Mutational patterns and their correlation to CHIP-related mutations and age in hematological malignancies

Tracking no: ADV-2021-004668R2

Anna Stengel (MLL Munich Leukemia Laboratory, Germany) Constance Baer (Munich Leukemia Laboratory, Germany) Wencke Walter (Munich Leukemia Laboratory, Germany) Manja Meggendorfer (MLL Munich Leukemia Laboratory, Germany) Wolfgang Kern (MLL Munich Leukemia Laboratory, Germany) Torsten Haferlach (MLL Munich Leukemia Laboratory, Germany) Claudia Haferlach (MLL Munich Leukemia Laboratory, Germany)

Abstract:
Acquired somatic mutations are crucial for the development of the majority of cancers. We performed a comprehensive comparative analysis of the mutational landscapes and their correlation to CHIP-related (clonal hematopoiesis of indeterminate potential) mutations and patient age of 122 genes in 3096 cases with 28 different hematological malignancies. Differences were observed regarding (i) the median number of mutations (highest, median n=4: aCML, CMML, MDS/MPN-U, s-AML; lowest, n=0, CML, MLN-eo, MGUS, PPBL), (ii) specificity of certain mutations (high frequencies in e.g. aCML (ASXL1, 86%), FL (KMT2D, 87%; CREBBP, 73%), HCL (BRAF, 100%), LPL (MYD88, 98%; CXCR4, 51%), MPN (JAK2, 68%)), (iii) distribution of mutations (broad distribution within/across the myeloid/lymphoid lineage e.g. for TET2, ASXL1, DNMT3A, TP53, BCR, ETV6), (iv) correlation of mutations to patient age (correlated to older age across entities: e.g. TET2, DNMT3A, ASXL1, TP53, EZH2, BCR, GATA2, IDH2; younger age: e.g. KIT, POT1, RAD21, U2AF2; WT1), (v) correlation of mutation number per patient with age (total cohort (p<0.001), AML (p<0.001), B-ALL (p=0.015), CLL (0.039), MDS (p<0.001), MPN (p<0.001), T-ALL (p=0.005)). Moreover, we observed high frequencies of mutations in RUNX1, SRSF2, IDH2, NRAS, EZH2 in cases comprising at least one DTA mutation (DNMT3A, TET2, ASXL1), while in cases without DTA mutations TP53, KRAS, WT1, SF3B1 were more frequent across entities, suggesting differences in pathophysiology. These results give further insight into the complex genetic landscape and the role of DTA mutations in hematological neoplasms and define mutation-driven entities (MDS/MPN overlap, s-AML) in comparison to entities defined by chromosomal fusions (CML, MLN-eo).

Conflict of interest: COI declared - see note

COI notes: CH, WK, and TH declare part ownership of Munich Leukemia Laboratory (MLL). AS, WW, CB and MM are employed by the MLL.

Preprint server: No;

Author contributions and disclosures: AS and CH designed the study, AS interpreted the data, AS wrote the manuscript. AS and CH were responsible for chromosome banding and FISH analyses, WW, CB and MM for molecular and bioinformatic analyses, WK for immunophenotyping and TH for cytomorphologic analyses. All authors read and contributed to the final version of the manuscript.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: For original data, please contact anna.stengel@mll.com.

Clinical trial registration information (if any):
Mutational patterns and their correlation to CHIP-related mutations and age in hematological malignancies

Running Head: Mutational patterns in hematological neoplasms

Anna Stengel, Constance Baer, Wencke Walter, Manja Meggendorfer, Wolfgang Kern, Torsten Haferlach, Claudia Haferlach

Affiliations:
MLL Munich Leukemia Laboratory, Max-Lebsche-Platz 31, 81377 Munich, Germany

Corresponding Author:
Dr. rer. nat. Anna Stengel, PhD
MLL Munich Leukemia Laboratory
Max-Lebsche-Platz 31, 81377 München, Germany
T: +49-89-990-17-471
F: +49-89-990-17-13471
Email: anna.stengel@mll.com

Word count:
Abstract: 250
Main text: 3661
Tables: 1
Figures: 4
Reference count: 38
Key points

Comparison of the mutation frequencies and numbers of 122 genes in 3096 cases allows identification of “mutation-driven” entities.

Differences in mutation patterns in cases with and without CHIP-associated mutations across entities suggest differences in pathophysiology.
Abstract

Acquired somatic mutations are crucial for the development of the majority of cancers. We performed a comprehensive comparative analysis of the mutational landscapes and their correlation to CHIP-related (clonal hematopoiesis of indeterminate potential) mutations and patient age of 122 genes in 3096 cases with 28 different hematological malignancies. Differences were observed regarding (i) the median number of mutations (highest, median n=4: aCML, CMML, MDS/MPN-U, s-AML; lowest, n=0, CML, MLN-eo, MGUS, PPBL), (ii) specificity of certain mutations (high frequencies in e.g. aCML (ASXL1, 86%), FL (KMT2D, 87%; CREBBP, 73%), HCL (BRAF, 100%), LPL (MYD88, 98%; CXCR4, 51%), MPN (JAK2, 68%)), (iii) distribution of mutations (broad distribution within/across the myeloid/lymphoid lineage e.g. for TET2, ASXL1, DNMT3A, TP53, BCOR, ETV6), (iv) correlation of mutations to patient age (correlated to older age across entities: e.g. TET2, DNMT3A, ASXL1, TP53, EZH2, BCOR, GATA2, IDH2; younger age: e.g. KIT, POT1, RAD21, U2AF2; WT1), (v) correlation of mutation number per patient with age (total cohort (p<0.001), AML (p<0.001), B-ALL (p=0.015), CLL (0.039), MDS (p<0.001), MPN (p<0.001), T-ALL (p=0.005)). Moreover, we observed high frequencies of mutations in RUNX1, SRSF2, IDH2, NRAS, EZH2 in cases comprising at least one DTA mutation (DNMT3A, TET2, ASXL1), while in cases without DTA mutations TP53, KRAS, WT1, SF3B1 were more frequent across entities, suggesting differences in pathophysiology. These results give further insight into the complex genetic landscape and the role of DTA mutations in hematological neoplasms and define mutation-driven entities (MDS/MPN overlap, s-AML) in comparison to entities defined by chromosomal fusions (CML, MLN-eo).
Introduction

Genetic changes causing aberrant proliferation constitute one of the hallmarks of cancer cells, which can be caused by molecular mutations (single nucleotide level) or by translocations, deletions, amplifications and/or whole chromosome aneuploidy (chromosomal level). The type of aberration causing malignant cell transformation depends on the respective cancer. In hematological malignancies, molecular mutations play crucial roles for pathogenesis and classification/diagnosis of many types of leukemias and lymphomas, but also chromosomal translocations (e.g. *BCR-ABL1* in chronic myeloid leukemia, CML) or aneuploidy (e.g. hyperdiploid or hypodiploid acute lymphoblastic leukemia, ALL) are known to be fundamental for development of the respective diseases. While some molecular mutations are known to be very specific for certain entities (e.g. *BRAF* in hairy cell leukemia (HCL), *MYD88* in lymphoplasmacytic lymphoma (LPL)), others were detected in a variety of malignancies (e.g. mutations in *TP53, TET2, DNMT3A, RUNX1*), often even across the myeloid/lymphoid lineage. Generally, accumulation of mutations in somatic cells during development is also known to be not only one of the reasons for cancer development, but also regarded as a major cause of aging. Moreover, mutations in some genes (mainly affecting *DNMT3A, TET2, ASXL1*) that are known to be present with high frequencies in patients with acute myeloid leukemia (AML) or myelodysplastic syndrome (MDS) were also detected in individuals with blood cell parameters in the normal range (clonal hematopoiesis of indeterminate potential, CHIP) in an age-related manner: CHIP was detected rarely in persons younger than 40 years, but in ~10% of persons older than 70 years. Moreover, individuals with CHIP showed an increased risk of developing hematological neoplasms (~14 fold higher than in individuals without CHIP), which was found to be dependent on clone size, the specific mutation detected and the number of mutations of the genes associated with CHIP. CHIP was additionally discussed to possibly increase the risk of heart disease, vascular diseases, inflammation and diabetes as well as functions as a risk factor for therapy-related AML/MDS.

Thus, to gain further insight into the mutational landscapes of hematological malignancies and their relation to CHIP and aging, we performed a comprehensive large scale evaluation of the mutational landscape of most hematological cancers. In more detail, we performed (i) a comprehensive analysis and comparison of the
mutation frequencies of 122 selected genes in 3096 cases with 28 different hematological malignancies for identification of “mutation-driven” entities, (ii) a correlation analysis of DTA mutations (i.e. mutations in $DNMT3A$, $TET2$, and $ASXL1$) with mutational landscapes and (iii) correlation of the mutation frequencies with age of the respective patients. The project aims at performing a “Real-World analysis” providing a reference for the mutational landscape in hematological malignancies obtained from an unselected group of patients. Beside this potential diagnostics implication, age-specific mutations should be characterized in several entities, which to our knowledge has not been performed in such a large number of hematological malignancies before.

Material and methods

Patients and samples
For diagnostic work-up, 3096 patients diagnosed with 28 different hematological malignancies (Table 1) sent to the MLL Munich Leukemia Laboratory between 08/2007 and 05/2020 were selected based on sample availability for WGS. Diagnoses (from peripheral blood and/or bone marrow) were confirmed based on morphology, immunophenotype, cytogenetics and molecular genetics as previously published.\textsuperscript{15-17} For analysis of mutation frequencies and for correlation to DTA mutations, the total cohort of 3096 cases was used, for correlation with age of the patients, only cohorts with a number of cases ≥50 were selected for adequate statistical power (11 different hematological malignancies, 2656 cases, Table 1). For abbreviations of entities used in the text, see Table 1. All patients had given written informed consent to the use of genetic and clinical data according to the Declaration of Helsinki. The study has been approved by the internal review board of the MLL Munich Leukemia Laboratory.

Whole Genome Sequencing (WGS)
Total RNA and genomic DNA were extracted from lysed cell pellet of diagnostic bone marrow or peripheral blood of all 3096 patients using the MagNA Pure 96 with DNA and Viral Nucleic Acid Large Volume Kit and Cellular RNA Large Volume Kit (Roche, Basel, Switzerland). Library preparation was performed using TruSeq DNA PCR-Free
HT sample preparation kit (Illumina, San Diego, CA, USA) according to manufacturer’s protocol and 151 bp paired-end sequences were generated on NovaSeq6000 and HiSeqX sequencing instruments (Illumina) with 90x coverage. As matched-normal samples were not available, a mixture of genomic DNA from multiple anonymous donors was used as normal controls. Read mapping and tumor/normal variant calling were performed using Illumina’s WGS app version 5.0 and tumor normal app 3.0 for analysis of genomic data. WGS reads were mapped to human reference genome (Ensembl GRCh37) using Illumina’s Isaac aligner (iSAAC-03.16.02.19). Structural variations were called using Manta (version 0.28.0). To remove false positive fusion events, a customized WGS filtering pipeline was applied with each variant requiring at least 2 paired read support and PASS or MGE10kb flag. Final analysis was performed only on protein-altering and splice-site variants as annotated by the Illumina annotation engine.

**Mutational analysis and variant filtering**

For mutational analysis, data on 122 selected genes was used (Supplemental Table 1). Of note, for multiple myeloma patients only 29 genes were evaluated due to low coverage (only genes with a coverage per exon >15 were used for analysis). To exclude artefacts, only variants in regions described as a high confidence by the Genome in a Bottle Consortium were analyzed further. Moreover, genes prone for generating artefacts by their location in problematic genomic regions were excluded from further analysis. To remove potential germline variants, each variant was queried against the gnomAD database and variants with global population frequencies >0.0005 were excluded. Moreover, variants were eliminated if a germline origin was suggested by ClinVar and/or COSMIC database and simultaneously a variant allele frequency (VAF) ranging from 0.45 to 0.55 was detected. Additionally, only mutations with a HePPY score of >0.5 were used for this cohort.

**Statistical analysis**

Statistical analyses were performed using SPSS (version 19.0.0) software (IBM Corporation, Armonk, NY). For correlation analysis, the Pearson correlation coefficient (bivariate correlation) was used. For comparison of median values, the
independent samples t test was applied (SPSS). All reported p-values are two-sided and were considered significant at p<0.05.

Data Sharing Statement
The data used in this manuscript is part of a larger data set containing 5000 whole genomes the MLL Munich Leukemia Laboratory. Data security is guaranteed by the MLL with respect to storage and scientific analyses and strictly follows the General Protection Regulation in Europe (GDPR, https://gdpr.info.eu/). Any request for scientific use of the data should be addressed to corresponding author of the manuscript at anna.stengel@mll.com.

Results

Numbers of mutations and mutational patterns
Entities with the highest numbers of mutations detected in the 122 selected genes (median n=4, respectively) and thus potentially with the largest impact of mutations on pathogenesis comprised a number of MDS/MPN overlap entities (aCML (range: 1-7 mutations), CMML (1-6), MDS/MPN-U (2-5)) as well as s-AML cases (2-8). By contrast, the lowest numbers (median n=0) of mutations were observed for CML (range: 0-3), MGUS (0-2), MLN-oeo (0-3), NK cell neoplasm (0-3) and PPBL (0-2). In the total cohort of 3096 cases, the most frequently mutated genes were TET2 (14%), ASXL1 (13%), TP53 (10%), SF3B1 (9%), DNMT3A (9%) and SRSF2 (9%). Entities with very high frequencies of specific mutations (> 50%) comprised aCML (ASXL1, 86%), BPDCN (TET2, 67%), BL (TP53, 60%), CMML (TET2, 67%; ASXL1, 58%), FL (KMT2D, 87%; CREBBP, 73%), HCL (BRAF, 100%), LPL (MYD88, 98%; CXCR4, 51%), MDS/MPN-U (ASXL1, 60%), MPN (JAK2, 68%), B-NHL (TP53, 50%) and T-NHL (STAT3, 52%) (Figures 1, 2). Mutations additionally enriched in distinct entities included SETBP1 (predominantly in MDS/MPN overlaps, 26%), CSF3R (30% in MDS/MPN-U), STAT3 (only in T-NHL and NK cell neoplasm, 52% and 23%), NOTCH1 and PHF6 (T-ALL, 38% and 30%) and MYC and ID3 (almost exclusively in BL, 30% each) (Figures 1, 2). Regarding the occurrence of mutations in myeloid and lymphoid neoplasms, genes predominantly mutated in myeloid neoplasms comprised e.g. SF3B1 (with the exception of CLL), JAK2, NPM1, RUNX1, IDH2, CEBPA, STAG2, NF1 and GATA2. By contrast, mutations in KMT2D, MYD88, ARID1A, ATM,
CXCR4, BIRC3 and CD79B were detected almost exclusively in lymphoid malignancies (Figure 1). Genes with high mutation frequencies in plasma cell neoplasms include mainly NRAS and KRAS, although mutations in these two genes were also detected in a variety of other entities. Of note, only 29 genes were evaluated in MM samples due to low coverage, hence conclusions on the mutational landscape in MM patients are limited. A broad distribution across entities was observed for mutations in e.g. TET2, ASXL1 (although both are enriched in myeloid neoplasms, especially in MDS/MPN overlaps), DNMT3A, TP53 (with a high mutation frequency in B-NHL and BL), BCO R and ETV6. Thus, the first three, i.e. DTA genes, were mutated with high frequencies also in lymphoid neoplasms. In line with this, gene mutations found in the largest number of entities comprise DNMT3A (n=23 entities), TET2 (n=21), ASXL1, TP53, NRAS (n=19, respectively), KRAS and BCO R (n=17, respectively).

We moreover analyzed a potential correlation of mutation frequencies with gender of patients (only mutations with a total mutation frequency of >2% were used for this investigation). Interestingly, a number of mutations were found to occur predominantly in females (DNMT3A, 12% in females vs. 7% in males, p<0.001; NPM1, 7% vs. 5%, p=0.002), while other were correlated with male gender, comprising a number of genes functioning in splicing (ASXL1, 9% vs. 15%, p<0.001; PHF6, 1% vs 3%, p=0.008; SF3B1, 8% vs. 10%, p=0.036 ; SRSF2, 5% vs. 11%, p<0.001; U2AF1, 2% vs. 4%, p=0.007 ; ZRSR2, 0% vs. 3%, p<0.001).

**Mutation patterns in cases with and without DTA mutations**

Further, we compared the mutational patterns of cases with at least one DTA mutation (DNMT3A, TET2, ASXL1, n=920 cases in the total cohort) with cases without such mutations (n=2176) to decipher CHIP-correlated patterns of accompanying mutations. Significant differences with respect to accompanying mutations were mainly detected for myeloid neoplasms (MDS, mutations in n=12 genes significantly different in cases with at least one DTA mutation compared to cases without DTA mutations, without DTA genes themselves; AML, n=7; MPN, n=5; aCML, n=2; CMML, n=1) but also for MPAL (n=3), T-ALL (n=2), B-ALL, FL and LPL (n=1, respectively) (Figure 3). In more detail regarding the affected genes, mutations in TP53 were found significantly enriched in cases without DTA mutations in 4 different entities (aCML, AML, MDS, MPAL), moreover mutations in KRAS, WT1 and
SF3B1 were more abundant in cases without DTA mutations (in CMML, AML and aCML, respectively). By contrast, cases with DTA mutations were characterized by high frequencies of mutations in RUNX1 (in n=4 entities; AML, B-ALL, MDS, MPN), SRSF2 (n=3; AML, MDS, MPN), IDH2 (n=3; AML, MDS, T-ALL), NRAS (n=3; MDS, MPAL, T-ALL) and EZH2 (n=2; MDS, MPN). Mutations that were found to be more abundant in cases with DTA mutations in n=1 entity each included IDH1, U2AF1 (both in AML), FAT4 (FL), UBR5 (LPL), BRCC3, CBL, DHX29, NF1, SH2B3, STAG2 (all in MDS), FLT3 (MPAL), CSF3R, MPL (both in MPN) (Figure 3). The median VAF (variant allele frequency) of DNMT3A, TET2 and ASXL1 in the total cohort was 42.7, 40.5 and 41.3, respectively (range 1.2 – 93.5, 1.1 – 66.7, 2.7 – 100, respectively). Median VAF of the genes that were recurrently found to be differentially associated with and without DTA genes comprised: TP53 (46.9 in cases with at least one DTA mutation vs. 59.8. in cases without DTA mutations), RUNX1 (43.9 vs. 43.2), SRSF2 (45.8 vs. 42.1), IDH2 (43.2 vs. 42.2), NRAS (36.5 vs. 27.1), and EZH2 (48.0 vs. 15.4). In more detail, comparison of the medium VAFs of certain genes with DTA genes in distinct entities suggest that e.g. mutations in EZH2 in MDS, mutations in NRAS in AML, B-ALL, MDS and MPN and TP53 mutations in T-ALL might have occurred after preceding CHIP, whereas mutations in EZH2 in T-ALL and mutations in TP53 in AML showed a high VAF and thus unlikely evolved after CHIP (see Supplemental Figure 1).

**Correlation of number and frequencies of molecular mutations with patient age**

As expected, the age of the analyzed cohorts differed strongly, ranging from 37 years (median; range: 11 – 91 years) in T-ALL to 81 years in MDS/MPN-U (77 – 85 years) (Table 1). For further analysis on age relation and distribution, only cohorts with ≥50 cases were used (2656 patients, 11 entities: aCML, AML, B-ALL, CLL, CML, FL, LPL, MDS, MPN, MM, T-ALL; see Table 1). Analysis of the relative number of patients in each decade in these 11 entities revealed that for many entities a peak in decade 7 (70.0 – 79.9 years; aCML, AML, MDS, LPL, MPN, MM) or decade 6 (B-ALL, CLL) was detected. By contrast, FL patients showed a peak in decade 4 and T-ALL patients in decade 2, while age distribution was found to be quite uniform in CML patients (Supplemental Figure 2). In this selected cohort of 2656 patients, totally 5709 mutations in the 122 analyzed genes were detected. Here, younger patients showed a lower median number of
mutations than older patients (median number of mutations for patients ≤60 years: 1; >60 years: 2; p<0.001). An association between age and the number of mutations per patient was detected for the total cohort (p<0.001) and for AML (p<0.001), B-ALL (p=0.015), CLL (0.039), MDS (p<0.001), MPN (p<0.001) and T-ALL (p=0.005), but not for aCML, CML, FL, LPL and MM. This effect was less pronounced when CHIP-related genes (ASXL1, DNMT3A, TET2) were omitted from the cohort and was than merely detected for the total cohort, AML and MDS cases (total cohort: p=0.025; AML: p=0.001; MDS: p=0.001). When correlating the mutation frequencies of the 122 analyzed genes with age, the strongest influence was found for AML, as 17 genes were mutated in an age-related manner: mutations in KIT, POT1, RAD21, U2AF2 and WT1 were significantly correlated with younger age, whereas for ASXL1, BCOR, BCRR3, DNMT3A, ETV6, IDH2, RUNX1, SRSR2, TET2, TP53, U2AF1 and PHF6, a significant correlation with older age was observed (Figure 4, Supplemental Figure 3). Of note, a number of mutations correlating with younger age were associated to an aberration known to occur itself more frequently in younger age, e.g. the vast majority of KIT mutations in AML were associated with t(8;21)(q22;q22)/RUNX1-RUNX1T1 or inv(16)(p13q22)/CBFB-MYH11 (22/25 cases, 88%), a similar observation was detected for RAD21 in AML (associated with t(8;21) in 8/19 cases, 42%). Age-dependent mutational profiles were also detected for CLL, B-ALL, MDS, MPN, T-ALL, FL, MM and LPL (Figure 4, Supplemental Figure 3). No age-dependent differences in mutational patterns were detected for aCML and CML. Thus, the genes that were found to be mutated in most entities in older patients comprised TET2 (in 6 entities), TP53 (n=4), DNMT3A (n=3), ASXL1, BCOR, EZH2, GATA2 and IDH2 (n=2, respectively). By contrast, POT1 was found to be preferentially mutated in younger patients in 2 entities (AML and LPL). Moreover, for some mutations the relation to patient age was entity-specific: PHF6 mutations were correlated to younger age in CLL (median age mutated vs. unmutated: 43 years vs. 67 years, p=0.04), but to older age in AML (74 years vs. 68 years, p=0.007), similar to CD79B (older age: CLL, 93 years vs. 67 years, p=0.015; younger age: LPL, 52 years vs. 71 years, p=0.023) and NRAS (older age: T-ALL, 75 years vs. 34 years, p<0.001; younger age: B-ALL, 63 years vs. 69 years, p=0.009; MM, 65 years vs. 70 years, 0.042) (Supplemental Figures 4,5). Of note, T-ALL patients with mutations in DNMT3A or TET2 were clearly older than patients without such mutations (DNMT3A: median age 75 vs. 33 years; TET2: 74 vs. 37 years).
Discussion

A number of genetic alterations are known to be crucial for the development of the majority of cancers, including acquired somatic mutations, structural variations i.e. chromosomal translocations leading to certain gene fusions and copy number variations. In the present study, we comprehensively analyzed the frequencies and distribution of somatic mutations in hematological malignancies (28 different entities, 122 selected genes) to decipher entities that seem to be mainly mutation-driven and further investigated the correlation of these mutations to patient age. Thus we provide a comprehensive large scale evaluation of the mutational landscape of most hematological cancers that has to our knowledge not been performed for such a large number of cases and entities before. Although molecular mutations are known to play important roles for pathogenesis in the vast majority of hematological neoplasms, our comparative analysis allowed identification of certain entities that seem to be mainly driven by the acquisition of mutations as primary event causing such diseases, as a high number of molecular mutations was found, including many MDS/MPN overlap cases (comprising aCML, CMML, MDS/MPN-U) as well as s-AML cases (median number of detected mutations: 4). Moreover, also in FL and AUL cases a high number (median: n=3) of mutations were detected. The high number of observed mutations in FL cases is noteworthy, as the hallmark of FL is known to be a chromosomal translocation (t(14;18)(q32;q21)/IGH-BCL2), detectable in ~90% of all FL cases and leading to overexpression of BCL2. However, as the IGH-BCL2 rearrangement was also detected by RT-PCR in healthy individuals, it was already discussed that this aberration alone is not sufficient for FL development. This is underlined by our data, showing an overall large number of mutations in FL patients in comparison to other hematological entities, especially a very high frequency of mutations in KMT2D (87%) and CREBBP (73%) which is in line with previous analysis. On the other hand, a very low number of mutations was detected in CML and MLN-eo cases, which are known to be defined by certain translocations as primary events for pathogenesis (CML: t(9;22)(q34;q11)/BCR-ABL1; MLN-eo: rearrangements involving PDGFRα (4q12), PDGFRB (5q31-33), FGFR1 (8p11) and t(8;9)(p22;p24.1)/PMC1-JAK2, although molecular mutations might play important
roles in e.g. progression of these diseases as well. Thus, a substantial role of molecular mutations in development of these diseases is not likely. Further, molecular mutations were found to be infrequent in MGUS and PPBL patients, constituting two entities that are known or discussed to be benign or “pre-malignant”, explaining the observed low number of mutations. Moreover, while certain mutations were found to be very specific or very frequent in certain entities, others showed a broad distribution within or even across the myeloid/lymphoid lineage. Specific/frequent mutations included well-known examples such as \textit{MYD88} in LPL, \textit{BRAF} in HCL and \textit{JAK2} in MPN, but also \textit{KMT2D} and \textit{CREBBP} in FL (see above), \textit{TP53} in BL, \textit{ASXL1} in MDS-MPN-U and aCML and \textit{TET2} in CMML and BPDCN. Gene mutations or mutational patterns that occur very frequently in a certain entity could be used in (differential) diagnostics for assignment of a correct diagnosis e.g. in difficult cases, as it is already applied for \textit{MYD88} mutations in LPL or \textit{BRAF} mutations in HCL.\textsuperscript{5,6} Moreover, some of these mutational patterns could in the future potentially be used – together with data on gene fusions and copy number changes – for a WGS-only classification and/or diagnosis of cases. On the other hand, the identification of certain mutations could – independently of the respective diagnosis – potentially be used therapeutically for so-called basket trials, a new type of clinical trial for which eligibility is based on the presence of a specific genomic alteration, irrespective of histology.\textsuperscript{31} These biomarker-driven trials were already conducted for e.g. patients with mutations in \textit{KRAS}, \textit{NRAS} or \textit{BRAF} (in patients with e.g. multiple myeloma, ovarian cancer, colorectal cancer, non-small cell lung cancer) or for mutations in \textit{IDH1}, \textit{IDH2} and \textit{FGFR2} (in patients with e.g. advanced biliary tract carcinoma or glioma).\textsuperscript{32-34} Thus, data as generated in the present study could provide further therapeutical options for patients otherwise lacking specific treatment possibilities.

It has to be noted, that although \textit{TET2} and \textit{ASXL1} showed a very high frequency in certain entities in our study, they belong to the genes that depicted the broadest distribution across entities and were detected in substantial amounts also in lymphoid malignancies. As \textit{TET2} and \textit{ASXL1} belong (together with \textit{DNMT3A}) to the genes that are known to be associated with CHIP\textsuperscript{9}, it can be postulated that CHIP might also play a role in these entities. Interestingly, previous analysis\textsuperscript{35} of the genomic profiles of ultra-stable CLL, low-count MBL (monoclonal B-cell lymphocytosis) and high-count MBL patients revealed striking similarities, and further suggested that at least some
somatic mutations may have occurred before the onset of CLL or even before MBL, similar to CHIP occurring before MDS. Additionally, mutational profiles of these ultra-stable CLL cases seemed to differ from CLLs with a more rapid progression and were characterized by infrequent CLL driver gene mutations and by non-coding variants targeting key pathways/cellular processes relevant to normal and neoplastic B-cell development.\textsuperscript{35} This observation of different molecular patterns is also supported by our data, as we deciphered differences in the mutation patterns of cases with vs. without CHIP-associated mutations across all entities. Cases with at least one DTA mutation showed an enrichment of mutations in \textit{RUNX1}, \textit{SRSF2}, \textit{IDH2}, \textit{NRAS}, \textit{EZH2}, while cases without DTA mutations were mainly characterized by frequent mutations in \textit{TP53}, but also in \textit{KRAS}, \textit{WT1} and \textit{SF3B1}, suggesting differences in pathophysiology of cases with and without mutations in DTA genes. Of note, these observations were not only made for myeloid malignancies such as MDS and AML, but also for other entities including T-ALL and B-ALL. Correlation of mutation frequencies to patient age revealed the well-known association of DTA mutations (\textit{DNMT3A}, \textit{TET2}, \textit{ASXL1}) to older age across entities, but also other genes were found to show this correlation (e.g. \textit{TP53}, \textit{EZH2}, \textit{BCOR}, \textit{GATA2}, \textit{IDH2}). A lower number of mutations were correlated with younger patient age, which could in many cases be explained by association with other age-related aberrations. Interestingly, some mutations show varying age-relation dependent on the entity, e.g. \textit{PHF6} mutations (CLL, younger age; AML, older age; T-ALL, equal distribution; see Supplemental Figure 5), proposing differences in mutation mechanisms or selection dependent on the respective entity. \textit{PHF6} is a tumor suppressor gene with a proposed role in transcriptional regulation and/or chromatin remodeling and was found to be mutated quite frequently in T-ALL (~20\% of cases) and more rarely in AML cases (3\%) were it was found to be associated with an immature morphology (FAB subtypes M0–M2).\textsuperscript{36–38} So far, \textit{PHF6} mutations were not described in CLL cases, thus their role in CLL pathogenesis and association to younger age in these patients could be further evaluated. 

Taken together, our data give further insight into the different mutational landscapes of leukemias and lymphomas and into their interrelation with aging, providing additional insight on different pathogenesis pathways of hematological cancer. It moreover indicates that CHIP might play a greater role in a higher number of hematological malignancies than so far anticipated and it characterizes the
relationship between DTA mutations and other mutations in more detail. Additionally, we hope that this “real-world analysis” of a large number of patients comprising most hematological cancers might help improving classification and/or diagnosis of particularly ambiguous cases on the one hand, as well as adding therapeutical molecular-based options for patients otherwise lacking specific treatment possibilities on the other hand (basket trials), thus providing a source for the researcher as well as the clinician.

Prior presentation footnote
Presented in abstract form at the 62nd American Society of Hematology Annual Meeting in December 2020.

Acknowledgements
The authors would like to thank all co-workers at the MLL Munich Leukemia Laboratory for their dedicated work. The authors would also like to thank all physicians for providing samples and caring for patients as well as collecting data.

Authorship Contributions
AS and CH designed the study, AS interpreted the data, AS wrote the manuscript. AS and CH were responsible for chromosome banding and FISH analyses, WW, CB and MM for molecular and bioinformatic analyses, WK for immunophenotyping and TH for cytomorphologic analyses. All authors read and contributed to the final version of the manuscript.

Disclosure of Conflicts of Interest
CH, WK, and TH declare part ownership of Munich Leukemia Laboratory (MLL). AS, WW, CB and MM are employed by the MLL.
References

1. Loeb KR, Loeb LA. Significance of multiple mutations in cancer. *Carcinogenesis*. 2000;21(3):379–385.

2. Orr-Weaver TL, Weinberg RA. A checkpoint on the road to cancer. *Nature*. 1998;392(6673):223–224.

3. Pui CH, Relling MV, Downing JR. Acute lymphoblastic leukemia. *N Engl J Med*. 2004;350(15):1535–1548.

4. Swerdlow SH, Compo E, Harris NL, et al. WHO classification of tumours of haematopoetic and lymphoid tissue. International Agency of Research on Cancer 2017; Revised 4th Edition, Volume 2.

5. Tiacci E, Trifonov V, Schiavoni G, et al. BRAF mutations in hairy-cell leukemia. *N Engl J Med*. 2011;364(24):2305-2315.

6. Varettoni M, Arcaini L, Zibellini S, et al. Prevalence and clinical significance of the MYD88 (L265P) somatic mutation in Waldenstrom's macroglobulinemia and related lymphoid neoplasms. *Blood*. 2013;121(13):2522-2528.

7. Freitas AA, de Magalhães JP. A review and appraisal of the DNA damage theory of ageing. *Mutation Research*. 2011;728(1-2):12–22.

8. Risques RA, Kennedy SR. Aging and the rise of somatic cancer-associated mutations in normal tissues. *PLoS Genet*. 2018;14(1):e1007108.

9. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 2015;126(1):9-16.

10. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014;371(26):2488-2498.
11. Malcovati L, Galli A, Travaglino E, et al. Clinical significance of somatic mutation in unexplained blood cytopenia. *Blood* 2017;129(25):3371-3378.

12. Busque L, Buscarlet M, Mollica L, Levine RL. Concise Review: Age-Related Clonal Hematopoiesis: Stem Cells Tempting the Devil. *Stem Cells.* 2018;36(9):1287-1294.

13. Jaiswal S, Natarajan P, Silver AJ. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N Engl J Med.* 2017;377(2):111-121.

14. Takahashi K, Wang F, Kantarjian H. Preleukaemic clonal haemopoiesis and risk of therapy-related myeloid neoplasms: a case-control study. *Lancet Oncol.* 2017;18(1):100-111.

15. 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.

16. 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.

17. 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.

18. Zook J, Catoe D, McDaniel J. Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci Data.* 2016;3:160025.
19. Fuentes Fajardo K, Adams D, Mason CE, et al. Detecting False-Positive Signals in Exome Sequencing. *Hum. Mutat.* 2012;33(4):609-613.

20. Karczewski K, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020; 581(7809):434-443.

21. Landrum M, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acid Res.* 2018;46(D1):D1062-D1067.

22. Hutter S, Baer C, Walter W, Kern W, Haferlach C, Haferlach T (2019). A Novel Machine Learning Based in silico Pathogenicity Predictor for Missense Variants in a Hematological Setting. *Blood.* 2019;134(Supplement_1):2090.

23. Horseman DE, Gascoyne RD, Coupland RW, Coldman AJ, Adomat SA. Comparison of cytogenetic analysis, southern analysis, and polymerase chain reaction for the detection of t(14;18) in follicular lymphoma. *Am J Clin Pathol.* 1995;103(4):472-488.

24. Rowley JD. Chromosome studies in the non-Hodgkin's lymphomas: the role of the 14;18 translocation. *J Clin Oncol.* 1988;6(5):919-925.

25. Roulland S, Lebailly P, Roussel G, et al. BCL-2/JH translocation in peripheral blood lymphocytes of unexposed individuals: lack of seasonal variations in frequency and molecular features. *Int J Cancer.* 2003;104(6):695-698.

26. Schmitt C, Balogh B, Grundt A, et al. The bcl-2/IgH rearrangement in a population of 204 healthy individuals: occurrence, age and gender distribution, breakpoints, and detection method validity. *Leuk Res.* 2006;30(6):745-750.

27. Green MR, Kihira S, Liu CL, et al. Mutations in early follicular lymphoma progenitors are associated with suppressed antigen presentation. *Proc Natl Acad Sci U S A.* 2015;112(10):E1116-1125.
28. 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.

29. Tesson B, Huet S, Grange B, et al. Absence of driver mutations in persistent polyclonal B-cell lymphocytosis with binucleated lymphocytes. *Blood*. 2017;130(10):1267-1269.

30. Lakshman A, Paul S, Rajkumar SV, et al. Prognostic significance of interphase FISH in monoclonal gammopathy of undetermined significance. *Leukemia*. 2018;32(8):1811-1815.

31. Qin B-D, Jiao X-D, Liu K, et al. Basket Trials for Intractable Cancer. *Front Oncol.* 2019;9:229.

32. Chenard-Poirier M, Kaiser M, Boyd K, et al. Results from the biomarker-driven basket trial of RO5126766 (CH5127566), a potent RAF/MEK inhibitor, in RAS- or RAF-mutated malignancies including multiple myeloma. *J of Clin Oncol.* 2017;35(15_suppl):2506.

33. Passiglia F, Malapelle U, Del Rec M, et al. KRAS inhibition in non–small cell lung cancer: Past failures, new findings and upcoming challenges. *Eur J Cancer*. 2020;137:57-68.

34. Xie C, McGrath NA, Monge Bonilla C, Fu J. Systemic treatment options for advanced biliary tract carcinoma. *J Gastroenterology*. 2020;55(10):944–957.

35. Agathangelidis A, Ljungström V, Scarfò L, et al. Highly similar genomic landscapes in monoclonal B-cell lymphocytosis and ultra-stable chronic lymphocytic leukemia with low frequency of driver mutations. *Haematologica*. 2018;103(5):865-873.

36. Van Vlierberghe P, Palomero T, Khiabanian H, et al. PHF6 mutations in T-cell acute lymphoblastic leukemia. *Nat Genet*. 2010;42(4):338–342.
37. Voss AK, Gamble R, Collin C, et al. Protein and gene expression analysis of Phf6, the gene mutated in the Borjeson-Forssman-Lehmann Syndrome of intellectual disability and obesity. *Gene Expr Patterns.* 2007;7(8):858–871.

38. Van Vlierberghe P, Patel J, Abdel-Wahab O, et al. PHF6 mutations in adult acute myeloid leukemia. *Leukemia.* 2011;25(1):130-134.
## Table 1: Summary of analyzed entities and abbreviations used.

| Entity                                                                 | Abbreviation | Number of cases | Age in years (median, range) | Number of mutations (median) |
|------------------------------------------------------------------------|--------------|----------------|-----------------------------|-----------------------------|
| Myelodysplastic syndrome                                              | MDS          | 577            | 73 (23 – 93)                | 2                           |
| Myelodysplastic/myeloproliferative neoplasms, unclassifiable          | MDS/MPN-U    | 10             | 81 (77 – 85)                | 4                           |
| Myeloproliferative neoplasms                                          | MPN          | 306            | 68 (22 – 91)                | 2                           |
| Chronic myeloid leukemia                                              | CML          | 94             | 55 (16 – 85)                | 0                           |
| Atypical CML                                                          | aCML         | 50             | 74 (50 – 92)                | 4                           |
| Chronic myelomonocytic leukemia                                       | CMML         | 24             | 75 (59 – 89)                | 4                           |
| Myeloid or lymphoid neoplasms associated with eosinophilia            | MLN-eo       | 36             | 51 (24 – 82)                | 0                           |
| Acute myeloid leukemia                                                | AML          | 675            | 68 (18 – 93)                | 2                           |
| Secondary acute myeloid leukemia                                      | s-AML        | 16             | 70 (48 – 83)                | 4                           |
| Therapy-associated acute myeloid leukemia                             | t-AML        | 16             | 53 (33 – 73)                | 1                           |
| Acute undifferentiated leukemia                                       | AUL          | 24             | 74 (18 – 88)                | 3                           |
| Mixed phenotype acute leukemia                                        | MPAL         | 30             | 64 (19 – 89)                | 2                           |
| B-cell acute lymphoblastic leukemia                                   | B-ALL        | 224            | 54 (0.1 – 93)               | 1                           |
| T-cell acute lymphoblastic leukemia                                   | T-ALL        | 101            | 37 (11 – 91)                | 2                           |
| T-cell neoplasm                                                       | T-NHL        | 33             | 62 (41 – 82)                | 1                           |
| NK cell neoplasms                                                    | NK           | 40             | 68 (44 – 87)                | 0                           |
| B-cell neoplasm                                                       | B-NHL        | 30             | 71 (46 – 83)                | 1                           |
| Burkitt’s lymphoma                                                    | BL           | 20             | 69 (26 – 82)                | 2                           |
| Chronic lymphocytic leukemia                                          | CLL          | 263            | 67 (30 – 94)                | 1                           |
| Follicular lymphoma                                                   | FL           | 63             | 54 (43 – 88)                | 3                           |
| Mantle cell lymphoma                                                  | MCL          | 17             | 66 (29 – 81)                | 1                           |
| Lymphoplasmacytic lymphoma                                            | LPL          | 55             | 70 (42 – 88)                | 2                           |
| Hairy cell leukemia                                                   | HCL          | 45             | 74 (35 – 86)                | 1                           |
| Hairy cell leukemia variant                                            | HCL-v        | 43             | 71 (42 – 81)                | 1                           |
| Monoclonal gammopathy of undetermined significance                    | MGUS         | 12             | 61 (35 – 87)                | 0                           |
| Multiple myeloma                                                      | MM           | 248            | 69 (26 – 91)                | 1                           |
| Persistent polyclonal B-cell lymphocytosis                            | PPBL         | 29             | 45 (30 – 56)                | 0                           |
| Blastic plasmacytoid dendritic cell neoplasm                          | BPDCN        | 15             | 74 (16 – 88)                | 2                           |

Entities highlighted in bold were used for association analysis of patient age (entities comprising at least 50 cases)
Figure Legends

Figure 1: Overview of the detected mutation frequencies in the 28 analyzed entities. The colour indicates the frequency of the respective mutation in each entity (colour of relative frequency, see below). Grey: n.a. The upper barplot indicates the number of entities, in which the respective gene shows a mutation frequency of >10%, the genes are sorted accordingly. The barplot on the right depicts the total number of genes that are mutated with a frequency of >10% in the respective entity.

Figure 2: Summary of most frequently mutated genes per entity. All genes with a mutation frequency of >5% are depicted. Colour code corresponds to the frequency level (see legend on the right). Entities are sorted according to classification into myeloid, lymphoid, mixed, plasma cell neoplasm and other.

Figure 3: Mutational patterns in relation to DTA mutations. Comparison of the mutational patterns of cases with at least one DTA mutation (DNMT3A, TET2, ASXL1; left panel) with cases without such mutations (right panel). The depicted entities are the ones in which certain mutations (shown on the bottom, script size indicates the number of entities with changes in this mutation) with significantly different frequencies dependent on presence of DTA mutations were found.

Figure 4: Graphical summary of age-associated molecular mutations. The circles include all mutations across all entities that were found to be associated with younger age (grey, left side), older age (light red, right side) or for which the age association was found to differ between entities (middle). Script size indicates the number of entities with changes in this mutation.
Cases with at least one mutation in CHIP genes (*DNMT3A*, *TET2*, *ASXL1*)

Cases without CHIP-associated mutations

**Figure 3**

**Cases with at least one mutation in CHIP genes (DNMT3A, TET2, ASXL1)**

- NRAS
- SRSF2
- BRCC3
- CSF3R
- FLT3
- IDH1
- IDH2
- MPL
- UBR5
- SH2B3
- EZH2
- FAT4
- STAG2

**Cases without CHIP-associated mutations**

- TP53
- KRAS
- SF3B1
- WT1
associated with younger age

age association differs between entities

associated with older age

Figure 4