Aberrant transcriptome analysis of 3,760 hematologic malignancies reveals *LRP1B* as hairy cell leukemia variant biomarker

Xueqi Cao[1,2], Sandra Huber[3], Ata Jadid Ahari[1], Franziska R. Traube[1,4], Marc Seifert[5], Christopher C. Oakes[6], Polina Secheyko[1,7], Ines Scheller[1,8], Nils Wagner[1,9], Vicente A. Yépez[1], Piers Blombery[10,11,12], Torsten Haferlach[3], Matthias Heinig[1,8], Leonhard Wachutka[1]*, Stephan Hutter[3]*, Julien Gagneur[1,2,8,13]*

[1] School of Computation, Information and Technology, Technical University of Munich, Garching, Germany

[2] Graduate School of Quantitative Biosciences (QBM), Munich, Germany

[3] Münchner Leukämielabor GmbH (MLL), Munich, Germany

[4] Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Stuttgart, Germany

[5] Department of Haematology, Oncology and Clinical Immunology, University Hospital Düsseldorf, Düsseldorf, Germany

[6] Division of Hematology, Department of Internal Medicine, The Ohio State University, Columbus, Ohio, USA

[7] Faculty of Biology, Ludwig-Maximilians-University Munich, Munich, Germany

[8] Computational Health Center, Helmholtz Center Munich, Neuherberg, Germany

[9] Helmholtz Association - Munich School for Data Science (MUDS), Munich, Germany

[10] Peter MacCallum Cancer Centre, Melbourne, Australia

[11] University of Melbourne, Melbourne, Australia

[12] Torsten Haferlach Leukaemiediagnostik Stiftung, Munich, Germany

[13] Institute of Human Genetics, School of Medicine, Technical University of Munich, Munich, Germany

* Corresponding authors

* Corresponding authors

Julien Gagneur: gagneur@in.tum.de

Stephan Hutter: stephan.hutter@mll.com

Leonhard Wachutka: wachutka@cit.tum.de

Abstract

Rare driver mutations are suspected to substantially contribute to the large heterogeneity of hematologic malignancies, but their identification remains challenging. To address this issue, we generated the largest dataset to date of matched whole genome sequencing and total RNA sequencing of hematologic malignancies from 3,760 patients spanning 24 disease entities. To discover rare and large regulatory aberrations, we analyzed aberrant expression and splicing events using an extension of DROP (Detection of RNA Outliers Pipeline) and AbSplice, an algorithm that identifies genetic variants causing aberrant splicing. We found a median of seven aberrantly expressed genes, two aberrantly spliced genes, and two rare splice-affecting variants per sample. Each category showed significant enrichment for well-characterized driver genes, with odds ratios exceeding three among genes called in more than five samples. We next trained disease-specific driver gene prediction models integrating these data with recurrent mutation analyses. On held-out data, integrative modeling significantly outperformed modeling based solely on genomic data and revealed promising novel candidate driver genes. Moreover, we found a truncated form of the low density lipoprotein receptor *LRP1B* to be aberrantly overexpressed in about half of hairy cell leukemia variant (HCL-V) samples and, to a lesser extent, in closely related B-cell neoplasms. This observation, which was confirmed in an independent cohort, suggests *LRP1B* be a novel biomarker and a yet unreported functional role of *LRP1B* within these rare entities. Altogether, this dataset and the companion computational workflow constitute unique resources to deepen our understanding of rare oncogenic events in hematologic cancers.

Introduction

Hematologic malignancies are characterized by abnormal blood cells in the bone marrow, peripheral blood, or lymphatic organs. In 2020, hematologic malignancies accounted for approximately 2.5% of new cancer cases globally and were responsible for 3.1% of cancer-associated mortality[1]. While some subtypes exhibit a high degree of uniformity in their manifestation and genetic profile, others display a significantly broader spectrum, hampering correct diagnosis and therapy decisions with a negative impact on treatment outcomes and survival. Thus, better understanding the complexity of each disease entity is of utmost interest to refine diagnostics and facilitate the development of new therapeutic options.

Within the last decade, the identification of driver genes in hematologic malignancies has been dramatically enhanced by the utilization of next-generation sequencing. This has provided valuable insights into the underlying genetic landscape of each entity and triggered a revision of the classification systems, which now emphasize genomics-based categorization of various leukemia and lymphoma entities[2–5]. However, despite significant progress in understanding prevalent driver mutations in hematologic malignancies and advancements in targeted treatments, there is still much to learn about the rare events that drive their development and progression[6,7].

Abnormal gene expression and splicing play a key role in cancer[8] and particularly in hematologic malignancy pathogenesis[9–19]. Abnormal gene expression, which could result in deregulated pathways, is suggested to be involved in development of hematologic malignancies. Examples include acute myeloid leukemia (AML) with defining genetic alterations[20], *BCR::ABL1*-positive chronic myeloid leukemia (CML)[21], or *CCND1* rearrangements in mantle cell lymphoma (MCL)[22]. On the other hand, abnormal splicing can generate non-functional or oncogenic transcript isoforms, and their dysregulation could be a potential key driver of hematologic malignancy[8–10,23–25]. Examples include aberrant splicing of *CD13* in AML[25] and *TP53* in CML[26]. However, a systematic analysis of rare aberrant expression and splicing among hematologic malignancies is still lacking.

To address this gap, we here conduct a comprehensive analysis of genomes (Whole Genome Sequencing, WGS) and matched transcriptomes (total RNA sequencing, RNA-Seq) of tumor tissues from 3,760 patients spanning 24 hematologic malignancy entities (Figure 1). We analyze this data using RNA-seq-based aberrant expression and splicing callers, as well as AbSplice, a tool we recently published to predict rare genetic variants causing aberrant splicing. We demonstrate how this resource can be utilized to identify a novel marker for a rare entity and enhance the prediction of hematologic malignancy driver genes beyond the commonly used mutational recurrence, thereby revealing novel candidate genes.

Methods

Further details about the methods are provided in Supplementary Materials and Methods.

Dataset

Patients

We used genomic and transcriptomic data from the Munich Leukemia Laboratory (MLL). We included a total of 3,760 samples sent to the MLL between September 2005 and April 2019 for routine diagnostic workup (Table 1). Diagnoses from peripheral blood or bone marrow were based on cytomorphology, immunophenotype, cytogenetics, and molecular genetics, as previously described[26–28]. All patients or their legal guardians gave written informed consent for genetic analyses and to the use of laboratory results and clinical data for research purposes, according to the Declaration of Helsinki. The study was approved by the MLL's institutional review board.

Table 1. Disease entities enrolled and corresponding study groups

Disease entity	**Abbreviation**	**Number of samples per disease entity**	**Study group**	**Number of samples per study group**
Acute myeloid leukemia	AML	730	AML	730
Myelodysplastic neoplasm	MDS	713	MDS	713
Marginal zone lymphoma	MZL	71	MatureB group	375
Mantle cell lymphoma	MCL	71		
High-grade B-cell lymphoma	HGBL	61		
Lymphoplasmacytic lymphoma	LPL	59		
Follicular lymphoma	FL	57		
Other B-cell non-Hodgkin lymphomas	B-NHL	56		
Myeloproliferative neoplasm	MPN	344	MPN	344
Chronic lymphocytic leukemia	CLL	279	CLL	279
Multiple myeloma	MM	236	PCN group	247
Monoclonal gammopathy of undetermined significance	MGUS	11		
Myelodysplastic/myeloproliferative neoplasm, unclassifiable	MDS/MPN-U	81	MDS/MPN group	207
Atypical chronic myeloid leukemia	aCML	67		
Myelodysplastic/myeloproliferative neoplasm with ring sideroblasts and thrombocytosis	MDS/MPN-RS-T	59		
B-cell precursor acute lymphoblastic leukemia	BCP-ALL	204	BCP-ALL	204
T-cell non-Hodgkin lymphoma	T-NHL	90	T-cell group	174

T-cell acute lymphoblastic leukemia	T-ALL	84		
Chronic myelomonocytic leukemia	CMML	142	CMML	142
Mastocytosis		98	Mastocytosis	98
Hairy cell leukemia	HCL	68	Hairy-cell group	97
Hairy cell leukemia variant	HCL-V	29		
Chronic myeloid leukemia	CML	92	CML	92
Chronic lymphoproliferative disorder of natural killer cells	CLPD-NK	58	NK	58

Bioinformatic analysis

WGS and RNA-seq followed by variant calling of single nucleotide variants, short insertions and deletions, structural variants, copy number variations, and gene fusions were performed on all samples as described previously[29–32]. The analysis was based on GENCODE[33] v33 annotation and using the reference genome GRCh37. AbSplice (v1.0.0)[34], Ensembl VEP (v81)[35], IntOGen (commit 437a047)[36], and MutSigCV (v1.41)[37] were applied to the filtered variants to reduce common germline variants and artifacts.

OUTRIDER (v1.17.2)[38] and FRASER (v.1.99.0)[39] were applied to expression from RNA-seq using DROP (v1.2.3)[40], calling rare abnormal expression and splicing events. We grouped the samples into 14 study groups based on hematopoietic cell origins and pathologies[2–5] to ensure sufficient sample sizes (Table 1, N at least 58).

We introduced a novel method, NB-act (Negative Binomial activation), to detect aberrant activation of genes usually not expressed and therefore omitted by OUTRIDER. NB-act models background read counts given the expected baseline expression of 1 FPKM and negative binomial distribution.

Statistical analysis

Overall survival analyses were performed according to Kaplan-Meier and compared using two-sided log-rank tests with R v4.2.2 (R Foundation for Statistical Computing, Vienna, Austria) with the survival[41] and survminer[42] packages. The overall survival was calculated as the time from diagnosis to death or last follow-up.

Driver gene prediction

Using features extracted from seven intOGen tools, AbSplice, OUTRIDER, NB-act, and FRASER results, random forest classifiers were trained by cross-validation to predict the probability of a gene being a hematologic malignancy driver gene.

Validation HCL-V dataset

Patients

The validation HCL-V dataset included a total of 42 patients, including 14 HCL-V and 28 HCL patients. The HCL-V patients were diagnosed based on immunophenotype and morphological characteristics consistent with HCL-V, including all CD5 negative, CD11c positive, CD123 negative markers. All HCL-V patients were confirmed as *BRAF*-V600E negative.

Gene expression analysis

RNA-seq was performed on sorted tumor cells. Gene expression was analyzed with DESeq2[43] and R. The size factors were calculated using DEseq2. The counts were normalized by size factors. The activation of *LRP1B* was determined with a two-component Gaussian mixture clustering with equal variance on size-factor-normalized and log-transformed counts, using the R package 'mclust' (v.6.0.0)[44].

Data Sharing Statement

For original data, contact the corresponding authors.

Results

We investigated WGS and RNA-seq from 3,760 samples representing 24 different types of leukemia and lymphoma (Table 1). In order to restrict our analysis to putative rare germline and somatic variants, we filtered variants called on WGS with stringent quality filters and population allelic frequency[45]. We next annotated the genes using the seven features from the software intOGen, which include positional recurrence of mutations in genome sequence (OncodriveCLUSTL), positional recurrence of mutations in protein conformation (HotMAPS), enrichment of mutations in functional domains (smRegions), three alternative measures of selection strength inferred from synonymous and nonsynonymous mutations (CBaSE, MutPanning, and dNdScv), and OncodriveFML, a method identifying excess of somatic

mutations across tumors in both coding and non-coding genomic regions[36,46–52]. Moreover, we annotated genetic variants falling into gene bodies, including deep intronic variants, with AbSplice-DNA, a tool predicting variants causing aberrant splicing[34]. On the RNA-seq data, we used OUTRIDER on protein-coding genes commonly expressed across the dataset to call aberrantly high or low expression and FRASER to call rare splicing events[39,40]. We also introduced a new method, NB-act, to call aberrant activation of genes usually not expressed (Methods). Combining all these results, we established a unique resource of transcriptomic and genomic aberrations spanning 3,760 hematologic malignancy samples (Figure 1).

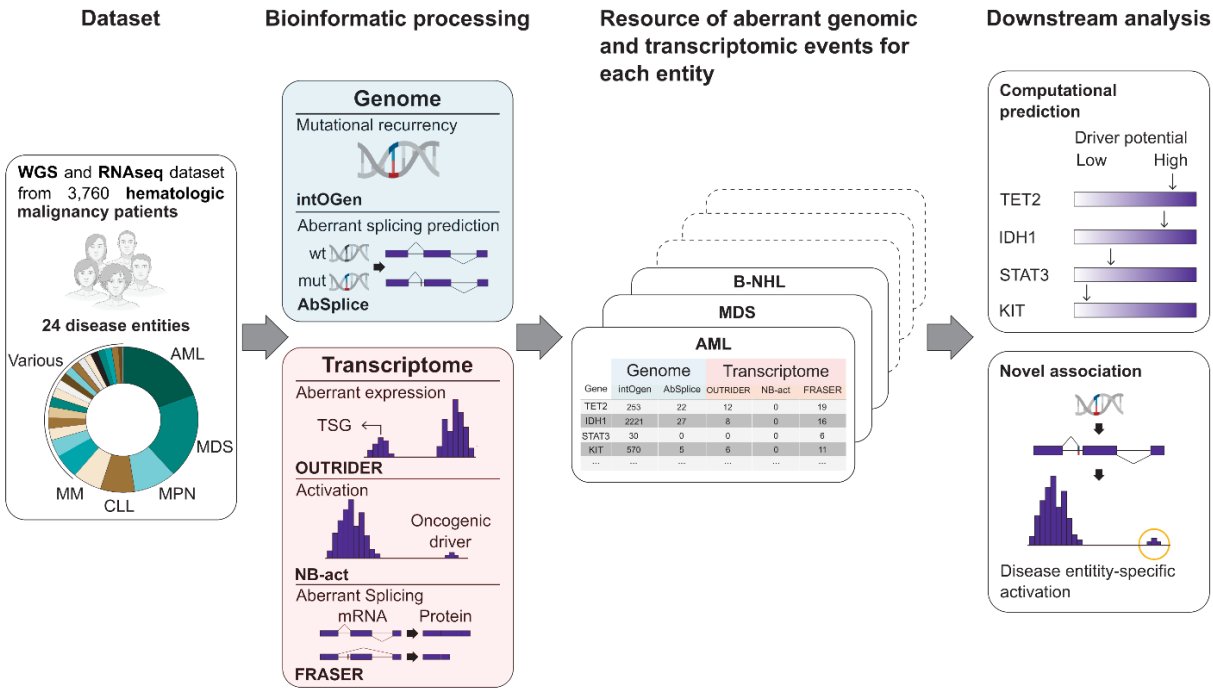

Figure 1. Overview of the study. Dataset: Whole genome sequencing and total RNA sequencing of 3,760 hematologic malignancies spanning 24 different disease entities. Bioinformatic processing: On genomic data, intOGen captures recurrent mutational patterns[37], and AbSplice predicts variants causing aberrant splicing[35]. Working on RNA-seq data, OUTRIDER calls aberrant expression levels of commonly-expressed genes[39], NB-act calls overexpression of rarely expressed genes (Methods), and FRASER calls aberrant splicing events[40]. Resource: a unique collection of transcriptomic and genomic aberrations for 24 hematologic malignancy entities. Downstream analysis: enrichment analysis to predict hematologic malignancy driver genes and reveal novel gene-disease associations.

Expression outliers are enriched for hematologic malignancy driver genes

We first called expression outliers using OUTRIDER, which detected 21,264 underexpression outliers (median of 2 per sample) and 14,041 overexpression outliers (median of 2 per sample) on 10,193 different protein-coding genes (Supplementary Figure 1-2). We observed a strong enrichment for hematologic tumor suppressor genes of the cancer gene census (CGC[54]) among underexpression outliers and for CGC hematologic oncogenes among overexpression outliers (Figure 2A-B). Notably, the genes called as outliers in more than five samples exhibited the highest enrichment, indicating that genes frequently called as outliers are more likely to be oncogenic. Moreover, we found that the number of expression outliers per sample was unevenly distributed (Supplementary Figure 1-2). Samples with numerous outliers represent cases where OUTRIDER could not adequately fit the data, either due to technical reasons or because the gene regulatory network is globally affected resulting in widespread expression aberrations throughout the genome. We reasoned that the enrichment for driver genes among expression outliers could be lower in all those samples with numerous outliers. Indeed, the enrichment for driver genes increased when focusing on the three most significant outliers (at most three) in each sample (Figure 2A-B).

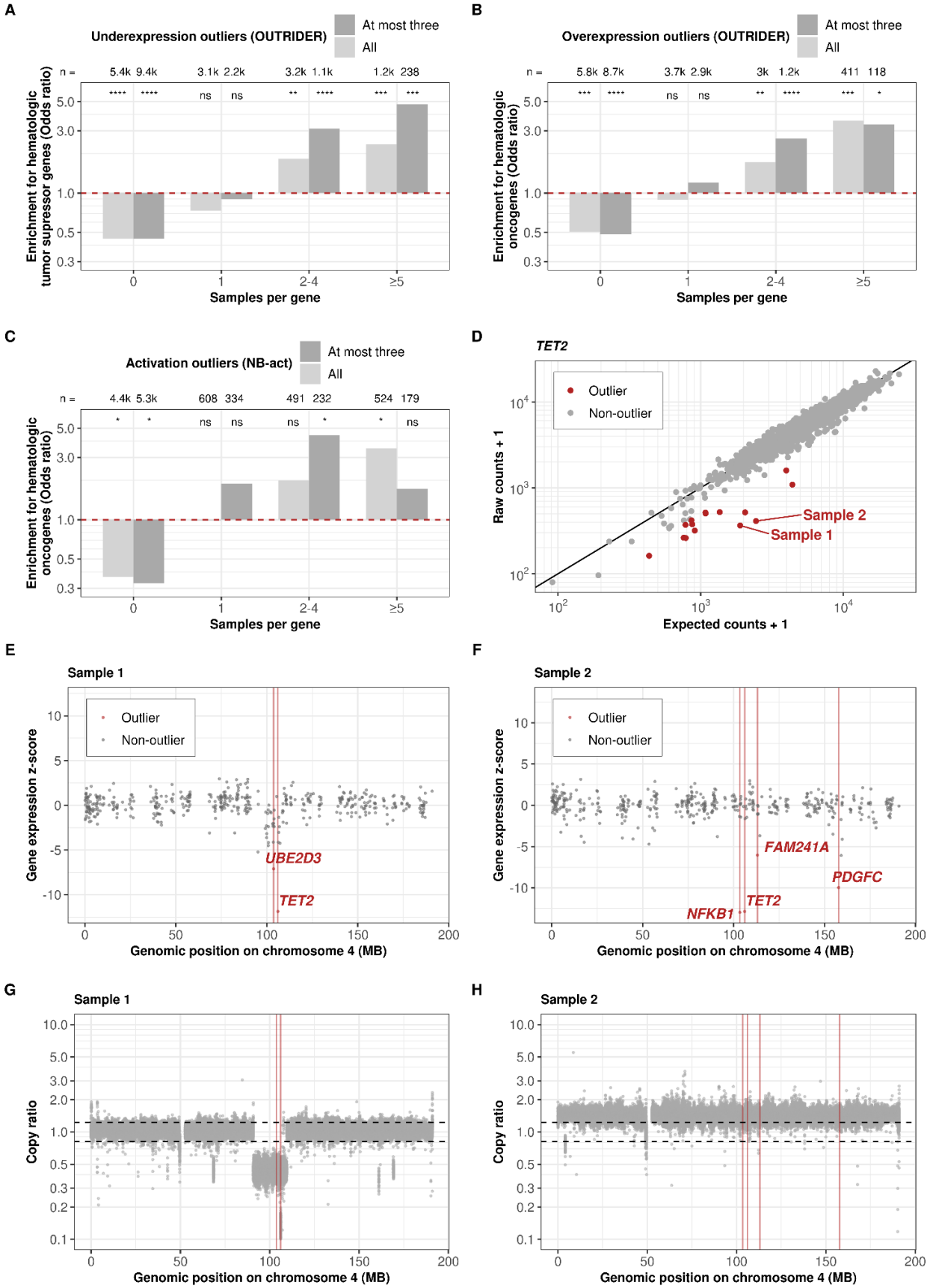
A
Underexpression outliers (OUTRIDER)
At most three
All
n =
5.4k 9.4k 3.1k 2.2k 3.2k 1.1k 1.2k 238
Enrichment for hematologic tumor supressor genes (Odds ratio)
Samples per gene
B
Overexpression outliers (OUTRIDER)
n =
5.8k 8.7k 3.7k 2.9k 3k 1.2k 411 118
Enrichment for hematologic oncogenes (Odds ratio)
Samples per gene
C
Activation outliers (NB-act)
n =
4.4k 5.3k 608 334 491 232 524 179
Enrichment for hematologic oncogenes (Odds ratio)
Samples per gene
D
TET2
Outlier
Non-outlier
Raw counts + 1
Expected counts + 1
Sample 2
Sample 1
E
Sample 1
Outlier
Non-outlier
Gene expression z-score
UBE2D3
TET2
Genomic position on chromosome 4 (MB)
F
Sample 2
Outlier
Non-outlier
Gene expression z-score
FAM241A
PDGFC
TET2
NFKB1
Genomic position on chromosome 4 (MB)
G
Sample 1
Copy ratio
Genomic position on chromosome 4 (MB)
H
Sample 2
Copy ratio
Genomic position on chromosome 4 (MB)

Figure 2. Expression outliers are enriched for hematologic malignancy driver genes. (A) Enrichment for CGC hematologic tumor suppressor genes among all genes called by OUTRIDER as well as at most three significant genes per sample that were called to be underexpression outliers. The genes are stratified by the number of samples in which the gene is called as an outlier. Numbers of the genes and significances from the Fisher test are labeled at the top of the bars (ns: not significant; *: $P \leq 0.05$; **: $P \leq 0.01$; ***: $P \leq 0.001$; ****: $P \leq 0.0001$). **(B-C)** As in A) for CGC hematologic oncogenes among overexpression and activation outliers, respectively. **(D)** Raw counts per sample against expected RNA-seq read counts for *TET2*. For fifteen samples, *TET2* was called as underexpression outlier, including Sample 1 and Sample 2, showcased in the panels (E-H). **(E-F)** Gene expression z-score among all samples against the genomic position of the genes on chromosome 4. **(G-H)** Copy ratio within chromosome 4 binned into one kb region. Red vertical lines mark the genomic position of the expression outliers. Black dashed lines mark the estimated region of no copy number variation. The underexpression outliers in *TET2* and *UBE2D3* in sample Sample 1 reflect the consequence of copy number loss that is very specific to the *TET2* locus. In contrast, an extra copy of whole chromosome 4 (karyotype: 47,XY,+4) is found in Sample 2. *TET2* reduced expression must have a different cause in this sample.

OUTRIDER filters out genes expressed in less than 5% of samples due to statistical modeling limitations, leaving a gap in detecting rare gene activation. We developed a complementary algorithm, NB-act (Methods), to fill this gap. We applied it to the 6,017 rarely-expressed protein-coding genes filtered out by OUTRIDER. NB-act identified 10,263 activation outliers among 1,623 genes (with a median of 0 and 75% quantile of 2 per sample, Supplementary Figure 3). We observed a notable enrichment for CGC hematologic oncogenes among all activation outliers (Figure 2C). Here too, restricting to at most three outliers per sample increased the enrichment. Altogether, this analysis provides a unique resource of aberrantly expressed genes in hematologic malignancies with strong enrichment for driver genes.

As an example, we found *TET2* as an underexpression outlier in 15 samples and never as an overexpression outlier (Figure 2D), consistent with its role as a tumor suppressor gene in myeloid leukemia[55–57]. The two samples with the lowest fold changes illustrate insights that can be gained following expression outlier analyses. For these two samples, *TET2* was among the most extreme outliers (Supplementary Figure 4-5). However, neither single nucleotide variants, short insertions and deletions, structural variants, nor gene fusions explained the observed decrease in *TET2* expression. Inspection of the genomic coverage indicated *TET2* locus nested within a double-copy loss of a small region in chromosome 4 for one sample. This explanation was further supported by the decreased expression of the neighbor gene *UBE2D3* (Figure 2 E, G). In contrast, genomic coverage investigations did not provide an explanation for the other sample. Across the dataset, 18.8% of the underexpression outliers could be explained by a

copy number loss, and 10.2 % of the overexpression outliers and 111.4% of the activation outliers could be explained by a copy number gain (Supplementary Figure 6). Further outliers, including *TET2* in that sample, could reflect yet-to-be-interpreted genetic variants, epigenetic causes, or outlier caller false positives[58].

Rare splicing aberrations are enriched for hematologic malignancy driver genes

Aberrant splicing, which is frequently caused by rare cis- or trans-effect of splice-affecting regulatory variants, often disrupts gene function as it typically results in non-functional transcripts. FRASER yielded 35,410 aberrantly spliced genes on a total of 7,591 genes in 2,854 samples (Supplementary Figure 7). We observed a substantial enrichment for CGC hematologic tumor suppressor genes among these splicing outliers (Figure 3A). As for expression outliers, restricting to at most three outliers per sample increased the enrichment. Moreover, the genes called as splicing outliers in more than five samples exhibited the highest enrichment, suggesting a higher potential of being oncogenic.

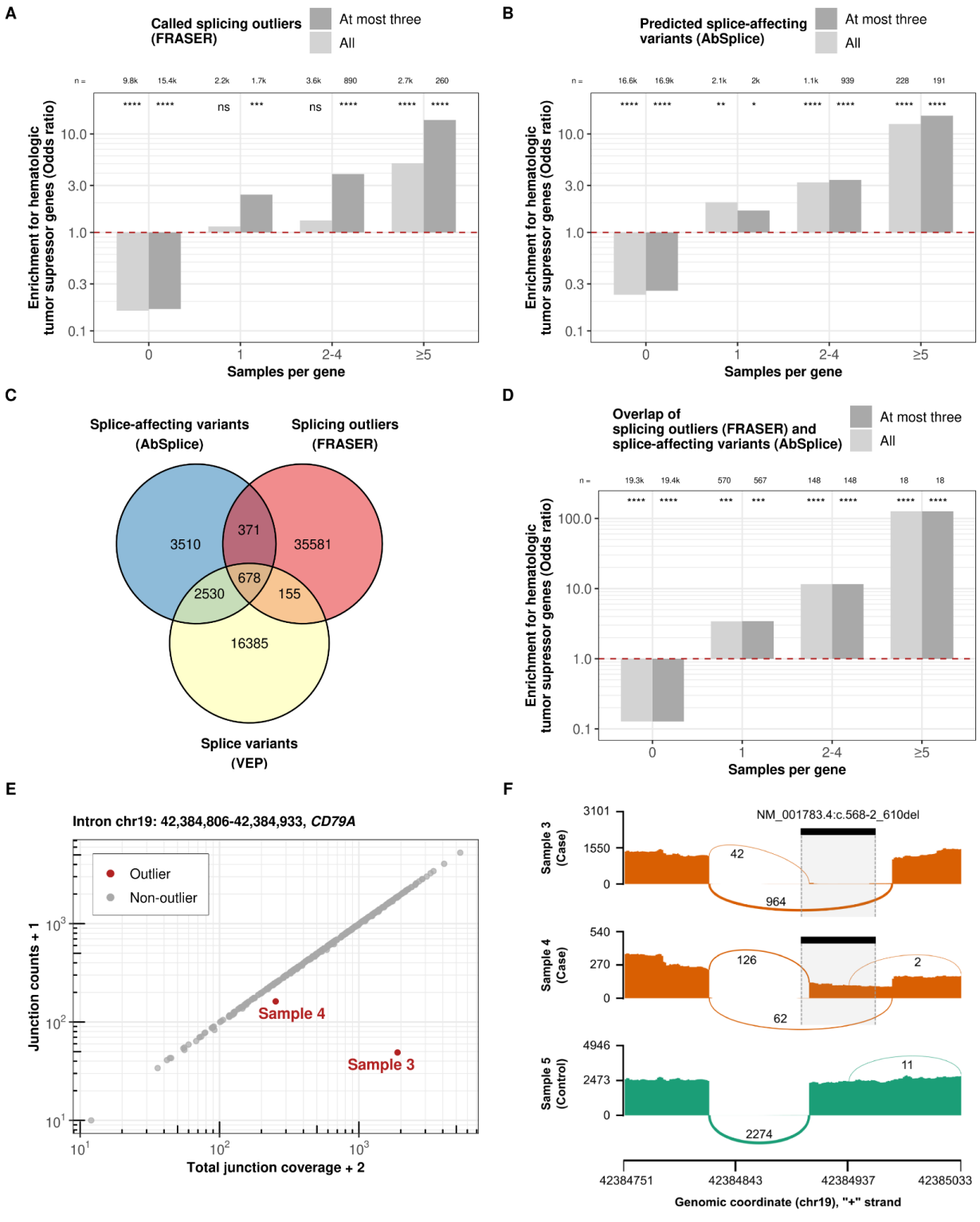

Figure 3. Transcriptomic splicing outliers and genomic splice-affecting variants are enriched for hematologic malignancy driver genes. (A) Enrichment for CGC hematologic tumor suppressor genes among all genes called by FRASER and at most three significant genes per sample that were called aberrantly spliced from RNA-seq (FRASER). The genes are stratified by the number of samples in which a gene is called as splicing outlier. Numbers

of the genes and significances from the Fisher test are labeled at the top of the bars. **(B)** As in (A) among splice-affecting variants (AbSplice). **(C)** Overlap of splice-affecting variants, VEP splice variants, and splicing outliers on the sample-gene level. **(D)** As in (A) among splice-affecting variants with corresponding splicing outliers. **(E)** Junction counts (split reads) against total junction coverage (exon-intron or intron-exon spanning reads) of the displayed intron. The displayed intron of the *CD79A* only shows abnormal splicing in samples Sample 3 and Sample 4. **(F)** Sashimi plots showing RNA-seq read coverage (y-axis) and the numbers of split reads spanning an intron indicated on the exon-connecting line (using pysashimi50) for two aberrant splicing events. Two case individuals using a unique splice-site in *CD79A* 5th exon acceptor site and one control individual are displayed. The rare splice-affecting deletion (NM_001783.4:c.568-2_610del) predicted by AbSplice, which existed exclusively in the two case individuals, is shown as black bars.

As a complementary approach to RNA-seq-based splicing outlier calling, we considered AbSplice[35], a recently published algorithm that predicts whether a variant causes aberrant splicing. Here we applied it for the first time on hematologic malignancy samples. AbSplice integrates two sequence-based machine learning tools, MMSplice[59], and SpliceAI[60], with so-called SpliceMaps, which are quantified tissue-specific usages of splice sites, including non-annotated and weak splice sites. We derived SpliceMaps from the raw RNA-seq data (Supplementary Materials and Methods). AbSplice classified 7,160 variants as splice-affecting out of 275,899,978 pre-filtered variants, resulting in 7,074 genes predicted to be affected (Supplementary Figure 8). These genes were strongly enriched for CGC hematologic tumor suppressor genes (Figure 3B). Furthermore, genes with AbSplice-predicted splice-affecting variants in multiple samples showed an even higher enrichment. In contrast to the RNA-seq-based splicing outlier calls, restricting AbSplice predictions to at most three calls per sample did not lead to a significantly higher enrichment, likely due to the limited number of samples with a high number of variants displaying strong AbSplice scores. Collectively, these results indicate that the aberrant splicing events predicted from the genome can contribute to identifying potential driver genes in hematologic malignancies.

Moreover, we identified 1,049 gene-sample combinations shared by RNA-seq called splicing outliers and predicted splice-affecting variants (Figure 3C). Among these combinations, 678 were called by both AbSplice and VEP, while AbSplice identified an additional 371 combinations, surpassing VEP which found 155 combinations. Within those 1,049 overlaps, we observed an even higher enrichment of CGC hematologic tumor suppressor genes (Figure 3D). As an example, we found splicing outliers in the oncogene *CD79A*[54,61–63] in two samples, resulting in a truncation of 18 amino acids at the beginning of the 5th exon (Figure 3E-F). We identified a unique, rare 45-base-pair deletion spanning the acceptor site

(NM_001783.4:c.568-2_610del) exclusively in these two samples, which AbSplice predicted this variant to be splice-affecting (AbSplice score = 0.36, Figure 3F, Supplementary Figure 9). Notably, similar deletions in the 4th and 5th exon of *CD79A* have been observed in diffuse large B-cell lymphoma, which has been shown to impact the ITAM signaling modules[61]. AbSplice successfully predicted the splice effect of a variant, and subsequent analysis confirmed the corresponding splicing aberrations in the transcriptome.

Additionally, some splicing aberrations were first noticed using the RNA-seq data and overlooked using the genomic data alone. As an example, we identified an unusual exon-skipping event of the 20th exon in the tumor suppressor gene *RB1*[64] (Supplementary Figure 10-11). Re-analysis of the genome revealed a deletion of 118-base-pair in the region of the 20th exon, which led to an exon-skipping-induced frameshift (Supplementary Figure 11). This structural variant was discarded initially in our variant calling pipeline due to quality control, but we eventually recognized its significant transcriptomic impact and rescued it. Altogether, calling splicing outliers from transcriptomic data offered substantial additional information on top of predicting splice-affecting variants. This approach proves invaluable in understanding individual samples and sheds light on the underlying mechanisms of splicing abnormalities.

All our results from splice-affecting variants, expression outliers, splicing outliers, and recurrent mutational patterns grouped by disease entities are provided in Supplementary Tables 1 and 2.

<u>Integration of rare transcriptomic and genomic events improves hematologic malignancy driver gene prediction</u>

We next trained models to predict driver genes based on transcriptomic and genomic features, including the seven intOGen tools, AbSplice, OUTRIDER, NB-act, and FRASER. To this end, we used the complete dataset on the one hand and used the dataset stratified by 14 study groups on the other hand (Table 1). As ground truth, we considered the 322 CGC hematologic malignancy driver genes complemented with 55 additional hematologic panel genes (Supplementary Materials and Methods). The performance of all 15 models was evaluated using 5-fold cross-validation.

Using the complete dataset, we found that the transcriptomic and genomic features exhibited complementary predictive power for hematologic malignancy driver genes (Figure 4A-B, Supplementary Figure 12-13). Specifically, integrating AbSplice predictions significantly

enhanced the genomic-based model trained on the seven intOGen tools. Moreover, we found that the transcriptomic features further significantly improved the model (Figure 4B). These findings underscore the relevance of incorporating aberrant expression and splicing analyses to predict driver genes. Among the 100 top-ranked genes, 63 were known from the ground truth, vastly greater than expectation (odds ratio = 106.3, P = 1.6×10^{-84}, Fisher test; Figure 4C, Supplementary Table 3). These enrichments for well-known driver genes demonstrated the reliability of the model predictions. Four genes, *CDNKN1B*, *EIF3E*, *HLA-A,* and *IL6ST*, are CGC driver genes that have not been annotated as hematologic relevant yet, indicating a broader role for those. The remaining 33 genes were promising candidate drivers whose role remains to be further assessed.

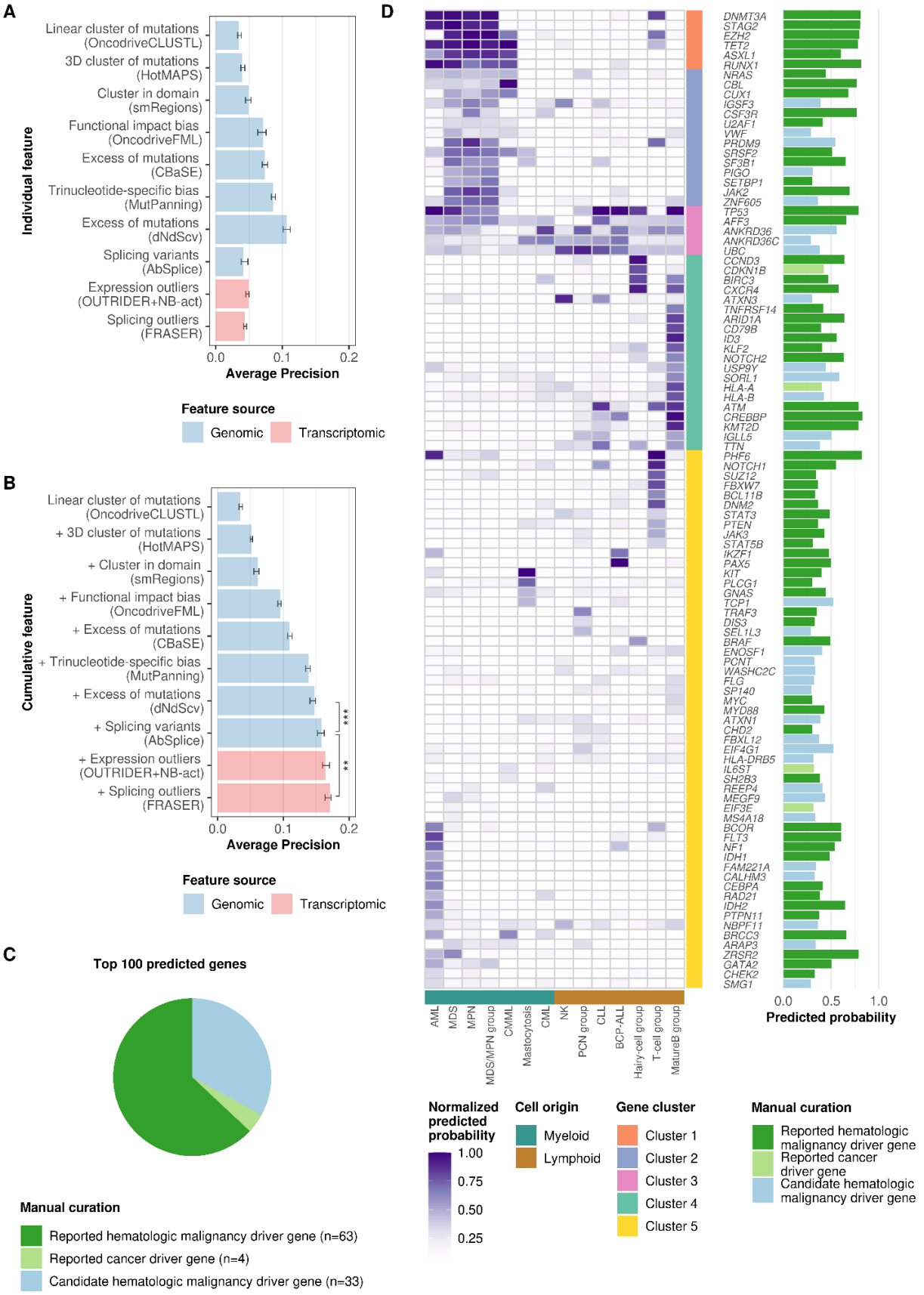
A
Individual feature
Linear cluster of mutations (OncodriveCLUSTL)
3D cluster of mutations (HotMAPS)
Cluster in domain (smRegions)
Functional impact bias (OncodriveFML)
Excess of mutations (CBaSE)
Trinucleotide-specific bias (MutPanning)
Excess of mutations (dNdScv)
Splicing variants (AbSplice)
Expression outliers (OUTRIDER+NB-act)
Splicing outliers (FRASER)
0.0
0.1
0.2
Average Precision
Feature source
Genomic
Transcriptomic
B
Cumulative feature
Linear cluster of mutations (OncodriveCLUSTL)
+ 3D cluster of mutations (HotMAPS)
+ Cluster in domain (smRegions)
+ Functional impact bias (OncodriveFML)
+ Excess of mutations (CBaSE)
+ Trinucleotide-specific bias (MutPanning)
+ Excess of mutations (dNdScv)
+ Splicing variants (AbSplice)
+ Expression outliers (OUTRIDER+NB-act)
+ Splicing outliers (FRASER)
***
**
0.0
0.1
0.2
Average Precision
Feature source
Genomic
Transcriptomic
C
Top 100 predicted genes
Manual curation
Reported hematologic malignancy driver gene (n=63)
Reported cancer driver gene (n=4)
Candidate hematologic malignancy driver gene (n=33)
D
DNMT3A
STAG2
EZH2
TET2
ASXL1
RUNX1
NRAS
CBL
CUX1
IGSF3
CSF3R
U2AF1
VWF
PRDM9
SRSF2
SF3B1
PIGQ
SETBP1
JAK2
ZNF605
TP53
AFF3
ANKRD36
ANKRD36C
UBC
CCND3
CDKN1B
BIRC3
CXCR4
ATXN3
TNFRSF14
ARID1A
CD79B
ID3
KLF2
NOTCH2
USP9Y
SORL1
HLA-A
HLA-B
ATM
CREBBP
KMT2D
IGLL5
TTN
PHF6
NOTCH1
SUZ12
FBXW7
BCL11B
DNM2
STAT3
PTEN
JAK3
STAT5B
IKZF1
PAX5
KIT
PLCG1
GNAS
TCP1
TRAF3
DIS3
SEL1L3
BRAF
ENOSF1
PCNT
WASCH2C
FLG
SP140
MYC
MYD88
ATXN1
CHD2
FBXL12
EIF4G1
HLA-DRB5
IL6ST
SH2B3
REEP4
MEGF9
EIF3E
MS4A18
BCOR
FLT3
NF1
IDH1
FAM221A
CALHM3
CEBPA
RAD21
IDH2
PTPN11
NBPF11
BRCC3
ARAP3
ZRSR2
GATA2
CHEK2
SMG1
0.0
0.5
1.0
Predicted probability
AML
MDS
MPN
MDS/MPN group
CMML
Mastocytosis
CML
NK
PCN group
CLL
BCP-ALL
Hairy-cell group
T-cell group
MatureB group
Normalized predicted probability
1.00
0.75
0.50
0.25
Cell origin
Myeloid
Lymphoid
Gene cluster
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Manual curation
Reported hematologic malignancy driver gene
Reported cancer driver gene
Candidate hematologic malignancy driver gene

Figure 4. Rare transcriptomic and genomic events add value to hematologic malignancy driver gene prediction on top of methods detecting mutational recurrence. (A) Individual features and the corresponding model's performances (measured by average precision) using a random forest classifier **(B)** Order as in (A) for cumulative features. The performance constantly improved when adding additional features. Significances from the Wilcoxon test are labeled at the right. **(C)** Numbers of genes in each category among the top 100 predicted genes when using all features and the complete dataset. Reported hematologic malignancy driver genes are the genes listed in either CGC hematologic malignancy driver genes or hematologic panel genes. Reported cancer driver genes are the genes documented by CGC. The rest of the genes are categorized as candidate hematologic malignancy driver genes. **(D)** The heatmap shows the predicted driver gene probability per gene (rows) and study group (columns) relative to the column-wise maximum value. Myeloid and lymphoid entities clustered due to the shared hematologic malignancy driver gene profiles (bottom track). The barplot shows the predicted probabilities for the model trained on the complete dataset. Bar colors as in (C).

In line with the results on the complete dataset, across all study groups, our integrative model outperformed or was on par with intOGen or MutSigCV - a driver gene predictor based on mutation frequency (Supplementary Figure 14). The study-group-wise models allowed further insights into disease entities (Figure 4D, Supplementary Table 4). Clustering the 100 top-ranked genes according to their study-group prediction probabilities revealed various disease-entity specificities. Cluster 3 consisted of genes that exhibited high predicted probabilities across all disease entities, including the pan-cancer tumor suppressor *TP53*[65]. Clusters 1 and 2 comprised genes that scored highly specifically in myeloid disease entities, such as *ASXL1* and *SRSF2*[4,66]. Conversely, the genes in the remaining clusters were associated with specific disease entities. Examples include *NOTCH1, PHF6, and FBXW7,* which are well-acknowledged for their role in T-cell leukemia, and *BRAF*, which was exclusively predicted in the hairy-cell group, in accordance with its recognized function[4,67–69].

Among the candidate genes, we identified several promising genes whose roles in hematologic malignancies remain to be established. Our analysis predicted *SORL1* as a candidate in multiple study groups, including myelodysplastic neoplasm (precursor of AML), B-cell precursor ALL, and a study group comprising T-cell non-Hodgkin lymphoma and T-cell acute lymphoblastic leukemia. Consistent with these observations, *SORL1* was found to be expressed on the leukemic cell surface and released into plasma in AML and ALL, with its level decreasing during remission[70]. *EIF4G1* stood out as another interesting candidate driver in AML and lymphoid entities, supported by previous analyses suggesting that *EIF4G1* is involved in cell survival in AML as a downstream target of *MYCN*, a known oncogene in neuroblastoma[71]. We also found *TCP1* to be predicted in Mastocytosis and B-cell precursor-ALL, whose high

expression has been associated with AML drug resistance and poor survival through the activation of *AKT/mTOR* signaling[72]. A fourth exciting candidate, *ATXN1,* was predicted as a candidate driver in multiple disease entities. Its important paralog *ATXN1L* is known as a novel regulator of hematopoietic stem cell quiescence[73]. Overall, our prediction model based on 3,760 genomes and transcriptomes reveals promising hematologic malignancy driver candidates.

Outlier clustering identifies *LRP1B* as a novel biomarker in HCL-V

Finally, we performed a detailed investigation into each disease entity, examining its association with expression outliers, splicing outliers, and variants predicted to cause aberrant splicing. Overall, we found 2,716 significant associations between 11,273 genes and 24 disease entities (Supplementary Table 5, FDR < 0.05, Fisher test). Focusing on activation outliers and established cancer driver genes, we found 43 associations between 37 CGC cancer driver genes and 12 disease entities (Figure 5A). While most of these cancer genes were specific for one disease entity (e.g., *NKX2-1* for T-cell acute lymphoblastic leukemia, T-ALL), few were associated with two different entities (*FAT3*, *HOXA11*, *LRP1B,* and *TERT*). Notably, B-cell precursor acute lymphoblastic leukemia, T-ALL, and multiple myeloma showed the most associated outlier cancer genes. Some associations were already described (Supplementary Table 6), including the activations of the transcription factors *TLX1* and *TLX3* in T-ALL[74,75], and the associations of *HOXA11*, *PREX2,* and *RET* with AML[76–78]. Novel associations included overexpression of *WNK2* in High-grade B-cell lymphoma, of *FAT4* in multiple myeloma, and of *LRP1B* in hairy cell leukemia (HCL), hairy cell leukemia variant (HCL-V) and marginal zone lymphoma (MZL).

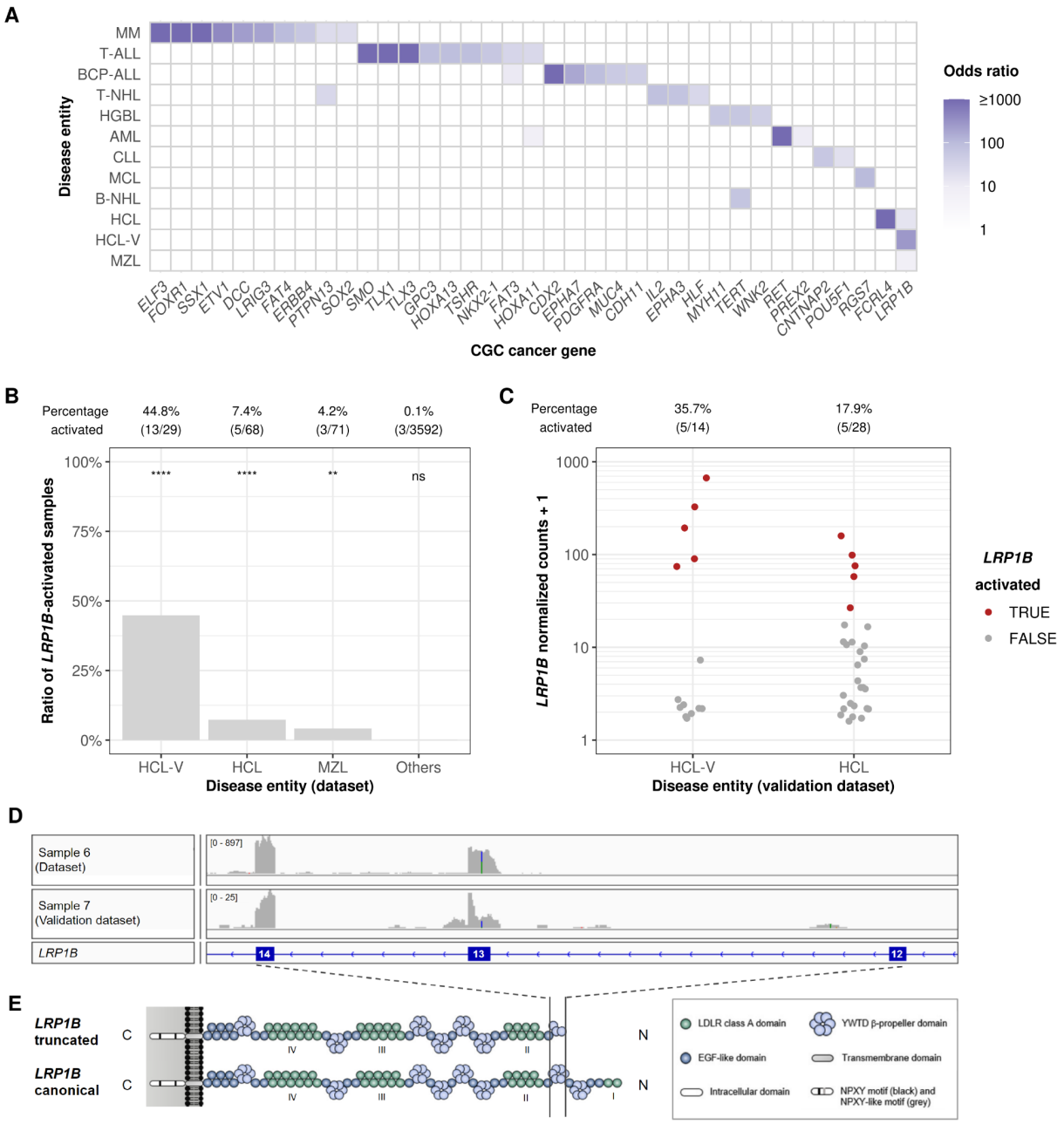

Figure 5. Aberrant expression of *LRP1B* is specific for HCL-V. (A) Disease entities (full names in Table 1) against aberrantly activated CGC genes, colored by odds ratio from Fisher test. **(B)** Percentage of *LRP1B*-activated samples per disease entity. **(C)** *LRP1B*-activated samples within corresponding disease entities. Percentages of samples showing *LPR1B* activation (NB-act) and significances from the one-sided Fisher test are labeled at the top. **(C)** Normalized counts of *LRP1B*-activated samples of the dataset with regard to the different disease entities in the validation dataset. Percentages of samples showing *LPR1B* activation (Gaussian mixture clustering) are labeled at the top. **(D)** Transcriptomic coverage tracks showing *LRP1B* truncated isoform (exon 12 to 14) of samples Sample 6 (Dataset) and Sample 7 (Validation dataset). **(E)** Domain organization of truncated and canonical *LRP1B* (adapted from Príncipe et al.[92]).

Remarkably, *LRP1B* was associated with very rare entities. *LRP1B*, or Low-Density Lipoprotein Receptor-related protein 1B, is a frequently altered gene in multiple cancer types, but its role remains unclear[79]. *LRP1B* was an activation outlier in 24 samples, out of which 21 were found within HCL-V, HCL, and MZL, where *LRP1B*-activated samples made up 44.8% (HCL-V), 7.4% (HCL) and 4.2% (MZL) of each entity (Figure 5B). Conversely, among all *LRP1B*-activated samples, more than half of the cases (13 out of 24) were found in HCL-V patients, indicating that aberrant *LRP1B* expression might play an important role in HCL-V. *LRP1B* activation associated, yet not significantly, with shorter survival in HCL-V patients (median OS: 3.4 vs. 6.3 years, $P = 0.45$, two-sided log-rank test; Supplementary Figure 15). We then evaluated *LRP1B* expression in an independent validation dataset of 42 samples. Consistently, we detected 10 *LRP1B*-activated samples in HCL-V (5/14; 36%) and HCL (5/28; 18%) in the validation dataset (Figure 5C). Moreover, the RNA-seq coverage for both datasets showed that samples overexpressing *LRP1B* expressed a truncated isoform in the majority (32/34; 94%) of samples (22/24 in the dataset; 10/10 in the validation dataset; Figure 5D). In these cases, *LRP1B* expression started at exon 13, thus lacking the first 636 amino acids of exon 1 to 12. Assuming this transcript is translated, it yields a shortened *LRP1B* protein truncated at half of the second β-propeller domain (Figure 5E). Three start codons located within exon 13 could enable the start of translation using the canonical open reading frame. However, the mechanistic cause for this activation of the truncated form of LRP1B remains to be discovered, as no single nucleotide variants, short insertions, and deletions, structural variants, or gene fusions involving *LRP1B* were found to be specific to the affected samples.

Discussion

The study of rare cancer aberrations is an emerging field that greatly benefits from the growing large-scale next-generation sequencing[80,81]. We have generated an extensive catalog of aberrant transcriptomic and genomic events across 3,760 patients spanning 24 hematologic malignancy disease entities, the largest of its kind to date. Furthermore, we built pan-leukemia and entity-specific driver gene predictors by integrating this data, successfully recovering known drivers, and proposing promising novel candidates. Our dataset, to our knowledge, currently represents the largest collection of hematologic malignancy samples that have undergone WGS along with matched RNA-seq data, which also includes rare disease entities.

One notable finding is the identification of *LRP1B* as a novel marker for HCL-V. In our RNA-seq samples, *LRP1B* expression was rarely detected, occurring in only approximately 1% of cases.

However, it was highly expressed in some cases of MZL, HCL, and approximately 50% of HCL-V, which we confirmed in an independent cohort. These three disease entities are mature B-cell malignancies and were previously regarded as separate entities in the revised 4th edition of the WHO classification of hematolymphoid tumors, distinguished based on immunophenotypic markers and molecular genetics[82]. HCL-V is typically resistant to conventional HCL therapy and does not show the HCL-specific *BRAF*-V600E mutation. However, as HCL, HCL-V, and MZL arise from malignant mature B-cells showing similar morphology, clear discrimination using conventional diagnostic techniques is often not possible. Thus, in the recently published 5th edition of WHO classification, the term “HCL-V" has been removed, recognizing that the biology of this disease is unrelated to HCL[2]. Instead, these cases are now considered splenic B-cell lymphoma/leukemia with prominent nucleoli (SBLPN), which also comprises rare cases of splenic MZL and B-prolymphocytic leukemia based on similar cytomorphological features. SBLPN rather serves as a placeholder for those morphologically defined cases of B-cell lymphoma not being classifiable into biologically distinct entities based on current evidence-based knowledge. We observed a trend for a worse prognosis for HCL-V samples expressing *LRP1B*, though a larger sample size is needed to establish a statistical significance. Overall, our results suggest a potential subcategorization of HCL-V/SBLPN based on LRP1B expression, whose functional implications remain to be elucidated.

LRP1B, also known as Low-Density Lipoprotein Receptor-related protein 1B, is broadly expressed in multiple normal tissues but not in blood or bone marrow[83–85]. It plays a role in various biological processes such as angiogenesis, chemotaxis, proliferation, adhesion, apoptosis, endocytosis, immunity, host-virus interaction, and protein folding. Additionally, *LRP1B* is among the most frequently altered genes in human cancer overall[86–91]. For tissues where *LRP1B* is normally expressed, *LRP1B* is often inactivated in cancer through several genetic and epigenetic mechanisms, making it a putative tumor suppressor gene. Overexpression of a truncated isoform of *LRP1B* in cells that normally do not express it, as we observed here, could play a role in oncogenesis by disrupting similar biological processes. Despite the challenges posed by the enormous size of LRP1B, as highlighted in the study by Príncipe et al.[92], our findings encourage further investigations to unravel the potential role of *LRP1B* in B-cell malignancies.

In conclusion, we established a unique and comprehensive resource encompassing the transcriptomic and genomic landscape of 3,760 hematologic malignancy samples, covering a

wide range of disease entities. This comprehensive resource can be leveraged to identify novel biomarkers, propose therapeutic decisions, and unravel the molecular underpinning of the heterogeneity of hematologic cancers.

Acknowledgments

The authors thank everyone from the MLL 5k project involved in patient inclusion, sample collection, and sequencing data generation; everyone involved in patient inclusion, sample collection, and sequencing data generation of the HCL-V validation dataset; Luise Schuller, and Kim Anh Lilian Le for their contributions to the initial work of this project; Felix Brechtmann, and Christian Mertes for the enlightening discussions.

This study was supported by the German Bundesministerium für Bildung und Forschung (BMBF) supported through the VALE (Entdeckung und Vorhersage der Wirkung von genetischen Varianten durch Artifizielle Intelligenz für LEukämie Diagnose und Subtypidentifizierung) project [031L0203B to X.Q., J.G.; 031L0203C to Sa.H., St.H.; and 031L0203A to M.H.]

Authorship Contributions

Contribution: X.C., L.W., St.H., and J.G. designed the research; X.C., Sa.H., A.J.A., M.S., P.S., and St.H. performed the research; C.O., and St.H. collected the data; X.C., Sa.H., A.J.A., M.S., I.S., and N.W. analyzed and interpreted the data; X.C., and Sa.H. performed statistical analysis; X.C., and A.J.A. developed the software; M.H., L.W., St.H., and J.G. supervised the research; M.H., T.H., St.H., and J.G. acquired the funding; X.C., Sa.H., A.J.A., F.R.T., M.S., L.W., St.H., and J.G. wrote the manuscript; and all authors reviewed and agreed to the final version of the manuscript.

Correspondence: Julien Gagneur, School of Computation, Information and Technology, Technical University of Munich, Boltzmannstraße 3, 85748 Garching bei München, Germany; e-mail: gagneur@in.tum.de; and Stephan Hutter, Münchner Leukämielabor GmbH (MLL), Max-Lebsche-Platz 31, 81377 München, Germany; e-mail: stephan.hutter@mll.com; and Leonhard Wachutka, School of Computation, Information and Technology, Technical University of Munich, Boltzmannstraße 3, 85748 Garching bei München, Germany; e-mail: wachutka@cit.tum.de.

Conflict of Interest Disclosures

T.H. declares part ownership of Munich Leukemia Laboratory (MLL). Sa.H. and St.H. are employed by the MLL.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660
2. Alaggio R, Amador C, Anagnostopoulos I, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia*. 2022;36(7):1720-1748. doi:10.1038/s41375-022-01620-2
3. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia*. 2022;36(7):1703-1719. doi:10.1038/s41375-022-01613-1
4. Arber DA, Orazi A, Hasserjian RP, et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood*. 2022;140(11):1200-1228. doi:10.1182/blood.2022015850
5. Campo E, Jaffe ES, Cook JR, et al. The International Consensus Classification of Mature Lymphoid Neoplasms: a report from the Clinical Advisory Committee. *Blood*. 2022;140(11):1229-1253. doi:10.1182/blood.2022015851
6. Ostroverkhova D, Przytycka TM, Panchenko AR. Cancer driver mutations: predictions and reality. *Trends Mol Med*. 2023;29(7):554-566. doi:10.1016/j.molmed.2023.03.007
7. Nussinov R, Tsai CJ, Jang H. Why Are Some Driver Mutations Rare? *Trends Pharmacol Sci*. 2019;40(12):919-929. doi:10.1016/j.tips.2019.10.003
8. Calabrese C, Davidson NR, Demircioğlu D, et al. Genomic basis for RNA alterations in cancer. *Nature*. 2020;578(7793):129-136. doi:10.1038/s41586-020-1970-0
9. Hautin M, Mornet C, Chauveau A, Bernard D, Corcos L, Lippert E. Splicing Anomalies in Myeloproliferative Neoplasms: Paving the Way for New Therapeutic Venues. *Cancers*. 2020;12(8):2216. doi:10.3390/cancers12082216
10. Grinev VV, Barneh F, Ilyushonak IM, et al. RUNX1/RUNX1T1 mediates alternative splicing and reorganises the transcriptional landscape in leukemia. *Nat Commun*. 2021;12(1):520. doi:10.1038/s41467-020-20848-z
11. Tanaka A, Nakano TA, Nomura M, et al. Aberrant EVI1 splicing contributes to EVI1-rearranged leukemia. *Blood*. 2022;140(8):875-888. doi:10.1182/blood.2021015325
12. Obeng EA. Mutant SF3B1 splices a more leukemogenic EVI1. *Blood*. 2022;140(8):800-801. doi:10.1182/blood.2022017380
13. Huber S, Haferlach T, Meggendorfer M, et al. SF3B1 mutations in AML are strongly associated with MECOM rearrangements and may be indicative of an MDS pre-phase. *Leukemia*. 2022;36(12):2927-2930. doi:10.1038/s41375-022-01734-7
14. Yang YT, Chiu YC, Kao CJ, et al. The prognostic significance of global aberrant alternative splicing in patients with myelodysplastic syndrome. *Blood Cancer J*. 2018;8(8):1-11. doi:10.1038/s41408-018-0115-2
15. Szelest M, Masternak M, Zając M, et al. The role of NPM1 alternative splicing in patients with chronic lymphocytic leukemia. *PLOS ONE*. 2022;17(10):e0276674. doi:10.1371/journal.pone.0276674
16. Dlamini Z, Shoba B, Hull R. Splicing machinery genomics events in acute myeloid

leukaemia (AML): in search for therapeutic targets, diagnostic and prognostic biomarkers. *Am J Cancer Res*. 2020;10(9):2690-2704.
17. Rivera OD, Mallory MJ, Quesnel-Vallières M, et al. Alternative splicing redefines landscape of commonly mutated genes in acute myeloid leukemia. *Proc Natl Acad Sci*. 2021;118(15):e2014967118. doi:10.1073/pnas.2014967118
18. Black KL, Naqvi AS, Asnani M, et al. Aberrant splicing in B-cell acute lymphoblastic leukemia. *Nucleic Acids Res*. 2018;46(21):11357-11369. doi:10.1093/nar/gky946
19. Huber S, Haferlach T, Meggendorfer M, et al. Mutations in spliceosome genes in myelodysplastic neoplasms and their association to ring sideroblasts. *Leukemia*. 2023;37(2):500-502. doi:10.1038/s41375-022-01783-y
20. Lewis RE, Cruse JM, Sanders CM, Webb RN, Suggs JL. Aberrant expression of T-cell markers in acute myeloid leukemia. *Exp Mol Pathol*. 2007;83(3):462-463. doi:10.1016/j.yexmp.2007.08.010
21. Expression of bcr And bcr-abl Fusion Transcripts in Normal and Leukemic Cells: Molecular and Cellular Biology: Vol 7, No 8. Accessed August 7, 2023. https://www.tandfonline.com/doi/abs/10.1128/mcb.7.8.2870-2876.1987?casa_token=_3tfip--FjQAAAAA:B_VOsgHNqTGranTRcyHe8lcC2Og4cX8yXMz3sQlJeDx43zDfEEkrj6l4W0p5VuxPU5NFLSNDd_s9fg
22. Rimokh R, Berger F, Bastard C, et al. Rearrangement of CCND1 (BCL1/PRAD1) 3' untranslated region in mantle- cell lymphomas and t(11q13)-associated leukemias. *Blood*. 1994;83(12):3689-3696. doi:10.1182/blood.V83.12.3689.3689
23. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011;478(7367):64-69. doi:10.1038/nature10496
24. Damm F, Kosmider O, Gelsi-Boyer V, et al. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood*. 2012;119(14):3211-3218. doi:10.1182/blood-2011-12-400994
25. Adamia S, Haibe-Kains B, Pilarski PM, et al. A Genome-Wide Aberrant RNA Splicing in Patients with Acute Myeloid Leukemia Identifies Novel Potential Disease Markers and Therapeutic Targets. *Clin Cancer Res*. 2014;20(5):1135-1145. doi:10.1158/1078-0432.CCR-13-0956
26. Nakai H, Kaneko H, Horiike S, et al. Multiple aberrant splicing of the p53 transcript without genomic mutations around exon-intron junctions in a case of chronic myelogenous leukaemia in blast crisis: a possible novel mechanism of p53 inactivation. *Br J Haematol*. 1994;87(4):839-842. doi:10.1111/j.1365-2141.1994.tb06747.x
27. Kern W, Voskova D, Schoch C, Hiddemann W, Schnittger S, Haferlach T. Determination of relapse risk based on assessment of minimal residual disease during complete remission by multiparameter flow cytometry in unselected patients with acute myeloid leukemia. *Blood*. 2004;104(10):3078-3085. doi:10.1182/blood-2004-03-1036
28. Haferlach T, Schoch C, Löffler H, et al. Morphologic Dysplasia in De Novo Acute Myeloid Leukemia (AML) Is Related to Unfavorable Cytogenetics but Has No Independent Prognostic Relevance Under the Conditions of Intensive Induction Therapy: Results of a Multiparameter Analysis From the German AML Cooperative Group Studies. *J Clin Oncol*. 2003;21(2):256-265. doi:10.1200/JCO.2003.08.005
29. Schoch C, Schnittger S, Bursch S, et al. Comparison of chromosome banding analysis, interphase- and hypermetaphase-FISH, qualitative and quantitative PCR for diagnosis and for follow-up in chronic myeloid leukemia: a study on 350 cases. *Leukemia*. 2002;16(1):53-59. doi:10.1038/sj.leu.2402329
30. Haferlach C, Walter W, Stengel A, et al. The diverse landscape of fusion transcripts in 25 different hematological entities. *Leuk Lymphoma*. 2021;62(13):3292-3295. doi:10.1080/10428194.2021.1953009
31. Hershberger CE, Moyer DC, Adema V, et al. Complex landscape of alternative splicing in

myeloid neoplasms. *Leukemia*. 2021;35(4):1108-1120. doi:10.1038/s41375-020-1002-y
32. Stengel A, Baer C, Walter W, et al. Mutational patterns and their correlation to CHIP-related mutations and age in hematological malignancies. *Blood Adv*. 2021;5(21):4426-4434. doi:10.1182/bloodadvances.2021004668
33. Höllein A, Twardziok SO, Walter W, et al. The combination of WGS and RNA-Seq is superior to conventional diagnostic tests in multiple myeloma: Ready for prime time? *Cancer Genet*. 2020;242:15-24. doi:10.1016/j.cancergen.2020.01.001
34. Frankish A, Diekhans M, Ferreira AM, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. 2019;47(D1):D766-D773. doi:10.1093/nar/gky955
35. Wagner N, Çelik MH, Hölzlwimmer FR, et al. Aberrant splicing prediction across human tissues. *Nat Genet*. Published online May 4, 2023:1-10. doi:10.1038/s41588-023-01373-3
36. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122. doi:10.1186/s13059-016-0974-4
37. Martínez-Jiménez F, Muiños F, Sentís I, et al. A compendium of mutational cancer driver genes. *Nat Rev Cancer*. 2020;20(10):555-572. doi:10.1038/s41568-020-0290-x
38. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499(7457):214-218. doi:10.1038/nature12213
39. Brechtmann F, Mertes C, Matusevičiūtė A, et al. OUTRIDER: A Statistical Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data. *Am J Hum Genet*. 2018;103(6):907-917. doi:10.1016/j.ajhg.2018.10.025
40. Scheller IF, Lutz K, Mertes C, Yépez VA, Gagneur J. Improved detection of aberrant splicing using the Intron Jaccard Index. Published online April 3, 2023:2023.03.31.23287997. doi:10.1101/2023.03.31.23287997
41. Yépez VA, Mertes C, Müller MF, et al. Detection of aberrant gene expression events in RNA sequencing data. *Nat Protoc*. 2021;16(2):1276-1296. doi:10.1038/s41596-020-00462-5
42. Therneau TM. *A Package for Survival Analysis in R*.; 2023. https://CRAN.R-project.org/package=survival
43. Kassambara A, Kosinski M, Biecek P. *Survminer: Drawing Survival Curves Using "Ggplot2."*; 2021. https://CRAN.R-project.org/package=survminer
44. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8
45. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J*. 2016;8(1):289-317.
46. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7
47. Arnedo-Pac C, Mularoni L, Muiños F, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUSTL: a sequence-based clustering method to identify cancer drivers. *Bioinformatics*. 2019;35(22):4788-4790. doi:10.1093/bioinformatics/btz501
48. Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol*. 2016;17(1):128. doi:10.1186/s13059-016-0994-0
49. Tokheim C, Bhattacharya R, Niknafs N, et al. Exome-Scale Discovery of Hotspot Mutation Regions in Human Cancer Using 3D Protein Structure. *Cancer Res*. 2016;76(13):3719-3731. doi:10.1158/0008-5472.CAN-15-3190
50. Martínez-Jiménez F, Muiños F, López-Arribillaga E, Lopez-Bigas N, Gonzalez-Perez A. Systematic analysis of alterations in the ubiquitin proteolysis system reveals its contribution to driver mutations in cancer. *Nat Cancer*. 2020;1(1):122-135. doi:10.1038/s43018-019-0001-2

51. Weghorn D, Sunyaev S. Bayesian inference of negative and positive selection in human cancers. *Nat Genet*. 2017;49(12):1785-1788. doi:10.1038/ng.3987
52. Dietlein F, Weghorn D, Taylor-Weiner A, et al. Identification of cancer driver genes based on nucleotide context. *Nat Genet*. 2020;52(2):208-218. doi:10.1038/s41588-019-0572-y
53. Martincorena I, Raine KM, Gerstung M, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*. 2017;171(5):1029-1041.e21. doi:10.1016/j.cell.2017.09.042
54. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer*. 2018;18(11):696-705. doi:10.1038/s41568-018-0060-1
55. Jiang S. Tet2 at the interface between cancer and immunity. *Commun Biol*. 2020;3(1):1-6. doi:10.1038/s42003-020-01391-5
56. Weissmann S, Alpermann T, Grossmann V, et al. Landscape of TET2 mutations in acute myeloid leukemia. *Leukemia*. 2012;26(5):934-942. doi:10.1038/leu.2011.326
57. Feng Y, Li X, Cassady K, Zou Z, Zhang X. TET2 Function in Hematopoietic Malignancies, Immune Regulation, and DNA Repair. *Front Oncol*. 2019;9. Accessed May 17, 2023. https://www.frontiersin.org/articles/10.3389/fonc.2019.00210
58. Zeisig BB, Kulasekararaj AG, Mufti GJ, Eric So CW. SnapShot: Acute Myeloid Leukemia. *Cancer Cell*. 2012;22(5):698-698.e1. doi:10.1016/j.ccr.2012.10.017
59. Cheng J, Nguyen TYD, Cygan KJ, et al. MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol*. 2019;20(1):48. doi:10.1186/s13059-019-1653-z
60. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell*. 2019;176(3):535-548.e24. doi:10.1016/j.cell.2018.12.015
61. Davis RE, Ngo VN, Lenz G, et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature*. 2010;463(7277):88-92. doi:10.1038/nature08638
62. Sakatani A, Igawa T, Okatani T, et al. Clinicopathological significance of CD79a expression in classic Hodgkin lymphoma. *J Clin Exp Hematop*. 2020;60(3):78-86. doi:10.3960/jslrt.20010
63. Lenk L, Carlet M, Vogiatzi F, et al. CD79a promotes CNS-infiltration and leukemia engraftment in pediatric B-cell precursor acute lymphoblastic leukemia. *Commun Biol*. 2021;4(1):1-6. doi:10.1038/s42003-020-01591-z
64. Di Fiore R, D'Anneo A, Tesoriere G, Vento R. RB1 in cancer: Different mechanisms of RB1 inactivation and alterations of pRb pathway in tumorigenesis: RB1/pRb INACTIVATION IN CANCER. *J Cell Physiol*. 2013;228(8):1676-1687. doi:10.1002/jcp.24329
65. Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb Perspect Biol*. 2010;2(1):a001008. doi:10.1101/cshperspect.a001008
66. Lindsley RC, Mar BG, Mazzola E, et al. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood*. 2015;125(9):1367-1376. doi:10.1182/blood-2014-11-610543
67. Falini B, Martelli MP, Tiacci E. BRAF V600E mutation in hairy cell leukemia: from bench to bedside. *Blood*. 2016;128(15):1918-1927. doi:10.1182/blood-2016-07-418434
68. Mullighan CG. Mutations of NOTCH1, FBXW7, and prognosis in T-lineage acute lymphoblastic leukemia. *Haematologica*. 2009;94(10):1338-1340. doi:10.3324/haematol.2009.012047
69. Van Vlierberghe P, Palomero T, Khiabanian H, et al. PHF6 mutations in T-cell acute lymphoblastic leukemia. *Nat Genet*. 2010;42(4):338-342. doi:10.1038/ng.542
70. Sakai S, Nakaseko C, Takeuchi M, et al. Circulating soluble LR11/SorLA levels are highly increased and ameliorated by chemotherapy in acute leukemias. *Clin Chim Acta*. 2012;413(19):1542-1548. doi:10.1016/j.cca.2012.06.025

71. Peramangalam PS, Surapally S, Zheng S, et al. MYCN Regulates Cell Survival Via EIF4G1 in Inv(16) Acute Myeloid Leukemia. *Blood*. 2022;140(Supplement 1):9117-9118. doi:10.1182/blood-2022-170112
72. Chen X, Chen X, Huang Y, Lin J, Wu Y, Chen Y. TCP1 increases drug resistance in acute myeloid leukemia by suppressing autophagy via activating AKT/mTOR signaling. *Cell Death Dis*. 2021;12(11):1058. doi:10.1038/s41419-021-04336-w
73. Kahle JJ, Souroullas GP, Yu P, et al. Ataxin1L Is a Regulator of HSC Function Highlighting the Utility of Cross-Tissue Comparisons for Gene Discovery. *PLOS Genet*. 2013;9(3):e1003359. doi:10.1371/journal.pgen.1003359
74. Girardi T, Vicente C, Cools J, De Keersmaecker K. The genetics and molecular biology of T-ALL. *Blood*. 2017;129(9):1113-1123. doi:10.1182/blood-2016-10-706465
75. Riz I, Hawley TS, Johnston H, Hawley RG. Role of TLX1 in T-cell acute lymphoblastic leukemia pathogenesis. *Br J Haematol*. 2009;145(1):140-143. doi:10.1111/j.1365-2141.2008.07556.x
76. Rudat S, Pfaus A, Bühler C, et al. The RET Receptor Tyrosine Kinase Promotes Acute Myeloid Leukemia through Protection of FLT3-ITD Mutants from Autophagic Degradation. *Blood*. 2016;128(22):2849. doi:10.1182/blood.V128.22.2849.2849
77. McNeer NA, Philip J, Geiger H, et al. Genetic mechanisms of primary chemotherapy resistance in pediatric acute myeloid leukemia. *Leukemia*. 2019;33(8):1934-1943. doi:10.1038/s41375-019-0402-3
78. Fu JF, Shih LY, Yen TH. HOXA11 plays critical roles in disease progression and response to cytarabine in AML. *Oncol Rep*. 2021;46(1):150. doi:10.3892/or.2021.8101
79. Príncipe C, Dionísio de Sousa IJ, Prazeres H, Soares P, Lima RT. LRP1B: A Giant Lost in Cancer Translation. *Pharmaceuticals*. 2021;14(9):836. doi:10.3390/ph14090836
80. Weinstein JN, Collisson EA, Mills GB, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45(10):1113-1120. doi:10.1038/ng.2764
81. Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*. 2018;173(2):371-385.e18. doi:10.1016/j.cell.2018.02.060
82. SH S, E C, NL H, et al. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. Accessed June 26, 2023. https://publications.iarc.fr/Book-And-Report-Series/Who-Classification-Of-Tumours/WHO-Classification-Of-Tumours-Of-Haematopoietic-And-Lymphoid-Tissues-2017
83. Strickland DK, Gonias SL, Argraves WS. Diverse roles for the LDL receptor family. *Trends Endocrinol Metab*. 2002;13(2):66-74. doi:10.1016/S1043-2760(01)00526-4
84. May P, Woldt E, Matz RL, Boucher P. The LDL receptor‐related protein (LRP) family: An old family of proteins with new physiological functions. *Ann Med*. 2007;39(3):219-228. doi:10.1080/07853890701214881
85. Dieckmann M, Dietrich MF, Herz J. Lipoprotein receptors – an evolutionarily ancient multifunctional receptor family. 2010;391(11):1341-1363. doi:10.1515/bc.2010.129
86. Beer AG, Zenzmaier C, Schreinlechner M, et al. Expression of a recombinant full-length LRP1B receptor in human non-small cell lung cancer cells confirms the postulated growth-suppressing function of this large LDL receptor family member. *Oncotarget*. 2016;7(42):68721-68733. doi:10.18632/oncotarget.11897
87. Haas J, Beer AG, Widschwendter P, et al. LRP1b shows restricted expression in human tissues and binds to several extracellular ligands, including fibrinogen and apoE – carrying lipoproteins. *Atherosclerosis*. 2011;216(2):342-347. doi:10.1016/j.atherosclerosis.2011.02.030
88. Liu CX, Ranganathan S, Robinson S, Strickland DK. γ-Secretase-mediated Release of the Low Density Lipoprotein Receptor-related Protein 1B Intracellular Domain Suppresses Anchorage-independent Growth of Neuroglioma Cells*. *J Biol Chem*. 2007;282(10):7504-7511. doi:10.1074/jbc.M608088200

89. Brown LC, Tucker MD, Sedhom R, et al. LRP1B mutations are associated with favorable outcomes to immune checkpoint inhibitors across multiple cancer types. *J Immunother Cancer*. 2021;9(3):e001792. doi:10.1136/jitc-2020-001792
90. Cowin PA, George J, Fereday S, et al. LRP1B Deletion in High-Grade Serous Ovarian Cancers Is Associated with Acquired Chemotherapy Resistance to Liposomal Doxorubicin. *Cancer Res*. 2012;72(16):4060-4073. doi:10.1158/0008-5472.CAN-12-0203
91. Liu F, Hou W, Liang J, Zhu L, Luo C. LRP1B mutation: a novel independent prognostic factor and a predictive tumor mutation burden in hepatocellular carcinoma. *J Cancer*. 2021;12(13):4039-4048. doi:10.7150/jca.53124
92. Príncipe C, Dionísio de Sousa IJ, Prazeres H, Soares P, Lima RT. LRP1B: A Giant Lost in Cancer Translation. *Pharmaceuticals*. 2021;14(9):836. doi:10.3390/ph14090836