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

Mohsen Abolfathi, Biologist (Master of Science)

born in Abdanan, Iran

Oral examination:

Recurrent mutations, expression analysis and functional characterization of cohesin subunits in myelodysplastic syndromes and acute myeloid leukemia

Referees:

Prof. Dr. Stefan Wiemann

Prof. Dr. Alwin Krämer

Acknowledgment

First of all, I would like to thank my supervisor Prof. Alwin Krämer for his kind support. Secondly, I want to thank the TAC members and the referees. Thirdly, I want to thank our collaborators from Düsseldorf, Prof. Rainer Haas and Dr. Thomas Schröder and the Düsseldorf School of Oncology (DSO) for granting me my stipend. Additionally, I would like to thank Prof. Anna Jauch, Mutlu Kartal-Kaess, Annik Roßberg and all the lab members. Finally, I want to thank all my friends and my family for being there for me.

Table of contents

Figures

Tables

1 Abstract

Myelodysplastic syndromes (MDS) are common hematopoietic disorders that are associated with bone marrow failure and the possibility of developing leukemia *(1)*. MDS cells often contain chromosomal abnormalities, which significantly impact on the prognosis of the disease as documented by the contribution of chromosomal aberrations to the International Prognostic Scoring System (IPSS) used to prognostically classify MDS cases. Our understanding of the molecular mechanism of MDS has been increasing especially due to the advancements in genomics and next generation sequencing. An increasing list of mutated genes is being described in MDS including hematopoietic transcription factors (ETV6, CEBPA, RUNX1, SPI1 (PU.1)) *(2-5)*, epigenetic regulators (ASXL1, TET2, DNMT3A, IDH1, IDH2, EZH2, SUZ12) *(3, 6, 7)* and microRNAs *(8, 9)*, RNA splicing factors (SF3B1, SRSF2, U2AF1, ZRSR2) *(10)*, cell cycle regulators (CDKN1A,TP53,BCL2, AURKA, AURKB, CDC20, MAD2L1, TUBG1) *(10-14)*, members of the cohesin complex (STAG2, RAD21, SMC1, SMC3A) *(10, 15, 16)*, members of other signaling pathways (JAK2, IRAK1, CTNNB1, NOTCH1, NPM1, SMAD7, TGFB1, NF-κB) *(9, 16-20)*, proteins involved in immunological processes (TLR2, STAT3) *(21, 22)*, and others factors (CBL, CALR, BCOR, BCORL1, SETBP1, GNAS, CDKN2B, Nup98, HoxD13) *(2, 3, 10, 23-25)*. Identification of mutations in these genes has increased our understanding of the disease but there is a lot to be done in order to gain insights into the mechanisms of MDS pathogenesis. Mutations in the members of the cohesin complex including RAD21 and STAG2 have been found in MDS and other types of cancer such as bladder *(26-28)*, breast *(29)*, and colorectal cancer *(30)*. These mutations are associated with chromosomal instability and aneuploidy in some cancer types such as bladder cancer *(31)* but this finding remains controversial in other cancer types *(32)* and requires further mechanistic studies and patient data analysis in order to be validated. Genome engineering has been improved greatly over the past couple of years with the recent introduction of the CRISPR/Cas9 system by the Zhang group *(33)*, making functional validation of the genomic data obtained from massively parallel sequencing studies feasible.

In this study, we characterized a panel of 63 genes that has been reported to be frequently mutated in MDS, in 90 patient samples from MDS patients with and without chromosomal aberrations in their bone marrow mononuclear cells using a targeted re-sequencing approach to assess the frequency of mutations in these genes. This approach allowed us to determine whether the mutation spectrum is different in MDS cases with and without chromosomal aberrations, especially with regard to the occurrence of cohesin complex subunit mutations. In addition, we functionally characterized the cohesin complex subunit STAG2 that has been reported to be mutated in several cancers including MDS and reported to be

associated with chromosomal instability at least in some studies. To do this, we used CRISPR/Cas9 genome engineering to knock out this gene and analyzed for signs of chromosomal instability. We have used HCT116-p53$^{+/+}$ and HCT116-p53$^{-/-}$ cell lines for these functional analyses. Finally, we also analyzed the expression of STAG2 in AML samples using immunofluorescence microscopy and Western blotting.

TP53 somatic SNVs were found in 7/90 (7.8%) of the cases and were mainly associated with complex karyotypes, which is in accordance with previous reports. STAG2 was mutated in only 4/90 (4.4%) of samples. No mutations in other cohesin components were found. On the other hand, STAG2 expression was lost in 18 out of 74 (24.3%) AML samples due to STAG2 mutations in 20% (2/10) and promoter methylation in 58.3% (7/12) of cases. In addition, we used CRISPR/Cas9 genome editing to knock out STAG2 in diploid, chromosomally stable HCT116-p53$^{+/+}$ and HCT116-p53$^{-/-}$ cells. Whereas loss of STAG2 led to alterations in gene expression profiles in both cell lines, chromosome aberrations were only induced in the HCT116-p53$^{-/-}$ background. We conclude that the expression of STAG2 is lost in about one quarter of AML cases, frequently as a consequence of promoter methylation. Depending on the genetic background, both disturbed gene expression and chromosomal aberrations are associated with loss of STAG2.

2 Zusammenfassung

Myelodysplastische Syndrome (MDS) sind hämatopoetische Erkrankungen, die auf Knochenmarksfehlfunktionen zurückzuführen sind und zu einer Leukämie führen können. MDS-Zellen weisen häufig chromosomale Aberrationen auf, die einen erheblichen Einfluss auf die Krankheitsprognose haben. Dies wird durch das International Prognostic Scoring System (IPSS) deutlich, welches die MDS-Klassen anhand chromosomaler Aberrationen prognostisch einteilt. Unser Verständnis des molekularen Mechanismus von MDS hat vor allem aufgrund der Fortschritte in Mutationsanalyse mittels Next Generation Sequenzierung (NGS) zugenommen. Eine wachsende Liste von mutierten Genen sind in MDS beschrieben, darzu gehören hämatopoetische Transkriptionsfaktoren (ETV6, CEBPA, RUNX1, SPI1, PU.1), epigenetische Regulatoren (ASXL1, TET2, DNMT3A, IDH1, IDH2, EZH2, SUZ12), microRNAs, RNA-Splicing Faktoren (SF3B1, SRSF2, U2AF1, ZRSR2), Zellzyklus-Regulatoren (CDKN1A, TP53, BCL2, AURKA, AURKB, CDC20, MAD2L1, TUBG1), Mitglieder des Cohesin-Komplexes (STAG2, RAD21, SMC1, SMC3A), Mitglieder anderer Signalwege (JAK2, IRAK1, CTNNB1, NOTCH1, NPM1, SMAD7, TGFB1, NF-□B), Proteine immunologischer Prozesse (TLR2, STAT3), und andere Faktoren

(CBL, CALR, BCOR, BCORL1, SETBP1, GNAS, CDKN2B, Nup98, HoxD13). Die Identifikation dieser Genmutationen hat unser Verständnis der Krankheit verbessert, aber es gibt noch einiges zu tun, um einen detaillierteren Einblick in den Mechanismus der MDS-Pathogenese zu gewinnen. Mutationen in Mitgliedern des Cohesin-Komplexes wie RAD21 und STAG2 wurden in MDS und in anderen Krebsarten wie Blasen-, Brust- und Darmkrebs gefunden. Diese Mutationen werden in einigen Krebsarten wie Blasenkrebs mit chromosomaler Instabilität und Aneuplodie in Verbindung gebracht, während bei anderen Krebsarten diese Hypothese umstritten ist und weiterer mechanistischer Untersuchungen und Analysen von Patientendaten bedarf. In den letzten Jahren hat sich das Genome-Engineering durch die Einführung der CRISPR/Cas9-Technologie durch die Zhang-Gruppe stark verbessert, wodurch eine funktionelle Validierung von genomischen Datensätzen aus Sequenzierungsstudien praktikabel wurde.

In der vorliegenden Arbeit haben wir 90 Proben von MDS-Patienten mit oder ohne chromosomale Aberrationen in mononukleären Zellen des Knochenmarkes einer gezielten Re-Sequenzierungsanalyse unterzogen. Dabei analysierten wir eine Gruppe von 63 Genen, welche häufig beim MDS mutiert vorliegen, und bewerteten die Häufigkeiten der nachgewiesenen Genmutationen. Dieser Ansatz ermöglichte die Untersuchung des Mutationsspektrums in Hinblick auf Unterschiede zwischen MDS-Fällen mit und ohne chromosomale Aberrationen, wobei wir vor allem an Mutationen in Komponenten des Cohesin-Komplexes interessiert waren. Zusätzlich charakterisierten wir die Cohesin-Komplex-Untereinheit STAG2 funktionell. STAG2 ist in mehreren Krebsarten darunter MDS mutiert und wird zumindest in einigen Studien mit chromosomaler Instabilität in Verbindung gebracht. Zur funktionellen Charakterisierung schalteten wir STAG2 in den diploiden, chromosomal stabilen Zelllinien HCT116-p53$^{+/+}$ und HCT116-p53$^{-/-}$ mit Hilfe der CRISPR/Cas9 Genome Engineering Technologie aus, und untersuchten den Grad der chromosomalen Instabilität. Außerdem analysierten wir mittels Immunfluoreszenzmikroskopie und Western Blot-Analysen die STAG2-Expression in mehreren AML-Patientenproben.

Somatische SNVs in TP53 wurden in sieben der 90 MDS-Patientenproben (7.8%) gefunden, wobei diese Proben häufig einen komplexen Karyotyp aufwiesen; dies stimmt mit früheren Studien überein. STAG2 war nur in vier der 90 Proben (4.4%) mutiert, während wir in den anderen Cohesin-Komplex-Komponenten keine Mutationen detektieren konnten. Auf der anderen Seite konnten wir keine STAG2-Expression in 18 der 74 AML-Patientenproben (24.3%) beobachten, was auf STAG2-Mutationen (2/10; 20%) oder STAG2-Promotormethylierungen (7/12; 58.3%) zurückgeführt werden konnte. Der Verlust der STAG2-Expression in den beiden Zelllinien HCT116-p53$^{+/+}$ und HCT116-p53$^{-/-}$ führte zu veränderten Genexpressionsprofilen in beiden Zelllinien, während chromosomale Aberrationen nur in den STAG2-deletierten HCT116-p53$^{-/-}$ induziert wurden.

3 Introduction

3.1 Normal hematopoiesis

Blood is composed of blood cells and plasma. Blood cellular components are differentiated from hematopoietic stem cells (HSCs) during hematopoiesis. HSCs are multipotent stem cells that are able to undergo self-renewal and differentiate into all blood cell lineages. Hematopoiesis occurs for the first time in human fetus at about on month of age in the yolk sac in the forms of blood islands. During prenatal development, hematopoiesis occurs also in the liver, spleen and lymph nodes. From about month four in human fetus, hematopoiesis starts in the bone marrow which is the only postnatal place for hematopoiesis however, maturation of lymphoid cells occurs in secondary lymphoid organs. hematopoiesis occurs mainly in tibia and femur till the age of about 30 years after the age of 30 it exclusively occurs in the vertebral and pelvis, sternum and ribs *(34)* (**Figure 1**).

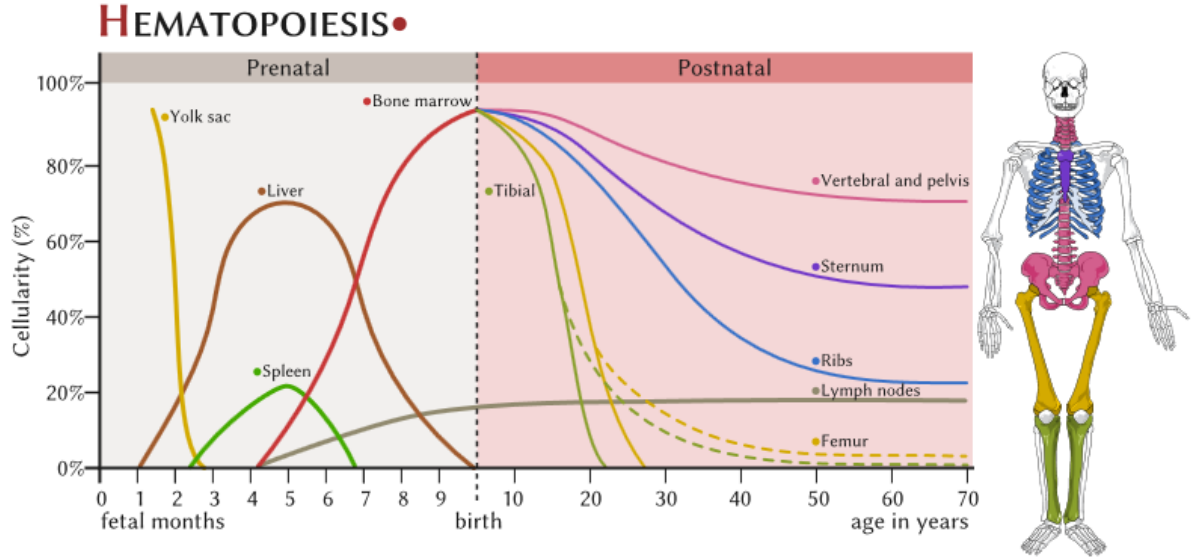


Figure 1: Hematopoiesis sites during development

Hematopoiesis starts in one month old embryos in the yolk sac and continues in liver, spleen and bone marrow. At birth hematopoiesis becomes limited to the bone marrow as indicated. Figure taken from wikipedia

HSCs differentiate into common myeloid progenitors (CMP) and common lymphoid progenitors (CLPs). CMPs produce all myeloid lineage cells including thrombocytes, erythrocytes, monocytes, basophils, neutrophils, and eosinophils. CLPs differentiate into B- and T-lymphocytes. Each differentiation step is tightly regulated and dependent on different growth factors and cytokines (**Figure 2**). The

microenvironment in bone marrow in which HSCs reside is called niche which includes all the cells and the factors that affect self-renewal and differentiation of HSCs. Changes in the niche or mutations in cytokines and growth factors or HSC genes can result in malignant hematopoiesis.

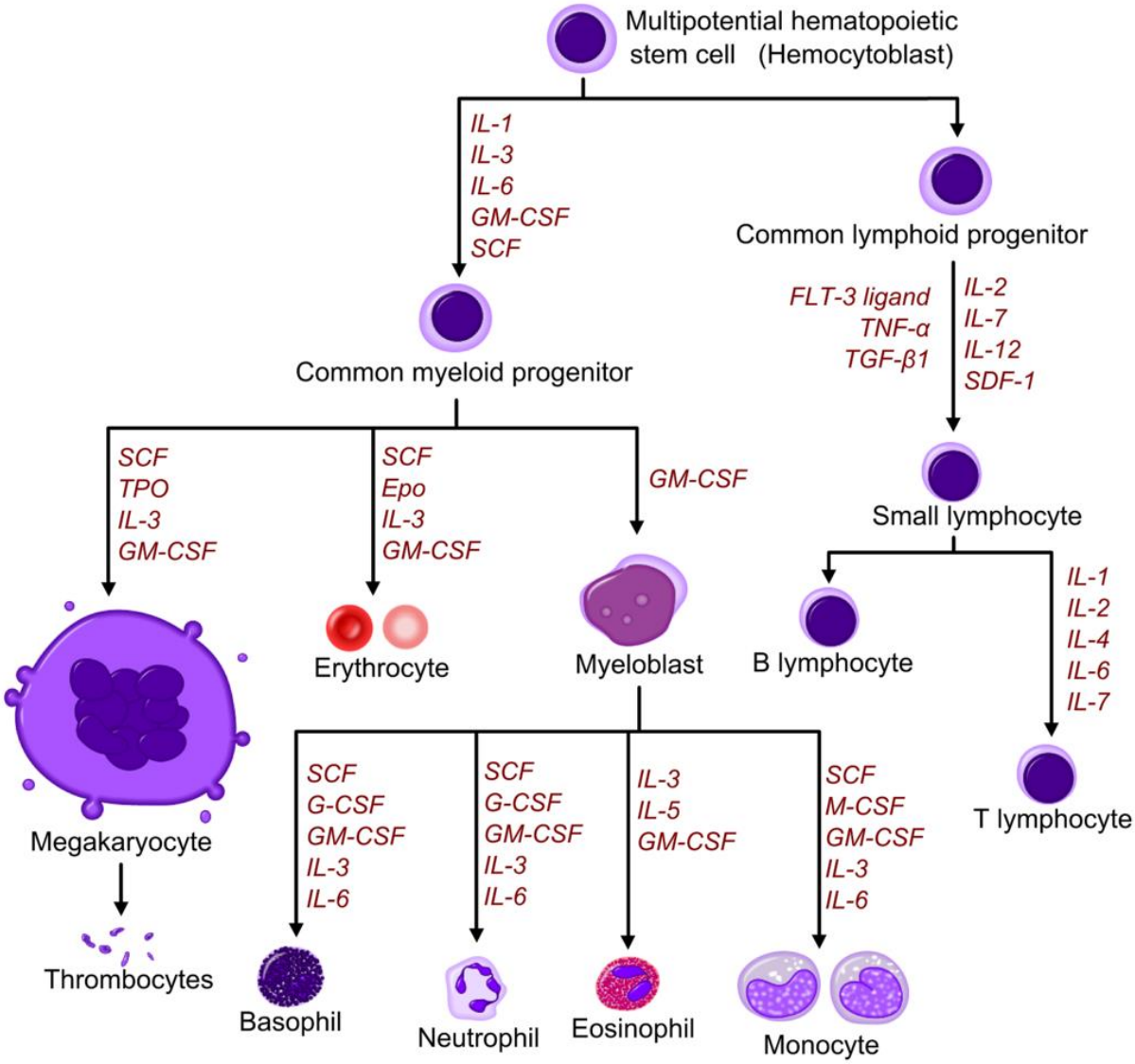


Figure 2: Differentiation of hematopoietic stem cells

Multipotent hematopoietic stem cells differentiate into common lymphoid progenitors, which produce all lymphocytes upon differentiation, and common myeloid progenitors which further differentiate into all myeloid lineage cells. Different growth factors for each differentiation step are shown. SCF, Stem Cell Factor; TPO, Thrombopoietin; IL, Interleukin; GM-CSF, Granulocyte Macrophage-Colony Stimulating Factor; EPO, Erythropoietin; M-CSF, Macrophage-Colony Stimulating Factor; G-CSF, Granulocyte-Colony Stimulating Factor; SDF-1, Stromal cell-derived factor-1; FLT-3 ligand, FMS-like tyrosine kinase 3 ligand; TNF-□, Tumour necrosis factor-alpha; TGFβ, Transforming growth factor beta. Figure taken from wikipedia

3.2 Malignant hematopoiesis

Hematopoiesis is a sophisticated process that is tightly controlled at different steps during differentiation. Malignant hematopoiesis can occur due to any change in the normal hematopoiesis process. For instance, any mutation in the growth factors regulating different steps of hematopoiesis can lead to malignant hematopoiesis. It is not clear whether malignant hematopoiesis occurs in HSCs or in other cells during differentiation. According to the literature, there are two main scenarios: it can either occur in the common myeloid progenitors in the myeloid lineage or in the common lymphoid progenitors in the lymphoid lineage. In general, myeloid malignancies are known as leukemias and lymphoid malignancies are called lymphomas. Malignant hematopoiesis is also categorized into acute and chronic malignancies based on the clinical course of the disorders. Myeloid malignancies include acute myeloid leukemia (AML) and chronic myeloid malignancies. Myelodysplastic syndrome (MDS) is a chronic myeloid malignancy which in 30% of the cases leads to secondary AML *(3)*.

3.3 Myelodysplastic syndromes (MDS)

3.3.1 MDS

MDS are clonal myeloid disorders in which abnormal blood cells are produced in the bone marrow mainly as a result of mutations in genes regulating the stem cell fate. Infection, anemia, shortness of breath, fatigue or bleeding may occur as a result of MDS especially in the late phases of the disease. About 30% of MDS transform into AML. MDS occurs rarely below the age of 50 but is common in more than 70 years old patients. MDS occurs annually in more than 20 per 100,000 people *(35)*. The overall survival rate is on average about 2.5 years from initial diagnosis. Like other types of cancer, the risk factors for MDS are not quite clear but exposure to chemotherapy and radiation are among them. Cytopenias including anemia are common in MDS, which are first identified in routine blood cell counts as the first step in the diagnosis of MDS. The next step in the diagnosis of MDS is usually looking at blood and bone marrow smears, which reveal the typical morphologic characteristics of MDS. Bone marrow dysplasia, dimorphic red blood cells, and ring sideroblasts are among the common morphologic features of MDS *(1)*. Karyotyping of the bone marrow cells using conventional cytogenetic methods and fluorescence in situ hybridization (FISH) is the next step in the diagnosis of MDS, which is used for further classification. Nowadays, sequencing of a panel of about 50 genes that are recurrently mutated in MDS based on next generation sequencing studies has become part of the diagnosis in some centers. However, the clinical outcomes associated with these mutations are still mainly unknown and need to be further studied.

3.3.2 MDS classification

MDS pathogenesis includes cytopenias, morphological and cytogenetic abnormalities, genetic and epigenetic dysregulation, and a deregulated immune system. Since the pathogenesis of MDS is complicated, it is difficult to classify MDS like other myeloid malignancies *(36)*. There are different classification systems which use a different set of criteria for MDS classification. However, percentage of blasts in bone marrow, cytopenia, dysplasia, and more recently cytogenetics are the main criteria for MDS classification. MDS are highly heterogeneous disorders regarding karyotype. 30 to 80% of patients have chromosomal abnormalities while the remaining 20 to 70 percent have a normal karyotype. Chromosomal abnormalities occur as a single abnormality, in combination with another abnormality, or together with two or more other abnormalities (complex karyotype). Deletion (5q) is the most frequent chromosomal abnormality either alone or in combination with other abnormalities. **Figure 3** represents the karyotype abnormalities in 1080 MDS patients *(36)*.

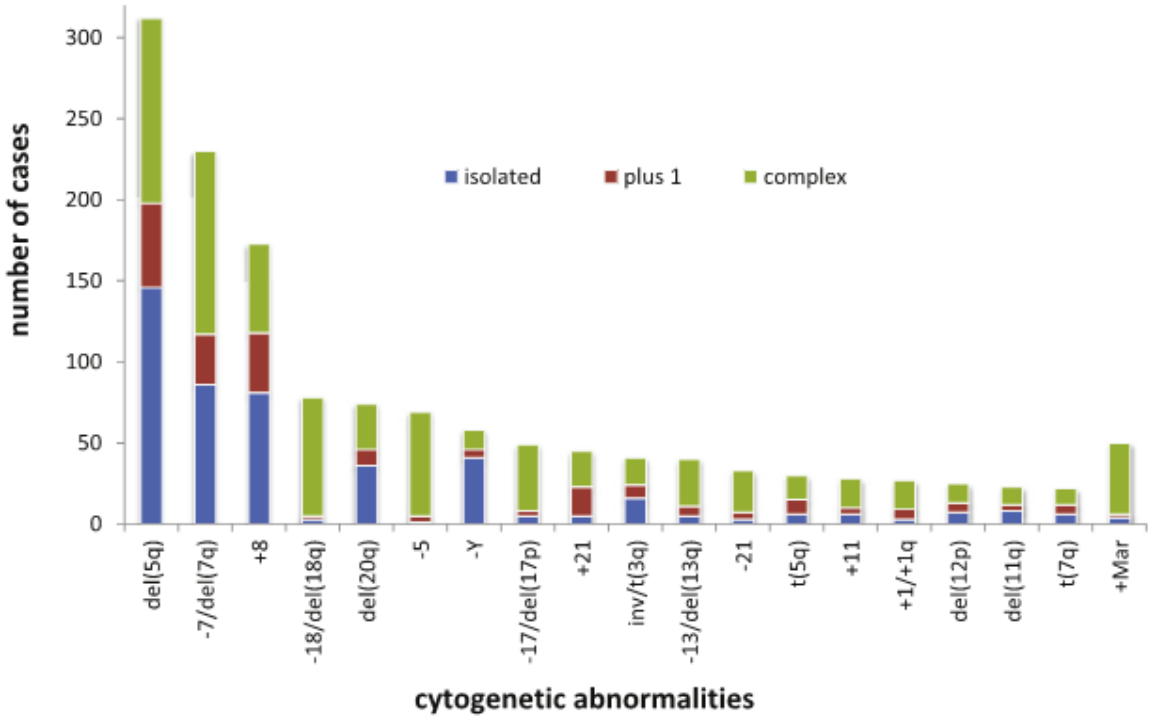


Figure 3: Representative karyotype abnormalities in MDS

Frequency distribution of chromosome abnormalities in 1080 patients with MDS. Taken from *(37)*

.

The most up to date system for putting chromosomal abnormalities in MDS into prognostic categories is the so-called revised international prognostic system (IPSS-R) which classifies chromosomal abnormalities in five prognostic subcategories: very good (-Y, del(11q)), good (normal, del (5q), del (12p), del(20q), double including del (5q)), intermediate (del(7q), +8, +19, i(17q), or any other single or double clones), poor (-7, inv(3)/t(3q)/del(3q), double including -7/del(7q), complex: 3 abnormalities), and very poor (complex: > 3 abnormalities) *(45)* (**Table 1)**.

Prognostic subgroups, % of patients	Cytogenetic abnormalities	Median survival,* y	Median AML evolution, 25%,* y	Hazard ratios OS/AML*	Hazard ratios OS/AML†
Very good (4%*/3%†)	−Y, del(11q)	5.4	NR	0.7/0.4	0.5/0.5
Good (72%*/66%†)	Normal, del(5q), del(12p), del(20q), double including del(5q)	4.8	9.4	1/1	1/1
Intermediate (13%*/19%†)	del(7q), +8, +19, i(17q), any other single or double independent clones	2.7	2.5	1.5/1.8	1.6/2.2
Poor (4%*/5%†)	−7, inv(3)/t(3q)/del(3q), double including −7/del(7q), complex: 3 abnormalities	1.5	1.7	2.3/2.3	2.6/3.4
Very poor (7%*/7%†)	Complex: > 3 abnormalities	0.7	0.7	3.8/3.6	4.2/4.9

OS indicates overall survival; and NR, not reached.
*Data from patients in this IWG-PM database, multivariate analysis (n = 7012).
†Data from Schanz et al[8] (n = 2754).

Table 1: IPSS-R MDS cytogenetic scoring system

Taken from (42).

There are four major systems for MDS classification: the French-American-British (FAB) system*(1, 38)*, the world health organization system (WHO)*(39, 40)*, the international prognostic scoring system (IPSS), and the revised international prognostic scoring system (IPSS-R). FAB being the oldest system and IPSS-R is the most up to date system. Each system uses a different set of criteria for classifying MDS. However, percentage of blasts in bone marrow, cytopenia, dysplasia, and more recently cytogenetics are the main criteria for MDS classification. For the purpose of the current introduction the IPSS and IPSS-R systems will be explained below.

In 1997 a group of scientists came together in a workshop in order to generate an international prognostic scoring system (IPSS) for MDS. The IPSS was supposed to be more precise than other classification systems in terms of prognostic power and clinical outcomes. The IPSS uses three criteria to classify MDS patients: the bone marrow blast percentage, the number of cytopenias, and the cytogenetics. The first two criteria are the same as in the FAB system. However, the IPSS puts the cytogenetic abnormalities into prognostic categories for the first time which is quit challenging due to the fact that chromosomal abnormalities are highly heterogeneous and they occur either alone or in combination with two or more other abnormalities. The IPSS puts the cytogenetic abnormalities into three groups good (normal karyotype, isolated del(5q), del(20q) and loss of the Y-chromosome), poor (any chromosome 7-abnormality, complex (≥ 3 cytogenetic changes)) and intermediate (any other abnormality) *(41)*. Using the risk score for bone marrow blasts, number of cytopenias and the cytogenetic subcategory, the IPSS puts MDS

paitients into 4 risk groups in terms of overall survival and AML evolution: low, 0; intermediate-1 (INT-1), 0.5 to 1.0; intermediate-2 (INT-2), 1.5 to 2.0; and high, ≥2.5 **Table 2**.

	Score Value				
Prognostic Variable	0	0.5	1.0	1.5	2.0
BM blasts (%)	<5	5-10	—	11-20	21-30
Karyotype*	Good	Intermediate	Poor		
Cytopenias	0/1	2/3			

Scores for risk groups are as follows: Low, 0; INT-1, 0.5-1.0; INT-2, 1.5-2.0; and High, ≥2.5.

* Good, normal, −Y, del(5q), del(20q); Poor, complex (≥3 abnormalities) or chromosome 7 anomalies; Intermediate, other abnormalities.

Table 2: IPSS for MDS

Taken from (41)

The four major IPSS groups were clearly distinct in terms of survival and AML evolution which shows the reliability of IPSS classification system **Figure 4**.

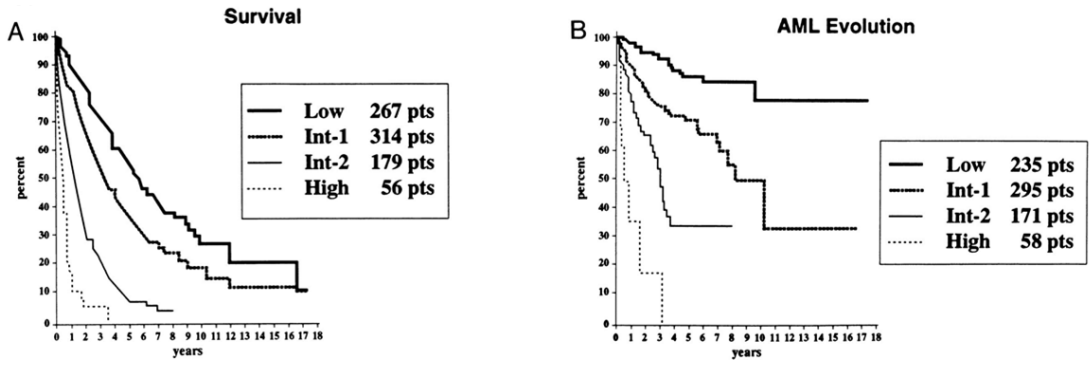


Figure 4: Age-related survival and AML evolution of MDS patients

Survival (A) and AML evolution (B) of MDS patients related to their classification by the IPSS for MDS: Low, INT-1, INT-2, and High. Taken from (*41*).

Although IPSS provided the gold standard for classifying MDS, it continued to be updated with increasing the amount of data available. Using a much larger international MDS patient database, a panel of scientists provide a revised version of

IPSS (IPSS-R) (IPSS-R, n = 7012, IPSS, n = 816). Although bone marrow blast percentage, cytopenias, and cytogenetics remain the main criteria for MDS classification in IPSS-R, there are some modifications compared to IPSS. The bone marrow blast percentage is divided into 3 subcategories by splitting the < 5% into two categories 0%-2%, > 2-< 5%, and putting the rest into one group > 10%. Cytogenetic abnormalities have been divided into 5 rather than 3 subcategories as mentioned above in **Table 2**. Cytopenias are evaluated based on their depth rather than their number. Furthermore, differentiating features such as age, performance, serum ferritin, and LDH have been considered for survival. IPSS-R divides the MDS patients into 5 prognostic subcategories very low, low, intermediate, high, and very high **Table 3**.

Risk category	Risk score
Very low	≤ 1.5
Low	> 1.5-3
Intermediate	> 3-4.5
High	> 4.5-6
Very high	> 6

Table 3: IPSS-R prognostic risk categories/scores

Taken from (42)

So the IPSS-R differs from IPSS in a number of criteria used for MDS classification **Table 4***(42)*.

Taken together, there are four major systems for MDS classification the FAB, the WHO, the IPSS, and the IPSS-R. FAB is the oldest and IPSS-R the most up to date MDS classification system. Which one is the best system for MDS classification is not clear. Scientists around the world still use all the systems to some extent. However, IPSS and in recent year IPSS-R are becoming more or less the gold standard for MDS classification.

1. New marrow blast categories
≤ 2%, > 2%-< 5%, 5%-10%, > 10%-30%
2. Refined cytogenetic abnormalities and risk groups
16 (vs 6) specific abnormalities, 5 (vs 3) subgroups
3. Evaluation of depth of cytopenias
Clinically and statistically relevant cutpoints used
4. Inclusion of differentiating features*
Age, Performance Status, serum ferritin, LDH; β_2-microglobulin†
5. Prognostic model with 5 (vs 4) risk categories
Improved predictive power

*For survival.
†Provisional.

Table 4: Refinements of the IPSS-R beyond the IPSS

Taken from (42).

3.4 Molecular mechanisms of MDS

Molecular mechanisms of MDS and other hematological malignancies have been conventionally studied through cytogenetics, and cloning and sequencing of single genes *(43)*. Over the past couple of years, the molecular abnormalities of MDS especially mutations in several gene families have been revealed by means of massive parallel sequencing studies. An increasing list of mutated genes is being described in MDS including hematopoietic transcription factors (ETV6, CEBPA, RUNX1, SPI1 (PU.1)) *(2-5)*, epigenetic regulators (ASXL1, TET2, DNMT3A, IDH1, IDH2, EZH2, SUZ12) *(3, 6, 7)* and microRNAs *(8, 9)*, RNA splicing factors (SF3B1, SRSF2, U2AF1, ZRSR2) *(10)*, cell cycle regulators (CDKN1A,TP53,BCL2, AURKA, AURKB, CDC20, MAD2L1, TUBG1) *(10-14)*, members of the cohesin complex (STAG2, RAD21, SMC1, SMC3A) *(10, 15, 16)*, members of other signaling pathways (JAK2, IRAK1, CTNNB1, NOTCH1, NPM1, SMAD7, TGFB1, NF-κB) *(9, 16-20)*, proteins involved in immunological processes (TLR2, STAT3) *(21, 22)*, and other factors (CBL, CALR, BCOR, BCORL1, SETBP1, GNAS, CDKN2B, Nup98, HoxD13) *(2, 3, 10, 23-25)* (**Table 5)**.

Genes
Hematopoietic Transcription factors
ETV6
CEBPA
RUNX1
SPI1(PU.1)
Epigenetic regulators
ASXL1
TET2
DNMT3A
IDH1
IDH2
EZH2
SUZ12
MicroRNAs
miR-21
RNA Splicing factors
SF3B1
SRSF2
U2AF1
ZRSR2
Cell Cycle regulators
CDKN1A (p21)
TP53
BCL2
AURKA
AURKB
CDC20
MAD2L1
TUBG1
Cohesin complex
STAG2
RAD21
SMC1A
SMC3
Signaling pathways
JAK2
IRAK1
CTNNB1(β-catenin)
NOTCH1
NPM1
SMAD7
TGFB1
NF-κB
Immunology
TLR2
STAT3
Others
CBL
CALR
BCOR
BCORL1
SETBP1
GNAS
CDKN2B
Nup98
HoxD13
Chromosomal aberrations
11q amplifications, deletions within 5q, 17p, and 7q
5q−, plus 8, −7

Table 5: Frequently mutated genes in MDS in the literature

Cohesin complex components are among the genes that have been described in recent years to be frequently mutated in MDS. The cohesin complex is a highly conserved 4-subunit ring structure that encircles sister chromatids, allowing their cohesion, and also plays critical roles in transcriptional regulation, DNA replication, heterochromatin formation and post-replicative DNA repair *(44)*. Somatic mutations in STAG2, a component of the cohesin complex, have been found in about 6% of MDS patients *(45)*. In a recent work, Kon et al. *(15)* detected mutations and deletions involving various components of the cohesin complex (STAG2, RAD21, SMC1A, and SMC3) in 8% of patients with MDS, 10% of those with CMML, and 12% of those with AML (**Figure 5**).

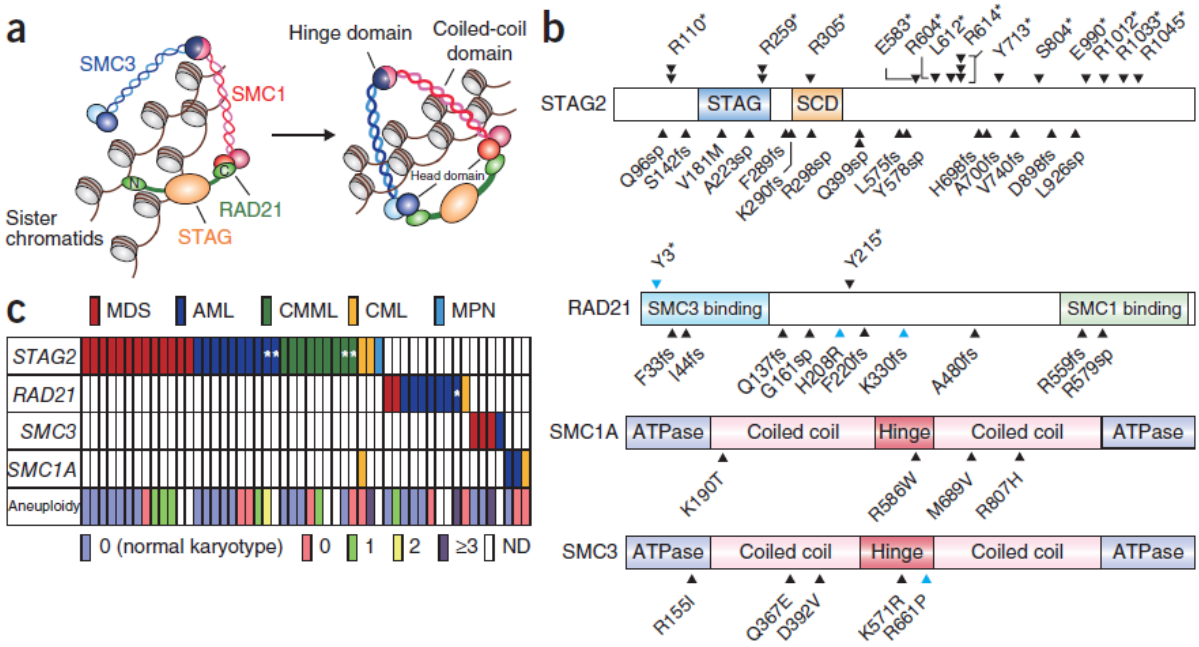


Figure 5: Recurrent cohesin mutations in MDS

(a) Cohesin holds chromatin strands together within a ring-like structure that is composed of the four core components STAG, RAD21, SMC1 and SMC3. (b) Mutations in the core components of the cohesin complex found in myeloid malignancies (black arrowheads) and myeloid leukemia-derived cell lines (blue arrowheads). The amino acids in the alterations are referred to using their one-letter abbreviations (for example, R110* represents p.Arg110*). (c) Distribution of cohesin mutations and deletions showing a nearly mutually exclusive pattern among different myeloid neoplasms. Gene deletions are indicated by asterisks. The number of numerical chromosome abnormalities in each cohesin-mutated or deleted case is shown at the bottom. ND, not determined. Taken from (*15*).

A similar mutation frequency was reported in AML patients by others *(46)*, suggesting that altered cohesin function plays a role in myeloid leukemogenesis. Cohesin mutations have been found in other types of cancer as well. In glioblastoma, the function of STAG2 has been related to maintenance of euploidy via its role in the cohesin complex. In a screen of a large series of bladder tumors and cell lines, Taylor et al. found inactivating mutations (nonsense, frameshift and splicing) in 67 of

307 tumors (21.8%) and 6 of 47 cell lines *(26)*. Functional assays in glioblastoma cell lines have linked loss of STAG2 expression to chromatid cohesion defects and aneuploidy *(31)*. In a siRNA screen of 101 candidate driver genes in breast cancer cell lines, Mahmood et al. found eight driver genes including RAD21 that were amplified, overexpressed and critical for breast tumor cell proliferation or survival *(29)*. In a retrospective observation study, Deb et al. examined RAD21 expression in 652 colorectal cancers using a tissue microarray approach. The results showed that RAD21 expression is a novel prognostic marker, particularly in the context of KRAS mutations and most likely within cancers arising through chromosomal instability *(30)* . In a screen of a large series of early colorectal adenomas, a precocious step during colorectal tumorigenesis, Cucco et al. identified eleven mutations in the SMC1A cohesin subunit. They showed that chromosomal instability is induced in normal human fibroblasts after either transfection of the SMC1A mutations identified in early adenomas or wild-type SMC1A gene silencing *(47)*. STAG2 is reported to be targeted by somatic aberrations in a subset (4%) of human pancreatic ductal adenocarcinoma (PDAs) *(48)*. Loss of STAG2 protein expression is seen in human PDAs tumor tissue with complete absence of STAG2 staining in 4.3% of patients. STAG2 expression is disrupted in these tumors suggesting a tumor suppressor role for STAG2 in human PDAs *(48)*. So, the cohesin complex members and especially STAG2 are mutated in different cancer entities. However, different cohesin members are mutated in different cancer entities and the functional consequences of cohesin mutations in different cancers is not quite clear and likely differ by tumor type. There are some reports on the role of cohesin mutations in chromosomal instability in cancer but results are controversial. To our knowledge, there is no report on the functional role of cohesin mutations in MDS especially in terms of chromosomal instability. Therefore, in the current work, we studied the frequency of STAG2 mutations in MDS and the consequence of its loss of function with regard to chromosomal instability in vitro.

3.5 Therapy

Currently, MDS diagnosis includes evaluation of cytopenias and dysplasia by assessing peripheral blood and bone marrow morphology, bone marrow biopsy to assess marrow cellularity and dysplasia, fibrosis, and topography, and cytogenetics. Somatic mutations are increasingly evaluated in MDS by massive parallel sequencing and will be part of the approach to diagnosis of MDS in the future (**Figure 6**) *(10)*. Nowadays, IPSS is the golden standard for diagnostic categorization and, correspondingly, for the choice of therapy. Once diagnosed as MDS, the therapy is decided based on whether it is low risk or high risk MDS. For low risk MDS which usually do not receive allogenic stem cell transplantation (alloSCT), some chemotherapeutic agents are currently available. These agents include hematopoietic growth factors, lenalidomide, and azanucleosides (5-azacitidine and

5-aza-2′-deoxycitidine (decitabine)). Erythroid growth factors (erythropoetin) are commonly used in MDS although there are few studies showing its efficacy and it has not been approved by FDA for treating MDS and anemia. Lenalidomide used in patients with lower risk MDS and del(5q) and has been shown to be quite effective. Most low risk MDS patients, however, are treated with 5-azacytidine and decitabine. In high risk MDS patients the following agents and approaches are available: Azanucleosides (5-azacitidine and 5-aza-2′-deoxycitidine (decitabine)), AML-like chemotherapy, and alloSCT. 5-azacitidine and decitabine are standard care for high risk MDS treatment. Response rates for both agents seem to be similar and both seem to improve survival but data from randomized clinical trials are available only for 5-azacitidine. AML-like chemotherapy might be used in younger patients that are candidates for alloSCT. Although usually restricted to young patients with a suitable donor, alloSCT is supposed to be the only curative treatment in MDS *(49)*. There is an increasing amount of data available on somatic mutations in MDS. However, they have still not been integrated into therapeutic approaches in this disorder.

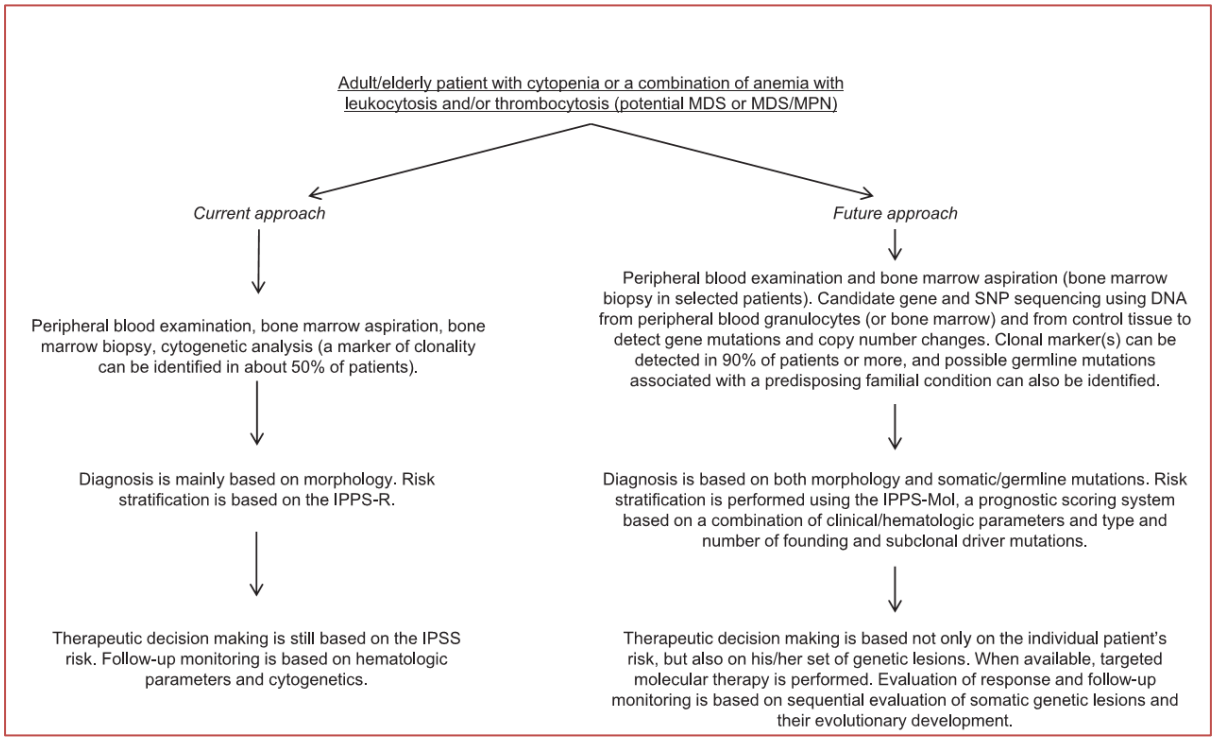


Figure 6: Current and future approach to diagnosis and prognostication of MDS

Taken from (10).

4 Materials and Methods

4.1 Materials

All the required substances and equipments for the Haloplex Target enrichment procedure are given in **Table 6**. The list of genes in the panel for targeted resequencing is given in **Table 7**.

Haloplex reuired reagents	company	Catalog #
Agencourt AMPure XP 60 mL Kit	Beckman Coulter Genomics	A63881
Agencourt SPRIPlate 96R - Ring Super Magnet Plate	Beckman Coulter Genomics	A32782
DynaMag-2 magnet	Life Technologies	12321D
Qubit 2.0 Fluorometer	Life Technologies	Q32866
Qubit assay tubes	Life Technologies	Q32856
Qubit® dsDNA BR Assay Kit	Life Technologies	Q32850
Eppendorf Research® (multi-channel) 8 channel (0.5-10)	Eppendorf	
Eppendorf Research® (multi-channel) 8 channel (10-100)	Eppendorf	
2 M acetic acid		
10 M NaOH		
10 mM Tris-HCl, pH 8.0		
Adhesive seals for 96-well PCR plates	Agilent	410186
Herculase II Fusion DNA Polymerase, 200 rxn	Agilent	600677
VWR® MiniFuge	VWR	93000-196
MPS1000 Mini Plate spinner	Labnet International	C1000
Microplate Foam Insert (2)	Scientific industries	504-0235-00
6-inch Platform	Scientific industries	146-6005-00

Table 6: Reagents and equipment required for Haloplex Target enrichment

Target ID	Regions	Coverage	High Coverage (>= 90%)	Low Coverage (< 90%)
AGBL1	25	99.87 %	25	0
AKAP9	52	99.19 %	50	2
ASXL1	17	100.00 %	17	0
AURKA	8	98.92 %	7	1
AURKB	8	100.00 %	8	0
B3GALT6	1	100.00 %	1	0
BCL2	2	100.00 %	2	0
BCOR	16	100.00 %	16	0
BCORL1	13	100.00 %	13	0
BUB1B	24	99.65 %	24	0
C19ORF80	4	100.00 %	4	0
CALR	9	100.00 %	9	0
CBL	16	100.00 %	16	0
CDC20	10	100.00 %	10	0
CDKN1A	3	100.00 %	3	0
CDKN2B	2	100.00 %	2	0
CEBPA	1	98.81 %	1	0
CTNNB1	14	99.50 %	14	0
DIDO1	15	98.14 %	15	0
DNMT3A	25	100.00 %	25	0
ETV6	8	100.00 %	8	0
EZH2	21	96.87 %	20	1
FANCA	44	99.81 %	44	0
FANCD2	45	98.08 %	42	3
FOXP1	20	100.00 %	20	0
GNAS	17	99.28 %	16	1
HAUS3	5	99.54 %	5	0
HAUS7	14	100.00 %	14	0
HOXD13	2	99.72 %	2	0
IDH1	8	100.00 %	8	0
IDH2	11	100.00 %	11	0
IRAK1	14	100.00 %	14	0

Target ID	Regions	Coverage	High Coverage (>= 90%)	Low Coverage (< 90%)
JAK2	23	99.61 %	23	0
KRAS	5	100.00 %	5	0
LATS1	7	99.53 %	7	0
MAD2L1	5	98.61 %	5	0
MAP2K3	12	100.00 %	12	0
NFKB1	24	100.00 %	24	0
NIPBL	46	99.69 %	46	0
NOTCH1	34	99.95 %	34	0
NPM1	12	100.00 %	12	0
NRAS	4	100.00 %	4	0
NUP98	34	99.51 %	34	0
RAD21	13	100.00 %	13	0
RUNX1	12	97.03 %	11	1
SETBP1	6	100.00 %	6	0
SF3B1	27	99.85 %	27	0
SGOL1	8	100.00 %	8	0
SMAD7	4	100.00 %	4	0
SMC1A	26	100.00 %	26	0
SMC3	29	98.56 %	27	2
SPERT	4	100.00 %	4	0
SPI1	5	99.14 %	5	0
SRSF2	3	97.48 %	3	0
STAG2	34	100.00 %	34	0
STAT3	23	99.86 %	23	0
SUZ12	16	92.48 %	14	2
TET2	10	99.34 %	10	0
TGFB1	7	100.00 %	7	0
TLR2	1	100.00 %	1	0
TP53	14	97.53 %	13	1
TUBG1	11	95.05 %	9	2
U2AF1	9	100.00 %	9	0
ZRSR2	12	94.77 %	10	2

Table 7: The gene panel for targeted resequencing

4.2 Patient samples

Genomic DNA (gDNA) was extracted from 90 patient samples by our collaborator at the University Hospital of Düsseldorf and was sent to the German Cancer Research center (DKFZ) in Heidelberg on dry ice. DNA concentration was measured by a Qubit fluorometer which showed a significant difference in comparison with NanoDrop (**Figure 7**). All further analysis steps were therefore based on Qubit fluorometer results. 225 ng of each sample were used for library preparation.

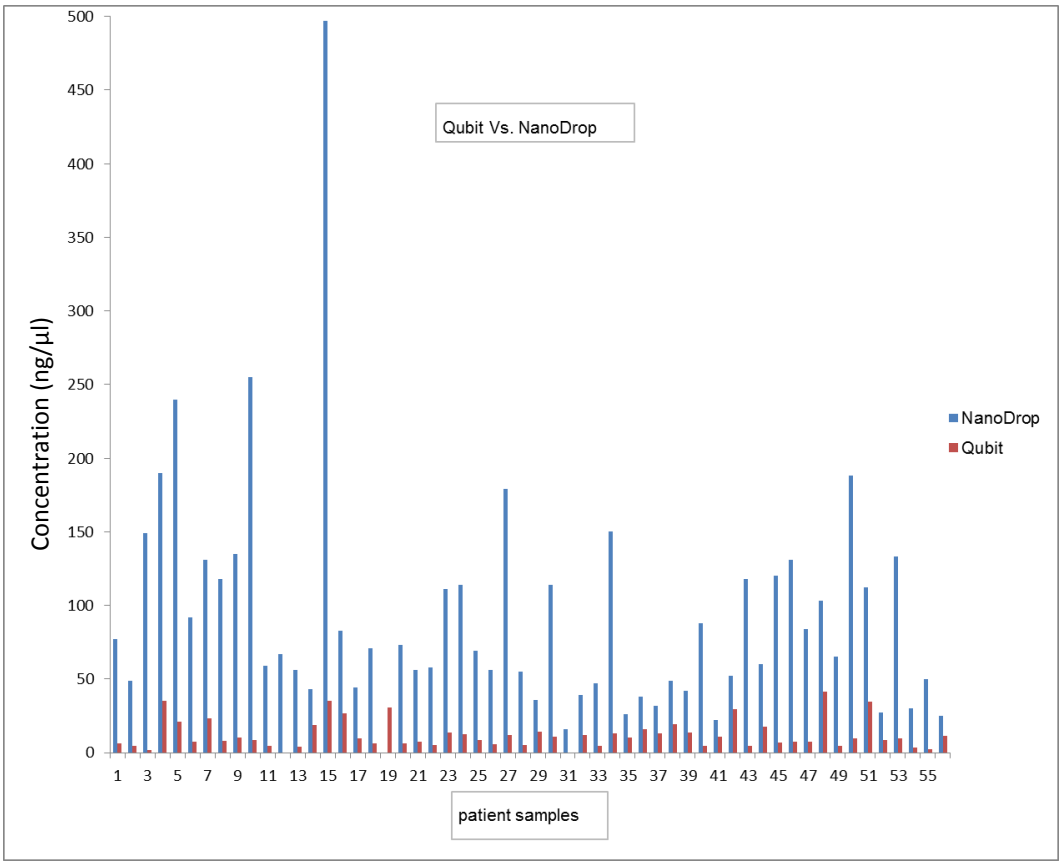


Figure 7: DNA concentrations measured by Qubit versus NanoDrop

4.3 Library preparation

We have created a list of 63 genes with recurrent mutations in MDS by literature search. The genes cover a broad range of biological phenomena as described in the introduction section. We have created a customized target enrichment kit using Agilent SureDesign software by following the instructions for this 63 gene panel covering their complete exome and some parts of the surrounding introns. Briefly, gDNA was digested by 16 different restriction enzymes. Then, the collection of digested restriction fragments was hybridized to the Haloplex probe capture library. After that, the hybridized DNA-Haloplex probes, containing biotin, was captured on streptavidin beads. Finally, the captured target regions were PCR amplified to produce a sequencing ready region (**Figure 8**).

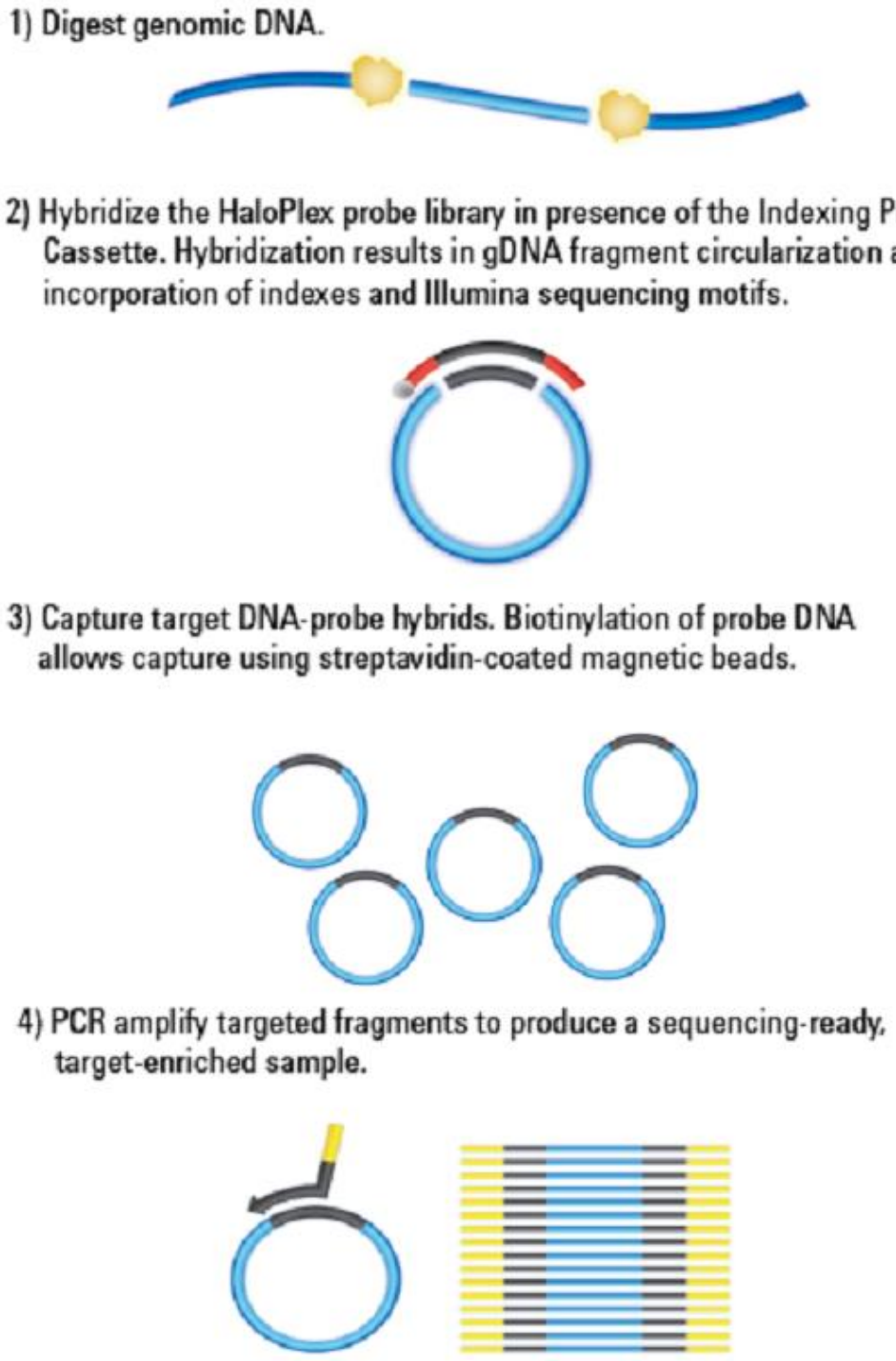


Figure 8: Overall Haloplex target-enriched sequencing sample preparation workflow

Step 1. Digest genomic DNA with restriction enzymes

In this step, gDNA samples are digested by 16 different restriction enzymes to create a library of gDNA restriction fragments.

1) Use the Qubit dsDNA BR Assay or PicoGreen staining kit to determine the concentration of your gDNA samples.
2) Prepare the DNA samples for the run. For 12-reaction runs, prepare 11 gDNA samples and one Enrichment Control DNA sample.

a. In separate 0.2-mL PCR tubes, dilute 225 ng of each gDNA sample in 45 µL nuclease-free water, for a final DNA concentration of 5 ng/µL. Store on ice
b. In a separate 0.2-mL PCR tube, dispense 45 µL of the supplied Enrichment Control DNA (ECD). Store on ice

3) Prepare the Restriction Enzyme Master Mix strip

The gDNA is digested in eight different restriction reactions, each containing two restriction enzymes. The 16 restriction enzymes are provided in two 8-well strip tubes that are distinguished by red and green color markers. Enzymes are combined from corresponding wells of the red- and green-marked strip tubes, along with restriction buffer and BSA to make eight different RE Master Mixes. **Figure 9** illustrates how to prepare the 8-well Restriction Enzyme Master Mix strip for a 12-sample run using the steps detailed below.

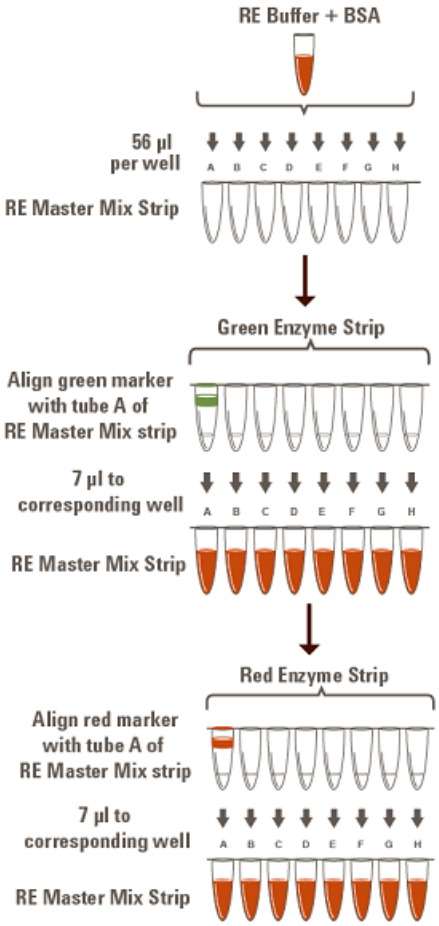


Figure 9: Preparation of the Restriction Enzyme Master Mix Strip for 12-sample run

a. Combine the amounts of RE Buffer and BSA Solution indicated in the table below in a 1.5-mL tube. Mix by vortexing briefly

Reagent	Volume for 1 Reaction	Volume for 12 Reactions (includes excess)
RE Buffer	34.0 µL	476 µL
BSA Solution	0.85 µL	11.9 µL
Total Volume	**34.85 µL**	**487.9 µL**

b. To begin preparation of the Restriction Enzyme Master Mix Strip, dispense the appropriate volume of the RE Buffer/BSA mixture to each well of an 8-well strip tube.

Reagent	Volume for 1 Reaction	Volume for 12 Reactions (includes excess)
RE Buffer/BSA mixture	4 µL	56 µL

c. Using a multichannel pipette, add the appropriate volume of each enzyme from the Green Enzyme Strip, with green marker aligned with tube A, to corresponding tubes A to H of the Restriction Enzyme Master Mix Strip.

Reagent	Volume for 1 Reaction	Volume for 12 Reactions (includes excess)
RE Enzymes from Green Enzyme Strip	0.5 µL	7 µL

d. Using a multichannel pipette, add the appropriate volume of each enzyme from the Red Enzyme Strip, with red marker aligned with tube A, to each corresponding tube A to H of the same Restriction Enzyme Master Mix Strip.

Reagent	Volume for 1 Reaction	Volume for 12 Reactions (includes excess)
RE Enzymes from Red Enzyme Strip	0.5 µL	7 µL

e. Mix by gentle vortexing and then spin briefly.
f. Keep the Restriction Enzyme Master Mix Strip on ice until it is used in step 4.

4) Aliquot the Restriction Enzyme Master Mixes to the rows of a 96-well plate to be used as the restriction digest reaction plate.

a. Align the Restriction Enzyme Master Mix Strip, prepared in step 3, along the vertical side of a 96-well PCR plate as shown below.
b. Using a multichannel pipette, carefully distribute 5 µL of each RE master mix row-wise into each well of the plate.

Figure 10: Distribution of the restriction enzyme mastermix into 96-well plate

Each row of the 96-well plate now contains 5 µL per well of the same restriction enzyme combination.

5) Aliquot DNA samples into the 96-well Restriction Digest Reaction Plate(s).
a. Align the DNA samples (11 gDNA samples and the ECD sample), prepared in step 2, along the horizontal side of the digestion reaction plate(s) as shown below.

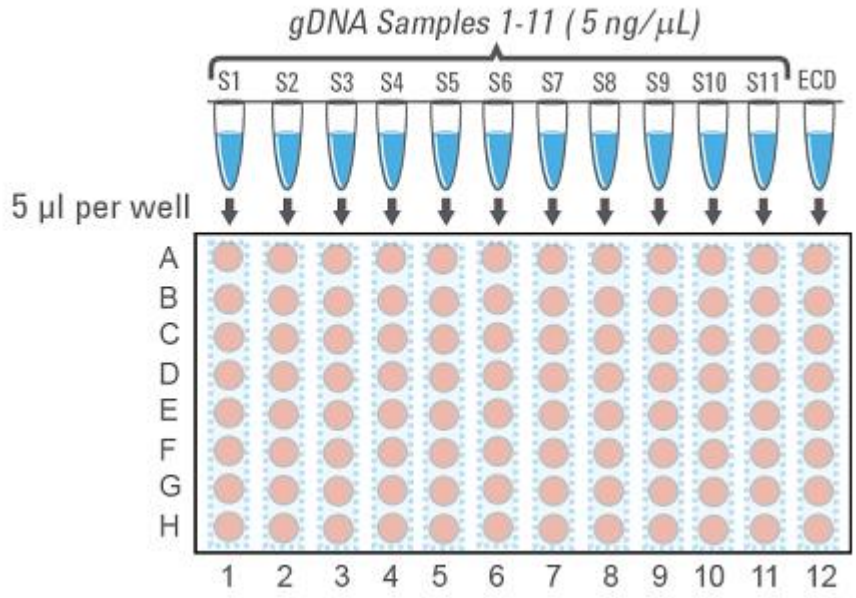


Figure 11: Distribution of the samples into the 96-well plate

b. Carefully distribute 5 µL of DNA samples column-wise into each well of the digestion reaction plate. If using a multichannel pipette, visually inspect pipette tips for equal volumes before dispensing.
c. Seal the plate thoroughly with adhesive plastic film

6) Carefully vortex the plate to mix the digestion reactions.
7) Briefly spin the plate in a plate centrifuge.
8) Place the Restriction Digest Reaction Plate in a thermal cycler and run the program in Table below, using a heated lid.

Table 5 Thermal cycler program for HaloPlex restriction digestion

Step	Temperature	Time
Step 1	37°C	30 minutes
Step 2	8°C	Hold

9) Validate the restriction digestion reaction by electrophoretic analysis of the Enrichment Control DNA (ECD) reactions. Keep the Restriction Digest Reaction Plate on ice during validation.
a. Transfer 4 µL of each ECD digestion reaction from wells of the digestion reaction plate to fresh 0.2-mL PCR tubes.
b. Incubate the removed 4-µL samples at 80°C for 5 minutes to inactivate the restriction enzymes.
c. Analyze the prepared samples using gel electrophoresis as follow:

I. Use The E-Gel® iBase™ Power System and E-Gel®4% and the corresponding program as shown below

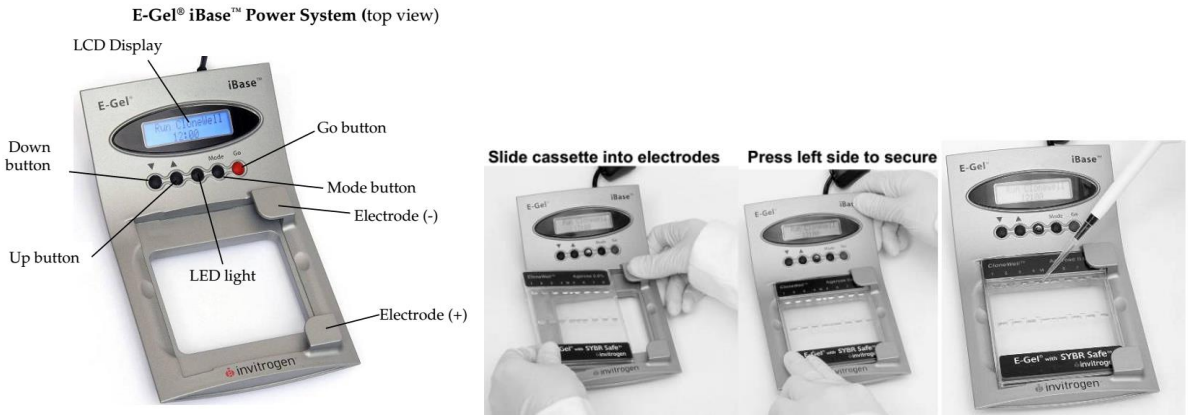


Figure 12: E-Gel® iBase™ Power System and E-Gel®4% and how to load samples

II. Select the corresponding program for E-Gel®4% and set the time for 30 minute
III. Open the package and remove the gel. Gently remove the comb(s) from the gel.
IV. Slide the cassette into the two electrode connections on the E-Gel® iBase™ device. Press on the left side of the cassette to secure it into the iBase™ Power System. The two electrodes on the right side of the gel cassette must be in contact with the two electrode connections on the base as shown above. The LED illuminates with a steady red light to show that the cassette is correctly inserted.
V. Load your samples and the appropriate molecular weight markers, and add water to any empty wells as shown above.
VI. To start electrophoresis press the Go button, a green light illuminates to show that the run is in progress. The LCD displays the countdown time while the run is in progress.
VII. The run will stop automatically when the programmed time has elapsed.
VIII. Press and release the Go button to stop the beeping.
IX. The ECD sample contains genomic DNA mixed with an 800-bp PCR product that contains restriction sites for all the enzymes used in the digestion protocol. When analyzing validation results, the undigested control should have gDNA bands at >2.5 kbp and a PCR product band at 800 bp. Each of the eight digested ECD samples should have a smear of gDNA restriction fragments between 100 and 2500 bp, overlaid with three predominant bands at approximately 125, 225 and 450 bp. These three bands correspond to the 800-bp PCR product-derived restriction fragments, and precise sizes will differ after digestion in each of the eight RE master mixes.
X. Look at the gel under UV light and it should look like the picture below

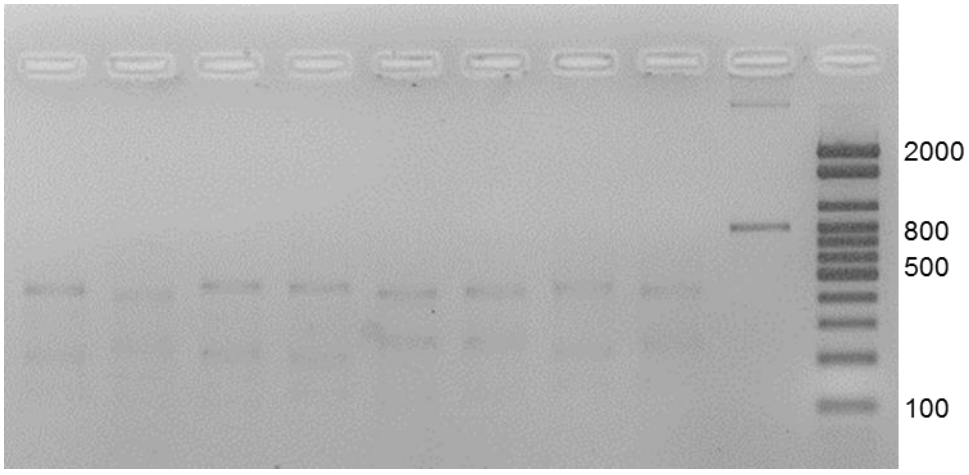


Figure 13: Validation of restriction digestion by gel electrophoresis

Lanes 1–8: ECD digestion reactions A–H, Lane 9: Undigested Enrichment Control DNA, Lane 10: LowRanger 100 bp DNA Ladder.

XI. If you do not have access to the E-Gel® iBase™ Power System and E-Gel®4% you can add 4 µl 6x TAE loading dye to each of the samples and load it on a 4% Agarose gel and run it for one hour. But you might not see the bands and smear as clearly as by E-Gel®4% and the run takes one hour. So I suggest using E-Gel® iBase™ Power System and E-Gel®4% since, you will need it later on for library purification.
If you do not continue to the next step, samples may be stored at –20°C for long-term storage. There are no more long-term stopping points until after the PCR amplification step.

Step 2. Hybridize digested DNA to HaloPlex probes for target enrichment and sample indexing

In this step, the collection of gDNA restriction fragments is hybridized to the HaloPlex probe capture library. The duration of the hybridization reaction is determined by the probe density of your design. Refer to the Certificate of Analysis provided with Box 1 of your kit to determine the hybridization conditions appropriate for your design. HaloPlex probes are designed to hybridize selectively to fragments originating from target regions of the genome and to direct circularization of the targeted DNA fragments. During the hybridization process, Illumina sequencing motifs including index sequences are incorporated into the targeted fragments.

1) Prepare a Hybridization Master Mix by combining the reagents in the Table below. Mix well by gentle vortexing, then spin the tube briefly.

Reagent	Volume for 1 Reaction	Volume for 12 Reactions (includes excess)
Hybridization Solution	50 µL	650 µL
HaloPlex Probe	20 µL	260 µL
Total Volume	**70 µL**	**910 µL**

2) Distribute 70 µL of the Hybridization Master Mix to each of 12 0.2-mL tubes
3) Add 10 µL of the appropriate Indexing Primer Cassette to each tube containing Hybridization Master Mix.
Be sure to add only one specific Indexing Primer Cassette to each hybridization tube, using different indexes for each sample to be multiplexed. Record the

identity of the Indexing Primer Cassette added to each tube for later sequence analysis.

4) Transfer digested DNA samples from the 96-well Restriction Digest Reaction Plate(s) directly into the hybridization reaction tubes prepared in step 3. Transfer all eight digestion reactions that correspond to one DNA sample into the appropriate hybridization reaction tube. After addition of each individual digest reaction to the hybridization solution, mix by pipetting before adding the next digest reaction to ensure complete inactivation of the enzymes.

For the ECD sample, add 32 µL of nuclease-free water, in addition to the digested DNA samples, to compensate for the volume removed for digest validation.

After pooling, each hybridization reaction contains the following components:

- 70 µL Hybridization Master Mix
- 10 µL Indexing Primer Cassette
- approximately 80 µL pooled digested DNA samples

5) Vortex the mixtures briefly and then spin tubes briefly.
6) Place the hybridization reaction tubes in a thermal cycler. Run the appropriate program in Table below, using the hybridization duration listed on the Certificate of Analysis. Use a heated lid. Do not include a low-temperature hold step in the thermal cycler program. Incubation at 54°C for more than the indicated time is not recommended.

Step	Temperature	Time (Duration of Step)
		Custom Designs with <20,000 probes (see Certificate of Analysis)†
Step 1	95°C	10 minutes
Step 2	54°C	3 hours

* Thermal cyclers that use calculated temperature methods cannot be set to 160 µL reaction volumes. In that case, enter the maximum possible volume.

Step 3. Capture the target DNA

In this step, the circularized target DNA-HaloPlex probe hybrids, containing biotin, are captured on streptavidin beads.

1) Remove reagents to be used in upcoming protocol steps from cold storage and allow the solutions to reach room temperature:

•From –20°C storage, remove the Capture Solution, Wash Solution, Ligation Solution and SSC Buffer.

•From +4°C storage, remove the HaloPlex Magnetic Beads.

2) Obtain or prepare 0.5 µL per sample, plus excess, of 2 M acetic acid, for use on page 33.
3) Prepare 25 µL per sample, plus excess, of fresh 50 mM NaOH for use in the DNA elution step on page 34. Prepare the 50 mM NaOH solution from a 10M NaOH stock solution.
4) Vigorously resuspend the provided HaloPlex Magnetic Beads on a vortex mixer. The magnetic beads settle during storage.
5) Prepare 40 µL (1 Volume) of HaloPlex Magnetic Beads per hybridization sample, plus excess, for the capture reaction:

a. Transfer the appropriate volume of bead suspension to a 1.5-mL tube.

Reagent	Volume for 1 Reaction	Volume for 12 Reactions (includes excess)
HaloPlex Magnetic Bead suspension	40 µL	520 µL

b. Put the tube into a 1.5 mL tube-compatible magnetic rack for 5 minutes.
c. After verifying that the solution has cleared, carefully remove and discard the supernatant using a pipette.
d. Add an equivalent volume of Capture Solution to the beads and resuspend by pipetting up and down.

Reagent	Volume for 1 Reaction	Volume for 12 Reactions (includes excess)
Capture Solution	40 µL	520 µL

e. Remove the hybridization reactions from the thermal cycler and immediately add 40 µL of the prepared bead suspension to each 160-µL hybridization reaction.
6) Remove the hybridization reactions from the thermal cycler and immediately add 40 µL of the prepared bead suspension to each 160-µL hybridization reaction
7) After adding the magnetic beads, mix the capture reactions thoroughly by pipetting up and down 15 times using a 100-µL pipette set to 80 µL.
8) Incubate the capture reactions at room temperature for 15 minutes.
9) Briefly spin the tubes in a desktop centrifuge and then transfer the tubes to the Agencourt SPRIPlate Super Magnet magnetic plate.
10) Wait for the solution to clear (about 30 seconds), then remove and discard the supernatant using a pipette set to 200 µL.
11) Wash the bead-bound samples:

a. Remove the capture reaction tubes from the magnetic plate and add 100 µL of Wash Solution to each tube.
b. Resuspend the beads thoroughly by pipetting up and down 10 times using a 100-µL multichannel pipette set to 80 µL.
c. Incubate the tubes in a thermal cycler at 46°C for 10 minutes, using a heated lid.

Do not include a low-temperature hold step in the thermal cycler program following the 10-minute incubation.

d. Briefly spin the tubes in a desktop centrifuge at room temperature and then transfer the tubes to the magnetic plate.
e. Wait for the solution to clear (about 30 seconds), then carefully remove and discard the supernatant using a pipette set to 120 µL. If necessary, carefully remove any residual liquid with a 20-µL volume pipette.

Step 4. Ligate the captured, circularized fragments

In this step, DNA ligase is added to the capture reaction to close nicks in the circularized HaloPlex probe-target DNA hybrids.

1) Prepare a DNA ligation master mix by combining the reagents in the following table. Mix the components thoroughly by gentle vortexing then spin the tube briefly.

Reagent	Volume for 1 Reaction	Volume for 12 Reactions (includes excess)
Ligation Solution	47.5 µL	617.5 µL
DNA Ligase	2.5 µL	32.5 µL
Total Volume	**50 µL**	**650 µL**

2) Add 50 µL of the DNA ligation master mix to the beads in each DNA capture reaction tube.
3) Resuspend the beads thoroughly by pipetting up and down 15 times using a 100-µL multichannel pipette set to 40 µL
4) Incubate the tubes in a thermal cycler at 55°C for 10 minutes, using a heated lid. The thermal cycler may be programmed to include a 4°C hold step following the 10-minute incubation. During the 10-minute incubation, prepare the PCR master mix as specified in the following step.

Step 5. Prepare the PCR Master Mix

In this step, you prepare a PCR master mix for the captured target DNA amplification step on page 35.

1) Prepare the PCR master mix by combining the reagents in the following table.

Reagent	Volume for 1 reaction	Volume for 12 reactions (includes excess)
Nuclease-free water	16.1 µL	209.3 µL
5X Herculase II Reaction Buffer	10 µL	130 µL
dNTPs (100 mM, 25 mM for each dNTP)	0.4 µL	5.2 µL
Primer 1 (25 µM)	1 µL	13 µL
Primer 2 (25 µM)	1 µL	13 µL
2 M Acetic acid	0.5 µL	6.5 µL
Herculase II Fusion DNA Polymerase	1 µL	13 µL
Total	**30 µL**	**390 µL**

2) Mix the master mix components by gentle vortexing, then distribute 30-µL aliquots to fresh 0.2-mL reaction tubes.
3) Store the tubes on ice until they are used in “Step 7. PCR amplify the captured target libraries” on page 35.

Step 6. Elute captured DNA with NaOH

When the 10-minute ligation reaction period is complete, proceed with the following steps to elute the captured DNA libaries.

1) Briefly spin the ligation reaction tubes in a desktop centrifuge and then transfer the tubes to the magnetic plate.
2) Wait for the solution to clear (about 30 seconds), then carefully remove and discard the supernatant using a pipette set to 50 µL.
3) Remove the tubes from the magnetic plate and add 100 µL of the SSC Buffer provided with the kit to each tube.
4) Resuspend the beads thoroughly by pipetting up and down 10 times using a 100-µL multichannel pipette set to 80 µL.
5) Briefly spin the tubes and then return the tubes to the magnetic plate.
6) Wait for the solution to clear (about 30 seconds), then carefully remove and discard the SSC Buffer using a multichannel pipette set to 120 µL. If necessary, carefully remove any residual liquid with a 20-µL volume pipette.
7) Add 25 µL of 50 mM NaOH, which was freshly-prepared on page 29, to each tube.
8) Resuspend the beads thoroughly by pipetting up and down 10 times using a 100-µL multichannel pipette set to 15 µL.
9) Incubate samples for 1 minute at room temperature to allow elution of the captured DNA.

10) Briefly spin the tubes and then transfer the tubes to the magnetic plate. Proceed immediately to PCR amplification in the following section.

Step 7. PCR amplify the captured target libraries

1) Prepare amplification reactions by transferring 20 µL of cleared supernatant from each tube on the magnetic plate to a PCR Master Mix tube held on ice (from page 33).
2) Mix by gentle vortexing and then spin briefly to collect the liquid.
3) Place the amplification reaction tubes in a thermal cycler and run the program in Table below, using a heated lid.

The optimal amplification cycle number varies for each HaloPlex Probe design. Consult the Certificate of Analysis (provided with HaloPlex Target Enrichment System Box 1) for the PCR cycling recommendation for your probe. In this case 20 cycles.

Segment	Number of Cycles	Temperature	Time
1	1	98°C	2 minutes
2	20	98°C	30 seconds
		60°C	30 seconds
		72°C	1 minute
3	1	72°C	10 minutes
4	1	8°C	Hold

If you do not continue to the next step, PCR products may be stored at –20°C for up to 72 hours or at 8°C overnight. For best results, however, purify PCR products as soon as possible.

Step 8. Purify the amplified target libraries

Purification can be done either with Agencourt AMPure beads following the protocol or gel purification using the following protocol. In author's experience, gel purification is preferred therefore it is brought below. The libraries were prepared in tow round each round 48 libraries. If the PCR product can be visualized on the gel, then the gel purification should work. In the first round some of libraries were purified with AMPure beads and if the adapter peak was high, the libraries were purified with gel purification (**Figure 14** and **Table 8**).

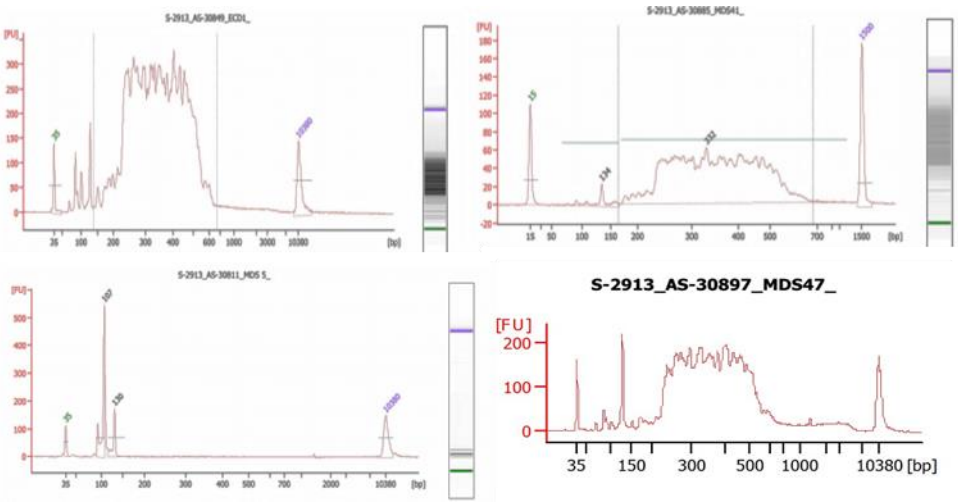


Figure 14: Representative enriched libraries

Libraries 1 and 2 in this figure are fine but, as you can see library 3 failed completely and library 4 is good but there is an adapter peak that could cause problem during sequencing and has to be removed

Name	Index	Peak Size range indicated (bp)	average size	library MW	Concentration found [ng/µl]	volume	final ng	purified?
MDS6	6	158-698	358	232700	3,5	20,0	70,9	yes perfect!
MDS9	9	162-656	358	232700	2,8	20,0	56,6	yes perfect!
MDS35	37	138-644	338	219700	5,4	20,0	107,7	yes perfect!
MDS23	25	160-565	334	217100	4,5	20,0	89,8	yes perfect!
MDS37	39	163-663	366	237900	7,6	20,0	152,1	yes perfect!
MDS44	46	140-641	355	230750	6,4	20,0	128,7	yes perfect!
MDS46	48	200-653	371	241150	2,8	20,0	56,0	yes perfect!
MDS20	21	168-572	351	228150	28,4	20,0	568,8	yes perfect!
MDS22	23	168-592	356	231400	22,8	20,0	456,6	yes perfect!
MDS29	31	165-666	369	239850	28,0	20,0	560,0	yes perfect!
MDS28	30	164-584	354	230100	18,6	20,0	372,4	yes perfect!

Table 8: Gel purified libraries

The libraries with a big adapter peak were gel purified in order to remove the peak

In the second round, all the libraries were purified using gel purification as follow:

1) Spin the PCR product tubes briefly
2) Transfer the PCR product to E-Gel®4% and run it using Gel® iBase™ Power System for 30 minutes
3) Visualize the amplicons using UV light and like the Figure below:

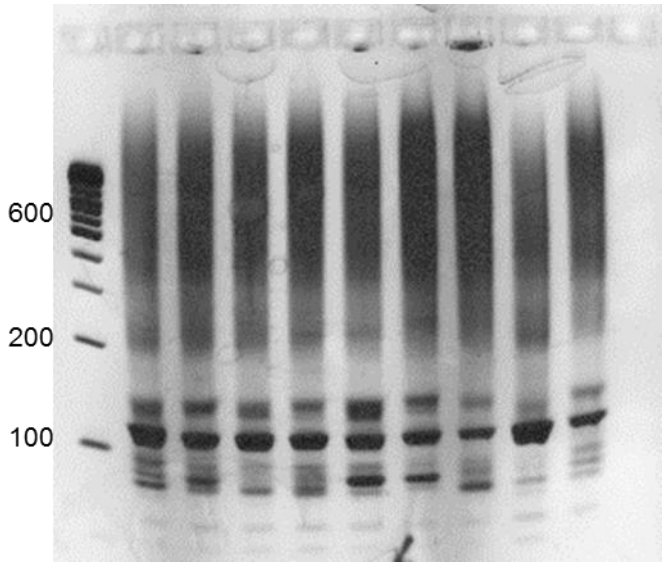


Figure 15: Sample libraries visualized under UV

4) cut the gel from 200 to 600 bp as follow:

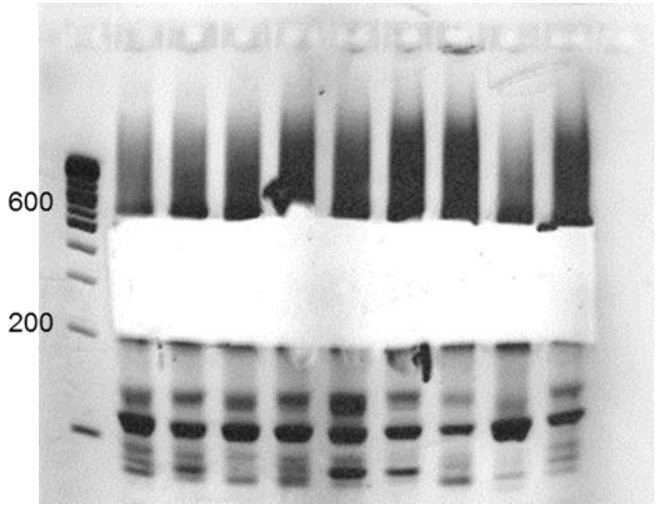


Figure 16: Sample libraries visualized and cut under UV

5) Purify the excised gels with Qiagen MinElute Gel Extraction Kit by following the protocol
6) Check the libraries on Agilent Bioanalyzer 2100. Representative libraries are brought in
7)

8) **Figure** 14**.**
9) Select the libraries with good quality for equimolar pooling
10) Calculate the required sample amount for 10 nmol/L equimolar pooling using the Average library size and the concentration given by Bioanalyzer 2100 using (**Table 9**)**.**

383	bp	Average Size of Library
248950	daltons	650 daltons/bp
0,24895	ug/pmol	10E6 daltons = 1 ug/pmol
248,95	ng/pmol	
13,1	ng/ul	Template Concentration
0,053	pmol/ul	
0,053	nmol/ml	
52,621	nmol/L	
10	nmol/L	Required for ~30K clusters
5,26	Dilution Factor	
1,90	**ul template**	
8,10	**EB Buffer**	
10,00	**Total**	
RED	**User Input**	
Purple	**Sample Prep**	

Table 9: Template amount required for equimolar pooling

This table is an example of the excel table used to calculate the amount of template and buffer in order to reach an equimolar of 10 nmol/ liter. The parts written in red are user input including Average library size in base pair (bp) in this case 383, template concentration ng/µl here 13.1, and the required equimolar concentration for ~30K clusters here 10 nmol/L. The parts written in purple are given results which include the amount of template and buffer. For each library Average library size and template concentration in given and the amount of template and buffer in microliter will be calculated and appears. For pulling, the given amounts of template and buffer will be pulled together.

The required template for equimolar pooling can be calculated using the Average library size and template concentration which is measured by Bioanalyzer 2100.

11) Pool the libraries and check the quality on a Bioanalyzer 2100 like in **Figure 17** and **Figure 18.**

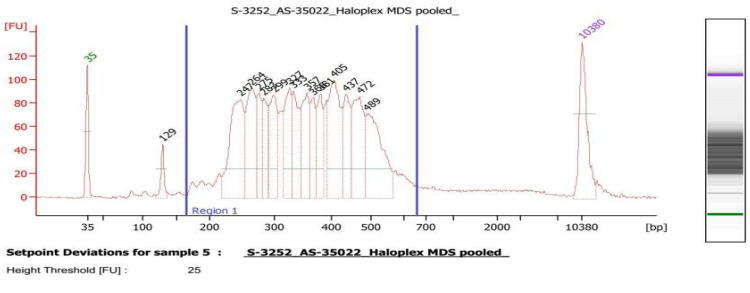


Figure 17: 31 pooled libraries

Libraries were pooled together at equimolar ratio of 10 picomolar and checked for quality on Agilent Bioanalyzer 2100

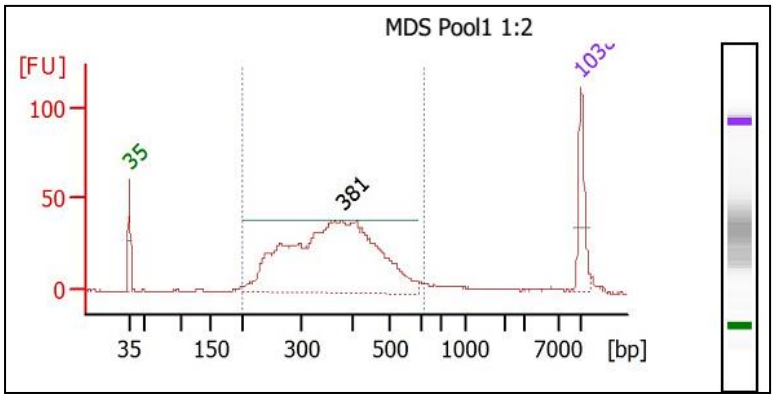


Figure 18: 32 pooled libraries

Libraries were pooled together at equimolar ratio of 10 picomolar and checked for quality on Agilent Bioanalyzer 2100

12) Send the pools for sequencing on an Ilumina HiSeq 2000 for paired end 100 base pair sequencing

Libraries were prepared as above and sent for quality control on an Agilent Bioanalyzer 2100. The library pools were sent for sequencing on an Ilumina HiSeq 2000 for paired end 100 base pair sequencing.

4.4 Data analysis

We received a total of 90 paired FASTQ files with removed indexes. These data were analyzed by established bioinformatics pipelines developed by our collaborators at the DKFZ.

4.5 Sanger sequencing

The single nucleotide variations (SNVs) in TP53, STAG2, and AKAP9 were verified by Sanger sequencing as follow:

The location of the SNVs in transcript was found in Ensemble. The gene was searched on Ensemble and the corresponding transcript was selected for cDNA sequence. The genomic sequence or, if the SNV was in one exon, the cDNA sequence was copied in APE plasmid software. The location of the mutation was found in the sequence and PCR primers were designed approximately 300 base pair upstream and downstream of the SNV. The list of PCR primers is brought in **Table 10.**

TP53 Mutation verification	**sequence**
TP53_SNVs_Fw	AGGCCTCTGATTCCTCACTGATTG
TP53_SNVs_Rev	GTCAGAGGCAAGCAGAGGCT
SA2 Mutation verification	
SA2_S1075X_Fw	GTTCATGACCTTTCAGATGTCACTCCG
SA2_S1075X_Rev	tccctatgcatacagtgtagcacaga
SA2_R146X_Fw	CTTGTGCATACTGAGAATAGATAGCA
SA2_R146X_Rev	CCCAGCCTAATGCTTACAATTTAATAAT
SA2_Y433H_Fw	GTACTGTTAATATGCTTAGAATTAGGACGT
SA2_Y433H_Rev	TGTGAAAGCTTCGATATGATCTGTAGT
AKAP9 mutation verification	
AKAP9_N408H_Fw	ACTAGGAGAATTACAAGAACAGATTGTGC
AKAP9_N408H_Rev	CTCCATCTGTGCCATGTGTTGT

Table 10.PCR primers for Sanger sequencing

Five microliter PCR product was run on a gel. When one single band with the right size was observed, the PCR product was purified using The MinElute PCR Purification Kit from QIAGEN following the protocol. Importantly, in the first step 3 Molar sodium acetate wit PH 5 was used as suggested in the protocol otherwise the product was lost. The purified PCR product was sent to GATC for sequencing. The sequence was blasted against the cDNA or gDNA sequence and the mismatches were found. The SNVs were verified by looking at the sequencing chromatograms and comparing them with the control.

4.6 Characterization of the created knockout cell lines

4.6.1 Protein Expression by Western Blotting

Western blot was done according to the established protocol in the lab. First, protein lysates were prepared, followed by SDS-page gel separation and finally blotting. The following materials were used:

A: Preparation of protein/cell lysates

Solutions:

RIPA:

50mM Tris-Cl pH7.4
1% NP-40
0.25% Sodium deoxycholate
150 mM NaCl
1mM EDTA
1 complete tablet (Protease inhibitor cocktail, Roche) per 50 ml
5 Phospho-Stop tablets (Roche) per 50 ml

6xSDS-sample buffer:

7ml 1M Tris-Cl/0,4% SDS, pH6.8
3ml glycerol
1g SDS
0.93g DTT
1,2 mg bromphenol blue
Add VE-H2O to 10ml

B: SDS gel electrophoresis

Solutions and Chemicals:

- 30 % Acrylamide/Bisacrylamide
- TEMED
- 10 % APS
- 1,5 M Tris pH 8.8 0.4 % SDS (SDS-PAGE)
- 1,0 M Tris pH 6.8 0.4 % SDS (SDS-PAGE)

- 20 % SDS (SDS-PAGE)

10x Running Buffer for SDS-PAGE:

30.3 g Tris
144.1 g Glycine
10 g SDS
dissolve in 1000 ml H2O
to obtain working solution take 100 ml 10x buffer
Adjust volume with H2O to 1000 ml

C: Western Blot

Materials:

- TBS supplemented with 0.1% Tween-20 (TBS-T; common container: refill when empty!)
- TBS-T supplemented with 5% (w/v) dry milk (TBS-T/5% milk; prepare freshly!)
- Membrane (PVDF or nitrocellulose)
- Methanol (in case PVDF membranes are to be used)
- Ponceau S working solution (usually commercial, ready-to-use, 0.1% solution)
- Primary and secondary antibodies
- ECL reagents
- Films for exposure

Buffers:

20X Borate buffer

25 g Boric acid
7.45 g EDTA
Dissolve in 1000 ml H2O pH 8.8 (adjust with NaOH: about 30 pellets)
Dilute 1:20 to obtain working solution (1x Borate Buffer)

Antibodies:

The following antibodies were used: RAD21 (D213) Antibody (Cell Signaling Technologies 4321), SA-2 Antibody (J-12) (Santa Cruz Biotechnologies sc-81852**).** The antibodies were diluted according to the company recommendations.

4.6.2 Protein Expression by Immunofluorescence

In this procedure the cells were first cultured on coverslips and after reaching confluency the cells were fixed and stained as follows:

Fixation Methods:

1) Methanol:

Carefully add 100% methanol (cold, stored at -20°C) to the coverslips, they should be covered by methanol incubate for 10 min at room temperature dry coverslips on filter paper fixed coverslips can be stored at -20°C

2) Methanol/Acetone:

Add methanol/acetone-mixture (1:1, cold, stored at -20°C) to the coverslips, they should be covered by methanol/acentone incubate for 7 min at room temperature dry coverslips on filter paper fixed coverslips can be stored at -20°C

3) 4% PFA:

Carefully add 4% PFA/PBS to the coverslips, they should be covered incubate for 15 minutes at room temperature wash coverslips twice with 1xPBS fixed coverslips, covered by 1xPBS, can be stored at 4°C for up to two weeks permeabilise cells prior to immunofluorescence staining add PBS/0,2%Triton-X-100 to coverslips incubate for 5min at room temperature wash coverslips twice with 1xPBS

4) PHEM-Extraction:

Using this method you extract cytosolic components prior to fixation (reduced cytosolic background)

1xPHEM-buffer: 60mM PIPES, 25mM HEPES, 8mM EGTA, 2mM MgCl2, pH = 6.9

Carefully add PHEM-buffer supplemented with 0.5%TritonX-100 to the coverslips, they should be covered incubate for up to 5 minute depending on the cells at room temperature wash coverslips once with PHEM-buffer without TritonX-100, they should be covered fix cells with one of described fixation methods above

Immunofluorescence staining:

The cells are extracted by PHEM if necessary and fixed using one of the above methods. Staining was done according to the established protocol in the lab.

4.7 Crisper/Cas9 knock out

We used Crisper/Cas9 technology in order to knock out STAG2. Cas9 is a nuclease guided by small RNAs (sgRNA) through Watson-Crick base pairing with target DNA (**Figure 19**) *(33)*. SgRNA consists of a 20-nucleotide guide sequence and a scaffold. The guide sequence recognizes the target cleavage sites through protospacer adjacent motif (PAM) which is a 5´-NGG-3´ motif and binds directly upstream of it and Cas9 mediates a double strand cleavage 3 bp upstream of PAM sequence.

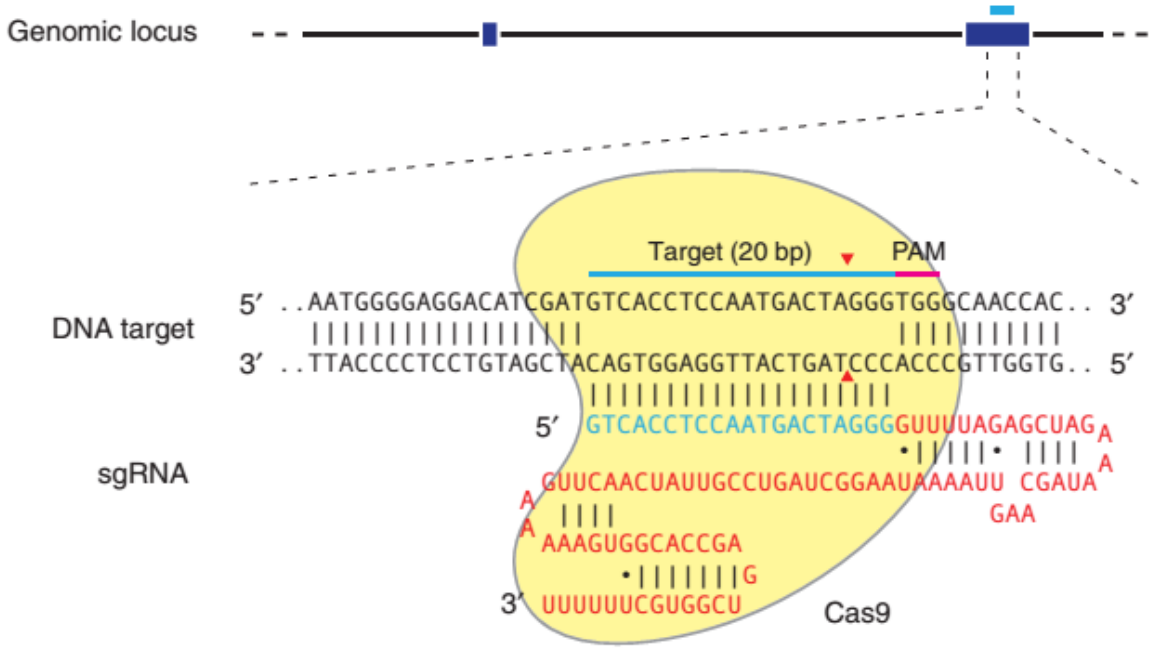


Figure 19: Schematic of the RNA-guided Cas9 nuclease

The Cas9 nuclease from S. pyogenes (in yellow) is targeted to genomic DNA (shown as example is the human EMX1locus) by an sgRNA consisting of a 20-nt guide sequence (blue) and a scaffold (red). The guide sequence pairs with the DNA target (blue bar on top strand), directly upstream of a requisite 5′-NGG adjacent motif (PAM; pink). Cas9 mediates a DSB ~3 bp upstream of the PAM (red triangle). Adapted from (33).

Cas9 promotes genome editing by stimulating a DSB at a target genomic locus. Upon cleavage by Cas9, the target locus typically undergoes one of two major pathways for DNA damage repair (**Figure 20**): the error-prone nonhomologous end joining (NHEJ) or the high-fidelity homology-directed repair (HDR) pathway, both of which can be used to achieve a desired editing outcome. In the absence of a repair template, DSBs are re-ligated through the NHEJ process, which leaves scars in the form of insertion/deletion (indel) mutations. NHEJ can be harnessed to mediate gene knockouts, as indels occurring within a coding exon can lead to frameshift mutations and premature stop codons *(33)*.

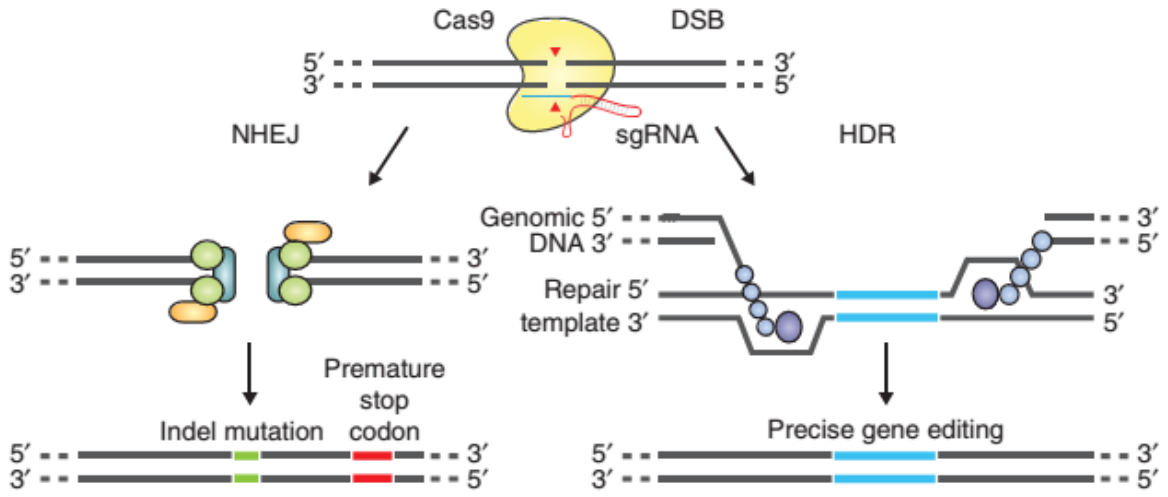


Figure 20: DSB repair promotes gene editing

DSBs induced by Cas9 (yellow) can be repaired in one of two ways. In the error-prone NHEJ pathway, the ends of a DSB are processed by endogenous DNA repair machinery and rejoined, which can result in random indel mutations at the site of junction. Indel mutations occurring within the coding region of a gene can result in frameshifts and the creation of a premature stop codon, resulting in gene knockout. Alternatively, a repair template in the form of a plasmid or a single stranded oligodeoxy nucleotide (ssODN) can be supplied to leverage the HDR pathway, which allows high fidelity and precise editing. Single-stranded nicks to the DNA can also induce HDR. Adapted from (33).

Here, we describe considerations for designing the 20-nt guide sequence, protocols for rapid construction and finally the use of the Cas9 nuclease to mediate NHEJ-based genome modifications in the HCT116-p53$^{+/+}$ and HCT116-p53$^{-/-}$ cell lines (**Figure 21**).

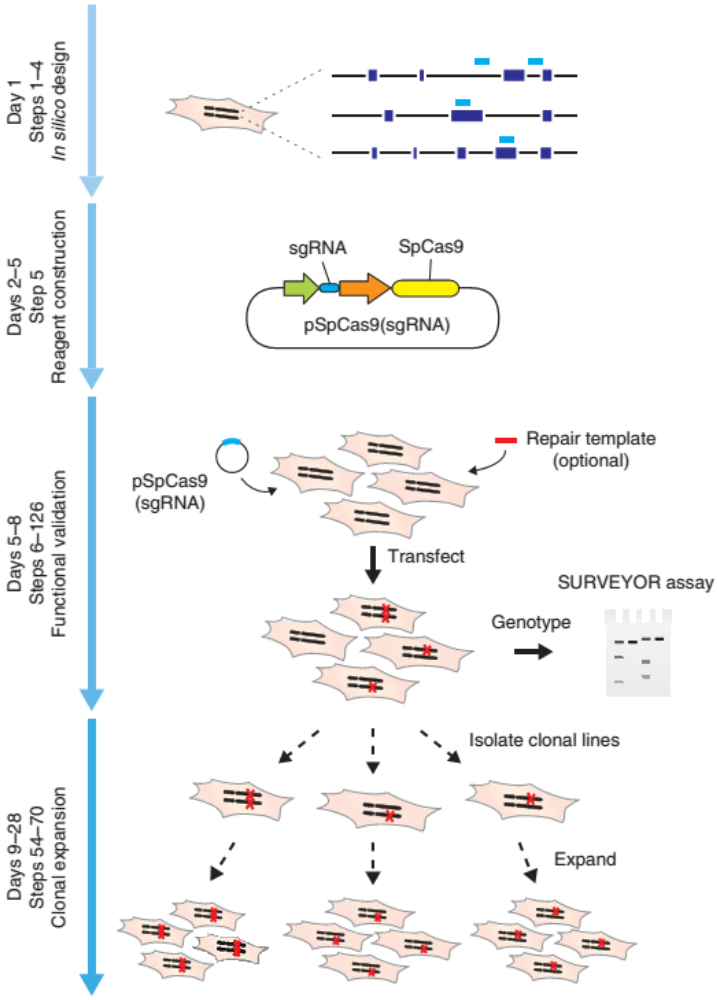


Figure 21: Creating knockout cell lines - timeline and overview

Steps for reagent design, construction, validation and cell line expansion are depicted. Custom sgRNAs (light blue bars) for each target, as well as genotyping primers, are designed in silico via the CRISPR Design Tool (http://tools.genomeengineering.org). sgRNA guide sequences can be cloned into an expression plasmid bearing both sgRNA scaffold backbone (BB) and Cas9, pSpCas9 (BB). The resulting plasmid is annotated as pSpCas9 (sgRNA). Completed and sequence-verified pSpCas9 (sgRNA) plasmids and optional repair templates for facilitating HDR are then transfected into cells and assayed for their ability to mediate targeted cleavage. Finally, transfected cells can be clonally expanded to derive isogenic cell lines with defined mutations. Adapted from (33).

Protocol: taken and modified from reference 33.

A: Design of targeting components and the use of the crispr Design tool

1. Input target genomic DNA sequence. An online CRISPR Design Tool (http://tools.genome-engineering.org) is provided that takes an input sequence (for example, a 1-kb genomic fragment from the region of interest), identifies and ranks

suitable target sites and computationally predicts off-target sites for each intended target. Alternatively, one can manually select guide sequences by identifying the 20-bp sequence directly upstream of any 5′-NGG.

2. Order necessary oligos and primers as specified by the online tool. If the cleavage site is chosen manually, the oligos or primers should be designed as described in **Figure 22**.

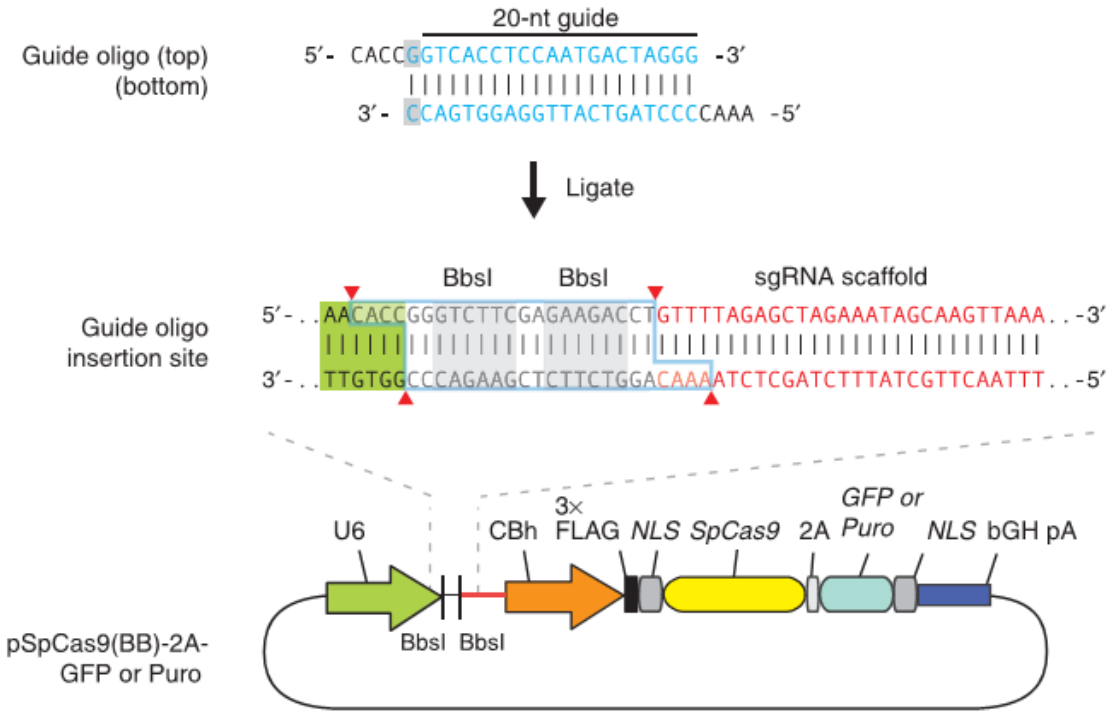


Figure 22. Cloning strategy

The guide oligos for the top strand example (blue) contain overhangs for ligation into the pair of BbsI sites in pSpCas9(BB), with the top and bottom strand orientations matching those of the genomic target (i.e., the top oligo is the 20-bp sequence preceding 5′-NGG in genomic DNA). Digestion of pSpCas9(BB) with BbsI allows the replacement of the Type II restriction sites (blue outline) with direct insertion of annealed oligos. Likewise, a G-C base pair (gray rectangle) is added at the 5′end of the guide sequence for U6 transcription, which does not adversely affect targeting efficiency. Alternate versions of pSpCas9(BB) also contain markers such as GFP or a puromycin resistance gene to aid the selection of transfected cells. Adapted from (33).

B: Cloning sgrna into the pspcas9(BB) vector for co-expression with cas9

NOTE:

Due to the simultaneous digestion-ligation step, the guide oligos CANNOT contain any BbsI enzyme site (i.e. the nucleotide sequence ‘GAAGAC’ or ‘GTCTTC’).

1. Phosphorylate and anneal each pair of oligos:

1 ul oligo 1 (100µM)
1 ul oligo 2 (100µM)
1 ul 10X T4 Ligation Buffer (NEB)
6.5 ul ddH2O
0.5 ul T4 PNK (NEB)
10 ul total

Anneal in a thermocycler using the following parameters:

37 °C 30 min

95°C 5 min and then ramp down to 25°C at 5°C /min

Dilute the annealed oligo 1:250 (250-fold).

2. Set up digestion-ligation reaction:

X ul pX330 or other backbone vector (100ng)
2 ul phosphorylated and annealed oligo duplex from step 1 (1:250 dilution)
2 ul 10X Tango buffer (or FastDigest Buffer)
1 ul DTT (10mM to a final concentration of 1mM)
1 ul ATP (10mM to a final concentration of 1mM)
1 ul FastDigest BbsI (Thermo Fisher Fermentas)
0.5 ul T7 DNA ligase
Y ul ddH2O
20 ul total

Incubate the ligation reaction in a thermocycler:

37°C 5 min

23°C 5 min Cycle the previous two steps for 6 cycles (total run time 1h)

4°C hold until ready to proceed

3. Treat ligation reaction with PlasmidSafe exonuclease to prevent unwanted recombination products:

11 ul ligation reaction from step 4
1.5 ul 10X PlasmidSafe Buffer
1.5 ul 10mM ATP
1 ul PlasmidSafe exonuclease
15 ul total
Incubate reaction at 37°C for 30 min

4. Transformation with 1-2 ul of the final product into competent cells

5. Pick colony and sequence verify the clones.

Transfecting the sequence verified plasmid into target cells

1. Seed the cells 24 hours before transfection in 6 well plates
2. Seed enough plates and Check if the confluency is above 50%
3. Choose the best wells for transfection
4. Lable 8 eppendorf tubes with 1A-4A and 1B-4B respectively
5. Add corresponding µl serum Free and Antiboiotic free Optimem to tubes labelled 1A-4A
6. Add corresponding µl Fugene directly to Optimem in tubes 1A-4A and mix breifly by inverting 2 times
7. Add corresponding µl Plasmid DNA to tubes 1B-4B
8. Incubate for 5 min at RT
9. Dropwise add Optimem:Fugene mix in Tube A to the Plasmid DNA in tube B and mix briefly by inverting 2 times
10. Incubate the mixture for 15 min
11. Dropwise add 100 µl of the mix to each well containing 2000 µl media mix gently
12. Incubate and check after 24 or 48 hours

Puromycin selection for creating single cell colonies

1. After transfecting the cells for 48 hours split the cells from one 6 well plate into 10 centimeter plates as follow
2. Trypsinize the cells and add 10 ml medium to each 6 well plate
3. Add respectively 4 ml , 3 ml, 2 ml, and 1 ml of the cell suspension to four 10 cm plates
4. Cultivate the cells without Puromycin for 48 hours
5. Add 0.5 µg/ml Puromycin to the HCT116 cells
6. Change the media every 3 to 4 days in order to remove dead cells
7. Look for single colonies to appear after about 2 weeks
8. Pick up the single colonies and check for STAG2 expression by western blot and Immunofluorescence
9. Expand and freeze down the knockout clones for further experiments

5 Results

5.1 Somatic single nucleotides variations (SNVs)

In order to assess the SNVs present in 63 genes that are reported to be frequently mutated in MDS and AML, we used a custom target enrichment approach. Oligos were designed against the exome and parts of surrounding introns of these 63 genes using Agilent SureDesign custom design software. Target regions were enriched through library preparation steps as detailed in the material method section. Libraries were pooled together and sequenced on Ilumina Highseq 2000 using a 100 bp paired end sequencing assay. A total of 90 paired raw reads were produced which were analyzed using established pipelines at the DKFZ core facility. After bioinformatics analysis, an excel spreadsheet was produced containing the chromosome coordinates, sample IDs, mutated genes, amino acid changes, Ensemble transcript changes, variant allele frequencies, whether or not the mutation has been described in the Catalog of Somatic Mutations in Cancer (COSMIC) and dbSNPs. To find the somatic single nucleotide variations (SNVs), all the mutations with dbSNPs were excluded. The data analysis was done by Ivo Buchhalter from the department of computational oncology. The SNVs with variant allele frequency of less than 20 percent were excluded. By applying the above mentioned filters, a final list of samples with SNVs was produced. An excel sheet was produced containing the patients IDs on the top horizontal side, gene IDs at the right side and karyotypes on the bottom side. 21 out of 90 patients showed no mutation or mutation with allele frequency below 20%, which were considered not mutated. These patients are brought at the horizontal top left side of the table and the patients IDs are written in red. Surprisingly, all these non-mutated patients had aberrant karyotypes as shown on the right side of the table. 34 genes out of 63 including SMAD7, TUBG1, ZRSR2, STAT3, SMC1A, SMC3, SPERT, SPI1, NPM1, NRAS, RAD21, LATS1,MAD2L1, MAP2K3, NFKB1, HAUS7, HOXD13, IDH1, IDH2, IRAK1, FANCD2, FOXP1, GNAS, ETV6,CDKN2B, CEBPA,CTNNB1, CDC20, C19orf80, CALR, AURKA, AURKB, B3GALT6, BCL2, and AGBL1 showed no mutation. 4 genes including TET2, SRSF2, SETBP1, and HAUS3 were mutated only in patients with normal karyotype. BCOR, BCORL1, CBL, NOTCH1, SF3B1, and TGFB1 were mutated in both patients with normal and abnormal karyotypes. DNMT3A, NIPBL, BUB1B, RUNX1, EZH2, SUZ12, FANCA, JAK2, ASXL1, CDKN1A, AKAP9, STAG2, and TP53 were exclusively mutated in patients with abnormal karyotype. TP53 was most frequently mutated in the patients with abnormal karyotypes and after that STAG2, AKAP9 and CDKN1A (**Table 11)**. Interestingly, all these four genes are related to chromosomal instability. The mutations in TP53, STAG2 and AKAP9 were verified by Sanger sequencing.

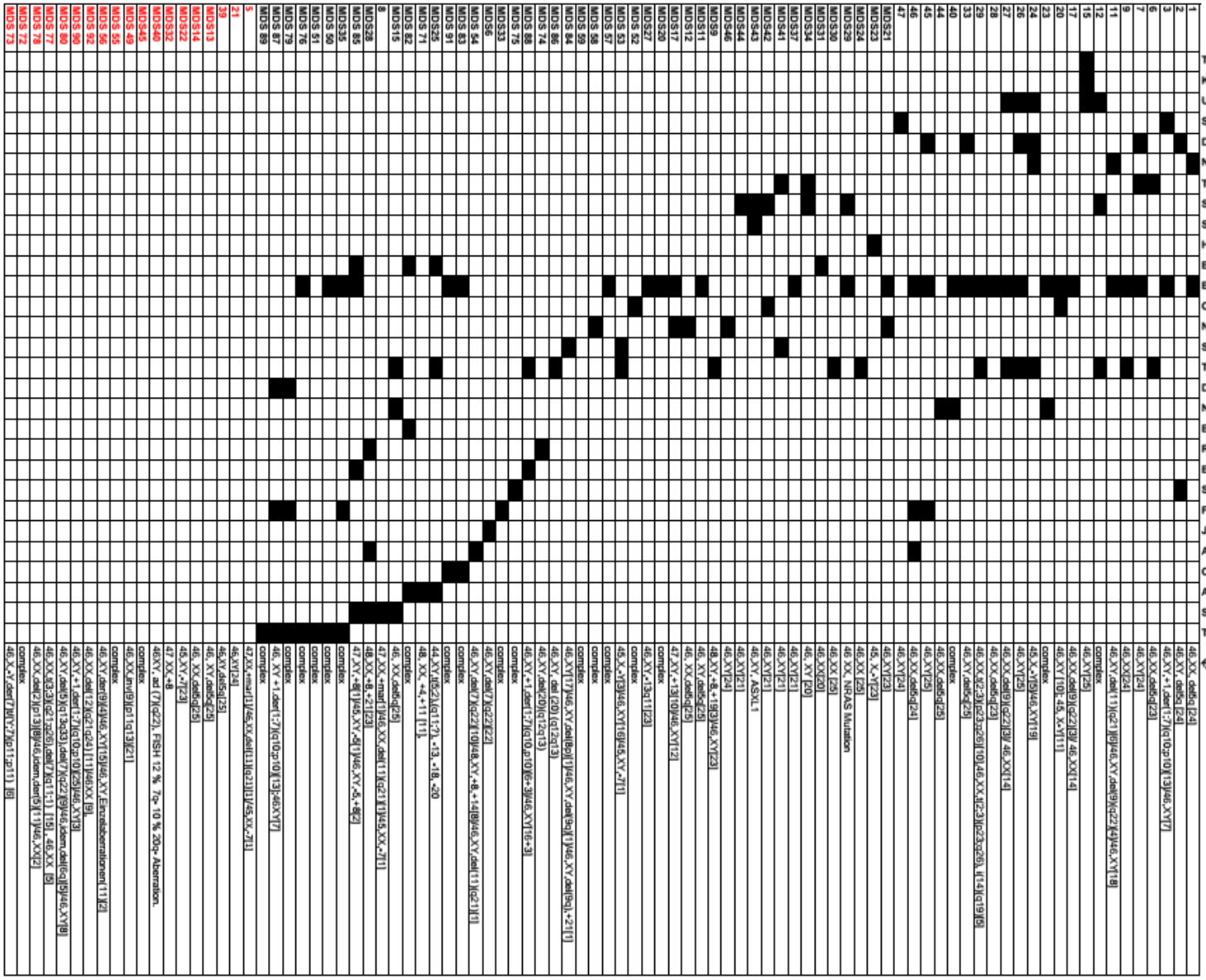


Table 11: Mutational analysis of 90 MDS patient bone marrow and blood samples

The mutational status of 29 genes in 90 MDS samples together with their karyotypes is shown. Filled boxes show SNVs. The patient IDs written in red, harbor no mutation.

5.2 Mutation verification

SNVs were found in the gene sequence in the Ensemble database. The corresponding sequence containing the SNV was copied to Ape plasmid software. After marking the SNV, PCR primers were designed manually at 300 bp upstream and downstream of the SNV. The target region was PCR amplified and ran on an agarose gel. If a single band was observed, the PCR product was purified using the Minelute PCR purification kit following the protocol. The purified PCR product was sent to GATC for sequencing. The result was visualized in Ape plasmid software.

Three SNVs in TP53, one SNV in STAG2, and one SNV in AKAP9 were verified by PCR followed by Sanger sequencing.

TP53

Three SNVs in TP53 were verified by Sanger sequencing. In MDS patients 76 and 79, cysteine 238 (C) was exchanged with tyrosine (Y) and in MDS patient 51 valine 216 (V) was replaced by methionine (M) (**Figure 23)**. HCT116 WT serves as a negative control.

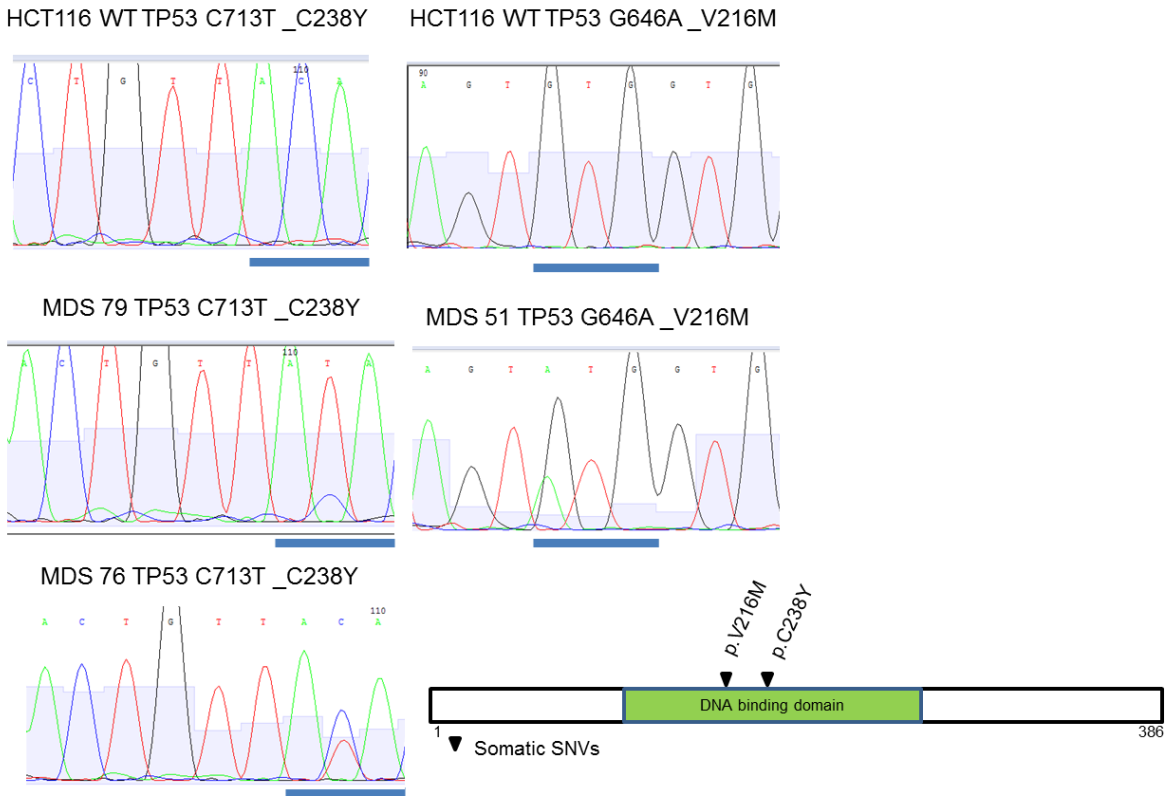


Figure 23: The SNVs in TP53 verified by Sanger sequencing

SNVs can be observed in MDS samples in comparison with the HCT116 control cell line. The sample, gene and the corresponding base and amino acid change is brought at the top of each chromatogram. cysteine (C), tyrosine (Y), methionine (M), valine (V).

STAG2

In total, four out of 90 samples had a SNV in STAG2. Arginine 146 (R) and serine 1075 (S) were replaced by a stop codon and tyrosine 433 (Y) was exchanged with histidine (H). Due to lack of material only one of the three SNVs was verified by Sanger sequencing (**Figure 24)**. HCT116 WT serves as a negative control.

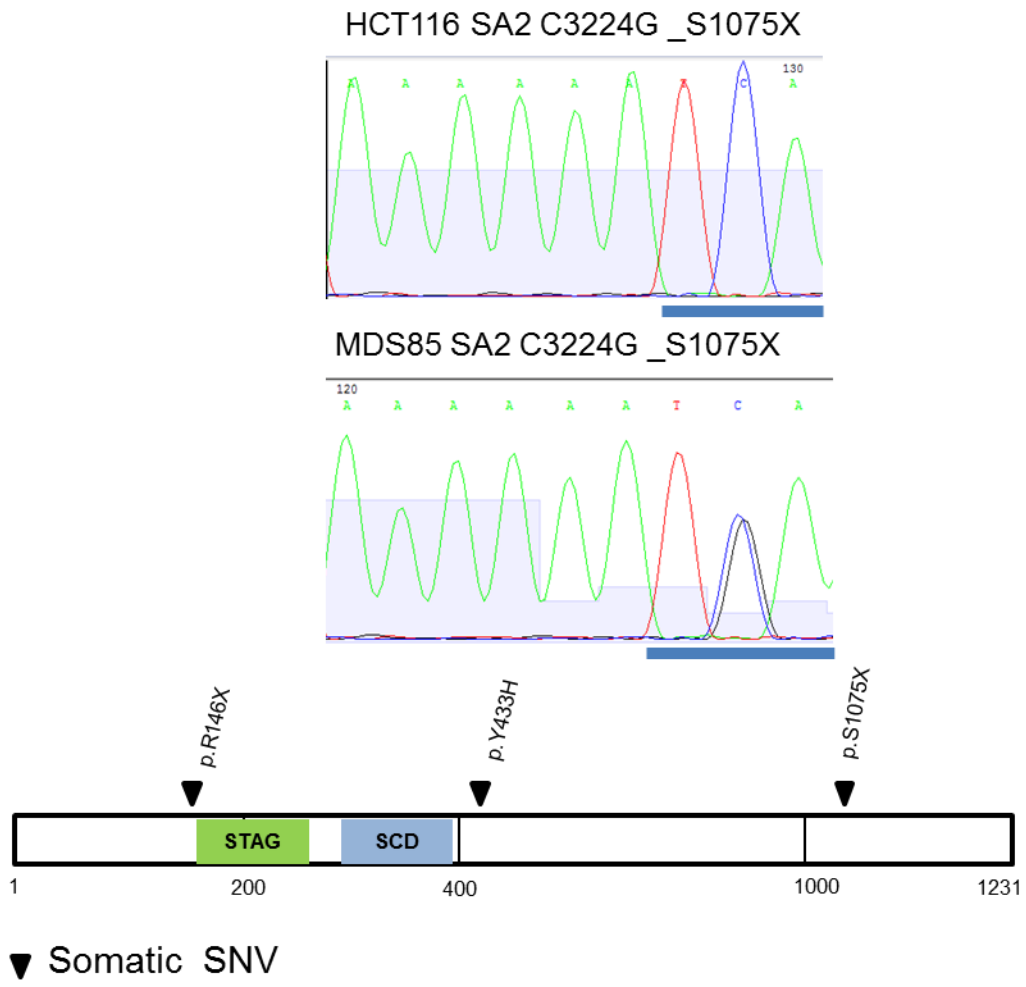


Figure 24: Somatic SNVs found in STAG2

Arginine 146 (R) and serine 1075 (S) were replaced by a stop codon and tyrosine 433 (Y) was exchanged with histidine (H). Only serine replacement was verified by Sanger sequencing due to material availability.

AKAP9

In total, 3 out of 90 samples were mutated in AKAP9. Due to lack of material only one of the SNVs was verified by Sanger sequencing **(Figure 25)**. HCT116 WT serves as a negative control.

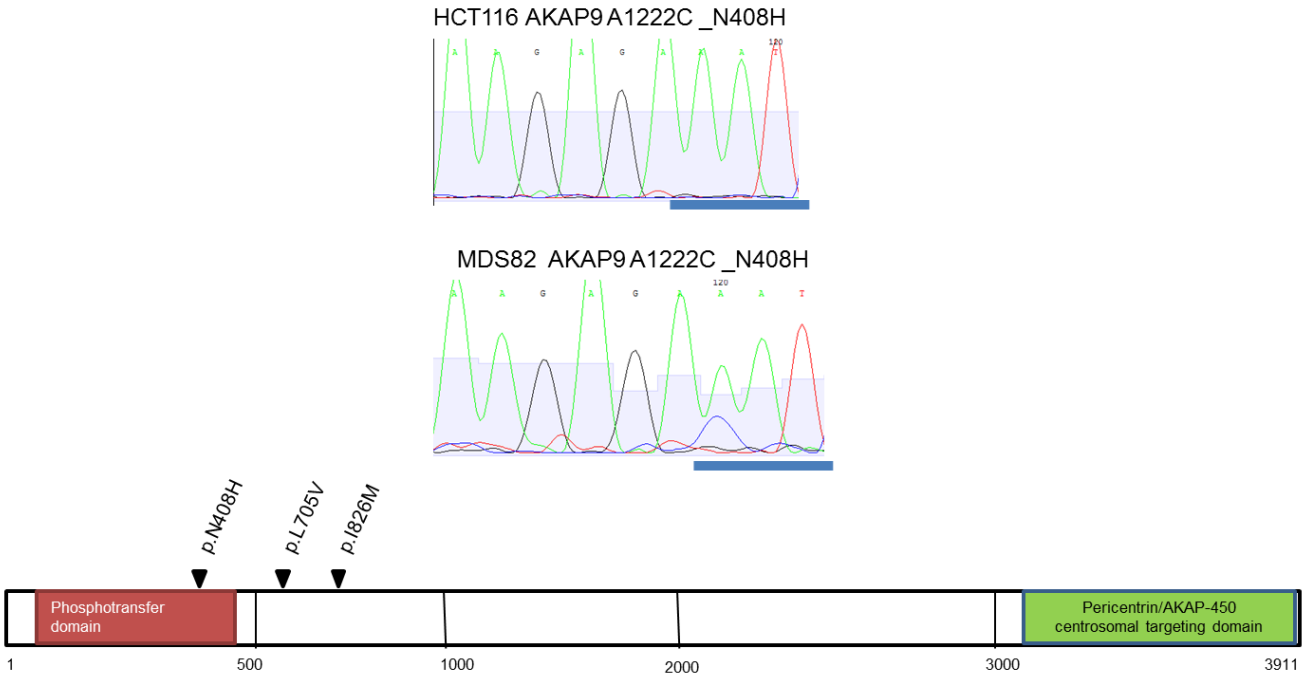


Figure 25: Somatic SNVs found in AKAP9

Only one SNV was verified by Sanger sequencing due to material availability. Asparagine 408 (N) was replaced by histidine (H), leucine 705 (L) with valine (V), and isoleucine 826 (I) with methionine (M).

5.3 STAG2 expression in AML

STAG2 has been reported to be affected by nonsense mutations leading to loss of protein expression in AML as well. Therefore, we used immunofluorescence staining and Western blotting in order to assay STAG2 expression in AML. For immuno-fluorescence staining the blood or bone marrow mononuclear cells were brought on coverslips using cytospinning and were fixed with methanol and kept at -20°C till staining. Cell pellets were frozen at -80°C and were used for preparing protein lysates following the protocol given in the materials and methods section. Expression of STAG2 was lost in 18 out of 74 AML cases by immunofluorescence staining. Representative samples are shown in **Figure 26A**. STAG2 expression was measured by Western blotting as shown in **Figure 26B**. In both assays patient 070 showed normal STAG2 expression and STAG2 expression was lost in patient 107.

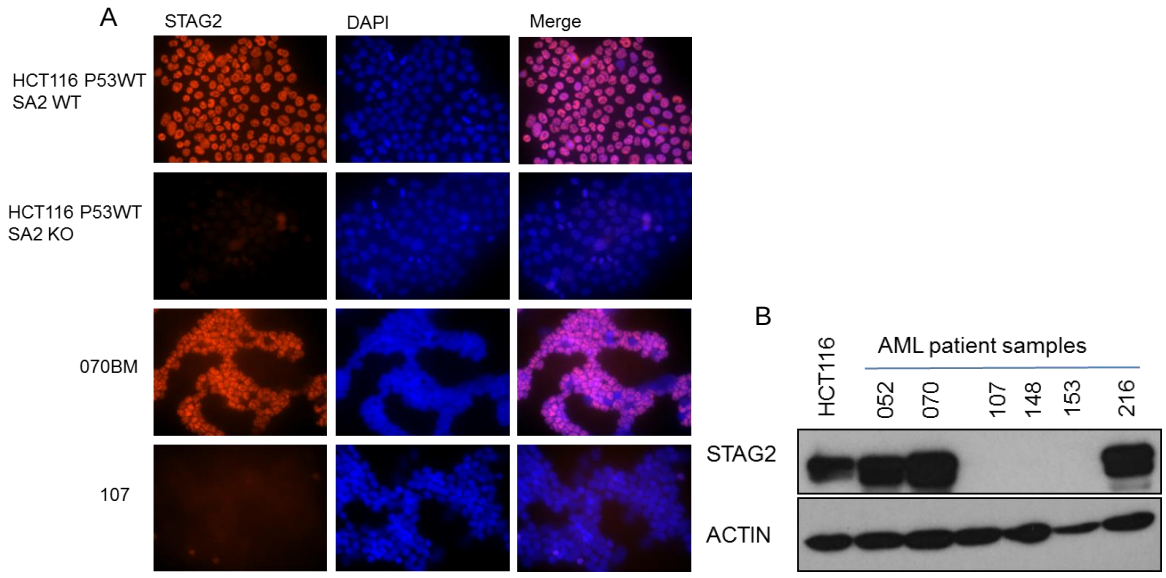


Figure 26: STAG2 expression in AML

(A) Representative images showing AML cases with normal (070) and lost (107) STAG2 expression. HCT116 STAG2 wildtype and knockout cells serve as positive and negative controls, respectively. (B) Expression analysis of STAG2 by Western blotting in wildtype HCT116 as a positive control and six AML cases.

5.4 STAG2 mutation in AML

In order to determine the reason for loss of STAG2 expression in AML, we used Sanger sequencing of the whole cDNA to identify mutations that could be responsible for STAG2 loss of expression. PCR primers were designed in a way that produced four amplicons covering the whole STAG2 cDNA. These four amplicons were PCR amplified separately and sequenced by Sanger sequencing. The results were blasted against STAG2 cDNA in order to find mutations. In total 18 out of 74 samples showed loss of STAG2 expression. Cell pellets were available for 10 out of these 18 samples. RNA was extracted using Qiagen RNeasy Mini Kit from these 10 samples and cDNA was generated using Roche Transcriptor High Fidelity cDNA Synthesis Kit. PCR was done using CloneAmp™ HiFi PCR Premix following the protocol but the annealing temperature was changed to 58°C. The PCR product was purified using QIAGEN MinElute PCR Purification Kit and sent for sequencing to the GATC Company. The results were visualized in ApE plasmid software. Two out of 10 samples showed a mutation. Patient 148 harbored cysteine 769 to stop codon mutation and patient 277 harbored lysine 128 to threonine mutation **(Figure 27)**.

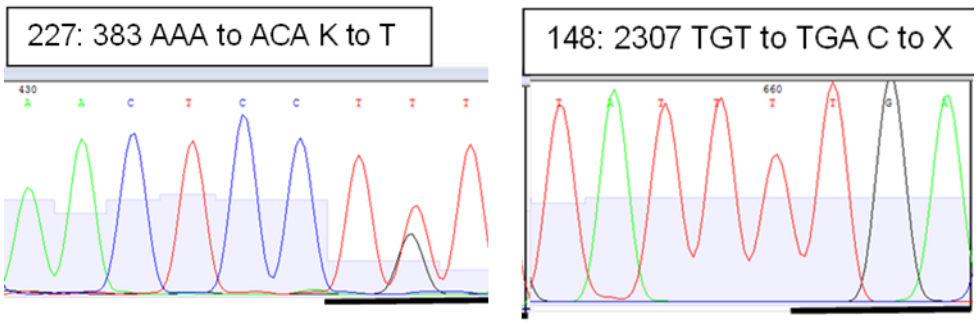


Figure 27: STAG2 mutations in two AML samples

Mutations in 277 and 148 patient samples are shown. K (lysine), T (threonine), C (cysteine), X (stop codon).

5.5 STAG2 promoter methylation in AML

We speculated that in patients that showed no mutation STAG2 loss of expression could be due to promoter methylation. Therefore, we planned to analyze the above mentioned 10 samples for promoter methylation. Only 7 out of 10 were analyzed due to material availability. The promoter region of STAG2 was identified by searching for STAG2 genomic sequence in the UCSC genome browser. 1000 base pairs upstream of the first coding exon were chosen as promoter region. This sequence was pasted in the online Meth Primer software (http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi) for designing primers. PCR was done using the Qiagen PyroMark PCR Kit following the protocol. PCR product was purified using QIAGEN MinElute PCR Purification Kit and sent for sequencing to GATC. The STAG2 promoter was found methylated in 4 out of the 7 AML samples analysed **(Figure 28).**

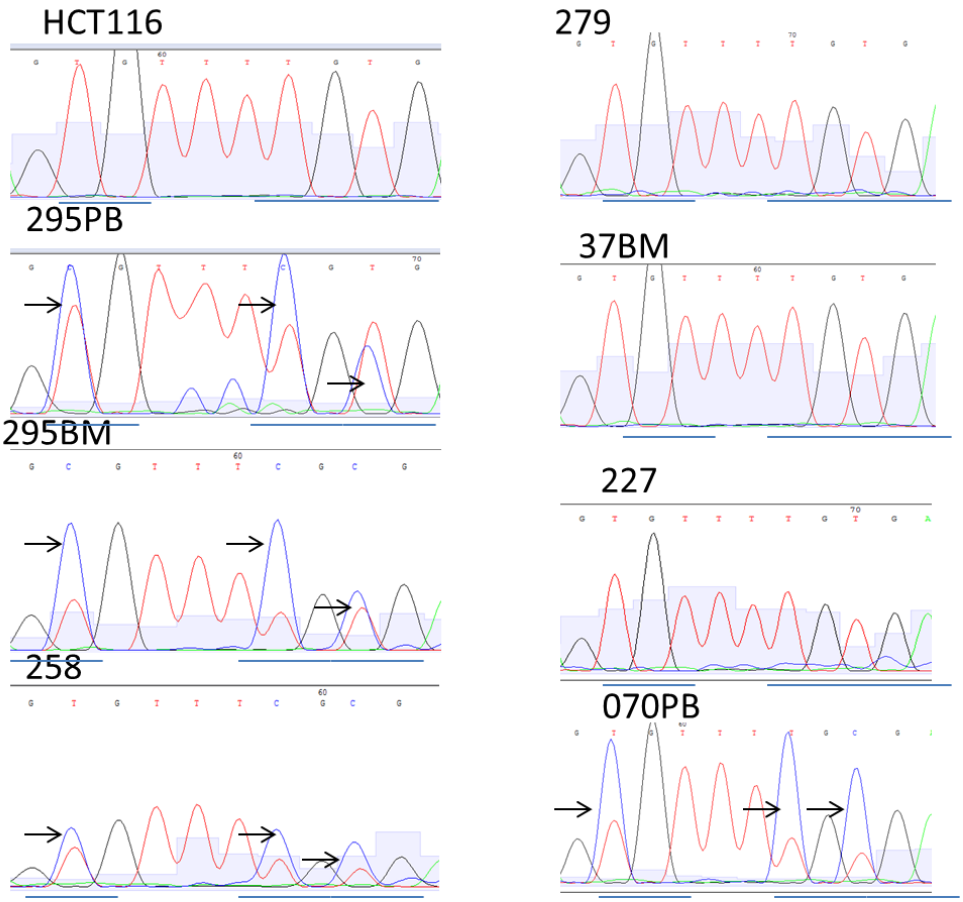


Figure 28: Promoter methylation in 7 AML patients

The same part of the STAG2 promoter with three CpG islands is shown. HCT116 serves as a control since STAG2 is highly expressed in that. The patients with blue C peak (arrow) show methylated promoter (samples 295PB, 295 BM, 258, 070PB). The patients without C peak in the CpG island have unmethylated promoter in which C is converted to T after bisulfite treatment (samples 227, 37BM, 279).

5.6 CRISPR/Cas9-based knockout

In order to understand the consequences of STAG2 loss of expression, we used CRISPR/Cas9 to knock out STAG2 in HCT116 cells. The detailed procedure is described in the materials and methods section. For CRISPR/Cas9-based knockout of STAG2, we have designed two sgRNAs using an online design tool from the Zhang laboratory (http://crispr.mit.edu/). The oligos were synthesized, annealed and cloned into the Psp-Cas9-Puro vector (Addgene PX459) following the protocol guidelines (http://www.genome-engineering.org/ crispr/?page_id=23). Cloned oligos were verified by restriction digestion and Sanger sequencing. The HCT116-p53$^{+/+}$ and HCT116-p53$^{-/-}$ cell lines were authenticated and characterized for expression of RAD21, STAG2 and p53 **(**

Figure 29**)**.

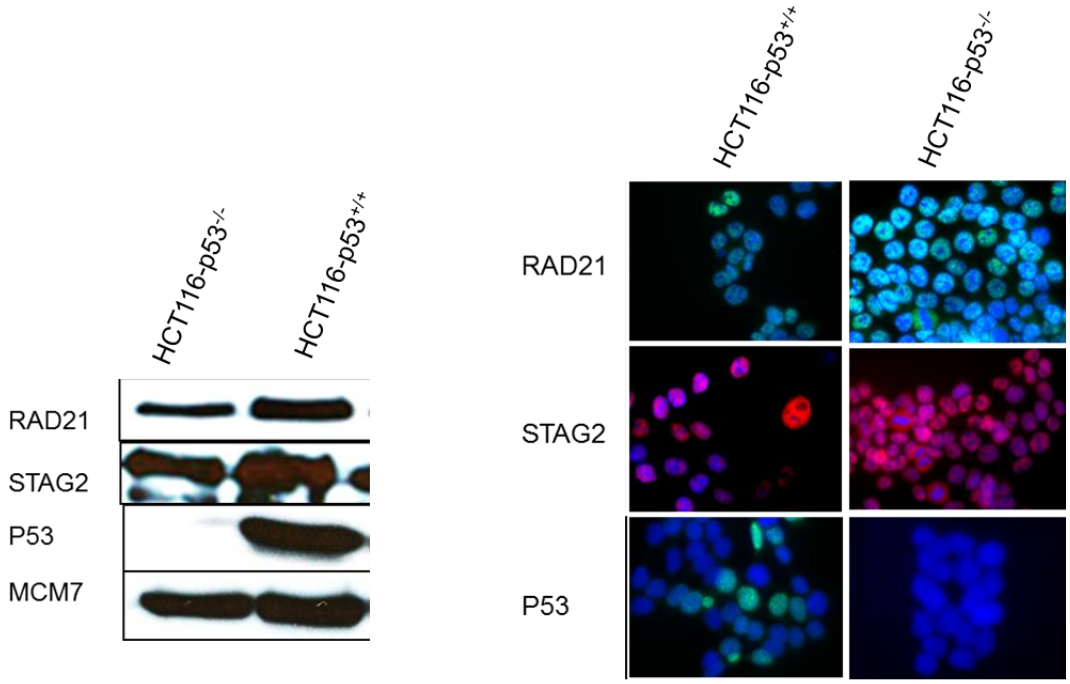


Figure 29: Characterization of hTERT-RPE, HCT116-p53$^{+/+}$ and HCT116-p53$^{-/-}$ cell lines

Characterization of HCT116-p53$^{+/+}$ (p53 wild type), and HCT116-p53$^{-/-}$ (p53 null) cell lines with Western Blotting (left panel) and immunofluorescence staining (right panel) using indicated antibodies.

The HCT116-p53$^{+/+}$ and HCT116-p53$^{-/-}$ cell lines were transfected with the pSpCas9(BB)-2A-Puro (PX459) containing STAG2 sgRNAs and selected for single clones using puromycin for about one month. Single clones were picked and assayed for the expression of STAG2 using Western blotting and immunofluorescence staining. Four clonal cell lines were created using CRISPR/Cas9 knockout of STAG2 **(Figure 30)**. K48 and K72 clones were created from the HCT116-p53$^{+/+}$ cell line and K17 and K51 from the HCT116-p53$^{-/-}$ cell line. The expression of STAG2 was completely lost in clones K48, K72, and K51. In K17, however, only one band of STAG2 expression was lost as shown by Western blotting. The Western blot and immunofluorescence staining for verifying the created clones was done by Annik Rossberg.

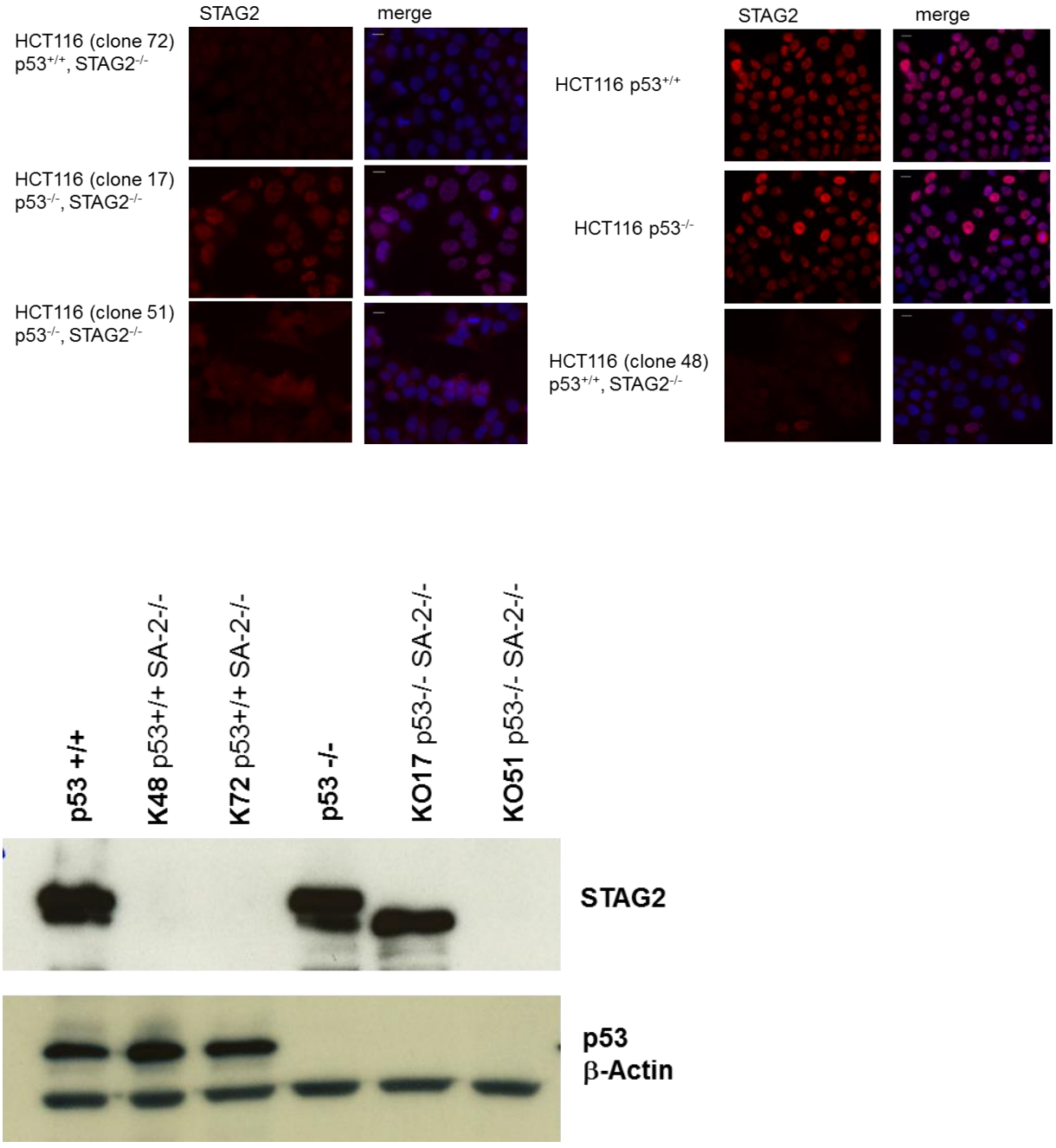


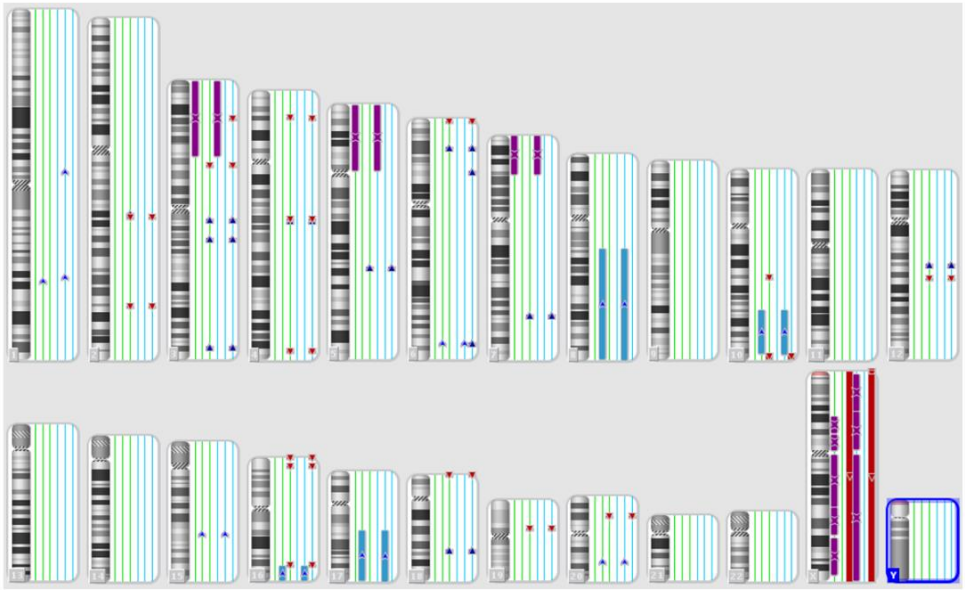
Figure 30: CRISPR/Cas9 knockout of STAG2

Created knockout cell lines characterized by Western blotting and immunofluorescence staining. HCT116-p53$^{+/+}$ and HCT116-p53$^{-/-}$ cell lines were transfected with pSpCas9(BB)-2A-Puro (PX459) containing STAG2 sgRNAs. After 2 days of transfection, the cells were passaged in media without Puromycin for 48 hours. Then, puromycin was added to media and the media were changed every 4 to 5 days. Puromycin selection was continued for about 4 weeks until single colonies appeared. The single colonies were picked and expanded and used for immunofluorescence staining and Western blotting to analyse for STAG2 expression.

5.7 Array-CGH of HCT116 *STAG2* wildtype versus knockout clones

In order to compare the created clones in terms of copy number variation and ploidy, array CGH assays were performed. HCT116-p53$^{+/+}$STAG2$^{+/+}$ was compared with HCT116-p53$^{+/+}$STAG2$^{-/-}$ and HCT116-p53$^{-/-}$STAG2$^{+/+}$ was compared to HCT116-p53$^{-/-}$ STAG2$^{-/-}$. No major differences were detected between the karyotypes of the compared clones (**Figure 31)**. The array-CGH was done by Mutlu Kartal-Kaess and Anna Jauch.

HCT116-p53$^{+/+}$STAG2$^{+/+}$ vs HCT116-p53$^{+/+}$STAG2$^{-/-}$



HCT116-p53$^{-/-}$STAG2$^{+/+}$ vs HCT116-p53$^{-/-}$STAG2$^{-/-}$

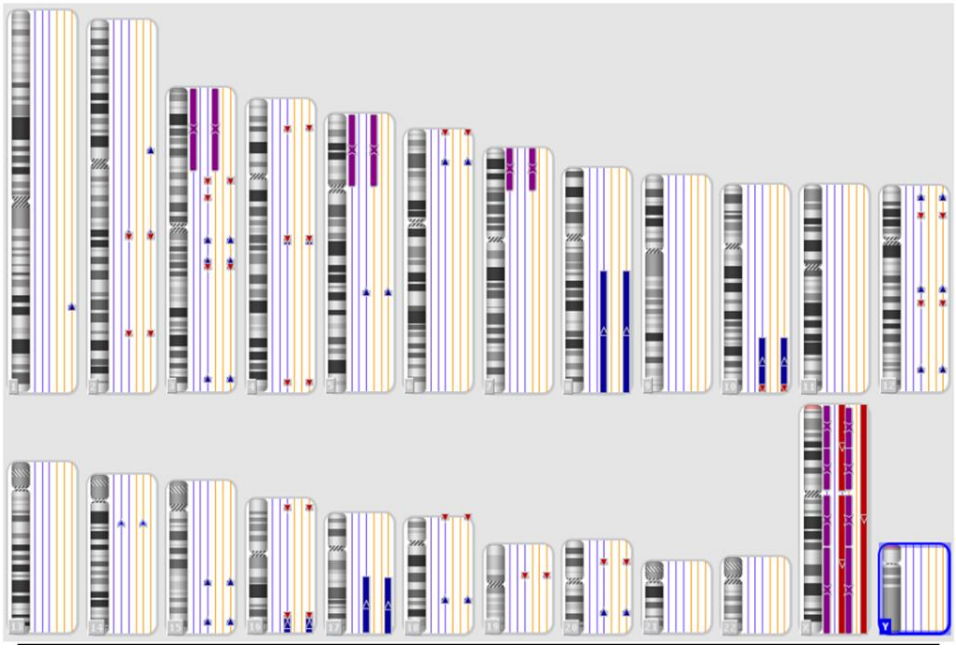


Figure 31: Array-CGH of HCT116 STAG2 wildtype versus STAG2 knockout clones

Array-CGH was used to compare the karyotypes of HCT116 STAG2 wildtype with STAG2 knockout clones in TP53 wildtype and knockout backgrounds, respectively.

5.8 Telomeric associations in STAG2 knockout clones

We also used multicolor fluorescence in situ hybridization (M-FISH) in order to analyse for karyotypic aberrations of the created cell lines. There were no differences between HCT116-p53$^{+/+}$STAG2$^{+/+}$ and HCT116-p53$^{+/+}$STAG2$^{-/-}$ (**Figure 32**). However, telomeric associations (tas) were observed between acrocentric chromosomes 22 and chromosomes 13 or 15 ((tas(13;22) and tas(15;22)) in HCT116-p53$^{-/-}$STAG2$^{-/-}$ but not HCT116-p53$^{-/-}$STAG2$^{+/+}$ (**Figure 33**). Tas is considered an early sign for chromosomal instability. The M-FISH was done by Mutlu Kartal-Kaess and Anna Jauch.

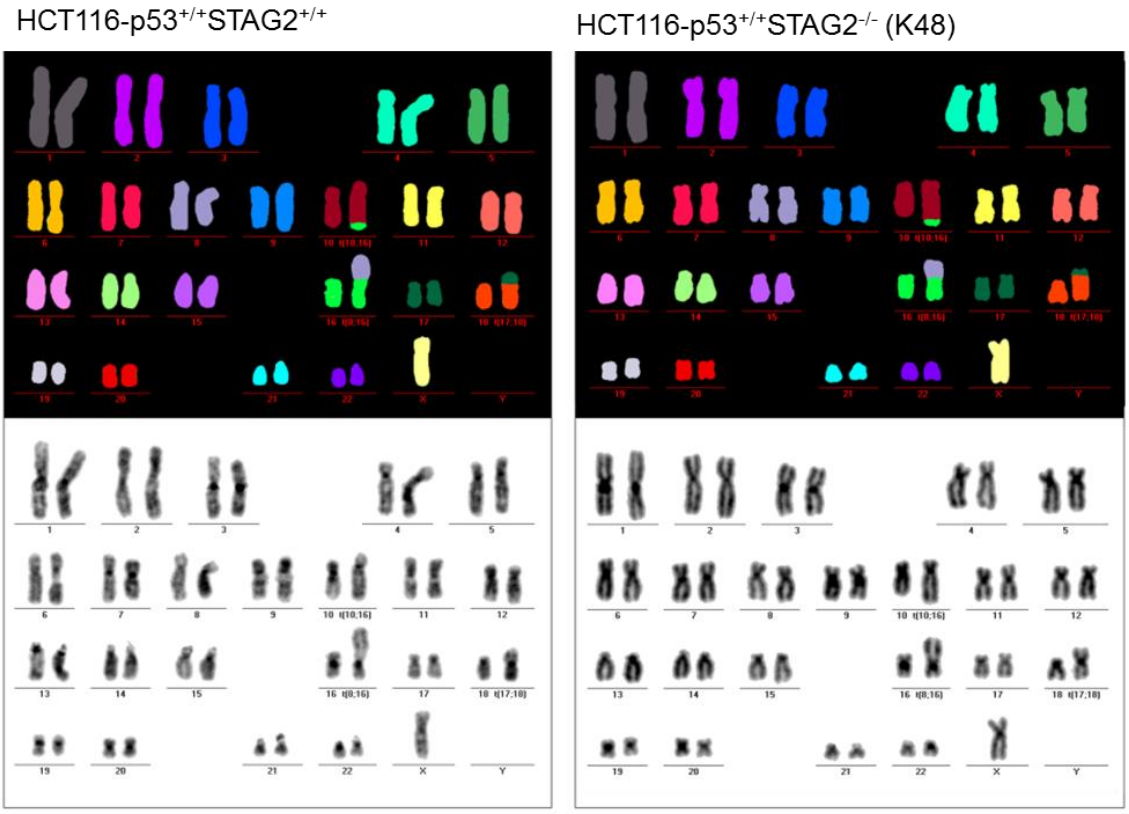


Figure 32: HCT116 STAG2 wildtype and knockout clones in a TP53 wildtype background

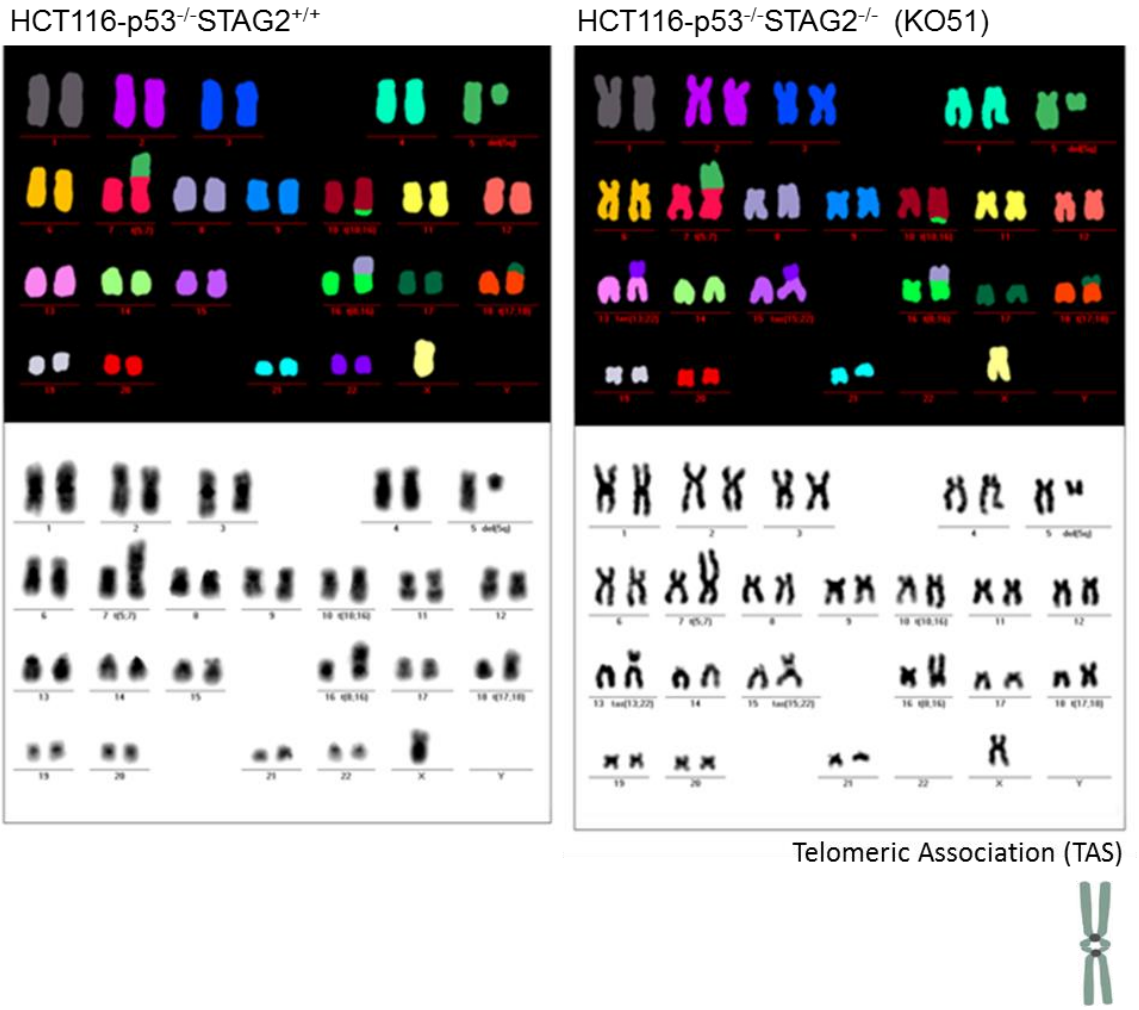


Figure 33: Telomeric association in HCT116-p53$^{-/-}$STAG2$^{-/-}$ cells

Tas(13;22) and tas(15;22) was observed in HCT116-p53$^{-/-}$STAG2$^{-/-}$ (KO51) but not in HCT116-p53$^{-/-}$ STAG2$^{+/+}$ (KO) cells

5.9 Gene expression profiling of HCT116 STAG2 wildtype versus STAG2 knockout clones

We used gene expression profiling in order to assess whether there is a difference between HCT116 STAG2 wildtype versus knockout cells in terms of gene expression. The results showed only mild differences between STAG2 knockout and STAG2 wildtype clones (**Figure 34)**. The top 20 different genes in terms of gene expression fold change did not belong to any known signaling pathway (**Table 12**).

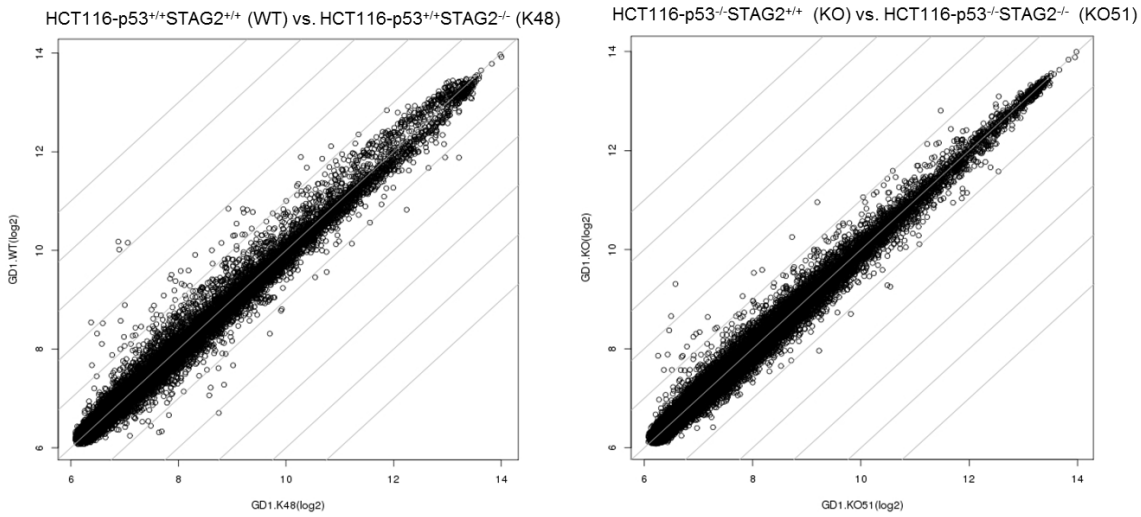


Figure 34: Gene expression profiling of HCT116 STAG2 wildtype and STAG2 knockout clones

The following clones were compared: HCT116-p53$^{+/+}$STAG2$^{+/+}$ (WT) with HCT116-p53$^{+/+}$STAG2$^{-/-}$ (K48) and HCT116-p53$^{-/-}$STAG2$^{-/-}$ (KO51) with HCT116-p53$^{-/-}$STAG2$^{+/+}$ (KO). Mild differences can be observed between the compared clones.

Symbol	K48 vs. WT fold_change
SH3KBP1	0,1
SH3KBP1	0,12
SCG2	0,12
STAG2	0,22
IL8	4,13
M160	0,26
RPS29	0,26
HAS3	0,27
ANXA10	0,28
SEC31A	0,32
RPL14	0,32
MT1X	0,32
MT1E	0,33
ITGB4	0,33
KIAA1199	2,97
HS3ST1	0,34
CD163L1	0,35
DENND2C	2,85
MTE	0,37
LOC392437	2,67

Symbol	KO51 vs. KO fold_change
TNFRSF6B	0.15
TNFRSF6B	0.27
GDF15	0.29
PLAU	0.35
LOC652097	0.35
STAG2	0.36
SLC2A3	0.38
SOX4	0.39
PRKACB	0.4
SAT1	0.4
CD24	0.4
ANXA10	2,44
TMEM200A	2,41
LOC644322	0.42
PRPH	0.42
IRF7	0.43
LOC653924	0.45
LOC653458	0.45
BDNF	2,16
AKAP12	0.47

Table 12: The top 20 genes in terms of gene expression fold change

HCT116-p53$^{+/+}$STAG2$^{+/+}$ (WT) versus HCT116-p53$^{+/+}$STAG2$^{-/-}$ (K48) and HCT116-p53$^{-/-}$STAG2$^{-/-}$ (KO51) with HCT116-p53$^{-/-}$STAG2$^{+/+}$ (KO) are respectively compared.

5.10 Cell proliferation of HCT116 STAG2 wildtype versus knockout clones

In order to compare cell proliferation rates, 500,000 cells were seeded in triplicates in 6-well plates and were counted every 24 hours for 7 days in culture. HCT116 STAG2

knockout clones proliferated slower than their corresponding wildtype clones both in a TP53 wildtype and knockout background **(Figure 35).**

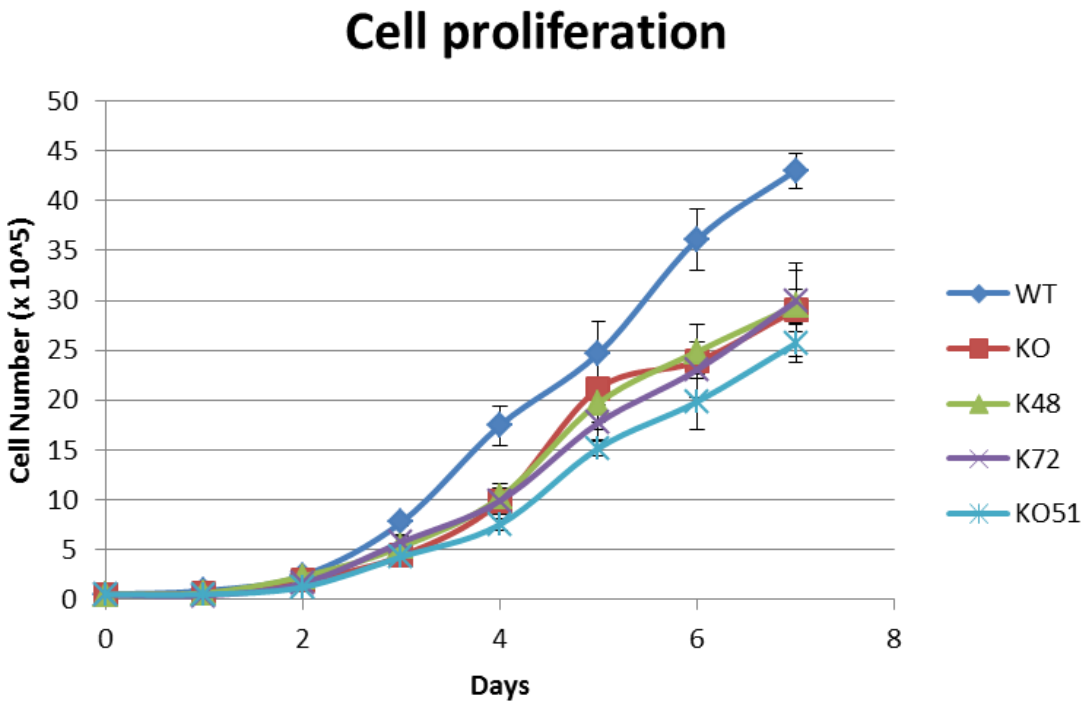


Figure 35: comparison of proliferation rate in STAG2 knockout vs. wildtype

HCT116-p53$^{+/+}$STAG2$^{+/+}$ (WT) proliferate slightly faster than HCT116-p53$^{+/+}$STAG2$^{-/-}$ (K48) and HCT116-p53$^{+/+}$STAG2$^{-/-}$ (K72). HCT116-p53$^{-/-}$ STAG2$^{-/-}$ (KO51) proliferate slightly slower than HCT116-p53$^{-/-}$STAG2$^{+/+}$ (KO).

6 Discussion

6.1 STAG2 is the only cohesin complex component found to be mutated in MDS

Current improvements in next generation sequencing have detected mutations in genes involved in a broad range of biological phenomena in different cancer entities. These findings include mutations in the genes encoding for the members of the cohesin complex (STAG2, RAD21, SMC1A, SMC3) which have been described for the first time in 2013 to be mutated in myeloid neoplasms suggesting a role for the cohesin complex in leukemogenesis *(15)*. Recurrent mutations and deletions involving multiple components of the cohesin complex, including STAG2 (5.8%), RAD21 (0.9%), SMC1A (0.0%) and SMC3 (1.3%), were reported in MDS. These mutations and deletions were mostly mutually exclusive and overall occurred in 12.1% of AML and 8.0% of MDS cases. Cohesin is composed of the four core subunits STAG2, RAD21, SMC1A, and SMC3 which form a ring-like structure. The cohesin complex has many roles including cohesion of sister chromatids, DNA repair, control of gene expression, chromosome segregation, DNA damage repair, DNA replication, and the control of heterochromatin and centromere formation *(50)*. In the current work we investigated the mutational spectrum of cohesin complex components in MDS and if these mutations are associated with karyotype aberrations. Using a targeted resequencing approach we studied the occurrence of somatic SNVs in 90 MDS patients. We found SNVs only in STAG2 in 4 out of 90 (4.4%) patients. The bone marrow mononuclear cells of all four patients harbored karyotype abnormalities. The SNVs were confirmed by Sanger sequencing. In contrast to previous studies, we found no mutations in other members of the cohesin complex. We also found somatic SNVs in TP53 in 7 out of 90 (7.8%) MDS patients. The SNVs in TP53 were mainly associated with complex karyotype aberrations, which is in accordance with previous studies. AKAP9, a centrosomal protein potentially involved in mitotic spindle formation and chromosome segregation, was also mutated in 3 out of 90 (3.3%) patients with aberrant karyotypes. Whether cohesin mutations are driver mutations remains unclear. Mossner and colleagues demonstrated on a patient-individual basis that mutations affecting epigenetic modifiers (eg, TET2, ASXL1) and RNA splicing factors (eg, SF3B1, SRSF2) are predominantly “founder” events in MDS and genes involved in signaling cascades (eg, JAK2 and CBL), transcription factors (eg, RUNX1 and ETV6), and cytogenetic lesions (eg, monosomy 7, trisomy 8, and del(5q)) are almost exclusively acquired as late events in MDS, emphasizing their potential use as indicators of disease progression *(51)*. Their data therefore suggest that STAG2 mutations are late events in MDS pathogenesis, although, to our knowledge no studies have specifically examined the timing of the occurrence of cohesin mutations in a hierarchical manner in any cancer entity. Clinical outcome associated with cohesin mutations in MDS has

not been studied so far in detail. Data by Montalban-Bravo et al. suggest that STAG2 mutations are an independent prognostic factor in MDS*(52)*. According to their findings, the presence of STAG2 mutations is associated with decreased overall survival in MDS, particularly in those cases classified as lower-risk MDS by IPSS. In AML however, according to Thol and colleagues overall survival, relapse-free survival, and complete remission rates are not influenced by the presence of cohesin mutations *(16)*. They found that the majority of patients with cohesin gene mutations had intermediate risk cytogenetics. Thol and colleagues found a strong correlation between cohesin mutations and NPM1 mutations in AML. STAG2 is reported to be mutated in 15.6% (12/77) of bladder cancer patients and its loss of function is associated with better prognosis *(27)*. One of the pitfalls of our study is that we have not performed a correlative analysis between clinical outcomes and STAG2 mutational status.

6.2 STAG2 expression is lost in AML

As mentioned above cohesin complex components are mutated in myeloid malignancies including AML. Here we investigated the protein expression of STAG2 in 74 AML samples. STAG2 expression was lost in 24.3% of AML samples. STAG2 expression is lost in other types of cancer as well. Solomon and colleagues observed complete loss of STAG2 expression in 3 out of 21 glioblastomas, 5 out of 9 Ewing sarcomas, and 1 out of 10 melanoma cell lines *(31)*. We then investigated whether the AML samples that lost STAG2 expression harbor any mutation. For that, we sequenced the entire cDNA of the corresponding samples using Sanger sequencing. Due to sample availability we could only sequence 10 samples. Two out of 10 samples (20%) harbored mutations. This shows that STAG2 loss is only in part due to STAG2 mutations. Kon and colleagues observed severely reduced expression of one or more cohesin components in KG-1 (STAG2) and MOLM-13 (STAG1, STAG2, RAD21 and NIPBL) cells without any accompanying mutations in the respective genes *(15)*. They found no significant differences in protein expression of the cohesin components in cohesin-mutated and non-mutated cell lines in whole-cell extracts. However, they found that the expression of one or more cohesin components, including SMC1, SMC3, RAD21 and STAG2, was significantly reduced in the chromatin-bound fractions of cell lines with mutated or reduced expression of cohesin components compared with the cell lines with no known cohesin mutations or abnormal cohesin expression, suggesting a substantial loss of cohesin-bound sites on chromatin. We speculated that STAG2 loss might be due to promoter methylation. In 7 out of 12 (58.3%) samples the STAG2 promoter was methylated which shows that loss of STAG2 expression is in part due to promoter methylation.

6.3 STAG2 knockout is associated with TAS in a TP53 null background

Only few studies have investigated the role of cohesin mutations with regard to the induction of chromosomal instability in cancer. Some studies suggest that cohesin mutations are associated with chromosomal instability in some cancer types. Solomon et al. observed that targeted inactivation of STAG2 led to sister chromatid cohesion defects and aneuploidy, whereas in two aneuploid human glioblastoma cell lines, targeted correction of the endogenous mutant alleles of STAG2 led to enhanced chromosomal stability *(31)*. To the contrary, others have shown that cohesin mutations are not associated with chromosomal instability. Balbas-Martinez and coworkers observed a loss of STAG2 expression in chromosomally stable tumors; STAG2 knockdown in bladder cancer cells did not increase aneuploidy *(27)*. Therefore, whether cohesin mutations are associated with chromosomal instability or not remains controversial. To our knowledge, whether cohesin mutations are associated with chromosomal instability in MDS and AML has not been investigated so far. We are the first to study the role of STAG2 with regard to chromosomal instability in myeloid malignancies. To do this, we used a Crispr/Cas9 genome editing approach to knockout STAG2 in HCT116-p53$^{+/+}$ and HCT116-p53$^{-/-}$ cells. Although array-CGH showed no differences between STAG2 wildtype and STAG2 knockout karyotypes, M-FISH analysis revealed telomeric associations between acrocentric chromosomes (tas(13;22) and tas(15;22)) to specifically occur in STAG2 knockout cells in a TP53 knockout background. Telomeric associations have been reported to be precursor leasion that subsequently develop into additional chromosomal aberrations *(53)*. Gelot and colleagues observed that at the genome level ablation of RAD21 or sororin produces large chromosomal rearrangements (translocation, duplication, deletion) *(54)*, which is in part in accordance with our observation of TAS although in our case TAS occurred only after STAG2 knockout in a TP53 null background. Tirode and colleagues observed in 299 Ewing sarcoma patients that the concurrent occurrence of STAG2 and TP53 mutations is associated with a poor prognosis *(55)*.

6.4 Gene expression profiling differences between STAG2 wildtype and knockout clones

STAG2 has been reported to be involved in the regulation of gene expression. We investigated whether the loss of STAG2 results in changes in gene expression. We observed mild differences between STAG2 wildtype and STAG2 knockout clones. The top 20 genes with different expression did not belong to any known signaling pathway. To explore whether STAG2 regulates transcription in human cancer cells, Solomon and colleagues used expression microarrays to measure global gene expression profiles of isogenic STAG2-proficient and STAG2-knockout cells. Expression profiles of STAG2-proficient and -deficient cells were remarkably similar

[i.e., only 16 of 28,869 genes (0.06%) were modulated >1.5-fold in STAG2-proficient 42MGBA cells], which indicated that STAG2 is not likely to be a major regulator of global gene expression *(31)*. However, Thota et al. observed a reduction in the expression of NRAS, JAK1, CBL, and HIF1A, among 28 differentially expressed genes, in MDS patients with reduced expression or mutation of STAG2 compared to patients with intact STAG2 expression *(56)*. Additionally, data by Mullenders et al. showed that in vivo knockdown of cohesin complex members alters hematopoiesis and leads to myeloproliferative neoplasms, suggesting their role in functionally controlling gene expression *(57)*.

6.5 Proliferation of STAG2 wildtype and knockout clones

We investigated the effects of STAG2 loss on proliferation of the HCT116 cell line. We observed no significant difference among the STAG2 null and STAG2 wildtype cell lines. However, the cell lines with STAG2 loss proliferated mildly slower. Kon et al. observed that forced expression of wild-type RAD21 and/or STAG2 induced significant growth suppression of the Kasumi-1 (with mutated RAD21) and MOLM-13 (with severe reduction of RAD21 and STAG2 expression) cell lines but not the K562 and TF1 (with wild-type RAD21) cell lines.

7 Conclusion and perspective

We found STAG2 to be mutated in 4 out of 90 (4.4%) of samples. No mutations were found in other members of the cohesin complex. STAG2 expression was lost in 17 out of 74 AML samples. This was in part due to mutation and in part due to promoter methylation. We observed TAS in STAG2 knockout HCT116 cells in a TP53 null background which might be an early sign of chromosomal instability. We speculate that STAG2 loss at centromeres could be replaced with STAG1 which is normally associated with telomeres. When STAG1 is competed away from telomeres to centromeres, this might result in telomere instability resulting in TAS.

References

1. A. Tefferi, J. W. Vardiman, Myelodysplastic syndromes. *N. Engl. J. Med.* **361**, 1872-1885 (2009).

2. R. Bejar *et al.*, Clinical effect of point mutations in myelodysplastic syndromes. *N. Engl. J. Med.* **364**, 2496-2506 (2011).

3. A. Murati *et al.*, Myeloid malignancies: mutations, models and management. *BMC Cancer.* **12**, 304-2407-12-304 (2012).

4. H. Harada *et al.*, High incidence of somatic mutations in the AML1/RUNX1 gene in myelodysplastic syndrome and low blast percentage myeloid leukemia with myelodysplasia. *Blood.* **103**, 2316-2324 (2004).

5. J. X. Cheng *et al.*, Genome-wide profiling reveals epigenetic inactivation of the PU.1 pathway by histone H3 lysine 27 trimethylation in cytogenetically normal myelodysplastic syndrome. *Leukemia.* **27**, 1291-1300 (2013).

6. J. Wang *et al.*, Loss of Asxl1 leads to myelodysplastic syndrome-like disease in mice. *Blood.* **123**, 541-553 (2014).

7. R. Itzykson, P. Fenaux, Epigenetics of myelodysplastic syndromes. *Leukemia.* **28**, 497-506 (2014).

8. G. W. Rhyasen, D. T. Starczynowski, Deregulation of microRNAs in myelodysplastic syndrome. *Leukemia.* **26**, 13-22 (2012).

9. T. D. Bhagat *et al.*, miR-21 mediates hematopoietic suppression in MDS by activating TGF-beta signaling. *Blood.* **121**, 2875-2881 (2013).

10. M. Cazzola, M. G. Della Porta, L. Malcovati, The genetic basis of myelodysplasia and its clinical relevance. *Blood.* **122**, 4021-4034 (2013).

11. Y. Zhao *et al.*, Downregulation of p21 in myelodysplastic syndrome is associated with p73 promoter hypermethylation and indicates poor prognosis. *Am. J. Clin. Pathol.* **140**, 819-827 (2013).

12. N. Droin, L. Guery, N. Benikhlef, E. Solary, Targeting apoptosis proteins in hematological malignancies. *Cancer Lett.* **332**, 325-334 (2013).

13. F. F. Heredia *et al.*, Proteins related to the spindle and checkpoint mitotic emphasize the different pathogenesis of hypoplastic MDS. *Leuk. Res.* **38**, 218-224 (2014).

14. F. Nolte *et al.*, Centrosome aberrations in bone marrow cells from patients with myelodysplastic syndromes correlate with chromosomal instability. *Ann. Hematol.* **92**, 1325-1333 (2013).

15. A. Kon *et al.*, Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat. Genet.* **45**, 1232-1237 (2013).

16. F. Thol *et al.*, Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. *Blood.*(2013).

17. J. Nangalia *et al.*, Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N. Engl. J. Med.* **369**, 2391-2405 (2013).

18. A. Kode *et al.*, Leukaemogenesis induced by an activating beta-catenin mutation in osteoblasts. *Nature.*(2014).

19. G. W. Rhyasen *et al.*, Targeting IRAK1 as a therapeutic approach for myelodysplastic syndrome. *Cancer. Cell.* **24**, 90-104 (2013).

20. M. Breccia, G. Alimena, NF-kappaB as a potential therapeutic target in myelodysplastic syndromes and acute myeloid leukemia. *Expert Opin. Ther. Targets.* **14**, 1157-1176 (2010).

21. Y. Wei *et al.*, Toll-like receptor alterations in myelodysplastic syndrome. *Leukemia.* **27**, 1832-1840 (2013).

22. A. Jerez *et al.*, STAT3 mutations indicate the presence of subclinical T-cell clones in a subset of aplastic anemia and myelodysplastic syndrome patients. *Blood.* **122**, 2453-2459 (2013).

23. F. Damm *et al.*, BCOR and BCORL1 mutations in myelodysplastic syndromes and related disorders. *Blood.* **122**, 3169-3177 (2013).

24. F. Thol *et al.*, SETBP1 mutation analysis in 944 patients with MDS and AML. *Leukemia.* **27**, 2072-2075 (2013).

25. R. Humeniuk, R. Koller, J. Bies, P. Aplan, L. Wolff, Loss of p15Ink4b accelerates development of myeloid neoplasms in Nup98-HoxD13 transgenic mice. *Stem Cells.*(2014).

26. C. F. Taylor, F. M. Platt, C. D. Hurst, H. H. Thygesen, M. A. Knowles, Frequent inactivating mutations of STAG2 in bladder cancer are associated with low tumour grade and stage and inversely related to chromosomal copy number changes. *Hum. Mol. Genet.* **23**, 1964-1974 (2014).

27. C. Balbas-Martinez *et al.*, Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. *Nat. Genet.* **45**, 1464-1469 (2013).

28. G. Guo *et al.*, Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat. Genet.* **45**, 1459-1463 (2013).

29. S. F. Mahmood *et al.*, A siRNA screen identifies RAD21, EIF3H, CHRAC1 and TANC2 as driver genes within the 8q23, 8q24.3 and 17q23 amplicons in breast

cancer with effects on cell growth, survival and transformation. *Carcinogenesis.* **35**, 670-682 (2014).

30. S. Deb *et al.*, RAD21 cohesin overexpression is a prognostic and predictive marker exacerbating poor prognosis in KRAS mutant colorectal carcinomas. *Br. J. Cancer.* **110**, 1606-1613 (2014).

31. D. A. Solomon *et al.*, Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science.* **333**, 1039-1043 (2011).

32. A. Djos, S. Fransson, P. Kogner, T. Martinsson, Aneuploidy in neuroblastoma tumors is not associated with inactivating point mutations in the STAG2 gene. *BMC Med. Genet.* **14**, 102-2350-14-102 (2013).

33. F. A. Ran *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281-2308 (2013).

34. K. S. Fernandez, P. A. de Alarcon, Development of the hematopoietic system and disorders of hematopoiesis that present during infancy and early childhood. *Pediatr. Clin. North Am.* **60**, 1273-1289 (2013).

35. D. E. Rollison *et al.*, Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001-2004, using data from the NAACCR and SEER programs. *Blood.* **112**, 45-52 (2008).

36. A. Giagounidis, D. Haase, Morphology, cytogenetics and classification of MDS. *Best Pract. Res. Clin. Haematol.* **26**, 337-353 (2013).

37. D. Haase *et al.*, New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood.* **110**, 4385-4395 (2007).

38. J. M. Bennett *et al.*, Proposals for the classification of the myelodysplastic syndromes. *Br. J. Haematol.* **51**, 189-199 (1982).

39. D. A. Arber *et al.*, The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood.* **127**, 2391-2405 (2016).

40. J. W. Vardiman *et al.*, The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood.* **114**, 937-951 (2009).

41. P. Greenberg *et al.*, International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood.* **89**, 2079-2088 (1997).

42. P. L. Greenberg *et al.*, Revised international prognostic scoring system for myelodysplastic syndromes. *Blood.* **120**, 2454-2465 (2012).

43. A. G. Kulasekararaj, A. M. Mohamedali, G. J. Mufti, Recent advances in understanding the molecular pathogenesis of myelodysplastic syndromes. *Br. J. Haematol.* **162**, 587-605 (2013).

44. D. Dorsett, L. Strom, The ancient and evolving roles of cohesin in gene expression and DNA repair. *Curr. Biol.* **22**, R240-50 (2012).

45. M. J. Walter *et al.*, Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* **366**, 1090-1098 (2012).

46. J. S. Welch *et al.*, The origin and evolution of mutations in acute myeloid leukemia. *Cell.* **150**, 264-278 (2012).

47. F. Cucco *et al.*, Mutant cohesin drives chromosomal instability in early colorectal adenomas. *Hum. Mol. Genet.* **23**, 6773-6778 (2014).

48. L. Evers *et al.*, STAG2 is a clinically relevant tumor suppressor in pancreatic ductal adenocarcinoma. *Genome Med.* **6**, 9 (2014).

49. G. Garcia-Manero, Myelodysplastic syndromes: 2015 Update on diagnosis, risk-stratification and management. *Am. J. Hematol.* **90**, 831-841 (2015).

50. M. Hahn *et al.*, Suv4-20h2 mediates chromatin compaction and is important for cohesin recruitment to heterochromatin. *Genes Dev.* **27**, 859-872 (2013).

51. M. Mossner *et al.*, Mutational hierarchies in myelodysplastic syndromes dynamically adapt and evolve upon therapy response and failure. *Blood.* **128**, 1246-1259 (2016).

52. G. Montalban-Bravo *et al.*, STAG2 Mutations Are an Independent Prognostic Factor in Patients with Myelodysplastic Syndromes. **128**, 3182 (2016).

53. L. Gorunova *et al.*, Cytogenetic analysis of 101 giant cell tumors of bone: nonrandom patterns of telomeric associations and other structural aberrations. *Genes Chromosomes Cancer.* **48**, 583-602 (2009).

54. C. Gelot *et al.*, The Cohesin Complex Prevents the End Joining of Distant DNA Double-Strand Ends. *Mol. Cell.* **61**, 15-26 (2016).

55. F. Tirode *et al.*, Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer. Discov.* **4**, 1342-1353 (2014).

56. S. Thota *et al.*, Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood.*(2014).

57. J. Mullenders *et al.*, Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. *J. Exp. Med.* **212**, 1833-1850 (2015).