Systems medicine in the scope of multi omics analyses

Inaugural-Dissertation
zur Erlangung der Doktorwürde
(Doctor Rerum Naturalium)
der Fakultät für Biologie

vorgelegt von

Patrick Metzger
geboren in Emmendingen

an der
Albert-Ludwigs-Universität Freiburg

Freiburg im Breisgau, August 2020

angefertigt am
Institut für Medizinische Bioinformatik und Systemmedizin
Universitätsklinikum Freiburg – Medizinische Fakultät

Dekan der Fakultät für Biologie: Prof. Dr. Dierk Reiff
Promotionsvorsitzender: Prof. Dr. Andreas Hiltbrunner

Betreuer der Arbeit: Prof. Dr. Dr. Melanie Börries

Referent: Prof. Dr. Dr. Melanie Börries
Koreferent: Prof. Dr. Tilman Brummer
Drittprüfer: PD Dr. Lena Illert

Datum der mündlichen Prüfung: 22.10.2020

Eidesstaatliche Erklärung

i) Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quellen gekennzeichnet. Insbesondere habe ich hierfür nicht die entgeldliche Hilfe von Vermittlungs- beziehungsweise Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit den Inhalten der vorgelegten Dissertation stehen.

ii) Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

iii) Die Bestimmungen der Promotionsordnung der Fakultät für Biologie sind mir bekannt, insbesondere weiß ich, dass ich vor Vollzug der Promotion zur Führung des Doktortitels nicht berechtigt bin.

Freiburg im Breisgau, den 03. August 2020

Patrick Metzger

"If my mind can conceive it, and my heart can believe it - then I can achieve it." –
Muhammad Ali

Danksagung

Ein besonderer Dank gilt meiner Betreuerin Melanie Börries, die mich in allen Situationen immer untersützt hat, zu jeder Zeit mit offenen Ohren und Armen da war und mich mit vielen Tipps, Ideen und Ratschlägen durch meine Doktorandenzeit geleitet hat. Ohne sie wäre die Arbeit in dieser Form nicht möglich gewesen. Zusätzlich möchte ich mich bei ihr bedanken, dass ich meine Doktorarbeit in ihrer Arbeitsgruppe machen durfte. Die unzähligen Gespräche, die Gruppenausflüge, das Skifahren am Feldberg, die Weihnachtsessen bei ihr zu Hause und viele Events mehr, werden mir für immer in Erinnerung bleiben. Es ware eine sehr tolle Zeit.

Ein herzliches Dankschön geht an Hauke Busch, der am Anfang meiner Doktorarbeit die Arbeitsgruppe zusammen mit Melanie geleitet hat. Auch er hatte immer ein offenes Ohr für jegliche Anliegen und stand immer mit Rat und Tat zu allen bioinformatischen Problemstellungen und darüber hinaus zur Seite.

Ebenso möchte ich mich bei allen jetzigen und ehemaligen Kolleginnen und Kollegen in der Arbeitsgruppe bedanken. Ohne eueren Zuspruch und die ergiebigen Diskussionen wäre die Arbeit nicht möglich gewesen. Insbesondere möchte ich mich bei Geoffroy, Hagen, Silke und Maria bedanken.

Zuletzt gilt der Dank meiner Familie, insbesondere Ariane und meinen Eltern. Euer Zuspruch und eure Motivation, die richtigen Worte in schwierigen Situationen und den ab und zu benötigten Tritt in den Hintern lieferten einen sehr wichtigen Beitrag zu dieser Arbeit.

Abstract

Health care is currently undergoing a radical change from a evidenced based medicine towards a more personalized, precise and systematic approach. Therefore, a variety of data sources, including conventional clinical data, imaging, molecular and genetic data from high-throughput methods, etc, are needed to provide accurate results. But it is the interplay of different research areas, such as biology, medicine and bioinformatics, that enables competent data analysis and interpretation.

Within this thesis a broad spectrum of tools and methods has been applied to answer specific scientific questions ranging from *in vitro* studies on an acute myeloid leukemia (AML) cell line, to studies within a mouse model for clear cell renal cell carcinoma (ccRCC), and actual in *in vivo* studies in patients with familial myelodysplastic syndrome (MDS). Additionally, a bioinformatics pipeline was developed to analyse the patients of the molecular tumor board (MTB) Freiburg followed by a subsequent stakeholder analysis to clarify the needs of clinicians on a platform that supports the MTB.

In summary, this thesis consists of a compendium of manifold applied bioinformatics approaches that use multi omics high-throughput data sets as well as conventional clinical and biological data to answer different research questions with the aim to foster a more personalized, precise and systematic way in patient care.

During the time of this thesis the following manuscripts were accepted for publication or are under preparation of being published:

- Hoefflin, R., Geißler, A.-L., Fritsch, R., Claus, R., Wehrle, J., Metzger, P., Reiser, M., Mehmed, L., Fauth, L., Heiland, D. H., Erbes, T., Stock, F., Csanadi, A., Miething, C., Weddeling, B., Meiss, F., von Bubnoff, D., Dierks, C., Ge, I., von Bubnoff, N. **Personalized Clinical Decision Making Through Implementation of a Molecular Tumor Board: A German Single-Center Experience**. JCO Precision Oncology. 2018

- Dittmann, J., Haydn, T., Metzger, P., Ward, G. A., Boerries, M., Vogler, M., Fulda, S. **Next-generation hypomethylating agent SGI-110 primes acute myeloid leukemia cells to IAP antagonist by activating extrinsic and intrinsic apoptosis pathways**. Cell Death and Differentiation. 2019

- Buechner, P., Hinderer, M., Unberath, P., Metzger, P., Boeker, M., Acker, T., Haller, F., Mack, E., Nowak, D., Paret, C., Schanze, D., von Bubnoff, N., Wagner, S., Busch, H., Boerries, M., Christoph, J. **Requirements analysis and specification for a molecular tumor board platform based on cBioPortal**. Diagnostics. 2020

- Kozyra, E. J., Pastor, V. B., Lefkopoulos, S., Sahoo, S. S., Busch, H., Voss, R. K., Erlacher, M., Lebrecht, D., Szvetnik, E. A., Hirabayashi, S., Pasaulienė, R., Pedace, L., Tartaglia, M., Klemann, C., Metzger, P., Boerries, M., Catala, A., Hasle, H., de Haas, V., Wlodarski, M. W. **Synonymous *GATA2* mutations result in selective loss of mutated RNA and are common in patients with *GATA2* deficiency**. Leukemia. 2020

- Hoefflin, R., Harlander, S., Schäfer, S., Metzger, P., Kuo, F., Schönenberger, D., Adlesic, M., Peighambari, A., Seidel, P., Chen, C., Consenza-Contreras, M., Jud, A., Lahrmann, B., Grabe, N., Heide, D., Uhl, F., Chan, T. A., Duyster, J., Zeiser, R., Schell, C., Heikenwalder, M., Schilling, O., Hakimi, A. A., Boerries, M., Frew, I. J. ***HIF-1*α and *HIF-2*α differently regulate tumour development, metabolism and inflammation of clear cell renal cell carcinoma in mice**. Nature Communications. 2020

- Metzger, P., Hess, M. E., Scheible R., Beoker, M., Geoffroy, A., Boerries, M. **MIRACUM-Pipe: an adaptable pipeline for whole exome sequencing and reporting for clinical decision making**. (under preparations, 2020).

Additionally to the publications enclosed in this cumulative thesis, the following publications are under consideration for publication:

- Ketterer, S., Mitschke, J., Ketscher, A., Schlimpert, M., Reichardt ,W., Baeuerle, N., Hess, M. E., Metzger, P., Boerries, B., Peters, C., Kammerer, B., Brummer, T., Steinberg, F., Reinheckel, T. **Cathepsin D deficiency in mammary epithelium transiently stalls breast cancer by interference with mTORC1 signaling**. (under review at Nature Communications, 2020).
- Zharavin, V., Balmford, J., Metzger, P., Boerries, M., Binder, H., Boeker, M. **Annotation of human exome gene variants with consensus pathogenicity**. (under review at MDPI Genes, 2020).

Zusammenfassung

Die medizinische Versorgung vollzieht derzeit einen radikalen Wandel von einer evidenzbasierten Medizin hin zu einem stärker personalisierten, präziseren und systematischeren Ansatz. Daher wird eine Vielzahl von Datenquellen benötigt, darunter konventionelle klinische Daten, Bildgebung, molekulare und genetische Daten aus Hochdurchsatzverfahren usw., um präzise Ergebnisse liefern zu können. Aber erst das Zusammenspiel verschiedener Forschungsgebiete, wie Biologie, Medizin und Bioinformatik, ermöglicht eine kompetente Datenanalyse und -interpretation.

Im Rahmen dieser Arbeit wurde ein breites Spektrum an Werkzeugen und Methoden angewandt, um spezifische wissenschaftliche Fragen zu beantworten, die von *in vitro*-Studien an einer Zelllinie der akuten myeloischen Leukämie (AML) über Studien mit einem Mausmodell für das klarzellige Nierenzellkarzinom (ccRCC) bis hin zu tatsächlichen *in vivo*-Studien an Patienten mit dem familiären myelodysplastischen Syndrom (MDS) reichen. Zusätzlich wurde eine Bioinformatik-Pipeline entwickelt, um die Patienten des Molekularen Tumorboards (MTB) Freiburg zu analysieren, gefolgt von einer anschließenden Stakeholder-Analyse zur Klärung der Bedürfnisse der Kliniker an eine Plattform, die das MTB unterstützt.

Zusammenfassend besteht diese Arbeit aus einem Kompendium vielfältig angewandter bioinformatischer Ansätze, die sich sowohl mit Multi Omics Hochdurchsatz-Datensätzen als auch mit konventionellen klinischen und biologischen Daten befassen, um verschiedenste Forschungsfragen zu beantworten, mit dem Ziel, eine individuellere, präzisere und systematischere Art der Patientenversorgung zu fördern.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Systems medicine

Despite the fact that medicine has always been personalized, different terms or terminologies exist nowadays trying to describe a more personalized, precise and systematic approach to tackle various diseases. Terms like personalized medicine, P4 medicine (P4 = predictive, preventive, personalized and participatory) or precision medicine and systems medicine illustrate the desire to establish a novel and advanced approach in medicine [Apweiler et al., 2018]. The goal, all the approaches have in common, is to improve diagnosis, target therapies, prognosis and prevention. The term 'systems medicine' includes the word *systems*, which intuitively describes a set of newly linked objects. In the context of medicine or biology this could be networks of connected molecules or populations of interacting cells. The first task in a systems approach is to identify the relevant yet measurable parameters. Afterwards, the system needs to be modeled such that the interactions in the system are described in a meaningful way and lead to the desired output. In a final step, the consequences of a particular interaction are studied and transferred to disease treatment and prevention [Apweiler et al., 2018].

For all the steps in systems medicine the combination of statistical, mathematical and computational tools and methods is for the appropriate data sets needed. In systems medicine all parts of different data sources are incorporated, like the complex biochemical processes described above as well as physiological and environmental interactions that sustain living organisms [Federoff and Gostin, 2009].

Modern medicine is rooted on fundamental biological discoveries, elucidation of underlying mechanisms, and is called evidence-based medicine. With this precedence, symptoms are parsed by organ system, cellular dysfunction, and molecular defect, often neglecting, due to its complexity, the dynamic interaction of all elements and how they affect the system as a whole. Therefore, the patient is often viewed as collection of visceral organs and a nervous system. This leads to an increasingly reduction frame-

work [Federoff and Gostin, 2009] and to a one-gene-one-disease relationship. However, for instance the presence of a disease-associated mutation does not always result in manifestation of the disease [Yadav et al., 2020]. Systems medicine tries to get beyond this reductionism. Type 1 diabetes mellitus and cystic fibrosis are two examples breaking this paradigm. In diabetes, the genes *HLA-DQB1*, *HLA-DRB1*, and 21 non-HLA genes are known to be contributory, but there are many environmental risk factors affecting the disease [Federoff and Gostin, 2009]. For cystic fibrosis, the *CFTR* gene is mainly responsible for the onset of the disease. Although almost all patients having cystic fibrosis carrying a *CFTR* mutation in both alleles. But on the other side, not all individuals with disease-associated mutations in the *CFTR* alleles obtain cystic fibrosis [Yadav et al., 2020]. Systems medicine, in contrasts, integrates conventional clinical data like hematologic, clinico-pathological, and imaging data with molecular and genetic data from multi omics high-throughput methods to achieve greater precision in diagnosis, the opportunity for earlier intervention, risk-based prevention, individualization of care, and the optimization of the patient-clinician interface [Federoff and Gostin, 2009].

Only the high availability of high-throughput omics data has made the step towards systems medicine possible. However, all the aforementioned procedures require a wide variety of data sources of the highest quality in order to deliver precise results. Additionally, there is a highly developed understanding of the underlying processes as well as methods and algorithms from the interdisciplinary fields of biology, systems biology, medicine, bioinformatics and medical informatics needed.

In the scope of this thesis, six research projects are presented in which numerous tools and algorithms from all the above-mentioned areas are acquired, adapted, and applied in order to gain insights and answers to open questions in the context of biological and medical research. Therefore, handling of multi omics data from genomics and transcriptomics as well as conventional clinical data was required.

1.2 Genomics

Genomics is the interdisciplinary field of biology that studies the structure, function, evolution, mapping, and editing of genomes. A genome is the complete set of deoxyribonucleic acid (DNA) molecules of an organism that contains all genes and thus the complete blueprint including instructions for growth, development, function and reproduction. The DNA molecule consists of two complementary polynucleotide chains, called strands, coiled around each other to form a double helix. The two strands are composed

of one of the four nucleobases adenine (A), thymine (T), guanine (G), cytosine (C), deoxyribose (sugar) and a phosphate group [Alberts et al., 2014]. The nucleotide chain is connected by a covalent bond between the sugar and the phosphate group whereas the strands are connected via hydrogen bonds between the bases to from the double-stranded DNA. The genetic code of the human genome is organized in 23 chromosomes containing approximately 3 billion nucleotides [Craig Venter et al., 2001]. Every single cell in the human body, with the exception of red blood cells, carries at least one copy of the code in its nucleus. In 1975 Frederick Sanger and Alan Coulson published the first sequencing method for DNA molecules, a refined version of which is still known and used today as the Sanger method [Sanger and Coulson, 1975, Sanger et al., 1977]. This breakthrough in deciphering the genetic code of many species gave birth to the field of genomics and provided insights into the development and functioning of an organism as well as into the course and development of diseases, which led to enormous progress in biological and medical understanding. To meet the high demand for sequencing data, the Sanger sequencing was superseded by next-generation or high-throughput sequencing (NGS) methods. NGS technologies produce millions of sequences in a single pass and enable the sequencing of the entire human genome in less than a day. In 1990, a huge effort was undertaken from many scientists to decipher the human genome and after 13 years a first version was published. After another 4 years the human reference genome was declared "finished" containing less than one error in 20000 base-pairs (bp) and all chromosomes were assembled [McElheny, 2010].

The genetic information of DNA is contained within genes. A complete set of information in an organism is called genotype. A gene is an inheritance unit that contains transcribable regions, so called open reading frames (ORF), and regulatory regions such as promoters and enhancers. The regulatory regions control the transcription of the ORF. The genes are mainly responsible for the functioning and development of an organism. But only a small fraction of sequences of the genome encode for proteins, e.g. only about 1.5% of the human genome consists of protein-coding exons [Wolfsberg et al., 2001].

1.2.1 Gene expression and regulation

During gene expression the genetic code is translated, i.e. the nucleotide sequence of a gene is translated into the amino acid sequence of a polypeptide that forms a protein. Two main steps are necessary for this: The first step in formation of proteins is the transcription of the genetic code contained in the DNA into ribonucleic acid (RNA). Be-

sides the exchange of thymine (T) to uracil (U), the RNA is a perfect copy of the DNA. The RNA carrying the protein information is called messenger RNA (mRNA). The start site of the transcription is defined by a promoter, a small base sequence that marks the transcription start site (TSS) of a gene. The end of transcription is also marked through a defined base sequence. In an eukaryote, the mature mRNA is usually shorter than the original DNA due to a process known as splicing. During RNA-splicing, parts of a gene sequence (introns) were removed and only the exons remain and are connected to each other. This process takes place during or immediately after transcription. The second step is the translation of nucleotide sequences to amino acid sequences. The amino acid sequence distinguishes the resulting structure and function of the protein. The translators of the nucleotide sequence into the amino acid sequence are the so called transfer RNAs (tRNAs), which are carried out on the ribosome. The pathway of the genetic information is depicted in Fig. 1.1. Apart from mRNAs and tRNAs more RNA classes exist as described in Tab. 1.1.

The genetic code which carries the blueprint for creating amino acid sequences from nucleotide sequences has some generic properties related to its functions:

Triplets Three nucleotides, called codons, build the elementary information unit and encode for a single amino acid. There exist $4^3 = 64$ possible codons. 61 of the 64 codons encode for amino acids, three (UAA, UAG, UGA) encode for stop codons, i.e. marking the end of a reading section. Two codons, AUG and GUG, encode for an amino acid (methionine and valine) as well as a start codons, i.e. marking the beginning of a reading section.

Continuity The triplets are continuously translated. No separators nor overlap of codons exist.

Degeneracy Although 61 codons encode for amino acids, there exist only 20 different amino acids. Therefore, an amino acid can be encode by various codons.

Universality Almost all organism use the same genetic code.

In humans exist roughly 21000 coding genes (mRNAs) and about 9000 non-coding genes. Depending on the cell type, it varies, which genes or how much RNA is transcribed. In general, the cells express between $30 - 60\%$ of the total gene pool. The amount of transcribed RNA and the selection of genes is not static and could be altered

Table 1.1: Major types of RNAs and their function [Alberts et al., 2014]

Type of RNA	RNA function
mRNAs	messenger RNAs encode for protein.
tRNAs	transfer RNAs serve as link between mRNA and amino acids.
rRNAs	ribosomal RNAs are the catalytic components of the ribosomes and essential for protein synthesis.
snRNAs	small nuclear RNAs have various nuclear functions, e.g. splicing of pre-mRNA.
snoRNAs	small nucleolar RNAs process and chemically modify rRNAs.
scaRNAs	small cajal RNAs modify snoRNAs and snRNAs.
miRNAs	micro RNAs regulate gene expression.
siRNAs	small interfering RNAs turn off gene expression.
Other non-coding RNAs	functions in many cell processes, e.g. protein transport and X-chromosome inactivation.

through external signals. The more differentiated respectively specialized a cell is the more specific is their expression pattern.

1.2.2 DNA variations

Since DNA contains the blueprint of the genes that are mainly responsible for the development and function of an organism, a specific section that correlates with a certain phenotype is called a quantitative trait loci (QTL). Additionally, a region influencing the gene expression level is termed expression QTL (eQTL). These regions are of particular interest in genetics and variations within these regions can change the phenotype by altering gene expression or the resulting protein structure. Although the organism has various control and repair mechanisms during DNA replication, transcription or translation, variations occur with a probability of $1e^{-8}$ % per nucleotide. External influences, such as UV-radiation or chemical substances increase the probability of variations by breaking or weakening DNA bonds. A disturbed repair mechanism of the organism could therefore lead to variations within the DNA. In multi-cellular organisms, two forms of variation are distinguished depending on their heritability:

somatic variations occur in body cells. All cells that originate from these cells via mitoses carry the variation. Therefore, an organism could consist of cells carrying the variation and cells without. Somatic variations are not inheritable.

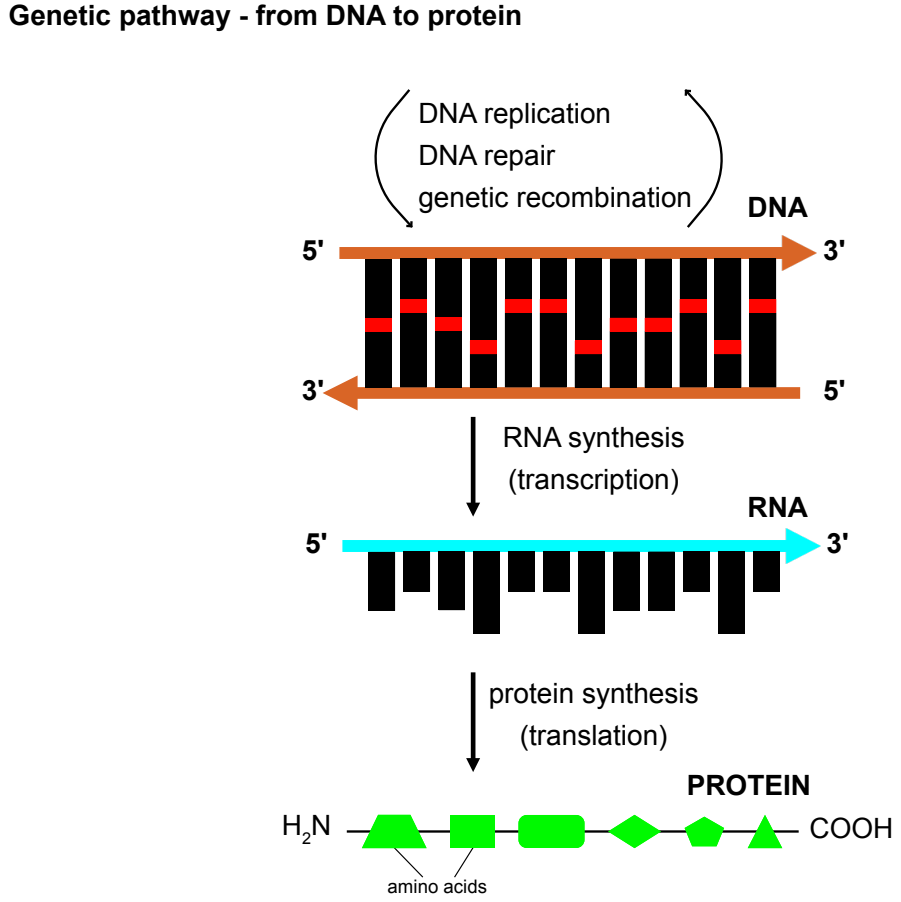


Figure 1.1: Genetic pathway: From DNA to protein. The process of protein synthesis via transcription from DNA to RNA and translation from RNA to protein occurring in living cells.

germline variations occur in germ cells and are inherited by.

Depending on the size of the DNA variations a distinction is made between:

- gene variants - point mutations
- chromosomal or structural variations
- numerical chromosomal aberrations

1.2.2.1 Gene variants - point mutations

The most common variations are so called gene variations or point mutations and occur in individual gene regions. Single nucleotides can be deleted (deletion), inserted (insertion), or replaced by another nucleotide (single nucleotide variation (SNV)), as seen in Fig. 1.2. This can lead to serious changes in the gene product, e.g. when one base is exchanged for another, a wrong amino acid is incorporated into a protein at the corresponding position or a newly created stop-codon prematurely terminates the translation of a protein. Due to the degeneracy of the genetic code, not all mutations lead to consequences in the resulting protein. Those variations are called 'silent' or 'synonymous' mutations. The different variant classifications of SNVs are depicted in Fig. 1.3.

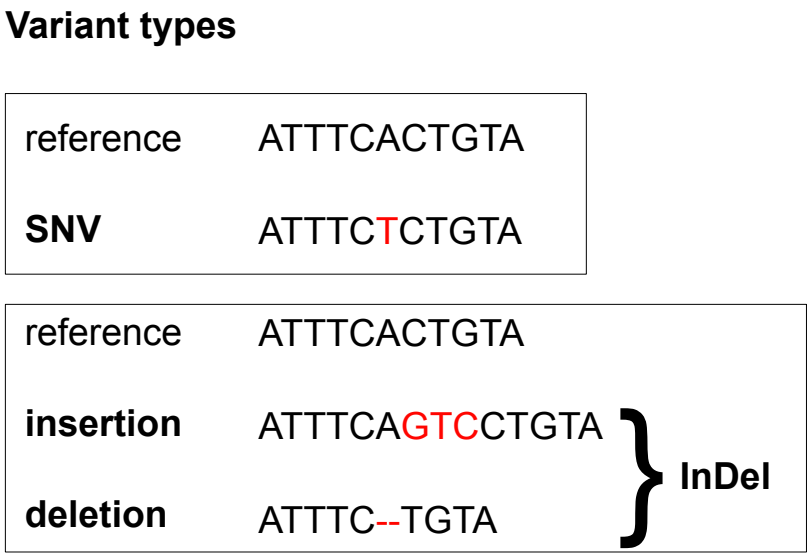


Figure 1.2: Gene variant types. SNV = single nucleotide variation, InDel = Insertion/Deletion.

If one, two or a number unequal to a multiple of three nucleotides are inserted or deleted at a position, a frame-shift mutation occurs, i.e. the reading frame of the triplets is shifted and all consecutive amino acids are affected.

In diploid organism, like humans, every cell carries two complementary copies of the complete DNA. If both alleles are identical, the state is called homozygous, and if they differ, it is called heterozygous, as shown in Fig. 1.4. As a result, one can distinguish between dominant and recessive mutations. For a dominat mutation a heterozygous condition is sufficient to cause a phenotype, whereas for a recessive mutation a homozygous condition is required. Additionally, loss of heterozygosity (LoH) can occur if a heterozy-

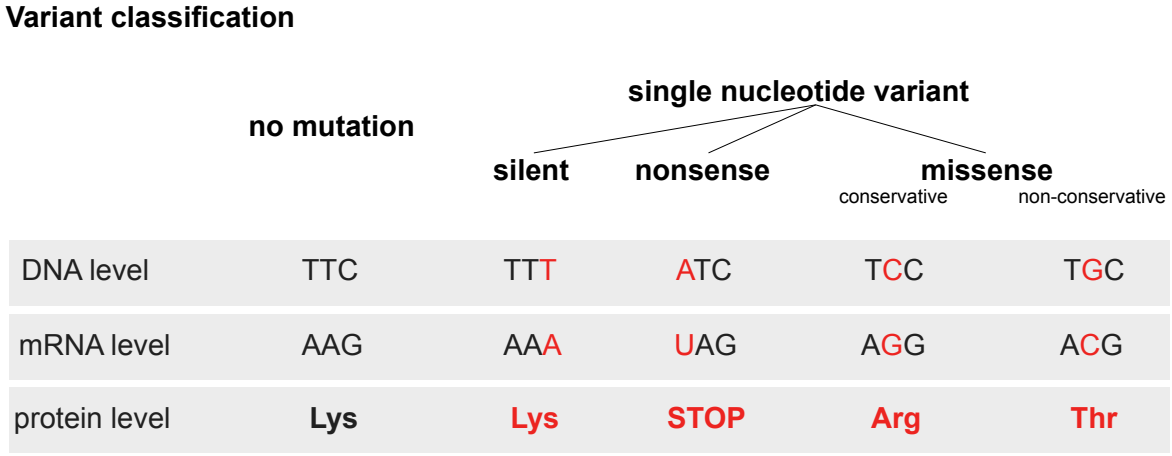


Figure 1.3: Variant classification. The classification of the possible single nucleotide variations. Lys = lysine, Arg = arginine, Thr = threonine.

gous individual acquires a mutation in the remaining allele leading to a homozygous state and thus to an observable phenotype.

If a SNV occurs with a frequency of $> 0.5\%$ in the population its referred to as a single nucleotide polymorphism (SNP). Due to the high frequency of SNPs, it is assumed that they are most likely not the drivers or origin of a rare disease and instead, the focus is on rare disease variants, especially in cancer research.

Gain-of-function mutations in genes associated with cell growth or cell proliferation (so-called proto-oncogenes) together with loss-of-function mutations in genes controlling for cell cycle processes (so-called tumor suppressor genes) play an important role in the development of cancer through uncontrolled cell growth. Oncogenes and tumor suppressor genes are grouped together as cancer genes and are frequently studied in the research field of oncology. Apart from cancer, SNVs on a single gene are also known to cause various other diseases like Alzheimer's disease [Wolf et al., 2013], Huntington' disease [Dayalu and Albin, 2015], or Phenylketonuria [Blau et al., 2010].

1.2.2.2 Chromosomal or structural variations

Chromosomal or structural variations affect the structure of chromosomes and are 1 kilo base (kb) or larger in size [Freeman, 2006]. A distinction is made between the following types, as shown in Fig. 1.5,

Deletion Loss of a chromosome fragment.

Duplication Doubling of a section within a chromosome.

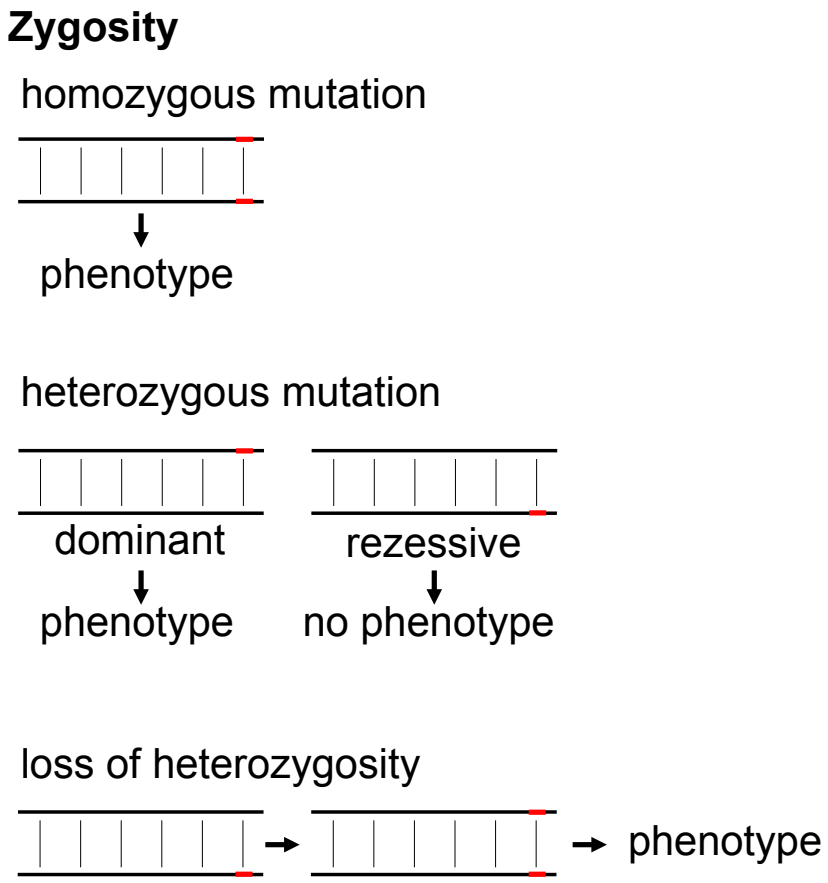


Figure 1.4: Zygosity. Homozygous and heterozygous mutations and their impact on phenotype.

Inversion One section rotates 180 degree.

Translocation A chromosomal piece gets transmitted to another, non homologous, chromosome.

Those structural changes are commonly called copy number variations (CNVs). A CNV refers to a change in the number of copies of a particular gene or region. In a diploid organism, each gene consists of two copies and structural variations are thought to produce gains, three or more copies, or losses, less than one copy.

Structural variations can as well lead to various diseases, such as the Charcot–Marie–Tooth disease (CMT) caused by a duplication on chromosome 17 around the *PMP22* gene [Lupski et al., 1991] or the Cri du chat syndrome caused by a deletion on the p-arm (short arm) of chromosome 5 [Lejeune et al., 1963].

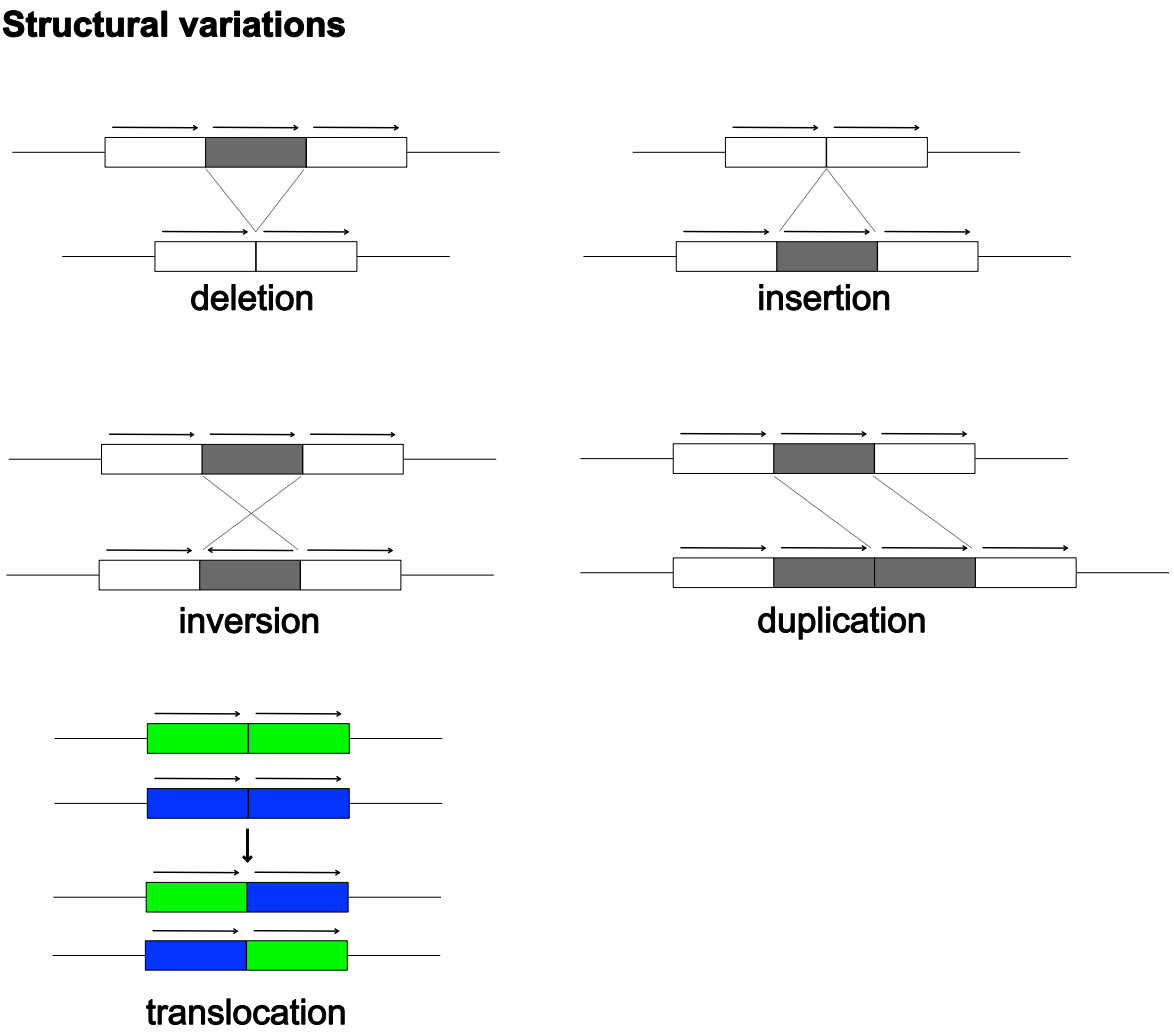


Figure 1.5: Structural variations. Different types of structural variations.

1.2.2.3 Numerical chromosomal aberration

If the number of chromosomes deviates from the normal configuration, the entire genome of an organism is affected. This type of variation is known as numerical chromosomal aberration or chromosome abnormality. The most common chromosome abnormality in humans is trisomy 21 also known as the Down syndrome [Malt et al., 2013], caused by the presence of a third copy of chromosome 21.

1.2.3 Fusion genes

A special type of a structural variation is a fusion gene, i.e. a hybrid gene formed from two independent genes. Fusion genes can occur from translocations, deletions or

inversions. The first fusion gene described in the early 1980s in the context of cancer was later termed Philadelphia chromosome; a translocation between chromosomes 9 and 22. The chromosomal translocation led to the fusion gene *BCR/ABL1* which could induce chronic myeloid leukemia (CML) [Mitelman et al., 2007].

2 Material and Methods

In the following chapter the methods used in this thesis to answer medical and biological questions were described.

2.1 Genome analysis

Before DNA can be analysed bioinformatically it has to be sequenced. Prior to sequencing, the DNA needs to be extracted, purified, and the DNA library prepared. In general the preparation includes five major steps after DNA procedures:

1. Fragmentation
2. Phosphorylation of 5' ends
3. A-tailing of 3' ends for adapter ligation
4. Adapter ligation
5. Amplification

Illumina has practically become standard in the field for high-throughput sequencing. Once the DNA library is prepared, the DNA will be applied to the surface of a so called flow cell and amplified to form local DNA colonies. The DNA colonies, mixed together with nucleotides labeled with a fluorophore, are than measured with a high-resolution camera. This technique is called sequencing by synthesis and the different bases are distinguished by their fluorescent tag. The output of a sequencing run is stored as a text file in the FASTQ format. It contains the measured bases and the corresponding quality scores in ASCII format. One sequence of bases from the sequencer is termed read and the FASTQ file consists of many such reads. One further distinguishes between single-end and paired-end reads. For single-end reads, the DNA fragment is only read from on side whereas for paired-end reads the fragment is read from both directions leading to

a higher accuracy and confidence about the measured bases. Sequencing with the latest high-throughput devices, like Illumina HighSeq or NovaSeq sequencers, produce around 3 billion reads with a paired-end read length of 150 base pairs (bp). The billion reads need now to be aligned to their exact location in the reference genome as depicted in Fig. 2.1 A. But before the process of alignment, the quality of each base contained in a read has to be analysed and low quality bases were removed. Additionally, adapters and bases contained in the low quality segment regions of a read, i.e. at the end of a read, were trimmed to obtain reads with sufficient quality for further processing. In this work, if not stated differently, a base quality higher than phred-score > 30, i.e. the probability of a incorrect base is 1 to 1000 or the base call accuracy is 99.9%, is kept. Trimmomatic [Bolger et al., 2014] was used to perform the task of quality control and adapter trimming. It uses a sliding window approach to cut the bases once the average quality fails to reach the quality threshold. For alignment of the reads the method of 'local seed and extend' using an indexed genome was used as depicted in Fig. 2.1 A. Therefore the assumption is made that the beginning of a read has few errors. In general, one has to be aware that sequence alignment is always a trade off between accuracy and speed and an alignment is never exact.

In this work, the Burrows-Wheeler aligner (bwa) [Li and Durbin, 2010], which relies on the Burrows-Wheeler transformation, was used for DNA alignment. The Burrows-Wheeler transformation transforms strings to have repeated characters in sorted order and than searches whether the DNA fragment is a sub-string of the reference. The alignment information, including position and mapping quality, is, together with the information of the read, stored as a text file in the sequence alignment format (SAM or in a compressed version as BAM).

From an information contextual perspective, a distinction is made between whole genome sequencing (WGS) and whole exome sequencing (WES). WGS covers the complete genome, including coding and non-coding elements whereas WES focuses only on coding and functional elements, like protein-coding genes, and approximately 1.2% of the entire human genome [Sims et al., 2014] is covered. The resulting file size of NGS depends on the desired coverage or read depth, the read-length and the total size of the genome. For a complete human genome with a read-length of 100bp and paired-end sequencing at a coverage of 30x, i.e. each base pair is on average contained in 30 reads, sums up to a size of 90 GB for WGS respectively 8 GB for WES.

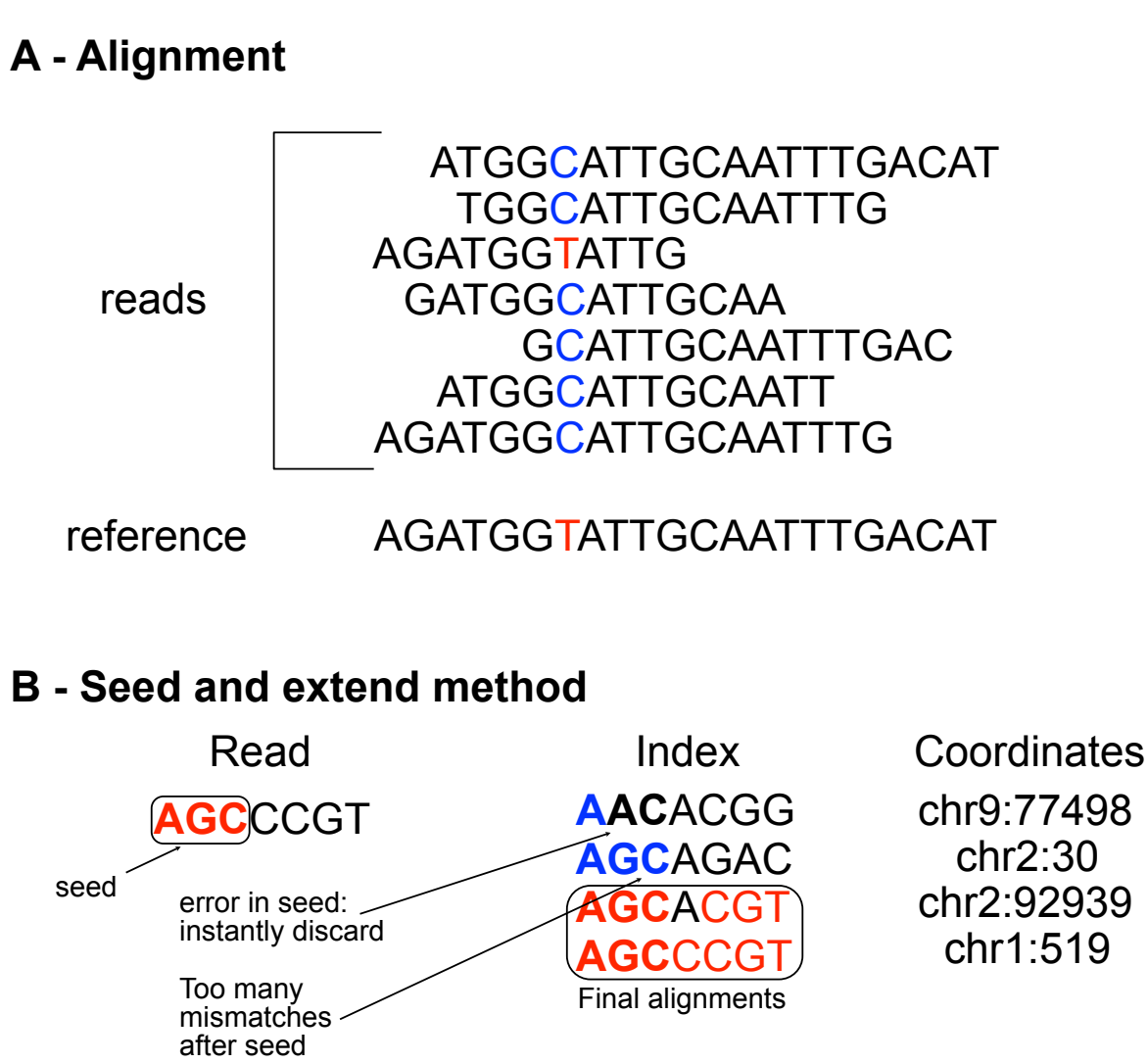


Figure 2.1: Alignment. **A** Alignment of the reads to the reference genome. **B** Schema of the 'seed and extend' method.

2.1.1 Variant calling

The process of identifying consistent differences between the sequenced reads and the reference genome is called variant calling. These differences include single nucleotide variations, small insertions and deletions, and larger copy number variants. Subsequently to the identification of variants, a proper annotation is necessary and very important. Depending on the classification of the variants, severe functional implications can occur, e.g. gain-of-function or loss-of-function mutations. Additionally, the origin of the mutation, i.e. is it a somatic or a germline mutation, the population frequency or minor allele frequency (MAF), the variant allele frequency (VAF), the potential treatment, etc. is

inferred or annotated. The variant calling results including the annotation are saved in a standardized variant calling format (VCF).

In this work, VarScan2 [Koboldt et al., 2012, Koboldt et al., 2013] was used for identifying the variants and subsequently annotated with annovar [Wang et al., 2010] including various databases, e.g. dbNSFP [Liu et al., 2016b] for functional implications, the Genome Aggregation Database (gnomAD) [Karczewski et al., 2020] for population frequencies and the drug gene interaction database (DGIdb) [Cotto et al., 2017] for possible treatment options. If not stated differently, the cutoffs for reporting variants were VAF $> 10\%$, MAF $< 0.1\%$ and only non-silent, i.e. missense and nonsense, variants were kept. VarScan2 uses the tumor and normal (germline) samples simultaneously as a pileup file performing pair wise comparisons of sequenced bases and the normalized read-depth at each position. A heuristic algorithm is used to independently derive the genotype for the tumor or germline based on the aforementioned limits for VAF, minimum coverage (8), baseline quality (30) and adjusted p-value (<0.05). The adjusted p-value is calculated based on Fisher's exact test comparing the read counts supporting either reference or alternative to the expected distribution based on the sequencing error. A variant with a VAF higher than 75% is considered homozygous. If, at a position, one or both, tumor and normal, carry a variant, a direct comparison of the genotype and supporting reads is performed to determine the somatic status. Is a variant present in the tumor sample it is called somatic, if its present heterozygous in the normal but homozygous in the tumor it is usually classified as LoH, or if the mutation is shared by both samples, it is classified as germline. Consecutively, a false-positive filter is applied to refine the variant call results. The filter aims to correct for the most common sequencing- or alignment-related artifacts.

2.1.2 Copy number variation calling

Copy number variation calling was performed with the tool Control-FREEC [Boeva et al., 2011, Boeva et al., 2012] within this work. The used algorithm calculates raw copy number profiles (CNP) by counting reads in non-overlapping windows. The window size is dependent and automatically determined by the coverage. Hereafter, the profiles are normalized either to the supplied control sample or, if not available, by GC content. The next step is the segmentation of normalized CNP and the analysis of the segmented profiles. The analyses include the identification of regions of genomic gains or losses and the prediction of copy number changes in these regions.

2.2 Trancriptome analysis

The analysis of all transcribed gens within a cell is called transcriptome analysis and is the foundation for qualitative and quantitative statements about the gene activity of the studied cells. Depending on the type of data generation, one distinguishes between microarrays and next-generation sequencing. Both methods have in common that the RNA is first converted into complementary DNA (cDNA) using reverse transcription.

2.2.1 DNA microarray

A DNA microarray consists of many thousand probes attached to its solid surface. Each probe carries a specifically designed cDNA sequence belonging to a known transcript. For a quantitative measurement, the test samples, labeled with a red and green fluorescent dye, bind to the corresponding cDNA sequence on the microarray. The position, intensity, and the emitted wavelength are measured by a high-resolution camera and provide the information about the amount of bound cDNA. The two most common manufacturers of microarrays are Affymetrix and Illumina.

The Affymetrix HGU 133 Plus 2 GeneChip microarray was used in this work and its pre-processing is explained in the following.

The raw Affymetrix CEL Files are imported with the R/Bioconductor [R Development Core Team, 2008,Gentleman et al., 2004,Huber et al., 2015] package affy [Gautier et al., 2004]. The CEL files contain the intensity values together with the standard deviation of each measured probe. After the import, the intensity values are background corrected and further normalized with the robust multi-array average (RMA) method. RMA fits a linear model to quantile normalized and log2 transformed intensity values to estimate the average expression of a gene from multiple probes [Irizarry et al., 2012]. Afterwards, the probe identifiers are converted to the corresponding gene identifiers. Within this work, if more than one probe matches to the same gene identifier, the probe with the largest inter-quartile-range (IQR) across the data set is kept. The resulting matrix after pre-processing contains gene expression values, i.e. normalized log2-transformed intensity values, for each measured gene identifier and each sample.

2.2.2 Next-generation sequencing

RNA sequencing follows the same steps as DNA sequencing described above, due to the reverse transcription of the RNA to cDNA. After the FASTQ files of the sequencing

are obtained, the reads have to be aligned to the reference genome together with the annotation of known transcripts with a splice-aware aligner.

In the scope of this thesis, the Spliced Transcripts Alignemnt to a Reference (STAR) software [Dobin et al., 2013] was used for this purpose. The STAR algorithm consists of two major steps:

1. seed searching step

2. clustering, stitching and scoring

For each read, the algorithm searches for the longest, exactly matching sequence in the reference genome. These part of a read is called maximal mapping prefix (MMP). All the differently aligned parts of a read are called seeds. In a subsequent step STAR searches the unmapped portion of a read to identify the next longest sequence exactly matching the reference genome. If a non exact match is identified, STAR allows for short mismatches compared to the reference and extends the MMP accordingly. But if the extension does not give a good alignment, the poor quality sequence is soft clipped. In the second step, the reads are clustered according to proximity to a set of anchor seeds or seeds without multi-mapping. Afterwards the seeds are stitched together based on the best alignment, i.e. a score based on mismatches, gaps, etc., for the read. The aligned reads are subsequently annotated with a gene transfer file (GTF) including the chromosomal coordinates for every transcript/gene. The STAR output contains the read counts for every gene within the GTF. A necessary step before down-stream processing of the raw counts is filtering of low expressed genes. Low expressed genes across all studied samples in a data set have no biological relevance because a certain amount of mRNA must be present in a cell for it to be transcribed into a protein.

In this work, a threshold of one count per million (cpm) in at least n samples, whereas n is equal to the smallest group size in the data set was used. The raw counts need to be normalized because they are affected by gene-length and total library size, i.e. the total number of counts in the sequencing run. The two most popular measures for gene-length normalized counts are either fragments per kilobase million (FPKM) values for paired-end reads respectively reads per kilobase million (RPKM) values for single-end reads or transcripts per million (TPM) values. But one major drawback of such measures is the cross-sample comparison because if the counts are heterogeneously distributed, e.g. if in one sample only a few highly abundant genes contribute to the biggest part of the library size, the low expressed genes are underestimated. Therefore, the use of normalization algorithms addressing such restrictions are recommended. The normalization method

almost solely used in the community is the weighted trimmed mean of M-value (TMM) normalization [Head et al., 2014, Conesa et al., 2016] which calculates scaling factors adjusting for the library size. Although the TMM methods is not correcting for GC bias and gene length, those effects are negligible for comparative analyses like identifying differentially expressed genes (DEGs).

2.2.3 Gene expression analysis

Although the two described methods for obtaining gene expression data, namely DNA microarrays and RNA sequencing (RNA-seq), are based on two completely different measurements techniques, i.e. DNA microarray intensity values follow a normal distribution whereas the count data from RNA-seq follow a negative binomial distribution, the same tools can be used for down-stream analyses if the RNA-seq data is used as log2-transformed CPMs which sufficiently well follow a normal distribution.

In this work, either the R/Bioconductor package limma (including voom if applicable) [Law et al., 2014, Ritchie et al., 2015] or DESeq2 [Love et al., 2014a, Love et al., 2014b] were applied for gene expression analysis. The main focus on gene expression analysis lies in the identification of relative differences in the expression of mRNAs between two, or more, given groups. More specifically, in the analysis for differentially expressed genes (DEGs) one tries to distinguish between significant differences between two experimental conditions or differences occurring just by chance. In the following, the limma workflow for a RNA-seq experiment is explained in greater detail. After importing the raw counts and the calculation of normalization factors, the low-expressed genes are filtered out. Subsequently a dimensionality reduction method, like multidimensional scaling (MDS) or a principal component analysis (PCA) is used to get an overview about the data and to identify potential outliers or mislabeled samples. The next step is to define the model one wants to fit to the data, i.e. define the experimental conditions. The model has to be defined before applying the voom transformation because voom uses the variance of the model residuals together with the log2-transformed CPMs to fit a linear model to each gene obtaining residuals for each. Further, the residuals are used to fit a smoothed curve to the square root of the residual standard deviation by the average expression which is later used to obtain weights for each gene and samples that are passed into limma along with the log2 CPM expression values. Next limma fits a linear model using weighted least squares to each gene, calculates the contrasts, i.e. log2 fold-changes (FC) between the groups, and performing empirical Bayes smoothing for standard errors. As results, for each gene the log2FC, average expression, and some statistical measures like

p-value or adjusted p-value are reported.

2.2.4 Gene set enrichment analysis

Despite identifying DEGs the normalized expression values can be used to perform a gene set enrichment analysis (GSEA). The GSEA identifies altered gene sets or pathways between two experimental conditions. As gene sets, public available databases like the Gene Ontology (GO) [Ashburner et al., 2000, Carbon et al., 2019], the Kyoto Encyclopedia of Genes and Genomes (KEGG) [Ogata et al., 1999], the Reactome Pathway Database [Jassal et al., 2020], ConsensusPathDB [Kamburov et al., 2012, Kamburov et al., 2013], the Molecular Signatures Database (MSigDB) [Subramanian et al., 2005, Liberzon et al., 2011, Liberzon et al., 2015], etc. can be used. For the GSEA itself the R/Bioconductor package GAGE [Luo et al., 2009] was used in this work. GAGE uses a parametric gene randomization procedure based on the log2FC statistic of each gene in the data set. The method assumes that the gene set comes from a different distribution than the background and therefore applies a two-sample t-test accounting for the variance within the gene set as well as the variance within the background. Additionally, GAGE calculates a global p-value for each gene set based on a meta-test on the negative log sum of the p-values from all one-on-one comparisons.

2.2.5 RNA fusions

For the identification of RNA fusions the tool FusionCatcher [Nicorici et al., 2014] was used within this work. FusionCatcher uses the raw FASTQ files and an ensemble of four different methods with four different aligners two identify RNA fusions. The reads passing the quality and filter criteria are aligned on the transcriptome and additionally on the genome. All reads mapping to the transcriptome are kept and searched for pairs such that one read maps on gene A and its paired-read maps on gene B. The identified gene pairs are than the potential candidate fusions. All unmapped reads are kept as well and further processed to shrink the candidate fusion list. A unmapped read is counted as evidence for a candidate fusion if its either found to map on exon-exon junction belonging to one of the candidate fusion pairs or if its corresponding paired-read maps to one of the genes forming the candidate fusion. In a further step, the reads are aligned to the before build database of candidate fusions. A read is counted as evidence for a candidate fusion if its either found to map on a gene-gene sequence belonging to a candidate fusion such as the first part of the read maps to the first gene and the second part maps to

the second gene forming the candidate fusion or the corresponding paired-read maps to the transcriptome of one of the genes forming the candidate fusion. The resulting list contains all candidate fusions found with the before mentioned methods.

3 Results

In this chapter, the thesis contains a compendium of research projects answering questions in the field of systems medicine. A variety of bioinformatics and medical informatics tools are used to elucidate the corresponding data and to clarify specific scientific questions. In total, there are six projects ranging from *in vitro* studies of a cancer cell line, over studies of a cancer entity in a model organism (mouse), and *in vivo* studies of a cancer entity in patients, to a bioinformatics approach for molecular tumor boards which is applied in personalized extended molecular-guided patient care and a subsequent stakeholder analysis. The research has led to publication, acceptance for publication or submission of manuscripts to peer-reviewed journals. The manuscripts include

1. Research on target therapies in acute myeloid leukemia focusing on hypomethylating agents triggering apoptosis pathways [Dittmann et al., 2019]
2. Research on clear cell renal cell carcinoma focusing on the role of two specific genes on the regulation of tumor development, metabolism and inflammation (accepted in Nature Communications)
3. Research in patients with *GATA2* deficiency in the context of myelodysplastic syndrome [Kozyra et al., 2020]
4. The development of a bioinformatics workflow to tackle the needs of molecular tumor boards (in preparations)
5. A retrospective case series of the molecular tumor board Freiburg [Hoefflin et al., 2018]
6. A stakeholder analysis of the requirements and specifications for a molecular tumor board to support clinicians [Buechner et al., 2020]

In the next sections the manuscripts are attached together with a short summary and my contributions to the work.

3.1 Next-generation hypomethylating agent SGI-110 primes acute myeloid leukemia cells to IAP antagonist by activating extrinsic and intrinsic apoptosis pathways

Acute myeloid leukemia (AML) is a disease occurring primarily in older adults manifesting itself as a heterogeneous disease characterized by clonal proliferation of poorly differentiated myeloid cells [O'Donnell et al., 2017]. Prognosis for elderly patients remain poor due to ineligibility for intensive treatments although some improvements in outcomes could be achieved for younger AML patients. Therefore, hypomethylating agents (HMA) are limited and the need for combination strategies is required. A priming of AML cells with SGI-110 (HMA) sensitizes them to the subsequent treatment with an antagonist of cellular inhibitor of apoptosis protein 1 and 2 (*cIAP1/2*) and x-linked IAP (*XIAP*). Both drugs synergistically induce cell death. In a transcriptome analysis it could be shown that SGI-110 alone or in combination with ASTX-660 up-regulated the expression of key regulators of extrinsic as well as intrinsic apoptosis signaling pathways. Both drugs are acting in a nice interplay providing a link between the two apoptotic pathways giving rise to a therapeutic potential in AML.

Dittmann, J., Haydn, T., Metzger, P., Ward, G. A., Boerries, M., Vogler, M., Fulda, S. (2019). **Next-generation hypomethylating agent SGI-110 primes acute myeloid leukemia cells to IAP antagonist by activating extrinsic and intrinsic apoptosis pathways**. Cell Death and Differentiation.

Contribution: I conducted the transcriptome analyses, including identification of differentially expressed genes and gene set enrichment analyses. I wrote the materials and methods parts *Transcriptome analysis* and *Gene set enrichment analysis (GSEA)*. I produced figure 3 and revised the whole manuscript.

Cell Death & Differentiation
https://doi.org/10.1038/s41418-019-0465-8

Cell Death & Differentiation

ARTICLE

Next-generation hypomethylating agent SGI-110 primes acute myeloid leukemia cells to IAP antagonist by activating extrinsic and intrinsic apoptosis pathways

Jessica Dittmann[1] · Tinka Haydn[1] · Patrick Metzger[2,3,4] · George A. Ward[5] · Melanie Boerries[2,3,6,7] · Meike Vogler[1] · Simone Fulda[1,7,8]

Received: 20 May 2019 / Revised: 21 November 2019 / Accepted: 21 November 2019

Abstract

Therapeutic efficacy of first-generation hypomethylating agents (HMAs) is limited in elderly acute myeloid leukemia (AML) patients. Therefore, combination strategies with targeted therapies are urgently needed. Here, we discover that priming with SGI-110 (guadecitabine), a next-generation HMA, sensitizes AML cells to ASTX660, a novel antagonist of cellular inhibitor of apoptosis protein 1 and 2 (cIAP1/2) and X-linked IAP (XIAP). Importantly, SGI-110 and ASTX660 synergistically induced cell death in a panel of AML cell lines as well as in primary AML samples while largely sparing normal CD34+ human progenitor cells, underlining the translational relevance of this combination. Unbiased transcriptome analysis revealed that SGI-110 alone or in combination with ASTX660 upregulated the expression of key regulators of both extrinsic and intrinsic apoptosis signaling pathways such as *TNFRSF10B* (DR5), *FAS*, and *BAX*. Individual knockdown of the death receptors TNFR1, DR5, and FAS significantly reduced SGI-110/ASTX660-mediated cell death, whereas blocking antibodies for tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL) or FAS ligand (FASLG) failed to provide protection. Also, TNFα-blocking antibody Enbrel had little protective effect on SGI-110/ASTX660-induced cell death. Further, SGI-110 and ASTX660 acted in concert to promote cleavage of caspase-8 and BID, thereby providing a link between extrinsic and intrinsic apoptotic pathways. Consistently, sequential treatment with SGI-110 and ASTX660-triggered loss of mitochondrial membrane potential (MMP) and BAX activation which contributes to cell death, as BAX silencing significantly protected from SGI-110/ASTX660-mediated apoptosis. Together, these events culminated in the activation of caspases-3/-7, nuclear fragmentation, and cell death. In conclusion, SGI-110 and ASTX660 cooperatively induced apoptosis in AML cells by engaging extrinsic and intrinsic apoptosis pathways, highlighting the therapeutic potential of this combination for AML.

Introduction

Acute myeloid leukemia (AML) is a heterogeneous disease characterized by clonal proliferation of poorly differentiated myeloid cells and occurs primarily in older adults

Edited by M. Piacentini Licence

Supplementary information The online version of this article (https://doi.org/10.1038/s41418-019-0465-8) contains supplementary material, which is available to authorized users.

✉ Simone Fulda
simone.fulda@kgu.de

[1] Institute for Experimental Cancer Research in Pediatrics, Goethe-University Frankfurt, Frankfurt am Main, Germany

[2] Institute of Medical Bioinformatics and Systems Medicine, Medical Center, Faculty of Medicine, University Freiburg, Freiburg im Breisgau, Germany

[3] Institute of Molecular Medicine and Cell Research (IMMZ), Albert Ludwigs-University Freiburg, Freiburg im Breisgau, Germany

[4] Faculty of Biology, University Freiburg, Freiburg im Breisgau, Germany

[5] Astex Pharmaceuticals, Cambridge, UK

[6] German Cancer Consortium (DKTK), Partner Site Freiburg, Freiburg im Breisgau, Germany

[7] German Cancer Research Center (DKFZ), Heidelberg, Germany

[8] German Cancer Consortium (DKTK), Partner Site Frankfurt, Frankfurt am Main, Germany

SPRINGER NATURE
CDDpress

(age ≥ 60 years) [1]. Despite improvements in outcomes for younger AML patients in recent decades, the prognosis for older patients, who are ineligible for intensive treatment, remains dismal [2], highlighting the urgent need for better therapeutic options.

Dysfunction of epigenetic modifiers, such as DNA methyltransferases (DNMTs), contributes to AML pathogenesis through aberrant epigenetic silencing of tumor suppressor genes (TSGs) involved in differentiation and apoptosis [3]. *DNMT3A* mutations are found in ~22% of AML patients, and around 60% of these mutations affect the R882 codon, which is highly associated with poor prognosis [3, 4]. First-generation hypomethylating agents (HMAs), such as azacytidine and decitabine, have been approved for the treatment of older AML patients [5]. SGI-110 (guadecitabine), a dinucleotide of decitabine and deoxyguanosine, is a next-generation HMA that is resistant to degradation by cytidine deaminase and provides a prolonged in vivo exposure compared with decitabine [6]. Recently, safety and clinical activity of SGI-110 in both elderly treatment-naive and relapsed/refractory AML patients have been shown in phase II trials [7, 8]. The proposed central mechanism of action of HMAs is the depletion of DNMTs, thus inducing hypomethylation of global DNA and CpG-island promoters, which might lead to gene expression of silenced TSGs [6, 9], and may sensitize tumor cells to other anticancer agents, including chemotherapeutics [10, 11], immunotherapeutics [12], or apoptosis-inducing agents [13].

Apoptosis plays an important role in the hematopoietic system. There are two well-defined pathways of apoptosis: the extrinsic and the intrinsic pathway [14]. The extrinsic (death receptor-mediated) pathway is activated upon interaction of death receptor ligands, such as tumor necrosis factor (TNF), TNF-related apoptosis-inducing ligand (TRAIL), and FAS ligand (FASLG) with their cognate death receptors TNF receptor 1 (TNFR1), TRAIL-R1 (DR4), TRAIL-R2 (DR5), and FAS, resulting in the activation of caspase-8, which can cleave downstream effector caspases [15]. The intrinsic (mitochondrial-mediated) pathway involves loss of mitochondrial membrane potential (MMP) due to mitochondrial outer membrane permeabilization (MOMP) that is controlled by proapoptotic (e.g., BAX, BAK, and BID) and anti-apoptotic proteins of the BCL-2 family. This leads to caspase activation, nuclear fragmentation, and apoptotic cell death [14]. A crosslink between extrinsic and intrinsic pathways is provided by caspase-8-mediated cleavage of BID into its active form (tBID) [16].

Evasion of apoptosis is a major cause of treatment resistance and is often caused by overexpression of antiapoptotic proteins such as inhibitor of apoptosis proteins (IAPs) [17]. Overexpression of XIAP has been associated with poor outcome in AML [18] and high cIAP1 expression has been reported in pediatric AML [19]. XIAP exerts its antiapoptotic activity by inhibiting caspases [20], while cIAP1/2 can regulate proapoptotic signaling complexes [17]. Therefore, IAPs are considered attractive targets for anticancer therapy. Various IAP antagonists have been developed which trigger autoubiquitination and proteasomal degradation of IAPs [21]. ASTX660, a novel orally bioavailable, nonpeptidomimetic antagonist of cIAP1/2 and XIAP [22], is currently under evaluation in a phase I/II clinical trial for advanced solid tumors and lymphomas (NCT: 02503423). However, several IAP antagonists have so far shown only weak single-agent efficacy in clinical trials [22].

Since the therapeutic efficacy of HMAs is limited in AML patients [13], there is a high medical need to identify novel combinations. Therefore, in the present study we investigated whether priming with next-generation HMA SGI-110 increases the sensitivity of AML cells towards cell death triggered by the novel IAP antagonist ASTX660.

Methods and materials

AML cell lines

AML cell lines were obtained from German Collection of Microorganisms and Cell cultures (DSMZ, Braunschweig, Germany), except ML-2 and PLB-985 cells that were kindly provided by T. Oellerich, Department of Medicine II, University Hospital Frankfurt, Germany. All cell lines were authenticated by STR profiling and routinely checked for mycoplasma contamination. All AML cell lines except OCI-AML-3 were cultured in RPMI 1640 medium (Life Technologies, Inc., Darmstadt, Germany) supplemented with 10% fetal calf serum (FCS), 1% penicillin/streptomycin, and 1% sodium pyruvate (Invitrogen, Karlsruhe, Germany). OCI-AML3 cells were cultured in alpha-MEM medium supplemented with 20% FCS, 1% penicillin/streptomycin, and 1% sodium pyruvate.

Primary samples

Bone marrow specimens from AML patients were obtained at diagnosis before the onset of therapy after obtaining written informed consent from patients according to the declaration of Helsinki and after approval by the local ethics committee of the University Hospital Frankfurt (Approval No. SHO-05-2014). Mononuclear cell (MNC) fractions were obtained by density gradient centrifugation using Ficoll Isopaque (Amersham Bioscience, Freiburg, Germany) and maintained as previously described [23]. The clinical characteristics of the AML patients are summarized in Supplementary Table S1. Primary samples with spontaneous cell death ≥40% at the time point of measurement were excluded from the analysis. MNCs and clinical data

were obtained from the hematological biobank and the tumor documentation of the UCT Frankfurt, Germany.

Human G-CSF-mobilized CD34+ hematopoietic progenitor cells (HPCs) from healthy donors were kindly provided by H. Bönig, Institute for Transfusion Medicine and Immunohematology, Frankfurt, Germany. After thawing, HPCs were cultured in IMDM enriched with 20% FCS, 1% penicillin/streptomycin, 50 ng/ml rh-SCF, 50 ng/ml IL-3, 100 ng/ml FLT3-Ligand (AF-300-19, PeproTech), and 20 ng/ml GM-CSF.

Determination of cell death and apoptosis

For measurements of cell death and apoptosis, cells were cultured at 1×10^5 cells/ml (AML cell lines) or 1×10^6 cells/ml (primary AML, CD34+ HPCs). Cell death of suspension cells was assessed by forward/side scatter (FSC/SSC) and flow cytometry (FACS Canto II, BD Biosciences, Heidelberg, Germany). Apoptosis was determined by FACS analysis following staining of the cells with Annexin V-FITC and propidium iodide (PI). Early-apoptotic (Annexin V+/PI−) and late-apoptotic cells (Annexin V+/PI+) were summarized as Annexin V+ cells. Primary AML cells were stained with Annexin V-FITC and 0.5 µl anti-CD45-APC antibody (17-0459-42, eBioscience, San Diego, CA, USA). Flow cytometry was used to determine cell death of primary AML blasts identified by using a CD45/SSC gating strategy as described previously [24]. Flow cytometric analysis using a PI-containing Nicoletti buffer was performed to determine the proportion of apoptotic sub-G1 hypodiploid cells with internucleosomal DNA fragmentation as previously described [25].

DNA isolation and global methylation assay

DNA was isolated using the PureLink Genomic DNA Mini Kit (Invitrogen, Carlsbad, CA, USA). Global DNA Methylation-long interspersed nuclear element-1 (LINE-1) kit (Active Motif, Carlsbad, CA, USA) was employed in accordance with the manufacturer's instructions.

Human genome microarray profiling

RNA was isolated using the PeqGold Total RNA Kit and the PeqGOLD DNAase digest kit (Peqlab, Erlangen, Germany) following the manufacturer's protocol. Human genome microarray profiling was performed at the DKFZ Genomics and Proteomics Core Facility (Heidelberg, Germany) using the Affymetrix Human Genome U133 Plus 2.0 Array.

Transcriptome analysis

Raw Affymetrix cell files were analyzed using the R/Bioconductor [26, 27] package affy [28] followed by quantile, robust multichip analysis (RMA) background normalization and log2 transformation. After preprocessing and filtering, 20352 genes were further analyzed. Differentially expressed genes (DEGs) were identified with the R/Bioconductor package limma [29]. For the comparisons of SGI-110/ASTX660-Ctrl, SGI-110-Ctrl, SGI-110-ASTX660 and SGI-110/ASTX660-ASTX660 the least-squares method and for the comparisons of ASTX-Ctrl and SGI-110/ASTX660-SGI-110 the robust method for linear model fitting were used. Genes were considered significant with an adjusted p value < 0.05 (Benjamini–Hochberg). Microarray data were deposited in the publicly accessible database Gene Expression Omnibus under accession number GSE138322.

Gene-set enrichment analysis (GSEA)

Enrichment of signaling pathways was performed as implemented in the R/Bioconductor package [30] with Gene ontology (GO) terms [31, 32], Reactome pathways [33] and ConsensusPathDB pathways [34]. Pathways are considered significant with an adjusted p value < 0.05 (Benjamini–Hochberg).

Cell surface expression of death receptors

Flow cytometric analysis of cell surface expression of TNFR1, DR4, DR5, and FAS was performed using following phycoerythrin (PE)-conjugated antibodies: anti-TNFR1 (130-106-286, Miltenyi Biotech, Bergisch Gladbach, Germany), anti-DR4 (FAB347P), anti-DR5 (FAB6311P, R & D Systems, Wiesbaden, Germany) and anti-FAS (556641, BD BioSciences, San Diego, CA, USA) and their respective isotype controls (IgG1 (IC002P), IgG2b (IC0041P, R & D systems, Wiesbaden, Germany)), REA control (130-113-462, Miltenyi Biotech, Bergisch Gladbach, Germany). In brief, treated cells were harvested and washed twice with ice-cold staining buffer (2% FCS in PBS). Afterwards, cells were incubated with PE-conjugated antibodies to the death receptors or respective isotype controls for 30 min at 4 °C in the dark. Cells were then washed twice with staining buffer and resuspended in PBS for flow cytometric analysis. Data were analyzed using FlowJo software.

Immunoprecipitation

BAX activation was determined by immunoprecipitation using active conformation-specific antibodies. Briefly, cells were lysed in CHAPS buffer (1% CHAPS, 150 mM NaCl, 10 mM HEPES, pH 7.4) supplemented with protease inhibitor cocktail (1169749800, Roche, Mannheim, Germany). Briefly, 500 µg protein was immunoprecipitated with 2 µl mouse anti-BAX antibody (6A7, Sigma-Aldrich,

Munich, Germany) and 10 µl pan-mouse IgG Dynabeads (Dako, Hamburg, Germany) overnight at 4 °C and washed with CHAPS buffer. The precipitate was analyzed for BAX expression by Western blotting, using the rabbit anti-BAX NT antibody (ABC11, Merck-Millipore, Burlington, MA, USA).

XIAP immunoprecipitation was performed as previously described [22] with the following exception: cell lysates were incubated with streptavidin-coated Dynabeads (60210, Thermo Fisher Scientific). The resulting precipitate was analyzed by Western blotting and the membranes probed with antibodies against second mitochondrial activator of caspases (SMAC) (#2954) and XIAP (#14334, Cell Signaling Technology) to detect interactions.

Determination of MMP

To measure MMP cells were incubated with 100 ng/ml tetramethylrhodamine methyl ester (TMRM Reagent; Thermo Fisher Scientific, Waltham, MA, USA) for 10 min at 37 °C, washed and directly analyzed by flow cytometry.

Western blot analysis

Cells were lysed using a Triton X-100 lysis buffer (30 mM TrisHCl, 150 mM NaCl, 10% Glycerol, 0.5 mM PMSF, 2 mM DTT, 1% Triton X-100, and 1× Protease Inhibitor Cocktail). Western blot analysis was carried out using the following primary antibodies: rabbit anti-BID (2002S, Cell Signaling), rabbit anti-caspase-8 (ab32125, Abcam), goat anti-cIAP1 (#AF8181, R & D systems), rabbit anti-BAX (ABC11, Merck-Millipore), rabbit anti-BAK (06-536, Upstate/Merck), mouse anti-XIAP (610716, BD Bioscience), rabbit anti-DNMT1 (39905, Active motif), mouse anti-DNMT3A (sc-365769, Santa Cruz Biotechnology), mouse anti-GAPDH (5G4-6C5, BioTrend (Hy Test Ltd)), mouse anti-β-ACTIN (A5441, Sigma), and rat anti-cIAP2 (ALX-803-341-C100, Enzo Life Sciences, Farmingdale, NY, USA). Goat anti-mouse IgG, goat anti-rabbit IgG (Abcam), goat anti-rat IgG, donkey anti-goat IgG conjugated to horseradish peroxidase (Santa Cruz Biotechnology, Santa Cruz, CA, USA), and enhanced chemiluminescence (Amersham Bioscience, Freiburg, Germany) or infrared dye-labeled secondary antibodies and infrared imaging (Odyssey Imaging System, LI-COR Bioscience, Bad Homburg, Germany) were used for detection. Representative blots of at least two independent experiments are shown.

RNA interference

Gene silencing with small interfering RNA (siRNA) was achieved using Silencer Select siRNA (Thermo Fisher Scientific, Waltham, MA, USA) and Neon Transfection System (Invitrogen, Karlsruhe, Germany) following the manufacturer's protocol. The following constructs were used: nontargeting control siRNA (4390843) or targeting siRNAs for BAX (#1: s1888, #2: s1890), BAK (#1: s1880, #2: s1881), cIAP1 (#1: s1449, #2: s1450, #3: s1448), XIAP (#1: s1454, 2: s1555, #3: s1556), TNFR1 (#1: s14266, #2: s14267), DR5 (#1: s16756, #2: s16758), and FAS (#1: s1506, #2: s1507, #3: s1508). For knockdown of cIAP1 and XIAP, cells were transfected once with 200 nM siRNA and immediately treated with SGI-110. For all other knockdowns, cells were transfected twice with 100 nM siRNA and treated 24 h after the second transfection.

Analysis of caspase-3/-7 activity and of morphological changes of the nucleus

Caspase activity was determined using Cell Event Caspase-3/-7 Green Detection Reagent (Thermo Fisher Scientific) according to manufacturer's instructions. For analysis by ImageXpress Micro XLS system (Molecular Devices, Biberach an der Riss, Germany), cells were additionally counterstained with Hoechst-33342 (Invitrogen, Carlsbad, CA, USA). For evaluation of the nuclear morphology, cells were stained with Hoechst-33342 and examined using an inverted fluorescence microscope (ImageXpress Micro XLS system) followed by analysis with MetaXpress Software (Molecular Devices Sunnyvale, CA, USA) using the nuclear fragmentation scoring tool.

Statistical analysis

All results are expressed as mean and standard deviation (SD). The numbers of independent repetitions and replicates for each experiment are indicated in the respective figure legends. Experiments were considered as reliable if the SD did not exceed 10% within the replicates and repetitions. For each in vitro independent experiment using cancer cell lines, technical triplicates were used and three experiments were performed to ensure adequate statistical power. Statistical significance was verified by using *t*-test in Microsoft Excel (two-samples, two-tailed distribution, unequal variance), unless otherwise indicated. For drug combination dose–response studies in multiple AML cell lines, data were analyzed by one-way ANOVA followed by Tukey's multiple comparisons test, using GraphPad Prism as specified in the table legend. Drug interaction was analyzed using CalcuSyn software (Biosoft, Cambridge, UK) [35]. The obtained values were rated as follows: $CI < 0.9$ indicates synergism, 0.9–1.1 additivity and >1.1 antagonism.

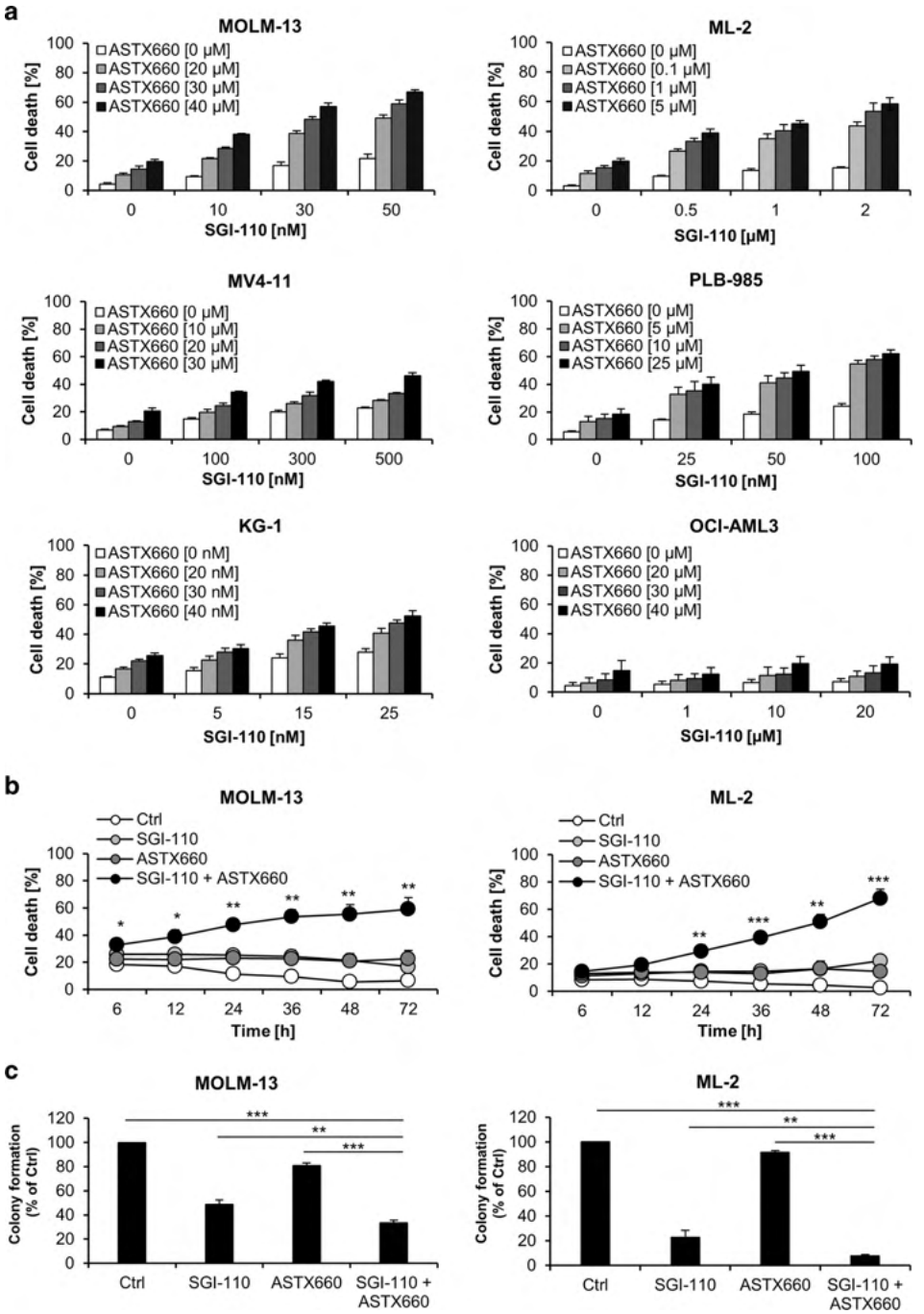
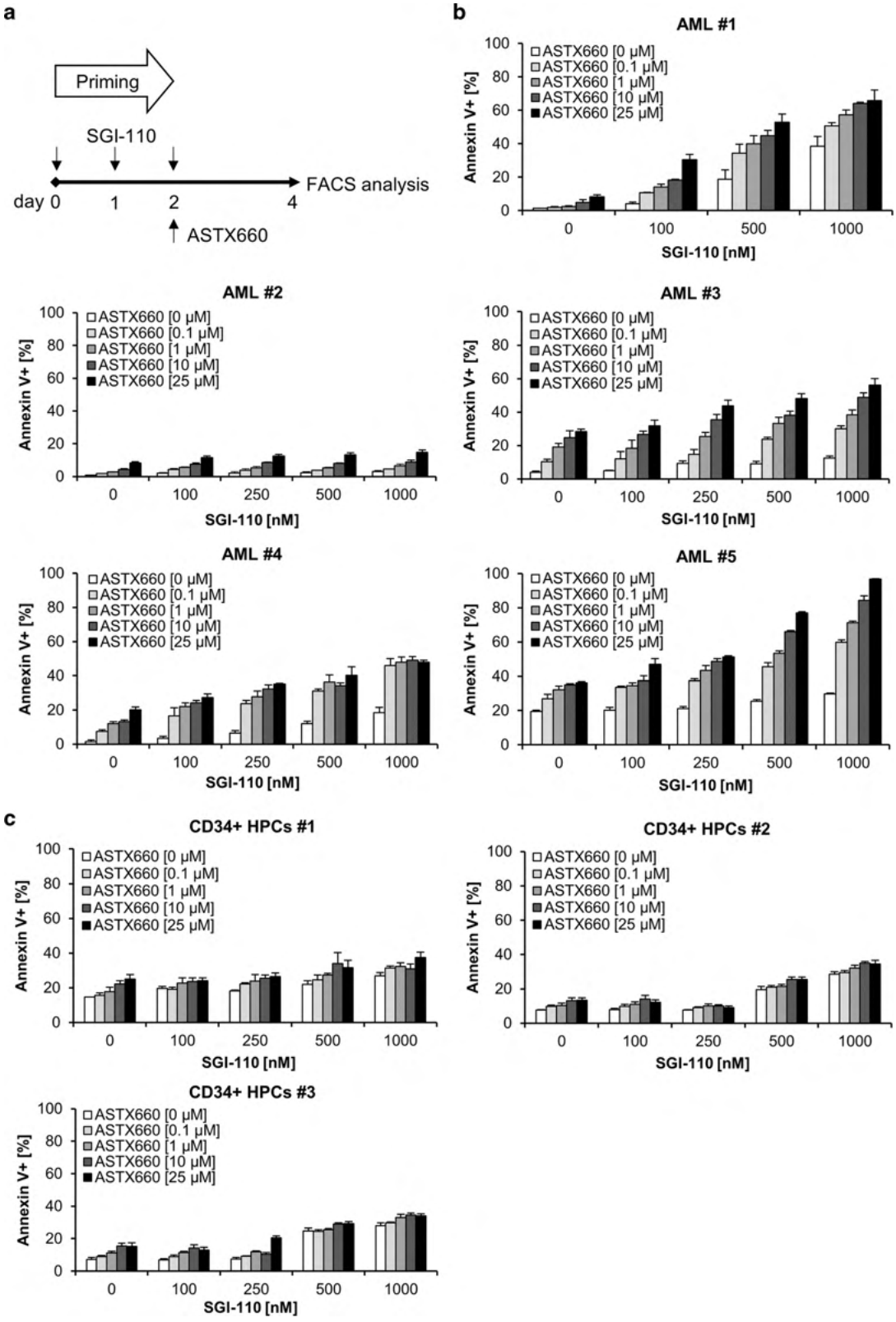


Fig. 1 SGI-110 synergizes with ASTX660 to induce cell death and to suppress clonogenic growth of AML cell lines. a Following pretreatment with indicated concentrations of SGI-110 for 24 h, AML cells were treated with indicated concentrations of ASTX660 for another 48 h. Cell death was determined by FSC/SSC analysis and flow cytometry. **b** After 24 h of pretreatment with SGI-110 (MOLM-13: 50 nM, ML-2: 2 μM), AML cells were treated with ASTX660 (MOLM-13: 40 μM, ML-2: 5 μM) for indicated time points. Cell death was determined by FSC/SSC analysis and flow cytometry. Significances after combination treatment are calculated versus single-treated cells. **c** AML cells were exposed to SGI-110 (MOLM-13: 50 nM, ML-2: 0.5 μM) for 24 h followed by ASTX660 (MOLM-13: 40 μM, ML-2: 5 μM) for 15 h before incubation in methylcellulose. Colony formation was assessed after 10–13 days and the number of colonies is expressed as percentage of solvent-treated controls. Mean and SD of three independent experiments performed in triplicate are shown. $*p < 0.05$, $**p < 0.01$; $***p < 0.001$.

Results

SGI-110 synergizes with ASTX660 to induce cell death and to suppress clonogenic growth of AML cell lines

Searching for new clinically relevant drug combinations for the treatment of AML, we asked the question whether the next-generation HMA SGI-110 can prime AML cells towards cell death induced by the novel IAP antagonist ASTX660. Interestingly, dose–response experiments revealed that SGI-110 pretreatment synergistically interacted with ASTX660 to induce cell death in five out of six AML cell lines with various genetic backgrounds and representing different AML subtypes (Fig. 1a, Supplementary Table 2–3). Only OCI-AML3 cells that harbor the hotspot *DNMT3A*^{R882C} mutation did not



respond even to high concentrations of ASTX660 and SGI-110. Synergistic drug interaction was evidenced by calculation of combination index (CI) (Supplementary Table S4). MOLM-13 and ML-2 cells, representing the most common AML subtypes, were selected for further experiments. To investigate whether the timing of drug treatment schedules

◀ **Fig. 2 SGI-110 cooperates with ASTX660 to induce cell death in a large proportion of primary AML blasts while showing minimal toxicity against normal CD34+ HPCs. a** Treatment schedule for primary AML blast cells and CD34+ HPCs. Primary cells were primed with different concentrations of SGI-110 (0.1–1 μM) three times at 24 h intervals, followed by different concentrations of ASTX660 (0.1–25 μM) on day 2. Cell death was determined by flow cytometric analysis on day 4. Stimulation of the primary AML samples (**b**) and CD34+ HPCs (**c**) with SGI-110 and/or ASTX660 was performed according to our treatment schedule (**a**). **b** Combined Annexin V-FITC/CD45-APC staining and flow cytometry were used to determine cell death of primary AML blasts identified by CD45/SSC gating procedure. **c** Apoptotic cell death in normal CD34+ HPCs was determined by Annexin V-FITC/PI staining and flow cytometry. **b**, **c** Mean and SD of single experiments performed in triplicate are shown.

influences drug interactions, drug combinations were hence further tested using simultaneous application. This treatment schedule induced up to 20% less cell death in MOLM-13 and ML-2 cells than the sequential treatment with SGI-110 and ASTX660 (Supplementary Fig. S1, Supplementary Table S3, S5), demonstrating that priming by SGI-110 increases the efficacy of ASTX660 treatment.

Time-dependent analysis of drug-induced cell death demonstrated that sequential treatment with SGI-110 and ASTX660 increased cell death over time (Fig. 1b). Besides these short-term assays, we tested the drug combination on long-term clonogenic survival by performing methylcellulose colony forming assays. Importantly, SGI-110 and ASTX660 cooperated to significantly reduce colony formation compared with single-agent treatments (Fig. 1c; Supplementary Fig. S2). In order to further investigate the effect of sequential SGI-110/ASTX660 treatment on cell proliferation, we analyzed cell growth by viable cell counting as well as immunofluorescence staining of the proliferation marker Ki-67 [36]. In line with the results of the colony formation assay, SGI-110 alone, or even more pronounced in combination with ASTX660, dramatically reduced proliferation of ML-2 cells, whereas MOLM-13 cells were only slightly affected (Supplementary Fig. S3A–B). Taken together, SGI-110 synergistically acted in concert with ASTX660 to induce cell death and to suppress proliferation and long-term clonogenic growth.

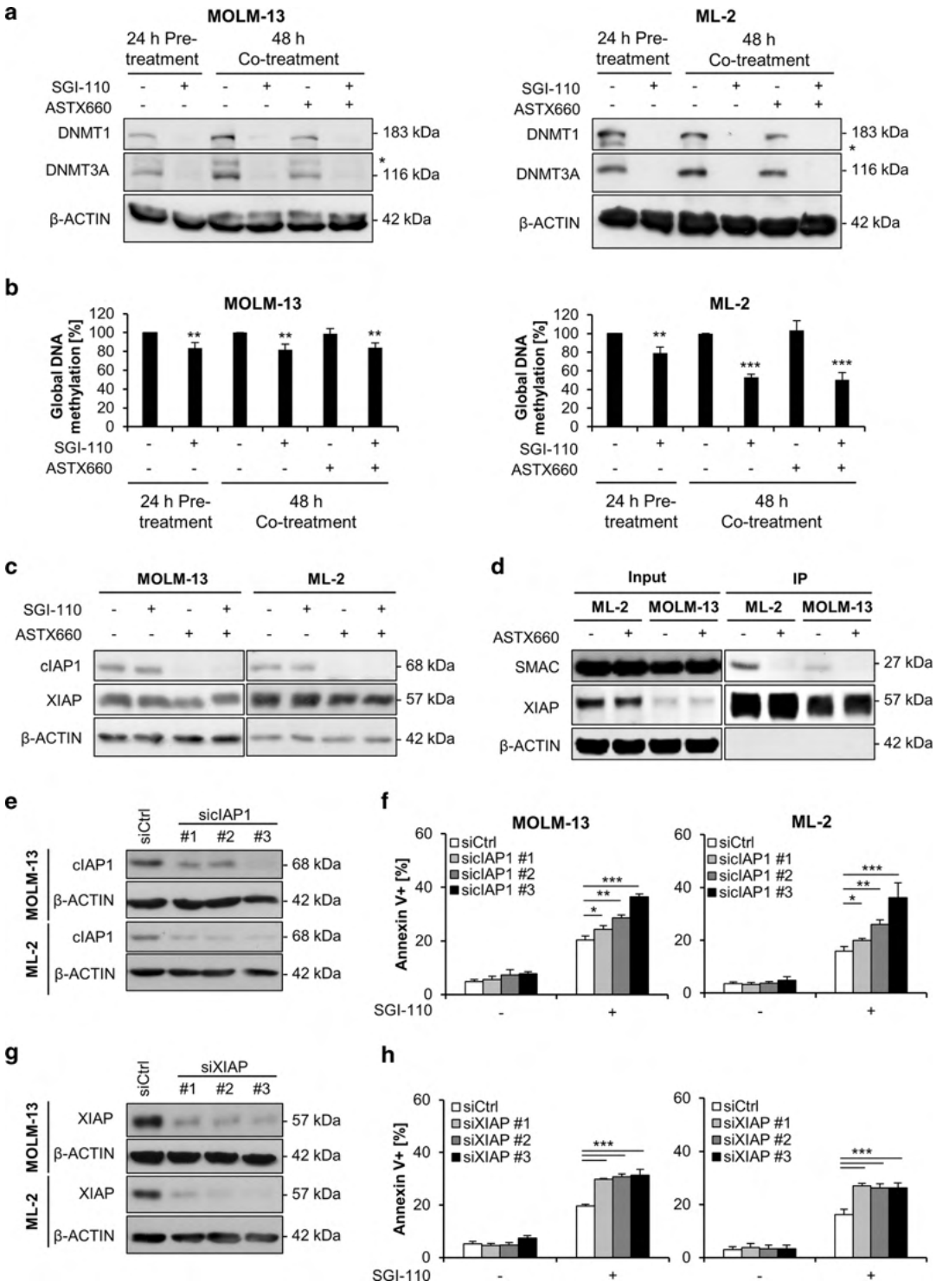
SGI-110/ASTX660 regimen is synergistically active against primary AML blasts, whereas it shows minimal toxicity against CD34+ HPCs

To assess the translational relevance of our findings we extended our studies to freshly isolated leukemic blasts from treatment-naive AML patients, which were pretreated with SGI-110 for 3 consecutive days, whereby ASTX660 was added on day 2 (Fig. 2a). SGI-110 concentrations were chosen according to plasma levels reported in patients [37]. Importantly, combined exposure to SGI-110 and ASTX660 substantially increased cell death in four out of five AML samples, whereas single-agent treatment with either SGI-110 or ASTX660 had minimal to moderate effects on cell death (Fig. 2b). Calculation of CI values revealed that sequential treatment with SGI-110 and ASTX660 synergistically induced cell death in most primary AML samples (Supplementary Table S6).

In contrast to primary AML blasts, the identical treatment failed to synergize in the induction of cell death in normal CD34+ HPCs from three healthy donors (Fig. 2c), whereas standard and high doses of Cytarabine, based on therapeutically achievable plasma concentrations during standard and high-dose therapy [38, 39], were highly toxic to normal CD34+ HPCs (Supplementary Fig. S4). Also, the sequential SGI-110/ASTX660 treatment did not affect long-term clonogenic survival of CD34+ HPC (Supplementary Fig. S5A), while Cytarabine inhibited colony formation in a dose-dependent manner (Supplementary Fig. S5B). Taken together, these data suggest that sequential SGI-110/ASTX660 treatment exerts some tumor selectivity against AML blasts with minimal effects on normal CD34+ HPCs pointing to a potential therapeutic window.

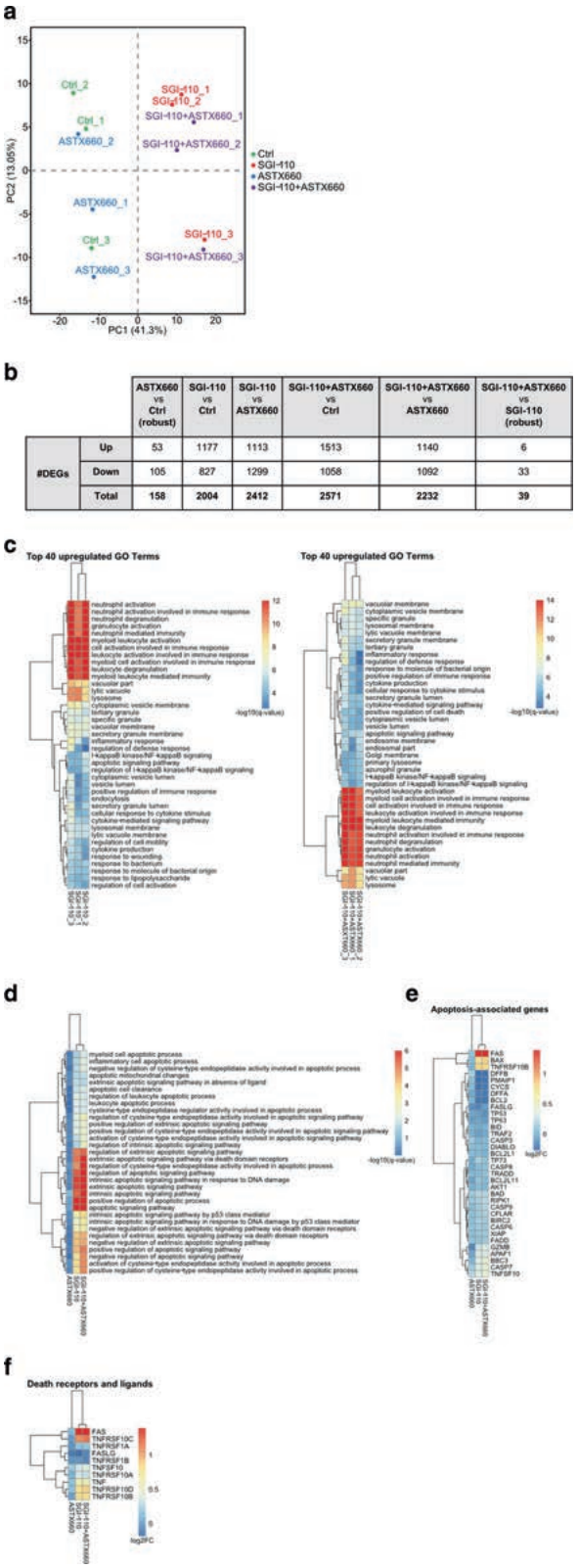
SGI-110 and ASTX660 demonstrate on-target activity in AML cell lines

Next, we monitored on-target activity of SGI-110 and ASTX660. DNMT1 and DNMT3A protein levels were depleted within 24 h of SGI-110 pretreatment and did not recover until 48 h of sequential treatment with ASTX660, confirming that SGI-110 directly targets DNMTs for protein degradation (Fig. 3a). As depletion of DNMT activity results in global hypomethylation of the genome, we further investigated changes in global DNA methylation. Indeed, SGI-110 significantly caused hypomethylation of LINE-1, which serves as a surrogate marker for global DNA methylation [40], within 24 h of SGI-110 pretreatment (Fig. 3b). As IAP antagonists have been reported to stimulate autoubiquitination and proteasomal degradation of IAP proteins [21], we examined the effect of ASTX660 on IAP protein levels. As both model cell lines do not express cIAP2 (Supplementary Fig. S6), we focused on the antagonism of cIAP1 and XIAP by ASTX660 treatment. ASTX660 induced degradation of cIAP1 in the absence and presence of SGI-110, while XIAP protein levels remained relatively unaffected, as expected (Fig. 3c) [22]. To further confirm the direct antagonism of XIAP by ASTX660 we also assessed the displacement of SMAC from XIAP by



immunoprecipitation using anti-XIAP antibody. Following ASTX660 treatment for 4 h, a complete disappearance of SMAC previously immunoprecipitated with XIAP was observed in MOLM-13 and ML-2 cells (Fig. 3d). These results confirm on-target activity of SGI-110 as well as of ASTX660 in AML cell lines.

Since ASTX660 can antagonize both cIAP1 and XIAP [22], we individually silenced cIAP1 and XIAP to define their contribution to the observed synergistic interaction with SGI-110. The efficiency of cIAP1 and XIAP knockdown by several independent siRNAs was confirmed by Western blotting (Fig. 3e, g). Of note, silencing of cIAP1



		ASTX660 vs Ctrl (robust)	SGI-110 vs Ctrl	SGI-110 vs ASTX660	SGI-110+ASTX660 vs Ctrl	SGI-110+ASTX660 vs ASTX660	SGI-110+ASTX660 vs SGI-110 (robust)
#DEGs	Up	53	1177	1113	1513	1140	6
	Down	105	827	1299	1058	1092	33
	Total	158	2004	2412	2571	2232	39

◀ **Fig. 3 SGI-110 and ASTX660 demonstrate on-target activity in AML cell lines. a**, **b** Following 24 h of pretreatment with SGI-110 (MOLM-13: 50 nM, ML-2: 2 µM), AML cells were treated with ASTX660 (MOLM-13: 40 µM, ML-2: 5 µM) for 48 h. **a** Protein expression of DNMT1 and DNMT3A was determined at indicated time points by Western blotting. β-ACTIN was used as loading control. Asterisks (*) mark unspecific bands. **b** Global DNA methylation was determined by colorimetric quantification of methylation levels of LINE-1, a widely used surrogate marker of global DNA methylation, and is shown as percentage of solvent-treated control cells. **c** Following 24 h of pretreatment with SGI-110 (MOLM-13: 50 nM, ML-2: 2 µM), AML cells were treated with ASTX660 (MOLM-13: 40 µM, ML-2: 5 µM) for 24 h. Protein levels of cIAP1 and XIAP were assessed by Western blotting. β-ACTIN served as loading control. **d** AML cells were treated for 4 h with ASTX660 (MOLM-13: 40 µM, ML-2: 5 µM). Cell lysates were subjected to immunoprecipitation using anti-XIAP antibody. Input lysates and precipitates were analyzed for XIAP and SMAC expression by Western blotting. β-Actin served as loading control. AML cells were transfected with siRNA against cIAP1 (**e**, **f**), XIAP (**g**, **h**) or non-targeting siRNA (siCtrl). Expression levels of cIAP1 (**e**) and XIAP (**g**) were assessed by Western blotting, with β-ACTIN serving as loading control. **f**, **h** To analyze the effect of siRNA-mediated loss of cIAP1 and XIAP on cell death, transfected cells were immediately treated with SGI-110 (MOLM-13: 50 nM, ML-2: 2 µM) for 72 h and apoptotic cell death was determined by Annexin V-FITC/PI staining and flow cytometry. (**b**, **f**, **h**) Mean and SD of three independent experiments carried out in triplicate are shown. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

(Fig. 3f) or XIAP (Fig. 3h) significantly increased SGI-110-induced cell death. These results indicate that targeting of both, cIAP1 and XIAP, contributes to enhance SGI-110-mediated cell death in AML cells.

Sequential SGI-110/ASTX660 treatment upregulates extrinsic and intrinsic apoptosis signaling genes

To gain further insights into the molecular mechanisms of the synergistic activity of SGI-110 and ASTX660 we performed transcriptome analysis prior to the onset of cell death. A principal component analysis was used to obtain an overview of treatment-induced changes in transcription (Fig. 4a). The first principal component (PC1) clearly separated the control and ASTX660-treated cells from those treated with SGI-110 alone or in combination with ASTX660. Analysis of DEGs revealed 2004 and 2571 DEGs upon SGI-110 alone or sequential SGI-110/ASTX660 treatment, respectively, whereas only 158 DEGs were detectable following ASTX660 treatment compared with control (Fig. 4b). These results show that SGI-110 was primarily responsible for the global transcriptional changes following sequential SGI-110/ASTX660 treatment. We further performed GO enrichment analysis to investigate the biological functions of the identified DEGs (Fig. 4c and Supplementary Fig. S7). Interestingly, in both SGI-110 alone and sequential SGI-110/ASTX660 treatment, upregulated DEGs were significantly enriched in the apoptotic signaling pathway and IκB kinase/Nuclear factor-kappa B signaling. Focusing our analysis on significantly upregulated apoptotic GO terms, we found several GO terms associated with apoptosis, including the extrinsic apoptotic

◀ **Fig. 4 Sequential SGI-110/ASTX660 treatment upregulates extrinsic and intrinsic apoptosis signaling genes.** ML-2 cells were treated with solvent alone or with 2 µM SGI-110 for 24 h followed by 5 µM ASTX660 for 9 h. RNA expression levels of three independent experiments were analyzed by microarray. **a** Principal component analysis (PCA) of ML-2 cells after drug treatment with SGI-110 and/or ASTX660 or solvent control (Ctrl). **b** Number of total, up-, and downregulated DEGs from multiple comparisons among the four treatment groups (adjusted p value < 0.05). For two comparisons, labeled with robust, the robust method for linear model fitting was used. The rest of the comparisons was fitted with the least-squares method. **c** The Top 40 enriched GO terms (biological process, cellular component, and molecular function, combined to one set for the analysis) of upregulated genes upon SGI-110 alone (left panel) or sequential SGI-110/ASTX660 treatment (right panel) compared with control treatment. **d** Heatmap showing significantly upregulated apoptotic GO terms in at least one condition after SGI-110 and/or ASTX660 treatment, normalized to solvent-treated control cells. Pathways were considered significant with an adjusted p value < 0.05. **e** Heatmap showing apoptosis-related genes regulated in SGI-110 and/or ASTX660-treated ML-2 cells, normalized to solvent-treated cells. Genes were selected based on the "Apoptosis" pathway of ConsensusPathDB and the "Apoptosis" pathway of Reactome. Only the genes annotated in both pathways were considered. **f** Heatmap displaying log2 fold expression changes of death receptors and their ligands upon SGI-110 and/or ASTX660 treatment compared with solvent-treated control cells. (**c**–**f**) Columns and rows were hierarchically clustered by their Euclidian distance using complete linkage. The colors correspond to the −log10 transformed adjusted p values from the gene-set enrichment analysis (**c**, **d**) or log2 fold changes (FC) (**e**, **f**) obtained by comparing SGI-110 and/or ASTX660-treated versus solvent-treated control cells.

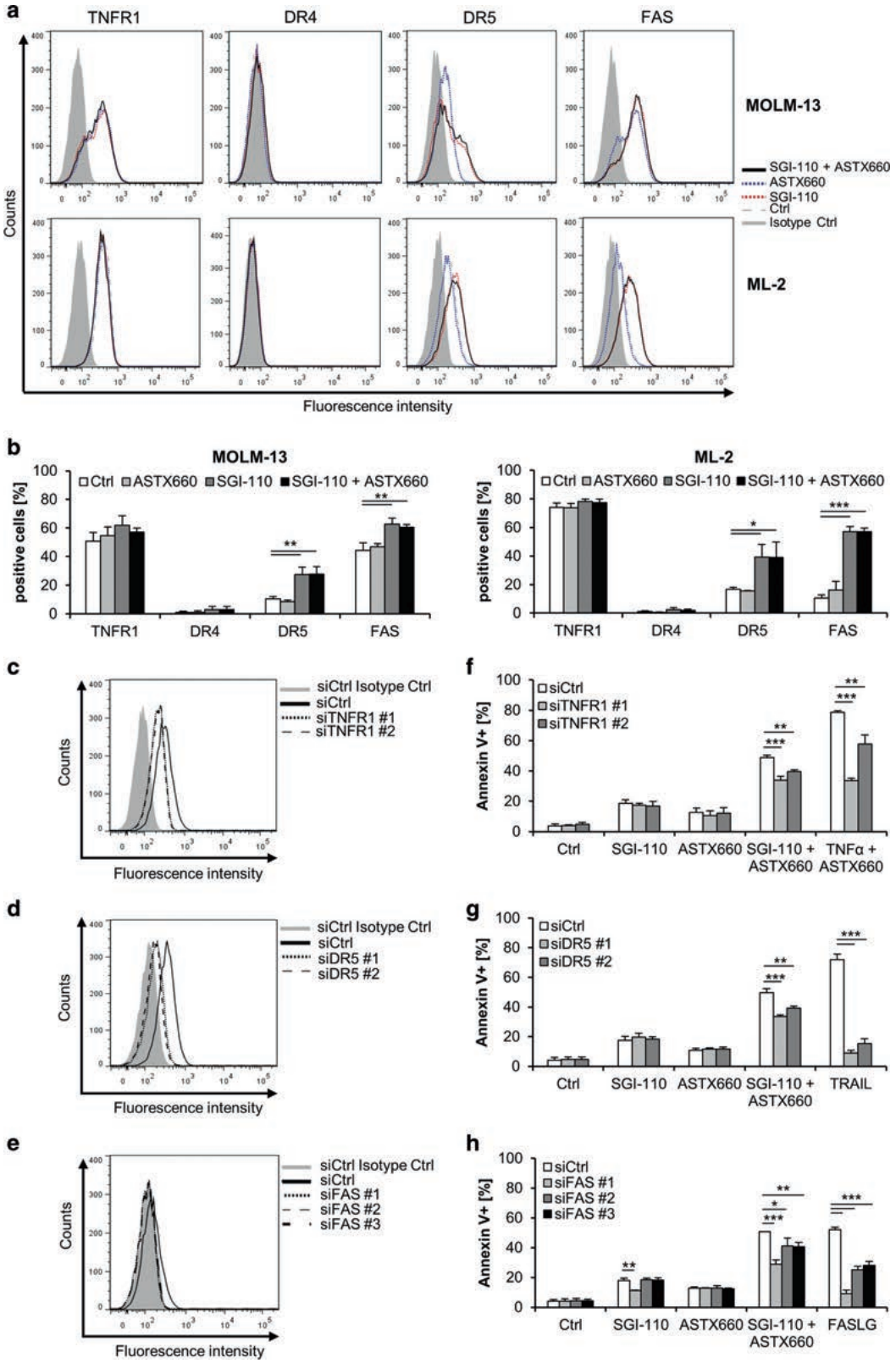
signaling pathway via death receptors, the intrinsic apoptotic signaling pathway and regulation of cysteine-type endopeptidase activity involved in apoptotic processes (Fig. 4d). Thus, we concentrated on apoptosis-related genes which were selected based on the "Apoptosis" pathway of ConsensusPathDB and Reactome (Fig. 4e). Of note, SGI-110 alone or in combination with ASTX660 significantly increased gene expression of *FAS*, *BAX*, and *TNFRSF10B* (DR5). Since this analysis indicated an involvement of death receptor-mediated apoptosis, we analyzed in more detail the transcriptional changes in death receptors and their ligands (Fig. 4f). Interestingly, *FAS*, several TRAIL receptors (i.e., *TNFRSF10B* (DR5), *TNFRSF10C* (DcR1), and *TNFRSF10D* (DcR2)) as well as *TNF* (TNFα) were significantly upregulated (adjusted p value < 0.05 and log2 fold change > 0.5) upon SGI-110 alone or sequential SGI-110/ASTX660 treatment. With regard to the additionally observed growth-inhibitory effects of SGI-110 alone or in combination with ASTX660, we extended our GSEA on proliferation-related GO terms. Thereby, we identified the cell cycle regulator *CDKN1A* (p21) as the strongest upregulated gene in the most significantly enriched proliferation GO term "negative regulation of proliferation" (Supplementary Fig. S8A–B). The upregulation of *CDKN1A* by SGI-110 alone or when combined with ASTX660 was confirmed by qRT-PCR (Supplementary Fig. S8C). Taken together, our transcriptome analysis showed that SGI-110 alone or in combination with ASTX660 upregulated signaling pathways in extrinsic and intrinsic apoptosis and regulated genes involved in proliferative signaling.

Sequential SGI-110/ASTX660 treatment initiates death receptor-dependent apoptosis

As our transcriptome analysis identified death receptors and their ligands as top hits following SGI-110 alone or sequential SGI-110/ASTX660 treatment, we next determined the functional relevance of the death receptor pathway for SGI-110/ASTX660-induced apoptosis. Since DcR1 and DcR2 are unable to transmit death signals [41], we focused our experiments on TNFR1, DR4, DR5, FAS, and their respective ligands. Validation experiments using qRT-PCR confirmed differential expression of selected death receptors and ligands in both AML cell lines (Supplementary Fig. S9). To investigate whether these drug-induced changes in gene expression resulted in altered protein expression of death receptors we analyzed cell surface expression by flow cytometry. Indeed, SGI-110 alone or sequential treatment with SGI-110 and ASTX660 significantly upregulated the surface expression of DR5 and FAS, while TNFR1 expression remained nearly unchanged (Fig. 5a, b). By comparison, DR4 was hardly detectable on the surface of MOLM-13 and ML-2 cells.

To test the functional relevance of TNFR1, FAS, and DR5 for SGI-110/ASTX660-induced apoptosis we genetically silenced these receptors by siRNA and confirmed knockdown efficiency by flow cytometry (Fig. 5c–e; Supplementary Fig. S10A–C). Importantly, knockdown of either TNFR1 or FAS significantly rescued SGI-110/ASTX660-mediated apoptosis in both AML cell lines. By comparison, DR5 knockdown significantly protected only ML-2 cells from SGI-110/ASTX660-induced apoptosis (Fig. 5f–h; Supplementary Fig. S10D–F). This set of experiments indicates that SGI-110/ASTX660-induced cell death depended, at least partly, on death receptors such as TNFR1 and FAS.

Next, we investigated whether the observed upregulation of *TNFα* (Fig. 4f) was required for SGI-110/ASTX660-mediated apoptosis as described previously for similar settings [21, 42–44]. To this end, we used the TNFα-blocking antibody Enbrel. While Enbrel slightly reduced SGI-110/ASTX660-induced apoptosis in ML-2 cells, it had no effect in MOLM-13 cells (Supplementary Fig. S11A). Control experiments confirmed that Enbrel significantly decreased TNFα/ASTX660-mediated apoptosis. In addition, TRAIL- or FASLG-blocking antibodies failed to rescue SGI-110/



ASTX660-induced apoptosis, whereas they potently suppressed TRAIL or FASLG-mediated apoptosis used as positive controls (Supplementary Fig. S11B–C). This indicates that SGI-110/ASTX660 induced cell death in AML cells in a TRAIL- and FASLG-independent fashion, while the effect of TNFα seems to be cell line-dependent.

◀ **Fig. 5 Sequential SGI-110/ASTX660 treatment initiates death receptor-dependent apoptosis. a–b** Following 24 h of pretreatment with SGI-110 (MOLM-13: 50 nM, ML-2: 2 µM), AML cells were treated with ASTX660 (MOLM-13: 40 µM, ML-2: 5 µM) for 15 h. Death receptor expression on cell surface was determined by flow cytometric analysis after staining with PE-conjugated antibodies specific to each death receptor (open histograms) or with isotype-matched IgG controls (shaded gray histogram). Representative overlay histograms (**a**) and quantification of cell surface expression of death receptors (**b**) from three independent experiments performed in triplicate are shown. (**c–h**) ML-2 cells were transfected with non-targeting control siRNA (siCtrl) or siRNA against TNFR1, DR5, or FAS. (**c–e**) Cell surface expression of death receptors (open histograms) was analyzed by flow cytometry. Gray shaded histograms represent respective isotype controls. Representative overlay histograms are shown. (**f–h**) Following 24 h of pretreatment with 2 µM SGI-110, ML-2 cells were treated with 5 µM ASTX660 for 36 h. Treatments with 1 ng/ml TNFα and 5 µM ASTX660 (**f**), 10 ng/ml TRAIL (**g**) or 750 ng/ml hexameric FAS ligand (FASLG) (**h**) for 36 h were used as positive controls to demonstrate the efficacy of gene silencing. Cell death was determined by Annexin V-FITC staining and flow cytometry. In (**b**, **f–h**), mean and SD of three independent experiments performed in triplicate are shown. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Sequential SGI-110/ASTX660 treatment engages intrinsic apoptosis as well as a crosstalk between extrinsic and intrinsic apoptotic pathways

As our transcriptome analysis highlighted the engagement of the intrinsic apoptosis pathway beyond the extrinsic pathway, we examined whether a crosstalk between these pathways may exist during SGI-110/ASTX660-induced apoptosis. One of the best-characterized connections between the two pathways is BID, which translocates to mitochondria after caspase-8-mediated cleavage [45]. Indeed, sequential SGI-110/ASTX660 treatment-induced cleavage of caspase-8 and BID (Fig. 6a), suggesting a crosstalk between extrinsic and intrinsic apoptosis pathways.

tBID has been reported to facilitate apoptosis by engaging lysosomal permeabilization [46] in addition to promoting activation of BAX/BAK and MOMP [47]. Since our microarray analysis indicated a potential role of lysosomes upon SGI-110/ASTX660 treatment (Fig. 4c, e), we examined the effect of lysosomal protease inhibitors on SGI-110/ASTX660-induced apoptosis. However, the addition of E64D/Pepstatin A failed to rescue AML cells from SGI-110/ASTX660-induced apoptosis, while control experiments confirmed the functionality of these inhibitors (Supplementary Fig. S12).

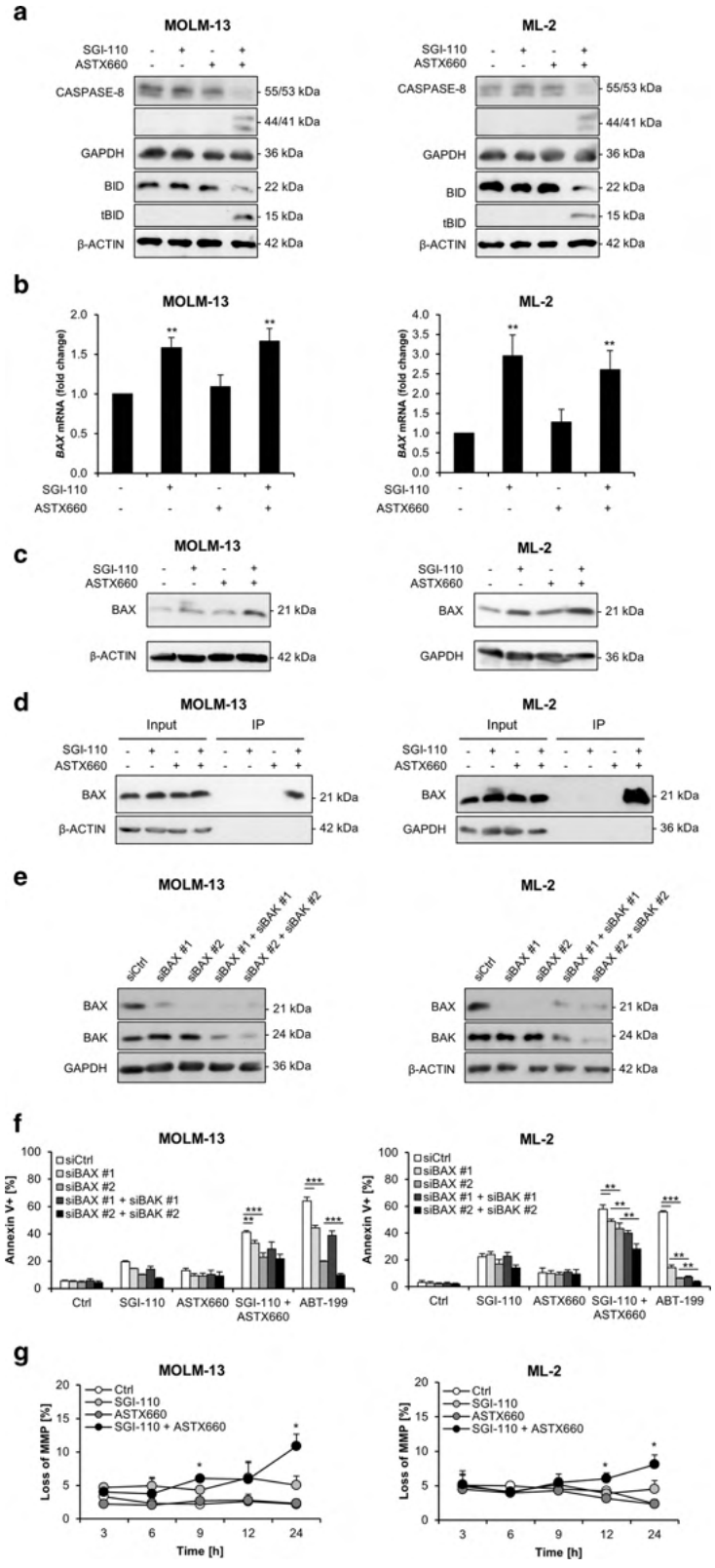
As our microarray analysis revealed increased BAX levels by sequential SGI-110/ASTX660 treatment, we then investigated the role of BAX in SGI-110/ASTX660-induced apoptosis. SGI-110/ASTX660-stimulated upregulation of BAX was confirmed by qRT-PCR (Fig. 6b) and Western blotting (Fig. 6c). To investigate whether BAX is activated upon sequential SGI-110/ASTX660 treatment we performed immunoprecipitation experiments with a conformation-specific antibody for activated BAX. Intriguingly, SGI-110 and ASTX660 acted in concert to activate BAX compared with either drug alone (Fig. 6d). To determine the functional relevance of BAX for SGI-110/ASTX660-induced apoptosis, we performed BAX knockdown using siRNA (Fig. 6e). Notably, silencing of BAX significantly reduced SGI-110/ASTX660-mediated apoptosis (Fig. 6f). Furthermore, concomitant knockdown of BAX and BAK was significantly more effective to inhibit SGI-110/ASTX660-mediated apoptosis in ML-2 cells compared with BAX knockdown alone (Fig. 6f). ABT-199 treatment, reported to kill AML cells in a BAX/BAK-dependent fashion [48], was used as a positive control (Fig. 6f).

As activated BAX can cause MOMP leading to cytochrome c release into the cytosol [49], we next assessed MMP. SGI-110 and ASTX660 cooperated to increase loss of MMP in a time-dependent manner (Fig. 6g). Together, these data suggest that sequential SGI-110/ASTX660 treatment engages intrinsic apoptosis as well as a crosstalk between extrinsic and intrinsic apoptotic pathways.

SGI-110 and ASTX660 cooperate to induce biochemical and morphological hallmarks of apoptosis

To examine effector mechanisms of SGI-110/ASTX660-induced cell death we determined activation of the executioner caspases-3 and -7 by fluorescence microscopy. Interestingly, sequential treatment with SGI-110 and ASTX660 significantly increased activation of caspases-3/-7 compared with either treatment alone in a time-dependent manner (Fig. 7a). Furthermore, SGI-110 acted together with ASTX660 to significantly increase DNA fragmentation (Fig. 7b). Consistently, the analysis of nuclear morphology revealed that AML cells underwent morphological changes typical of apoptosis such as chromatin condensation and nuclear fragmentation (Fig. 7c). Quantification revealed a significant increase in nuclei fragmentation upon combination treatment compared with either single treatment or to untreated cells (Fig. 7d). Together, this set of experiments demonstrates that SGI-110 and ASTX660 cooperated to trigger typical apoptotic events such as caspase activation and nuclear fragmentation.

However, addition of the pan-caspase inhibitor zVAD.fmk alone failed to protect from SGI-110/ASTX660-induced apoptosis and even increased cell death, in line with reports showing that caspase inhibition can cause a switch from apoptotic to necroptotic cell death in AML cells [23, 42]. In line with this notion, simultaneous treatment with zVAD.fmk and the necroptosis inhibitors Necrostatin-1s (Nec-1s), Necrosulfonamide (NSA), GSK'872, or Dabrafenib significantly reduced SGI-110/ASTX660-induced



cell death compared with SGI-110/ASTX660-treated cells in the presence of zVAD.fmk alone (Supplementary Fig. S13). This indicates that SGI-110/ASTX660 primarily induced caspase-dependent apoptosis in apoptosis-proficient AML cells, while it engaged necroptosis upon caspase inhibition.

◀ **Fig. 6 Sequential SGI-110/ASTX660 treatment engages intrinsic apoptosis as well as a crosstalk between extrinsic and intrinsic apoptotic pathways.** **a** AML cells were pretreated with SGI-110 (MOLM-13: 50 nM, ML-2: 2 µM) for 24 h followed by ASTX660 (MOLM-13: 40 µM, ML-2: 5 µM) for 15 h. **a** Cleavage of caspase-8 and BID was assessed by Western blotting, GAPDH, and β-ACTIN served as loading controls. **b**, **c** AML cells were pretreated with SGI-110 (MOLM-13: 50 nM, ML-2: 2 µM) for 24 h followed by ASTX660 (MOLM-13: 40 µM, ML-2: 5 µM) for 9 h (**b**) or 15 h (**c**). **b** *BAX* mRNA expression was analyzed by qRT-PCR and fold changes relative to untreated control are shown with mean and SD of three independent experiments performed in triplicate. Significances are calculated versus control cells. **c** BAX expression was determined by Western blotting, GAPDH, and β-ACTIN served as loading controls. **d** After 24 h of pretreatment with SGI-110 (MOLM-13: 50 nM, ML-2: 2 µM), AML cells were treated with ASTX660 (MOLM-13: 40 µM, ML-2: 5 µM) for 15 h (MOLM-13) or 24 h (ML-2). Activation of BAX was assessed by immunoprecipitation using active conformation-specific antibodies and protein expression of BAX was detected by Western blotting. GAPDH and β-ACTIN served as loading controls. **e**, **f** AML cells were transiently transfected with siRNA against non-targeting siRNA (siCtrl) or BAX, BAK, or both. **e** Expression of BAX and BAK were assessed by Western blotting, with GAPDH serving as loading control. **f** Transfected cells were pretreated with SGI-110 (MOLM-13: 50 nM, ML-2: 2 µM) for 24 h, followed by addition of ASTX660 (MOLM-13: 40 µM, ML-2: 5 µM) for 24 h (MOLM-13) or 36 h (ML-2). Treatment with 1 µM ABT-199 for 24 h (MOLM-13) or 36 h (ML-2) was used as positive control to demonstrate the efficacy of gene silencing. Apoptotic cell death was measured by Annexin V-FITC/PI staining and flow cytometry. **g** Following 24 h of pretreatment with SGI-110 (MOLM-13: 50 M, ML-2: 2 µM), AML cells were treated with ASTX660 (MOLM-13: 40 µM, ML-2: 5 µM) for indicated time points. Loss of MMP was analyzed by flow cytometry using the fluorescent dye TMRM. Significances after sequential SGI-110/ASTX660 treatment are calculated versus single-treated cells. In **b**, **f**, **g**, mean and SD of three independent experiments carried out in triplicate are shown. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

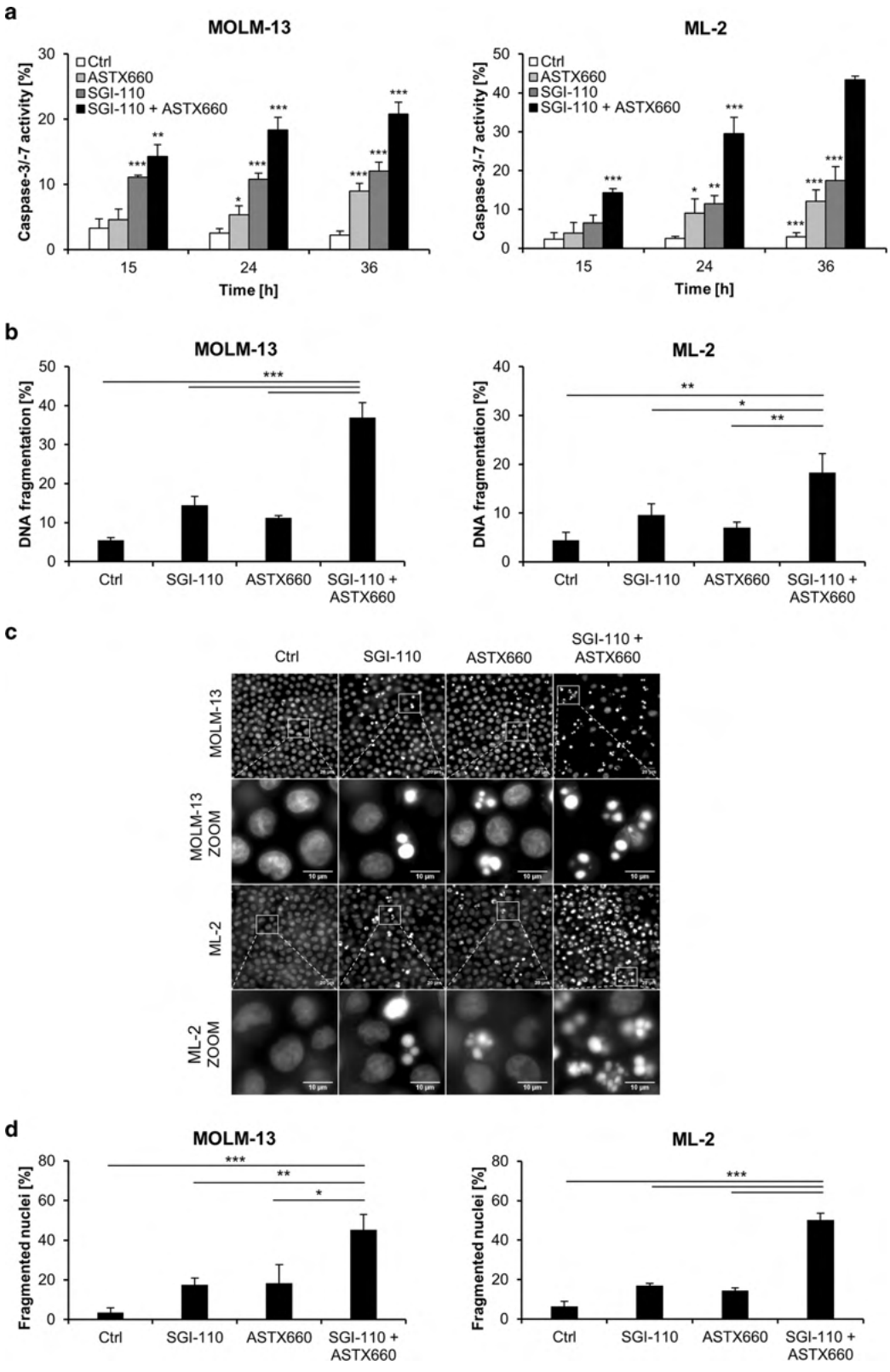
Discussion

Given the limited treatment options for elderly patients, the development of more effective and well-tolerated therapies remains a major unmet medical need [50]. Approval of decitabine for the first-line treatment of elderly AML patients highlights the general potential of HMAs in this setting [51]. However, as HMAs demonstrated only limited efficacy as monotherapy, combination strategies with either chemotherapy, immunotherapy, or targeted therapies such as the BCL-2 inhibitor ABT-199 are currently in different stages of clinical testing [13]. As evasion of apoptosis, e.g., by overexpression of IAPs, contributes to treatment resistance and poor outcome in AML [52], in the present study we investigated the combination of SGI-110 and the IAP antagonist ASTX660. Here, we discover that the combination of SGI-110 and ASTX660 synergistically induced cell death in a range of AML cell lines as well as in patient-derived AML blasts. By comparison, sequential treatment with SGI-110 and ASTX660 exhibited limited toxicity to normal CD34+ HPCs, pointing to a favorable therapeutic window. Concentrations of SGI-110 (0.1–1 µM) required for synergism with ASTX660 in our study might be achievable in AML patients, as clinically achievable plasma levels have been reported to range from 20 to 400 nM for SGI-110 and from 0.3 to 1.6 µM for parental compound decitabine [37, 53].

While the *DNMT3A* mutant AML cell line OCI-AML3 proved to be resistant to sequential SGI-110/ASTX660 treatment, SGI-110 together with ASTX660 synergistically induced cell death in the tested primary AML sample harboring *DNMT3A* mutation. In line with this finding, Metzeler et al. reported that *DNMT3A*-mutated patients benefit from HMA treatment [54]. A possible explanation might be that $DNMT3A^{R882}$-associated hypomethylation patterns are preserved in primary AML samples, but not in the $DNMT3A^{R882C}$-mutated AML cell line OCI-AML3 [55]. Further studies are required to unravel the mechanisms for SGI-110/ASTX660 resistance in OCI-AML3 cells.

Importantly, our study provides new insights into the molecular mechanisms underlying the synergistic induction of cell death by SGI-110 and ASTX660. We identify activation of both extrinsic and intrinsic apoptotic pathways as key events during SGI-110/ASTX660-induced cell death in AML cells. This conclusion is supported by several lines of evidence.

First, unbiased transcriptome analysis revealed that SGI-110 alone and sequential SGI-110/ASTX660 treatment upregulated the death receptors *DR5* and *FAS*, which was accompanied by increased protein expression on the cell surface. *FAS* upregulation might be a direct effect of HMA-induced *FAS* promoter demethylation [56]. By comparison, DR4 was hardly detectable on the cell surface of both AML cell lines, which is in line with previous studies demonstrating the absence or low expression of DR4 for different cancer entities [57, 58]. Second, TNFR1, DR5, and FAS were all required for cell death induction, since individual silencing of these death receptors significantly reduced SGI-110/ASTX660-induced apoptosis. The fact that SGI-110/ASTX660-mediated cell death did not depend on a single death receptor underlines the importance of the death receptor pathway as a whole rather than single gene activation. Third, experiments with blocking antibodies revealed a contribution of TNFα-driven autocrine/paracrine signaling to SGI-110/ASTX660-mediated apoptosis in a cell line-dependent manner, which is consistent with our previous studies showing that Enbrel rescued leukemia cells from HMA/IAP antagonist-induced cell death [42, 43]. However, SGI-110/ASTX660-induced cell death occurred independently of TRAIL and FASLG, suggesting an activation of DR5 and FAS in a ligand-independent fashion. This is in line with previous evidence suggesting that anticancer drug-mediated apoptosis involved ligand-independent death receptor clustering and activation of caspases in leukemia as well as solid cancer cells [59–61].



Fourth, SGI-110/ASTX660-induced death signals are transmitted by caspase-8 and tBID to mitochondria, leading to activation of BAX and finally mitochondrial apoptosis. Fifth, SGI-110 increases BAX expression, in line with a previous study using decitabine [62], thus shifting the balance of pro- and antiapoptotic factors towards apoptosis. The crucial role of BAX for SGI-110/ASTX660-induced apoptosis is emphasized by genetic silencing of BAX,

◀ **Fig. 7 SGI-110 and ASTX660 cooperate to induce biochemical and morphological hallmarks of apoptosis. a** After 24 h of pretreatment with SGI-110 (MOLM-13: 50 nM, ML-2: 2 μM), AML cells were additionally treated with ASTX660 (MOLM-13: 40 μM, ML-2: 5 μM) for indicated time points. Caspase-3/-7 activity was determined by Cell Event Caspase-3/-7 Green Detection Reagent and ImageXpress Micro XLS system. Significances are calculated versus control cells. **b–d** Following 24 h of SGI-110 pretreatment (MOLM-13: 50 nM, ML-2: 2 μM), AML cells were treated with ASTX660 (MOLM-13: 40 μM, ML-2: 5 μM) for 48 h. **b** DNA fragmentation was analyzed by flow cytometry after PI staining of nuclei. **c** Nuclear fragmentation was imaged using Hoechst-33342 staining and ImageXpress Micro XLS system. Representative images per sample are shown. **d** Quantification of fragmented nuclei upon indicated treatment. Mean and SD of three independent experiments performed in triplicate are shown. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

which significantly protected AML cells from SGI-110/ASTX660-mediated apoptosis. Taken together, our in-depth molecular studies highlight the importance of both extrinsic and intrinsic apoptotic pathways in mediating SGI-110/ASTX660-induced apoptosis.

Since both SGI-110 and ASTX660 as single agents are currently under evaluation in early clinical trials, our findings have important implications for the development of new combination strategies for AML. SGI-110 may offer several advantages compared with first-generation HMAs, as subcutaneous administration of SGI-110 exhibited increased efficacy and reduced toxicity compared with decitabine due to reduced peak plasma levels and prolonged half-life [8]. In addition, response rates to SGI-110 were reported in AML patients after azacytidine failure [63]. Previously, only first-generation HMAs such as azacytidine and decitabine have been shown to synergistically induce cell death together with IAP antagonist in AML [42], acute lymphoblastic leukemia [43] as well as in many NCI-60 cell lines from different cancer types [64], indicating a potential broad application of this combination strategy in cancer therapy. The success of current antileukemic therapies is often limited by evasion of apoptosis [52] highlighting the potential of this combination strategy of SGI-110 and ASTX660 for the treatment of AML.

Acknowledgements We thank S. Jordan (UCT Biobank, Frankfurt) for providing primary AML samples, H. Bönig (Institute for Transfusion Medicine and Immunohematology, University Hospital Frankfurt) for providing CD34+ HPCs, Astex Pharmaceuticals (Cambridge, UK) for providing SGI-110 and ASTX660, C. Hugenberg for expert secretarial assistance and Dr M. Bewerunge-Hudler and her team (Genomics and Proteomics Core Facility, German Cancer Research Center/DKFZ, Heidelberg, Germany) for their microarray service. We thank V. Ehrenfeld for assistance with image preparation. This work was supported by grants from the BMBF (to SF) and German Cancer Aid (to SF), and by Astex Pharmaceuticals, Cambridge, UK (to SF). MB is supported by grants from the DFG (CRC850), BMBF (CoNfirm; FKZ 01ZX1708F), and MIRACUM within the Medical Informatics Funding Scheme (FKZ 01ZZ1606A-H). PM is funded by MIRACUM.

Compliance with ethical standards

Conflict of interest This work was supported in part by Astex Pharmaceuticals, Cambridge, UK (to SF). GAW is employee of Astex Pharmaceuticals. JD, TH, PM, MB, and MV declare no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. O'Donnell MR, Tallman MS, Abboud CN, Altman JK, Appelbaum FR, Arber DA, et al. Acute myeloid leukemia, version 3.2017, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2017;15:926–57.
2. Dohner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. N. Engl J Med. 2015;373:1136–52.
3. Chaudry SF, Chevassut TJ. Epigenetic guardian: a review of the DNA methyltransferase DNMT3A in acute myeloid leukaemia and clonal haematopoiesis. Biomed Res Int 2017;2017:5473197.
4. Marcucci G, Metzeler KH, Schwind S, Becker H, Maharry K, Mrozek K, et al. Age-related prognostic impact of different types of DNMT3A mutations in adults with primary cytogenetically normal acute myeloid leukemia. J Clin Oncol. 2012;30:742–50.
5. Nieto M, Demolis P, Béhanzin E, Moreau A, Hudson I, Flores B, et al. The European Medicines Agency review of decitabine (Dacogen) for the treatment of adult patients with acute myeloid leukemia: summary of the scientific assessment of the committee for medicinal products for human use. Oncologist. 2016;21:692–700.
6. Griffiths E, Choy G, Redkar S, Taverna P, Azab M, Karpf AR. SGI-110: DNA methyltransferase inhibitor oncolytic. Drugs future. 2013;38:535.
7. Roboz GJ, Kantarjian HM, Yee KWL, Kropf PL, O'Connell CL, Griffiths EA, et al. Dose, schedule, safety, and efficacy of guadecitabine in relapsed or refractory acute myeloid leukemia. Cancer. 2018;124:325–34.
8. Kantarjian HM, Roboz GJ, Kropf PL, Yee KWL, O'Connell CL, Tibes R, et al. Guadecitabine (SGI-110) in treatment-naive patients with acute myeloid leukaemia: phase 2 results from a multicentre, randomised, phase 1/2 trial. Lancet Oncol. 2017;18:1317–26.
9. Wolff F, Leisch M, Greil R, Risch A, Pleyer L. The double-edged sword of (re)expression of genes by hypomethylating agents: from viral mimicry to exploitation as priming agents for targeted immune checkpoint modulation. Cell Commun Signal. 2017;15:13.
10. Fang F, Munck J, Tang J, Taverna P, Wang Y, Miller DF, et al. The novel, small-molecule DNA methylation inhibitor SGI-110 as an ovarian cancer chemosensitizer. Clin Cancer Res. 2014;20:6504–16.
11. Kuang Y, El-Khoueiry A, Taverna P, Ljungman M, Neamati N. Guadecitabine (SGI-110) priming sensitizes hepatocellular carcinoma cells to oxaliplatin. Mol Oncol. 2015;9:1799–814.
12. Srivastava P, Paluch BE, Matsuzaki J, James SR, Collamat-Lai G, Karbach J, et al. Immunomodulatory action of SGI-110, a hypomethylating agent, in acute myeloid leukemia cells and xenografts. Leuk Res. 2014;38:1332–41.
13. Bewersdorf JP, Shallis R, Stahl M, Zeidan AM. Epigenetic therapy combinations in acute myeloid leukemia: what are the options? Ther Adv Hematol. 2019;10:2040620718816698.
14. Fulda S. Cell death in hematological tumors. Apoptosis. 2009;14:409–23.

15. Ashkenazi A. Targeting the extrinsic apoptosis pathway in cancer. Cytokine Growth Factor Rev. 2008;19:325–31.
16. Li H, Zhu H, Xu C-j, Yuan J. Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. Cell. 1998;94:491–501.
17. Fulda S, Vucic D. Targeting IAP proteins for therapeutic intervention in cancer. Nat Rev Drug Discov. 2012;11:109–24.
18. Tamm I, Richter S, Scholz F, Schmelz K, Oltersdorf D, Karawajew L, et al. XIAP expression correlates with monocytic differentiation in adult de novo AML: impact on prognosis. Hematol J. 2004;5:489–95.
19. Wuchter C, Richter S, Oltersdorf D, Karawajew L, Ludwig WD, Tamm I. Differences in the expression pattern of apoptosis-related molecules between childhood and adult de novo acute myeloid leukemia. Haematologica. 2004;89:363–4.
20. Eckelman BP, Salvesen GS, Scott FL. Human inhibitor of apoptosis proteins: why XIAP is the black sheep of the family. EMBO Rep. 2006;7:988–94.
21. Varfolomeev E, Blankenship JW, Wayson SM, Fedorova AV, Kayagaki N, Garg P, et al. IAP antagonists induce autoubiquitination of c-IAPs, NF-kappaB activation, and TNFalpha-dependent apoptosis. Cell. 2007;131:669–81.
22. Ward GA, Lewis EJ, Ahn JS, Johnson CN, Lyons JF, Martins V, et al. ASTX660, a novel non-peptidomimetic antagonist of cIAP1/2 and XIAP, potently induces TNFalpha-dependent apoptosis in cancer cell lines and inhibits tumor growth. Mol Cancer Ther. 2018;17:1381–91.
23. Safferthal C, Rohde K, Fulda S. Therapeutic targeting of necroptosis by Smac mimetic bypasses apoptosis resistance in acute myeloid leukemia cells. Oncogene. 2017;36:1487–502.
24. Lacombe F, Durrieu F, Briais A, Dumain P, Belloc F, Bascans E, et al. Flow cytometry CD45 gating for immunophenotyping of acute myeloid leukemia. Leukemia 1997;11:1878–86.
25. Nicoletti I, Migliorati G, Pagliacci MC, Grignani F, Riccardi C. A rapid and simple method for measuring thymocyte apoptosis by propidium iodide staining and flow cytometry. J Immunol Methods. 1991;139:271–9.
26. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.
27. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods. 2015;12:115–21.
28. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy-analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004;20:307–15.
29. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics. 2005;21:2067–75.
30. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinf. 2009;10:161.
31. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–29.
32. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015;43:D1049–1056.
33. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The reactome pathway knowledgebase. Nucleic Acids Res. 2018;46:D649–D655.
34. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res. 2013;41:D793–800.
35. Chou T. The median-effect principle and the combination index for quantitation of synergism and antagonism. In: Chou TC and Rideout DC, editors. Synergism and antagonism in chemotherapy, Academic Press, San Diego, CA, 1991. 61–102.
36. Scholzen T, Gerdes J. The Ki-67 protein: from the known and the unknown. J Cell Physiol. 2000;182:311–22.
37. Issa JJ, Roboz G, Rizzieri D, Jabbour E, Stock W, O'Connell C, et al. Safety and tolerability of guadecitabine (SGI-110) in patients with myelodysplastic syndrome and acute myeloid leukaemia: a multicentre, randomised, dose-escalation phase 1 study. Lancet Oncol. 2015;16:1099–110.
38. Weinstein HJ, Griffin TW, Feeney J, Cohen HJ, Propper RD, Sallan SE. Pharmacokinetics of continuous intravenous and subcutaneous infusions of cytosine arabinoside. Blood. 1982;59:1351–3.
39. Sampath D, Cortes J, Estrov Z, Du M, Shi Z, Andreeff M, et al. Pharmacodynamics of cytarabine alone and in combination with 7-hydroxystaurosporine (UCN-01) in AML blasts in vitro and during a clinical trial. Blood. 2006;107:2517–24.
40. Yang AS, Estecio MR, Doshi K, Kondo Y, Tajara EH, Issa JP. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. Nucleic Acids Res. 2004;32:e38.
41. Marsters SA, Pitti RA, Sheridan JP, Ashkenazi A. Control of apoptosis signaling by Apo2 ligand. Recent Prog Horm Res. 1999;54:225–34.
42. Steinhart L, Belz K, Fulda S. Smac mimetic and demethylating agents synergistically trigger cell death in acute myeloid leukemia cells and overcome apoptosis resistance by inducing necroptosis. Cell Death Dis. 2013;4:e802.
43. Gerges S, Rohde K, Fulda S. Cotreatment with Smac mimetics and demethylating agents induces both apoptotic and necroptotic cell death pathways in acute lymphoblastic leukemia cells. Cancer Lett. 2016;375:127–32.
44. Chromik J, Safferthal C, Serve H, Fulda S. Smac mimetic primes apoptosis-resistant acute myeloid leukaemia cells for cytarabine-induced cell death by triggering necroptosis. Cancer Lett. 2014;344:101–9.
45. Billen LP, Shamas-Din A, Andrews DW. Bid: a Bax-like BH3 protein. Oncogene. 2008;27:S93–104.
46. Guicciardi ME, Bronk SF, Werneburg NW, Yin X-M, Gores GJ. Bid is upstream of lysosome-mediated caspase 2 activation in tumor necrosis factor α-induced hepatocyte apoptosis. Gastroenterology 2005;129:269–84.
47. Delbridge AR, Grabow S, Strasser A, Vaux DL. Thirty years of BCL-2: translating cell death discoveries into novel cancer therapies. Nat Rev Cancer. 2016;16:99–109.
48. Luedtke DA, Niu X, Pan Y, Zhao J, Liu S, Edwards H, et al. Inhibition of Mcl-1 enhances cell death induced by the Bcl-2-selective inhibitor ABT-199 in acute myeloid leukemia cells. Signal Transduct Target Ther. 2017;2:17012.
49. Garrido C, Galluzzi L, Brunet M, Puig PE, Didelot C, Kroemer G. Mechanisms of cytochrome c release from mitochondria. Cell Death Differ. 2006;13:1423–33.
50. Bell JA, Galaznik A, Huelin R, Stokes M, Guo Y, Fram RJ, et al. Effectiveness and safety of therapeutic regimens for elderly patients with acute myeloid leukemia: a systematic literature review. Clin Lymphoma Myeloma Leuk. 2018;18:e303–e314.
51. He P-F, Zhou J-D, Yao D-M, Ma J-C, Wen X-M, Zhang Z-H, et al. Efficacy and safety of decitabine in treatment of elderly patients with acute myeloid leukemia: a systematic review and meta-analysis. Oncotarget. 2017;8:41498.
52. Fulda S. Exploiting inhibitor of apoptosis proteins as therapeutic targets in hematological malignancies. Leukemia. 2012;26:1155–65.
53. Hollenbach PW, Nguyen AN, Brady H, Williams M, Ning Y, Richard N, et al. A comparison of azacitidine and decitabine activities in acute myeloid leukemia cell lines. PLoS ONE. 2010;5:e9001.

54. Metzeler KH, Walker A, Geyer S, Garzon R, Klisovic RB, Bloomfield CD, et al. DNMT3A mutations and response to the hypomethylating agent decitabine in acute myeloid leukemia. Leukemia. 2011;26:1106.
55. Chen D, Christopher M, Helton NM, Ferguson I, Ley TJ, Spencer DH. DNMT3A(R882)-associated hypomethylation patterns are maintained in primary AML xenografts, but not in the DNMT3A (R882C) OCI-AML3 leukemia cell line. Blood Cancer J. 2018;8:38.
56. Karlic H, Herrmann H, Varga F, Thaler R, Reitermaier R, Spitzer S, et al. The role of epigenetics in the regulation of apoptosis in myelodysplastic syndromes and acute myeloid leukemia. Crit Rev Oncol Hematol. 2014;90:1–16.
57. Chen J-J, Shen HCJ, Rivera Rosado LA, Zhang Y, Di X, Zhang B. Mislocalization of death receptors correlates with cellular resistance to their cognate ligands in human breast cancer cells. Oncotarget. 2012;3:833–42.
58. Bousserouel S, Le Grandois J, Gosse F, Werner D, Barth SW, Marchioni E, et al. Methanolic extract of white asparagus shoots activates TRAIL apoptotic death pathway in human cancer cells and inhibits colon carcinogenesis in a preclinical model. Int J Oncol. 2013;43:394–404.
59. Bush JA, Cheung KJ Jr., Li G. Curcumin induces apoptosis in human melanoma cells through a Fas receptor/caspase-8 pathway independent of p53. Exp Cell Res. 2001;271:305–14.
60. Day TW, Huang S, Safa AR. c-FLIP knockdown induces ligand-independent DR5-, FADD-, caspase-8-, and caspase-9-dependent apoptosis in breast cancer cells. Biochem Pharmacol. 2008;76:1694–704.
61. Micheau O, Solary E, Hammann A, Dimanche-Boitrel MT. Fas ligand-independent, FADD-mediated activation of the Fas death pathway by anticancer drugs. J Biol Chem. 1999;274:7987–92.
62. Shin DY, Park YS, Yang K, Kim GY, Kim WJ, Han MH, et al. Decitabine, a DNA methyltransferase inhibitor, induces apoptosis in human leukemia cells through intracellular reactive oxygen species generation. Int J Oncol. 2012;41:910–8.
63. Sebert M, Renneville A, Bally C, Peterlin P, Beyne-Rauzy O, Legros L, et al. A phase II study of guadecitabine in higher-risk myelodysplastic syndrome and low blast count acute myeloid leukemia after azacitidine failure. Haematologica. 2019;104:1565–71.
64. Carter BZ, Mak PY, Mak DH, Shi Y, Qiu Y, Bogenberger JM, et al. Synergistic targeting of AML stem/progenitor cells with IAP antagonist birinapant and demethylating agents. J Natl Cancer Inst. 2014;106:djt440.

3.2 *HIF-1α* and *HIF-2α* differently regulate tumour development, metabolism and inflammation of clear cell renal cell carcinoma in mice

In kidney carcinoma, clear cell renal cell carcinoma (ccRCC) represents $70-80\%$ of all observed carcinomas [Frew and Moch, 2015]. In the majority of cases a homozygeous inactivation of the *VHL* tumor suppressor gene is present and represents a central element in the early tumor onset [Gerlinger et al., 2014]. Subsequent mutations or copy number variations in epigenetic regulators, cell cycle regulators or in the *PI3K* pathway further promote the development of ccRCC. It is known that a mutational inactivation of *VHL* leads to accumulation of the *HIF-1α* and *HIF-2α* transcription factors [Maxwell et al., 1999]. Various studies reported *HIF-1α* as tumor suppressor inhibiting, and *HIF-2α* as an oncogene promoting the aggressiveness of ccRCC behaviour [Monzon et al., 2011, Turajlic et al., 2018, Gordan et al., 2008, Shen et al., 2011]. In this research we could show that *HIF-1α* is essential for tumor formation and that the deletion of *HIF-2α* has only minor effects on tumor initiation and growth. Nevertheless, both genes are required for the clear cell phenotype of ccRCC. Therefore, we concluded that *HIF-1α* has an oncogenic role in ccRCC initiation and suggested alterations in the balance of *HIF-1α* and *HIF-2α* which affects different aspects of ccRCC biology and disease aggressiveness.

Hoefflin, R., Harlander, S., Schäfer, S., Metzger, P., Kuo, F., Schönenberger, D., Adlesic, M., Peighambari, A., Seidel, P., Chen, C., Consenza-Contreras, M., Jud, A., Lahrmann, B., Grabe, N., Heide, D., Uhl, F., Chan, T. A., Duyster, J., Zeiser, R., Schell, C., Heikenwalder, M., Schilling, O., Hakimi, A. A., Boerries, M., Frew, I. J. (2020). ***HIF-1α* and *HIF-2α* differently regulate tumour development, metabolism and inflammation of clear cell renal cell carcinoma in mice**. Nature Communications.

Contribution: I conducted the transcriptome analyses, including identification of differentially expressed genes and gene set enrichment analyses. Additionally, I performed the correlation analyses between the transriptome and proteome data sets. I wrote the paragraphs *RNA-sequencing* and *Gene Set Enrichment Analysis*. I created figures 3, 4, 5, and supplemental figures S6, S7, S9, and edited the manuscript.

HIF-1α and HIF-2α differently regulate tumour development and inflammation of clear cell renal cell carcinoma in mice.

Rouven Hoefflin [1,*], Sabine Harlander [2,3,*], Silvia Schäfer [1,4,5], Patrick Metzger [5,6], Fengshen Kuo [7], Désirée Schönenberger [2,3], Mojca Adlesic [1,4], Asin Peighambari [1,4,5], Philipp Seidel [1,4], Chia-yi Chen [8], Miguel Consenza-Contreras [8], Andreas Jud [9], Bernd Lahrmann [10], Niels Grabe [10], Danijela Heide [11], Franziska M. Uhl [1,5], Timothy A. Chan [7], Justus Duyster [1,12], Robert Zeiser [1,4,12], Christoph Schell [8], Mathias Heikenwalder [11], Oliver Schilling [8,12,13], A. Ari Hakimi [7,14], Melanie Boerries [6,12,13] and Ian J. Frew [1,2,3,4,5,12]

[1] Clinic of Internal Medicine I, Hematology, Oncology and Stem Cell Transplantation, Medical Centre - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[2] Institute of Physiology, University of Zurich, Zurich, Switzerland
[3] Zurich Center for Integrative Human Physiology, University of Zurich, Zurich, Switzerland
[4] Signalling Research Centres BIOSS and CIBSS, University of Freiburg, Freiburg, Germany
[5] Faculty of Biology, University of Freiburg, Freiburg, Germany
[6] Institute of Medical Bioinformatics and Systems Medicine, Medical Centre - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[7] Immunogenomics & Precision Oncology Platform (IPOP), Memorial Sloan Kettering Cancer Center, New York, NY.
[8] Institute for Surgical Pathology, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[9] Department of General and Visceral Surgery, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany.
[10] Hamamatsu Tissue Imaging and Analysis (TIGA) Center, BioQuant, University of Heidelberg, Heidelberg, Germany.
[11] Division of Chronic Inflammation and Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany
[12] Comprehensive Cancer Center Freiburg (CCCF), Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[13] German Cancer Consortium (DKTK), Partner Site Freiburg; and German Cancer Research Center (DKFZ), Heidelberg, Germany
[14] Urology Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY.
* These authors contributed equally

Corresponding Author and Address: Ian Frew

Clinic of Internal Medicine I, Hematology, Oncology and Stem Cell Transplantation
Medical Centre - University of Freiburg
Breisacherstraße 115
79106 Freiburg
Germany
Phone: +49 761 270 71831
Email: ian.frew@uniklinik-freiburg.de

ABSTRACT

Mutational inactivation of *VHL* is the earliest genetic event in the majority of clear cell renal cell carcinomas (ccRCC), leading to accumulation of the HIF-1α and HIF-2α transcription factors. While correlative studies of human ccRCC and functional studies using human ccRCC cell lines have implicated HIF-1α as an inhibitor and HIF-2α as a promoter of aggressive tumour behaviours, their roles in tumour onset have not been functionally addressed. Herein we show using an autochthonous ccRCC model that *Hif1a* is essential for tumour formation whereas *Hif2a* deletion has only minor effects on tumour initiation and growth. Both HIF-1α and HIF-2α are required for the clear cell phenotype. Transcriptomic and proteomic analyses reveal that HIF-1α regulates glycolysis while HIF-2α regulates genes associated with lipoprotein metabolism, ribosome biogenesis and E2F and MYC transcriptional activities. HIF-2α-deficient tumours are characterised by increased antigen presentation, interferon signalling and $CD8^+$ T cell infiltration and activation. Single copy loss of *HIF1A* or high levels of *HIF2A* mRNA expression correlate with altered immune microenvironments in human ccRCC. These studies reveal an oncogenic role of HIF-1α in ccRCC initiation and suggest that alterations in the balance of HIF-1α and HIF-2α activities can affect different aspects of ccRCC biology and disease aggressiveness.

INTRODUCTION

More than 400,000 new cases of kidney cancer arose worldwide in 2018 [1]. Clear cell renal cell carcinoma (ccRCC) represents 70-80% of all cancers of the kidneys [2]. Biallelic inactivation of the von Hippel-Lindau (*VHL*) tumour suppressor gene is a truncal genetic event that arises in the majority of cases of ccRCC [3–6], demonstrating that loss of one or more of the various tumour suppressor functions of the pVHL protein isoforms [2,7] is central to the earliest steps in the initiation of tumour formation. Subsequent mutations or chromosomal copy number alterations in epigenetic regulatory genes (including *PBRM1*, *BAP1*, *SETD2*, *KDM5C*), cell cycle regulatory genes (including *TP53*, *CDKN2A*, *MYC*) or PI3K pathway genes (including *PIK3CA*, *PTEN*, *MTOR*, *TSC1*) arise recurrently in ccRCC and are believed to cooperate with *VHL* inactivation to promote the development and evolution of ccRCC tumours [8,9]. Numerous mouse models have supported this notion of genetic cooperation by showing that renal epithelial cell-specific inactivation of different combinations of *Vhl* together with *Pten* [10], *Tsc1* [11], *Pbrm1* [11–13], *Bap1* [11,14], *Trp53* [15], *Trp53/Rb1* [16], *Cdkn2a* [17] or with *Myc* [17] overexpression causes the formation of cystic and solid precursor lesions or ccRCC tumours.

The best characterised tumour suppressor function of pVHL relates to its role in targeting the alpha subunits of the hypoxia-inducible transcription factors (HIF-1α and HIF-2α) for oxygen-dependent, ubiquitin-mediated proteolytic degradation [18]. Genetic inactivation of *VHL* causes the constitutive stabilisation of HIF-1α and HIF-2α, which induce gene expression programs that play a central role in the pathogenesis of ccRCC by altering cellular metabolism, inducing angiogenesis, promoting epithelial to mesenchymal transition, invasion and metastatic spread. Numerous lines of evidence argue that HIF-2α plays a major pro-tumourigenic role in established human ccRCCs, whereas HIF-1α appears to function rather to inhibit aggressive tumour behaviour. Loss of the region of chromosome 14q harbouring *HIF1A* correlates with poor survival [19] and is commonly found in ccRCC metastases [20]. ccRCC tumours that express only HIF-2α have higher proliferation rates than those expressing HIF-1α and HIF-2α [21]. ccRCC tumour cell lines frequently display intragenic deletions of *HIF1A* but express wildtype HIF-2α [22]. HIF-2α is necessary for the formation of ccRCC xenografts [23,24] while knockdown of HIF-1α enhances xenograft tumour formation in cell lines that express both HIF-1α and HIF-2α [22]. These observations have given rise to the concept that HIF-2α functions as a ccRCC oncogene and HIF-1α as a tumour suppressor. This prompted the development of HIF-2α-specific inhibitors which show excellent on-target efficacy in ccRCC xenograft models, efficacy in a subset of patient-derived xenograft models and clinical responses in some patients in phase I clinical trials [25–27]. These pharmacological studies in patient-derived xenograft models however also indicate that HIF-2α specific inhibition is not sufficient to inhibit the growth of all ccRCCs [25], suggesting that other oncogenic drivers may be important in some or all tumours. It should be noted that all of the functional and genetic data described above largely relates to either studies of established, later stage ccRCC human tumours or to the somewhat artificial setting of xenograft tumour formation by cultured ccRCC cell lines or patient-derived xenograft models. These studies have necessarily been unable to adequately assess the involvement of HIF-1α and HIF-2α throughout the entire process of tumour evolution beginning with *VHL* mutant cells in the context of a normal renal tubular epithelium.

To address the roles of HIF-1α and HIF-2α in the development of ccRCC we take advantage of an accurate mouse model of ccRCC based on tamoxifen-inducible renal epithelial cell-specific deletion (Ksp-CreER^{T2}) of *Vhl*, *Trp53* and *Rb1* [16]. This mouse model at least partly reflects the complex patterns of chromosomal copy number gains and losses of cell cycle regulatory genes in human ccRCC and reproduces many aspects of the evolution of human ccRCC by first developing cystic and solid precursor lesions that progress to tumours over the course of 5-12 months following gene deletion in adult mice [16]. Tumours arising in this model exhibit histological, immunohistochemical, transcriptional and mutational similarities to human ccRCC [16]. We introduce floxed alleles of *Hif1a* and *Hif2a* (also known as *Epas1*) into this genetic background and show that HIF-1α is essential for tumour formation whereas deletion of HIF-2α has only moderate effects on tumour onset and growth rate but leads to increased intratumoural immune activation. This study defines differing roles of HIF-1α and HIF-2α in ccRCC formation and progression and suggests a model in which alterations in their relative activities affect different aspects of tumour biology and immunology.

RESULTS

ccRCC formation is strongly dependent on *Hif1a*

We fed 6 week-old mice tamoxifen-containing food for two weeks to induce gene deletion in cohorts of *Ksp-CreER^{T2}; Vhl$^{fl/fl}$; Trp53$^{fl/fl}$; Rb1$^{fl/fl}$* (hereafter termed Vhl$^{\Delta/\Delta}$Trp53$^{\Delta/\Delta}$Rb1$^{\Delta/\Delta}$ in the text and VpR in figures), *Ksp-CreER^{T2}; Vhl$^{fl/fl}$; Trp53$^{fl/fl}$; Rb1$^{fl/fl}$; Hif1a$^{fl/fl}$* (hereafter termed Vhl$^{\Delta/\Delta}$Trp53$^{\Delta/\Delta}$Rb1$^{\Delta/\Delta}$Hif1a$^{\Delta/\Delta}$ in the text and VpRH1 in figures) and *Ksp-CreER^{T2}; Vhl$^{fl/fl}$; Trp53$^{fl/fl}$; Rb1$^{fl/fl}$; Hif2a$^{fl/fl}$* (hereafter termed Vhl$^{\Delta/\Delta}$Trp53$^{\Delta/\Delta}$Rb1$^{\Delta/\Delta}$Hif2a$^{\Delta/\Delta}$ in the text and VpRH2 in figures) mice. Tumour onset, volume and numbers were monitored over time using contrast-assisted μCT imaging and mice were sacrificed at individual timepoints based on the presence of rapid tumour growth. These data were added to, or compared to, our previously published [16] analyses of separate Vhl$^{\Delta/\Delta}$Trp53$^{\Delta/\Delta}$Rb1$^{\Delta/\Delta}$ and Trp53$^{\Delta/\Delta}$Rb1$^{\Delta/\Delta}$ (termed pR in figures) cohorts, respectively. All animals from both cohorts were housed in the same animal facility. We first determined that tumour growth curves (Supplementary Fig. 1a) showed an excellent goodness of fit (Supplementary Fig. 1b) to the exponential linear regression $e^{\alpha t}$ where α describes the co-efficient of exponential growth, a mathematical description of the tumour growth rate, and t represents time in days after gene deletion. These analyses showed that *Vhl* deletion accelerates tumour onset (Fig. 1a), increases tumour number (Fig. 1b) and increases tumour growth rate (Fig. 1c) in the Trp53$^{\Delta/\Delta}$Rb1$^{\Delta/\Delta}$ background. *Hif1a* co-deletion completely abolished these tumour-promoting effects of *Vhl* deletion (Fig. 1a) and these mice developed very few tumours (Fig. 1b), which grew slowly when they did develop (Fig. 1c). In contrast, *Hif2a* deletion caused more moderate, yet statistically significant effects, partly delaying tumour onset (Fig. 1a), partly reducing the number of tumours per mouse (Fig. 1b) and average tumour growth rates (Fig. 1c). Metastases were not observed in any of the genotypes. These data indicate that HIF-1α is very important for the efficient evolution and growth of *Vhl* mutant ccRCCs, while HIF-2α is only partly required and many tumours still develop in the *Vhl/Trp53/Rb1/Hif2a* quadruple mutant background.

Since *Hif1a* deletion provided such a strong phenotypic rescue we next investigated whether the *Hif1a* and *Hif2a* genes are indeed deleted in the relevant tumours to exclude that the tumours might be escapers in which Cre activity failed to correctly recombine the floxed *Hif1a* or *Hif2a* alleles. PCRs specific for the recombined *Hif1a* and *Hif2a* alleles revealed that tumour DNA exhibited Cre-induced recombined alleles of these genes in tumours from the relevant mice (Supplementary Fig. 2a). To compare the extent of Cre-mediated deletion of *Vhl*, *Trp53*, *Rb1*, *Hif1a* and *Hif2a* we conducted quantitative real time PCR using primers to specifically amplify floxed exons of each gene as well as non-floxed exons of the *Vhl*, *Trp53* and *Hif2a* genes (which served as normalisation controls) from genomic DNA from cortex samples from non-Cre mice (WT cortex), as well as tumours from Vhl$^{\Delta/\Delta}$Trp53$^{\Delta/\Delta}$Rb1$^{\Delta/\Delta}$, Vhl$^{\Delta/\Delta}$Trp53$^{\Delta/\Delta}$Rb1$^{\Delta/\Delta}$Hif1a$^{\Delta/\Delta}$ and Vhl$^{\Delta/\Delta}$Trp53$^{\Delta/\Delta}$Rb1$^{\Delta/\Delta}$Hif2a$^{\Delta/\Delta}$ mice. These analyses showed that the allelic ratios of floxed exons normalised to the average of the 3 non-floxed exons were reduced for *Vhl*, *Trp53* and *Rb1* compared to WT cortex and for *Hif1a* and *Hif2a* when compared to WT cortex or to mouse genotypes that did not harbour the floxed allele (Supplementary Fig. 2b). Residual floxed exons (approximately 10-15% allelic burden) in the genomic DNA of tumours likely reflect DNA derived from non-Cre-expressing cells of the tumour stroma. These analyses demonstrated that all genes are deleted by Cre and importantly that *Hif1a* and *Hif2a* are deleted to similar extents to *Vhl*, *Trp53* and *Rb1* in the relevant tumour samples. We additionally analysed RNA sequencing data (see experiments described below) which showed that Vhl$^{\Delta/\Delta}$Trp53$^{\Delta/\Delta}$Rb1$^{\Delta/\Delta}$ tumours displayed lower mRNA levels of *Hif1a* and *Hif2a* than WT cortex but that there was no compensatory

upregulation of *Hif2a* in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ tumours, nor of *Hif1a* in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours (Supplementary Fig. 3a,b). This data also revealed the specific reduction in the relative numbers of sequencing reads in the floxed exons compared to adjacent non-floxed exons. This was true for all floxed genes in the mouse genotypes that contain the floxed alleles but not in those that do not (Supplementary Fig. 2c), consistent with specific Cre-mediated recombination occurring equivalently for all genes in all genotypes. In the case of *Vhl*, the deletion of the first exon including the first intronic mRNA splice site leads to sequencing read-through into the intron. Since the intronic sequencing reads do not start at the same position in different tumour samples, it is difficult to assess the effect of this read-through on potential translation of the resulting mRNA transcript, however western blotting of primary cells derived from $Vhl^{fl/fl}$ mice demonstrated that Cre-mediated recombination results in complete loss of the pVHL protein isoforms ([15] and Supplementary Fig. 5). The slightly varying degrees of residual sequencing reads in the floxed exons of the different genes likely reflects gene expression in various types of tumour stromal cells, which likely differentially express the different genes.

Tumours in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ mice lacked the clear nuclear HIF-1α signal that was present in tumours from $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ mice when staining with an anti-HIF-1α antibody (Fig. 1d). RNA sequencing analyses identified *Car9* as a HIF-1α specific target gene (Supplementary Fig. 3c) and the protein product of this gene, CA9, showed membrane staining in tumours from $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ mice but not in tumours from $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ mice (Fig. 1d). Nuclear HIF-2α staining was present in tumours from $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ mice but absent in all tumours from $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ mice. Collectively these analyses demonstrate that Cre activity occurs equivalently for all floxed genes and that *Hif1a* and *Hif2a* are deleted in the tumours arising in the relevant mouse backgrounds.

Characterisation of *Hif1a*- and *Hif2a*-deficient mouse ccRCC

Histomorphological analyses and comparisons were performed for the different genetic backgrounds on a total of 26 ($Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$), 16 ($Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$) and 21 ($Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$) H&E stained tumours. In line with our previous report [16], all $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$-tumours were classified as mid- to high grade (46% grade 2; 54% grade 3) tumours. The malignant lesions in the $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ (19% grade 2; 75% grade 3; 6% grade 4) or $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$-background (14% grade 2; 81% grade 3; 5% grade 4) displayed on average higher grades. Since tumour grade up to grade 3 is classified mostly based on nucleolus size, these data hint that loss of HIF-1α or HIF-2α may modify processes such as transcription of ribosomal DNA genes that affect the nucleolus [28]. Potentially relevant mechanisms that have been previously linked to HIF-α activities and that might contribute to nucleolar alterations include metabolic generation of ATP and deoxynucleotides to fuel transcription, epigenetic regulatory mechanisms and DNA repair [28]. In order to analyse similarities to the classical human ccRCC clear cell phenotype, we established a scoring system (Fig. 1e) based on a three-tiered classification of tumours with completely clear cytoplasm (score 1), partly clear or weakly stained cytoplasm (score 2) or stronger cytoplasmic eosin staining (score 3). Fifty-seven percent of the $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours were classified with a clear cell score of one or two, whereas the vast majority of tumors in the other genetic backgrounds showed a score of three (86% of $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ and 83% of $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$) (Fig. 1f). This observation is consistent

with our previous findings that HIF-1α is necessary for the clear cell phenotype of normal renal epithelial cells following *Vhl* deletion [29] but also implicates HIF-2α in the clear cell phenotype in this tumour model. Intratumoural histomorphological heterogeneity was observed mainly in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ (23%) and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumors (28%) (Supplementary Fig. 4a) and is a well-known characteristic of human ccRCC [30]. Other typical histopathological features of ccRCC like necrosis (Supplementary Fig. 4b) or intratumoural hemorrhage (Supplementary Fig. 4c) were equally distributed throughout the different genotypes and were mainly detected in larger tumours. The vast majority of tumours showed a solid and spherical growth pattern with pushing rather than infiltrating borders. Hemangioinvasion with direct tumour infiltration of blood vessels or extra-parenchymal invasion of the perirenal fat tissue was not observed in any of the cases. A subset of $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ (15%), $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ (25%) and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours (19%) exhibited cystic features (Supplementary Fig. 4d). While all ccRCC in the $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ background showed strong phospho-4E-BP1-staining, indicative of PI3K/mTOR-pathway activation, only 50% of the $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ and 88% of the $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours were positive (Fig. 1d). Irrespective of the genetic background, all malignant lesions stained positively for the proximal tubule marker CD10 (Fig. 1d).

Cancer assays do not reflect HIF-1α's *in vivo* oncogenic role

Since our genetic experiments demonstrated that HIF-1α is necessary for the efficient initiation of ccRCC formation we wondered firstly if HIF-1α is generally required for cellular proliferation following loss of *Vhl* and secondly whether established $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours remain dependent on HIF-1α or whether they might lose this dependency during tumour evolution. To mimic the earliest events in ccRCC formation in a genetically tractable cellular system that allows long term proliferation assays, we derived mouse embryo fibroblasts (MEFs) from wild type, $Vhl^{fl/fl}$, $Vhl^{fl/fl}Hif1a^{fl/fl}$, $Vhl^{fl/fl}Hif2a^{fl/fl}$ and $Vhl^{fl/fl}Hif1a^{fl/fl}Hif2a^{fl/fl}$ embryos, infected them with Adeno-GFP as control or Adeno-Cre-GFP and confirmed the deletion of the floxed genes by real time PCR and western blotting (Supplementary Fig. 5a,b). Long term proliferation assays confirmed previous findings that loss of *Vhl* in MEFs induces an early loss of proliferative capacity, however, in contrast to the initial claim that this senescence phenotype is independent of HIF-α activity [31], our results clearly demonstrate the dependency on *Hif1a* but not on *Hif2a* (Fig. 2a). This proliferative rescue due to *Hif1a* deletion contrasts with the suppression of ccRCC initiation by *Hif1a* deletion *in vivo*. To remove the potential confounding factor of senescence we took advantage of the fact that deletion of *Trp53* overcomes the phenotype of loss of proliferative capacity associated with loss of *Vhl* [15,32]. To investigate the effect of loss of HIF-1α function in immortalised cells we infected $Vhl^{fl/fl}Trp53^{fl/fl}$ MEFs with lentiviruses expressing either non-silencing control shRNA or expressing two different shRNAs against *Hif1a* and infected the cells with Adeno-GFP or Adeno-Cre-GFP. Western blotting confirmed the reduced abundance of pVHL, p53 and HIF-1α (Supplementary Fig. 5c). Knockdown of HIF-1α further increased the proliferation rate of immortalised *Vhl/Trp53* null MEFs (Fig. 2b), furthering illustrating that HIF-1α generally acts to inhibit proliferation in the context of *Vhl* deletion.

We next used a cell line derived from a mouse $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ ccRCC (termed 2020 cells) (Supplementary Fig. 5d) and introduced human pVHL30 to rescue *Vhl* function (Supplementary Fig. 5e) as well as knocked down *Hif1a* with two independent shRNAs (Supplementary Fig. 5f). We confirmed the

efficient functional re-introduction of pVHL30 and knockdown of *Hif1a* mRNA in reducing HIF-1α protein (Supplementary Fig. 5g) and showed that the knockdowns reduced the abundance of the PDK1 and LDH-A proteins, that are encoded by the HIF-1α transcriptional target genes *Pdk1* and *Ldha*, equivalently to pVHL30 reintroduction (Supplementary Fig. 5h). Proliferation assays revealed that neither pVHL30 reintroduction (Fig. 2c), nor *Hif1a* (Fig. 2d) knockdown affected cellular proliferation of 2020 cells growing in renal epithelial medium on cell culture plastic but that either of these manipulations were sufficient to increase the growth of 2020 cells as spheroids in non-adherent cell culture conditions (Fig. 2e, 2f and 2g), a common readout of cellular transformation. Since re-introduction of pVHL into human ccRCC cell lines does not affect proliferation rates in culture, but does inhibit tumour formation in the xenograft setting [33], we conducted allograft studies in SCID-Beige mice. pVHL re-introduction into 2020 cells significantly delayed tumour growth (Fig. 2h) but HIF-1α knockdown did not (Fig. 2i).

Collectively, these studies show that HIF-1α in fact antagonises cellular proliferation of normal mouse cells lacking *Vhl* and is dispensable for proliferation and allograft tumour formation of a mouse ccRCC cell line, highlighting the specificity of the requirement for HIF-1α for tumour onset in the autochthonous setting. This argues that the oncogenic role of HIF-1α is evident only in the context of the physiological environment of the renal epithelium.

Impact of HIF-1α and HIF-2α on mouse ccRCC transcriptome

Our previous analyses demonstrated that the mouse ccRCC model exhibits an excellent overlap with human ccRCC at the global transcriptional level [16]. To gain further insight into *in vivo* relevant functions of HIF-1α and HIF-2α we compared the molecular features of tumours that developed in the presence of both HIF-1α and HIF-2α to those that were genetically restricted to develop in the absence of either HIF-1α or HIF-2α. We conducted RNA sequencing of 6 wild type (WT) cortex samples, 6 $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumour samples, 8 $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ tumour samples and 10 $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumour samples, and combined these data with our previously-obtained RNA sequencing data from 3 WT cortex samples and 6 $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumour samples. After read trimming and mapping (Supplementary Fig. 6a,b) the mRNA abundance of 19,723 genes was determined in each sample. All normalised gene expression values are provided in Supplementary Data 1. Transcriptomic profile principal component analysis and unsupervised hierarchical clustering by sample Euclidean distance matrix (Supplementary Fig. 6c) suggested minimal batch effect amongst different sequencing runs. Principal component analysis (Fig. 3a) also revealed clear separation of WT cortex from all tumour samples on the PC1 axis and this accounted for 36% of the overall variability. $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours tended to segregate from one another on the PC2 axis, which represented 9% of total variability, suggesting that they are the most distinct in terms of gene expression patterns, whereas $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ tumours were more widely distributed along the entire axis. These analyses are consistent with the deletion of *Vhl*, *Trp53* and *Rb1* in all three tumour genotypes inducing large transcriptional changes, with more limited and specific contributions of HIF-1α and HIF-2α to the regulation of specific sets of genes.

We focused analyses on genes that were differentially expressed between $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours and the $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumour genotypes and identified 396

differentially expressed genes that are dependent on HIF-1α (Supplementary Fig. 7a), 804 differentially expressed genes that are dependent on HIF-2α (Supplementary Fig. 7b) and 131 differentially expressed genes that are dependent on both HIF-1α and HIF-2α (Supplementary Fig. 7c). To begin to identify biological processes that these sets of genes are likely to reflect or regulate, we conducted Generally Applicable Gene-set Enrichment (GAGE) analyses using the pathway databases from ConsensusPathDB, the Biological Processes from Gene Ontology and MSigDB terms for Chemical and Genetic Perturbations and Transcription Factor Targets. The full list of statistically significantly altered (*P adj.* < 0.05) gene sets of these GAGE analyses are provided in Supplementary Data 2. $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ tumours in comparison to $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours show low expression of glycolytic genes (Fig. 3b), as well as signatures associated with hypoxia and known HIF-1α targets. This is consistent with a large body of previous work in human ccRCC cells [34] and in mouse models [29] that shows that HIF-1α is the primary transcription factor that promotes Warburg-like metabolism of high rates of glycolysis and low oxidative phosphorylation. Additional HIF-1α-dependent signatures include reduced expression of genes involved in cell adhesion (Fig. 3c) and focal adhesion and receptor signalling (Fig. 3d). Genes that were expressed at low levels in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours compared to the other two tumour genotypes included those involved in different DNA repair processes (*e.g. Fancf* – DNA interstrand cross link repair, *Rad52* – homologous recombination repair, *Ogg1* – oxidative stress induced base excision repair, *Ercc2* – transcription coupled nucleotide excision repair) (Fig. 3e), cholesterol uptake and lipoprotein metabolism (Fig. 3f), which may relate to the observed dependency of the clear cell phenotype on HIF-2α, and ribosome biogenesis (Fig. 3g), potentially consistent with the slower rate of proliferation of $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours. Other HIF-2α-dependent GAGE terms that might be relevant to the evolution and proliferation of $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours included genes that are targets of the MYC and E2F transcription factors, consistent with previous studies showing that HIF-2α promotes the activity of the MYC transcription factor [21]. GAGE analyses also identified a significant downregulation of the gene set HIF-2α Transcription Network, including *Epo*, *Egln1*, *Egln3*, *Igfbp1* and *Pfkfb3*. To further investigate potential overlap with recently-defined HIF-2α-dependent genes in human ccRCC, we used a set of 277 genes that were identified as being inhibited specifically in tumour cells in ccRCC tumorgrafts in mice treated with the HIF-2α inihibitor PT2399 [25,35]. Analyses of the expression levels of the mouse orthologues of this set of HIF-2α target genes revealed that many of these genes are highly upregulated in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours compared to WT cortex, but that the loss of either HIF-1α or HIF-2α did not broadly affect the up-regulation of these genes (Supplementary Fig. 7d). Nonetheless, 11 genes, marked in red in Supplementary Fig. 7d and shown in Supplementary Fig. 7e, were expressed at significantly lower levels (fold change < -1.7, *P* < 0.05) in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours than in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours. The relatively small overlap between mouse and human HIF-2α-dependent genes may be due to inherent differences between mice and humans, to the very different experimental settings of acute pharmacological inhibition versus tumour evolution in the genetic absence of *Hif2a*, or to specific features of this particular model of ccRCC. In this latter context, it is noteworthy that many human HIF-2α-dependent ccRCC genes are related to the cell cycle and to DNA damage responses. These signatures are highly represented in the comparison $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ vs WT cortex (see GAGE signatures in Supplementary Data 2). We speculate that it is likely that these genes are not dependent on HIF-2α in the mouse model due to the fact that the deletion of *Rb1* and *Trp53* already strongly affects these classes of genes.

Interestingly, genes that were upregulated in HIF-2α-deficient tumours include those enriched in diverse GAGE terms for interferon signalling, T cell activation, innate immunity, adaptive immunity, antigen processing and presentation and NF-κB as well as IRF transcription factor targets, suggestive of an altered immune environment in these tumours. Supplementary Fig. 8a shows a selection of these enriched immune signatures and highlights that the signatures are upregulated in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours compared to WT cortex, that there are very few or no statistically significant differences in these signatures between $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours (*i.e.* that HIF-1α deficiency does not strongly alter the inflammatory tumour environment in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours) and that all of these signatures are further highly statistically significantly upregulated in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours in comparison to $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours. Gene expression heatmaps of differentially expressed genes associated with GAGE terms for T cell activation (Fig. 3h), response to IFN-β (Fig. 3i) and IFN-γ production (Supplementary Fig. 8b) are shown as examples of these inflammatory signatures. We conclude that these analyses suggest that there is a complex inflammatory response in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours that is further modified by HIF-2α deficiency. These phenotypes were further investigated in experiments described in the following sections.

Impact of HIF-1α and HIF-2α on mouse ccRCC proteome

In order to further explore whether the biological alterations predicted by transcriptomic analyses are also reflected at the protein expression level, as well as to attempt to capture differences in the proteomes of the tumours that might not be reflected in their transcriptomes, we used exploratory quantitative proteomic analyses of 6 samples of WT cortex and 6 tumours each from $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$, $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ mice as an independent discovery tool. These analyses allowed the quantification of 4257 proteins that were present in at least 4 of 6 samples of each genotype (Supplementary Data 3). As is commonly observed in comparisons of proteome and transcriptome data, the overall correlations of protein abundance and mRNA abundance were low (Supplementary Fig. 9a). However, there were strong correlations between fold changes in mRNA abundance and fold changes in protein abundances when analysing only those proteins that showed differential expression between genotypes (Supplementary Fig. 9b-d). Through comparison with previously-conducted analyses of the proteome of 8 human ccRCC tumours [36,37], we identified a strong correlation in the relative abundance of proteins in mouse $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ ccRCC and in human ccRCC (Fig. 4a). Of the differentially expressed proteins identified in the comparison between mouse $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ ccRCC and wild type cortex, 82% were also identified as differentially expressed proteins in comparisons of human ccRCC to normal kidney (Figure 4b), further emphasising that the $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ model accurately reflects the molecular features of human ccRCC. Using a less stringent cut-off for statistical significance ($P < 0.01$), we identified 884 proteins that are upregulated in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ ccRCCs compared to WT cortex (Fig. 4c). To characterise biological pathways that are altered in tumour compared to normal tissue, we conducted two complementary analyses; ROAST (rotation gene set testing) analysis [38] was used to assess gene set enrichment based on the expression levels of all measured proteins and gene set enrichment analysis using the online platform of MSigDB (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) was performed using only the lists of statistically differentially upregulated proteins. These analyses revealed many overlaps with one another as well as with GAGE gene set terms that emerged from the analyses of the transcriptome,

including glycolysis, hypoxia, DNA repair, mTORC1 signalling, E2F and MYC targets and IFNγ response (Supplementary Data 3 and Supplementary Fig. 10a).

To compare the effect of the absence of HIF-1α or HIF-2α on the proteome, we first conducted principal components analysis (Supplementary Fig. 10b), which revealed that all tumour samples clustered separately from the WT cortex samples, but that the tumour samples of all of the different genotypes largely overlapped with one another, suggesting a relatively high degree of similarity in the overall protein expression patterns of tumours from the different genetic backgrounds. ROAST analyses as well as gene set enrichment analyses of the lists of proteins that are differentially expressed between $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours (Fig. 4d) and between $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours (Fig. 4d,e and Supplementary Data 3) both also highlighted numerous similarities to the transcriptomic analyses. HIF-1α deficiency reduced expression of glycolytic enzymes and increased expression of proteins associated with oxidative phosphorylation and respiratory electron transport (Fig. 4f), while HIF-2α deficiency reduced the expression of MYC targets and resulted in increased expression of genes associated with immune responses, interferon signaling, cytokine signalling and antigen presentation (Fig. 4g). In conclusion, the analyses of the proteomes strongly align with the analyses of the transcriptomes, providing independent validation for the predicted biological differences between the tumour genotypes.

HIF-2α deficiency alters antigen presentation in ccRCC

Our transcriptomic and proteomic analyses implicated antigen presentation as being upregulated in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours. Heatmaps of RNA sequencing data revealed higher levels of expression of many MHC class I (Fig. 5a) and MHC class II (Fig. 5b) genes, as well as other genes involved in antigen processing and presentation (Fig. 5c) in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours than in the tumours of the other genotypes. Immunohistochemical staining with an anti-MHC class II antibody revealed that all tumour genotypes displayed cases in which the tumour cells were either negative, partly positive or almost entirely positive (Fig. 5d). However, tumour cells in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours were more frequently positive than the other genotypes (Fig. 5e). Since MHC class II expression is upregulated by interferon-γ signalling, these results are consistent with the fact that interferon signalling terms were upregulated in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours in our transcriptomic and proteomic analyses. Tumour cells in at least half of ccRCCs are immunohistochemically positive for MHC class II expression [39,40] and ccRCC cells have been shown to present class II ligands [41]. Building upon these observations, we next sought to determine in human ccRCC whether *HIF2A* (also known as *EPAS1*) expression correlates with expression of MHC class I and class II genes, as well as other genes involved in antigen processing and presenting. Analyses of TCGA mRNA expression data revealed that *HIF2A*, but not *HIF1A*, is more highly abundant in ccRCC in comparison to normal kidney and that *HIF2A* shows a wide distribution of expression levels amongst tumours (Fig. 5f,g). This upregulation and wide expression level distribution is not observed in chromophobe RCC or papillary RCC (Fig. 5f,g). The wide expression distribution provided a good basis to investigate potential correlations between *HIF2A* mRNA abundance and the abundance of mRNAs involved in antigen presentation. Examples of correlations of MHC class I (Fig. 5h,i), MHC class II (Fig. j,k) and antigen processing and presentation (Fig. 5l,m) genes are shown. The Spearman correlation analyses of the full lists of these classes of genes are provided in Supplementary Data 4. Consistent with our mouse tumour data, Spearman correlation analyses revealed statistically significant negative correlations between *HIF2A*

expression and the expression of 3 of 6 MHC class I genes, 12 of 15 MHC class II genes and 26 of 51 non-MHC genes from the GO Antigen Processing and Presentation gene set. While highly statistically significant, the relatively small magnitudes of some of these correlations suggest that *HIF2A* expression levels may be one of several factors that influence the overall expression of antigen presenting genes in ccRCC. Collectively the mice and human data suggest that HIF-2α suppresses antigen presentation.

HIF-1α and HIF-2α alter the ccRCC immune microenvironment

Since diverse and complex inflammatory signatures were observed in the transcriptomes and proteomes of ccRCC tumours, we next applied three different bioinformatic methods to our RNA sequencing data to attempt to further deconvolve the relative abundance of different types of immune cells or specific gene signatures associated with inflammation in the different tumour genotypes. We applied two methods that were previously used to deconvolve the immune microenvironment of human ccRCC [42,43]. These methods are based on single sample gene set enrichment analyses (ssGSEA) using the mouse orthologues of an expanded number of immune-specific gene signatures to those initially described by Bindea et al. [44] , which we term Bindea&Others (described in [43]), and a set of gene signatures that were identified by analyses of human ccRCC tumourgrafts, termed eTME (described in [42]). The genes in these signatures and their overlap are listed in Supplementary Data 5. The third method is CIBERSORT with the mouse specific ImmuCC gene panel, which uses a matrix-weighted score, based on both high and low expressed genes in each immune subset, to assess the relative abundance of each immune cell population [45]. For all three methods we generated z-scores representing the degree of enrichment of each signature and statistically compared the immune infiltration scores of all tumours of a given genotype in a series of pairwise comparisons to WT cortex and to the other genotypes (Fig. 6a). In general, comparisons of all three tumour genotypes to WT cortex revealed enrichment of gene sets associated with myeloid cell inflammation, including dendritic cells, monocytes, macrophages and neutrophils, but not mast cells or eosinophils. Terms associated with different types of T cells and B cells revealed inconsistent results, varying depending on the deconvolution method used. $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours showed greater enrichment of a number of different immune cell signatures when compared to $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours and to $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ tumours, including several types of T cells, monocytes and macrophages, and notably for interferon-γ signalling (REACTOME.IFNG), consistent with the previous GAGE analyses. $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ tumours showed statistically lower average z-scores, or trends to lower scores, for myeloid cell signatures than $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours.

To further characterise the immune microenvironments of the three different tumour genotypes, we next conducted immunohistochemical stainings for a series of markers of different types of immune cells to permit analyses of a larger set of tumours of each genotype (n = 14-26 tumours). We stained sections of whole tumour-bearing kidneys with antibodies against CD3 to label T cells, CD4 to label helper T cells, CD8 to label effector T cells, CD69 as an early activation marker of T cells and NK cells, Perforin to label activated cytotoxic T cells and NK cells, PD-1 to label antigen-exposed activated or exhausted T-cells, B220 to label B cells, CD68 to label monocytes and macrophages, F4/80 to label differentiated macrophages and Ly-6G to label granulocytes and neutrophils. These markers revealed considerable inter-tumoural heterogeneity in terms of the density of infiltrating cells, even within the same kidney (Supplementary Fig. 11a-j). We quantified the densities of positively stained cells either by manual counting, via automated detection and

quantification algorithms, or we calculated the average relative staining intensity for F4/80 where it was not possible to identify individual cells in the network of macrophages, within the tumours as well as in unaffected regions of kidney tissue (normal) within the same mouse (Supplementary Fig. 11k,l). Consistent with HIF-2α deficient tumours showing the highest GAGE mRNA signatures of T cell inflammation, $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours displayed increased densities of CD3 (Fig. 6b), CD4 (Fig. 6c) and CD8 (Fig. 6d) positive T cells compared to normal tissue, whereas only CD8 positive T cell densities were significantly increased in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ tumours compared to the respective normal tissues. Notably, both $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours exhibited higher densities of CD8 positive T cells than $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours, in line with the Bindea&Others ssGSEA CD8 T cell signature results. $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ but not $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours showed increased densities of CD4 positive cells compared to $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours. This result is not reflected by any of the bioinformatic immune deconvolution methods. Analyses of the T cell activation markers CD69 (Fig. 6e) and Perforin (Fig. 6f) revealed that all tumours showed increased T cell activation compared to normal tissue, and that $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours showed higher levels of T cell activation than $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ or $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ tumours, consistent with our conclusions from the GAGE analyses. There was no statistically significant enrichment of PD-1 positive cells, a marker of exhausted T cells, in any of the tumour genotypes (Fig. 6g). B220 staining revealed increased B cell density in all tumour genotypes compared to normal tissue, and higher densities in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ tumours than in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ or $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours (Fig. 6h). Interestingly, these observations do not reflect the results of the bioinformatic immune cell deconvolutions for B cells. In contrast to the relatively low numbers of tumour-infiltrating lymphocytes, myeloid lineage cells were much more abundant within tumours. CD68 positive monocytes/macrophages (Fig. 6i), F4/80 positive macrophages (Fig. 6j) and Ly-6G-labelled granulocytes/neutrophils (Fig. 6k) were highly abundant in tumours compared to normal tissue but there were no differences in abundance of these cells between tumour genotypes.

Given that HIF-1α and HIF-2α deficiencies increase CD8 positive T cell infiltration, we first sought to gain insight into potential mechanisms that might explain these observations by examining our RNA sequencing data. We previously demonstrated that $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours show upregulation of numerous cytokines in comparison to WT cortex [16], however analyses of the new dataset did not reveal any cytokines that were specifically altered by the absence of HIF-1α or HIF-2α (Supplementary Fig. 12a) that might be expected to influence T cell infiltration or activation. To functionally investigate whether signalling molecules or metabolic factors that are released by mouse or human ccRCC cells might directly influence the proliferation of $CD8^+$ T cells, we activated mouse splenic CD8+ T cells and incubated them for 3 days in conditioned medium from mouse 2020 ccRCC cells (including VHL30 rescue and *Hif1a* knockdown cells), human renal proximal tubule epithelial cells (RPTEC), human A498 ccRCC cells or human 786-O (including VHL 30 rescue) ccRCC cells, compared to non-conditioned medium. However, none of the conditioned media altered $CD8^+$ T cell proliferation (Supplementary Fig. 12c,d), arguing against a direct, soluble factor-mediated, VHL- or HIF-α-dependent cross-talk between ccRCC and T cells as being the mechanism that underlies the altered immune microenvironment in the $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ and $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours.

To investigate whether genetic alterations in *HIF1A* or *HIF2A* might influence immune cell infiltration in human ccRCC, we analysed data from the TCGA KIRC study (Firehose-legacy data set) [9] using cBioPortal [46,47]. ccRCC tumours frequently lose one copy of *HIF1A* and less frequently gain one copy of *HIF2A* (Fig. 7a). Loss of one copy of either gene correlated with lower mRNA levels but gain of a copy did not correlate with increased mRNA abundance (Supplementary Fig. 13a,b). ccRCC tumours that exhibit mono- or bi-allelic loss of *HIF1A* (collectively *HIF1A* loss) show worse overall survival (Fig. 7b) and progression-free survival (Supplementary Fig. 13c) than unaffected tumours, whereas there are no overall or progression-free survival differences between tumours with a copy number gain of *HIF2A* and unaffected tumours (Fig. 7c and Supplementary Fig 13d). Previous studies have identified that loss of larger regions of chromosome 14q, including the *HIF1A* gene, correlates with poor prognosis [19]. We took advantage of the extensive clinical and whole exome sequencing data of the TCGA dataset to investigate whether co-variants of *HIF1A* loss may account for the observed survival differences. Tumours with *HIF1A* loss were statistically more likely to have higher grade and stage, and display lymph node positivity and metastases (Supplementary Table 1), consistent with this subgroup representing a more aggressive form of ccRCC. The only mutation that occurred more frequently in the *HIF1A* loss subgroup than the unaltered subgroup was *BAP1*, which was detected in 9% of all ccRCC tumours (Supplementary Fig. 13e,f). However, *BAP1* mutation status alone did not significantly affect survival (Supplementary Fig. 13g) in this cohort and removal of all *BAP1* mutant tumours from the cohort did not alter the correlation of *HIF1A* loss with poor prognosis (Supplementary Fig. 13h,i). The conclusion that *BAP1* mutation status is not a relevant co-variant that affects survival outcome in the *HIF1A* loss cohort was also demonstrated by COX univariate (HR 1.839, 95% CI 1.332-2.539, P = 0.00021) and multivariate proportional hazards analyses (HR 1.776, 95% CI 1.274-2.474, P = 0.00069). These findings suggest that loss of one allele of *HIF1A*, which is predicted to lead to diminished HIF-1α abundance, may be selected for during the evolution of some ccRCC tumours and that this correlates with aggressive disease.

To investigate whether *HIF1A* loss correlates with altered inflammation, we first demonstrated that *HIF1A* loss tumours exhibit on average between 1.9- and 2.1-fold higher levels of mRNA of *CD3D*, *CD3E*, *CD8A* and *CD8B* and 1.4-fold higher levels of *CD4* than unaltered tumours (Fig. 7d,f,h,j,l), suggesting that this group of tumours has higher $CD8^+$, and to a lesser extent $CD4^+$, T cell infiltration. This observation is consistent with the mouse analyses in which *Hif1a*-deficient tumours on average display approximately double the number of $CD8^+$ and $CD4^+$ T cells. In contrast, *HIF2A* gain tumours show no differences in the expression of any of these T cell marker genes compared to unaltered tumours (Fig. 7e,g,l,k,m). To gain a more in-depth overview of the effects of *HIF* gene copy number or expression level alterations on the immune microenvironment, we performed immune deconvolution analyses of RNA-seq data, again using three independent methods of immune cell deconvolution; ssGSEA using the Bindea&Others and eTME gene signatures and using the CIBERSORT method [48]. We compared *HIF1A* loss and *HIF2A* gain tumours to diploid human ccRCCs (Fig. 7n) and also took advantage of the wide distribution of mRNA expression levels of *HIF2A* to compare tumours in the top (Q4) and bottom (Q1) quartiles of *HIF2A* mRNA abundance. While *HIF2A* gain tumours exhibited very few alterations in immune scores, *HIF1A* loss tumours showed statistically significant increases or decreases in 57 of 83 immune signatures of a variety of lymphoid and myeloid lineage cells. Notable amongst these are consistently upregulated scores for T helper cells and for B cells, mirroring our immunohistochemical findings of the comparison of $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ and

534 $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ tumours. It should however be noted that the magnitude of the z-scores are
535 generally low, suggesting that this group of ccRCCs on average exhibits numerous subtle differences in
536 immune inflammation compared to tumours with normal *HIF1A* copy number. In contrast, high *HIF2A* mRNA
537 expressing tumours displayed downregulation of scores for interferon-γ and for APM2 (measuring MHC class
538 II antigen presentation), consistent with the upregulation of these features in mouse ccRCC tumours lacking
539 HIF-2α. Somewhat paradoxically, high *HIF2A* expressing tumours also displayed general upregulation of
540 CD8 T cell scores and some NK cell scores and downregulation of all three scores for regulatory T cells and
541 for the immune checkpoint proteins PD1 and CTLA4. These scores might be predicted to reflect elevated
542 CD8 T cell activity. However, these tumours also display elevated scores for monocytes, neutrophils and
543 mast cells, which in some settings contribute to suppression of anti-tumour CD8 T cell responses.
544 Immunosuppressive mast cells were shown to correlate with *HIF2A* mRNA abundance in human ccRCC [49].
545 Thus, the extent of T cell mediated anti-tumour immunity is likely to be determined by the balance of the
546 abundance and activities of several different immune cell types in a manner that is partly influenced by
547 *HIF2A* expression. Finally, it is also noteworthy that *HIF1A* copy loss and *HIF2A* mRNA high tumours
548 showed opposite effects on signatures for pericytes, endothelial cells and angiogenesis, implying that both
549 HIF-1α and HIF-2α may act as positive factors that promote blood vessel formation in ccRCC tumours.
550
551 Collectively, the mouse and human analyses demonstrate that the genetic copy number or expression level
552 status of the *HIF* genes in ccRCC tumour cells correlate with the composition and activation state of the
553 innate and adaptive immune systems in the tumour microenvironment.
554
555
556
557
558
559

DISCUSSION

Herein we show by introducing floxed alleles of *Hif1a* or *Hif2a* into an autochthonous mouse model of ccRCC that HIF-1α is necessary for tumour formation whereas HIF-2α deficiency has only a moderate effect on tumour initiation and growth. While it cannot be excluded that there might be differences between mice and humans, and the findings may be contextual to the *Vhl/Trp53/Rb1* mutant background, this result seemingly contrasts with several independent lines of evidence from the study of human ccRCC tumours and ccRCC cell lines, which have demonstrated that HIF-2α possesses strong oncogenic activity and HIF-1α acts in the manner of a tumour suppressor to suppress aggressive tumour behaviour. How can these apparent discrepancies be reconciled? There are several possible explanations which suggest that these observations may not in fact be discrepancies. Additionally, our current study also suggests that there may be caveats to the interpretation of previous studies, arguing against an oversimplified, binary oncogene/tumour suppressor model of the contributions of these proteins to ccRCC development. We argue that HIF-1α and HIF-2α are likely to play different roles at different stages of tumour formation and progression and that these roles might potentially be modulated by the spectrum of mutations present in each individual ccRCC. We further argue that it is possible that the balance of the relative strengths of the activities of the different HIF-α proteins might be dynamically modulated via different mechanisms throughout the lifetime of a ccRCC tumour to tailor their combined transcriptional outputs to provide an evolutionary advantage at the given stage of the tumour.

In addition to genetic mechanisms that can irreversibly affect the balance of HIF-1α and HIF-2α activities, such as loss of one copy of chromosome 14q, encoding *HIF1A*, or intragenic deletions of *HIF1A* [22], several mechanisms exist that potentially provide tumour cells with a more dynamic mode of fine-tuning the relative strengths of HIF-1α and HIF-2α expression and activities. These include mutual epigenetic suppression [50], mutual suppression of protein levels [23], HIF-2α-mediated suppression of HIF-1α translation [51] and HAF-mediated ubiquitination and degradation of HIF-1α to promote the switch from HIF-1α towards HIF-2α activity [52]. Analyses of expression patterns in human ccRCC support the notion that HIF-1α plays an oncogenic role at early and late stages of ccRCC development and progression. In VHL patients, the earliest *VHL*-null multi-cellular renal tubule lesions tend to strongly express HIF-1α and weakly express HIF-2α, but later lesions such as cysts and tumours express HIF-1α as well as higher levels of HIF-2α [23,53]. While HIF-2α protein is present in the vast majority of all sporadic ccRCC tumours, HIF-1α protein expression is not detected in about 30% of cases and this correlates with increased tumour cell proliferation [21]. Nonetheless, HIF-1α is detected in about 70% of ccRCC tumours and several studies have correlated higher HIF-1α expression levels with poor patient survival (reviewed in [54]). Consistent with our current findings of an obligate oncogenic role of HIF-1α, transgenic overexpression of constitutively stabilised HIF-1α [55], but not HIF-2α [56] in mouse renal tubules causes the formation of small lesions that have some features of precursor lesions of ccRCC. Interestingly, combined overexpression of HIF-1α and HIF-2α did not cause a more severe phenotype than HIF-1α overexpression alone [56], demonstrating that even the combined actions of both HIF-1α or HIF-2α are insufficient to induce tumour formation. In agreement with this conclusion, numerous previous studies showed that deletion of *Vhl* in mouse renal epithelial cells *in vivo* [10,57–62], resulting in abrogation of the many different tumour suppressor functions of pVHL [2,7,63], was insufficient to cause tumour initiation either when both HIF-1α and HIF-2α were stabilised, or when the balance of HIF-1α and

HIF-2α activities were genetically altered by co-deletion of *Hif1a* or *Hif2a* [29]. Nonetheless, deletion of either *Hif1a* or *Hif2a* was sufficient to inhibit the formation of cysts and tumours induced by *Vhl/Trp53* double mutation [29], demonstrating that both HIF-1α and HIF-2α have pro-tumourigenic activities. In this context, it is noteworthy that we now show that *Hif2a* deletion fails to strongly inhibit tumour formation in the *Vhl/Trp53/Rb1* deletion model. One explanation for these different findings may relate to our RNA sequencing observations which highlighted that HIF-2α increases expression of MYC and E2F target genes, consistent with previous findings that HIF-2α stimulates MYC activity [21]. This predicted cell-cycle promoting activity of HIF-2α is likely to be necessary for tumour formation in the *Vhl/Trp53* background, but may become at least partly redundant due to the additional mutation of *Rb1*, which promotes the cell cycle by removing the negative regulation of E2F transcription factors.

The patterns of copy number alterations that arise in the TGCA ccRCC dataset are consistent with the idea that the balance of HIF-1α and HIF-2α activities may be differently selected for, or tolerated, depending on the status of the network of G1-S cell cycle controlling genes. We have previously described a copy number signature in which multiple genes involved in the p53 pathway and G1-S checkpoint are altered in about two thirds of ccRCC tumours [16]. While *HIF1A* copy number losses or mutations (45%) occur in ccRCC more frequently than gains (2.5%), *HIF2A* losses are very rare (4%) and gains more common (15%) (Supplementary Fig. 14a). Importantly, of the rare tumours with *HIF2A* losses or mutations, only 2 of 16 did not exhibit a copy number alteration in one or more genes that control the G1-S checkpoint (Supplementary Fig. 14b). Thus, rare genetic events that might be predicted to result in lowered HIF-2α activity almost always arise in the background of genetic alterations that are predicted to abrogate or weaken normal cell cycle control. We speculate that this may allow the tumours to grow in the absence of a lowered cell cycle promoting activity of HIF-2α. In contrast, the distribution of copy number losses of *HIF1A* does not correlate with the presence or absence of alterations in the G1-S network: 52 tumours harbouring *HIF1A* copy loss showed no alteration in G1-S genes, while 146 tumours with *HIF1A* copy loss did display alterations (Supplementary Fig. 14c).

Since our current and previous mouse genetic studies collectively show that the requirement for HIF-1α and HIF-2α is dependent on the underlying genotype of the tumour (*Vhl/Trp53* mutations versus *Vhl/Trp53/Rb1* mutations), it will be important to test the generality of our findings in other genetic mouse models of ccRCC, particularly as epigenetic modifications resulting from mutations in *Pbrm1* have been shown to alter HIF-α transcriptional outputs [12,64,65]. It is likely that other ccRCC-relevant epigenetic tumour suppressors such as BAP1, SETD2 and KDM5C also influence HIF-α transcriptional outputs. Thus, the combinations of mutations present in ccRCC cells might represent another mechanism for altering the balance of the relative activities of HIF-1α and HIF-2α, potentially affecting the genetic dependency on the two HIF-α proteins.

Our present study also highlights a potential general caveat to the interpretation of the results of studies of human ccRCC cell lines, or patient-derived tumour lines, in cell culture and in xenograft assays. In contrast to the clear requirement for *Hif1a* in tumour formation in the autochthonous setting, we find that HIF-1α rather exhibits putative tumour suppressor activities in cell culture-based assays, including inducing the early loss of proliferative capacity of MEFs following *Vhl* deletion, inhibiting the proliferation of immortalised

Vhl/Trp53 null MEFs and inhibiting anchorage independent growth of a *Vhl/Trp53/Rb1* mutant mouse ccRCC cell line. Furthermore, HIF-1α knockdown did not affect the growth of allograft tumours generated with this ccRCC cell line. This latter experiment argues against the idea that removal of HIF-1α activity in an established mouse ccRCC tumour, to mimic the loss of HIF-1α function that arises as a later event in tumour progression in a subset of human ccRCCs, has a potent effect on tumour aggressiveness. Collectively, our studies show that in the same genetic ccRCC tumour model, the cell culture and allograft assays do not reflect the *in vivo* requirement in the autochthonous setting. Previous studies which have interpreted the oncogenic and tumour suppressive roles of HIF-α factors in ccRCC based on studies of cultured ccRCC cells and xenograft models may therefore not necessarily be reflective of the true *in vivo* functions of HIF-1α and HIF-2α at different stages of tumour development in the physiological context of the kidney.

Our transcriptomic and proteomic analyses suggest that one potential oncogenic contribution of HIF-1α to tumour initiation is the induction of expression of genes that induce Warburg metabolism of high rates of conversion of glucose to lactate and low rates of pyruvate entry into mitochondrial respiration. These findings of HIF-1α dependency are in accordance with numerous studies in mice and humans [29,66–68]. This mode of Warburg tumour metabolism has recently been proven to occur *in vivo* in human ccRCC tumours [69] and is in stark contrast to the normal mode of metabolism of renal proximal tubule epithelial cells, the cell of origin of ccRCC, which generate ATP predominantly through mitochondrial oxidation of substrates other than glucose, allowing them to fuel solute transport, including the transport of glucose, from the tubular fluid back to the blood stream [70,71]. It remains to be functionally elucidated if and how this shift to glucose metabolism contributes to ccRCC formation by proximal tubule cells and whether this represents a true driver of tumour formation or acts as an enabler of tumour formation once additional transforming genetic mutations arise. Previously published mouse genetic data suggest that the latter is likely to be the case as induction of Warburg metabolism in mice through *Vhl* deletion alone did not cause tumour formation. We identified that both HIF-1α and HIF-2α are necessary for the clear cell phenotype of *Vhl/Trp53/Rb1*-deficient ccRCCs, implying that they might differently contribute to the accumulation of cytoplasmic glycogen and lipids. HIF-1α-dependent alterations in glucose and glycogen utilisation and reduction of mitochondrial abundance and activity [29] might lead to accumulation of glycogen and failure to efficiently metabolise fatty acids. We also identified a HIF-2α-dependent metabolic signature involving genes involved in cholesterol and lipoprotein uptake and metabolism, which may account for the requirement of HIF-2α in the clear cell phenotype.

A major advantage of the study of autochthonous models of ccRCC is that the tumours develop in the presence of a functioning immune system. In general, the activities of cytotoxic innate and adaptive immune cells can act to inhibit or delay tumour formation and progression, but an inflamed tumour microenvironment can also promote tumour formation through a variety of mechanisms [72]. The immune microenvironment of human ccRCC appears to be unusual in some respects as higher levels of T cell infiltration have been shown to correlate with disease recurrence and worse survival [43,73,74]. In almost all other tumour types the degree of T cell inflammation is a good prognostic factor [14]. ccRCC tumours are also frequently inflamed with immature myeloid lineage cells that at least partly reflect different states of differentiation along the path from myeloid progenitor to differentiated dendritic cells, macrophages and neutrophils. At least some of these myeloid lineage cells have been proposed to promote tumour formation by suppressing T cell activation [75–77].

Our bioinformatic immune deconvolution and immunohistochemical analyses of the immune microenvironment of mouse ccRCC tumours is broadly consistent with this latter observation, namely that the immune microenvironment is dominated by myeloid lineage cells with a modest degree of T cell inflammation and activation. An unexpected finding was that tumours with *Hif2a* deletion showed stronger mRNA signatures associated with tumour immune cell infiltration, antigen presentation and interferon activity, as well as higher densities of CD8 T cells and cells expressing the T cell activation markers CD69 and Perforin, compared to the other two tumour genotypes. Furthermore, MHC class I and II genes, as well as other genes involved in antigen processing and presentation, were upregulated and tumour cells in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours were more frequently immunohistochemically positive for MHC class II expression, suggestive of increased presentation of intracellular and extracellular epitopes in this tumour genotype. Collectively, these observations reflect a generally higher degree of immune activity directed against *Hif2a*-deficient tumour cells and it is plausible that this immune control may at least partly contribute to the reduced numbers of tumours and delayed tumour onset seen in this genotype. Interestingly, we demonstrated not only in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif2a^{\Delta/\Delta}$ tumours but also in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}Hif1a^{\Delta/\Delta}$ tumours that the density of intra-tumoural $CD8^+$ T cells were approximately double the density in $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ tumours. The densities of $CD8^+$ T cells observed in the mouse ccRCC tumours (approximately 40-80 cells/mm^2) are within the range seen in most cases of human ccRCC (approximately 20-160 cells/mm^2) [49], providing support for the relevance of our mouse model. Consistent with our findings of HIF-2α acting as a suppressor of T cell inflammation in mouse ccRCC, it was shown that HIF-2α expression levels in human ccRCCs anti-correlate with T cell abundance and markers of T cell activation [49]. We speculate that this apparent suppression of anti-tumour immune responses by HIF-2α might reflect a mechanism of positive selection that maintains HIF-2α expression in ccRCCs. Moreover, we show that loss of one copy of *HIF1A* correlates with a worse survival outcome, higher mRNA signatures of T cell abundance and a broadly altered immune microenvironment in human ccRCCs. This observation is consistent with the fact that higher levels of T cell infiltration have been shown to correlate with disease recurrence and worse survival in ccRCC [43,73,74].

Thus, genetic and likely transcriptional and translational mechanisms that alter the balance of HIF-1α and HIF-2α abundance and activities appear to affect T cell inflammation. The mechanisms that underlie the increased T cell infiltration and/or activity in the absence of HIF-1α or HIF-2α will require further study. Analyses of RNA sequencing revealed that the mRNA levels of many cytokine-encoding genes are upregulated in tumours compared to normal cortex, but none of these genes were differentially regulated by HIF-1α or HIF-2α. Our functional tests to investigate if T cell proliferation might be suppressed by ccRCC cells in general, for example through the secretion of immunosuppressive glycolytic metabolic products such as lactate or H^+ [78], did not reveal any crosstalk between mouse or human ccRCC cells and mouse $CD8^+$ T cells under cell culture conditions. It is possible that these simplified assays failed to reproduce the metabolic conditions that are present *in vivo*. It is also likely that many other complex factors such as the presence and activation states of other immune microenvironmental cells (such as macrophages, MDSCs, dendritic cells, regulatory T cells), antigen presentation and presence or absence of co-activating or inhibitory ligands by ccRCC cells, as well as the composition of the extracellular matrix might all play a role in the trafficking and activation of T cells. Indeed, immune deconvolution analyses revealed that the group of tumours that exhibit loss of one copy of *HIF1A* or that show high levels of mRNA expression of *HIF2A* show many differences in

signatures of different types of cells in the immune microenvironment. Since our own studies comparing multiple bioinformatic immune cell deconvolution methods showed that different methods and different signatures can lead to different results for the same immune cell type, as well as the fact that we observed both consistencies and inconsistencies between bioinformatic predictions and direct immune cell enumeration using immunohistochemistry, it will be important to treat these bioinformatic observations as hypothesis-generating starting points that will need to be orthogonally tested by staining for specific immune cell markers in larger cohorts of human ccRCC tumours. Given the clinical importance of immune checkpoint-based therapies for ccRCC and the fact that not all patients respond to these regimes, we believe that further investigation of the relationship between HIF-1α and HIF-2α status and the immune microenvironment is of potential therapeutic relevance. A further corollary of our findings is that inhibition of HIF-1α and HIF-2α transcription factor activities in ccRCC cells could be investigated therapeutically to inhibit tumour cell proliferation and simultaneously to attempt to increase T cell infiltration and activation. The potential direct effects of pharmacological inhibition of HIF-α factors on different immune cells would also have to be considered in this strategy. Finally, while specific inhibitors of HIF-2α are available and are currently being tested in clinical trials [25–27], our findings demonstrating the importance of HIF-1α for ccRCC formation argue that the development of specific inhibitors of HIF-1α or of new specific dual HIF-1α/HIF-2α inhibitors would also be desirable and may have therapeutic benefit in ccRCC. Proof-of-principle that these approaches are likely to be tolerable and effective comes from previously reported therapeutic effects of Acriflavine, which inhibits the binding of both HIF-1α and HIF-2α to HIF-1β [79,80], in the $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ ccRCC model [16], as well as in xenograft and autochthonous mouse models of several different types of tumours [79,81–83].

METHODS

Mice

Previously-described Ksp1.3-CreER^{T2};*Vhl$^{fl/fl}$;Trp53$^{fl/fl}$;Rb1$^{fl/fl}$* mice [16] were intercrossed with previously-described Ksp1.3-CreER^{T2};*Vhl$^{fl/fl}$;Trp53$^{fl/fl}$;Hif1a$^{fl/fl}$* and Ksp1.3-CreER^{T2};*Vhl$^{fl/fl}$;Trp53$^{fl/fl}$;Hif2a$^{fl/fl}$* mice [29] to generate the experimental Ksp1.3-CreER^{T2};*Vhl$^{fl/fl}$;Trp53$^{fl/fl}$;Rb1$^{fl/fl}$*, Ksp1.3-CreER^{T2};*Vhl$^{fl/fl}$;Trp53$^{fl/fl}$;Rb1$^{fl/fl}$Hif1a$^{fl/fl}$* and Ksp1.3-CreER^{T2};*Vhl$^{fl/fl}$;Trp53$^{fl/fl}$;Rb1$^{fl/fl}$Hif2a$^{fl/fl}$* mouse lines. Littermate mice that lacked the Cre transgene served as wild type controls. Gene deletion in 6 week-old mice was achieved by feeding with food containing tamoxifen (400 parts per million) for 2 weeks. Mouse crosses and phenotyping were conducted under the breeding license of the Laboratory Animal Services Center, University of Zurich and tumour monitoring studies were conducted under license ZH116/16 of the Canton of Zurich. Investigators were not blinded to the genotype of the mice.

µCT imaging

Monitoring of tumour growth in mice was performed on a monthly basis using µCT as previously described [16]. Tumour size was assessed by measuring the maximum diameter of all three dimensions in the respective planes (x, y and z-plane). The volumes were then calculated using the mathematical formula of an ellipsoid:

(1) $V = \frac{4}{3} * \pi * radius\,(x) * radius\,(y) * radius\,(z)$

Generation of MEFs and mouse ccRCC cell line

Isolation of MEF lines [29] and preparation of primary renal epithelial cells [15] have been previously described. MEFs were infected with adenovirus expressing Cre recombinase and GFP (Ad-Cre-GFP; Vector Biolabs; #1700) or GFP only (Ad-CMV-GFP; Vector Biolabs; #1060). The mouse ccRCC cell line 2020 was isolated from a piece of tumour tissue from a Vhl$^{\Delta/\Delta}$Trp53$^{\Delta/\Delta}$Rb1$^{\Delta/\Delta}$ mouse, minced with a scalpel blade and digested for 70 minutes at 37°C with 1 mg/ml Collagenase II solution in 1x HBSS. The digestion was inactivated with 20 ml K1 medium (Dulbecco's modified Eagle's medium (DMEM) and Hams F12 mixed 1:1, 2 mM glutamine, 10 kU/ml penicillin, 10 mg/ml streptomycin, hormone mix (5 µg/ml insulin, 1.25 ng/ml prostaglandin E_1 (PGE_1), 34 pg/ml triiodothyronine (T3), 5µg/ml transferrin, 1.73 ng/ml sodium selenite, 18 ng/ml of hydrocortisone and 25 ng/ml epidermal growth factor (EGF)) + 10% FCS. The cell solution was subsequentially filtered through a 70-µm cell strainer, pelleted, and plated in K-1 medium + 10% FCS in a humidified 5% (v/v) CO_2 and 20% O_2 incubator at 37°C. Medium was changed 48 hours after plating. Cells were split 1:5 when sub confluent.

Retroviral and lentiviral infections

Retroviral and lentiviral infections and cell selection were carried out as previously described [84]. Cells were infected with the retroviruses pBabe-PURO (Vector) or pBabe-PURO-VHL30 (VHL30) or the lentiviruses LKO.1 expressing non-silencing control shRNA (shRNA-ns), or shRNAs against *Hif1a* (TRCN0000232220, TRCN0000232222 or TRCN0000232223), respectively termed shRNA-*Hif1a* #220, shRNA-*Hif1a* #222 and shRNA-*Hif1a* #223.

MEF and ccRCC cell proliferation assays

3T3 proliferation assays of MEFs [15] have been previously described. 3,000 mouse ccRCC 2020 cells per well were seeded in 96 well plates in 6 replicates and incubated for six days. Cells were cultivated in K-1

medium + 10% FCS, medium was changed after 3 days of incubation. Cell proliferation was measured using a sulforhodamine B (SRB) colorimetric assay [15]. After the indicated time points the cells were fixed with 10 % (w/v) trichloroacetic acid and stained with 0.057 % (w/v) SRB solution. 10 mM Tris base solution (pH 10.5) was used to solubilize SRB, followed by OD measurement at 510 nm in a microplate reader (Tecan Spark 10M plate reader). For sphere-forming assays, cell suspensions were filtered through a 40-µm cell strainer and 1,000 cells were seeded in 6 well low attachment plates (Corning). The cells were cultivated in K-1 medium + 10% FCS. Every three days fresh medium was added to the wells. After 14 days, microscopy pictures of the formed spheres were captured with DKM 23U274 camera connected to Eclipse Ts2R-FL microscope (Nikon) at 20x magnification. Images were acquired with IC Capture 2.4 software and analysed using ImageJ software.

Allograft assay

Single cell suspensions were prepared with Accutase (Gibco) and $5x10^6$ cells were re-suspended in 75 µl RPMI following transfer into a precooled 30G insulin syringe mixed with 75 µl Matrigel (Corning). Syringes with cell suspension were kept on ice to avoid hardening of the Matrigel. SCID Beige mice (Charles River Laboratories) were anesthetised by inhalation of 3 % isoflurane using oxygen as carrier gas. Mice were shaved and cells were injected subcutaneously in the flank. Tumour volumes were measured weekly with a calliper. Experiments were conducted under license G-17/165 of the Regierungspräsidium Freiburg.

Conditioned medium T cell proliferation assay

10,000 mouse ccRCC 2020, human RPTEC (from Dr. Jiing-Kuan Yee), 786-O (ATCC) or A498 (ATCC) ccRCC cells were seeded in triplicates in a 6-well-plate with 2 ml RPMI + 10% FCS and kept incubated in a humidified 5% (v/v) CO_2 and 20% O_2 incubator at 37°C for two days. Two days later, spleens from C57BL/6 mice were extracted, washed in PBS and mashed through a 100 µm cell strainer in MACS buffer (PBS 1x + 2 % FCS + 2 mM EDTA). The mashed spleen was filtered again through a 100 µm cell strainer into a 50 ml conical tube and centrifugated for 10 minutes at 290g. The pellet was labelled manually with magnetic CD8a (Ly-2) MicroBeads (Miltenyi Biotech). Isolated $CD8a^+$ T cells were centrifuged, resuspended in proliferation medium (RPMI + 10% FCS supplemented with 25 µM β-Mercaptoethanol) and counted. $CD8a^+$ T cells were then stained with the CellTrace Violet Proliferation Dye (Thermo Fisher). Stained $CD8a^+$ T cells were stimulated with CD3/CD28 Dynabeads (Thermo Fisher) and activated with interleukin-2 (IL-2). The conditioned medium was distributed into fresh 6-well-plates and $2x10^5$ of stained, stimulated and activated $CD8a^+$ T cells were added. The mix of conditioned medium and T-cells was incubated for three days in a humidified 5% (v/v) CO_2 and 20% O_2 incubator at 37°C. After the incubation time the T cells were resuspended and centrifuged in a 2 ml reaction tube for 5 minutes at 1600 rpm and 4 °C. The dead cells within the pellet were stained with the Live/Dead Fixable Aqua Dead Cell Stain Kit (Thermo Fisher), washed with 200 µl MACS buffer and centrifuged for 5 minutes at 515g and 4 °C, 25 µl CD16/32 antibody (Fisher Scientific, 14016185, diluted 1:25 in MACS buffer) was added to the pellet to block Fc-mediated reactions. After 10 minutes of incubation at 4 °C in the dark, 25 µl of CD8a antibody (APC-conjugated, Biolegend, 100712, diluted 1:100 in MACS buffer) was added to the suspension and incubated for 30 minutes at 4 °C in the dark. Afterwards T cells were washed twice with MACS buffer and the pellet was resuspended in 100 µl MACS buffer. Via flowcytometry (BD LSRFortessa) dead/living cells were measured with a 405 nm Extinction Laser (AmCyan), T cells were measured with a 640 nm Extinction Laser (APC) and the

Proliferation Dye was measured with a 405 nm Extinction Laser (Pacific Blue). Data was gated and analysed using FlowJo Software.

Immunohistochemistry

Immunohistochemical stainings were performed as previously described[10]. Primary antibodies against the following proteins or epitopes were used at the following dilutions and antigen retrieval conditions: B220 (1:3,000, BD Biosciences, 553084, Tris/EDTA 20 min, 100°C), CA9 (1:2,000, Invitrogen, PA1-16592, citrate, 10 min, 110°C), CD3 (1:250, Zytomed, RBK024, citrate 30 min, 95°C), CD4 (1:1,000, eBioscience, 14-9766, citrate, 30 min, 100°C), CD8a (1:200, Invitrogen, 14-0808-82, citrate buffer, 15 min, 114°C), CD10 (1:2,000, Thermo Fisher Scientific, PA5-47075, citrate buffer, 10 min, 110°C), CD68 (1:100, abcam ab125212, citrate buffer, 30 min, 95°C), CD69 (1:1,000, Bioss, bs-2499R, Tris/EDTA, 15 min, 114 °C), F4/80 (1:250, Linaris Biologische Produkte, T-2006, BOND Enzyme Pretreatment Kit (Leica AR9551), 10 min, 37 °C), HIF-1α (1:20,000, Novus Biotechnologies, NB-100-105, citrate buffer, 10 min, 110°C, Catalyzed Signal Amplification Kit (DakoCytomation)), HIF-2α (1:1,000, abcam ab109616, Tris/EDTA 15 min, 114 °C), Ly-6G (1:800, BD, 551459), MHC II (1:500, Novus Biotechnologies, NBP1-43312, BOND Enzyme Pretreatment Kit (Leica AR9551), 10 min, 37 °C), PD-1 (1:100, R&D systems, AF1021, Tris/EDTA 20 min, 100°C), Perforin (1:100, Biorbyt, orb312827, Tris/EDTA, 15 min, 114 °C), phospho-Thr37/Thr46-4E-BP1 (1:800, Cell Signaling Technologies, 2855, citrate buffer, 10 min, 110°C). The following anti-HIF-2α antibodies did not provide specific nuclear signals in immunohistochemical staining using either Citrate or Tris/EDTA antigen retrieval methods: abcam ab199, Aviva Systems Biology ARP32253, Biorbyt orb96817, Sigma MAB3472, GeneTex GTX30114. For analyses of immune cell markers sections were scanned using a Nanozoomer Scansystem (Hamamatsu Photonics). Automatic quantifications of B220, CD3, CD4, CD8a and CD68 positive cells were carried out from duplicate stains (average values were determined) as previously described [85] using the VIS software suite (Visiopharm, Hoersholm, Denmark). Each tumor was outlined manually. Immune cell densities were calculated as cells per mm² based on surface area and immune cell quantification. The quantifications of cells stained with F4/80 was performed using a positive pixel count and presented as percentage positive pixel (%PP). PD-1, Perforin, Ly-6G and CD69 stains were quantified by manual annotation of positively stained cells.

Real-time PCR of mRNA and genomic DNA and recombination-specific genomic DNA PCR

RNA was isolated from powdered frozen samples using the NucleoSpin RNA kit (Machery Nagel), cDNA prepared using random hexamer primers and Ready-To-Go You-Prime First-Strand Beads (GE Healthcare). Real-time PCR was performed using the LightCycler 480 SYBR Green Master mix (Roche) using a LightCycler 480 (Roche). The following sets of primer pairs (sequences provided as 5′-3′) were used:

18s (fwd GTTCCGACCATAAACGATGCC, rev TGGTGGTGCCCTTCCGTCAAT)

Hif1a (fwd GGTTCCAGCAGACCCAGTTA, rev AGGCTCCTTGGATGAGCTTT)

Hif2a (fwd GAGGAAGGAGAAATCCCGTGA, rev CTGATGGCCAGGCGCATGATG)

Vhl (fwd CAGCTACCGAGGTCATCTTTG, rev CTGTCCATCGACATTGAGGGA)

Genomic DNA was isolated from powdered frozen samples using the GeneElute Mammalian Genomic DNA Miniprep kit (Sigma-Aldrich). 60ng genomic DNA per reaction was subjected to real time PCR using the LightCycler 480 SYBR Green Master mix (Roche) using a LightCycler 480 (Roche) and annealing

temperature of 55°C. The following sets of primer pairs (sequences provided as 5′-3′) were used to amplify the following floxed and non-floxed exons:

Vhl Exon 1 (floxed) (fwd ATAATGCCCCGGAAGGCAG, rev TGAGCCACAAAGGCAGCAC)
Vhl Exon 3 (not floxed) (fwd ACCCTGAAAGAGCGGTGCCTTC, rev CGCTGTATGTCCTTCCGCACAC
Hif1a Exon 2 (floxed) (fwd CGGCGAAGCAAAGAGTCTGAAG, rev CGGCATCCAGAAGTTTTCTCACAC)
Hif2 Exon 2 (floxed) (fwd GCTGAGGAAGGAGAAATCCCG, rev CTTATGTGTCCGAAGGAAGCTG)
Hif2a Exon 1 (not floxed) (fwd TGGCGTCTTACAACCTCCTCCC, TCCGAGAGTCCCGCTCAATCAG)
Trp53 Exon 4 (floxed) (fwd TGAAGCCCTCCGAGTGTCAG, rev AGCCCAGGTGGAAGCCATAG)
Trp53 Exon 11 (not floxed) (fwd AGAAGGGCCAGTCTACTTCCCG, rev AAAAGGCAGCAGAAGGGACCG)
Rb1 Exon 19 (floxed) (fwd AATACAGAGACACAAGCAGCC, rev GAGCCACAACTTAACCTAGTCC)

The following sets of primers were used for PCR amplification of DNA products that are specific to Cre-recombined alleles of the *Hif1a* [86] and *Hif2a* [87] genes.

Hif1a (Fwd II GCAGTTAAGAGCACTAGTTG, Rev GGAGCTATCTCTCTAGACC,)
Hif2a (P1 CAGGCAGTATGCCTGGCTAATTCCAGTT, P3 GCTAACACTGTACTGTCTGAAAGAGTAGC)

Western blotting

Antibodies against the following proteins were used for Western blotting: β-ACTIN (1:5,000, Sigma Aldrich, A2228), HIF-1α (1:500, Novus Biologicals, NB-100-479), LAMIN-A/C (1:500, Santa Cruz, sc-376248), LDH-A (1:500, Santa Cruz Biotechnology, sc-27230), PDK1 (1:1,000, Assay Designs, KAP-PK112-0), VHL (1:1,000, Cell Signaling Technologies, #68547), VINCULIN (1:5,000, Abcam, ab130007).

RNA-sequencing

RNA was isolated from powdered frozen samples of wild-type kidney cortex controls from Cre negative mice in the $Vhl^{fl/fl}Trp53^{fl/fl}Rb1^{fl/fl}$ background and from tumours of the different genetic backgrounds using the NucleoSpin RNA kit (Machery Nagel). Paired-end RNA-sequencing was performed on an Illumina HISEQ4000 device by the core facility of the German Cancer Research Center (DKFZ) in Heidelberg with the Illumina TruSeq Stranded RNA library preparation kit. Previously published sequencing data of wild type cortex and tumours from the $Vhl^{\Delta/\Delta}Trp53^{\Delta/\Delta}Rb1^{\Delta/\Delta}$ mouse model was also included for subsequent analyses [16]. Raw data fastq-files were pre-processed with trimmomatic [88] to assure sufficient read quality by removing adapters and bases in the low quality segment regions (end of the reads) with a base quality below 20. Before trimming the average number of reads was 48309915 ± 12246964 [26804116,70620322], after trimming the average number of reads was 45451780 ± 13975818 [20629675,70032834]. Hence, an average of 93.2% ± 10.5% of the raw reads survived the trimming step (Supplementary Fig. 6a). The overall quality of the bases and reads was good. After quality control and trimming the reads were 2-pass aligned using the STAR aligner [89] and the GRCm38 reference genome from Ensembl. 85.1% ± 3.3% of the reads were uniquely mapped and considered (Supplementary Fig. 6b). The alignment step was followed by normalization and differential expression analysis with the R/Bioconductor [90] package DESeq2 [91]. The normalisation of the raw read counts was performed with DESeq2 by considering the library size. Additionally, all genes with a low count across all samples, i.e. the row sum of a gene was below 5 in a gene by sample matrix, were removed from the dataset. After pre-processing and filtering 19,723 genes were further analysed and fitted with a negative binomial generalised linear model followed by Wald statistics to

identify differentially expressed genes. Genes were considered significant with an adjusted p-value < 0.001 (Benjamini-Hochberg). Raw RNA sequencing data have been uploaded to GEO with identifier GSE150983 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150983].

Gene Set Enrichment Analysis

Enrichment of signalling pathways was performed as implemented in the R/Bioconductor package GAGE (Generally Applicable Gene-Set Analysis) [92] with signalling pathways from Gene Ontology [93,94], ConsensusPathDB [95] and MSigDB [96]. The human gene identifiers from the MSigDB pathways were mapped on mouse homologs with the R/Bioconductor package GeneAnswers (R package version 2.28.0). Pathways were considered significant with an adjusted p-value < 0.05 (Benjamini-Hochberg).

ssGSEA Immune Deconvolution Analysis

RNA-seq raw read sequences were aligned against mouse genome assembly mm10 by STAR 2-pass alignment [89]. QC metrics, for example general sequencing statistics, gene feature and body coverage, were then calculated based on the alignment result through RSeQC. RNA-seq gene level count values were computed by using the R package GenomicAlignments [97] over aligned reads with UCSC KnownGene [98] in mm10 as the base gene model. The Union counting mode was used and only mapped paired reads after alignment quality filtering were considered. Gene level FPKM (Fragments Per Kilobase Million) and raw read count values were computed by the R package DESeq2 [91]. Single-Sample GSEA [99] was utilised for immune deconvolution analyses based on FPKM expression values to estimate the abundance of immune cell types [44], MHC class I antigen presenting machinery expression, T-cell infiltration score (TIS), Immune Infiltration Score (IIS) [43] and immune cytolytic score (CYT) [100] as well as the eTME signatures [42] which was developed from leveraging RCC patient-derived xenograft RNA-sequencing data. In addition to the gene signature-based deconvolution approach, CIBERSORT [48] which is a regression-based method using Support Vector Machine algorithm was also employed using either the human gene panel or the mouse specific reference panel, ImmuCC [45].

Mass spectrometry-based proteome profiling

Tissue sections of wild-type kidney cortex controls from Cre negative mice in the $Vhl^{fl/fl}Trp53^{fl/fl}Rb1^{fl/fl}$ background and from tumours of the different genetic backgrounds were homogenised by sonication (Diagenode, 4102 Seraing, Belgium) in 100 mM HEPES, pH 8.0, 4 % (w/v) SDS, 10 mM DTT, followed by heat incubation (95°C, 10 min), centrifugation (16,000 g, 10 min), cysteine alkylation, and buffer exchange to 100 mM HEPES, pH 8.0 with subsequent trypsination of 100 µg proteome based on the filter aided sample preparation protocol [101]. 24 samples (six mice from four genotypes) were distributed across three sets and differentially labeled by amine-reactive tandem-mass tags (TMT11plex, Thermo/Pierce, Rockford, IL, USA) including a pooled normalisation sample. A summary of the labeling scheme can be found online as part of the ProteomeXchange submission (see below for details). Each batch was fractionated by high pH reversed phase chromatography (XBridge C18, 3.5µm, 150 mm x 4.5 mm column (Waters, MA, USA)). Both eluents A (water) and B (70 % acetonitrile) contained 10 mM ammonium formate, adjusted to pH 10 with ammonium hydroxide. Flow rate was 0.3 ml/min. After washing with 16 % B, samples were eluted by a linear gradient from 16 – 55 % B in 40 minutes. Peptide elution was monitored by UV/VIS absorption at 214 nm. 16 fractions were collected and concatenated into eight final fractions (pool scheme was 1+9, 2+10, 3+11,

4+12, 5+13, 6+14, 7+15, 8+16). For analysis by liquid chromatography – tandem mass spectrometry (LC-MS/MS), fractions were separated by an EASY nano-LC system 1000 (Thermo Fisher Scientific, Waltham, MA, USA) and using an EASY-Spray™ C18 column (250 mm x 75 µm, 2µm particles heated at 50°C, Thermo Fisher Scientific, Waltham, MA, USA). Both eluents A (water) and B (acetonitrile) contained 0.1% formic acid. The gradient program consisted of the following steps: linear 2 - 25% B increase in 60 minutes and 25 - 60% B in 30 minutes, providing a 90 min separation window at 300 nl/min flow rate. Peptides were analyzed using an Oribtrap Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) operating in data dependent acquisition (DDA) mode. Survey scans were performed at 70,000 resolution, an AGC target of 3e6 and a maximum injection time of 50 ms followed by targeting the top 10 precursor ions for fragmentation scans at 35,000 resolution with 1.2 m/z isolation windows, an NCE of 32 and a dynamic exclusion time of 40 s. For all MS2 scans, the intensity threshold was set to 1000, the AGC to 1e5, maximum injection time of 100 ms and the fixed first mass to 100 m/z. Data were analyzed by MaxQuant v 1.6.013 with the following settings: tryptic specificity, up to two missed cleavages, TMT-modification of peptide N-termini and lysine side chains; cysteine carbamidomethylation, mouse reviewed sequences (downloaded from Uniprot on Aug 26th, 2019), 1 % FDR for peptides and proteins, precursor intensity fraction = 0.5, one or more unique peptides for protein quantitation. MaxQuant output was further processed by MSStatsTMT [102] for normalisation, batch removal, and protein assembly. Differential protein abundance was assessed using linear models of microarray analysis. Data are available via PRIDE / ProteomeXchange with identifier PXD016630 [http://www.ebi.ac.uk/pride/archive/projects/PXD016630] [103]. Of note, the dataset is associated with the ontology term “TMT6plex” since an ontology term for the isobaric TMT11plex has not yet been established.

DATA AVAILABILITY

Raw proteomics data are available via PRIDE / ProteomeXchange with identifier PXD016630 [http://www.ebi.ac.uk/pride/archive/projects/PXD016630]. Raw RNA sequencing data have been uploaded to GEO with identifier GSE150983 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150983]. The source data underlying analyses in Figs 3b-i, 4c-g, 5a-c, 6a, 7a and Supplementary Figs S3, S7a-e, S8a,b, S9a-d, S10a and S12a are provided as Supplementary Data 1-6. Full scan blots for Supplementary Figure 2 and 5 are provided as Source Data file. All remaining relevant data are available in the article, supplementary information, or from the corresponding author upon reasonable request.

REFERENCES

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **68**, 394–424 (2018).
2. Frew, I. J. & Moch, H. A Clearer View of the Molecular Complexity of Clear Cell Renal Cell Carcinoma. *Annu. Rev. Pathol. Mech. Dis.* **10**, 263–289 (2015).
3. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* **46**, 225–233 (2014).
4. Turajlic, S. *et al.* Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell* **173**, 595-610.e11 (2018).
5. Mitchell, T. J. *et al.* Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. *Cell* **173**, 611-623.e17 (2018).
6. Batavia, A. A., Schraml, P. & Moch, H. Clear cell renal cell carcinoma with wild-type *von Hippel-Lindau* gene: a non-existent or new tumour entity? *Histopathology* **74**, 60–67 (2019).
7. Frew, I. J. & Krek, W. pVHL: A Multipurpose Adaptor Protein. *Science Signaling* **1**, pe30–pe30 (2008).
8. Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet* **45**, 860–867 (2013).
9. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
10. Frew, I. J. *et al.* pVHL and PTEN tumour suppressor proteins cooperatively suppress kidney cyst formation. *EMBO J* **27**, 1747–1757 (2008).
11. Gu, Y.-F. *et al.* Modeling Renal Cell Carcinoma in Mice: *Bap1* and *Pbrm1* Inactivation Drive Tumor Grade. *Cancer Discov* **7**, 900–917 (2017).
12. Nargund, A. M. *et al.* The SWI/SNF Protein PBRM1 Restrains VHL-Loss-Driven Clear Cell Renal Cell Carcinoma. *Cell Reports* **18**, 2893–2906 (2017).
13. Espana-Agusti, J., Warren, A., Chew, S. K., Adams, D. J. & Matakidou, A. Loss of PBRM1 rescues VHL dependent replication stress to promote renal carcinogenesis. *Nat Commun* **8**, 2026 (2017).
14. Wang, S.-S. *et al. Bap1* is essential for kidney function and cooperates with *Vhl* in renal tumorigenesis. *Proc Natl Acad Sci USA* **111**, 16538–16543 (2014).
15. Albers, J. *et al.* Combined mutation of *Vhl* and *Trp53* causes renal cysts and tumours in mice. *EMBO Mol Med* **5**, 949–964 (2013).
16. Harlander, S. *et al.* Combined mutation in Vhl, Trp53 and Rb1 causes clear cell renal cell carcinoma in mice. *Nat Med* **23**, 869–877 (2017).
17. Bailey, S. T. *et al.* MYC activation cooperates with Vhl and Ink4a/Arf loss to induce clear cell renal cell carcinoma. *Nat Commun* **8**, 15770 (2017).
18. Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271–275 (1999).
19. Monzon, F. A. *et al.* Chromosome 14q loss defines a molecular subtype of clear-cell renal cell carcinoma associated with poor prognosis. *Mod Pathol* **24**, 1470–1479 (2011).
20. Turajlic, S. *et al.* Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell* **173**, 581-594.e12 (2018).
21. Gordan, J. D. *et al.* HIF-α Effects on c-Myc Distinguish Two Subtypes of Sporadic VHL-Deficient Clear Cell Renal Carcinoma. *Cancer Cell* **14**, 435–446 (2008).
22. Shen, C. *et al.* Genetic and Functional Studies Implicate HIF1 as a 14q Kidney Cancer Suppressor Gene. *Cancer Discovery* **1**, 222–235 (2011).
23. Raval, R. R. *et al.* Contrasting Properties of Hypoxia-Inducible Factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-Associated Renal Cell Carcinoma. *Molecular and Cellular Biology* **25**, 5675–5686 (2005).
24. Kondo, K., Kim, W. Y., Lechpammer, M. & Kaelin, W. G. Inhibition of HIF2a Is Sufficient to Suppress pVHL-Defective Tumor Growth. 6.
25. Chen, W. *et al.* Targeting renal cell carcinoma with a HIF-2 antagonist. *Nature* **539**, 112–117 (2016).
26. Courtney, K. D. *et al.* Phase I Dose-Escalation Trial of PT2385, a First-in-Class Hypoxia-Inducible Factor-2α Antagonist in Patients With Previously Treated Advanced Clear Cell Renal Cell Carcinoma. *JCO* **36**, 867–874 (2018).
27. Cho, H. *et al.* On-target efficacy of a HIF-2α antagonist in preclinical kidney cancer models. *Nature* **539**, 107–111 (2016).
28. Lindström, M. S. *et al.* Nucleolus as an emerging hub in maintenance of genome stability and cancer pathogenesis. *Oncogene* **37**, 2351–2366 (2018).
29. Schönenberger, D. *et al.* Formation of Renal Cysts and Tumors in *Vhl/Trp53* -Deficient Mice Requires HIF1α and HIF2α. *Cancer Res* **76**, 2025–2036 (2016).
30. Zaldumbide, L., Erramuzpe, A., Guarch, R., Cortés, J. M. & López, J. I. Large (>3.8 cm) clear cell renal cell carcinomas are morphologically and immunohistochemically heterogeneous. *Virchows Arch* **466**, 61–66 (2015).

31. Young, A. P. *et al.* VHL loss actuates a HIF-independent senescence programme mediated by Rb and p400. *Nat Cell Biol* **10**, 361–369 (2008).
32. Welford, S. M., Dorie, M. J., Li, X., Haase, V. H. & Giaccia, A. J. Renal Oxygenation Suppresses VHL Loss-Induced Senescence That Is Caused by Increased Sensitivity to Oxidative Stress. *Molecular and Cellular Biology* **30**, 4595–4603 (2010).
33. Kondo, K., Klco, J., Nakamura, E., Lechpammer, M. & Kaelin, W. G. Inhibition of HIF is necessary for tumor suppression by the von Hippel-Lindau protein. *Cancer Cell* **1**, 237–246 (2002).
34. Semenza, G. L. HIF-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations. *J. Clin. Invest.* **123**, 3664–3671 (2013).
35. Courtney, K. D. *et al.* HIF-2 Complex Dissociation, Target Inhibition, and Acquired Resistance with PT2385, a First-in-Class HIF-2 Inhibitor, in Patients with Clear Cell Renal Cell Carcinoma. *Clin Cancer Res* **26**, 793–803 (2020).
36. Drendel, V. *et al.* Proteome profiling of clear cell renal cell carcinoma in von Hippel-Lindau patients highlights upregulation of Xaa-Pro aminopeptidase-1, an anti-proliferative and anti-migratory exoprotease. *Oncotarget* **8**, 100066–100078 (2017).
37. Weißer, J. *et al.* Quantitative proteomic analysis of formalin-fixed, paraffin-embedded clear cell renal cell carcinoma tissue using stable isotopic dimethylation of primary amines. *BMC Genomics* **16**, 559 (2015).
38. Wu, D. *et al.* ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* **26**, 2176–2182 (2010).
39. Saito, T., Kimura, M., Kawasaki, T., Sato, S. & Tomita, Y. MHC class II antigen-associated invariant chain on renal cell cancer may contribute to the anti-tumor immune response of the host. *Cancer Lett.* **115**, 121–127 (1997).
40. Gastl, G. *et al.* Major histocompatibility complex class I and class II expression in renal cell carcinoma and modulation by interferon gamma. *J. Urol.* **155**, 361–367 (1996).
41. Dengjel, J. Unexpected Abundance of HLA Class II Presented Peptides in Primary Renal Cell Carcinomas. *Clinical Cancer Research* **12**, 4163–4170 (2006).
42. Wang, T. *et al.* An Empirical Approach Leveraging Tumorgrafts to Dissect the Tumor Microenvironment in Renal Cell Carcinoma Identifies Missing Link to Prognostic Inflammatory Factors. *Cancer Discov* **8**, 1142–1155 (2018).
43. Şenbabaoğlu, Y. *et al.* Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol* **17**, 231 (2016).
44. Bindea, G. *et al.* Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782–795 (2013).
45. Chen, Z. *et al.* seq-ImmuCC: Cell-Centric View of Tissue Transcriptome Measuring Cellular Compositions of Immune Microenvironment From Mouse RNA-Seq Data. *Front Immunol* **9**, 1286 (2018).
46. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discovery* **2**, 401–404 (2012).
47. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).
48. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453–457 (2015).
49. Xiong, Y. *et al.* Tumor infiltrating mast cells determine oncogenic HIF-2α-conferred immune evasion in clear cell renal cell carcinoma. *Cancer Immunol Immunother* **68**, 731–741 (2019).
50. Xu, J. *et al.* Epigenetic regulation of HIF-1α in renal cancer cells involves HIF-1α/2α binding to a reverse hypoxia-response element. *Oncogene* **31**, 1065–1072 (2012).
51. Schulz, K. *et al.* HIF-1α protein is upregulated in HIF-2α depleted cells via enhanced translation. *FEBS Letters* **586**, 1652–1657 (2012).
52. Koh, M. Y., Lemos, R., Liu, X. & Powis, G. The Hypoxia-Associated Factor Switches Cells from HIF-1 -to HIF-2 -Dependent Signaling Promoting Stem Cell Characteristics, Aggressive Tumor Growth and Invasion. *Cancer Research* **71**, 4015–4027 (2011).
53. Mandriota, S. J. *et al.* HIF activation identifies early lesions in VHL kidneys: Evidence for site-specific tumor suppressor function in the nephron. *Cancer Cell* 10 (2002).
54. Gudas, L. J., Fu, L., Minton, D. R., Mongan, N. P. & Nanus, D. M. The role of HIF1α in renal cell carcinoma tumorigenesis. *J Mol Med* **92**, 825–836 (2014).
55. Fu, L., Wang, G., Shevchuk, M. M., Nanus, D. M. & Gudas, L. J. Generation of a Mouse Model of Von Hippel–Lindau Kidney Disease Leading to Renal Cancers by Expression of a Constitutively Active Mutant of HIF1a. *Cancer Research* **71**, 6848-6856 (2011).
56. Fu, L., Wang, G., Shevchuk, M. M., Nanus, D. M. & Gudas, L. J. Activation of HIF2a in Kidney Proximal Tubule Cells Causes Abnormal Glycogen Deposition but not Tumorigenesis. *Cancer Research*, **73**, 2916-2925 (2013).
57. Farsijani, N. M. *et al.* Renal epithelium regulates erythropoiesis via HIF-dependent suppression of erythropoietin. *Journal of Clinical Investigation* **126**, 1425–1437 (2016).

58. Pritchett, T. L., Bader, H. L., Henderson, J. & Hsu, T. Conditional inactivation of the mouse von Hippel–Lindau tumor suppressor gene results in wide-spread hyperplastic, inflammatory and fibrotic lesions in the kidney. *Oncogene* **34**, 2631–2639 (2015).
59. Rankin, E. B., Tomaszewski, J. E. & Haase, V. H. Renal Cyst Development in Mice with Conditional Inactivation of the von Hippel-Lindau Tumor Suppressor. *Cancer Res* **66**, 2576–2583 (2006).
60. Iguchi, M. *et al.* Acute Inactivation of the *VHL* Gene Contributes to Protective Effects of Ischemic Preconditioning in the Mouse Kidney. *Nephron Exp Nephrol* **110**, e82–e90 (2008).
61. Schietke, R. E. *et al.* Renal Tubular HIF-2a Expression Requires VHL Inactivation and Causes Fibrosis and Cysts. *PLoS ONE* **7**, 12 (2012).
62. Mathia, S. *et al.* Action of hypoxia-inducible factor in liver and kidney from mice with Pax8-rtTA-based deletion of von Hippel-Lindau protein. *Acta Physiol* **207**, 565–576 (2013).
63. Frew, I. J. & Krek, W. Multitasking by pVHL in tumour suppression. *Current Opinion in Cell Biology* **19**, 685–690 (2007).
64. Gao, W., Li, W., Xiao, T., Liu, X. S. & Kaelin, W. G. Inactivation of the PBRM1 tumor suppressor gene amplifies the HIF-response in VHL $^{-/-}$ clear cell renal carcinoma. *Proc Natl Acad Sci USA* **114**, 1027–1032 (2017).
65. Chowdhury, B. *et al.* PBRM1 Regulates the Expression of Genes Involved in Metabolism and Cell Adhesion in Renal Clear Cell Carcinoma. *PLoS ONE* **11**, e0153718 (2016).
66. Fu, L., Minton, D. R., Zhang, T., Nanus, D. M. & Gudas, L. J. Genome-Wide Profiling of TRACK Kidneys Shows Similarity to the Human ccRCC Transcriptome. *Molecular Cancer Research* **13**, 870–878 (2015).
67. Minton, D. R. *et al.* Analyses of the Transcriptome and Metabolome Demonstrate That HIF1α Mediates Altered Tumor Metabolism in Clear Cell Renal Cell Carcinoma. *PLoS ONE* **10**, e0120649 (2015).
68. Hakimi, A. A. *et al.* An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma. *Cancer Cell* **29**, 104–116 (2016).
69. Courtney, K. D. *et al.* Isotope Tracing of Human Clear Cell Renal Cell Carcinomas Demonstrates Suppressed Glucose Oxidation In Vivo. *Cell Metabolism* **28**, 793-800.e2 (2018).
70. Curthoys, N. P. & Moe, O. W. Proximal tubule function and response to acidosis. *Clin J Am Soc Nephrol* **9**, 1627–1638 (2014).
71. Lee, J. W., Chou, C.-L. & Knepper, M. A. Deep Sequencing in Microdissected Renal Tubules Identifies Nephron Segment-Specific Transcriptomes. *J. Am. Soc. Nephrol.* **26**, 2669–2677 (2015).
72. Gonzalez, H., Hagerling, C. & Werb, Z. Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes Dev.* **32**, 1267–1284 (2018).
73. Giraldo, N. A. *et al.* Orchestration and Prognostic Significance of Immune Checkpoints in the Microenvironment of Primary and Metastatic Renal Cell Cancer. *Clinical Cancer Research* **21**, 3031–3040 (2015).
74. Giraldo, N. A. *et al.* Tumor-Infiltrating and Peripheral Blood T-cell Immunophenotypes Predict Early Relapse in Localized Clear Cell Renal Cell Carcinoma. *Clin Cancer Res* **23**, 4416–4428 (2017).
75. Najjar, Y. G. *et al.* Myeloid-Derived Suppressor Cell Subset Accumulation in Renal Cell Carcinoma Parenchyma Is Associated with Intratumoral Expression of IL1β, IL8, CXCL5, and Mip-1α. *Clin Cancer Res* **23**, 2346–2355 (2017).
76. Chevrier, S. *et al.* An Immune Atlas of Clear Cell Renal Cell Carcinoma. *Cell* **169**, 736-749.e18 (2017).
77. Dannenmann, S. R. *et al.* Tumor-associated macrophages subvert T-cell function and correlate with reduced survival in clear cell renal cell carcinoma. *OncoImmunology* **2**, e23562 (2013).
78. O'Sullivan, D., Sanin, D. E., Pearce, E. J. & Pearce, E. L. Metabolic interventions in the immune response to cancer. *Nat. Rev. Immunol.* **19**, 324–335 (2019).
79. Lee, K. *et al.* Acriflavine inhibits HIF-1 dimerization, tumor growth, and vascularization. *Proceedings of the National Academy of Sciences* **106**, 17910–17915 (2009).
80. Wu, D., Potluri, N., Lu, J., Kim, Y. & Rastinejad, F. Structural integration in hypoxia-inducible factors. *Nature* **524**, 303–308 (2015).
81. Shay, J. E. S. *et al.* Inhibition of hypoxia-inducible factors limits tumor progression in a mouse model of colorectal cancer. *Carcinogenesis* **35**, 1067–1077 (2014).
82. Yin, T., He, S., Shen, G. & Wang, Y. HIF-1 Dimerization Inhibitor Acriflavine Enhances Antitumor Activity of Sunitinib in Breast Cancer Model. *oncol res* **22**, 139–145 (2015).
83. Mangraviti, A. *et al.* HIF-1α- Targeting Acriflavine Provides Long Term Survival and Radiological Tumor Response in Brain Cancer Therapy. *Sci Rep* **7**, 14978 (2017).
84. Thoma, C. R. *et al.* pVHL and GSK3β are components of a primary cilium-maintenance signalling network. *Nat Cell Biol* **9**, 588–595 (2007).
85. Hoefflin, R. *et al.* Spatial niche formation but not malignant progression is a driving force for intratumoural heterogeneity. *Nat Commun* **7**, ncomms11845 (2016).
86. Rankin, E. B. *et al.* Inactivation of the arylhydrocarbon receptor nuclear translocator (Arnt) suppresses von Hippel-Lindau disease-associated vascular tumors in mice. *Mol. Cell. Biol.* **25**, 3163–3172 (2005).
87. Gruber, M. *et al.* Acute postnatal ablation of Hif-2alpha results in anemia. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2301–2306 (2007).

88. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
89. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
90. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
91. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
92. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 161 (2009).
93. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
94. Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049-1056 (2015).
95. Kamburov, A., Stelzl, U., Lehrach, H. & Herwig, R. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.* **41**, D793-800 (2013).
96. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545–15550 (2005).
97. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* **9**, e1003118 (2013).
98. Rosenbloom, K. R. *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**, D670-681 (2015).
99. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
100. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
101. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
102. Choi, M. *et al.* MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–2526 (2014).
103. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

ACKNOWLEDGEMENTS

We are most grateful to the team of Stephan Wolf from the Genomics and Proteomics Core Facility, German Cancer Research Center/DKFZ, Heidelberg, Germany for their sequencing service, to Martin Biniossek for support with the proteomic analysis and to the members of the Frew laboratory for helpful discussions. This work was supported by grants from the Deutsche Forschungsgemeinschaft to IJF (BIOSS Excellence Cluster and CRC 850), to RZ and MB (CRC 850), to OS (SCHI 871/11-1), from the Else-Kröner-Fresenius Stiftung and Berta-Ottenstein Programme for Clinician Scientists to RH, and to MB and PM from the German Federal Ministry of Education and Research (BMBF) for MIRACUM within the Medical Informatics Funding Scheme (FKZ 01ZZ1801B).

AUTHOR CONTRIBUTIONS

RH, SH, SS, DS, MA, AP, FU, PS, DH, CC and IF, designed and conducted experiments, PM, CC, MC-C, RH, OS, FK and IF analysed RNA seq and proteomic data, CS contributed pathological analyses, AJ, BL, NG, TC, JD, RZ, MH, OS, AH, MB and IF provided technical assistance and scientific advice, IF designed the study and wrote the manuscript in consultation with all authors.

COMPETING INTERESTS

The authors declare no potential conflicts of interest.

Figure 1. ccRCC formation is strongly dependent on *Hif1a* and only moderately affected by *Hif2a* deletion. **a** Tumor onset in cohorts of pR, VpR, VpRH1 and VpRH2 mice. *P* values were calculated by two-sided log-rank Mantel–Cox test. **b** Number of tumours per mouse at the time of sacrifice based on μ-CT imaging (VpR n=65, pR n=25, VpRH1 n=36 and VpRH2 n=65 mice). Mean ± SEM are shown, *P* values were calculated by Dunn's multiple comparisons test. **c** Tumour growth rates based on μ-CT imaging (VpR n=48, pR n=5, VpRH1 n=7 and VpRH2 n=56 tumours). Box-whisker plots depict median, bounded by Q1 (25% lower quartile) and Q3 (75% upper quartile) and whiskers depict 1.5 times the Q3-Q1 interquartile range. *P* values were calculated by two-sided Student's t-test without adjustment for multiple comparisons. **d** Representative immunohistochemical stainings for the indicated antibodies in samples from WT cortex, a non-tumour region of VpR cortex, VpR, VpRH1 and VpRH2 tumours. All panels are the same magnification, scale bar = 50 μm. The number of positive tumours / number of tumours examined are indicated. **e** Representative examples of the histological appearance of tumours assigned clear cell scores of 1, 2 or 3. Scale bars = 100 μm. **f** Distribution of clear cell scores between VpR (n = 10 mice, 23 tumours), VpRH1 (n = 8 mice, 14 tumours) and VpRH2 (n = 9 mice, 18 tumours) tumour cohorts. *P* values were calculated using the two-sided Mann-Whitney U test without adjustments for multiple comparisons.

Figure 2. HIF-1α is dispensable for cellular proliferation and for allograft tumour formation. **a** 3T3 proliferation assays of MEFs derived from mice of the indicated genotypes infected with adenoviruses expressing GFP or Cre. Mean ± std. dev. are derived from three independent cultures. **b** 3T3 proliferation assays of MEFs derived from *Vhl*$^{fl/fl}$*Trp53*$^{fl/fl}$ mice infected with non-silencing shRNA (shRNA-ns) or shRNA against *Hif1a* (shRNA-*Hif1a* #1 and shRNA-*Hif1a* #2), followed by infection with adenoviruses expressing GFP or Cre. Mean ± std. dev. are derived from three independent cultures. **c,d** Proliferation assays of mouse ccRCC cell line 2020 expressing empty vector control or human pVHL30 (**c**) or non-silencing shRNA (shRNA-ns) or shRNA against *Hif1a* (shRNA-*Hif1a* #1 and shRNA-*Hif1a* #2) (**d**). Mean ± std. dev. are derived from two independent experiments each with replicates of six cultures. **e-g** Representative images (scale bars depict 200 μm) (**e**) and size distributions (**f,g**) of spheres formed by the cells described in c,d when grown in non-adherent cell culture plates. Mean ± std. dev. of the total number of colonies pooled from three independent experiments are shown, *P* values were calculated by two-sided Student's t-test. **h,i** Survival of mice following subcutaneous allograft tumour assays of the cells described in c into SCID-Beige mice. *P* values were calculated by two-sided log-rank Mantel Cox test.

Figure 3. HIF-1α and HIF-2α deletion affect different transcriptional programs and inflammatory responses. **a** Principal component analysis of RNA sequencing of WT Cortex and VpR, VpRH1 and VpRH2 tumours **b-i** Gene expression heatmaps for selected differentially regulated genes from the indicated GSEA terms glycolysis (**b**), cell adhesion molecules (**c**), focal adhesion and receptor signalling (**d**), DNA repair (**e**), lipoprotein metabolism (**f**), ribosome biogenesis (**g**), T cell activation (**h**) and response to IFN-β (**i**). Rows represent row-normalised z-scores of mRNA abundance, each column represents an individual sample from WT cortex or VpR, VpRH1 and VpRH2 tumours. Source data is provided in Supplementary Data 1 and 2.

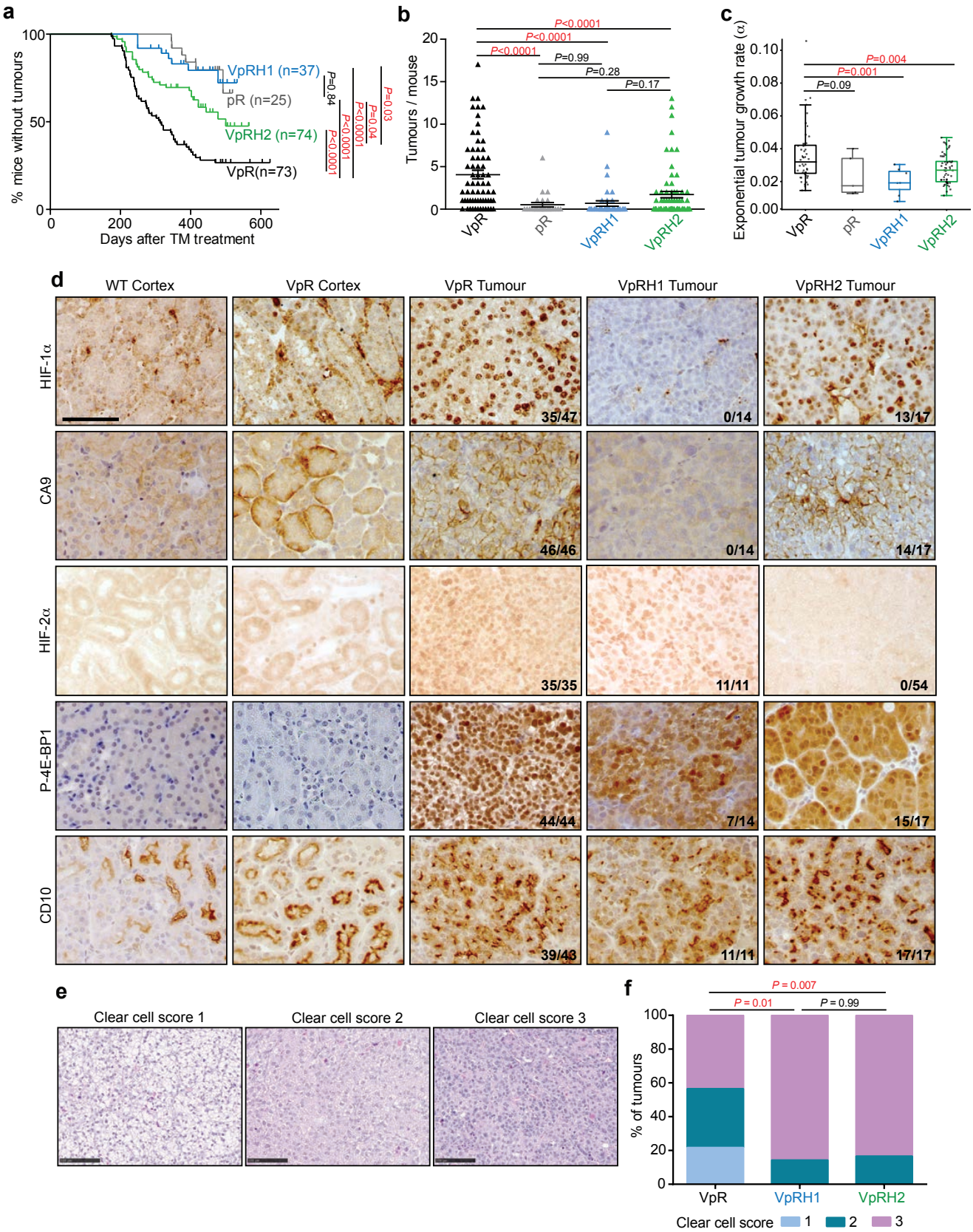
Figure 4. Proteomic analyses of the effects of HIF-1α and HIF-2α deletion in mouse ccRCCs. **a** Correlation of protein abundance in mouse and human ccRCC samples. Each dot represents a unique pair of orthologous proteins between the two species. Spemann's correlation coefficient is depicted. **b** Venn diagram showing the overlap of differentially expressed proteins derived from comparison of mouse ccRCC with WT cortex and human ccRCC with normal kidney. **c-e** Volcano plots showing differentially expressed proteins (green dots) between VpR and WT cortex (**c**), VpRH1 and VpR (**d**) and VpRH2 and VpR (**e**). **f,g** Protein expression heatmaps showing differentially expressed proteins between VpRH1 and VpR (**f**) and VpRH2 and VpR (**g**) as well as a summary of GSEA terms associated with the down- and up-regulated proteins. Rows represent row-normalised z-scores of protein abundance, each column represents an individual sample from WT cortex or VpR, VpRH1 and VpRH2 tumours. Source data is provided in Supplementary Data 3.

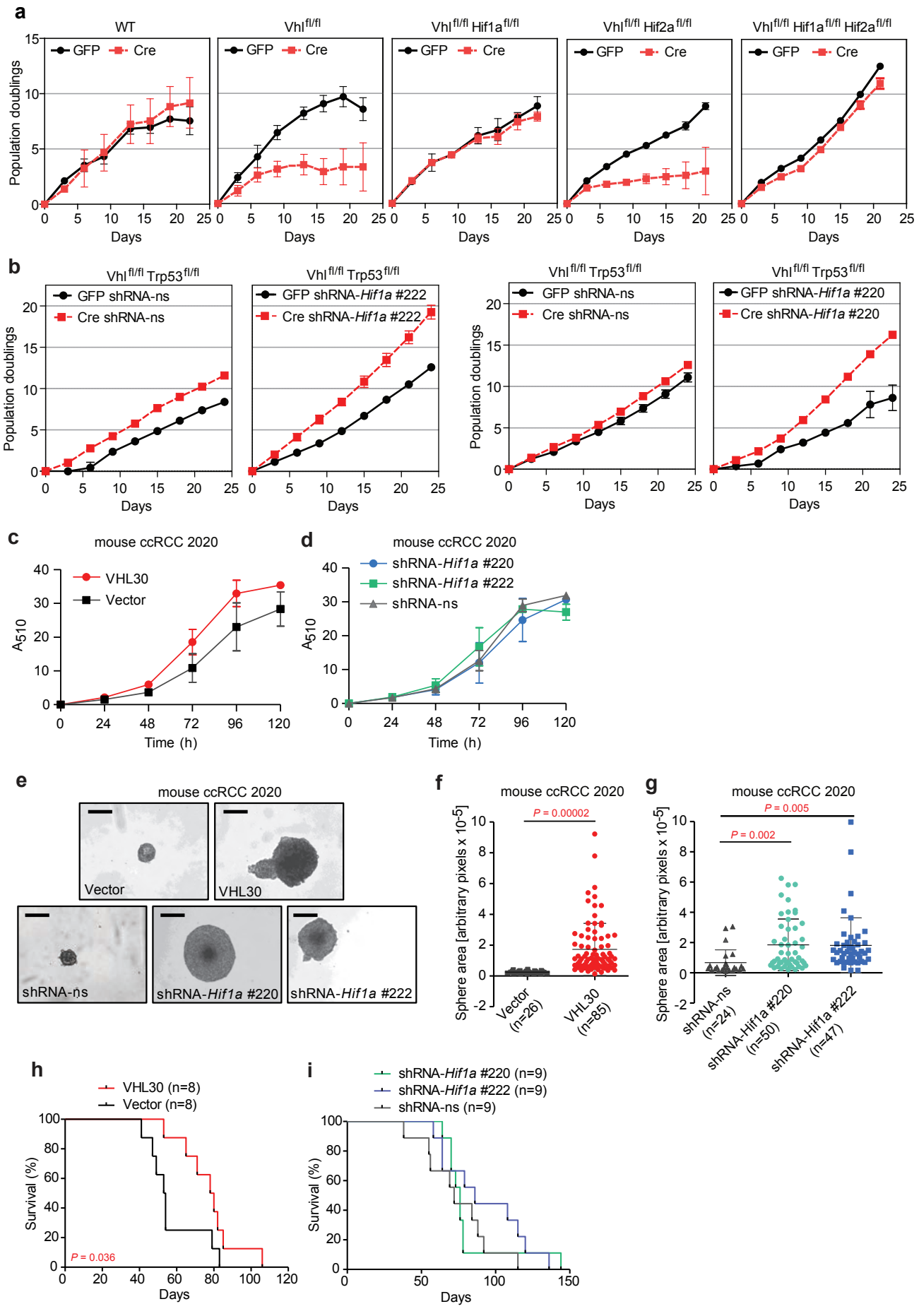
Figure 5. HIF-2α influences the expression of MHC class I and II genes. **a-c** Gene expression heatmaps for MHC class I (**a**), class II (**b**) and other antigen processing and presenting (**c**) genes. Rows represent row-normalised z-scores of mRNA abundance, each column represents an individual sample from WT cortex or VpR, VpRH1 and VpRH2 tumours. Source data is provided in Supplementary Data 1 and 2. **d** Examples of different scores for MHC class II immunohistochemical staining. All panels are the same magnification, scale bar = 100 μm. **e** Distribution of MHC class II staining scores in the indicated (n) number of VpR, VpRH1 and VpRH2 tumours. *P* values are derived from the two-sided Mann-Whitney U test without adjustments for multiple comparisons. **f,g** Relative *HIF1A* and *HIF2A* mRNA abundance in TCGA datasets of human chromophobe (KICH, n=66), clear cell (KIRC, n=533) and papillary (KIRP, n=290) renal cell carcinomas and associated normal renal tissues (Normal_KICH n=25, Normal_KIRC n=72, Normal_KIRP n=32). Box-whisker plots depict median, bounded by Q1 (25% lower quartile) and Q3 (75% upper quartile) and whiskers depict

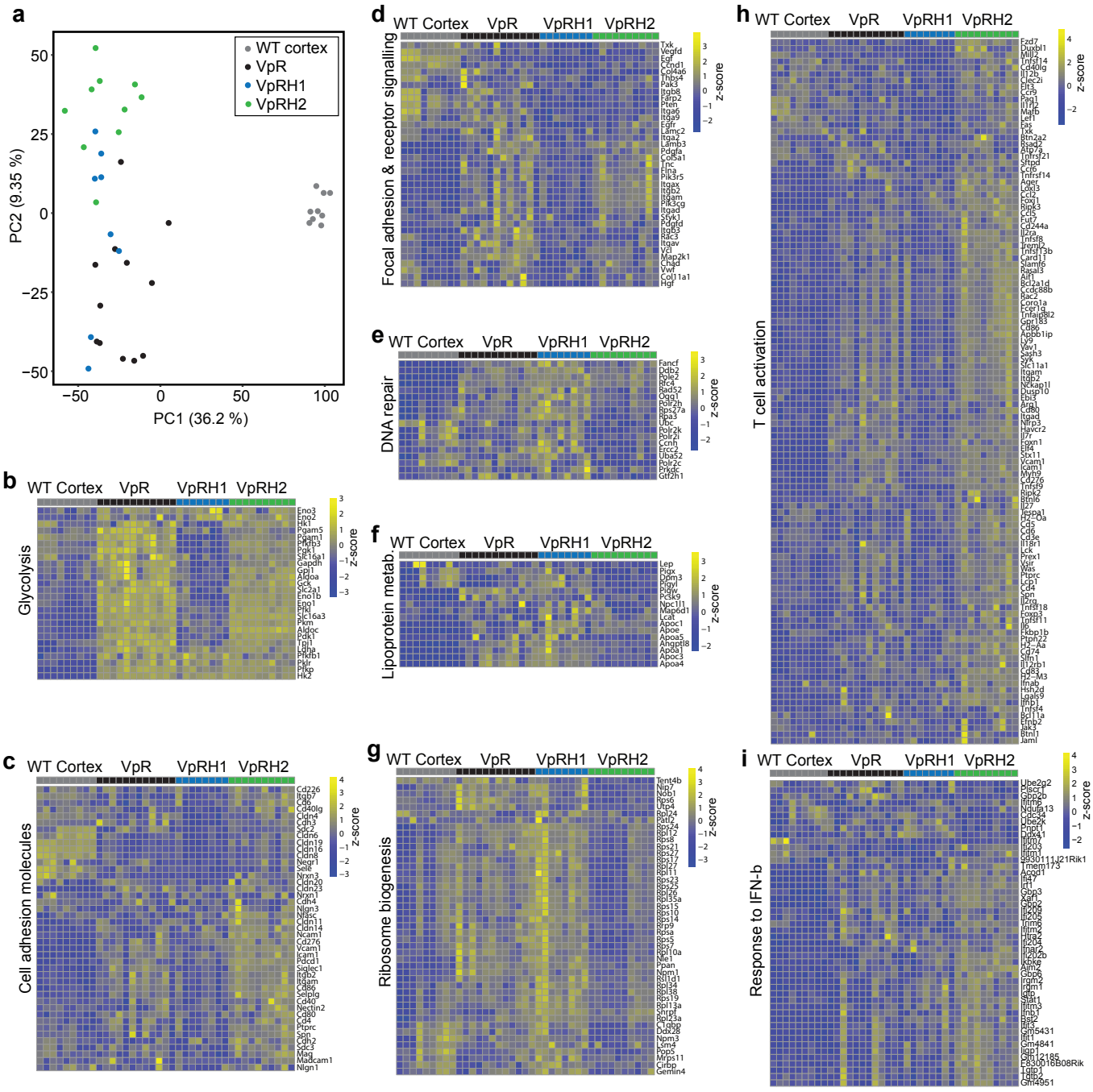
1.5 times the Q3-Q1 interquartile range. **h-m** Spearman's correlation analyses between *HIF2A* mRNA abundance and mRNA abundance of two MHC class I (**h,i**), class II (**j,k**) and antigen processing/presentation (**l,m**) genes in ccRCC (TCGA KIRC dataset). Source data is provided in Supplementary Data 4.

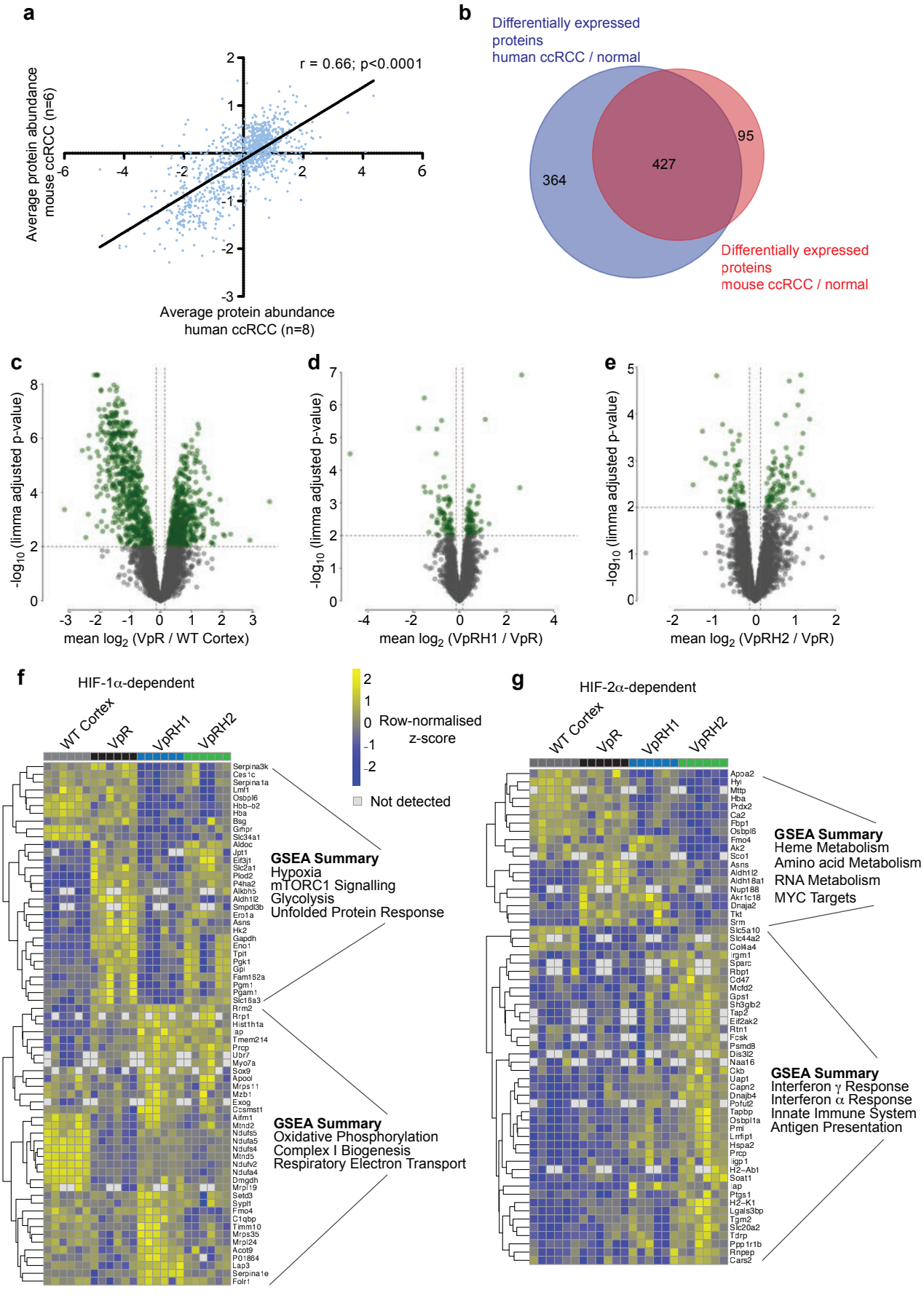
Figure 6. Deconvolution of the immune microenvironment of VpR, VpRH1 and VpRH2 tumours. **a** Summary of immune deconvolution results using ssGSEA with the Bindea&Others and eTME gene signatures, as well as the ImmuCC method. Pairwise comparisons of the expression levels of each immune cell-specific gene set between WT cortex, VpR, VpRH1 and VpRH2 tumours are shown. Columns depict the comparison between the genotypes and rows depict the gene set. Heatmap colours represent the mean differences in the z-scores. Comparisons marked with an asterix show *P* values < 0.05 between each genotype, two-sided Mann-Whitney U test without multiple comparisons. Gene signatures and source data with z-scores and *P* values are provided in Supplementary Data 5 and 6, respectively. **b-k** Quantification of the densities of immunohistochemically positive cells stained with the indicated antibodies in unaffected normal renal tissue and VpR (n=26), VpRH1 (n=14) and VpRH2 (n=21) tumours. Mean ± SEM are shown, *P* values for pairwise comparisons were calculated by 1-way ANOVA followed by two-sided Mann-Whitney U test without adjustments for multiple comparisons.

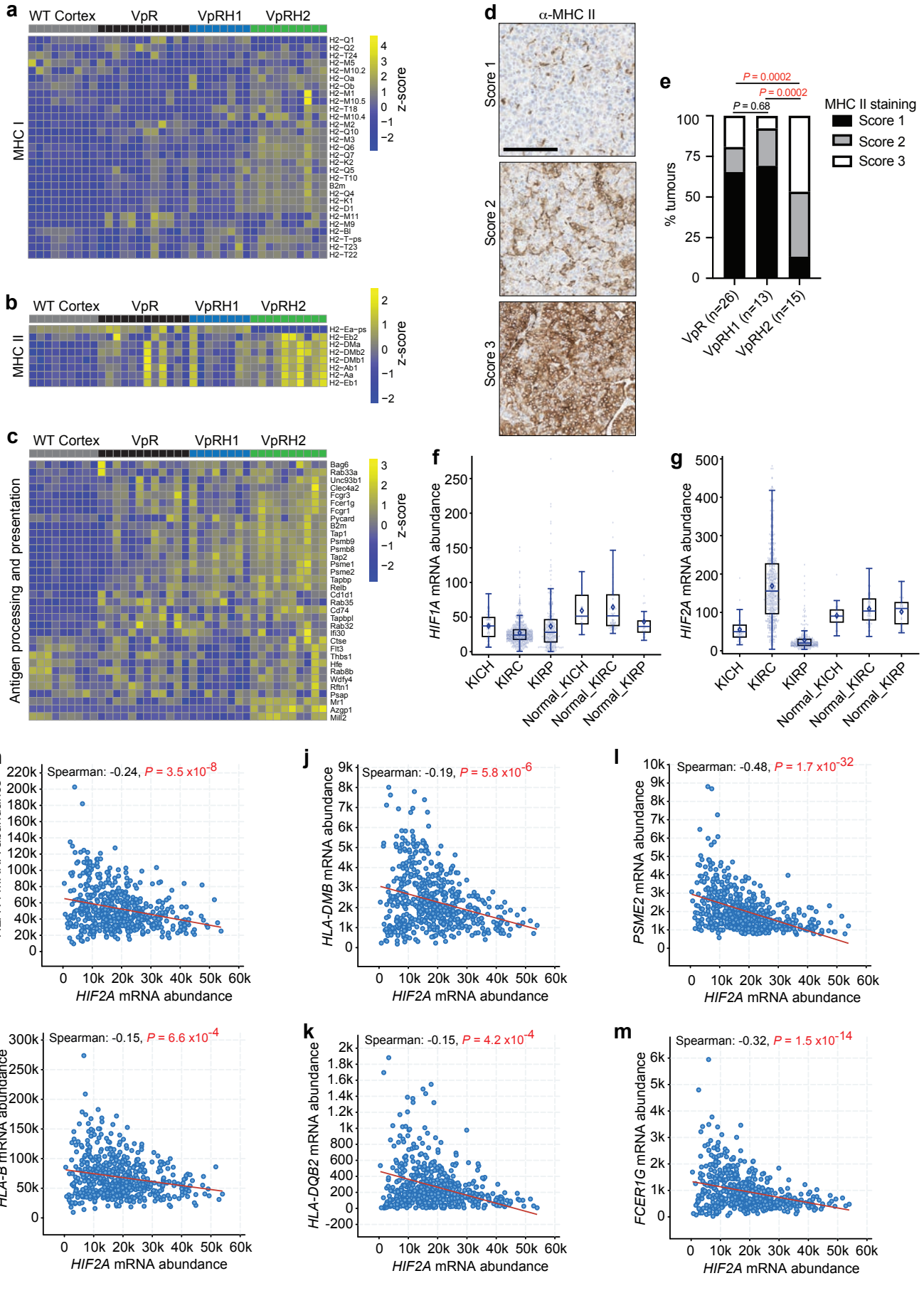
Figure 7. *HIF1A* copy number loss and *HIF2A* mRNA expression levels correlate with altered immune microenvironments in human ccRCC. **a** Oncoprint showing the genetic alterations in *HIF1*A and *HIF2A* in human ccRCC tumours based on GISTIC. **b,c** Kaplan-Meier curves showing overall survival of ccRCC patients whose tumours exhibit loss of one or two copies of *HIF1A* (loss) or gain of a copy of *HIF2A* (gain) versus patients without these copy number alterations (Unaltered). *P* values are derived from the two-sided log-rank test. **d-m** mRNA abundance (log_2 transformed, normalised, RNAseq v2 RSEM) of *CD3D* (**d,e**), *CD3E* (**f,g**), *CD4* (**h,i**), *CD8A* (**j,k**), *CD8B* (**l,m**) in *HIF1A* loss and *HIF2A* gain ccRCC tumours compared to Unaltered tumours. *P* values are derived from two-sided Student's t-test. **n** Summary of immune deconvolution results in ccRCC (TCGA KIRC dataset) using ssGSEA with the Bindea&Others and eTME gene signatures, as well as the CIBERSORT method. Depicted are pairwise comparisons of the expression levels of each gene set between *HIF2A* gain (n=65) and *HIF2A* diploid (Unaltered, n=349), between *HIF1A* loss (n=188) and *HIF1A* diploid (Unaltered, n=220) tumours and between tumours in the top quartile (Q4) and the lowest quartile (Q1) of *HIF2A* expression. Columns depict the comparison between the genotypes and rows depict the gene set. Heatmap colours represent the mean differences in the z-scores. Comparisons marked with an asterix show *P* values < 0.05 between each genotype, two-sided Mann-Whitney U test without multiple comparisons. Source data with z-scores and *P* values is provided in Supplementary Data 6.

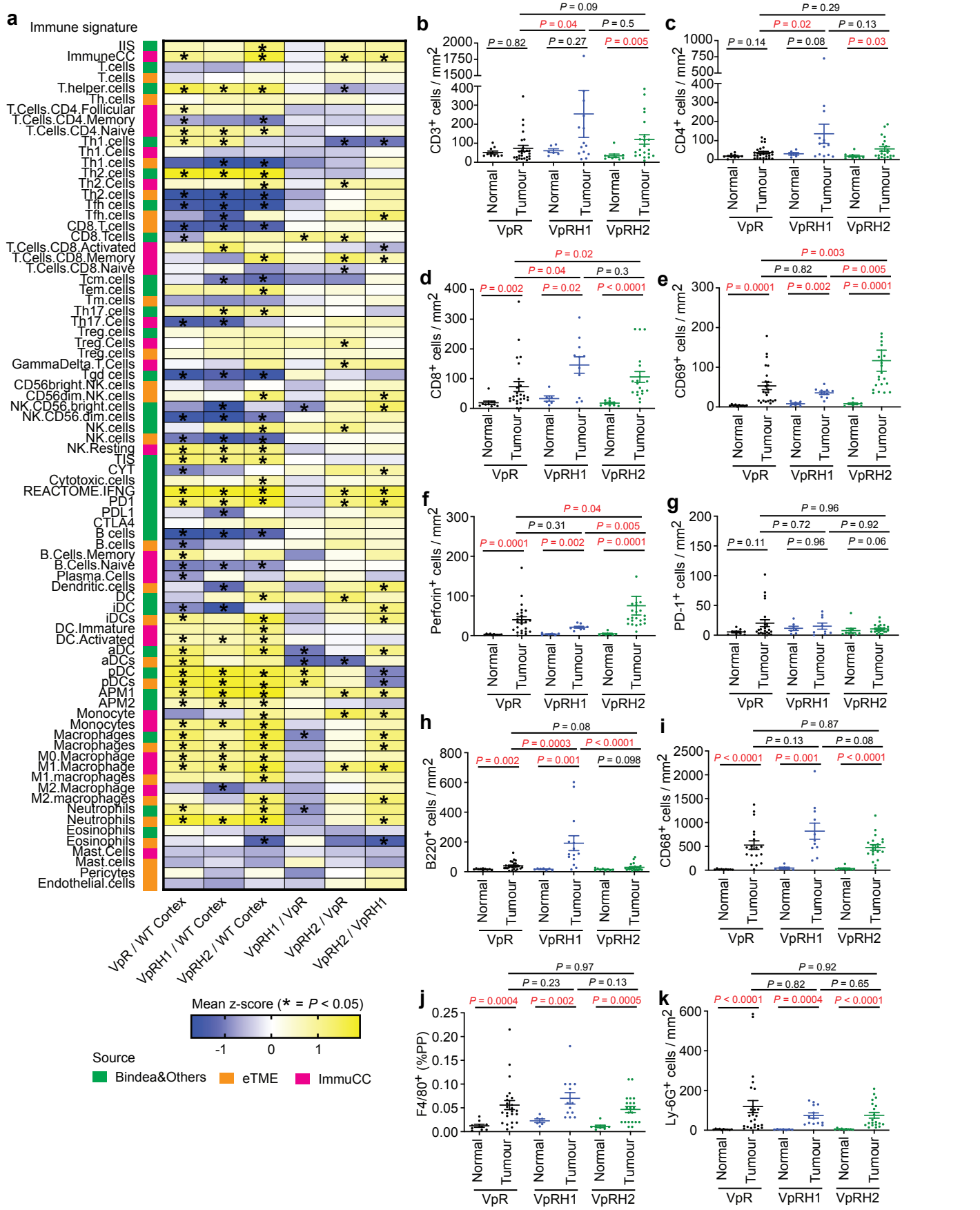

a
% mice without tumours
Days after TM treatment
VpRH1 (n=37)
pR (n=25)
VpRH2 (n=74)
VpR(n=73)
P=0.84
P<0.0001
P<0.0001
P=0.04
P<0.0001
P=0.03
b
Tumours / mouse
P<0.0001
P<0.0001
P<0.0001
P=0.99
P=0.28
P=0.17
VpR
pR
VpRH1
VpRH2
c
Exponential tumour growth rate (α)
P=0.004
P=0.001
P=0.09
d
WT Cortex
VpR Cortex
VpR Tumour
VpRH1 Tumour
VpRH2 Tumour
HIF-1α
CA9
HIF-2α
P-4E-BP1
CD10
35/47
0/14
13/17
46/46
0/14
14/17
35/35
11/11
0/54
44/44
7/14
15/17
39/43
11/11
17/17
e
Clear cell score 1
Clear cell score 2
Clear cell score 3
f
P = 0.007
P = 0.01
P = 0.99
% of tumours
Clear cell score 1 2 3

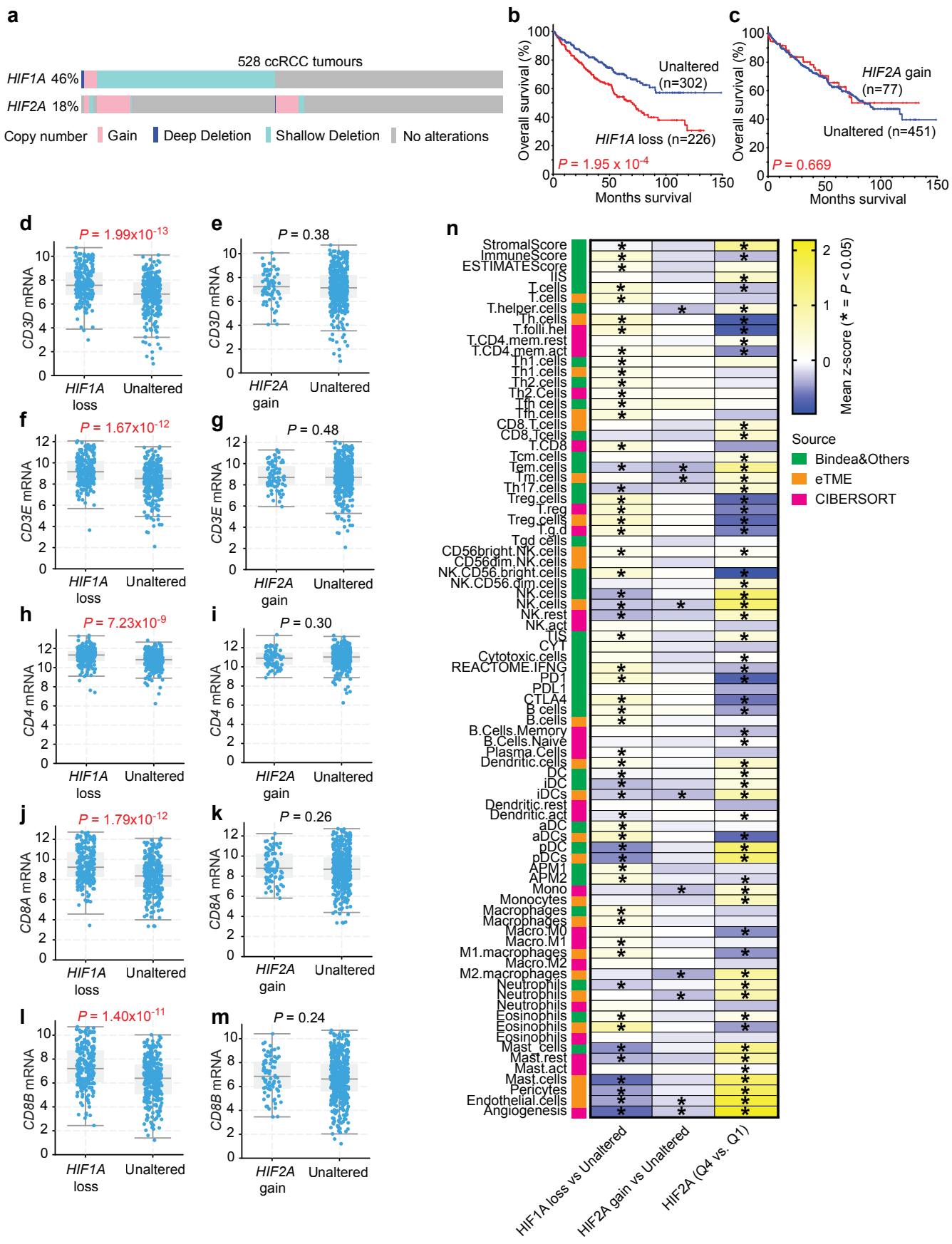

a
WT
Vhl fl/fl
Vhl fl/fl Hif1a fl/fl
Vhl fl/fl Hif2a fl/fl
Vhl fl/fl Hif1a fl/fl Hif2a fl/fl
GFP
Cre
Population doublings
Days
b
Vhl fl/fl Trp53 fl/fl
GFP shRNA-ns
Cre shRNA-ns
GFP shRNA-Hif1a #222
Cre shRNA-Hif1a #222
GFP shRNA-Hif1a #220
Cre shRNA-Hif1a #220
c
mouse ccRCC 2020
VHL30
Vector
A510
Time (h)
d
mouse ccRCC 2020
shRNA-Hif1a #220
shRNA-Hif1a #222
shRNA-ns
e
mouse ccRCC 2020
Vector
VHL30
shRNA-ns
shRNA-Hif1a #220
shRNA-Hif1a #222
f
mouse ccRCC 2020
P = 0.00002
Sphere area [arbitrary pixels x 10-5]
Vector (n=26)
VHL30 (n=85)
g
mouse ccRCC 2020
P = 0.005
P = 0.002
shRNA-ns (n=24)
shRNA-Hif1a #220 (n=50)
shRNA-Hif1a #222 (n=47)
h
VHL30 (n=8)
Vector (n=8)
P = 0.036
Survival (%)
Days
i
shRNA-Hif1a #220 (n=9)
shRNA-Hif1a #222 (n=9)
shRNA-ns (n=9)


a
WT cortex
VpR
VpRH1
VpRH2
PC2 (9.35 %)
PC1 (36.2 %)
b
Glycolysis
c
Cell adhesion molecules
d
Focal adhesion & receptor signalling
e
DNA repair
f
Lipoprotein metab.
g
Ribosome biogenesis
h
T cell activation
i
Response to IFN-b
z-score


a
r = 0.66; p<0.0001
Average protein abundance mouse ccRCC (n=6)
Average protein abundance human ccRCC (n=8)
b
Differentially expressed proteins human ccRCC / normal
364
427
95
Differentially expressed proteins mouse ccRCC / normal
c
-log10 (limma adjusted p-value)
mean log2 (VpR / WT Cortex)
d
-log10 (limma adjusted p-value)
mean log2 (VpRH1 / VpR)
e
-log10 (limma adjusted p-value)
mean log2 (VpRH2 / VpR)
f
HIF-1α-dependent
WT Cortex
VpR
VpRH1
VpRH2
Row-normalised z-score
Not detected
GSEA Summary
Hypoxia
mTORC1 Signalling
Glycolysis
Unfolded Protein Response
GSEA Summary
Oxidative Phosphorylation
Complex I Biogenesis
Respiratory Electron Transport
g
HIF-2α-dependent
WT Cortex
VpR
VpRH1
VpRH2
GSEA Summary
Heme Metabolism
Amino acid Metabolism
RNA Metabolism
MYC Targets
GSEA Summary
Interferon γ Response
Interferon α Response
Innate Immune System
Antigen Presentation


a
WT Cortex
VpR
VpRH1
VpRH2
MHC I
z-score
b
MHC II
c
Antigen processing and presentation
d
α-MHC II
Score 1
Score 2
Score 3
e
P = 0.0002
P = 0.0002
P = 0.68
% tumours
MHC II staining
Score 1
Score 2
Score 3
VpR (n=26)
VpRH1 (n=13)
VpRH2 (n=15)
f
HIF1A mRNA abundance
KICH
KIRC
KIRP
Normal_KICH
Normal_KIRC
Normal_KIRP
g
HIF2A mRNA abundance
h
Spearman: -0.24, P = 3.5 x10-8
HLA-A mRNA abundance
HIF2A mRNA abundance
j
Spearman: -0.19, P = 5.8 x10-6
HLA-DMB mRNA abundance
l
Spearman: -0.48, P = 1.7 x10-32
PSME2 mRNA abundance
i
Spearman: -0.15, P = 6.6 x10-4
HLA-B mRNA abundance
k
Spearman: -0.15, P = 4.2 x10-4
HLA-DQB2 mRNA abundance
m
Spearman: -0.32, P = 1.5 x10-14
FCER1G mRNA abundance


a
Immune signature
IIS
ImmuneCC
T.cells
T.cells
T.helper.cells
Th.cells
T.Cells.CD4.Follicular
T.Cells.CD4.Memory
T.Cells.CD4.Naive
Th1.cells
Th1.Cells
Th1.cells
Th2.cells
Th2.Cells
Th2.cells
Tfh.cells
Tfh.cells
CD8.T.cells
CD8.Tcells
T.Cells.CD8.Activated
T.Cells.CD8.Memory
T.Cells.CD8.Naive
Tcm.cells
Tem.cells
Tm.cells
Th17.cells
Th17.Cells
Treg.cells
Treg.Cells
Treg.cells
GammaDelta.T.Cells
Tgd.cells
CD56bright.NK.cells
CD56dim.NK.cells
NK.CD56.bright.cells
NK.CD56.dim.cells
NK.cells
NK.cells
NK.Resting
TIS
CYT
Cytotoxic.cells
REACTOME.IFNG
PD1
PDL1
CTLA4
B.cells
B.cells
B.Cells.Memory
B.Cells.Naive
Plasma.Cells
Dendritic.cells
DC
iDC
iDCs
DC.Immature
DC.Activated
aDC
aDCs
pDC
pDCs
APM1
APM2
Monocyte
Monocytes
Macrophages
Macrophages
M0.Macrophage
M1.Macrophage
M1.macrophages
M2.Macrophage
M2.macrophages
Neutrophils
Neutrophils
Eosinophils
Eosinophils
Mast.Cells
Mast.cells
Pericytes
Endothelial.cells
VpR / WT Cortex
VpRH1 / WT Cortex
VpRH2 / WT Cortex
VpRH1 / VpR
VpRH2 / VpR
VpRH2 / VpRH1
Mean z-score (* = P < 0.05)
-1
0
1
Source
Bindea&Others
eTME
ImmuCC
b
CD3+ cells / mm2
P = 0.09
P = 0.04
P = 0.5
P = 0.82
P = 0.27
P = 0.005
c
CD4+ cells / mm2
P = 0.29
P = 0.02
P = 0.13
P = 0.14
P = 0.08
P = 0.03
d
CD8+ cells / mm2
P = 0.02
P = 0.04
P = 0.3
P = 0.002
P = 0.02
P < 0.0001
e
CD69+ cells / mm2
P = 0.003
P = 0.82
P = 0.005
P = 0.0001
P = 0.002
P = 0.0001
f
Perforin+ cells / mm2
P = 0.04
P = 0.31
P = 0.005
P = 0.0001
P = 0.002
P = 0.0001
g
PD-1+ cells / mm2
P = 0.96
P = 0.72
P = 0.92
P = 0.11
P = 0.96
P = 0.06
h
B220+ cells / mm2
P = 0.08
P = 0.0003
P < 0.0001
P = 0.002
P = 0.001
P = 0.098
i
CD68+ cells / mm2
P = 0.87
P = 0.13
P = 0.08
P < 0.0001
P = 0.001
P < 0.0001
j
F4/80+ (%PP)
P = 0.97
P = 0.23
P = 0.13
P = 0.0004
P = 0.002
P = 0.0005
k
Ly-6G+ cells / mm2
P = 0.92
P = 0.82
P = 0.65
P < 0.0001
P = 0.0004
P < 0.0001
Normal
Tumour
VpR
VpRH1
VpRH2


a
528 ccRCC tumours
HIF1A 46%
HIF2A 18%
Copy number
Gain
Deep Deletion
Shallow Deletion
No alterations
b
Overall survival (%)
Unaltered (n=302)
HIF1A loss (n=226)
P = 1.95 x 10-4
Months survival
c
Overall survival (%)
HIF2A gain (n=77)
Unaltered (n=451)
P = 0.669
Months survival
d
P = 1.99x10-13
CD3D mRNA
HIF1A loss
Unaltered
e
P = 0.38
CD3D mRNA
HIF2A gain
Unaltered
f
P = 1.67x10-12
CD3E mRNA
HIF1A loss
Unaltered
g
P = 0.48
CD3E mRNA
HIF2A gain
Unaltered
h
P = 7.23x10-9
CD4 mRNA
HIF1A loss
Unaltered
i
P = 0.30
CD4 mRNA
HIF2A gain
Unaltered
j
P = 1.79x10-12
CD8A mRNA
HIF1A loss
Unaltered
k
P = 0.26
CD8A mRNA
HIF2A gain
Unaltered
l
P = 1.40x10-11
CD8B mRNA
HIF1A loss
Unaltered
m
P = 0.24
CD8B mRNA
HIF2A gain
Unaltered
n
StromalScore
ImmuneScore
ESTIMATEScore
IIS
T.cells
T.cells
T.helper.cells
Th.cells
T.folli.hel
T.CD4.mem.rest
T.CD4.mem.act
Th1.cells
Th1.cells
Th2.cells
Th2.Cells
Tfh cells
Tfh.cells
CD8.T.cells
CD8.Tcells
T.CD8
Tcm.cells
Tem.cells
Tm.cells
Th17.cells
Treg.cells
T.reg
Treg.cells
T.g.d
Tgd cells
CD56bright.NK.cells
CD56dim.NK.cells
NK.CD56.bright.cells
NK.CD56.dim.cells
NK.cells
NK.cells
NK.rest
NK.act
CYT
CYT
Cytotoxic.cells
REACTOME.IFNG
PD1
PDL1
CTLA4
B.cells
B.cells
B.Cells.Memory
B.Cells.Naive
Plasma.Cells
Dendritic.cells
DC
iDC
iDCs
Dendritic.rest
Dendritic.act
aDC
aDCs
pDC
pDCs
APM1
APM2
Mono
Monocytes
Macrophages
Macrophages
Macro.M0
Macro.M1
M1.macrophages
Macro.M2
M2.macrophages
Neutrophils
Neutrophils
Neutrophils
Eosinophils
Eosinophils
Eosinophils
Mast.cells
Mast.rest
Mast.act
Mast.cells
Pericytes
Endothelial.cells
Angiogenesis
HIF1A loss vs Unaltered
HIF2A gain vs Unaltered
HIF2A (Q4 vs. Q1)
Mean z-score (* = P < 0.05)
Source
Bindea&Others
eTME
CIBERSORT

3.3 Synonymous *GATA2* Mutations Result in Selective Loss of Mutated RNA and are Common in Patients with *GATA2* Deficiency

One of the main genetic disorders causing an immunodeficiency / myelodysplasia syndrome manifesting in a multitude of clinical phenotypes, like acute myeloid leukemia (AML) or familial myelodysplastic syndrome (MDS), is a germline mutation in the *GATA2* gene [Hahn et al., 2011]. *GATA2* deficiency is considered the most hereditary predisposition to pediatric MDS, accounting for 15% of MDS cases [Wlodarski et al., 2016]. Three major types of pathogenic *GATA2* mutations are described in literature: (i) missense mutations in the zinc finger 2 domain, (ii) splice site, nonsense, frameshift, and whole gene deletions, and (iii) noncoding substitutions in intron 4 leading to haploinsufficiency [Hirabayashi et al., 2017, Wlodarski et al., 2017, Hsu et al., 2013]. Silent mutations are mostly overlooked and only little is known about them. Although silent mutations are not affecting the protein structure or their stability it is reported that such variants can alter RNA by pre-mRNA splicing, mRNA stability and structure, miRNA binding, and translation [D'Souza et al., 1999, Macaya et al., 2009]. Initial a synonymous substitution in exon 3 of the *GATA2* gene with the clinical phenotype of *GATA2* deficiency was identified [Wehr et al., 2018]. Further analyses of our cohort with the focus on silent *GATA2* substitutions revealed five distinct silent *GATA2* mutations leading to a RNA deleterious effect in nine patients. We concluded that silent *GATA2* substitutions are a common, mostly overlooked, cause of *GATA2* deficiency and should affect the evaluation of pathogenic variant discovery and genetic counseling.

Kozyra, E. J., Pastor, V. B., Lefkopoulos, S., Sahoo, S. S., Busch, H., Voss, R. K., Erlacher, M., Lebrecht, D., Szvetnik, E. A., Hirabayashi, S., Pasaulienė, R., Pedace, L., Tartaglia, M., Klemann, C., Metzger, P., Boerries, M., Catala, A., Hasle, H., de Haas, V., Wlodarski, M. W. (2020). **Synonymous *GATA2* mutations result in selective loss of mutated RNA and are common in patients with *GATA2* deficiency**. Leukemia.

Contribution: I supported the analyses of the whole exome sequencing data of the patients. This included the quality control steps of the raw sequencing files, followed

by the alignment to the reference genome, the variant calling and the annotation of the identified variants. Adjustments to the pipeline were necessary to reliable detect the silent *GATA2* substitutions. I revised the manuscript.

Leukemia
https://doi.org/10.1038/s41375-020-0899-5

ARTICLE

Genomics and gene therapy

Synonymous *GATA2* mutations result in selective loss of mutated RNA and are common in patients with GATA2 deficiency

Emilia J. Kozyra[1,2] · Victor B. Pastor[1] · Stylianos Lefkopoulos[2,3] · Sushree S. Sahoo[1,4] · Hauke Busch[5,6,7] · Rebecca K. Voss[1] · Miriam Erlacher[1,8,9] · Dirk Lebrecht[1] · Enikoe A. Szvetnik[1] · Shinsuke Hirabayashi[1,10] · Ramunė Pasaulienė[11] · Lucia Pedace[12] · Marco Tartaglia[13] · Christian Klemann[14] · Patrick Metzger[2,15] · Melanie Boerries[8,9,15] · Albert Catala[16] · Henrik Hasle[17] · Valerie de Haas[18] · Krisztián Kállay[19] · Riccardo Masetti[20] · Barbara De Moerloose[21] · Michael Dworzak[22] · Markus Schmugge[23] · Owen Smith[24] · Jan Starý[25] · Ester Mejstrikova[25] · Marek Ussowicz[26] · Emma Morris[27,28,29] · Preeti Singh[30,31] · Matthew Collin[30,31] · Marta Derecka[3] · Gudrun Göhring[32] · Christian Flotho[1,8,9] · Brigitte Strahm[1] · Franco Locatelli[12,33] · Charlotte M. Niemeyer[1,8,9] · Eirini Trompouki[3,34] · Marcin W. Wlodarski[1,4] · European Working Group of MDS in Childhood (EWOG-MDS)

Received: 18 September 2019 / Revised: 19 May 2020 / Accepted: 29 May 2020

Abstract

Deficiency of the transcription factor GATA2 is a highly penetrant genetic disorder predisposing to myelodysplastic syndromes (MDS) and immunodeficiency. It has been recognized as the most common cause underlying primary MDS in children. Triggered by the discovery of a recurrent synonymous *GATA2* variant, we systematically investigated 911 patients with phenotype of pediatric MDS or cellular deficiencies for the presence of synonymous alterations in *GATA2*. In total, we identified nine individuals with five heterozygous synonymous mutations: c.351C>G, p.T117T ($N = 4$); c.649C>T, p.L217L; c.981G>A, p.G327G; c.1023C>T, p.A341A; and c.1416G>A, p.P472P ($N = 2$). They accounted for 8.2% (9/110) of cases with GATA2 deficiency in our cohort and resulted in selective loss of mutant RNA. While for the hotspot mutation (c.351C>G) a splicing error leading to RNA and protein reduction was identified, severe, likely late stage RNA loss without splicing disruption was found for other mutations. Finally, the synonymous mutations did not alter protein function or stability. In summary, synonymous *GATA2* substitutions are a new common cause of GATA2 deficiency. These findings have broad implications for genetic counseling and pathogenic variant discovery in Mendelian disorders.

Introduction

Germline mutations in the *GATA2* gene, mostly arising de novo, had been reported to cause an immunodeficiency/myelodysplasia syndrome manifesting with a multitude of clinical phenotypes. These include monocytopenia and mycobacterial infections syndrome (MonoMAC syndrome) [1], dendritic cell, monocyte, B and NK lymphoid deficiency (DCML deficiency) [2], familial myelodysplastic syndrome (MDS)/acute myeloid leukemia (AML) [3], chronic neutropenia [4], Emberger syndrome [5] and warts, immunodeficiency, lymphedema and anogenital dysplasia syndrome (WILD syndrome) [6]. Finally, GATA2 deficiency is considered the most common hereditary predisposition to pediatric MDS, accounting for as much as 15% of MDS with excess of blasts (MDS-EB), with a particularly high prevalence among MDS patients carrying monosomy 7 (37%) [7]. To date, more than 400 GATA2-deficient cases have been published [8, 9] with three major types of pathogenic *GATA2* mutations: (1) missense mutations within zinc finger 2 (ZnF2), (2) null mutations (splice site, nonsense, frameshift, and whole gene deletions), and (3) noncoding substitutions in the EBOX-GATA-ETS regulatory region in intron 4 (hg19, g.128202128-128202173, NM_032638.4) [8–10]. Overall, germline *GATA2* mutations are thought to

Supplementary information The online version of this article (https://doi.org/10.1038/s41375-020-0899-5) contains supplementary material, which is available to authorized users.

✉ Marcin W. Wlodarski
marcin.wlodarski@stjude.org

Extended author information available on the last page of the article

result in haploinsufficiency and context-dependent loss of essential transcription factor activity [3, 5, 11–14].

Genomic studies typically focus on the discovery of nonsynonymous variants that alter coding regions or canonical splice sites because their effect is predictable. Conversely, due to codon degeneracy, synonymous substitutions do not alter the amino acid composition of the encoded protein and are usually not reported as pathogenic. However, previous studies revealed that such variants can alter RNA or protein on multiple levels including pre-mRNA splicing, messenger RNA (mRNA) stability and structure, miRNA binding, and translation [15–24].

Here, we initially identified a synonymous substitution in exon 3 of the *GATA2* gene (c.351C>G, p.T117T) in two unrelated pedigrees, with the clinical phenotype of GATA2 deficiency. The variant was recently reported in an adult patient (the mother of two siblings studied here) presenting with immunodeficiency, severe infections and lung disease [25]. This prompted us to study the contribution of synonymous alterations to the genetic spectrum of GATA2 deficiency and to assess their pathogenic role. We discovered and characterized five distinct synonymous mutations with RNA-deleterious effect in nine patients. They represent a new type of mutation in GATA2 deficiency and have broad implications for both the discovery of disease-causing mutations and genetic counseling.

Methods

Patient cohort and genomics

The screening cohort consisted of 911 patients (Fig. 1a): 729 children and adolescents with primary MDS classified according to WHO criteria [26–28] enrolled in the studies 1998 and 2006 of the European Working Group of MDS in Childhood (EWOG-MDS, #NCT00662090), and 182 patients with cytopenias and/or GATA2-specific clinical problems, referred to our diagnostic laboratory. *GATA2* gene sequence, including intron 4 was analyzed in bone marrow (BM) samples using targeted deep sequencing with Sanger sequence validation, and subsequent confirmation of germline mutational status in nonmyeloid tissues as previously reported [7, 29]. Whole exome/genome sequencing (WES/WGS) was performed in patients with synonymous *GATA2* variants to rule out other hereditary causes (Supplementary methods, Supplementary Table 1).

Targeted investigations of GATA2 transcript expression

We analyzed RNA expression in blood, BM or fibroblasts using Sanger, deep sequencing, and TA cloning-based sequencing (Supplementary methods, Supplementary Fig. 1 and Supplementary Table 2). In addition, *GATA2* expression in various hematopoietic compartments of healthy controls was measured (Supplementary methods).

Studies of GATA2 protein stability and function

In order to explore the influence of synonymous mutations on protein stability and function, in vitro analysis of exogenously expressed *GATA2* was performed in 293T cells. To further investigate the protein function, in vivo studies in zebrafish were accomplished (for details see Supplementary methods). Experiments were performed in duplicates or triplicates as indicated in the figure legends.

Statistics

For reporter assay, data from biological and technical triplicate experiments were presented as the mean values ± standard deviation (SD). Statistical significance was assessed using GraphPad Prism v 7.04 software employing either standard one-way ANOVA test (reporter assay, thermodynamic effect of GATA2 variants) or Student's *t* test (allele quantification in patients' cDNA by deep sequencing, frequency of zebrafish phenotypes). P values < 0.05 were considered statistically significant.

Results

Identification of synonymous GATA2 variants

We initially discovered two unrelated individuals (P1, P3) with GATA2 deficiency carrying an identical synonymous *GATA2* variant. This prompted a systematic evaluation of the *GATA2* gene sequence in our screening cohort of patients presenting for the most part with the phenotype of pediatric MDS (Fig. 1a). At first, we categorized "classical" disease-causing alterations and identified 101 patients with 62 distinct pathogenic *GATA2* mutations (Fig. 1a). The distribution of mutations corroborated data reported in previous studies [9]. The most common were null mutations affecting the N-terminal part of the protein: stop-gain, frameshift, splice site ($N = 52$), followed by missense mutations within or adjacent to ZnF2 ($N = 36$), intron 4 EBOX-GATA-ETS site alterations ($N = 10$), and other aberrations ($N = 3$): one in-frame and two whole gene deletions (Fig. 1b).

Next, we searched *GATA2* coding sequence for the presence of synonymous substitutions. Variants that are either not reported or very rare (<0.05% allele frequency) in the gnomAD population database were found in nine patients. These variants were present in 8.2% (9/110) of all patients with *GATA2* alterations, and 14.8% (9/61) of cases with

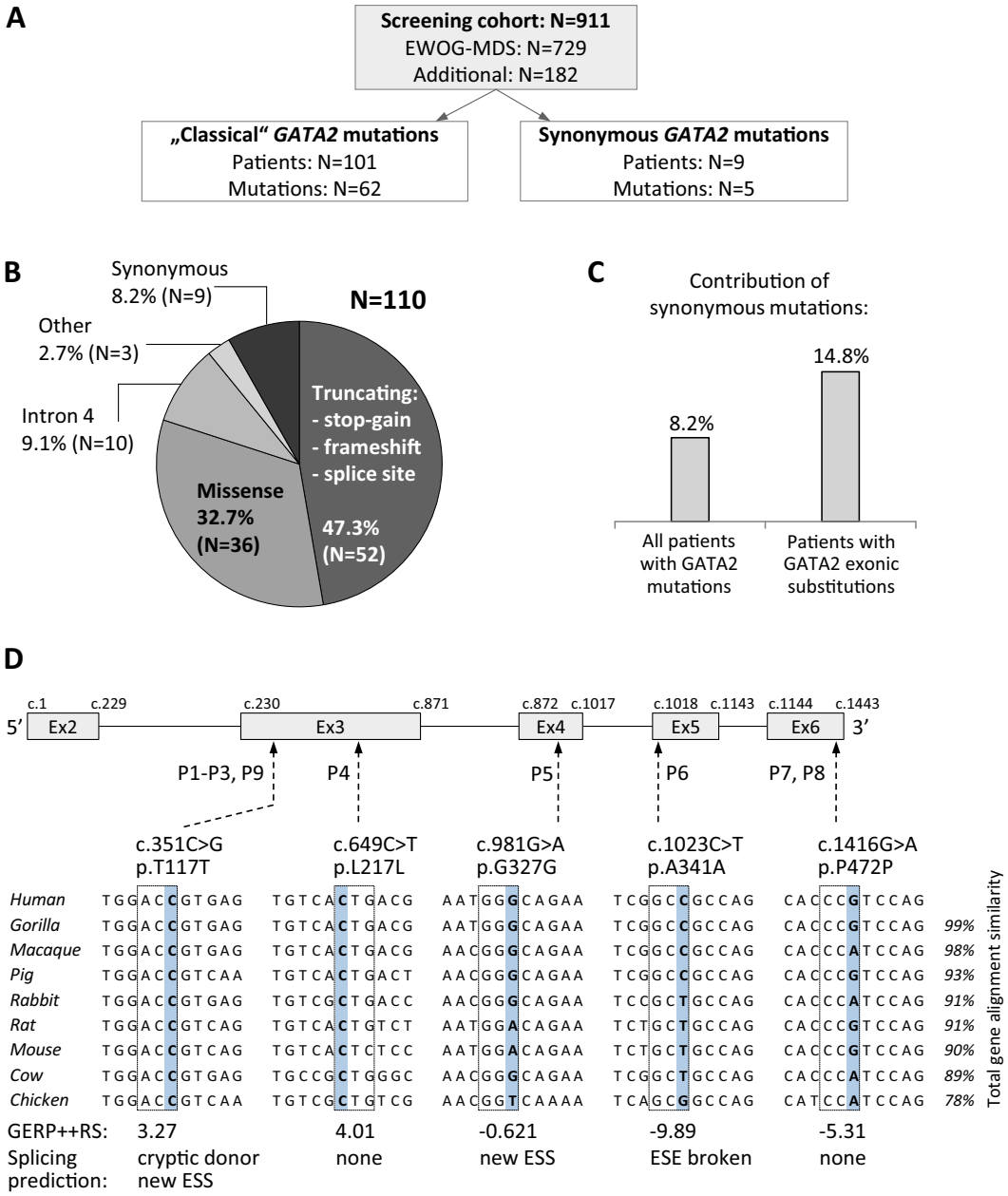


Fig. 1 Composition and genetics of the study cohort. a Flow diagram depicts the screening cohort and *GATA2* mutations identified. **b** Overall distribution of genotypes among 110 patients with GATA2 deficiency. Truncating variants are localized prior to or within zinc finger 2; missense mutations cluster mainly to zinc finger 2 region; intron 4 mutations affect the EBOX-GATA-ETS regulatory region (+9.5 kb) of *GATA2*; other: one in-frame deletion and two whole gene deletions; synonymous variants are proposed as a new group of pathogenic *GATA2* mutations. Numbers in parentheses refer to individual patients. **c** Frequency of patients with synonymous mutations among all *GATA2* positive cases and among the group of patients carrying exonic substitutions. **d** Schematic representation of the *GATA2* gene (NM_032638.4) with synonymous variants identified. Affected nucleotide is shaded blue and dashed line boxes indicate respective codon triplet. Nucleotide conservation is presented for nine species. Evolutionary conservation is depicted on the bottom as Genomic Evolutionary Rate Profiling (GERP++RS) score with values ranging from −12.36 to 6.18, and 6.18 being the most conserved. Splicing prediction was performed with Human Splicing Finder v.3.0. ESS exonic splicing silencer, ESE exonic splicing enhancer.

GATA2 exonic substitutions only (Fig. 1c). In comparison, common polymorphisms with synonymous effect p.P5P, p.P22P, p.Q38Q, p.T188T, and p.A411A were not significantly enriched in our cohort (not shown), arguing against their disease-causing role in MDS.

The synonymous substitutions encountered in P1–P9 were predicted to have a likely benign effect using the combined annotation-dependent depletion score (CADD) and gene-specific calibration by Gene-Aware Variant INterpretation (GAVIN) (Table 1). The evolutionary

Table 1 Synonymous variants identified in MDS patients.

Patient no. (ID)	*GATA2* mutation	Genomic DNA VAF% (total depth)	cDNA VAF% WT/Mut (total depth)	Evol conser/ PhysChem diff	CADD/GAVIN (C2; $P > 26$, $B <$ 19)	gnomAD browser MAF% (mutant/ total)	Splicing prediction
P1 (D 1239)	c.351C>G; p.T117T	^{WES}50% (160), ^{DS}47% (766)	^{DS}99.88%/0.006% (320215)	Medium/none	10.5/B	None	Cryptic donor, new ESS site
P2 (sister of P1)	c.351C>G; p.T117T	Heterozygous (Sanger)	Not done	As above	As above	As above	As above
P3 (D 749)	c.351C>G; p.T117T	^{WES}47% (130), ^{DS}51% (703)	^{DS}99.89%/0.004% (838094)	As above	As above	As above	As above
P4 (LT)	c.649C>T; p.L217L	^{WES}48% (196), ^{DS}51% (725)	^{DS}79.43%/20.56% (429616)	High/none	10.7/B	0.001% (2/246096)	None
P5 (D 722)	c.981G>A; p.G327G	^{WES}52% (156), ^{DS}48% (788)	^{DS}99.96%/0.014% (7190)	Medium/none	18.5/B	None	New ESS site
P6 (D 1142)	c.1023C>T; p.A341A	^{WES}47% (296), ^{DS}49% (1449)	^{DS}99.91%/0.095% (4213)	Weak/none	15.4/B	0.002% (6/275438)	ESE site broken
P7 (D)	c.1416G>A; p.P472P	^{WES}49% (63)	^{DS}99.91%/0.085% (388784)	Medium/none	12.4/B	0.027% (70/256322)	None
P8 (I 386)	c.1416G>A; p.P472P	Heterozygous (Sanger)	Not done	As above	As above	As above	As above
P9 (UKA2604)	c.351C>G; p.T117T	Heterozygous (Sanger)	Not done	Medium/none	10.5/B	None	Cryptic donor, new ESS site

Gene annotation: *GATA2* (NM_032638.4).

VAF variant allelic frequency, *WT* wild-type allele, *Mut* mutated allele, *WES* whole exome sequencing, *DS* deep sequencing, *Sanger* identified by Sanger sequencing, *Evol* (evolutionary) conservation assessed using Phylop and PhastCons, *PhysChem diff* physicochemical difference between amino acids, *CADD* combined annotation-dependent depletion score, *GAVIN* Gene-Aware Variant Interpretation (C2: CADD scores significantly predictive for pathogenicity ($p < 0.05$), *P* pathogenic if CADD > 26, *B* benign if CADD lower than 19), *MAF* minor allelic frequency, *Splicing prediction* Human Splicing Finder v. 3.0. *ESS* exonic splicing silencer, *ESE* exonic splicing enhancer.

nucleotide conservation was high for c.351 and c.649 nucleotides (Fig. 1d), suggesting their resistance to evolutionary change. Splicing prediction tools assigned a high chance of splice defects to c.351C>G, c.981G>A, and c.1023C>T variants either via activation of a cryptic donor, introduction of an alternative splicing silencer or disruption of an existing splicing enhancer (Fig. 1d, Supplementary Fig. 2, and Table 1).

Phenotype of patients with synonymous *GATA2* mutations

Patients with synonymous *GATA2* mutations were diagnosed at a median age of 11.5 (3–24) years. Hematologic and immunological phenotypes were consistent with the heterogeneous clinical picture of GATA2 deficiency and included varying degrees of immune cytopenias (low B/NK, DC cells, monocytopenia), immunodeficiency, neutropenia, and/or pancytopenia (supplemental case descriptions). P2 is the sibling of GATA2-deficient patient P1 and was categorized as a silent *GATA2* mutation carrier with a reduction of B- and NK-cells. Their mother was previously reported with pulmonary alveolar proteinosis [25]. P7 and P8 (unrelated, carrying the same mutation), initially presented with thrombocytopenia and while P8 developed transfusion-dependent refractory cytopenia of childhood (RCC), P7 remained stable with BM morphology suspicious for RCC. P9 was first seen with complications of immunodeficiency and clinically evolved to MDS. Monosomy 7 in BM was detected at diagnosis in four patients (P1, P3, P4, and P6), normal karyotypes were present in four (P5, P7, P8, and P9), while no marrow exam was performed in P2 (Table 2). According to the WHO classification, P1 and P4–P8 were diagnosed with RCC, and P9 with MDS with multilineage dysplasia as a young adult. Initial disease of P3 was MDS-EB, which progressed to AML after 6 months. Other clinical problems in the affected patients were transient organ dysfunction after birth and facial abnormalities in P4, hepatosplenomegaly in P5, hypospadias in P6, and Crohn's colitis as well as HPV-driven neoplasia in P9. The majority of patients (6/9) underwent allogeneic hematopoietic stem cell transplantation (HSCT) with favorable outcome: 5/6 patients were alive at last follow up (at a median of 1.9 years after HSCT) and 1/6 (P3) died from infection 7 months following HSCT (Table 2).

Exclusion of other hereditary causes

We next aimed to determine if other genetic conditions predisposing to inherited bone marrow failure (IBMF) or

Table 2 Clinical characteristics of patients with synonymous *GATA2* mutations.

Patient no. (ID)	Age at Dx	Sex	Hematological presentation and other features	Karyotype	Therapies	Age and status at last FUP
P1 (D 1239)	12	F	RCC, low IgG, low monocytes/B/DC	−7	MUD-HSCT	13.5 years: alive
P2 (sister of P1)	11	F	B/NK-cell lymphopenia, low IgA/G	Not done	Observation	12 years: alive
P3 (D 749)	14	F	MDS-EB	−7	CB-HSCT	15.4 years: died from infection 7 months after HSCT
P4 (LT)	3	M	RCC, facial abnormalities, skin hypopigmentation, joint hypermobility	−7	MSD-HSCT	7.6 years: alive
P5 (D 722)	11	M	RCC, hepatosplenomegaly	Normal	Observation	18.5 years: alive
P6 (D 1142)	11.5	M	RCC, hypospadias	−7	MUD-HSCT	15.2 years: alive
P7 (D)	14	F	Suspicious for RCC	Normal	–	20.9 years: alive
P8 (I 386)	4	F	RCC	Normal	MUD-HSCT	4 years: alive
P9 (UKA2604)	24	F	MDS-MLD, low monocytes/B/NK/DC, recurrent viral warts, mycobacterium avium infections, HPV-driven neoplasia, Crohn's colitis	Normal	MUD-HSCT	32 years: alive

Dx diagnosis, *RCC* refractory cytopenia of childhood, *DC* dendritic cells, *NK* natural killer cells, *MDS-EB* myelodysplastic syndrome with excess blast, *MDS-MLD* myelodysplastic syndrome with multilineage dysplasia, *HSCT* allogeneic hematopoietic stem cell transplantation, *−7* monosomy 7, *MUD* matched unrelated donor, *CB* cord blood, *MSD* matched sibling donor, *FUP* follow-up.

MDS might have been previously missed in our patients. WES/WGS was performed in all families with exception of P8 who was assessed by a 135 IBMFS/MDS gene panel. The WES analysis focused on known IBMF/MDS and pancancer genes (300 genes) [26, 30]. Multiple heterozygous variants of uncertain significance (VUS) were identified (Supplementary Table 1). After comprehensive review by a multidisciplinary board representing pediatric hematology, genetic counseling, and molecular biology, only P6 remained with additional potentially pathogenic *SAMD9* variants p.K877E and p.F366LfsX33. This patient did not have features typical for MIRAGE syndrome, which was initially ascribed to *SAMD9* mutations [31, 32]. Notably, we discovered VUS in the Fanconi anemia (FA) genes *FANCD1, FANCD2*, and *FANCS* in three patients. However, these VUS were heterozygous and FA was ruled out in all three patients by means of chromosomal breakage studies and clinical phenotyping.

Synonymous *GATA2* variants result in selective loss of mRNA expression

Building on the assumption that synonymous variants detected in our patients were associated with degradation of the mutant (Mut) mature mRNA, we first sequenced cDNA transcribed from polyadenylated RNA transcripts (equivalent to mRNA) using Sanger method. Compared with genomic DNA, cDNA sequences showed loss of heterozygosity manifested by complete lack of the Mut allele in five out of seven cases: P1, P3, P5, P6, and P7, and a substantial reduction in P4 and P9 (Fig. 2a upper panel). Compared with hematopoietic specimens, Mut allele expression was slightly higher in skin fibroblasts of P1 and P4 (Fig. 2a lower panel). Because it is not known if monoallelic *GATA2* expression might be a general phenomenon in normal hematopoiesis, we sequenced three healthy controls who carried a common heterozygous polymorphism (rs2335052: c.490G>A; p.A164T). Both the genomic DNA and cDNA showed an equal ratio of alternative to reference alleles (Fig. 2b).

Deep sequencing based quantification of allelic frequency showed nearly total absence of Mut alleles in P1, P3, P5–P7, and a reduction of Mut expression to 21% in P4 (Table 1 and Fig. 2c). Combined across all samples, we observed median values of 27 reads for Mut, versus 330,544 reads for wild-type (WT) alleles. Lastly, TA cloning of P5's and P6's cDNA followed by sequencing of an average of 345 single colonies was a third independent method confirming the RNA reduction (0% and 11% of Mut amplicons for P5 and P6, respectively, not shown). In order to address at which stage of RNA maturation the Mut alleles were lost, we deep sequenced products that were reverse-transcribed using alternative priming approaches. While oligo(dT) that are specific to mature transcripts (mRNA) produced almost exclusively *GATA2* WT reads, the use of random hexamers (enriching both pre-mRNA and mRNA) resulted in an increase of Mut reads to ~30% for P1 and P6 (Fig. 2d).

Splicing analysis of the *GATA2* gene

In order to ascertain the mechanism of monoallelic *GATA2* expression, RNA sequencing (RNAseq) was performed in sorted CD34+ BM cells of five patients (P1, P4–P7).

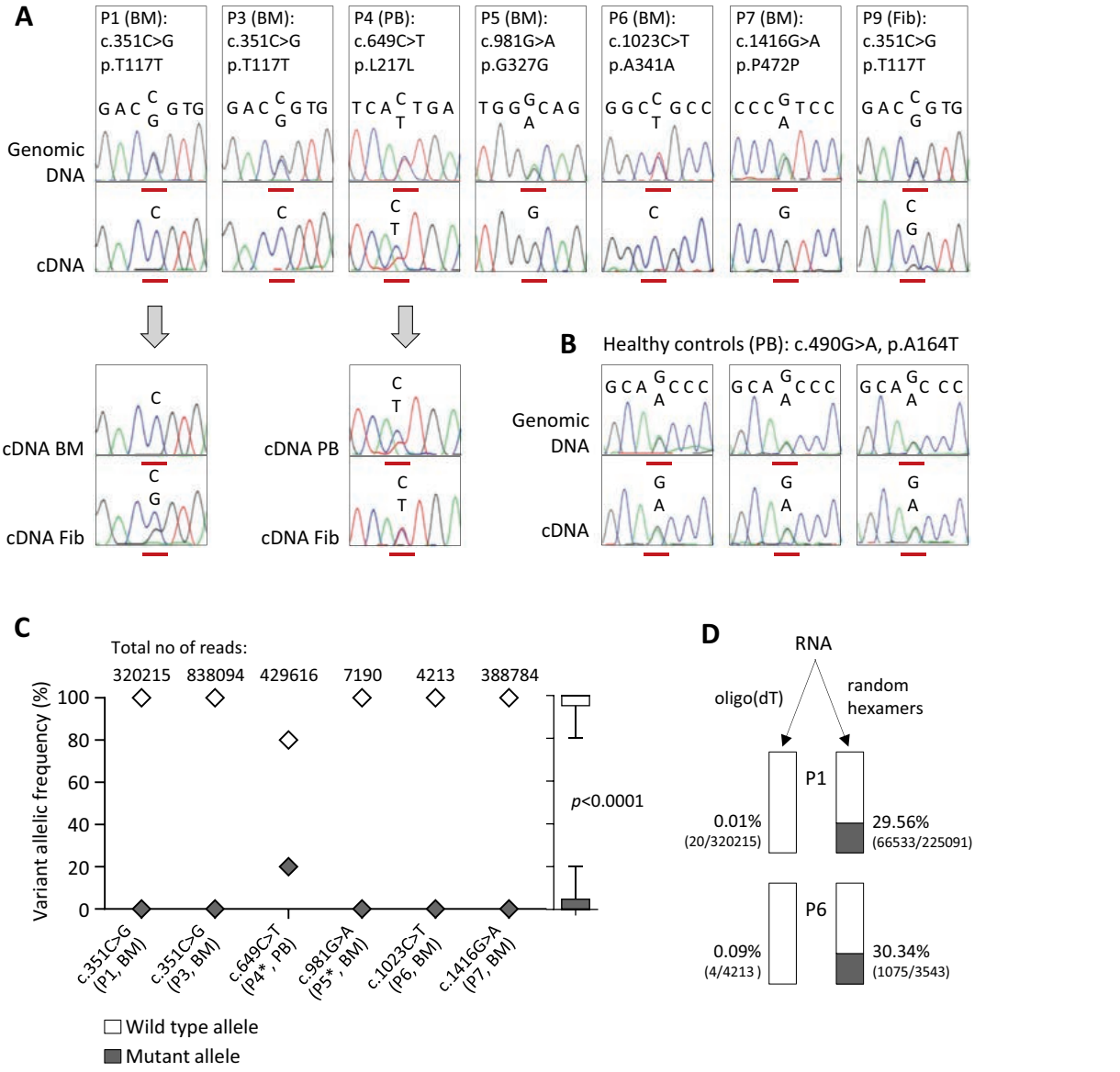


Fig. 2 Sequencing analysis of patients with synonymous *GATA2* mutations. a Upper panel: Representative electropherograms with genomic and cDNA sequence surrounding the affected nucleotide (red line). All five distinct synonymous mutations are represented. Lower panel: Comparison of allelic expression in the hematopoietic and nonhematopoietic tissue of P1 and P4. **b** Genetic testing of healthy controls carrying a common nonsynonymous *GATA2* polymorphism c.490G>A (19.5% minor allele frequency in gnomAD). **c** Allelic frequency of *GATA2* alleles determined by targeted deep sequencing of patients' cDNA; numbers of reads taken from one representative replicate. In all cases oligo(dT) priming was used with exception of P4 and P5 (*) where the mixture of random hexamers and oligo(dT) was utilized. Outside right: Boxplot depicts combined allelic contribution in all patients. Calculation of *p* value was performed using Student's *t* test (mean ± SD values). **d** Frequency of mutated alleles determined by deep sequencing of cDNA obtained from bone marrow RNA using two different reverse transcription priming methods. Mutant vs. total read counts are shown in parentheses, and percentage represents the proportion of mutated alleles in the sample. BM bone marrow, Fib fibroblasts, PB peripheral blood.

Isoform analysis revealed two novel splice junctions in P1, not observed in the Ensembl database and healthy controls (Fig. 3a and Supplementary Fig. 3). In both new transcripts in P1 the c.351C>G mutation acts as a new splice donor that joins to alternative acceptors either at c.488 or at c.608. Long range RT-PCR and sequencing in P1′ BM and fibroblasts (Fig. 3b) confirmed the presence of the transcript with c.488 alternative acceptor. Finally, TA cloning of the cDNA PCR products of P1 and sequencing of 348 colonies revealed the presence of three novel transcripts (Fig. 3c). Two of these were identical as detected by RNAseq; the third transcript found in only nine colonies harbors the c.351 donor that joins to a new splice acceptor at position c.539. All three transcripts resulted in sequence frameshift

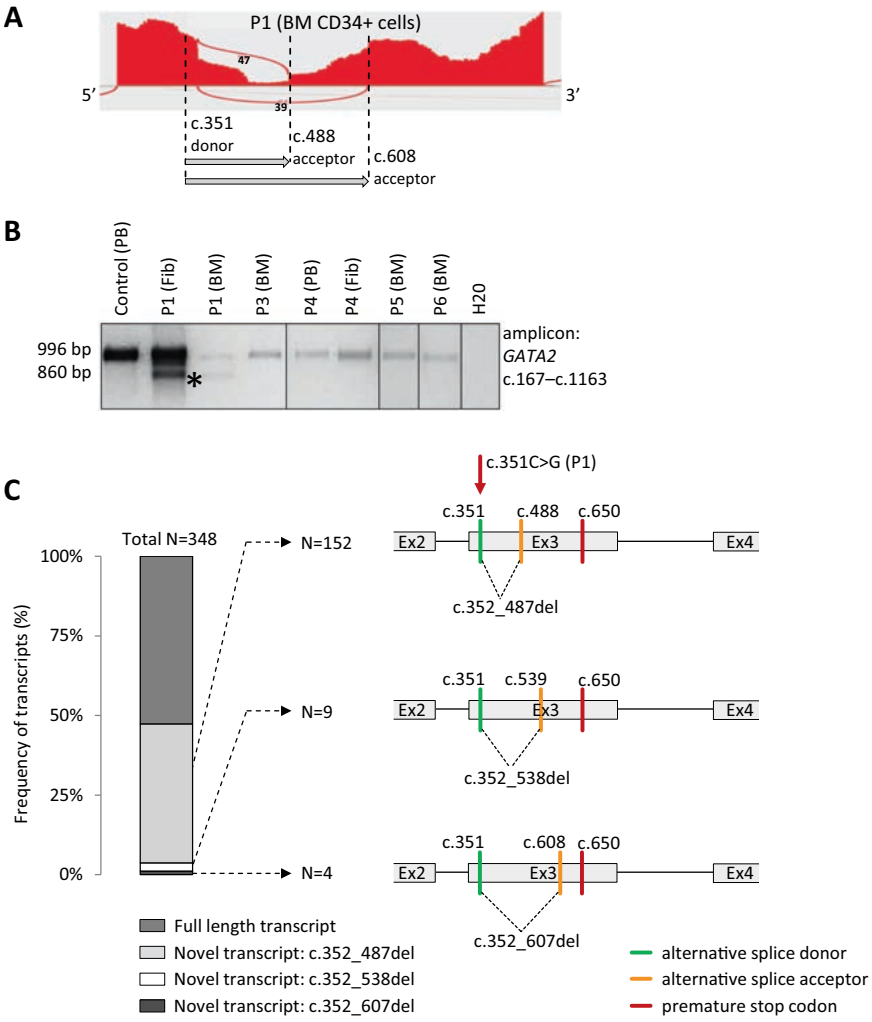


Fig. 3 Alternative splicing. a Sashimi plot of *GATA2* exon 3 of P1 depicting two novel splicing events (represented by arcs) detected by RNAseq; dashed lines indicate the positions of an alternative donor and two new acceptors. **b** Long range RT-PCR spanning exon 2–5 of *GATA2* transcript revealed the presence of an additional shorter product of 860 bp in P1 (indicated by asterisk (*)) corresponding to one of the transcripts found by RNAseq (donor: c.351—acceptor: c.488). Only wild-type allele (996 bp) was detected in P3–P6. **c** Frequency and schematic representation of novel splicing patterns in P1 detected by TA cloning of the RT-PCR product. It confirmed the presence of three novel transcripts, of which two were also identified by RNAseq. BM bone marrow, Fib fibroblasts, PB peripheral blood. *GATA2* (NM_032638.4).

and occurrence of a premature stop codon at c.650. No new isoforms were found in P4–P7 by RNAseq; additional TA cloning and sequencing of cDNA in P5 and P6 detected only properly spliced full length transcripts.

Synonymous variants are predicted not to affect RNA stability

The impact of synonymous variants on mRNA stability and secondary structure was determined using Mfold, RNAfold, and Quickfold tools. Synonymous substitutions were predicted not to significantly affect secondary structure of mRNA (Supplementary Fig. 4a). In addition, no relevant energy change (ΔG) was observed between Mut and WT (Supplementary Fig. 4b). As a comparison, five common synonymous polymorphisms from GnomAD and five nonsynonymous pathogenic *GATA2* mutations were included in the analysis. None of these variants had influence on the mRNA structure and thermodynamic characteristics.

Analysis of protein stability and function

We investigated the levels of endogenous GATA2 protein in P9 who carried RNA-deleterious mutation c.351C>G, p.T117T and had sufficient primary specimen. Analysis was performed in patient-derived platelets since GATA2 was previously found to be highly expressed in this hematopoietic subpopulation (Supplementary Fig. 5) [33, 34].

GATA2 protein levels were severely reduced, similarly to other known pathogenic GATA2 mutations (Fig. 4a). Next, to determine the effect of the synonymous variants on GATA2 transcriptional function, GATA-specific reporter assay was performed. Transactivation activity was comparable between synonymous Mut and WT (Fig. 4b). We subsequently assessed the protein:DNA-binding ability using the electrophoretic mobility shift assay (EMSA) for the mutation c.649C>T, p.L217L. Using this limited approach, no significant difference in DNA binding between Mut and WT GATA2 proteins was seen (Fig. 4c). Of note, both functional experiments were

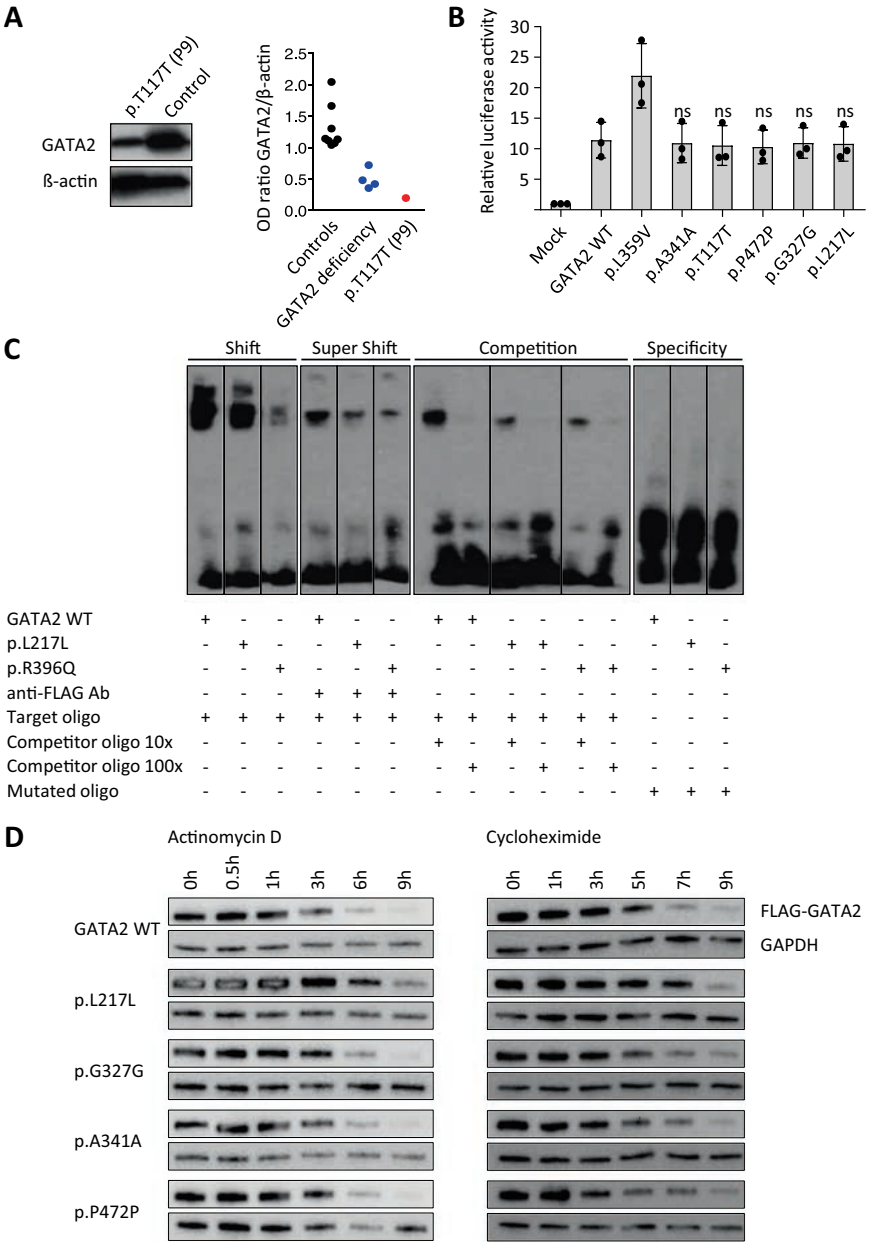


Fig. 4 Assessment of protein expression and function. a Expression of endogenous GATA2 protein in platelets of P9 (c.351C>G, p.T117T). Left: Representative western blot image for the expression level of GATA2 and ß-actin. Right: Relative optical density of GATA2 protein normalized with ß-actin. GATA2 deficiency group comprises four pathogenic mutations in *GATA2* coding region. **b** Transactivation activity measured by GATA reporter assay. Previously reported p.L359V mutant was used as a positive control. Fold activation of luciferase reporter was determined from the triplicate values from three independent experiments and presented as the mean ± SD values. Comparison between the GATA2 WT and each synonymous Mut, as well as *p* value calculation was performed using a standard one-way ANOVA test. **c** The electrophoretic mobility shift assay (EMSA) depicting DNA-binding ability of GATA2 p.L217L, compared with WT and a known loss-of-function mutation p.R396Q, lanes spliced from the same gel runs (experiment repeated three times). Shift: semiquantitative comparison of GATA2 binding strength to the target oligo (biotin-labeled probe containing wild-type target sequence); Super Shift: specificity of FLAG-GATA2 binding to DNA sequence with anti-FLAG antibody (Ab); Competition: includes 10× and 100× excess of unlabeled wild-type probe competing for binding of the protein; Specificity: contains biotin-labeled probe with mutated target sequence which cannot be bound by the analyzed protein. Three latter setups verify that the signal obtained in the shift reaction is the result of specific DNA-protein interaction. **d** Assessment of translation efficiency (left) and protein stability (right) of GATA2 synonymous mutants in transiently transfected 293T cells treated with 1 µg/ml actinomycin D and 10 µg/ml cycloheximide, respectively. Experiments were performed in duplicates. Blots from representative experiments are shown.

performed at steady state with a high level of ectopic protein expression.

Because it is known that synonymous variants can impair translation, we aimed to analyze the effect of Muts on protein levels. We ectopically expressed cDNA under the principle that splicing effect will not be expected due to missing introns, and observed protein changes will result from altered translation. We blocked the transcription with actinomycin D in transfected 293T cells and analyzed protein levels over time (Fig. 4d). Expectedly, protein content decreased during the course of treatment for all genotypes resulting from exhaustion of mRNA reserves. However, p.L217L showed slightly higher protein content as compared with WT. To better delineate the cause for the relative increase in protein levels after transcription blockade, we then quantified the proteins after translation inhibition (cycloheximide). The p.L217L variant was associated with a slowdown of protein degradation visible after 5–7 h of treatment (Fig. 4d).

Effect of synonymous *GATA2* c.649C>T variant on zebrafish hematopoiesis

For further analysis we selected the c.649C>T, p.L217L variant due to only partial reduction of the mutated allele expression in hematopoietic specimen of the P4. We hypothesized that the mutation may exert its effect on the protein level and aimed to determine if it alters zebrafish hematopoiesis. We used a previously published MO against *gata2b* [35] and visualized hematopoietic stem and progenitor cell (HSPC) in zebrafish embryos by whole-mount in situ hybridization of the HSPC marker *c-myb* at 28 h post fertilization, when HSPCs arise from the dorsal aorta. Expectedly, *gata2b* inhibition resulted in a reduction of HSPCs in zebrafish embryos (Fig. 5a, top right) [35]. We then performed a phenotype rescue experiment by co-injecting *gata2b* MO with human *GATA2* WT or Mut mRNA. Phenotype rescue (defined as medium/high phenotype; Supplementary Fig. 6) was achieved in 83% and 98% of embryos injected with WT and Mut mRNA, respectively, (Fig. 5b–d). However, we observed a significantly higher proportion of high phenotypes in animals rescued with Mut mRNA (42%) as compared with WT (19%), $p < 0.05$ (Fig. 5d right panel).

Discussion

GATA2 deficiency is a monogenic disorder known so far to be caused by heterozygous nonsynonymous mutations, whole gene deletions or intronic enhancer mutations, all of which result in haploinsufficiency. In this study, we report the identification of synonymous, RNA-deleterious mutations in *GATA2* that accounted for 8.2% of all *GATA2*–mutated patients and 14.8% of cases with *GATA2* exonic substitutions. In total, we identified nine patients harboring five distinct synonymous *GATA2* variants that are either absent or exceedingly rare in general population: p.T117T, p.P472P, p.L217L, p.G327G, and p.A341A. Two of these (p.T117T and p.P472P) were encountered in multiple unrelated pedigrees, suggesting either independent mutational events or rare founders in the European population (which is possible at least for p.P472P present in gnomAD in 23 individuals of non-Finnish European ancestry). The phenotype of patients carrying synonymous variants resembled GATA2 deficiency. All of the patients were alive at last follow-up with exception of one patient who died from HSCT-related complications. Additional mutations in *GATA2* were not identified. Other MDS-predisposing conditions were excluded based on clinical studies and WES/WGS in all patients with the exception of P6 who carried two VUS in the *SAMD9* gene. No specific features demarcating GATA2 from SAMD9 syndrome were present in this patient; hypospadias are unspecific and had been reported in both conditions [36, 37]. At this point, we cannot rule out that in P6 both gene defects acted in a synergistic manner facilitating MDS development.

Computational prediction assigned an increased probability of missplicing to three of the five variants. Further assessment of mutation deleteriousness with existing *in silico* tools failed to ascribe pathogenic effects. Because of the difficulty in predicting deleteriousness, synonymous mutations have been generally left out in genomic studies. However, it is likely that many disease-causing mutations are being consistently overlooked—including mutations located in noncoding regions of the genome as well as synonymous variants. So far, little is known about the role of such mutations in hematopoietic malignancies due to lack of routine screening of the inter-/intragenic regions. Besides known recurrent deleterious mutations in the regulatory element of *GATA2* [10] there are only few examples of noncoding mutations associated with BMF. Recently, two patients were reported with dyserythropoietic anemia and an intronic substitution in *GATA1* gene that is 24 nucleotides upstream of the canonical splice acceptor site. This alteration resulted in reduced canonical splicing and increased use of an alternative splice acceptor site that causes a partial intron retention event [38]. Moreover, mutations in 5'UTR and deep intronic region of *ELANE* gene have been reported to be associated with severe congenital neutropenia [39]. Due to lack of studies integrating functional evaluation, the prevalence of such variants in Mendelian disorders is yet to be determined. It is remarkable that recent pancancer studies report acquired synonymous driver mutations at a rate of ~6–8% among all single-nucleotide changes found in human cancers [40]. This is strikingly similar to the proportion of

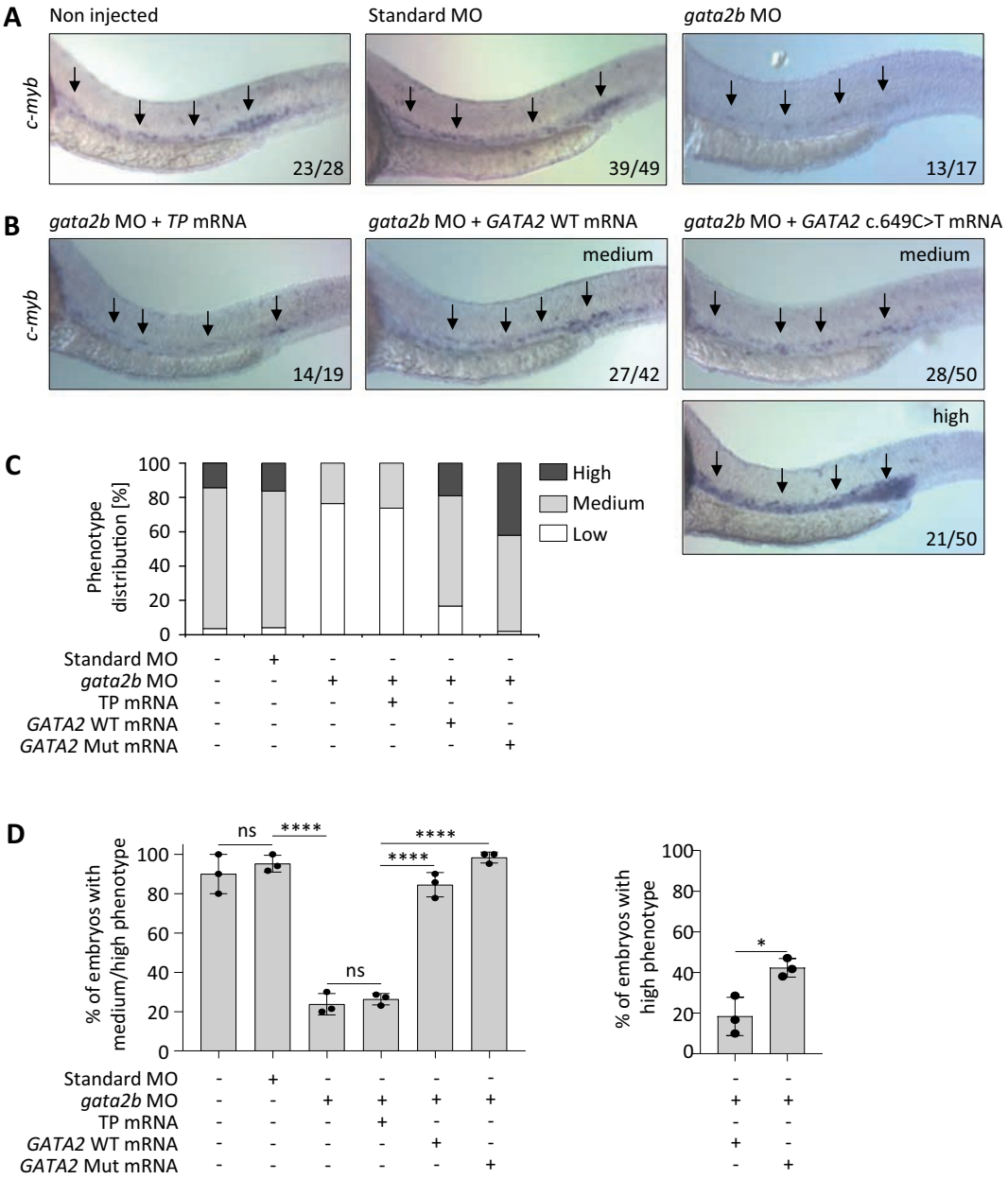


Fig. 5 In vivo analysis of zebrafish hematopoiesis. a Whole-mount in situ hybridization (WISH) performed at 28 h post fertilization (hpf) for *c-myb* in embryos injected with control or *gata2b* morpholino (MO). The arrows indicate the location of hematopoietic stem and progenitor cells (HSPC) in the zebrafish aorta. The numbers in each picture indicate the number of embryos with depicted phenotype out of all embryos analyzed. **b** Rescue experiment performed on embryos co-injected with *gata2b* MO and human *GATA2* WT or Mut (c.649C>T) mRNA. Transposase (*TP*) mRNA was used as a random control RNA with no expected impact. A substantial number of the Mut (c.649C>T) mRNA-injected morphants show an even stronger (high) staining than control *TP* mRNA-injected siblings. **c** Graphical representation of the rescue experiment. Distribution of phenotypes in all tested animals: high, indicates a stronger staining to the one of noninjected and standard control MO-injected embryos; medium, indicates a staining equivalent or approximately equivalent to the staining of noninjected and standard control MO-injected embryos; low, indicates a weaker staining compared with the staining of noninjected and standard control MO-injected embryos (deficiency of HSPCs). **d** Left panel: Graphical representation of percentage of embryos with medium or high staining for each treatment category. Right panel: Percentage of morphant embryos depicting high staining when injected with either WT or Mut (c.649C>T) mRNA. Each experiment was performed in triplicates ($N = 3$; mean ± SD, Student's t test, $^{*}p < 0.05$, $^{****}p < 0.0001$).

(germline) synonymous mutations identified in our study. Mutations causing phenotypically severe hereditary disease are mainly introduced as random de novo events, and it is well accepted that purifying selection will eventually eliminate these deleterious alleles. This is especially valid in high-penetrance conditions, such as GATA2 deficiency that

often manifests before the reproductive age and thus results in reduced fecundity.

There are multiple ways how synonymous substitutions can exert deleteriousness even though the amino acid sequence is not changed. As confirmed using three orthogonal approaches, all of the mutations found here resulted in a nearly complete and selective loss of the Mut transcript in hematopoietic cells, with the exception of c.649C>T, p.L217L that showed Mut allele reduction to ~20%. In contrast, paired analysis of two patients revealed a higher Mut allele expression in skin fibroblasts versus hematopoietic cells. Potential explanations for this discrepancy might be the variability of allelic expression across different tissues [41, 42] or the notion of context-dependent monoallelic expression observed for ~20% of human genes [43]. In addition, we observed that divergence in allelic ratio depends not only on the tissue analyzed but also on the stage of RNA processing. Strikingly, the mutation frequency in BM of two patients (c.351C>G and c.1023C>T) increased from nearly absent in mRNA to 30% in total RNA transcripts, implying that the defect manifests at a late stage of RNA maturation, at least for these two variants. Splicing disruption was predicted for three variants (c.351C>G, c.981G>A, and c.1023C>T); however, splicing analysis confirmed novel splicing pattern only for c.351C>G. This mutation resulted in aberrant transcripts with premature stop codon which makes it functionally equivalent to a frameshift-truncating mutation causing nonsense-mediated decay. For the remaining four mutations, no abnormal splicing was detected. It is conceivable that these Mut mRNAs are extremely unstable and subjected to a very rapid sequestration. Another potential explanation for loss of allelic expression is epigenetic silencing that could arise from aberrant promoter methylation. Supporting this, allelic disbalance due to hypermethylation was recently observed in one patient with GATA2 p.T354M mutation [44]. Synonymous variants can also affect translation and thus result in increased or decreased protein stability or function. Surprisingly, p.L217L Mut protein was slightly more stable in vitro, although its function (tested in vitro using the EMSA DNA gel shift assay at steady state, with ectopic *GATA2* overexpression) seemed not to be affected. Further, this mutation not only rescued the GATA2-deficient phenotype in zebrafish, but also resulted in a significantly higher number of HSPCs in comparison with control animals. Higher stability of this mutated protein might potentially explain the relative increase in its functional properties in vivo. In analogy, it is known that moderate *GATA2* overexpression enhances proliferation and self-renewal of progenitor cells [45]. We reason that the more efficient rescue of the morphant phenotype can be associated with higher stability of the p.L217L Mut, which is seen when transiently overexpressed in 293T cell line (Fig. 4d). Because of the challenging data (decrease of Mut RNA expression but higher protein stability of protein) we do question the pathogenicity of this mutation until additional biological data or patients are reported. Limited availability of patients' primary specimens as well as instability of the transcripts with synonymous mutations precluded further mechanistic studies.

Reported diagnostic yields for WES/WGS in single individuals can reach ~40% and heavily rely on computational predictions [46, 47] which are difficult to achieve for synonymous mutations. Moreover, WES is limited to the analysis of coding regions only. Even though genome sequencing overcomes this constraint, it generates an enormous output of alterations within coding and non-coding regions of the genes. In setting of GATA2 deficiency, WGS would facilitate the detection of pathogenic intronic mutations in regulatory region in intron 4 (corresponding to +9.5 kb enhancer region) as well as whole gene and partial gene deletions. However, allelic loss on RNA level would be missed. The utility of transcriptome analysis was previously highlighted by the identification of disease-causing mutations in patients with negative exome or genome sequencing results, increasing the diagnostic rate by as much as 35% [48, 49]. Hence, we propose that diagnostic sequencing should incorporate a cascade approach where RNA sequencing follows inconclusive DNA analysis in patients with suspected disease. This approach is feasible not only for patients with GATA2 deficiency but also in patients with high index of suspicion for a specific Mendelian disorder but without a known pathogenic mutation. Our findings suggest that a straightforward Sanger or deep sequencing of cDNA would be sufficient to confirm the RNA-deleteriousness of a synonymous variant.

In summary, we demonstrate that a significant proportion of GATA2-deficient patients carry damaging synonymous alterations. These genetic changes, previously excluded from analysis due to their likely silent effect, should be incorporated into standard diagnostic pipeline for individuals with GATA2 disease phenotype. However, patients with other hereditary BM failure and MDS syndromes might also benefit from this extended diagnostic approach. In the long term, identification of pathogenic synonymous variants has the potential to improve genetic counseling, HSCT donor selection, and clinical outcomes.

Data availability

WES data have been deposited at the European Genome–phenome Archive (EGA), which is hosted by the

EBI and the CRG, under accession number EGAS00001003817. Further information about EGA can be found on https://ega-archive.org "The European Genome–phenome Archive of human data consented for biomedical research".

Acknowledgements This work was supported by grants from Fritz Thyssen Foundation 10.17.1.026MN (to MWW and ET); Deutsche Kinderkrebsstiftung DKS 2017.03, ERAPerMed by German Federal Ministry of Education and Research (BMBF) 2018-123/01KU1904, and Deutsche Krebshilfe 109005 (to MWW); José Carreras Leukemia Foundation (to VBP); Deutsche Forschungsgemeinschaft (DFG) 322977937/GRK2344 "MeInBio–BioInMe" (to ET and MB); Marie Curie Career Integration 631432 "Bloody Signals", and DFG CIBSS-EXC-2189 390939984 (to ET), Baden-Württemberg LGFG stipend (to EJK), DFG EXC 22167-390884018 (to HB) and CRC850 (to MB), BMBF CoNfirm 01ZX1708F (to MB), and BMBF MyPred Network for young individuals with syndromes predisposing to myeloid malignancies (to MWW, ME, GG, CF, BS, and CMN). We are very grateful to Ayami-Yoshimi Nöllke, Lheanna Klaeyle, Sophie Krüger, Sandra Zolles, Christina Jäger, Sophia Hollander, Marco Teller, Ali-Riza Kaya, Alexandra Fischer, Wilfried Truckenmüller, Peter Nöllke, and Anne Strauss (Freiburg) for excellent assistance in diagnostic workup, laboratory work, and data management, and Prof. Dr. Rudolf Grosschedl (Max Planck Institute Freiburg), Dr. Claudia Wehr, and Dr. Ulrich Salzer (Freiburg) for helpful discussions. The authors also thank all members of the European Working Group of MDS in Childhood (EWOG-MDS) for performing reference examinations (pathology, cytogenetics, molecular genetics), HSCT or other forms of patient care. This work was performed within the European Reference Network for Paediatric Cancer (ERN-PAEDCAN). The authors acknowledge the contribution of the Center of Inborn and Acquired Blood Diseases at the Freiburg Center for Rare Diseases, and the Hilda Biobank at the Department of Pediatrics and Adolescent Medicine, Freiburg, Germany.

Author contributions EJK and MWW designed the study. EJK, MWW, VBP, ME, and ET contributed to manuscript conception. EJK, VBP, EAS, RKV, SH, DL, LP, MT, PM, MB, HB, PS, and MWW performed genomic studies and analyzed data. EJK, ET, SSS, SL, PS, and MD accomplished functional studies. MC, EM, ME, RP, LP, MT, CK, AC, HH, VH, KK, RM, BM, MD, MS, OS, JS, EM, MU, GG, CF, BS, FL, CMN, and MWW were involved in the patient care and testing. All authors contributed to the manuscript and approved the final version.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Primary patients' samples were obtained after written informed consent in accordance with the Declaration of Helsinki. The study was approved by the local Ethics Committee (CPMP/ICH/135/95). All animal experiments were performed in accordance with relevant guidelines and regulations, approved by the review committee of the Max Planck Institute of Immunobiology and Epigenetics and the Regierungspraesidium Freiburg, Germany (license Az 35-9185.81/G-14/95).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Hsu AP, Sampaio EP, Khan J, Calvo KR, Lemieux JE, Patel SY, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. Blood. 2011;118:2653–5.
2. Dickinson RE, Griffin H, Bigley V, Reynard LN, Hussain R, Haniffa M, et al. Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. Blood. 2011;118:2656–8.
3. Hahn CN, Chong CE, Carmichael CL, Wilkins EJ, Brautigan PJ, Li XC, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. Nat Genet. 2011;43:1012–7.
4. Pasquet M, Bellanne-Chantelot C, Tavitian S, Prade N, Beaupain B, Larochelle O, et al. High frequency of GATA2 mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia. Blood. 2013;121:822–9.
5. Ostergaard P, Simpson MA, Connell FC, Steward CG, Brice G, Woollard WJ, et al. Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). Nat Genet. 2011;43:929–31.
6. Dorn JM, Patnaik MS, Van Hee M, Smith MJ, Lagerstedt SA, Newman CC, et al. WILD syndrome is GATA2 deficiency: a novel deletion in the GATA2 gene. J Allergy Clin Immunol Pract. 2017;5:1149–52.
7. Wlodarski MW, Hirabayashi S, Pastor V, Stary J, Hasle H, Masetti R, et al. Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. Blood 2016;127:1387–97.
8. Hirabayashi S, Wlodarski MW, Kozyra E, Niemeyer CM. Heterogeneity of GATA2-related myeloid neoplasms. Int J Hematol. 2017;106:175–82.
9. Wlodarski M, Collin M, Horwitz MS. GATA2 deficiency and related myeloid neoplasms. Seminars in Hematology. 2017;54:81–6.
10. Hsu AP, Johnson KD, Falcone EL, Sanalkumar R, Sanchez L, Hickstein DD, et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. Blood 2013;121:3830–7. S1-7
11. Hahn CN, Brautigan PJ, Chong CE, Janssan A, Venugopal P, Lee Y, et al. Characterisation of a compound in-cis GATA2 germline mutation in a pedigree presenting with myelodysplastic syndrome/acute myeloid leukemia with concurrent thrombocytopenia. Leukemia 2015;29:1795–7.
12. Cortes-Lavaud X, Landecho MF, Maicas M, Urquiza L, Merino J, Moreno-Miralles I, et al. GATA2 Germline Mutations Impair GATA2 Transcription, Causing Haploinsufficiency: Functional Analysis of the p.Arg396Gln Mutation. J Immunol. 2015;194:2190–8.

13. Sologuren I, Martinez-Saavedra MT, Sole-Violan J, de Borges de Oliveira E Jr, Betancor E, Casas I, et al. Lethal Influenza in Two Related Adults with Inherited GATA2 Deficiency. J Clin Immunol. 2018;38:513–26.
14. Chong CE, Venugopal P, Stokes PH, Lee YK, Brautigan PJ, Yeung DTO, et al. Differential effects on gene transcription and hematopoietic differentiation correlate with GATA2 mutant disease phenotypes. Leukemia 2018;32:194–202.
15. D'Souza I, Poorkaj P, Hong M, Nochlin D, Lee VM, Bird TD, et al. Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. Proc Natl Acad Sci USA. 1999;96:5598–603.
16. Cartegni L, Krainer AR. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. Nat Genet 2002;30:377–84.
17. Macaya D, Katsanis SH, Hefferon TW, Audlin S, Mendelsohn NJ, Roggenbuck J, et al. A synonymous mutation in TCOF1 causes Treacher Collins syndrome due to mis-splicing of a constitutive exon. Am J Med Genet A 2009;149A:1624–7.
18. Vidal C, Cachia A, Xuereb-Anastasi A. Effects of a synonymous variant in exon 9 of the CD44 gene on pre-mRNA splicing in a family with osteoporosis. Bone 2009;45:736–42.
19. Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR, Gelernter J, et al. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. Hum Mol Genet. 2003;12:205–16.
20. Wang D, Johnson AD, Papp AC, Kroetz DL, Sadee W. Multidrug resistance polypeptide 1 (MDR1, ABCB1) variant 3435C>T affects mRNA stability. Pharmacogenet Genomics 2005;15:693–704.
21. Nackley AG, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskyi O, Makarov SS, et al. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. Science 2006;314:1930–3.
22. Buhr F, Jha S, Thommen M, Mittelstaet J, Kutz F, Schwalbe H, et al. Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. Mol cell 2016;61:341–51.
23. Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, Vouret-Craviari V, et al. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. Nat Genet 2011;43:242–5.
24. Simhadri VL, Hamasaki-Katagiri N, Lin BC, Hunt R, Jha S, Tseng SC, et al. Single synonymous mutation in factor IX alters protein properties and underlies haemophilia B. J Med Genet. 2017;54:338–45.
25. Wehr C, Grotius K, Casadei S, Bleckmann D, Bode SFN, Frye BC, et al. A novel disease-causing synonymous exonic mutation in GATA2 affecting RNA splicing. Blood 2018;132:1211–5.
26. Baumann I, Niemeyer CMBJ, Shannon K WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon: IARC Press; 2008. p. 104–7.
27. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood 2009;114:937–51.
28. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016;127:2391–405.
29. Pastor V, Hirabayashi S, Karow A, Wehrle J, Kozyra EJ, Nienhold R, et al. Mutational landscape in children with myelodysplastic syndromes is distinct from adults: specific somatic drivers and novel germline variants. Leukemia. 2016;31:759–62.
30. Bluteau O, Sebert M, Leblanc T, Peffault de Latour R, Quentin S, Lainey E, et al. A landscape of germ line mutations in a cohort of inherited bone marrow failure patients. Blood 2018;131:717–32.
31. Buonocore F, Kuhnen P, Suntharalingham JP, Del Valle I, Digweed M, Stachelscheid H, et al. Somatic mutations and progressive monosomy modify SAMD9-related phenotypes in humans. J Clin Invest. 2017;127:1700–13.
32. Narumi S, Amano N, Ishii T, Katsumata N, Muroya K, Adachi M, et al. SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. Nat Genet 2016;48:792–7.
33. Huang Z, Dore LC, Li Z, Orkin SH, Feng G, Lin S, et al. GATA-2 reinforces megakaryocyte development in the absence of GATA-1. Mol Cell Biol. 2009;29:5168–80.
34. Goardon N, Marchi E, Atzberger A, Quek L, Schuh A, Soneji S, et al. Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. Cancer Cell 2011;19:138–52.
35. Butko E, Distel M, Pouget C, Weijts B, Kobayashi I, Ng K, et al. Gata2b is a restricted early regulator of hemogenic endothelium in the zebrafish embryo. Development 2015;142:1050–61.
36. Schwartz JR, Wang S, Ma J, Lamprecht T, Walsh M, Song G, et al. Germline SAMD9 mutation in siblings with monosomy 7 and myelodysplastic syndrome. Leukemia 2017;31:1827–30.
37. Novakova M, Zaliova M, Sukova M, Wlodarski M, Janda A, Fronkova E, et al. Loss of B cells and their precursors is the most constant feature of GATA-2 deficiency in childhood myelodysplastic syndrome. Haematologica 2016;101:707–16.
38. Abdulhay NJ, Fiorini C, Verboon JM, Ludwig LS, Ulirsch JC, Zieger B, et al. Impaired human hematopoiesis due to a cryptic intronic GATA1 splicing mutation. J Exp Med. 2019;216:1050–60.
39. Makaryan V, Zeidler C, Bolyard AA, Skokowa J, Rodger E, Kelley ML, et al. The diversity of mutations and clinical outcomes for ELANE-associated neutropenia. Curr Opin Hematol. 2015;22:3–11.
40. Supek F, Minana B, Valcarcel J, Gabaldon T, Lehner B. Synonymous mutations frequently act as driver mutations in human cancers. Cell 2014;156:1324–35.
41. Pirinen M, Lappalainen T, Zaitlen NA, Consortium GT, Dermitzakis ET, Donnelly P, et al. Assessing allele-specific expression across multiple tissues from RNA-seq read data. Bioinformatics 2015;31:2497–504.
42. Wilkins JM, Southam L, Price AJ, Mustafa Z, Carr A, Loughlin J. Extreme context specificity in differential allelic expression. Hum Mol Genet. 2007;16:537–46.
43. Serre D, Gurd S, Ge B, Sladek R, Sinnett D, Harmsen E, et al. Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. PLoS Genet 2008;4:e1000006.
44. Al Seraihi AF, Rio-Machin A, Tawana K, Bodor C, Wang J, Nagano A, et al. GATA2 monoallelic expression underlies reduced penetrance in inherited GATA2-mutated MDS/AML. Leukemia. 2018;32:2502–7.
45. Nandakumar SK, Johnson K, Throm SL, Pestina TI, Neale G, Persons DA. Low-level GATA2 overexpression promotes myeloid progenitor self-renewal and blocks lymphoid differentiation in mice. Exp Hematol 2015;43:565–77.e1-10.
46. Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat Genet 2015;47:717–26.
47. Ji J, Shen L, Bootwalla M, Quindipan C, Tatarinova T, Maglinte DT, et al. A semi-automated whole exome sequencing workflow

leads to increased diagnostic yield and identification of novel candidate variants. Cold Spring Harb Mol Case Stud. 2019;5: a003756.
48. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci Transl Med. 2017;9: eaal5209.
49. Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat Commun 2017;8:15824.

Affiliations

Emilia J. Kozyra[1,2] · Victor B. Pastor[1] · Stylianos Lefkopoulos[2,3] · Sushree S. Sahoo[1,4] · Hauke Busch[5,6,7] · Rebecca K. Voss[1] · Miriam Erlacher[1,8,9] · Dirk Lebrecht[1] · Enikoe A. Szvetnik[1] · Shinsuke Hirabayashi[1,10] · Ramunė Pasaulienė[11] · Lucia Pedace[12] · Marco Tartaglia[13] · Christian Klemann[14] · Patrick Metzger[2,15] · Melanie Boerries[8,9,15] · Albert Catala[16] · Henrik Hasle[17] · Valerie de Haas[18] · Krisztián Kállay[19] · Riccardo Masetti[20] · Barbara De Moerloose[21] · Michael Dworzak[22] · Markus Schmugge[23] · Owen Smith[24] · Jan Starý[25] · Ester Mejstrikova[25] · Marek Ussowicz[26] · Emma Morris[27,28,29] · Preeti Singh[30,31] · Matthew Collin[30,31] · Marta Derecka[3] · Gudrun Göhring[32] · Christian Flotho[1,8,9] · Brigitte Strahm[1] · Franco Locatelli[12,33] · Charlotte M. Niemeyer[1,8,9] · Eirini Trompouki[3,34] · Marcin W. Wlodarski[1,4] · European Working Group of MDS in Childhood (EWOG-MDS)

1 Division of Pediatric Hematology and Oncology, Department of Pediatrics and Adolescent Medicine, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

2 Faculty of Biology, University of Freiburg, Schänzlestraße 1, 79104 Freiburg, Germany

3 Department of Cellular and Molecular Immunology, Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany

4 Department of Hematology, St. Jude Children´s Research Hospital, Memphis, USA

5 Institute of Molecular Medicine and Cell Research, University of Freiburg, Freiburg, Germany

6 Lübeck Institute of Experimental Dermatology and Institute of Cardiogenetics, University of Lübeck, Lübeck, Germany

7 Comprehensive Cancer Center Freiburg (CCCF), University Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

8 German Cancer Consortium (DKTK), Freiburg, Germany

9 German Cancer Research Center (DKFZ), Heidelberg, Germany

10 Department of Pediatrics, Hokkaido University Graduate School of Medicine, Sapporo, Japan

11 Vilnius University Hospital Santaros Klinikos, Center for Pediatric Oncology and Hematology, Bone Marrow Transplantations Unit, Vilnius, Lithuania

12 Department of Pediatric Hematology and Oncology, Istituto di Ricovero e Cura a Carattere Scientifico Ospedale Pediatrico Bambino Gesù, Rome, Italy

13 Genetics and Rare Diseases Research Division, Istituto di Ricovero e Cura a Carattere Scientifico Ospedale Pediatrico Bambino Gesù, Rome, Italy

14 Department of Pediatric Pneumology, Allergy and Neonatology, Hannover Medical School, Hannover, Germany

15 Institute of Medical Bioinformatics and Systems Medicine, Medical Center—University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

16 Department of Hematology and Oncology, Hospital Sant Joan de Déu, Barcelona, Spain

17 Department of Pediatrics, Aarhus University Hospital Skejby, Aarhus, Denmark

18 Dutch Childhood Oncology Group (DCOG), Princess Máxima Centre, Utrecht, The Netherlands

19 Central Hospital of Southern Pest—National Institute of Hematology and Infectious Diseases, Budapest, Hungary

20 Department of Pediatric Oncology and Hematology, University of Bologna, Bologna, Italy

21 Department of Pediatric Hematology-Oncology and Stem Cell Transplantation, Ghent University Hospital, Ghent, Belgium

22 St. Anna Children´s Hospital and Cancer Research Institute, Pediatric Clinic, Medical University of Vienna, Vienna, Austria

23 Department of Hematology and Oncology, University Children's Hospital, Zurich, Switzerland

24 Paediatric Oncology and Haematology, Our Lady's Children's Hospital Crumlin, Dublin, Ireland

25 Department of Pediatric Hematology and Oncology, Charles University and University Hospital Motol, Prague, Czech Republic

26 Department of Paediatric Bone Marrow Transplantation, Oncology and Hematology, Medical University of Wroclaw, Wroclaw, Poland

27 Institute of Immunity and Transplantation, University College London (UCL), London, UK

28 Bone Marrow Transplant (BMT) Programme, UCL Hospital National Health Service Foundation Trust (NHS FT), London, UK

29 Department of Immunology, Royal Free London NHS FT, London, UK

[30] Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK

[31] NIHR Newcastle Biomedical Research Centre at Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

[32] Institute of Human Genetics, Hannover Medical School, Hannover, Germany

[33] Department of Pediatrics, Sapienza University of Rome, Rome, Italy

[34] CIBSS—Centre for Integrative Biological Signaling Studies, Freiburg, Germany

3.4 MIRACUM-Pipe

The increasing demand of high-throughput data in medical centers, especially for therapy recommendations for advanced tumor patients within an interdisciplinary molecular tumor board (MTB) requires the development of bioinformatics workflows for their standardized processing. Therefore, the use and exchange of standardized and harmonized data is needed. The Medical Informatics Initiative (MI-I) has set itself the task of solving the standardization and integration of data on a large scale. MIRACUM (Medical Informatics for Research And Care in University Medicine) [Prokosch et al., 2018], as one of the founded consortia in Germany, supports interdisciplinary MTBs combining extensive molecular pathology with state-of-the art sequencing and subsequent bioinformatics analyses for personalized tumor diagnostics and therapy recommendations in a multi-center setting. Hence, transparent data integration and decision making is key. As a result, the MIRACUM-Pipe was developed to provide an automated analyses workflow for whole exome and targeted next-generation sequencing data to produce reliable and reproducible results. For easier installation and a flawless distribution MIRACUM-Pipe is deployed as a Docker container.

Metzger, P., Hess, M. E., Scheible R., Beoker, M., Geoffroy, A., Boerries, M. **MIRACUM-Pipe: an adaptable pipeline for whole exome sequencing and reporting for clinical decision making**. (under preparations, 2020).

Contribution: I developed and tested the workflow of the pipeline and wrote the manuscript.

Bioinformatics
doi.10.1093/bioinformatics/xxxxxx
Advance Access Publication Date: Day Month Year
Application Note

OXFORD

Genome Analysis

MIRACUM-Pipe: an adaptable pipeline for whole exome sequencing and reporting for clinical decision making

Patrick Metzger [1,2,*], Maria Hess [1,2,4], Raphael Scheible [3], Martin Boeker [3], Geoffroy Andrieux [1,4,5] and Melanie Boerries [1,4,5,*]

[1]Institute of Medical Bioinformatics and Systems Medicine, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, [2]Faculty of Biology, University of Freiburg, [3]Institute of Medical Biometry and Statistics, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg [4]Comprehensive Cancer Center Freiburg (CCCF), Faculty of Medicine and Medical Center-University of Freiburg [5]German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Partner Site Freiburg

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Whole exome sequencing of patients with advanced tumors is becoming an established method in medical centers. However, somatic variant calling and interpretation as well as report creation requires in-depth knowledge in both bioinformatics and oncology. Since most medical centers rely on in-house solutions for quality control and data analysis, reporting can vary substantially which hinders reliable data exchange of results between medical centers.
Results: MIRACUM-Pipe provides a simple to use, "one-click" standardized solution to analyzing whole exome sequencing data including quality control, variant calling, copy number calling, annotation and report generation.
Availability: https://github.com/AG-Boerries/MIRACUM-Pipe-docker
Contact: patrick.metzger@uniklinik-freiburg.de

1 Introduction

The increasing use of and demand for high-throughput data in medical centers, particularly those generated for therapy recommendations of advanced tumor patients of a Molecular Tumor Board, requires the development of bioinformatics tools for their standardized handling, analysis and interpretation. An important prerequisite for multi-center clinical studies is the exchange and usage of standardized and harmonization data in order to provide the best comprehensive insight into and understanding of the disease. As one of its objectives, the Medical Informatics Initiative (MI-I) of the German Ministry of Education and Research (BMBF) has established this large-scale data standardization and integration scenario. MIRACUM (Medical Informatics for Research And Care in University Medicine) [3], as one of the four funded MI-I consortia, is rolling out three use cases in twelve university hospitals in Germany. One use case is the establishment of interdisciplinary Molecular Tumor Boards (MTBs), which combine extensive molecular pathology with state-of-the-art sequencing and subsequent bioinformatics analysis for personalized tumor diagnostics and therapy recommendations. To support transparent data integration and decision making across the individual MIRACUM tumor boards we have developed MIRACUM-Pipe, an automated analysis workflow for whole exome sequencing (WES) that produces reliable and reproducible results across different facilities. In order to facilitate the installation and flawless distribution of MIRACUM-Pipe within the MIRACUM consortium and beyond, we put the entire application into a docker container [1].

2 MIRACUM-Pipe

MIRACUM-Pipe incorporates tools for detecting single nucleotide variants (SNVs), insertions and deletions (InDels), loss of heterozygosity (LoH), copy number variations (CNVs) as well as for determining quality and statistics. Various functional prediction and annotation databases are integrated to automatically annotate the identified variants. The workflow

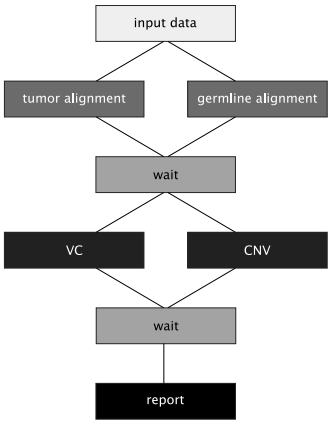


Fig. 1. This illustration shows the architecture of the MIRACUM-Pipe. The pipeline pre-processes and aligns tumor DNA and germline DNA in parallel. Subsequently, VS and CNV are also performed in parallel. The pipeline again waits for these two tasks to be computed and finally generates the PDF report.

is designed as a fully automated "one-click" solution from the raw sequencing files to the PDF report containing quality assessments, the identified and annotated variants (SNVs, InDels and LoH), CNVs, and a gene-set enrichment analysis both of the SNVs and CNVs, respectively.

MIRACUM-Pipe comprises bash and R code to perform WES data processing and basic annotations as well as complex functional annotations and downstream analysis of the results. The pipeline is divided into three main parts, namely (1) alignment and quality control; (2) analysis, annotation and interpretation including variant calling (VC) and copy number calling; and (3) assembly of results into a PDF report. The main focus lies on the annotation, from various available sources, of identified variants and results to facilitate its interpretation.

Performance: The pipeline requires potent computing power for the individual computation processes and is therefore recommended for operation on a high-performance computer cluster. At the same time, all its internally provided tools are desgined to use the available hardware resources. Most tools we use internally run in parallel and require a configurable amount of memory. Moreover, we have also implemented certain abstract steps of MIRACUM-Pipe, which by default run in parallel (see Figure 1) with the aim of distributing resources evenly. As an example, a WES dataset, consisting of a tumor and a matched germline sample each with 80 million paired-end reads each, was analyzed in 10.5 h on a high-performance computer cluster [Two 18-Core Intel Xeon E5-2697v4 processors (2.3-3.6 GHz) with 1 TB of RAM].

Usability: To make the installation as simple and error-free as possible, we offer a docker container. This container includes a shell script that implements the pipeline processes, certain tools and several databases. Since the MIRACUM-Pipe requires a large number of additional user-specific databases and files, we designed an environment which is wrapped around the docker container. This simplifies the application and setup of the MIRACUM-Pipe. Furthermore, some tools cannot be delivered with the docker container due to existing license rights. To address these issues, our software is split into two GitHub repositories: one for the pipeline itself which is intended to be used as a docker container, and another for its application and setup. Alternatively, the MIRACUM-Pipe can be installed and used without the docker container, which might be interesting for further development purposes.

Configuration: MIRACUM-Pipe contains several tools which are adjustable via several parameters. We offer customization of the recommended default settings within the configuration files. Two levels of customization are possible: (1) for the entire MIRACUM-Pipe installation and (2) individually for selected patients. The latter allows for a high customization of the application settings within one installation. In addition to the parameters of the used tools, resource limits and file names are also adjustable.

Flexibility: The result of a pipeline run is a report in the form of a PDF document with several plots and data based on various analyses of several used tools. This PDF is created based on a generated LaTeX-document. Therefore, the report can be easily adapted and recompiled to a PDF at any time, and parts of it can additionally be extracted and re-used in different documents. This applies to texts, tables and to each individually generated image or graph within the report.

3 Conclusions

Next-generation sequencing, and WES in particular, is increasingly used to guide physicians in making personalized therapy recommendations in oncology. Common standards for in-depth analysis strategies and their medical translation are still lacking. To overcome these problems, we have developed MIRACUM-Pipe to simplify analysis and to make the data interoperable and reusable. This is the current standard for analyzing and reporting for patients in the Molecular Tumor Board of the Comprehensive Cancer Center Freiburg, Medical Center University of Freiburg [2] as well as for the Use Case 3 "From Knowledge to Action - Support for Molecular Tumor Boards" within the MIRACUM consortium. The pipeline is easily adaptable to incorporate or merge future databases, analysis tools and workflows, which will help to unify personalized oncology endeavors within the German health system. The creation of a docker container simplified the distribution in general. The approach of providing an environment wrapped around the docker container further simplifies the application. Parallel computation allows a high degree of capacity utilization. Configuration files limit the hardware resources that are required and allow the tools' parameters to be adjusted, both for the entire installation and individually for selected patients. The resulting PDF report is based on a LaTeX-document including all generated images, which can subsequently be used in many ways.

Funding

This work was funded by MIRACUM within the Medical Informatics Funding Scheme (01ZZ1801B, BMBF).
Conflicts of interest: None declared.

Acknowledgements

The authors would like to thank James Balmford for English proofreading. We are grateful to the many members of the MIRACUM team and the Molecular Tumor Board Freiburg Team who have actively participated in the development of MIRACUM-Pipe.

References

[1]C. Anderson. Docker [software engineering]. 32(3):102–c3, may 2015.
[2]Rouven Hoefflin, Anna-Lena Geißler, Ralph Fritsch, et al. Personalized clinical decision making through implementation of a molecular tumor board: A german single-center experience. *JCO Precision Oncology*, (2):1–16, aug 2018.
[3]Hans-Ulrich Prokosch, Till Acker, Johannes Bernarding, et al. Miracum: Medical informatics in research and care in university medicine. 57:e82–e91, 2018.

3.5 Personalized Clinical Decision Making Through Implementation of a Molecular Tumor Board: A German Single-Center Experience

Due to improved availability and decreasing sequencing costs, oncology is undergoing a paradigm shift towards personalized, biomarker-based cancer treatment. Additionally, the availability of an increased number of molecular targeted therapies is beneficial for this process. For example, in melanoma cases, the use of molecular biomarkers for standard treatment stratification in a systems medicine approach, such as the activation of *BRAF* mutations for treatment with *BRAF* inhibitors, has shown beneficial results [Chapman et al., 2011]. It has been shown that the one-mutation-one-drug approach may not be sufficient and thus the molecular profiling of a tumor for treatment decision increases the impact and the tendency of combination therapies [Hoadley et al., 2014, Hyman et al., 2015,Borghaei et al., 2015]. However, the challenge remains to distinguish relevant mutations from passenger mutations [Lawrence et al., 2013], or clonal molecular intra- and intertumoral heterogeneity [Gerlinger et al., 2012], or the dynamic changes in the molecular composition of cancer in the background of different treatments. In our retrospective case series we analysed 198 patients from the molecular tumor board (MTB) Freiburg in a two year time frame. The performed genetic tests included targeted NGS, i.e. gene panels, whole exome sequencing, and RNA sequencing. We were able to show that such a personalized treatment concept, like the MTB Freiburg, which integrates clinical, molecular, and high-throughput results for clinical decision making, is feasible and that patients significantly benefit from therapy recommendations based on a systems medicine approach.

Hoefflin, R., Geißler, A.-L., Fritsch, R., Claus, R., Wehrle, J., Metzger, P., Reiser, M., Mehmed, L., Fauth, L., Heiland, D. H., Erbes, T., Stock, F., Csanadi, A., Miething, C., Weddeling, B., Meiss, F., von Bubnoff, D., Dierks, C., Ge, I., von Bubnoff, N. (2018). **Personalized Clinical Decision Making Through Implementation of a Molecular Tumor Board: A German Single-Center Experience**. JCO Precision Oncology.

Contribution: I performed the whole exome analyses, the RNA sequencing analyses

and the gene fusion analyses of the patients discussed in the molecular tumor board Freiburg. I wrote the part *Investigational Genetic Tumor Characterization*. I created figure 1 and revised the entire manuscript.

original report

Personalized Clinical Decision Making Through Implementation of a Molecular Tumor Board: A German Single-Center Experience

Rouven Hoefflin
Anna-Lena Geißler
Ralph Fritsch
Rainer Claus
Julius Wehrle
Patrick Metzger
Meike Reiser
Leman Mehmed
Lisa Fauth
Dieter Henrik Heiland
Thalia Erbes
Friedrich Stock
Agnes Csanadi
Cornelius Miething
Britta Weddeling
Frank Meiss
Dagmar von Bubnoff
Christine Dierks
Isabell Ge
Volker Brass
Steffen Heeg
Henning Schäfer
Martin Boeker
Justyna Rawluk
Elke Maria Botzenhart
Gian Kayser
Simone Hettmer
Hauke Busch
Christoph Peters
Martin Werner
Justus Duyster
Tilman Brummer
Melanie Boerries
(continued)

abstract

Purpose **Dramatic advances in our understanding of the molecular pathophysiology of cancer, along with a rapidly expanding portfolio of molecular targeted drugs, have led to a paradigm shift toward personalized, biomarker-driven cancer treatment. Here, we report the 2-year experience of the Comprehensive Cancer Center Freiburg Molecular Tumor Board (MTB), one of the first interdisciplinary molecular tumor conferences established in Europe. The role of the MTB is to recommend personalized therapy for patients with cancer beyond standard-of-care treatment.**

Methods **This retrospective case series includes 198 patients discussed from March 2015 through February 2017. The MTB guided individual molecular diagnostics, assessed evidence of actionability of molecular alterations, and provided therapy recommendations, including approved and off-label treatments as well as available matched clinical trials.**

Results **The majority of patients had metastatic solid tumors (73.7%), mostly progressive (77.3%) after a mean of 2.0 lines of standard treatment. Diagnostic recommendations resulted in 867 molecular diagnostic tests for 172 patients (five per case), including exome analysis in 36 cases (18.2%). With a median turnaround time of 28 days, treatment recommendations were given to 104 patients (52.5%). These included single-agent targeted therapies (42.3%), checkpoint inhibitors (37.5%), and combination therapies (18.3%). Treatment recommendations were implemented in 33 of 104 patients (31.7%), of whom 19 (57.6%) showed stable disease or partial response, including 14 patients (7.1% of the entire population) receiving off-label treatments.**

Conclusion **Personalized extended molecular-guided patient care is effective for a small but clinically meaningful proportion of patients in challenging clinical situations. Limited access to targeted drugs, lack of trials, and submission at late disease stage prevents broader applicability, whereas genome-wide analyses are not a strict requirement for predictive molecular testing.**

INTRODUCTION

Personalized cancer medicine uses molecular biomarkers for standard-of-care treatment stratification, such as activating *BRAF* mutations for the treatment of melanoma with BRAF inhibitors.[1] In parallel, it has become evident that therapeutic strategies with targeted drugs are no longer specific for the treatment of distinct entities but rather for particular molecular profiles across different cancers.[2-4] Thus, testing for single-drug targets can provide therapeutic information, but its predictive value may vary between entities. Although an activating *BRAF* V600E mutation will predict response to BRAF inhibitors in melanoma,[1] it may not do so in colorectal cancers because of epidermal growth factor receptor (EGFR) feedback activation with requirement of additional EGFR targeting.[5,6] Moreover, non-V600 *BRAF* mutations might not be responsive to BRAF inhibition at all.[7] Thus, one-mutation–one-drug approaches may be ineffective, especially in heavily pretreated patients with cancer. Underlying causes include

Silke Lassmann

Nikolas von Bubnoff

Author affiliations and support information (if applicable) appear at the end of this article.

M.B., S.L., and N.v.B. share authorship.

Corresponding author: Nikolas von Bubnoff, MD, Division Hematology, Oncology, and Stem Cell Transplantation, Medical Center, University of Freiburg, Hugstetter Str 55, D-79106 Freiburg, Germany; Twitter: @Uniklinik_Fr; e-mail: nikolas.bubnoff@uniklinik-freiburg.de.

the challenge to discriminate relevant mutations and pathway aberrations from background and passenger mutations,[8] the clonal molecular intra- and intertumoral heterogeneity,[9,10] and dynamic changes in the molecular composition of cancer, especially if treatment leads to selection of resistant subclones. Examples include the selection of *RAS* mutant clones in colorectal cancer treated with EGFR antibodies, such as cetuximab or panitumumab,[11] or the acquisition of a secondary *EGFR* T790M kinase domain mutation mediating resistance to EGFR kinase inhibitors, such as gefitinib or erlotinib in non–small-cell lung cancer.[12,13]

This increasing amount of complexity requires tools to translate individual information into personalized treatment concepts. A molecular tumor board (MTB) represents a platform that integrates clinical and molecular parameters for clinical decision making. Here, we report the 2-year experience of the Comprehensive Cancer Center Freiburg MTB that provides personalized treatment recommendations on the basis of individual molecular diagnostics. We hereby present detailed data on patient characteristics, treatment recommendations, clinical adherence to recommendations, and outcomes of treated patients.

METHODS

MTB Implementation and Organization

The MTB is run by an interdisciplinary team of medical and scientific experts with a focus on clinical and translational oncology and computational and molecular biology. Cases are submitted using an online registration and documentation system (Appendix Fig A1). Each case is assigned to a clinician scientist with expertise in the specific cancer type (entity expert), who reviews the literature and available clinical trials. In parallel, the molecular pathology team reviews the individual tumor pathology and sets up a presentation of already performed and suggested diagnostic tests. The initial discussion includes a clinical case presentation, review of the pathology data and the tumor-specific genetic landscape, known molecular predictive or prognostic markers, active clinical trials, and potential in- and off-label molecular targeted treatments. The molecular diagnostic requests are performed using certified and standard operating procedure (SOP)–driven processes. Diagnostic results are presented to the MTB by the molecular pathology and/or the computational biologist team. After discussion, treatment recommendations are given and are supported by levels of evidence (Data Supplement). These are based on published molecular biomarker recommendations.[14]

Patients and Patient Informed Consent

All patients discussed (n = 198) were included in this retrospective single-center case series. All molecular diagnostic tests were conducted in accordance with the medical treatment contract signed by each patient. Patient tissue was stored in the local biobank and required a signed informed consent, approved by the University of Freiburg institutional review board. Patients with individual or family history indicative of germline disease-causing mutations were referred to the Institute of Human Genetics for counseling and possibly germline genetic analyses.

Diagnostic Molecular Pathology

Appropriate tissues were subjected to molecular analyses as recommended by the MTB (Fig 1). All analyses were carried out according to routine pathology laboratory testing procedures, with assays being nationally validated and certified. Targeted next-generation sequencing (tNGS) included a custom-designed hotspot eight-gene panel (designed by S.L. and produced by Illumina, San Diego, CA), a *BRCA1/2* panel (produced by Illumina), a hotspot 48-gene panel (TruSeq Amplicon Cancer Panel, Illumina), and a 54-gene myeloid panel (TruSight Myeloid Sequencing Panel, Illumina).[15-17]

Investigational Genetic Tumor Characterization

Whole-exome sequencing (WES) and RNA sequencing (RNA-Seq) were performed on tumor tissue. Complementary germline DNA was obtained from peripheral blood or healthy tissue. Only nonsynonymous mutations detected with a variant allele frequency > 10% and listed with a minor allele frequency < 0.001% by the Exome Aggregation Consortium[18] were reported. Single nucleotide variations were classified according to ClinVar,[19] COSMIC,[20] dbSNP,[21,22]

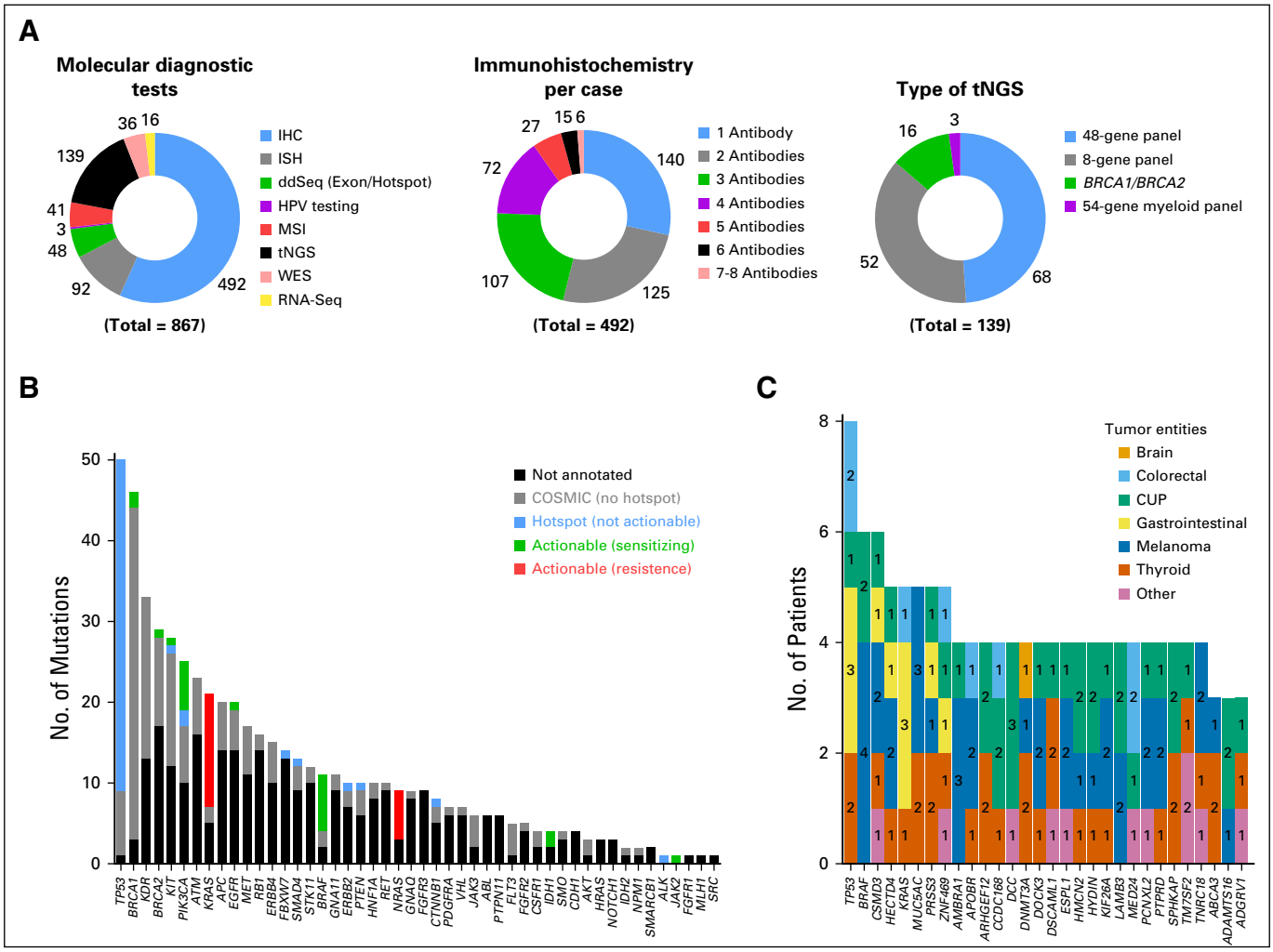


Fig 1. Molecular diagnostic testing. (A) The panels depict the type of molecular diagnostic testing performed (left panel) and specify the number of immunohistochemical stains (one to eight antibodies) per case (middle panel) as well as the type of targeted next-generation sequencing (tNGS) library sequenced (right panel). tNGS was performed either by a custom panel (eight-gene panel), a 48-gene panel (TruSeq Amplicon Cancer Panel, Illumina, San Diego, CA), a 54-gene myeloid panel (TruSight Myeloid Sequencing Panel, Illumina) or a custom *BRCA1/2* consortium panel. (B) The bar plot depicts the number of sequence variants detected in tumor DNA of 139 patients using tNGS. The bars indicate the numbers of mutations in a given gene (black) and sequence variants that are annotated in COSMIC (gray). The numbers of actionable mutations is shown in green (drug sensitizing) and red (drug resistance). (C) The bar plot depicts the 30 most frequently somatic mutated genes of 36 patients analyzed by whole-exome sequencing (WES). The colors indicate different tumor entities. Mutations with a variant allele frequency > 10% and a minor allele frequency < 0.001 were considered. The GI tumor category includes liver, pancreas, stomach, and esophagus. CUP, carcinoma of unknown primary; HPV, human papillomavirus; IHC, immunohistochemistry; ISH, in situ hybridization; MSI, microsatellite instability.

hotspot mutation[23,24] (http://cancerhotspots.org/#/home), TARGET db (http://archive.broadinstitute.org/cancer/cga/target), drug-gene interaction (DGIdb; http://www.dgidb.org),[25] and CADD (http://cadd.gs.washington.edu), and categorized according to the predicted impact on protein function by Condel.[26-31] Copy number alteration analysis was performed using Control-FREEC.[32] The STAR[33] aligner was used to align and infer the gene expression level. FusionCatcher (https://doi.org/10.1101/011650) was used to predict gene fusions. Differentially expressed genes were identified using the limma-voom package from R/Bioconductor.[34,35]

RESULTS

From March 2015 through February 2017, 49 MTB meetings were attended by a median of 16 physicians and scientists, ensuring continuous interdisciplinary data interpretation and discussions with diagnostic and therapeutic decision making. The workflow of the MTB included a case and literature review, molecular diagnostic recommendations, and follow-up discussions of the molecular diagnostic results, including treatment recommendations (Appendix Fig A1). Thus, a total of 385 case discussions were held for 198 patients (1.9 discussions per patient;

Table 1. Results

Recommendations	No. (%)
Meetings, No.	49
Case discussions, total No. (per-patient average)	385 (1.9)
Recommendations, No.	505
Diagnostic	305 (60.4)
Therapeutic	104 (20.6)
No treatment recommendation	77 (15.2)
Conditional recommendation	14 (2.8)
Genetic counseling	4 (0.8)
Referral to organ board	1 (0.2)
Diagnostic recommendations, total No. (per-patient average)	305 (1.5)
Patients with diagnostic recommendations	172 (86.9)
Routine pathology	153 (89.0)
Extended genetic analysis*	69 (40.1)
Rebiopsy	15 (8.7)
Other	6 (3.5)
Diagnostic recommendations, not implemented	71 (23.3)
Patients with treatment recommendations	104 (52.5)
Implemented	33 (31.7)
Partial response	11 (33.3)
Stable disease	8 (24.2)
No. of treatment recommendations	104
Treatment recommendations, not implemented	71 (68.3)

NOTE: Data presented as No. (%) unless otherwise noted. See Figure 2 and Data Supplement for details of treatment recommendations and treatment recommendations not implemented.
*Whole-exome sequencing, RNA-seq.

Table 1). In total, 505 structured recommendations were given (2.5 per patient; Table 1). These included 305 diagnostic and 104 treatment recommendations.

Patient Characteristics

The average patient age at the time of the initial MTB presentation was 58 years (range, 1 to 85 years). Detailed patient characteristics are listed in Table 2. One hundred ninety-one of 198 patients (96.4%) had an underlying malignant condition. Patients with solid tumors largely outbalanced hematopoietic malignancies (95.5% *v* 4.5%). Soft tissue tumors (12.6%), CNS tumors (11.1%), and carcinoma of unknown primary (CUP; 10.1%) were the most frequent tumor entities. The majority of patients (n = 146; 73.7%) suffered from metastatic disease, and 77.3% (n = 153) showed disease progression while receiving the standard treatment (Table 2). The mean time interval from diagnosis to first MTB discussion was 33.6 months (range, 1 to 541 months). Patients with treatment-refractory metastatic disease had undergone a mean of 2.0 (range, one to 11) lines of systemic pretreatments. A minority of the patients was referred to the board with rare tumors (n = 33; 16.7%) or because of young age (n = 3; 1.5%).

Molecular Diagnostic Testing

The distribution of molecular diagnostic recommendations is shown in Table 1. For 172 patients (86.9%), 305 recommendations were given and included routine molecular tests in 153 (89%), extended genetic analysis in 69 (40.1%), and both in 53 (30.8%) patients. Rebiopsies were recommended in 15 cases, mostly because of lack of adequate tissue. Of all diagnostic recommendations, 234 (76.7%) were implemented, resulting in 867 single diagnostic tests (mean, five per patient), including 815 routine molecular tests and 52 extended genetic analyses (Fig 1A, left panel).

Routine molecular diagnostics included immunohistochemical (IHC) staining for biomarkers (n = 492; Fig 1A, middle panel), such as programmed death-ligand 1 (PD-L1) and mismatch repair proteins, in situ hybridizations (ISH) for gene copy number analyses (n = 92), and testing for microsatellite instability and/or gene hotspot variations (n = 89) and tNGS (n = 139; Fig 1A). The latter included libraries of different gene panels (Fig 1A; right panel). The most frequent COSMIC annotated sequence variants detected by tNGS occurred in *TP53*, *BRCA1*, *KDR*, *KIT*, *KRAS*, *PIK3CA*, *BRCA2*, and *BRAF* (Fig 1B; Data Supplement). Therapeutically relevant mutations in hotspot regions were identified in 41 of 139 patients (29.5%), including drug-sensitizing variants in *BRAF*, *PIK3CA*, *IDH1*, *EGFR*, and *KIT*, as well as drug resistance variants in *KRAS* and *NRAS*.

Extended genetic analyses including exome and transcriptome assays were performed for 36 patients (18.2%; WES and RNA-Seq: n = 35; RNA-Seq only: n = 1). In those, we identified a total of 5,335 variants, including 18 COSMIC annotated hotspot mutations (Data Supplement). Sixteen were classified as therapy relevant according to the DGI and the TARGET databases. Among the remaining non-hotspot mutations, 1,518 were annotated in COSMIC, including 288 and 28 mutations annotated in

Table 2. Patient Characteristics

Characteristic	
Total, No.	198
Sex	
Female	84 (42.4)
Male	114 (57.6)
Age, years, median (range)	58 (1-85)
Time interval from diagnosis to first MTB, months, median (range)	33.6 (1-541)
Tumor type	
Soft tissue	25 (12.6)
CNS	22 (11.1)
Unknown primary site	20 (10.1)
Colorectal	19 (9.6)
Urogenital	15 (7.5)
Thyroid	14 (7.0)
Breast	10 (5.1)
Lung	9 (4.5)
Hepatobiliary	8 (4.0)
Skin	8 (4.0)
Upper GI tract	8 (4.0)
Hematologic	8 (4.0)
Neuroendocrine	6 (3.0)
Pediatric	5 (2.5)
Head and neck	1 (0.5)
Others	13 (6.6)
Solid tumors (n = 189): stage at presentation	
Complete remission	1 (0.5)
Localized disease	42 (22.2)
Metastatic disease	146 (77.3)
No. of previous lines of therapy, mean (range)	2.0 (0-11)
0	25 (12.6)
1	66 (33.3)
2 to 3	80 (40.4)
> 3	27 (13.6)
Reason for referral	
Progressive disease after standard treatment	153 (77.3)
Rare tumor	33 (16.7)
Others	12 (6.1)

NOTE: Data presented as No. (%) unless otherwise noted.
Abbreviation: MTB, Molecular Tumor Board.

DGI and TARGET databases, respectively (Data Supplement). A total of 3,799 mutations were unknown to COSMIC (Data Supplement). The disease impact of non-hotspot mutations is more difficult to evaluate; however, it can lead to additional therapy-relevant insights. For example, the *ERBB2* S656F mutation might, according to TARGET and DGI databases, constitute an activating mutation, therefore targetable by trastuzumab or lapatinib. The most frequently mutated genes were *TP53* and *BRAF* (Fig 1C).

Overall, 71 of 305 diagnostic recommendations (23.3%) were not pursued. As shown in the Data Supplement, reasons for nonadherence included technical reasons (53.5%; mostly lack of sufficient tissue or DNA/RNA), patient death (12.7%), loss to follow-up (11.3%), medical reasons (9.9%), or patient will (9.9%).

Treatment Recommendations

Specific treatment recommendations were given to 104 patients (Table 1; Fig 2) and mainly included off-label immune checkpoint inhibitor (CPI; n = 36; 34.6%), off-label targeted therapy (n = 19; 18.3%) with tyrosine kinase inhibitors, small molecules or antibodies that were not CPI (AB), trial inclusions (n = 13; 12.5%), and off-label combination treatments (n = 18; 17.3%; Data Supplement; Fig 2). Ninety of 104 treatment recommendations (86.5%) were either off-label therapies (n = 77) or trial inclusions (n = 13).

The implementation rate of treatment recommendations was 31.7% (33 of 104). In-label recommendations were pursued in nine of 14 cases (64.3%), whereas off-label recommendations and trial inclusions were implemented in only 28.6% (22 of 77) and 15.4% (two of 13) of the cases, respectively. Intended trial inclusion in 11 patients failed because of poor performance status or patient death (n = 5), closed trial arm (n = 4), or patient will (n = 2). Main reasons for nonimplementation of treatment recommendations included loss to follow-up (22.5%), recommendation in the future (19.7%), patient death (16.9%), patient will (14.1%), and medical reasons (14.1%; Data Supplement). Of note, evidence level of individual off-label recommendations did not affect implementation rates (data not shown).

Clinical Outcome

In 33 patients with implemented treatment recommendations, partial remissions (PR) and stable diseases (SD) were seen in 11 (33.3%) and eight patients (24.2%; Table 1), respectively. After excluding in-label therapies, nine patients

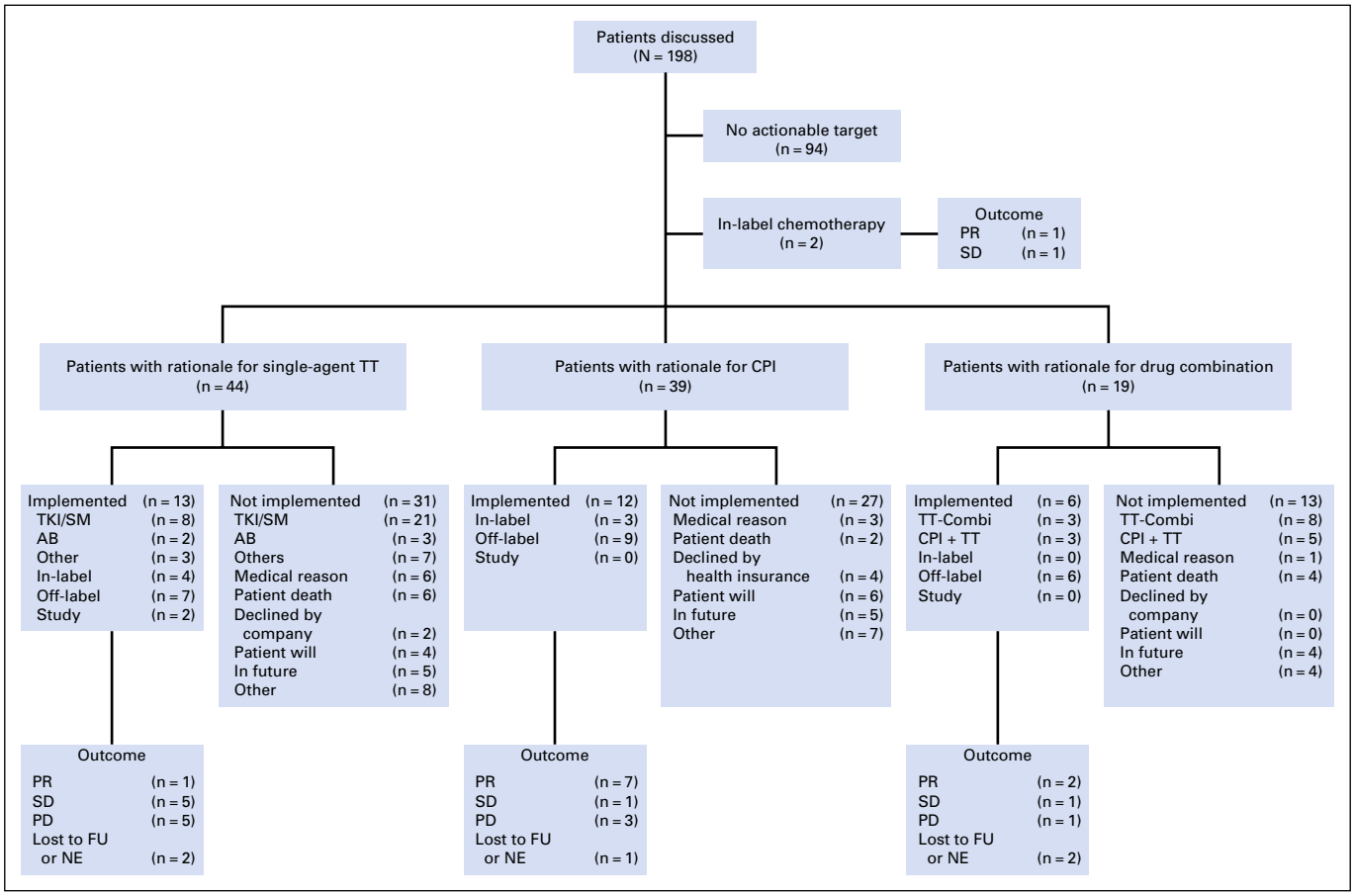


Fig 2. Flow diagram of patients discussed at the Molecular Tumor Board. Responses were determined according to Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1. AB, antibody; Combi, combination; CPI, checkpoint inhibitor; FU, follow-up; NE, not evaluable; PD, progressive disease; PR, partial remission; SD, stable disease; SM, small molecule; TKI, tyrosine kinase inhibitor; TT, targeted therapy.

achieved PR and five patients SD, resulting in an overall response rate of 4.6% (nine of 198 patients) and a disease control rate (DCR) of 7.1% (14 of 198 patients). Of note, all five patients experiencing SD experienced disease progression while receiving the previous treatment. Of 14 responders receiving off-label therapies, eight (57.1%) showed a progression-free survival (PFS) ratio (PFS2/PFS1; PFSr) > 1.3, supporting the impact of the recommended therapies.[36] Three patients had a PFSr < 1.3 with ongoing responses, meaning that their PFSr is still increasing. Details about the outcome of responding patients are shown in Table 3. Two individual cases are shown in the Data Supplement. Adherence to recommendations and outcome according to type of treatment is shown in Fig 2. To assess whether implementation of treatment recommendations affected overall survival from first MTB discussion, we analyzed all patients with stage IV malignancies according to three subgroups (n = 148; Fig 3). The median survival was not reached for patients with implemented treatment recommendations (n = 33 recommendations pursued; 95% CI, 9 months to not reached), 8 months for patients for whom treatment recommendations were not implemented (n = 43 recommendations not pursued; 95% CI, 3 to 10 months), and 10 months for patients who did not receive a treatment recommendation (n = 72 no recommendations; 95% CI, 7 to 17 months). Patients who did not receive the recommended therapy because of death before treatment initiation (n = 12) were excluded from analysis.

DISCUSSION

In a cohort of 198 patients with mostly advanced malignancies beyond standard-of-care treatment, the Comprehensive Cancer Center Freiburg MTB identified actionable targets in 52.5% of the cases. Thirty-two percent received the recommended treatment. In 33 patients with implemented treatment recommendations the disease-control rate was 57.6%; it was 9.6% (19 of 198 patients) for the entire cohort. Because the primary goal of an MTB is to give treatment recommendations beyond standard of care, we excluded five responders who received in-label

Table 3. Patients With Tumor Response

Cancer Type	Rationale for Treatment Recommendation	Board Recommendation	EL	L	R	PFS2 (week)	PFS1 (week)	PFSr	Outcome
Adeno-CUP	Reclassification: histopathological most likely from GI tract	Chemotherapy with FOLFIRI	A1	In	PR	36	NE	NE	PR for 9 months, then PD and death after 14 months
	Immunogenic tumor (PD-L1 positive in cancer cells, abundant TILs, high mutational burden with 358 SNVs and 7 LoH)	Checkpoint inhibition with nivolumab	B3	Off	PR	> 76	70	> 1.1	PR since 17 months and ongoing
	Reclassification: histopathological most likely pancreatic cancer	Second-line treatment with gemcitabine and NAB-paclitaxel	A1	In	SD	61	17	3.6	For 14 months, then PD
ALL	Combination with JAK inhibitor for synergistic antistromal effect[37]	Combination of ruxolitinib and nilotinib analog to CoRNea trial (ClinicalTrials.gov identifier: NCT02253277)	C1	Off	PR	22	39	0.6	Marked decrease of *BCR-ABL* transcript ratio for 4 months, then PD
Ependymoma	Activation of PI3K/mTOR pathway suggested by positive pS6K staining (70% of tumor cells)	Targeted therapy with sirolimus	C3	Off	SD	65	11	5.9	For 16 months, then PD
CRC	Immunogenic tumor (PD-L1 positive in 5% of cancer cells, abundant TILs, MSI deficiency)	Checkpoint inhibition with pembrolizumab	A3	Off	PR	> 99	17	> 5.8	Since 22 months and ongoing
GIST	In-label first-line treatment	Imatinib	A1	In	SD	NE	154	NE	Since 35 months and ongoing
GBM	Immunogenic tumor (PD-L1 positive in 20% of cancer cells, TILs in tumor periphery)	Combination of nivolumab and bevacizumab	B1	Off	PD + CI	12	35	0.3	Significant CI after 2 months. Imaging at month 4 demonstrated PD
Lymphoma plus NBCCS	After chemotherapy (Hodgkin lymphoma) the patient is under risk to develop basalioma or other secondary neoplasia because of the underlying NBCCS.	Prophylaxis with hedgehog inhibitor cholecalciferol	C3	Off	NE	NE	NE	NE	No basaliomas since start of chemotherapy for 13 months and ongoing
Melanoma	BRAF exon 15 mutation D594G described as inhibitory mutation[38]	Combination of targeted therapy with sorafenib and trametinib	C3	Off	SD	11	11	1.0	SD at week 11 with regressive liver metastases. Clinical progression at week 21. Patient died at month 7.
NSCLC	Mutation in EGFR exon 19	Targeted therapy with gefitinib	A1	In	PR	NE	32	NE	Initially good PR, then PD after 8 months
Ovary	No molecular rationale but trial availability	Trial (anti-VEGF/ANG2 nanobody BI 836880; ClinicalTrials.gov identifier: NCT02674152)	C1	Off	SD	69	11	6.3	For 16 months, then PD
Squamous-CUP	Immunogenic tumor (51 SNVs and 13 LoH)	Checkpoint inhibition with nivolumab	B3	Off	SD	13	22	0.6	Stable disease and significant clinical improvement for 4 months. Then PD and death.

(Continued on following page)

Table 3. Patients With Tumor Response (Continued)

Cancer Type	Rationale for Treatment Recommendation	Board Recommendation	EL	L	R	PFS2 (week)	PFS1 (week)	PFSr	Outcome
SCLC	Results from Checkmate-032 demonstrating responses to nivolumab independent of PD-L1 expression (PD-L1 < 1% on cancer cells)	Checkpoint inhibition with nivolumab	A3	Off	PR	68	16	4.3	Since 16 months and ongoing
PNET	Study availability for solid tumors	Trial (anti-VEGF/ANG2 nanobody BI 836880; ClinicalTrials.gov identifier: NCT02674152)	C1	Off	SD	> 61	169	> 0.4	Since 14 months and ongoing
Sarcoma	Despite nonimmunogenic tumor (15 SNVs, 1 LoH, no PD-L1 on cancer cells, few TILs), recent data show response to checkpoint inhibition (SARC028 study)	Checkpoint inhibition with pembrolizumab	A3	Off	PR	78	228	> 0.3	Since 18 months and ongoing
Thyroid (medullary)	RET M918T	Targeted therapy with vandetanib	A1	In	SD	NE	78	NE	For 18 months, then PD
Thyroid (anaplastic)	Immunogenic tumor (PD-L1 positive in 80% of cancer cells, abundant TILs, 412 SNVs, and 112 LoH, suggesting genetic instability)	Checkpoint inhibition with pembrolizumab	A3	Off	PR	98	5	19.6	PR at month 2 and ongoing since 22 months. Normalization of thyroglobulin at month 5 and ongoing. Discontinuation of pembrolizumab at month 8
Thyroid (anaplastic)	Immunogenic tumor (PD-L1 positive in 60% of cancer cells, 1,234 SNVs, and 16 LoH, suggesting genetic instability)	Combination of lenvatinib and pembrolizumab	B3	Off	PR	96	16	6.0	Since 22 months and ongoing
Thyroid (anaplastic)	Immunogenic tumor (PD-L1 positive in 50% of cancer cells, moderate infiltration of TILs, high mutational burden, with 99 SNVs and 12 LoH), SNVs and CNVs indicating enrichment in MAPK-PI3K signaling	Combination of lenvatinib and pembrolizumab	B3	Off	PR	70	6	11.7	For 16 months, then PD
Thyroid (papillary)	Immunogenic tumor (PD-L1 positive in 40% of cancer cells, abundant TILs); results of Keynote-012 trial.	Checkpoint inhibition with pembrolizumab	C3	Off	PR	25	8	3.1	PR for 5 months, then PD of the primary tumor

NOTE. Twenty-one of 198 patients received the recommended therapy and showed at least SD (19 patients) and/or significant CI (one patient). One patient was treated pre-emptively; therefore, the response was not evaluable. Listed are the diagnostic results that constituted the rationale for treatment recommendation, the board recommendation including level of evidence (EL), in- or off-label status (L), treatment response (R), PFS of the recommended treatment (PFS2; PFS1 if first line), PFS of the previous systemic treatment (PFS1), PFS ratio (PFSr, PFS2/PFS1), and outcome.

Abbreviations: ALL, acute lymphoblastic leukemia; ANG2, angiopoietin-2; CI, clinical improvement; CNV, copy number variation; CRC, colorectal cancer; CUP, cancer of unknown primary; FOLFIRI, fluorouracil, leucovorin, and irinotecan; GBM, glioblastoma; GIST, GI stromal tumor; LoH, loss of heterozygosity; MAPK, mitogen-activated protein kinase; MSI, microsatellite instability; mTOR, mammalian target of rapamycin; NAB-paclitaxel, nanoparticle albumin-bound paclitaxel; NBCCS, nevoid basal cell carcinoma syndrome; NE, not evaluable; NSCLC, non–small-cell lung cancer; PD, progressive disease; PD-L1, programmed death-ligand 1; PFS, progression-free survival; PI3K, phosphatidylinositide 3-kinase; PNET, primitive neuroectodermal tumor; PR, partial response; SCLC, small-cell lung cancer; SD, stable disease; SNV, single-nucleotide variant; TIL, tumor-infiltrating lymphocyte; VEGF, vascular endothelial growth factor.

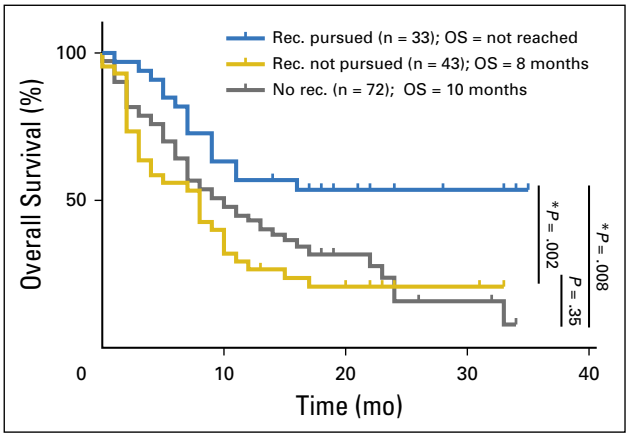


Fig 3. Survival analysis. The Kaplan-Meier curve shows the survival of the following three subgroups of patients with stage IV malignancies (n = 148): patients who implemented the treatment recommendation (Rec. pursued, n = 33), patients who did not implement the treatment recommendation (Rec. not pursued, n = 43; of note: patients who did not receive the recommended therapy because of death before treatment initiation [n = 12] were excluded from analysis), and patients who did not receive a treatment recommendation (n = 72). The curve comparison with the log-rank (Mantel-Cox) test revealed statistical significant differences as shown on graph. OS, overall survival. (*) $P < .01$.

therapies resulting in a DCR of 7.1% (14 of 198 patients). Other MTB case series reported DCRs in 3.2%, 7.8%, 9%, and 23.3% of the patients,[38-41] suggesting that approximately 10% of patients might benefit from advanced personalized decision making.

Although molecular heterogeneity will limit the effect of therapeutic kinase inhibitors, higher nonsynonymous mutational burden can create more neoantigens and therefore improve response rates to CPI.[42,43] In our series, eight of 11 patients (72.7%) showing PR received CPI, including seven off-label uses. Predictive biomarkers for individualized immunotherapies are emerging and changing rapidly, with strong differences between entities.[44] Here, we used IHC for programmed cell death protein 1 (PD-1)/PD-L1, tumor-infiltrating lymphocytes, microsatellite instability testing, and mutational burden assessment as predictive biomarkers. In the near future, identifying individual cancer neoantigens might allow a more precise prediction of responses to immunotherapies.[45] This highlights the importance of an interdisciplinary MTB team that analyzes and interprets biomarkers to identify patients who might benefit from off-label immuno-oncology treatments.

In an MTB workflow, the portfolio of molecular diagnostic tests, as well as criteria to match and prioritize targeted therapies to molecular biomarkers, affects the probability to identify patients with actionable targets. Here, we used customized molecular diagnostics, including IHC/ISH and tNGS, in 153 out of 198 patients (77.3%) We implemented WES or RNA-Seq analyses for patients with carcinomas of unknown primary and rare cancers and with diseases in which routine molecular diagnostics did not reveal any actionable target (18.2% of patients).

Multidimensional data have not been implemented successfully to clinical routine, partly because of the complexity of developing and evaluating mathematical predictive models.[46,47] A recent analysis showed that an MTB workflow including WES/whole-genome sequencing, RNA-Seq, and data interpretation by a multidisciplinary board required a turnaround time of 6 weeks.[48] Using high-dimensional molecular data, the Molecular Screening for Cancer Treatment Optimization (MOSCATO-01) trial reported actionable mutations in less than half of the patients with advanced solid tumors,[49] and in the National Cancer Institute Molecular Analysis for Therapy Choice (NCI-MATCH) trial, only 9% of the patients could be assigned to one of the prespecified treatment arms.[50] In contrast, our approach of customized molecular diagnostic testing with restricted use of extended genetic analyses (WES, RNA-Seq) allows a faster turnover with comparable rates of genetically matched treatment recommendations. Therefore, average costs per case can be reduced at least by half when compared with performing extended molecular analysis for each patient. We identified actionable targets in 52.5% of cases and provided treatment recommendations with a median turnaround time of 28 days. To improve standardization and turnaround time, we recently implemented SOPs for diagnostic workups (Data Supplement). Our approach shares similarities with Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT), focusing on therapeutically targetable biomarkers for fast clinical decision making and referral of patients to available clinical trials.[51]

Targeted drug combinations might offer better DCR over single-agent therapies.[52-55] In part, this is due to crosstalk between signaling pathways as well as spatial and temporal clonal heterogeneity, especially in patients with advanced cancer who failed standard-of-care treatment.[56,57] Most current programs for precision oncology use prespecified, genetically matched, single-agent treatments (NCI-MATCH, ClinicalTrials.gov identifier: NCT02465060; or Targeted Agent and Profiling Utilization Registry [TAPUR], ClinicalTrials.gov identifier: NCT02693535). In our series, three of 19

treatments that successfully controlled disease (15.8%) included molecular combination treatments (Fig 2). These patients did not suffer from grade 3 to 4 adverse effects, although treatment combinations may bear a higher risk of toxicity.[58]

Earlier referral to an MTB (eg, after failure of first-line treatment) might prevent the institution of ineffective treatments, improve the implementation rate, and increase the likelihood of success of molecular biomarker–matched treatments. In our series, patient death, patient preference, or medical reasons precluded implementation in 23.3% of diagnostic and 68.3% of treatment recommendations. The survival analysis revealed a significant overall survival advantage for patients with implemented MTB treatment recommendations (median overall survival not reached; 95% CI, 9 months to not reached) compared with patients where recommendations were not pursued (8 months; 95% CI, 3 to 10 months; $P = .002$) as well as for patients without treatment recommendation (10 months; 95% CI, 7 to 17 months; $P = .008$). Because of the low sample size and the heterogeneous composition of patients in the cohorts, the validity of this survival analysis is limited.

Access to molecular biomarker–matched, off-label agents for cancer treatment is limited. In a recent single-center study, only 5% of molecular biomarker–matched treatment recommendations were implemented, mainly because of limited access to clinical trials or to restricted use of drugs outside their marketed label.[59] Thus, it is crucial to build up platforms for patients and treating physicians to link individual molecular information of the tumor to appropriate nonapproved drugs and available clinical trials. To this end, MTB networks might implement SOPs for diagnostic work-ups and data interpretation and build alliances to governmental institutions and insurance companies to generate criteria for the financial coverage of molecular analyses and off-label treatments. Finally, an MTB is predestined to generate knowledge and evidence in oncology via single-person trials instead of large, time- and cost-intensive clinical trials. In case of sequence variants with undetermined significance, precision oncology workflows should allow fast reverse translation of sequence variants into informative preclinical models. In a patient with melanoma, we identified a kinase-inactivating *BRAF* mutation (Data Supplement). In vitro characterization indicated antitumor activity of combined pan-RAF and mitogen-activated protein kinase kinase inhibition and guided successful treatment with sorafenib and trametinib. In rare entities, and especially in the setting of treatment-refractory cancers, precision oncology networks should allow hypothesis-driven in vitro studies and validation in small sets of individuals. Thus, within the concept of patient-centric, biomarker-driven trial designs,[60] an MTB might constitute a critical tool to identify informative patients for clinical trials of targeted therapies in rare molecular subgroups.

In summary, this MTB experience illustrates that patient management, on the basis of individual molecular biomarker profiling and analysis, is feasible in patients beyond standard-of-care treatment. We show a high proportion of trial- and off-label treatment recommendations (86.5%) and a DCR for off-label treatments of 7.1%. In cases where no approved treatment is available, an MTB might allow molecular biomarker–matched off-label use of approved drugs across entity barriers or alternatively facilitate the access to therapeutic basket trials.

DOI: https://doi.org/10.1200/PO.18.00105
Published online on ascopubs.org/journal/po on August 16, 2018.

AUTHOR CONTRIBUTIONS

Conception and design: Rouven Hoefflin, Ralph Fritsch, Rainer Claus, Julius Wehrle, Meike Reiser, Cornelius Miething, Christoph Peters, Martin Werner, Justus Duyster, Tilman Brummer, Silke Lassmann, Nikolas von Bubnoff

Financial support: Martin Werner, Silke Lassmann

Administrative support: Rouven Hoefflin, Rainer Claus, Martin Boeker, Melanie Boerries

Provision of study material or patients: Julius Wehrle, Dagmar von Bubnoff, Isabell Ge, Steffen Heeg, Martin Boeker, Simone Hettmer, Silke Lassmann

Collection and assembly of data: Rouven Hoefflin, Anna-Lena Geißler, Ralph Fritsch, Rainer Claus, Meike Reiser, Leman Mehmed, Lisa Fauth, Dieter Henrik Heiland, Friedrich Stock, Agnes Csanadi, Cornelius Miething, Britta Weddeling, Frank Meiss, Dagmar von Bubnoff, Christine Dierks, Isabell Ge, Steffen Heeg, Henning Schäfer, Justyna Rawluk, Gian Kayser, Simone Hettmer, Justus Duyster, Melanie Boerries, Silke Lassmann, Nikolas von Bubnoff

Data analysis and interpretation: Rouven Hoefflin, Anna-Lena Geißler, Ralph Fritsch, Rainer Claus, Julius Wehrle, Patrick Metzger, Meike Reiser, Lisa Fauth, Thalia Erbes, Cornelius Miething, Britta Weddeling, Volker Brass, Henning Schäfer, Martin Boeker, Justyna Rawluk, Elke Maria Botzenhart, Gian Kayser, Hauke Busch, Martin Werner, Justus Duyster, Tilman Brummer, Melanie Boerries, Silke Lassmann, Nikolas von Bubnoff

Manuscript writing: All authors

Final approval of manuscript: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/po/author-center.

Rouven Hoefflin
Employment: Roche Pharma AG (I)

Anna-Lena Geißler
No relationship to disclose

Ralph Fritsch
Honoraria: Genentech, Servier, Merck, Amgen

Consulting or Advisory Role: Genentech

Travel, Accommodations, Expenses: Amgen, Celgene, Servier, Merck

Rainer Claus
Stock and Other Ownership Interests: Medigene, Evotec

Honoraria: Roche, Abbvie, Novartis, Janssen Oncology, Gilead Sciences

Consulting or Advisory Role: Abbvie, Roche, Janssen Oncology

Speakers' Bureau: Janssen Oncology, Roche

Travel, Accommodations, Expenses: Janssen Oncology, Abbvie

Julius Wehrle
Travel, Accommodations, Expenses: PharmaMar, Novartis

Patrick Metzger
No relationship to disclose

Meike Reiser
Travel, Accommodations, Expenses: AstraZeneca

Leman Mehmed
No relationship to disclose

Lisa Fauth
No relationship to disclose

Dieter Henrik Heiland
No relationship to disclose

Thalia Erbes
Honoraria: Roche, Novartis, Teva, Pfizer, AstraZeneca, Eisai

Friedrich Stock
No relationship to disclose

Agnes Csanadi
No relationship to disclose

Cornelius Miething
Consulting or Advisory Role: Roche Pharma AG, Novartis

Travel, Accommodations, Expenses: Celgene

Britta Weddeling
No relationship to disclose

Frank Meiss
Honoraria: Bristol-Myers Squibb, Novartis, MSD Oncology

Consulting or Advisory Role: Bristol-Myers Squibb, Novartis

Travel, Accommodations, Expenses: Novartis, Bristol-Myers Squibb

Dagmar von Bubnoff
Honoraria: Roche, iOMEDICO AG

Travel, Accommodations, Expenses: Bristol-Myers Squibb

Christine Dierks
No relationship to disclose

Isabell Ge
No relationship to disclose

Volker Brass
No relationship to disclose

Steffen Heeg
No relationship to disclose

Henning Schäfer
Honoraria: Genentech

Consulting or Advisory Role: Bristol-Myers Squibb

Travel, Accommodations, Expenses: Bristol-Myers Squibb

Martin Boeker
No relationship to disclose

Justyna Rawluk
Consulting or Advisory Role: Roche Pharma AG, Takeda, Boehringer Ingelheim, MSD, Bristol-Myers Squibb, Chugai Pharma, AstraZeneca

Elke Maria Botzenhart
No relationship to disclose

Gian Kayser
No relationship to disclose

Simone Hettmer
No relationship to disclose

Hauke Busch
No relationship to disclose

Christoph Peters
Leadership: University Medical Center Freiburg
Honoraria: University Medical Center Freiburg
Travel, Accommodations, Expenses: Novartis

Martin Werner
Honoraria: Diakovere
Consulting or Advisory Role: Roche, Novartis, Johnson & Johnson
Research Funding: Agilent (Inst), Novartis (Inst), Roche (Inst)

Justus Duyster
Honoraria: Novartis, Genentech, Pfizer
Consulting or Advisory Role: Novartis, Roche
Travel, Accommodations, Expenses: Novartis

Tilman Brummer
No relationship to disclose

Melanie Boerries
No relationship to disclose

Silke Lassmann
Honoraria: Novartis, AstraZeneca
Travel, Accommodations, Expenses: Novartis, AstraZeneca

Nikolas von Bubnoff
Honoraria: AstraZeneca, Amgen, Bristol-Myers Squibb
Consulting or Advisory Role: Novartis
Research Funding: Novartis
Travel, Accommodations, Expenses: Novartis

ACKNOWLEDGMENT

We thank the team of the Genomics and Proteomics Core Facility, German Cancer Research Center/DKFZ, Heidelberg, Germany, for their sequencing service.

Affiliations

All authors: University of Freiburg, Freiburg; **Ralph Fritsch**, **Julius Wehrle**, **Cornelius Miething**, **Christoph Peters**, **Martin Werner**, **Justus Duyster**, **Tilman Brummer**, **Melanie Boerries**, **Silke Lassmann**, and **Nikolas von Bubnoff**, German Cancer Consortium, partner site Freiburg, and German Cancer Research Center, Heidelberg; **Rainer Claus**, Augsburg Medical Center, Augsburg; and **Hauke Busch**, University of Lübeck, Lübeck, Germany.

Support

Supported by Comprehensive Cancer Center Freiburg and by Deutsches Konsortium für Translationale Krebsforschung Grant No. L628 (N.v.B., S.L.). This work was also supported by Deutsche Forschungsgemeinschaft through Grant No. SFB 850, Z1 (H.B., M. Boerries, S.L.) and a Heisenberg professorship (T.B.); by the Federal Ministry for Education and Research through the framework of the e:Med research and funding concept, DeCaRe Grant No. FKZ 01ZX1409B (M. Boerries); the Federal Ministry for Education and Research Medical Informatics Initiative Consortium MIRACUM FKZ 01ZZ1606H (P.M., M. Boerries, M. Boeker); and the Berta-Ottenstein-Programme, Faculty of Medicine, University of Freiburg (R.H., J.W.).

REFERENCES

1. Chapman PB, Hauschild A, Robert C, et al: Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 364:2507-2516, 2011
2. Hoadley KA, Yau C, Wolf DM, et al: Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell 158:929-944, 2014
3. Hyman DM, Puzanov I, Subbiah V, et al: Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. N Engl J Med 373:726-736, 2015
4. Borghaei H, Paz-Ares L, Horn L, et al: Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. N Engl J Med 373:1627-1639, 2015
5. Prahallad A, Sun C, Huang S, et al: Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. Nature 483:100-103, 2012
6. Hong DS, Morris VK, El Osta B, et al: Phase IB study of vemurafenib in combination with irinotecan and cetuximab in patients with metastatic colorectal cancer with BRAFV600E mutation. Cancer Discov 6:1352-1365, 2016
7. Stones CJ, Kim JE, Joseph WR, et al: Comparison of responses of human melanoma cell lines to MEK and BRAF inhibitors. Front Genet 4:66, 2013
8. Lawrence MS, Stojanov P, Polak P, et al: Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499:214-218, 2013

9. Gerlinger M, Rowan AJ, Horswell S, et al: Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 366:883-892, 2012

10. Xu X, Hou Y, Yin X, et al: Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell 148:886-895, 2012

11. Siravegna G, Mussolin B, Buscarino M, et al: Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. Nat Med 21:795-801, 2015 [Erratum: Nat Med 21:827, 2015]

12. Kobayashi S, Boggon TJ, Dayaram T, et al: EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N Engl J Med 352:786-792, 2005

13. Goss G, Tsai CM, Shepherd FA, et al: Osimertinib for pretreated EGFR Thr790Met-positive advanced non-small-cell lung cancer (AURA2): A multicentre, open-label, single-arm, phase 2 study. Lancet Oncol 17:1643-1652, 2016

14. Li MM, Datto M, Duncavage EJ, et al: Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. J Mol Diagn 19:4-23, 2017

15. Hirsch B, Endris V, Lassmann S, et al: Multicenter validation of cancer gene panel-based next-generation sequencing for translational research and molecular diagnostics. Virchows Arch 472:557-565, 2018

16. Geißler AL, Geißler M, Kottmann D, et al: ATM mutations and E-cadherin expression define sensitivity to EGFR-targeted therapy in colorectal cancer. Oncotarget 8:17164-17190, 2017

17. Kovaleva V, Geissler AL, Lutz L, et al: Spatio-temporal mutation profiles of case-matched colorectal carcinomas and their metastases reveal unique de novo mutations in metachronous lung metastases by targeted next generation sequencing. Mol Cancer 15:63, 2016

18. Lek M, Karczewski KJ, Minikel EV, et al: Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285-291, 2016

19. Landrum MJ, Lee JM, Riley GR, et al: ClinVar: Public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42:D980-D985, 2014

20. Forbes SA, Beare D, Gunasekaran P, et al: COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res 43:D805-D811, 2015

21. Sherry ST, Ward M, Sirotkin K: dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res 9:677-679, 1999

22. Sherry ST, Ward MH, Kholodov M, et al: dbSNP: The NCBI database of genetic variation. Nucleic Acids Res 29:308-311, 2001

23. Chang MT, Asthana S, Gao SP, et al: Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat Biotechnol 34:155-163, 2016

24. Chang MT, Bhattarai TS, Schram AM, et al: Accelerating discovery of functional mutant alleles in cancer. Cancer Discov 8:174-183, 2018

25. Cotto KC, Wagner AH, Feng YY, et al: DGIdb 3.0: A redesign and expansion of the drug-gene interaction database. Nucleic Acids Res 46:D1068-D1073, 2017

26. Shihab HA, Gough J, Cooper DN, et al: Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat 34:57-65, 2013

27. Reva B, Antipin Y, Sander C: Predicting the functional impact of protein mutations: Application to cancer genomics. Nucleic Acids Res 39:e118, 2011

28. González-Pérez A, López-Bigas N: Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 88:440-449, 2011

29. Chen Y, Cunningham F, Rios D, et al: Ensembl variation resources. BMC Genomics 11:293, 2010
30. Adzhubei IA, Schmidt S, Peshkin L, et al: A method and server for predicting damaging missense mutations. Nat Methods 7:248-249, 2010
31. Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073-1081, 2009
32. Boeva V, Zinovyev A, Bleakley K, et al: Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. Bioinformatics 27:268-269, 2011
33. Dobin A, Davis CA, Schlesinger F, et al: STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29:15-21, 2013
34. Ritchie ME, Phipson B, Wu D, et al: limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47, 2015
35. Law CW, Chen Y, Shi W, et al: voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 15:R29, 2014
36. Von Hoff DD, Stephenson JJ Jr, Rosen P, et al: Pilot study using molecular profiling of patients' tumors to find potential targets and select treatments for their refractory cancers. J Clin Oncol 28:4877-4883, 2010
37. Traer E, MacKenzie R, Snead J, et al: Blockade of JAK2-mediated extrinsic survival signals restores sensitivity of CML cells to ABL inhibitors. Leukemia 26:1140–1143, 2012
38. Larkin J, Ascierto PA, Dréno B, et al: Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. N Engl J Med 371:1867-76, 2014
39. Sohal DP, Rini BI, Khorana AA, et al: Prospective clinical study of precision oncology in solid tumors. J Natl Cancer Inst 108:djv332, 2015
40. Johnson DB, Dahlman KH, Knol J, et al: Enabling a genetically informed approach to cancer medicine: A retrospective evaluation of the impact of comprehensive tumor profiling using a targeted next-generation sequencing panel. Oncologist 19:616-622, 2014
41. Dalton WB, Forde PM, Kang H, et al: Personalized medicine in the oncology clinic: Implementation and outcomes of the Johns Hopkins Molecular Tumor Board. JCO Precision Oncol: 10.1200/PO.16.00046, 2017
42. Rizvi NA, Hellmann MD, Snyder A, et al: Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 348:124-128, 2015
43. McGranahan N, Furness AJ, Rosenthal R, et al: Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science 351:1463-1469, 2016
44. Gibney GT, Weiner LM, Atkins MB: Predictive biomarkers for checkpoint inhibitor-based immunotherapy. Lancet Oncol 17:e542-e551, 2016
45. Gfeller D, Bassani-Sternberg M, Schmidt J, et al: Current tools for predicting cancer-specific T cell immunity. OncoImmunology 5:e1177691, 2016
46. McShane LM, Cavenagh MM, Lively TG, et al: Criteria for the use of omics-based predictors in clinical trials. Nature 502:317-320, 2013
47. Singer J, Irmisch A, Ruscheweyh HJ, et al: Bioinformatics for precision oncology. Brief Bioinform: bbx143, 2017
48. Horak P, Klink B, Heining C, et al: Precision oncology based on omics data: The NCT Heidelberg experience. Int J Cancer 141:877-886, 2017
49. Massard C, Michiels S, Ferté C, et al: High-throughput genomics and clinical outcome in hard-to-treat advanced cancers: Results of the MOSCATO 01 trial. Cancer Discov 7:586-595, 2017
50. Conley BA, Gray R, Chen A, et al: NCI-molecular analysis for therapy choice (NCI-MATCH) clinical trial: interim analysis. Presented at American Association for Cancer Research 107th Annual Meeting 2016, New Orleans, LA, April 16-20, 2016

51. Cheng DT, Mitchell TN, Zehir A, et al: Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. J Mol Diagn 17:251-264, 2015
52. Tannock IF, Hickman JA: Limits to personalized cancer medicine. N Engl J Med 375:1289-1294, 2016
53. Yap TA, Omlin A, de Bono JS: Development of therapeutic combinations targeting major cancer signaling pathways. J Clin Oncol 31:1592-1605, 2013
54. Liu S, Nikanjam M, Kurzrock R: Dosing de novo combinations of two targeted drugs: Towards a customized precision medicine approach to advanced cancers. Oncotarget 7:11310-11320, 2016
55. Turner NC, Reis-Filho JS: Genetic heterogeneity and cancer drug resistance. Lancet Oncol 13:e178-e185, 2012
56. Swanton C, Govindan R: Clinical implications of genomic discoveries in lung cancer. N Engl J Med 374:1864-1873, 2016
57. Shi H, Hugo W, Kong X, et al: Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. Cancer Discov 4:80-93, 2014
58. Soria JC, Massard C, Izzedine H: From theoretical synergy to clinical supra-additive toxicity. J Clin Oncol 27:1359-1361, 2009
59. Beltran H, Eng K, Mosquera JM, et al: Whole-exome sequencing of metastatic cancer and biomarkers of treatment response. JAMA Oncol 1:466-474, 2015
60. Biankin AV, Piantadosi S, Hollingsworth SJ: Patient-centric trials for therapeutic development in precision oncology. Nature 526:361-370, 2015

Appendix

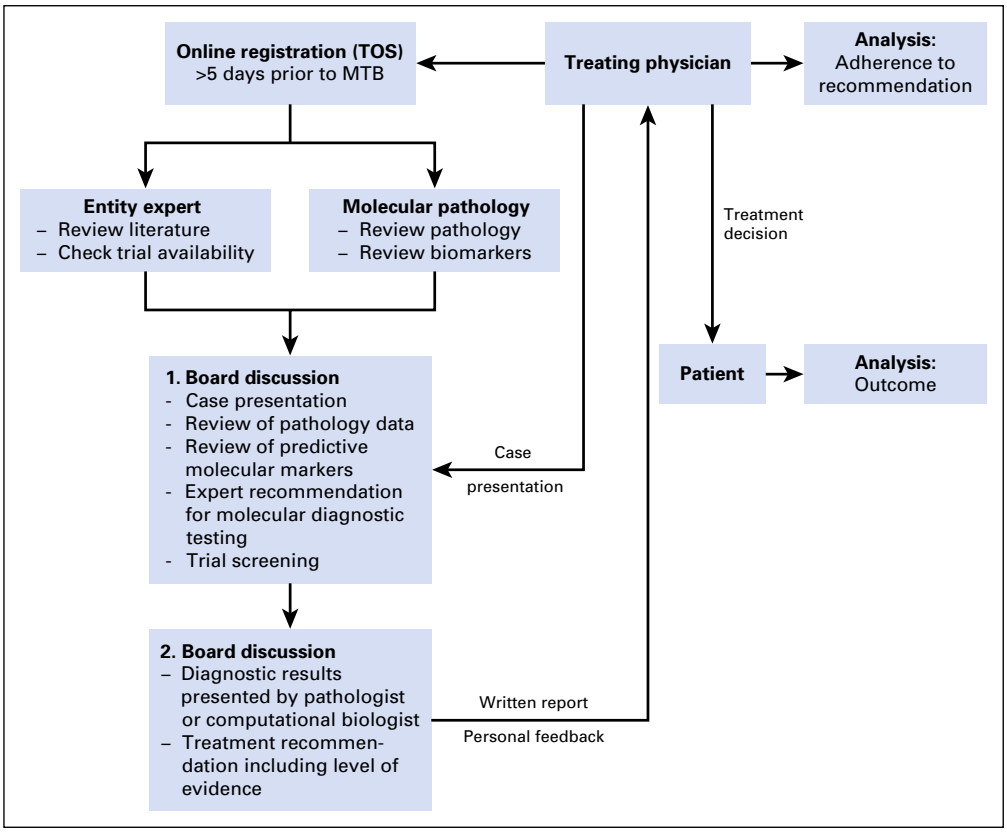


Fig A1. Molecular Tumor Board (MTB) workflow. TOS, Tumorboard Online System.

3.6 Requirements Analysis and Specification for a Molecular Tumor Board Platform Based on cBioPortal

Due to the growing amount of high-throughput sequencing data, clinicians involved in the Interdisciplinary Molecular Tumor Board (MTB) need a platform that supports the processes and visualization of these results within an MTB. Recent data suggest that such MTBs are beneficial for patients as they improve therapy and subsequent care [Hoefflin et al., 2018, Parker et al., 2015, Bryce et al., 2017]. However, there is a great need for standardized tools that support the interpretation and presentation of such large omics data [Hinderer et al., 2017]. To address this issue the MIRACUM (Medical Informatics in Reseach and Care in University Medicine) consortium has established Use Case 3 which focuses on supporting and providing IT, bioinformatics and medical informatics tools to translate and visualize the resulting data [Prokosch et al., 2018] for MTBs. The cBio Cancer Genomics Portal (cBioPortal) was identified as the platform, and two rounds of interviews, supported by descriptive screenshot mockups, allowed additional requirements to be identified for the platform. A total of 24 new requirements were identified that had not been implemented in cBioPortal before. Based on the clinical requirements, the study provides important information to support the members of the MTB in interpreting the complex heterogeneous data for a personalized therapy recommendation.

Buechner, P., Hinderer, M., Unberath, P., Metzger, P., Boeker, M., Acker, T., Haller, F., Mack, E., Nowak, D., Paret, C., Schanze, D., von Bubnoff, N., Wagner, S., Busch, H., Boerries, M., Christoph, J. (2020). **Requirements analysis and specification for a molecular tumor board platform based on CBIoPortal**. Diagnostics.

Contribution: I wrote, reviewed and edited together with the other authors the manuscript.

diagnostics

MDPI

Article

Requirements Analysis and Specification for a Molecular Tumor Board Platform Based on cBioPortal

Philipp Buechner [1], Marc Hinderer [1], Philipp Unberath [1], Patrick Metzger [2,3], Martin Boeker [4], Till Acker [5], Florian Haller [6], Elisabeth Mack [7], Daniel Nowak [8,9], Claudia Paret [10,11], Denny Schanze [12], Nikolas von Bubnoff [13,14,15], Sebastian Wagner [16], Hauke Busch [17], Melanie Boerries [2,18,19,†,*], Jan Christoph [1,†,*]

1 Department of Medical Informatics, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), 91058 Erlangen-Tennenlohe, Germany; philipp.buechner@fau.de (P.B.); marc.hinderer@fau.de (M.H.); philipp.unberath@fau.de (P.U.)
2 Institute of Medical Bioinformatics and Systems Medicine, Faculty of Medicine and Medical Center-University of Freiburg, 79110 Freiburg, Germany; patrick.metzger@mol-med.uni-freiburg.de
3 Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany
4 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, 79104 Freiburg, Germany; martin.boeker@imbi.uni-freiburg.de
5 Institute of Neuropathology, Justus-Liebig-University Giessen, 35392 Giessen, Germany; till.acker@patho.med.uni-giessen.de
6 Institute of Pathology, University Hospital Erlangen, 91054 Erlangen, Germany; florian.haller@uk-erlangen.de
7 Department of Hematology, Oncology and Immunology, Philipps-University Marburg, and University Hospital Gießen and Marburg, Marburg, Germany Baldingerstraße; 35043 Marburg, Germany; elisabeth.mack@staff.uni-marburg.de
8 Department of Hematology and Oncology, Medical Faculty Mannheim, Heidelberg University, 68167 Mannheim, Germany; daniel.nowak@medma.uni-heidelberg.de
9 Heinrich-Lanz-Center for digital health, Medical Faculty Mannheim, Heidelberg University, 68167 Mannheim, Germany
10 Pediatric Hematology/Oncology, Children's Hospital, University Medical Center of the Johannes Gutenberg-University Mainz, 55131 Mainz, Germany; paretc@uni-mainz.de
11 University Cancer Center (UCT) of the University Medical Center of the Johannes Gutenberg-University Mainz, 55131 Mainz, Germany
12 Institute of Human Genetics, University Hospital Magdeburg, Faculty of Medicine, Otto-von-Guericke University, 39120 Magdeburg, Germany; denny.schanze@med.ovgu.de
13 Department of Hematology and Oncology, Medical Center, University of Schleswig-Holstein, Campus Lübeck, 23538 Lübeck, Germany; nikolas.bubnoff@uniklinik-freiburg.de
14 German Cancer Consortium (DKTK), partner site Freiburg, 79106 Freiburg, Germany
15 Department of Hematology, Oncology and Stem Cell Transplantation, Medical Center, Faculty of Medicine, University of Freiburg, 79106 Freiburg, Germany
16 Department of Medicine 2, Hematology/Oncology, Goethe University Hospital, 60590 Frankfurt am Main, Germany; swagner@med.uni-frankfurt.de
17 Institute of Experimental Dermatology and Institute of Cardiogenetics, University of Lübeck, 23562 Lübeck, Germany; hauke.busch@uni-luebeck.de
18 Comprehensive Cancer Center Freiburg (CCCF), Faculty of Medicine and Medical Center – University of Freiburg, 79106 Freiburg, Germany
19 German Cancer Consortium (DKTK), partner site Freiburg; and German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
† These authors contributed equally to this work.
* Correspondence: melanie.boerries@uniklinik-freiburg.de (Melanie Boerries); jan.christoph@fau.de (J.C.)

Received: 15 December 2019; Accepted: 7 February 2020; Published: 10 February 2020

Abstract: Clinicians in molecular tumor boards (MTB) are confronted with a growing amount of genetic high-throughput sequencing data. Today, at German university hospitals, these data are usually handled in complex spreadsheets from which clinicians have to obtain the necessary information. The aim of this work was to gather a comprehensive list of requirements to be met by cBioPortal to support processes in MTBs according to clinical needs. Therefore, oncology experts at nine German university hospitals were surveyed in two rounds of interviews. To generate an interview guideline a scoping review was conducted. For visual support in the second round, screenshot mockups illustrating the requirements from the first round were created. Requirements that cBioPortal already meets were skipped during the second round. In the end, 24 requirements with sometimes several conceivable options were identified and 54 screenshot mockups were created. Some of the identified requirements have already been suggested to the community by other users or are currently being implemented in cBioPortal. This shows, that the results are in line with the needs expressed by various disciplines. According to our findings, cBioPortal has the potential to significantly improve the processes and analyses of an MTB after the implementation of the identified requirements.

Keywords: decision making, computer-assisted; decision support systems, clinical; precision medicine; computational biology; molecular tumor board; cBioPortal; requirements analysis; neoplasms

1. Introduction

Advances in next-generation sequencing (NGS) technology and the resulting decrease of costs hold large promises for personalized medicine, currently revolutionizing cancer diagnostics in particular. The sequencing of whole tumor exomes, genomes and transcriptomes of patients allows physicians to make individual molecular-guided decisions. However, the complex nature of cancer and its large interindividual heterogeneity require an interdisciplinary board composed of medical and scientific experts to review and interpret the equally complex analysis results. Recent data suggest that such molecular tumor boards (MTBs) have the potential to improve therapy and care for patients that have run out of guideline-based treatment options or have rare tumors [1,2].

Several German medical centers have already started to implement MTBs in their clinical environment, all working with various types of omics data from, e.g., NGS and other technologies [3,4]. To handle the many results from a large amount of omics data there is a high need for a standardized toolset that supports clinicians in analyzing and interpreting these data and creating high-quality presentations of complex multi-dimensional data effectively. Moreover, it requires both the integration of clinical with molecular and genomic data and the visualization of joint analysis results. However, experiences with the implementation and establishment of information technology (IT) and bioinformatics support for MTBs are still rare in Germany and probably world-wide, thus in need of improvement and optimization [5].

To address this issue, the MIRACUM consortium (Medical Informatics in Research and Care in University Medicine) has established the Use Case 3 which focuses on the provision of IT and bioinformatics support for translation and visualization of data analyzed in MTBs. As part of this use case, we will establish a generic, open-source framework that supports the analysis, interpretation and visualization of both clinical and omics data [6]. Data analysis is handled through MIRACUM-Pipe [7] which provides automatic, parameter-controlled processing of omics data with alignment, variant calling, annotation and analysis. The second aspect of the framework will be the data visualization and documentation of the results of the MTB. Both, the pipeline and the visualization, will be provided as separate and open-source components developed in a user-centered design process.

The cBio Cancer Genomics Portal (cBioPortal) was selected as a suitable platform to visualize the generated data supplied by the MIRACUM-Pipe [7] and to support the decision-making

processes in an MTB. cBioPortal provides an extensive set of tools for exploring, visualizing and analyzing multi-dimensional and large-scale cancer genomics data sets [8,9]. Within the context of an MTB, cBioPortal may support case preparation, case review, and the documentation and communication of treatment recommendations in the near future. Therefore, it is well suited to replace the current practice at some German university hospitals of managing complex mutation data in huge spreadsheets by providing comprehensible visualizations [10,11]. At the participating clinics, cBioPortal could optimize the processing of up to 200 cases per year with sometimes hundreds of identified but not necessarily relevant mutations, and thus improve the decision making [5].

The integration of cBioPortal into the workflow of MTBs requires adjustments regarding different functionalities and needs (requirements). For instance, to meet some of our requirements the user needs to have write access to the data stored in cBioPortal. Therefore, we must find a proper solution to accomplish this in line with the concepts cBioPortal currently pursues as a (read-only) data warehouse.

The objective of this work is to provide a requirements specification for an IT platform based on cBioPortal that supports processes in molecular tumor boards to find and document a therapy recommendation. To our knowledge, there is so far no systematic assessment of such requirements from the point of view of MTB participants in different hospitals, even though there are already existing tools that integrate numerous data sources and even support the documentation process in a uniform MTB tool [12,13]. Our work could serve as a blueprint for the development of further tools based on cBioPortal for MTBs in Germany and worldwide.

2. Materials and Methods

To identify the requirements, we conducted a qualitative research to assess a set of potential requirements in two consecutive rounds of interviews. The second round of interviews became necessary because the first round was developed iteratively and therefore not all participants had the chance to comment on all mentioned potential requirements.

In preparation, we reviewed the literature published between 1997 and 2017 (scoped review) for existing systems, tools and knowledgebases that support molecular tumor boards (see Figure 1), resulting in an interview guideline for the first round of interviews.

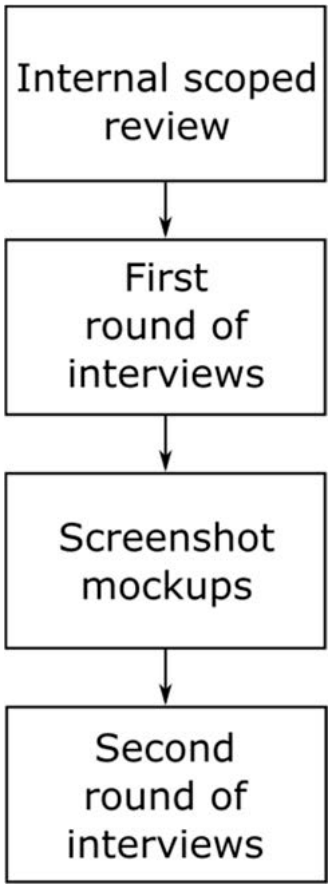


Figure 1. Outline of the process of requirements analysis.

Based on almost all assessments from the first round of interviews, for the second round, we created screenshot mockups for better understanding and visualization of possible implementations in cBioPortal.

2.1. Details about Scoping Review

We conducted a review of literature for potential MTB tools following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines [14,15] as far as appropriate for the requirements analysis. Therefore, we searched MEDLINE and Web of Science (all databases) for articles focusing on MTBs or equivalent clinical decision-making structures published between 1997 and 2017. We captured several features of different potential MTB tools which are either described in the literature or used by physicians at our MIRACUM sites. We used these findings to prepare for the first round of interviews. A detailed description of the methods we used can be found in Appendix 1.

2.2. First and Second Round of Interviews

Based on this prior knowledge, we conducted one group interview per partner site and per round from a constructivist point of view. We hypothesized that each site would have different views and visions on requirements for supporting a local MTB. Therefore, we took all suggestions regarding the demands of the participants into account.

The interviews were conducted in local focus groups, in which the interviewers served as moderators. This approach allowed discussions between the participants and thus as many requirements and their potential variants as possible could be identified. At each site, all participants

were interviewed together in one session. The interviewers had an interdisciplinary background: medicine (Philipp Buechner, Melanie Boerries), medical informatics (Jan Christoph, Marc Hinderer), bioinformatics (Jan Christoph, Melanie Boerries) and biology (Melanie Boerries).

All interviews, except the one at the first author's local university hospital, took place as web conferences with transmission of voices and screen contents. In addition to the option of easily recording the session, the major argument for this setting was that the participants at the various university hospitals were spread all over Germany.

All interview participants were members of the MIRACUM Use Case 3 and thus known to us. The members responsible for the use case at each site arranged an appointment and invited additional local experts, who all were also members of the use case. All participants had to have experiences with the processes related to an MTB in order to join the interviews.

2.2.1. Structure and Purpose of the First Round of Interviews

PB, MH and JC conducted the interviews of the first round together as members of MIRACUM's Use Case 3 between June 2018 and August 2018. It comprised a short guideline with questions (see Appendix 2) we developed from the results of the scoped review. We also demonstrated the main functionalities (see Appendices 3 and 4) of the following potential MTB tools:

- cBioPortal [9],
- OncoKB [16],
- Sophia Genetics [17],
- Clarivate "Key Pathway Advisor" [18],
- Clarivate "MetaCore" [18] and
- "CIViC" (Clinical Interpretations of Variants in Cancer) [19].

We collected all mentioned requirements cBioPortal must meet (including details about potential options for implementation) that were mentioned during the meetings. The interview process was developed iteratively and the information gained was immediately incorporated into the subsequent interviews with other partner sites during this first round. We used the web conference system "Adobe Connect" [20] to conduct, record and subsequently analyze these interview sessions.

2.2.2. Structure and Purpose of the Second Round of Interviews

The second round of interviews was performed by Philipp Buechner, Melanie Boerries and Jan Christoph between November 2018 and December 2018. In order to make optimal use of the limited time during the interviews, Philipp Buechner developed a comprehensive interview guideline describing the requirements and their potential options identified in the first round with text and screenshot mockups (see Appendix 5). However, this round did not cover the requirements mentioned in the first round of interviews, that are already implemented in cBioPortal or are generally out of the scope of MIRACUM Use Case 3. To familiarize the participants with the requirements, this guideline—once it was finally validated by Melanie Boerries, Jan Christoph and Philipp Unberath —was handed out to them prior the meetings.

Since some requirements had more than one potential option regarding implementation and visualization, sites were asked to select one during this round of interviews. In case they had different opinions, they were encouraged to find a compromise.

For the final software specification—after all interviews have been conducted and analyzed - we grouped related features into larger meta-categories to account for individual requirements and yet to keep the assessment structured. For example, the term "sample metadata", comprises six (individual) data features:

- Localization and time of the sampling;
- Type of sampling (e.g., fine-needle aspiration biopsy);
- Distinction between fresh-frozen and formalin-fixed paraffin-embedded samples;
- Scope of sequencing (e.g., gene panel or whole-exome sequencing);
- Name and version of both the used panel and kit;

- Hyperlink to the corresponding product-specific website of the manufacturer.

When calculating the total number of requirements identified by us, we only counted those combined meta categories. Therefore, the above-mentioned example of "sample metadata" counts as one requirement instead of six individual ones.

We used the web conference system "Zoom" [21] to conduct, record and subsequently analyze these interview sessions.

2.3. Low-Fidelity Mockup Demonstrator

We created 54 descriptive screenshot mockups for almost all options of the identified requirements from the first round of interviews using the image-editing tool GNU Image Manipulation Program (GIMP), version 2.10.8 [22]. These low-fidelity mockups are based on full-screen screenshots of the cBioPortal graphical user interface and have been manipulated to give the realistic appearance of providing the respective functions. To quickly direct the viewer's focus to the part of the image that represents the demanded function we indirectly highlighted the area by darkening the rest of the image with a black overlay (opacity: 20%).

2.4. Consultation with Main Developers of MSKCC

After all interviews have been conducted, we discussed the requirements with the main developers of cBioPortal from the Memorial Sloan Kettering Cancer Center (MSKCC) in New York, USA, in an online audio conference. Prior to that, we had detailed the most important and far-reaching changes in a letter including excerpts from our mockups.

The aim of this was to increase the chances of merging our planned implementations into the main development branch of cBioPortal and to maintain contact with the main developers right from the beginning.

2.5. Ethical Approval

This study was ethically approved by the ethics committee of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) (see Appendix 6).

3. Results

3.1. Overview of Scoped Review

Based on our keyword search we selected 306 unique articles out of which 27 dealt with MTBs and were kept for further review. Next, two papers were discarded because their full texts were not available. From the remaining 25 publications, we excluded another 13 since they did not describe IT support in MTBs, which resulted in a total of twelve articles for our review. For details see Appendix 1.

3.2. Details about Interviewees

We conducted the first round of interviews with a total of 18 participants at nine different university hospitals to determine all requirements for an MTB software tool based on cBioPortal. Up to four participants were interviewed simultaneously at each site. Representatives of the following disciplines were involved: oncology (10), pathology (4), systems medicine/systems biology (2), bioinformatics (1) and human genetics (1).

The second round of interviews discussed and evaluated the requirements identified in the first round in detail with a total of 16 participants at the same university hospitals as above. Up to three participants were interviewed simultaneously at each site. Representatives of the following disciplines were present: oncology (7), pathology (4), systems medicine/systems biology (2), bioinformatics (1), human genetics (1) and urology (1).

3.3. Requirements from the First Round of Interviews and Screenshot Mockups

During the first round of interviews, a total of 49 requirements with up to seven potential options for implementation each were surveyed (see Appendix 7). Ten requirements either already implemented in cBioPortal or out of scope of our Use Case were dropped prior to the second round of interviews. This included features to:

- Highlight mutations with existing treatment options;
- Display information about general availability of a specific drug in Germany;
- Point out mutations causing treatment resistance;
- Mark germline mutations;
- Display variant allele frequencies alongside corresponding coverage;
- Integrate the database "Clinical Interpretations of Variants in Cancer" (CIViC) [19];
- Visualize mRNA expression data.

A platform to discuss individual mutations across hospitals, for example in the context of a forum, is outside the scope of MIRACUM's Use Case 3 and was, therefore, not included in the second round of interviews either.

The choice of the file format to be used for the import of the mutation data into cBioPortal was made independently of the interviews by all use case members (MAF: "mutation annotation format"). In addition, a feature to permanently hide mutations in a specific sample in cBioPortal was denied since this filtering should be done by MIRACUM-Pipe [7] only.

In preparation for the second round of interviews, we created a total of 54 screenshot mockups demonstrating almost all surveyed requirements and their respective options (see Appendix 8).

3.4. Consolidated Requirements from the Second Round of Interviews

Below we provide a rough overview of the consolidated requirements we surveyed during the second round of interviews. For a list and detailed description of all final requirements and their respective options see Appendix 9 (All Appendix files can be found in the Supplementary files).

3.4.1. Improving Patient Case Analysis

Since case analysis in personalized medicine relies on various information such as details about the patient or—in case of MTBs—the underlying tumor, cBioPortal should provide those by integrating various information and knowledge sources in a single tool. Therefore, the participants requested, amongst others, clinical patient data to be stored in cBioPortal. Displaying sample metadata and their subsequent analysis was also requested. This includes, for example, the type and location of a biopsy as well as the exact specifications of the sequencer used.

Furthermore, cBioPortal seems to lack important (calculated) values for its usage in molecular tumor boards. Ranging from the tumor mutational burden (see Figure 2A) up to values that indicate the pathogenicity of individual mutations (see Figure 2C). Mutations are automatically annotated with the latter by the MIRACUM-Pipe [7] and this information should be displayed in cBioPortal. We also discovered potential improvements for already existing features in cBioPortal, like the visualization of copy number variations (see Figure 2E).

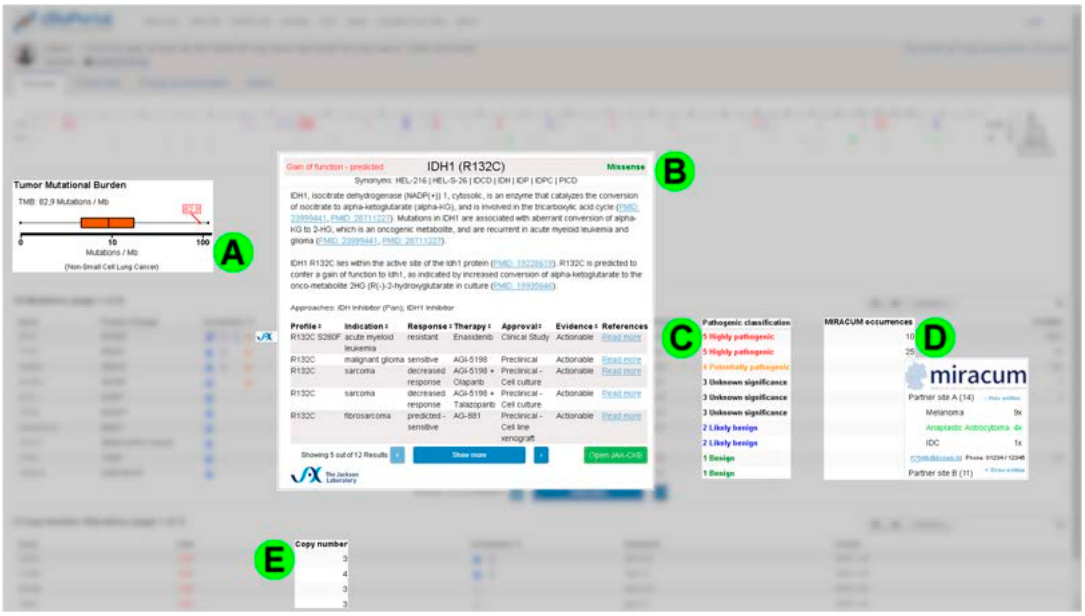


Figure 2. Collage demonstrating some requirements for the case analysis. This collage is a synthesis of different screenshot mockups, which exemplarily illustrates some requirements for individual case analysis: (**A**) A visual representation of the tumor mutational burden relative to the expected range of the corresponding tumor entity using a box plot. (**B**) Integration of data from the JAX Clinical Knowledgebase (JAX-CKB) via annotation and pop-up. (**C**) Classification of pathogenicity as an example of data extension in the mutation table of the patient view. (**D**) Tool for an entity-specific display of how often a certain mutation was found and classified as relevant for the therapy recommendation at the partner sites. If necessary, the contact data may be used to exchange experiences. (**E**) Extension of the table with copy number variation (CNV) data by the exact number of copies. Example data adopted from the public cBioPortal (https://cbioportal.org) and JAX-CKB knowledgebase [23].

Since the interviewees also use numerous different databases when evaluating a patient case, the integration of additional knowledgebases is highly desired. In this context, the JAX Clinical Knowledgebase [23] was mentioned explicitly and considered to be beneficial when integrated (see Figure 2B).

Besides that, the visualization of molecular pathways can be an important tool that links individual mutations to molecular function and pathway in the search for a therapy option.

In order to improve cooperation, there were requests for a central service to (automatically) report a mutation that has led to a therapy recommendation before. If other participating hospitals similarly identify this mutation in one of their samples, it should be highlighted and contact details should be displayed (see Figure 2D) for detailed, personal exchange of expertise.

3.4.2. Supporting the Development and Recording of a Therapy Recommendation

Once the patient's data has been reviewed, the members of the molecular tumor board determine—if possible—the potentially relevant mutations for a therapy recommendation based on their previous analysis. This selection should be recorded in cBioPortal (see Figure 3A) and serve as the basis of the following features.

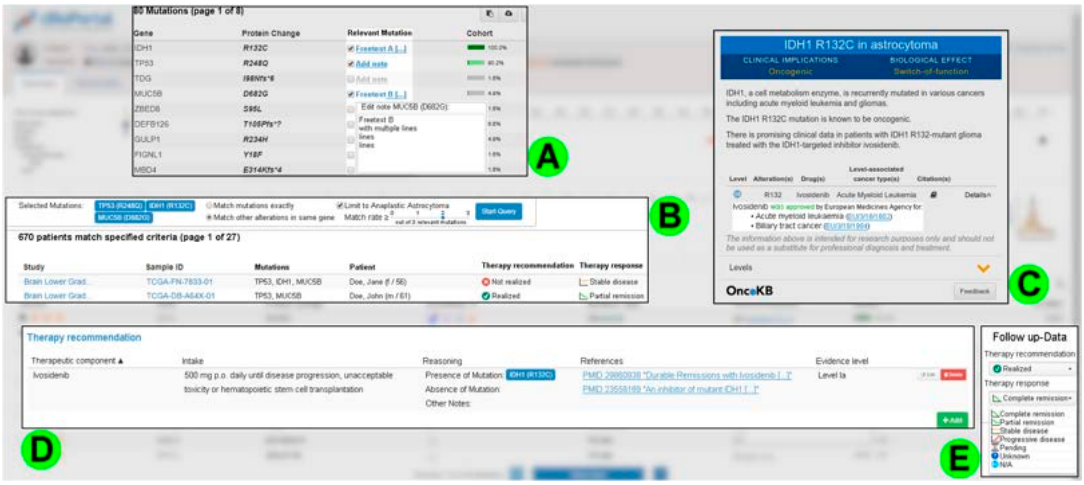


Figure 3. Collage demonstrating some requirements for the development and recording of a therapy recommendation. This collage depicts image sections from the screenshot mockups we created based on the original interface of cBioPortal: (**A**) Checkboxes and text fields in the mutation table of the patient view to mark potentially relevant mutations for the therapy recommendation. (**B**) Search functionality with automatic parameter transfer to find previous patient cases with a mutation pattern similar to that of the current patient. (**C**) Extension of OncoKB's information to cover the European Medicines Agency's (EMA) approval status of a given drug. (**D**) Summary of already entered components of the therapy recommendation for the current patient. (**E**) Option to record follow-up data for the current patient. Example data adopted from the public cBioPortal (https://cbioportal.org) and OncoKB [16].

This set of relevant mutations is used to search for similar patient cases that have been analyzed in the local hospital before. A search should be comprehensively parameterizable and values from the current patient case (e.g., tumor entity, relevant mutations, etc.) should be automatically applied (see Figure 3B). The interviewees considered the gathering of information related to therapy recommendations for previous patients as the main goal of this functionality. This requirement was complemented by the need to document follow-up data (see Figure 3E). Therefore, details on the progression status of similar patient cases can be reviewed and included in the evaluation of the current case more easily.

Building on this, further information on a possible therapy approach is required. In addition to the already available functions in cBioPortal provided by OncoKB [16], the interviewees requested a way to easily query the approval status of a drug in their respective country (in the case of MIRACUM: Germany) (see Figure 3C).

Since, according to the participants, some therapy components (e.g., a drug) are only available in the context of (pre-) clinical studies, integration of clinical trial databases such as ClinicalTrials.gov [24] were requested. The interviewees expect a more efficient search for suitable studies through the automatic transfer of relevant search parameters, which are taken from the cBioPortal data record.

Once a patient case has been prepared based on its individual data, it will be presented and discussed during a meeting of the MTB in order to jointly develop a therapy recommendation. There was no consensus as to what extent cBioPortal must support such a presentation. Some considered an automatically generated set of slides with individualized content as helpful. Others preferred to use the cBioPortal graphical user interface itself during the presentation with an option to hide irrelevant content. However, some interviewees also considered such assistance to be completely unnecessary.

After a therapy recommendation has been decided upon within the MTB, it must be recorded in detail in cBioPortal. Besides information on the therapy itself (e.g., the name of a drug), the molecular and clinical rationale for the recommendation needs to be documented, too. As a justification, the mutations earlier classified as relevant or the tumor mutational burden (TMB) may be referenced.

To submit the results of the MTB to the client (e.g., the treating physician), the interviewees requested a function to generate a PDF report. Besides the actual therapy recommendation, this report should also contain extracts from the consulted databases and thus also be used to archive the current state of knowledge that led to the decisions made.

3.4.3. Requirements for IT Infrastructure

In order to integrate cBioPortal in the various clinical system landscapes, a standardized application programming interface (API) like FHIR (Fast Healthcare Interoperability Resources [25]) should be used by the respective components of the hospital information system to feed a local cBioPortal instance with various and comprehensive data (e.g., clinical data). In addition, the information, which is created or altered within cBioPortal, should be accessible via this interface for export into the local hospital information system.

The existing user system in cBioPortal must be extended by a comprehensive and flexible user and rights management. Of course, only authenticated and authorized users should be permitted to use the cBioPortal instance for an MTB at all. Furthermore, it should be possible to "freeze" a therapy recommendation made by the molecular tumor board and to allow subsequent changes only for certain users and only in a reproducible way. This implies that all alterations to the data records in cBioPortal can be traced, thus guaranteeing their integrity and authenticity.

In addition, some sites indicated that patient data may be stored in a pseudonymized manner if this becomes necessary for legal reasons (like it is done in the public available cBioPortal instance hosted by the Memorial Sloan Kettering Cancer Center). Of course, in this case, it must be guaranteed that the data can be reassigned to the patient at the end. For this purpose, an identification code assigned to the patient by the hospital information system could be used for pseudonymization.

4. Discussion

Due to the huge amount of data produced by NGS technologies, the growing number of clinical studies and released targeted therapies to address treatment of cancer, new tools are required to identify relevant molecular alterations and matching therapy options for individual patients. The present paper outlines which of those are essential and which functionalities exactly have to be provided to support processes in MTBs. To our knowledge, there is no comparable work so far tackling this issue from the point of view of participants in MTBs.

4.1. Results and Future Work

We worked out a requirements specification for the software supporting the processes in an MTB based on the open source project cBioPortal, an integrated database, which allows to visualize clinical parameters and molecular findings of individual patients in the context of current knowledge about cancer diseases. It turned out that the screenshot mockups we had created prior to the second round of interviews played an important role in the further requirements analysis. They allowed us to describe the different requirements and their potential options identified in the first round of interviews briefly and succinctly and thus to quickly start a discussion in the second round. They may also serve as blueprints for the future implementation of the requirements as it was the case with the integration of the Genome Aggregation Database (gnomAD [26]), where our mockups served as inspiration for the core cBioPortal development [27].

This conversation with the participants turned out to be the most important part of this work at all, as on the one hand questions that occurred could be resolved immediately and on the other hand, requirements that had not yet been considered by neither the interviewees nor us came to light by intensive discussions. Therefore, we could finally identify 24 requirements that were not yet integrated into cBioPortal at the time of the survey. Some of them have been proposed to the community of cBioPortal with requests for comments (RFC) filed by other users or even have been implemented before the finalization of this paper [27–34]. RFCs are publicly available documents everyone can create with proposals for new features in cBioPortal that often also include descriptions

of a potential way to implement these. They are an important step in the development of new functionalities in cBioPortal and help to involve the community in the process of implementation. This demonstrates, how important those requirements—even outside of an MTB context—seem to be for the future of cBioPortal.

Besides that, some of our requirements identified are also met by tools not directly associated with cBioPortal, like MatchMiner [35] published by the Dana–Farber Cancer Institute. This tool, which has recently been prototypically integrated in cBioPortal [36], offers a service to match clinical trials to a patient case. At the moment, all studies have to be cured manually, so to use this feature to prepare MTBs in Germany, the ongoing challenge of an automatic import from databases like ClinicalTrials.gov [24] still remains to be resolved.

For the future we plan to integrate cBioPortal in the hospital information systems (HIS) of our partner sites. This means the cBioPortal instance imports clinical data related to patients that have been registered for the MTB directly from the corresponding electronic health record (EHR). This data comprises at least age, gender and tumor entity and ideally also recent diagnoses and therapy attempts as well as data from cancer registries. The sequencing data is automatically processed by the MIRACUM-Pipe [7] under parameter control, including alignment, variant calling, annotation and analyses, in order to finally be passed directly to the cBioPortal instance.

In addition to the import of existing data into cBioPortal, users also need to add persistent information to patient cases in cBioPortal to mark relevant mutations in a patient's sample or to finally document the therapy recommendation. Since we are in ongoing contact with the main developers of cBioPortal, who finally determine which new features are merged into their project and provide future maintenance of them, we identified a considerable conflict with this requirement: cBioPortal's primary focus is to support research in form of a read-only data warehouse. Therefore, the main developers stressed, that having direct write access by the user is currently not intended. A solution for this problem would be to place a hyperlink in cBioPortal to an input mask not hosted in cBioPortal itself, where the user can enter the data (e.g., the therapy recommendation). This form forwards the data to the patient's EHR in the HIS via a standardized application programming interface (API) like FHIR [25], which is supported by most modern systems of an EHR (e.g., MEONA [37]). In order to make this data available again in cBioPortal, an (automatic) import must take place on a regular basis (e.g., twice daily). Thus, the data warehouse concept of cBioPortal would remain and no user write access in cBioPortal itself is necessary.

We also discussed another problem we came across: the rigid process to import mutation data into cBioPortal. At the moment, there is no way to (dynamically) import additional annotation data on individual mutations without fundamental changes to the backend, which would be necessary for the integration of the demanded scores and similar. The team in Use Case 3 of the MIRACUM consortium responsible for the implementation already took the first steps and submitted a request for comments to the community to address this problem. This request deals with a flexible integration of further mutation data by means of an additional database column with the data stored in JavaScript object notation (JSON) format [38].

In general, it is necessary to store data in a structured manner rather than in free text in order to achieve a high grade of automation during import. For example, the International Classification of Diseases for Oncology (ICD-O) can be used to encode the tumor entity, the German Federal Ministry of Health has stated in its announcement [39]. The use of a standardized ontology, such as that provided by OncoTree [40], can also be useful.

However, not only technical hurdles have to be tackled, but also legal ones. For example, when integrating further databases, particular attention must be paid to licensing as some restrict use to research purposes only. For example, even though the JAX Clinical Knowledgebase (JAX-CKB) was developed "to support clinical decision-making" [23], in the disclaimer of their website they allow usage "only for research and educational purposes" [14]. Use Case 3 of the MIRACUM consortium currently uses cBioPortal in clinical research. However, if it is used outside of a research context in the future, of course, this and also existing laws and regulations must be taken into account during

the development, as well. Besides compliance with data protection regulations [15,17] and the Medical Devices Act [18], this is a very broad field.

As for future works, the implementation of the collected requirements must address these problems and find viable ways in close coordination with the main developers and the community around cBioPortal. This is the only way to integrate the features into the project permanently and to ensure their further maintenance and support.

4.2. Related Work

We came across two related works taking place in Germany. Halfmann et al. report that they developed a tool that aims to support the preparation process of a molecular tumor board as well as the presentation of a patient case during a meeting [12]. A video published by them demonstrates how different tools, including cBioPortal, can be combined in a single user interface [20]. Among other requirements, they discovered, like we did, that a function "to search for comparable local cases" [12] is demanded by clinical experts.

Fegeler et al. also describe a software solution for the support of molecular tumor boards. In addition to the option of planning and managing the processes in an MTB, they also describe an integrated video conferencing system. They plan to integrate cBioPortal and knowledge databases, to support the development of a therapy recommendation [13].

4.3. Limitations

Since we collected the requirements based on the expertise of clinicians who could only spend a certain amount of time for the survey and interviews (timeframe ranged between 40 min and two hours per interview and per round), we concentrated on the requirements which had not yet been implemented in cBioPortal. Therefore, we cannot make any statements about already existing features of the software that are useful for an MTB. As we iteratively improved and extended the interview guideline in the first round, the second round of interviews for consolidation had to be conducted so that in the end all sites had the opportunity to comment on every single requirement.

We also limited the number of sites interviewed to nine university hospitals spread all over Germany. Furthermore, not all disciplines and their representatives involved in a molecular tumor board were interviewed. Although we tried to incorporate as many disciplines as possible, this may not be a representative sample for Germany.

As a by-product of our requirements analysis and in preparation for the interviews, we performed a scoped review to provide an overview of already existing tools and systems that support molecular tumor boards. A limitation of it is its methodological rigor as compared to a full systematic review. We limited our database search to the MEDLINE and Web of Science databases. Therefore, we might have neglected relevant articles neither listed in MEDLINE nor in Web of Science. In addition, only one author performed the review, so this might also reduce the quality of results since misinterpretations cannot be systematically excluded. However, we believe that for our purposes we achieved a high degree of methodological quality throughout this scoping review by following the PRISMA statement [21,22] as far as appropriate for the requirements analysis.

5. Conclusions

By interviewing experts at our partner sites, we identified and consolidated for the first time a list of requirements for IT-supported preparation of molecular tumor boards based on cBioPortal. This list comprises a total of 24 requirements that had not yet been implemented during the time of the interviews. For almost all of them and their subordinated features, we have created descriptive screenshot mockups (54 in total) which supported the interview process and may contribute to the further development of cBioPortal. This work provides important information based on the clinical needs that will ultimately support the members of an MTB interpret the complex data for a personalized therapy recommendation.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/xxx/s1.

Author Contributions: Conceptualization, P.B., M.H., P.U., M.B. (Melanie Boerries) and J.C.; data curation, P.B. and M.H.; formal analysis, P.B. and M.H.; investigation, P.B., M.H., M.B. (Melanie Boerries) and J.C.; methodology, P.B., M.H. and J.C.; project administration, P.B., M.H. and J.C.; supervision, M.B. (Melanie Boerries) and J.C.; Validation, P.B., M.H., P.U., M.B. (Melanie Boerries) and J.C.; visualization, P.B.; writing—original draft, P.B.; writing—review and editing, P.B., M.H., P.U., P.M., M.B. (Martin Boeker), T.A., F.H., E.M., D.N., C.P., D.S., N.v.B., S.W., H.B., M.B. (Melanie Boerries) and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by MIRACUM that is funded by the German Federal Ministry of Education and Research (BMBF) within the "Medical Informatics Funding Scheme" (FKZ 01ZZ1801A and FKZ 01ZZ1801B).

Acknowledgments: We especially thank all interview participants for their willingness and time to conduct the interviews. In addition, we would like to thank Nikolaus Schultz and Jianjiong Gao from the Memorial Sloan Kettering Cancer Center (MSKCC), who provided us with their feedback on the identified requirements and our planned changes to cBioPortal. The present work was performed in (partial) fulfillment of the requirements for obtaining the degree "Dr. med." from the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) (Philipp Buechner).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

API	Application programming interface
cBioPortal	cBio Cancer Genomics Portal
CIViC	Clinical Interpretation of Variants in Cancer (knowledgebase)
CNV	Copy Number Variation
EHR	Electronic Health Record
EMA	European Medicines Agency
FHIR	Fast Healthcare Interoperability Resources
GIMP	GNU Image Manipulation Program
gnomAD	Genome Aggregation Database
HIS	Hospital Information System
IT	Information technology
JAX-CKB	Jackson Laboratory Clinical Knowledgebase
JSON	JavaScript Object Notation
MIRACUM	Medical Informatics in Research and Care in University Medicine
MSKCC	Memorial Sloan Kettering Cancer Center
MTB	Molecular Tumor Board
NGS	Next-generation sequencing
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RFC	Request for Comments
TMB	Tumor Mutational Burden

References

1. Parker, B.A.; Schwaederlé, M.; Scur, M.D.; Boles, S.G.; Helsten, T.; Subramanian, R.; Schwab, R.B.; Kurzrock, R. Breast Cancer Experience of the Molecular Tumor Board at the University of California, San Diego Moores Cancer Center. *J. Oncol. Pract.* **2015**, *11*, 442–449, doi:10.1200/JOP.2015.004127.
2. Bryce, A.H.; Egan, J.B.; Borad, M.J.; Stewart, A.K.; Nowakowski, G.S.; Chanan-Khan, A.; Patnaik, M.M.; Ansell, S.M.; Banck, M.S.; Robinson, S.I.; et al. Experience with precision genomics and tumor board, indicates frequent target identification, but barriers to delivery. *Oncotarget* **2017**, *8*, 27145–27154, doi:10.18632/oncotarget.16057.
3. Hoefflin, R.; Geißler, A.-L.; Fritsch, R.; Claus, R.; Wehrle, J.; Metzger, P.; Reiser, M.; Mehmed, L.; Fauth, L.; Heiland, D.H.; et al. Personalized Clinical Decision Making Through Implementation of a Molecular Tumor Board: A German Single-Center Experience. *JCO Precision Oncology* **2018**, 1–16, doi:10.1200/PO.18.00105.

4. Perera-Bel, J.; Hutter, B.; Heining, C.; Bleckmann, A.; Fröhlich, M.; Fröhling, S.; Glimm, H.; Brors, B.; Beißbarth, T. From somatic variants towards precision oncology: Evidence-driven reporting of treatment options in molecular tumor boards. *Genome Med.* **2018**, *10*, 18:1-18:15, doi:10.1186/s13073-018-0529-2.
5. Hinderer, M.; Boerries, M.; Haller, F.; Wagner, S.; Sollfrank, S.; Acker, T.; Prokosch, H.-U.; Christoph, J. Supporting Molecular Tumor Boards in Molecular-Guided Decision-Making - The Current Status of Five German University Hospitals. *Stud. Health Technol. Inform.* **2017**, *236*, 48–54, doi:10.3233/978-1-61499-759-7-48.
6. Prokosch, H.-U.; Acker, T.; Bernarding, J.; Binder, H.; Boeker, M.; Boerries, M.; Daumke, P.; Ganslandt, T.; Hesser, J.; Höning, G.; et al. MIRACUM: Medical Informatics in Research and Care in University Medicine. *Methods Inf. Med.* **2018**, *57*, e82-e91, doi:10.3414/ME17-02-0025.
7. Metzger, P.; et al. MIRACUM-Pipe (AG-Boerries/MIRACUM-Pipe repository). Available online: https://github.com/AG-Boerries/MIRACUM-Pipe (accessed on 11 August 2019).
8. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **2013**, *6*, *doi:*10.1126/scisignal.2004088.
9. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2012**, *2*, 401–404, doi:10.1158/2159-8290.CD-12-0095.
10. Shneiderman, B. Inventing Discovery Tools: Combining Information Visualization with Data Mining. In *Discovery science*: *4th international conference, DS 2001, Washington, DC, USA, November 25 - 28, 2001 ; proceedings*; Jantke, K.P., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2001; 17–28, ISBN 978-3-540-42956-2.
11. Omta, W.A.; Nobel, J. de; Klumperman, J.; Egan, D.A.; Spruit, M.R.; Brinkhuis, M.J.S. Improving Comprehension Efficiency of High Content Screening Data Through Interactive Visualizations. *Assay Drug Dev. Technol.* **2017**, *15*, 247–256, doi:10.1089/adt.2017.794.
12. Halfmann, M.; Stenzhorn, H.; Gerjets, P.; Kohlbacher, O.; Oestermeier, U. User-Driven Development of a Novel Molecular Tumor Board Support Tool. In *Data Integration in the Life Sciences*; Auer, S., Vidal, M.-E., Eds.; Springer International Publishing: Cham, 2019; 195–199, ISBN 978-3-030-06015-2.
13. Fegeler, C.; Zsebedits, D.; Bochum, S.; Finkeisen, D.; Martens, U.M. Implementierung eines IT-gestützten molekularen Tumorboards in der Regelversorgung. *FORUM* **2018**, *5*, 322–328, doi:10.1007/s12312-018-0459-3.
14. The Jackson Laboratory. JAX Clinical Knowledgebase - Disclaimer. Available online: https://ckb.jax.org/about/disclaimer (accessed on 18 August 2019).
15. German Federal Ministry of Justice and Consumer Protection. German Federal Data Protection Act ("Bundesdatenschutzgesetz / BDSG"). Available online: https://www.gesetze-im-internet.de/englisch_bdsg/englisch_bdsg.pdf (accessed on 4 April 2019).
16. Chakravarty, D.; Gao, J.; Phillips, S.M.; Kundra, R.; Zhang, H.; Wang, J.; Rudolph, J.E.; Yaeger, R.; Soumerai, T.; Nissan, M.H.; et al. OncoKB: A Precision Oncology Knowledge Base. *JCO Precision Oncology* **2017**, *2017*, *doi:*10.1200/PO.17.00011.
17. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - of 27 April 2016 - on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). GDPR, 2016.
18. German Federal Ministry of Justice and Consumer Protection. German Medical Devices Act ("Gesetz über Medizinprodukte - MPG"). Available online: https://www.gesetze-im-internet.de/mpg/ (accessed on 18 August 2019).
19. Griffith, M.; Spies, N.C.; Krysiak, K.; McMichael, J.F.; Coffman, A.C.; Danos, A.M.; Ainscough, B.J.; Ramirez, C.A.; Rieke, D.T.; Kujan, L.; et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat. Genet.* **2017**, *49*, 170–174, doi:10.1038/ng.3774.
20. YouTube-User PersOnS. Interface Prototype Demo - YouTube. Available online: https://www.youtube.com/watch?v=VXD3Rap11zg (accessed on 19 August 2019).
21. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097:1-e1000097:6, doi:10.1371/journal.pmed.1000097.

22. Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis, J.P.A.; Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* **2009**, *6*, e1000100:1-e1000100:28, doi:10.1371/journal.pmed.1000100.
23. Patterson, S.E.; Liu, R.; Statz, C.M.; Durkin, D.; Lakshminarayana, A.; Mockus, S.M. The clinical trial landscape in oncology and connectivity of somatic mutational profiles to targeted therapies. *Hum. Genomics* **2016**, *10*, 4:1–4:13, doi:10.1186/s40246-016-0061-7.
24. U.S. National Library of Medicine. ClinicalTrials.gov. Available online: https://clinicaltrials.gov/ (accessed on 10 August 2019).
25. HL7. Welcome to FHIR. Available online: https://www.hl7.org/fhir/ (accessed on 30 August 2019).
26. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Alföldi, J.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, D.P.; et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes **2019, doi:**10.1101/531210.
27. GitHub user leexgh. Pull request: add gnomad column #2064. Available online: https://github.com/cBioPortal/cbioportal-frontend/pull/2064 (accessed on 15 August 2019).
28. GitHub user leexgh. Pull request: add dbsnp column to mutation table #2502. Available online: https://github.com/cBioPortal/cbioportal-frontend/pull/2502 (accessed on 15 August 2019).
29. GitHub user leexgh. Pull request: add clinvar to patient page #2596. Available online: https://github.com/cBioPortal/cbioportal-frontend/pull/2596 (accessed on 15 August 2019).
30. GitHub user jjgao. Issue: Adding clinical trials matching in patient view #6444. Available online: https://github.com/cBioPortal/cbioportal/issues/6444 (accessed on 15 August 2019).
31. GitHub user pvannierop. Pull request: Integration of treatment response data in OncoPrint tab #2053. Available online: https://github.com/cBioPortal/cbioportal-frontend/pull/2053 (accessed on 15 August 2019).
32. GitHub user pvannierop. Pull request: Integration of treatment response data in PlotsTab (incl. new Waterfall plot) #2055. Available online: https://github.com/cBioPortal/cbioportal-frontend/pull/2055 (accessed on 15 August 2019).
33. GitHub user jjgao. Issue: PDF of patient view page #6446. Available online: https://github.com/cBioPortal/cbioportal/issues/6446 (accessed on 15 August 2019).
34. Lukasse, P.; van Hagen, S. RFC45: Gene panel information in Patient View. Available online: https://docs.google.com/document/d/1X7dA_wJFtv5xJO1oHCSt8DUdTmk07RexvHUpjCJsSM4/edit (accessed on 15 August 2019).
35. Lindsay, J.; Fitz, C.D.V.; Zwiesler, Z.; Kumari, P.; van der Veen, B.; Monrose, T.; Mazor, T.; Barry, S.; Albayrak, A.; Tung, M.; et al. MatchMiner: An open source computational platform for real-time matching of cancer patients to precision medicine clinical trials using genomic and clinical criteria. Available online: https://www.biorxiv.org/content/10.1101/199489v3 (accessed on 20 January 2020).
36. GitHub user victoria34. Pull request: Matchminer Proxy controller #5679. Available online: https://github.com/cBioPortal/cbioportal/pull/5679 (accessed on 21 January 2019).
37. MEONA GmbH. Meona. Available online: https://www.meona.de/ (accessed on 20 January 2020).
38. Unberath, P. RFC50: Add support for additional arbitrary variant annotation. Available online: https://docs.google.com/document/d/1Pybk4_-lrirKJZ_cH64riZBRdWXdkJnCQqzx1O2fjRo/edit#heading=h.oyj1ec8k7lgx (accessed on 18 August 2019).
39. German Federal Ministry of Health. Updated standardized oncological basic data set of the Consortium of German Tumor Centers e.V. (ADT) and the Society of Epidemiological Cancer Registries in Germany e.V. (GEKID) ("Aktualisierter einheitlicher onkologischer Basisdatensatz der Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID)"). Available online: https://www.bundesanzeiger.de/ (accessed on 9 March 2019).
40. OncoTree. Available online: https://github.com/cBioPortal/oncotree (accessed on 19 March 2019).

4 Discussion

In this thesis, different clinical, biological, bioinforamtics and medical informatics approaches were applied in six research projects to answer scientific questions in the field of systems medicine. The projects covered a broad spectrum from *in vitro* over *in vivo* to retrospective cases studies, including research on various diseases like AML, ccRCC and MDS.

One of the biggest challenges in the field of medical research is the limited treatment options which are effective and well-balanced, especially for elderly patients with AML [Bell et al., 2018]. Hypomethylating agent (HMA) treatments although combined with immunotherapy, chemotherapy or targeted-therapy have to be shown less effective [Bewersdorf et al., 2019] and the over expression of the inhibitor of apoptosis proteins (IAPs) family could be linked to a poor outcome in AML [Fulda, 2012]. In our HMA and IAP antagonist focused AML research ([Dittmann et al., 2019]) we could show that the synergism between the HMA SGI-110 and the IAP ASTX-660 induced both, the extrinsic as well as the intrinsic apoptosis pathways leading to a reliable cell death. Therefore, both pathways are essential and necessary by shifting the interplay of pro- and anti-apoptotic factors towards apoptosis. This is of major importance for developing new treatment combination strategies. The findings are important not only for AML but also for many cancer therapies, as they are often limited by the circumvention of apoptosis [Fulda, 2012].

In the ccRCC study (ccRCC; Höfflin et al.) it became obvious that a simplification or binarization of genes in only two categories, like tumor suppressors and oncogenes, is not sufficient to explain the complex mechanisms of tumor development. Although *HIF-1α* is reported as inhibitor and *HIF-2α* as promoter of aggressive tumor behaviour and a higher *HIF-1α* expression correlates with poor outcome of renal cell carcinoma [Gudas et al., 2014] we could show that *HIF-1α* is necessary for tumor formation but *HIF-2α* has only moderate effects on tumor initiation and growth. We assume that the *HIF-2α* deletion fails to inhibit tumor formation in our *Vhl/Trp53/Rb1* deletion model since the *Rb1* deletion eliminates the negative regulation of E2F transcription factors by cell cycle

promotion. Therefore, an additional cell cycle stimulus by *HIF-2α*, which increases the expression of *Myc* and E2F target genes [Gordan et al., 2008], may no longer be necessary. In summary, both investigated genes show pro tumorigenic activities and don't fit in the categorical thinking of tumor suppressors and oncogenes.

Further, we studied the influence of synonymous or silent mutations in the *GATA2* gene. Mutations in *GATA2* lead to a *GATA2* deficiency. The deficiency is a monogenetic disorder caused by heterozygeous missense mutations, whole gene deletions, or intronic enhancer mutations, which causes haploinsufficiency. In our *GATA2* research ([Kozyra et al., 2020]) we identified silent *GATA2* variants causing RNA deleteriousness and showing the same clinical *GATA2* deficiency phenotype in patients. Silent, intergenic, and intronic variants are usually overlooked and routine diagnostic procedures do not include routine screening for such variants. Therefore, only a few regions are known [Hsu et al., 2013]. We confirmed that silent mutations affect translation and plead for integration of silent *GATA2* mutations in the context of *GATA2* deficiency in routine diagnostic. Furthermore, analysing and understanding of silent, intergenic and intronic variants will improve genetic counseling, donor selection for hematopoietic stem cell transplantation and thus the clinical outcome.

In our retrospective case study over two years and nearly 200 patients of the MTB Freiburg ([Hoefflin et al., 2018]) we showed the positive effects for patients with advanced personalized decision making in accordance with comparable boards [Larkin et al., 2014, Sohal et al., 2016, Johnson et al., 2014, Dalton et al., 2017]. Additionally, we showed an increased response rate to checkpoint inhibitors by combining NGS results, e.g. tumor mutational burden (TMB), with molecular pathological measurements, like *PD-1* and *PD-L1* expression or the microsatellite stability status (MSI/MSS). It is known that a higher TMB can create more neoantigens which is beneficial for the treatment with checkpoint inhibitors [Rizvi et al., 2015, McGranahan et al., 2016]. Due to resource constraints, particularly time and money, a NGS approach is not feasible for all patients [Horak et al., 2017] and therefore selection is made based on clinical background and available treatment options. If extended genetic analyses are carried out only in appropriate cases, processing time and costs could be reduced and the outcome improved. A similar study at the Memorial Sloan Kettering Cancer Center (MSK-CC), which focused on the targeted combination of drugs, concluded with a better outcome for patients if they followed their therapy recommendations [Cheng et al., 2015]. In general, targeted-drug combinations provide better disease control rates than single-agent therapies [Tannock and Hickman, 2016, Yap et al., 2013, Liu et al., 2016a, Turner and

Reis-Filho, 2012], due to cross-talk between signaling pathways [Swanton and Govindan, 2016,Shi et al., 2014] and at the same time without increased toxicity [Soria et al., 2009]. Earlier consultations of such an interdisciplinary tumor board, e.g. after failure of first-line treatment, could prevent ineffective treatments, improve the implementation rate, and increase the likelihood of success of molecular biomarker-matched treatments. And foremost, a MTB is predestined for knowledge generation and evidence generation in oncology through one-person studies instead of the traditional cost- and time-intensive clinical trials.

For a long-term lasting success and the proper integration of high-throughput omics data into clinical routine, integration into the electronic patient record (EHR) and the hospital information system (HIS) is essential. The possible integration interfaces, e.g. with FHIR, are already available for most systems. With our stakeholder analysis ([Buechner et al., 2020]) further requirements for a platform supporting MTBs were identified, including sequencing results together with clinical and molecular-pathological findings, the proper annotation and interpretation of variants and last but not least the meaningful visualization. The platform of choice is cBioPortal as for two other tumor boards [Fegeler et al., 2018] and [Halfmann et al., 2019]. Together with the help of the core developers of cBioPortal, we plan to integrate the requirements not yet available in the tool. Additionally, we have to consider licensing issues, as many databases may only be used in a research environment, but not in a medical environment. Further implications for data security, privacy and ethics have to be be considered.

With the increasing amount of high-throughput omics data and the integration of corresponding molecular pathological and conventional clinical data, a deeper understanding of the underlying processes, not only in the filed of oncology, could be gained. Mutational profiles, expression patterns or methylation motifs of different cancer entities or diseases could help to specify and discover further cancer/disease sub-classes [Shyr and Liu, 2013] and shed light on tissue or disease diversity which, in the case of oncology, leads to tumor heterogeneity and heterogeneous therapy responses. The closer connection between the above-mentioned data sources contribute to improve the three key elements in medicine, namely diagnostics, prognostics, and prediction. Diagnostics will be improved through the establishment of robust disease biomarkers in clinical practice [Goossens et al., 2015] which could be also used to predict prognosis [Ludwig and Weinstein, 2005] and therapy efficiency [La Thangue and Kerr, 2011]. One of the remaining problems is no longer the availability of such data sets but comparability between different laboratories or institutions. Each laboratory has its own standard

of processing or generation, especially high-throughput omics data. The subsequent bioinformatics pipeline adds an additionally layer of complexity and possible sources of error to the process. Therefore, the research area needs standardized processes to harmonize the produced data and to foster data sharing, which we have tackled with the development of the MIRACUM-Pipe. Only by sharing data, the amount of generated information can reach a peak and meaningful knowledge to support clinical decision making is produced. In Germany many research projects aim to solve precisely these issues. One of the projects is MIRACUM, one of the consortia funded by the Medical Informatics Initiative. MIRACUM designed three use cases to address the aforementioned problems. Use Case 1 supports the enrolment of patients into clinical trials, Use Case 2 uses the tremendous amount of clinico-molecular data to build a predictive knowledge tool and Use Case 3 aims to support molecular tumor boards and to establish a generic framework for all necessary steps from the analysis, interpretation, and visualization of multi omics data to the final therapy recommendation [Prokosch et al., 2018]. Furthermore the federal state of Baden-Württemberg is establishing centers for personalized medicine (Zentren für Personalisierte Medizin; ZPM). The aim is to network and to harmonize all medical centers within Baden-Württemberg and to use the generated data for a broader understanding of the diseases. However, the extensive sequencing also poses ethical challenges, in particular with regard to the sequencing of germline material, which might involve predispositions for different diseases. Sharing between different institutions must be carried out in accordance with the applicable data protection laws, which must be taken into account and implemented. Last but not least, the clinicians have to be trained in reading, understanding, and interpreting multi omics analyses.

Bibliography

[Alberts et al., 2014] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., et al. (2014). *Molecular Biology of the Cell*. Norton & Company, New York, 6th revise edition.

[Apweiler et al., 2018] Apweiler, R., Beissbarth, T., Berthold, M. R., Blüthgen, N., Burmeister, Y., et al. (2018). Whither systems medicine? *Experimental and Molecular Medicine*, 50(3):e453–6.

[Ashburner et al., 2000] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., et al. (2000). Gene ontology: Tool for the unification of biology.

[Bell et al., 2018] Bell, J. A., Galaznik, A., Huelin, R., Stokes, M., Guo, Y., et al. (2018). Effectiveness and Safety of Therapeutic Regimens for Elderly Patients With Acute Myeloid Leukemia: A Systematic Literature Review.

[Bewersdorf et al., 2019] Bewersdorf, J. P., Shallis, R., Stahl, M., and Zeidan, A. M. (2019). Epigenetic therapy combinations in acute myeloid leukemia: what are the options? *Therapeutic Advances in Hematology*, 10:204062071881669.

[Blau et al., 2010] Blau, N., van Spronsen, F. J., and Levy, H. L. (2010). Phenylketonuria. *The Lancet*, 376(9750):1417–1427.

[Boeva et al., 2012] Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., et al. (2012). Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*, 28(3):423–425.

[Boeva et al., 2011] Boeva, V., Zinovyev, A., Bleakley, K., Vert, J. P., Janoueix-Lerosey, I., Delattre, O., and Barillot, E. (2011). Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. *Bioinformatics*, 27(2):268–269.

[Bolger et al., 2014] Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120.

[Borghaei et al., 2015] Borghaei, H., Paz-Ares, L., Horn, L., Spigel, D. R., Steins, M., et al. (2015). Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer. *New England Journal of Medicine*, 373(17):1627–1639.

[Bryce et al., 2017] Bryce, A. H., Egan, J. B., Borad, M. J., Keith Stewart, A., Nowakowski, G. S., et al. (2017). Experience with precision genomics and tumor board, indicates frequent target identification, but barriers to delivery. *Oncotarget*, 8(16):27145–27154.

[Buechner et al., 2020] Buechner, P., Hinderer, M., Unberath, P., Metzger, P., Boeker, M., et al. (2020). Requirements analysis and specification for a molecular tumor board platform based on CBIoPortal. *Diagnostics*, 10(2):1–15.

[Carbon et al., 2019] Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N. L., et al. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, 47(D1):D330–D338.

[Chapman et al., 2011] Chapman, P. B., Hauschild, A., Robert, C., Haanen, J. B., Ascierto, P., et al. (2011). Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *New England Journal of Medicine*, 364(26):2507–2516.

[Cheng et al., 2015] Cheng, D. T., Mitchell, T. N., Zehir, A., Shah, R. H., Benayed, R., et al. (2015). Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): A hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *Journal of Molecular Diagnostics*, 17(3):251–264.

[Conesa et al., 2016] Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., and Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1):13.

[Cotto et al., 2017] Cotto, K. C., Wagner, A. H., Feng, Y.-Y., Kiwala, S., Coffman, A. C., et al. (2017). DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Research*, 46(July):1068–1073.

[Craig Venter et al., 2001] Craig Venter, J., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., et al. (2001). The sequence of the human genome. *Science*, 291(5507):1304–1351.

[Dalton et al., 2017] Dalton, W. B., Forde, P. M., Kang, H., Connolly, R. M., Stearns, V., et al. (2017). Personalized Medicine in the Oncology Clinic: Implementation and Outcomes of the Johns Hopkins Molecular Tumor Board. *JCO Precision Oncology*, (1):1–19.

[Dayalu and Albin, 2015] Dayalu, P. and Albin, R. L. (2015). Huntington Disease. *Neurologic Clinics*, 33(1):101–114.

[Dittmann et al., 2019] Dittmann, J., Haydn, T., Metzger, P., Ward, G. A., Boerries, M., et al. (2019). Next-generation hypomethylating agent SGI-110 primes acute myeloid leukemia cells to IAP antagonist by activating extrinsic and intrinsic apoptosis pathways. *Cell Death and Differentiation*.

[Dobin et al., 2013] Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., et al. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21.

[D'Souza et al., 1999] D'Souza, I., Poorkaj, P., Hong, M., Nochlin, D., Lee, V. M., et al. (1999). Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. *Proceedings of the National Academy of Sciences of the United States of America*, 96(10):5598–5603.

[Federoff and Gostin, 2009] Federoff, H. J. and Gostin, L. O. (2009). Evolving from reductionism to holism: Is there a future for systems medicine?

[Fegeler et al., 2018] Fegeler, C., Zsebedits, D., Bochum, S., Finkeisen, D., and Martens, U. M. (2018). Implementierung eines IT-gestützten molekularen Tumorboards in der Regelversorgung. *Forum*, 33(5):322–328.

[Freeman, 2006] Freeman, J. L. (2006). Copy number variation: New insights in genome diversity. *Genome Research*, 16(8):949–961.

[Frew and Moch, 2015] Frew, I. J. and Moch, H. (2015). A Clearer View of the Molecular Complexity of Clear Cell Renal Cell Carcinoma. *Annual Review of Pathology: Mechanisms of Disease*, 10(1):263–289.

[Fulda, 2012] Fulda, S. (2012). Exploiting inhibitor of apoptosis proteins as therapeutic targets in hematological malignancies. *Leukemia*, 26(6):1155–1165.

[Gautier et al., 2004] Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20(3):307–315.

[Gentleman et al., 2004] Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10):R80.

[Gerlinger et al., 2014] Gerlinger, M., Horswell, S., Larkin, J., Rowan, A. J., Salm, M. P., et al. (2014). Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature Genetics*, 46(3):225–233.

[Gerlinger et al., 2012] Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., et al. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New England Journal of Medicine*, 366(10):883–892.

[Goossens et al., 2015] Goossens, N., Nakagawa, S., Sun, X., and Hoshida, Y. (2015). Cancer biomarker discovery and validation.

[Gordan et al., 2008] Gordan, J. D., Lal, P., Dondeti, V. R., Letrero, R., Parekh, K. N., et al. (2008). HIF-α Effects on c-Myc Distinguish Two Subtypes of Sporadic VHL-Deficient Clear Cell Renal Carcinoma. *Cancer Cell*, 14(6):435–446.

[Gudas et al., 2014] Gudas, L. J., Fu, L., Minton, D. R., Mongan, N. P., and Nanus, D. M. (2014). The role of HIF1α in renal cell carcinoma tumorigenesis.

[Hahn et al., 2011] Hahn, C. N., Chong, C.-E., Carmichael, C. L., Wilkins, E. J., Brautigan, P. J., et al. (2011). Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nature Genetics*, 43(10):1012–1017.

[Halfmann et al., 2019] Halfmann, M., Stenzhorn, H., Gerjets, P., Kohlbacher, O., and Oestermeier, U. (2019). User-driven development of a novel molecular tumor board support tool. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11371 LNBI, pages 195–199. Springer Verlag.

[Head et al., 2014] Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., et al. (2014). Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques*, 56(2):61–77.

[Hinderer et al., 2017] Hinderer, M., Boerries, M., Haller, F., Wagner, S., Sollfrank, S., et al. (2017). Supporting molecular tumor boards in molecular-guided decision-making -The current status of five German university hospitals. *Studies in Health Technology and Informatics*, 236:48–54.

[Hirabayashi et al., 2017] Hirabayashi, S., Wlodarski, M. W., Kozyra, E., and Niemeyer, C. M. (2017). Heterogeneity of GATA2-related myeloid neoplasms.

[Hoadley et al., 2014] Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., et al. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, 158(4):929–944.

[Hoefflin et al., 2018] Hoefflin, R., Geißler, A.-L., Fritsch, R., Claus, R., Wehrle, J., et al. (2018). Personalized Clinical Decision Making Through Implementation of a Molecular Tumor Board: A German Single-Center Experience. *JCO Precision Oncology*, (2):1–16.

[Horak et al., 2017] Horak, P., Klink, B., Heining, C., Gröschel, S., Hutter, B., et al. (2017). Precision oncology based on omics data: The NCT Heidelberg experience. *International Journal of Cancer*, 141(5):877–886.

[Hsu et al., 2013] Hsu, A. P., Johnson, K. D., Falcone, E. L., Sanalkumar, R., Sanchez, L., et al. (2013). GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood*, 121(19):3830–3837.

[Huber et al., 2015] Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, 12(2):115–121.

[Hyman et al., 2015] Hyman, D. M., Puzanov, I., Subbiah, V., Faris, J. E., Chau, I., et al. (2015). Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. *New England Journal of Medicine*, 373(8):726–736.

[Irizarry et al., 2012] Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., et al. (2012). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Selected Works of Terry Speed*, pages 601–616.

[Jassal et al., 2020] Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., et al. (2020). The reactome pathway knowledgebase. *Nucleic Acids Research*, 48(D1):D498–D503.

[Johnson et al., 2014] Johnson, D. B., Dahlman, K. H., Knol, J., Gilbert, J., Puzanov, I., et al. (2014). Enabling a Genetically Informed Approach to Cancer Medicine: A Retrospective Evaluation of the Impact of Comprehensive Tumor Profiling Using a Targeted Next-Generation Sequencing Panel. *The Oncologist*, 19(6):616–622.

[Kamburov et al., 2012] Kamburov, A., Galicka, H., Lehrach, H., and Herwig, R. (2012). ConsensusPathDB : assembling a more complete picture of cell biology. (2010):2797.

[Kamburov et al., 2013] Kamburov, A., Stelzl, U., Lehrach, H., and Herwig, R. (2013). The ConsensusPathDB interaction database: 2013 update. *Nucleic acids research*, 41(Database issue):D793–800.

[Karczewski et al., 2020] Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809):434–443.

[Koboldt et al., 2013] Koboldt, D. C., Larson, D. E., and Wilson, R. K. (2013). Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. In *Current Protocols in Bioinformatics*, pages 15.4.1–15.4.17. John Wiley & Sons, Inc., Hoboken, NJ, USA.

[Koboldt et al., 2012] Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., et al. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, 22(3):568–576.

[Kozyra et al., 2020] Kozyra, E. J., Pastor, V. B., Lefkopoulos, S., Sahoo, S. S., Busch, H., et al. (2020). Synonymous GATA2 mutations result in selective loss of mutated RNA and are common in patients with GATA2 deficiency. *Leukemia*, 2(1).

[La Thangue and Kerr, 2011] La Thangue, N. B. and Kerr, D. J. (2011). Predictive biomarkers: A paradigm shift towards personalized cancer medicine. *Nature Reviews Clinical Oncology*, 8(10):587–596.

[Larkin et al., 2014] Larkin, J., Ascierto, P. A., Dréno, B., Atkinson, V., Liszkay, G., et al. (2014). Combined Vemurafenib and Cobimetinib in BRAF -Mutated Melanoma. *New England Journal of Medicine*, 371(20):1867–1876.

[Law et al., 2014] Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, 15(2):R29.

[Lawrence et al., 2013] Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, 499(7457):214–218.

[Lejeune et al., 1963] Lejeune, J., Lafourcade, J., Berger, R., Vialatte, J., Boeswillwald, M., et al. (1963). [3 CASES OF PARTIAL DELETION OF THE SHORT ARM OF A 5 CHROMOSOME]. *Comptes rendus hebdomadaires des seances de l'Academie des sciences*, 257:3098–102.

[Li and Durbin, 2010] Li, H. and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, 26(5):589–595.

[Liberzon et al., 2015] Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., et al. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*, 1(6):417–425.

[Liberzon et al., 2011] Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., et al. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, 27(12):1739–1740.

[Liu et al., 2016a] Liu, S., Nikanjam, M., and Kurzrock, R. (2016a). Dosing de novo combinations of two targeted drugs: Towards a customized precision medicine approach to advanced cancers. *Oncotarget*, 7(10):11310–11320.

[Liu et al., 2016b] Liu, X., Wu, C., Li, C., and Boerwinkle, E. (2016b). dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Human Mutation*, 37(3):235–241.

[Love et al., 2014a] Love, M., Anders, S., and Huber, W. (2014a). *Differential analysis of count data – the DESeq2 package.*

[Love et al., 2014b] Love, M. I., Huber, W., and Anders, S. (2014b). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12):550.

[Ludwig and Weinstein, 2005] Ludwig, J. A. and Weinstein, J. N. (2005). Biomarkers in Cancer Staging, Prognosis and Treatment Selection. *Nature Reviews Cancer*, 5(11):845–856.

[Luo et al., 2009] Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D., and Woolf, P. J. (2009). GAGE: generally applicable gene set enrichment for pathway analysis. *BMC bioinformatics*, 10:161.

[Lupski et al., 1991] Lupski, J. R., de Oca-Luna, R. M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., et al. (1991). DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell*, 66(2):219–232.

[Macaya et al., 2009] Macaya, D., Katsanis, S. H., Hefferon, T. W., Audlin, S., Mendelsohn, N. J., et al. (2009). A synonymous mutation in TCOF1 causes treacher collins syndrome due to mis-splicing of a constitutive exon. *American Journal of Medical Genetics, Part A*, 149(8):1624–1627.

[Malt et al., 2013] Malt, E., Dahl, R. C., Haugsand, T. M., Ulvestad, I. H., Emilsen, N. M., et al. (2013). Health and disease in adults with Down syndrome. *Tidsskrift for den Norske Laegeforening*, 133(3):290–294.

[Maxwell et al., 1999] Maxwell, P. H., Wlesener, M. S., Chang, G. W., Clifford, S. C., Vaux, E. C., et al. (1999). The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature*, 399(6733):271–275.

[McElheny, 2010] McElheny, V. K. (2010). *Drawing the Map of Life*. Basic Book, New York.

[McGranahan et al., 2016] McGranahan, N., Furness, A. J., Rosenthal, R., Ramskov, S., Lyngaa, R., et al. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*, 351(6280):1463–1469.

[Mitelman et al., 2007] Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation.

[Monzon et al., 2011] Monzon, F. A., Alvarez, K., Peterson, L., Truong, L., Amato, R. J., et al. (2011). Chromosome 14q loss defines a molecular subtype of clear-cell renal cell carcinoma associated with poor prognosis. *Modern Pathology*, 24(11):1470–1479.

[Nicorici et al., 2014] Nicorici, D., Satalan, M., Edgren, H., Kangaspeska, S., Murumagi, A., et al. (2014). FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data. Technical report.

[O'Donnell et al., 2017] O'Donnell, M. R., Tallman, M. S., Abboud, C. N., Altman, J. K., Appelbaum, F. R., et al. (2017). Acute myeloid leukemia, version 3.2017: Clinical practice guidelines in oncology.

[Ogata et al., 1999] Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes.

[Parker et al., 2015] Parker, B. A., Schwaederlé, M., Scur, M. D., Boles, S. G., Helsten, T., et al. (2015). Breast Cancer Experience of the Molecular Tumor Board at the University of California, San Diego Moores Cancer Center. *Journal of Oncology Practice*, 11(6):442–449.

[Prokosch et al., 2018] Prokosch, H. U., Acker, T., Bernarding, J., Binder, H., Boeker, M., et al. (2018). MIRACUM: Medical Informatics in Research and Care in University Medicine. *Methods of information in medicine*, 57(S 01):e82–e91.

[R Development Core Team, 2008] R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

[Ritchie et al., 2015] Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, 43(7):e47.

[Rizvi et al., 2015] Rizvi, N. A., Hellmann, M. D., Snyder, A., Kvistborg, P., Makarov, V., et al. (2015). Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*, 348(6230):124–128.

[Sanger and Coulson, 1975] Sanger, F. and Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3):441–448.

[Sanger et al., 1977] Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12):5463–5467.

[Shen et al., 2011] Shen, C., Beroukhim, R., Schumacher, S. E., Zhou, J., Chang, M., et al. (2011). Genetic and Functional Studies Implicate HIF1 as a 14q Kidney Cancer Suppressor Gene. *Cancer Discovery*, 1(3):222–235.

[Shi et al., 2014] Shi, H., Hugo, W., Kong, X., Hong, A., Koya, R. C., et al. (2014). Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. *Cancer Discovery*, 4(1).

[Shyr and Liu, 2013] Shyr, D. and Liu, Q. (2013). Next generation sequencing in cancer research and clinical application. *Biological Procedures Online*, 15(1):4.

[Sims et al., 2014] Sims, D., Sudbery, I., Ilott, N. E., Heger, A., and Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2):121–132.

[Sohal et al., 2016] Sohal, D. P., Rini, B. I., Khorana, A. A., Dreicer, R., Abraham, J., et al. (2016). Prospective Clinical Study of Precision Oncology in Solid Tumors. *Journal of the National Cancer Institute*, 108(3):10–12.

[Soria et al., 2009] Soria, J. C., Massard, C., and Izzedine, H. (2009). From theoretical synergy to clinical supra-additive toxicity. *Journal of Clinical Oncology*, 27(9):1359–1361.

[Subramanian et al., 2005] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550.

[Swanton and Govindan, 2016] Swanton, C. and Govindan, R. (2016). Clinical implications of genomic discoveries in lung cancer.

[Tannock and Hickman, 2016] Tannock, I. F. and Hickman, J. A. (2016). Limits to Personalized Cancer Medicine. *New England Journal of Medicine*, 375(13):1289–1294.

[Turajlic et al., 2018] Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Chambers, T., et al. (2018). Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell*, 173(3):581–594.e12.

[Turner and Reis-Filho, 2012] Turner, N. C. and Reis-Filho, J. S. (2012). Genetic heterogeneity and cancer drug resistance. *The Lancet Oncology*, 13(4):e178–e185.

[Wang et al., 2010] Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, 38(16):e164.

[Wehr et al., 2018] Wehr, C., Grotius, K., Casadei, S., Bleckmann, D., Bode, S. F., et al. (2018). A novel disease-causing synonymous exonic mutation in GATA2 affecting RNA splicing.

[Wlodarski et al., 2017] Wlodarski, M. W., Collin, M., and Horwitz, M. S. (2017). GATA2 deficiency and related myeloid neoplasms. *Seminars in Hematology*, 54(2):81–86.

[Wlodarski et al., 2016] Wlodarski, M. W., Hirabayashi, S., Pastor, V., Starý, J., Hasle, H., et al. (2016). Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. *Blood*, 127(11):1387–1397.

[Wolf et al., 2013] Wolf, A. B., Caselli, R. J., Reiman, E. M., and Valla, J. (2013). APOE and neuroenergetics: an emerging paradigm in Alzheimer's disease. *Neurobiology of Aging*, 34(4):1007–1017.

[Wolfsberg et al., 2001] Wolfsberg, T. G., McEntyre, J., and Schuler, G. D. (2001). Guide to the draft human genome. *Nature*, 409(6822):824–826.

[Yadav et al., 2020] Yadav, A., Vidal, M., and Luck, K. (2020). Precision medicine — networks to the rescue. *Current Opinion in Biotechnology*, 63:177–189.

[Yap et al., 2013] Yap, T. A., Omlin, A., and De Bono, J. S. (2013). Development of therapeutic combinations targeting major cancer signaling pathways. *Journal of Clinical Oncology*, 31(12):1592–1605.