPAN8 overexpression cell line (**Figure 71 & Figure 72**). Although the knockdown efficiency of the two shRNAs was relatively high on RNA level, on protein level, the knockdown in the cell lines with high basal TSPAN8 expression only reduced the amount of absolute protein abundancy on the cell surface. The vector used for the overexpression of *TSPAN8* was leaky and therefore, we observed upregulated basal *TSPAN8* levels in the not induced cell lines.

We subsequently transplanted these KIKA cell lines orthotopially into NSG mice (**Figure 73**). We saw no doxycycline induced effect on tumor growth when using a non-targeting shRNA control (**Figure 74a, Figure 75a, Figure 76a, and Figure 77a**). Also, an overexpression of TSPAN8 did not influence tumor aggressiveness or tumor growth (**Figure 78**). However, a

reduction of TSPAN8 stopped tumor growth of both the shTSPAN8(1) KIKA27 cell line (**Figure 75b**) and the shTSPAN8(2) KIKA75 cell line (**Figure 77c**). The shTSPAN8(2) KIKA27 (**Figure 75c**) cell line, the shTSPAN8(1) KIKA24 cell line (**Figure 74b**) and the shTSPAN8(1) KIKA75 cell line (**Figure 77b**) showed delayed tumor growth. Knockdown of *TSPAN8* had no effect on the shTSPAN8(2) KIKA24 (**Figure 74c**) and on both of the shTSPAN8 KIKA38 cell lines (**Figure 76b,c**). By inducing TSPAN8 knockdown on already established KIKA27 tumors (**Figure 79**), we could stop (**Figure 80b**) or delay (**Figure 80d**) tumor growth.

Our study is the first to show the effect of TSPAN8 *in vivo* on primary patient xenografts. The data so far suggest that low amounts of TSPAN8 on the surface of tumor cells may be sufficient to mediate tumor aggressiveness *in vivo*, but a loss of protein stops or delays tumor growth.

In mouse xenograft models with commercial cell lines, TSPAN8 has already been identified as an attractive target molecule to target colorectal and ovarian cancer[288,360,374]. For this, monoclonal antibodies have been developed and especially the Ts29.2 mAB generated in the lab of Claude Boucheix proved highly effective[288]. They could show that intraperitoneal treatment of subcutaneous transplanted colorectal cancer cell lines with the antibody alone delayed primary tumor growth significantly, if treated early. These results recapitulate the effect of TSPAN8 knockdown that we observed for ccRCC patient xenografts. In contrast to previously published data, Ailane *et al.* did not observe differences of cell proliferation *in vitro* and could not observe differences in vessel formation. Aurélie Maisonial-Besset and her team coupled the same antibody with a radioactive ^{111}In or ^{177}Lu and treated nude mice xenografted with colorectal cancer cell lines. This treatment resulted in inhibited tumor growth using 100-fold reduced amounts of antibodies in comparison to treatment with an uncoupled antibody[375]. Another antibody targeting *TSPAN8* has been developed by Park *et al.*[360]. Epithelial ovarian cancer (EOC) cell lines were intraperitoneally injected into BALB/c-nude mice and simultaneously treated intravenously with their antibody, which resulted in a reduced number of metastases. Consistent with the study by Ailane *et al*., *in vitro* treatment of cell lines with the antibody by Park *et al.* did not influence cellular viability *in vitro* nor endothelial cell activation *in vivo*. However, they observed that treatment of tumor cells lead to an increased internalization of TSPAN8.

In accordance with the published data for the *in vitro* function of TSPAN8 (**Figure 83**), we observed no effect on cell proliferation[288,360,363,365]. This is also supported by the fact that we

could not observe any differentially expressed genes between overexpression or knockdown of TSPAN8 in vitro (**Figure 81 & Figure 82**).

Surprisingly, normal bulk kidney samples showed higher *TSPAN8* expression than ccRCC tumor tissue (**Figure 67c & Figure 69**). In normal kidney tissue, TSPAN8 is mainly expressed in the endothelium of the vasculature and in the apical membrane of cells of the proximal tubules of the kidney, the proposed ccRCC cell of origin[376,377] (see **section 1.2.1**). Therefore, high basal TSPAN8 expression in normal tissue may be an obstacle for a therapeutic use of mABs targeting TSPAN8. When treating mice with radiolabeled α-TSPAN8 antibodies, Bonnet *et al.* saw off-target binding to spleen, liver, lung, heart and kidney tissue, which show basal TSPAN8 expression in normal tissue[374]. Additionally, the use of radioactive labeled antibodies may lead to excessive blood radiotoxicity, as TSPAN8-rich exosomes are secreted from tumor tissue into the circulation[351,352]. Interestingly, the *Tspan8* knockout mouse shows only a weak phenotype with reduced male body weight and reduced bone mineral density[378]

TSPAN8 is a scaffold protein that forms tetraspanin-enriched microdomain clusters with cell-matrix interaction proteins. Studies suggest that TSPAN8 may influence invasiveness of tumor cells into the microenvironment by interacting and crosslinking cell surface molecules[363,365]. Especially its role in inhibition of E-cadherin[356] and in clustering with CD44[355] have been previously reported as essential regulators of tumor invasiveness, EMT and metastasis[371,379]. By elucidating the interaction network of TSPAN8, future studies will eventually decipher a mechanism, by which TSPAN8 influences tumor aggressiveness.

TSPAN8 has been identified as an early marker for kidney injury that modulates migration and invasiveness of renal tubule cells to repair renal tissue[380]. Acute kidney injury is a complex and multifactorial disease, but it is mainly caused by exposure to toxins or ischemic stress[381]. The injury leads to an acute reduction and malfunction of the glomerular filtration, followed by hypoxia and inflammation of the surrounding tissue. In contrast to the glomerulus as the primary filtration unit of the kidney, the renal tubules can regenerate after injury. Upon kidney injury, a cellular repair program is activated[382]. From uninjured proximal tubule cells, so called scattered tubular cells (STCs) emerge in order to facilitate tubular regeneration[383]. These cells express de novo CD44 [384], CD24 [385], vimentin and CD133 [386], which are markers of which we also found some upregulated in the KIKA ccRCC cell lines (**Figure 16**). Subsequently, the tubules regenerate and kidney function is slowly reestablished[381]. Patients with acute kidney injury may potentially develop a chronic kidney disease and subsequently renal cell carcinoma[387,388]. Hirukawa *et al.* identified TSPAN8 as a marker for acute kidney

injury, whereas other tetraspanins like CD151 and CD9 did not change their expression[380]. The authors observed that *TSPAN8* expression was independent of spatial proximity to the injury and suggested a systemic activation, which is in line with our observation when characterizing TSPAN8 expression patterns (**Figure 65**). Moreover, another study by Penas *et al.* found that tetraspanins were involved in regulating wound healing of normal keratinocytes *in vitro*[389]. Uncontrolled wound healing has often been proposed as a mechanism to promote cancer progression[390]. Many aspects of wound healings resemble aspects of tumor progression. In healthy tissue, the re-epithelialization is self-limiting as soon as the wound is closed, whereas tumor cells lose their control over proliferation and migration. The kidney injury molecule-1 (KIM-1) is highly upregulated in STCs and during kidney repair processes. Ghislaine Scelo and her group investigated into whether blood levels of KIM-1 is a predictive marker for kidney cancer[391]. They could show that high plasma levels correlate with the incidence of RCC even 5 years before cancer diagnosis.

In the progression of kidney cancer, TSPAN8 might support invasiveness of tumor cells by forming TEM signaling hubs and activate processes that resemble uncontrolled wound healing processes.

4.3 FOXA2 is a potential transcriptional regulator of *TSPAN8*

Agaësse *et al.* have shown that in cutaneous melanoma, the tumor suppressor tumor protein 53 (TP53) is a transcriptional repressor of *TSPAN8*. In contrast to many other tumors, *TP53* is mutated in only 2.8 % of the ccRCC patients of the ccRCC patients[31,63] and both gene and protein levels did not correlate with TSPAN8 expression (**Figure 84**).

We could show that FOXA2 target genes were enriched in the *in vivo* selection for FOXA2 target genes (**Figure 85**), even if the transcription factor itself was not differentially expressed. The helix-turn-helix transcription factor FOXA2 or HNF3β is a member of the forkhead box protein family[392]. Their activity as transcriptional activators of genes in dense chromatin regions has been associated with many processes, such as cellular growth, proliferation and differentiation[392-395]. Forkhead box proteins play an essential role in the embryonic development and are therefore prone to play a role in cancer development[392,396,397]. In the developing embryo for example, FOXA2 is essential to form the notochord that regulates the dorsoventral pattern of somites and neural tube[398,399].

The role of FOXA2 in cancer development is ambivalent. On the one hand it has been shown to act as a tumor and metastasis suppressor in liver[400-403] and pancreatic cancer[404]. On the

other hand, it promotes neuroendocrine tumors of the prostate[405], drives uterine carcinomas[296] and ovarian cancer[295], where it is required for the maintenance of cancer stem cells[406].

Except for the Wilms tumor, which is a kidney tumor distinct from renal cell carcinoma and ccRCC, where *FOXA2* expression was detected[407], little is known about its function in the carcinogenesis in other subtypes of kidney cancer. As described in **section 1.2.1**, the HIF-pathway is essential for ccRCC development. A HIF-FOXA2 axis has been found to promote neuroendocrine prostate tumors, it might also play a role in the tumorigenesis of ccRCC[408]. CNAs with a gain of *FOXA2* have been reported in late grade ccRCC[409] and Jia *et al.* associated high *FOXA2* expression with bad overall survival prognosis when stratifying patients by median *FOXA2* levels[297]. We analyzed the more appropriate cancer specific survival statistics (**see section 3.6.1**) and could confirm their assumption when applying maximally selected rank statistics to find an optimal FOXA2 level to stratify patients (**Figure 86c**). Later stages of ccRCC show significantly higher FOXA2 expression levels (**Figure 86b**). This finding is in line with the high basal expression levels of *FOXA2* in the KIKA cell lines (**Figure 86a**), which were derived from late patient stages[1].

As overexpression of FOXA2 showed deleterious effects in the cells, KIKA27 *FOXA2* knockdown cell line were orthotopically transplanted into NSG mice and tumor growth monitored. Knockdown of *FOXA2* led to a significantly delayed tumor growth in mice (**Figure 89**). This result indicates that FOXA2 regulated gene expression has an effect on KIKA27 derived tumor progression. Nevertheless, whether this is a cell line intrinsic effect or plays a global role in ccRCC has to be validated in more cell lines in future studies.

FOXA2 binds the promotor of *TSPAN8* in HepG2 and liver hepatocytes[299,300,410] (**Figure 88a**) and therefore we wondered whether differential *FOXA2* levels influence also *TSPAN8* levels in our ccRCC cell lines. By inducing a *FOXA2* knockdown in the *TSPAN8* low expressing KIKA27 cell line, we observed a reduction of *TSPAN8* levels according to *FOXA2* levels (**Figure 87a,c**). Overexpression of *FOXA2* was leaky resulting in increased *FOXA2* basal levels without induction of the overexpression construct (**Figure 87a**). This slight overexpression had no effect on *TSPAN8* levels, but the extensive overexpression of *FOXA2* also increased *TSPAN8* levels significantly (**Figure 87c**).

Cytosine methylation to 5-methylcytosine of CpG dinucleotides is a common epigenetic process in vertebrates in order to regulate gene expression, with hypomethylated promotors as being active and hypermethylated promkotors rendering the gene expression inactive[411-413]. In cancer, tumor suppressor genes are commonly hypermethylated, whereas oncogenes tend to

be hypomethylated[414]. *TSPAN8* has been shown to be epigenetically regulated[415] and cell lines derived from the lung metastases of the last passage of the *in vivo* selection slowly downregulated *TSPAN8* in cell culture conditions (see **Figure 66**), which is in accordance with reported demethylation and therefore epigenetic regulation kinetics[416].

We did not analyze the differentially methylated regions between the different passages of the passage 4 lung metastases derived cell lines, but we could show that the genomic region of *TSPAN8* was highly methylated in the first passage of the *in vivo* selection and gets demethylated during the selection process (**Figure 88b**). The epic array covers about 850.000 CpGs, including ENCODE open chromatin and enhancer regions[417]. Among them are those bound by the transcription factor FOXA2. We observed that the promoter regions of *TSPAN8* that have FOXA2 consensus sequences and are bound in hepatic cell lines by FOXA2 were hypomethylated in the last passage of the in vivo section.

In summary, we identified a simple signature that can predict patients' survival robustly. Tetraspanin-8 is among the genes of the signature playing an essential role in tumor progression *in vivo*. Our data indicates that *TSPAN8* is regulated by a differentially methylated promoter which is bound by the transcription factor FOXA2. Future studies may examine the transcriptional regulation and mechanism of TSPAN8 in the progression of clear cell renal cell carcinoma.

5 Material and Methods

5.1 Primary Patient Material

Primary patient clear cell renal cell carcinomas were obtained from PD Dr. Sascha Pahernik from the Department of Urology (University Clinic Heidelberg/Nürnberg). The study was approved by the ethical committee of the University of Heidelberg (case number 207/2005) and conducted in accordance with the Helsinki Declaration. A written informed consent was obtained from every patient. Tumor material was classified as clear cell renal cell carcinoma by the pathology department and during xenotransplantation morphological properties of the parental patient tumors were preserved. Human origin of the xenografts was verified by hKi67 staining. All tumors were highly vascularized, pervaded by a branched blood vessel network

5.2 Xenograft mouse model

Animal experiments were previously approved by the national authorities (Regierungspräsidium Karlsruhe; authorization number G-233/11 and G28/17) and studies were conducted according to the GV-SOLAS regulations. Mice were housed and bred in the DKFZ animal facility under specific pathogen free (SPF) conditions and maintained in individually ventilated cages (IVCs). Xenograft mouse experiments were carried out in NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl} (NSG) mice.

5.3 Cell Culture of primary cell lines

Primary patient derived cell lines were cultured in Renal CSC medium at 37°C and 5 % CO_2 (**Table 10**). The medium was partly replaced twice a week and cells were splitted in a ratio of 1:2 to 1:8 when cell density exceeded the culture flasks. Passaging was performed by dissociating the cells with Accutase (Thermo Scientific) and replating according to **Table 11**. For cryopreservation, cells were aliquoted, re-suspended in CryoStor CS10 (Sigma) and stored in liquid nitrogen.

Table 10 – Renal CSC Medium

Compound	Amount	Company
Advanced DMEM/F12	500 mL	Thermo Scientific
bFGF	50 ng/mL	PeproTech
BSA	1 %	Thermo Scientific
Epinephrine	1 µM	Sigma
Glucose	0.6 %	Sigma
Glutamine	2 mM	Sigma
H_2O	25 mL	Thermo Scientific
hEGF	20 ng/mL	PeproTech
Heparine	2 µg/mL	Sigma
Hepes	5 mM	Sigma
Hydrocortisone	0.2 µg/mL	Sigma
L-glutathione	1 µg/mL	Sigma
Lipid Mix	1 mL	Sigma
LONG R3 IGF-I	10 ng/mL	Sigma
N2 Supplement	5 mL	Thermo Scientific
Trace Elements A	250 µL	Mediatech
Trace Elements B	500 µL	Mediatech
Trace Elements C	500 µL	Mediatech
Triiodothryonine	10 nM	Sigma
β-mercaptoethanol	100 µM	Thermo Scientific

Table 11 – Cell lines and their culture conditions

Cell Line	Culture Flask	Growth condition
KIKA12	Corning® Costar® Ultra-Low Attachment Flasks	Spheres
KIKA24	Corning® Costar® Ultra-Low Attachment Flasks	Spheres
KIKA27	Corning® CellBIND® Surface Flasks	Adherent
KIKA38	Corning® Costar® Ultra-Low Attachment Flasks	Spheres
KIKA75	Corning® Costar® Ultra-Low Attachment Flasks	Spheres

5.4 *In vitro* drug treatment

To obtain the drug response rates of cell lines *in vitro*, each cell line was seeded in three biological and four technical replicates for each drug concentration and control one day prior treatment (KIKA75, KIKA38: 7000 cells/well, Corning 96-well clear flat bottom ultra-low attachment plates; KIKA27: 7000 cells/well, Corning 96-well clear flat bottom CellBind

plates). Drugs and the solvent DMSO as control were titrated in four technical replicates each onto the cells. As positive control, staurosporine (Sigma, 1µM) was added to eight technical replicates per cell line and replicate. The cells were incubated for 48 hours (KIKA27) or 72 hours (KIKA38, KIKA75) in the incubator (37°C, 5 % CO_2). Subsequently, cell viability was measured with the Cell Titer-Blue Cell Viability Assay (Promega). CellTiter-Blue (20µl CTB per 200µl medium) was incubated for 4 hours in the incubator (37°C and 5 % CO_2) and measured with a fluorescent plate reader at 560(20)Ex/590(10)Em.

5.5 Tumor cell injection into the kidney capsule

20 minutes before the operation Carprofen s.c. (5mg/kg) was given to the mice. Mice were anesthetized by injecting ketamin (90 mg/kg) and xylazinhydrochloride (14.5 mg/kg) intraperitoneally and as soon as the pedal withdrawal reflex was absent, the animals were placed on heating pads. Eyes were covered with dexpanthenol cream to prevent dehydration. The right kidney was exposed with a small incision on the right flank and tumor cells were injected together with Matrigel (3 mg/ml, maximum 30 µl) into the kidney capsule. As postoperative pain treatment a carprofen-gel (10 mg/kg/day) was placed into the cage.

5.6 In vivo bioluminescence imaging

In order to monitor tumor growth, ccRCC primary patient derived KIKA models were transduced with a luciferase transgene (Renilla-luciferin 2-monooxygenase) and monitored weekly using the IVIS 200 imager system. For imaging, mice were injected intraperitoneally with D-Luciferine Firefly Potassium salt (15 mg/mL, 5 µl/g/mouse in PBS) 6 minutes prior measuring and anesthetized with isoflurane (4.5 % during initiation, 1-2 % to uphold anesthesia). Data was analyzed by the Living Image software.

5.7 In vivo doxycycline-inducible knockdown

The knockdown of genes via the pTRIPZ-vector system was induced by applying doxycycline to the animals via their drinking water (2 mg/mL)

5.8 Stopping criteria for animal experiments

The general condition of the animals is monitored daily. When animals suffer middle afflictions (in accordance with Annex VIII der RL 2010/63/EU) they were monitored twice a day and as soon as their health condition worsened (body weight loss of > 20 %, anorexia, apathy, signs of dehydration, fur anomalies, paralysis, scoliosis, shortness of breath, paleness, in accordance with GV-SOLAS) or after an observation period of more than one year, they were euthanized by cervical dislocation.

Measurement of tumor growth with bioluminescence allows identification of stopping criteria before the appearance of e.g. dyspnea. As stopping criteria a 1000-fold increase of the initial signaling was defined.

5.9 Tumor dissociation

5.9.1 Kidney tumor

The primary tumor was excised and cut into small pieces and transferred into gentleMACS C-tubes with CO2 independent medium (Thermo Fischer Scientific) supplemented with the human Tumor Dissociation Kit (Miltenyi Biotec). Dissociation of the tumor was performed using the "kidney" program of the gentleMACS following an incubation at 37°C using the MACSMix tube rotator for 15 minutes. The process was repeated until the tumor was completely dissociated, which was visually determined under the microscope.

5.9.2 Lung metastases

The lung metastases were excised by removing as much normal lung material as possible and transferred into gentleMACS C-tubes together with CO2 independent medium (ThermoFisher Scientific) supplemented with 2.5 mg/mL Collagenase D (Sigma), 1 mgP/mL Elastase (Worthington) and 100 µg/mL DNase (Sigma). For dissociation, the "lung" program of the gentleMACS was used following incubation at 37°C using the MACSMix tube rotator for 15 minutes. The process was repeated until the tumor was completely dissociated, which was visually determined under the microscope.

5.9.3 Generation of cell lines

The dissociation suspension was filtered through 100 µm and 70 µm strainers and lysed with ACK lysis buffer to remove erythrocytes. Tumor cells were cultured in Renal-CSC medium together with 5 µg/ml Fungizone (Thermo Scientific), 50 U/ml Penicillin, 50 µg/ml Streptomycin (Sigma) and Y-27632 (Sigma). After 24h, the medium was replaced by Renal-CSC medium.

5.10 Measurement of tumor size and metastases scoring

Tumor size was measured by measuring the diameter of the tumor with a ruler. Metastasis scoring was performed as followed: Score 0 = no lung metastases discernible by eye or histology of the median cut of the lung. Score 1 = micrometastasis detectable by microscopy (< 100 µm). Score 2 = metastases visible by eye or by microscopy (> 100µm). Score 3 = Multifocal metastasis >1 mm each.

5.11 Flow cytometry

For flow cytometry stainings, dissociated cells were resuspended in PBS (Sigma), supplemented with 1 % BSA (Thermo Fischer) and 2mM EDTA (Invitrogen), and incubated with antibodies or fluorophore-coupled antibodies (**Table 12**) for 30 minutes in the dark at 4°C. After staining, samples were washed twice with the staining buffer. If required, the cells were incubated with a secondary antibody for 15 minutes in the dark (**Table 13**) and subsequently washed twice. For analysis, cells were filtered and co-stained with 0.1 µg/mL DAPI (Thermo Fisher) and recorderd on the LSRForessa analyzer (BD Biosciences). Final analysis was performed using FlowJo v10 (Tree Star; Ashland, OR).

Table 12 – Primary Antibodies used for FACS

Name	Manufacturer (Cat. No.)	Dilution
APC-Rat IgG1 kappa Isotype (eBRG1)	Thermo Fisher (17-4301-82)	1:20
APC-Mouse IgG2b kappa Isotype (eBMG2b)	Thermo Fisher (17-4732-42)	1:20
APC Mouse Isotype Control	BD Biosciences (554681)	1:20
BB515 Mouse Anti-Human CD54 (ICAM1)	BD Biosciences (564685)	1:100
BB515 Mouse IgG1 kappa Isotype Control	BD Biosciences (564416)	1:100
hTSPAN8	Ts29.2, Boucheix *et al.*[288]	1µg/1*10^6 cells

Additional, the Human Cell Surface Marker Screening Panel (BD Lyoplate) was used following the manufacturer's recommendations to immunophenotype the KIKA cell lines. 25000 cells per antibody were used and counterstained with secondary antibodies (**Table 13**). Cells were co-stained with 0.1 µg/mL DAPI (Thermo Fisher) and recorderd on the Guava easyCyte System (Millipore). Final analysis was performed using FlowJo v10 (Tree Star; Ashland, OR).

Table 13 – Secondary Antibodies used for FACS

Name	Manufacturer (Cat. No.)	Dilution
F(ab')2 Donkey Anti-Mo IgG APC	eBiosciences	1:100
Goat Anti-Mouse IgG Human ads-PE	SouthernBiotech	1:300
Goat Anti-Rat IgG (H+L) Mouse ads-PE	SouthernBiotech	1:300

5.12 Hematoxilin and Eosin (HE) and Immunohistochemistry (IHC) staining

Hematoxilin and Eosin (HE) and immunohistochemistry (IHC) staining was performed by Vanessa Vogel and Ornella Kossi (Hi-STEM). Tumor samples were fixed at room temperature in 10 % formalin (Sigma) for at least 48h and dehydrated with increasing ethanol and xylene (Sigma) concentrations. Tumor samples were embedded in paraffin and cut for staining.

For IHC and HE staining, slides were deparaffinized, rehydrated by xylene and decreasing concentrations of ethanol.

HE staining was done either automatically using a tissue stainer or manually. For both methods, following protocol is applied. Slides were incubated in hematoxylin according to Mayer (Sigma), washed with water and subsequent staining with Eosine Y (Sigma). Staining was fixated by acetic acid followed by increasing concentrations of ethanol. Slides were covered with xylene based mounting medium (ThermoFisher) and cover slips.

For IHC, antigens were unmasked by proteolytic induced epitope retrieval (Proteinase K, 1.5 mAU/ml, Qiagen) or heat induced using a steam pot with citrate buffer of either pH 6.0 or pH 9.0. Primary antibodies were incubated 30 minutes at RT or overnight at 4°C, washed and endogenous peroxidase blocked by H_2O_2. Subsequently, the slides were incubated with the HRP dual link rabbit/mouse (Dako) polymer system for 20 minutes, washed and stained with the DAB chromogen (Dako kit). The slides were counterstained with hematoxylin (Sigma) and washed with water. Slides were covered with aqueous based mounting medium (Sigma) and cover slips.

Table 14 – Primary Antibodies used for IHC

Antigen	Manufacturer (Cat. No.)	Dilution
ICAM1	Sigma (HPA004877)	1:100
hKi67	Dako (M7240)	1:1000
hTSPAN8	Ts29.2, Boucheix *et al.*[288]	4 µg/mL

5.13 Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

RNA was extracted from both tumor tissue and cell lines with the help of the miRNeasy mini kit (Qiagen) and subsequently reverse-transcribed with the high capacity cDNA reverse-transcription kit (Applied Biosystems) according to the manufacturers protocol.

Triplicates of 10 ng of reverse transcribed RNA were analyzed by quantitative real-time polymerase chain reaction (qRT-PCR) using TaqMan probes (**Table 15**; Thermo Fischer Scientific, 6-carbofluorescein (FAM) labeled) and TaqMan™ Fast Advanced Master Mix (Thermo Fischer Scientific) in a 10 µl reaction according to the fast protocol using the Viia Real-Time PCR System (Applied Biosystems).

Table 15 – TaqMan probes used for qRT-PCR

TaqMan Probe	Assay ID
CXCL5	Hs01099660_g1
GAPDH	Hs02786624_g1
IL13RA2	Hs00152924_m1
LOC644936 (AC026410.1)	Hs04978470_g1
ONECUT2	Hs00191477_m1
PPIA	Hs04194521_s1
PRDX2	Hs00853603_s1
PTMA	Hs02339492_g1
PTPN13	Hs01106214_m1
SLC2A9	Hs01119178_m1
TSPAN8	Hs00610327_m1
ZMIZ1-AS1	Hs05005383_m1

The data was acquired and analyzed using the QuantStudio™ Real-Time PCR Software (Applied Biosystems). The ΔΔCT method was applied using GAPDH and PPIA as multiple endogenous controls. Normal kidney was used as positive control (Clonetech).

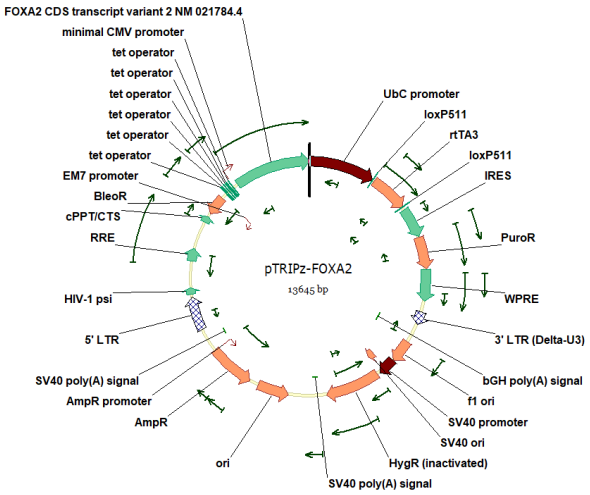
5.14 Gibson cloning of FOXA2 and TSPAN8 overexpression vectors

The pTRIPz backbone vector was linearized with the restriction enzymes AgeI and MluI and purified by gel extraction (QIAquick Gel Extraction Kit). FOXA2 transcript variant 1 (NM_021784.4) and TSPAN8 (NM_004616.2) mRNA was synthesized by GeneArt Strings (ThermoFisher). Overhangs for the Gibson assembly with AgeI and MluI restriction sites, stop codon and pTRIPz homologous regions were added by using following primers (Sigma) in a touchdown PCR (Q5 High-Fidelity DNA Polymerase, NEB):

Table 16 – Primer to anneal overhangs (small characters: homologous regions; capitals: overhang)

Primer	Primer Sequence
FOXA2 Fwd	tcgtttagtgaaccgtcagatcgcaaccggtATGCACTCGGCTTCCAGTATG
FOXA2 Rv	gcgccaaaacccggcgcggaggccaacgcgtTTAAGAGGAGTTCATAATGGGCC
TSPAN8 Fwd	tcgtttagtgaaccgtcagatcgcaaccggtATGGCAGGTGTGAGTGCC
TSPAN8 Rv	gcgccaaaacccggcgcggaggccaacgcgttcaTTTGTTCCCGATCTGGCAATAC

a



b

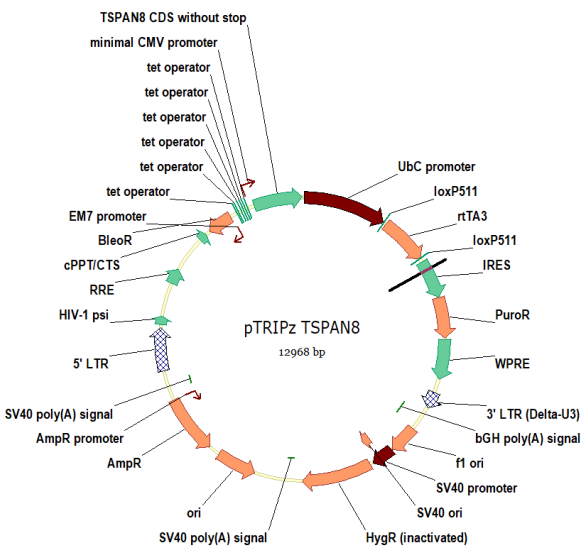


Figure 91 – Vector Maps of the cloned overexpression constructs

(a) pTRIPz-FOXA2 overexpression vector
(b) pTRIPz-TSPAN8 overexpression vector

Amplified DNA fragments were purified by gel extraction and subsequently cloned into the pTRIPz-backbone vector with the Gibson assembly master mix (NEB).

5.15 Gibson cloning of luciferase expression vector

The luciferase vector was cloned similarily as the overexpression vectors. The Lego-iCer2 backbone was linearized by BamH1 and BsrG1 digest and purified by gel extraction (QIAquick Gel Extraction Kit). The luc2 gene[418] was amplified per PCR from an in-house Lego-luc2-venus vector construct, the IRES sequence from the Lego-iCer2 vector and the zeocin resistance cassette from an in-house miRE_Ren713_stblTomatoEz_Zeo vector. Following primers were used to amplify the fragments together with homologous regions for the Gibson assembly:

Table 17 – Primer to anneal overhangs (small characters: homologous regions; capitals: overhang)

Primer	Primer Sequence
Luc2 Fw	gtcctccgattgactgagtcgcccgGATCCCAGTGTGGTGGTACGGGAATTCATGGAAGATG CCAAAAACATTAAGAAGGGC
Luc2 Rv	ttacgtagcggccGCGGCGCGCCGGCCCTCG
IRES Fw	tcgtttagtgaaccgtcagatcgcaaccggtATGGCAGGTGTGAGTGCC
IRES Rv	gccggcgcgccgcGGCCGCTACGTAAATTCC
ZeoR Fw	ataatatggccactcgagATGGCCAAGTTGACCAGTG
ZeoR Rv	gtgctggcggccggccgctttacttgcatgcTCAGTCCTGCTCCTCGGC

Amplified DNA fragments were purified by gel extraction and subsequently cloned into the pLego-backbone vector with the Gibson assembly master mix (NEB).

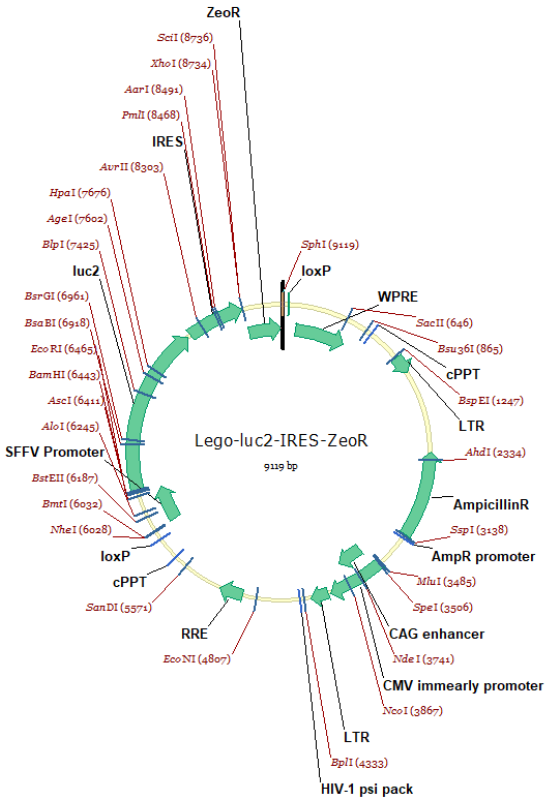


Figure 92 – Vector Maps of the cloned luciferase vector construct

5.16 Plasmid amplification

Constructs were transformed into one shot stbl3 chemically competent e.coli (Invitrogen) and grown on selective antibiotics agarose plates. Colonies were picked and grown in 3 mL LB-Cultures containing carbenicillin (37°C, 160 rpm, 8h). Plasmids were isolated using QIAprep Spin Miniprep Kit (Qiagen) and DNA sequence analyzed by Sanger sequencing (see **section 5.17**). Plasmids were further amplified by growing in 150 mL LB-cultures (37°C, 160 rpm, 14h) and subsequent isolation using QIAprep Spin Maxiprep Kit (Qiagen).

5.17 Sanger Sequencing

Plasmid DNA was isolated from stbl3 bacteria by the Plasmid Maxi Kit (QIAGEN) and sent for Sanger sequencing (SUPREMERUN, GATC, Eurofins) together with applicable primers (**Table 18**). Sequencing results were analyzed with ApE (v2.0.55, M. Wayne Davis, May 4, 2018).

Table 18 – Sequencing Primer list

Name	Direction	Primer
pTRIPz shRNA	fw	GGAAAGAATCAAGGAGG
pTRIPz EM7	rv	GGTCCGAGGTTCTAGACGAG
pTRIPz UBC	rv	GGGTTCTAAGGCCGAGTCTT
Lego	fw	AAAGAGCTCACAACCCCTCA
Luc2	fw	CGAGGTGCCTAAAGGACTGA
Luc2	rv	TCGATATGTGCGTCGGTAAA
IRES	rv	GGAAAGACCCCTAGGAATGC
ZeoR	fw	GTGGTCGGAGGTCGTGTC
ZeoR	rv	AAGTCGTCCTCCACGAAGTC
pTRIPz std	fw	GCCTCTTCGCTATTACGCCAG
pTRIPz std	rv	TTAGGCACCCCAGGCTTTAC
pTRIPz M13-21	fw	TGTAAAACGACGGCCAGT
pTRIPz M13-R	rv	CAGGAAACAGCTATGAC

5.18 Lentiviral knockdown and overexpression of target genes

To generate lentiviral particles, HEK-293T cells (ATCC) were cultured to a confluence of 60 % in T150 flasks (TPP) in IMDM medium supplemented with 10 % heat-inactivated FCS (Thermo Scientific). Cells were pre-treated 1h before transfection with Chloroquine (25 µM). With the help of the calcium-phosphate transfection method as described before[248], the packaging plasmid (psPAX2; 37.5 µg), the envelope plasmid (pMD2.G; 5 µg) and the lentiviral vector (50 µg, **Table 19**, **section 5.14, 5.15**) were transfected into HEK-293T cells. 24 h post-transfection, medium was replaced with the collection medium (IMDM, 10 % FCS, 1 mM Sodium Butyrate). After 24 to 36 h later, the supernatant containing the produced virus particles was collected, filtered (0.45 µm (Millipore) and ultra-centrifuged (SW32, 21 000 rpm, 2 h, 4°C; Beckman Coulter). The viral pellet was resuspended in 1/500th of the starting supernatant volume PBS and stored at -80°C.

The virus titer was determined by transduction of target cells in serial dilutions using 10 µg/mL polybrene (Sigma) together with 1 µg/mL doxycycline (Sigma) and flow cytometer analysis 72 h later. A virus concentration of 30% transduction efficiency was chosen and subsequently used for cell transduction. To select for positive transduced cells, puromycin (2 µg/mL; ThermoFisher) or Zeozine (100 µg/mL; ThermoFisher) was added to the medium and replaced together with the medium every other day. Untransduced cells, missing the resistance gene on the viral vector, were used as a selection control. Selection was performed until no live cells were detected in the selection control anymore. Selection efficiency was evaluated by inducing the cells with doxycycline 72 h and subsequent flow cytometry analysis of the RFP reporter protein.

Table 19 – shRNA Vectors

Vector	Short hairpin Sequence (5'-3')
pTRIPz shFOXA2 V2THS_86206	AAGAGGAGTTCATAATGGG
pTRIPz shFOXA2 V3THS_306420	TGAGGTCCATTTTGTGGGG
pTRIPz shTSPAN8 V2THS_42312	ATAGCTTTGGCATGGTCTC
pTRIPz shTSPAN8 V3THS_341934	AGATTTGAAAACAGCTCCT
pTRIPz non-targeting	CTTACTCTCGCCCAAGCGAGAG

5.19 Growth Curves

To obtain the growth rates of cell lines *in vitro*, each cell line was seeded in six technical replicates for each time point (KIKA75, KIKA24, KIKA38: 3000 cells/well, Corning 96-well clear flat bottom ultra-low attachment plates; KIKA27: 1000 cells/well, Corning 96-well clear flat bottom CellBind plates). Knockdown conditions were pretreated for three days with 1 µg/ml Doxycycline prior seeding. Cell viability as an indicator for cell number was measured with the Cell Titer-Blue Cell Viability Assay (Promega). CellTiter-Blue (20µl CTB per 200µl medium) was incubated for 4 h at 37°C and measured with a fluorescent plate reader at 560(20)Ex/590(10)Em.

Growth rates were calculated with the help of the growthcurver package[419] and growth rate differences tested for significance using the two-sided Student's *t*-test. With the help of mean regression factors, a combined growth model was calculated for each condition.

5.20 Statistical Tests

To compare unpaired normally distributed data, the Student's *t*-test was applied. If unpaired data was not normally distributed (not significant Shapiro-Wilk normality test) data was compared using the unpaired two-samples Wilcoxon test. If not stated otherwise, Pearson

correlation was used to test for correlation between unordered numeric covariates. To test for correlation between ordered covariates, the non-parametric tau test was calculated to estimate the Kendall rank correlation coefficients and subsequently plotted with the help of the *stats*[420] and *corrplot* package[421].

5.21 Statistical Software and Extensions

For statistical analysis, both GraphPad Prism 7.03 and R (3.5.1)[422] with *Bioconductor*[423] were used. Following R extensions were used: *readr* and *readxl* to import data into R[424,425], *ggplot2*[426], *ggrepel*[427] and *ggpubr*[428] for graphics, *dplyr* and *reshape2* for data transformation[429,430], *doParallel* and *foreach* for task parallelization and parallel computing[431], *knitr*[432], *BiocStyle*[433] and *Rmarkdown*[434] for rendering the R code into a html document.

5.22 Gene Expression profiling

Gene expression profiling of Genomic DNA free total RNA was performed at the Genomics and Proteomics DKFZ Core Facility (GPCF). RNA, extracted as stated above, with a RIN > 7 was spotted on Illumina HumanHT12v4 BeadChips and analyzed. Sample probe profiles, control probe profiles and probe feature data were exported from the raw bead array scans by using the Gene Expression extension of GenomeStudio V2011.1 (Illumina).

5.22.1 Import and Quality Control of Expression Profiling

Expression data was further analyzed using R (3.5.1)[422] and Bioconductor[423], following the *limma* pipeline[271]. In short: The background was corrected and the samples were quantile normalized by using the control probes of the Illumina BeadArray and subsequently log2 transformed[435] (**Supplemental Figure 5A-D**). To reduce the multiple testing problem[436], probe sets within the 0.5 quantile of the interquartile range (IQR) and duplicated entrez probes were removed[437]. The quality of this final dataset was analyzed to remove potential outliers[438]. Outliers were classified when samples were outliers in at least two of the following metrics: Distances between arrays, boxplots (Kolmogorov-Smirnov statistic K_a between each array's distribution and the distribution of the pooled data) and MA plots (Hoeffing's *D*-statistic). One sample was classified as outlier in at least two metrics and removed subsequently (**Supplemental Figure 5E-G**). Gene symbol annotation was updated using the BioMart database (column Illumina_HumanHT_12_v4)[439]. The final expression set was subsetted for differential gene expression analysis and standardized ($\text{z-score} = \frac{x-\mu}{\sigma}$) for supervised clustering and principal component analyses.

Background expression threshold was estimated by making use of the negative control probes on the arrays[440]. We defined probes to be expressed when their signal intensity was 10% above the 95 % percentile of the negative control probes.

5.22.2 Hierarchical Clustering

Clustering method used, if not stated otherwise, is Ward's minimum variance method[441,442] and the distance measured using the *L1* norm (absolute distance, between two vectors, also called Manhattan) to calculate the distance matrix.

5.22.3 Coefficient of Determination

To compare whole gene expression profiles of bulk lung metastases and kidney tumors with the respective cell lines derived from the lung metastases of passage 4, the coefficient of determination (R^2) was calculated by the square of the sample correlation coefficient of the linear regression. The mean of the coefficients of determination was calculated for each group and compared using Student's *t*-test (unpaired, two-sided).

5.22.4 Principal Component Analysis

To reduce the multidimensional gene expression profiles, principal component analysis was performed on the top 500 variant genes[443,444]. Only samples of interest were included in this analysis to maximize the explanations of variance between the samples.

5.22.5 Methylation Array

DNA Methylation status of 850 million probes was assessed using the Infinium MethylationEPIC BeadChip (Illumina). The Genomics and Proteomics DKFZ Core Facility (GPCF) treated the DNA with sodium bisulfite and performed the infinium assay. Raw data was imported and analyzed with *RnBeads*[445] and *minfi*[446]. Methylation profiles of seleced genomic regions were plotted with the Methylation plotter[298].

5.22.6 Differential Gene Expression Analysis

Differential gene expression analysis of the multifactor designed experiments was performed fitting linear models for each probe by the *limma* package[270,271]. Samples were group-means parametrized, contrasted for the comparison of interest and statistical analysis was performed by an empirical Bayes method.

5.23 ccRCC TCGA Dataset

Level 3 RNA-Seq mRNA expression data of 532 Clear Cell Renal Cell Carcinoma patients (TCGA, KIRC), RSEM quantified, was downloaded from cBioportal[447,448]. Counts were TMM normalized, preprocessed using the voom method[449] and standardized (z-score $= \frac{x-\mu}{\sigma}$).

Additionally, *z*-scored TP53 protein expression data, measured by reverse-phase protein array, was downloaded from cBioportal.

5.23.1 SSIGN Score

The composite Mayo SSIGN score was calculated for each patient according to the scoring system suggested by Frank *et al.*[142]. Tumor stage, size, grade and necrosis was obtained from the clinical annotation of the dataset. Not available data was interpreted as 0 for the respective category.

5.23.2 Survival Analysis

Survival probabilities were analyzed by estimate a survival function on patient survival data[251]. Time point of death because of the cancer ("VITAL_STATUS = "Dead" and "TUMOR_STATUS" = "WITH TUMOR") was recorded as an event and patients alive ("VITAL_STATUS" = "Alive") after the observation period or death of non-cancer reasons ("VITAL_STATUS = "Dead" and "TUMOR_STATUS" = "TUMOR FREE") were censored (right-censoring). Patients were assigned to a group, Kaplan Meier estimators calculated[450] and plotted with the help of the *survminer* R package[451]. The log rank test was performed to test for significant differences of Kaplan Meier Estimators[452]. Hazard ratios were calculated using Cox's proportional hazard model.

5.23.3 Maximally Selected Rank Statistics

With the help of the outcome-oriented *survminer*[451] and *maxstat* package[286], the cutpoint for stratification of low and high risk groups of patients was calculated.

5.24 LASSO Regression

LASSO regression in the context of Cox's proportional hazard model was performed on standardized gene expression data using the *coxnet* function of the *glmnet* R package[265,273].

To avoid overfitting, LASSO regression was performed in 10000 iterations with 70 % randomly selected patients of the cohort each. The λ-value with minimal cross-validated error in 10-folds for each regression was selected. Only factors with non-zero regression coefficients in more than 50 % of the LASSO-models were selected to calculate the beta coefficients for the ccRCC score. C-statistics were calculated to evaluate the fit of the regression[453,454]. Patient's scores were calculated using the ccRCC score as the weighted sum of standardized gene expression.

5.24.1 ODCF PBS Cluster

The iterations were calculated on the PBS cluster of the DKFZ Omics IT and Data Management Core Facility (ODCF). Each node of the cluster has up to 64 cores and a clock frequency of 2.0 - 2.7 Ghz. One node with 64 assigned cores and a maximum of 120 GB RAM was utilized to run the script and create the result file (loaded modules: pandoc/2.2.1 and R/3.5.0).

5.25 Validation ccRCC Dataset Sato *et al.*

The aligned 100 bp paired-end reads (Illumina HiSeq2000) of the Sato *et al.*[31] RNA Sequencing were downloaded from the European Genome-phenome Archive (EGA) with the accession number EGAS00001000509 (Data Set ID: EGAD00001000597). The expressions counts matrix was calculated following the workflow package *rnaseqGene* from Bioconductor[455]: The sorted bam files were imported in R[456] and paired reads associated to the UCSC hg19 gene model[457] for only those genes that are covered by the TCGA KIRC dataset using the *GenomicAlignment* package[458]. Counts were TMM normalized, preprocessed using the voom method[449] and standardized using the scaling factors used for the TCGA Dataset. All patient samples had complete clinical features and were used for the analysis[31].

5.26 Multivariate Analysis

To evaluate whether the ccRCC score is an independent prognostic variable, a proportional hazard model was calculated with all additional collected clinical relevant phenotypes that were present in > 90 % of the cohort. Model coefficients and statistics are plotted using the *ggforest* function of the *survminer* package[451].

5.27 Consensus NMF clustering

All prognostic signatures for ccRCC that are based on gene expression data were applied on both the TCGA-KIRC Dataset as well as the Sato *et al.* dataset as proposed by Gulati *et al.*[267] by unsupervised NMF clustering using the Broad Institute's GenePattern server (http://cloud.genepattern.org) with the NMFConsensus Module Version 5 and a cluster number from two to 10 in 2000 iterations[283].

5.28 C and log rank statistics

The rank correlation (concordance index) or the log rank statistics for the censored survival data was calculated using the *survival* package[459].

5.29 Intratumor heterogeneity of the ccRCC score – Subclonal analysis

Subclonal microarray expression data GSE31610 and GSE53000 were downloaded from the Gene Expression Omnibus (GEO) archive. Clinical data and non-synonymous mutations were accessed from the supplemental data of Gerlinger *et al.*[73]. Expression data was imported in R, gene annotation updated using the BioMart database (column affy_hugene_1_0_st_v1)[439] and expression data standardized as described before. The ccRCC Score and a corresponding color was calculated for each available subclonal region. Subsequently, the number of non-synonymous was plotted against the corresponding ccRCC score of each region. When more than one branch was present for a region (e.g. R3min and R3dom), both branches were colored according to the corresponding expression data (e.g. R3).

5.30 Gene Set Enrichment Analysis (GSEA)

GSEA was performed by using differential gene expression data ordered by logFC as pre-ranked list for the *fgsea* package[460].

5.31 Karyogram and Chromosome Painting

Cells were incubated 2h with vinblastine sulfate (2 µg/mL). Next, cells were spun down (5min, 150 g) and resuspended in a hypotonic solution (0.275 % KCl & 0.5 % NaCitrate). After 20 minute incubation at 37°C, cells were spun down (5 min, 150 g) and fixed using 10 mL of a 3:1 mixture of Methanol and Acetic Acid. Cells were incubated for 10 minutes in the fixative, spun down (5 min, 150 g) and resuspended in 2 mL fresh fixative. The cells were spread by dropping the fixated cell solution on microscope slides heated above a water bath (70°C). Nuclei were either stained with DAPI for a karyogram or chromosomes individually stained, performed by Dr. Larisa Savelyeva.

5.32 SNP Array

SNPs array hybridization was performed at the Genomics and Proteomics DKFZ Core Facility (GPCF). Genomic DNA, extracted as stated above, with a RIN > 7 was spotted on Illumina InfiniumOmni2-5-8v1-3 BeadChips and analyzed. Signal and feature data were exported from the raw bead array scans by using the Genotyping extension of GenomeStudio 2.0 (Illumina).

CNV analysis was performed by the CNV Analysis tool of GenomeStudio 2.0 and CNVs plotted by the *Circos* software[461].

5.33 Graphics Software

Adobe Illustrator CS6 and BioRender was used to design figures, Adobe Photoshop CS6 to resize histological images.

Bibliography

1 Watermeier, T. R. *A novel patient-derived renal cancer model platform enables the identification and functional characterization of tumor-initiating cells*, Ruperto-Carola University of Heidelberg, (2014).

2 Geist, F. *et al.* Abstract 1063: A clear cell renal cancer metastasis model identifies novel mediators of tumor aggressiveness and predictors of patient survival. *Cancer Research* **78**, 1063-1063, doi:10.1158/1538-7445.Am2018-1063 (2018).

3 Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **68**, 394-424 (2018).

4 Robert Koch-Institut. *Krebs in Deutschland für 2013/2014*. Vol. 11 (Robert Koch-Institut, 2017).

5 Jemal, A. *et al.* Global cancer statistics. *CA: a cancer journal for clinicians* **61**, 69-90 (2011).

6 Ahmad, A., Ormiston-Smith, N. & Sasieni, P. Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960. *British journal of cancer* **112**, 943 (2015).

7 Smittenaar, C., Petersen, K., Stewart, K. & Moitt, N. Cancer incidence and mortality projections in the UK until 2035. *British journal of cancer* **115**, 1147 (2016).

8 Lauby-Secretan, B. *et al.* Body fatness and cancer—viewpoint of the IARC Working Group. *New England Journal of Medicine* **375**, 794-798 (2016).

9 Cumberbatch, M. G., Rota, M., Catto, J. W. & La Vecchia, C. The role of tobacco smoke in bladder and kidney carcinogenesis: a comparison of exposures and meta-analysis of incidence and mortality risks. *European urology* **70**, 458-466 (2016).

10 Corrao, G., Scotti, L., Bagnardi, V. & Sega, R. Hypertension, antihypertensive therapy and renal-cell cancer: a meta-analysis. *Current drug safety* **2**, 125-133 (2007).

11 McClellan, S. *Kidney cancer rates are increasing, so what's fuelling the surge?*, <https://scienceblog.cancerresearchuk.org/2017/04/24/kidney-cancer-rates-are-increasing-so-whats-fuelling-the-surge/> (2017).

12 Sachdeva, K. J., Bagi; Curti, Brendan; Abel, E Jason. *Renal Cell Carcinoma - Physical Examination*, <https://emedicine.medscape.com/article/281340-clinical#b3> (2018).

13 Chandrasekar, T. *et al.* Metastatic renal cell carcinoma: Patterns and predictors of metastases-A contemporary population-based series. *Urologic oncology* **35**, 661 e667-661 e614, doi:10.1016/j.urolonc.2017.06.060 (2017).

14 Mai, K. T. *et al.* A comparative study of metastatic renal cell carcinoma with correlation to subtype and primary tumor. *Pathol Res Pract* **197**, 671-675, doi:10.1078/0344-0338-00144 (2001).

15 Neuzillet, Y., Lechevallier, E., Andre, M., Daniel, L. & Coulange, C. Accuracy and clinical role of fine needle percutaneous biopsy with computerized tomography guidance of small (less than 4.0 cm) renal masses. *The Journal of urology* **171**, 1802-1805 (2004).

16 Onkologie, Leitlinienprogramm (Langversion). (Deutsche Krebsgesellschaft e.V.,, 2017).

17 Rini, B. I., Campbell, S. C. & Escudier, B. Renal cell carcinoma. *Lancet* **373**, 1119-1132, doi:10.1016/S0140-6736(09)60229-4 (2009).

18 Eckel-Passow, J. E. *et al.* Assessing the clinical use of clear cell renal cell carcinoma molecular subtypes identified by RNA expression analysis. *Urologic oncology* **33**, 68.e17-23, doi:10.1016/j.urolonc.2014.07.019 (2015).

19 Sachdeva, K. J., Bagi; Curti, Brendan; Abel, E Jason. *Renal Cell Carcinoma - Histology*, <https://emedicine.medscape.com/article/281340-workup#c1> (2018).

20 Decastro, G. J. & McKiernan, J. M. Epidemiology, clinical staging, and presentation of renal cell carcinoma. *Urol Clin North Am* **35**, 581-592, doi:10.1016/j.ucl.2008.07.005 (2008).

21 Gebhard, R. *et al.* Abnormal cholesterol metabolism in renal clear cell carcinoma. *Journal of lipid research* **28**, 1177-1184 (1987).

22 Grignon, D. J. & Che, M. Clear cell renal cell carcinoma. *Clin Lab Med* **25**, 305-316, doi:10.1016/j.cll.2005.01.012 (2005).

23 Chevrier, S. *et al.* An Immune Atlas of Clear Cell Renal Cell Carcinoma. *Cell* **169**, 736-749 e718, doi:10.1016/j.cell.2017.04.016 (2017).

24 Yamashita, Y., Honda, S., Nishiharu, T., Urata, J. & Takahashi, M. Detection of pseudocapsule of renal cell carcinoma with MR imaging and CT. *AJR Am J Roentgenol* **166**, 1151-1155, doi:10.2214/ajr.166.5.8615260 (1996).

25 Gudbjartsson, T. *et al.* Histological subtyping and nuclear grading of renal cell carcinoma and their implications for survival: a retrospective nation-wide study of 629 patients. *Eur Urol* **48**, 593-600, doi:10.1016/j.eururo.2005.04.016 (2005).

26 Zhou, M. & He, H. in *Renal Cell Carcinoma: Clinical Management* (eds Steven C. Campbell & Brian I. Rini) 23-41 (Humana Press, 2013).

27 Frew, I. J. & Moch, H. A clearer view of the molecular complexity of clear cell renal cell carcinoma. *Annu Rev Pathol* **10**, 263-289, doi:10.1146/annurev-pathol-012414-040306 (2015).

28 Turajlic, S. *et al.* Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell* **173**, 595-610 e511, doi:10.1016/j.cell.2018.03.043 (2018).

29 Herman, J. G. *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A* **91**, 9700-9704 (1994).

30 Iliopoulos, O., Kibel, A., Gray, S. & Kaelin, W. G., Jr. Tumour suppression by the human von Hippel-Lindau gene product. *Nat Med* **1**, 822-826 (1995).

31 Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet* **45**, 860-867, doi:10.1038/ng.2699 (2013).

32 Kibel, A., Iliopoulos, O., DeCaprio, J. A. & Kaelin, W. G., Jr. Binding of the von Hippel-Lindau tumor suppressor protein to Elongin B and C. *Science* **269**, 1444-1446 (1995).

33 Hon, W. C. *et al.* Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL. *Nature* **417**, 975-978, doi:10.1038/nature00767 (2002).

34 Ivan, M. *et al.* HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science* **292**, 464-468, doi:10.1126/science.1059817 (2001).

35 Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271-275, doi:10.1038/20459 (1999).

36 Chachami, G. *et al.* Transport of hypoxia-inducible factor HIF-1alpha into the nucleus involves importins 4 and 7. *Biochem Biophys Res Commun* **390**, 235-240, doi:10.1016/j.bbrc.2009.09.093 (2009).

37 Schodel, J. *et al.* High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. *Blood* **117**, e207-217, doi:10.1182/blood-2010-10-314427 (2011).

38 Wenger, R. H., Stiehl, D. P. & Camenisch, G. Integration of oxygen signaling at the consensus HRE. *Sci STKE* **2005**, re12, doi:10.1126/stke.3062005re12 (2005).

39 Maxwell, P. H. *et al.* Hypoxia-inducible factor-1 modulates gene expression in solid tumors and influences both angiogenesis and tumor growth. *Proc Natl Acad Sci U S A* **94**, 8104-8109 (1997).

40 Ravi, R. *et al.* Regulation of tumor angiogenesis by p53-induced degradation of hypoxia-inducible factor 1alpha. *Genes Dev* **14**, 34-44 (2000).

41 Seagroves, T. N. *et al.* Transcription factor HIF-1 is a necessary mediator of the pasteur effect in mammalian cells. *Mol Cell Biol* **21**, 3436-3444, doi:10.1128/MCB.21.10.3436-3444.2001 (2001).

42 Biswas, S. *et al.* Effects of HIF-1alpha and HIF2alpha on Growth and Metabolism of Clear-Cell Renal Cell Carcinoma 786-0 Xenografts. *J Oncol* **2010**, 757908, doi:10.1155/2010/757908 (2010).

43 Metallo, C. M. *et al.* Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* **481**, 380-384, doi:10.1038/nature10602 (2011).

44 Rankin, E. B. *et al.* Hypoxia-inducible factor 2 regulates hepatic lipid metabolism. *Mol Cell Biol* **29**, 4527-4538, doi:10.1128/MCB.00200-09 (2009).

45 Schodel, J. *et al.* Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nat Genet* **44**, 420-425, S421-422, doi:10.1038/ng.2204 (2012).

46 Roe, J. S. *et al.* p53 stabilization and transactivation by a von Hippel-Lindau protein. *Mol Cell* **22**, 395-405, doi:10.1016/j.molcel.2006.04.006 (2006).

47 Krieg, M. *et al.* Up-regulation of hypoxia-inducible factors HIF-1alpha and HIF-2alpha under normoxic conditions in renal carcinoma cells by von Hippel-Lindau tumor suppressor gene loss of function. *Oncogene* **19**, 5435-5443, doi:10.1038/sj.onc.1203938 (2000).

48 Raval, R. R. *et al.* Contrasting properties of hypoxia-inducible factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-associated renal cell carcinoma. *Mol Cell Biol* **25**, 5675-5686, doi:10.1128/MCB.25.13.5675-5686.2005 (2005).

49 Rosenberger, C. *et al.* Expression of hypoxia-inducible factor-1alpha and -2alpha in hypoxic and ischemic rat kidneys. *J Am Soc Nephrol* **13**, 1721-1732 (2002).

50 Mandriota, S. J. *et al.* HIF activation identifies early lesions in VHL kidneys: evidence for site-specific tumor suppressor function in the nephron. *Cancer Cell* **1**, 459-468 (2002).

51 Lachance, G. *et al.* DNMT3a epigenetic program regulates the HIF-2alpha oxygen-sensing pathway and the cellular response to hypoxia. *Proc Natl Acad Sci U S A* **111**, 7783-7788, doi:10.1073/pnas.1322909111 (2014).

52 Keith, B., Johnson, R. S. & Simon, M. C. HIF1alpha and HIF2alpha: sibling rivalry in hypoxic tumour growth and progression. *Nat Rev Cancer* **12**, 9-22, doi:10.1038/nrc3183 (2011).

53 Shen, C. *et al.* Genetic and functional studies implicate HIF1alpha as a 14q kidney cancer suppressor gene. *Cancer Discov* **1**, 222-235, doi:10.1158/2159-8290.CD-11-0098 (2011).

54 Monzon, F. A. *et al.* Chromosome 14q loss defines a molecular subtype of clear-cell renal cell carcinoma associated with poor prognosis. *Mod Pathol* **24**, 1470-1479, doi:10.1038/modpathol.2011.107 (2011).

55 Adam, J. *et al.* Renal cyst formation in Fh1-deficient mice is independent of the Hif/Phd pathway: roles for fumarate in KEAP1 succination and Nrf2 signaling. *Cancer Cell* **20**, 524-537, doi:10.1016/j.ccr.2011.09.006 (2011).

56 Kondo, K., Klco, J., Nakamura, E., Lechpammer, M. & Kaelin, W. G., Jr. Inhibition of HIF is necessary for tumor suppression by the von Hippel-Lindau protein. *Cancer Cell* **1**, 237-246 (2002).

57 Kondo, K., Kim, W. Y., Lechpammer, M. & Kaelin, W. G., Jr. Inhibition of HIF2alpha is sufficient to suppress pVHL-defective tumor growth. *PLoS Biol* **1**, E83, doi:10.1371/journal.pbio.0000083 (2003).

58 Koshiji, M. *et al.* HIF-1alpha induces cell cycle arrest by functionally counteracting Myc. *EMBO J* **23**, 1949-1956, doi:10.1038/sj.emboj.7600196 (2004).

59 Gordan, J. D. *et al.* HIF-alpha effects on c-Myc distinguish two subtypes of sporadic VHL-deficient clear cell renal carcinoma. *Cancer Cell* **14**, 435-446, doi:10.1016/j.ccr.2008.10.016 (2008).

60 Gordan, J. D., Bertout, J. A., Hu, C. J., Diehl, J. A. & Simon, M. C. HIF-2alpha promotes hypoxic cell proliferation by enhancing c-myc transcriptional activity. *Cancer Cell* **11**, 335-347, doi:10.1016/j.ccr.2007.02.006 (2007).

61 Covello, K. L. *et al.* HIF-2alpha regulates Oct-4: effects of hypoxia on stem cell function, embryonic development, and tumor growth. *Genes Dev* **20**, 557-570, doi:10.1101/gad.1399906 (2006).

62 Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43-49, doi:10.1038/nature12222 (2013).

63 Ricketts, C. J. *et al.* The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep* **23**, 313-326 e315, doi:10.1016/j.celrep.2018.03.075 (2018).

64 Bertout, J. A. *et al.* HIF2alpha inhibition promotes p53 pathway activity, tumor cell death, and radiation responses. *Proc Natl Acad Sci U S A* **106**, 14391-14396, doi:10.1073/pnas.0907357106 (2009).

65 Obacz, J., Pastorekova, S., Vojtesek, B. & Hrstka, R. Cross-talk between HIF and p53 as mediators of molecular responses to physiological and genotoxic stresses. *Mol Cancer* **12**, 93, doi:10.1186/1476-4598-12-93 (2013).

66 Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899-905, doi:10.1038/nature08822 (2010).

67 Mitchell, T. J. *et al.* Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. *Cell* **173**, 611-623 e617, doi:10.1016/j.cell.2018.02.020 (2018).

68 Klatte, T. *et al.* Cytogenetic profile predicts prognosis of patients with clear cell renal cell carcinoma. *J Clin Oncol* **27**, 746-753, doi:10.1200/JCO.2007.15.8345 (2009).

69 Kovacs, G., Szucs, S., De Riese, W. & Baumgartel, H. Specific chromosome aberration in human renal cell carcinoma. *Int J Cancer* **40**, 171-178 (1987).

70 Pavlovich, C. P. *et al.* Patterns of aneuploidy in stage IV clear cell renal cell carcinoma revealed by comparative genomic hybridization and spectral karyotyping. *Genes Chromosomes Cancer* **37**, 252-260, doi:10.1002/gcc.10209 (2003).

71 Knudson, A. G., Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**, 820-823 (1971).

72 Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883-892, doi:10.1056/NEJMoa1113205 (2012).

73 Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* **46**, 225-233, doi:10.1038/ng.2891 (2014).

74 Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **469**, 539-542, doi:10.1038/nature09639 (2011).

75 Pena-Llopis, S. *et al.* BAP1 loss defines a new class of renal cell carcinoma. *Nat Genet* **44**, 751-759, doi:10.1038/ng.2323 (2012).

76 Hakimi, A. A. *et al.* Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. *Eur Urol* **63**, 848-854, doi:10.1016/j.eururo.2012.09.005 (2013).

77 Pena-Llopis, S., Christie, A., Xie, X. J. & Brugarolas, J. Cooperation and antagonism among cancer genes: the renal cancer paradigm. *Cancer Res* **73**, 4173-4179, doi:10.1158/0008-5472.CAN-13-0360 (2013).

78 Anderson, K. *et al.* Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* **469**, 356-361, doi:10.1038/nature09650 (2011).

79 Bashashati, A. *et al.* Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *The Journal of pathology* **231**, 21-34, doi:10.1002/path.4230 (2013).

80 Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109-1113, doi:10.1038/nature09460 (2010).

81 Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-94, doi:10.1038/nature09807 (2011).

82 Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A* **110**, 4009-4014, doi:10.1073/pnas.1219747110 (2013).

83 Hoglund, M. *et al.* Dissecting karyotypic patterns in renal cell carcinoma: an analysis of the accumulated cytogenetic data. *Cancer Genet Cytogenet* **153**, 1-9, doi:10.1016/j.cancergencyto.2003.12.019 (2004).

84 Wang, Z. *et al.* Sarcomatoid Renal Cell Carcinoma Has a Distinct Molecular Pathogenesis, Driver Mutation Profile, and Transcriptional Landscape. *Clinical cancer research : an official journal of the American Association for Cancer Research* **23**, 6686-6696, doi:10.1158/1078-0432.CCR-17-1057 (2017).

85 Scelo, G. *et al.* Variation in genomic landscape of clear cell renal cell carcinoma across Europe. *Nat Commun* **5**, 5135, doi:10.1038/ncomms6135 (2014).

86 Turajlic, S. *et al.* Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell* **173**, 581-594 e512, doi:10.1016/j.cell.2018.03.057 (2018).

87 Thompson, M. Polybromo-1: the chromatin targeting subunit of the PBAF complex. *Biochimie* **91**, 309-319, doi:10.1016/j.biochi.2008.10.019 (2009).

88 Brugarolas, J. PBRM1 and BAP1 as novel targets for renal cell carcinoma. *Cancer J* **19**, 324-332, doi:10.1097/PPO.0b013e3182a102d1 (2013).

89 Kanu, N. *et al.* SETD2 loss-of-function promotes renal cancer branched evolution through replication stress and impaired DNA repair. *Oncogene* **34**, 5699-5708, doi:10.1038/onc.2015.24 (2015).

90 Li, L., Miao, W., Huang, M., Williams, P. & Wang, Y. Integrated Genomic and Proteomic Analyses Reveal Novel Mechanisms of the Methyltransferase SETD2 in Renal Cell Carcinoma Development. *Mol Cell Proteomics*, doi:10.1074/mcp.RA118.000957 (2018).

91 Chiang, Y. C. *et al.* SETD2 Haploinsufficiency for Microtubule Methylation Is an Early Driver of Genomic Instability in Renal Cell Carcinoma. *Cancer Res* **78**, 3135-3146, doi:10.1158/0008-5472.CAN-17-3460 (2018).

92 Weinberg, R. *The biology of cancer*. (Garland science, 2013).

93 Sarbassov, D. D., Guertin, D. A., Ali, S. M. & Sabatini, D. M. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307**, 1098-1101, doi:10.1126/science.1106148 (2005).

94 Jacinto, E. *et al.* SIN1/MIP1 maintains rictor-mTOR complex integrity and regulates Akt phosphorylation and substrate specificity. *Cell* **127**, 125-137, doi:10.1016/j.cell.2006.08.033 (2006).

95 Mora, A., Komander, D., van Aalten, D. M. & Alessi, D. R. PDK1, the master regulator of AGC kinase signal transduction. *Semin Cell Dev Biol* **15**, 161-170 (2004).

96 Beauchamp, E. M. & Platanias, L. C. The evolution of the TOR pathway and its role in cancer. *Oncogene* **32**, 3923-3932, doi:10.1038/onc.2012.567 (2013).

97 Porstmann, T. *et al.* SREBP activity is regulated by mTORC1 and contributes to Akt-dependent cell growth. *Cell Metab* **8**, 224-236, doi:10.1016/j.cmet.2008.07.007 (2008).

98 Hudson, C. C. *et al.* Regulation of hypoxia-inducible factor 1alpha expression and function by the mammalian target of rapamycin. *Mol Cell Biol* **22**, 7004-7014 (2002).

99 Kucejova, B. *et al.* Interplay between pVHL and mTORC1 pathways in clear-cell renal cell carcinoma. *Mol Cancer Res* **9**, 1255-1265, doi:10.1158/1541-7786.MCR-11-0302 (2011).

100 Brugarolas, J. *et al.* Regulation of mTOR function in response to hypoxia by REDD1 and the TSC1/TSC2 tumor suppressor complex. *Genes Dev* **18**, 2893-2904, doi:10.1101/gad.1256804 (2004).

101 Nargund, A. M. *et al.* The SWI/SNF Protein PBRM1 Restrains VHL-Loss-Driven Clear Cell Renal Cell Carcinoma. *Cell Rep* **18**, 2893-2906, doi:10.1016/j.celrep.2017.02.074 (2017).

102 Elorza, A. *et al.* HIF2alpha acts as an mTORC1 activator through the amino acid carrier SLC7A5. *Mol Cell* **48**, 681-691, doi:10.1016/j.molcel.2012.09.017 (2012).

103 Yue, M., Jiang, J., Gao, P., Liu, H. & Qing, G. Oncogenic MYC Activates a Feedforward Regulatory Loop Promoting Essential Amino Acid Metabolism and Tumorigenesis. *Cell Rep* **21**, 3819-3832, doi:10.1016/j.celrep.2017.12.002 (2017).

104 Mossmann, D., Park, S. & Hall, M. N. mTOR signalling and cellular metabolism are mutual determinants in cancer. *Nature Reviews Cancer* **18**, 744-757, doi:10.1038/s41568-018-0074-8 (2018).

105 He, C. L. *et al.* Pyruvate Kinase M2 Activates mTORC1 by Phosphorylating AKT1S1. *Sci Rep* **6**, 21524, doi:10.1038/srep21524 (2016).

106 Santulli, G. *Angiogenesis: insights from a systematic overview*. (Nova Biomedical, 2013).

107 Ferrara, N. Role of vascular endothelial growth factor in regulation of physiological angiogenesis. *Am J Physiol Cell Physiol* **280**, C1358-1366, doi:10.1152/ajpcell.2001.280.6.C1358 (2001).

108 Prior, B. M., Yang, H. T. & Terjung, R. L. What makes vessels grow with exercise training? *J Appl Physiol (1985)* **97**, 1119-1128, doi:10.1152/japplphysiol.00035.2004 (2004).

109 Maglione, D. *et al.* Two alternative mRNAs coding for the angiogenic factor, placenta growth factor (PlGF), are transcribed from a single gene of chromosome 14. *Oncogene* **8**, 925-931 (1993).

110 Raica, M. & Cimpean, A. M. Platelet-Derived Growth Factor (PDGF)/PDGF Receptors (PDGFR) Axis as Target for Antitumor and Antiangiogenic Therapy. *Pharmaceuticals (Basel)* **3**, 572-599, doi:10.3390/ph3030572 (2010).

111 Wang, W. *et al.* Effect of platelet-derived growth factor-B on renal cell carcinoma growth and progression. *Urologic oncology* **33**, 168 e117-127, doi:10.1016/j.urolonc.2014.12.015 (2015).

112 Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).

113 Hanahan, D. & Folkman, J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* **86**, 353-364 (1996).

114 Sato, K. *et al.* Increased serum levels of vascular endothelial growth factor in patients with renal cell carcinoma. *Jpn J Cancer Res* **90**, 874-879 (1999).

115 Bienes-Martinez, R. *et al.* Autocrine stimulation of clear-cell renal carcinoma cell migration in hypoxia via HIF-independent suppression of thrombospondin-1. *Sci Rep* **2**, 788, doi:10.1038/srep00788 (2012).

116 Yao, X. *et al.* Two distinct types of blood vessels in clear cell renal cell carcinoma have contrasting prognostic implications. *Clinical cancer research : an official journal of the American Association for Cancer Research* **13**, 161-169, doi:10.1158/1078-0432.CCR-06-0774 (2007).

117 Bonsib, S. M. The renal sinus is the principal invasive pathway: a prospective study of 100 renal cell carcinomas. *The American journal of surgical pathology* **28**, 1594-1600 (2004).

118 Bonsib, S. M. Renal lymphatics, and lymphatic involvement in sinus vein invasive (pT3b) clear cell renal cell carcinoma: a study of 40 cases. *Modern pathology* **19**, 746 (2006).

119 Bianchi, M. *et al.* Distribution of metastatic sites in renal cell carcinoma: a population-based analysis. *Annals of Oncology* **23**, 973-980 (2011).

120 Hosseini, H. *et al.* Early dissemination seeds metastasis in breast cancer. *Nature*, doi:10.1038/nature20785 (2016).

121 Husemann, Y. *et al.* Systemic spread is an early step in breast cancer. *Cancer Cell* **13**, 58-68, doi:10.1016/j.ccr.2007.12.003 (2008).

122 Podsypanina, K. *et al.* Seeding and propagation of untransformed mouse mammary cells in the lung. *Science* **321**, 1841-1844, doi:10.1126/science.1161621 (2008).

123 Giancotti, F. G. Mechanisms governing metastatic dormancy and reactivation. *Cell* **155**, 750-764, doi:10.1016/j.cell.2013.10.029 (2013).

124 Rankin, E. B. & Giaccia, A. J. Hypoxic control of metastasis. *Science* **352**, 175-180, doi:10.1126/science.aaf4405 (2016).

125 Haase, V. H. Oxygen regulates epithelial-to-mesenchymal transition: insights into molecular mechanisms and relevance to disease. *Kidney Int* **76**, 492-499, doi:10.1038/ki.2009.222 (2009).

126 Lu, X. & Kang, Y. Hypoxia and hypoxia-inducible factors: master regulators of metastasis. *Clinical cancer research : an official journal of the American Association for Cancer Research* **16**, 5928-5935, doi:10.1158/1078-0432.CCR-10-1360 (2010).

127 Hay, E. D. An overview of epithelio-mesenchymal transformation. *Acta Anat (Basel)* **154**, 8-20 (1995).

128 Kong, D., Li, Y., Wang, Z. & Sarkar, F. H. Cancer Stem Cells and Epithelial-to-Mesenchymal Transition (EMT)-Phenotypic Cells: Are They Cousins or Twins? *Cancers (Basel)* **3**, 716-729, doi:10.3390/cancers30100716 (2011).

129 Erler, J. T. *et al.* Lysyl oxidase is essential for hypoxia-induced metastasis. *Nature* **440**, 1222-1226, doi:10.1038/nature04695 (2006).

130 Krishnamachary, B. *et al.* Regulation of colon carcinoma cell invasion by hypoxia-inducible factor 1. *Cancer Res* **63**, 1138-1143 (2003).

131 Sullivan, R. & Graham, C. H. Hypoxia-driven selection of the metastatic phenotype. *Cancer Metastasis Rev* **26**, 319-331, doi:10.1007/s10555-007-9062-2 (2007).

132 Staller, P. *et al.* Chemokine receptor CXCR4 downregulated by von Hippel-Lindau tumour suppressor pVHL. *Nature* **425**, 307-311, doi:10.1038/nature01874 (2003).

133 Amin, M. B. *et al. AJCC Cancer Staging Manual*. Vol. 8 (Springer International Publishing, 2017).

134 Taneja, K. & Williamson, S. R. Updates in Pathologic Staging and Histologic Grading of Renal Cell Carcinoma. *Surgical pathology clinics* **11**, 797-812 (2018).

135 Munich Cancer Registry. *Survival ICD-10 C64: Kidney cancer. Updated 22nd Aug 2018* <https://www.tumorregister-muenchen.de/en/facts/surv/sC64__E-ICD-10-C64-Kidney-cancer-survival.pdf > (2018).

136 Fuhrman, S. A., Lasky, L. C. & Limas, C. Prognostic significance of morphologic parameters in renal cell carcinoma. *Am J Surg Pathol* **6**, 655-663 (1982).

137 Zisman, A. *et al.* Improved prognostication of renal cell carcinoma using an integrated staging system. *J Clin Oncol* **19**, 1649-1657, doi:10.1200/JCO.2001.19.6.1649 (2001).

138 Ficarra, V. *et al.* Original and reviewed nuclear grading according to the Fuhrman system: a multivariate analysis of 388 patients with conventional renal cell carcinoma. *Cancer* **103**, 68-75, doi:10.1002/cncr.20749 (2005).

139 Rioux-Leclercq, N. *et al.* Prognostic ability of simplified nuclear grading of renal cell carcinoma. *Cancer* **109**, 868-874, doi:10.1002/cncr.22463 (2007).

140 Becker, A. *et al.* Critical analysis of a simplified Fuhrman grading scheme for prediction of cancer specific mortality in patients with clear cell renal cell carcinoma--Impact on prognosis. *Eur J Surg Oncol* **42**, 419-425, doi:10.1016/j.ejso.2015.09.023 (2016).

141 Smith, Z. L. *et al.* Simplification of the Fuhrman grading system for renal cell carcinoma. *Can J Urol* **22**, 8069-8073 (2015).

142 Frank, I. *et al.* An outcome prediction model for patients with clear cell renal cell carcinoma treated with radical nephrectomy based on tumor stage, size, grade and necrosis: the SSIGN score. *J Urol* **168**, 2395-2400, doi:10.1097/01.ju.0000035885.91935.d5 (2002).

143 Zigeuner, R. *et al.* External validation of the Mayo Clinic stage, size, grade, and necrosis (SSIGN) score for clear-cell renal cell carcinoma in a single European centre applying routine pathology. *Eur Urol* **57**, 102-109, doi:10.1016/j.eururo.2008.11.033 (2010).

144 Motzer, R. J. *et al.* Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. *J Clin Oncol* **17**, 2530-2540, doi:10.1200/JCO.1999.17.8.2530 (1999).

145 Albiges, L. *et al.* Body Mass Index and Metastatic Renal Cell Carcinoma: Clinical and Biological Correlations. *J Clin Oncol* **34**, 3655-3663, doi:10.1200/JCO.2016.66.7311 (2016).

146 Motzer, R. J. *et al.* Kidney cancer, version 2.2019, NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network* (2018).

147 Campbell, S. C. *et al.* Guideline for management of the clinical T1 renal mass. *J Urol* **182**, 1271-1279, doi:10.1016/j.juro.2009.07.004 (2009).

148 Leibovich, B. C. *et al.* Nephron sparing surgery for appropriately selected renal cell carcinoma between 4 and 7 cm results in outcome similar to radical nephrectomy. *J Urol* **171**, 1066-1070, doi:10.1097/01.ju.0000113274.40885.db (2004).

149 Scherr, A. J., Lima, J. P., Sasse, E. C., Lima, C. S. & Sasse, A. D. Adjuvant therapy for locally advanced renal cell cancer: a systematic review with meta-analysis. *BMC Cancer* **11**, 115, doi:10.1186/1471-2407-11-115 (2011).

150 Blom, J. H. *et al.* Radical nephrectomy with and without lymph-node dissection: final results of European Organization for Research and Treatment of Cancer (EORTC) randomized phase 3 trial 30881. *Eur Urol* **55**, 28-34, doi:10.1016/j.eururo.2008.09.052 (2009).

151 Bekema, H. J. *et al.* Systematic review of adrenalectomy and lymph node dissection in locally advanced renal cell carcinoma. *Eur Urol* **64**, 799-810, doi:10.1016/j.eururo.2013.04.033 (2013).

152 Amato, R. J. Chemotherapy for renal cell carcinoma. *Semin Oncol* **27**, 177-186 (2000).

153 Motzer, R. J. *et al.* Effect of cytokine therapy on survival for patients with advanced renal cell carcinoma. *J Clin Oncol* **18**, 1928-1935, doi:10.1200/JCO.2000.18.9.1928 (2000).

154 Buti, S. *et al.* Chemotherapy in metastatic renal cell carcinoma today? A systematic review. *Anticancer Drugs* **24**, 535-554, doi:10.1097/CAD.0b013e3283609ec1 (2013).

155 Naito, S. *et al.* Expression of P-glycoprotein and multidrug resistance in renal cell carcinoma. *Eur Urol* **24**, 156-160 (1993).

156 Dan, S. *et al.* An integrated database of chemosensitivity to 55 anticancer drugs and gene expression profiles of 39 human cancer cell lines. *Cancer Res* **62**, 1139-1147 (2002).

157 Walsh, N. *et al.* Expression of multidrug resistance markers ABCB1 (MDR-1/P-gp) and ABCC1 (MRP-1) in renal cell carcinoma. *BMC Urol* **9**, 6, doi:10.1186/1471-2490-9-6 (2009).

158 Gottesman, M. M., Fojo, T. & Bates, S. E. Multidrug resistance in cancer: role of ATP-dependent transporters. *Nat Rev Cancer* **2**, 48-58, doi:10.1038/nrc706 (2002).

159 Chen, J. *et al.* Low expression of phosphatase and tensin homolog in clearcell renal cell carcinoma contributes to chemoresistance through activating the Akt/HDM2 signaling pathway. *Mol Med Rep* **12**, 2622-2628, doi:10.3892/mmr.2015.3740 (2015).

160 Deschavanne, P. J. & Fertil, B. A review of human cell radiosensitivity in vitro. *Int J Radiat Oncol Biol Phys* **34**, 251-266 (1996).

161 Blanco, A. I., Teh, B. S. & Amato, R. J. Role of radiation therapy in the management of renal cell cancer. *Cancers (Basel)* **3**, 4010-4023, doi:10.3390/cancers3044010 (2011).

162 Zhao, J. *et al.* The emerging role of hypoxia-inducible factor-2 involved in chemo/radioresistance in solid tumors. *Cancer Treat Rev* **41**, 623-633, doi:10.1016/j.ctrv.2015.05.004 (2015).

163 Bhatt, R. S. *et al.* Hypoxia-inducible factor-2alpha: effect on radiation sensitivity and differential regulation by an mTOR inhibitor. *BJU Int* **102**, 358-363, doi:10.1111/j.1464-410X.2008.07558.x (2008).

164 Wrann, S., Kaufmann, M. R., Wirthner, R., Stiehl, D. P. & Wenger, R. H. HIF mediated and DNA damage independent histone H2AX phosphorylation in chronic hypoxia. *Biological chemistry* **394**, 519-528 (2013).

165 Hui, Z. *et al.* Radiosensitization by inhibiting STAT1 in renal cell carcinoma. *Int J Radiat Oncol Biol Phys* **73**, 288-295, doi:10.1016/j.ijrobp.2008.08.043 (2009).

166 Onufrey, V. & Mohiuddin, M. Radiation therapy in the treatment of metastatic renal cell carcinoma. *Int J Radiat Oncol Biol Phys* **11**, 2007-2009 (1985).

167 DiBiase, S. J. *et al.* Palliative irradiation for focally symptomatic metastatic renal cell carcinoma: support for dose escalation based on a biological model. *J Urol* **158**, 746-749 (1997).

168 Halperin, E. C. & Harisiadis, L. The role of radiation therapy in the management of metastatic renal cell carcinoma. *Cancer* **51**, 614-617 (1983).

169 Mickisch, G. H. *et al.* Radical nephrectomy plus interferon-alfa-based immunotherapy compared with interferon alfa alone in metastatic renal-cell carcinoma: a randomised trial. *Lancet* **358**, 966-970 (2001).

170 Dariane, C., Timsit, M. O. & Mejean, A. Position of cytoreductive nephrectomy in the setting of metastatic renal cell carcinoma patients: does the CARMENA trial lead to a paradigm shift? *Bull Cancer* **105 Suppl 3**, S229-S234, doi:10.1016/S0007-4551(18)30377-1 (2018).

171 Cole, W. H. & Everson, T. C. Spontaneous regression of cancer: preliminary report. *Ann Surg* **144**, 366-383 (1956).

172 Challis, G. B. & Stam, H. J. The spontaneous regression of cancer. A review of cases from 1900 to 1987. *Acta Oncol* **29**, 545-550 (1990).

173 Rini, B. I. *et al.* Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of renal cell carcinoma. *Journal for immunotherapy of cancer* **4**, 81 (2016).

174 McDermott, D. F. Update on the application of interleukin-2 in the treatment of renal cell carcinoma. *Clinical Cancer Research* **13**, 716s-720s (2007).

175 Escudier, B. *et al.* Sorafenib in advanced clear-cell renal-cell carcinoma. *New England Journal of Medicine* **356**, 125-134 (2007).

176 U.S. Food & Drug Administration (FDA). *Nexavar - sorafenib*, <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2005/021923_s000_nexavartoc.cfm> (2005).

177 European Medicines Agency (EMA). *Nexavar - sorafenib*, <https://www.ema.europa.eu/en/medicines/human/EPAR/nexavar> (2006).

178 Motzer, R. J. *et al.* Pazopanib versus sunitinib in metastatic renal-cell carcinoma. *N Engl J Med* **369**, 722-731, doi:10.1056/NEJMoa1303989 (2013).

179 Escudier, B. *et al.* Phase III trial of bevacizumab plus interferon alfa-2a in patients with metastatic renal cell carcinoma (AVOREN): final analysis of overall survival. *J Clin Oncol* **28**, 2144-2150, doi:10.1200/JCO.2009.26.7849 (2010).

180 Hudes, G. *et al.* Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *N Engl J Med* **356**, 2271-2281, doi:10.1056/NEJMoa066838 (2007).

181 Motzer, R. J. *et al.* Phase 3 trial of everolimus for metastatic renal cell carcinoma : final results and analysis of prognostic factors. *Cancer* **116**, 4256-4265, doi:10.1002/cncr.25219 (2010).

182 Ishida, Y., Agata, Y., Shibahara, K. & Honjo, T. Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. *EMBO J* **11**, 3887-3895 (1992).

183 Keir, M. E., Butte, M. J., Freeman, G. J. & Sharpe, A. H. PD-1 and its ligands in tolerance and immunity. *Annu Rev Immunol* **26**, 677-704, doi:10.1146/annurev.immunol.26.021607.090331 (2008).

184 Zou, W. & Chen, L. Inhibitory B7-family molecules in the tumour microenvironment. *Nat Rev Immunol* **8**, 467-477, doi:10.1038/nri2326 (2008).

185 Thompson, R. H. *et al.* Tumor B7-H1 is associated with poor prognosis in renal cell carcinoma patients with long-term follow-up. *Cancer Res* **66**, 3381-3385, doi:10.1158/0008-5472.CAN-05-4303 (2006).

186 Teft, W. A., Kirchhof, M. G. & Madrenas, J. A molecular perspective of CTLA-4 function. *Annu Rev Immunol* **24**, 65-97, doi:10.1146/annurev.immunol.24.021605.090535 (2006).

187 Linsley, P. S. *et al.* Human B7-1 (CD80) and B7-2 (CD86) bind with similar avidities but distinct kinetics to CD28 and CTLA-4 receptors. *Immunity* **1**, 793-801 (1994).

188 Krummel, M. F. & Allison, J. P. CD28 and CTLA-4 have opposing effects on the response of T cells to stimulation. *J Exp Med* **182**, 459-465 (1995).

189 Hamid, O. & Carvajal, R. D. Anti-programmed death-1 and anti-programmed death-ligand 1 antibodies in cancer therapy. *Expert Opin Biol Ther* **13**, 847-861, doi:10.1517/14712598.2013.770836 (2013).

190 Zarrabi, K. & Wu, S. An evaluation of nivolumab for the treatment of metastatic renal cell carcinoma. *Expert Opin Biol Ther* **18**, 695-705, doi:10.1080/14712598.2018.1478962 (2018).

191 Escudier, B. *et al.* CheckMate 025 Randomized Phase 3 Study: Outcomes by Key Baseline Factors and Prior Therapy for Nivolumab Versus Everolimus in Advanced Renal Cell Carcinoma. *Eur Urol* **72**, 962-971, doi:10.1016/j.eururo.2017.02.010 (2017).

192 Escudier, B. *et al.* Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, doi:10.1093/annonc/mdz056 (2019).

193 Motzer, R. J. *et al.* Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. *N Engl J Med* **378**, 1277-1290, doi:10.1056/NEJMoa1712126 (2018).

194 Motzer, R. J. *et al.* IMmotion151: A Randomized Phase III Study of Atezolizumab Plus Bevacizumab vs Sunitinib in Untreated Metastatic Renal Cell Carcinoma (mRCC). *Journal of Clinical Oncology* **36**, 578-578, doi:10.1200/JCO.2018.36.6_suppl.578 (2018).

195 Motzer, R. J. *et al.* LBA6_PRJAVELIN renal 101: A randomized, phase III study of avelumab + axitinib vs sunitinib as first-line treatment of advanced renal cell carcinoma (aRCC). *Annals of Oncology* **29**, mdy424.036-mdy424.036, doi:10.1093/annonc/mdy424.036 (2018).

196 Merck (MSD). *Merck's KEYTRUDA® (pembrolizumab) in Combination with Pfizer's Inlyta® (axitinib) Significantly Improved Overall Survival (OS) and Progression-free Survival (PFS) as First-Line Therapy for Advanced or Metastatic Renal Cell Carcinoma.*, <https://www.mrknewsroom.com/news-release/oncology-newsroom/mercks-keytruda-pembrolizumab-combination-pfizers-inlyta-axitinib-sig> (2018).

197 ClinicalTrials.gov. *Study of Pembrolizumab (MK-3475) Monotherapy in Locally Advanced/Metastatic Renal Cell Carcinoma (MK-3475-427/KEYNOTE-427)*, <https://clinicaltrials.gov/ct2/show/NCT02853344> (2018).

198 ClinicalTrials.gov. *Safety and Efficacy Study of Pembrolizumab (MK-3475) as Monotherapy in the Adjuvant Treatment of Renal Cell Carcinoma Post Nephrectomy (MK-3475-564/KEYNOTE-564)*, <https://clinicaltrials.gov/ct2/show/NCT03142334> (2018).

199 ClinicalTrials.gov. *Lenvatinib/Everolimus or Lenvatinib/Pembrolizumab Versus Sunitinib Alone as Treatment of Advanced Renal Cell Carcinoma (CLEAR)*, <https://clinicaltrials.gov/ct2/show/NCT02811861> (2018).

200 Rini, B. I. *et al.* IMA901, a multipeptide cancer vaccine, plus sunitinib versus sunitinib alone, as first-line therapy for advanced or metastatic renal cell carcinoma (IMPRINT): a multicentre, open-label, randomised, controlled, phase 3 trial. *Lancet Oncol* **17**, 1599-1611, doi:10.1016/S1470-2045(16)30408-9 (2016).

201 Figlin, R. *et al.* 1137O - Interim analysis of the phase 3 ADAPT trial evaluating rocapuldencel-T (AGS-003), an individualized immunotherapy for the treatment of newly-diagnosed patients with metastatic renal cell carcinoma (mRCC). *Annals of Oncology* **28** (2017).

202 ClinicalTrials.gov. *Phase 3 Trial of Autologous Dendritic Cell Immunotherapy Plus Standard Treatment of Advanced Renal Cell Carcinoma (ADAPT)*, <https://clinicaltrials.gov/ct2/show/NCT01582672> (2018).

203 ClinicalTrials.gov. *166 Studies found for: Recruiting, Not yet recruiting, Active, not recruiting, Enrolling by invitation Studies | Interventional Studies | Renal Cell Carcinoma, Metastatic*, <https://clinicaltrials.gov/ct2/results?cond=Renal+Cell+Carcinoma%2C+Metastatic&recrs=b&recrs=a&recrs=f&recrs=d&age_v=&gndr=&type=Intr&rslt=&Search=Apply> (2019).

204 Martinez-Saez, O., Gajate Borau, P., Alonso-Gordoa, T., Molina-Cerrillo, J. & Grande, E. Targeting HIF-2 alpha in clear cell renal cell carcinoma: A promising

therapeutic strategy. *Crit Rev Oncol Hematol* **111**, 117-123, doi:10.1016/j.critrevonc.2017.01.013 (2017).

205 ClinicalTrials.gov. *A Trial of PT2977 in Combination With Cabozantinib in Patients With Clear Cell Renal Cell Carcinoma (ccRCC)*, <https://clinicaltrials.gov/ct2/show/NCT03634540?term=HIF2&recrs=abdf&type=Intr&cond=Renal+Cell+Carcinoma%2C+Metastatic&phase=01&rank=2> (2018).

206 Russell, W. M. S., Burch, R. L. & Hume, C. W. *The principles of humane experimental technique*. Vol. 238 (Methuen London, 1959).

207 Kleinman, H. K., Philp, D. & Hoffman, M. P. Role of the extracellular matrix in morphogenesis. *Curr Opin Biotechnol* **14**, 526-532 (2003).

208 Scherer, W. F., Syverton, J. T. & Gey, G. O. Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. *J Exp Med* **97**, 695-710 (1953).

209 Ertel, A., Verghese, A., Byers, S. W., Ochs, M. & Tozeren, A. Pathway-specific differences between tumor cell lines and normal and tumor tissue cells. *Mol Cancer* **5**, 55, doi:10.1186/1476-4598-5-55 (2006).

210 Domcke, S., Sinha, R., Levine, D. A., Sander, C. & Schultz, N. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat Commun* **4**, 2126, doi:10.1038/ncomms3126 (2013).

211 Bjare, U. Serum-free cell culture. *Pharmacol Ther* **53**, 355-374 (1992).

212 Ben-David, U. *et al.* Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* **560**, 325-330, doi:10.1038/s41586-018-0409-3 (2018).

213 Dufour, J. M. *et al.* Sertoli cell line lacks the immunoprotective properties associated with primary Sertoli cells. *Cell Transplant* **17**, 525-534 (2008).

214 Pan, C., Kumar, C., Bohl, S., Klingmueller, U. & Mann, M. Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. *Mol Cell Proteomics* **8**, 443-450, doi:10.1074/mcp.M800258-MCP200 (2009).

215 Alge, C. S., Hauck, S. M., Priglinger, S. G., Kampik, A. & Ueffing, M. Differential protein profiling of primary versus immortalized human RPE cells identifies expression patterns associated with cytoskeletal remodeling and cell survival. *J Proteome Res* **5**, 862-878, doi:10.1021/pr050420t (2006).

216 Lee, J. *et al.* Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell* **9**, 391-403, doi:10.1016/j.ccr.2006.03.030 (2006).

217 Freedman, L. P., Cockburn, I. M. & Simcoe, T. S. The Economics of Reproducibility in Preclinical Research. *PLoS Biol* **13**, e1002165, doi:10.1371/journal.pbio.1002165 (2015).

218 Gillet, J. P. *et al.* Redefining the relevance of established cancer cell lines to the study of mechanisms of clinical anti-cancer drug resistance. *Proc Natl Acad Sci U S A* **108**, 18708-18713, doi:10.1073/pnas.1111840108 (2011).

219 Gazdar, A. F., Gao, B. & Minna, J. D. Lung cancer cell lines: Useless artifacts or invaluable tools for medical science? *Lung Cancer* **68**, 309-318, doi:10.1016/j.lungcan.2009.12.005 (2010).

220 Brodaczewska, K. K., Szczylik, C., Fiedorowicz, M., Porta, C. & Czarnecka, A. M. Choosing the right cell line for renal cell cancer research. *Mol Cancer* **15**, 83, doi:10.1186/s12943-016-0565-8 (2016).

221 Beniers, A. J. *et al.* Establishment and characterization of five new human renal tumor xenografts. *The American journal of pathology* **140**, 483-495 (1992).

222 Beroukhim, R. *et al.* Patterns of gene expression and copy-number alterations in von-hippel lindau disease-associated and sporadic clear cell carcinoma of the kidney. *Cancer Res* **69**, 4674-4681, doi:10.1158/0008-5472.CAN-09-0146 (2009).

223 Huijbers, I. J. Generating Genetically Modified Mice: A Decision Guide. *Methods Mol Biol* **1642**, 1-19, doi:10.1007/978-1-4939-7169-5_1 (2017).

224 Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* **89**, 5547-5551 (1992).

225 Rudolph, K. L. Telomeres and telomerase influence the course of human diseases, aging and carcinogenesis. *Curr Mol Med* **5**, 133-134 (2005).

226 Gnarra, J. R. *et al.* Defective placental vasculogenesis causes embryonic lethality in VHL-deficient mice. *Proc Natl Acad Sci U S A* **94**, 9102-9107 (1997).

227 Kleymenova, E. *et al.* Susceptibility to vascular neoplasms but no increased susceptibility to renal carcinogenesis in Vhl knockout mice. *Carcinogenesis* **25**, 309-315, doi:10.1093/carcin/bgh017 (2004).

228 Rankin, E. B., Tomaszewski, J. E. & Haase, V. H. Renal cyst development in mice with conditional inactivation of the von Hippel-Lindau tumor suppressor. *Cancer Res* **66**, 2576-2583, doi:10.1158/0008-5472.CAN-05-3241 (2006).

229 Fu, L., Wang, G., Shevchuk, M. M., Nanus, D. M. & Gudas, L. J. Generation of a mouse model of Von Hippel-Lindau kidney disease leading to renal cancers by expression of a constitutively active mutant of HIF1alpha. *Cancer Res* **71**, 6848-6856, doi:10.1158/0008-5472.CAN-11-1745 (2011).

230 Harlander, S. *et al.* Combined mutation in Vhl, Trp53 and Rb1 causes clear cell renal cell carcinoma in mice. *Nat Med* **23**, 869-877, doi:10.1038/nm.4343 (2017).

231 Gu, Y. F. *et al.* Modeling Renal Cell Carcinoma in Mice: Bap1 and Pbrm1 Inactivation Drive Tumor Grade. *Cancer Discov* **7**, 900-917, doi:10.1158/2159-8290.CD-17-0292 (2017).

232 Flanagan, S. P. 'Nude', a new hairless gene with pleiotropic effects in the mouse. *Genet Res* **8**, 295-309 (1966).

233 Shultz, L. D. *et al.* Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. *J Immunol* **174**, 6477-6489 (2005).

234 Shultz, L. D., Brehm, M. A., Garcia-Martinez, J. V. & Greiner, D. L. Humanized mice for immune system investigation: progress, promise and challenges. *Nat Rev Immunol* **12**, 786-798, doi:10.1038/nri3311 (2012).

235 Ito, M. *et al.* NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells. *Blood* **100**, 3175-3182, doi:10.1182/blood-2001-12-0207 (2002).

236 Ruggeri, B. A., Camp, F. & Miknyoczki, S. Animal models of disease: pre-clinical animal models of cancer and their applications and utility in drug discovery. *Biochem Pharmacol* **87**, 150-161, doi:10.1016/j.bcp.2013.06.020 (2014).

237 An, Z., Jiang, P., Wang, X., Moossa, A. R. & Hoffman, R. M. Development of a high metastatic orthotopic model of human renal cell carcinoma in nude mice: benefits of fragment implantation compared to cell-suspension injection. *Clin Exp Metastasis* **17**, 265-270 (1999).

238 Naito, S., von Eschenbach, A. C., Giavazzi, R. & Fidler, I. J. Growth and metastasis of tumor cells isolated from a human renal cell carcinoma implanted into different organs of nude mice. *Cancer Res* **46**, 4109-4115 (1986).

239 Garofalo, A. *et al.* Comparative study on the metastatic behavior of human tumors in nude, beige/nude/xid and severe combined immunodeficient mice. *Invasion Metastasis* **13**, 82-91 (1993).

240 Grisanzio, C. *et al.* Orthotopic xenografts of RCC retain histological, immunophenotypic and genetic features of tumours in patients. *The Journal of pathology* **225**, 212-221, doi:10.1002/path.2929 (2011).

241 Sivanand, S. *et al.* A validated tumorgraft model reveals activity of dovitinib against renal cell carcinoma. *Sci Transl Med* **4**, 137ra175, doi:10.1126/scitranslmed.3003643 (2012).

242 Pavia-Jimenez, A., Tcheuyap, V. T. & Brugarolas, J. Establishing a human renal cell carcinoma tumorgraft platform for preclinical drug testing. *Nat Protoc* **9**, 1848-1859, doi:10.1038/nprot.2014.108 (2014).

243 Dong, Y. *et al.* Tumor Xenografts of Human Clear Cell Renal Cell Carcinoma But Not Corresponding Cell Lines Recapitulate Clinical Response to Sunitinib: Feasibility of Using Biopsy Samples. *Eur Urol Focus* **3**, 590-598, doi:10.1016/j.euf.2016.08.005 (2017).

244 CrownBio. *Kidney Cancer PDX Models*, <https://models.crownbio.com/kidney-cancer/> (2019).

245 Hong, B. *et al.* Intra-tumour molecular heterogeneity of clear cell renal cell carcinoma reveals the diversity of the response to targeted therapies using patient-derived xenograft models. *Oncotarget* **8**, 49839-49850, doi:10.18632/oncotarget.17765 (2017).

246 Murphy, G. P. & Hrushesky, W. J. A murine renal cell carcinoma. *J Natl Cancer Inst* **50**, 1013-1025 (1973).

247 Salup, R. R. & Wiltrout, R. H. Adjuvant immunotherapy of established murine renal cancer by interleukin 2-stimulated cytotoxic lymphocytes. *Cancer Res* **46**, 3358-3363 (1986).

248 Noll, E. M. *et al.* CYP3A5 mediates basal and acquired therapy resistance in different subtypes of pancreatic ductal adenocarcinoma. *Nat Med* **22**, 278-287, doi:10.1038/nm.4038 (2016).

249 Vermeulen, L. *et al.* Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. *Proc Natl Acad Sci U S A* **105**, 13427-13432, doi:10.1073/pnas.0805706105 (2008).

250 Kleinbaum, D. G. & Klein, M. *Survival analysis*. Vol. 3 (Springer, 2010).

251 Kaplan, E. L. & Meier, P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association* **53**, 457-481 (1958).

252 Mantel, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* **50**, 163-170 (1966).

253 Weiß, C. *Basiswissen medizinische statistik*. (Springer-Verlag, 2010).

254 Cox, D. R. Regression models and life- tables. *Journal of the Royal Statistical Society: Series B (Methodological)* **34**, 187-202 (1972).

255 Marubini, E. & Valsecchi, M. G. *Analysing survival data from clinical trials and observational studies*. Vol. 15 (John Wiley & Sons, 2004).

256 Breslow, N. E. Analysis of survival data under the proportional hazards model. *International Statistical Review/Revue Internationale de Statistique*, 45-57 (1975).

257 Efron, B. The efficiency of Cox's likelihood function for censored data. *Journal of the American statistical Association* **72**, 557-565 (1977).

258 Hertz-Picciotto, I. & Rockhill, B. Validity and efficiency of approximation methods for tied survival times in Cox regression. *Biometrics* **53**, 1151-1156 (1997).

259 Pawitan, Y. *In all likelihood: statistical modelling and inference using likelihood*. (Oxford University Press, 2001).

260 Harrell, F. E. Regression modeling strategies. 2001. *Nashville: Springer CrossRef Google Scholar* (2001).

261 Pencina, M. J. & D'Agostino, R. B. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* **23**, 2109-2123, doi:10.1002/sim.1802 (2004).

262 Friedman, J., Hastie, T. & Tibshirani, R. *The elements of statistical learning*. Vol. 1 (Springer series in statistics New York, NY, USA:, 2001).

263 EliteDataScience.com. *WTF is the Bias-Variance Tradeoff? (Infographic)*, <https://elitedatascience.com/bias-variance-tradeoff> (2016-2018).

264 Goeman, J. J. L1 penalized estimation in the Cox proportional hazards model. *Biom J* **52**, 70-84, doi:10.1002/bimj.200900028 (2010).

265 Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw* **39**, 1-13, doi:10.18637/jss.v039.i05 (2011).

266 Tibshirani, R. The lasso method for variable selection in the Cox model. *Stat Med* **16**, 385-395 (1997).

267 Gulati, S. *et al.* Systematic evaluation of the prognostic impact and intratumour heterogeneity of clear cell renal cell carcinoma biomarkers. *European urology* **66**, 936-948 (2014).

268 Andrieu, C., De Freitas, N., Doucet, A. & Jordan, M. I. An introduction to MCMC for machine learning. *Machine learning* **50**, 5-43 (2003).

269 Jiang, Y., Qiu, Y., Minn, A. J. & Zhang, N. R. Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. *Proceedings of the National Academy of Sciences* **113**, E5528-E5537 (2016).

270 Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust Hyperparameter Estimation Protects against Hypervariable Genes and Improves Power to Detect Differential Expression. *Ann Appl Stat* **10**, 946-963, doi:10.1214/16-AOAS920 (2016).

271 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).

272 De Bin, R., Sauerbrei, W. & Boulesteix, A. L. Investigating the prediction ability of survival models based on both clinical and omics data: two case studies. *Statistics in medicine* **33**, 5310-5329 (2014).

273 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).

274 Sauerbrei, W. & Schumacher, M. A bootstrap resampling procedure for model building: application to the Cox regression model. *Stat Med* **11**, 2093-2109 (1992).

275 phenoTest: Tools to test association between gene expression and phenotype in a way that is efficient, structured, fast and scalable. We also provide tools to do GSEA (Gene set enrichment analysis) and copy number variation. R package version 1.30.0 (2018).

276 Beleut, M. *et al.* Integrative genome-wide expression profiling identifies three distinct molecular subgroups of renal cell carcinoma with different patient outcome. *BMC cancer* **12**, 310 (2012).

277 Boström, A.-K., Lindgren, D., Johansson, M. E. & Axelson, H. Effects of TGF-β signaling in clear cell renal cell carcinoma cells. *Biochemical and biophysical research communications* **435**, 126-133 (2013).

278 Brooks, S. A. *et al.* ClearCode34: A prognostic risk predictor for localized clear cell renal cell carcinoma. *European urology* **66**, 77-84 (2014).

279 Brannon, A. R. *et al.* Molecular Stratification of Clear Cell Renal Cell Carcinoma by Consensus Clustering Reveals Distinct Subtypes and Survival Patterns. *Genes Cancer* **1**, 152-163, doi:10.1177/1947601909359929 (2010).

280 Kosari, F. *et al.* Clear cell renal cell carcinoma: gene expression analyses identify a potential signature for tumor aggressiveness. *Clinical cancer research : an official journal of the American Association for Cancer Research* **11**, 5128-5139, doi:10.1158/1078-0432.Ccr-05-0073 (2005).

281 Lane, B. R. *et al.* Differential expression in clear cell renal cell carcinoma identified by gene expression profiling. *The Journal of urology* **181**, 849-860 (2009).

282 Zhao, H. *et al.* Gene expression profiling predicts survival in conventional renal cell carcinoma. *PLoS medicine* **3**, e13 (2005).

283 Reich, M. *et al.* GenePattern 2.0. *Nature genetics* **38**, 500 (2006).

284 Kendall, M. G. A new measure of rank correlation. *Biometrika* **30**, 81-93 (1938).

285 Altman, D. G. *Practical statistics for medical research*. (CRC press, 1990).

286 Lausen, B., Hothorn, T., Bretz, F. & Schumacher, M. Assessment of optimal selected prognostic factors. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **46**, 364-374 (2004).

287 Cotto, K. C. *et al.* DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res* **46**, D1068-D1073, doi:10.1093/nar/gkx1143 (2018).

288 Ailane, N. *et al.* Effect of an anti-human Co-029/tspan8 mouse monoclonal antibody on tumor growth in a nude mouse model. *Front Physiol* **5**, 364, doi:10.3389/fphys.2014.00364 (2014).

289 Boucheix, C. & Rubinstein, E. Tetraspanins. *Cell Mol Life Sci* **58**, 1189-1205, doi:10.1007/PL00000933 (2001).

290 Charrin, S., Jouannet, S., Boucheix, C. & Rubinstein, E. Tetraspanins at a glance. *J Cell Sci* **127**, 3641-3648, doi:10.1242/jcs.154906 (2014).

291 Hemler, M. E. Tetraspanin proteins promote multiple cancer stages. *Nat Rev Cancer* **14**, 49-60 (2014).

292 Zoller, M. Tetraspanins: push and pull in suppressing and promoting metastasis. *Nat Rev Cancer* **9**, 40-55, doi:10.1038/nrc2543 (2009).

293 Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev* **16**, 948-958, doi:10.1101/gad.981002 (2002).

294 Agaesse, G., Barbollat-Boutrand, L., El Kharbili, M., Berthier-Vergnes, O. & Masse, I. p53 targets TSPAN8 to prevent invasion in melanoma cells. *Oncogenesis* **6**, e309, doi:10.1038/oncsis.2017.11 (2017).

295 Salem, M. *et al.* miR-590-3p Promotes Ovarian Cancer Growth and Metastasis via a Novel FOXA2-Versican Pathway. *Cancer Res* **78**, 4175-4190, doi:10.1158/0008-5472.CAN-17-3014 (2018).

296 Le Gallo, M. *et al.* The FOXA2 transcription factor is frequently somatically mutated in uterine carcinosarcomas and carcinomas. *Cancer* **124**, 65-73, doi:10.1002/cncr.30971 (2018).

297 Jia, Z. *et al.* Forkhead-box series expression network is associated with outcome of clear-cell renal cell carcinoma. *Oncol Lett* **15**, 8669-8680, doi:10.3892/ol.2018.8405 (2018).

298 Mallona, I., Diez-Villanueva, A. & Peinado, M. A. Methylation plotter: a web tool for dynamic visualization of DNA methylation data. *Source Code Biol Med* **9**, 11, doi:10.1186/1751-0473-9-11 (2014).

299 Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

300 Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **46**, D794-D801, doi:10.1093/nar/gkx1081 (2018).

301 Wang, Y., Wang, Y. & Liu, F. A 44-gene set constructed for predicting the prognosis of clear cell renal cell carcinoma. *International journal of molecular medicine* (2018).

302 Buttner, F. *et al.* Survival Prediction of Clear Cell Renal Cell Carcinoma Based on Gene Expression Similarity to the Proximal Tubule of the Nephron. *Eur Urol* **68**, 1016-1020, doi:10.1016/j.eururo.2015.05.045 (2015).

303 Buttner, F. *et al.* Clinical utility of the S3-score for molecular prediction of outcome in non-metastatic and metastatic clear cell renal cell carcinoma. *BMC Med* **16**, 108, doi:10.1186/s12916-018-1088-5 (2018).

304 Moch, H. [The WHO/ISUP grading system for renal carcinoma]. *Pathologe* **37**, 355-360, doi:10.1007/s00292-016-0171-y (2016).

305 Katoh, H. *et al.* CXCR2-expressing myeloid-derived suppressor cells are essential to promote colitis-associated tumorigenesis. *Cancer Cell* **24**, 631-644, doi:10.1016/j.ccr.2013.10.009 (2013).

306 Jamieson, T. *et al.* Inhibition of CXCR2 profoundly suppresses inflammation-driven and spontaneous tumorigenesis. *J Clin Invest* **122**, 3127-3144, doi:10.1172/JCI61067 (2012).

307 Coffelt, S. B., Wellenstein, M. D. & de Visser, K. E. Neutrophils in cancer: neutral no more. *Nat Rev Cancer* **16**, 431-446, doi:10.1038/nrc.2016.52 (2016).

308 Szczerba, B. M. *et al.* Neutrophils escort circulating tumour cells to enable cell cycle progression. *Nature* **566**, 553-557, doi:10.1038/s41586-019-0915-y (2019).

309 Shibasaki, N. *et al.* Role of IL13RA2 in Sunitinib Resistance in Clear Cell Renal Cell Carcinoma. *PLoS One* **10**, e0130980, doi:10.1371/journal.pone.0130980 (2015).

310 Kelly-Welch, A., Hanson, E. M. & Keegan, A. D. Interleukin-13 (IL-13) pathway. *Sci STKE* **2005**, cm8, doi:10.1126/stke.2932005cm8 (2005).

311 Tu, M. *et al.* IL-13 receptor alpha2 stimulates human glioma cell growth and metastasis through the Src/PI3K/Akt/mTOR signaling pathway. *Tumour Biol* **37**, 14701-14709, doi:10.1007/s13277-016-5346-x (2016).

312 Terabe, M., Park, J. M. & Berzofsky, J. A. Role of IL-13 in regulation of anti-tumor immunity and tumor growth. *Cancer Immunol Immunother* **53**, 79-85, doi:10.1007/s00262-003-0445-0 (2004).

313 Augustyn, A. *et al.* ASCL1 is a lineage oncogene providing therapeutic targets for high-grade neuroendocrine lung cancers. *Proc Natl Acad Sci U S A* **111**, 14788-14793, doi:10.1073/pnas.1410419111 (2014).

314 Fujino, K. *et al.* Insulinoma-Associated Protein 1 Is a Crucial Regulator of Neuroendocrine Differentiation in Lung Cancer. *The American journal of pathology* **185**, 3164-3177, doi:10.1016/j.ajpath.2015.08.018 (2015).

315 Guo, H. *et al.* ONECUT2 is a driver of neuroendocrine prostate cancer. *Nat Commun* **10**, 278, doi:10.1038/s41467-018-08133-6 (2019).

316 Sun, Y. *et al.* MiR-429 inhibits cells growth and invasion and regulates EMT-related marker genes by targeting Onecut2 in colorectal carcinoma. *Mol Cell Biochem* **390**, 19-30, doi:10.1007/s11010-013-1950-x (2014).

317 Rotinen, M. *et al.* ONECUT2 is a targetable master regulator of lethal prostate cancer that suppresses the androgen axis. *Nat Med* **24**, 1887-1898, doi:10.1038/s41591-018-0241-1 (2018).

318 Claas, C. *et al.* Association between the rat homologue of CO-029, a metastasis-associated tetraspanin molecule and consumption coagulopathy. *J Cell Biol* **141**, 267-280 (1998).

319 Kanetaka, K. *et al.* Possible involvement of tetraspanin CO-029 in hematogenous intrahepatic metastasis of liver cancer cells. *J Gastroenterol Hepatol* **18**, 1309-1314 (2003).

320 Sadej, R., Grudowska, A., Turczyk, L., Kordek, R. & Romanska, H. M. CD151 in cancer progression and metastasis: a complex scenario. *Lab Invest* **94**, 41-51, doi:10.1038/labinvest.2013.136 (2014).

321 Zhao, K., Wang, Z., Hackert, T., Pitzer, C. & Zoller, M. Tspan8 and Tspan8/CD151 knockout mice unravel the contribution of tumor and host exosomes to tumor progression. *J Exp Clin Cancer Res* **37**, 312, doi:10.1186/s13046-018-0961-6 (2018).

322 Hall, A., Karplus, P. A. & Poole, L. B. Typical 2-Cys peroxiredoxins--structures, mechanisms and functions. *FEBS J* **276**, 2469-2477, doi:10.1111/j.1742-4658.2009.06985.x (2009).

323 Gill, J. G., Piskounova, E. & Morrison, S. J. Cancer, Oxidative Stress, and Metastasis. *Cold Spring Harb Symp Quant Biol* **81**, 163-175, doi:10.1101/sqb.2016.81.030791 (2016).

324 West, J. D., Roston, T. J., David, J. B., Allan, K. M. & Loberg, M. A. Piecing Together How Peroxiredoxins Maintain Genomic Stability. *Antioxidants (Basel)* **7**, doi:10.3390/antiox7120177 (2018).

325 Freiss, G. & Chalbos, D. PTPN13/PTPL1: an important regulator of tumor aggressiveness. *Anticancer Agents Med Chem* **11**, 78-88 (2011).

326 Sotelo, N. S., Schepens, J. T., Valiente, M., Hendriks, W. J. & Pulido, R. PTEN-PDZ domain interactions: binding of PTEN to PDZ domains of PTPN13. *Methods* **77-78**, 147-156, doi:10.1016/j.ymeth.2014.10.017 (2015).

327 Zhan, H. *et al.* Tumour-suppressive role of PTPN13 in hepatocellular carcinoma and its clinical significance. *Tumour Biol* **37**, 9691-9698, doi:10.1007/s13277-016-4843-2 (2016).

328 Glondu-Lassis, M. *et al.* PTPL1/PTPN13 regulates breast cancer cell aggressiveness through direct inactivation of Src kinase. *Cancer Res* **70**, 5116-5126, doi:10.1158/0008-5472.CAN-09-4368 (2010).

329 Dromard, M. *et al.* The putative tumor suppressor gene PTPN13/PTPL1 induces apoptosis through insulin receptor substrate-1 dephosphorylation. *Cancer Res* **67**, 6806-6813, doi:10.1158/0008-5472.CAN-07-0513 (2007).

330 Anzai, N., Jutabha, P., Amonpatumrat-Takahashi, S. & Sakurai, H. Recent advances in renal urate transport: characterization of candidate transporters indicated by genome-wide association studies. *Clin Exp Nephrol* **16**, 89-95, doi:10.1007/s10157-011-0532-z (2012).

331 Caulfield, M. J. *et al.* SLC2A9 is a high-capacity urate transporter in humans. *PLoS Med* **5**, e197, doi:10.1371/journal.pmed.0050197 (2008).

332 Matsuo, H. *et al.* Mutations in glucose transporter 9 gene SLC2A9 cause renal hypouricemia. *Am J Hum Genet* **83**, 744-751, doi:10.1016/j.ajhg.2008.11.001 (2008).

333 Itahana, Y. *et al.* The uric acid transporter SLC2A9 is a direct target gene of the tumor suppressor p53 contributing to antioxidant defense. *Oncogene* **34**, 1799-1810, doi:10.1038/onc.2014.119 (2015).

334 Han, X., Yang, J., Li, D. & Guo, Z. Overexpression of uric acid transporter SLC2A9 inhibits proliferation of hepatocellular carcinoma cells. *Oncol Res*, doi:10.3727/096504018X15199489058224 (2018).

335 Sangkop, F., Singh, G., Rodrigues, E., Gold, E. & Bahn, A. Uric acid: a modulator of prostate cells and activin sensitivity. *Mol Cell Biochem* **414**, 187-199, doi:10.1007/s11010-016-2671-8 (2016).

336 Merkle, E. C. & Steyvers, M. Choosing a strictly proper scoring rule. *Decision Analysis* **10**, 292-304 (2013).

337 Matthay, K. K., George, R. E. & Yu, A. L. Promising therapeutic targets in neuroblastoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* **18**, 2740-2753, doi:10.1158/1078-0432.CCR-11-1939 (2012).

338 Hemler, M. E. Tetraspanin functions and associated microdomains. *Nat Rev Mol Cell Biol* **6**, 801-811, doi:10.1038/nrm1736 (2005).

339 Andre, M. *et al.* Proteomic analysis of the tetraspanin web using LC-ESI-MS/MS and MALDI-FTICR-MS. *Proteomics* **6**, 1437-1449, doi:10.1002/pmic.200500180 (2006).

340 Le Naour, F., Andre, M., Boucheix, C. & Rubinstein, E. Membrane microdomains and proteomics: lessons from tetraspanin microdomains and comparison with lipid rafts. *Proteomics* **6**, 6447-6454, doi:10.1002/pmic.200600282 (2006).

341 Berditchevski, F. Complexes of tetraspanins with integrins: more than meets the eye. *J Cell Sci* **114**, 4143-4151 (2001).

342 Little, K. D., Hemler, M. E. & Stipp, C. S. Dynamic regulation of a GPCR-tetraspanin-G protein complex on intact cells: central role of CD81 in facilitating GPR56-Galpha q/11 association. *Mol Biol Cell* **15**, 2375-2387, doi:10.1091/mbc.e03-12-0886 (2004).

343 Sridhar, S. C. & Miranti, C. K. Tetraspanin KAI1/CD82 suppresses invasion by inhibiting integrin-dependent crosstalk with c-Met receptor and Src kinases. *Oncogene* **25**, 2367-2378, doi:10.1038/sj.onc.1209269 (2006).

344 Stipp, C. S., Kolesnikova, T. V. & Hemler, M. E. EWI-2 is a major CD9 and CD81 partner and member of a novel Ig protein subfamily. *J Biol Chem* **276**, 40545-40554, doi:10.1074/jbc.M107338200 (2001).

345 Claas, C. *et al.* The tetraspanin D6.1A and its molecular partners on rat carcinoma cells. *Biochem J* **389**, 99-110, doi:10.1042/BJ20041287 (2005).

346 Maecker, H. T., Todd, S. C. & Levy, S. The tetraspanin superfamily: molecular facilitators. *FASEB J* **11**, 428-442 (1997).

347 Schorey, J. S. & Bhatnagar, S. Exosome function: from tumor immunology to pathogen biology. *Traffic* **9**, 871-881, doi:10.1111/j.1600-0854.2008.00734.x (2008).

348 French, K. C., Antonyak, M. A. & Cerione, R. A. Extracellular vesicle docking at the cellular port: Extracellular vesicle binding and uptake. *Semin Cell Dev Biol* **67**, 48-55, doi:10.1016/j.semcdb.2017.01.002 (2017).

349 Lakkaraju, A. & Rodriguez-Boulan, E. Itinerant exosomes: emerging roles in cell and tissue polarity. *Trends Cell Biol* **18**, 199-209, doi:10.1016/j.tcb.2008.03.002 (2008).

350 Andre, F. *et al.* Tumor-derived exosomes: a new source of tumor rejection antigens. *Vaccine* **20 Suppl 4**, A28-31 (2002).

351 Hemler, M. E. Tetraspanin proteins mediate cellular penetration, invasion, and fusion events and define a novel type of membrane microdomain. *Annu Rev Cell Dev Biol* **19**, 397-422, doi:10.1146/annurev.cellbio.19.111301.153609 (2003).

352 Zoller, M. Gastrointestinal tumors: metastasis and tetraspanins. *Z Gastroenterol* **44**, 573-586, doi:10.1055/s-2006-926795 (2006).

353 Kwon, H. J. *et al.* Expression of CD9 and CD82 in clear cell renal cell carcinoma and its clinical significance. *Pathol Res Pract* **210**, 285-290, doi:10.1016/j.prp.2014.01.004 (2014).

354 Gesierich, S., Berezovskiy, I., Ryschich, E. & Zoller, M. Systemic induction of the angiogenesis switch by the tetraspanin D6.1A/CO-029. *Cancer Res* **66**, 7083-7094, doi:10.1158/0008-5472.CAN-06-0391 (2006).

355 Kuhn, S. *et al.* A complex of EpCAM, claudin-7, CD44 variant isoforms, and tetraspanins promotes colorectal cancer progression. *Mol Cancer Res* **5**, 553-567, doi:10.1158/1541-7786.MCR-06-0384 (2007).

356 Greco, C. *et al.* E-cadherin/p120-catenin and tetraspanin Co-029 cooperate for cell motility control in human colon carcinoma. *Cancer Res* **70**, 7674-7683, doi:10.1158/0008-5472.CAN-09-4482 (2010).

357 Guo, Q. *et al.* Tetraspanin CO-029 inhibits colorectal cancer cell movement by deregulating cell-matrix and cell-cell adhesions. *PLoS One* **7**, e38464, doi:10.1371/journal.pone.0038464 (2012).

358 Kanetaka, K. *et al.* Overexpression of tetraspanin CO-029 in hepatocellular carcinoma. *J Hepatol* **35**, 637-642 (2001).

359 Herlevsen, M., Schmidt, D. S., Miyazaki, K. & Zoller, M. The association of the tetraspanin D6.1A with the alpha6beta4 integrin supports cell motility and liver metastasis formation. *J Cell Sci* **116**, 4373-4390, doi:10.1242/jcs.00760 (2003).

360 Park, C. S. *et al.* Therapeutic targeting of tetraspanin8 in epithelial ovarian cancer invasion and metastasis. *Oncogene* **35**, 4540-4548, doi:10.1038/onc.2015.520 (2016).

361 Gesierich, S. *et al.* Colocalization of the tetraspanins, CO-029 and CD151, with integrins in human pancreatic adenocarcinoma: impact on cell motility. *Clinical cancer research : an official journal of the American Association for Cancer Research* **11**, 2840-2852, doi:10.1158/1078-0432.CCR-04-1935 (2005).

362 Bhansali, M., Zhou, J. & Shemshedini, L. TM4SF3 and AR: A Nuclear Complex that Stabilizes Both Proteins. *Mol Endocrinol* **30**, 13-25, doi:10.1210/me.2015-1075 (2016).

363 Zhou, Z. *et al.* TM4SF3 promotes esophageal carcinoma metastasis via upregulating ADAM12m expression. *Clin Exp Metastasis* **25**, 537-548, doi:10.1007/s10585-008-9168-0 (2008).

364 Pan, S. J. *et al.* Over-expression of tetraspanin 8 in malignant glioma regulates tumor cell progression. *Biochem Biophys Res Commun* **458**, 476-482, doi:10.1016/j.bbrc.2015.01.128 (2015).

365 Berthier-Vergnes, O. *et al.* Gene expression profiles of human melanoma cells with different invasive potential reveal TSPAN8 as a novel mediator of invasion. *Br J Cancer* **104**, 155-165, doi:10.1038/sj.bjc.6605994 (2011).

366 El Kharbili, M. *et al.* Tspan8-beta-catenin positive feedback loop promotes melanoma invasion. *Oncogene*, doi:10.1038/s41388-019-0691-z (2019).

367 Yue, S., Mu, W., Erb, U. & Zoller, M. The tetraspanins CD151 and Tspan8 are essential exosome components for the crosstalk between cancer initiating cells and their surrounding. *Oncotarget* **6**, 2366-2384, doi:10.18632/oncotarget.2958 (2015).

368 Lu, J. *et al.* Exosomal tetraspanins mediate cancer metastasis by altering host microenvironment. *Oncotarget* **8**, 62803-62815, doi:10.18632/oncotarget.19119 (2017).

369 Yue, S., Mu, W. & Zoller, M. Tspan8 and CD151 promote metastasis by distinct mechanisms. *Eur J Cancer* **49**, 2934-2948, doi:10.1016/j.ejca.2013.03.032 (2013).

370 Zhu, Y. *et al.* Multi-factorial modulation of colorectal carcinoma cells motility - partial coordination by the tetraspanin Co-029/tspan8. *Oncotarget* **8**, 27454-27470, doi:10.18632/oncotarget.16247 (2017).

371 Zoller, M. CD44: can a cancer-initiating cell profit from an abundantly expressed molecule? *Nat Rev Cancer* **11**, 254-267, doi:10.1038/nrc3023 (2011).

372 Zanjani, L. S. *et al.* Increased expression of CD44 is associated with more aggressive behavior in clear cell renal cell carcinoma. *Biomark Med* **12**, 45-61, doi:10.2217/bmm-2017-0142 (2018).

373 Mizumoto, A. *et al.* Induction of epithelial-mesenchymal transition via activation of epidermal growth factor receptor contributes to sunitinib resistance in human renal cell carcinoma cell lines. *J Pharmacol Exp Ther* **355**, 152-158, doi:10.1124/jpet.115.226639 (2015).

374 Bonnet, M. *et al.* Targeting the Tetraspanins with Monoclonal Antibodies in Oncology: Focus on Tspan8/Co-029. *Cancers (Basel)* **11**, doi:10.3390/cancers11020179 (2019).

375 Maisonial-Besset, A. *et al.* Tetraspanin 8 (TSPAN 8) as a potential target for radio-immunotherapy of colorectal cancer. *Oncotarget* **8**, 22034-22047, doi:10.18632/oncotarget.15787 (2017).

376 The Human Protein Atlas. *Tissue expression of TSPAN8 - Staining in kidney - The Human Protein Atlas*, <https://www.proteinatlas.org/ENSG00000127324-TSPAN8/tissue/kidney> (2019).

377 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419, doi:10.1126/science.1260419 (2015).

378 Champy, M. F. *et al.* Reduced body weight in male Tspan8-deficient mice. *Int J Obes (Lond)* **35**, 605-617, doi:10.1038/ijo.2010.165 (2011).

379 Beavon, I. R. The E-cadherin-catenin complex in tumour metastasis: structure, function and regulation. *Eur J Cancer* **36**, 1607-1620 (2000).

380 Hirukawa, T. *et al.* Kidney Diseases Enhance Expression of Tetraspanin-8: A Possible Protective Effect against Tubular Injury. *Nephron Extra* **4**, 70-81, doi:10.1159/000362451 (2014).

381 Hanif, M. O. & Ramphul, K. in *StatPearls* (2018).

382 Smeets, B. *et al.* Proximal tubular cells contain a phenotypically distinct, scattered cell population involved in tubular regeneration. *The Journal of pathology* **229**, 645-659, doi:10.1002/path.4125 (2013).

383 Berger, K. & Moeller, M. J. Mechanisms of epithelial repair and regeneration after acute kidney injury. *Semin Nephrol* **34**, 394-403, doi:10.1016/j.semnephrol.2014.06.006 (2014).

384 Shankland, S. J., Smeets, B., Pippin, J. W. & Moeller, M. J. The emergence of the glomerular parietal epithelial cell. *Nature reviews Nephrology* **10**, 158 (2014).

385 Lazzeri, E. *et al.* Regenerative potential of embryonic renal multipotent progenitors in acute renal failure. *Journal of the American Society of Nephrology* **18**, 3128-3138 (2007).

386 Lindgren, D. *et al.* Isolation and characterization of progenitor-like cells from human renal proximal tubules. *The American journal of pathology* **178**, 828-837, doi:10.1016/j.ajpath.2010.10.026 (2011).

387 Coca, S. G., Singanamala, S. & Parikh, C. R. Chronic kidney disease after acute kidney injury: a systematic review and meta-analysis. *Kidney Int* **81**, 442-448, doi:10.1038/ki.2011.379 (2012).

388 Lin, M. Y. *et al.* Association of dialysis with the risks of cancers. *PLoS One* **10**, e0122856, doi:10.1371/journal.pone.0122856 (2015).

389 Penas, P. F., Garcia-Diez, A., Sanchez-Madrid, F. & Yanez-Mo, M. Tetraspanins are localized at motility-related structures and involved in normal human keratinocyte wound healing migration. *J Invest Dermatol* **114**, 1126-1135, doi:10.1046/j.1523-1747.2000.00998.x (2000).

390 Sundaram, G. M., Quah, S. & Sampath, P. Cancer: the dark side of wound healing. *FEBS J* **285**, 4516-4534, doi:10.1111/febs.14586 (2018).

391 Scelo, G. *et al.* KIM-1 as a Blood-Based Marker for Early Detection of Kidney Cancer: A Prospective Nested Case-Control Study. *Clinical cancer research : an official journal of the American Association for Cancer Research* **24**, 5594-5601, doi:10.1158/1078-0432.CCR-18-1496 (2018).

392 Carlsson, P. & Mahlapuu, M. Forkhead transcription factors: key players in development and metabolism. *Dev Biol* **250**, 1-23 (2002).

393 Wang, Y. P. & Lei, Q. Y. Metabolic recoding of epigenetics in cancer. *Cancer Commun (Lond)* **38**, 25, doi:10.1186/s40880-018-0302-3 (2018).

394 Cirillo, L. A. *et al.* Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J* **17**, 244-254, doi:10.1093/emboj/17.1.244 (1998).

395 Cirillo, L. A. & Zaret, K. S. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell* **4**, 961-969 (1999).

396 Tuteja, G. & Kaestner, K. H. SnapShot: forkhead transcription factors I. *Cell* **130**, 1160, doi:10.1016/j.cell.2007.09.005 (2007).

397 Tuteja, G. & Kaestner, K. H. Forkhead transcription factors II. *Cell* **131**, 192, doi:10.1016/j.cell.2007.09.016 (2007).

398 Ang, S. L. & Rossant, J. HNF-3 beta is essential for node and notochord formation in mouse development. *Cell* **78**, 561-574 (1994).

399 Weinstein, D. C. *et al.* The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. *Cell* **78**, 575-588 (1994).

400 Wang, J. *et al.* FOXA2 suppresses the metastasis of hepatocellular carcinoma partially through matrix metalloproteinase-9 inhibition. *Carcinogenesis* **35**, 2576-2583, doi:10.1093/carcin/bgu180 (2014).

401 Wang, L., Wu, J. & Xie, C. miR-92a promotes hepatocellular carcinoma cells proliferation and invasion by FOXA2 targeting. *Iran J Basic Med Sci* **20**, 783-790, doi:10.22038/IJBMS.2017.9010 (2017).

402 Liu, J. *et al.* Coordination of FOXA2 and SIRT6 suppresses the hepatocellular carcinoma progression through ZEB2 inhibition. *Cancer Manag Res* **10**, 391-402, doi:10.2147/CMAR.S150552 (2018).

403 Chand, V., Pandey, A., Kopanja, D., Guzman, G. & Raychaudhuri, P. Opposing Roles of the Fork-head box genes FoxM1 and FoxA2 in Liver Cancer. *Mol Cancer Res*, doi:10.1158/1541-7786.MCR-18-0968 (2019).

404 Vorvis, C. *et al.* Transcriptomic and CRISPR/Cas9 technologies reveal FOXA2 as a tumor suppressor gene in pancreatic cancer. *Am J Physiol Gastrointest Liver Physiol* **310**, G1124-1137, doi:10.1152/ajpgi.00035.2016 (2016).

405 Park, J. W., Lee, J. K., Witte, O. N. & Huang, J. FOXA2 is a sensitive and specific marker for small cell neuroendocrine carcinoma of the prostate. *Mod Pathol* **30**, 1262-1272, doi:10.1038/modpathol.2017.44 (2017).

406 Peng, Q. *et al.* Autophagy maintains the stemness of ovarian cancer stem cells by FOXA2. *J Exp Clin Cancer Res* **36**, 171, doi:10.1186/s13046-017-0644-8 (2017).

407 Yi, Y. *et al.* A Murine Model of K-RAS and beta-Catenin Induced Renal Tumors Expresses High Levels of E2F1 and Resembles Human Wilms Tumor. *J Urol* **194**, 1762-1770, doi:10.1016/j.juro.2015.04.090 (2015).

408 Qi, J. *et al.* Siah2-dependent concerted activity of HIF and FoxA2 regulates formation of neuroendocrine phenotype and neuroendocrine prostate tumors. *Cancer Cell* **18**, 23-38, doi:10.1016/j.ccr.2010.05.024 (2010).

409 Thiesen, H. J. *et al.* Stratification of clear cell renal cell carcinoma (ccRCC) genomes by gene-directed copy number alteration (CNA) analysis. *PLoS One* **12**, e0176659, doi:10.1371/journal.pone.0176659 (2017).

410 Robertson, A. G. *et al.* Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. *Genome Res* **18**, 1906-1917, doi:10.1101/gr.078519.108 (2008).

411 Ehrlich, M. *et al.* Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* **10**, 2709-2721 (1982).

412 Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* **39**, 457-466, doi:10.1038/ng1990 (2007).

413 Schubeler, D. Function and information content of DNA methylation. *Nature* **517**, 321-326, doi:10.1038/nature14192 (2015).

414 Gonzalo, S. Epigenetic alterations in aging. *J Appl Physiol (1985)* **109**, 586-597, doi:10.1152/japplphysiol.00238.2010 (2010).

415 Koch, C. M. *et al.* Specific age-associated DNA methylation changes in human dermal fibroblasts. *PLoS One* **6**, e16679, doi:10.1371/journal.pone.0016679 (2011).

416 Wodarz, D., Boland, C. R., Goel, A. & Komarova, N. L. Methylation kinetics and CpG-island methylator phenotype status in colorectal cancer cell lines. *Biol Direct* **8**, 14, doi:10.1186/1745-6150-8-14 (2013).

417 Illumina. *Infinium MethylationEPIC Kit*, <https://www.illumina.com/products/by-type/microarray-kits/infinium-methylation-epic.html> (2019).

418 Mašek, T., Vopalenský, V. & Pospíšek, M. The Luc2 gene enhances reliability of bicistronic assays. *Open Life Sciences* **8**, 423-431 (2013).

419 Sprouffske, K. & Wagner, A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC bioinformatics* **17**, 172 (2016).

420 R: A Language and Environment for Statistical Computing (2018).

421 corrplot: Visualization of a correlation matrix. R package version 0.84 (2017).

422 Ihaka, R. & Gentleman, R. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**, 299-314, doi:10.2307/1390807 (1996).

423 Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* **12**, 115-121, doi:10.1038/nmeth.3252 (2015).

424 readr: read rectangular text data. R package version 1.1.1 (2017).

425 readxl: Read Excel Files. R package version 1.1.0. (2018).

426 ggplot2: elegant graphics for data analysis (Springer, 2016).

427 ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.8.0 (2018).

428 ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.1.8. (2018).

429 Wickham, H. Reshaping data with the reshape package. *Journal of Statistical Software* **21**, 1-20 (2007).

430 dplyr: A Grammar of Data Manipulation. R package version 0.7.6 (R Core Development Team Vienna, 2018).

431 Foreach: provides foreach looping construct for R. R package version 1.4. 4 (Vienna, Austria: The R Foundation. Retrieved from https://CRAN.R-project.org/package=foreach, 2017).

432 knitr: A general-purpose package for dynamic report generation in R, R package version 1.20 (2018).

433 BiocStyle: Standard styles for vignettes and other Bioconductor documents. R package version 2.9.6 (2018).

434 rmarkdown: dynamic documents for R. R package version 1.10 (2018).

435 Shi, W., Oshlack, A. & Smyth, G. K. Optimizing the noise versus bias trade-off for Illumina whole genome expression BeadChips. *Nucleic Acids Res* **38**, e204, doi:10.1093/nar/gkq871 (2010).

436 Hahne, F., Huber, W., Gentleman, R. & Falcon, S. *Bioconductor case studies*. (Springer Science & Business Media, 2010).

437 genefilter: genefilter: methods for filtering genes from high-throughput experiments. R package version 1.63.2. (2018).

438 Kauffmann, A. & Huber, W. Microarray data quality control improves the detection of differentially expressed genes. *Genomics* **95**, 138-142 (2010).

439 Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature protocols* **4**, 1184 (2009).

440 Smyth, G. K. *et al.* (2018).

441 Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of classification* **31**, 274-295 (2014).

442 Ward Jr, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* **58**, 236-244 (1963).

443 Mardia, K., J, K. & Bibby, J. *Multivariate Analysis*. (Academic Press, 1979).

444 Venables, W. N. & Ripley, B. D. in *Modern Applied Statistics with S* 251-269 (Springer, 2002).

445 Assenov, Y. *et al.* Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods* **11**, 1138-1140, doi:10.1038/nmeth.3115 (2014).

446 Fortin, J.-P., Triche Jr, T. J. & Hansen, K. D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**, 558-560 (2016).

447 Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404, doi:10.1158/2159-8290.CD-12-0095 (2012).

448 Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1-pl1 (2013).

449 Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29, doi:10.1186/gb-2014-15-2-r29 (2014).

450 Therneau, T. M. & Grambsch, P. M. *Modeling survival data: extending the Cox model*. (Springer Science & Business Media, 2013).

451 Kassambara, A., Kosinski, M. & Biecek, P. survminer: Drawing Survival Curves using'ggplot2'. *R package version 0.3* **1** (2017).

452 Harrington, D. P. & Fleming, T. R. A class of rank test procedures for censored survival data. *Biometrika* **69**, 553-566 (1982).

453 Harrell, F. E., Jr., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. Evaluating the yield of medical tests. *JAMA* **247**, 2543-2546 (1982).

454 Harrell, F. E., Jr., Lee, K. L. & Mark, D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* **15**, 361-387, doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4 (1996).

455 Love, M. I., Anders, S., Kim, V. & Huber, W. RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Research* **4** (2015).

456 Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 1.33.5 (2018).

457 TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation package for TxDb object(s). R package version 3.2.2. (2018).

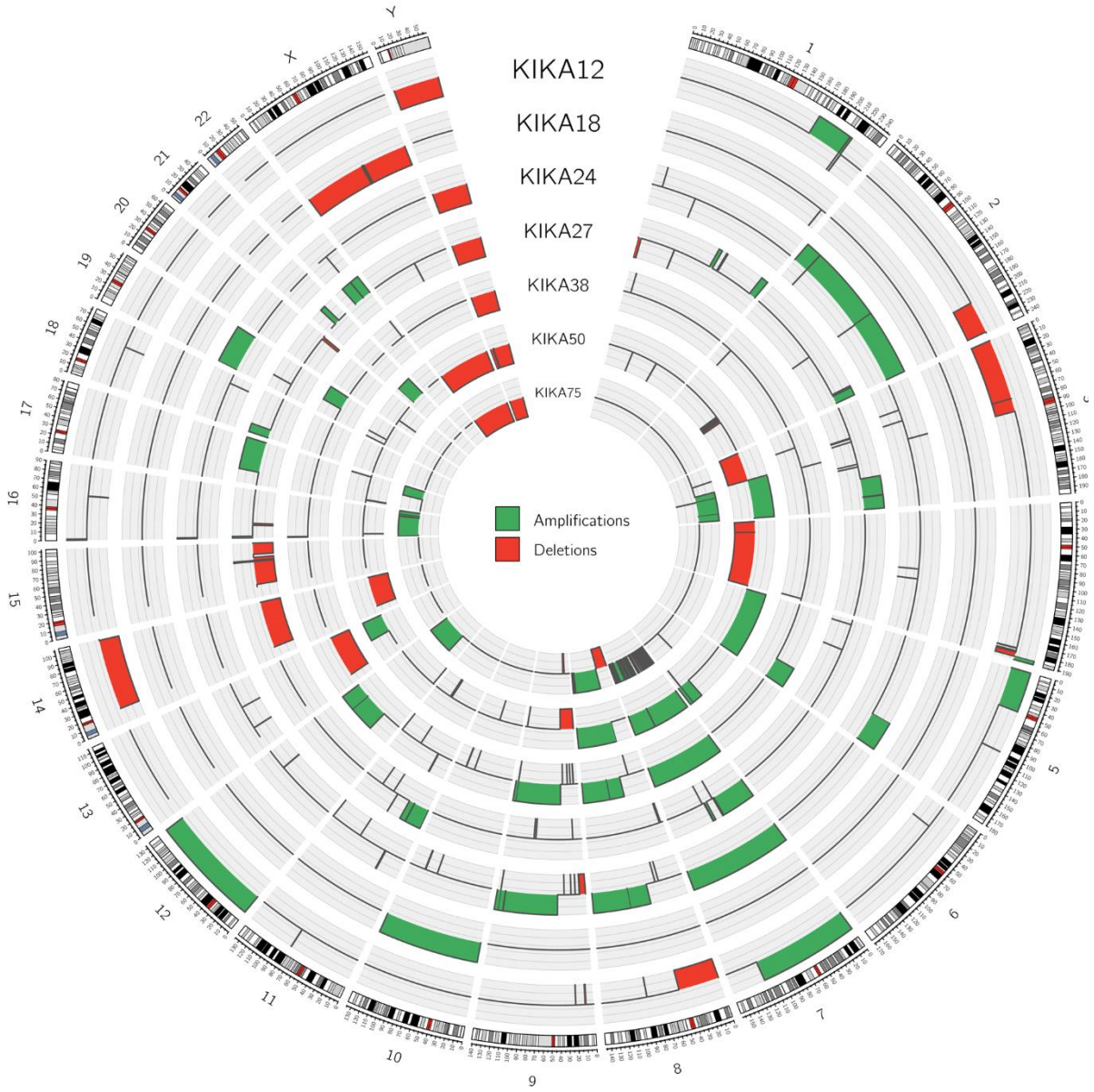
458 Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS computational biology* **9**, e1003118 (2013).

459 Hmisc: Harrell Miscellaneous. R package version 4.1-1 (2018).

460 Sergushichev, A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *BioRxiv*, 060012 (2016).

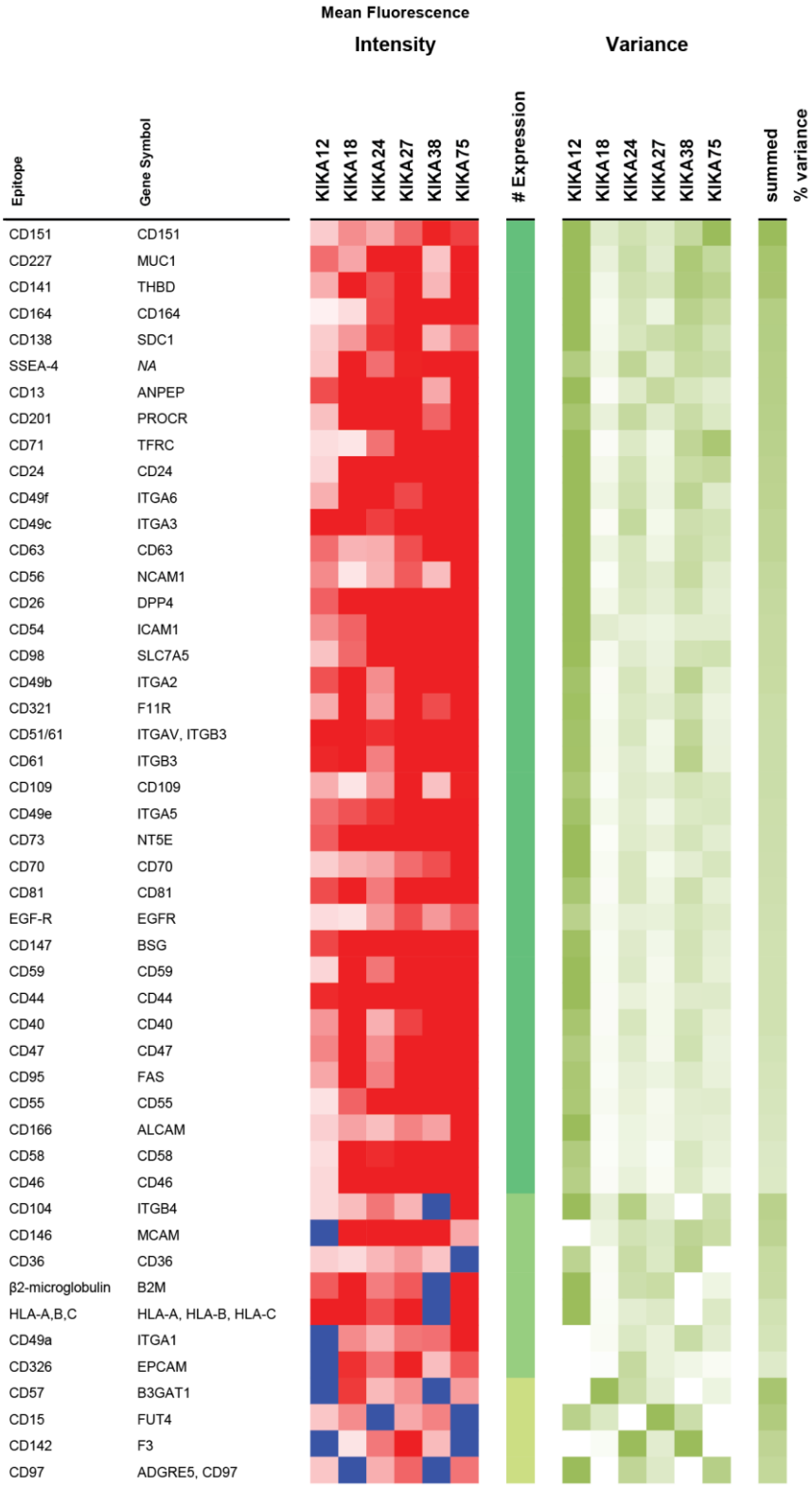
461 Krzywinski, M. I. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Research*, doi:10.1101/gr.092759.109 (2009).

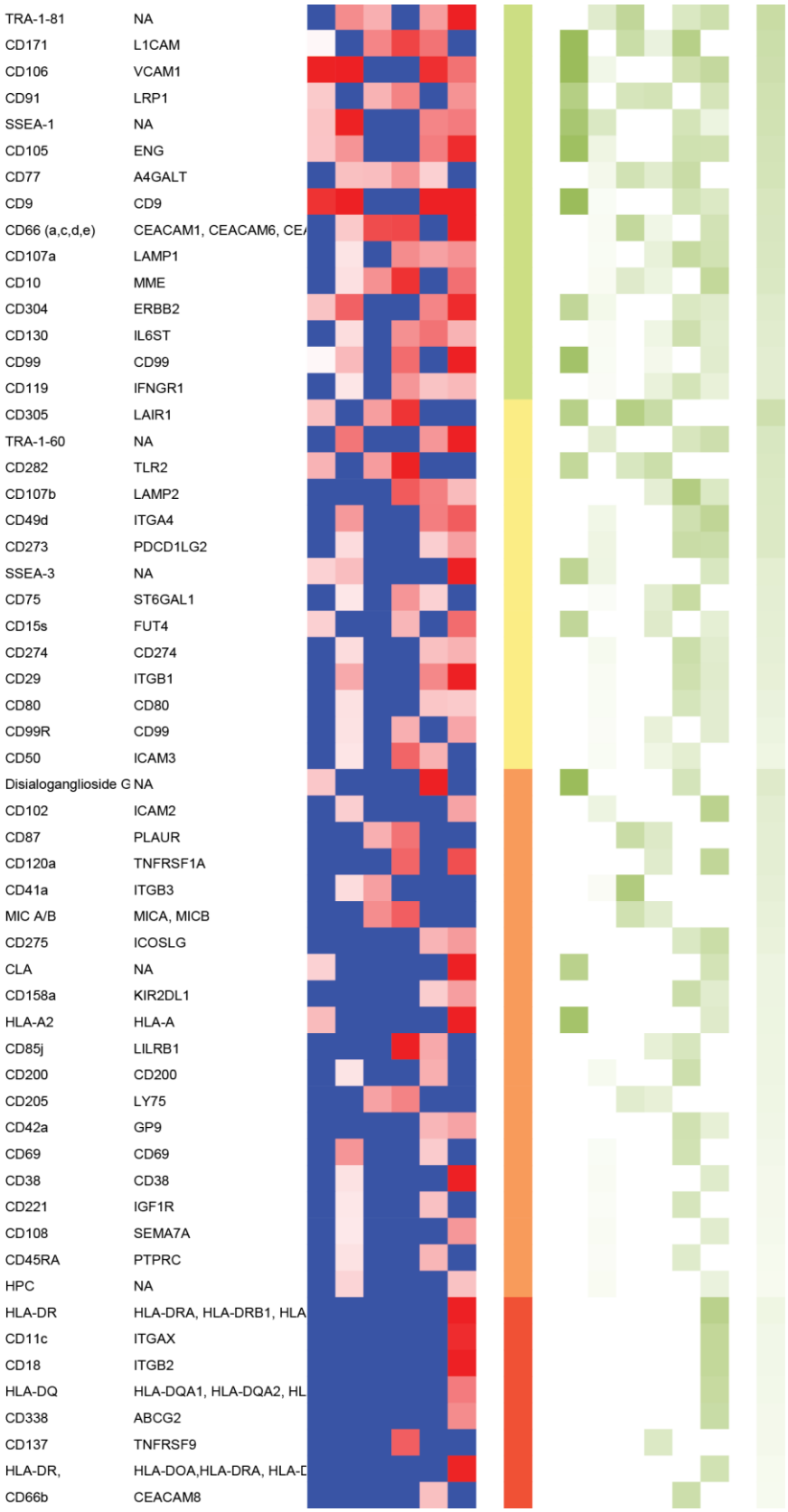
Supplements

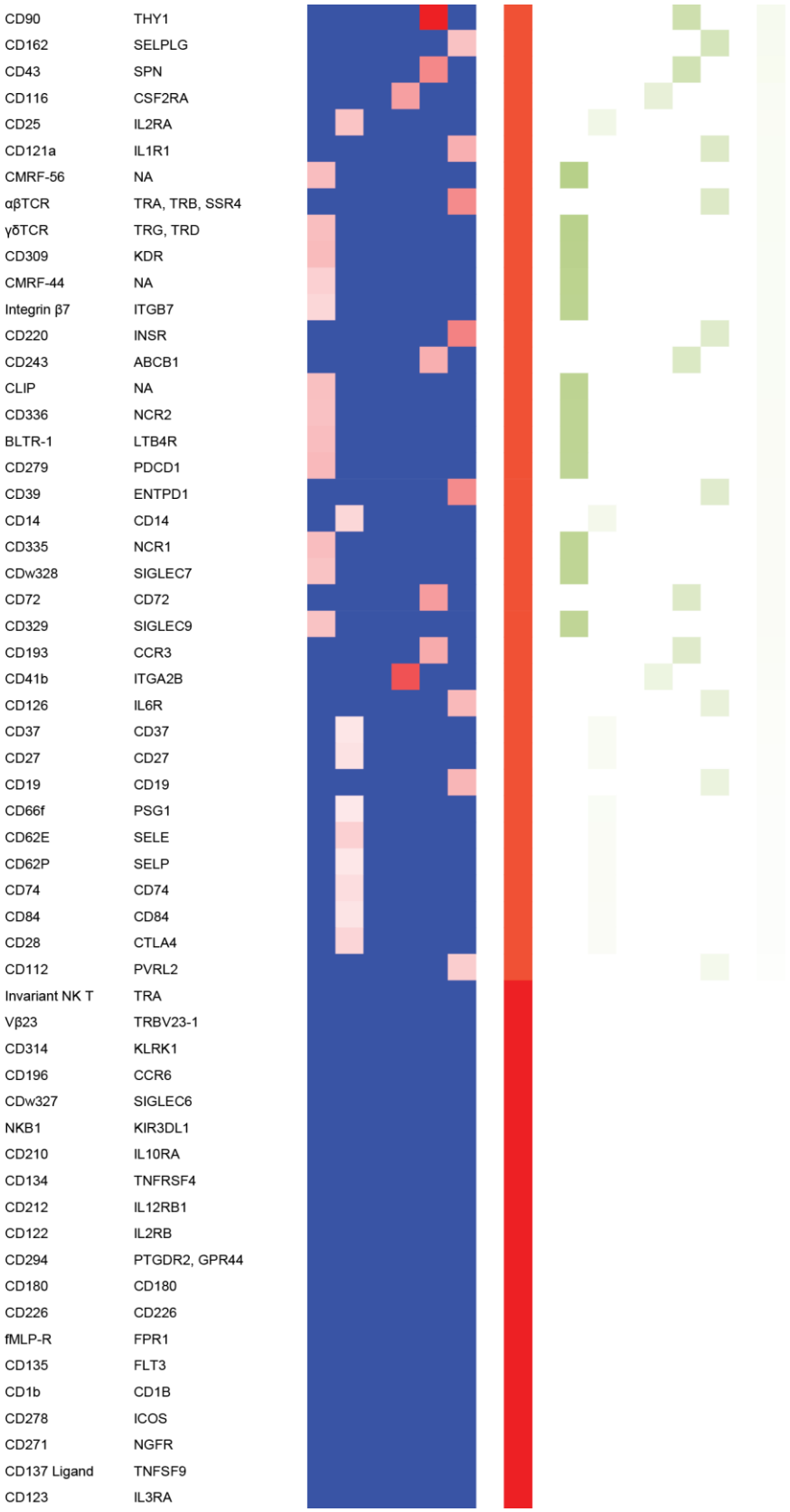


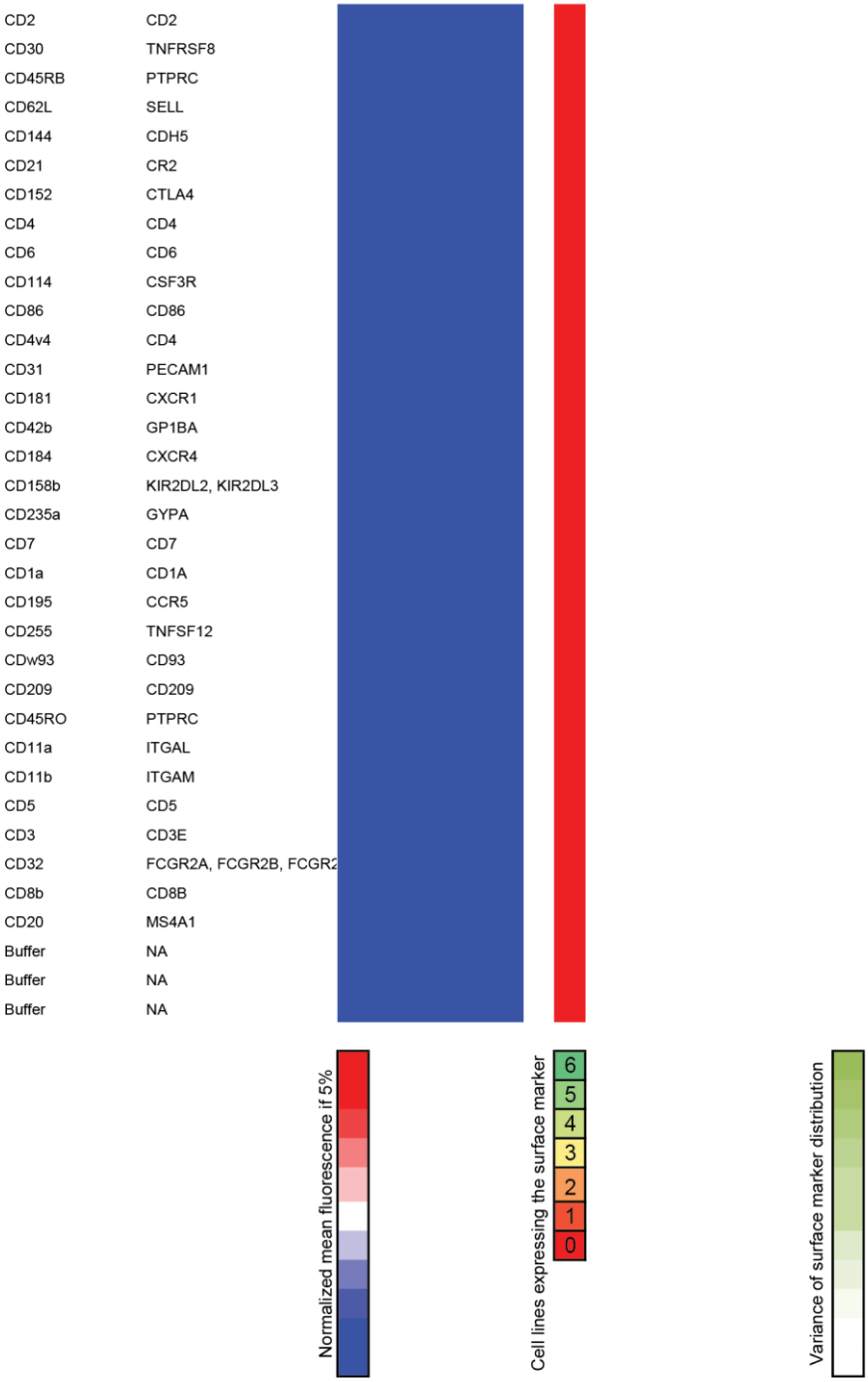
Supplemental Figure 1 – Circos plot of the CNA of all KIKA models

The copy number alterations were identified by the Infinium Omni2.5-8 SNP array.


Mean Fluorescence
Intensity
Variance
Epitope
Gene Symbol
KIKA12
KIKA18
KIKA24
KIKA27
KIKA38
KIKA75
Expression
KIKA12
KIKA18
KIKA24
KIKA27
KIKA38
KIKA75
summed
% variance
CD151 CD151
CD227 MUC1
CD141 THBD
CD164 CD164
CD138 SDC1
SSEA-4 NA
CD13 ANPEP
CD201 PROCR
CD71 TFRC
CD24 CD24
CD49f ITGA6
CD49c ITGA3
CD63 CD63
CD56 NCAM1
CD26 DPP4
CD54 ICAM1
CD98 SLC7A5
CD49b ITGA2
CD321 F11R
CD51/61 ITGAV, ITGB3
CD61 ITGB3
CD109 CD109
CD49e ITGA5
CD73 NT5E
CD70 CD70
CD81 CD81
EGF-R EGFR
CD147 BSG
CD59 CD59
CD44 CD44
CD40 CD40
CD47 CD47
CD95 FAS
CD55 CD55
CD166 ALCAM
CD58 CD58
CD46 CD46
CD104 ITGB4
CD146 MCAM
CD36 CD36
β2-microglobulin B2M
HLA-A,B,C HLA-A, HLA-B, HLA-C
CD49a ITGA1
CD326 EPCAM
CD57 B3GAT1
CD15 FUT4
CD142 F3
CD97 ADGRE5, CD97

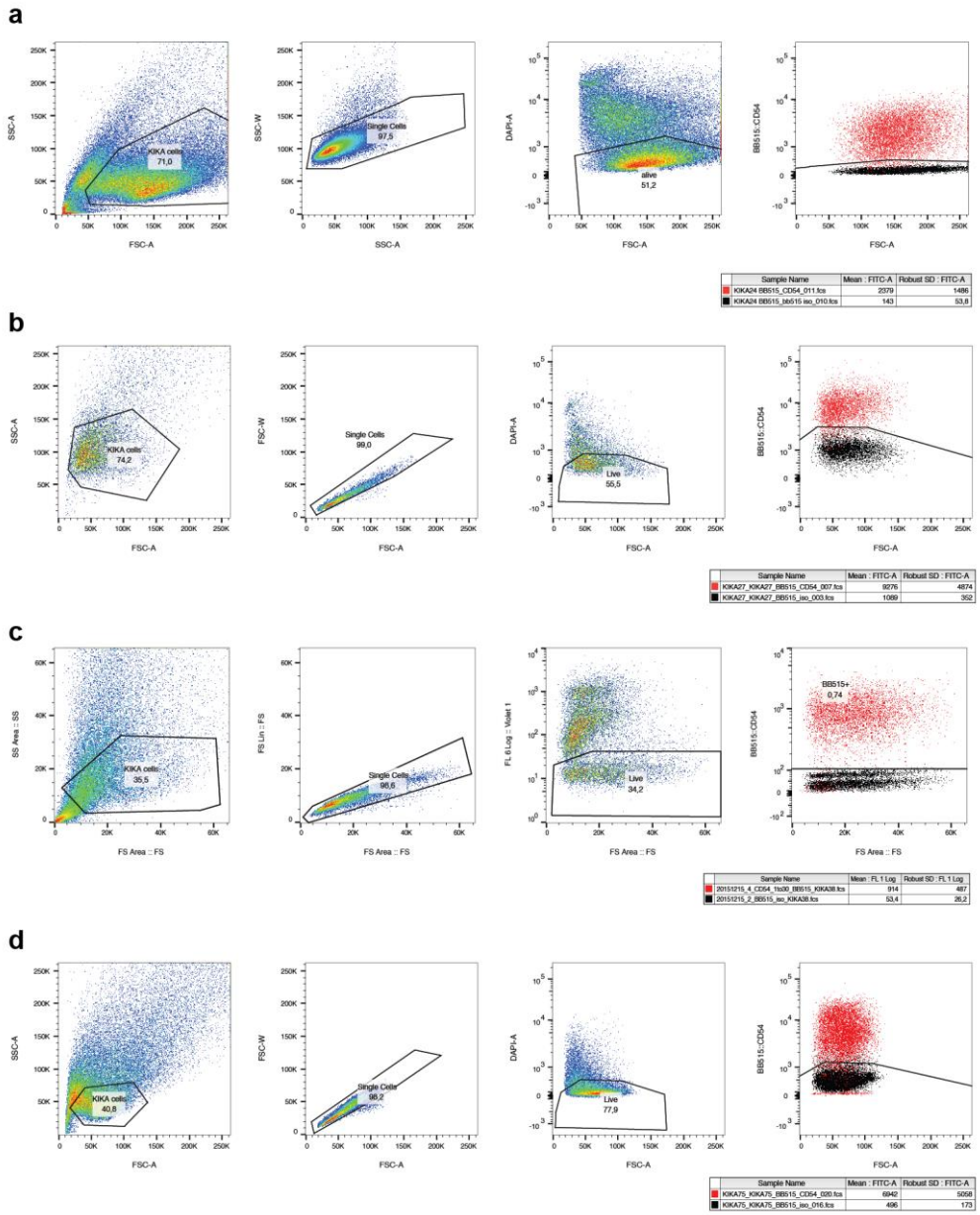

TRA-1-81 NA
CD171 L1CAM
CD106 VCAM1
CD91 LRP1
SSEA-1 NA
CD105 ENG
CD77 A4GALT
CD9 CD9
CD66 (a,c,d,e) CEACAM1, CEACAM6, CE
CD107a LAMP1
CD10 MME
CD304 ERBB2
CD130 IL6ST
CD99 CD99
CD119 IFNGR1
CD305 LAIR1
TRA-1-60 NA
CD282 TLR2
CD107b LAMP2
CD49d ITGA4
CD273 PDCD1LG2
SSEA-3 NA
CD75 ST6GAL1
CD15s FUT4
CD274 CD274
CD29 ITGB1
CD80 CD80
CD99R CD99
CD50 ICAM3
Disialoganglioside G NA
CD102 ICAM2
CD87 PLAUR
CD120a TNFRSF1A
CD41a ITGB3
MIC A/B MICA, MICB
CD275 ICOSLG
CLA NA
CD158a KIR2DL1
HLA-A2 HLA-A
CD85j LILRB1
CD200 CD200
CD205 LY75
CD42a GP9
CD69 CD69
CD38 CD38
CD221 IGF1R
CD108 SEMA7A
CD45RA PTPRC
HPC NA
HLA-DR HLA-DRA, HLA-DRB1, HLA
CD11c ITGAX
CD18 ITGB2
HLA-DQ HLA-DQA1, HLA-DQA2, HL
CD338 ABCG2
CD137 TNFRSF9
HLA-DR, HLA-DOA,HLA-DRA, HLA-D
CD66b CEACAM8


CD90 THY1
CD162 SELPLG
CD43 SPN
CD116 CSF2RA
CD25 IL2RA
CD121a IL1R1
CMRF-56 NA
αβTCR TRA, TRB, SSR4
γδTCR TRG, TRD
CD309 KDR
CMRF-44 NA
Integrin β7 ITGB7
CD220 INSR
CD243 ABCB1
CLIP NA
CD336 NCR2
BLTR-1 LTB4R
CD279 PDCD1
CD39 ENTPD1
CD14 CD14
CD335 NCR1
CDw328 SIGLEC7
CD72 CD72
CD329 SIGLEC9
CD193 CCR3
CD41b ITGA2B
CD126 IL6R
CD37 CD37
CD27 CD27
CD19 CD19
CD66f PSG1
CD62E SELE
CD62P SELP
CD74 CD74
CD84 CD84
CD28 CTLA4
CD112 PVRL2
Invariant NK T TRA
Vβ23 TRBV23-1
CD314 KLRK1
CD196 CCR6
CDw327 SIGLEC6
NKB1 KIR3DL1
CD210 IL10RA
CD134 TNFRSF4
CD212 IL12RB1
CD122 IL2RB
CD294 PTGDR2, GPR44
CD180 CD180
CD226 CD226
fMLP-R FPR1
CD135 FLT3
CD1b CD1B
CD278 ICOS
CD271 NGFR
CD137 Ligand TNFSF9
CD123 IL3RA



Supplemental Figure 2 – Immunophenotypisation of the KIKA cell lines

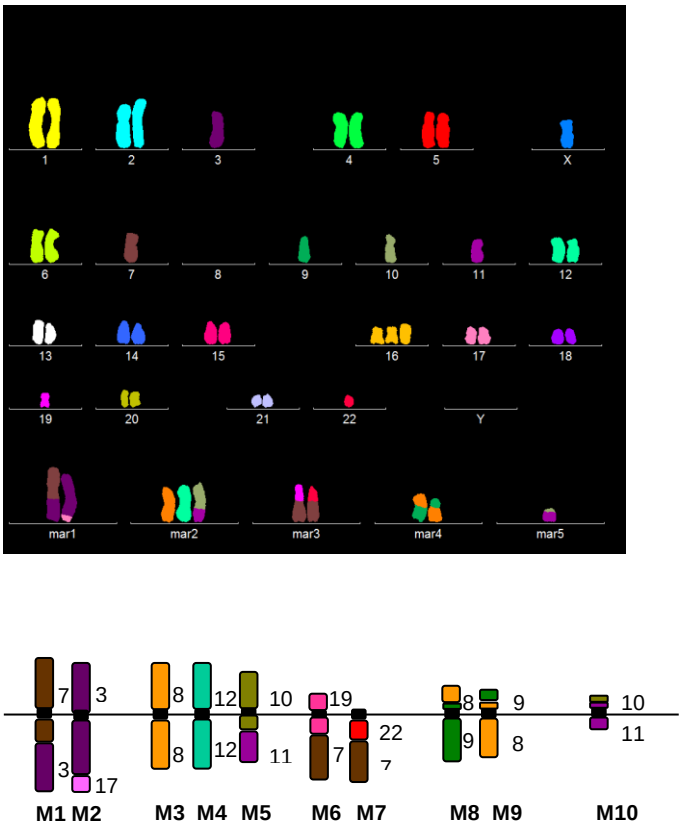
Cell surface markers measured by flow cytometry were sorted according to the number of cell lines expressing a surface marker (5% above respective isotype) and according to the standardized robust coefficient of variation (CV).



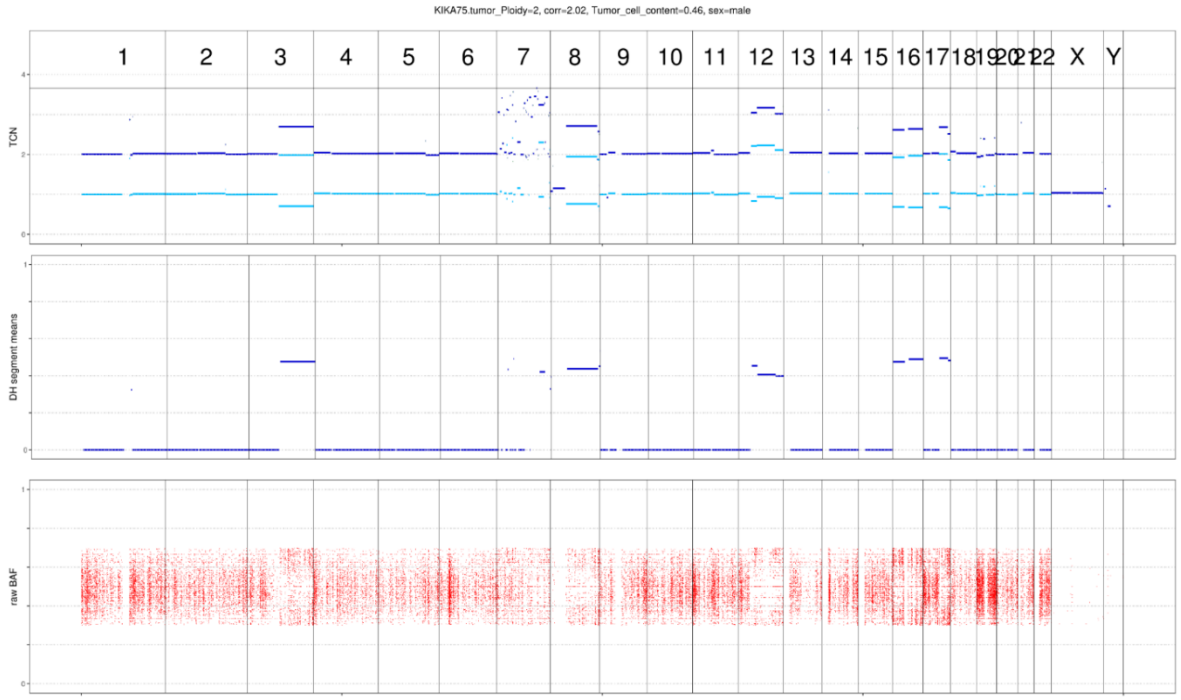
Supplemental Figure 3 – Gating Scheme BB515::CD54/ICAM1

(a) KIKA24, **(b)** KIKA27, **(c)** KIKA38, **(d)** KIKA75

a

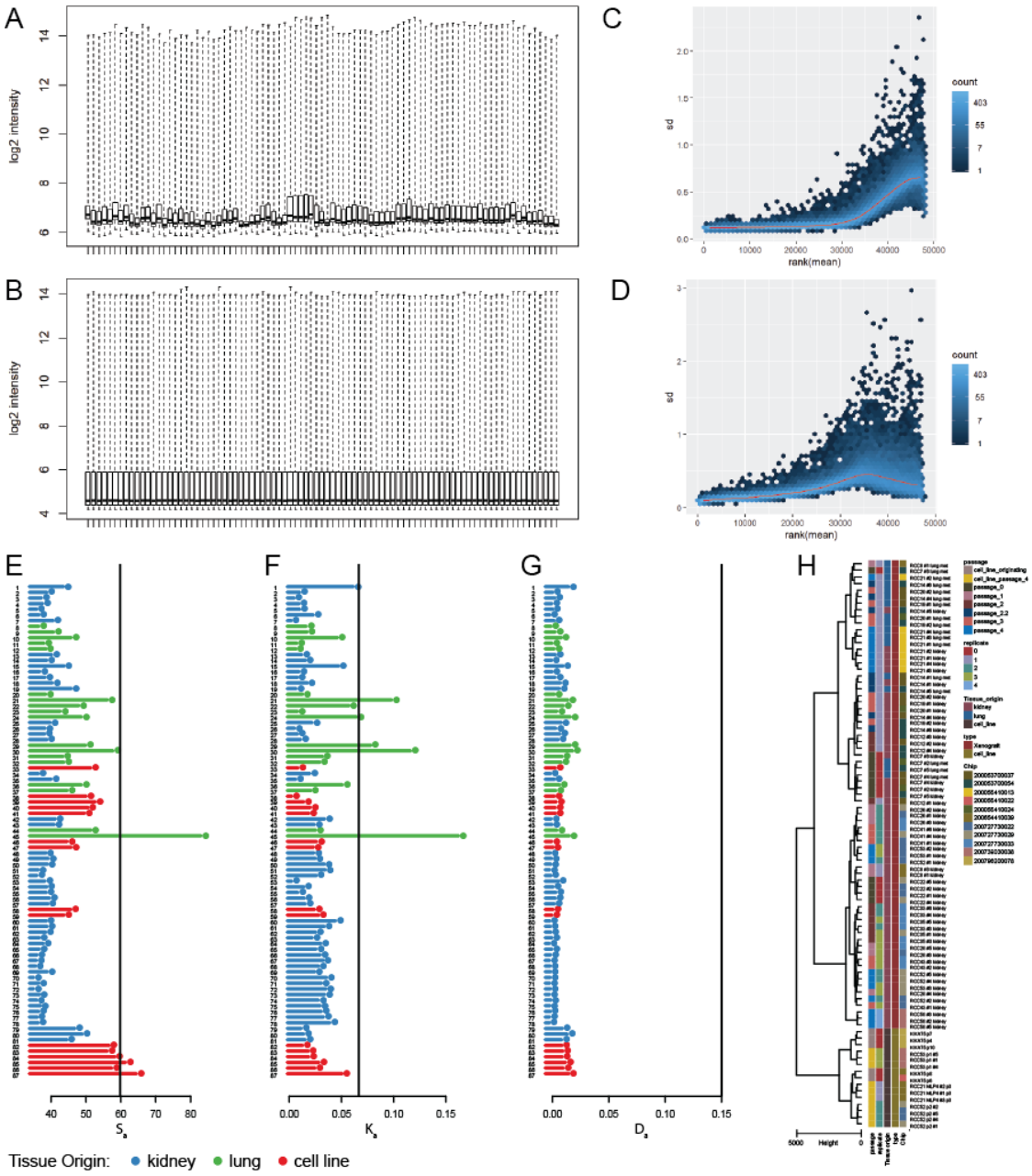


b



Supplemental Figure 4 – Genome of KIKA75

a) Karyogram and chromosome painting of the KIKA75 cell line, made by Larisa Savelyeva
b) Copy number alterations of KIKA75 by whole exome sequencing, analyzed by Gregor Warsow



Supplemental Figure 5 – Quality control of microarray data

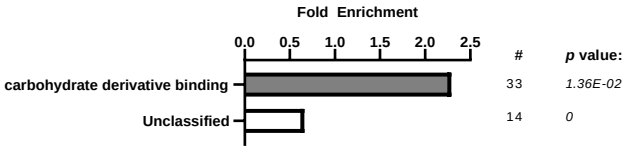
Boxplot of Log2 probe intensities of the 87 samples analyzed Illumina HT12v4 BeadArrays before **(A)** and after **(B)** background correction and quantile-normalization. **(C)** Not normalized and **(D)** background corrected and quantile-normalized ranked log2 means of all probes on all arrays against their standard deviations

ArrayQualityMetrics quality control: Bar chart and figure legends were adapted from the package output

(E) Outlier detection for Distances between arrays: Sum of distances to other arrays S_a. Based on the distribution of the values across all arrays, a threshold of 59.8 was determined, 3 arrays exceeded the threshold and were considered outliers. **(F)** Outlier detection for Boxplots: Kolmogorov-Smirnov statistic K_a. Based on the distribution of the values across all arrays, a threshold of 0.0667 was determined, 5 arrays exceeded the threshold and were considered outliers. **(G)** Outlier detection for MA plots: D_a. A threshold of 0.15 was used, None of the arrays exceeded the threshold and was considered an outlier. Vertical lines indicate outlier thresholds. **(H)** Dendrogram of the standardized samples (z-scores) with Manhattan distance and Ward's linkage. Cell line derived cluster apart from xenograft derived RNA expression profiles.

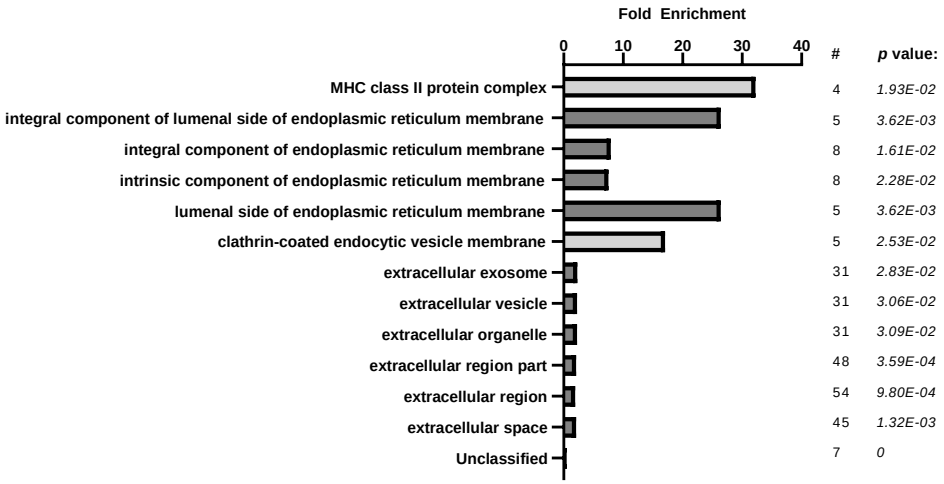
a

GO molecular function



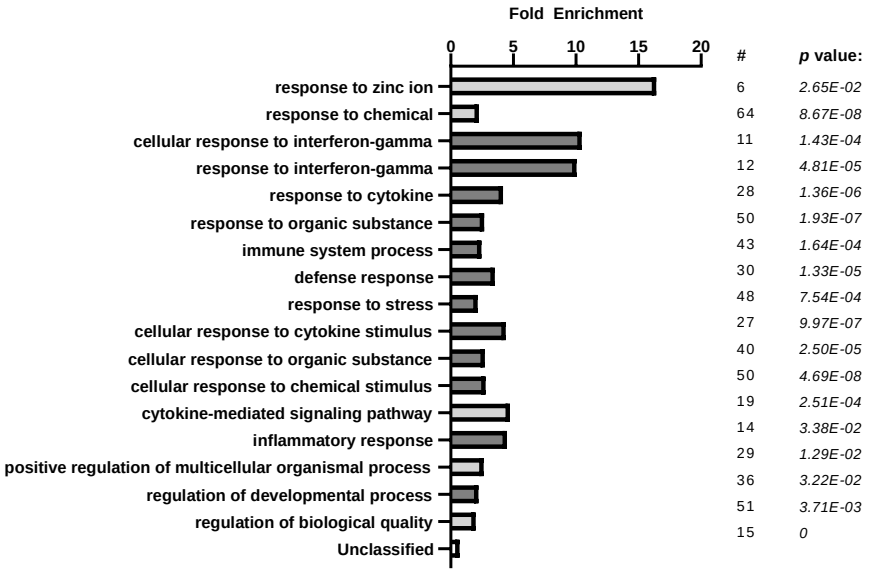
a

GO cellular component



c

GO biological process

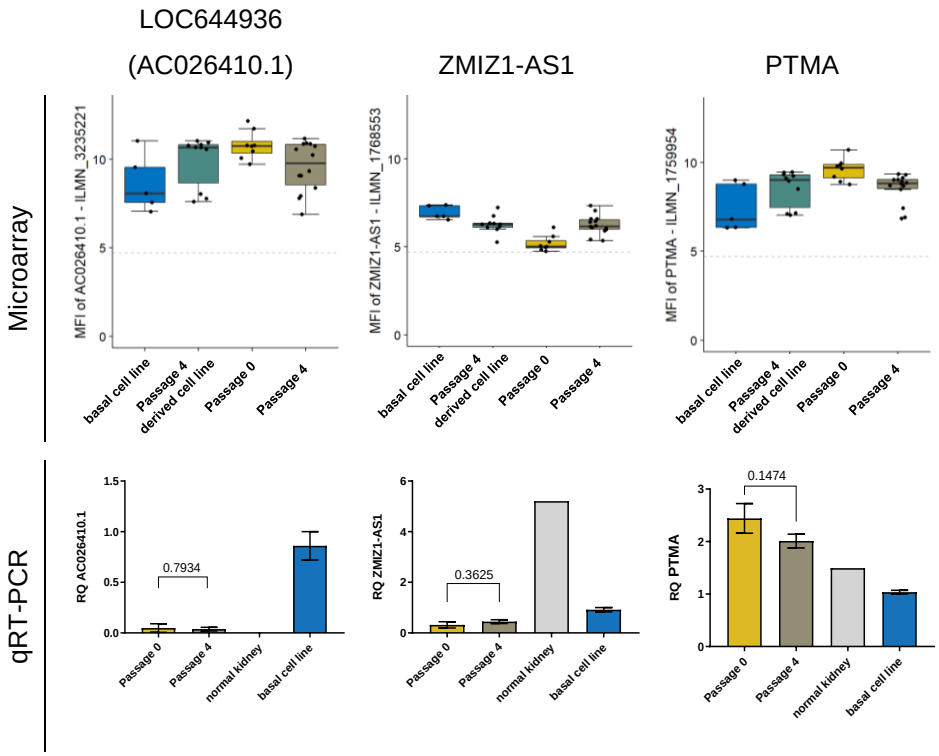


Supplemental Figure 6 – PANTHER overrepresentation test for the gene ontology of the differentially expressed genes, that are not differentially expressed between passage 4 lung metastases derived cell lines and the originating cell line

Terms of the same biological items are colored identically. PANTHER Overrepresentation Test (Released 2018-11-13), GO Ontology database released 2018-12-01, Fisher's exact test, Bonferroni correction for multiple testing.

Characteristic	TCGA KIRC (n = 532)	Sato et al. (n = 100)
Age – yr		
Mean	60.6 ± 12.1	63.7 ± 11.4
Median (range)	61.0 (26-90)	64 (25-91)
Sex – no. (%)		
Female	188 (35.3)	23 (23)
Male	344 (64.7)	77 (77)
Race – no. (%)		
Asian	8 (1.5)	
Black or African American	55 (10.3)	
Caucasian	462 (86.8)	
Not Available	7 (1.3)	
Ethnicity – no. (%)		
Hispanic or Latino	26 (4.9)	
Other	354 (66.5)	
Not Available	152 (28.6)	
TNM Classification – no. (%)		
Stage I	266 (50.0)	65 (65)
Stage II	57 (10.7)	10 (10)
Stage III	123 (23.1)	13 (13)
Stage IV	83 (15.6)	12 (12)
Discrepancy	3 (0.6)	
Fuhrman Grade – no. (%)		
G1	14 (2.6)	13 (13)
G2	228 (42.9)	58 (58)
G3	206 (38.7)	22 (22)
G4	76 (14.3)	5 (5)
GX / Not Available	8 (1.5)	2 (2)
Laterality – no. (%)		
Left	251 (47.2)	
Right	280 (52.6)	
Bilateral	1 (0.2)	
VHL mutated – no. (%)	225 (42.3)	66 (66)
History of Neoadjuvant Treatment – no. (%)	17 (3.2)	100 (100)
White Blood Cell counts – no. (%)		
Normal	266 (50.0)	
Elevated	164 (30.8)	
Low	8 (1.5)	
Not Available	94 (17.7)	
Hemoglobin level – no. (%)		
Normal	184 (15.2)	
Elevated	5 (0.9)	
Low	262 (49.2)	
Not Available	81 (34.6)	
Platelet Count – no. (%)		
Normal	358 (67.3)	
Elevated	37 (7.0)	
Low	46 (8.6)	
Not Available	91 (17.1)	
IDH Level – no. (%)		
Normal	72 (13.5)	
Elevated	12 (2.2)	
Not Available	448 (84.2)	
Median follow-up	1199 days	1536 days
Total no. of deaths	175	23
No. of deaths from ccRCC	104	

Supplemental Figure 7 – Patient classification of the TCGA KIRC and Sato *et al.* ccRCC cohort



Supplemental Figure 8 – Excluded differential expressed genes

Genes of the in vivo selection that were differentially expressed between the microarrays, but could not be validated by quantitative TaqMan real-time PCR. *P* values calculated using an unpaired, two-sided Student's *t*-test. Error bars depict mean ± s.e.m.

Symbol (Entrez ID)	Center	Scale
IL13RA2 (3598)	-0.215532994647864	1.94353172038516
SLC2A9 (56606)	4.0200715670196	1.22002756911761
PTPN13 (5783)	5.36338272647119	0.854811787225463
CXCL5 (6374)	0.363609006323805	3.24779314225545
PRDX2 (7001)	7.29049780755846	0.802574629971008
TSPAN8 (7103)	-0.878408470041345	2.79943988032313
ONECUT2 (9480)	-1.57348772908843	2.19490878291467

Supplemental Figure 9 – Scaling factors from the TCGA-KIRC dataset used to standardize microarray data

Characteristic	Heidelberg (n = 13)
Age – yr	
Mean	64.0 ± 11.63
Median (range)	64.0 (49-87)
Sex – no. (%)	
Female	2 (15,4)
Male	11 (84.6)
TNM Classification – no. (%)	
Stage I	3 (23.1)
Stage II	0 (0)
Stage III	3 (23.1)
Stage IV	7 (53.8)
Fuhrman Grade – no. (%)	
G1	1 (7.7)
G2	5 (38.5)
G3	5 (38.5)
G4	2 (15.4)
Laterality – no. (%)	
Left	6 (46.2)
Right	5 (38.5)
Bilateral	1 (7.7)
NA	1 (7.7)
Median follow-up	431 days
Total no. of deaths	2 (15.4)
No. of deaths from ccRCC	1 (7.7)

Supplemental Figure 10 – Patient classification of the Heidelberg ccRCC cohort

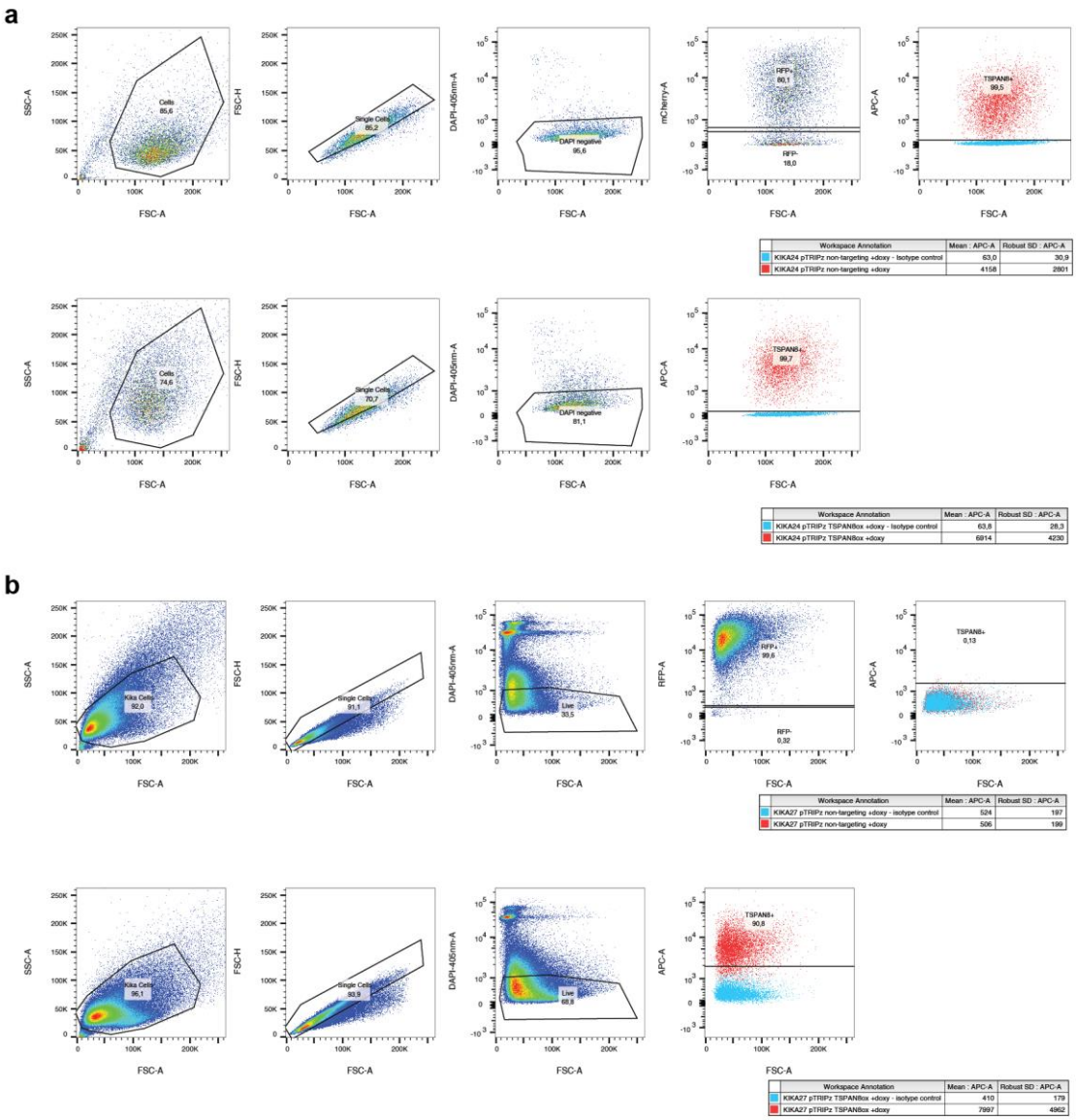
Patient	PFS event	PFS days	Stage	ccRCC Score	ccRCC Score - 3 risk groups
RCC8	0	1443	III	0.3968	intermediate
RCC12	1	77	III	1.4033	high
RCC15	0	5	IV	1.7475	high
RCC18	0	431	IV	1.8590	high
RCC20	1	462	IV	1.8051	high
RCC23	0	1862	I	-1.0635	intermediate
RCC24	1	16	IV	3.7268	high
RCC27	1	22	IV	2.9468	high
RCC28	1	530	I	-1.0825	intermediate
RCC38	0	157	IV	3.6582	high
RCC40	0	1145	I	-1.3892	intermediate
RCC50	1	9	IV	2.9063	high
RCC75	1	90	III	1.7267	high

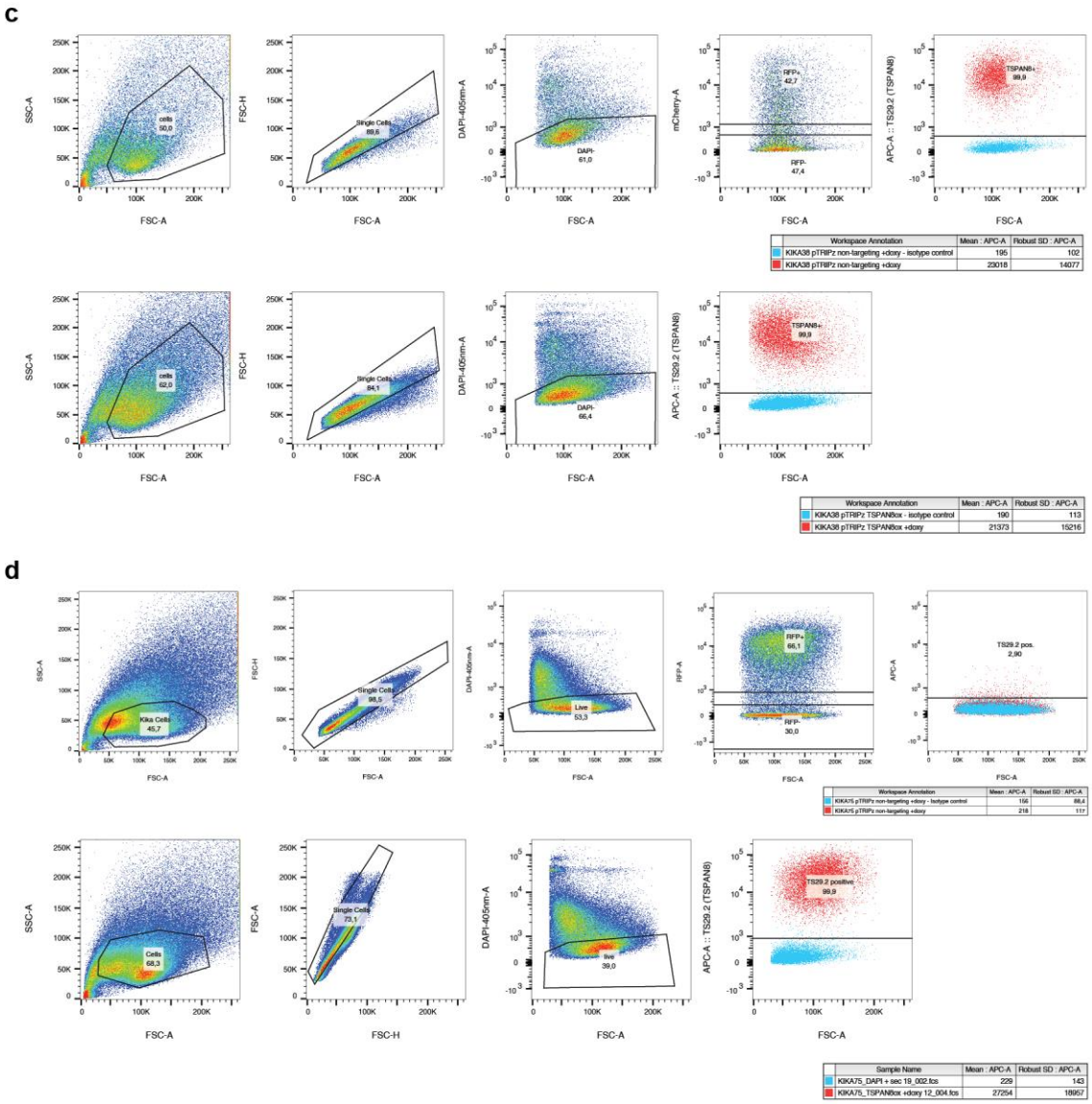
Supplemental Figure 11 – Heidelberg Mini Cohort Patient Classification

Patient	Region	Distance	Risk Catagory	5-year survival rate [%]	Hazard Rate	ccRCC Score	Hex-Color-Code
EV001	M2a	68	High Risk	24.52	2.79	1.985	#B25700
EV001	M2b	61	Intermediate Risk	78.02	1.17	0.251	#DDB400
EV001	R1	71	High Risk	32.85	2.74	1.752	#B86400
EV001	R2	72	High Risk	43.08	2.63	1.473	#BE7100
EV001	R3	70	High Risk	12.18	2.95	2.389	#A84000
EV001	R4a	48	Intermediate Risk	82.28	0.77	0.010	#E2C100
EV001	R5	68	High Risk	28.44	2.76	1.873	#B55C00
EV001	R8	69	High Risk	17.15	2.86	2.211	#AC4A00
EV002	R1	41	High Risk	0.06	4.50	3.642	#8B0000
EV002	R3	53	High Risk	0.31	4.14	3.396	#900C00
EV002	R4	46	High Risk	23.59	2.79	2.012	#B15400
EV002	R7	47	High Risk	1.55	3.65	3.071	#971C00
EV002	R9	52	High Risk	1.56	3.62	3.070	#971C00
EV003	R1	35	High Risk	29.04	2.76	1.857	#B55C00
EV003	R2	38	High Risk	31.31	2.75	1.794	#B76100
EV003	R5	39	High Risk	0.56	3.94	3.290	#931200
EV003	R6	45	High Risk	0.24	4.14	3.441	#8F0A00
EV003	R7	36	High Risk	32.43	2.74	1.763	#B76100
EV005	R1	43	High Risk	18.76	2.85	2.159	#AE4D00
EV005	R2	46	High Risk	1.62	3.62	3.061	#991E00
EV005	R3	43	High Risk	2.65	3.46	2.934	#9B2400
EV005	R4	41	High Risk	40.85	2.67	1.534	#BD6E00
EV005	R5	42	High Risk	60.92	2.22	0.943	#CB8D00
EV005	R6dom	45	High Risk	44.25	2.63	1.440	#BF7300
EV005	R7	48	High Risk	45.22	2.61	1.413	#C07600
EV006	R1	59	Intermediate Risk	75.84	1.38	0.359	#DAAF00
EV006	R2	60	Intermediate Risk	74.28	1.52	0.431	#D8AA00
EV006	R3	58	High Risk	46.85	2.58	1.368	#C27900
EV006	R4	59	High Risk	66.51	1.97	0.748	#D09700
EV006	R7	64	Intermediate Risk	74.29	1.52	0.431	#D8AA00
EV006	R15	60	High Risk	73.67	1.54	0.459	#D7A700
EV006	N1a	57	Intermediate Risk	81.36	0.88	0.066	#E1BE00
EV006	N1b	57	High Risk	64.87	2.02	0.807	#CE9500
EV007	R1	34	High Risk	7.49	3.09	2.597	#A33600
EV007	R2	36	High Risk	2.79	3.46	2.919	#9B2400
EV007	R3min	32	High Risk	12.73	2.92	2.368	#A94200
EV007	R4	33	High Risk	39.15	2.69	1.580	#BC6C00
EV007	R5	29	High Risk	17.55	2.85	2.198	#AC4A00
EV007	R6	45	High Risk	12.63	2.92	2.372	#A94200
EV007	R9dom	32	High Risk	49.65	2.52	1.288	#C37B00
RMH002	R1	24	High Risk	49.27	2.55	1.299	#C37B00
RMH002	R2	17	High Risk	19.18	2.85	2.146	#AE4D00
RMH002	R3	30	High Risk	44.20	2.63	1.442	#BF7300
RMH002	R6	29	High Risk	22.58	2.81	2.042	#B15400
RMH004	R2	70	High Risk	1.18	3.65	3.135	#961900
RMH004	R3	41	High Risk	51.85	2.49	1.224	#C58000
RMH004	R4	60	High Risk	2.99	3.46	2.900	#9C2600
RMH004	R8	54	High Risk	7.81	3.09	2.581	#A33600

RMH004	VT	42	High Risk	52.54	2.49	1.204	#C58000
RMH008	R1	37	Intermediate Risk	88.60	0.03	-0.467	#EEDA00
RMH008	R2	45	Intermediate Risk	88.14	0.10	-0.426	#EDD800
RMH008	R3	38	Intermediate Risk	90.17	-0.18	-0.624	#F2E200
RMH008	R4min	41	Intermediate Risk	89.08	-0.03	-0.513	#EFDD00
RMH008	R6min	40	Intermediate Risk	88.64	0.02	-0.471	#EEDA00
RMH008	R8	38	Intermediate Risk	92.12	-0.48	-0.856	#F7EF00
RK26	R1	31	Intermediate Risk	94.26	-0.86	-1.184	#FFFF00
RK26	R2	34	Intermediate Risk	94.11	-0.82	-1.158	#FFFF00
RK26	R3	30	Intermediate Risk	93.37	-0.68	-1.035	#FBF700
RK26	R4	30	Intermediate Risk	93.13	-0.65	-0.999	#FBF700
RK26	R5dom	13	Intermediate Risk	91.87	-0.43	-0.822	#F6EC00
RK26	R7	9	Intermediate Risk	87.86	0.12	-0.400	#ECD500
RK26	R8	18	Intermediate Risk	89.72	-0.12	-0.577	#F0E000
RK26	R10	17	Intermediate Risk	86.21	0.34	-0.264	#E8CE00
RK26	R11	19	Intermediate Risk	87.25	0.18	-0.348	#EBD300

Supplemental Figure 12 – Multiregion ccRCC Score analysis


a
SSC-A
FSC-A
Cells
85,6
FSC-H
Single Cells
86,2
DAPI-405nm-A
DAPI negative
95,6
mCherry-A
RFP+
99,1
RFP-
18,0
APC-A
TSPAN8+
99,5
Workspace Annotation | Mean : APC-A | Robust SD : APC-A
KIKA24 pTRIPz non-targeting +doxy - Isotype control | 63,0 | 30,9
KIKA24 pTRIPz non-targeting +doxy | 4158 | 2801
Cells
74,6
Single Cells
70,7
DAPI negative
81,1
TSPAN8+
99,7
Workspace Annotation | Mean : APC-A | Robust SD : APC-A
KIKA24 pTRIPz TSPAN8ox +doxy - Isotype control | 63,8 | 28,3
KIKA24 pTRIPz TSPAN8ox +doxy | 6914 | 4230
b
Kika Cells
92,0
Single Cells
91,1
Live
33,5
RFP-A
RFP+
99,6
RFP-
0,32
TSPAN8+
0,13
Workspace Annotation | Mean : APC-A | Robust SD : APC-A
KIKA27 pTRIPz non-targeting +doxy - isotype control | 524 | 197
KIKA27 pTRIPz non-targeting +doxy | 506 | 199
Kika Cells
96,1
Single Cells
93,9
Live
68,8
TSPAN8+
90,8
Workspace Annotation | Mean : APC-A | Robust SD : APC-A
KIKA27 pTRIPz TSPAN8ox +doxy - isotype control | 410 | 179
KIKA27 pTRIPz TSPAN8ox +doxy | 7997 | 4962



Supplemental Figure 13 – Gating Scheme of TS29.2

(a) KIKA24, **(b)** KIKA27, **(c)** KIKA38 and **(d)** KIKA75

Patient	CSS days	CSS event	ccRCC early score	ccRCC early Score - 3 risk groups	ccRCC early Score - 2 risk groups	ccRCC Score	ccRCC Score - 3 risk groups	ccRCC Score - 2 risk groups	Grade	Gender
TCGA-6D-AA2E	362	0	-1.4140	Intermediate Risk	Low Risk	-1.5342	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-A3-3306	1119	0	-2.8345	Low Risk	Low Risk	-3.0754	Low Risk	Low Risk	G3	MALE
TCGA-A3-3307	1435	0	-0.5873	Intermediate Risk	High Risk	-0.6372	Intermediate Risk	High Risk	G3	MALE
TCGA-A3-3308	16	0	0.4782	High Risk	High Risk	0.5188	High Risk	High Risk	G2	FEMALE
TCGA-A3-3311	1190	0	-1.3597	Intermediate Risk	Low Risk	-1.4752	Intermediate Risk	Low Risk	G2	MALE
TCGA-A3-3313	735	0	-2.3533	Low Risk	Low Risk	-2.5533	Low Risk	Low Risk	G3	MALE
TCGA-A3-3316	1492	0	0.8343	High Risk	High Risk	0.9052	High Risk	High Risk	G3	MALE
TCGA-A3-3317	1490	0	-0.2335	Intermediate Risk	High Risk	-0.2533	Intermediate Risk	High Risk	G2	MALE
TCGA-A3-3319	1129	0	-1.0845	Intermediate Risk	Low Risk	-1.1766	Intermediate Risk	Low Risk	G2	MALE
TCGA-A3-3320	1507	0	-1.8050	Low Risk	Low Risk	-1.9584	Low Risk	Low Risk	G1	FEMALE
TCGA-A3-3322	1477	0	-1.7014	Low Risk	Low Risk	-1.8460	Low Risk	Low Risk	G2	MALE
TCGA-A3-3323	1105	0	-1.4207	Intermediate Risk	Low Risk	-1.5415	Intermediate Risk	Low Risk	G1	MALE
TCGA-A3-3324	1185	0	-2.1367	Low Risk	Low Risk	-2.3183	Low Risk	Low Risk	G3	MALE
TCGA-A3-3325	1169	0	-1.3895	Intermediate Risk	Low Risk	-1.5076	Intermediate Risk	Low Risk	G2	MALE
TCGA-A3-3326	1136	0	-2.0203	Low Risk	Low Risk	-2.1920	Low Risk	Low Risk	G1	MALE
TCGA-A3-3328	1384	0	-1.2453	Intermediate Risk	Low Risk	-1.3512	Intermediate Risk	Low Risk	G2	MALE
TCGA-A3-3329	1623	0	-1.9442	Low Risk	Low Risk	-2.1094	Low Risk	Low Risk	G2	MALE
TCGA-A3-3331	1484	0	-1.7507	Low Risk	Low Risk	-1.8994	Low Risk	Low Risk	G2	FEMALE
TCGA-A3-3335	1885	0	-0.3869	Intermediate Risk	High Risk	-0.4197	Intermediate Risk	High Risk	G4	MALE
TCGA-A3-3343	944	0	0.0692	High Risk	High Risk	0.0751	Intermediate Risk	High Risk	G3	MALE
TCGA-A3-3346	137	1	0.6787	High Risk	High Risk	0.7364	High Risk	High Risk	G3	MALE
TCGA-A3-3347	1609	0	0.5500	High Risk	High Risk	0.5968	High Risk	High Risk	G2	FEMALE
TCGA-A3-3349	1384	0	-1.5425	Intermediate Risk	Low Risk	-1.6736	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-A3-3351	909	0	0.2123	High Risk	High Risk	0.2304	Intermediate Risk	High Risk	G2	MALE
TCGA-A3-3352	561	0	-0.0955	Intermediate Risk	High Risk	-0.1036	Intermediate Risk	High Risk	G3	MALE
TCGA-A3-3357	2686	0	-0.6345	Intermediate Risk	High Risk	-0.6884	Intermediate Risk	High Risk	G3	MALE
TCGA-A3-3358	1306	0	-1.5430	Intermediate Risk	Low Risk	-1.6741	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-A3-3359	2502	0	-1.9511	Low Risk	Low Risk	-2.1169	Low Risk	Low Risk	G2	FEMALE
TCGA-A3-3362	1558	0	-1.8580	Low Risk	Low Risk	-2.0159	Low Risk	Low Risk	G2	FEMALE
TCGA-A3-3363	319	0	-0.2076	Intermediate Risk	High Risk	-0.2252	Intermediate Risk	High Risk	G2	MALE
TCGA-A3-3365	872	0	-0.8871	Intermediate Risk	Low Risk	-0.9625	Intermediate Risk	Low Risk	G2	MALE
TCGA-A3-3367	2268	0	-2.0275	Low Risk	Low Risk	-2.1998	Low Risk	Low Risk	G3	MALE
TCGA-A3-3370	2272	0	-2.2710	Low Risk	Low Risk	-2.4640	Low Risk	Low Risk	G2	FEMALE
TCGA-A3-3372	735	0	-0.2470	Intermediate Risk	High Risk	-0.2680	Intermediate Risk	High Risk	G2	MALE
TCGA-A3-3373	1620	0	-1.6052	Low Risk	Low Risk	-1.7416	Low Risk	Low Risk	G3	FEMALE
TCGA-A3-3374	1313	0	-1.2375	Intermediate Risk	Low Risk	-1.3427	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-A3-3376	1695	0	-1.6540	Low Risk	Low Risk	-1.7945	Low Risk	Low Risk	G2	MALE
TCGA-A3-3378	630	0	-1.2560	Intermediate Risk	Low Risk	-1.3628	Intermediate Risk	Low Risk	G3	MALE
TCGA-A3-3380	567	0	-1.7500	Low Risk	Low Risk	-1.8988	Low Risk	Low Risk	G2	MALE
TCGA-A3-3382	574	0	-0.8323	Intermediate Risk	Low Risk	-0.9031	Intermediate Risk	Low Risk	G3	MALE
TCGA-A3-3383	860	0	-1.9153	Low Risk	Low Risk	-2.0781	Low Risk	Low Risk	G2	MALE
TCGA-A3-3385	1991	0	-2.0145	Low Risk	Low Risk	-2.1857	Low Risk	Low Risk	G2	FEMALE
TCGA-A3-3387	617	0	-1.6678	Low Risk	Low Risk	-1.8096	Low Risk	Low Risk	G2	MALE
TCGA-A3-A6NI	1017	0	-1.9423	Low Risk	Low Risk	-2.1073	Low Risk	Low Risk	G3	MALE
TCGA-A3-A6NJ	468	0	-1.8329	Low Risk	Low Risk	-1.9887	Low Risk	Low Risk	G1	FEMALE
TCGA-A3-A6NL	688	0	-1.9185	Low Risk	Low Risk	-2.0815	Low Risk	Low Risk	G2	FEMALE
TCGA-A3-A6NN	3	0	-0.9085	Intermediate Risk	Low Risk	-0.9857	Intermediate Risk	Low Risk	G2	MALE
TCGA-A3-A8CQ	3	0	-2.0304	Low Risk	Low Risk	-2.2029	Low Risk	Low Risk	G2	FEMALE
TCGA-A3-A8OU	0	0	-1.4157	Intermediate Risk	Low Risk	-1.5360	Intermediate Risk	Low Risk	G1	FEMALE
TCGA-A3-A8OV	340	0	-2.1105	Low Risk	Low Risk	-2.2899	Low Risk	Low Risk	G2	MALE
TCGA-A3-A8OW	323	0	-0.9255	Intermediate Risk	Low Risk	-1.0041	Intermediate Risk	Low Risk	G2	MALE
TCGA-A3-A8OX	0	0	-1.3873	Intermediate Risk	Low Risk	-1.5052	Intermediate Risk	Low Risk	G3	FEMALE
TCGA-AK-3425	3340	0	-0.9485	Intermediate Risk	Low Risk	-1.0291	Intermediate Risk	Low Risk	G2	MALE
TCGA-AK-3426	884	1	1.1683	High Risk	High Risk	1.2675	High Risk	High Risk	G3	MALE
TCGA-AK-3427	3580	0	-0.9834	Intermediate Risk	Low Risk	-1.0669	Intermediate Risk	Low Risk	GX	MALE
TCGA-AK-3428	3725	0	-0.2411	Intermediate Risk	High Risk	-0.2615	Intermediate Risk	High Risk	G2	MALE
TCGA-AK-3429	3325	0	-0.2396	Intermediate Risk	High Risk	-0.2600	Intermediate Risk	High Risk	G2	FEMALE
TCGA-AK-3431	2239	0	0.8230	High Risk	High Risk	0.8930	High Risk	High Risk	G3	FEMALE
TCGA-AK-3433	3406	0	-0.4970	Intermediate Risk	High Risk	-0.5393	Intermediate Risk	High Risk	GX	FEMALE
TCGA-AK-3434	2085	0	-1.9926	Low Risk	Low Risk	-2.1619	Low Risk	Low Risk	G2	MALE
TCGA-AK-3436	3328	0	-0.6786	Intermediate Risk	High Risk	0.8845	High Risk	High Risk	G2	MALE
TCGA-AK-3440	2863	0	-1.5611	Intermediate Risk	Low Risk	-1.6937	Intermediate Risk	Low Risk	G3	MALE
TCGA-AK-3443	1422	0	-0.5340	Intermediate Risk	High Risk	-0.5794	Intermediate Risk	High Risk	GX	MALE
TCGA-AK-3444	1470	0	-2.3560	Low Risk	Low Risk	-2.5563	Low Risk	Low Risk	G2	FEMALE
TCGA-AK-3445	2390	0	-0.1280	Intermediate Risk	High Risk	-0.1389	Intermediate Risk	High Risk	G3	MALE
TCGA-AK-3447	1216	0	0.0281	Intermediate Risk	High Risk	0.0305	Intermediate Risk	High Risk	G2	MALE
TCGA-AK-3450	1778	0	-1.5721	Low Risk	Low Risk	-1.7057	Low Risk	Low Risk	G2	FEMALE
TCGA-AK-3451	2866	0	-0.2055	Intermediate Risk	High Risk	-0.2230	Intermediate Risk	High Risk	G3	MALE
TCGA-AK-3453	2529	0	0.3691	High Risk	High Risk	0.4004	Intermediate Risk	High Risk	G2	FEMALE
TCGA-AK-3454	873	0	-2.2949	Low Risk	Low Risk	-2.4899	Low Risk	Low Risk	G3	MALE
TCGA-AK-3455	683	0	-0.0201	Intermediate Risk	High Risk	-0.0218	Intermediate Risk	High Risk	G3	FEMALE
TCGA-AK-3456	1142	0	-0.8695	Intermediate Risk	Low Risk	-0.9434	Intermediate Risk	Low Risk	G3	MALE
TCGA-AK-3458	1167	0	-1.2427	Intermediate Risk	Low Risk	-1.3484	Intermediate Risk	Low Risk	G3	MALE
TCGA-AK-3460	2506	0	-1.7010	Low Risk	Low Risk	-1.8456	Low Risk	Low Risk	G2	MALE
TCGA-AK-3461	2215	0	-1.2685	Intermediate Risk	Low Risk	-1.3763	Intermediate Risk	Low Risk	G2	MALE
TCGA-AK-3465	369	0	-0.6426	Intermediate Risk	High Risk	-0.6972	Intermediate Risk	High Risk	GX	FEMALE
TCGA-AS-3777	1237	0	-0.2147	Intermediate Risk	High Risk	-0.2329	Intermediate Risk	High Risk	[Not Available]	MALE
TCGA-AS-3778	43	0	-2.3841	Low Risk	Low Risk	-2.5867	Low Risk	Low Risk	G1	MALE
TCGA-B0-4688	101	1	1.3040	High Risk	High Risk	3.0356	High Risk	High Risk	G4	MALE
TCGA-B0-4690	43	1	0.9824	High Risk	High Risk	2.6868	High Risk	High Risk	G3	MALE
TCGA-B0-4691	139	1	-0.0660	Intermediate Risk	High Risk	1.5492	High Risk	High Risk	G3	MALE
TCGA-B0-4693	77	0	-0.8734	Intermediate Risk	Low Risk	-0.9476	Intermediate Risk	Low Risk	G4	FEMALE
TCGA-B0-4694	106	0	0.1581	High Risk	High Risk	0.1715	Intermediate Risk	High Risk	G4	MALE
TCGA-B0-4696	865	1	3.1386	High Risk	High Risk	3.4053	High Risk	High Risk	G3	MALE
TCGA-B0-4697	578	1	0.8889	High Risk	High Risk	2.5853	High Risk	High Risk	G4	FEMALE
TCGA-B0-4698	42	1	1.8446	High Risk	High Risk	3.6222	High Risk	High Risk	G4	MALE
TCGA-B0-4699	110	0	0.8049	High Risk	High Risk	2.4942	High Risk	High Risk	G4	MALE
TCGA-B0-4700	1979	1	0.7547	High Risk	High Risk	2.4397	High Risk	High Risk	G4	MALE
TCGA-B0-4701	238	1	0.0261	Intermediate Risk	High Risk	1.6491	High Risk	High Risk	G3	FEMALE
TCGA-B0-4703	182	1	0.2521	High Risk	High Risk	1.8944	High Risk	High Risk	G4	MALE
TCGA-B0-4706	65	0	0.3563	High Risk	High Risk	0.3866	Intermediate Risk	High Risk	G4	MALE
TCGA-B0-4707	600	1	0.8212	High Risk	High Risk	0.8909	High Risk	High Risk	G4	MALE
TCGA-B0-4710	1754	0	-0.3385	Intermediate Risk	High Risk	-0.3673	Intermediate Risk	High Risk	G3	FEMALE
TCGA-B0-4712	1336	1	1.3250	High Risk	High Risk	3.0585	High Risk	High Risk	G3	MALE
TCGA-B0-4713	202	1	0.5719	High Risk	High Risk	0.6206	High Risk	High Risk	G2	FEMALE
TCGA-B0-4714	99	1	0.1173	High Risk	High Risk	1.7481	High Risk	High Risk	G3	MALE
TCGA-B0-4718	1777	0	0.5897	High Risk	High Risk	0.6399	High Risk	High Risk	G2	MALE
TCGA-B0-4810	478	1	0.4743	High Risk	High Risk	0.5146	High Risk	High Risk	G3	MALE
TCGA-B0-4811	1416	1	0.3546	High Risk	High Risk	0.3848	Intermediate Risk	High Risk	G3	MALE
TCGA-B0-4813	18	0	2.0188	High Risk	High Risk	2.1904	High Risk	High Risk	G3	MALE
TCGA-B0-4814	168	1	0.0101	Intermediate Risk	High Risk	1.6318	High Risk	High Risk	G3	MALE
TCGA-B0-4815	1587	1	1.6551	High Risk	High Risk	1.7957	High Risk	High Risk	G4	MALE
TCGA-B0-4816	1370	1	0.3745	High Risk	High Risk	0.4063	Intermediate Risk	High Risk	G3	MALE
TCGA-B0-4817	1018	0	1.0691	High Risk	High Risk	1.1599	High Risk	High Risk	G3	MALE
TCGA-B0-4818	509	1	-0.2116	Intermediate Risk	High Risk	-0.2296	Intermediate Risk	High Risk	G3	FEMALE
TCGA-B0-4819	183	0	0.7898	High Risk	High Risk	2.4778	High Risk	High Risk	G4	FEMALE
TCGA-B0-4821	1229	1	0.9584	High Risk	High Risk	1.0398	High Risk	High Risk	G3	FEMALE
TCGA-B0-4822	1110	1	0.4385	High Risk	High Risk	0.4757	High Risk	High Risk	G4	MALE
TCGA-B0-4823	454	0	-1.7767	Low Risk	Low Risk	-1.9277	Low Risk	Low Risk	G2	MALE
TCGA-B0-4824	1656	0	-1.6055	Low Risk	Low Risk	-1.7419	Low Risk	Low Risk	G3	FEMALE
TCGA-B0-4827	884	1	0.9361	High Risk	High Risk	1.0157	High Risk	High Risk	G4	FEMALE
TCGA-B0-4828	307	1	-0.4322	Intermediate Risk	High Risk	1.1519	High Risk	High Risk	G3	MALE
TCGA-B0-4833	2384	1	-1.1976	Intermediate Risk	Low Risk	-1.2993	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-B0-4834	2088	0	-1.0909	Intermediate Risk	Low Risk	-1.1836	Intermediate Risk	Low Risk	G3	MALE
TCGA-B0-4836	1237	1	0.9835	High Risk	High Risk	2.6879	High Risk	High Risk	G3	MALE
TCGA-B0-4837	1377	1	0.6228	High Risk	High Risk	0.6757	High Risk	High Risk	G3	MALE

TCGA-B0-4838	833	0	-1.7009	Low Risk	Low Risk	-1.8455	Low Risk	Low Risk	G3	FEMALE
TCGA-B0-4839	1638	0	-1.1809	Intermediate Risk	Low Risk	-1.2813	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-B0-4841	204	1	0.5836	High Risk	High Risk	2.2540	High Risk	High Risk	G3	MALE
TCGA-B0-4842	1723	1	0.7983	High Risk	High Risk	0.8662	High Risk	High Risk	G4	FEMALE
TCGA-B0-4843	320	1	1.3081	High Risk	High Risk	1.4193	High Risk	High Risk	G3	MALE
TCGA-B0-4844	313	1	0.6625	High Risk	High Risk	2.3397	High Risk	High Risk	G3	MALE
TCGA-B0-4845	1984	1	0.0528	Intermediate Risk	High Risk	1.6781	High Risk	High Risk	G2	MALE
TCGA-B0-4846	1199	1	0.0656	High Risk	High Risk	1.6920	High Risk	High Risk	G2	MALE
TCGA-B0-4847	792	0	-0.0203	Intermediate Risk	High Risk	1.5988	High Risk	High Risk	G3	MALE
TCGA-B0-4848	882	1	0.8808	High Risk	High Risk	0.9556	High Risk	High Risk	G3	MALE
TCGA-B0-4849	69	1	-0.0018	Intermediate Risk	High Risk	-0.0019	Intermediate Risk	High Risk	G3	MALE
TCGA-B0-4852	1120	1	-0.1085	Intermediate Risk	High Risk	-0.1177	Intermediate Risk	High Risk	G2	FEMALE
TCGA-B0-4945	2143	0	-1.3919	Intermediate Risk	Low Risk	-1.5102	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-B0-5075	637	0	-0.2105	Intermediate Risk	High Risk	-0.2284	Intermediate Risk	High Risk	G2	FEMALE
TCGA-B0-5077	1316	0	-1.2352	Intermediate Risk	Low Risk	-1.3401	Intermediate Risk	Low Risk	G3	MALE
TCGA-B0-5080	342	1	-0.1186	Intermediate Risk	High Risk	1.4922	High Risk	High Risk	G3	MALE
TCGA-B0-5081	362	1	0.5210	High Risk	High Risk	0.5652	High Risk	High Risk	G2	FEMALE
TCGA-B0-5083	1044	1	-0.8838	Intermediate Risk	Low Risk	-0.9589	Intermediate Risk	Low Risk	G3	MALE
TCGA-B0-5084	222	1	1.5728	High Risk	High Risk	3.3273	High Risk	High Risk	G3	MALE
TCGA-B0-5085	770	0	-0.4087	Intermediate Risk	High Risk	-0.4434	Intermediate Risk	High Risk	G3	FEMALE
TCGA-B0-5088	563	0	-1.7844	Low Risk	Low Risk	-1.9361	Low Risk	Low Risk	G3	MALE
TCGA-B0-5092	459	1	-0.1500	Intermediate Risk	High Risk	1.4581	High Risk	High Risk	G3	FEMALE
TCGA-B0-5094	333	1	0.6731	High Risk	High Risk	2.3511	High Risk	High Risk	G2	MALE
TCGA-B0-5095	245	0	0.2816	High Risk	High Risk	0.3055	Intermediate Risk	High Risk	G3	MALE
TCGA-B0-5096	68	0	1.0803	High Risk	High Risk	1.1721	High Risk	High Risk	GX	FEMALE
TCGA-B0-5097	665	0	0.9733	High Risk	High Risk	1.0560	High Risk	High Risk	G2	FEMALE
TCGA-B0-5098	1583	0	0.3707	High Risk	High Risk	0.4022	Intermediate Risk	High Risk	G3	FEMALE
TCGA-B0-5099	485	0	-0.6952	Intermediate Risk	High Risk	-0.7543	Intermediate Risk	High Risk	G3	FEMALE
TCGA-B0-5100	1911	1	0.8562	High Risk	High Risk	0.9289	High Risk	High Risk	G3	MALE
TCGA-B0-5102	2762	0	-1.3689	Intermediate Risk	Low Risk	-1.4852	Intermediate Risk	Low Risk	G3	FEMALE
TCGA-B0-5104	2750	0	-1.5655	Intermediate Risk	Low Risk	-1.6985	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-B0-5106	1597	0	-1.0251	Intermediate Risk	Low Risk	-1.1123	Intermediate Risk	Low Risk	G2	MALE
TCGA-B0-5107	926	1	0.0043	Intermediate Risk	High Risk	1.6255	High Risk	High Risk	G4	FEMALE
TCGA-B0-5108	1781	0	0.3549	High Risk	High Risk	0.3851	Intermediate Risk	High Risk	G2	MALE
TCGA-B0-5109	586	1	2.4685	High Risk	High Risk	2.6783	High Risk	High Risk	G4	MALE
TCGA-B0-5110	2008	0	-1.4766	Intermediate Risk	Low Risk	-1.6020	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-B0-5113	1174	0	-0.0427	Intermediate Risk	High Risk	-0.0463	Intermediate Risk	High Risk	G2	FEMALE
TCGA-B0-5115	1603	0	-0.0929	Intermediate Risk	High Risk	1.5200	High Risk	High Risk	G3	MALE
TCGA-B0-5116	1273	0	-0.4293	Intermediate Risk	High Risk	-0.4658	Intermediate Risk	High Risk	G3	MALE
TCGA-B0-5117	1607	0	-1.2101	Intermediate Risk	Low Risk	-1.3129	Intermediate Risk	Low Risk	G2	MALE
TCGA-B0-5119	1551	0	-1.7734	Low Risk	Low Risk	-1.9242	Low Risk	Low Risk	G2	FEMALE
TCGA-B0-5120	1168	0	-1.2359	Intermediate Risk	Low Risk	-1.3409	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-B0-5121	1484	0	-1.4845	Intermediate Risk	Low Risk	-1.6106	Intermediate Risk	Low Risk	G2	MALE
TCGA-B0-5399	1410	0	-1.5174	Intermediate Risk	Low Risk	-1.6464	Intermediate Risk	Low Risk	G2	MALE
TCGA-B0-5400	1732	0	0.8731	High Risk	High Risk	0.9473	High Risk	High Risk	G4	FEMALE
TCGA-B0-5402	1289	0	-0.8212	Intermediate Risk	Low Risk	0.7298	High Risk	High Risk	G4	MALE
TCGA-B0-5690	3389	0	-1.4822	Intermediate Risk	Low Risk	-1.6082	Intermediate Risk	Low Risk	G1	FEMALE
TCGA-B0-5691	3428	0	-1.6741	Low Risk	Low Risk	-1.8164	Low Risk	Low Risk	G3	FEMALE
TCGA-B0-5692	3941	0	-0.5727	Intermediate Risk	High Risk	-0.6214	Intermediate Risk	High Risk	G3	FEMALE
TCGA-B0-5693	4071	0	-2.1441	Low Risk	Low Risk	-2.3263	Low Risk	Low Risk	G2	FEMALE
TCGA-B0-5694	480	1	0.6125	High Risk	High Risk	0.6646	High Risk	High Risk	G3	MALE
TCGA-B0-5695	2148	0	-1.3415	Intermediate Risk	Low Risk	-1.4555	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-B0-5696	2607	0	-0.2390	Intermediate Risk	High Risk	-0.2593	Intermediate Risk	High Risk	G4	MALE
TCGA-B0-5697	2628	0	-0.9564	Intermediate Risk	Low Risk	-1.0376	Intermediate Risk	Low Risk	G2	MALE
TCGA-B0-5698	3628	0	-1.8350	Low Risk	Low Risk	-1.9909	Low Risk	Low Risk	G3	MALE
TCGA-B0-5699	3838	0	-2.3218	Low Risk	Low Risk	-2.5191	Low Risk	Low Risk	G2	MALE
TCGA-B0-5700	1789	0	-1.9618	Low Risk	Low Risk	-2.1285	Low Risk	Low Risk	G2	MALE
TCGA-B0-5701	2459	0	0.4756	High Risk	High Risk	0.5161	High Risk	High Risk	G4	MALE
TCGA-B0-5702	2170	0	-0.2506	Intermediate Risk	High Risk	-0.2719	Intermediate Risk	High Risk	G2	MALE
TCGA-B0-5703	2244	0	-1.8546	Low Risk	Low Risk	-2.0122	Low Risk	Low Risk	G3	MALE
TCGA-B0-5705	4534	0	-1.7150	Low Risk	Low Risk	-1.8608	Low Risk	Low Risk	G2	FEMALE
TCGA-B0-5706	3203	0	0.5053	High Risk	High Risk	0.5482	High Risk	High Risk	G2	MALE
TCGA-B0-5707	3741	0	-0.8451	Intermediate Risk	Low Risk	-0.9169	Intermediate Risk	Low Risk	G3	FEMALE
TCGA-B0-5709	3971	0	-0.2541	Intermediate Risk	High Risk	-0.2757	Intermediate Risk	High Risk	G3	FEMALE
TCGA-B0-5710	2428	0	-1.6105	Low Risk	Low Risk	-1.7473	Low Risk	Low Risk	G2	MALE
TCGA-B0-5711	3986	0	-0.6728	Intermediate Risk	High Risk	-0.7300	Intermediate Risk	High Risk	G3	MALE
TCGA-B0-5712	2720	0	-0.6938	Intermediate Risk	High Risk	0.8680	High Risk	High Risk	G3	FEMALE
TCGA-B0-5713	2780	0	-0.6309	Intermediate Risk	High Risk	-0.6846	Intermediate Risk	High Risk	G3	FEMALE
TCGA-B0-5812	3831	0	-1.4832	Intermediate Risk	Low Risk	-1.6093	Intermediate Risk	Low Risk	G3	MALE
TCGA-B2-3923	991	0	0.1928	High Risk	High Risk	0.2092	Intermediate Risk	High Risk	G2	MALE
TCGA-B2-3924	1091	0	-1.8982	Low Risk	Low Risk	-2.0595	Low Risk	Low Risk	G2	MALE
TCGA-B2-4098	51	0	-1.4175	Intermediate Risk	Low Risk	-1.5380	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-B2-4099	971	0	-1.3740	Intermediate Risk	Low Risk	-1.4907	Intermediate Risk	Low Risk	G3	MALE
TCGA-B2-4101	648	0	-0.1049	Intermediate Risk	High Risk	-0.1138	Intermediate Risk	High Risk	G3	MALE
TCGA-B2-4102	951	0	-1.9652	Low Risk	Low Risk	-2.1322	Low Risk	Low Risk	G2	MALE
TCGA-B2-5633	962	0	-1.2916	Intermediate Risk	Low Risk	-1.4014	Intermediate Risk	Low Risk	G2	MALE
TCGA-B2-5635	754	0	-1.4462	Intermediate Risk	Low Risk	-1.5691	Intermediate Risk	Low Risk	G2	MALE
TCGA-B2-5636	918	0	-0.4735	Intermediate Risk	High Risk	-0.5138	Intermediate Risk	High Risk	G2	MALE
TCGA-B2-5639	1002	1	-0.3640	Intermediate Risk	High Risk	1.2259	High Risk	High Risk	G3	MALE
TCGA-B2-5641	655	0	-1.5365	Intermediate Risk	Low Risk	-1.6671	Intermediate Risk	Low Risk	G3	MALE
TCGA-B2-A4SR	507	0	-0.5760	Intermediate Risk	High Risk	-0.6250	Intermediate Risk	High Risk	[Not Available]	MALE
TCGA-B4-5377	365	0	-0.8657	Intermediate Risk	Low Risk	0.6816	High Risk	High Risk	G3	FEMALE
TCGA-B4-5378	175	0	-1.4161	Intermediate Risk	Low Risk	-1.5365	Intermediate Risk	Low Risk	G2	MALE
TCGA-B4-5832	155	0	1.9223	High Risk	High Risk	2.0857	High Risk	High Risk	G2	MALE
TCGA-B4-5834	38	0	-1.8218	Low Risk	Low Risk	-1.9767	Low Risk	Low Risk	G1	MALE
TCGA-B4-5835	16	0	-1.1715	Intermediate Risk	Low Risk	-1.2710	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-B4-5836	141	0	-0.8484	Intermediate Risk	Low Risk	-0.9205	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-B4-5838	166	0	-0.4999	Intermediate Risk	High Risk	-0.5424	Intermediate Risk	High Risk	G2	MALE
TCGA-B4-5843	11	0	-1.9128	Low Risk	Low Risk	-2.0753	Low Risk	Low Risk	G2	MALE
TCGA-B4-5844	7	0	-0.3725	Intermediate Risk	High Risk	-0.4042	Intermediate Risk	High Risk	G1	FEMALE
TCGA-B8-4143	708	1	0.6450	High Risk	High Risk	2.3207	High Risk	High Risk	G3	FEMALE
TCGA-B8-4146	511	0	-2.0623	Low Risk	Low Risk	-2.2376	Low Risk	Low Risk	G2	FEMALE
TCGA-B8-4148	1519	0	-1.8958	Low Risk	Low Risk	-2.0569	Low Risk	Low Risk	G3	FEMALE
TCGA-B8-4151	1298	0	-0.3458	Intermediate Risk	High Risk	-0.3752	Intermediate Risk	High Risk	G2	FEMALE
TCGA-B8-4153	761	0	-0.1421	Intermediate Risk	High Risk	-0.1542	Intermediate Risk	High Risk	G3	MALE
TCGA-B8-4154	1379	0	-2.0097	Low Risk	Low Risk	-2.1805	Low Risk	Low Risk	G2	FEMALE
TCGA-B8-4619	523	0	-0.8204	Intermediate Risk	Low Risk	-0.8902	Intermediate Risk	Low Risk	G2	MALE
TCGA-B8-4620	777	0	-0.0014	Intermediate Risk	High Risk	-0.0016	Intermediate Risk	High Risk	G2	FEMALE
TCGA-B8-4621	787	0	-1.5386	Intermediate Risk	Low Risk	-1.6694	Intermediate Risk	Low Risk	G3	MALE
TCGA-B8-4622	1524	0	-0.2343	Intermediate Risk	High Risk	1.3666	High Risk	High Risk	G3	MALE
TCGA-B8-5158	1217	0	0.5513	High Risk	High Risk	0.5981	High Risk	High Risk	G4	MALE
TCGA-B8-5159	721	0	-2.2719	Low Risk	Low Risk	-2.4650	Low Risk	Low Risk	G3	FEMALE
TCGA-B8-5162	36	0	0.3234	High Risk	High Risk	0.3509	Intermediate Risk	High Risk	G2	MALE
TCGA-B8-5163	821	0	0.2027	High Risk	High Risk	0.2199	Intermediate Risk	High Risk	G3	FEMALE
TCGA-B8-5164	26	0	-0.9700	Intermediate Risk	Low Risk	-1.0524	Intermediate Risk	Low Risk	G3	MALE
TCGA-B8-5165	736	0	-2.1207	Low Risk	Low Risk	-2.3009	Low Risk	Low Risk	G2	MALE
TCGA-B8-5545	1524	0	-1.3593	Intermediate Risk	Low Risk	-1.4748	Intermediate Risk	Low Risk	G2	MALE
TCGA-B8-5546	505	0	-2.2357	Low Risk	Low Risk	-2.4257	Low Risk	Low Risk	G2	FEMALE
TCGA-B8-5549	194	0	-2.6278	Low Risk	Low Risk	-2.8512	Low Risk	Low Risk	G3	MALE
TCGA-B8-5550	1475	0	0.0402	Intermediate Risk	High Risk	0.0436	Intermediate Risk	High Risk	G3	MALE
TCGA-B8-5551	16	0	-0.2779	Intermediate Risk	High Risk	-0.3015	Intermediate Risk	High Risk	G3	FEMALE
TCGA-B8-5552	1045	0	-1.8082	Low Risk	Low Risk	-1.9619	Low Risk	Low Risk	G2	FEMALE
TCGA-B8-5553	435	0	-1.7956	Low Risk	Low Risk	-1.9482	Low Risk	Low Risk	G2	FEMALE
TCGA-B8-A54D	829	0	-0.1508	Intermediate Risk	High Risk	-0.1637	Intermediate Risk	High Risk	G2	MALE
TCGA-B8-A54E	908	0	-0.7081	Intermediate Risk	High Risk	-0.7683	Intermediate Risk	Low Risk	G3	FEMALE
TCGA-B8-A54F	519	0	-1.7160	Low Risk	Low Risk	-1.8618	Low Risk	Low Risk	G2	FEMALE
TCGA-B8-A54G	53	0	-1.2192	Intermediate Risk	Low Risk	-1.3228	Intermediate Risk	Low Risk	G3	MALE
TCGA-B8-A54H	256	0	-0.5815	Intermediate Risk	High Risk	-0.6309	Intermediate Risk	High Risk	G3	FEMALE
TCGA-B8-A54I	150	0	-0.9119	Intermediate Risk	Low Risk	-0.9894	Intermediate Risk	Low Risk	G3	MALE

TCGA-B8-A54J	528	0	-0.4135	Intermediate Risk	High Risk	-0.4487	Intermediate Risk	High Risk	G2	MALE
TCGA-B8-A54K	469	0	-0.6169	Intermediate Risk	High Risk	-0.6693	Intermediate Risk	High Risk	G1	MALE
TCGA-B8-A7U6	495	0	-1.2780	Intermediate Risk	Low Risk	-1.3866	Intermediate Risk	Low Risk	G3	FEMALE
TCGA-B8-A8YJ	431	0	-1.8596	Low Risk	Low Risk	-2.0176	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4158	3374	0	-1.8989	Low Risk	Low Risk	-2.0603	Low Risk	Low Risk	G2	MALE
TCGA-BP-4159	2599	1	-1.1660	Intermediate Risk	Low Risk	-1.2651	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-4160	2879	0	-0.0213	Intermediate Risk	High Risk	-0.0231	Intermediate Risk	High Risk	G2	MALE
TCGA-BP-4161	2744	0	-1.2300	Intermediate Risk	Low Risk	-1.3345	Intermediate Risk	Low Risk	G3	MALE
TCGA-BP-4162	3072	0	-1.7387	Low Risk	Low Risk	-1.8864	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4163	2837	0	-0.2725	Intermediate Risk	High Risk	-0.2957	Intermediate Risk	High Risk	G3	FEMALE
TCGA-BP-4164	991	0	-0.6055	Intermediate Risk	High Risk	-0.6570	Intermediate Risk	High Risk	G2	FEMALE
TCGA-BP-4165	3035	0	-1.4046	Intermediate Risk	Low Risk	-1.5240	Intermediate Risk	Low Risk	G1	FEMALE
TCGA-BP-4166	13	0	-0.4254	Intermediate Risk	High Risk	-0.4616	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4167	2716	0	0.2721	High Risk	High Risk	0.2952	Intermediate Risk	High Risk	G2	MALE
TCGA-BP-4169	700	1	1.4563	High Risk	High Risk	1.5801	High Risk	High Risk	G2	FEMALE
TCGA-BP-4170	2341	0	-1.8624	Low Risk	Low Risk	-2.0207	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4173	1892	0	-0.2207	Intermediate Risk	High Risk	-0.2395	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4174	1878	0	-0.5086	Intermediate Risk	High Risk	-0.5518	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4176	1953	0	-0.9938	Intermediate Risk	Low Risk	-1.0783	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-4177	1669	0	-0.4550	Intermediate Risk	High Risk	-0.4937	Intermediate Risk	High Risk	G2	MALE
TCGA-BP-4325	2962	0	-2.0762	Low Risk	Low Risk	-2.2526	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4326	1624	0	-0.0847	Intermediate Risk	High Risk	-0.0919	Intermediate Risk	High Risk	G2	FEMALE
TCGA-BP-4327	109	0	-0.2592	Intermediate Risk	High Risk	-0.2812	Intermediate Risk	High Risk	G2	FEMALE
TCGA-BP-4329	844	0	-0.0733	Intermediate Risk	High Risk	-0.0795	Intermediate Risk	High Risk	G2	MALE
TCGA-BP-4330	1886	0	-0.7512	Intermediate Risk	Low Risk	-0.8151	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-BP-4331	2452	1	-1.7862	Low Risk	Low Risk	-1.9380	Low Risk	Low Risk	G2	MALE
TCGA-BP-4332	1132	0	0.0652	High Risk	High Risk	0.0707	Intermediate Risk	High Risk	G2	MALE
TCGA-BP-4334	645	0	0.3733	High Risk	High Risk	0.4050	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4335	475	1	0.1849	High Risk	High Risk	1.8215	High Risk	High Risk	G3	FEMALE
TCGA-BP-4337	2	0	0.7769	High Risk	High Risk	0.8429	High Risk	High Risk	G4	FEMALE
TCGA-BP-4338	2857	0	-1.3785	Intermediate Risk	Low Risk	-1.4957	Intermediate Risk	Low Risk	G3	MALE
TCGA-BP-4340	561	0	-2.1099	Low Risk	Low Risk	-2.2892	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4341	1588	1	-0.2218	Intermediate Risk	High Risk	-0.2406	Intermediate Risk	High Risk	G2	MALE
TCGA-BP-4342	2254	1	1.0171	High Risk	High Risk	1.1036	High Risk	High Risk	G3	MALE
TCGA-BP-4343	1910	1	0.3103	High Risk	High Risk	0.3366	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4344	1665	0	-1.0239	Intermediate Risk	Low Risk	-1.1110	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-BP-4345	1515	0	0.3550	High Risk	High Risk	0.3852	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4346	1492	0	-0.2098	Intermediate Risk	High Risk	-0.2276	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4347	1366	0	-0.8480	Intermediate Risk	Low Risk	-0.9201	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-4349	372	0	-2.0950	Low Risk	Low Risk	-2.2730	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4351	969	0	0.4115	High Risk	High Risk	0.4465	Intermediate Risk	High Risk	G2	FEMALE
TCGA-BP-4352	344	1	1.5285	High Risk	High Risk	3.2792	High Risk	High Risk	G4	FEMALE
TCGA-BP-4353	375	0	-1.3895	Intermediate Risk	Low Risk	-1.5075	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-4354	1033	1	0.7575	High Risk	High Risk	2.4428	High Risk	High Risk	G4	MALE
TCGA-BP-4355	952	0	-0.1941	Intermediate Risk	High Risk	-0.2106	Intermediate Risk	High Risk	G4	FEMALE
TCGA-BP-4756	374	0	-2.0561	Low Risk	Low Risk	-2.2308	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4758	2206	0	-2.0388	Low Risk	Low Risk	-2.2121	Low Risk	Low Risk	G2	MALE
TCGA-BP-4759	2370	0	-1.6872	Low Risk	Low Risk	-1.8306	Low Risk	Low Risk	G2	MALE
TCGA-BP-4760	2359	0	-0.5428	Intermediate Risk	High Risk	-0.5889	Intermediate Risk	High Risk	G2	MALE
TCGA-BP-4761	182	0	-0.7031	Intermediate Risk	High Risk	-0.7629	Intermediate Risk	Low Risk	G4	MALE
TCGA-BP-4762	1342	0	-1.4495	Intermediate Risk	Low Risk	-1.5727	Intermediate Risk	Low Risk	G3	MALE
TCGA-BP-4763	1269	0	-0.9524	Intermediate Risk	Low Risk	-1.0333	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-BP-4765	2182	0	-1.9144	Low Risk	Low Risk	-2.0771	Low Risk	Low Risk	G2	MALE
TCGA-BP-4766	1461	0	-2.2355	Low Risk	Low Risk	-2.4254	Low Risk	Low Risk	G3	FEMALE
TCGA-BP-4768	400	0	-1.7873	Low Risk	Low Risk	-1.9392	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4769	1875	0	-0.5652	Intermediate Risk	High Risk	-0.6133	Intermediate Risk	High Risk	G2	MALE
TCGA-BP-4770	329	1	1.2150	High Risk	High Risk	2.9391	High Risk	High Risk	G4	FEMALE
TCGA-BP-4771	162	0	0.1965	High Risk	High Risk	1.8341	High Risk	High Risk	G4	MALE
TCGA-BP-4774	1884	0	-1.7002	Low Risk	Low Risk	-1.8447	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4775	1842	0	-2.7778	Low Risk	Low Risk	-3.0138	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4776	411	0	-1.5907	Low Risk	Low Risk	-1.7258	Low Risk	Low Risk	G2	MALE
TCGA-BP-4777	1730	0	-1.6858	Low Risk	Low Risk	-1.8291	Low Risk	Low Risk	G3	MALE
TCGA-BP-4781	2078	0	-1.6693	Low Risk	Low Risk	-1.8111	Low Risk	Low Risk	G3	MALE
TCGA-BP-4782	354	0	-1.6054	Low Risk	Low Risk	-1.7419	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4784	1853	0	-0.5296	Intermediate Risk	High Risk	-0.5746	Intermediate Risk	High Risk	G2	FEMALE
TCGA-BP-4787	480	1	0.0372	Intermediate Risk	High Risk	1.6612	High Risk	High Risk	G4	FEMALE
TCGA-BP-4789	1488	0	-2.2487	Low Risk	Low Risk	-2.4398	Low Risk	Low Risk	G2	MALE
TCGA-BP-4790	1110	0	-1.5457	Intermediate Risk	Low Risk	-1.6771	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-4795	620	0	-0.5158	Intermediate Risk	High Risk	-0.5596	Intermediate Risk	High Risk	G2	FEMALE
TCGA-BP-4797	1106	0	-0.3418	Intermediate Risk	High Risk	-0.3708	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4798	334	1	1.0755	High Risk	High Risk	1.1669	High Risk	High Risk	G4	MALE
TCGA-BP-4799	1132	1	0.9190	High Risk	High Risk	0.9971	High Risk	High Risk	G3	MALE
TCGA-BP-4801	1123	0	-1.5464	Intermediate Risk	Low Risk	-1.6778	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-4803	204	0	0.2845	High Risk	High Risk	0.3087	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4804	1458	0	-1.1179	Intermediate Risk	Low Risk	-1.2129	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-4807	211	0	-1.9918	Low Risk	Low Risk	-2.1611	Low Risk	Low Risk	G3	MALE
TCGA-BP-4959	2658	0	-1.5710	Low Risk	Low Risk	-1.7045	Low Risk	Low Risk	G3	MALE
TCGA-BP-4960	2170	0	-0.1502	Intermediate Risk	High Risk	-0.1630	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4961	1934	0	-2.2418	Low Risk	Low Risk	-2.4323	Low Risk	Low Risk	G2	MALE
TCGA-BP-4962	1784	0	-0.8955	Intermediate Risk	Low Risk	-0.9716	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-4963	1833	0	-2.1527	Low Risk	Low Risk	-2.3356	Low Risk	Low Risk	G3	MALE
TCGA-BP-4964	1861	0	-1.6128	Low Risk	Low Risk	-1.7499	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4965	1870	0	-1.6903	Low Risk	Low Risk	-1.8340	Low Risk	Low Risk	G2	MALE
TCGA-BP-4967	205	0	0.3263	High Risk	High Risk	0.3541	Intermediate Risk	High Risk	G2	MALE
TCGA-BP-4968	1745	0	-1.6017	Low Risk	Low Risk	-1.7378	Low Risk	Low Risk	G3	MALE
TCGA-BP-4969	1793	0	-1.4585	Intermediate Risk	Low Risk	-1.5825	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-BP-4970	433	0	0.1906	High Risk	High Risk	0.2068	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4971	1486	0	0.1744	High Risk	High Risk	0.1892	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4972	1501	0	-0.5430	Intermediate Risk	High Risk	-0.5891	Intermediate Risk	High Risk	G3	FEMALE
TCGA-BP-4973	1383	0	-0.3810	Intermediate Risk	High Risk	-0.4134	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4974	211	1	-0.1869	Intermediate Risk	High Risk	1.4180	High Risk	High Risk	G4	MALE
TCGA-BP-4975	1432	0	-2.1524	Low Risk	Low Risk	-2.3353	Low Risk	Low Risk	G3	MALE
TCGA-BP-4976	1631	0	-2.6366	Low Risk	Low Risk	-2.8607	Low Risk	Low Risk	G3	MALE
TCGA-BP-4977	454	0	-1.4991	Intermediate Risk	Low Risk	-1.6265	Intermediate Risk	Low Risk	G3	MALE
TCGA-BP-4981	1096	0	0.4061	High Risk	High Risk	0.4406	Intermediate Risk	High Risk	G3	FEMALE
TCGA-BP-4982	1013	0	-1.9759	Low Risk	Low Risk	-2.1439	Low Risk	Low Risk	G3	MALE
TCGA-BP-4983	1412	0	1.2238	High Risk	High Risk	1.3278	High Risk	High Risk	G4	FEMALE
TCGA-BP-4985	951	1	0.9043	High Risk	High Risk	0.9811	High Risk	High Risk	G4	MALE
TCGA-BP-4986	784	0	-1.4374	Intermediate Risk	Low Risk	-1.5596	Intermediate Risk	Low Risk	G3	MALE
TCGA-BP-4987	1123	0	-2.1280	Low Risk	Low Risk	-2.3088	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-4988	827	0	-1.0231	Intermediate Risk	Low Risk	-1.1100	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-4989	118	0	0.0955	High Risk	High Risk	0.1036	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-4991	1412	0	-2.3166	Low Risk	Low Risk	-2.5135	Low Risk	Low Risk	G2	MALE
TCGA-BP-4992	501	0	-0.8557	Intermediate Risk	Low Risk	-0.9285	Intermediate Risk	Low Risk	G4	MALE
TCGA-BP-4993	177	0	-1.5259	Intermediate Risk	Low Risk	-1.6556	Intermediate Risk	Low Risk	G3	MALE
TCGA-BP-4994	1307	0	-1.5121	Intermediate Risk	Low Risk	-1.6407	Intermediate Risk	Low Risk	G3	MALE
TCGA-BP-4995	1370	0	-1.5108	Intermediate Risk	Low Risk	-1.6392	Intermediate Risk	Low Risk	G3	MALE
TCGA-BP-4998	931	0	-1.3703	Intermediate Risk	Low Risk	-1.4867	Intermediate Risk	Low Risk	G3	MALE
TCGA-BP-4999	1265	0	-1.8144	Low Risk	Low Risk	-1.9686	Low Risk	Low Risk	G2	MALE
TCGA-BP-5000	563	0	-1.4935	Intermediate Risk	Low Risk	-1.6204	Intermediate Risk	Low Risk	G3	MALE
TCGA-BP-5001	1176	0	-1.6715	Low Risk	Low Risk	-1.8135	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-5004	1125	0	-2.7117	Low Risk	Low Risk	-2.9422	Low Risk	Low Risk	G3	MALE
TCGA-BP-5006	840	0	-1.9452	Low Risk	Low Risk	-2.1105	Low Risk	Low Risk	G2	MALE
TCGA-BP-5007	1139	0	-0.6544	Intermediate Risk	High Risk	-0.7100	Intermediate Risk	High Risk	G2	MALE
TCGA-BP-5008	1070	0	-1.5149	Intermediate Risk	Low Risk	-1.6437	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-5009	1091	1	-0.7715	Intermediate Risk	Low Risk	-0.8371	Intermediate Risk	Low Risk	G3	MALE
TCGA-BP-5010	877	0	0.7225	High Risk	High Risk	0.7839	High Risk	High Risk	G4	MALE
TCGA-BP-5168	1462	0	-2.0185	Low Risk	Low Risk	-2.1901	Low Risk	Low Risk	G2	MALE

TCGA-BP-5169	193	0	-1.0118	Intermediate Risk	Low Risk	-1.0978	Intermediate Risk	Low Risk	G4	MALE
TCGA-BP-5170	2410	0	-1.3256	Intermediate Risk	Low Risk	-1.4382	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-5173	62	0	-1.8166	Low Risk	Low Risk	-1.9710	Low Risk	Low Risk	G2	MALE
TCGA-BP-5174	2255	0	-1.5729	Low Risk	Low Risk	-1.7066	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-5175	931	0	-0.6631	Intermediate Risk	High Risk	-0.7195	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-5176	1589	0	-1.1979	Intermediate Risk	Low Risk	-1.2997	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-BP-5177	293	0	-1.6415	Low Risk	Low Risk	-1.7810	Low Risk	Low Risk	G3	FEMALE
TCGA-BP-5178	1910	1	-0.5342	Intermediate Risk	High Risk	1.0413	High Risk	High Risk	G4	MALE
TCGA-BP-5180	2261	0	-1.3825	Intermediate Risk	Low Risk	-1.5000	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-5181	1494	0	-1.8567	Low Risk	Low Risk	-2.0145	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-5182	1164	0	-1.8998	Low Risk	Low Risk	-2.0613	Low Risk	Low Risk	G3	MALE
TCGA-BP-5183	1290	0	-0.3644	Intermediate Risk	High Risk	-0.3953	Intermediate Risk	High Risk	G3	MALE
TCGA-BP-5184	1132	0	-1.9604	Low Risk	Low Risk	-2.1270	Low Risk	Low Risk	G3	MALE
TCGA-BP-5185	1131	0	-2.0698	Low Risk	Low Risk	-2.2457	Low Risk	Low Risk	G3	MALE
TCGA-BP-5186	693	0	-1.8407	Low Risk	Low Risk	-1.9971	Low Risk	Low Risk	G2	FEMALE
TCGA-BP-5187	406	0	-1.3701	Intermediate Risk	Low Risk	-1.4865	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-5189	821	1	-1.2617	Intermediate Risk	Low Risk	-1.3690	Intermediate Risk	Low Risk	G4	MALE
TCGA-BP-5190	1010	0	-2.5437	Low Risk	Low Risk	-2.7599	Low Risk	Low Risk	G3	MALE
TCGA-BP-5191	966	0	0.4054	High Risk	High Risk	0.4398	Intermediate Risk	High Risk	G2	MALE
TCGA-BP-5192	714	0	-2.4756	Low Risk	Low Risk	-2.6859	Low Risk	Low Risk	G2	MALE
TCGA-BP-5194	408	0	-1.9481	Low Risk	Low Risk	-2.1136	Low Risk	Low Risk	G2	MALE
TCGA-BP-5195	749	0	-1.4136	Intermediate Risk	Low Risk	-1.5338	Intermediate Risk	Low Risk	G2	MALE
TCGA-BP-5196	1017	0	-1.7957	Low Risk	Low Risk	-1.9483	Low Risk	Low Risk	G2	MALE
TCGA-BP-5198	603	0	0.4274	High Risk	High Risk	0.4637	High Risk	High Risk	G3	MALE
TCGA-BP-5199	1354	0	0.3112	High Risk	High Risk	0.3377	Intermediate Risk	High Risk	G4	MALE
TCGA-BP-5200	1062	0	-0.0123	Intermediate Risk	High Risk	-0.0133	Intermediate Risk	High Risk	G4	MALE
TCGA-BP-5201	950	0	0.1804	High Risk	High Risk	1.8166	High Risk	High Risk	G4	MALE
TCGA-BP-5202	29	0	-0.2198	Intermediate Risk	High Risk	-0.2385	Intermediate Risk	High Risk	G2	MALE
TCGA-CJ-4634	3495	0	-1.7592	Low Risk	Low Risk	-1.9087	Low Risk	Low Risk	G2	FEMALE
TCGA-CJ-4635	1415	0	-1.0602	Intermediate Risk	Low Risk	-1.1503	Intermediate Risk	Low Risk	G3	MALE
TCGA-CJ-4636	1923	0	-0.1124	Intermediate Risk	High Risk	-0.1220	Intermediate Risk	High Risk	G3	MALE
TCGA-CJ-4637	2225	1	0.6830	High Risk	High Risk	2.3619	High Risk	High Risk	G4	FEMALE
TCGA-CJ-4638	431	1	0.4311	High Risk	High Risk	2.0886	High Risk	High Risk	G4	FEMALE
TCGA-CJ-4639	3227	0	-0.6030	Intermediate Risk	High Risk	-0.6543	Intermediate Risk	High Risk	G2	FEMALE
TCGA-CJ-4640	3477	0	-0.2735	Intermediate Risk	High Risk	-0.2968	Intermediate Risk	High Risk	G4	MALE
TCGA-CJ-4641	1660	1	-0.1106	Intermediate Risk	High Risk	1.5008	High Risk	High Risk	G4	FEMALE
TCGA-CJ-4642	3203	0	0.4197	High Risk	High Risk	0.4553	High Risk	High Risk	G2	MALE
TCGA-CJ-4643	1792	0	0.3683	High Risk	High Risk	0.3996	Intermediate Risk	High Risk	G3	FEMALE
TCGA-CJ-4644	336	1	0.0798	High Risk	High Risk	1.7074	High Risk	High Risk	G3	FEMALE
TCGA-CJ-4868	645	1	0.8004	High Risk	High Risk	2.4893	High Risk	High Risk	G3	MALE
TCGA-CJ-4869	2552	0	-0.9575	Intermediate Risk	Low Risk	-1.0389	Intermediate Risk	Low Risk	G2	MALE
TCGA-CJ-4870	1497	0	-0.6147	Intermediate Risk	High Risk	-0.6669	Intermediate Risk	High Risk	G2	FEMALE
TCGA-CJ-4871	2421	0	0.6348	High Risk	High Risk	2.3096	High Risk	High Risk	G4	MALE
TCGA-CJ-4872	1434	0	-0.8853	Intermediate Risk	Low Risk	-0.9605	Intermediate Risk	Low Risk	G4	MALE
TCGA-CJ-4873	2257	0	0.0591	High Risk	High Risk	0.0642	Intermediate Risk	High Risk	G3	FEMALE
TCGA-CJ-4874	2281	0	-1.8579	Low Risk	Low Risk	-2.0157	Low Risk	Low Risk	G3	FEMALE
TCGA-CJ-4875	3551	1	-0.0049	Intermediate Risk	High Risk	1.6155	High Risk	High Risk	G3	MALE
TCGA-CJ-4876	1953	0	-0.6946	Intermediate Risk	High Risk	-0.7537	Intermediate Risk	High Risk	G3	MALE
TCGA-CJ-4878	2184	0	0.0928	High Risk	High Risk	0.1007	Intermediate Risk	High Risk	G2	FEMALE
TCGA-CJ-4881	2012	0	0.6844	High Risk	High Risk	0.7426	High Risk	High Risk	G3	MALE
TCGA-CJ-4882	1882	0	1.3792	High Risk	High Risk	1.4964	High Risk	High Risk	G3	MALE
TCGA-CJ-4884	1758	0	-0.1432	Intermediate Risk	High Risk	-0.1554	Intermediate Risk	High Risk	G3	FEMALE
TCGA-CJ-4885	3448	0	0.7291	High Risk	High Risk	2.4118	High Risk	High Risk	G3	MALE
TCGA-CJ-4886	1951	0	-1.5931	Low Risk	Low Risk	-1.7285	Low Risk	Low Risk	G3	FEMALE
TCGA-CJ-4887	931	1	-1.0172	Intermediate Risk	Low Risk	0.5172	High Risk	High Risk	G3	MALE
TCGA-CJ-4888	1566	1	0.7347	High Risk	High Risk	2.4180	High Risk	High Risk	G4	MALE
TCGA-CJ-4889	1945	0	-1.1045	Intermediate Risk	Low Risk	-1.1984	Intermediate Risk	Low Risk	G4	FEMALE
TCGA-CJ-4890	3516	0	0.6313	High Risk	High Risk	2.3058	High Risk	High Risk	G4	MALE
TCGA-CJ-4891	819	0	0.7890	High Risk	High Risk	0.8561	High Risk	High Risk	G4	FEMALE
TCGA-CJ-4892	1520	0	-1.9200	Low Risk	Low Risk	-2.0831	Low Risk	Low Risk	G2	FEMALE
TCGA-CJ-4893	749	0	-2.0463	Low Risk	Low Risk	-2.2202	Low Risk	Low Risk	G3	FEMALE
TCGA-CJ-4894	840	1	0.1029	High Risk	High Risk	0.1117	Intermediate Risk	High Risk	G3	MALE
TCGA-CJ-4895	1199	1	-0.0073	Intermediate Risk	High Risk	1.6129	High Risk	High Risk	G4	MALE
TCGA-CJ-4897	3339	0	-0.9566	Intermediate Risk	Low Risk	-1.0379	Intermediate Risk	Low Risk	G3	FEMALE
TCGA-CJ-4899	1527	0	-1.9723	Low Risk	Low Risk	-2.1400	Low Risk	Low Risk	G2	MALE
TCGA-CJ-4900	1713	1	0.9613	High Risk	High Risk	2.6638	High Risk	High Risk	G4	FEMALE
TCGA-CJ-4901	1449	0	-0.0917	Intermediate Risk	High Risk	-0.0995	Intermediate Risk	High Risk	G3	MALE
TCGA-CJ-4902	1519	0	-0.0132	Intermediate Risk	High Risk	-0.0143	Intermediate Risk	High Risk	G3	MALE
TCGA-CJ-4903	1559	0	-1.5654	Intermediate Risk	Low Risk	-1.6984	Intermediate Risk	Low Risk	G3	MALE
TCGA-CJ-4904	3300	0	-0.0709	Intermediate Risk	High Risk	1.5439	High Risk	High Risk	G3	FEMALE
TCGA-CJ-4905	1495	0	-1.2563	Intermediate Risk	Low Risk	-1.3630	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-CJ-4907	1498	0	-0.5526	Intermediate Risk	High Risk	-0.5996	Intermediate Risk	High Risk	G3	MALE
TCGA-CJ-4908	1530	0	-1.4288	Intermediate Risk	Low Risk	-1.5502	Intermediate Risk	Low Risk	G2	MALE
TCGA-CJ-4912	1656	0	0.4655	High Risk	High Risk	0.5051	High Risk	High Risk	G3	MALE
TCGA-CJ-4916	1372	0	0.0327	Intermediate Risk	High Risk	0.0355	Intermediate Risk	High Risk	G3	FEMALE
TCGA-CJ-4918	93	1	0.0331	Intermediate Risk	High Risk	1.6567	High Risk	High Risk	G4	MALE
TCGA-CJ-4920	139	0	-1.2932	Intermediate Risk	Low Risk	-1.4031	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-CJ-4923	572	1	1.3184	High Risk	High Risk	3.0512	High Risk	High Risk	G4	FEMALE
TCGA-CJ-5671	3984	0	-1.1644	Intermediate Risk	Low Risk	-1.2633	Intermediate Risk	Low Risk	G3	MALE
TCGA-CJ-5672	2188	0	-1.7282	Low Risk	Low Risk	-1.8751	Low Risk	Low Risk	G3	MALE
TCGA-CJ-5675	3933	0	-0.9015	Intermediate Risk	Low Risk	-0.9781	Intermediate Risk	Low Risk	G3	MALE
TCGA-CJ-5676	4064	0	-0.1420	Intermediate Risk	High Risk	-0.1541	Intermediate Risk	High Risk	G3	MALE
TCGA-CJ-5677	781	1	0.1121	High Risk	High Risk	1.7425	High Risk	High Risk	G4	FEMALE
TCGA-CJ-5678	574	1	-0.6390	Intermediate Risk	High Risk	0.9275	High Risk	High Risk	G3	MALE
TCGA-CJ-5679	679	1	0.8832	High Risk	High Risk	0.9582	High Risk	High Risk	G4	MALE
TCGA-CJ-5680	767	1	0.1630	High Risk	High Risk	1.7977	High Risk	High Risk	G4	FEMALE
TCGA-CJ-5681	551	1	-0.2523	Intermediate Risk	High Risk	1.3471	High Risk	High Risk	G3	FEMALE
TCGA-CJ-5682	3733	0	0.0603	High Risk	High Risk	1.6863	High Risk	High Risk	G4	MALE
TCGA-CJ-5683	1888	0	-0.8006	Intermediate Risk	Low Risk	-0.8687	Intermediate Risk	Low Risk	G3	MALE
TCGA-CJ-5684	2229	0	-0.5402	Intermediate Risk	High Risk	-0.5861	Intermediate Risk	High Risk	G2	MALE
TCGA-CJ-5686	2036	0	-1.9800	Low Risk	Low Risk	-2.1482	Low Risk	Low Risk	G3	FEMALE
TCGA-CJ-5689	1619	0	-1.3122	Intermediate Risk	Low Risk	-1.4237	Intermediate Risk	Low Risk	G4	MALE
TCGA-CJ-6027	3612	0	-0.9297	Intermediate Risk	Low Risk	-1.0087	Intermediate Risk	Low Risk	G4	MALE
TCGA-CJ-6028	1624	1	0.2164	High Risk	High Risk	1.8556	High Risk	High Risk	G4	MALE
TCGA-CJ-6030	2297	0	-0.7560	Intermediate Risk	Low Risk	-0.8202	Intermediate Risk	Low Risk	G3	MALE
TCGA-CJ-6031	1904	0	-1.1167	Intermediate Risk	Low Risk	-1.2116	Intermediate Risk	Low Risk	G3	MALE
TCGA-CJ-6032	3636	0	-0.7508	Intermediate Risk	High Risk	-0.8146	Intermediate Risk	Low Risk	G3	FEMALE
TCGA-CJ-6033	224	1	1.3472	High Risk	High Risk	3.0825	High Risk	High Risk	G4	FEMALE
TCGA-CW-5580	1962	0	-0.2751	Intermediate Risk	High Risk	1.3223	High Risk	High Risk	G3	FEMALE
TCGA-CW-5581	2797	0	-2.2405	Low Risk	Low Risk	-2.4310	Low Risk	Low Risk	G3	MALE
TCGA-CW-5583	2487	0	-2.2297	Low Risk	Low Risk	-2.4192	Low Risk	Low Risk	G2	FEMALE
TCGA-CW-5584	164	0	-0.2657	Intermediate Risk	High Risk	-0.2883	Intermediate Risk	High Risk	G3	MALE
TCGA-CW-5585	2607	0	-1.2824	Intermediate Risk	Low Risk	0.2294	Intermediate Risk	High Risk	G2	MALE
TCGA-CW-5587	2224	0	-0.9119	Intermediate Risk	Low Risk	-0.9893	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-CW-5588	2015	0	-1.3696	Intermediate Risk	Low Risk	-1.4860	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-CW-5589	2376	0	-1.6644	Low Risk	Low Risk	-1.8059	Low Risk	Low Risk	G2	MALE
TCGA-CW-5590	1074	0	-0.7284	Intermediate Risk	High Risk	0.8305	High Risk	High Risk	G3	MALE
TCGA-CW-5591	2269	0	-0.7161	Intermediate Risk	High Risk	0.8439	High Risk	High Risk	G2	MALE
TCGA-CW-6087	41	0	0.7379	High Risk	High Risk	2.4214	High Risk	High Risk	G4	MALE
TCGA-CW-6088	3220	0	-1.9818	Low Risk	Low Risk	-2.1503	Low Risk	Low Risk	G2	MALE
TCGA-CW-6090	2550	0	-1.1864	Intermediate Risk	Low Risk	-1.2872	Intermediate Risk	Low Risk	G3	MALE
TCGA-CW-6093	3144	0	-1.6907	Low Risk	Low Risk	-1.8344	Low Risk	Low Risk	G1	MALE
TCGA-CW-6097	571	1	0.2541	High Risk	High Risk	0.2757	Intermediate Risk	High Risk	G4	MALE
TCGA-CZ-4853	773	0	-2.1728	Low Risk	Low Risk	-2.3574	Low Risk	Low Risk	G2	MALE
TCGA-CZ-4854	1403	0	-0.8379	Intermediate Risk	Low Risk	-0.9091	Intermediate Risk	Low Risk	G2	MALE
TCGA-CZ-4856	18	0	-1.5576	Intermediate Risk	Low Risk	-1.6900	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-CZ-4857	1431	1	0.7263	High Risk	High Risk	2.4089	High Risk	High Risk	G3	MALE
TCGA-CZ-4858	2103	1	1.8190	High Risk	High Risk	1.9736	High Risk	High Risk	G4	MALE
TCGA-CZ-4859	1786	0	-2.4463	Low Risk	Low Risk	-2.6542	Low Risk	Low Risk	G2	FEMALE

TCGA-CZ-4860	206	1	1.4880	High Risk	High Risk	3.2353	High Risk	High Risk	G4	MALE
TCGA-CZ-4861	446	1	0.1347	High Risk	High Risk	0.1461	Intermediate Risk	High Risk	G2	MALE
TCGA-CZ-4862	3269	0	-1.6899	Low Risk	Low Risk	-1.8336	Low Risk	Low Risk	G2	MALE
TCGA-CZ-4863	1927	0	-0.5059	Intermediate Risk	High Risk	-0.5489	Intermediate Risk	High Risk	G3	FEMALE
TCGA-CZ-4864	2828	0	-1.1484	Intermediate Risk	Low Risk	-1.2460	Intermediate Risk	Low Risk	G3	MALE
TCGA-CZ-4865	166	0	-1.2758	Intermediate Risk	Low Risk	-1.3843	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-CZ-4866	3265	0	-1.3097	Intermediate Risk	Low Risk	-1.4210	Intermediate Risk	Low Risk	G3	FEMALE
TCGA-CZ-5451	1928	0	-0.5649	Intermediate Risk	High Risk	-0.6129	Intermediate Risk	High Risk	G3	MALE
TCGA-CZ-5452	1788	0	-0.4140	Intermediate Risk	High Risk	-0.4492	Intermediate Risk	High Risk	G2	MALE
TCGA-CZ-5453	2417	0	-0.7078	Intermediate Risk	High Risk	-0.7680	Intermediate Risk	Low Risk	G2	MALE
TCGA-CZ-5454	721	1	-0.1581	Intermediate Risk	High Risk	1.4493	High Risk	High Risk	G2	MALE
TCGA-CZ-5455	561	1	-0.2002	Intermediate Risk	High Risk	1.4036	High Risk	High Risk	G4	MALE
TCGA-CZ-5456	2420	0	-0.2187	Intermediate Risk	High Risk	-0.2372	Intermediate Risk	High Risk	G3	MALE
TCGA-CZ-5457	2752	0	-0.8906	Intermediate Risk	Low Risk	-0.9663	Intermediate Risk	Low Risk	G4	MALE
TCGA-CZ-5458	2787	0	-0.8219	Intermediate Risk	Low Risk	-0.8918	Intermediate Risk	Low Risk	G3	MALE
TCGA-CZ-5459	1682	0	0.0366	Intermediate Risk	High Risk	0.0397	Intermediate Risk	High Risk	G3	MALE
TCGA-CZ-5460	2871	0	-0.2667	Intermediate Risk	High Risk	1.3314	High Risk	High Risk	G2	MALE
TCGA-CZ-5461	330	1	0.2691	High Risk	High Risk	1.9128	High Risk	High Risk	G4	MALE
TCGA-CZ-5462	311	1	-0.7214	Intermediate Risk	High Risk	0.8382	High Risk	High Risk	G3	MALE
TCGA-CZ-5463	662	0	-0.2070	Intermediate Risk	High Risk	-0.2246	Intermediate Risk	High Risk	G2	MALE
TCGA-CZ-5464	2126	0	0.1349	High Risk	High Risk	1.7672	High Risk	High Risk	G2	MALE
TCGA-CZ-5465	2562	0	-0.2107	Intermediate Risk	High Risk	-0.2287	Intermediate Risk	High Risk	G2	FEMALE
TCGA-CZ-5466	684	0	-0.4363	Intermediate Risk	High Risk	-0.4734	Intermediate Risk	High Risk	G2	MALE
TCGA-CZ-5467	73	1	0.1462	High Risk	High Risk	0.1586	Intermediate Risk	High Risk	G4	FEMALE
TCGA-CZ-5468	59	1	0.4091	High Risk	High Risk	2.0647	High Risk	High Risk	G4	MALE
TCGA-CZ-5469	945	1	0.3671	High Risk	High Risk	0.3983	Intermediate Risk	High Risk	G2	MALE
TCGA-CZ-5470	386	0	0.1560	High Risk	High Risk	0.1692	Intermediate Risk	High Risk	G3	FEMALE
TCGA-CZ-5982	2437	0	-2.1271	Low Risk	Low Risk	-2.3079	Low Risk	Low Risk	G2	FEMALE
TCGA-CZ-5984	2065	0	-1.3594	Intermediate Risk	Low Risk	-1.4749	Intermediate Risk	Low Risk	G3	MALE
TCGA-CZ-5985	1995	0	-0.8297	Intermediate Risk	Low Risk	-0.9002	Intermediate Risk	Low Risk	G2	MALE
TCGA-CZ-5986	373	0	-2.4471	Low Risk	Low Risk	-2.6551	Low Risk	Low Risk	G3	MALE
TCGA-CZ-5987	445	1	0.5565	High Risk	High Risk	2.2246	High Risk	High Risk	G2	MALE
TCGA-CZ-5988	693	0	-1.3559	Intermediate Risk	Low Risk	-1.4711	Intermediate Risk	Low Risk	G2	MALE
TCGA-CZ-5989	1903	0	0.4452	High Risk	High Risk	0.4830	High Risk	High Risk	G2	MALE
TCGA-DV-5565	1328	0	-1.4028	Intermediate Risk	Low Risk	-1.5220	Intermediate Risk	Low Risk	G2	MALE
TCGA-DV-5566	1397	0	-1.3074	Intermediate Risk	Low Risk	-1.4186	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-DV-5567	2002	0	-0.8350	Intermediate Risk	Low Risk	-0.9060	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-DV-5568	370	0	-1.1709	Intermediate Risk	Low Risk	-1.2704	Intermediate Risk	Low Risk	G2	MALE
TCGA-DV-5569	355	0	-1.7009	Low Risk	Low Risk	-1.8455	Low Risk	Low Risk	G2	FEMALE
TCGA-DV-5573	1129	0	-1.7872	Low Risk	Low Risk	-1.9391	Low Risk	Low Risk	G2	MALE
TCGA-DV-5574	2014	0	-0.9975	Intermediate Risk	Low Risk	-1.0823	Intermediate Risk	Low Risk	G2	MALE
TCGA-DV-5575	1728	0	-1.9738	Low Risk	Low Risk	-2.1416	Low Risk	Low Risk	G2	FEMALE
TCGA-DV-5576	726	1	-0.9260	Intermediate Risk	Low Risk	-1.0047	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-DV-A4VX	1625	1	-0.4520	Intermediate Risk	High Risk	1.1304	High Risk	High Risk	G4	MALE
TCGA-DV-A4VZ	365	0	-0.2393	Intermediate Risk	High Risk	-0.2596	Intermediate Risk	High Risk	G2	MALE
TCGA-DV-A4W0	2468	0	-1.1585	Intermediate Risk	Low Risk	-1.2570	Intermediate Risk	Low Risk	G3	MALE
TCGA-EU-5904	551	0	-1.6399	Low Risk	Low Risk	-1.7792	Low Risk	Low Risk	G1	FEMALE
TCGA-EU-5905	119	0	-1.9689	Low Risk	Low Risk	-2.1362	Low Risk	Low Risk	G3	FEMALE
TCGA-EU-5906	206	0	-1.9905	Low Risk	Low Risk	-2.1597	Low Risk	Low Risk	G2	MALE
TCGA-EU-5907	127	0	-0.1297	Intermediate Risk	High Risk	-0.1407	Intermediate Risk	High Risk	G3	MALE
TCGA-G6-A5PC	242	1	0.0795	High Risk	High Risk	1.7071	High Risk	High Risk	G4	FEMALE
TCGA-G6-A8L6	313	1	-0.3322	Intermediate Risk	High Risk	1.2604	High Risk	High Risk	G3	MALE
TCGA-G6-A8L7	2131	0	-0.6502	Intermediate Risk	High Risk	-0.7055	Intermediate Risk	High Risk	G3	FEMALE
TCGA-G6-A8L8	1090	0	-1.1994	Intermediate Risk	Low Risk	-1.3014	Intermediate Risk	Low Risk	G3	FEMALE
TCGA-GK-A6C7	61	0	-2.1249	Low Risk	Low Risk	-2.3055	Low Risk	Low Risk	[Not Available]	FEMALE
TCGA-MM-A563	591	0	0.4284	High Risk	High Risk	0.4648	High Risk	High Risk	G2	MALE
TCGA-MM-A564	607	0	0.7728	High Risk	High Risk	0.8385	High Risk	High Risk	G2	MALE
TCGA-MM-A84U	700	0	-0.8557	Intermediate Risk	Low Risk	-0.9284	Intermediate Risk	Low Risk	G2	FEMALE
TCGA-MW-A4EC	498	0	-1.6150	Low Risk	Low Risk	-1.7523	Low Risk	Low Risk	G2	FEMALE
TCGA-T7-A92I	356	0	-1.7467	Low Risk	Low Risk	-1.8951	Low Risk	Low Risk	G1	FEMALE

Patient	Stage	Standardized Age	VHL mutation	SSIGN Score	Beleut signature (A vs. B vs. C)	Bostrom signature	ClearCode34 signature	Kosari signature	Lane signature	Zhao signature
TCGA-6D-AA2E	Stage I	0.608	not mutated	4	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-A3-3306	Stage I	0.526	not mutated	3	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3307	Stage III	0.444	not mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3308	Stage III	1.349	mutated	4	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-A3-3311	Stage I	-0.297	not mutated	2	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3313	Stage I	-0.133	not mutated	3	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-A3-3316	Stage II	-0.297	mutated	4	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-A3-3317	Stage II	0.526	not mutated	3	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3319	Stage I	0.773	not mutated	2	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-A3-3320	Stage I	-0.709	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3322	Stage I	-0.791	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3323	Stage I	-0.627	not mutated	2	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3324	Stage I	-0.791	not mutated	3	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3325	Stage I	-0.709	not mutated	0	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3326	Stage I	-1.120	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3328	Stage I	1.514	not mutated	4	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-A3-3329	Stage I	1.184	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3331	Stage I	2.090	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3335	Stage II	-1.614	not mutated	3	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-A3-3343	Stage II	1.514	not mutated	4	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3346	Stage I	0.608	not mutated	3	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-A3-3347	Stage III	1.267	not mutated	4	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-A3-3349	Stage I	-2.191	mutated	2	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-A3-3351	Stage II	-1.532	not mutated	3	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3352	Stage III	1.102	not mutated	5	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3357	Stage II	0.114	not mutated	4	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3358	Stage I	-0.297	not mutated	2	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-A3-3359	Stage I	1.761	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3362	Stage I	-0.050	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3363	Stage II	-0.873	not mutated	3	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-A3-3365	Stage I	-1.203	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3367	Stage I	0.938	not mutated	3	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3370	Stage I	-1.038	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3372	Stage III	0.279	mutated	4	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3373	Stage I	-0.544	mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3374	Stage I	-0.791	not mutated	2	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-A3-3376	Stage I	-0.791	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3378	Stage I	-0.050	mutated	3	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-A3-3380	Stage I	-0.544	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3382	Stage I	0.691	mutated	3	Cluster B	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-A3-3383	Stage I	-0.709	mutated	2	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-A3-3385	Stage I	-1.203	mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-3387	Stage I	-0.956	not mutated	2	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-A3-A6NI	Stage I	-1.120	not mutated	3	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-A3-A6NJ	Stage I	-0.297	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-A6NL	Stage I	-0.956	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-A3-A6NN	Stage I	1.431	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-A8CQ	Stage I	-0.133	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-A8OU	Stage I	1.102	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-A8OV	Stage I	1.184	not mutated	2	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-A8OW	Stage III	-1.944	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-A3-A8OX	Stage I	0.361	not mutated	3	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-AK-3425	Stage I	0.608	not mutated	2	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-AK-3426	Stage III	-1.944	not mutated	9	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-AK-3427	Stage I	0.361	not mutated	2	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AK-3428	Stage III	0.114	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-AK-3429	Stage II	-0.544	not mutated	5	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AK-3431	Stage II	0.114	not mutated	4	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-AK-3433	Stage II	-1.038	not mutated	3	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AK-3434	Stage I	0.938	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-AK-3436	Stage IV	-1.697	not mutated	7	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-AK-3440	Stage I	-0.215	not mutated	3	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AK-3443	Stage II	-1.285	not mutated	3	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AK-3444	Stage I	1.596	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-AK-3445	Stage III	0.691	mutated	7	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-AK-3447	Stage II	1.843	not mutated	3	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AK-3450	Stage I	2.008	mutated	2	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-AK-3451	Stage II	-1.038	mutated	4	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AK-3453	Stage II	-0.215	not mutated	3	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AK-3454	Stage I	1.925	not mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-AK-3455	Stage III	0.855	mutated	5	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AK-3456	Stage II	-1.038	not mutated	4	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AK-3458	Stage I	-1.038	mutated	3	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AK-3460	Stage I	-0.215	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-AK-3461	Stage I	0.938	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-AK-3465	Stage I	0.855	not mutated	2	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AS-3777	Stage I	0.197	not mutated	2	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-AS-3778	Stage I	-2.108	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-4688	Stage IV	-1.203	not mutated	12	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4690	Stage IV	0.361	mutated	13	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4691	Stage IV	-0.462	not mutated	10	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4693	Stage III	0.938	not mutated	6	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-4694	Stage III	0.938	not mutated	6	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4696	Stage III	-0.215	not mutated	7	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4697	Stage IV	-1.203	not mutated	10	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4698	Stage IV	1.184	not mutated	8	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4699	Stage IV	1.102	not mutated	12	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4700	Stage IV	-0.050	mutated	12	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4701	Stage IV	0.444	not mutated	11	Cluster C	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-4703	Stage IV	-0.791	mutated	10	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4706	Stage III	0.032	mutated	6	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4707	Stage III	0.197	not mutated	6	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4710	Stage III	1.184	mutated	7	Cluster C	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-4712	Stage IV	1.267	mutated	11	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4713	Stage III	1.267	mutated	6	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4714	Stage IV	1.678	mutated	11	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B0-4718	Stage III	-0.297	not mutated	6	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-4810	Stage III	-1.120	mutated	9	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4811	Stage III	-1.038	not mutated	7	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4813	Stage III	0.608	not mutated	7	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4814	Stage IV	-0.215	mutated	13	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-4815	Stage III	0.361	not mutated	6	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4816	Stage II	-0.956	mutated	6	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-4817	Stage III	1.678	not mutated	7	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4818	Stage II	0.608	mutated	6	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-4819	Stage IV	-0.050	mutated	10	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4821	Stage III	0.608	not mutated	7	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4822	Stage II	1.431	mutated	5	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4823	Stage I	2.255	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-4824	Stage I	-0.956	mutated	5	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-4827	Stage III	1.349	mutated	6	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4828	Stage IV	1.514	mutated	10	Cluster C	TGF-b signature score high	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4833	Stage I	1.761	mutated	4	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-4834	Stage I	-0.956	not mutated	5	Cluster B	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-4836	Stage IV	0.032	not mutated	11	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4837	Stage I	0.197	not mutated	5	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4838	Stage I	0.691	not mutated	5	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)

TCGA-B0-4839	Stage I	1.596	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4841	Stage IV	0.197	not mutated	10	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-4842	Stage III	1.020	mutated	6	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4843	Stage III	-0.297	not mutated	7	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4844	Stage IV	-0.050	not mutated	11	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4845	Stage IV	0.773	mutated	10	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-4846	Stage IV	-0.709	not mutated	10	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-4847	Stage IV	-0.050	mutated	11	Cluster C	TGF-b signature score low	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4848	Stage III	-0.544	not mutated	7	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-4849	Stage III	-0.791	not mutated	7	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-4852	Stage II	1.431	mutated	5	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-4945	Stage I	1.184	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5075	Stage III	1.349	mutated	6	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5077	Stage I	1.349	mutated	5	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5080	Stage IV	0.197	not mutated	11	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-5081	Stage III	1.514	mutated	6	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5083	Stage I	0.197	not mutated	5	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B0-5084	Stage IV	-2.273	not mutated	13	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5085	Stage III	1.267	mutated	7	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B0-5088	Stage I	-0.627	mutated	5	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5092	Stage IV	-0.627	not mutated	9	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5094	Stage IV	0.114	not mutated	10	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5095	Stage III	1.678	mutated	7	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-5096	Stage III	0.938	mutated	8	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5097	Stage III	-0.133	mutated	6	Cluster A	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-5098	Stage I	-0.627	not mutated	5	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5099	Stage III	2.255	mutated	7	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5100	Stage III	0.938	not mutated	7	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5102	Stage I	1.102	mutated	5	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5104	Stage I	2.419	mutated	4	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5106	Stage I	0.279	mutated	4	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5107	Stage IV	0.361	mutated	9	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5108	Stage III	-0.544	mutated	6	Cluster A	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5109	Stage III	0.691	not mutated	8	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5110	Stage I	0.855	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5113	Stage III	0.691	mutated	6	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5115	Stage IV	-1.450	mutated	10	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5116	Stage III	-0.709	mutated	7	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-5117	Stage I	-1.697	not mutated	4	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B0-5119	Stage I	0.032	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5120	Stage I	0.938	not mutated	4	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-5121	Stage I	-0.380	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-5399	Stage I	-1.203	mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5400	Stage III	-0.133	not mutated	4	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5402	Stage IV	0.279	mutated	6	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5690	Stage I	-0.627	not mutated	4	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5691	Stage I	0.444	mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-5692	Stage III	0.444	mutated	7	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5693	Stage I	-1.120	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5694	Stage III	0.855	not mutated	7	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-5695	Stage I	0.032	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5696	Stage III	0.691	mutated	6	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5697	Stage I	-0.873	not mutated	4	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5698	Stage I	1.349	not mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5699	Stage I	-0.627	mutated	4	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5700	Stage I	1.349	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5701	Stage III	0.361	not mutated	6	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5702	Stage I	0.855	not mutated	4	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5703	Stage I	1.020	mutated	5	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5705	Stage I	0.361	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5706	Stage II	-1.285	not mutated	5	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5707	Stage I	-1.779	not mutated	5	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B0-5709	Stage III	0.114	mutated	7	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B0-5710	Stage I	-0.297	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5711	Stage III	-0.873	mutated	7	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5712	Stage IV	0.608	not mutated	10	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5713	Stage III	1.184	mutated	7	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B0-5812	Stage I	-0.627	mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B2-3923	Stage II	-0.133	not mutated	5	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B2-3924	Stage I	1.020	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B2-4098	Stage I	0.938	not mutated	4	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B2-4099	Stage I	1.843	not mutated	5	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B2-4101	Stage II	-0.709	not mutated	6	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B2-4102	Stage I	0.032	not mutated	4	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B2-5633	Stage I	-0.380	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B2-5635	Stage I	1.102	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B2-5636	Stage I	1.514	not mutated	4	Cluster A	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B2-5639	Stage IV	-1.203	not mutated	11	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B2-5641	Stage I	1.514	not mutated	5	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B2-A4SR	Stage II	0.032	not mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B4-5377	Stage IV	0.608	not mutated	11	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B4-5378	Stage I	0.114	not mutated	4	Cluster A	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B4-5832	Stage III	0.361	mutated	4	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B4-5834	Stage I	-0.133	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B4-5835	Stage I	0.279	not mutated	2	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B4-5836	Stage I	0.032	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B4-5838	[Discrepancy]	-0.709	not mutated	6	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B4-5843	Stage I	-1.285	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B4-5844	Stage II	0.032	not mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-4143	Stage IV	0.444	not mutated	9	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B8-4146	Stage I	-1.614	not mutated	0	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-4148	Stage I	0.197	mutated	3	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-4151	Stage III	-0.791	mutated	4	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-4153	Stage III	1.102	mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B8-4154	Stage I	1.020	mutated	2	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-4619	Stage I	-0.215	not mutated	4	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B8-4620	Stage III	0.773	not mutated	6	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B8-4621	Stage I	0.197	mutated	5	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B8-4622	Stage IV	-0.297	not mutated	11	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-5158	Stage III	-0.380	not mutated	6	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B8-5159	Stage I	0.032	mutated	3	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-5162	Stage II	0.114	mutated	3	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B8-5163	Stage III	0.197	mutated	5	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B8-5164	Stage III	0.361	mutated	5	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B8-5165	Stage I	-1.450	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-5545	Stage I	-1.532	not mutated	2	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-5546	Stage I	-1.861	not mutated	2	Cluster B	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-5549	Stage I	-0.627	mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-5550	Stage III	0.855	mutated	5	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B8-5551	Stage I	0.361	mutated	3	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B8-5552	Stage I	-1.614	not mutated	0	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-5553	Stage I	0.526	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-A54D	Stage III	0.691	not mutated	6	Cluster C	TGF-b signature score low	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B8-A54E	Stage I	0.114	not mutated	5	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-B8-A54F	Stage I	-0.956	not mutated	2	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B8-A54G	Stage I	-0.873	not mutated	5	Cluster C	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-B8-A54H	Stage II	0.691	not mutated	6	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-B8-A54I	Stage I	-1.038	not mutated	3	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)

TCGA-B8-A54J	Stage II	-0.050	not mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B8-A54K	Stage I	0.032	not mutated	4	Cluster C	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B8-A7U6	Stage I	-0.544	not mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-B8-A8YJ	Stage I	-0.050	not mutated	4	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4158	Stage I	0.691	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4159	Stage I	0.773	mutated	2	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4160	Stage III	0.526	not mutated	6	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4161	Stage I	1.102	mutated	5	Cluster C	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4162	Stage I	0.361	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4163	Stage III	-0.050	not mutated	7	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4164	Stage III	-0.791	mutated	4	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4165	Stage I	0.279	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4166	Stage III	0.691	mutated	5	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4167	Stage III	-0.133	mutated	4	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4169	Stage II	1.267	not mutated	5	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4170	Stage I	0.938	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4173	Stage II	-1.120	not mutated	6	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4174	Stage II	-0.956	not mutated	4	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4176	Stage I	0.279	not mutated	2	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4177	Stage I	0.361	not mutated	2	Cluster A	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4325	Stage I	0.279	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4326	Stage I	-0.627	not mutated	2	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4327	Stage II	1.184	not mutated	3	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4329	Stage III	1.184	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4330	Stage III	-0.050	not mutated	4	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4331	Stage I	-0.709	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4332	Stage III	-2.026	not mutated	4	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4334	Stage III	-0.380	not mutated	5	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4335	Stage IV	0.361	not mutated	11	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4337	Stage III	1.267	not mutated	6	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4338	Stage I	-1.450	not mutated	3	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4340	Stage I	0.773	mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4341	Stage III	0.526	mutated	4	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4342	Stage II	1.514	mutated	6	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4343	Stage III	0.279	not mutated	7	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4344	Stage I	1.184	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4345	Stage III	0.114	not mutated	7	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4346	Stage III	-0.297	not mutated	7	Cluster C	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4347	Stage III	1.102	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4349	Stage I	0.608	mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4351	Stage III	-0.791	not mutated	4	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4352	Stage IV	1.102	not mutated	10	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4353	Stage I	0.032	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4354	Stage IV	-1.697	not mutated	14	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4355	Stage III	-0.133	mutated	4	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4756	Stage I	0.114	not mutated	2	Cluster B	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4758	Stage I	-1.697	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4759	Stage I	-0.873	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4760	Stage I	0.691	not mutated	0	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4761	Stage III	-0.297	not mutated	6	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4762	Stage I	-1.532	mutated	5	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4763	Stage I	1.514	mutated	2	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4765	Stage I	-1.450	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4766	Stage I	-1.450	not mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4768	Stage I	0.938	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4769	Stage I	0.197	not mutated	2	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4770	Stage IV	1.020	not mutated	6	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4771	Stage IV	0.114	mutated	10	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4774	Stage I	-0.297	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4775	Stage I	-0.462	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4776	Stage I	-0.709	not mutated	2	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4777	Stage I	-1.203	not mutated	3	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4781	Stage I	1.431	not mutated	3	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4782	Stage I	-0.462	mutated	2	Cluster B	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4784	Stage I	0.526	not mutated	2	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4787	Stage IV	-0.133	not mutated	10	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4789	Stage I	-1.038	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4790	Stage I	1.267	not mutated	2	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4795	Stage I	1.102	not mutated	2	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4797	Stage III	-2.191	not mutated	5	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4798	[Discrepancy]	1.102	mutated	8	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4799	Stage III	0.773	not mutated	7	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4801	Stage I	-0.297	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4803	Stage III	1.514	not mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4804	Stage I	-0.133	not mutated	2	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4807	Stage I	-1.532	mutated	3	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4959	Stage I	-0.956	not mutated	3	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4960	Stage II	-1.203	mutated	4	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4961	Stage I	-1.120	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4962	Stage II	-0.215	not mutated	5	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4963	Stage I	0.197	mutated	3	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4964	Stage I	-0.544	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4965	Stage I	-1.203	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4967	Stage III	1.267	mutated	4	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4968	Stage I	-1.697	mutated	5	Cluster A	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4969	Stage I	0.197	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4970	Stage III	-1.367	mutated	7	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4971	Stage III	-1.697	not mutated	5	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4972	Stage III	-1.450	not mutated	7	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4973	Stage III	-1.120	mutated	5	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4974	Stage IV	-0.215	mutated	10	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4975	Stage I	-1.697	mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4976	Stage I	1.349	mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4977	Stage I	-0.297	mutated	3	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4981	Stage III	1.184	not mutated	5	Cluster C	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4982	Stage I	-1.532	mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4983	Stage III	0.526	not mutated	6	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4985	Stage III	0.938	not mutated	6	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4986	Stage I	1.184	mutated	3	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4987	Stage I	-1.614	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4988	Stage I	0.938	mutated	4	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4989	Stage III	-0.215	mutated	7	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4991	Stage I	-0.544	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4992	Stage I	0.444	mutated	2	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-4993	Stage I	-0.215	not mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4994	Stage I	-0.544	not mutated	5	Cluster B	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-4995	Stage I	0.608	mutated	5	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-4998	Stage I	-0.956	mutated	3	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-4999	Stage I	-0.380	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5000	Stage I	-1.697	not mutated	3	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5001	Stage I	-1.450	mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5004	Stage I	-0.627	mutated	3	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5006	Stage I	0.032	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5007	Stage II	-1.285	mutated	5	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5008	Stage I	-1.203	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5009	Stage I	-0.709	mutated	5	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5010	Stage III	0.197	not mutated	4	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5168	Stage I	1.184	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5169	Stage I	0.773	mutated	2	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)

TCGA-BP-5170	Stage I	-0.462	mutated	2	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-5173	Stage I	1.184	mutated	2	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5174	Stage I	-1.285	mutated	2	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-5175	Stage I	-0.050	mutated	3	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5176	Stage I	1.431	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5177	Stage I	-1.203	mutated	3	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5178	Stage IV	0.855	not mutated	10	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5180	Stage I	-0.627	mutated	2	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5181	Stage I	-0.215	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5182	Stage I	-0.380	not mutated	3	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5183	Stage III	-0.297	mutated	5	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5184	Stage I	-0.544	mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5185	Stage I	-0.380	not mutated	3	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-BP-5186	Stage I	-0.873	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5187	Stage I	-0.544	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5189	Stage I	-0.050	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-BP-5190	Stage I	0.032	mutated	3	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5191	Stage III	1.514	not mutated	4	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5192	Stage I	-0.133	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5194	Stage I	-1.779	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5195	Stage I	1.184	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5196	Stage I	-0.627	mutated	2	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5198	Stage III	0.938	mutated	7	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5199	Stage II	-0.215	mutated	3	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5200	Stage II	-1.367	mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-BP-5201	Stage IV	0.197	not mutated	8	Cluster C	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-BP-5202	Stage III	1.184	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4634	Stage I	-0.050	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4635	Stage I	-1.038	not mutated	3	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4636	Stage III	-0.791	mutated	5	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4637	Stage IV	-0.709	mutated	7	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4638	Stage IV	-1.203	mutated	10	Cluster C	TGF-b signature score low	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-CJ-4639	Stage II	-0.956	not mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4640	Stage III	-0.956	not mutated	4	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4641	Stage IV	-0.462	not mutated	8	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4642	Stage II	-1.120	not mutated	1	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4643	Stage II	0.526	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4644	Stage IV	-1.038	mutated	9	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4868	Stage IV	-1.532	not mutated	11	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4869	Stage III	-0.956	not mutated	7	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4870	Stage III	-0.215	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4871	Stage IV	0.197	not mutated	8	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4872	Stage I	-0.791	mutated	2	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-CJ-4873	Stage III	2.008	not mutated	7	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4874	Stage I	1.020	not mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4875	Stage IV	0.526	mutated	9	Cluster C	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4876	Stage II	-0.297	not mutated	4	Cluster C	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-CJ-4878	Stage III	0.855	not mutated	4	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4881	Stage III	-1.614	not mutated	7	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4882	Stage III	-0.297	not mutated	7	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4884	Stage III	0.938	mutated	5	Cluster C	TGF-b signature score low	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-CJ-4885	Stage IV	0.279	mutated	11	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4886	Stage I	-1.532	not mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4887	Stage IV	-1.038	not mutated	9	Cluster C	TGF-b signature score high	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4888	Stage IV	-0.133	not mutated	10	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4889	Stage I	0.197	not mutated	2	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-CJ-4890	Stage IV	0.938	not mutated	8	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4891	Stage III	-0.297	not mutated	6	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4892	Stage I	0.361	mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4893	Stage I	1.267	not mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4894	Stage III	-0.215	mutated	5	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4895	Stage IV	0.114	not mutated	8	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4897	Stage III	1.514	not mutated	7	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4899	Stage I	-1.532	not mutated	4	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4900	Stage IV	0.691	mutated	14	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4901	Stage III	-1.120	not mutated	7	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4902	Stage III	0.032	not mutated	7	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-CJ-4903	Stage I	-0.873	mutated	3	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4904	Stage IV	-0.050	not mutated	9	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4905	Stage I	0.114	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4907	Stage III	-0.215	not mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4908	Stage I	-1.861	mutated	2	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-4912	Stage II	0.032	mutated	4	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4916	Stage III	0.691	mutated	5	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-4918	Stage IV	0.279	mutated	10	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-CJ-4920	Stage I	0.279	mutated	4	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-CJ-4923	Stage IV	0.197	not mutated	10	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-5671	Stage I	-0.791	not mutated	3	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-5672	Stage I	1.925	mutated	3	Cluster B	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-5675	Stage II	0.773	mutated	6	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-5676	Stage III	-1.120	mutated	7	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-5677	Stage IV	-0.544	mutated	10	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-5678	Stage IV	0.114	not mutated	8	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-5679	Stage III	1.020	mutated	6	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-5680	Stage IV	0.361	mutated	10	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-5681	Stage IV	-1.367	not mutated	11	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-5682	Stage IV	-0.050	not mutated	10	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-5683	Stage I	1.431	not mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-CJ-5684	Stage III	0.032	mutated	6	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-5686	Stage I	-0.133	mutated	3	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-5689	Stage I	2.419	not mutated	4	Cluster A	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-6027	Stage I	1.349	mutated	2	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-6028	Stage IV	-0.215	not mutated	10	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-CJ-6030	Stage I	0.361	mutated	5	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-6031	Stage I	-0.544	mutated	3	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CJ-6032	Stage II	0.197	mutated	6	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CJ-6033	Stage IV	-0.544	mutated	10	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CW-5580	Stage IV	1.020	mutated	9	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CW-5581	Stage I	-1.367	mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CW-5583	Stage I	-0.791	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CW-5584	Stage III	1.102	not mutated	7	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CW-5585	Stage IV	-0.791	not mutated	8	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CW-5587	Stage III	0.114	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CW-5588	Stage I	1.431	mutated	2	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-CW-5589	Stage I	-0.709	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CW-5590	Stage IV	-0.791	not mutated	9	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-CW-5591	Stage IV	-0.380	mutated	8	Cluster B	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CW-6087	Stage IV	0.032	not mutated	12	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CW-6088	Stage I	-0.050	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CW-6090	Stage I	0.608	mutated	5	Cluster B	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CW-6093	Stage I	1.020	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CW-6097	Stage III	-2.355	not mutated	6	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-4853	Stage I	1.761	mutated	2	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-4854	Stage I	0.608	not mutated	2	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-4856	Stage I	0.114	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-4857	Stage IV	-0.380	not mutated	9	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-4858	Stage II	-1.779	mutated	5	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-4859	Stage I	-0.133	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)

TCGA-CZ-4860	Stage IV	-0.050	not mutated	12	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-4861	Stage II	0.197	mutated	3	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-4862	Stage I	-1.203	not mutated	2	Cluster C	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-CZ-4863	Stage III	-0.791	not mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-CZ-4864	Stage II	2.090	not mutated	2	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-CZ-4865	Stage I	0.773	not mutated	2	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-4866	Stage I	1.514	not mutated	3	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-CZ-5451	Stage II	1.102	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-5452	Stage II	0.691	not mutated	5	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-5453	Stage II	0.526	not mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-5454	Stage IV	0.197	not mutated	9	Cluster A	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-5455	Stage IV	0.197	mutated	8	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-5456	Stage II	-0.297	not mutated	4	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-5457	Stage III	0.114	mutated	6	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-5458	Stage III	-1.450	mutated	5	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-5459	Stage III	0.197	not mutated	5	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-CZ-5460	Stage IV	-0.462	mutated	8	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-5461	Stage IV	-0.709	not mutated	6	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-5462	Stage IV	1.843	not mutated	9	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-CZ-5463	Stage II	1.267	not mutated	3	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-5464	Stage IV	0.691	not mutated	8	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-5465	Stage III	1.267	mutated	4	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-5466	Stage III	0.526	mutated	6	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-5467	Stage III	2.090	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-5468	Stage IV	1.925	not mutated	8	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-5469	Stage II	-1.614	mutated	5	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-5470	Stage II	0.938	mutated	4	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-CZ-5982	Stage I	-0.133	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-5984	Stage I	-0.791	mutated	3	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-5985	Stage II	-0.215	mutated	3	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-CZ-5986	Stage I	0.032	mutated	3	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-CZ-5987	Stage IV	-0.050	mutated	8	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-CZ-5988	Stage I	-1.861	not mutated	2	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-CZ-5989	Stage II	-0.050	mutated	3	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-DV-5565	Stage I	-0.133	not mutated	4	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-DV-5566	Stage I	0.526	not mutated	2	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-DV-5567	Stage I	-1.697	not mutated	2	Cluster A	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-DV-5568	Stage I	-2.849	not mutated	4	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-DV-5569	Stage I	-2.602	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-DV-5573	Stage I	-1.614	not mutated	2	Cluster B	TGF-b signature score low	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-DV-5574	Stage I	-1.944	not mutated	4	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-DV-5575	Stage I	-0.709	not mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-DV-5576	Stage I	-0.462	not mutated	2	Cluster A	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-DV-A4VX	Stage IV	-0.133	not mutated	4	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-DV-A4VZ	Stage I	-0.627	not mutated	2	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-DV-A4W0	Stage I	-0.462	not mutated	3	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-EU-5904	Stage I	-1.120	mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-EU-5905	Stage I	0.526	mutated	5	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-EU-5906	Stage I	-0.462	mutated	4	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-EU-5907	Stage III	1.678	mutated	7	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-G6-A5PC	Stage IV	-0.544	not mutated	6	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
TCGA-G6-A8L6	Stage IV	-0.462	not mutated	4	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-G6-A8L7	Stage I	1.678	not mutated	5	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-G6-A8L8	Stage I	0.114	not mutated	3	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-GK-A6C7	Stage I	1.267	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-MM-A563	[Discrepancy]	-1.614	not mutated	4	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-MM-A564	Stage II	0.608	not mutated	3	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
TCGA-MM-A84U	Stage I	-0.215	not mutated	4	Cluster C	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
TCGA-MW-A4EC	Stage I	0.938	not mutated	2	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
TCGA-T7-A92I	Stage I	-1.120	not mutated	2	Cluster B	TGF-b signature score low	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)

Supplemental Figure 14 – TCGA Patient Classification

Patient	CSS months	CSS event	ccRCC early score	ccRCC early Score - 3 risk groups	ccRCC early Score - 2 risk groups	ccRCC score	ccRCC Score - 3 risk groups	ccRCC Score - 2 risk groups	Grade	Gender
EGAR00001121490_ccRCC-1-tumor	111	0	-1.3265	Intermediate Risk	Low Risk	-1.4392	Intermediate Risk	Low Risk	G3	F
EGAR00001121491_ccRCC-2-tumor	73	1	-1.6033	Low Risk	Low Risk	-0.1187	Intermediate Risk	High Risk	G3	M
EGAR00001121492_ccRCC-3-tumor	8	1	0.0812	High Risk	High Risk	0.0881	Intermediate Risk	High Risk	G1 + G2	M
EGAR00001121493_ccRCC-4-tumor	9	1	-1.2060	Intermediate Risk	Low Risk	-1.3085	Intermediate Risk	Low Risk	G3	M
EGAR00001121494_ccRCC-5-tumor	98	0	-1.6090	Low Risk	Low Risk	-1.7457	Low Risk	Low Risk	G1 + G2	M
EGAR00001121495_ccRCC-6-tumor	94	0	-1.6239	Low Risk	Low Risk	-1.7620	Low Risk	Low Risk	G1 + G2	M
EGAR00001121496_ccRCC-7-tumor	11	0	-0.9036	Intermediate Risk	Low Risk	-0.9804	Intermediate Risk	Low Risk	G4	F
EGAR00001121497_ccRCC-8-tumor	80	0	-1.4044	Intermediate Risk	Low Risk	-1.5238	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121498_ccRCC-9-tumor	12	1	1.0406	High Risk	High Risk	2.7498	High Risk	High Risk	G4	M
EGAR00001121499_ccRCC-10-tumor	3	1	1.3218	High Risk	High Risk	3.0550	High Risk	High Risk	G4	M
EGAR00001121500_ccRCC-11-tumor	79	0	-1.7124	Low Risk	Low Risk	-1.8579	Low Risk	Low Risk	G1 + G2	M
EGAR00001121501_ccRCC-12-tumor	30	0	-1.5725	Low Risk	Low Risk	-1.7062	Low Risk	Low Risk	G1 + G2	M
EGAR00001121502_ccRCC-13-tumor	18	1	0.8895	High Risk	High Risk	0.9651	High Risk	High Risk	G3	F
EGAR00001121503_ccRCC-14-tumor	76	0	-0.2040	Intermediate Risk	High Risk	-0.2213	Intermediate Risk	High Risk	G1 + G2	M
EGAR00001121504_ccRCC-15-tumor	69	1	-0.9072	Intermediate Risk	Low Risk	-0.9844	Intermediate Risk	Low Risk	G1 + G2	F
EGAR00001121505_ccRCC-16-tumor	73	0	-0.9638	Intermediate Risk	Low Risk	-1.0457	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121506_ccRCC-17-tumor	74	0	-2.3761	Low Risk	Low Risk	-2.5781	Low Risk	Low Risk	G1 + G2	F
EGAR00001121507_ccRCC-18-tumor	71	0	-1.4420	Intermediate Risk	Low Risk	-1.5645	Intermediate Risk	Low Risk	G1 + G2	F
EGAR00001121508_ccRCC-19-tumor	70	0	-0.2870	Intermediate Risk	High Risk	-0.3114	Intermediate Risk	High Risk	NA	F
EGAR00001121509_ccRCC-20-tumor	31	0	-0.4391	Intermediate Risk	High Risk	1.1445	High Risk	High Risk	G1 + G2	F
EGAR00001121510_ccRCC-21-tumor	36	0	-1.6970	Low Risk	Low Risk	-1.8412	Low Risk	Low Risk	G1 + G2	M
EGAR00001121511_ccRCC-22-tumor	11	1	-0.8584	Intermediate Risk	Low Risk	-0.9313	Intermediate Risk	Low Risk	G3	M
EGAR00001121512_ccRCC-23-tumor	39	0	-1.5013	Intermediate Risk	Low Risk	-1.6289	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121513_ccRCC-24-tumor	25	1	0.1708	High Risk	High Risk	0.1853	Intermediate Risk	High Risk	G1 + G2	M
EGAR00001121514_ccRCC-25-tumor	45	0	-2.3976	Low Risk	Low Risk	-2.6014	Low Risk	Low Risk	G1 + G2	M
EGAR00001121515_ccRCC-26-tumor	60	0	-1.7103	Low Risk	Low Risk	-1.8557	Low Risk	Low Risk	G1 + G2	M
EGAR00001121516_ccRCC-27-tumor	53	0	-1.5494	Intermediate Risk	Low Risk	-1.6811	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121517_ccRCC-28-tumor	51	0	-1.2005	Intermediate Risk	Low Risk	-1.3025	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121518_ccRCC-29-tumor	57	0	-2.0845	Low Risk	Low Risk	-2.2616	Low Risk	Low Risk	G1 + G2	F
EGAR00001121519_ccRCC-30-tumor	14	1	0.3192	High Risk	High Risk	0.3463	Intermediate Risk	High Risk	G3	M
EGAR00001121520_ccRCC-31-tumor	52	0	-2.0950	Low Risk	Low Risk	-2.2731	Low Risk	Low Risk	G1 + G2	M
EGAR00001121521_ccRCC-32-tumor	44	1	0.2350	High Risk	High Risk	0.2550	Intermediate Risk	High Risk	G1 + G2	M
EGAR00001121522_ccRCC-34-tumor	51	0	-0.0921	Intermediate Risk	High Risk	1.5209	High Risk	High Risk	G1 + G2	M
EGAR00001121523_ccRCC-35-tumor	46	0	-1.2225	Intermediate Risk	Low Risk	-1.3264	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121524_ccRCC-36-tumor	47	0	0.3039	High Risk	High Risk	0.3297	Intermediate Risk	High Risk	G1 + G2	M
EGAR00001121525_ccRCC-37-tumor	50	0	-1.6083	Low Risk	Low Risk	-1.7449	Low Risk	Low Risk	G1 + G2	F
EGAR00001121526_ccRCC-38-tumor	46	0	-2.0146	Low Risk	Low Risk	-2.1858	Low Risk	Low Risk	G1 + G2	M
EGAR00001121527_ccRCC-39-tumor	30	0	-0.7494	Intermediate Risk	High Risk	-0.8131	Intermediate Risk	Low Risk	NA	M
EGAR00001121528_ccRCC-40-tumor	34	0	-1.8199	Low Risk	Low Risk	-1.9746	Low Risk	Low Risk	G1 + G2	M
EGAR00001121529_ccRCC-41-tumor	18	1	1.8056	High Risk	High Risk	1.9591	High Risk	High Risk	G4	M
EGAR00001121530_ccRCC-42-tumor	39	0	0.3163	High Risk	High Risk	0.3432	Intermediate Risk	High Risk	G1 + G2	M
EGAR00001121531_ccRCC-43-tumor	35	0	-1.7597	Low Risk	Low Risk	-1.9093	Low Risk	Low Risk	G1 + G2	F
EGAR00001121532_ccRCC-44-tumor	35	0	-1.5379	Intermediate Risk	Low Risk	-1.6686	Intermediate Risk	Low Risk	G1 + G2	F
EGAR00001121533_ccRCC-45-tumor	34	0	-1.9776	Low Risk	Low Risk	-2.1457	Low Risk	Low Risk	G4	M
EGAR00001121534_ccRCC-46-tumor	33	0	-1.2122	Intermediate Risk	Low Risk	-1.3152	Intermediate Risk	Low Risk	G3	M
EGAR00001121535_ccRCC-48-tumor	33	0	-1.7549	Low Risk	Low Risk	-1.9041	Low Risk	Low Risk	G1 + G2	M
EGAR00001121536_ccRCC-49-tumor	43	0	-0.8568	Intermediate Risk	Low Risk	-0.9297	Intermediate Risk	Low Risk	G3	M
EGAR00001121537_ccRCC-50-tumor	30	0	0.4298	High Risk	High Risk	0.4663	High Risk	High Risk	G1 + G2	M
EGAR00001121538_ccRCC-51-tumor	47	0	-0.7865	Intermediate Risk	Low Risk	-0.8534	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121539_ccRCC-52-tumor	45	0	-0.9151	Intermediate Risk	Low Risk	0.6280	High Risk	High Risk	G3	F
EGAR00001121540_ccRCC-53-tumor	42	0	-2.2533	Low Risk	Low Risk	-2.4448	Low Risk	Low Risk	G1 + G2	M
EGAR00001121541_ccRCC-54-tumor	40	0	-2.0502	Low Risk	Low Risk	-2.2245	Low Risk	Low Risk	G1 + G2	M
EGAR00001121542_ccRCC-55-tumor	42	0	0.7395	High Risk	High Risk	0.8024	High Risk	High Risk	G1 + G2	F
EGAR00001121543_ccRCC-56-tumor	65	0	-1.1911	Intermediate Risk	Low Risk	-1.2923	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121544_ccRCC-58-tumor	37	0	-1.1287	Intermediate Risk	Low Risk	-1.2246	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121545_ccRCC-59-tumor	31	0	-1.7811	Low Risk	Low Risk	-1.9325	Low Risk	Low Risk	G1 + G2	M
EGAR00001121546_ccRCC-60-tumor	49	0	-1.9838	Low Risk	Low Risk	-2.1524	Low Risk	Low Risk	G1 + G2	F
EGAR00001121547_ccRCC-61-tumor	49	0	-1.3827	Intermediate Risk	Low Risk	-1.5002	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121548_ccRCC-62-tumor	54	0	-1.2249	Intermediate Risk	Low Risk	-1.3290	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121549_ccRCC-64-tumor	81	0	0.9335	High Risk	High Risk	1.0128	High Risk	High Risk	G3	M
EGAR00001121550_ccRCC-65-tumor	10	1	1.5244	High Risk	High Risk	3.2747	High Risk	High Risk	G3	M
EGAR00001121551_ccRCC-66-tumor	84	0	0.3397	High Risk	High Risk	1.9894	High Risk	High Risk	G1 + G2	F
EGAR00001121552_ccRCC-67-tumor	42	1	-2.0308	Low Risk	Low Risk	-2.2034	Low Risk	Low Risk	G3	M
EGAR00001121553_ccRCC-68-tumor	88	0	-1.5059	Intermediate Risk	Low Risk	-1.6339	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121554_ccRCC-69-tumor	89	0	-1.1978	Intermediate Risk	Low Risk	-1.2996	Intermediate Risk	Low Risk	G3	M
EGAR00001121555_ccRCC-70-tumor	18	1	0.3527	High Risk	High Risk	2.0035	High Risk	High Risk	G3	M
EGAR00001121556_ccRCC-71-tumor	72	0	-2.3644	Low Risk	Low Risk	-2.5654	Low Risk	Low Risk	G1 + G2	M
EGAR00001121557_ccRCC-72-tumor	92	0	-1.8876	Low Risk	Low Risk	-2.0481	Low Risk	Low Risk	G3	M
EGAR00001121558_ccRCC-73-tumor	88	0	-1.8685	Low Risk	Low Risk	-2.0273	Low Risk	Low Risk	G1 + G2	M
EGAR00001121559_ccRCC-75-tumor	6	1	0.8896	High Risk	High Risk	2.5860	High Risk	High Risk	G1 + G2	M
EGAR00001121560_ccRCC-76-tumor	32	1	-0.5258	Intermediate Risk	High Risk	1.0503	High Risk	High Risk	G1 + G2	F
EGAR00001121561_ccRCC-77-tumor	62	0	-1.7710	Low Risk	Low Risk	-1.9215	Low Risk	Low Risk	G1 + G2	M
EGAR00001121562_ccRCC-78-tumor	87	0	-1.7984	Low Risk	Low Risk	-1.9512	Low Risk	Low Risk	G1 + G2	F
EGAR00001121563_ccRCC-79-tumor	89	0	-1.9604	Low Risk	Low Risk	-2.1271	Low Risk	Low Risk	G3	M
EGAR00001121564_ccRCC-80-tumor	100	0	-0.8139	Intermediate Risk	Low Risk	-0.8830	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121565_ccRCC-81-tumor	43	1	0.7416	High Risk	High Risk	0.8047	High Risk	High Risk	G3	M
EGAR00001121566_ccRCC-82-tumor	103	0	-1.8542	Low Risk	Low Risk	-2.0118	Low Risk	Low Risk	G3	M
EGAR00001121567_ccRCC-83-tumor	10	0	-2.4165	Low Risk	Low Risk	-2.6218	Low Risk	Low Risk	G1 + G2	M
EGAR00001121568_ccRCC-84-tumor	67	0	-1.9731	Low Risk	Low Risk	-2.1408	Low Risk	Low Risk	G3	M
EGAR00001121569_ccRCC-85-tumor	111	0	-1.5051	Intermediate Risk	Low Risk	-1.6330	Intermediate Risk	Low Risk	G3	F
EGAR00001121570_ccRCC-86-tumor	113	0	-2.6369	Low Risk	Low Risk	-2.8610	Low Risk	Low Risk	G3	F
EGAR00001121571_ccRCC-87-tumor	79	0	-2.3586	Low Risk	Low Risk	-2.5591	Low Risk	Low Risk	G1 + G2	M
EGAR00001121572_ccRCC-88-tumor	96	0	-1.6339	Low Risk	Low Risk	-1.7727	Low Risk	Low Risk	G1 + G2	M
EGAR00001121573_ccRCC-89-tumor	117	0	-1.3530	Intermediate Risk	Low Risk	-1.4680	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121574_ccRCC-90-tumor	39	1	-0.8332	Intermediate Risk	Low Risk	-0.9040	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121575_ccRCC-91-tumor	63	1	-1.3065	Intermediate Risk	Low Risk	-1.4175	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121576_ccRCC-92-tumor	122	0	-1.8884	Low Risk	Low Risk	-2.0489	Low Risk	Low Risk	G1 + G2	M
EGAR00001121577_ccRCC-93-tumor	122	0	-2.5353	Low Risk	Low Risk	-2.7508	Low Risk	Low Risk	G1 + G2	M
EGAR00001121578_ccRCC-94-tumor	16	1	-1.5433	Intermediate Risk	Low Risk	-1.6745	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121579_ccRCC-95-tumor	132	0	-0.3513	Intermediate Risk	High Risk	-0.3812	Intermediate Risk	High Risk	G1 + G2	M
EGAR00001121580_ccRCC-96-tumor	131	0	-2.0787	Low Risk	Low Risk	-2.2554	Low Risk	Low Risk	G1 + G2	M
EGAR00001121581_ccRCC-97-tumor	33	1	-0.3882	Intermediate Risk	High Risk	1.1996	High Risk	High Risk	G1 + G2	M
EGAR00001121582_ccRCC-98-tumor	1	0	-2.6796	Low Risk	Low Risk	-2.9073	Low Risk	Low Risk	G3	M
EGAR00001121583_ccRCC-99-tumor	111	0	0.4573	High Risk	High Risk	0.4962	High Risk	High Risk	G1 + G2	F
EGAR00001121584_ccRCC-100-tumor	142	0	-1.6210	Low Risk	Low Risk	-1.7587	Low Risk	Low Risk	G1 + G2	M
EGAR00001121585_ccRCC-102-tumor	142	0	-1.5620	Intermediate Risk	Low Risk	-1.6948	Intermediate Risk	Low Risk	G1 + G2	M
EGAR00001121586_ccRCC-103-tumor	145	0	-1.4122	Intermediate Risk	Low Risk	-1.5322	Intermediate Risk	Low Risk	G1 + G2	F
EGAR00001121587_ccRCC-104-tumor	116	0	-2.0876	Low Risk	Low Risk	-2.2650	Low Risk	Low Risk	G1 + G2	F
EGAR00001121588_ccRCC-105-tumor	73	0	-2.0656	Low Risk	Low Risk	-2.2411	Low Risk	Low Risk	G1 + G2	M
EGAR00001121589_ccRCC-106-tumor	63	1	-1.6829	Low Risk	Low Risk	-1.8259	Low Risk	Low Risk	G1 + G2	M

Patient	Stage	Standardized Age	New Metastases	VHL nonsyn mutation	Beleut signature (A vs. B vs. C)	Boström signature	ClearCode34 signature	Kosari signature	Lane signature	Zhao signature
EGAR00001121490_ccRCC-1-tumor	I	-0.215	yes	mutated	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121491_ccRCC-2-tumor	IV	0.855	yes	not mutated	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121492_ccRCC-3-tumor	III	0.855	yes	mutated	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121493_ccRCC-4-tumor	I	1.761	yes	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121494_ccRCC-5-tumor	I	1.349	yes	mutated	Cluster C	TGF-b signature score high	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 1 (good)
EGAR00001121495_ccRCC-6-tumor	I	1.184	no	mutated	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121496_ccRCC-7-tumor	II	2.172	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121497_ccRCC-8-tumor	I	1.020	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121498_ccRCC-9-tumor	IV	1.184	yes	mutated	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121499_ccRCC-10-tumor	IV	1.267	yes	mutated	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121500_ccRCC-11-tumor	I	1.267	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121501_ccRCC-12-tumor	I	1.596	no	mutated	Cluster B	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121502_ccRCC-13-tumor	III	-1.367	yes	mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121503_ccRCC-14-tumor	III	-0.215	yes	mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121504_ccRCC-15-tumor	II	0.361	no	mutated	Cluster B	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121505_ccRCC-16-tumor	II	-1.038	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121506_ccRCC-17-tumor	I	-0.544	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121507_ccRCC-18-tumor	I	-0.380	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121508_ccRCC-19-tumor	II	-1.779	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121509_ccRCC-20-tumor	IV	-0.380	yes	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121510_ccRCC-21-tumor	I	0.773	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121511_ccRCC-22-tumor	III	0.279	yes	mutated	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121512_ccRCC-23-tumor	I	0.938	no	not mutated	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121513_ccRCC-24-tumor	III	2.502	no	mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121514_ccRCC-25-tumor	I	0.361	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121515_ccRCC-26-tumor	I	-0.791	no	mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121516_ccRCC-27-tumor	I	-0.380	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121517_ccRCC-28-tumor	I	-0.133	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121518_ccRCC-29-tumor	I	1.020	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121519_ccRCC-30-tumor	II	0.279	yes	mutated	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121520_ccRCC-31-tumor	I	1.184	no	not mutated	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121521_ccRCC-32-tumor	II	-0.462	yes	not mutated	Cluster C	TGF-b signature score high	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121522_ccRCC-34-tumor	IV	0.361	yes	mutated	Cluster B	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121523_ccRCC-35-tumor	I	0.526	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121524_ccRCC-36-tumor	II	-0.133	no	mutated	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
EGAR00001121525_ccRCC-37-tumor	I	1.020	no	not mutated	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121526_ccRCC-38-tumor	I	-0.050	yes	not mutated	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121527_ccRCC-39-tumor	II	1.020	no	mutated	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
EGAR00001121528_ccRCC-40-tumor	I	0.361	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121529_ccRCC-41-tumor	III	1.761	yes	not mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121530_ccRCC-42-tumor	III	1.020	no	not mutated	Cluster A	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121531_ccRCC-43-tumor	I	-0.462	no	not mutated	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121532_ccRCC-44-tumor	I	-0.380	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121533_ccRCC-45-tumor	I	0.032	no	not mutated	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121534_ccRCC-46-tumor	I	0.032	no	mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121535_ccRCC-48-tumor	I	-1.532	no	not mutated	Cluster A	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121536_ccRCC-49-tumor	III	0.279	no	mutated	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121537_ccRCC-50-tumor	III	0.608	no	mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121538_ccRCC-51-tumor	III	0.773	no	mutated	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
EGAR00001121539_ccRCC-52-tumor	IV	0.444	yes	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121540_ccRCC-53-tumor	I	0.691	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121541_ccRCC-54-tumor	I	-0.297	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
EGAR00001121542_ccRCC-55-tumor	III	-0.215	yes	mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121543_ccRCC-56-tumor	I	1.431	no	mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121544_ccRCC-58-tumor	I	0.114	no	mutated	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121545_ccRCC-59-tumor	I	-0.380	no	mutated	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
EGAR00001121546_ccRCC-60-tumor	I	1.431	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121547_ccRCC-61-tumor	I	-0.380	no	mutated	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121548_ccRCC-62-tumor	I	1.431	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121549_ccRCC-64-tumor	III	-0.215	no	mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121550_ccRCC-65-tumor	IV	0.197	yes	mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121551_ccRCC-66-tumor	IV	1.020	yes	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
EGAR00001121552_ccRCC-67-tumor	I	-0.133	yes	mutated	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121553_ccRCC-68-tumor	I	0.691	no	mutated	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121554_ccRCC-69-tumor	I	1.596	yes	mutated	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121555_ccRCC-70-tumor	IV	0.773	yes	not mutated	Cluster A	TGF-b signature score low	ccB signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
EGAR00001121556_ccRCC-71-tumor	I	0.938	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121557_ccRCC-72-tumor	I	-2.108	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121558_ccRCC-73-tumor	I	-0.380	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121559_ccRCC-75-tumor	IV	-0.215	yes	mutated	Cluster C	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121560_ccRCC-76-tumor	IV	1.678	yes	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121561_ccRCC-77-tumor	I	-1.614	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121562_ccRCC-78-tumor	I	0.773	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121563_ccRCC-79-tumor	I	-2.108	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121564_ccRCC-80-tumor	II	0.773	yes	mutated	Cluster B	TGF-b signature score high	ccA signature	Aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
EGAR00001121565_ccRCC-81-tumor	III	1.020	yes	not mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121566_ccRCC-82-tumor	I	-0.627	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121567_ccRCC-83-tumor	I	0.938	no	mutated	Cluster C	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121568_ccRCC-84-tumor	I	0.691	no	not mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121569_ccRCC-85-tumor	I	1.020	yes	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121570_ccRCC-86-tumor	I	1.102	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121571_ccRCC-87-tumor	I	-1.697	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121572_ccRCC-88-tumor	I	0.279	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121573_ccRCC-89-tumor	I	-0.133	no	not mutated	Cluster B	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 2 (poor)
EGAR00001121574_ccRCC-90-tumor	I	0.938	no	not mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121575_ccRCC-91-tumor	I	0.526	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121576_ccRCC-92-tumor	I	-0.133	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121577_ccRCC-93-tumor	I	-0.050	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121578_ccRCC-94-tumor	I	0.691	yes	mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121579_ccRCC-95-tumor	I	-0.133	yes	mutated	Cluster A	TGF-b signature score low	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121580_ccRCC-96-tumor	I	-0.709	no	not mutated	Cluster B	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121581_ccRCC-97-tumor	IV	-1.697	yes	mutated	Cluster C	TGF-b signature score high	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121582_ccRCC-98-tumor	I	-0.544	no	not mutated	Cluster B	TGF-b signature score high	ccB signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121583_ccRCC-99-tumor	II	-0.050	no	mutated	Cluster A	TGF-b signature score low	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121584_ccRCC-100-tumor	I	0.361	yes	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121585_ccRCC-102-tumor	I	-0.627	no	mutated	Cluster B	TGF-b signature score high	ccB signature	Aggressive subgroup	Aggressive subgroup	Cluster 2 (poor)
EGAR00001121586_ccRCC-103-tumor	I	-0.462	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121587_ccRCC-104-tumor	I	-0.709	no	not mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121588_ccRCC-105-tumor	I	-0.133	no	mutated	Cluster A	TGF-b signature score low	ccA signature	Non-aggressive subgroup	Indolent subgroup	Cluster 1 (good)
EGAR00001121589_ccRCC-106-tumor	I	-0.215	no	mutated	Cluster A	TGF-b signature score low	ccB signature	Non-aggressive subgroup	Aggressive subgroup	Cluster 1 (good)

Supplemental Figure 15 – Sato *et al.* Patient Classification

List of Abbreviations

AJCC	American Joint Cancer Committee
ATCC	American Type Culture Collection
BAP1	BRCA1-associated protein
bFGF	Basic fibroblast growth factor
BH	Benjamini Hochberg
BSA	Bovine serum albumin
ccRCC	clear cell Renal Cell Carcinoma
CD	Cluster of differenciation
cDNA	Complementary DNA
CI	Confidence Interval
C-index	Concordance Index (or Harrel's C)
CG	CpG-islands (CGIs)
CNA	Copy number alteration
CNV	Copy number variation
CSS	Cancer-Specific Survival
CT	Computerized tomography
CTB	Cell Titer Blue
CTLA4	Cytotoxic T-lymphocyte-associated protein 4
Ctrl	Control
DAPI	4',6-diamidino-2-phenylindole
DKFZ	Deutsches Krebsforschungsinstitut
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EMT	Epithelial to mesenchymal transition
FACS	Fluorescence-activated cell sorting
FAM	6-carboxyfluorescein
FCS	Fetal calf serum
FDA	Food and Drug Administration
FDR	False discovery rate
FSC-A	Forward scatter area
FSC-H	Forward scatter height
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GEMMs	Genetically engineered mouse models
GEO	Gene Expression Omnibus

GFP	Green fluorescent protein
GSEA	Gene Set Enrichment Analysis
H&E	Hematoxylin and eosin
HIF	Hypoxia Induxible Factor
HI-STEM	Heidelberg Institute for Stem Cell Technology and Experimental Medicine
HR	Hazard Ratio
HRE	Hypoxia Response Element
IFN	Interferon
IGF	Insulin-like growth factor
IHC	Immunohistochemistry
IL-2	Interleukin-2
IMDM	Iscove's modified Dulbecco's medium
IQR	Interquartile range
LB	Lysogeny broth
logFC	Log fold change
mAB	Monoclonal antibody
MFI	Mean fluorescent intensity
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
mTOR	mechanistic Target of Rapamycin
NES	Normalized enrichment score
n.s.	Not significant
NSG	NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}
NT-control	Non-targeting control
nonsyn	Nonsynonymous
OS	Overall survival
PBRM1	Polybromo-1
PBS	Phosphate-buffered saline
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase chain reaction
PD-1	Programmed death 1
PDGF	Platelet-derived growth factor
PD-L1	Programmed death-ligand 1
PDX	Patient-derived xenograft
PE	Phycoerythrin
PFA	Paraformaldehyde

PI3K	Phosphatidylinositol-3-kinase
PIP2	phosphatidylinositol-4,5-bisphosphate
PIP	phosphatidylinositol-3,4,5-triphosphate
PKB	Protein kinase B
PPIA	Peptidylprolyl Isomerase A (cyclophilin A)
pPM	Primary patient material
PTEN	Phosphatase and tensin homolog
qRT-PCR	Quantitative real-time PCR
RFP	Red fluorescent protein
RHEB	Ras homolog enriched in brain
RNA	Ribonucleic acid
RNA-seq	RNA Sequencing
ROS	Reactive oxygen species
rpm	Revolution per minute
RT	Room temperature
RTK	Receptor Tyrosine Kinase
s.d.	Standard deviation
s.e.m.	Standard error of the mean
SETD2	Su(var), Enhancer of zeste, Trithorax-domain containing 2
shRNA	Short-hairpin ribonucleic acid
SNP	single nucleotide polymorphism
SSC	Side scatter
TCGA	The Cancer Genome Atlas
TKI	Tyrosine kinase inhibitor
TP53	Tumor Protein P53
TSC2	Tuberous sclerosis complex 2
TSPAN8	Tetraspanin 8
TSP-1	Thrombospondin-1
ULA	Ultra-low attachment
VEGF	Vascular endothelial growth factor
VHL	Von-Hippel-Lindau
WT	Wild type

List of Figures

List of Supplemental Figures

List of Tables

Contributions

The following people contributed to the work presented in this doctoral thesis:

The doctoral thesis was supervised by **Prof. Dr. Andreas Trumpp** and **Dr. Martin R. Sprick**.

The ccRCC patient derived xenograft model and cell lines were developed by **Dr. Teresa Dolt (formerly Rigo-Watermeier)** and ccRCC samples provided by **Prof. Dr. Klaus Höfner**, **Prof. Dr. Sascha Pahernik** and **Prof. Dr. Markus Hohenfellner**.

Corinna Klein, **Ornella Kossi** and **Vanessa Vogel** supported the work technically and experimentally.

Dr. Steffen Schmitt and **Ann Atzberger** from the DKFZ Flow Cytometry core facility supported in setting up flow cytometry experiments. **Manuela Brom** from the DKFZ Light Microscopy core facility provided support in all aspects of microscopy.

Tatjana Schmidt and **Sabine Henze** from the DKFZ Genomics and Proteomics core facility performed quality control and sample spotting onto bead arrays. **Oliver Heil** supported analysis of array data.

Whole exome sequencing and sequence alignment has been performed by the High Throughput Sequencing Unit of the Genomics & Proteomics Core Facility. **Dr. Gregor Warsow** performed the canopy and CNV analysis of the samples. **Frank Thommen** essentially supported with the cluster infrastructure.

Dr. Tim Holland-Letz contributed with the biostatistics of the *in vivo* experiments and was instrumental in statistical support of the regression and survival analyses.

Anja Rathgeb and her team of the Center for Preclinical Research monitored animal welfare.

Prof. Dr. Seishi Ogawa and **Yuseke Sato** from the Kyoto University made their gene expression patient cohort available to us.

Dr. Thomas Höfner and **Dr. Jan Philipp Radke** were involved in the clinical follow-up of Heidelberg ccRCC patients.

I want to thank especially **Andrea Barnert, Vera Thiel** and **Dr. Martin Sprick** for proofreading this doctoral thesis.

Acknowledgments

My greatest thanks go to **Dr. Martin Sprick**. On the first day of the selection, you convinced me to join you and the METICS, which was the best decision I could have made. You gave me the freedom to develop my PhD thesis to a project that I can now truly call my project. I am grateful that you always had an open door for me, supported me in whatever was needed, had great ideas and kept me on track. Thank you for the great time here in your METICS group.

I would like thank my PhD Supervisor **Prof. Dr. Andreas Trumpp** for the great opportunity to perform my PhD in your laboratory. I got the unique chance to develop my project and my scientific personality in such a great international environment. The DKFZ and especially HI-STEM are an extraordinary place to do science. I really enjoyed the opportunities you gave me to present my research at summer schools, international conferences and to develop into the person I am now.

Many thanks go to **Dr. Michael Milsom** for your role as my second referee and for many of your valuable contributions during lab meetings, retreats and car rides.

Thank you, **Prof. Dr. Jan Lohmann** and **Prof. Dr. Peter Angel** for being members of my defense committee.

I want to thank my TAC committee members **Prof. Dr. Stefan Fröhling** and **Dr. Albrecht Stenzinger** who supported me not only scientifically but encouraged me that my work is meaningful.

I would like to give a big thanks to the **METICS**-family, for the fun it was to work with you, for all your help and all of your support. **Lisa Becker, Magdalena Büscher, Teresa Dolt, Elisa Espinet, Corinna Klein, Ornella Kossi, Sarah-Jane Neuberth, Elisa Noll, Manuel Reitberger, Jonas Schwickert, Maximilian Seidel, Vera Thiel, Vanessa Vogel, Roberto Würth and Franziska Zickgraf.**

Corinna Klein, I don’t know what I would have done without you! You put me on the right track, are a divine mouse-surgeon and made this work possible. Thank you for giving me so much support and advice.

Vanessa Vogel and **Ornella Kossi** thank you for all the help with immunohistochemistry, qRT-PCR or cell culture.

Elisa Noll and **Manuel Reitberger**, thank you for the wonderful time in the greatest aquarium of all times. Elisa, you were always my inspiring example. For you, everything seemed to be so easy and playful, but you were always there for me and introduced me to PhD life. Manu, we made this office to a nerdy computer area. Together we learned programing, went to conferences, discussed and solved many scientific challenges and had a great time. You are a terrific colleague!

Maren Pein and **Mattia Falcone**, we met at the selection, started our PhD together at HI-STEM and now our journey together comes to an end. We experienced a truly intense time with many ups and downs and yet, even though we've grown apart, we are a great trio. You two have made me feel at home in Heidelberg, and I wish that no matter where the wind takes us, we will not lose what we had.

Thanks to all members of A010 and HI-STEM! Thanks especially to **Carsten Bahr, Julius Gräsel, Simon Haas, Pablo Hernández-Malmierca, Jasper Panten, Carmen Rothmund, Pia Sommerkamp, Shub Sood and Marc Thier**. Thank you, **Kristin Decker** for endless discussions about TSPAN8.

I want to thank **Marina Gilke**, **Erika Krückel** and **Dagmar Wolf** for all their indispensable administrative support.

My special thanks go to **Angela Laugsch** for your continuous support with coffee whenever I needed it.

Lena Wiedmann, we became very good friends during your endless stay at HI-STEM and I am glad to have you!

Alica Torkov, we were friends from the first moment on here in Heidelberg and together we had quite a ride! We founded interface, formed BioContact and went to crazy concerts, journeys or boat trips. Whenever you will ever have another great idea, I'm in!

I want to thank the whole **BioContact** association. You became my second Heidelberg family. I am grateful that I got to know you as friends **Sebastian Baars, Christin Elßner, Moritz Gartlgruber and Christina Lehrer**. **Daniel Baumann**, I still think that we have made our season something special, even though our ideas have been lost with the new generations.

Barbara Janssens, **Marion Gürth** and **Timo Kehl**, you supported me very much with all the career days, trainings and crazy projects ideas we had.

Thank you, my friends from the Freiburg clique, you have been supporting me for many years on my way to becoming a scientist: **Xavier Bemtgen, Stefan Bergmann, Lukas Dunkl, Anja and Lars Ellenrieder, Marvin Festag, Nicolas Gengenbacher, Sira Günther, Saskia Häbe, Jan Hülsdünker, Chris Kanne, Rhena Klar, Sebastian Kromer, Katharina Scharla, Dominik Wieland and Julia Wolanski.**

Ganz besonders danken möchte ich meinen Eltern, **Conny** und **Jörg**, und meinem Bruder **Max** für eure unermüdliche Unterstützung und euren Rückhalt, den ihr mir während meines Studiums und meiner Doktorarbeit gegeben habt. Vielen Dank, dass ihr immer für mich da gewesen seid, kritisch nachgefragt und verstanden habt, was mich wirklich beschäftigt. Ohne euch wäre ich nicht da, wo ich jetzt bin! DANKE!

Andrea, du bist wunderbar! Ich danke dir für Alles, für deine Unterstützung, dein Verständnis, deine Liebe! Ich freue mich so sehr, dass wir uns gefunden haben und dass du ein so wichtiger Teil meines Lebens geworden bist!