

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for
Mathematics of the
Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Marcel Waschow, Master of Science, Biology
born in Hamburg, Germany
Oral examination: December 5th, 2017

siRNA high-content screening of 3D cultured
MCF10A spheroids probing breast tumorigenesis

Referees: Prof. Dr. Roland Eils

Prof. Dr. Benedikt Brors

Prof. Dr. Elmar Schiebel

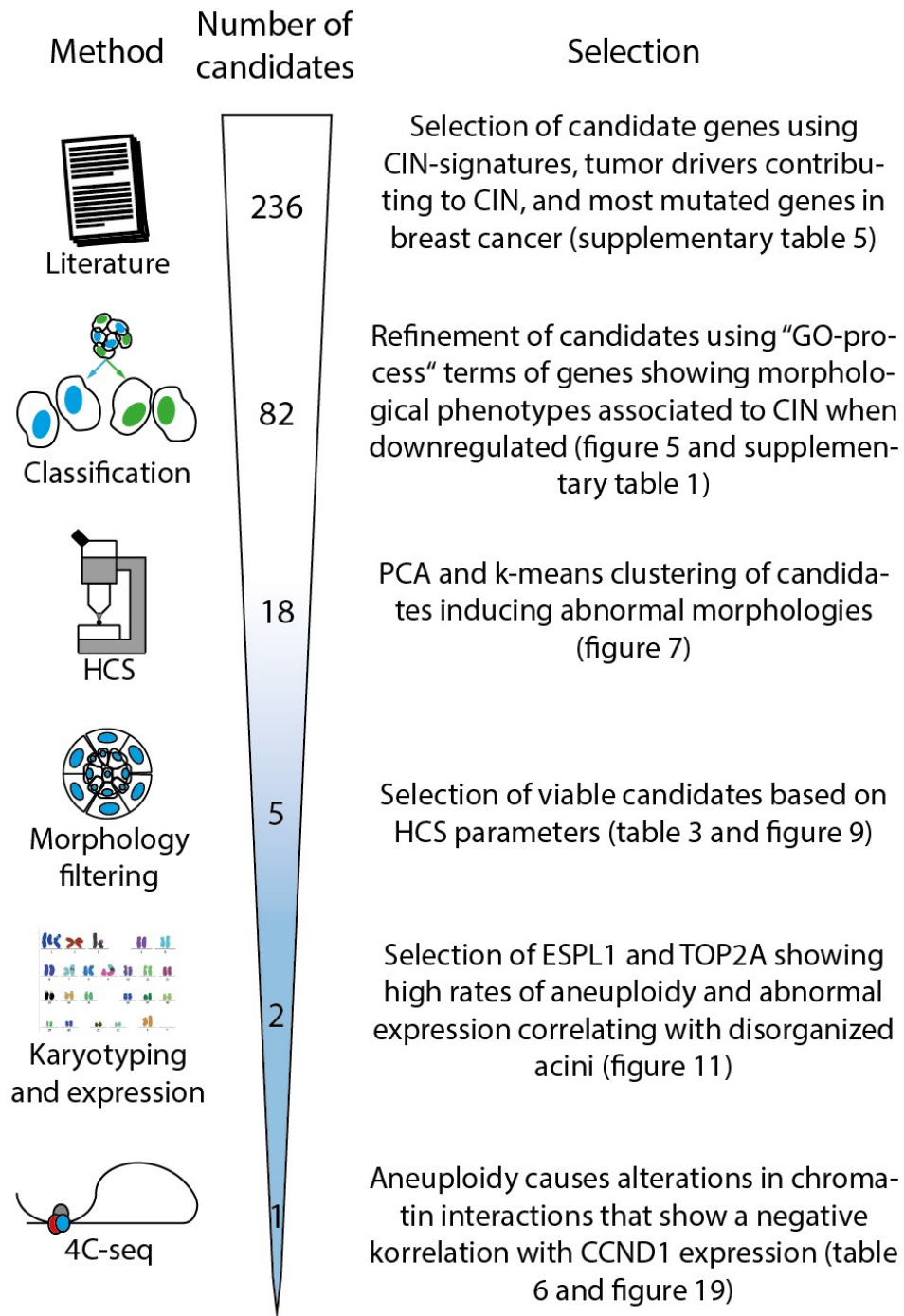
Prof. Dr. Stefan Wölfl

Abstract

Genome instability and its resulting phenotypes has long been studied and found to be a key factor for cancer development and treatment success. One of the key processes is chromosomal instability (CIN), which results in aneuploidy and is a recurring property of cancer cells. It is considered to be a main contributor to tumor heterogeneity. Breast cancer is the most common form of cancer in women today and shows aneuploidy in approximately 75 % of all cases. Forms with a lower association to aneuploidy, e.g. the luminal subtypes generally have a better prognosis when compared to basal-like or triple-negative ones, which frequently show high rates of aneuploidy. The precursor forms of breast cancer are ductal-, and lobular carcinoma in situ, whose early stages show similar rates of aneuploidy similar to their invasive later stages. This suggests that aneuploidy is a central mechanism in early cancer development and potentially initiation. To test how aneuploidy confers the ability to initiate and drive cancer, we use MCF10A cells in a 3D cell culture model, which reflects many of the physiological properties of human breast lobules. We introduce aneuploidy by downregulating a variety of tumor driver genes that are correlated to CIN in cancer and measure the effects on morphogenesis, expression, tumorigenicity, and chromatin conformation.

Using a high-content screening approach, we could identify two genes namely ESPL1 and TOP2A that upon knockdown induce abnormal mitosis resulting in aneuploidy. These cells develop into disorganized acini, which share features of lobular neoplasia, e.g. no hollow lumen and disorganized cell as well as acinus structure. Microarray expression profiles revealed a deregulation of breast cancer related genes including CD24, CD44, ALDH1A3 and CCND1. Interestingly, the chromosome conformation capture (4C-seq) experiments displayed a modified enhancer and CTCF binding landscape for CCND1, which correlated with the upregulation of CCND1. However, the deregulated expression was not sufficient to transform MCF10A cells to form persistent tumors in mice.

In conclusion, random aneuploidy influences the chromatin conformation and is capable of deregulating cancer driver genes. This highlights a new mechanism of cancer initiation and development; aneuploidy not only provides a tool for evolution but functions also as a tumor driver.



Abstract figure: Graphical overview of the candidate selection process.

Zusammenfassung

Genomische Instabilität und ihre Ursachen sind wichtige Faktoren in der Krebsentstehung zu sein, und beeinflussen den Behandlungserfolg. Häufig äußert sich GIN in Form von chromosomaler Instabilität (CIN), wodurch Aneuploidie verursacht wird. Aneuploidie wiederum ist eine der Hauptursachen für Krebsheterogenität. Brustkrebs beispielsweise ist in 75% der Fälle aneuploid und stellt die häufigste Form von Krebs bei Frauen dar. Spezielle Untergruppen wie luminaler Brustkrebs zeigen weniger Aneuploidie und haben eine bessere Prognose als basal- oder triple-negative Brusttumoren. Duktale- und lobuläre Karzinome in situ werden als Vorstufen von Brustkrebs anerkannt und haben schon früh ein Ausmaß an Aneuploidie, das mit invasiven Krebsformen vergleichbar ist. Hieraus lässt sich ein wesentlicher Stellenwert bei der Entwicklung, und potentiell auch bei der Entstehung von Krebs, ableiten. Um herauszufinden wie Aneuploidie die Krebsentwicklung und Entstehung begünstigt, verwenden wir MCF10A Zellen in einem 3D Zellkulturmodell, welches Ähnlichkeit mit der physiologischen Brustdrüse aufweist. In dem Modell induzieren wir Aneuploidie, indem wir Gene herunterregulieren die mit CIN und Krebs in Verbindung gebracht werden und messen die Effekte auf die Morphogenese, Genexpression, Tumorigenität und Chromatininteraktion.

In einem high-content screening Verfahren konnten wir die zwei Gene ESPL1 und TOP2A identifizieren, die die Mitose stören, wodurch Aneuploidie verursacht wird. Die aneuploiden Zellen entwickeln sich zu Azini, welche lobulären Neoplasien ähneln und ein fehlendes Lumen, sowie eine disorganisierte Zell- und Azinusstruktur aufweisen. Diese zeigten bei Expressionsversuchen eine Deregulation von CD24, CD44, ALDH1A3 und CCND1 gezeigt und zu einer Änderung der Interaktionen von Enhancern und CTCF-Bindestellen geführt. Diese veränderten Interaktionen weisen eine Korrelation mit der CCND1 Hochregulierung auf. Allerdings hat die induzierte Aneuploidie nicht zu einer Transformation geführt, um persistente Tumoren in Mäusen zu bilden.

Diese Ergebnisse zeigen, dass Aneuploidie die Chromatininteraktionen beeinflusst. Dies wiederum bedeutet, dass Aneuploidie nicht nur ein evolutionäres Werkzeug für Krebs ist, sondern aktiv krebserlevante Gene deregulieren kann.

List of contents

List of tables and figures	B
Tables	B
Figures	B
Abbreviations	E
1 Introduction	1
1.1 General introduction	1
1.2 Chromosomal instability	1
1.2.2 The converse nature of early CIN	2
1.2.3 Heterogeneity, drug resistance, and recurrence	3
1.2.4 CIN and epigenetics	4
1.3 Breast Cancer	5
1.3.1 Molecular properties of breast cancer	6
1.4 MCF10A 3D cell culture	7
1.4.1 MCF10 cell line series	8
1.5 High-content siRNA screening	9
1.6 Scope of the study	11
2 Material and Methods	12
2.1 Monolayer (2D) cell culture	12
2.2 3D cell culture	12
2.3 RNAi experiments	13
2.4 Quantitative PCR	13

2.4	Immunofluorescence for monolayer and 3D cell culture	14
2.4.1	SiR dyes (Lukinavičius et al., 2014)	15
2.5	Drug tests	15
2.6	Microarray expression profiling	15
2.7	Multiplex fluorescence in situ hybridization (M-FISH)	16
2.8	Mammosphere assay (modified after Shaw et al., 2012)	17
2.9	Xenografts	18
2.10	Chromatin conformation capture (4C)	18
2.11	Data analysis	20
2.11.1	High-content screen immunostainings	20
2.11.3	Time-lapse analysis	22
2.11.4	Breast cancer stem cell marker analysis	22
2.11.5	Drug sensitivity testing	22
2.11.6	Mammosphere formation	22
2.11.7	4C sequencing analysis	23
2.12	Statistical Methods	24
2.12.2	List similarity	24
3	Results	25
3.1	High-content screen	25
3.2	Karyotype and acinus architecture	32
3.3	Time-lapse imaging and analysis	34
3.4	Expression profiling	36

3.5	Validation of the expression changes of MCF10dE and MCF10dT.....	40
3.6	Drug tests	41
3.7	Testing tumorigenic potential in-vitro and in-vivo.....	42
3.8	Chromatin conformation capture (4C-seq).....	44
4	Discussion	48
4.1	Finding the connection of morphology, karyotype and cancer using siRNA HCS.....	51
4.2	Cancer marker expression and its effects	53
4.3	The effects of aneuploidy on the genome and its regulation	57
4.4	Conclusion	60
5	Outlook.....	61
6	Literature	63
7	Supplementary Data	79
7.1	Supplementary results.....	79
7.2	Supplementary figures	81
7.3	Supplementary tables.....	82
	Acknowledgements.....	106

List of tables and figures

Tables

Table 1: Prevalence of aneuploid cancers by entities (Cimini, 2008, modified).....	2
Table 2: Characteristics of BLCA related molecular subtypes compared to all other subtypes (Foulkes & Smith, 2010, modified).....	6
Table 3: Selected candidates with their respective visual selection criteria and cluster.	29
Table 4: Number of up- and downregulated genes of siScrambled, siESPL1 and siTOP2A cultured in 2D and 3D.	37
Table 5: Enrichment of top overexpressed genes and their associated enriched gene sets..	38
Table 6: Cis-chromatin interaction value by viewpoint and cell line.....	44

Figures

Figure 1: CCND1 expression is insensitive against copy number variations and ploidy in invasive breast carcinoma (data downloaded from COSMIC).	4
Figure 2: Cryo-slices of the MCF10 progression cell line and their increasingly malign properties.	9
Figure 3: Comparison of a healthy acinus with perturbed acini and ex-vivo DCIS and LCIS sections.	10
Figure 4: Restriction fragment definition and orientation (Klein et al., 2015).	23
Figure 5: Screening candidates categorized into 7 “GO-process” defined groups.....	26
Figure 6: Comparison of the parameters size, hollowness, and polarization of a random 25 % sample (n=20) of the candidates between single and double knockdown at day 3, 6 and 9.	27
Figure 7: Selection of candidates using PCA and k-means clustering plotted by PC1 vs. PC2 and PC1 vs. PC3.	28
Figure 8: Overview of examples for morphology filtering using visual phenotypes.	30

Figure 9: Final 5 candidates in comparison with the negative and scrambled control illustrating morphological changes of the acinus architecture defined by restructured nucleus (H2B-GFP) and Golgi (GM130) architecture and DNA-damage (H2AFX).	32
Figure 10: Comparison of a normal MCF10A karyotype with induced near tetraploid/aneuploid karyotypes.	33
Figure 11: Correlation of aneuploid cells and formation of disorganized acini.....	34
Figure 12: Acinus feature development tracked over time compared between a normal and two abnormal spheroids.....	35
Figure 13: Mean velocity of the nuclei that migrate in the acini during different stages of the development.	36
Figure 14: Abnormal mitotic phenotypes of siESPL1 and siTOP2A in comparison to a normal division.	36
Figure 15: Relative expression of breast cancer related genes.	40
Figure 16: Immunofluorescence of BCSC markers in abnormal spheroids in comparison to normal spheroids. 5 representative spheroids per condition were imaged.	41
Figure 17: Drug response curves of siScrambled, siESPL1 and siTOP2A cultured in 2D and 3D treated with paclitaxel.	42
Figure 18: Comparison of the mammosphere forming efficiency from cells previously cultured as monolayer or on Matrigel.	43
Figure 19: Correlation of cCIV and the expression level of CCND1 in the different cell lines.	45
Figure 20: Far-cis interaction probability.....	46
Figure 21: Near-cis interaction probability and alteration of the enhancer and CTCF binding site landscape.....	47
Figure 22: Hallmarks of cancer and potential treatment strategies as proposed by Hanahan and Weinberg (Hanahan & Weinberg 2011).	48

Figure 23: Cooperation of separase (ESPL1) and TOP2A during sister chromatid separation in the anaphase (Chen et al., 2015).....	52
Figure 24: The central role of Cyclin D (CCND1 and CCND2) during the G1 phase (Otto & Sicinski 2017).....	55
Figure 25: TOP2B and separase are involved in chromatin interaction rearrangements (Canela et al., 2017 and Petronczki et al., 2003, modified).....	60
Figure 26: Proposed model of CCND1 expression changes in abnormal mitosis and resulting development of CIS.	61

Abbreviations

°C:	Degree Celsius
2D/3D:	2-dimensional/3-dimensional
BCSC:	Breast cancer stem cell
BLCA:	Basal-like cancer / basal-like tumors
bp:	base pairs
CCA:	Connected component analysis
CCD:	Charge-coupled device
cCIV:	Cis-chromatine interaction value
cDNA:	Complementary DNA
CIN:	Chromosomal instability
COSMIC:	Catalogue Of Somatic Mutations In Cancer
CTCF:	CCCTC binding factor
DAPI:	4.6-diamidino-2-phenylindole
DCIS:	Ductal carcinoma in situ
DEL/del:	Deletion
DER/der:	Derivative
ddwater:	double distilled water or MilliQ water, water that was distilled twice
dKD:	Double-knockdown
DKFZ:	German cancer research center (Deutsches Krebsforschungsinstitut)
DMEM:	Dulbecco's Modified Eagle Medium

DNA:	Deoxyribonucleic acid
DOP-PCR:	Degenerative oligonucleotide primed PCR
EDTA:	Ethylenediaminetetraacetic acid
EGF:	Epidermal growth factor
EMBL:	The European Molecular Biology Laboratory
EMBL-CPD:	EMBL-cellular phenotype database
ER:	Estrogen receptor
ER α :	Estrogen receptor alpha
Et al.:	et alii, engl.: and others
FDR:	False discovery rate
Fig.:	Figure
FISH:	Fluorescence in situ hybridization
g:	Gravitational force
GAMMA:	Interleukin 2 receptor gamma
GFP:	Green fluorescent protein
GIN:	Genomic instability
GSEA:	Gene set enrichment analysis
h:	Hour(s)
HCS:	High-content screen
i:	Isochromosome
IC50:	Half maximal inhibitory concentration

IF:	Immunofluorescence
KCl:	Potassium chloride
KD:	Knockdown
KNIME:	The Konstanz information miner
LCIS:	Lobular carcinoma in situ
SiESPL1:	MCF10A cells treated with siRNA targeting ESPL1 RNA
SiScrambled:	MCF10A cells treated with siRNA targeting no specific RNA
SiTOP2A:	MCF10A cells treated with siRNA targeting TOP2A RNA
MFE:	Mammosphere forming efficiency
M-FISH:	Multiplex fluorescence in situ hybridization
Min:	Minutes
MIP:	Maximum intensity projection
ml:	Milliliters
n:	Number of...
nl:	Nano liter
NGS-mice:	NOD/GAMMA-SCID mice
NOD:	Non-obese diabetic
O/N:	Over night
p:	p-value
PBS:	Phosphate buffered saline
PC:	Principal component

PCA:	Principal component analysis
PCR:	Polymerase chain reaction
PgR:	Progesterone receptor
pH:	Potential of hydrogen
QPCR:	Quantitative polymerase chain reaction
RNA:	Ribonucleic acid
RPM:	Rotations per minute
RT:	Room temperature
s:	Second
SCID:	Severe combined immunodeficiency
shRNA:	Small hairpin RNA
SiR:	Silicone rhodamine
siRNA:	Small interfering RNA
sKD:	Single-knockdown
SRC:	Self-renewal capacity
SSC:	Saline-sodium citrate
T:	Time point
t:	Translocation
T/E:	Trypsin/EDTA
TNBC:	Triple-negative breast cancer
TNR-status:	Triple-negative receptor status

TSS: Transcription starting site

μ l: Micro-liter

1 Introduction

1.1 General introduction

In 1914 Theodor Boveri contributed to set the field of cancer genomics and made the first groundbreaking assumptions about the influence of aneuploidy on cancer development (Boveri, 2008). By examining sea urchin eggs, he discovered that chromosomes can be mis-segregated and accumulate in the daughter cells. Presuming these chromosomes are essential for cellular processes, he concluded that additional copies could lead to process deregulation and ultimately result in cancer (Boveri, 2008).

Today, the greatest obstacles in cancer therapy, besides metastasis, are the development of drug resistance upon treatment and non-responding cancers (Holohan et al., 2013). The number of patients suffering from such cancers indicates a lack of understanding of this disease and its diverse nature (Soto & Sonnenschein, 2012; Vogelstein et al., 2013; Hanahan & Weinberg, 2011). In the current view, this is to some extent attributable to tumor heterogeneity (Marusyk et al., 2012). Tumor heterogeneity results from a complex interplay between the individual genetic and epigenetic background as well as genomic instability (GIN). GIN includes structural, numerical, and sequence alterations, which are seen in nearly all cancers.

1.2 Chromosomal instability

Chromosomal instability (CIN) is a form of GIN where the healthy number and structure of whole chromosomes or parts of it cannot be maintained. Besides other mechanisms, this can be a result of missegregation of the chromosomes during division ultimately resulting in aneuploid cells (Gordon et al., 2012). Though aneuploid cells are more likely to suffer from CIN, not all cells that show aneuploidy are chromosomally instable. Many transcriptionally highly active cells like the puff-cells in *Drosophila melanogaster* (Korge, 1975) or somatic liver cells (Duncan et al., 2010) and neurons (Rehen et al., 2001) can show some form of aneuploidy, while being an important part of the healthy tissue. In a study by Knouse and colleagues using single cell sequencing less than 5% of all cells of healthy skin, brain and liver

samples showed aneuploidy (Knouse et al., 2014). In the majority of the cases, however, aneuploid cells pose a risk towards tumorigenesis (Clemente-Ruiz et al., 2016).

Aneuploidy is a common feature of many cancers (table 1, Cimini, 2008), and tumors diagnosed in early stages often contain aneuploid cells (Ottesen et al., 1995; Wang et al., 2014). Ottesen and colleagues could show that, independent of molecular subtype, cells from ductal carcinoma in situ (DCIS) tissues display high rates of aneuploidy. The frequency only slightly increases with progressing tumor stage suggesting an early occurrence.

Table 1: Prevalence of aneuploid cancers by entities (Cimini, 2008, modified)

Entity	Total aneuploid cancers
Bladder	157/192 (81.8 %)
Breast	598/800 (74.8 %)
Cervix uteri	75/84 (89.3 %)
Colon	301/340 (88.5 %)
Corpus uteri	116/165 (70.3 %)
Liver	110/155 (71.0 %)
Lung	413/435 (94.9 %)
Ovary	386/422 (91.5 %)
Prostate	151/186 (81.2 %)
Stomach	167/180 (92.8 %)

1.2.2 The converse nature of early CIN

High degrees of CIN as well as prolonged CIN have been shown to be potent cell death initiators (Clemente-Ruiz et al., 2016), even in absence of control mechanisms guided by p53 (Giam & Rancati, 2015). Conversely, it is proposed to be a key tool for evolution in cancer (Clemente-Ruiz et al., 2016). In the current understanding, CIN leads to the generation of a karyotypically diverse population that has a higher capability to adapt to environmental changes. The shaping of the aneuploid cancer landscape not only includes random gains and losses of chromosomes or part of them. Davoli and colleagues describe a misbalance of specific regions in the genome correlated to STOP-genes (genes that negatively regulate the cell cycle and that are enriched for tumor suppressors, e.g. PTEN, TP53, SMAD4, CDKN2A) and GO-genes (genes that positively regulate the cell cycle and are enriched for oncogenes e.g. IDH1, BRAF, KRAS, PIK3CA) that drive the selection; a phenomenon called cumulative

haploinsufficiency and triplosensitivity (Davoli et al., 2013). In this model the selection of specific aberrations in many regions of the genome collectively resulting in aneuploidy is driven by tumor suppressors that fail to operate, when there is only one allele left and oncogenes that disproportionally increase their oncogenic potential, as soon as there are more than 2 alleles.

The idea of CIN being an initiator, e.g. of colon cancer and retinoblastoma, is supported by mathematical models as proposed by Nowak and colleagues (Nowak et al., 2002). In another study, using single cell genome sequencing and copy number profiling, the authors suggest that TNBC cells acquire aneuploidy early (Wang et al., 2014). After initial selection of a stable clone with aneuploid karyotype from a previously unstable cell population, the cells start to accumulate mutations with a higher rate. The high mutation rate may then account for increased heterogeneity allowing some cells to better survive selective pressure like chemotherapy. What exactly drives the higher mutation rate was not further discussed.

Gao and colleagues propose a similar model where early short bursts of crisis lead to punctuated copy number changes (Gao et al., 2016). These changes remain stable, while the tumor mass expands. Other authors report a mutual dependence of CIN and mutations (Pino & Chung, 2014) or that certain mutations are necessary to induce aneuploidy for the progression of cancer (Drost et al., 2015).

1.2.3 Heterogeneity, drug resistance, and recurrence

The works of Selmecki and colleagues show that extra copies of key chromosomes can confer drug resistance in *candida albicans* against fluconazole (Selmecki, Forche & Berman, 2006). The same concept is applicable to mammalian cells. In 2001, Gorre and colleagues hypothesize that the mechanism of resistance in their chronic myeloid leukemia samples is attributable to a gene amplification (Gorre et al., 2001). Potentially, these amplifications are the result of early CIN and subsequent selection of cell populations harboring a variety of genetic changes. Such changes can include amplifications as well as translocations, fusions, deletions, and mutations that either circumvent cancer inhibition or activate cancer-driving processes giving rise to cells that appear to have a spectrum of new properties or abilities. The plasticity of such a cell population enables it to adapt to challenges like chemotherapy or targeted treatments, the immune system and finally normally lethal levels of CIN. Such a

“gain of function” of the cell population mediated by aneuploidy was shown by Sotillo and colleagues (Sotillo et al., 2010). Here, MAD2 overexpression in KRAS-driven lung tumors in mice leads to an increased tumor volume when compared to cells without MAD2 overexpression. In addition, aneuploid cells suffering from CIN relapsed significantly earlier after KRAS withdrawal and gave rise to highly aneuploid tumors. This suggests that aneuploidy is capable of conferring functions that mitigate the loss of previously necessary drivers like KRAS overexpression.

1.2.4 CIN and epigenetics

Genes are often deregulated when one allele is deleted or amplified. However, this is not always the case. One prominent example is CCND1, a gene mainly involved in cell cycle progression as well as a broad variety of other pathways. Importantly, CCND1 is often upregulated by overexpression or amplification in many cancer entities (Musgrove et al., 2011). A special case is invasive breast cancer where CCND1 expression does not correlate with copy number variations or ploidy here (figure 1, Forbes et al., 2017).

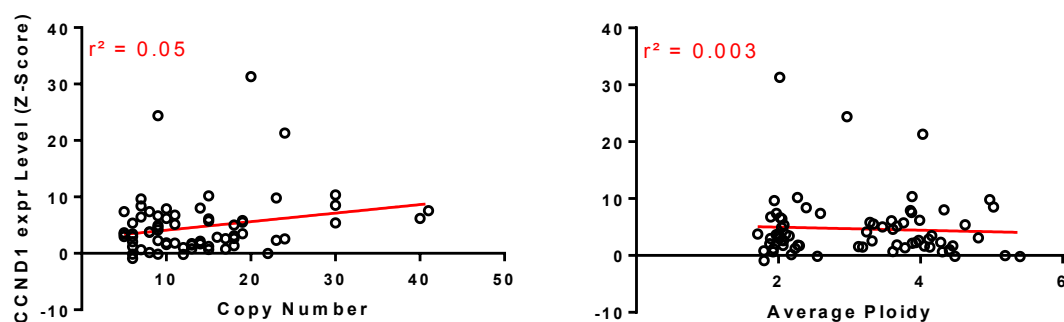


Figure 1: CCND1 expression is insensitive against copy number variations and ploidy in invasive breast carcinoma (data downloaded from COSMIC). Downregulation: Z-score < -2, upregulation: Z-score > 2.

This illustrates that aneuploidy alone does not dictate expression levels but relies on epigenetic regulation. Mechanistically, it is questionable how aneuploidy could directly affect epigenetic regulation, like DNA or histone modifications. However, there is another layer of epigenetic regulation, namely chromatin conformation, that is theoretically susceptible towards chromatin content changes. Since the development of chromatin conformation capture assays like 3C (Dekker et al., 2002), 4C (Simonis et al., 2006) and Hi-C (Lieberman-aiden et al., 2009) more insight about chromatin conformation and the

correlation to gene regulation could be gained. The theory behind these methods is that DNA-DNA interactions are mediated by proteins that bring two DNA loci in proximity to each other. These loci are cut by restriction enzymes, purified, ligated and then further analyzed using PCR or sequencing methods looking at either interactions of two single loci or even multiple loci versus multiple loci. Canela and colleagues could show that DNA-DNA interactions that are mediated by CCCTC binding factor (CTCF) and cohesin bound loop anchors correlate with translocation breakpoint regions in several cancers (Canela et al., 2017), thus showing the potential influence on tumorigenesis. Although, aneuploidy is a central property in cancer and chromatin capture technologies are extensively employed, the influence of aneuploidy on chromatin conformation and the consequences are elusive.

1.3 Breast Cancer

Breast cancer is a diverse group with subtypes, which have good examples for aneuploid tumors (table 1) that frequently relapse (Voduc et al., 2017, table 2). Worldwide, breast cancer is the most common cancer in women (Ferlay et al., 2013). Overall, it shows a 5 year survival rate of 80 - 90 % (Cancer Facts and Figures, 2016; Ferlay et al., 2013), but the prognosis strongly correlates with molecular subtype (Keegan et al., 2013, table 2). Basal-like tumors (BLCA) account for 15-20 % of all breast cancers and represent one subtype showing considerably higher aggressiveness (Foulkes & Smith, 2010; Hammond et al., 2010, table 2). However, these numbers vary between ethnical groups and age. Strongly correlated to this molecular breast cancer type is the lack of estrogen α -, progesterone- and HER2-receptor expression (table 2). 80 % of the basal-like cancers show this triple-negative status. Triple-negative breast cancers account for approximately 37 % of breast cancer related deaths in the first three years (Keegan et al., 2013). 75 % of breast cancers show aneuploidy (table 1, Cimini, 2008), and the molecular subtypes associated with basal-like subtypes generally show high degrees of aneuploidy (table 2, Foulkes & Smith, 2010).

Table 2: Characteristics of BLCA related molecular subtypes compared to all other subtypes (Foulkes & Smith, 2010, modified).

Characteristic	Subtype of Breast Cancer			
	Triple-Negative	Basal-like	BRCA1-Related	All Other Subtypes
<i>IHC expression</i>				
ER α	Negative, by definition	Usually negative	Usually negative	Usually positive
PgR	Negative, by definition	Almost always negative	Usually negative	Usually positive
HER2	Negative, by definition	Usually negative	Usually negative	Usually negative
<i>Other features</i>				
Degree of aneuploidy	Usually high	High	Very high	Variable
Gene-expression profile	Often basal-like and occasionally claudin-low	Basal-like, by definition	Usually basal-like	Not basal-like, by definition
Prognosis in first 5 yrs after diagnosis	Intermediate	Generally adverse	Generally adverse	Generally good

1.3.1 Molecular properties of breast cancer

Healthy estrogen receptor (ER) and progesterone receptor (PgR) expression not only influences menstrual cycle but also maturation of primary and secondary sex characteristics by modulating the expression of a broad range of genes (Klinge, 2001; Stepanova et al., 2006). Upon binding of the ligand, the receptors translocate into the nucleus, dimerize and bind to “hormone response elements” in the DNA to act as transcription factors. In breast cancer ER α (ESR1) and PgR expression are correlated to a better prognosis when compared to ER α -/PgR- cancers (Foulkes & Smith, 2010). Since PgR expression is modulated by ER α expression, ER α + /PgR- and ER α - /PgR+ cancers are rare (Bae et al., 2015). However, patients suffering from these cancer types might benefit from conventional treatment. Together, both receptors are one key factor for determining the treatment strategy. ER α is the only diagnostically relevant estrogen receptor today, although there are other ERs like ER β and mER (membrane bound ER), whose influence on the prognosis is not fully understood (Soltysik & Czekaj, 2013). Conventionally, ER α + cancers are treated with resection, chemotherapy, endocrine therapy, anti-HER2 therapy or a combination of those depending on tumor stage (Senkus et al., 2015). HER2 is a receptor tyrosine kinase that activates RAS- and AKT-signaling inducing proliferation and survival (Moasser, 2011). Overexpression and amplification are frequently found in breast cancer (Burstein, 2005) but also other entities (Merrick et al., 2006). Most HER2+ tumors can be treated with HER2 targeted treatments like Trastuzumab in combination with chemotherapy and resection (Senkus et al., 2015).

The absence of ER α , PgR and HER2 expression in TNBCs drastically limits the treatment options because antibodies, modulators or inhibitors against these proteins cannot be applied. The current standard therapy is resection and/or irradiation combined with adjuvant or neoadjuvant chemotherapy (Senkus et al., 2015). Patients with TNBC under neoadjuvant therapy tend to have better complete response rates when compared to patients with non-TNBC (22% vs. 11%, Liedtke et al., 2008) and have similar survival rates as non-TNBC patients. However, the majority of TNBC patients develop residual disease resulting in a significantly lower overall survival (88 % vs 68 %, $p = 0.001$, Liedtke et al., 2008). Additionally to the receptor status, TNBCs and BLCA show high intra- and intertumoral heterogeneity (Abramson & Mayer, 2014) and aneuploidy (table 2, Foulkes & Smith, 2010). A previous study showed that breast cancer stem cells (BCSC), defined by CD24^{low}/CD44⁺ status, are more abundant in TNBCs which could be associated with poor outcome (Idowu et al., 2012). The role of aneuploidy in cancer, especially in relation to BCSCs, is not fully understood. In the current understanding, BCSCs harboring CIN become aneuploid and accumulate genetic aberrations over time giving rise to progressively changing daughter cells that utilize CIN as evolutionary tool (Van Wely et al., 2012).

1.4 MCF10A 3D cell culture

Conventional monolayer (2D) cell culture models are well suited to quickly estimate the response to drugs or other treatments. However, they fail to account for more complex interactions, e.g. three-dimensional organization, secretion, cell-cell and cell-matrix interactions or nutrient, and oxygen gradients that are inherent for physiological tissue. A good example for the limited applicability of data collected with monolayer culture to physiological conditions is the epigenetic downregulation of FABP3 in monolayer cell culture. It results in a change of EGFR traffic, leading to altered drug sensitivity (Nevo et al., 2009). This relativizes conclusions drawn in studies on EGFR signaling as well as drug sensitivity using 2D cell culture.

Cells showing symptoms of CIN during division, namely lagging chromosomes, chromatin bridges, micronuclei, multipolar spindles or aneuploidy, usually divide with a reduced rate when compared to healthy cells (Santaguida & Amon, 2015). Adapting the model of CIN occurring early, these cells need to overcome growth arrest to be able to divide often

enough to utilize the genetic changes in the following steps. Such cells are overgrown by healthy cells in 2D culture. In physiological tissue, when they are not checked by the immune system, as well as in 3D cell culture these cells can form their own microenvironment. This makes 3D culture the optimal setup to test the influence of cells suffering from CIN on their environment.

1.4.1 MCF10 cell line series

One cell line that can form acini in 3D cell culture is MCF10A. The acini resemble physiological structures of the mammary gland by means of overall organization including size, polarization and lumen formation (figure 2). They show a basal-like expression profile (Neve et al., 2006), triple-negativity (Lawrence et al., 2015), and a CDKN2A deletion (Jonsson et al., 2007). However, they are non-tumorigenic and were isolated from a patient suffering from cystic fibrosis (Soule et al., 1990). Yoon and colleagues could show that MCF10A cell populations have a stable karyotype unlike TNBCs (Yoon et al., 2002). By overexpression of HRAS Dawson and colleagues created MCF10AT cells that form persistent pre-malign nodules in mice (Dawson et al., 1996), which occasionally develop into invasive adenocarcinoma. Some of these invasive adenomas show CIN, although MCF10AT cell are considered to be chromosomally stable (figure 2, Heppner & Wolman, 1999). RAS overexpression could be shown to induce replication stress that can result in CIN (Miron et al., 2015), which might explain the sudden appearance of CIN in MCF10AT xenografts. As stated before, the previously reported absence of CIN might be an artefact of monolayer culture. One of the invasive carcinomas formed by MCF10AT xenografts was dissociated and further cultured. One clone from this carcinoma showing increased rates of CIN gave rise to MCF10CA1 cells, that are highly tumorigenic (figure 2, Santner et al., 2001) correlating CIN to malignancy in MCF10AT cells.

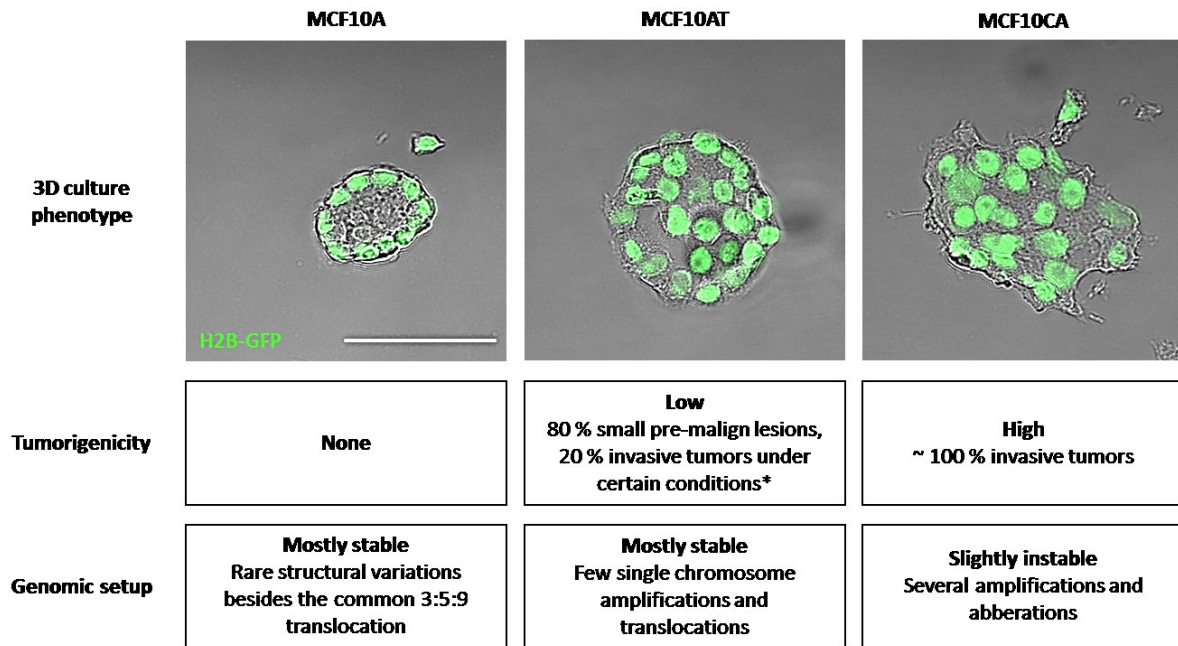


Figure 2: Cryo-slices of the MCF10 progression cell line and their increasingly malign properties. MCF10A cells show organized hollow spheroids when cultured on Matrigel, whereas MCF10AT spheroids lack a hollow lumen. MCF10CA spheroids develop into disorganized spheroids showing protrusions that indicate invasive behavior. The bars represent a length of 50 μm . * in the presence of EGF.

Originally, the overexpression of the mutated HRAS was introduced to increase proliferation (Dawson et al., 1996). However, mutated HRAS also effects apoptosis, the energy metabolism, angiogenesis (Pylayeva-gupta et al., 2011), and induces replication stress (Miron et al., 2015), which masks the origin of tumorigenicity (MCF10A -> MCF10AT) and what causes the increased malignancy (MCF10AT -> MCF10CA).

1.5 High-content siRNA screening

High-content RNAi screens utilizing a 3D cell culture setup enable testing hypotheses that require intercellular communication and cell-cell interactions. Britschgi and colleagues phenotypically screened primary breast cells cultured as 3D culture to find mechanistic hints how the cellular fate can be regulated using small-hairpin RNA (shRNA). By permanent downregulation of tumor suppressors, which also regulate cellular fate, they found that LATS1 and 2 play an important role in fate decision over estrogen signaling outside of the hippo pathway (Britschgi et al., 2017).

Unlike shRNA, small-interfering RNA (siRNA) mediated knockdowns confer a transient downregulation of genes by targeting certain RNAs for degradation before translation without additionally altering the genome (Elbashir et al., 2001). This is a good tool to acquire an optimal ratio of specificity, efficiency and versatility in a high-content screening setup (Semizarov et al., 2003) while avoiding prolonged downregulation that could be detrimental on cell fitness.

Evidence that a single knockdown can produce abnormal DCIS/LCIS-like (lobular carcinoma in situ) morphogenesis in MCF10A acini was given by Russ and colleagues (Russ et al., 2012). In their study the downregulation of LLGL1 and LLGL2 (alternatively named HUGL1 and HUGL2 in the paper, respectively) was shown to induce a reduced lumen formation as well as a change of polarity of MCF10A acini similar to MCF10AT spheroids (figure 3), though this was not induced by aneuploid cells.

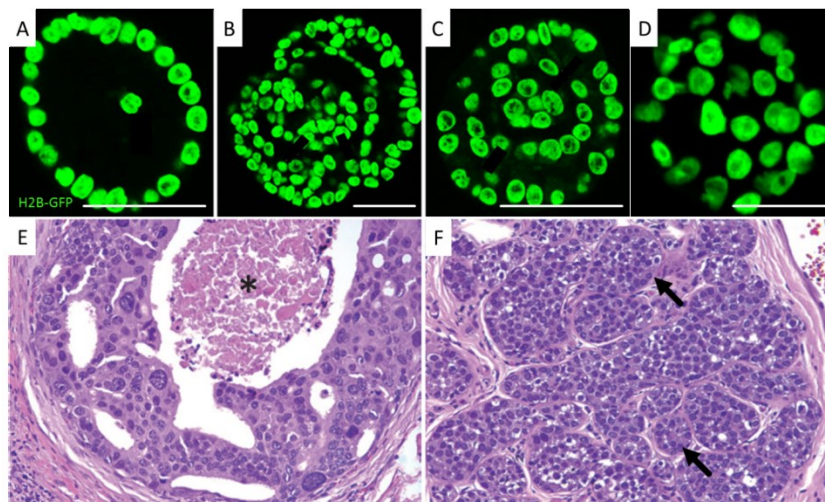


Figure 3: Comparison of a healthy acinus with perturbed acini and ex-vivo DCIS and LCIS sections. A) Hollow MCF10A acinus with a single cell inside the lumen (Russ et al., 2012, modified), B+C) shLLGL1 and shLLGL2 treated MCF10A acini lacking the hollow lumen (Russ et al., 2012, modified), D) MCF10AT filled spheroid resembling pre-malign DCIS/LCIS, E) Section of high nuclear grade DCIS with variably sized and shaped nuclei; the asterisk marks central necrosis; hematoxylin and eosin staining; 200x magnification (Morrow, Schnitt, & Norton, 2015, modified), F) Classic-type LCIS section (see arrows for examples of a small (bottom) and big (top) filled acinus); hematoxylin and eosin staining; 200x magnification (Morrow, Schnitt, & Norton, 2015, modified). The bars (A-D) represent a length of 50 μm .

In contrast to high-throughput screens, which primarily detect single parameters like cell death induced by drugs for a high number of conditions, high-content screens can detect a

higher number of parameters for a medium number of conditions. MCF10A acini show several physiologically relevant traits like a hollow lumen, polarization, growth arrest, and stage dependent rotation/cell migration (Debnath, Muthuswamy, & Brugge, 2003; Wang et al., 2012). By introducing aneuploidy any of these traits might be perturbed. Also, unexpected or new traits like escaping cells, protrusions, cell death or formation of new phenotypes might emerge, which makes multiparametric analysis necessary.

1.6 *Scope of the study*

Most BLCA show a high rate of aneuploid cells. In order to find the influence of aneuploidy on tumorigenesis in this context, we want to use non-tumorigenic, chromosomally stable MCF10A cells that share a basal-like expression profile as well as TNF-status with most BLCA. Cultured on Matrigel, they form structures that resemble healthy acini. We test if introducing aneuploidy to these cells results in changes of the morphology and expression reminiscent of cancer cells.

Specifically, we use a variety of CIN correlated cancer driver genes in a siRNA high-content screen. Referring to the argument that CIN is an early event, helping the cells to generate enough genetic variability to increase the population's plasticity, siRNA is especially suited here. Transiently inducing mitotic errors that lead to segregation defects simulates the early accumulation of genetic material by CIN without further manipulation. Subsequently, the aneuploid cells can form a niche and divide by interacting with the environment and other cells. This will show whether i) aneuploid cells can change the acinus morphogenesis and ii) differently induced aneuploidy leads to different morphological features that might resemble DCIS/LCIS. Next, we will follow aneuploid cells during altered acinus development using time-lapse imaging and identify how the cells interfere with normal morphogenesis. To correlate the abnormal development with a potentially malignant transformation, expression profiling and enrichment analysis will be conducted. The expression of relevant markers will be confirmed using immune-staining as well as QPCR. In addition, we will compare properties like tumor initiating potential (mammospheres and xenografts) and drug-sensitivity to infer the level of transformation. Finally, we will use chromatin conformation capturing (4C) to show how aneuploidy leads to altered chromatin interaction resulting in the deregulation of specific cancer related genes.

2 Material and Methods

2.1 Monolayer (2D) cell culture

MCF10A cells (Soule et al., 1990) were kindly provided by the Zev Gartner and colleagues (<http://www.gartnerlab.ucsf.edu/>) and transfected with a pBabe vector containing a construct of GFP labelled H2B. The MCF10A stock originates from ATCC (ATCC, CRL-10317). MCF10AT (Dawson et al., 1996) and MCF10CA (Santner et al., 2001) were acquired from the Karmanos cancer institute (KCI, <http://www.karmanos.org/home>).

DMEM/F12 medium (life technologies, ThermoFisher, Waltham, Massachusetts, USA, #11330-032) was acquired from GIBCO (ThermoFisher, Waltham, Massachusetts, USA). The standard medium was supplemented with EGF [20 ng/ml] (Sigma, St. Louis, Missouri, USA, E9644), Insulin [10 µg/ml] (Sigma, St. Louis, Missouri, USA, I1882), cholera toxin [100 ng/ml] (Sigma, St. Louis, Missouri, USA, C8052), hydrocortisone [500 µg/ml] (Sigma, St. Louis, Missouri, USA, H0888) and 5 % horse serum (life technologies, ThermoFisher, Waltham, Massachusetts, USA, #1605-122) to get the full medium.

MCF10A as well as MCF10AT and MCF10CA cells were cultured in polystyrene culture flasks (Greiner Bio-One GmbH, Kremsmünster, Oberösterreich, #690175) as monolayer. As soon as they reached a density of 80-90 % the cells were washed with phosphate buffered saline (PBS). Afterwards they were detached using 0.05 % trypsin/EDTA (T/E) for 15 minutes at 37 °C and agitation. Subsequently the cell suspension was either diluted in full medium at least 1/12 to obtain a T/E concentration of below 0.005% and cell density between 10.000 and 50.000 cells/ml depending on further culturing needs or first centrifuged at approximately 300 g for 2-3 minutes and resuspended in an appropriate volume of full medium and pipetted into a new culture flask.

2.2 3D cell culture

Basement membrane extract Matrigel® was acquired from Corning (New York, USA, #356231). After preconditioning the well of a 24-well plate (Greiner Bio-One GmbH, Kremsmünster, Oberösterreich, #662160) with 100 µl full medium and subsequent removal of the medium 70 µl Matrigel was dispensed into each well and let solidify for 30 minutes at

room temperature (RT). After diluting MCF10A cells in suspension (for procedure see chapter monolayer (2D) cell culture) to a concentration of 50.000 cells/ml they were enriched with 5 % Matrigel and pipetted into wells that were previously coated with Matrigel. The medium was exchanged every 3 days with full medium supplemented with 5 % Matrigel.

Previous studies showed that some contents of Matrigel vary strongly between different LOTs (Hughes et al., 2010). To ensure that the results are no batch dependent all results have been confirmed using three different LOTs of Matrigel (LOT #: 2323349, 4160003, 5047311).

2.3 RNAi experiments

The silencer select siRNAs were acquired from Ambion (ThermoFisher, Waltham, Massachusetts, USA). The full list of siRNAs and their IDs can be found in supplementary table 2. From the diluted cell culture suspension 20.000 cells were dispensed into each well of a 24-well plate and left over night. To prepare transfection complexes 1 μ l Lipofectamine 2000 (ThermoFisher, Waltham, Massachusetts, USA, #11668019) was mixed with 50 μ l nuclease free water and incubated for 5 minutes. In parallel 0.1 μ l siRNA with a concentration of 30 μ M was mixed with 50 μ l nuclease free water and incubated for 5 minutes. The Lipofectamin mix and the siRNA mix were added together and incubated for 15 minutes to get the transfection mix. The medium was removed from the cell containing wells and replaced with the transfection mix plus 300 μ l transfection medium (culture medium lacking serum) and incubated overnight.

2.4 Quantitative PCR

The total RNA was isolated using the QIAGEN RNeasy Kit (Qiagen, Venlo, Netherlands, #74106) after supplier recommendations. In short, the cells were lysed using RLT buffer, applied to silica membrane columns and centrifuged. The bound RNA was washed and then eluted using 50 μ l nuclease free water. The quality and concentration was confirmed using the Nanodrop 2000 and samples with a concentration of below 50 ng / μ l and a 260/280 absorption ratio of below 1.8 were discarded. The total RNA was treated with the ThermoFisher (previously Fermentas) DNase digestion kit (ThermoFisher, Waltham, Massachusetts, USA, #K1622) after supplier recommendations to get rid of DNA. As

proposed 500 mg RNA and DNase I were used for the DNA-digestion which was stopped using 1 µl 50 mM EDTA. In the next step the RNA was used for cDNA synthesis with the RevertAid First Strand cDNA synthesis kit (ThermoFisher, Waltham, Massachusetts, USA, #K1622) after supplier recommendations using poly-A-primers. The cDNA was diluted with nuclease-free water to an endconcentration of 1.4 ng/µl. 4 µl of diluted cDNA was then mixed with either 5 µl SYBR green MM (ThermoFisher, Waltham, Massachusetts, USA, #AB-1162) 0.2 µl 10µM forward and reverse primer and 0.6 µl nuclease-free water or 10 µl TaqMan Universal Master Mix II, with UNG (ThermoFisher, Waltham, Massachusetts, USA, #4440044) and 1 µl TaqMan Gene Expression Assay (ThermoFisher, Waltham, Massachusetts, USA, #4331182) with <100 ng cDNA with 5µl water in the designated well of a 96-well Q-PCR plate. The plate was measured and analyzed using the Fisher StepOnePlus RealTime PCR machine (ThermoFisher, Waltham, Massachusetts, USA, #4376599) and the Step One Software version 2.0. The experiment was set to quantitation-comparative Ct ($\Delta\Delta Ct$), SybrGreen or 6-FAM dye (for TaqMan assays) reagents with 95 °C 10 minutes denaturation, 40 cycles of 15 seconds at 95 °C, 30 seconds at 60 °C, melt curve: 95 °C for 15 seconds, 60 °C for 1 minute and a subsequent heating ramp of 0.3 °C/minute until 95 °C was reached.

The primers with their respective sequence or the Expression assay used for Q-PCR are: INCENP forward: CAAGAAGACTGCCGAAGAGC, INCENP reverse: TCAGAACCAACTTTCTGGGG, RPLP0 forward: GCGACCTGGAAGTCCAACCT, RPLP0 reverse: CCATCAGCACCCACAGCCTTC, additionally we used TaqMan® Gene Expression Assays for CD24, CD44, EPCAM, ALDH1A3, CCND1, TOP2A and ESPL1.

2.4 Immunofluorescence for monolayer and 3D cell culture

The full list of antibodies used for immunofluorescence staining can be found in supplementary table 3. The medium was aspirated and the cells were fixed using 2 % formaldehyde (Sigma, St. Louis, Missouri, USA, F8775) in PBS for 12 minutes. The fixative was removed and the cells were washed 3 times with PBS for 1-5 minutes at 4 °C (or RT). After fixation the cells were treated with pre-chilled 0.5 % triton x-100 in PBS-glycin for <10 minutes at 4 °C (or RT) and subsequently washed 3x for 1-5 minutes with 1x washing buffer (10x Washing buffer: 38 g NaCl, 9.38 g Na₂HPO₄, 2.07 g NaH₂PO₄, 2.5 g NaN₃, 5 g BSA, 10 ml Triton-X 100, 2.05 ml Tween-20, filled up to 500 ml with water and titrated to a pH of

7.4). In the next step the cells were blocked with 10 % serum (1:1 goat serum (Sigma, St. Louis, Missouri, USA, G9023) and donkey serum (Jackson ImmunoResearch, West Grove, PA, USA, # 017-000-121) in 1x washing buffer (=blocking buffer). The primary antibody (see supplementary table 3) is diluted in blocking buffer and incubated on the cells overnight (at least for 1 hour) at 4 °C (or RT). The cells were then washed 3x with 1x washing buffer for 10 minutes each at RT. After the washing the secondary antibody (see supplementary table 3) diluted in blocking buffer was incubated on the cells for 1 hour and the cells afterwards washed 2x with washing buffer. Confocal imaging was conducted using the Zeiss LSM 780 under the control of AutofocusScreen macro (<http://www.ellenberg.embl.de/apps/AFS/>, 24.02.2016).

2.4.1 SiR dyes (Lukinavičius et al., 2014)

To stain the Golgi and lysosomes we used the experimental SiR-pepstatin dye kindly provided by Kai Johnsson and Luc Reymond (Spirochrome, Stein am Rhein, Switzerland) in a concentration of ~0.5 µM in full medium which was added to the cells 4 h before imaging and was changed every 48 h. Pepstatin is an inhibitor of aspartyl proteases but shows a low inhibitory effect when coupled to SiR to ensure proper cell functioning while reliably delivering the probe to the intended organelle.

2.5 Drug tests

To test the functionality of the anti-H2AFX antibody (see supplementary table 3) upon drug treatment cells cultured in 2D and 3D were treated with increasing concentrations (0, 1, 3, 10, 30, 100 µM) of Etoposide in full medium for the entire time until staining.

In order to test drug sensitivity against paclitaxel we treated 3D and 2D cultured cells with increasing concentrations (0, 1, 3, 10, 30, 100 nM) for 48 h. Subsequently the cells were washed and imaged 24 h later.

2.6 Microarray expression profiling

The total RNA was isolated using the QUIAGEN RNeasy Kit (for procedure see chapter “Quantitative PCR”). The RNA was then labeled and hybridized in technical and biological duplicates with Illumina HumanHT-12 v4 Expression BeadChips (Illumina, San Diego, USA,

BD-901-1001) and quantile normalized against the non-treated MCF10A negative controls by the DKFZ Genomics and Proteomics core facility.

The top 100 deregulated (either up, down or up-and-down) as well as all upregulated (> 1.5 fold, $p < 0.05$) genes were analyzed using the gene set enrichment analysis tool (<http://software.broadinstitute.org/gsea/index.jsp>) of the Broad institute. Set enrichments were conducted using all known gene sets and selected based on the rank of the lowest p-value (hypergeometric distribution) and false-discovery-rate (q-value).

2.7 Multiplex fluorescence in situ hybridization (M-FISH)

Cells with a density of 40-60 % were treated with 40 ng/ml colcemid (Sigma, D7385) for 12-20 hours at 37 °C to enrich the culture for cells in the metaphase. In the next step, the cells were detached by adding T/E to the cells for 10 minutes at 37 °C and transferred into a reaction tube. The cells were then centrifuged at 300 g for 10 minutes at RT. The supernatant was aspirated until there was ~500 µl left and the cells resuspended by flicking the tube. Subsequently, the cells were treated with 5 ml hypotonic solution (0.55 % KCl in deionized water) which was applied dropwise for 10 minutes at RT. Afterwards, the cells were centrifuged at 300 g for 10 minutes at RT and the supernatant again aspirated until there was ~500 µl left so that the cells could be resuspended by flicking the tube. The cells in the suspension were then fixed by dropwise adding 5 ml of Carnoy's Fixative (75 % Methanol, 25 % acetic acid) and incubated for 10 minutes at RT. The fixation was repeated twice. The fixed solutions were then further processed and analyzed by Anna Jauch and Brigitte Schoell (Human genetics, DKFZ) and Ilse Chudoba from Metasystems (statistical analysis).

For statistical quantification of aneuploid nuclei 200 – 2000 metaphases were counted using the Metafer 4 v3.11.7 software (Metasystems, Altlußheim, Germany). The metaphases were classified using the "mFISH classifier" of the software and a sensitivity threshold of 20. The metaphases were then manually grouped into normal and aneuploid appearance based on the number of visible chromosomes.

M-FISH was performed as described by Geigl and colleagues (Geigl et al., 2006). Briefly, Landry and colleagues summarized the process as "*seven pools of flow-sorted human whole*

chromosome painting probes were amplified and directly labeled using seven different fluorochromes (DEAC, FITC, Cy3, Cy3.5, Cy5, Cy5.5, and Cy7) using degenerative oligonucleotide primed PCR (DOP-PCR). Metaphase chromosomes immobilized on glass slides were denatured in 70 % formamide/2x saline-sodium citrate (SSC) pH 7.0 at 72 °C for 2 minutes followed by dehydration in a degraded ethanol series. Hybridization mixture containing combinatorically labeled painting probes, an excess of unlabeled cot1 DNA, 50% formamide, 2xSSC, and 15% dextran sulfate were denatured for 7 minutes at 75 °C, pre-annealed at 37 °C for 20 minutes and hybridized at 37 °C to the denatured metaphase preparations. After 48 hours, the slides were washed in 2x SSC at room temperature for 3x 5 minutes followed by two washes in 0.2x SSC/0.2 % Tween-20 at 56 °C for 7 minutes, each. Metaphase spreads were counterstained with 4,6-diamidino-2-phenylindole (DAPI) and covered with antifade solution. Metaphase spreads were recorded using a DM RXA epifluorescence microscope (Leica Microsystems, Bensheim, Germany) equipped with a Sensys CCD camera (Photometrics, Tucson, Arizona, USA). Camera and microscope were controlled by the Leica Q-FISH software and images were processed on the basis of the Leica MCK software and presented as multicolor karyograms (Leica Microsystems Imaging solutions, Cambridge, United Kingdom)” (Landry et al., 2013).

2.8 Mammosphere assay (modified after Shaw et al., 2012)

To test the renewal capacity 1000 cells/cm² were seeded in Greiner ULA 96 well plates (Greiner Bio-One, Kremsmünster, Oberösterreich, # 655970) with tumorsphere medium (Standard medium without serum supplemented with 2% B27, ThermoFisher, Waltham, Massachusetts, USA, # 12587010). The cells were cultured for 5 days and the mammospheres counted. Subsequently, the first generation MFE (mammosphere formation efficiency) was calculated after $100 \times (\# \text{ seeded cell} / \# \text{ mammospheres} > 50\mu\text{m})$.

The mammospheres were harvested by removing the medium and added to a collection tube. After another rinsing of the well the collected material was centrifuged at 150 g for 3 minutes and the pelleted spheres were resuspended in accutase and incubated for 10 minutes at RT to get a single cell suspension. All cells were reseeded into the same amount of wells. Finally, the cells were cultured for 5 days and the mammospheres counted to calculate the second generation MFE.

2.9 Xenografts

MCF10A, MCF10AT and MCF10CA cells were treated with either scrambled, ESPL1, TOP2A or no siRNA and cultured until there were at least 1.5×10^6 cells. The cells were detached using 0.05 % T/E and dissolved in a mixture of full medium and Matrigel (3:1) to a density of 2×10^6 cells / ml. After anesthesia by gaseous narcosis (Isoflurane 2.5 % with Oxygen mixed by air flow) 100 ml ($=2 \times 10^5$ cells) of the mixture was injected into the mammary fat pad of 2-months old NSG-mice (NOD scid gamma mice, 6 mice per condition, all female). Additionally, one EGF-pellet (Innovative America, #NE-121) with a dose of 0.18 mg 17β -estradiol / pellet was transplanted. The xenografts were then grown for 4 weeks. The mice were treated according to the German authorization numbers G-240/11 and G-82/16. For reasons of animal safety mice with abnormally big tumors were killed before this deadline to minimize suffering. If no tumors were visible after 4 weeks, the xenografts were left untouched for two more weeks and then extracted. The brain, lungs and liver were inspected for signs of metastases. The extracted tumors were weighed and partially formalin-fixed (10 %, buffered). The mouse experiments and immunohistochemistry were conducted by Corinna Klein and Martin Sprick from the Hi-Stem, DKFZ.

2.10 Chromatin conformation capture (4C)

To assess the chromatin interactions upon induction of aneuploidy 4C sample preparation was performed as described by Van de Werken and colleagues (Van de Werken et al., 2012). In short, 10 million cells per viewpoint were isolated from Matrigel (80 % Matrigel, 20 % PBS) by incubating them with dispase for 10 min at 37 °C. After dissociation of the cells with 0.05 % T/E for 5 min at 37 °C the cells were stepwise washed with 5 ml Versene, trypsin and medium. Subsequently, they were counted, diluted to 10^7 cells and fixed using 4 % formaldehyde for 10 min at RT. The cells were washed with 1 M glycine, centrifuged for 8 min at 300 g, resuspended in 1 ml lysis buffer (50 mM TRIS, 150 mM NaCl, 5 mM EDTA, 0.5 % NP-40, 1 % TX-100 in H₂O) and incubated for 5 min at RT, for 5 min in a 65 °C water bath and then cooled on ice. In the next step, the cells were washed with PBS, transferred into 1.5 tubes and either frozen at -80 °C or further processed for the digestion using 10 % SDS at 37 °C for 1 hour. After adding 75 μ l 20 % TX-100 and incubation for 1 hour at 37 °C a 5 μ l aliquot as pre-digestion control was taken. In a first three-step digestion with each 200 U DpnII or NlaIII the samples were incubated for 4 hours at 37 °C, O/N at 37 °C and

again 4 hours at 37 °C. A post-digestion control was taken and all samples incubated at 65 °C for 20 minutes to inactivate the enzyme. The samples were transferred to a 50 ml tube and incubated at 16 °C in the water bath O/N with a ligation mix (700 µl 10x ligation buffer, 7 ml ddwater, and 10 µl T4 Ligase (Roche 5 U/µl)). The ligation control was taken and tested on a 0.6 % agarose gel. De-crosslinking was done by adding 15 µl Proteinase K (20 mg/ml) and incubation at 65 °C in the water bath O/N. The next morning 30 µl RNase A (10 mg/ml) was added and the mixture incubated at 37 °C for 45 min. After adding 7 ml phenol-chloroform and vigorously mixing the sample was centrifuged for 15 minutes at RT at 3000 g. The water phase was aspirated and transferred to a 50 ml tube and mixed with 7 ml ddwater, 7 µl glycogen, 1.5 ml 2 M NaAC pH 5.6, and 35 ml 100 % ethanol and frozen at -80 °C. In the next step, the samples were centrifuged for 30 minutes at 4 °C at 9000 g and the pellet resuspended in 10 ml cold 70 % ethanol. After a centrifugation for 15 min at 4 °C at 3000 g the supernatant was discarded and the pellet dried at RT for a few minutes. Subsequently the pellet was dissolved in 1.5 µl 10 mM Tris pH 7.5 at 37 °C and then either frozen at -20 °C or digested a second time with 200 U Csp6I or Bfal and incubated O/N at 37 °C. The digestion efficiency was tested on a 0.6 % agarose gel. The enzyme was then inactivated at 65 °C for 25 minutes and the sample transferred into a new 50 ml tube and mixed with 1.4 ml 10x ligation buffer, 20 µl ligase (Roche, Basel, Switzerland, 5 U/µl) and filled up with ddwater to 14 ml for a second ligation O/N at 16 °C. The next morning 1.4 ml 2 M NaAC pH 5.6, 14 µl glycogen and 35 ml 100 % ethanol were added and frozen at -80 °C. In the next step, the samples were centrifuged for 45 minutes at 4°C at 3000 g and the supernatant discarded. After resuspending the samples in 70 % ethanol and centrifugation for 15 at 20 °C at 3000 g the pellet was dried and dissolved in 150 µl 10 mM Tris pH 7.5 at 37 °C. The samples were then purified using the QIAquick PCR purification kit (Qiagen, Venlo, Netherlands, #28106) after supplier recommendations with 3 columns per sample and 50 µl 10 mM Tris pH 7.5 as elution buffer. The concentration was determined using a Nanodrop 2000 (Thermo Fisher, Waltham, Massachusetts, USA, ND-2000). The eluates were validated with a test PCR checking for a smear close to the expected size of fragments on a 1.5 % agarose gel. Subsequently the sequencing PCR was conducted (2.5 µl 10x PCR buffer, 0.5 µl 10 mM dNTPs, 4.5 µl 1 µg/µl forward primer 1:7 diluted, 3 µl 1 µg/µl reverse primer 1:7 diluted, 0.35 µl polymerase ELT, 5 µl 100 ng 4C DNA template, 9.15 µl ddwater) including no-template

control (2 min 94 °C, 35x- 15 s 94 °C, 1 min 55 °C, 3 min 68 °C- then 7 min 68 °C and cooling at 4 °C). 5 µl of the products were tested on a 1.5 % agarose gel and the rest was purified with the Roche-PCR-purification kit (Roche, #11732676001) and 10 µl checked on a 1.5 % agarose gel for dimers >100 bp. If present the annealing temperature or primers would needed to be modified. In the next step 16 PCR per viewpoint were prepared in the same way as before with double the amount. The products were purified using the Roche PCR purification kit and then again purified using the QIAquick PCR purification kit and the concentration was tested with the Nanodrop2000. The single end sequencing was conducted in the DKFZ sequencing core facility with < 20 pooled probes (10nmol combined) per lane on a HiSeq2000. The experiments were conducted by Sabine Aschenbrenner, technical assistant of the Eilslabs.

2.11 Data analysis

All images were automatically analyzed using custom KNIME version 3.2.1 workflows (Berthold et al., 2007). The list of workflows and their location can be found in supplementary table 6.

2.11.1 High-content screen immunostainings

Nine images per well (per condition there were three replicates) from the HCS immunostainings were analyzed using the “HCS 3D analyzer” KNIME workflow. Images were excluded if they were empty/black or out of focus. In detail the images were median filtered and segmented using a local threshold (otsu). After filtering objects on the edges of the images and a “fill holes” step, the objects were split by the ImageJ watershed. The segments were processed into labels using a connected component analysis (CCA) and then used to create three label-regions for each object. These regions were used to read out parameters in i) the center ii) the whole object iii) and the region outside the center region but within the whole region (donut shaped). The H2B-GFP as well as the GM130 intensity signal that were read out in the donut and the center region and used to calculate the hollowness (“H2B-GFP intensity donut”/ “H2B-GFP intensity center”) and the polarization (“GM130 intensity center”/ “GM130 intensity donut”). The other parameters (number and size of objects, circularity, Haralick and Tamura textures, H2B-GFP intensity, H2AFX intensity) were

measured in the whole region and the mean, standard deviation and count were calculated as mean per spheroid over all replicates and their 27 images.

The QPCR validation of the knockdowns was conducted by calculating the portion of cells showing a polylobed nucleus. The images were analyzed with the “image based knockdown efficiency” KNIME workflow (for the detailed list including input, output and server/GitHub location can be found in supplementary table 6). The workflow is based on global thresholding to segment the cells. The segmented cells were then classified by random forest classification that was trained using 20 % of the images as training set and subsequently cross validated and counted.

The nine parameters of each knockdown condition of the HCS were normalized against the respective negative control creating a percentage deviation to the negative control after the formula

$$\% \textit{ difference} = 100 - (\textit{Condition } y_x * 100) / \textit{Negative } x$$

X = Number of the plate the set of experiments were conducted on

Y = Position of the condition on the plate (6 conditions per plate)

The parameters were then used to perform a principal component analysis (PCA) by preserving 100 % of the information and calculating nine eigenvectors for each condition by spectral decomposition using the “PCA cluster” KNIME workflow. These were then used in a k-means clustering approach to identify groups that are furthest separated from the other groups. The “k” was set based on an optimal separation of negative control and scrambled control associated clusters from the other clusters. The three clusters furthest apart were chosen for further manual selection of suitable candidates. Dead cells/spheroids show a high difference to the negative control or other phenotypes. Such candidates were filtered from the three furthest clusters as well as candidates that were visually not/hard to be distinguished from the negative controls.

To automatically quantify the portion of abnormal spheroids in each condition the spheroids were classified using the “class identifier” KNIME workflow. After a global

thresholding (after Otsu's method, Otsu, 1979), watershed and CCA the cells were classified by a random forest classifier that was trained using 20 % of all images to be classified.

2.11.3 *Time-lapse analysis*

The time-lapse images were concatenated and automatically cropped with the "TL concatenator" KNIME workflow using global thresholding segmentation and a 3D CCA that was then used as a mask for the segment cropper. The resulting single spheroid time-lapses were then further analyzed with the "TopTL analyzer" KNIME workflow. After a Gaussian convolution, global thresholding (after Otsu's method), fill hole-step, contour smoothing, watershed and CCA similar to the HCS image analysis three regions (center, whole object, donut) were created using the initial segments. These regions were used to calculate the size, sphericity, hollowness and polarization as described before.

Images processed by the "TL concatenator" (see previous paragraph) were also used for a pseudo-single cell tracking inside the spheroids using the "MIP single cell tracking" KNIME workflow. Here, the single spheroid time-lapse images were z-projected using standard deviation approach. After a median filtering, local thresholding (Bernsen), contour smoothing, watershed and CCA the single cells were tracked in the segmented area using ImageJ's "Trackmate tracker". In this case, the imperfect tracking and resulting constant splitting of tracks after a few time points could be used to calculate an average speed of all nuclei during short time intervals.

2.11.4 *Breast cancer stem cell marker analysis*

The intensities were measured with the "BCSC analyzer" KNIME workflow that first z-projects the single spheroid stack images followed by a Gaussian convolution and global thresholding resulting in mean intensities per single spheroid.

2.11.5 *Drug sensitivity testing*

The images were processed using the "drug test analysis" KNIME workflow. After a Gaussian convolution, global thresholding (after Otsu's method, Otsu, 1979), watershed and CCA the size and the number of objects were calculated per well.

2.11.6 *Mammosphere formation*

The stitched whole-well images were first z-projected (maximum intensity), convolved, global thresholded (manual) and then processed by CCA using the “mammosphere_analysis” KNIME workflow. Objects that were bigger than four cells (85 pixels) were selected and counted.

2.11.7 4C sequencing analysis

The 4C sequencing data was analyzed by Qi Wang using custom in-house workflows developed by the AG Herrmann. The demultiplexing and trimming of primer sequences, a Python script from “FourCSeq” R package was used. The remaining sequences were aligned to the human reference genome GRCh37/hg19 using Bowtie 2 alignment program with default parameters and no clipping allowed. We performed an in silico cutting of the reference genome using FourCSeq. The definition of restriction fragments are illustrated in the figure 4. Only reads that start directly at a restriction enzyme cutting site and show the right orientation of the read at the fragment end following figure b have been assigned to each restriction fragment in each sample.

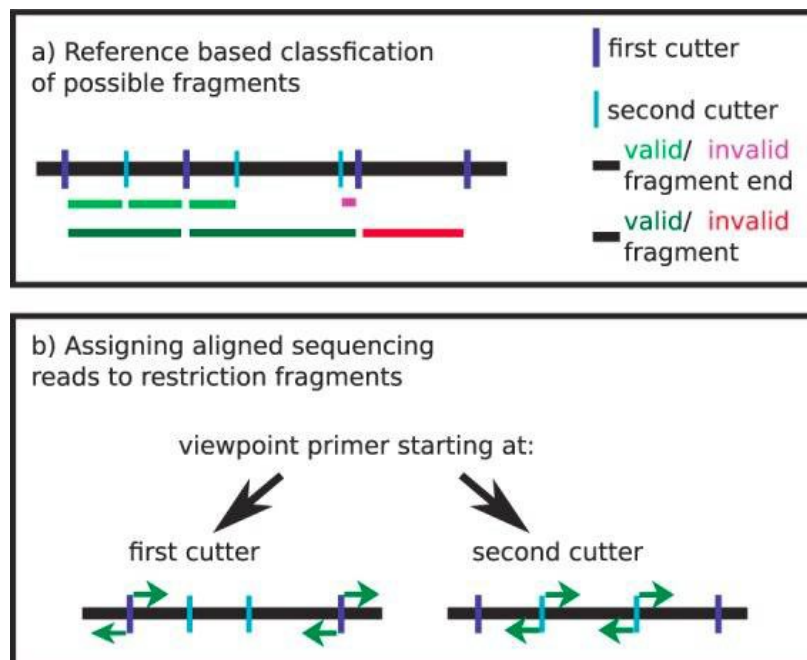


Figure 4: Restriction fragment definition and orientation (Klein et al., 2015).

Different samples were normalized according to their total library sizes. We transformed the data to unique coverage (more than one reads per fragment end is set to one) to avoid possible PCR artifacts which may originate from a single ligation event. A window size of 100

was used in counting the number of ligated sites to define inter-chromosomal interactions. To call out statistically significant interacting regions, z-scores were calculated based on the relative unique coverage in the background window of 3000. We randomly permuted the dataset 100 times and applied a false discovery rate (FDR) threshold of 0.01 to the z-scores. Windows with a z-score above the FDR are scored as “interacting region”. For each experiment, the percentage of reads on the viewpoint chromosome over all reads was calculated, where only experiments with more than 50 % intra-chromosomal reads are considered suitable for trans-interactions analysis. A window size of 500 unique fragment ends was used. Windows that exceed the threshold are scored as trans-interactions.

2.12 *Statistical Methods*

The data quantified using KNIME was further processed by integrated custom R scripts or Graphpad and statistically quantified using linear regression, linear/non-linear fitting or ANOVA where appropriate. The statistical level of significance was considered to be below a value of 0.05.

2.12.2 *List similarity*

To determine the overlapping elements of the list of genes from siESPL1 and siTOP2A expression profiles the “list comparison” tool from the Whitehead Institute, Cambridge was used (<http://jura.wi.mit.edu/bioc/tools/compare.php>).

3 Results

3.1 High-content screen

In a first literature research we focused on publications investigating i) CIN gene signatures that better predict the outcome of patients than morphological tumor classifications (for references, please see supplementary table 5) ii) tumor drivers that can function as tumor suppressors (TSGs) and contribute to CIN (for references please see supplementary table 5) iii) well known and/or highly mutated tumor drivers including the top 10 most mutated genes based on the COSMIC database (<http://cancer.sanger.ac.uk/cosmic>, Forbes et al., 2017) resulting in a list of 236 candidates (Abstract figure: Literature).

To further refine this list, we selected 82 candidates that are associated to seven “GO-process” groups (supplementary table 1, figure 5 and abstract figure: classification). The selection of the candidates was based on visible morphological phenotypes during knockdown (Genome RNAi, DKFZ and Cellular Phenotype Database, EMBL) as predictor for aneuploidy as shown by Sadaie and colleagues (Sadaie et al., 2015). The list is comprised of a spectrum from well-known to poorly-studied genes defined by the number of cited articles on Pubmed (<https://www.ncbi.nlm.nih.gov/pubmed/>). In total, the genes include ~31 % high (> 200 cited articles), ~36 % medium (< 200 and > 50 cited articles) and ~33 % low (< 50 cited articles) cited candidates.

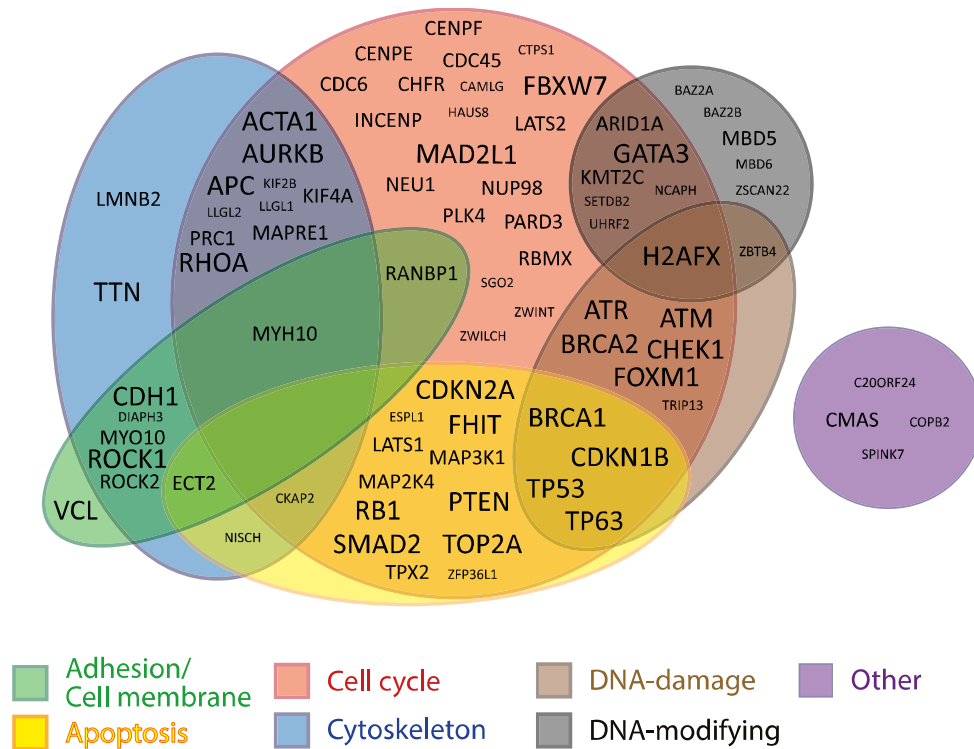


Figure 5: Screening candidates categorized into 7 “GO-process” defined groups. The character size of the names correlates with the referenced number of publications on PubMed (12 points: >200 publications, 10 points: between 200 and 50 publications, 8 points: < 50 publications; August 2016).

To be able to capture as many changes as possible we imaged the developing acini after 3, 6 and 9 days. At those time points, we measured nine parameters (number of acini, size, roundness, hollowness, polarization, Tamura and Haralick textures, H2AFX intensity, and H2B intensity). Using these three time points, there was a strong difference between day 3 and 6 when comparing most of the parameters. After day 6, the acini did only show slight changes over time (figure 6). The morphological phenotypes visible from day 6 were stable until day 21 (supplementary figure 1). Further culturing of acini on Matrigel was impractical due to gradual degradation and depletion of the Matrigel layer and subsequent monolayer formation. This indicates day 6 as the optimal time point for further analysis of the acini in this setup.

To test whether the additional knockdown of TP53 leads to more severe phenotypes, we conducted double-knockdowns (dKD) with TP53 and the respective candidate (knockdown using only one target: KD). In general, there was no detectable effect of the double knockdown (figure 6).

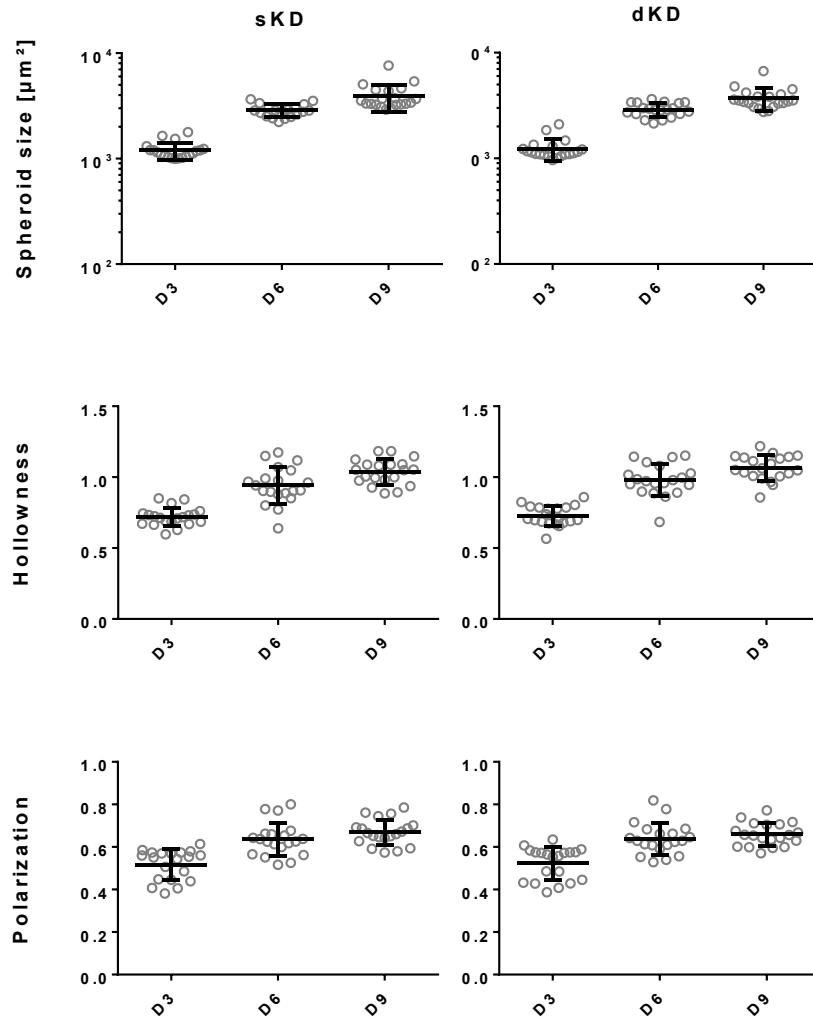


Figure 6: Comparison of the parameters size, hollowness, and polarization of a random 25 % sample (n=20) of the candidates between single and double knockdown at day 3, 6 and 9. Each data point represents a single candidate measured in triplicates. The black bars represents the mean of all data points and the error bars the standard deviation.

The knockdown of all candidates was compared with a 2D culture control to see if the 2D phenotypes were similar to results reported in the literature and RNAi morphology Databases (EMBL cellular phenotype database: “<http://www.ebi.ac.uk/fg/sym/search/result>”, DKFZ GenomeRNAi database: “<http://www.genomernai.org/>”). Differences were found in 16 cases, 21 were inconclusive and 45 were in conjunction with the literature (supplementary table 1).

Analysis of all candidates using all 9 parameters in a principal component analysis (PCA, for eigenvector distribution see supplementary figure 2) and subsequent k-means clustering

resulted in 8 groups of candidates (figure 7). The three groups with the highest distance (distance > 70) apart from the null vector in figure 7 highlighted as red squares (BRCA1, H2AFX, MBD5, MYH10, RB1), green diamonds (scr28, DIAPH3, ESPL1, LATS2, NEU1, ZWILCH, ZWINT) and blue triangles (ATR, CHFR, CMAS, COPB2, SETDB2, TOP2A) were selected for visual validation (Abstract figure: HCS). The other groups (circle that is crossed, white, boldly bordered, dotted, or black) were omitted because they were closer to the null vector or the negative control associated group (k-means cluster 1, figure 7).

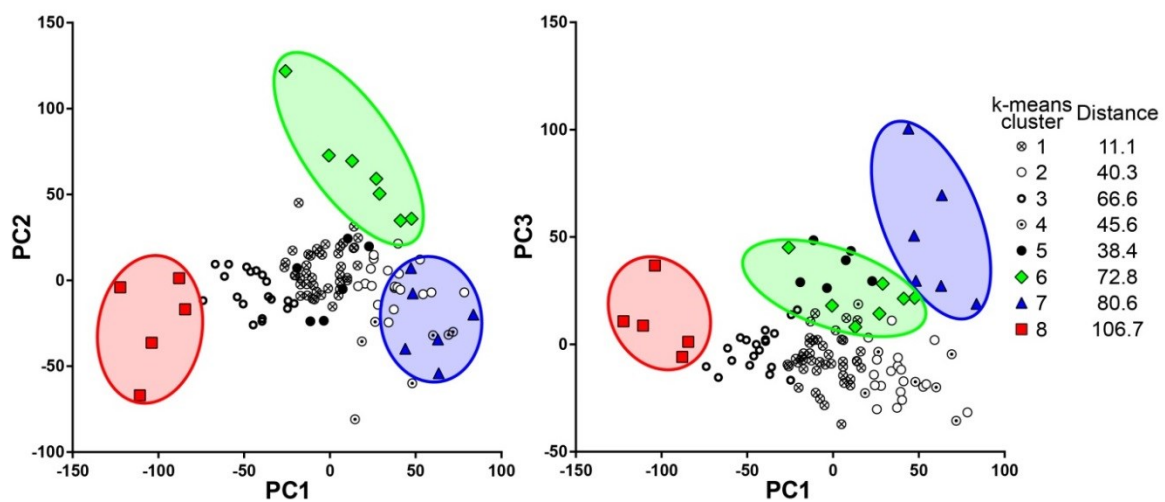


Figure 7: Selection of candidates using PCA and k-means clustering plotted by PC1 vs. PC2 and PC1 vs. PC3.

A further visual validation step resulted in five candidates (DIAPH3, ESPL1, H2AFX, SETDB2, TOP2A, Abstract figure: Morphology filtering). These candidates were selected from the candidates inside the three colored groups accounting for a sufficient amount of viable spheroids as well as a clearly abnormal appearance (figure 8, table 3) based on the parameters quantified during the HCS.

Table 3: Selected candidates with their respective visual selection criteria and cluster. Dead: none or a very low number of viable spheroids present (additionally to a reduction of the number of acini of more than 50 %), weak: the measured phenotype was visually difficult to differentiate from the negative control, not visible: the phenotype was not visible, abnormal: the spheroids showed altered acinus architecture (multiacinar structure, tubular/invasive structures, disorganized spheroids, filled spheroids, altered Golgi), asterisk: scrambled control that was visually not distinguishable from the negative control, the colors of the “Cluster” column correspond to the colors of the clusters in figure 7.

Gene	Visual selection	Cluster
scr28	not visible*	6
ATR	weak	7
BRCA1	weak	8
CHFR	weak	7
CMAS	dead	7
COPB2	dead	7
DIAPH3	abnormal	6
ESPL1	abnormal	6
H2AFX	abnormal	8
LATS2	weak	6
MBD5	weak	8
MYH10	not visible	8
NEU1	weak	6
RB1	weak	8
SETDB2	abnormal	7
TOP2A	abnormal	7
ZWILCH	dead	6
ZWINT	dead	6

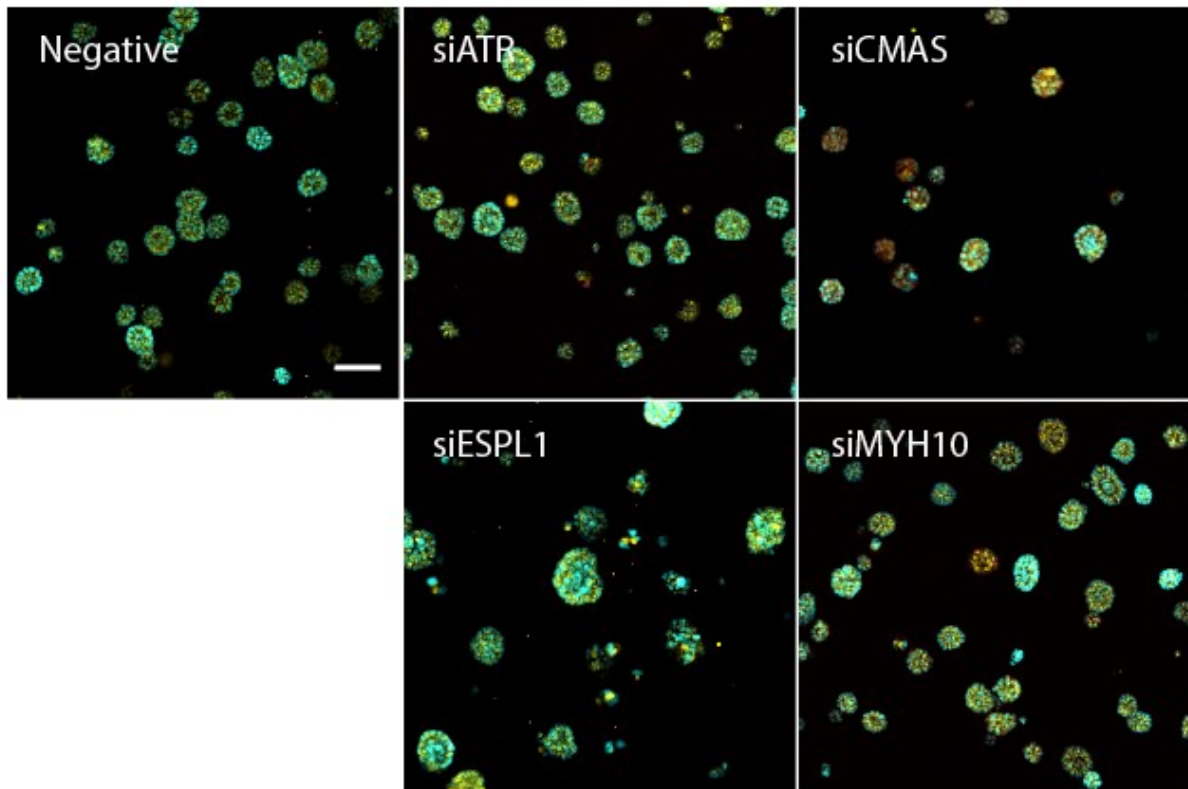


Figure 8: Overview of examples for morphology filtering using visual phenotypes. siATR represents the “weak” phenotype with only a few single cells lying around the otherwise normal acini, siCMAS shows an acinus reduction of > 50 % with a slight increase of H2AFX signal representing the “dead” phenotype, siESPL1 shows a high number of acini with abnormal nucleus and Golgi architecture which defines the “abnormal” phenotype, siMYH10 shows neither acinus architecture or acinus number alteration making them visually indistinguishable from the negative control. Nuclei: cyan, Golgi (GM130): yellow, DNA damage (H2AFX): red, the bar (top left image) represents a length of 100 μm .

Info box 1: Background of the selected 5 candidates

DIAPH3 is a member of the formin gene family which is involved in actin remodeling (Palazzo et al., 2001) and influence amoeboid tumor cell mechanics (Morley et al., 2015). DIAPH3 is overexpressed in approximately 9.06 % of all breast tumors (Forbes et al., 2017). It was reported to induce a decreased rate of intracellular protein transport (EMBL-CPD). A nuclear phenotype has not been described, which is in line with our findings in 2D. The acini, however, showed an altered architecture regarding nucleus positioning and Golgi organization giving the acini an unorganized morphology (figure 9). In addition, there was a lower density of acini as well as some single dead cells.

ESPL1 has a cysteine protease activity and cleaves cohesin subunits to initiate the final step of sister chromatid separation to further progress the mitosis machinery (Papi et al., 2005). The overexpression of ESPL1 was shown to increase tumorigenicity in an aneuploidy correlated fashion (Mukherjee et al., 2014) and it is overexpressed in 9.33 % of all breast cancers (Forbes et al., 2017). However, downregulation of ESPL1 in cooperation with non-functional p53 was shown to induce aneuploid tumors in mice as well (Mukherjee et al., 2011). In our hands, the knockdown of ESPL1 leads to polylobed nuclei with an aneuploid karyotype. In the 3D model, they develop into small, disorganized spheroids where polylobed nuclei pervade the structure and interfere with polarization and lumen formation (figure 9). Similar to DIAPH3 there was a reduced density as well as an increased number of dead cells.

H2AFX belongs to the family of histones that are involved in DNA organization. H2AFX is associated to sites with double-strand breaks and mediates DNA repair (Fernandez-capetillo et al., 2003). According to COSMIC H2AFX is overexpressed in 4.8 % of all breast cancers (Forbes et al., 2017). The downregulation was not correlated with any phenotypic changes in 2D by EMBL-CPD. However, in the 2D screen the cells showed an increase in nucleus size as well as reduced H2AFX intensity. In 3D, there were many dead cells, the spheroids were smaller and showed slight organizational abnormalities (figure 9).

SETDB2 is a member of SET domain proteins which contains a methyl-CPG binding domain and plays a role in segregation (Falandry et al., 2010). The COSMIC database reports an up- and downregulation in breast cancers of 2.08 % and 1.36 %, respectively. EMBL-CPD reported no morphological phenotypes upon knockdown in 2D which is in line with our findings. However, Falandry et al. reported nuclear as well as mitotic phenotypes for SETDB2 downregulation in 293T and HELA cells (Falandry et al., 2010). Spheroids growing on Matrigel lacked the developed of a hollow lumen (figure 9).

During the cell cycle the topoisomerase IIa (**TOP2A**) plays a role in chromatin condensation and chromatid separation by transiently breaking and subsequent rejoining of the DNA after the strands passed each other (Chen et al., 2015). TOP2A overexpression was

previously shown to correlate with poor prognosis (Chen et al., 2015) and the COMSIC database shows an upregulation in 18.75 % of all breast cancers. The knockdown of TOP2A in 2D results in a reduced number of cells and polylobed nuclei similar to the reported results of the EMBL-CPD. Comparable to ESPL1 the cells grown on Matrigel developed into disorganized, small spheroids consisting of or containing polylobed nuclei (figure 9).

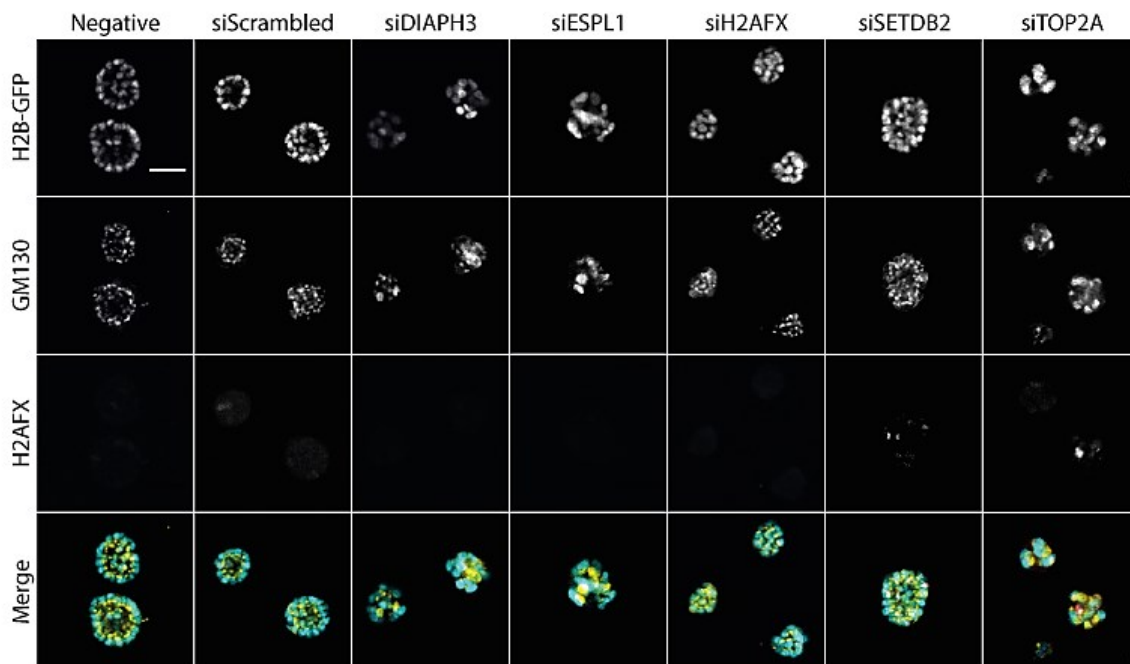


Figure 9: Final five candidates in comparison with the negative and scrambled control illustrating morphological changes of the acinus architecture defined by restructured nucleus (H2B-GFP) and Golgi (GM130) architecture and DNA-damage (H2AFX). The bar (top left image) represents a length of 50 μ m.

3.2 *Karyotype and acinus architecture*

Sadaie and colleagues showed that “irregular” or polylobed nuclei usually do not have a diploid chromosome set in 2D (Sadaie et al., 2015). Since we observed polylobed nuclei in the 2D and 3D knockdowns, we wanted to confirm that these interfere with acinus development. Most metaphases of the negative control showed no additional genomic abnormalities besides the chromosome 1q amplification and the 3, 5, 9 translocations, which were acquired during immortalization of the cell line (figure 10, Soule et al., 1990). However, in three out of 15 metaphases we detected gains or losses of single whole chromosomes and

one additional metaphase with a chromosome 14;22 translocation (supplementary table 4) indicating a consistent emerging and subsequent loss of subpopulations.

In general, MCF10A cells show an increased number of aneuploid cells when treated with Lipofectamine2000 and siRNA shown by the scrambled control (figure 11). The knockdown of ESPL1 and TOP2A leads to more severe division defects resulting in a high number of polylobed nuclei with polyploid karyotype in the cell populations (figure 11, x-axis) resulting in higher rates of 8-10 % aneuploid nuclei in comparison to 2-4 % for the others (figure 11). Also in some cases, ESPL1 and TOP2A downregulated cells acquired tetraploid or higher chromosome sets (figure 10).

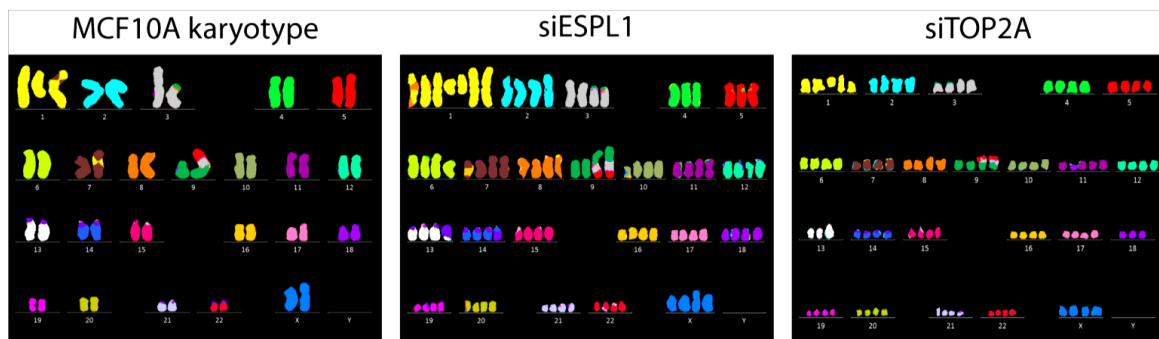


Figure 10: Comparison of a normal MCF10A karyotype with induced near tetraploid/aneuploid karyotypes. As described before MCF10A cells show a near diploid karyotype (Soule et al., 1990) with a chromosome 1q amplification and 3, 5, 9 translocations (i(1)(q10), del(1)(q12q32), der(3)t(3;9), i(8)(q10), der(9)t(3;5;9)).

After quantifying the amount of aneuploid nuclei, we correlated this with the relative number of abnormal acini. To do this we morphologically classified the acini into normal and abnormal (disrupted nuclei or Golgi organization, filled lumen, exceptionally big) classes. Subsequent plotting of the percental portion of aneuploid nuclei (x-axis) versus the ratio of abnormal and normal acini (y-axis) yielded a correlation of $r^2=0.49$ (figure 11) in a linear regression. The downregulation of ESPL1 and TOP2A lead to the highest number of aneuploid cells as well as abnormal spheroids (figure 11, red dotted box). Thus, all further experiments were focused on these conditions (Abstract figure: Karyotyping and expression).

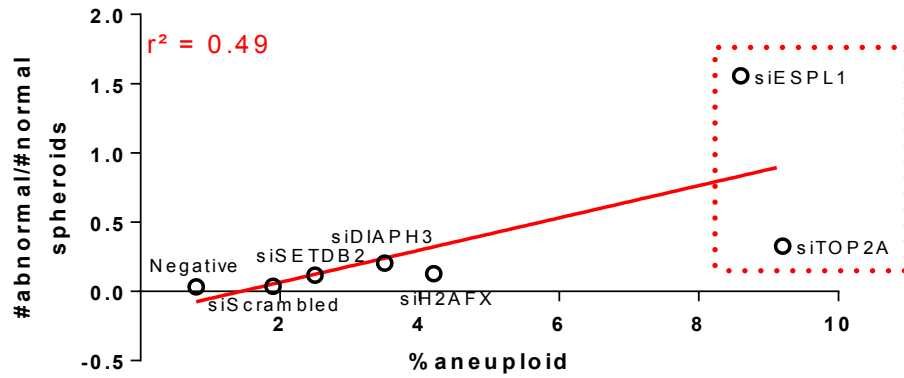


Figure 11: Correlation of aneuploid cells and formation of disorganized acini. Red line: linear regression ($r^2=0,49$), the red dotted box marks the candidates with the highest aneuploidy.

3.3 Time-lapse imaging and analysis

In a time-lapse over 7 days the nuclei were followed during the acinus development. The negative control cells developed into hollow and polarized spheroids (figure 12). The migration speed of cells inside the spheroids was fast at the beginning and rapidly slowed down. At the end, the cells only slowly moved which was similar for the scrambled control (linear regression slope significantly non-zero: $p < 0.0001$ and $p = 0.0116$ for Negative and siScrambled, respectively, figure 13). In addition, there were no aberrant mitotic events (figure 14).

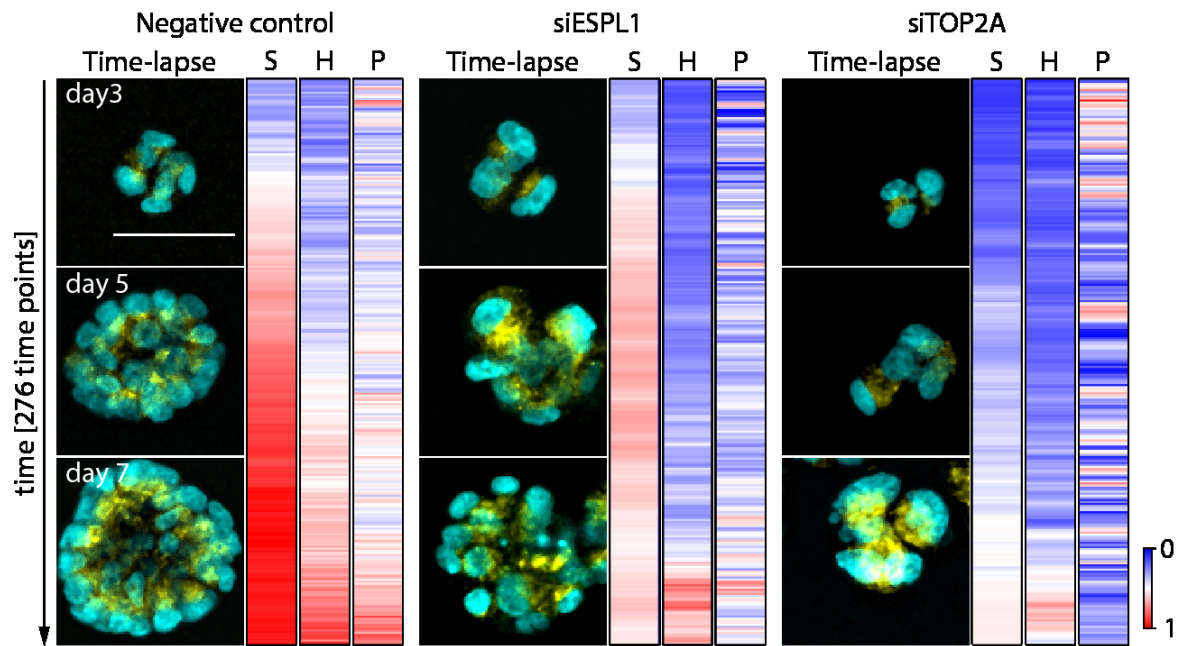


Figure 12: Acinus feature development tracked over time compared between one normal and two abnormal spheroids. Time-lapse (cyan: H2B-GFP, yellow: Golgi and lysosomes, the bar represents a length of 50 μm), S: size, H: hollowness, P: polarization, each bin of the 276 time points spans a time window of 20 minutes.

Disorganized spheroids of siESPL1 and siTOP2A were smaller and showed defects in polarization and hollow lumen formation over the whole development time when compared to the negative control (figure 12). The migration speed of the cells inside the acini remained constant for both knockdowns over the time-lapse period (linear regression slope significantly non-zero: $p = 0.72$ and $p = 0.57$ for siESPL1 and siTOP2A, respectively, figure 13). When cultured for up to 21 days, the disorganized spheroids maintained their morphology (supplementary figure 1). During division, cells of both conditions showed similar mitotic aberrations, e.g. delayed metaphases, unaligned metaphases and mitotic slippage (figure 14, t1-t5). The time-lapse additionally shows, how the polylobed nuclei do not align and sterically interfere with hollow lumen formation and polarization.

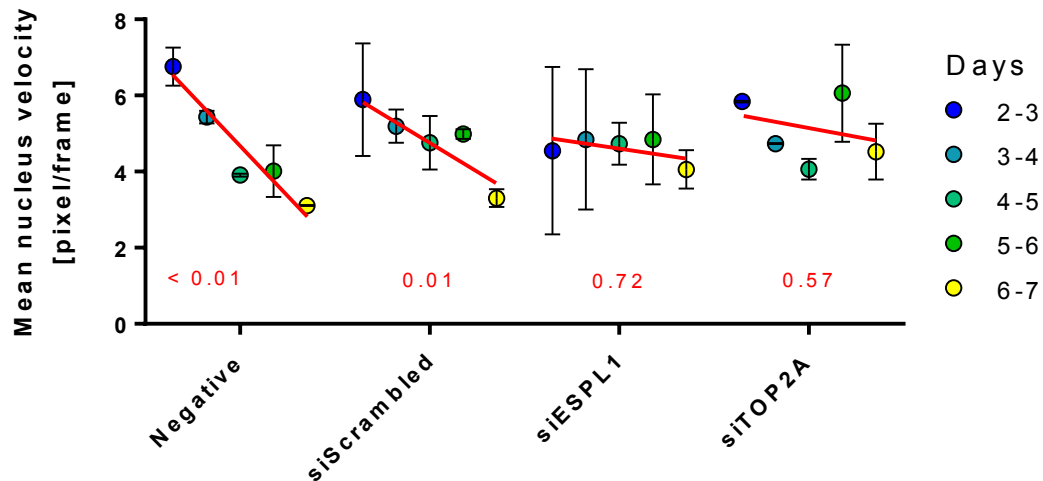


Figure 13: Mean velocity of the nuclei that migrate in the acini during different stages of the development. The red line represents a linear regression for each condition; the red number indicates the p-value for the regression line to have a slope of non-zero (slope $\neq 0$).

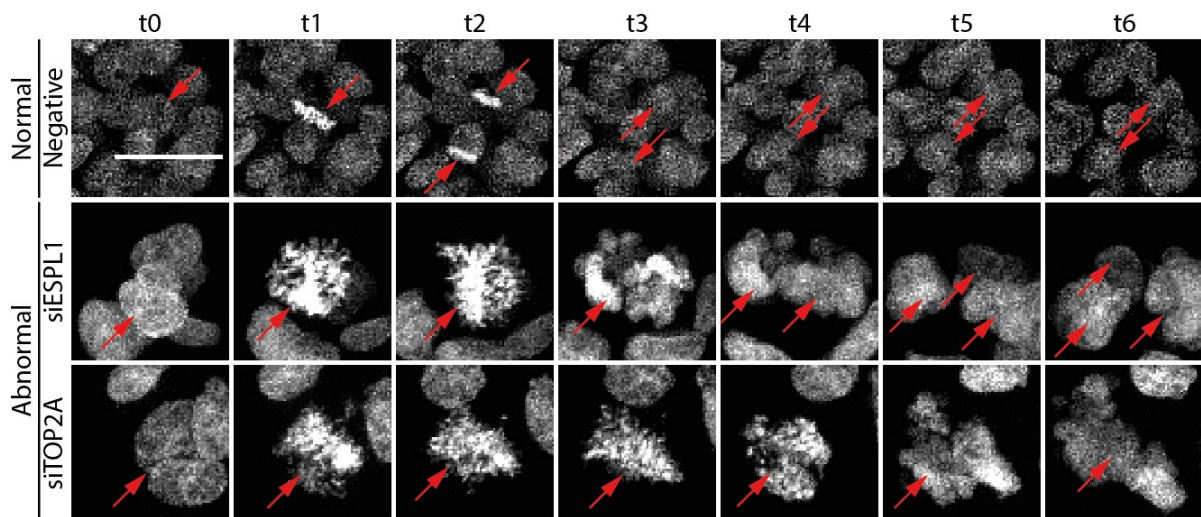


Figure 14: Abnormal mitotic phenotypes of siESPL1 and siTOP2A in comparison to a normal division. Grey: H2B-GFP; red arrows: single identifiable nuclei; time interval: 20 min; the bar represents a length of 25 μm .

3.4 Expression profiling

Testing the expression wise consequences of aneuploidy, we see the expression profiles of siESPL1 and siTOP2A to exhibit an upregulation of a higher number of genes, whereas the downregulation was similar for all three conditions in 2D, when compared to the scrambled control (table 4). When cultured on Matrigel, siScrambled showed a similar amount of

upregulated genes as in 2D, culture but fewer genes were downregulated. SiESPL1 showed a weaker level of deregulated genes on Matrigel, and siTOP2A showed a strong increase of upregulated genes but had a comparable number of downregulated genes to 2D culture (table 4). The overlap of deregulated genes of 2D and 3D cultured cells were minimal, and the only relevant common element was ALDH1A3, which was upregulated in siESPL1 and siTOP2A for 2D and 3D culture each.

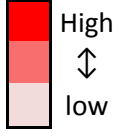
Table 4: Number of up- and downregulated genes of siScrambled, siESPL1 and siTOP2A cultured in 2D and 3D.

		siScrambled	siESPL1	siTOP2A
2D	Upregulated	9	108	128
	Downregulated	87	91	117
3D	Upregulated	18	67	283
	Downregulated	28	10	90

Gene enrichment analysis showed a deregulation of a broad range of cancer related genes in 2D as well as 3D knockdowns of siESPL1 and siTOP2A, while siScrambled was not enriched for any cancer related gene sets in 2D or 3D culture. Enrichment of genes that were > 1.5 fold overexpressed ($p < 0.05$) showed that siESPL1 gene sets were stronger associated to breast cancer gene sets than other gene sets, whereas siTOP2A showed a stronger association to leukemia gene sets (table 5). 3D profiles of siESPL1 and siTOP2A resulted in a higher rate of breast cancer related gene sets or cancer module sets when compared to 2D profiles (siESPL1: 8/20 vs. 4/20, siTOP2A: 4/20 vs. 0/20, respectively).

Table 5: Enrichment of top overexpressed genes and their associated enriched gene sets. The gradient codes for the number of enriched gene sets correlated with the respective cancer entity. O/L: enrichment of the overlapped upregulated genes of siESPL1 and siTOP2A, CA: cancer, Thyr.: thyroid, Panc.: pancreas, Melanom.: melanoma, Prost.: prostata.

Top 20 enrich.	siScrambled		siESPL1		siTOP2A		O/L
	3d	2d	3d	2d	3d	2d	-
Breast CA	0	0	6	3	1	0	5
CA module	0	0	2	1	3	0	2
Leukemia	0	0	0	0	4	0	1
Bone CA	0	0	0	1	0	2	0
Thyr. CA	0	0	1	0	0	1	1
Pancr. CA	0	0	0	1	0	0	0
Melanom.	0	0	0	1	0	0	0
Lung CA	0	0	0	0	1	0	1
Ovary CA	0	0	0	0	0	1	0
Prost. CA	0	0	0	0	0	1	0
Other CA	0	0	0	0	0	0	1
non-cancer	20	20	11	13	11	13	9



High
↕
low

Remarkably, siESPL1 and siTOP2A shared 46 genes in their profiles that were upregulated almost exclusively in 3D culture and relevant for cancer, e.g. keratins, interleukins, S100 calcium regulators, EGFR-effectors, BCSC-markers, cyclins and genes that share common transcription factor motifs. In a gene-set enrichment analysis the overlap list (O/L) showed a high association to breast cancer related gene sets (table 5).

While siESPL1 and siTOP2A showed an upregulation of ALDH1A3, siTOP2A also displayed a deregulation of additional BCSC markers, e.g. downregulation of CD24 and upregulation of CD44 (Musgrove et al., 2011; Munz et al., 2009).

Info box 2: Background of upregulated and cancer relevant genes

Keratins are widely expressed in many cancer entities and are correlated to invasion, metastasis and treatment responsiveness and serve as reliable marker proteins (Karantza, 2011). In the profiles of siESPL1 and siTOP2A Keratin 14, which is correlated to breast cancer metastasis (Cheung et al., 2016) as well as Keratin 16 and Keratin 81 could be identified to be overexpressed.

Besides immune regulation, interleukins have a broad range of functions in health but also disease (Brocker et al., 2010). Interleukin 1a and 1b as well as interleukin 8 are associated to breast cancer (Pantschenko et al., 2003; Singh et al., 2013) and were upregulated in both conditions.

S100A8 and its dimerization partner S100A9 were upregulated in siESPL1 and siTOP2A expression profiles. They are members of the S100 family, which are responsible for calcium sensing, and deregulation can be seen in many cancers (Bresnick et al., 2015).

The upregulated EGFR-effectors HBEGF and EGR1 were shown to contribute to metastasis (HBEGF, Zhou et al., 2014) and are regulated by c-JUN and JNK downstream of EGFR signaling (EGR1, Hoffmann et al., 2008).

ALDH1A3 is widely used as CSC/SC marker and correlates with a variety of cancers but especially breast cancer and breast cancer stem cells (Marcato et al., 2011; Duan et al., 2016). However, there are varying results reported in the literature regarding its association to stem cell populations, predictive power and prognostic value (Medema, 2013; Liu et al., 2015). ALDH1A3 was upregulated not only in the expression profiles of 3D culture but also in monolayer culture of siESPL1 and siTOP2A.

CCND1 is involved in a broad variety of pathways and upregulated by overexpression but also amplification in many cancer entities (Musgrove et al., 2011). According to COSMIC CCND1 is upregulated in ~17% of all breast cancers (Forbes et al., 2017). In a study that used melanoma cells, a CCND1 amplification was shown to confer sensitivity to certain drugs like paclitaxel (Wilson et al., 2016).

Additionally, the expression profiles of siESPL1 and siTOP2A share genes that have a MAZ (EGR1, ACTB, DAAM1, DUSP6, FGFBP1), ETS2 (EGR1, ACTB, HBEGF, ARHGDIB, CCL20), PAX4 (DAAM1, HBEGF, ARHGDIB) and/or TCF3 (DAAM1, DUSP6, HBEGF, CCL20) binding motif in the promoter region. These transcription factors are involved in RAS signaling (MAZ, Ray & Ray, 2015), cooperate with PTEN or mutant p53 to modify the chromatin

(ETS2, Trimboli et al., 2009; Prives & Lowe, 2015), miR regulation (PAX4, J Zhang et al., 2015) and finally Wnt-signaling and embryonic stem cell self-renewal (TCF3, Yi et al., 2011). However, these transcription factors themselves could not be shown to be upregulated.

3.5 Validation of the expression changes of MCF10dE and MCF10dT

Validating the findings of the microarray expression profile we see that the expression of CD24 in QPCR experiments was slightly downregulated for siScrambled and siTOP2A but not siESPL1. However, there was no significant change for any condition contradicting the expression-profiling results (figure 15). For CD44, ALDH1A3 and CCND1, siESPL1 and siTOP2A showed a significant upregulation, while the expression showed no difference for siScrambled (figure 15).

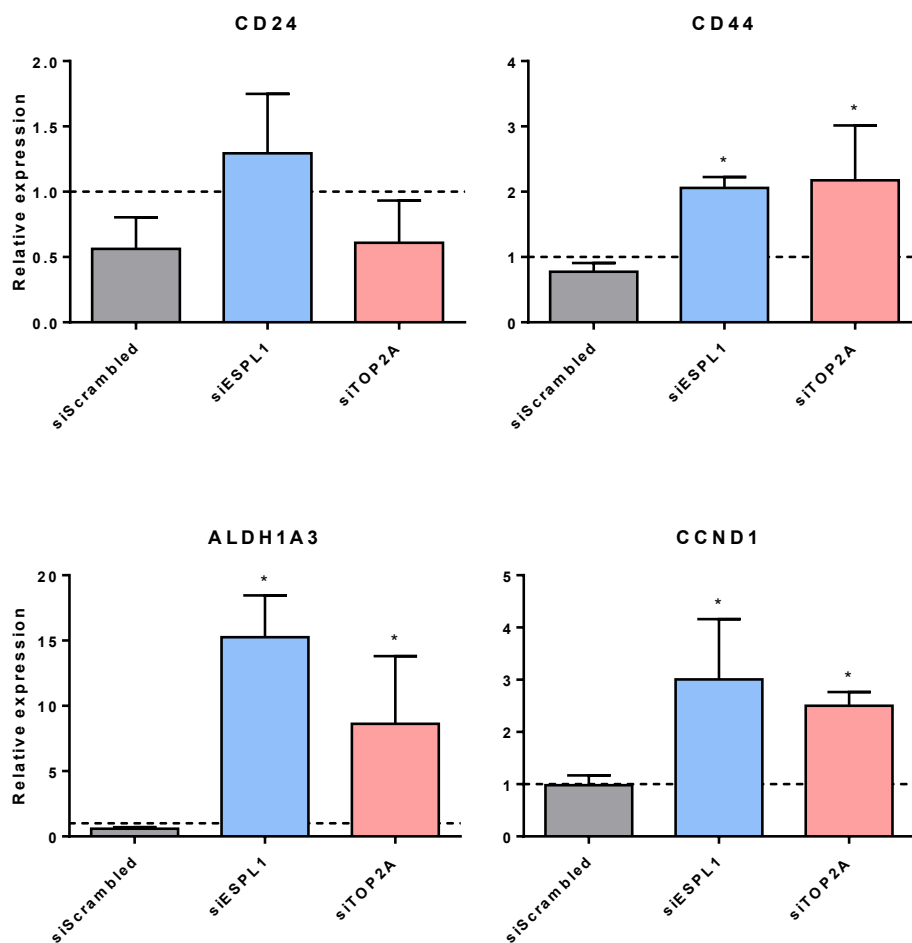


Figure 15: Relative expression of breast cancer related genes. The dotted line represents the expression level of the negative control that was used for normalization, asterisk: significant ($p \leq 0.05$).

In contrast to the QPCR results anti CD24 staining by immunofluorescence was significantly lower for siESPL1 and siTOP2A when compared to siScrambled or the negative control (figure 16). The BCSC markers CD44 and ALDH1A3 as well as the oncogene CCND1 showed a significantly increased fluorescence signal for siESPL1 and siTOP2A but not for siScrambled (figure 16).

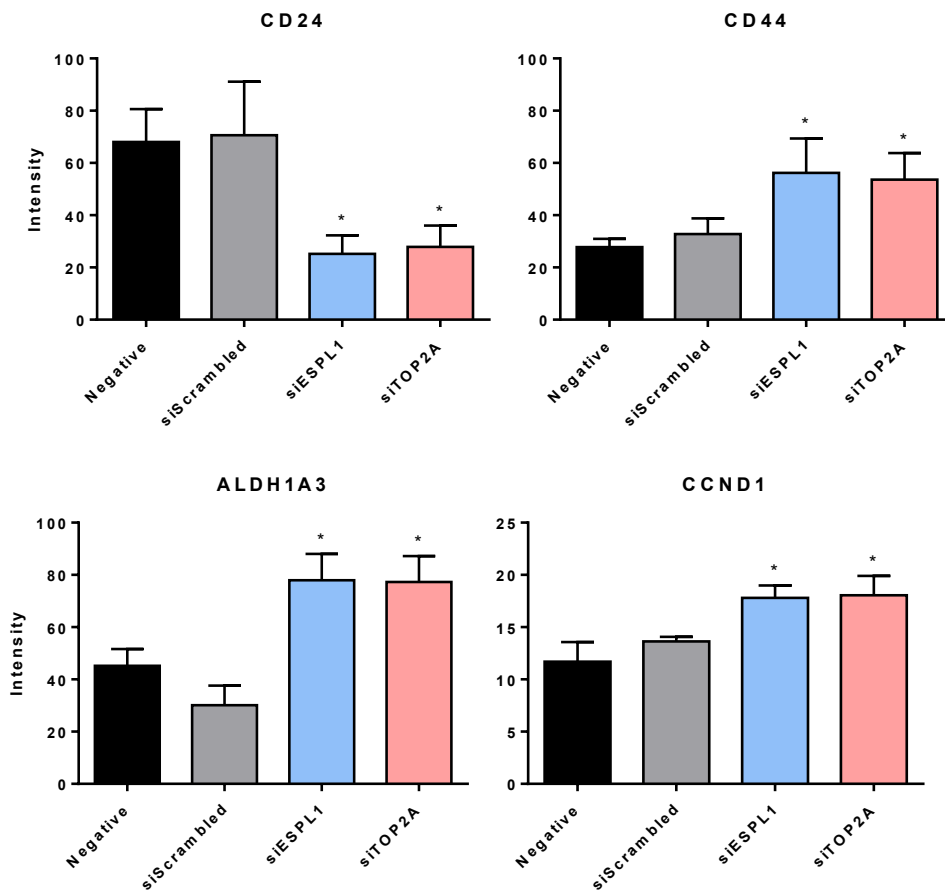


Figure 16: Immunofluorescence of BCSC markers in abnormal spheroids in comparison to normal spheroids. Five representative spheroids per condition were imaged, asterisk represents a significant difference to the negative control ($P = 0.05$).

3.6 Drug tests

By testing the sensitivity to paclitaxel, which can be altered in cancer cells, we detected in 2D culture, the IC₅₀ of the negative and scrambled control was 2.6 and 2.8 μM , respectively. The IC₅₀ changed from 2.6 to 9.5 for the negative control but did not change for the scrambled control in 3D culture. SiESPL1 cells showed an IC₅₀ of 10 μM as monolayer and increased their sensitivity to $<0.1 \mu\text{M}$ when cultured on Matrigel. The IC₅₀ of siTOP2A

cultured as monolayer was 0.2 μM and changed to $<0.1 \mu\text{M}$ similar to ESPL1 undercutting the lowest measurable value for this concentration range. Interestingly Bouchet and colleagues reported a reduced sensitivity of aneuploid cells against paclitaxel (Bouchet et al., 2007). However, Wilson and colleagues showed that CCND1 overexpression can sensitize cells against paclitaxel (Wilson et al., 2016).

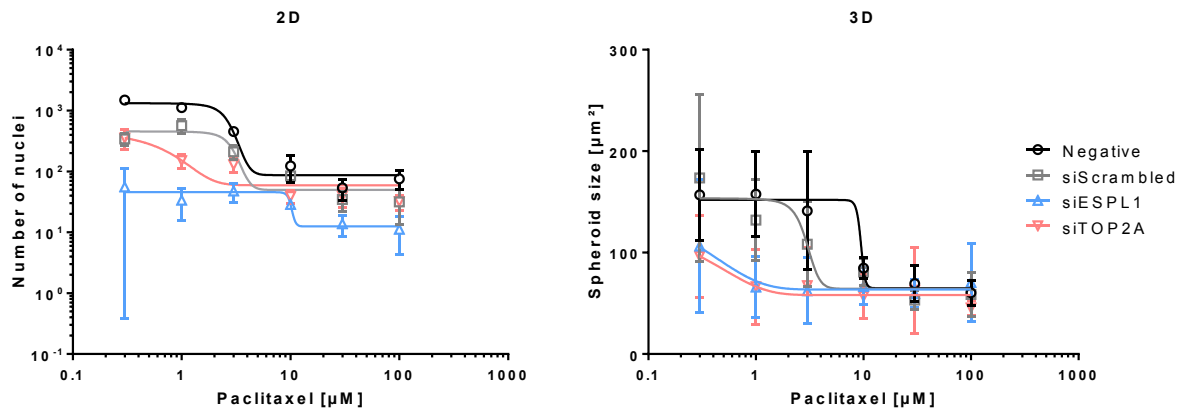


Figure 17: Drug response curves of siScrambled, siESPL1 and siTOP2A cultured in 2D and 3D treated with paclitaxel.

3.7 Testing tumorigenic potential in-vitro and in-vivo

Stem cells show an increased renewal capacity when compared to other cells. To assess whether the upregulation of stem cell marker upon ESPL1 and TOP2A downregulation leads to a functional increase in renewal capacity of MCF10A cells, we performed a clonogenic mammosphere formation assay. When previously cultured in 2D, the cells of the negative control showed a mammosphere forming efficiency (MFE) of approximately 1 %, which is in line with the data reported in previous studies (Fang et al., 2011; Liu et al., 2009). The scrambled controls as well as siESPL1 and siTOP2A showed an MFE of around 0.5 %. When previously cultured in 3D, the MFE decreased for the negative control to a MFE of approximately 0.5 % but did not change for the other conditions. Although the 2D cultured negative control showed a higher MFE than the other groups there were neither statistically significant differences between 2D and 3D cultured cells nor between the first and the second generation in any group.

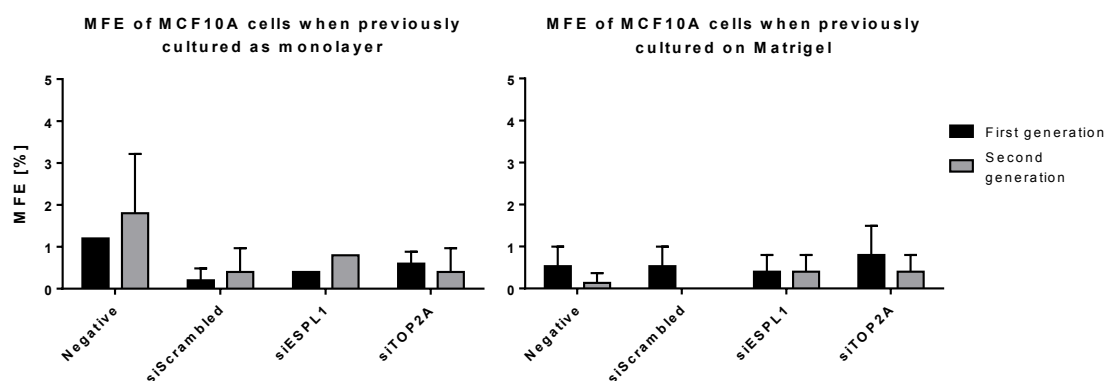


Figure 18: Comparison of the mammosphere forming efficiency from cells previously cultured as monolayer or on Matrigel.

Because of the abnormal morphology and related expression, we tested if these changes are sufficient to increase the tumorigenic potential of any of the MCF10 cell lines. After imbedding in Matrigel and injection into NGS mice, neither MCF10A nor any of the si- or shRNA transfected cell lines formed persistent tumors over 4 weeks.

MCF10AT cells, which were generated by overexpressing the active mutant HRAS also acquired the ability to form persistent tumors in mice with a low efficiency (Dawson et al., 1996). We exploit the low efficiency to gain a bigger dynamic range for testing the effects of aneuploidy on the tumorigenicity as well as the tumor initiation capacity. After eight weeks siScrambled treated MCF10AT formed tumors in five out of six mice with a tumor weight of 0.02 ± 0.01 gram when injecting 50.000 cells. SiESPL1 and siTOP2A treated MCF10AT formed tumors in six out of six mice each with the weight of 0.11 ± 0.21 and 0.18 ± 0.25 resulting in no statistical difference for tumorigenic potential or tumor weight.

MCF10CA cells, which served as positive control in this experiment, formed tumors in 100 % of the mice. SiTOP2A treated MCF10CA cells developed into significantly bigger tumors ($p = 0.05$) when compared to siScrambled treated MCF10CA (supplementary figure 3). Although siESPL1 treated MCF10CA cells on average formed slightly bigger tumors, there was no statistical significance (supplementary figure 3).


3.8 Chromatin conformation capture (4C-seq)

siESPL1 and siTOP2A showed a high degree of aneuploidy, which led to increased amounts of chromatin in the nucleus. To test if this interferes with the chromatin conformation that in turn influences the expression, we looked at the interaction of the DNA at the transcription start site (TSS) regions of breast cancer related genes (Chang et al., 2013), that were also deregulated in the expression profiles of siESPL1 and siTOP2A.

Interestingly, cells cultured as monolayer on average showed the highest cis-chromatin interaction value (cCIV), which estimates the number of interactions with a high probability of the viewpoint (in this case the transcription start site of each gene) with other loci on the same chromosome (Table 6). Using the cCIV, the viewpoints with the highest discrimination between 2D and 3D are CCND1, FGFBP1, MAZ and TCF3. HBB, which we used as a control (Guo et al., 2015), is fairly constant for monolayer cultures. In 3D however, the cCIV of HBB shows a higher variability between the cell lines.

Table 6: Cis-chromatin interaction value by viewpoint and cell line. 2D: cells cultured as monolayer, 3D: cells cultured on Matrigel.

	CCND1	DUSP6	EGR1	ETS2	FGF1	FGFBP1	FGFR2	HBB	MAZ	PAX4	RUNX3	TCF3	
2D	MCF10A	579	487	566	584	602	587	636	648	515	634	622	558
	siScrambled	534	565	519	586	568	583	595	662	512	591	602	452
	siESPL1	576	583	549	679	590	623	616	668	592	627	598	585
	siTOP2A	596	629	642	652	612	639	650	671	577	659	650	560
	MCF10CA	578	567	573	588	572	550	593	655	461	645	621	397
3D	MCF10A	542	524	450	494	451	469	497	566	403	511	477	182
	siScrambled	543	567	527	475	495	443	510	524	320	515	494	270
	siESPL1	326	418	491	506	563	402	531	628	419	602	580	319
	siTOP2A	360	334	414	482	539	438	504	616	410	621	507	168
	MCF10CA	56	546	485	516	573	446	525	485	405	576	567	307



Interestingly, CCND1 expression of our cell lines did not vary much for monolayer cultured cells but did show an upregulation with increasing aneuploidy in 3D (Abstract figure: 4C-seq). When the expression was plotted against the cCIV, we found a negative correlation with a coefficient of determination of $r^2 = 0.82$ (figure 19, $p < 0.001$). This

correlation is noteworthy with the background that Dalvai and colleagues could show that a binding of the 3' enhancer enh2 to the TSS region of CCND1 mediated by H2A.Z could downregulate the expression of CCND1 (Dalvai et al., 2012).

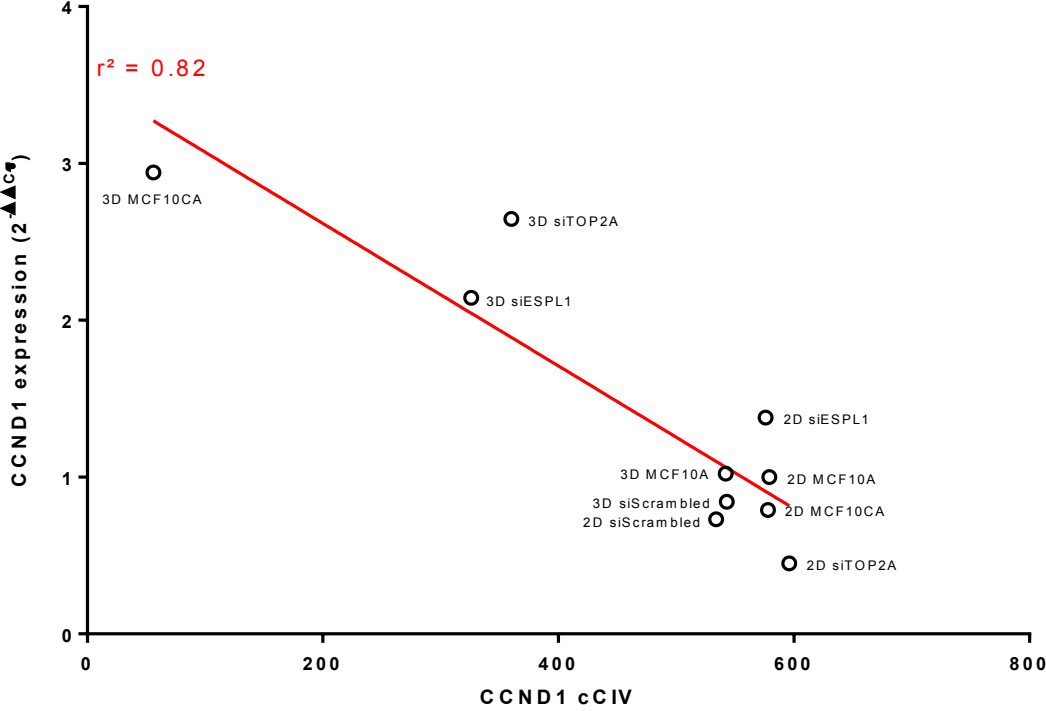


Figure 19: Correlation of cCIV and the expression level of CCND1 in the different cell lines.

Comparing the probability of chromatin interactions around the TSS of CCND1 between the different cell lines, we could see the highest number of reads between 600 kb upstream and 60 kb downstream (figure 20, near-cis region). The interactions inside and outside this window did not change much in 2D for any condition (figure 20, left). In 3D, the reads drastically dropped close to zero outside of the previously defined window (far-cis region) for MCF10dE and SiTOP2A, whereas we saw a clear reduction of reads within the near-cis region (figure 20, right). Remarkably, the MCF10CA sample showed very few reads around the TSS of CCND1.

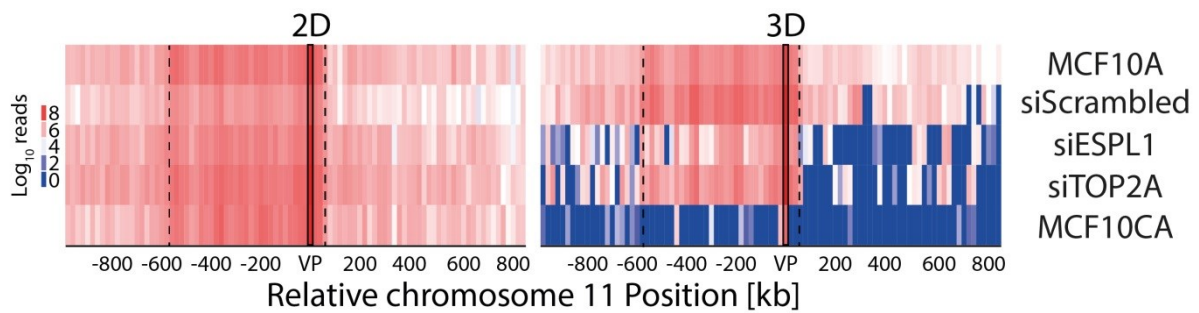


Figure 20: Far-cis interaction probability. The numbers are given as relative position on the chromosome 11 from the TSS of CCND1 (yellow VP) in kb and binned into 20 kb segments. Two dashed lines highlight the 600 kb upstream and 50 kb downstream window.

The annotation of the interaction in the near-cis region with H2A.Z acetylation, enhancer and CTCF-binding sites using the data generated by the Broad institute, ENCODE for HMEC, vHMEC, epithelial and myoepithelial breast cells and “Genecards.org” revealed a reorganization of interactions for 3D cultured siESPL1 and siTOP2A. After the identification of each major peak with H2A.Z acetylation, enhancer and/or CTCF binding sites, we saw that 3D cultured MCF10A as well as siScrambled cells showed a similar profile as the 2D conditions (figure 21). 3D siESPL1 showed reduced peaks at position 2, 4, 6, 7, 8, 10, 11, 13, 14, 15, 16, 17, 19, 23, 24, 25, 27, and 33. With alteration of interactions in the positions 2, 6, 8, 11, 13, 15, 18, 23, 24, 26, 27 and 33 siTOP2A cultured in 3D showed overall more interactions in this region than siESPL1. The position 34 does not only contain a predicted enhancer and CTCF binding site but is also located in the ORAOV1 locus/promoter and is annotated as the enhancer with the highest enhancer-score in the gene cards database for CCND1. Two out of three positions that are exclusively marked to be CTCF binding sites were altered in 3D siESPL1 and siTOP2A. Remarkably, none of the exclusive CTCF binding sites showed H2A.Z acetylation which is associated to active gene regulation (Valdes-Mora et al., 2012).

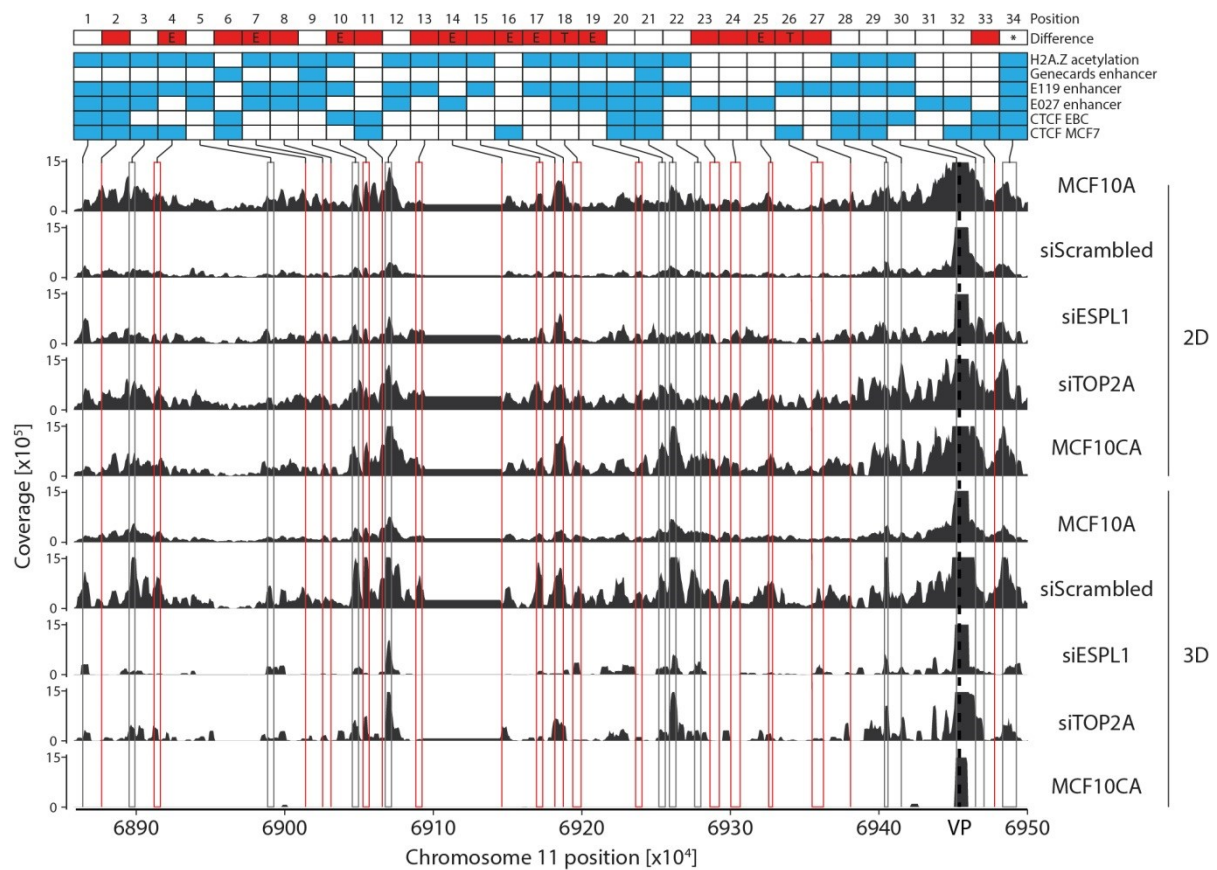


Figure 21: Near-cis interaction probability and alteration of the enhancer and CTCF binding site landscape. In the top row of the panel red indicates a difference of 3D siESPL1 (E) or 3D siTOP2A (T) or both if not further specified to all the other conditions. In the bottom rows of the panel, blue indicates a H2A.Z acetylation site, enhancer (Genecards, E119 enh, E028 enh, E027 enh) or CTCF binding site (CTCF EBC, CTCF MCF7) predicted by the ENCODE dataset or by “Genecards.org”. The asterisk indicates the overlapping region of ORAOV1 with the enhancer and CTCF binding sites. The viewpoint (black VP) is highlighted by a dashed line.

4 Discussion

At the beginning of the last century Theodor Boveri proposed cancer to be a chromosomal disease (Boveri, 2008). He could detect mitotic abnormalities in urchin eggs that resulted in chromosome mis-segregation. Unaware of genes and the regulation behind them, he argued that abnormal mitosis might lead to the accumulation of chromosomes and finally result in cancer. Since the connection of mutations and cancerogenesis in 1982 (Tabin et al., 1982; Reddy, Reynolds & Santos, 1982) the genetic origin of cancer is extensively explored. However, today we know of several more contributors to cancer. In 2011, Hanahan and Weinberg summarized many of the theories centered on genomics in the “Hallmarks of cancer” (figure 22, Hanahan & Weinberg, 2011). They explain how genetic changes convey new properties, which are also modified by the environment.

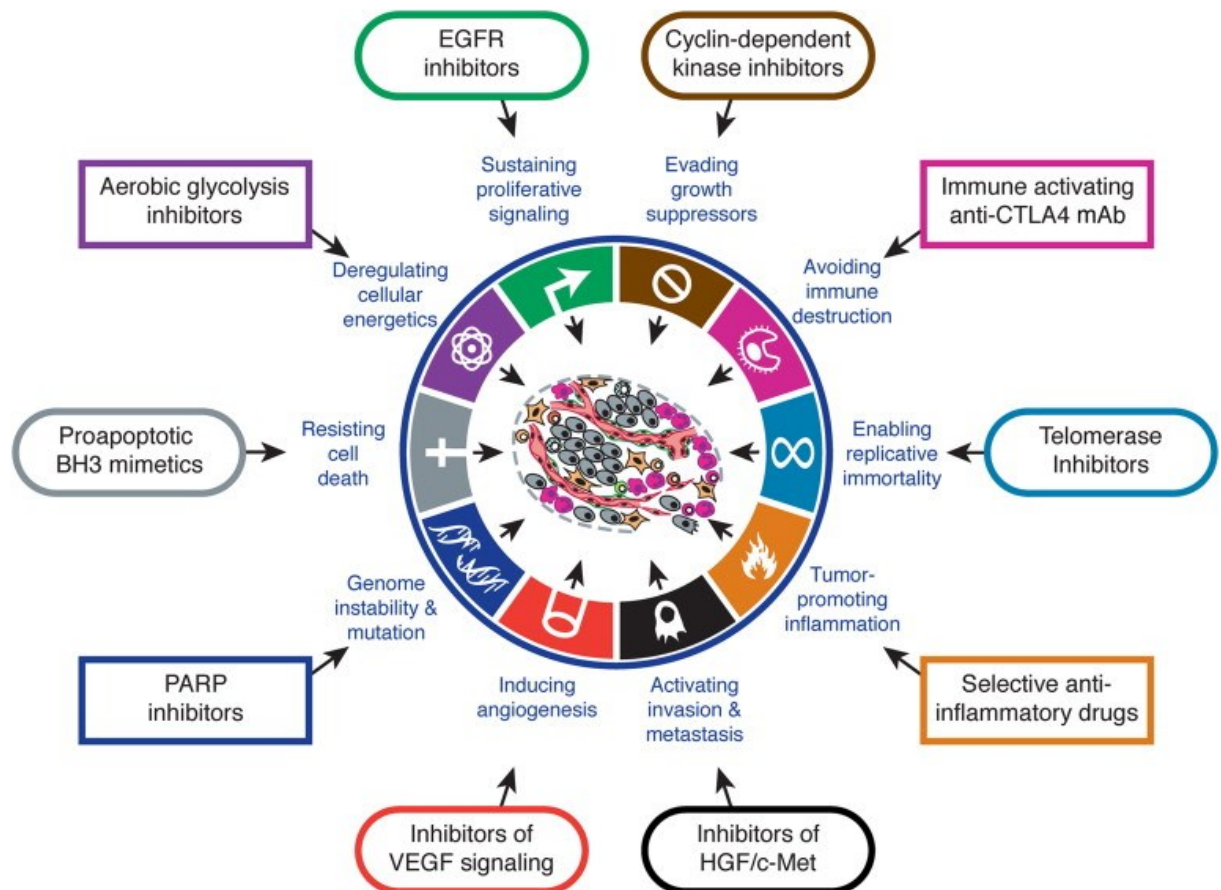


Figure 22: Hallmarks of cancer and potential treatment strategies as proposed by Hanahan and Weinberg (Hanahan & Weinberg, 2011).

The important influence of the microenvironment on tumorigenesis in detail was summarized by Mina Bissell (Bissell et al., 2011). While the hallmarks of cancer mainly focused on a genetic origin of the hallmarks Bissell & Hines argue that the microenvironment, e.g. stroma and extra cellular matrix is equally important for tissue homeostasis and tumorigenesis.

Another factor influencing tumorigenesis is epigenetic gene regulation. DNA- and histone-modifications as well as chromatin-conformation have been proven to effect central pathways, e.g. HIPPO signaling (Harvey et al., 2013), and to regulate the expression of nearly all genes.

Generally, during cancerogenesis these factors show a combinatorial effect. The majority of colorectal cancers develop after a mutation sequence of the APC-KRAS-DCC-P53 genes and often show micro-satellite instability and epigenetic gene alterations (Armaghany et al., 2012). Breast cancer shows aneuploidy in 75 % of the tumors, and recurring mutations appear in many genes including TP53, PIK3CA, PTEN and BRCA1 (Nik-zainal et al., 2016). However, the interplay of all of these factors is not clear.

Remarkably, aneuploidy is seen in nearly all cancer entities (Cimini, 2008). Several studies suggest that the acquisition of aneuploidy is a central, as well as early event in cancer (Nowak et al., 2002, Wang et al., 2014, Gao et al., 2016). In their recent work, Gao and colleagues raise the question how genome instability can be reversibly turned on. In their opinion, this is a critical question because aneuploidy shows detrimental effects on cell viability and proliferation (Gao et al., 2016). However, in the perspective that other data suggest that adult tissues show a certain degree of aneuploidy by chance (Knouse et al., 2014) a reversible modus not always seems to be necessary. Rather, it implies that there is a tendency for abnormal mitosis. In most cases, arising abnormal daughter cells are selected against in the tissue environment, eliminated by the immune system or they die because of the detrimental effects of aneuploidy. Several studies showed that additional chromosomes can lead to an overload of proteins that affect the cells by proteotoxic as well as oxidative stress and DNA-damage (Dodgson et al., 2016; Ohashi et al., 2015); an insight that might even be utilized for future treatments. These effects are further amplified in ongoing CIN and effectively lead to cell cycle arrest and death.

Some of the arising abnormal daughter cells survive this first wave of death however. Theoretically, they could achieve this by i) exhibiting only low levels of CIN and “fly under the radar” (Giam & Rancati, 2015), ii) the deletion of cell death initiators like p53, iii) excessive accumulation of genomic regions with survival genes overriding cell death initiators or a combination of these scenarios. However, scenarios ii and iii alone seem to be unlikely because ongoing genomic rearrangements would further impair proper cell division and viability. The central question now is how the daughter cells utilize the genetic material to fuel the evolution, despite suffering from impaired cell division. One explanation might be a phenomenon called triplosensitivity and haploinsufficiency as proposed by Davoli and colleagues (Davoli et al., 2013). The authors hypothesize, that the cells overcome the detrimental effects of aneuploidy by specifically amplifying chromosomes or chromosome arms that carry GO-genes while deleting regions with STOP- or tumor suppressor genes. However, the mechanism why some genes are haploinsufficient or triplosensitive, while others are not remains unclear. Additionally, this concept intrinsically requires cells to divide with a sufficient rate and/or live long enough to generate enough daughter cells to select for the required genomic setting. This assumption calls to dispute what mechanisms can alter expression upon aneuploidization to make the cells survive and also drive the cell cycle.

Changes in the epigenome could be shown to contribute to tumorigenesis. In some cancers there are very few or no detectable mutations and even no genomic alterations whatsoever (Mack, 2014; Parker et al., 2014; Rogier Versteeg, 2014); again highlighting the importance of epigenetics and microenvironment for tumor development. However, DNA methylation tends to be more stable and changes over longer periods of time (Talens et al., 2017; Vandiver et al., 2015). Some histone modifications are more readily changeable and can react to immediate situations (Lee et al., 2010). It is unclear though, how or whether aneuploidy can directly influence DNA methylation or histone modification and utilize their function for increased viability or cell cycle fidelity.

Recently, chromatin capture technologies revealed how CTCF plays an important role in the regulation of genes by chromatin-chromatin interactions, which was reviewed by Ong and Corces (Ong & Corces, 2014). Presuming the organization of chromatin interactions is tightly controlled and dependent on a correct chromosomal composition the accumulation

of additional genetic material potentially directly interferes with proper chromatin-chromatin interactions and could thus deregulate tumor drivers.

4.1 Finding the connection of morphology, karyotype and cancer using siRNA HCS

As Ottesen and colleagues described, early stages of breast DCIS show high rates of aneuploidy (Ottesen et al., 1995). It is very likely, that these genomic alterations lead to an abnormal behavior of cells rendering them immune against growth inhibitory cues exerted by the microenvironment while providing a niche the cells can properly grow in. Utilizing this phenomenon, we induced abnormal mitosis in 3D cultured MCF10A cells using siRNA targeted at different CIN associated genes. Thereby, we selected candidates that i) increase the rate to accumulate DNA as proposed before mimicking the tendency of normal tissues to acquire aneuploidy by chance and ii) induce abnormal morphological properties similar to CIS. It is important in this context to have a transient and specific effect with the right dose of aneuploidy. ShRNA and CRISPR are excellent tools to precisely and permanently downregulate genes. However, in our context they would introduce ongoing and high levels of CIN and lead to additional genomic changes. Also, most drugs transiently inhibit the function of a certain protein but often show side effects (Kayl & Meyers, 2006) which renders them impractical for our purpose.

Unexpectedly, a workaround over 2D KDs which were then reseeded onto Matrigel was necessary. In a first attempt of siRNA mediated knockdown under 3D culture conditions, we could detect mCherry labeled siRNA in the cytoplasm of cells but cells cultured on Matrigel did neither show knockdown-specific phenotypes nor mRNA depletion in QPCR assays. In contrast, KD in 2D and subsequent reseeded into 3D resulted in abnormal acini that resembled lobular neoplasia as defined by O'Malley (O'Malley, 2010). Zoldan and colleagues reported that mixing siRNA particles with Matrigel did not affect the efficiency of the siRNA (Zoldan et al., 2011). The question remains whether this is attributable to the cell line, because different cell lines react to siRNA with different efficiencies. To rule out problems with Matrigel we tested different LOTS because of previously reported high LOT-to-LOT variation with regard to the compound composition which might alter the efficiency of siRNAs delivery.

When matching our 2D knockdown phenotypes with the literature as well as RNAi databases, we saw only 45 of 82 candidates to reflect the findings. This might indicate that the siRNAs effects either in our hands and/or on MCF10A cells are different in comparison to other examiners/cell lines. Another argument is that the analyzed parameters were not exactly the same. Interestingly, while AURKB, INCENP, ECT2, and TPX2 induce similar phenotypes when compared to ESPL1 and TOP2A in 2D the knockdowns in 3D result in no or weak morphological changes or killed the cells. This might indicate that the modus or grade of aneuploidization in these conditions is distinct and the cells can either rescue the effects or die from it in 3D.

ESPL1 and TOP2A knockdowns resulted in whole genome and whole chromosome amplifications without additional alterations. This similar phenotype might be attributable to the related function and cooperation during sister chromatid separation (figure 23).

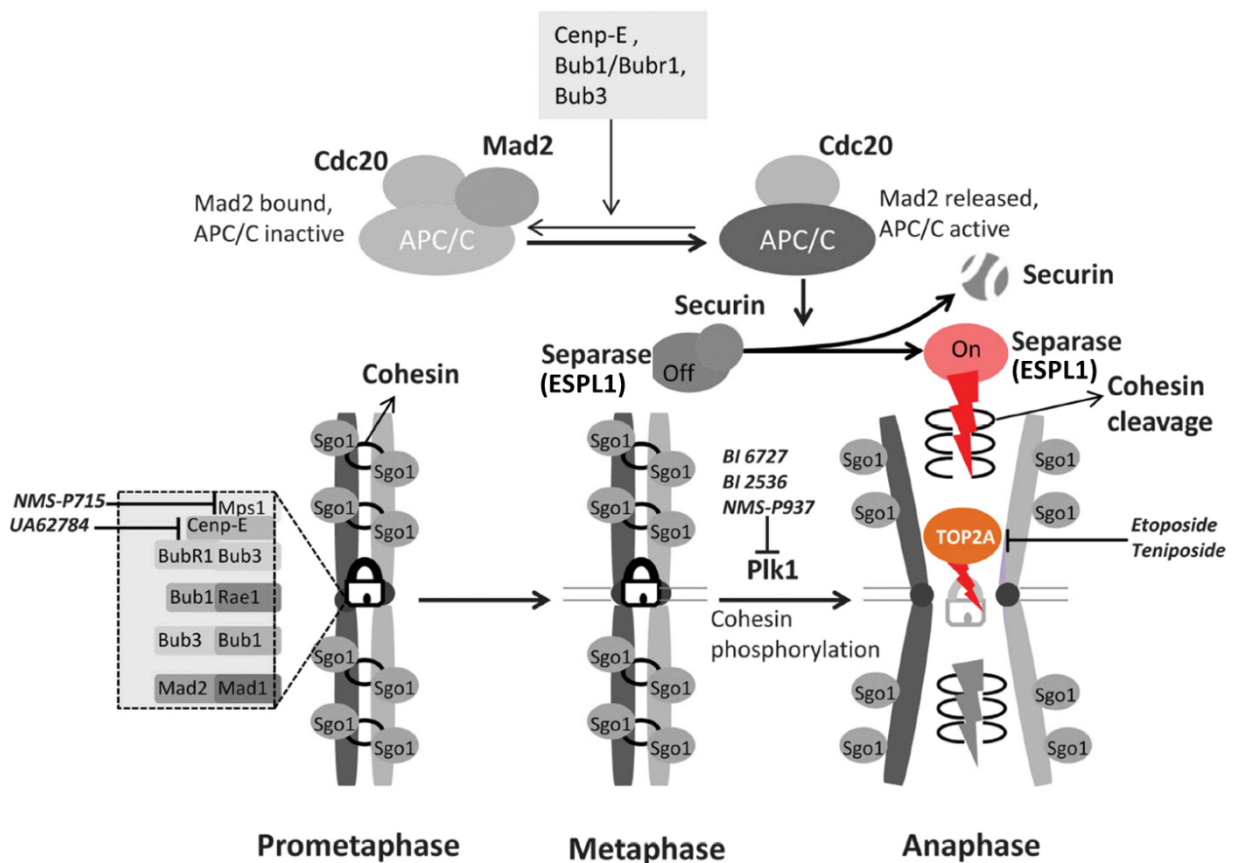


Figure 23: Cooperation of separase (ESPL1) and TOP2A during sister chromatid separation in the anaphase (Chen et al., 2015).

This shared phenotype is remarkable before the background that TOP2A influences many other processes that involve decatenation (Chen et al., 2015). Apparently, impaired sister chromatid separation is the most immediate effect. The transient downregulation and subsequent aneuploidy alters the acinus morphology. This is a demonstration that gene mutations are not imperative to induce abnormal morphologies reminiscent of DCIS/LCIS.

By anticipating the inhibitory effects of p53 on growth of aneuploid cells we additionally sensitized the cells by knocking down p53 as well. However, the abnormal acini kept growing and retained their morphology even without p53 knockdown. This brings into question why p53 did not fully inhibit the growth. When compared to the parameters of the single-KD there were no measurable differences of the double-KD. By monitoring aneuploid cells in 2D we could see a clear reduction of mitotic activity and these cells were eventually overgrown. In the 3D context, they could grow unhindered with a slower rate.

Time-lapse imaging revealed that aneuploid cells displayed an impaired rotational movement and loss of directed migration. This interfered with the usual motion of cells that develop into normal acini. Wang and colleagues reported that improper motion can lead to abnormal acinus morphologies (Wang et al., 2013). Seemingly, under the right circumstances even “unfit” MCF10A cells can continue to grow and form abnormal acini.

As reported by Yoon and colleagues, MCF10A exhibit a low level of chromosomal instability (Yoon et al., 2002). We saw four out of fifteen cells with either chromosome 14;22 translocation or gains of random single chromosomes (supplementary table 5). Taking into account that these changes only occurred rarely and our results were robust over several conditions and experiments we presume that this does not interfere with our findings.

4.2 *Cancer marker expression and its effects*

As reviewed by Yersal and Barutca, the molecular classification of cancer-subtypes using expression profiles has increased the number of treatment options (Yersal & Barutca, 2014) and considerably improved the therapy success for some sub-types (Perou et al., 2000). In order to classify the breast cancer subtypes, ER α , keratins, PgR and HER2 expression are quantified. However, there are many other markers that are deregulated and can help predict the treatment strategy and outcome (Kabel, 2017). The deregulation of cancer

markers can result from a variety of reasons, namely activating or deactivating mutations of upstream effectors, gene amplifications, deletions or translocation. Nevertheless, deriving transcription levels from the genetic setup alone is challenging because gene regulation heavily depends on the epigenetic setup.

Interestingly, independently of cell lines or tissue, several studies reported an upregulation of cancer related genes upon induction of aneuploidy (Zhang et al., 2013; Ben-david et al., 2014). Our expression profiles support this and share the expression of breast cancer markers (figure 15 and 16). The breast cancer markers CD24, CD44 and ALDH1A3 are frequently deregulated and correlate with aggressiveness. However, a recent study showed that the mentioned markers seem to be expressed by different subpopulations rather than providing a breast cancer stem cell profile (Liu et al., 2014). This idea is supported by many conflicting results regarding CD24⁻/CD44⁺ breast cancer stem cell profiles as discussed by Liu and colleagues. Another marker that was upregulated in our expression profiles is CCND1 (figure 15 and 16). CCND1 plays an important role during the G1 phase and connects several central pathways involved in tumorigenesis (figure 24). It is overexpressed in over 17 % of all breast cancers but shows copy number variations in only 5 % of the cancers (Forbes et al., 2017). Assuming that CCND1 expression is not sensitive to copy number variations (figure 1, left), the upregulation upon aneuploidization in both conditions (siESPL1 and siTOP2A) could be a hint that CCND1 silencing mechanisms were disrupted by an impaired chromatin conformation.

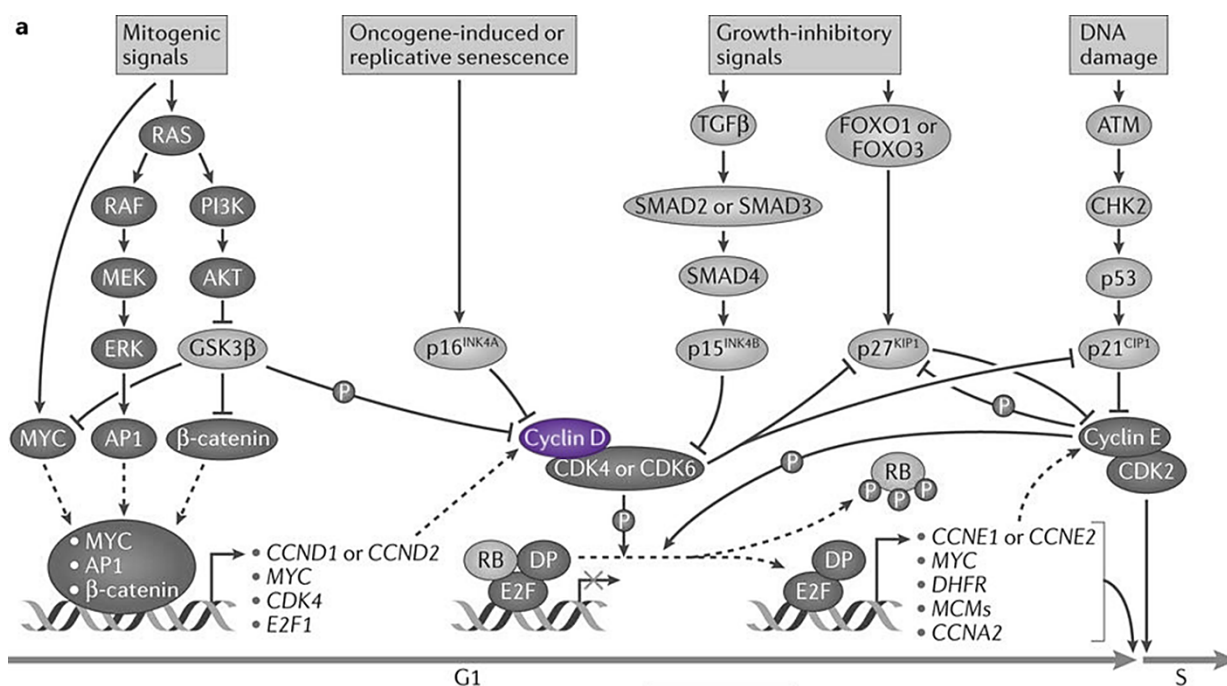


Figure 24: The central role of Cyclin D (CCND1 and CCND2) during the G1 phase (Otto & Sicinski, 2017).

Insensitivity of CCND1 to ploidy in invasive carcinoma (figure 1, right) does not necessarily contradict the upregulation of CCND1 in newly acquired aneuploidy of CIS-like acini. Potentially, CCND1 is upregulated early after aneuploidization to maintain the cell cycle until CIS further develops into more severe forms. This is supported by the study of Simpson and colleagues (Simpson et al., 1997) where they found a decreasing expression level of CCND1 with increasing DCIS grade. They pointed out that yet to be discovered mechanisms other than aneuploidy were responsible for the CCND1 upregulation in early DCIS. Combining the facts that DCIS shows aneuploidy early (Ottesen et al., 1995), CCND1 is upregulated in early DCIS (Simpson et al., 1997) and upon aneuploidization in our experiments (figure 15 & 16), we postulate that CCND1 upregulation occurs early after aneuploidization in CIS development and declines in later stages to be upregulated in invasive carcinoma again.

The transition of CIS to invasive carcinoma is extensively studied (Ma et al., 2003; Lee et al., 2012; Hannemann et al., 2006). However, there is very limited comprehensive data about differential expression from CIS when compared to normal tissue. Single mutations in genes like BRCA1 and BRCA2 as well as ATM, BRIP1, CDH1, CHEK2, MRE11A, NBN, PALB2, PTEN, RAD50, RAD51C, STK11 and TP53 are correlated to breast cancer and are used as

diagnostic markers (breastcancer.org). Nevertheless, it is not clear if these markers increase the risk of either tumor initiation or tumor development in the case CIS is already present.

In a study by So and colleagues (So et al., 2014), the authors argue that expression of CCND1 alongside c-MYC and IGF-IR is important for the initiation of cancer development. They also found a higher CD44 expression and its isoforms to be correlated with increasing malignant potential using MCF10 cell lines. Remarkably, the introduction of random aneuploidy to MCF10A cells led to the development of acini reminiscent of CIS regarding morphology and expression. Though, the upregulation of CD44 and CCND1 in our experiments was not sufficient to increase the tumorigenic potential. This might be due to a lack of de-regulation of other important tumor drivers that were discussed in the previous paragraphs. Moreover, it might be possible that engrafted cells require a larger time window to fully develop their tumorigenic potential. In addition, this raises the question, what is necessary to induce tumorigenicity.

MCF10AT cells which overexpress activated HRAS, are tumorigenic and develop into invasive carcinoma in around 20 % of the xenografts (Dawson et al., 1996). In our hands, MCF10AT cells formed xenografts with a similar size and frequency with siESPL1, siTOP2A or siScrambled treatment. The additional amplification of the karyotype did not affect the tumorigenic potential of MCF10AT. So and colleagues showed already that CCND1 was upregulated in MCF10AT xenografts and that these tumors shared many features with untreated CIS (So et al., 2014). Also, Simpson and colleagues demonstrated that higher grades of CIS did not show further CCND1 upregulation (Simpson et al., 1997). This brings the applicability of this setup using MCF10AT into question. However, further histological analysis similar to what Sadlonova and colleagues conducted is required to assess whether i) CCND1 expression correlates with tumor size, ii) the treatment induced tumors show traits of higher CIS stages or iii) there are more invasive structures (Sadlonova et al., 2007).

Using MCF10CA1 cells treated with siESPL1, siTOP2A or siScrambled in xenografts we detected a slight increase in size for siESPL1 and siTOP2A (supplementary figure 3). The increased size for siESPL1 was not significant, though. Importantly, the number of mice was low and the effects were small. It remains unclear if the lacking significance is a result of missing marker upregulation, additional aneuploidy or other effects.

4.3 *The effects of aneuploidy on the genome and its regulation*

In 2006, Carter and colleagues published a gene signature containing mostly genes whose deregulation correlates with aneuploidy (Carter et al., 2006). They were able to better predict clinical outcome of six cancer entities and reliably stratify breast cancers into grade 1 and 2. Their findings do not only demonstrate the influence of aneuploidy on cancer but confirm that aneuploidy plays a critical role in gene regulation. The mechanisms behind that were discussed to be caused by i) the DNA content itself changes expression and the mitotic competence of a cell, ii) a loss of the ability to maintain the “integrity of genetic information on the chromosomal level” and iii) an override of checkpoints during mitosis caused by the genes from their signature.

Beach and colleagues showed that aneuploid cells retain the variability of the mother population, and cells that shared the same aneuploidies could have drastically different phenotypes (Beach et al., 2017). This highlights that the genetic setup follows the regulation of epigenetics. Epigenetic regulation over chromatin-chromatin interactions make up one of the major epigenetic tools the cell has to regulate gene expression (Lieberman-aiden et al., 2009). As opposed to methylation/acetylation of the DNA or histones, chromatin-chromatin interactions are, at least theoretically, susceptible against changes of the chromatin content. This is supported by the fact that aneuploidy and changes of the expression profile do not always correlate.

By looking at different chromatin interactions of the promoter regions of genes that were upregulated in the siESPL1 and siTOP2A expression profiles (Info box 2) using 4C-seq, we could see a reduced number of interactions for the 3D conditions when compared to the 2D conditions. Remarkably, CCND1 showed a strong negative correlation of expression and the number of interactions (figure 19).

A central question of this work is how CCND1 could be upregulated upon aneuploidization, despite being copy number insensitive. As Whalen and colleagues showed, enhancer-promoter interactions independent of genomic distance are important for gene expression (Whalen et al., 2016). The detection of several potential CCND1 promoter interacting enhancer regions that had fewer interactions with the CCND1 promoter is an important argument that CCND1 is regulated by several enhancers forming

inhibitory loops. One of these inhibitory enhancer interactions was described by Dalvai and colleagues. They showed that the enhancer enh2 interacts with the CCND1 promoter over H2A.Z. Upon acetylation of H2A.Z by the Tip60 acetyltransferase, the H2A.Z-mediated inhibitory loop is released, which activates CCND1 for transcription (Dalvai et al., 2012). The fact that the upregulation of CCND1 in 2D was not detectable could be caused by i) the missing niche that cells growing in 3D have to form in order to develop into aneuploid acini and divide enough to bring the conformational changes into effect or ii) additional unknown cues in the 3D environment that enable the cells to alter chromatin conformation. Another interesting finding is that many of the interaction changes occurred in the far-cis regions, which also exhibit the fewest interactions (figure 20). The biological meaning of the low interacting regions for CCND1 expression needs to be further analyzed.

Our results suggest that either additional chromosomes reduce the accessibility of the chromosomal regions, or interactions of different chromosomes play a crucial role in aneuploidy driven CCND1 upregulation. The first argument would result in broader changes and would be less consistent. This contradicts our findings of similar changes between siESPL1 and siTOP2A and over different viewpoints. Different chromosomes interact with each other using CTCF motifs (Botta et al., 2010) but can seemingly also mediate intra-chromosomal interactions. Intra-chromosomal interactions would not change much with additional copies of other chromosomes if all of them would only self-interact or they are not modified by other processes. However, this is exactly what we see, a change of interactions based on increased chromatin content.

Our findings might be a good explanation for the deregulation of tumor drivers whose overexpression could not be correlated with copy number changes. However, the mechanism how aneuploidy confers the interaction changes in detail remains unclear. One approach to explain this might be that genes with specific enhancer landscapes around them are more susceptible against alterations of chromatin interactions upon chromosome gains or losses. This is supported by the fact that there are several genes in the expression profiles of siESPL1 and siTOP2A that share either of the transcription factor binding motifs of TCF3, PAX4, ETS2 and MAZ. All of these transcription factors have also been shown to play a role in CCND1 regulation (Hovanec et al., 2001; Mellado-gil et al., 2016; Guo et al., 2011, TRANSFAC

database for transcription factors, Broad institute). This is a hint that certain enhancer setups might predispose genes to be deregulated upon chromosome number changes. Another idea would be that the transcription factors themselves are deregulated and cause the abnormal expression. However, the transcription factors did not show altered levels in our expression profiles.

The implications of aneuploidy driven reshaping of the enhancer-landscape resulting in the deregulation of cancer relevant genes are immense. It implicates that aneuploidy would not only be a tool for diversification and subsequent cancer evolution but also be capable of actively driving further steps towards tumorigenicity. Additionally, this concept could strengthen the arguments made by Davoli and colleagues (Davoli et al., 2013) because this could explain why some genes are considered triplosensitive or haploinsufficient. In short, genes sharing either inhibitory elements or transcription factor binding sites similar to CCND1 are deregulated by the altered interactions rather than the copy number changes. In succession, they would appear to be triplosensitive because additional chromosomes would change the intra-chromosomal interaction landscape and increase the gene expression similar to CCND1. In the case of haploinsufficiency it is to be determined, whether chromosome number changes also result in reduced interactions in 3D. If this is the case, then classical enhancer function that works by assembling proteins to the promoter region of the target gene to enhance the expression might be impaired and result in downregulation of the gene.

Lastly, ESPL1 and TOP2A cooperate during sister chromatid separation, act together during chromatin condensation and potentially also during chromatin conformation processes. CTCF and cohesin are the most important proteins mediating chromatin conformation (Uhlmann, 2016). Chromatin interactions are stabilized by cohesin which is broken down by separase (Petronczki et al., 2003). TOP2B which is an isoform of TOP2A mediates DNA breaks that correlate with cohesin (Canela et al., 2017). Taken together this potential cooperation of cohesion/separase/topoisomerase and its disruption affecting chromatin interactions might be the reason for the observed results instead of the aneuploidy (Uhlmann, 2016).

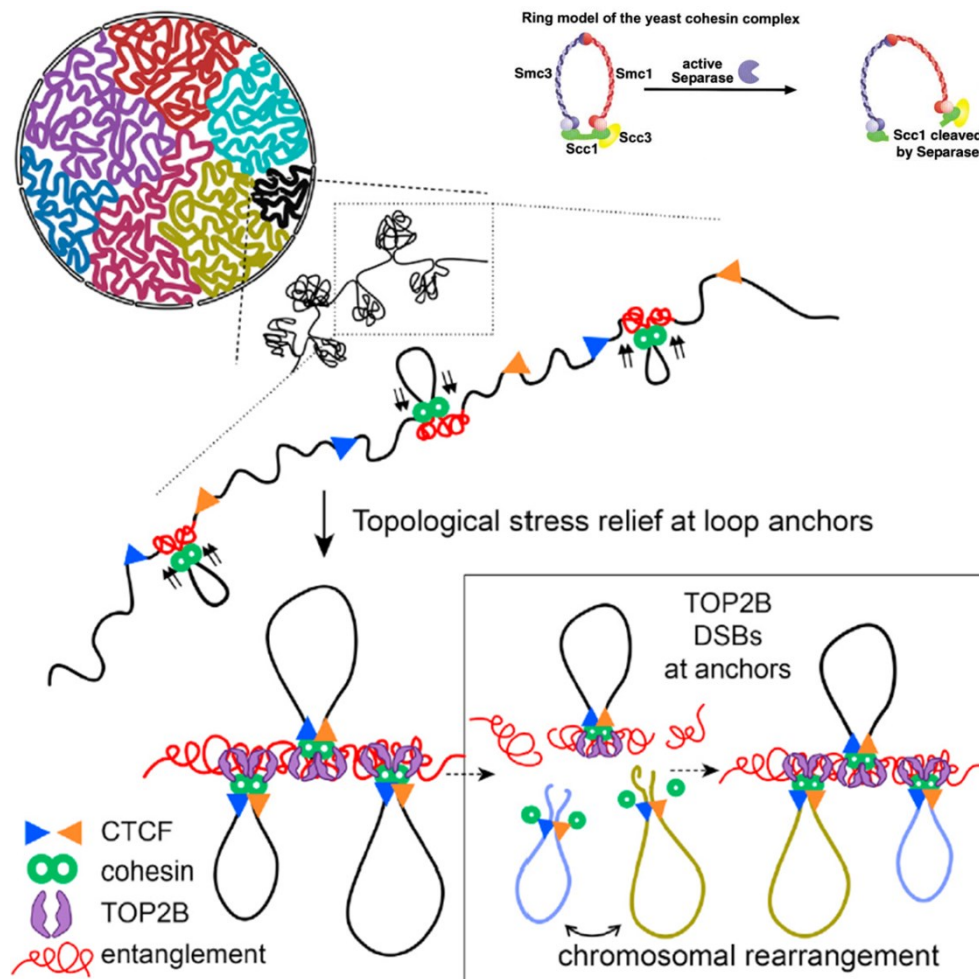


Figure 25: TOP2B and separate are involved in chromatin interaction rearrangements (Canela et al., 2017 and Petronczki et al., 2003, modified).

4.4 Conclusion

We can show that introduction of mitotic errors by downregulation of ESPL1 and TOP2A first lead to accumulation of genetic material (figure 10). In 3D, we can measure that this reliably triggers the upregulation of CCND1 (figure 15 and 16) by altering the chromatin conformation of CCND1 associated enhancers (figure 21). The upregulation of CCND1 leads to altered cell cycle progression and potentially increases the chance for abnormal cells to further develop into early forms of CIS (figure 26). This is supported by So and colleagues, who show that CCND1 is upregulated in low grade CIS (So et al., 2014). In further developmental steps it is progressively downregulated with increasing grade (Simpson et al., 1997). The upregulation of CCND1 in invasive breast cancers might then again further drive cancer development by keeping aneuploid cells dividing (figure 26).

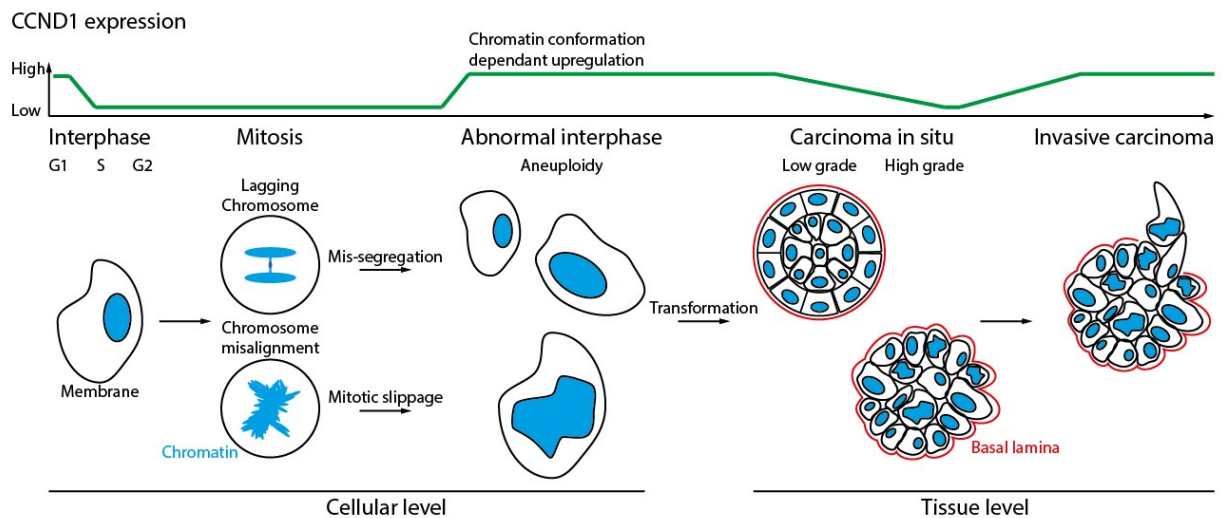


Figure 26: Proposed model of aneuploidy driven CCND1 upregulation and resulting development of CIS.

5 Outlook

This project centered on the effects of tumor drivers whose deregulation was correlated to CIN. In the initial HCS, we transiently downregulated genes to mimic a loss of function which resulted in abnormal mitosis and aneuploidy. In cancer, the upregulation of tumor drivers is seen just as often and might be an alternative way of testing the model resulting from this project. In this context, the effect of CCND1 upregulation would be especially interesting. To be able to better generalize the results, other cell models could be analyzed.

One remaining question is whether aneuploid cells alone in the acini show the altered expression or all cells within the acini deregulate the tumor markers. This could be tested using single cell expression profiles. The immunofluorescence staining suggests the latter alternative. If so, how do the aneuploid cells confer the expression to the other cells?

Although we tested for genomic alteration, we did not test if the induction of aneuploidy also caused smaller alterations that were not visible in MFISH. At least in theory, it is possible that the changes were caused by mutations that are only detectable by sequencing.

MCF10A cells that were induced to be aneuploid did not show increased tumorigenic potential in xenografts. There are several questions that arise in this context. What is necessary to transform MCF10A cells into a physiologically relevant xenograft model?

MCF10AT cells, which overexpress activated HRAS, seem like an overkill method with limited clinical applicability. Would TP53 downregulation in addition to the aneuploidy caused by ESPL1 and TOP2A downregulation be enough? The initial screen gives the impression that additional TP53 downregulation might not be enough. However, it is not clear if the lack of effects in the dKD HCS are a result of wrong dosage or timing.

To get a better understanding of how aneuploidy shapes the chromatin interactions, chromatin conformation capture technologies at single cell resolution (Hi-C, 4-C) might shed light on the effects of random aneuploidy or copy number alteration of single chromosomes and how consistent the changes are across cells of a single tumor. Here, a method could be utilized that leads to a selective miss-segregation of single chromosomes shown by Beach and colleagues (Beach et al., 2017). In yeast, the substitution of the centromeric region with a conditional centromere that can express GAL1 during mitosis stops kinetochore formation and proper segregation. This approach might be adaptable for mammalian cells in the future.

6 *Literature*

- Abramson, VG. & Mayer, IA., 2014. Molecular Heterogeneity of Triple Negative Breast Cancer. *Curr Breast Cancer Rep*, 6, pp.154–158.
- American Cancer Society, Cancer Facts & Figures 2016.
- Armaghany, T. et al., 2012. Genetic Alterations in Colorectal Cancer. *Gastrointest Cancer Research*, 5, pp.19–27.
- Bae, SY. et al, 2015. Poor prognosis of single hormone receptor- positive breast cancer : similar outcome as triple-negative breast cancer. *BMC Cancer*, 138, pp.1–9.
- Beach, RR. et al., Aneuploidy Causes Non-genetic Individuality Article Aneuploidy Causes Non-genetic Individuality. *Cell*, 169, p.229–242.
- Ben-david, U. et al., 2014. Aneuploidy induces profound changes in gene expression, proliferation and tumorigenicity of human pluripotent stem cells. *Nature Communications*, 5, pp.1–11.
- Berthold, MR. et al., 2007. KNIME – The Konstanz Information Miner. *Springer*, 11, pp.26–31.
- Bissell, MJ., Hines, WC. & Berenblum, I., 2011. Why don ' t we get more cancer ? A proposed role of the microenvironment in restraining cancer progression. *Nature medicine*, 17, pp.320–329.
- Botta, M. et al., 2010. Intra- and inter-chromosomal interactions correlate with CTCF binding genome wide. *Molecular Systems Biology*, 6, pp.1–6.
- Bouchet, BP. et al., 2007. Paclitaxel resistance in untransformed human mammary epithelial cells is associated with an aneuploidy-prone phenotype. *British Journal of Cancer*, 97, pp.1218–1224.
- Boveri, T., 2008. Concerning the Origin of Malignant Tumours by Theodor Boveri. Translated and annotated by Henry Harris. *Journal of Cell Science*, 121, pp.1–84.

- Bresnick, AR., Weber, DJ. & Zimmer, DB., 2015. S100 proteins in cancer. *Nature Reviews. Cancer*, 15, pp.96–109.
- Britschgi, A. et al., 2017. The Hippo kinases LATS1 and 2 control human breast cell fate via crosstalk with ER α . *Nature*, 541, pp.541–545.
- Brocker, C. et al., 2010. Evolutionary divergence and functions of the human interleukin (IL) gene family. *Human genomics*, 5, pp.30–55.
- Burstein, HJ., 2005. The Distinctive Nature of HER2-Positive Breast Cancers. *The New England journal of medicine*, 353, pp.1652–1654.
- Dutruef, C. et al., 2014. Early epigenetic downregulation of WNK2 kinase during pancreatic ductal adenocarcinoma development. *Oncogene*, 33, pp.3401–3410.
- Canela, A. et al., 2017. Article Genome Organization Drives Chromosome Fragility. *Cell*, 170, pp.507-521.
- Carraro, DM., Elias, EV & Victor, P., 2014. Ductal carcinoma in situ of the breast : morphological and molecular features implicated in progression. *Bioscience Reports*, 34, pp.19–28.
- Carter, SL. et al., 2006. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nature genetics*, 38, pp.1043–1048.
- Castro, NP. et al., 2008. Evidence that molecular changes in cells occur before morphological alterations during the progression of breast ductal carcinoma. *Breast cancer research*, 10, pp.1–14.
- Chang, CQ. et al., 2013. A systematic review of cancer GWAS and candidate gene meta-analyses reveals limited overlap but similar effect sizes. *European Journal of Human Genetics*, 22, pp.402–408.

- Chen, T. et al., 2015. Topoisomerase II α in chromosome instability and personalized cancer therapy. *Oncogene*, 34, pp.4019–31.
- Cheung, KJ. et al., 2016. Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proceedings of the National Academy of Sciences*, 113, p.201508541.
- Cimini, D., 2008. Merotelic kinetochore orientation, aneuploidy, and cancer. *Biochimica et Biophysica Acta - Reviews on Cancer*, 1786, pp.32–40.
- Clemente-Ruiz, M. et al., 2016. Gene Dosage Imbalance Contributes to Chromosomal Instability-Induced Tumorigenesis. *Developmental Cell*, 36, pp.290–302.
- Dalvai, M. et al., 2012. H2A . Z-dependent crosstalk between enhancer and promoter regulates Cyclin D1 expression. *Oncogene*, 32, pp.4243–4251.
- Davoli, T. et al., 2013. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell*, 155, pp.948–62.
- Dawson, P.J. et al., 1996. MCF10AT: a model for the evolution of cancer from proliferative breast disease. *The American journal of pathology*, 148, pp.313–319.
- Debnath, J., Muthuswamy, SK. & Brugge, JS., 2003. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods*, 30, pp.256–268.
- Dekker, J. et al., 2002. Capturing Chromosome Conformation. *Science*, 295, pp.1306–1311.
- Dodgson, SE. et al., 2016. The pleiotropic deubiquitinase Ubp3 confers aneuploidy tolerance. *Genes & development*, 30, pp.2259–2271.

- Drost, J. et al., 2015. Sequential cancer mutations in cultured human intestinal stem cells. *Nature*, 521, pp.43-47.
- Duan, JJ. et al., 2016. ALDH1A3, a metabolic target for cancer diagnosis and therapy. *International journal of cancer*, 139, pp.965–75.
- Duncan, AW. et al., 2010. The ploidy conveyor of mature hepatocytes as a source of genetic variation. *Nature*, 467, pp.707–10.
- Elbashir, SM., Lendeckel, W. & Tuschl, T., 2001. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & development*, 15, pp.188–200.
- Falandry, C. et al., 2010. CLLD8/KMT1F is a lysine methyltransferase that is important for chromosome segregation. *Journal of Biological Chemistry*, 285, pp.20234–20241.
- Fang, X. et al., 2011. Twist2 contributes to breast cancer progression by promoting an epithelial – mesenchymal transition and cancer stem-like cell self-renewal. *Oncogene*, 30, pp.4707–4720.
- Ferlay, J. et al., 2013. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012. *European Journal of Cancer*, 49, pp.1374–1403.
- Fernandez-capetillo, O. et al., 2003. DNA damage-induced G 2 – M checkpoint activation by histone H2AX and 53BP1. *Nature Cell Biology*, 4, pp.993–998.
- Forbes, SA. et al., 2017. COSMIC : somatic cancer genetics at high-resolution. *Nucleic acids research*, 45, pp.777–783.
- Foulkes, Smith, RF., 2010. Triple-negative breast cancer. *The New England journal of medicine review*, 363, pp.1938–1948.
- Gao, R. et al., 2016. Articles Punctuated copy number evolution and clonal stasis in triple-negative

- breast cancer. *Nature genetics*, 48, pp.1119-1130.
- Geigl, JB., Uhrig, S. & Speicher, MR., 2006. Multiplex-fluorescence in situ hybridization for chromosome karyotyping. *Nature protocols*, 1, pp.1172-1184.
- Giam, M. & Rancati, G., 2015. Aneuploidy and chromosomal instability in cancer : a jackpot to chaos. *Cell division*, 10, pp.1–12.
- Gordon, DJ., Resio, B. & Pellman, D., 2012. Causes and consequences of aneuploidy in cancer. *Nature reviews. Genetics*, 13, pp.189–203.
- Gorre, ME. et al., 2001. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science*, 293, pp.876–880.
- Guo, Y. et al., 2015. Article CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer / Promoter Function Article CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer / Promoter Function. *Cell*, 162, pp.900–910.
- Guo, Z. et al., 2011. The elements of human cyclin D1 promoter and regulation involved. *Clinical epigenetics* ,2 , pp.63–76.
- Hammond, MEH. et al., 2010. American society of clinical oncology/college of American pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Archives of Pathology and Laboratory Medicine*, 134, pp.e48–e72.
- Hanahan, D. & Weinberg, R., 2011. Hallmarks of cancer: the next generation. *Cell*, 144, pp.646–74.
- Hannemann, J. et al., 2006. Classification of ductal carcinoma in situ by gene expression profiling. *Breast cancer research*, 8, pp.1–20.

- Harvey, KF., Zhang, X. & Thomas, DM., 2013. The Hippo pathway and human cancer. *Nature review.Cancer*, 13, pp.246–257.
- Heppner, GH. & Wolman, SR., 1999. MCF-10AT: A Model for Human Breast Cancer Development. *The breast journal*, 5, pp.122–129.
- Hoffmann, E. et al., 2008. Transcriptional regulation of EGR-1 by the interleukin-1-JNK-MKK7-c-Jun pathway. *Journal of Biological Chemistry*, 283, pp.12120–12128.
- Holohan, C. et al., 2013. Cancer drug resistance: an evolving paradigm. *Nature reviews. Cancer*, 13, pp.714–726.
- Hovanes K. et al., 2001. Beta-catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer. *Nature genetics*, 28, pp.53–57.
- Hughes, CS., Postovit, LM. & Lajoie, GA., 2010. Matrigel: a complex protein mixture required for optimal growth of cell culture. *Proteomics*, 10, pp.1886–1890.
- Idowu, M.O. et al., 2012. CD44+/CD24-/low cancer stem/progenitor cells are more abundant in triple-negative invasive breast carcinoma phenotype and are associated with poor outcome. *Human Pathology*, 43, pp.364–373.
- Jonsson,G. et al., 2007. High-Resolution Genomic Profiles of Breast Cancer Cell Lines Assessed by Tiling BAC Array Comparative Genomic Hybridization. *Genes, chromosomes & cancer*, 46, pp.543–558.
- Kabel, A.M., 2017. Tumor markers of breast cancer : New prospectives. *Journal of Oncological Science*, 3, pp.5–11.
- Karantza, V., 2011. Keratins in health and cancer: more than mere epithelial cell markers. *Oncogene*, 30, pp.127–138.

- Kayl, AE. & Meyers, CA., 2006. Side-effects of chemotherapy and quality of life in ovarian and breast cancer patients. *Current opinion in obstetrics & gynecology*, 18, pp.24–28.
- Keegan, TH. et al., 2013. Impact of breast cancer subtypes on 3-year survival among adolescent and young adult women. *Breast Cancer Res*, 15, p.R95.
- Klein, FA. et al., 2015. FourCSeq : analysis of 4C sequencing data. *Bioinformatics*, 31, pp.3085–3091.
- Klinge, CM., 2001. Estrogen receptor interaction with estrogen response elements. *Nucleic acids research*, 29, pp.2905–2919.
- Knouse, KA. et al., 2014. Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. *Proceedings of the National Academy of Sciences*, 111, pp.13409–13414.
- Korge, G., 1975. Chromosome puff activity and protein synthesis in larval salivary glands of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, 72, pp.4550–4554.
- Landry, JJM. et al., 2013. The Genomic and Transcriptomic Landscape of a HeLa Cell Line. *G3 Genes, Genomes, Genetics*, 3, pp. 1213–1224.
- Lawrence, RT. et al., 2015. The Proteomic Landscape of Triple-Negative Breast Cancer. *Cell Reports*, 11, pp.630–644.
- Lee, J., Smith, E. & Shilatifard, A., 2010. Minireview The Language of Histone Crosstalk. *Cell*, 142, pp.682–685.
- Lee, S. et al., 2012. Differentially Expressed Genes Regulating the Progression of Ductal Carcinoma In Situ to Invasive Breast Cancer. *Cancer Research*, 72, pp.4574–4586.
- Lieberman-aiden, E. et al., 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 326, pp.289–294.
- Liedtke, C. et al., 2008. Response to Neoadjuvant Therapy and Long-Term Survival in Patients With Triple-Negative Breast Cancer. *Journal of clinical oncology*, 26, pp.1275–1281.

- Liu, G. et al., 2009. Genomic amplification and oncogenic properties of the GASC1 histone demethylase gene in breast cancer. *Oncogene*, 28, pp.4491–4500.
- Liu, Y. et al., 2015. ALDH1A1 mRNA expression in association with prognosis of triple-negative breast cancer. *Oncotarget*, 6, pp.41360–9.
- Liu, Y. et al., 2014. Lack of correlation of stem cell markers in breast cancer stem cells. *British journal of cancer*, 110, pp.2063–71.
- Lukinavičius, G. et al., 2014. Fluorogenic probes for live-cell imaging of the cytoskeleton. *Nature Methods*, 11, pp.731–3.
- Ma, X. et al., 2003. Gene expression profiles of human breast cancer progression. *Proceedings of the National Academy of Sciences*, 100, pp.5974-5979.
- Mack SC et al., 2014. Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature*, 506, pp.445–450.
- Marcato, P. et al., 2011. Aldehyde Dehydrogenase Activity of Breast Cancer Stem Cells is Primarily Due to Isoform ALDH1A3 and Its Expression is Predictive of Metastasis. *STEM CELLS*, 29, pp.32–45.
- Marella, NV. et al., 2009. Cytogenetic and cDNA microarray expression analysis of MCF10 human breast cancer progression cell lines. *Cancer research*, 69, pp.5946–53.
- Marusyk, A., Almendro, V. & Polyak, K., 2012. Intra-tumour heterogeneity: a looking glass for cancer? *Nature reviews. Cancer*, 12, pp.323–34.
- Medema, J.P., 2013. Cancer stem cells: The challenges ahead. *Nature Cell Biology*, 15, pp.338–344.
- Mellado-gil, JM. et al., 2016. PAX4 preserves endoplasmic reticulum integrity preventing beta cell

- degeneration in a mouse model of type 1 diabetes mellitus. *Diabetologia*, 59, pp.755–765.
- Merrick, DT. et al., 2006. Analysis of c-ErbB1 / Epidermal Growth Factor Receptor and c-ErbB2 / HER-2 Expression in Bronchial Dysplasia : Evaluation of Potential T argets for Chemoprevention of Lung Cancer. *Clinical cancer reserach*, 2, pp.2281–2289.
- Miron, K. et al., 2015. Oncogenes create a unique landscape of fragile sites. *Nature communications*, 6, p.7094.
- Moasser, MM., 2011. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene*, 26, pp.6469–6487.
- Morley, S. et al., 2015. Regulation of microtubule dynamics by DIAPH3 influences amoeboid tumor cell mechanics and sensitivity to taxanes. *Scientific Reports*, 5, pp.1–16.
- Morrow, M., Schnitt, SJ. & Norton, L., 2015. Current management of lesions associated with an increased risk of breast cancer. *Nature reviews. Clinical oncology*, 12, pp.227–238.
- Mukherjee, M. et al., 2014. MMTV-Espl1 Transgenic Mice Develop Aneuploid, Estrogen Receptor Alpha (ER α)-Positive Mammary Adenocarcinomas. *Oncogene*, 33, pp.5511–5522.
- Mukherjee, M. et al., 2011. Separase Loss of Function Cooperates with the Loss of p53 in the Initiation and Progression of T- and B-Cell Lymphoma , Leukemia and Aneuploidy in Mice. *PLoS One*, 6, e22167.
- Munz, M., Baeuerle, PA. & Gires, O., 2009. The emerging role of EpCAM in cancer and stem cell signaling. *Cancer Research*, 69, pp.5627–5629.
- Musgrove, EA. et al., 2011. Cyclin D as a therapeutic target in cancer. *Nature reviews. Cancer*, 11, pp.558–572.

- Neve, RM. et al., 2006. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer cell*, 10, pp.515–527.
- Nevo, J. et al., 2009. Mammary-derived growth inhibitor alters traffic of EGFR and induces a novel form of cetuximab resistance. *Clinical Cancer Research*, 15, pp.6570–6581.
- Nik-zainal, S. et al., 2016. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, 534, pp.47–54.
- Nowak, M.A. et al., 2002. The role of chromosomal instability in tumor initiation. *Proceedings of the National Academy of Sciences*, 99, pp.16226-16231.
- O'Malley, FP., 2010. Lobular neoplasia : morphology , biological potential and management in core biopsies. *Modern pathology*, pp.14–25.
- Ohashi, A. et al., 2015. Aneuploidy generates proteotoxic stress and DNA damage concurrently with p53-mediated post-mitotic apoptosis in SAC-impaired cells. *Nature Communications*, 6, pp.1–16.
- Ong, C. & Corces, VG., 2014. CTCF: An Architectural Protein Bridging Genome Topology and Function. *Nature reviews. Genetics*, 15, pp.234–246.
- Otsu, N., 1997. A Threshold Selection Method from Gray-Level Histograms. *IEEE*, 4, pp.661-665.
- Ottesen, G.L. et al., 1995. DNA aneuploidy in early breast cancer. *British journal of cancer*, 72, pp.832–9.
- Otto, T. & Sicinski, P., 2017. Cell cycle proteins as promising targets in cancer therapy. *Nature reviews. Cancer*, 17, pp.93–115.
- Palazzo, F. et al., 2001. mDia mediates Rho-regulated formation and orientation of stable

- microtubules. *Nature Cell Biology*, 3, pp.723–729.
- Pantschenko AG et al., 2003. The interleukin-1 family of cytokines and receptors in human breast cancer: Implications for tumor progression. *International Journal of Oncology*, 23, pp.269-284.
- Papi, M. et al., 2005. Multiple roles for separase auto-cleavage during the G2 / M transition. *Nature Cell Biology*, 7, pp.1029-1035.
- Parker, M. et al., 2014. NF- κ B signalling in ependymoma. *Nature*, 506, pp.451-455.
- Perou, CM. et al., 2000. Molecular portraits of human breast tumours. *Nature*, 406, pp.747–752.
- Petronczki, M. et al., 2003. Un ménage à quatre: the molecular biology of chromosome segregation in meiosis. *Cell*, 112, pp.423–440.
- Pino, MS. & Chung, DC., 2014. The chromosomal instability pathway in colon cancer. *Gastroenterology*, 138, pp.2059–2072.
- Prives, C. & Lowe, SW., 2015. Mutant p53 and chromatin regulation. *Nature*, 525, pp.199–200.
- Pylayeva-gupta, Y., Grabocka, E. & Bar-sagi, D., 2011. RAS oncogenes: weaving a tumorigenic web. *Nature reviews. Cancer*, 11, pp.761–774.
- Ray, A. & Ray, BK., 2015. Induction of Ras by SAF-1/MAZ through a feed-forward loop promotes angiogenesis in breast cancer. *Cancer medicine*, 4, pp.224–234.
- Reddy EP, Reynolds RK, Santos E, Barbacid M., 1982. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature*, 300, pp.149–152.
- Rehen, SK. et al., 2001. Chromosomal variation in neurons of the developing and adult mammalian

- nervous system. *Proceedings of the National Academy of Sciences*, 98, pp.13361–13366.
- Rogier Versteeg, 2014. Cancer: Tumours outside the mutation box. *Nature*, 506, pp.438–439.
- Russ, A. et al., 2012. Hugel1 and Hugel2 in Mammary Epithelial Cells: Polarity, Proliferation, and Differentiation. *PLoS ONE*, 7, pp.3–14.
- Sadaie, M. et al., 2015. Cell-based screen for altered nuclear phenotypes reveals senescence progression in polyploid cells after Aurora kinase B inhibition. *Molecular biology of the cell*, 26, pp.2971–85.
- Sadlonova, A. et al., 2007. Human Breast Fibroblasts Inhibit Growth of the MCF10AT Xenograft Model of Proliferative Breast Disease. *The American Journal of Pathology*, 170, pp.1064–1076.
- Santaguida, S. & Amon, A., 2015. Short - and long-term effects of chromosome mis-segregation and aneuploidy. *Nature reviews. Molecular cell biology*, 16, pp.473-485.
- Santner, SJ. et al., 2001. Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells. *Breast cancer research and treatment*, 65, pp.101–10.
- Selmecki, A., Forche, A. & Berman, J., 2006. Aneuploidy and Isochromosome Formation in Drug-Resistant *Candida albicans*. *Science*, 313, pp.367–370.
- Semizarov, D. et al., 2003. Specificity of short interfering RNA determined through gene expression signatures. *Proceedings of the National Academy of Sciences*, 100, pp.6347-6352.
- Senkus,E. et al., 2015. Primary Breast Cancer: ESMO Clinical Practice Guidelines. *Annals of oncology*, 26, pp.v8-v30.
- Shaw, FL. et al., 2012. A Detailed Mammosphere Assay Protocol for the Quantification of Breast Stem Cell Activity. *Journal of mammary gland biology and neoplasia*, 17, pp.111–117.

- Simonis, M. et al., 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nature genetics*, 38, pp.1348–1354.
- Simpson, JF. et al., 1997. Amplification of CCND1 and Expression of Its Protein Product , Cyclin D1 , in Ductal Carcinoma in Situ of the Breast. *The American journal of pathology*, 151, pp.161–168.
- Singh, JK. et al., 2013. Recent advances reveal IL-8 signaling as a potential key to targeting breast cancer stem cells. *Breast cancer research*, 15, pp.1–9.
- So, JY. et al., 2014. Differential Expression of Key Signaling Proteins in MCF10 Cell Lines, a Human Breast Cancer Progression Model. *Mol Cell Pharmacol*, 4, pp.31–40.
- Soltysik, K. & Czekaj, P., 2013. Membrane estrogen receptors - is it an alternative way of estrogen action? *Journal of physiology and pharmacology*, 64, pp.129–142.
- Sotillo, R. et al., 2010. Mad2-induced chromosome instability leads to lung tumour relapse after oncogene withdrawal. *Nature*, 464, pp.436–40.
- Soto, A.M. & Sonnenschein, C., 2012. Is systems biology a promising approach to resolve controversies in cancer research? *Cancer cell international*, 12, p.12.
- Soule, HD. et al., 1990. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer research*, 50, pp.6075–86.
- Stepanova, M., Lin, F. & Lin, VC., 2006. In silico modelling of hormone response elements. *BMC bioinformatics*, 46, pp.163-176.
- Tabin CJ, et al., 1982. Mechanism of activation of a human oncogene. *Nature*, 300, pp.143–149.
- Talens, RP. et al., 2017. Variation , patterns, and temporal stability of DNA methylation : considerations for epigenetic epidemiology. *FASEB journal*, 24, pp.3135–3144.

- Tammela, T. et al., 2017. potential and progression in lung adenocarcinoma. *Nature*, 545, pp.355–359.
- Trimboli, AJ. et al., 2009. Pten in stromal fibroblasts suppresses mammary epithelial tumours. *Nature*, 461, pp.1084–91.
- Uhlmann, F., 2016. SMC complexes : from DNA to chromosomes. *Nature reviews. Molecular cell biology*, 17, 399-412.
- Valde´s-Mora, F. et al., 2012. Acetylation of H2A . Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. *Genome Research*, 22, pp.307–321.
- Van de Werken, HJ. et al., 2012. 4C Technology : Protocols and Data Analysis. *Methods in enzymology*, 513, pp.89-112.
- Van Wely, KH. & Martinez, AC., 2012. Linking stem cells to chromosomal instability. *Oncolmmunology*, 1, pp.195–200.
- Vandiver, AR. et al., 2015. DNA methylation is stable during replication and cell cycle arrest. *Scientific reports*, 5, pp.1–8.
- Voduc, KD. et al., 2017. Breast Cancer Subtypes and the Risk of Local and Regional Relapse. *Journal of clinical oncology*, 28, 1684-1691.
- Vogelstein, B. et al., 2013. Cancer genome landscapes. *Science*, 339, pp.1546–58.
- Wang, H. et al., 2012. Rotational motion during three-dimensional morphogenesis of mammary epithelial acini relates to laminin matrix assembly. *Proceedings of the National Academy of Sciences*, 110.
- Wang, H. et al., 2013. Rotational motion during three-dimensional morphogenesis of mammary

- epithelial acini relates to laminin matrix assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 110, pp.163–8.
- Wang, Y. et al., 2014. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*, 512, pp.1–15.
- Whalen, S. et al., 2016. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nature genetics*, 48, pp.488–496.
- Wilson, MA. et al., 2016. Copy number changes are associated with response to treatment with carboplatin, paclitaxel, and sorafenib in melanoma. *Clinical Cancer Research*, 22, pp.374–382.
- Yersal, O. & Barutca, S., 2014. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World Journal of Clinical Oncology*, 5, pp.412–425.
- Yi, F. et al., 2011. Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. *Nature Cell Biology*, 13, pp.762–70.
- Yoon, DS. et al., 2002. Variable Levels of Chromosomal Instability and Mitotic Spindle Checkpoint Defects in Breast Cancer. *The American journal of pathology*, 161, pp.391–397.
- Yu, J. et al., 2017. REC8 functions as a tumor suppressor and is epigenetically downregulated in gastric cancer , especially in EBV-positive subtype. *Oncogene*, 36, pp.182–193.
- Zhang, J. et al., 2015. Transcriptional control of PAX4-regulated miR-144/451 modulates metastasis by suppressing ADAMs expression. *Oncogene*, 34, pp.3283–3295.
- Zhang, R. et al., 2013. Gene expression analysis of induced pluripotent stem cells from aneuploid chromosomal syndromes. *BMC Genomics*, 14, p.S8.
- Zhou, ZN. et al., 2014. Autocrine HBEGF expression promotes breast cancer intravasation, metastasis

and macrophage-independent invasion in vivo. *Oncogene*, 33, pp.3784–3793.

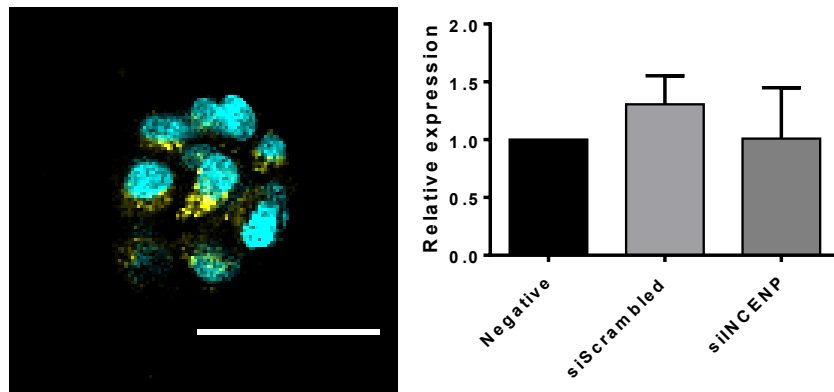
Zoldan, J. et al., 2011. Directing Human Embryonic Stem Cell Differentiation by NonViral Delivery of siRNA in 3D Culture. *Biomaterials*, 32, pp.7793–7800.

7 *Supplementary Data*

7.1 *Supplementary results*

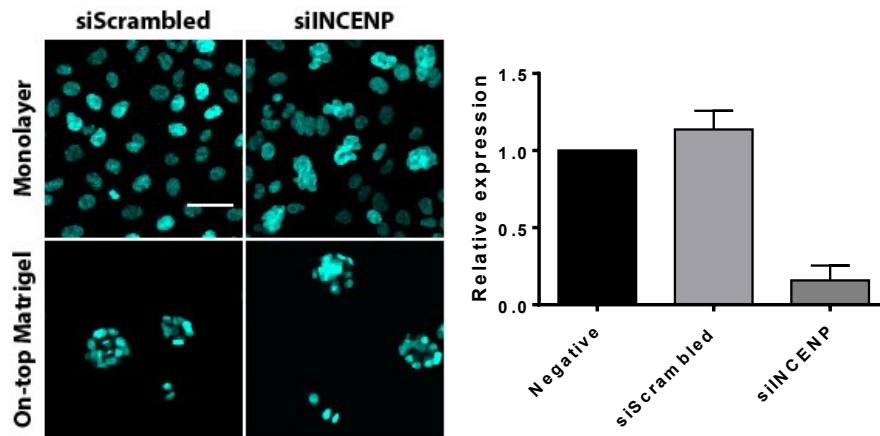
Knockdown protocol development

To test the downregulation of an example candidate in the 3D environment we transfected MCF10A cells growing on Matrigel with mCherry-INCENP siRNA after standard protocol and could detect the fluorescence inside the cells one day later. However, induction of morphological phenotypes characteristic for INCENP downregulation like polylobed shapes was absent. QPCR analysis also showed no downregulation of INCENP expression (Supplementary results figure 1).



Supplementary results figure 1: Cellular localization of mCherry reporter siRNA and QPCR quantified INCENP expression level after transfection of MCF10A cells with siINCENP growing on Matrigel. Cyan: H2B-GFP, yellow: mCherry reporter siRNA, the bar represents a length of 50 μ m.

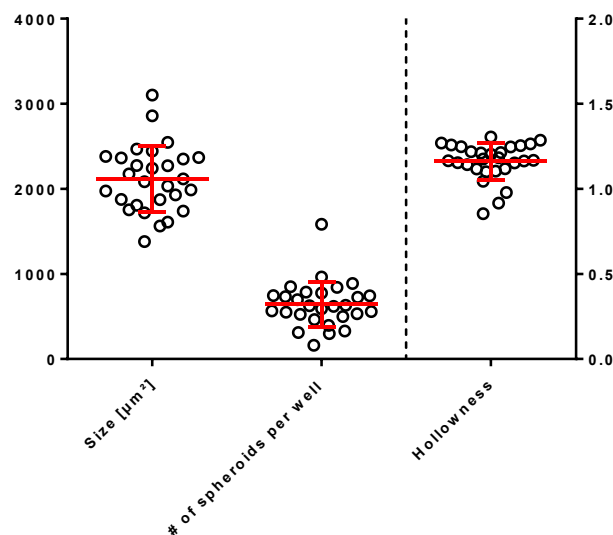
Attempting to improve the effects of the knockdown we transfected cells as monolayer and subsequently reseeded them into monolayer and 3D culture. A high abundance of morphologically abnormal nuclei was visible in monolayer culture. When cultured on Matrigel there were not only single polylobed nuclei in the well but also acini consisting of or at least containing polylobed nuclei. The QPCR analysis shows a strong downregulation of INCENP RNA of \sim 85% (Supplementary results figure 2) after the initial knockdown. Image analysis revealed a rate of $>$ 50 % polylobed nuclei in the whole population of cells in monolayer culture.



Supplementary results figure 2: Improved downregulation efficiency after a knockdown in monolayer culture with subsequent reseeding. Cyan: H2B-GFP, the bar (top left image) represents a length of 50 μm .

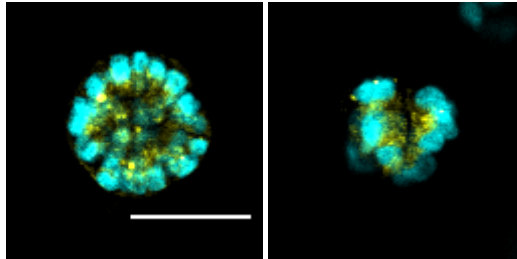
Screen consistency

Assessing the assays consistency all negative controls of all individual plates were compared with each other. All parameters of the negative controls appear in a similar dimension and have a small std. dev. (supplementary results figure 3).

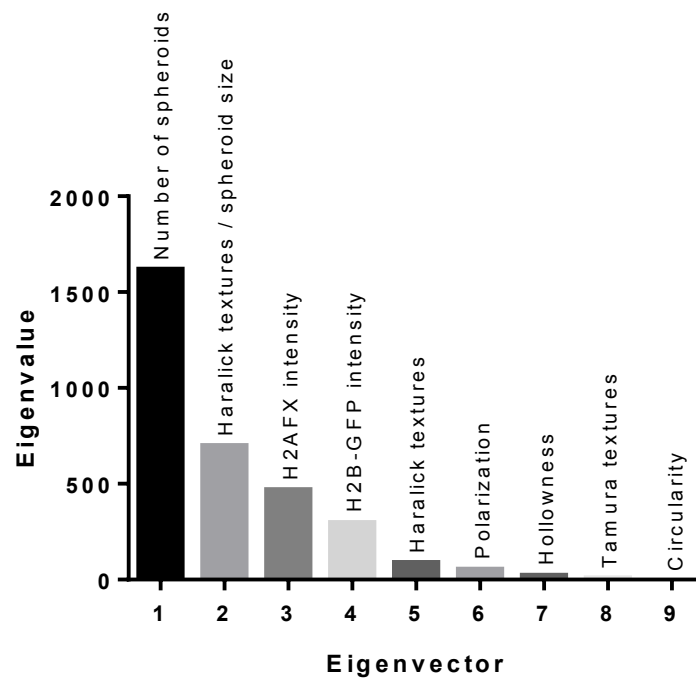


Supplementary results figure 3: Consistency assessment of the HCS comparing the negative controls of all 28 measured plates. Each point represents the mean of all replicates of the negative controls per well. The overall mean and standard deviation are represented by the red bar and the red error bars.

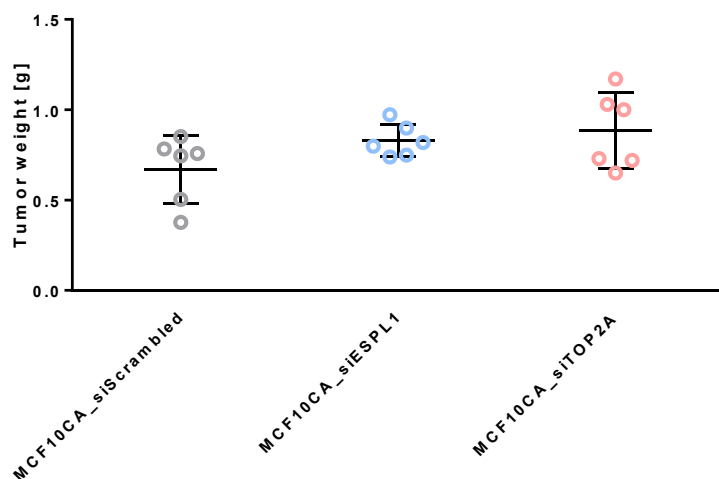
7.2 *Supplementary figures*



Supplementary figure 1: Morphology of a normal (left) and abnormal (right) spheroid after 21 days of culturing on Matrigel. Cyan: H2B-GFP, Yellow: Golgi and lysosomes. The bar represents a length of 50 μm .



Supplementary figure 2: Eigenvector distribution of the PCA of the nine screening parameters plotted against their respective eigenvalue and supplemented with their strongest correlating parameter.



Supplementary figure 3: Tumorigenicity of MCF10CA in after siScrambled, siESPL1 and siTOP2A

treatment.

7.3 Supplementary tables

Supplementary table 1: Screening candidates. The names and IDs are according to NCBI. The right column indicates if the measured 2D phenotype was in line (green), inconclusive (yellow) or not in line with the literature (red).

Gene	ID	Reference	# publications (august 2016)	Measured 2D phenotype in line with reported 2D phenotype
ACTA1	58	havaki 2007 cancer cell	346	Green
APC	324	thompson et al 2010 current biol, cosmic (most mutated genes)	710	Green
ARID1A	8289	cosmic (most mutated genes)	155	Yellow
ATM	472	pandita 2002 oncogene, cosmic (most mutated genes)	1027	Green
ATR	545	pino & chung 2010 gastroenterology	405	Yellow
AURKB	9212	carter et al 2006 nature	323	Green
BAZ2A	11176	parry and clarke 2011 genes and cancer	27	Green
BAZ2B	29994	parry and clarke 2011 genes and cancer	20	Green
BRCA1	672	thompson et al 2010 current biol, cosmic (most mutated genes)	2346	Green
BRCA2	675	thompson et al 2010 current biol	1408	Yellow
C20orf24	55969	carter et al 2006 nature	12	Green
CAMLG	819	thompson et al 2010 current biol	33	Green
CDC45	8318	chibon et al 2010 nature med	56	Green
CDC6	990	carter et al 2006 nature	119	Green

CDH1	999	cosmic (most mutated genes)	1509	
CDKN1B	1027	davoli et al 2013 cell	849	
CDKN2A	1029	davoli et al 2013 cell, cosmic (most mutated genes)	1930	
CENPE	1062	chibon et al 2010 nature med	74	
CENPF	1063	thompson et al 2010 current biol	73	
CHEK1	1111	carter et al 2006 nature	403	
CHFR	55743	perez de castro et al 2007 carcinogenesis	87	
CKAP2	26586	thompson et al 2010 current biol	39	
CMAS	55907	carter et al 2006 nature	25	
COPB2	9276	hutchins et al 2010 science	56	
CTPS1	1503	carter et al 2006 nature	49	
DIAPH3	81624	johansson 2013 plos one	38	
ECT2	1894	carter et al 2006 nature	79	
ESPL1	9700	carter et al 2006 nature	49	
FBXW7	55294	thompson et al 2010 current biol	259	
FHIT	2272	saldivar 2012 plos genet	306	
FOXM1	2305	carter et al 2006 nature	278	
GATA3	2625	cosmic most mutated genes	270	
H2AFX	3014	carter et al 2006 nature	337	
HAUS8	93323	thompson et al 2010 current biol	19	
INCENP	3619	neumann et al 2010 nature	66	
KIF2B	84643	thompson et al 2010 current biol	15	
KIF4A	24137	carter et al 2006 nature	57	
KMT2C	58508	cosmic (most mutated genes)	65	
LATS1	9113	sorrentino et al 2014 nat cell biol	94	
LATS2	26524	sorrentino et al 2014 nat cell biol	89	
LMNB2	84823	kuga et al 2014 oncogene	61	
LLGL1	3996	russ et al 2012 plos one	40	
LLGL2	3993	russ et al 2012 plos one	27	
MAD2L1	4085	carter et al 2006 nature	213	
MAP2K4	6416	carter et al 2006 nature, cosmic (most mutated genes)	165	
MAP3K1	4214	davoli et al 2013 cell	150	
MAPRE1	22919	thompson et al 2010 current biol	115	
MBD5	55777	parry and clarke 2011 genes and cancer	25	
MBD6	114785	parry and clarke 2011 genes and cancer	15	
MYH10	4628	overholzer 2007 cell	101	
MYO10	4651	overholzer 2007 cell	50	
NCAPH	23397	chibon et al 2010 nature med	32	
NEU1	4758	butler et al 2013 traffic	77	
NISCH	11188	johansson et al 2013 plos one	36	
NUP98	4928	perez de castro et al 2007 carcinogenesis	171	
PARD3	56288	wan et al 2013 mboc	94	
PLK4	10733	chibon et al 2010 nature med	75	

PRC1	9055	carter et al 2006 nature	58	
PTEN	5728	cosmic (most mutated genes)	1606	
RANBP1	5902	thompson et al 2010 current biol	71	
RB1	5925	thompson et al 2010 current biol, cosmic (most mutated genes)	1015	
RBMX	27316	davoli et al 2013 cell	101	
RHOA	387	overholzer 2007 cell	763	
ROCK1	6093	overholzer 2007 cell	318	
ROCK2	9475	overholzer 2007 cell	135	
SETDB2	83852	parry and clarke 2011 genes and cancer	13	
SGO2	151246	chibon et al 2010 nature med	26	
SMAD2	4087	petersen et al 2010 oncogene	459	
SPINK7	84651	thompson et al 2010 current biol	18	
TOP2A	7153	carter et al 2006 nature	363	
TP53	7157	thompson et al 2010 current biol, cosmic (most mutated genes)	7747	
TP63	8626	stefanou et al 2004 histol histopathol	642	
TPX2	22974	carter et al 2006 nature	93	
TRIP13	9319	carter et al 2006 nature	40	
TTN	7273	cosmic most mutated genes	237	
UHRF2	115426	parry and clarke 2011 genes and cancer	35	
VCL	7414	mierke et al 2010 jbc	172	
ZBTB4	57659	parry and clarke 2011 genes and cancer	13	
ZFP36L1	677	davoli et al 2013 cell	49	
ZSCAN22	342945	personal communication (Claudia Lukas, Uni Kopenhagen)	9	
ZWILCH	55055	perez de castro et al 2007 carcinogenesis	11	
ZWINT	11130	carter et al 2006 nature	38	

Supplementary table 2: Sirna library

HGNC symbol	siRNA ID #1	siRNA ID #2	siRNA ID #3
ACTA1	s941	s942	s943
APC	s1433	s1434	s1435
ARID1A	s15785	s15786	s15784
ATM	s1710	s1709	s1708
ATR	s536	s56826	s227305
AURKB	s17611	s17612	s17613
BAZ2A	s22058	s22056	s22057
BAZ2B	s26865	s26866	s26867
BRCA1	s458	s459	s457
BRCA2	s2085	s224694	s224695
C20orf24	s31821	n269281	n269282
CAMLG	s2371	s2372	s2370
CDC45	s15829	s15830	s15831

CDC6	s2744	s2745	s2746
CDH1	s2769	s2770	s2768
CDKN1B	s2837	s2838	s2839
CDKN2A	s216	s218	s217
CENPE	s2915	s2916	s2917
CENPF	s2918	s2919	s2920
CHEK1	s502	s503	s504
CHFR	s31392	s31393	s31394
CKAP2	s25563	s25564	s25565
CMAS	s31758	s31757	s31759
COPB1	s3371	s3372	s3373
CTPS1	s3731	s3732	s229529
DIAPH3	s37734	s37735	s37736
ECT2	s4444	s4445	s4446
ESPL1	s18686	s18687	s18688
FBXW7	s30663	s30664	s224357
FHIT	s5191	s5192	s5193
FOXM1	s5248	s5249	s5250
GATA3	s5600	s5601	s5599
H2AFX	s6412	s6413	s226270
HAUS8	s41118	s41119	s41120
INCENP	s7424	s7423	s7422
KIF2b	s39236	s39237	s39238
KIF4A	s24406	s24407	s24408
KMT2C	s33888	s33890	s33889
LATS1	s17392	s17393	s17392
LATS2	s25503	s25504	s25505
LLGL1	s8216	s8217	s8215
LLGL2	s8209	s8210	s8211
LMNB2	s39476	s39477	s39478
MAD2L1	s8392	s8393	s8391
MAP2K4	s12702	s12703	s12701
MAP3K1	s8667	s8668	s8669
MAPRE1	s22673	s22674	s22675
MBD5	s31485	s31486	s31487
MBD6	s41594	s41595	s41596
MYH10	s9170	s9171	s9169
MYO10	s9225	s9223	s9224
NCAPH	s23735	s225959	s225960
NEU1	s224109	s9455	s9454
NISCH	s22093	s22094	s223208
NUP98	s9782	s9783	s9784
PARD3	s32126	s32127	s32128
PLK4	s21085	s21084	s21083

PRC1	s17267	s17268	s17269
PTEN	s325	s326	s327
RANBP1	s11772	s11770	s11771
RB1	s523	s524	s522
RBMX	s26142	s26143	s223747
RHOA	s758	s759	s760
ROCK1	s12097	s12098	s12099
ROCK2	s18161	s18162	s18163
SETDB2	s38215	s38216	s38217
SGOL2	s45544	s45543	s45545
SMAD2	s8397	s8398	s8399
SPINK7	s39252	s39253	s39251
TAF4	s13733	s13734	s13735
TOP2a	s14308	s14309	s14307
TP53	s605	s606	s607
TP63	s16411	s229399	s229400
TPX2	s22745	s22746	s22747
TRIP13	s17805	s17806	s17807
TTN	s14484	s14485	s14486
UHRF2	s41816	s41817	s41818
VCL	s14762	s14763	s14764
ZBTB4	s33538	s33539	s33540
ZFP36L1	s2089	s2090	s2091
ZSCAN22	s50958	s50956	s50957
ZWILCH	s30072	s30073	s30074
ZWINT	s21949	s21950	s21951
SCRAMBL ED	s813	s814	

Supplementary table 3: List of antibodies

Experiments	Antibody	Species	Reactivity	Dilution
HCS immunofluorescence	Anti-GM130 (ABCAM, ab52649)	Rabbit	Cow, Dog, Human, Monkey	1:2 50
	Anti-γH2AFX (ABCAM, ab22551)	Mouse	Mouse, Rat, Cow, Human	1:2 50
BCSC-marker confirmation	Anti-CD24 (ThermoFisher, MA5-11833)	Mouse	Human	1:1 000
	Anti-CD44 (ThermoFisher, MA5-13890)	Mouse	Human	1:2 50

	Anti-ALDH1A3 (ThermoFisher, PA5-29188)	Ra bbit	Human	1:5 00
	Anti-CCND1 (ThermoFisher, AHF0082)	Mo use	Rat, Human	1:1 00
	Anti-EPCAM (ThermoFisher, MA5-12436)	Mo use	Human, Mouse	1:2 50
Secondary antibodies	Alexa-568 anti (ThermoFisher, a-21070)	Go at	Mouse	1:5 00
	Alexa-568 anti (ThermoFisher, a-21050)	Do nkey	Rabbit	1:5 00
	Alexa-633 anti (ThermoFisher, a-21052)	Go at	Mouse	1:5 00

Supplementary table 4: Chromosomal changes of 15 metaphases of untreated MCF10A cells using

MFISH. Green: gains, yellow: losses, blue: translocations, n: number of chromosomes, i: isochromosome, del: deletion, der: derivative, t: translocation.

Chromosome	image 1 (n=47)	image 2 (n=45)	image 3 (n=47)	image 4 (n=48)	image 5 (n=47)	image 6 (n=47)	image 7 (n=47)	image 8 (n=47)	image 9 (n=47)	image 10 (n=47)	image 11 (n=47)	image 12 (n=42)	image 13 (n=47)	image 14 (n=47)	image 15 (n=47)
1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1	1 x1 i(1q) x1 del(1q) x1
2	2 x2	2 x2	2 x2	2 x2	2 x2	2 x2	2 x2	2 x2	2 x2	2 x2	2 x2	2 x2	2 x2	2 x2	2 x2
3	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1	3 x1 der(3)t(3;9)) x1
4	4 x2	4 x2	4 x2	4 x2	4 x2	4 x2	4 x2	4 x2	4 x2	4 x2	4 x2	4 x2	4 x2	4 x2	4 x2
5	5 x2	5 x2	5 x2	5 x2	5 x2	5 x2	5 x2	5 x2	5 x2	5 x2	5 x2	5 x2	5 x2	5 x2	5 x2
6	6 x2	6 x2	6 x2	6 x2	6 x2	6 x2	6 x2	6 x2	6 x2	6 x2	6 x2	6 x1	6 x2	6 x2	6 x2
7	7 x2	7 x2	7 x2	7 x2	7 x2	7 x2	7 x2	7 x2	7 x2	7 x2	7 x2	7 x1	7 x2	7 x2	7 x2
8	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1	8 x1 i(8q)? x1
9	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1	9 x1 der(9)t(5;3)) x1
10	10 x2	10 x2	10 x2	10 x2	10 x2	10 x2	10 x2	10 x2	10 x2	10 x2	10 x2	10 x2	10 x2	10 x2	10 x2
11	11 x2	11 x2	11 x2	11 x2	11 x2	11 x2	11 x2	11 x2	11 x2	11 x2	11 x2	11 x2	11 x2	11 x2	11 x2
12	12 x2	12 x2	12 x2	12 x2	12 x2	12 x2	12 x2	12 x2	12 x2	12 x2	12 x2	12 x2	12 x2	12 x2	12 x2
13	13 x2	13 x2	13 x2	13 x2	13 x2	13 x2	13 x2	13 x2	13 x2	13 x2	13 x2	13 x2	13 x2	13 x2	13 x2
14	14 x2 14 x1	14 x2 14 x1	14 x2 14 x1	14 x2 14 x1	14 x2 14 x1	14 x2 14 x1	14 x2 14 x1	14 x2 14 x1	14 x2 14 x1	14 x2 14 x1	14 x2 14 x1	14 x2 14 x1	14 x2 14 x1	14 x2 14 x1	14 x1 t(14;22) x1
15	15 x2	15 x2	15 x2	15 x2	15 x2	15 x2	15 x2	15 x2	15 x2	15 x2	15 x2	15 x2	15 x2	15 x2	15 x2
16	16 x2	16 x2	16 x2	16 x2	16 x2	16 x2	16 x2	16 x2	16 x2	16 x2	16 x2	16 x1	16 x2	16 x2	16 x2
17	17 x2	17 x2	17 x2	17 x2	17 x2	17 x2	17 x2	17 x2	17 x2	17 x2	17 x2	17 x1	17 x2	17 x2	17 x2
18	18 x2	18 x2	18 x2	18 x2	18 x2	18 x2	18 x2	18 x2	18 x2	18 x2	18 x2	18 x2	18 x2	18 x2	18 x2
19	19 x2	19 x2	19 x2	19 x2	19 x2	19 x2	19 x2	19 x2	19 x2	19 x2	19 x2	19 x2	19 x2	19 x2	19 x2
20	20 x2	20 x2	20 x2	20 x2	20 x2	20 x2	20 x2	20 x2	20 x2	20 x2	20 x2	20 x2	20 x2	20 x2	20 x2
21	21 x2	21 x2	21 x2	21 x2	21 x2	21 x2	21 x2	21 x2	21 x2	21 x2	21 x2	21 x2	21 x2	21 x2	21 x2
22	22 x2	22 x2	22 x2	22 x2	22 x2	22 x2	22 x2	22 x2	22 x2	22 x2	22 x2	22 x2	22 x2	22 x2	22 x1 t(14;22) x1
X	X x2	X x2	X x2	X x2	X x2	X x2	X x2	X x2	X x2	X x2	X x2	X x2	X x2	X x2	X x2
Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Supplementary table 5: List of all potential screening candidates before “classification”

name	id	Literature	where from
ACTA1	58	- (overholzer 2007)	overholzer 2007 cell, havaki 2007 cancer cell
ACTL6A	86		carter, bao et al 2013 cell
AKT1	207	sasaki 2010 biochem biophy res commun	sasaki 2010 biochem biophy res commun
ANAPC 1	64682	-	ming liang he 2011 am j cancer research
ANLN	54443		carter, hall et al 2005 clinical canc res
APC	324	miclea. 2011. exp cell res, vitale et al 2011 cell death differ, thompson et al 2010 cell	vitale, thompson
ARHGA P19	84986		johansson
ARHGE F5	7984	kuroiva 2011 jcs	debily et al 2004 hum mol genet
ARID1A	8289		*
ASF1B	55723		carter
ASPM	259266	schaukat. 2012. plos one, buchmann. 2011. genes & dev	cinsarc
ATAD2	29028		carter, kalashnikova et al 2010 cancer res
ATM	472	schaukat. 2012. plos one, pandita. 2002. nature oncogene	
ATR	545	pino. 2010. gastroenterology	pino. 2010. gastroenterology
AURKA	6790	zhang. 2011. mol cell biochem, vitale et al 2011 cell death differ, thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	cinsarc, vitale, thompson, de castro

AURKB	9212	addepalli. 2010. gene ther, vitale et al 2011 cell death differ, thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	cinsarc+carter, vitale, thompson, de castro
AXIN2	8313	hadjihannas 2010 embo report, thompson et al 2010 cell	thompson
BIRC5	332	lamers 2011 endocr relat cancer, perez de castro et al 2007 carcinogenesis	cinsarc, de castro
BMI1	648	fasano 2007 cell stem cell	fasano 2007 cell stem cell
BORA	79866		carter
BRCA1	672	ballal 2009 j biol chem, vitale et al 2011 cell death differ, thompson et al 2010 cell	vitale, thompson
BRCA2	675	hattori 2011 mol cancer ther, vitale et al 2011 cell death differ, thompson et al 2010 cell	vitale, thompson
BUB1	699	shi 2010 Mol. Hum. Reprod. Advance oxford journal, thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	cinsarc, yuan et al 2005 clin canc res, thompson, de castro
BUB1B	701	miyamoto 2011 hum mol gen, thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	cinsarc, yuan et al 2005 clin canc res, thompson, de castro
BUB3	9184	li 2009 plos one, thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	yuan et al 2005 clin canc res, thomspn, de castro
C20orf2 4	55969		carter
CAMLG	819	thompson et al 2010 cell	lim et al 2011 breast canc res, thomspn
CCNA2	890		carter, tane et al 2009 cell cycle
CCNB1	891	perez de castro et al 2007 carcinogenesis	cinsarc+carter, de castro

CCNB2	9133	perez de castro et al 2007 carcinogenesis	cinsarc+carter, de castro
CCNE1	989	Etemadmogha 2010 plos one, thompson et al 2010 cell	thompson
CCT5	22948		carter
CDC20	991	taniguchi 2008 anticancer res, thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	cinsarc+carter, yuan et al 2005 clin canc res, thompson, perez de castro
CDC45	8318		cinsarc+carter
CDC6	990		cinsarc+carter
CDC7	8317		cinsarc, bonte et al 2008 neoplasia
CDCA2	157313		cinsarc, uchida 2013 plos one
CDCA3	83461		cinsarc+carter, uchida 2012 bmc cancer
CDCA8	55143	perez de castro et al 2007 carcinogenesis	cinsarc+carter, de castro
CDH1	999	naoe 2010 mol cell biol, ye 2012 mol med report	*
CDK1	983	zhang 2011 j cellular biochem, perez de castro et al 2007 carcinogenesis	westbrook et al 2007 cancer res, de castro
CDKN1 A	1026	vitale et al 2011 cell death differ	vitale
CDKN1 B	403429	sugihara 2006	sugihara 2006
CDKN2 A	1029	vitale et al 2011 cell death differ	*, vitale
CENPA	1058	mcgovern 2012 breast cancer res	carter, mcgovern 2012 breast cancer res
CENPE	1062	tanudji 2004 mboc, thompson et al 2010 cell	cinsarc, thompson

CENPF	1063	thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	thompson, de castro
CENPH	64946	orthaus 2006 biochem biophys res comm, thompson et al 2010 cell	thompson
CENPL	91687	-	cinsarc
CEP110	9738	perez de castro et al 2007 carcinogenesis	de castro
CEP55	55165	inoda et al 2009 j immunotherapy	cinsarc+carter
CHEK1	1111	höglund et al 2011 clinical cancer research	carter
CHEK2	11200	schaukat 2012, nagel et al 2012 breast cancer research treatment	
CHFR	55743	maddika 2009 JOURNAL OF BIOLOGICAL CHEMISTRY, perez de castro et al 2007 carcinogenesis	de castro
CIT	11113	whitworth et al 2012 plos one	whitworth et al 2012 plos one
CKAP2	26586	hong 2009 cell cycle, thompson et al 2010 cell	thompson
CKAP5	9793	perez de castro et al 2007 carcinogenesis	de castro
CKS1B	1163	shi 2010 onkotarget, perez de castro et al 2007 carcinogenesis	de castro
CKS2	1164	frontini 2012 dev cell, perez de castro et al 2007 carcinogenesis	cinsarc+carter, de castro
CLASP1	23332	Kiyosue 2005 jcb, thompson et al 2010 cell	thompson
CLASP2	23122	Kiyosue 2005 jcb, thompson et al 2010 cell	thompson
CMAS	55907	-	carter
COPB2	9276	-	

CTNNB 1	1499	zeng 2007 neoplasia, thompson et al 2010 cell	thompson
CTPS	1503	-	carter
CTSB	1508	bengsch et al 2013 oncogene	
DCC	1630	koren 2003 breast cancer treatme	
DHCR7	1717		carter
DIAPH3	81624	johansson 2013 plos one	johansson
DKC1	1736	montanaro 2006 j pathol	carter
ECT2	1894	kanada 2008 Molecular Biology of the Cell, perez de castro et al 2007 carcinogenesis	cinsarc+carter, de castro
ELAVL1	1994	zhu et al 2013 tumour biol	carter
ERBB2	2064	leung 2012 nature	
ESPL1	9700	thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	cinsarc+carter, thompson, de castro
ETS1	2113	verschuur 2013 cancer & met	
EZH2	2146	panousis 2011 eur j gynaec oncol	carter
FANCI	55215	garcia 2009 carcinogenesis	cinsarc
FBXO5	26271		CINSARC
FBXW7	55294	izumi 2012 plos one, vitale et al 2011 cell death differ, thompson et al 2010 cell	vitale, thompson
FEN1	2237	van pel 2013 plos genet	carter
FHIT	2272	fragile site FRA3B, arun 2005 Cancer Epidemiol Biomarkers Prev	fragile site

FLT1	2321	schaukat 2012, lee 2011 blood, lee 2008 plos med	
FOXM1	2305	xue 2010 plos one, thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	cinsarc+carter, thompson, de castro
FOXQ1	94234	kaneda 2010 cancer res, feuerborn 2011 j cell physiol	
GATA3	2625	vudoc et al 2008 cancer epidemiolo, dydensborg 2009 nature	*
GATA4	2626	hua 2009 mol cancer res	
GPI	2821	funasaka 2009 cancer res	carter
H2AFX	3014	perez de castro et al 2007 carcinogenesis	cinsarc+carter, de castro
H2AFZ	3015	svotelis 2010 cell cycle	carter
HAUS8	93323	thompson et al 2010 cell	thompson
HDGF	3068	chen 2012 j pathol	carter
HP1BP3	50809	? (housekeeping gene?!)	cinsarc
ID1	3397	vitale et al 2011 cell death differ, thompson et al 2010 cell	vitale, thompson
INCENP	3619	Jennifer C Hofmann, Alma Husedzinovic, and Oliver J Gruss. The function of spliceosome components in open mitosis. Nucleus. 2010 Nov- Dec; 1(6): 447–459. PMID: PMC3027046	
KIF11	3832	vitale et al 2011 cell death differ, thompson et al 2010 cell	cinsarc, vitale, thompson
KIF14	9928	singel 2012 cancer research	cinsarc
KIF15	56992	scanlan 2001 cancer immun	cinsarc

KIF18A	81930	zhang 2010 carcinogenesis	cinsarc
KIF20A	10112	Hill 2000 embo j	cinsarc+carter
KIF23	9493	?	cinsarc
KIF2a	3796	ganem 2004 jcb, thompson et al 2010 cell	thompson
KIF2b	84643	manning 2007 mboc, thompson et al 2010 cell	thompson
KIF2C	11004	Kline-Smith 2004 mol biol cell, thompson et al 2010 cell	cinsarc, thompson
KIF4A	300521	wandke 2012 jcb, thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	cinsarc+carter, thompson, de castro
KIFC1	3833	de 2009 cancer res	cinsarc
KLF4	9314	yu 2011 oncogene, thompson et al 2010 cell	thompson
KNTC1	9735	perez de castro et al 2007 carcinogenesis	de castro
KRAS	3845	sunaga 2011 mol cancer ther, chen et al 2010 Breast Cancer Research and Treatment	
LATS1	9113	vitale et al 2011 cell death differ, perez de castro et al 2007 carcinogenesis	vitale, de castro
LATS2	26524	vitale et al 2011 cell death differ, perez de castro et al 2007 carcinogenesis	vitale, de castro
LLGL1	3996	wan et al 2012 mboc	
LLGL2	3993	wan et al 2012 mboc	
LSM4	25804	carter et al 2006	carter
MAD1L 1	8379	chen 2012 oncol rep, vitale et al 2011 cell death differ, thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	vitale, thompson, de castro

MAD2L 1	4085	micel 2004 pnas, thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	cinsarc+carter, yuan et al 2005 clin canc res, thompson, de castro
MAP2K 1	5604	-	
MAP2K 3	5606	schaukat 2012, jia 2010 proteomics clin appl	
MAP2K 4	6416	*	*
MAPK1	5594	-	
MAPK8	5599	schaukat 2012, parra 2010 int j mol med	
MAPRE 1	22919	thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	thomspn, de castro
MCM10	55388	mark 2008 bichem biophys res comm	carter
MCM2	4171	cobanoglu 2010 clinical res	cinsarc
MCM7	4176	liu 2012 lung cancer	cinsarc+carter
MDM2	4193	brekmann 2012 breast cncr research, thompson et al 2010 cell	thompson
MDM4	4194	thompson et al 2010 cell	thompson
MELK	9833	hebbard 2010 cancer res	cinsarc+carter
MEX3C	51320	burrell 2013 nature	
MKL1	57591	scharenberg 2010 int j biochem cell biol	
MLH1	4292	iwaizuma 2013 mutat res	
KMT2C	58508	watanabe 2011 plos one, gupta 2012 cancer research	*

MMP2	4313	hillion 2009, mendes 2009 clinic exp met	
MMP9	4318	mendes 2009 clinic exp met	
MRE11 A	4361	yuan 2012 j nat canc inst	
MSH6	2956	bonadona 2011 jama	carter
MT1JP	4498	carter et al 2006	carter
MTCH2	23788	carter et al 2006	carter
MUS81	80198	wu et al 2010 anticancer research, franchitto 2008 jcb,	
MYC	4609	wang 2004 breast cancer research, vitale et al 2011 cell death differ	
MYH10	4628	overholzer 2007 cell, lordier 2012 nature	
MYO10	4651	overholzer 2007 cell, liu 2012 mol biol cell	
NCAPD 2	9918	perez de castro et al 2007 carcinogenesis	de castro
NCAPH	23397	perez de castro et al 2007 carcinogenesis	cinsarc, carter, de castro
NDC80	10403	tooley 2011 mboc, thompson et al 2010 cell	thompson
NDE1	54820		cisnarc
NDUFA B1	4706		carter
NEK2	4751	schaukat 2012, prigent 2005 exp cell res, perez de castro et al 2007 carcinogenesis	cinsarc+carter, de castro
NEU1	4758	butler 2013 trafic	
NIN	51199	perez de castro et al 2007 carcinogenesis	de castro

NISCH	11188	johansson	joahnsson
NPM1	4869	qin 2011 Int J Med Sci, perez de castro et al 2007 carcinogenesis	de castro
NUF2	83540	mattiuzzo 2011 plos one	cinsarc
NUMA1	4926	williams 2011 nature, perez de castro et al 2007 carcinogenesis	de castro
NUP20 5	23165	expression atlas-embl	carter
NUP98	4928	funasaka 2011 cel cycle, perez de castro et al 2007 carcinogenesis	de castro
NXT1	29107		carter
OIP5	11339	expression atlas embl	cinsarc+carter
PAK3	5063	schaukat 2012, parker 2013 plos one	cinsarc
PAK4	10298	schaukat 2012, tabusa 2012 mol cancer res	
PARD3	56288	wan 2013 mboc	
PARP1	142	rojo 2012 annual oncol	
PASK	23178	schaukat 2012, expression atlas embl	
PAX9	5083	hsu 2009 PNAS	
PBK	55872	kim 2012 cancer res	carter, cinsarc
PCM1	5108	perez de castro et al 2007 carcinogenesis	de castro
PCNA	5111	expression atlas embl,	carter
PIGN	23556	burrell 2013 nature	
PIK3CA	5290	zhou 2011 World J Gastroenterol,	*

		aleskandarani 2010 breast cancer res treat	
PIM1	5292	hsu 2012 cancer letters, vitale et al 2011 cell death differ	
PIN1	5300	matsuura 2010 THE JOURNAL OF BIOLOGICAL CHEMISTRY, perez de castro et al 2007 carcinogenesis	de castro
PINK1	56018	schaukat 2012, dagda 2009 JOURNAL OF BIOLOGICAL CHEMISTRY, expression atlas embl, berthier 2011 um pathol	
PLK1	5347	tyagi 2010 biochem phamacol, vitale et al 2011 cell death differ,perez de castro et al 2007 carcinogenesis	de castro
PLK4	10733	vitale et al 2011 cell death differ, perez de castro et al 2007 carcinogenesis	cinsarc, vitale, de castro
PRC1	9055	shrestha 2012 mbc, perez de castro et al 2007 carcinogenesis	cinsarc+carter, de castro
PRKCA	5578	schaukat 2012, expression atlas embl	
PRPF4	9128	thompson et al 2010 cell	thompson
PTEN	5728	bowen 2009 anticancer res	*
PTH1R	5745	expression atlas embl, liang 2012 med oncol	
PTPN11	5781	Liu 2012 cancer res, expression atlas embl, zhou 2009 histopathol	
PTTG1	9232	huang 2012 braz j med biol res, thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	cinsarc+carter, thompson, de castro
RAD21	5885	thompson et al 2010 cell, perez de castro et al	carter, thompson, de castro

		2007 carcinogenesis	
RAD51 AP1	10635	expression atlas embl, obama 2008 clin canc res	cinsarc+carter
RAE1	8480	thompson et al 2010 cell	thompson
RANBP 1	5902	thompson et al 2010 cell	thompson
RASSF1	11186	van der weyden 2005 mol cell biol, perez de castro et al 2007 carcinogenesis	de castro
RB1	5925	semizarrov 2004 nucl acid res, vitale et al 2011 cell death differ, thompson et al 2010 cell	*, vitale, thompson
REST	5978	thompson et al 2010 cell	thompson
RFC4	5984	expression atlas embl,	carter
RHOA	387	overholzer 2007 cell, vega 2011 jcb	overholzer
RNASE H2A	10535	flanagan 2009 mol canc ther	cinsarc+carter
ROCK1	6093	overholzer 2007 cell, liu 2011 j orthop res	overholzer
ROCK2	9475	overholzer 2007 cell	overholzer
RRM1	6240	kim 2011 j kor med sci	carter
RRM2	6241	expression atlas embl, yun 2008 exp & mol med	cinsarc+carter
SFRS2	6427	gout 2012 plos one, expression atlas embl	carter
SGOL1	151648	kahyo 2011 oncogene, thompson et al 2010 cell	thompson
SGOL2	151246	thompson et al 2010 cell	cinsarc, thompson

SLC16A 1	6566	thompson et al 2010 cell	thompson
SMAD2	4087	petersen 2012 oncogene	
SMAD4	4089	Deckers 2006 cancer res, stuelten 2006 bmc cancer	
SMC1A	8243	zhang 2013 oncology letters, thompson et al 2010 cell	thompson
SMC2	10592	expression atlas embl	cinsarc
SMC3	9126	ghiselli 2006 molecular cancer, thompson et al 2010 cell	thompson
SNCG	6623	shen 2011 chin med j, perez de castro et al 2007 carcinogenesis	de castro
SPAG5	10615	johansson 2013 plos one	cinsarc, johansson
SPC25	57405	wang 2013 jnci	cinsarc
SPINK7	84651	cheng et al 2009 jbc, thompson et al 2010 cell	thompson
STAG1	10274	thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	thompson, de castro
TACC3	10460	schneider 2008 oncogene, perez de castro et al 2007 carcinogenesis	de castro
TBCK	93627	schaukat 2012, -	
TCF4	6925	johansson 2013 plos one	
TERT	7015	beesley 2011 plos one	
THY1	7070	expression atlas embl	
TLN1	7094	lai 2011 j pathol	

TOP2a	7153	thompson et al 2010 cell	cinsarc+carter, thompson
TOR1AI P1	26092	expression atlas embl	
TP53	7157	vitale et al 2011 cell death differ, thompson et al 2010 cell	*, vitale, thompson
TP63	8626	stefanou 2004 histopathol	
TPX2	22974	warner 2009 clin canc res, perez de castro et al 2007 carcinogenesis	cinsarc+carter, de castro
TRIP13	9319	furnier 2006 cancer research	cinsarc+carter
TTK	7272	thompson et al 2010 cell, perez de castro et al 2007 carcinogenesis	cinsarc+carter, yuan et al 2005 clin canc res, thompson, de castro
TTN	7273	expression atlas embl	*
UBE2C	11065	lin 2006 neoplasia, perez de castro et al 2007 carcinogenesis	carter, de castro
UBE2I	7329	perez de castro et al 2007 carcinogenesis	carter, de castro
UNG	7374	pulukuri 2009 mol canc res	carter
VCL	7414	mierke 2010 jbc, expression atlas embl	
VHL	7428	zhou 2012 febs letters, thompson et al 2010 cell	thompson
VRK1	7443	schaukat 2012, molitor 2013 oncogenesis	
WNK1	65125	schaukat 2012, Tu 2010 pnas, expression atlas embl	
WWOX	51741	Fu 2011 blood	
YIF1B	90522	johansson 2013 plos one	johansson

ZNF516	9658	burrell 2013 nature	
ZSCAN2 2	342945	-	
ZW10	9183	varma 2008 diss, perez de castro et al 2007 carcinogenesis	de castro
Zwilch	55055	perez de castro et al 2007 carcinogenesis	carter, de castro
ZWINT	11130	perez de castro et al 2007 carcinogenesis	cinsarc+carter, de castro

Supplementary table 6: List of KNIME workflows used for data analysis.

Workflow name	Location	Input	Output
Image based knockdown efficiency	\\data.isilon.bioquant.uni-heidelberg.de\BQData\ag-conrad\Group\Marcel\Thesis\kname workflows	Image name = condition_replicate (e.g. negative_1, negative_2,...), monolayer cells, H2B-GFP, 1 plane, 850µm*850µm (512px*512px)	count of cells with the phenotypes "normal", "dead or artefact" and "INCENP" (polylobed) per condition
HCS 3D analyzer	\\data.isilon.bioquant.uni-heidelberg.de\BQData\ag-conrad\Group\Marcel\Thesis\kname workflows	Image name = "project_AQ_well#_position#_timepoint", monolayer cells or spheroids, H2B-GFP, 1 plane, 850µm*850µm (512px*512px)	Mean and standard deviation of 9 parameters per spheroid over all replicates
PCA cluster	\\data.isilon.bioquant.uni-heidelberg.de\BQData\ag-conrad\Group\Marcel\Thesis\kname workflows	Excel sheet with normalized parameters for each condition	PCA covariance matrix and spectral decomposition, clustering of all conditions using k-means
Class identifier	\\data.isilon.bioquant.uni-heidelberg.de\BQData\ag-conrad\Group\Marcel\Thesis\kname workflows	Image name = "project_AQ_well#_position#_timepoint", monolayer cells or spheroids, H2B-GFP, 1 plane, 850µm*850µm (512px*512px), all images of 1 condition per analysis	Count and % of abnormal and normal spheroids (also artefacts) of all analyzed images combined
TL concatenator	\\data.isilon.bioquant.uni-heidelberg.de\BQData\ag-conrad\Group\Marcel\Thesis\kname workflows	Single images (time points, H2B-GFP, 1 plane, 850µm*850µm (512px*512px)) that need to be concatenated and cropped	Time lapse images of each segment
Top TL analyzer	\\data.isilon.bioquant.uni-heidelberg.de\BQData\ag-conrad\Group\Marcel\Thesis\kname workflows	Time lapse images created by TL concatenator	Time resolved size, hollowness and polarization of single spheroids
MIP single cell tracking	\\data.isilon.bioquant.uni-heidelberg.de\BQData\ag-conrad\Group\Marcel\Thesis\kname workflows	Time lapse images created by TL concatenator	Mean nucleus velocity of all single cells during 5 different time windows
BCSC analyzer	\\data.isilon.bioquant.uni-heidelberg.de\BQData\ag-conrad\Group\Marcel\Thesis\kname workflows	Image name = condition_antibody_replicate, images with single spheroids with stacks (0.6 px/µm)	Size, mean intensity per spheroid of each condition and antibody
Drug test analysis	\\data.isilon.bioquant.uni-heidelberg.de\BQData\ag-conrad\Group\Marcel\Thesis\kname workflows	"project_AQ_well#_position#_timepoint", monolayer cells or spheroids, H2B-GFP, 1 plane, 850µm*850µm (512px*512px)	Size and number of cells or spheroids per well
Mammosphere analysis	\\data.isilon.bioquant.uni-heidelberg.de\BQData\ag-conrad\Group\Marcel\Thesis\kname workflows	3*3 stitched images consisting of H2B-GFP, 1 plane, 850µm*850µm (512px*512px)	Number of colonies > 4 cells

Acknowledgements

Herewith I want to thank Roland Eils and Christian Conrad for the opportunity to work in this challenging and innovative environment. Furthermore I want to thank the current and former members of our lab for making the time so enjoyable. Especially, I want to thank Lorenz Maier, Stephan Tirier, Fabian Braun, Timo Trefzer, Foo Wei Ten, Lorenz `Bob` Chua, Linda Klauss, Katharina Jechow, Teresa Krieger, Julia Neugebauer, Björn Eismann, Dominik Niopek, Mareike Hoffmann and Sabine Aschenbrenner. Also I want to thank Stephan Tirier, Fabian Braun, Pawel Gershkovich and especially Caro Waschow for proofreading this manuscript.

I feel a big gratitude for the people conducting or helping me conduct many of the challenging experiments and analyses: Lorenz Maier for teaching me how to use and also suffer from KNIME; Anna Jauch und Brigitte Schöll as well as Ilse Chudoba from Metasystems for helping me with the MFISH preparation and analysis; Corinna Klein and Martin Sprick for conducting the mouse experiments; Sabine Aschenbrenner, Qi Wang and Carl Herrmann for conducting and analyzing the 4C-seq experiments.

Lastly, I want to thank my family and friends who always supported me. I dedicate this work to my lovely wife Caro and my wonderful kids Maila and Arthur.