



Pathophysiologic and clinical implications of molecular profiles resultant from deletion 5q

Vera Adema,^a Laura Palomo,^b Wencke Walter,^c Mar Mallo,^b Stephan Hutter,^c Thomas La Framboise,^d Leonor Arenillas,^e Manja Meggendorfer,^c Tomas Radivoyevitch,^a Blanca Xicoy,^f Andrea Pellagatti,^g Claudia Haferlach,^c Jacqueline Boultonwood,^g Wolfgang Kern,^c Valeria Visconte,^a Mikkael Sekeres,^h John Barnard,ⁱ Torsten Haferlach,^c Francesc Solé,^{b,1} and Jaroslaw P. Maciejewski^{a,1*}

^aDepartment of Translational Hematology and Oncology Research, Lerner Research Institute Cleveland Clinic, Taussig Cancer Institute, 9500 Euclid Avenue, Cleveland, OH 44195, USA

^bMyelodysplastic Syndrome Research Group, Josep Carreras Leukaemia Research Institute, Institut Català d'Oncologia-Hospital Germans Trias i Pujol, Universitat Autònoma de Barcelona, Badalona, Spain

^cMLL Munich Leukemia Laboratory, Munich, Germany

^dDepartment of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, USA

^eLaboratori de Citologia Hematològica, Servei de Patologia, Hospital del Mar and GRETNHE, Cancer Research Program, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

^fHematology Service, Institut Català d'Oncologia (ICO)-Hospital Germans Trias i Pujol, Institut de Recerca Contra la Leucèmia Josep Carreras, Universitat Autònoma de Barcelona, Badalona, Spain

^gBlood Cancer UK Molecular Haematology Unit, Nuffield Division of Clinical Laboratory Sciences, Radcliffe Department of Medicine, University of Oxford and Oxford BRC Haematology Theme, Oxford, United Kingdom

^hLeukemia Program, Department of Hematology and Medical Oncology, Cleveland Clinic, Cleveland Clinic Taussig Cancer Institute, Cleveland, OH, USA

ⁱDepartment of Quantitative Health Sciences, Cleveland Clinic, Lerner Research Institute, Cleveland, OH, USA

Summary

Background Haploinsufficiency (HI) resulting from deletion of the long arm of chromosome 5 [del(5q)] and the accompanied loss of heterozygosity are likely key pathogenic factors in del(5q) myeloid neoplasia (MN) although the consequences of del(5q) have not been yet clarified.

Methods Here, we explored mutations, gene expression and clinical phenotypes of 388 del(5q) vs. 841 diploid cases with MN [82% myelodysplastic syndromes (MDS)].

Findings Del(5q) resulted as founder (better prognosis) or secondary hit (preceded by *TP53* mutations). Using Bayesian prediction analyses on 57 HI marker genes we established the minimal del(5q) gene signature that distinguishes del(5q) from diploid cases. Clusters of diploid cases mimicking the del(5q) signature support the overall importance of del(5q) genes in the pathogenesis of MDS in general. Sub-clusters within del(5q) patients pointed towards the inherent inpatient heterogeneity of HI genes.

Interpretation The underlying clonal expansion drive results from a balance between the “HI-driver” genes (e.g., *CSNK1A1*, *CTNNA1*, *TCERG1*) and the proapoptotic “HI-anti-drivers” (e.g., *RPS14*, *PURA*, *SIL1*). The residual essential clonal expansion drive allows for selection of accelerator mutations such as *TP53* (denominating poor) and *CSNK1A1* mutations (with a better prognosis) which overcome pro-apoptotic genes (e.g., *p21*, *BAD*, *BAX*), resulting in a clonal expansion. In summary, we describe the complete picture of del(5q) MN identifying the crucial genes, gene clusters and clonal hierarchy dictating the clinical course of del(5q) patients.

Funding Torsten Haferlach Leukemia Diagnostics Foundation. US National Institute of Health (NIH) grants R35 HL135795, R01HL123904, R01 HL118281, R01 HL128425, R01 HL132071, and a grant from Edward P. Evans Foundation.

Copyright © 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: Myelodysplastic syndromes; 5q deletion; Haploinsufficiency; *TP53*; *CSNK1A1*

*Corresponding author.

E-mail address: maciejj@ccf.org (J.P. Maciejewski).

¹ These authors contributed equally to this work.

eBioMedicine 2022;80:
104059

Published online xxx
<https://doi.org/10.1016/j.ebiom.2022.104059>

Research in context

Evidence before this study

The prevailing theory in del(5q) is that haploinsufficiency (HI) stemming from deletion and not simply loss of heterozygosity (LOH) is the culprit in clonal evolution. To date no haploinsufficient gene has been found to be the pivotal leukaemogenic factor conveying growth advantage, but various other genes have been found to be important for phenotypic features or for propensity to acquire subsequent specific lesions. This study represents the most comprehensive compendium of genomic and RNA sequencing data of patients with MDS carrying del(5q). Unlike previous studies, we based our analysis on an innovative copy number adjustment that mitigates the clonality effect on the haploinsufficient genes on 5q.

Added value of this study

Our study based on a large cohort, has defined distinct expression signatures of molecular subtypes of del(5q) according to the position in the clonal hierarchy and correcting previous assumptions that del(5q) was the ultimate ancestral lesion at the level of a stem cell. We have also provided a minimal haploinsufficient (HI) expression signature of del(5q) including most indicative genes and their pathophysiologic role. Furthermore, we defined HI-driver (clonal expansion) and HI-anti-drivers (brakes, proapoptotic) genes and their complex interactions in del(5q) neoplasia affecting the clinical course. Clonal drive overcomes proapoptotic pressure (HI-anti-drivers) through escape mechanisms promoted by accelerator events (e.g., *CSNK1A1*, *TP53* mutations).

Implications of all the available evidence

Our analysis of a compendium of del(5q) genomics confirms some of the previous findings, revises previous assumptions, discovers new haploinsufficient candidate genes and precisely circumscribes molecular relationships (e.g., del(5q) with *TP53* mutations) and generates a minimalistic expression signature of del(5q).

Introduction

Deletions in the long arm of chromosome 5 [del(5q)] can be found in different myeloid neoplasia (MN) and are the most common karyotypic alterations in myelodysplastic syndromes (MDS). Del(5q) can be found alone or with other cytogenetic abnormalities and is associated with specific morphologic features.¹ However, del(5q) is the only cytogenetic abnormality used to define a phenotypically distinct WHO MDS subtype,² originally described by Van den Berghe and colleagues in 1974.³ Del(5q) encompasses two commonly deleted regions (CDRs). The centromeric CDR on 5q31.2-5q31.3

(CDR-1)⁴ is associated with a higher-risk MDS or acute myeloid leukaemia (AML). The distal CDR spanning 1.5Mb on 5q32-q33.2 (CDR-2)⁵ is present in the classical 5q- Syndrome and conveys better prognosis.

The clinical del(5q) phenotype may also be affected by the actual boundaries of the deletion,⁶ along with other chromosomal defects and somatic mutations co-existing in a hierarchical relationship with del(5q).⁷ Significant progress in the understanding of the disease mechanisms resulting from del(5q) has been made. However, it has not been entirely clarified which haploinsufficient (HI) genes on 5q are associated with the disease development and malignant progression.

Here, we examine a large cohort of del(5q) patients to clarify the molecular relationships of these lesions with other somatic defects and determine their impact on the resultant expression patterns and the morphologic and clinical phenotypes.

Methods

Patients and samples: A total of 400 samples (388 patients and 12 follow-up samples) with MNs (myeloid neoplasms) with del(5q) (Table 1) determined by conventional G-banding cytogenetics and fluorescence *in situ* hybridization (FISH) were included in the study. A cohort of patients with MN diploid status for 5q [$n = 844$ samples (840 patients and four follow-up samples)] were also collected for comparison purposes. All samples included were profiled for gene sequencing and cytogenetics was available for all.

Ethics Information: All samples were collected with previous written informed consent in accordance with institutional ethics committee approvals following the revised Declaration of Helsinki. A total of 51 samples were collected at the Josep Carreras Leukaemia Research Institute (2010/4102/I) in between 2009 and 2014 on behalf of the MDS Spanish group, 185 samples were collected at the Cleveland Clinic (IRB-5024) in between 2002 and 2017, and 1,008 samples were collected at the MLL Munich Leukemia laboratory in between 2005 and 2018. We think our cohort is indeed representative because: (i) cohorts from 3 centres of excellence were collected in the same time period and all the diagnostic criteria and data collected were synchronized; (ii) our sample size is large enough to accommodate for the biologic diversity, including rare subtypes and; (iii) is not hampered by a single institutional bias.

Next-Generation Sequencing (NGS): Samples underwent targeted deep sequencing (tumour sample; $n = 141$), whole exome sequencing (WES; tumour and germline sample; $n = 95$) or whole genome sequencing (WGS; tumour sample; $n = 1,008$). To combine the data from the three platforms (Table S9) we selected 181 genes (on the basis of their implication in the

del(5q) cohort (n = 388)			
	isolated del(5q)	del(5q) + 1	del(5q) in CK
n (%)	188 (49)	36 (9)	164 (42)
Female, n (%)	132 (70)	23 (64)	71 (43)
Age, median years (min-max)	73 (35–94)	73 (42–88)	73 (24–89)
Classification, n (%)			
<i>Low Risk</i>			
MDS low risk	154 (82)	25 (69)	37 (22)
CMML-1	1 (0.5)	1 (3)	0
<i>High risk</i>			
MDS high risk	20 (11)	5 (14)	78 (48)
CMML-2	0	0	0
sAML	10 (5)	4 (11)	30 (18)
pAML	3 (1.5)	1 (3)	11 (7)
MDS/MPN-U	0	0	8 (5)
Blood counts (%)			
Anaemia (<10g/dL)	117 (69)	19 (58)	106 (68)
Thrombocytopenia (<100 × 10 ⁹ /L)	17 (10)	12 (36)	113 (72)
Thrombocytosis (>450 × 10 ⁹ /L)	24 (14)	2 (6)	2 (1)
Neutropenia (<1.8 × 10 ⁹ /L)	51 (46)	17 (55)	88 (72)
% BM blasts, median (min-max)	3 (0–76)	3 (0–45)	8 (0–87)

Table 1: Clinical features of del(5q) cohort.

CK: Complex Karyotype: ≥ 3 cytogenetic alterations; Low risk: MDS low-risk: MDS-SLD (MDS with single lineage dysplasia), MDS-RS (MDS with ring sideroblasts), MDS-MLD (MDS with multilineage dysplasia), MDS del(5q) (MDS with isolated del(5q)), MDS-U (unclassifiable MDS); CMML-1 (chronic myelomonocytic leukemia-1); High risk: MDS high-risk: MDS-EB-1 (MDS with excess blast-1), MDS-EB-2 (MDS with excess blast-2); CMML-2 (chronic myelomonocytic leukemia-2); sAML (secondary AML); pAML (primary AML); MDS/MPN-U (myelodysplastic/myeloproliferative neoplasm unclassifiable). BM: bone marrow. (CBC data available for: iso-del(5q): anemia: n = 169; thrombocytopenia: n = 169; thrombocytosis n = 169; neutropenia: n = 111; del(5q) +1: anemia: n = 33; thrombocytopenia: n = 33; thrombocytosis n = 33; neutropenia: n = 31; del(5q) in CK: anemia: n = 156; thrombocytopenia: n = 156; thrombocytosis n = 156; neutropenia: n = 122).

pathogenesis of MN; Table S2). After variant calling, all samples were analysed together based on a strict in-house variant filtering process in line with current scientific standards. There was no institutional biased in terms of distribution of somatic mutations. RNA-sequencing (RNA-seq) was performed only on samples collected at the MLL Munich Leukemia laboratory.

Clonal hierarchy: To determine dominant/ancestral and sub-clonal/secondary hits and del(5q) for each patient: *Copy Number Variation from WES data* was calculated based on the allelic imbalance of informative heterozygous SNPs within the 5q deleted region in WES germ line samples. *Copy Number Variation analysis from WGS data:* Copy number variant (CNV) calls were created using GATK4 (version 4.0.2.1), following the “best practices” somatic CNV workflow (<https://software.broadinstitute.org/gatk/best-practices/>).

RNA-Seq: Total RNA from whole bone marrow (BM) samples was extracted. RNA-Seq libraries were constructed from ribosomal RNA depleted RNA using Tru-Seq Total Stranded RNA kit (Illumina) and sequenced on a NovaSeq 6000 platform.

Haploinsufficiency expression analysis: RNA-Seq results from 995 patients (170 del(5q) and 825 diploid patients) with MN were analysed. Patients were further classified according to the karyotype, BM blast % and *TP53* mutational status. After data processing (supplemental methods), a total of 148 del(5q) and 752 diploid patients were included in the analysis. Gene expression was measured in counts per million (CPM), assigned to each gene (see supplemental methods). **Adjustments for del(5q) clonality:** Linear models were fit per gene using the R package limma⁸ with $\log_2(\text{CPM} + 0.1)$ as the response variable and del(5q) CNV clonality (centred at 50% clonality), del(5q) status (isolated, +1, CK), diploid karyotype (normal, 1+, 2+, CK), *TP53* mutational status (wild type, mutant) and blast risk [BM blasts %: low (<5%), high (>5%)] as covariates.

Del(5q) haploinsufficient signature: Sparse Prediction Modeling: models for predicting del(5q) vs. diploid status were built using a Bayesian sparse logistic regression approach as implemented in the R package HTLR.⁹ The htlr function from HTLR was used to train the prediction models with the default parameters (t prior),

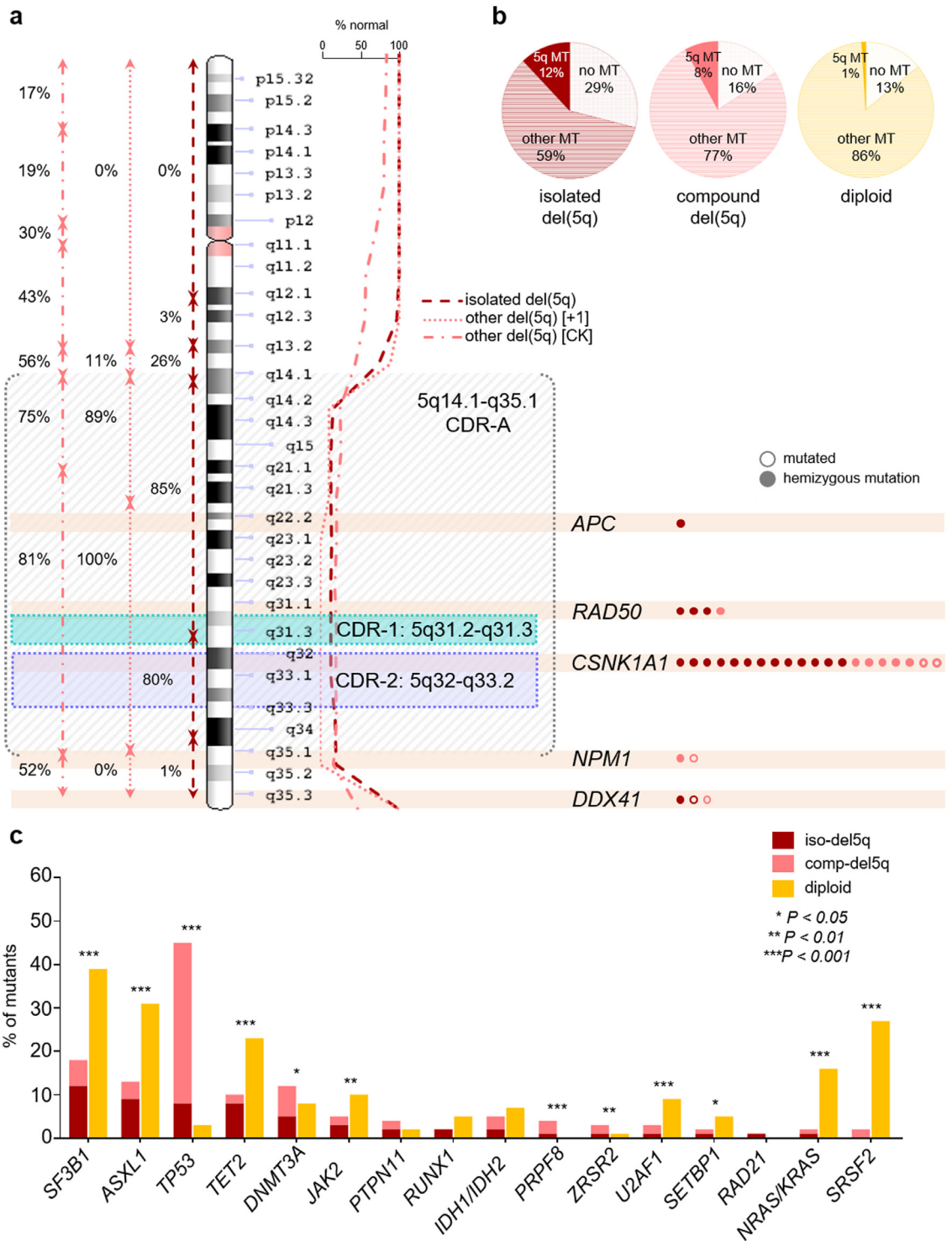


Figure 1. Cyto-genetic and molecular characterization of 5q: (a) CNV analysis based on the cytogenetic break points was used to determine the most commonly deleted region in del(5q).

(b) Mutation distribution in isolated del(5q), compound del(5q), and diploid patients.

(c) Percentage of mutant patients in the most common mutated genes (mutations on chromosome 5 are not included).

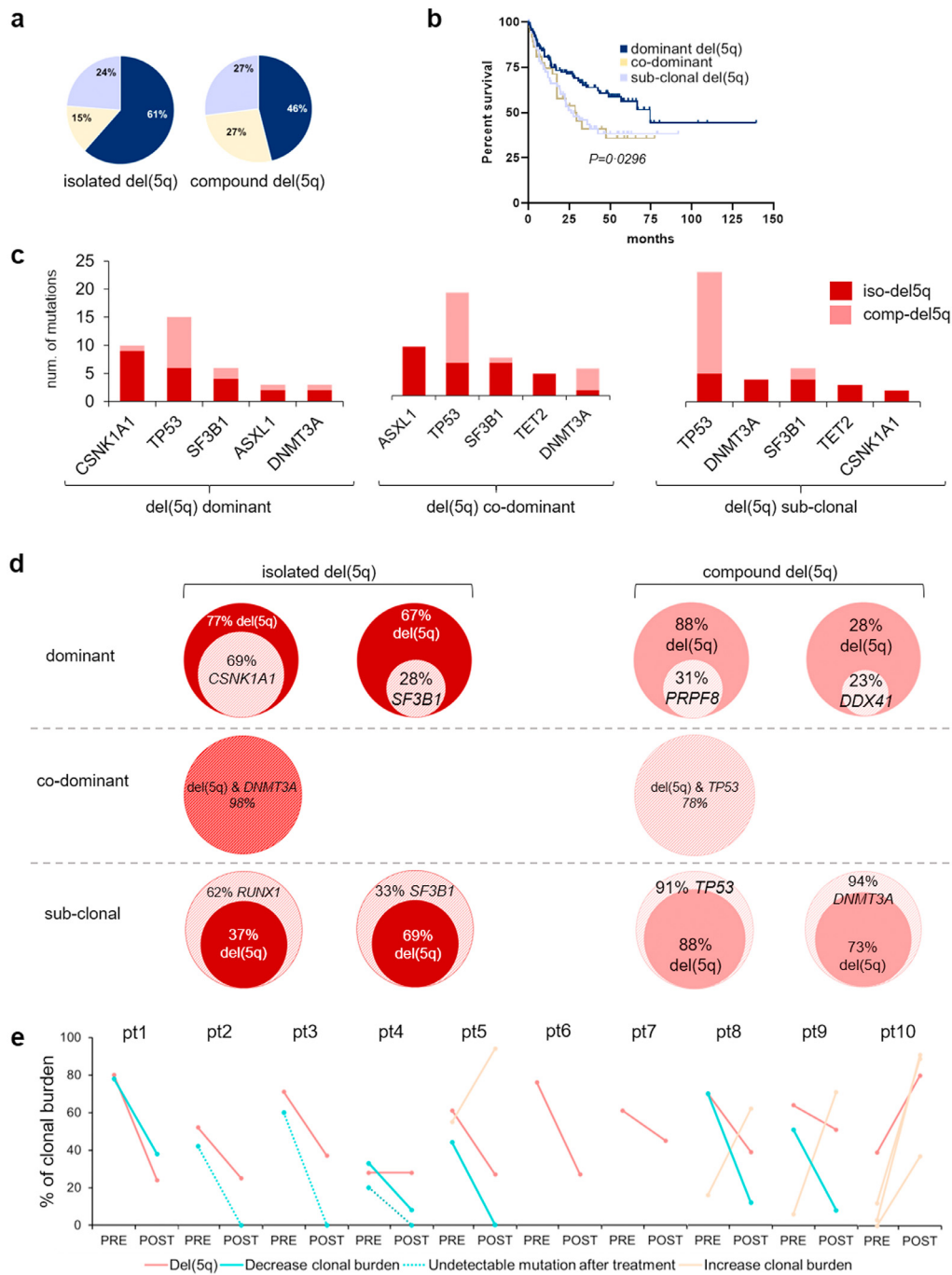


Figure 2. Clonal architecture and subclonal hierarchy: reconstruction of the clonal hierarchy using an allelic imbalance method for WES samples (with available paired sample) and by CNV analyses for WGS, were compared to the VAF of most common mutations adjusted by zygosity and copy number.

(a) Distribution of ancestral (navy blue), co-dominant (yellow) or subclonal (light purple) del(5q) in isolated del(5q) and compound del(5q).

(b) Kaplan-Meier curves for ancestral, secondary and co-dominant del(5q) event with significant differences (Log rank test).

(c) Mutational distribution of the most representative mutations in dominant, co-dominant or subclonal del(5q) in patients with isolated del(5q) or compound del(5q).

(d) Exemplary cases of clonal architecture of isolated del(5q) patients (left) and in compound del(5q) patients (right).

(e) Percentage of clonal burden of del(5q) (red), increased subclonal mutations (yellow), and decreased subclonal mutations (blue). Lines connect paired samples. Dotted lines indicate undetectable mutation after treatment.

except for initial states, which were chosen using the bias-corrected Bayesian classification approach option. Final gene predictors were those having non-zero model coefficients (calculated using the HTLR `nzero_idx` function). Models were trained using MHI(50%) and MHI(25%) data with 900 samples (148 del(5q) samples) and expression from 57 del(5q) genes. Class predictions were made using the posterior median of the regression coefficients. Prediction error was assessed using ten-fold cross validation and feature selection stability was assessed by counting the number of times the feature was selected across the ten fits.

Statistical analyses: Statistical analyses were performed on GraphPad Prism (version 8.0.2) and R statistical software packages (version 3.6.0). Chi-square or Fisher's exact test were performed for categorical variables, while Mann-Whitney and Wilcoxon tests were used for pairwise continuous variables. The overall survival (OS) for diagnosis samples was defined from diagnosis to death or last follow-up. All deaths, whether or not related to MDS, were considered as the endpoint of the follow-up interval. Kaplan-Meier method was used to estimate the OS curves, and log rank test was used for comparisons between groups. All p-values reported are two-sided and considered statistically significant when $P < 0.05$.

Role of the funders: Funding sources had no role in study design, data collection, data analyses or results interpretation, in writing of the manuscript, or in any aspect related to the study.

Results

Clinical and genetic features of del(5q) patients. We included 388 patients with del(5q)-associated (Table 1) and 844 MN without del(5q) (Table S1; Figure 1a-b); RNA-seq results from BM (obtained retrograde flushing of discarded de-identified BM harvest kits) of 64 healthy individuals [median age 59 (range, 26–85)] were included as reference. Patients were grouped according to del(5q) status: isolated del(5q) (iso-del5q) and compound del(5q) (comp-del5q): including del(5q) plus any other alteration (+1-del5q) or in the context of complex karyotype (CK-del5q) (Figure S1a-b). To delineate the boundaries of the del(5q), we conducted copy number variant (CNV) analyses using conventional cytogenetics. The CDRs in our del(5q) cohort (CDR-A; Figure 1a) included CDR-1 and CDR-2.^{4,5} The commonly retained regions⁶ (CRR) centromeric (p12-q13) and telomeric (q35.1-q35.3) were mostly co-deleted in CK-del5q (36% and 52% of patients, respectively; Figure 1a). NGS analysis for 181 selected genes (Table S2) yielded 2,251 somatic mutations in 1,034 (83%) cases (Table S3). While a total of 71% of patients with iso-del5q carried \geq one mutation (12% on 5q), 85% of patients with comp-del5q harboured \geq one mutation (8% on 5q; $P = 0.001$; [Fisher's exact test]) but were less mutated than

patients without 5q ($P = 0.0001$; [Fisher's exact test]) (Figures 1b; S1c).

For the mutational profile, we selected the most commonly mutated genes in MNs ($n = 181$). Comp-del5q showed a higher frequency of *TP53* mutations vs. iso-del5q and mutant diploid cases ($P < 0.0001$; [Chi²-test]), and a significantly lower frequency of hits in *SF3B1*, *TET2*, and *ASXL1*, vs. iso-del5q (Figure 1c). When we focused on mutations on 5q, *CSNK1A1* (5q32) was the most commonly mutated gene in iso-del5q cases. Canonical *CSNK1A1*^{E98} ($n = 13$) and *CSNK1A1*^{D140V} ($n = 2$) were identified.^{10,11} Neither canonical *DDX41* (5q35.3) frame-shift germ line, nor somatic missense ATP-binding domain mutations were found within deleted areas and the only three *DDX41* mutations in del(5q) were all heterozygous.^{12,13} *RAD50* (5q31.1; $n = 4$) was found in a hemizygous configuration, whereas *NPM1* (5q35.1; $n = 2$) mutations were always heterozygous and associated with comp-del5q in secondary-AML (s-AML).

Subclonal hierarchy and dynamics. Zygosity- and copy number-adjusted VAF of somatic mutations were used. To determine the clonal architecture, the CNV analyses were performed following allelic imbalance calculations for heterozygous SNPs located in the CDRs of del(5q)¹⁴ (Figure S2a-b). Del(5q) occurs in diverse configurations (Figure S2c-e) with coinciding somatic mutations: (i) as a dominant event, more frequent in iso-del5q (61%) than in comp-del5q (46%; $P = 0.03$; [Fisher's exact test]); (ii) as a sub-clonal event with similar frequencies in comp-del5q (27%) and iso-del5q (24%); or (iii) as a co-dominant event with unresolved rank of somatic alterations/del(5q), more commonly in comp-del5q (27%) than in iso-del5q (15%; $P = 0.04$; Figure 2a). Patients with a dominant del(5q) had better overall survival (OS; $P = 0.0296$; [Kaplan-Meier, log rank test]) than those with del(5q) in a co-dominant or subclonal configuration (Figure 2b). Del(5q) occurring in a dominant configuration coincided with secondary *CSNK1A1* mutations (ten%) in iso-del5q, and *TP53* (26%) in comp-del5q. However, when del(5q) was a sub-clonal event, it was often preceded by dominant *ASXL1* mutations (26%) in iso-del5q and dominant *TP53* (56%) in comp-del5q (Figure 2c). Figure 2d shows exemplary cases with ancestral iso-del5q and sub-clonal *CSNK1A1* and *SF3B1*. When a clear clonal hierarchy could not be mathematically resolved, del(5q) co-ranked with *TP53*, *ASXL1*, and *SF3B1* mutations in iso-del5q and *TP53* and *DNMT3A* in comp-del5q.

Sequential BM and granulocyte fraction samples were available from 7 and 3 patients, respectively. Nine patients received at least 6 cycles of LEN (LEN cycles: pt1:8, pt2: 9; pt3: 6, pt4:8; pt5: 9, pt6: 13; pt7:10; pt8:7; pt9:32), and pt10 was during disease progression. Most consistently, del(5q) decreased in response to LEN (pt1, pt2, pt3, pt4, pt6, pt7). Concurrent mutations showed diverse dynamics; (i) decrease/ disappearance of mutations (e.g., *DNMT3A*,

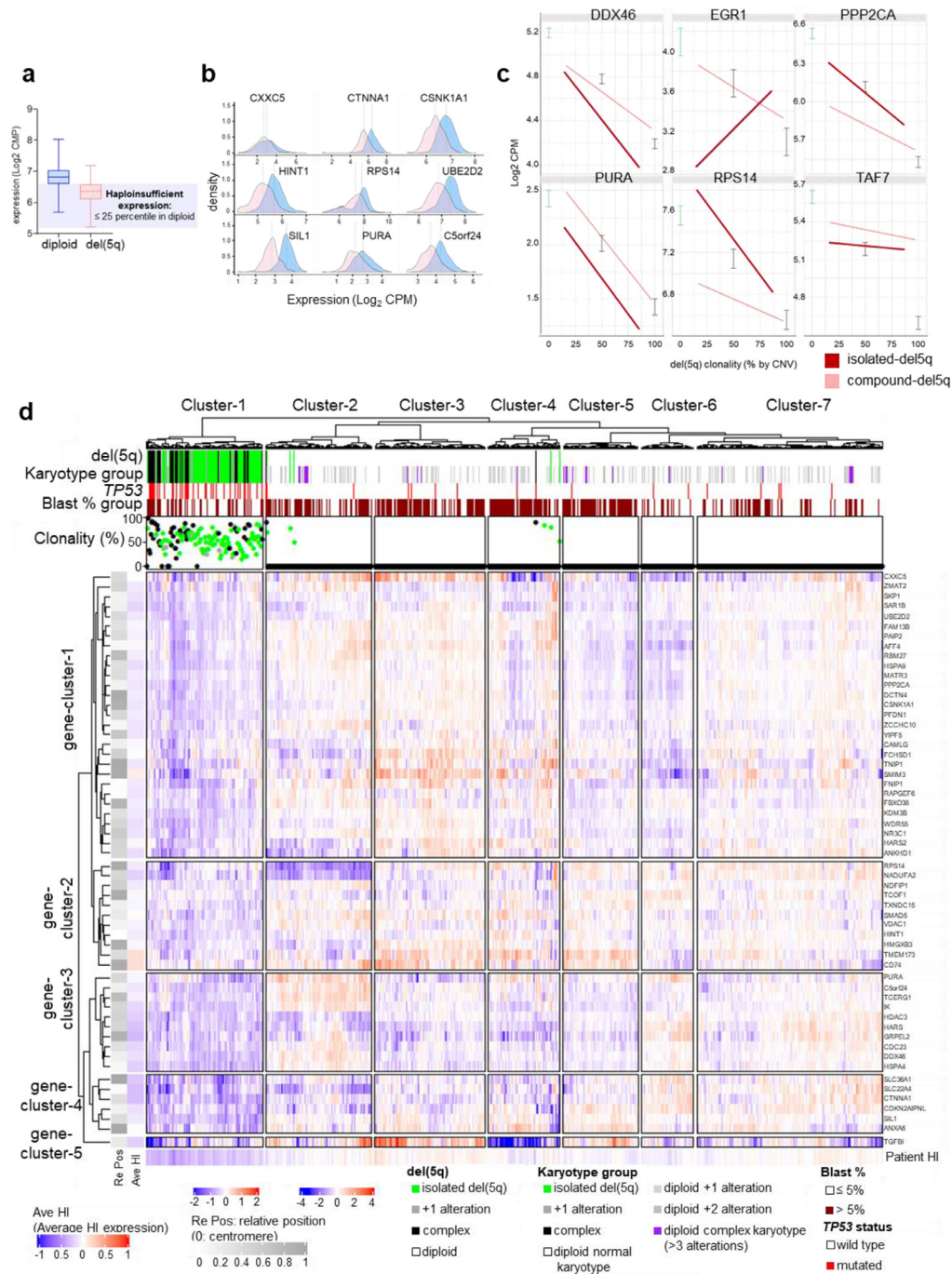


Figure 3. Haploinsufficiency (HI) analysis of selected genes on 5q. (a) Box plot showing the definition of HI expression defined when the expression in del(5q) was lower than the percentile 25th of the diploid expression level.

(b) Gene expression (Log₂ CPM) in del(5q) cases (pink) and diploid cases (blue). Dashed lines show median values.

(c) Exemplary cases of different correlation types between expression and clonality in isolated del(5q) and compound del(5q). Green error bar shows the expression of the given gene in diploid cases. Black error bar shows the expected expression at 50% and 100% clonality of del(5q).

(d) Heatmap using the sparse model-based clustering of del(5q) and diploid patients (see methods), including the 57 HI genes. Karyotype group, TP53 mutational status, BM Blast % group and % of clonality are also depicted.

CSNK1A1, *ASXL1*) paralleling del(5q) contraction (pt1, pt2, pt3, pt4); (ii) expansion of pre-existing mutations (e.g., *PRPF8*, *RUNX1*, *ASXL1*) during LEN treatment, while del(5q) clone decreased (pt5 was responder to LEN while pt8 and pt9 nine progressed despite therapy); (iii) progressing patients were characterized by acquisition of new independent subclones (e.g., pt10; *ASXL1*; Figure 2e).

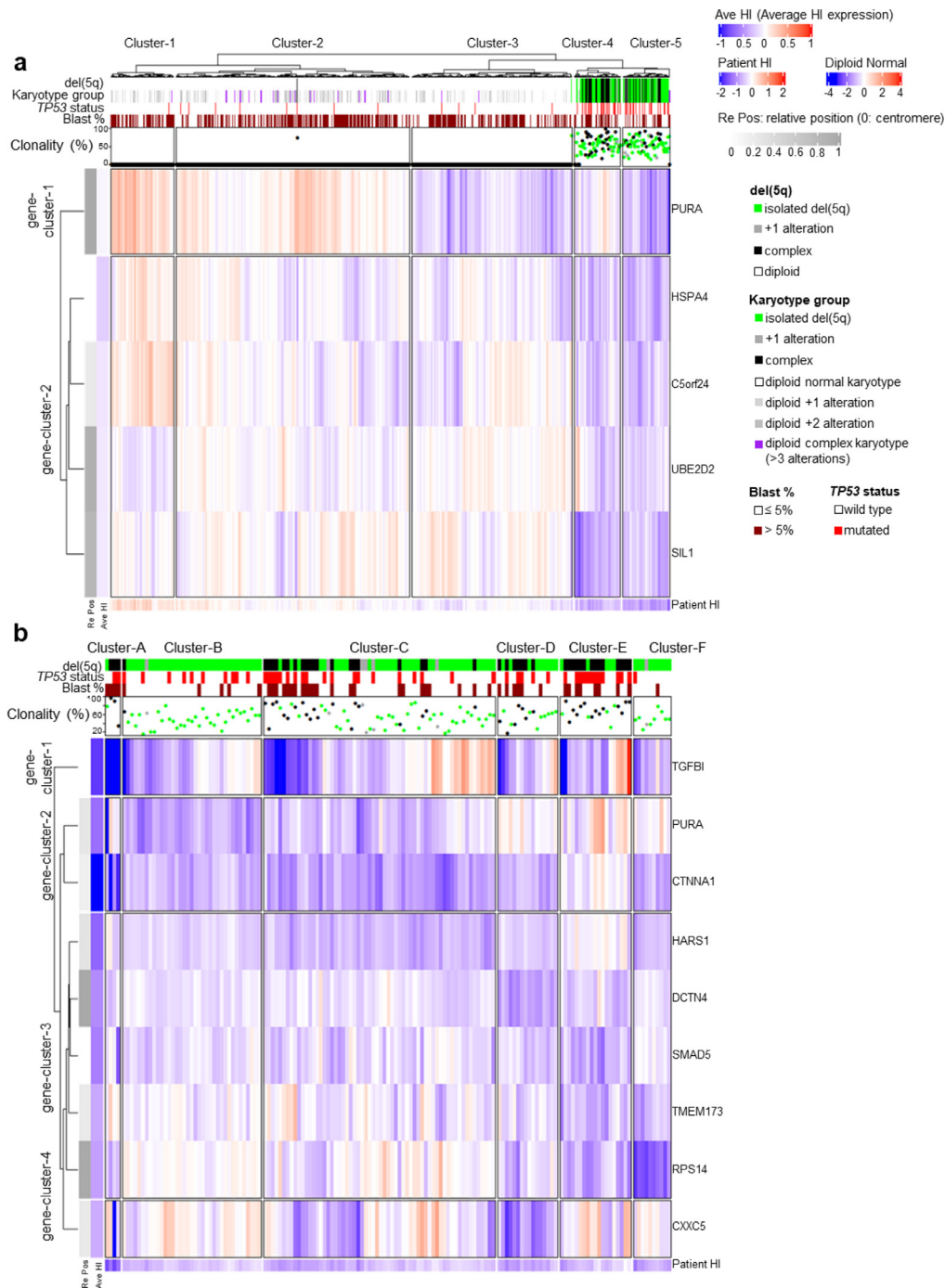
Haploinsufficiency of genes located on 5q. Of 405 genes on 5q, 188 were located within CDR-A (CDR-1: 41, CDR-2: 55 genes) and 128 were deleted across all patients (according to CNV analysis). For dichotomized threshold analysis, we defined mRNA expression to be HI in del(5q) when it was ≤ 25 th percentile of the diploid expression (Figure 3a-b). We noticed a variable landscape of HI expression in some of the classically HI genes, probably due to normal tissue contamination. We then queried the relationship between mRNA and corresponding 5q ploidy. Genes with a higher-than-expected expression, due to the del(5q), may show a partial compensation or deletion of a methylated allele (*PPP2AC*, *RPS14*), while those with a steeper slope are affected by opposing effects (*DDX46*, *PURA*), while some did not correlate with the del(5q) clonal burden (*TAF7*, *EGR1*) (Figure 3c). To identify causative genes, subsequent analyses were restricted to the 57 genes (q23.3-q33.1) showing evidence of HI and negative ploidy slope in isolated del(5q) (Figure S3, Table S4). Some previously implicated genes in the del(5q) region did not follow these rules. Remarkably, decline of expression of *CDC25C*, *DIAPH1*, *SPARC*, *RAD51* and *EGR1* genes either did not correlate with the del(5q) clonal burden or was too variable to be pathognomonic. Other genes were excluded because they were not deleted consistently enough [*APC*: 144/148, *DDX41* one/148, *NPM1* one/148 (deleted/total)] to assume the ubiquitous role in the pathogenesis of del(5q).

Ploidy-adjusted expression value of the 57 selected genes were used for unsupervised clustering to determine differences in the expression levels between del(5q) and diploid patients for genes on 5q (Figure 3d). As expected, del(5q) cases clustered together and showed consistent HI of 5q marker gene expression (Tables S5 and 6). Cluster-one ($n = 146$) included almost all del(5q) cases, except for eight patients (“mis-categorized” to other clusters). It was characterized by low risk MDS (LR-MDS), presence of anaemia /neutropenia and low mutational burden, with *TP53* being the most commonly mutated gene. Diploid cluster-two ($n=133$) featured a normal karyotype, frequent *ASXL1* and *TET2* mutations, and profound down-modulation of *RPS14* and *NDUFA2* mRNA. Clusters-three, -four, -five ($n = 138, 90, 94$, respectively) included most of the high risk MDS (HR-MDS). Cluster-three was enriched for thrombocytopenia and *SRSF2* mutations; cluster-four for anaemia, thrombocytopenia, and *ASXL1* and *SRSF2* mutations. Cluster-five was characterized by pancytopenia and frequent *ASXL1* mutations and CK. Cluster-six

($n = 66$) and -seven ($n = 233$) contained the majority of non-del(5q) LR-MDS. Cluster-six had the highest percentage of abnormal karyotypes. Cluster-seven, the largest cluster, included the majority of normal or non-CK karyotypes and showed a high frequency of *SF3B1* and *DNMT3A* mutations ($n = 29$).

In terms of marker HI genes, Cluster-three exhibited higher *CXXC5* mRNA levels vs. del(5q) defined Cluster-one. Cluster-four was characterized by low expression of *TGFBI*. Cluster-five showed overall differences on the expression pattern in gene Cluster-two (*RPS14*, *NDUFA2*, *SMAD5*) and gene Cluster-four (*SIL1*, *CTNNA1*, among others). The expression pattern of Cluster-7 was most distinct from that of Cluster-one.

Minimal haploinsufficient signature of del(5q) and haploinsufficient patterns within del(5q). Using the 57 HI marker genes, we performed a sparse Bayesian prediction analyses to identify the minimal del(5q) gene signature that could distinguish del(5q) from diploid cases. Using the 50% clonality adjusted data, the minimal gene signature contained *UBE2D2*, *C5orf24*, *PURA*, *SIL1*, and *HSPA4* genes and misclassified only one del(5q) sample (0.1% observed error rate for clustering within 5q deleted samples; Figures 4a; S4a). Even when this five-gene signature was applied to the 25% clonality adjusted data, the error rate was only 6.7%. When we conducted pairwise associations between the 57 genes, we found a striking correlation with distinct groups of highly correlated 5q genes. Even though diploid patients also exhibited some significant correlations, those correlations were stronger and more significant in del(5q). Repeating the signature search using the 25% clonality adjusted data resulted in a seven HI gene signature (*PURA*, *SIL1*, *HSPA4*, *CSNK1A1*, *TCOF1*, *CTNNA1*, and *FAM13B*) with only 11 del(5q) and eight diploid patients misclassified (one% error rate; Figure S4b-c). When we applied model-based sparse sub-clustering within del(5q) patients a nine gene signature was found (*TGFBI*, *CTNNA1*, *PURA*, *CXXC5*, *DCTN4*, *SMAD5*, *HARS*, *TMEM173*, and *RPS14*) to obtain six different expression clusters demonstrating heterogeneity of HI within del(5q) patients (Figure 4b; d, Tables S7 and 8). Cluster comparison showed that cluster-A and cluster-E had the majority of CK but cluster-A showed low mutational burden and lowest expression of *TGFBI*, *CTNNA1*, and *PURA*. Cluster-E showed more 5q as secondary hit, high number of *TP53* mutations ($n = 20$), and HI for *SMAD5*, *RPS14*, and *TMEM173* but less profound HI of *PURA* and *CXXC5*. Cluster-B and cluster-F included the majority of isolated del(5q). Ancestral 5q hits and deeply depressed expression of *TGFBI*, *CTNNA1*, and *PURA* (but not *CXXC5* or *RPS14*) were grouped in cluster-B. Cluster-C pattern was the most common ($n = 62$), with 50% of ancestral del(5q), frequent *TP53* mutations ($n = 23$) and low expression of *CTNNA1*, *PURA*, and *HARS*. Interestingly, *RPS14* mRNA was not uniformly HI in del(5q) patients. By comparison diploid cluster-



two contained patients with a down-modulation of RPS14 but no relationship with p21 levels or TP53 mutations was seen in diploid cases vs. those with del(5q) and low RPS14 expression.

Pathogenic interactions. We identified several subclasses according to their function with regard to clonal advantage: (i) directly or indirectly acting proapoptotic genes (*BAD*, *BAK1*, *BAX*, among others) (Figure 5a) whose HI facilitates subclonal escape via e.g., hypomorphic TP53 mutations or via CSNK1A1 mutations. HI levels of CSNK1A1 may promote apoptosis (of MDM2 dephosphorylation leading to p21 increase). UBE2D2 HI in del(5q) may result in accumulation of nuclear p53¹⁵ contributing to mutational pressure; and (ii) tumour suppressor genes whose HI directly provides selective pressure for del(5q) CSNK1A1, CTNNA1, C5orf24 (*TECRG1*) genes.

We then focused on selected interactions between well-established mutations and the newly defined HI 5q genes. CSNK1A1 mutations were only found in del(5q) patients (12% vs. none in diploid cases); CSNK1A1 and TP53 mutations did not coincide; only 15% were co-mutated with TP53 (two% vice versa, Figure 5b). All del(5q) showed CSNK1A1 HI expression [15% of diploid patients were CSNK1A1 low expressors (le; ≤ 75 th percentile; Figure 5c)]. Diploid CSNK1A1-le also had low expression of RPS14 ($P < 0.001$, [Fisher's exact test]; Figure 5d). Moreover, CSNK1A1 expression did not correlate with TP53 mutations (Figure 5e). Irrespective of CSNK1A1 mutations or wild type expression level, CDKN1A (p21) as a downstream effector of TP53 was generally up regulated in del(5q) except for TP53 mutants showing p21 mRNA levels comparable to those of diploid patients ($P = 0.125$; [Mann-Whitney test]; Figure 5f). Similarly, del(5q) cases with HI RPS14 expression also showed decreased p21 expression ($P < 0.0001$; [Mann-Whitney test]; Figure 5g).

While expression of *CDC25C*, *DIAPH1*, and *RAD50* genes did not correlate with ploidy (FDR for ploidy slope (SL): 0.988; 0.072; 0.307; respectively; Figure 5h), *EGR1* expression showed a positive ploidy slope (FDR SL: 0.211), but those with a relatively lower expression were 75% of TP53 mutants (40/53; $P < 0.0001$; [Chi²-test]) (Figure 5j). *APC*, when involved in the deleted region, was indeed HI ($P < 0.0001$) and correlated with the presence of CSNK1A1 mutants ($n = \text{eight/nine}$) (Figure 5k). When deleted (21/167) *DDX41* were indeed HI; all these patients had a CK and 18/21 of these patients also harboured TP53 mutations (Figure 5l).

Discussion

Del(5q) results in various biologic features due to the HI of individual genes contained in the affected region. We investigated the molecular mechanisms of del(5q) including HI genes and mutations affecting cellular function to elucidate the phenotype-genotype

associations leading to apoptosis/ proliferation/ differentiation and disassociation in MDS.

Using an integrative-molecular approach, we demonstrated: (i) del(5q) patients have similar but less complex mutational landscapes as other MDS subtypes with mutational signatures dominated by exclusively hemizygous CSNK1A1 mutations, along with an increased frequency of TP53 mutations; (ii) a subset of HI genes ($n = 57$) was selected on significant expression reduction with a negative correlation with 5q clonality. Minimal expression signature points towards the presence of marker genes for del(5q) irrespective of their functional importance. For analytic purposes, minimal signature genes may help to design RNA based tests for del(5q) using minimal number of target genes and to identify del(5q) cells in single cell RNA-seq. While reproducibility of minimal signature may be an issue, our study has identified 57 consistently HI genes, which can be added to improve the precision of identification of del(5q); (iii) del(5q) neoplasms result from complex interaction between HI-driver (clonal expansion) and HI-anti-drivers (brakes, proapoptotic) genes; (iv) clonal drive eventually overcomes proapoptotic pressure (HI-anti-drivers) through escape mechanisms promoted by accelerator events; (v) using gene clustering according to expression reduction, we identified subsets of diploid patients with 5q gene repression patterns similar to del(5q) patients; (vi) we determined a minimal HI gene signature of del(5q) involving essential genes. Integration of these findings led to a new and more complete concept of pathogenesis of del(5q).

In contrast to some previous studies postulating that del(5q) is an obligatory primary hit,¹⁶ we demonstrated that del(5q) is not always an initiating event. It may be preceded by other mutations (e.g., mutations in TP53), a succession characterized by an early advanced phenotype. Subclonal TP53 mutations likely represent a later escape mechanism resulting from apoptotic pressure produced by HI of various genes. When del(5q) was dominant, the most common secondary hits in isolated del(5q) were mutations in CSNK1A1^{10,17,18} and for compound del(5q) TP53 mutations. The individual fate of the clinical course of del(5q) is thus influenced by the hierarchy and the succession of molecular events, adding further complexity to del(5q) sub-entity.

For the purpose of this discussion, we focused on the ancestral del(5q) characterizing the classical del(5q) entity (iso-del5q). Among HI genes involved in ancestral del(5q) there must be some HI-drivers which provide the impetus for clonal expansion, thus enabling selection needed to eventually overcome HI-anti-driver genes producing the phenotypic dysplasia and apoptosis of del(5q). Such a pressure leads to acquisition of accelerator events such as CSNK1A1 mutations, TP53 mutations or deletion, and/or monosomy 7.

Putative HI-drivers include the previously "established" del(5q) genes: CSNK1A1,¹⁷ CTNNA1,¹⁹ and

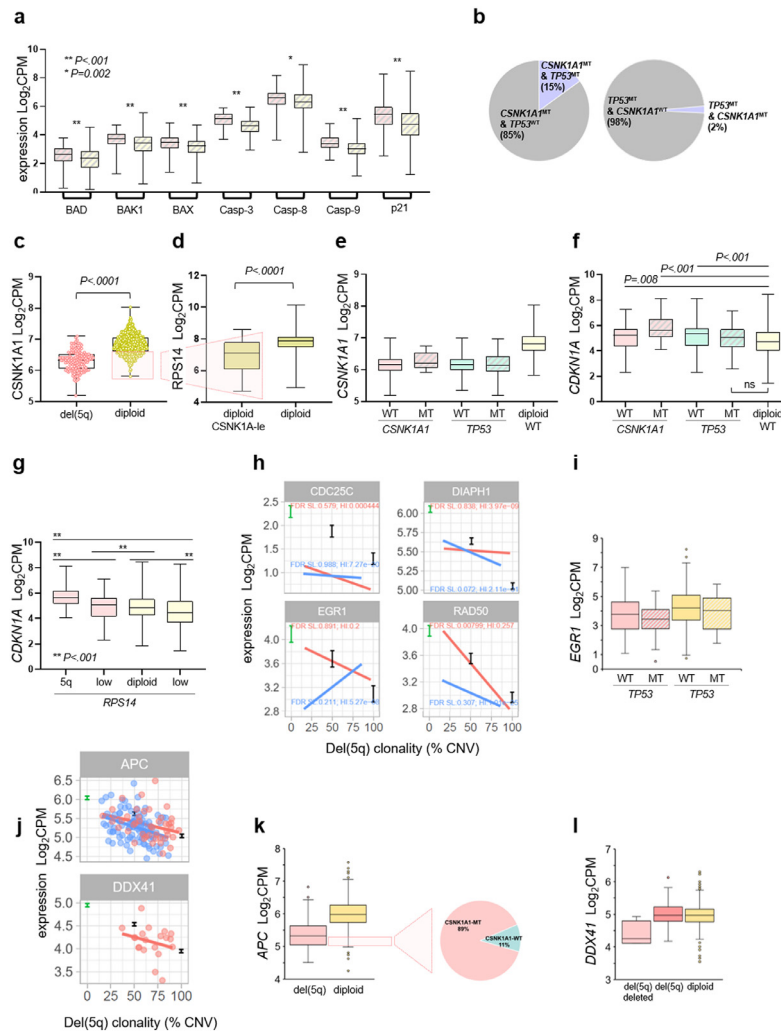


Figure 5. Pathogenic mutational interactions in del(5q): (a) Expression of selected pro-apoptotic genes showing a significant higher expression in del(5q) compared to diploid patients.

(b) Frequency of mutant *CSNK1A1* that are co-mutated with *TP53* (left) and frequency of *TP53* mutants that include *CSNK1A1* mutations (right).

(c) Expression (Log₂ CPM) of *CSNK1A1* in del(5q) patients and diploid patients.

(d) *RPS14* expression (Log₂ CPM) comparison between diploid patients with low expression (le) of *CSNK1A1* and diploid patients with normal expression of *CSNK1A1*.

(e) *CSNK1A1* expression comparison between del(5q) patients WT (solid colour) or MT (grey stripes) for *CSNK1A1*, *TP53*. Diploid WT patient expression is also depicted.

(f) *CDKN1A* expression comparison between del(5q) patients WT (solid colour) or MT (grey stripes) for the *CSNK1A1* and *TP53* genes. Diploid WT patient expression is also depicted.

(g) *CDKN1A* expression in del(5q) vs. del(5q) low *RPS14* expressors and diploid patients vs. diploid low *RPS14* expressors (cluster-2).

(h) Correlation between the percentage of del(5q) (according to CNV analysis) and expression (Log₂ CPM) of excluded genes located on 5q. Green error bar represents the expression of the given gene in diploid patients, *TP53* wild type, and <5% bone marrow blasts. Black error bar represents the estimated expression of the given gene at 50% and 100% of del(5q) clonality. [Isolated del(5q) patients (blue line), compound del(5q) (pink line)].

(i) *EGR1* expression in del(5q) and diploid cases *TP53* mutant (solid colour) or wild type (grey stripes).

(j) Correlation between % of del(5q) (according to CNV analysis) and expression (log₂ CPM). Dots represents expression and clonality of patients without deletion involving *APC* or *DDX41*.

(k) *APC* expression in del(5q) and diploid cases. On the right frequency of *CSNK1A1* mutants and wild type in del(5q) *APC* low expressors.

(l) *DDX41* expression in patients with del(5q) and *DDX41* deleted, del(5q), and diploid cases.

also some new genes such as *C5orf25* (*TCERG1*).²⁰ Thus, del(5q) results in a complex interaction between the HI-anti-drivers which resist the expansion promoted by the HI-driver genes. The acquisition of accelerator hits will shift the balance between the proapoptotic and pro-proliferative functions in favour to the last. Of note is that we included as accelerators *CSNK1A1* gain of function mutations which shift the balance towards pro-proliferative function Wnt/ β -catenin.¹⁸ *TP53* alterations may cooperate with the pro-tumorigenic effects or cytogenetic alterations like monosomy 7 or del(17p) may ultimately lead to a more aggressive phenotype. Finally, *EGR1* upregulation in response to apoptotic stress will be attenuated via *TP53* inactivation,^{21,22} consistent with our results.

HI-anti-driver genes include *e.g.*, *RPS14*, *HSPA4*, *SIL-1*, and *UBE2D2* and all promoted the increased apoptosis in del(5q).²³ Whereas we show that *RPS14* is not uniformly HI, our results are otherwise consistent with earlier reports,^{24,25} as del(5q) cases with *RPS14* deletion did harbour most of the *TP53* mutations. Together with or in the absence of *RPS14*, HI of *HSPA4* and *SIL1* genes may also contribute to cell cycle arrest and apoptotic pressure in del(5q).²⁶ *PURA* is another del(5q) gene which represses *TP53* and *BAX* and thus its HI promotes apoptosis.²⁷ Interestingly, *PURA* can either function as a homodimer (activating the transcription) or as a repressive heterodimer with *PURB* located on chromosome 7p. Simultaneous deletion of *PURA* and *PURB* correlates with an enhanced progression to sAML in MDS.²⁸ Another HI-anti driver gene, *UBE2D2* may also be consequential to del(5q) pathogenesis. Knock-down of *UBE2D2* decreases *TP53* ubiquitination and degradation.¹⁵ Finally, increased expression of *CDKN1A* (*p21*), *caspase-3*, *-8*, and *-9* and *BAX* genes supports notions of high apoptotic pressure in del(5q) cells.

The recently published single cell barcoding strategy by knockout of HI loci of *Csnk1a1*, *Apc*, and *Egr1* showed that only HI of *Csnk1a1* resulted in a clonal expansion as shown in serial transplantations, while HI of *Apc* or *Egr1* did not.²⁹ Our study is compatible with these experiments also assigning a prominent role of *CSNK1A1* in the clonal expansion, despite its duality in function (apoptosis vs. enhancing the Wnt/ β -catenin pathway). The latter pathway can be augmented by the *CSNK1A1* gain of function mutations. A similar effect may be a result of *APC* down-modulation in deleted cases, as previously described.³⁰

Despite considerable effort, our study may suffer from the obvious shortcomings inherent to retrospective study with samples collected at different times. Another limitation is that genomic and expression data were based on bulk DNA and RNA. Further studies with single cell technologies are needed to determine the differences during haematopoiesis of del(5q) and non-del(5q) cells.

In summary, we provide an integrated analysis of the genome of del(5q) patients. We identified key HI genes (drivers vs. anti-drivers) and gene clusters, that together with acquired mutational events and their hierarchy might determine the clinical course and complexity of del(5q). We also provide a substrate for future generation of mouse models for confirmatory and genetics studies to further characterize the HI-drivers and HI-anti-drivers in del(5q).

Contributors

V.A. carried out the experimental/analytic procedures, interpreted the data, and wrote the manuscript. L.P., M. M., and V.V. collected data and performed DNA molecular studies. W.W., M.M., S.H., T.L., T.R., and Jo.B. conducted bioinformatic analysis. Jo.B. established statistical analysis and conducted haploinsufficiency analysis. V.V., L.A., B.X., A.P., J.B., M.S., provided clinical specimens and important insights on the manuscript. T.H., C.H., and W.K. provided molecular data RNA-Seq and WGS and helped to interpret results and edit the manuscript. J.P.M., T.H., and F.S. supervised the study and provided patient specimens and data. V.V. and J.P.M. have verified the underlying data. J.P.M. formulated and edited the manuscript. J.P.M. and F.S. equally contributed. All authors read and approved the final manuscript.

Data sharing statement

All information are provided in the appendix. Requests for additional information can be sent to the corresponding author. RNA-sequencing results have been reported in Hershberger CE et al., *Leukemia* 2021.

Declaration of interests

Dr. Sekeres received consultation fees from BMS/Celgene and Kurome. All the other authors declare no conflicts interests.

Acknowledgements

This work was supported in part by Torsten Haferlach Leukemia Diagnostics Foundation. US National Institute of Health (NIH) grants R35 HL135795, R01HL123904, R01 HL118281, R01 HL128425, R01 HL132071 (J.P. Maciejewski), and a grant from Edward P. Evans Foundation (J.P. Maciejewski). US National Institute of Health (NIH) grants R01 CA217992, R01 LM013067, R21 CA248138 (T. LaFramboise). Blood Cancer UK, grants 13042 and 19004 (A. Pellagatti and J. Boulton). Instituto de Salud Carlos III, Ministerio de Economía y Competitividad PI/14/00013; PI/17/0575 (F. Sole); 2017 SGR288 (GRC) Generalitat de Catalunya (F. Sole); economical support from CERCA Program/

Generalitat de Catalunya, and Fundació Internacional Josep Carreras and “la Caixa” Foundation (F. Sole).

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2022.104059.

References

- 1 Giagounidis AAN, Germing U, Haase S, et al. Clinical, morphological, cytogenetic, and prognostic features of patients with myelodysplastic syndromes and del(5q) including band q31. *Leukemia*. 2004;18(1):113–119.
- 2 Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391–2405.
- 3 Van den Berghe H, Cassiman JJ, David G, Fryns JP, Michaux JL, Sokal G. Distinct haematological disorder with deletion of long arm of no. 5 chromosome. *Nature*. 1974;251(5474):437–438.
- 4 Le Beau MM, Espinosa R, Neuman WL, et al. Cytogenetic and molecular delineation of the smallest commonly deleted region of chromosome 5 in malignant myeloid diseases. *Proc Natl Acad Sci U S A*. 1993;90(12):5484–5488.
- 5 Boultonwood J, Fidler C, Soularue P, et al. Novel genes mapping to the critical region of the 5q- syndrome. *Genomics*. 1997;45(1):88–96.
- 6 Jerez A, Gondek LP, Jankowska AM, et al. Topography, clinical, and genomic correlates of 5q myeloid malignancies revisited. *J Clin Oncol Off J Am Soc Clin Oncol*. 2012;30(12):1343–1349.
- 7 Makishima H, Yoshizato T, Yoshida K, et al. Dynamics of clonal evolution in myelodysplastic syndromes. *Nat Genet*. 2017;49(2):204–212.
- 8 Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
- 9 Li L, Yao W. Fully Bayesian logistic regression with hyper-LASSO priors for high-dimensional feature selection. *J Stat Comput Simul*. 2018;88(14):2827–2851.
- 10 Schneider RK, Adema V, Heck D, et al. Role of casein kinase 1A1 in the biology and targeted therapy of del(5q) MDS. *Cancer Cell*. 2014;26(4):509–520.
- 11 Heuser M, Meggendorfer M, Cruz MMA, et al. Frequency and prognostic impact of casein kinase 1A1 mutations in MDS patients with deletion of chromosome 5q. *Leukemia*. 2015;29(9):1942–1945.
- 12 Polprasert C, Schulze I, Sekeres MA, et al. Inherited and somatic defects in DDX41 in myeloid neoplasms. *Cancer Cell*. 2015;27(5):658–670.
- 13 Maciejewski JP, Padgett RA, Brown AL, Müller-Tidow C. DDX41-related myeloid neoplasia. *Semin Hematol*. 2017;54(2):94–97.
- 14 McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303.
- 15 Saville MK, Sparks A, Xirodimas DP, et al. Regulation of p53 by the ubiquitin-conjugating enzymes UbcH5B/C *in vivo*. *J Biol Chem*. 2004;279(40):42169–42181.
- 16 Mossner M, Jann JC, Wittig J, et al. Mutational hierarchies in myelodysplastic syndromes dynamically adapt and evolve upon therapy response and failure. *Blood*. 2016;128(9):1246–1259.
- 17 Ribezzo F, Snoeren IAM, Ziegler S, et al. Rps14, Csnk1a1 and miRNA145/miRNA146a deficiency cooperate in the clinical phenotype and activation of the innate immune system in the 5q- syndrome. *Leukemia*. 2019;33(7):1759–1772.
- 18 Smith AE, Kulasekararaj AG, Jiang J, et al. CSNK1A1 mutations and isolated del(5q) abnormality in myelodysplastic syndrome: a retrospective mutational analysis. *Lancet Haematol*. 2015;2(5):e212–e221.
- 19 Liu TX, Becker MW, Jelinek J, et al. Chromosome 5q deletion and epigenetic suppression of the gene encoding alpha-catenin (CTNNA1) in myeloid cell transformation. *Nat Med*. 2007;13(1):78–83.
- 20 Montes M, Coiras M, Becerra S, et al. Functional consequences for apoptosis by transcription elongation regulator 1 (TCERG1)-mediated Bcl-x and Fas/CD95 alternative splicing. *PLoS One*. 2015;10(10):e0139812.
- 21 Stoddart A, Fernald AA, Wang J, et al. Haploinsufficiency of del(5q) genes, Egri and Apc, cooperate with Tp53 loss to induce acute myeloid leukemia in mice. *Blood*. 2014;123(7):1069–1078.
- 22 Yu J, Baron V, Mercola D, Mustelin T, Adamson ED. A network of p73, p53 and Egri is required for efficient apoptosis in tumor cells. *Cell Death Differ*. 2007;14(3):436–446.
- 23 Parker JE, Mufti GJ, Rasool F, Mijovic A, Devereux S, Pagliuca A. The role of apoptosis, proliferation, and the Bcl-2-related proteins in the myelodysplastic syndromes and acute myeloid leukemia secondary to MDS. *Blood*. 2000;96(12):3932–3938.
- 24 Schneider RK, Schenone M, Ferreira MV, et al. Rps14 haploinsufficiency causes a block in erythroid differentiation mediated by S100A8 and S100A9. *Nat Med*. 2016;22(3):288–297.
- 25 Boultonwood J. The role of haploinsufficiency of RPS14 and p53 activation in the molecular pathogenesis of the 5q- syndrome. *Pediatr Rep*. 2011;3(Suppl 2):e10.
- 26 Rosam M, Krader D, Nickels C, et al. Bap (Sil1) regulates the molecular chaperone BiP by coupling release of nucleotide and substrate. *Nat Struct Mol Biol*. 2018;25(1):90–100.
- 27 Kim K, Choi J, Heo K, et al. Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription. *J Biol Chem*. 2008;283(14):9113–9126.
- 28 Lezon-Geyda K, Najfeld V, Johnson EM. Deletions of PURA, at 5q31, and PURB, at 7p13, in myelodysplastic syndrome and progression to acute myelogenous leukemia. *Leukemia*. 2001;15(6):954–962.
- 29 Stoddart A, Fernald AA, Wang J, et al. Haploinsufficiency of del(5q) genes, Egri and Apc, cooperate with Tp53 loss to induce acute myeloid leukemia in mice. *Blood*. 2014;123(7):1069–1078.
- 30 Li L, Sheng Y, Li W, et al. β -Catenin is a candidate therapeutic target for myeloid neoplasms with del(5q). *Cancer Res*. 2017;77(15):4116–4126.