Transcriptional profiling reveals mixed epistasis in Chronic Lymphatic Leukemia

Draft

1 Abstract

While recurrent (driver) mutations in CLL have been extensively catalogued, how they affect the disease phenotypes individually or collaboratively remains poorly characterized. To address this need, we performed RNA sequencing on 184 CLL patients and linked gene expression changes to genomic variations.

We identified robust and previously unknown gene expression signatures for recurrent copy number variations (including trisomy12, del11q22.3, del17p13, del18p12 and gain8q24), gene mutations (TP53, BRAF and SF3B1) and the somatic hypermutation status of the immunoglobulin heavy-chain variable region (IGHV). The most profound gene expression changes are associated with IGHV (3275 genes), methylation groups and trisomy 12 (3557 genes), which were apparent on unsupervised clustering. Gene set enrichment... (ighv, Tri12). We also demonstrate genetic interactions reflected in the transcriptional profiles. IGHV showed a mixed epistasis interaction with trisomy 12, suggesting.... Different groups of genes showed coordinated changes representing buffering, suppression and inversion of their expression phenotypes. Mixed epistasis involving IGHV and trisomy was also supported by drug response phenotypes and global methylation profiles.

In summary, our study provides a comprehensive reference data set for gene expression in CLL. We show that IGHV mutation status, recurrent gene mutations and CNVs drive gene expression and exhibit specific expression pattern. This includes epistatic interaction between trisomy12 and IGHV. Using a novel way to describe coordinated changes we can group genes into sets related to buffering, inversion and suppression.

2 Introduction

Chronic lymphocytic leukemia (CLL) is a heterogeneous disease with variable course and tumour dynamics [Fabbri and Dalla-Favera, 2016; Zenz et al., 2010]. Large cohort studies using next generation sequencing technologies uncovered a highly variable genotypic and epigenetic landscape underlying this heterogeneity. The mutation status of IGHV and TP53 mutations have been established as major prognostic factors. The functional role of most genomic variants remain incompletely characterized, but the incidence of many gene mutations suggest a strong selective advantage. The majority of CLL patients have more than one recurrent mutation or copy number variation. Numerous characteristic patterns of co-occurrence have been identified between different genomic variations e.g. between the IGHV mutation status, Notch1, SF3B1 and ATM mutation. Nevertheless the functional impact and the interplay which determine phenotypic changes are still incompletely understood.

Transcriptome profiling techniques, such as RNA-sequencing, capture a snapshot of the dynamic RNA expression on a genome-wide scale, reflecting the diversity of cellular states and underlying regulatory mechanisms. Previous transcriptomic studies on CLL samples showed variability in gene expression, but suggested that major genetic subgroups may not be discernible by analyzing gene expression [Ferreira et al., 2014; Rosenwald et al., 2001].

A recent comparison between CLL and normal B cell samples showed large expression changes that were associated with metabolic pathways, ribosome and spliceosome [Ferreira et al., 2014]. While CLL showed high gene expression variability, few robust gene expression changes were associated with genotypes. Associations were found for the IGHV status, which accounted for 1.5% of the overall variance. The analysis revealed two subgroups termed c1/c2, which were independent of known molecular disease groups. These subgroups were independent of the genomic background, but showed clear expression changes, which were associated with clinical outcome [Ferreira et al., 2014]. Later expression changes were linked to potential technical effects of sample processing. In a large pan cancer analysis Dvinge et al. found that blood collection procedures may rapidly change the transcriptional and posttranscriptional landscapes of hematopoietic cells, resulting in biased activation of specific biological pathways [Dvinge et al., 2014].

Variable RNA expression of CLL patients could be a consequence of the cell of origin and e.g. heterogeneous genetic landscape including the impact of driver mutations on transcriptional programmes and thus provide the opportunity to dissect the functional roles through transcriptomic profiling. However, associations between recurrent genomic

variations and transcriptomic changes have been surprisingly rare considering the targeting of key processes including BCR signalling (IGHV), DNA damage pathways (p53, ATM) and NOTCH signalling (NOTCH1, FBXW7, MED12) [Landau et al., 2013; Puente et al., 2011; Rossi and Gaidano, 2016]. Even for major biological disease subgroups as the hypermutation status of IGHV an initial study could only find differences in the overall variance of gene expression between U-CLL and M-CLL, but no clear expression profiles [Ecker et al., 2015].

To understand the biology of CLL and study the impact of genetic variants on gene expression in CLL, we profiled 184 CLL samples using RNA-sequencing and searched for molecular features underlying gene expression variability in CLL. We aimed to identify the expression signatures for the most prevalent gene mutations and copy number changes, which includes the IGHV hypermutation status, trisomy12, gene mutations in TP53, ATM, BRAF, SF3B1, Notch1 and MED12 and CNVs as del17p13, del11q22.3, del8p12 and gain8q24. Besides transcriptional profiling of these single variants we looked for pattern of genomic interaction within gene expression.

3 Material and Methods

3.1 Data acquisition

3.1.1 RNA-sequencing

RNA isolation and library preparation were performed as described Dietrich et al. [2017]. In short, total RNA was isolated from patient blood samples using the RNA RNeasy mini kit (Qiagen). For quantification a Qubit 2.0 Fluorometer was used, and RNA integrity was evaluated with an Agilent 2100 Bioanalyzer. Samples with RNA integrity number (RIN) <8 were excluded from the study. Library preparation was performed according to the Illumina TruSeq RNA sample preparation v2 protocol. Samples were paired-end sequenced at the DKFZ Genomics and Proteomics Core Facility. Two to three samples were multiplexed per lane on Illumina HiSeq 2000, Illumina HiSeq3000/4000 or Illumina HiSeqX machines.

Raw RNA-seq reads were demultiplexed and quality control was performed using FastQC version 0.11.5 [Andrews, 2010]. Internal trimming with STAR version 2.5.2a [Dobin and Gingeras, 2015] was used to remove adapters before mapping [Dobin and Gingeras, 2015]. Mapping was performed using `STAR` version 2.5.2a [Dobin and Gingeras, 2015] against the Ensembl human reference genome release 75 (Homo sapiens GRCh37.75) [Flicek et al., 2014]. `STAR` was run in default mode with internal adapter trimming using the `clip3pAdapterSeq` option. Mapped reads were summarized into counts using `htseq-count` version 0.9.0 [Anders and Huber, 2010] with default parameters and `union` mode. Thus, only fragments unambiguously overlapping with one gene were counted [Anders and Huber, 2010]. The count data were then imported into R (version 3.4) for subsequent analysis.

Among the RNA-seq data from 184 CLL samples presented in the study, 136 samples were included in the Primary Blood Cancer Cell Encyclopedia (PACE) [Dietrich et al., 2017].

3.1.2 Somatic variants

Mutations, including 66 gene mutations and 22 structural variants, of 184 CLL samples were retreaved as previously described in the Primary Blood Cancer Cell Encyclopedia (PACE) project [Dietrich et al., 2017]. Statistical analyses were restricted to genetic variants with a prevalence $\geq$ 3.5% (7/184 samples), i.e., to 7 gene mutations and 5 SVs. In addition, the somatic hypermutation status of the immunoglobulin heavy-chain variable region (IGHV) and CLL subtype classifications defined by global levels of CpG methylation [Oakes et al., 2016] level were also obtained from PACE.

Genomic variations, including 66 gene mutations and 22 structural variants, of the 184 CLL samples were extensively profiled in one of our previous studies and can be queried

from the Primary Blood Cancer Cell Encyclopedia (PACE) project [Dietrich et al., 2017]. Statistical analyses were restricted to those genetic variants with a prevalence $\geq 3.5\%$ (7/184 samples), i.e., to 7 gene mutations and 5 SVs. In addition, the somatic hypermutation status of the immunoglobulin heavy-chain variable region (IGHV) and CLL subtype classifications defined by global levels of CpG methylation level [Oakes et al., 2016] were also obtained from PACE.

3.2 Statistical analysis

3.2.1 Statistical analysis

3.2.2 Exploratory data analysis

All statistical analyses were performed using R version 3.4. Raw count data were transformed using variance stablizing transformation (VST) [Love et al., 2014] before exploratory data analysis. Genes were ranked by overall variance, and the most variable genes were used in the subsequent analyses. PCAs were generated with the R function `prcomp` using the 1000 most variable genes. For hierarchical clustering of samples, expression data of the 1000 most variable genes were centred and scaled and distances were calculated by the R `dist` function with Euclidean distance. The `complexHeatmaps` package [Gu et al., 2016] was further used to cluster and display gene expression using `ward.D2` as method [Murtagh and Legendre, 2014]. Two-dimensional t-distributed stochastic neighbour embedding (t-SNE) was performed using the `Rtsne` function of the `Rtsne` package [Krijthe, 2017] on sample distances calculated based on the 150 most variable genes.

3.2.3 Batch effect estimation

Due to the large sample size, transcriptome data were generated over a period of four years and platforms were changed due to technological development during the period of sequencing, which lead to the changes in sequencing depth and read length (101, 125 and 150 nucleotides). Therefore, the batch effect due to platform differences were estimated. While the amount of sequences belonging to adapters differed depending on the read length, we could not find associations between the top 10 principle components and different batches after adapter trimming (data not shown).

3.2.4 Differential Expression Analysis

Differentially expressed genes between samples with and without each of the genetic variants were identified using `DESeq2` version 1.16.1 [Love et al., 2014]. Exploratory data analysis revealed IGHV hypermutation status and trisomy12 as dominating factors for gene expression. Both were used as co-variates in the GLMs of the remaining variants.

To account for multiple testing, the method of Benjamini and Hochberg [Benjamini and Hochberg, 1995] was applied to control FDR at $\alpha = 0.1$.

3.2.5 Gene set enrichment analysis

Gene set enrichment analysis was performed using the Bioconductor package `piano` version 1.18.0 [Väremo et al., 2013] using adjusted p.values from `DESeq2` and Fisher's combined probability test as method [Fisher, 1932]. `Hallmark` and `KEGG` gene set collections version 4.0 were downloaded from `MSigDB` [Kanehisa et al., 2017; Liberzon et al., 2015]. Significance of gene sets was determined based on a background distribution of sampled genes with 1000 permutations. P-values were corrected for multiple testing using the method of Benjamini and Hochberg, with a FDR = 0.1 [Benjamini and Hochberg, 1995].

3.2.6 Tumor epistasis model

Genetic interactions were modeled with an epistasis model describing observed gene expression g by a linear combination of the expression in the two interacting variants *variantA* and *variantB*.

$$g \sim variantA + variantB + variantA : variantB \tag{1}$$

The model was fit using `DESeq2` [Love et al., 2014]. P-values were adjusted for multiple testing with FDR = 0.1 [Benjamini and Hochberg, 1995].

4 Results

4.1 Gene mutations and IGHV are main drivers of gene expression variability in CLL

Previous studies found a limited impact of known disease subgroups on GEP [Rosenwald et al., 2001]. To obtain an overview of drivers for gene expression variability in CLL, we performed hierarchical clustering based on the 1000 most variable genes (figure 1A). Hierarchical clustering showed a clear separation of distinct subgroups which segregated with IGHV, methylation status and presence of trisomy 12. These results were supported by principal component analysis (PCA)(figure 1C) and suggest that genetic subgroups in CLL shape gene expression to a previously underappreciated extend. We found PC1, which represented 9.5% of variance, to be associated with IGHV status and PC2 and PC3 separated samples based on trisomy 12. Besides these main drivers, we also found numerous differentially expressed genes associated with further genetic variants including SF3B1, BRAF, NOTCH1 and TP53 mutation (figure 1B).

A large previous gene expression study in CLL described c1/c2 groups, which were found to be associated with B cell receptor activation and outcome [Ferreira et al., 2014]. We used hierarchical clustering with the differentially expressed genes characteristic of c1/c2, which indeed revealed two main clusters. In our data set, these differentially expressed genes showed very low variability across samples . In addition, none of PC1 - PC10 showed associations to the c1/2 subgroups and we could not identify an association of C1/C2 with outcome including time to next treatment or survival (data not shown). Recent data suggested the process associated with C1/2 gene expression being related to transcriptional changes linked to blood collection procedures and nonsense mediated decay [Dvinge et al., 2014]. In our analysis gene expression changes associated with C1/C2 add little variance compared to other sources of variability.

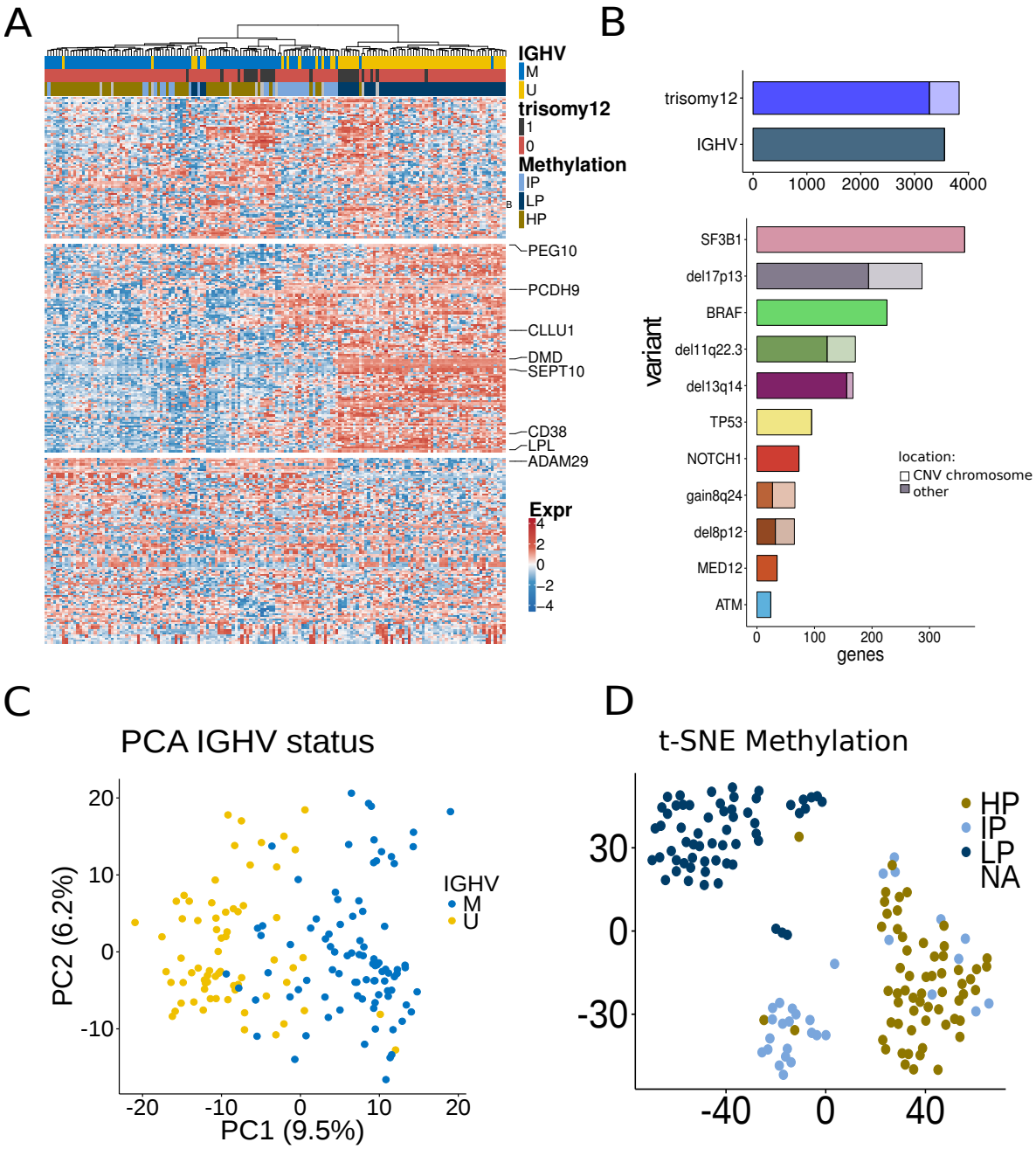

Figure 1: **Gene expression variability in CLL:** A) Hierarchical clustering of CLL samples based on the 1000 most variable genes. IGHV groups and Trisomy12 samples form major cluster. B) Number of differentially expressed genes ($p_{adj} < 0.01$) for genomic markers. C) IGHV status is associated with the 1st principal component, which explains 9.5% of variance. D) Two-dimensional t-distributed stochastic neighbour embedding (t-SNE) plot, based on sample distances using the 150 most variable genes. Distinct clusters are associated with the DNA methylation groups.

4.1.1 IGHV status is linked to distinct expression changes

To assess the influence of genetic drivers on gene expression in detail, we analysed gene expression changes for the most prevalent mutations and copy number variants of CLL. The somatic hypermutation status of IGHV was a main determinant of gene expression variability. We found 3275 genes significantly differentially expressed between M-CLL and U-CLL after adjustment for multiple testing using the method of Benjamini and Hochberg for FDR = 1%(figure 1B). In total 9.5 % of variance within gene expression was associated with the IGHV status (figure 1C). This revealed a larger impact on transcriptional changes than previously detected [Ferreira et al., 2014]. Genes previously found to be good markers related to the IGHV status include CD38, LPL, ZAP70, SEPT10, ADAM29 and PEG10 [Kienle et al., 2010], and were also associated with IGHV in our study (figure 2A/D). Besides this, we identified new marker genes associated with IGHV as PTCH1 and FGFR1 (figure 2D). PTCH1 is a Hedgehog signalling pathway component and was associated with CLL before [Decker et al., 2012]. The fibroblast growth factor receptor FGFR1 is related to cell growth and survival and is studied as potential target in the context of several other cancer types [Zhou et al., 2016]. Different IGHV genes were also found among the most differentially expressed genes, but showed heterogenous expression within the U and M-CLL groups. Commonly used IGHV Receptors (1-69 or 4-34) associated with U and M-CLL, respectively. We compared IGHV1-69 gene expression with data from IG gene analysis in the corresponding samples. Gene expression showed a strong relation to IG gene usage (figure 2B) and these data suggest, that RNA sequencing can be used to assess IG gene usage.

To understand pathways involved in U- and M-CLL we performed gene set enrichment analysis. Differentially expressed genes between IGHV genes were enriched in B cell receptor signalling, T cell receptor signalling as well as in ribosomal genes, several cancer pathways and chemokine signalling pathways (figure 2C). Detailed summaries of the deregulated genes within these pathways are shown in the supplements (Supplement figure 2).

Within the B cell receptor signalling gene set we identified cell surface molecules (CD19, CD22, CD81) and NFAT and NFkB signalling pathways to be downregulated. From the "T cell receptor signalling" gene set, ZAP70, PAK and p38 were upregulated in U-CLL, while IL10 and SHP1 were downregulated. For chemokine signalling pathways, we found downregulation of CXCR3 and CXCR4 in U-CLL, while a set of Interferons (IFNB1, INFA21) were upregulated.

In summary our data are in line with the major biological role of IGHV mutation status in CLL and provide a rich resource to assemble deregulated pathways in the disease.

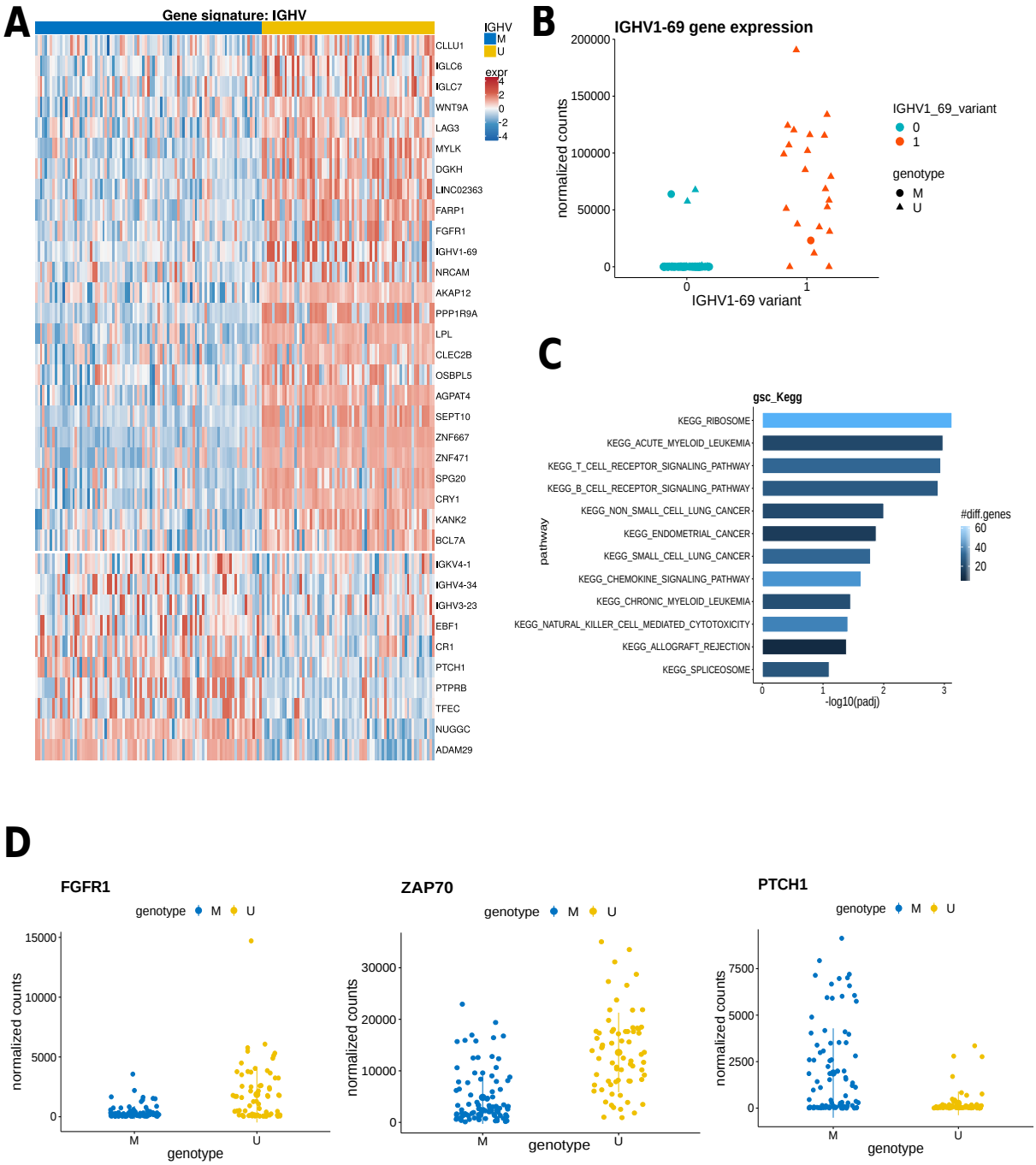

Figure 2: **Gene expression changes between IGHV groups:** A)Differentially expressed genes between IGHV groups with $p_{adj} < 0.01$, $\log_2$ fold change > 2 and basemean > 500. B) IGHV1-69 expression by corresponding IGHV1-69 variant usage determined in IG gene analysis. C) Enriched pathways between IGHV groups. D) Normalized gene counts for FGFR1, ZAP70 and PTCH1 separated by IGHV status.

4.1.2 Trisomy12 expression signature

We identified 3557 differentially expressed genes with $p_{adj} < 0.01$ in trisomy12 samples. To narrow them down, we filtered them by $\log_2$ fold change > 2 and basemean > 500. The basemean describes the mean of normalized counts of all samples, normalizing for sequencing depth [Love et al., 2014]. We find distinct expression pattern of up- and down-regulated genes for trisomy 12 samples (see figure 3A). Even though many upregulated genes are located on chromosome 12, the majority of differentially expressed genes are distributed among the other chromosomes and can not be ascribed to a dosage effect (see figure 3A,B). Among differentially expressed genes we found numerous marker of integrin signalling as SOCS3, ITGB2-AS1 and RAPGEF5. In addition endocytosis is one of the enriched pathways in trisomy12 (see table 1) and we find genes coding for adhesion molecules like GIT2 and ADD2 up regulated. This is in line with previous findings about increased lymph node homing via cellular adhesion and transendothelial migration of circulating cells into the lymph node in trisomy12 [Ganghammer et al., 2015; Riches et al., 2014]. We also find important checkpoint genes like the immune checkpoint CTLA4 and the cell cycle regulator CHFR differentially expressed in trisomy12. Both are associated with increased proliferation and tumor progression [Mittal et al., 2013; Oh et al., 2009] and reveal mechanisms used by trisomy12 to promote tumor growth. Altogether these results suggest that modulation of the microenviroment and deregulation of cell cycle checkpoint genes are important mechanism in trisomy12 tumorigenesis.

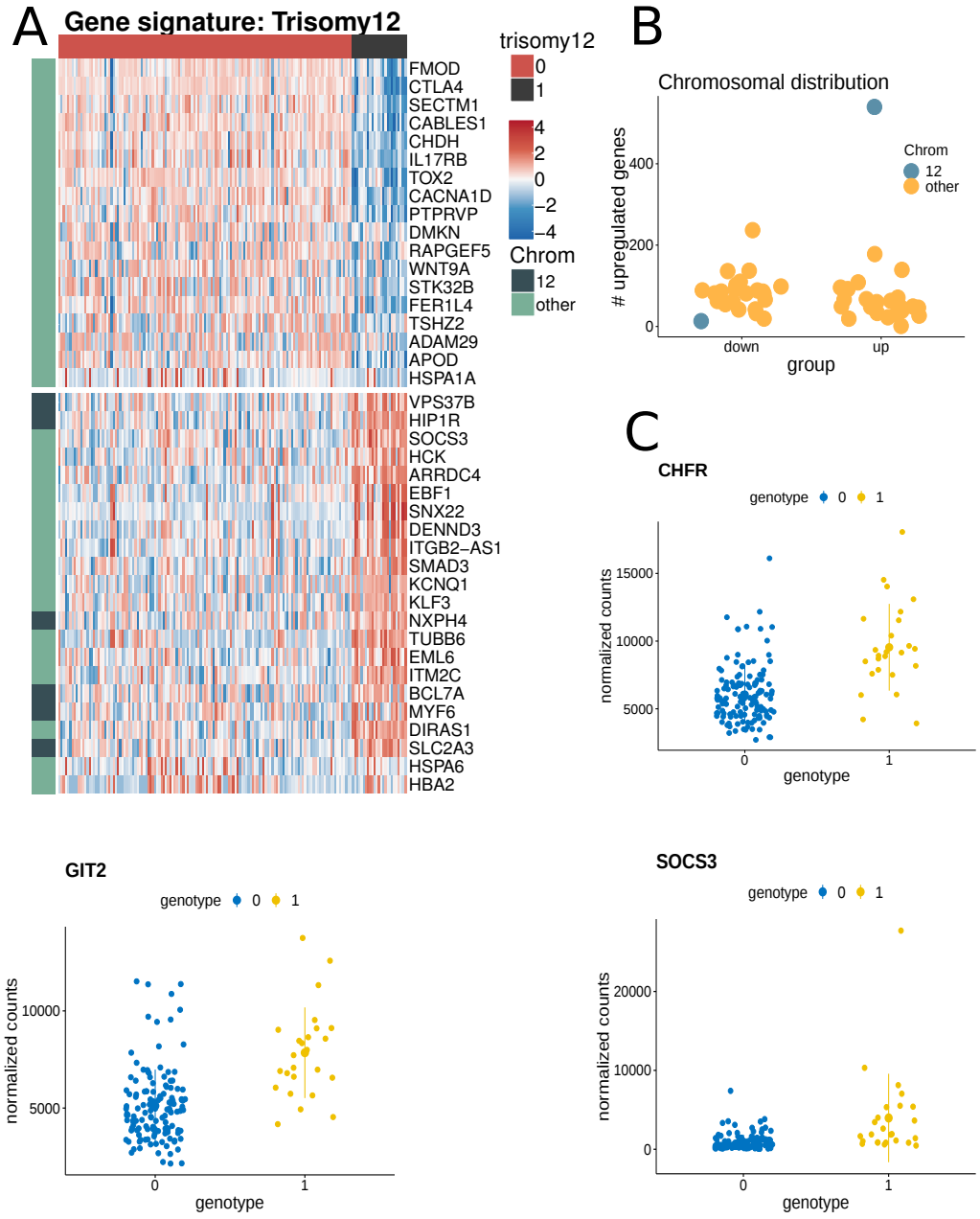

Figure 3: **Gene expression in Trisomy12:** A) Differentially expressed genes in Trisomy12 with $p_{adj} < 0.01$, $\log_2$ fold change > 2 and basemean > 500. B) Role of dosage effect: Chromosomal distribution of differentially expressed genes in Trisomy12. Most upregulated genes are located on chromosome 12.

4.1.3 Other genomic driver of gene expression

We also found numerous differentially expressed genes associated with other prevalent genetic variants (figure 1B). Mutation in the splicing factor SF3B1 gene showed more than 350 associations, indicating different layer of transcriptional aberrations within this variant. Among differentially expressed genes we found the chaperone gene UQCC. This gene has already been linked to SF3B1 mutations by differential isoform usage [Reyes et al., 2013]. Here we show another layer of differential regulation and highlight its role. In addition, we were able to identify more than 100 differentially expressed genes in driver mutations such as BRAF and common CNVs as del17p13, del13q14 and del11q22.3 (figure 1B). Top hits for each variant and enriched pathways of differentially expressed genes are shown in table 1. We further evaluated the extend of a direct dosage effect from the chromosomal distribution within differentially expressed genes of CNVs. Deletions showed a similar degree of dosage related genes on downregulated genes as variants related to a gain on upregulated genes. But between variants the dosage effect varied (see figure 1B). In line with this the sizes of lesions vary between CNVs and even within sample with the same CNV.

4.1.4 Intermediated programmed methylated samples form an independent cluster upon most variable genes.

Based on methylation pattern, distinction by IGHV status was recently refined by introducing a categorization into low- (LP), intermediate- (IP), and high- (HP) programmed samples. As mutation status of IGHV, methylation pattern were associated with the maturation status of tumor precursor cells within haematopoietic cell lineage [Oakes et al., 2016]. Using t-distributed stochastic neighbour embedding (t-SNE) we identified three cluster associated with these methylation groups based on the expression of the most variable genes (figure 1D). These findings confirm the relevance of methylation groups and a refined distinction of sample into these three groups to investigate CLL subtypes. Previous analysis of methylation pattern in CLL suggested a disease specific role of the transcription factors EGR, NFAT, AP1 and EGF by establishing aberrant methylation pattern [Oakes et al., 2016]. In line with this, we found EGF1, NFAT and EGR1 among genes whose expression patterns are associated with methylation groups.

Differentially expressed genes			
Variant	$\mathbf{p_{adj}}<$ 0.01	**Tophits**	**Kegg pathways**
trisomy12	3828	ANAPC5, CHFR, LIX1, GIT2, ITFG2, ADD2, NCKAP1L, UACA, SCARB1	Endocytosis, Regulation of Actin Cytoskeleton, Ubiquitin mediated proteolysis
IGHV	3557	SLC16A9,NETO1,PLD1, FRMD4B, PLEKHG4B,NPTX1, KCNK9,PON1,PRR18,	NK-cell mediated cytotoxicity, Calcium signaling pathyway, Pathways in cancer
SF3B1	361	APBB3,SRRM5,IFI27, TFCP2L1,UQCC1,FBN1, PSD2,FBLN2,HPCAL4	Melanogenesis, Cell cycle Neuroactive ligand receptor interaction
del17p13	287	CPT1C,NEURL4,MTTP, MTMR11,ZBTB4,SENP3, TFCP2L1,KIAA0753,RABEP1	Fatty acid metabolism, Endocytosis, Cell cycle
BRAF	226	TMPRSS3,SLC38A11,RAB25, DSP, ARHGEF37,PAGE2B, RNF157,ZFHX4,KIF14,	MAPK signaling pathway, Hematopoietic cell lineage, Wnt signaling pathway
del11q22	171	RNASE1, REXO2,USP28, CUL5,ATM, ALKBH8, TMPRSS5, NPAT,SIK2,	Cell cycle, Ubiquitin mediated proteolysis, Pyruvate metabolism
del13q14	167	ENPP3,TMPRSS4,CDCP1, TSPAN13, RGL3,SLC1A6 INHBA,RAI14,SOAT2	Pantheonate and COA biosyn., Starch and sucrose metabolism, Riboflavin metabolism
TP53	95	NOTCH4,PTPRB,MDM2 TEAD1,TFCP2L1,CMYA5, LRRC63, MAP2K4, PAX9	Notch signaling pathway, Adherens junction, Dorso ventral axis form.
Notch1	73	SPAG17,DSP,NOTCH4, GJB7,DNAH2,SH3RF1 POU6F2, ARHGEF37,SPACA9	Antigen processing, Spliceosome, Huntingtons disease
gain8q24	66	LGSN, ADGRG7, CBS, SNTB1,E2F5,RAD21, ZNF462,ZNF7,ADAMDEC1	Selenoamino acid metab., Glycine serine and threonine metab., Cystein and metheonine metab.
del8p12	65	TRPM2-AS,PSPHP1,INTS9, MTCO3P12,LGSN,HMBOX1, KIF13B, LINC01016, FERMT2	Basal cell carcinoma, Hedgehog signaling pathway, Alzheimers disease
MED12	35	CDH20,LGSN,CSPG5, MYO5C, ERRFI1,ADGRG7, FCRL4,CLCNKA,KLF4	Lysosome, Intest. immune network for IGA prod., Neurotrophin signaling pathway
ATM	24	RNASE1,SNTB1,PPM1E, RBFOX2,SAXO2,PLCB1, CYP51A1-AS1,CAMK2A,P4HA2	Gap junction, Vascular smooth muscle contr., Phosphatidylinositol sig. system

Table 1: **Summary differentially expressed genes in genomic variants**: Number of differentially expressed genes, tophits and top enriched pathways are shown.

4.2 IGHV status and Trisomy12 interact in an epistatic way determining gene expression

Most CLL samples show high mutational burden with around 2000 molecular lesions in the entire tumour genome [Gaidano and Rossi, 2017]. Even though most of them are rather related to genomic instability than being functional or tumour driving events, CLL is still characterized by an interplay of numerous genetic changes and environmental factors. In our cohort CLL samples carry about 3 of the tested variants on average (see Supplement figure 7A). To investigate the role of genetic interaction we tested their collaborative effect on gene expression phenotypes. We investigated the epistatic gene expression changes in the most severe genomic alterations the IGHV status and Trisomy12. Epistasis was defined as a non-linear effect on gene expression between sample with both variants co-occuring and the single variants alone [Fisher, 1919].

In total 893 genes showed specific expression pattern in a combined genotype ($p_{adj} < 0.1$). These expression changes differed from the expected change by simple combination of the single variant's effects. We observed different ways of epistatic interaction and clustered genes by them (see figure 4A). We distinguished between the following types of mixed epistasis (see figure 4B): Buffering, when the up or downregulation of a gene by a genetic variant was strongly enhanced in sample with the combined genotype. Inversion, when the effects in the single variants alone were reversed in the combined genotype. Suppression, when a strong up or downregulation of a gene in one or both variants alone was absent in samples with both variants. In total we identified five cluster of genes representing different ways of mixed epistasis as inversion down, suppression, different degrees of buffering and inversion up (see figure 4A from top to bottom). Different types of epistasis are also shown on the level of gene counts of single genes at the example of the adhesion mediating protein coding gene CHAD (inversion), the endonuclease encoding gene GEN1 (buffering up) and the transcription factor LEF1 (suppression)(see figure 4C).

These findings strongly suggest a functional association between the IGHV status and trisomy12 in CLL. The IGHV hypermutation status characterizes the B cell receptor signalling activity. This seem to impact trisomy12 related mechanisms as well. To further investigate this interaction we used enrichment tests for genes in the different mixed epistasis cluster. We found genes upregulated in trisomy12 U-CLL sample, but suppressed in M-CLL trisomy12 samples were enriched in Wnt beta catenin and Notch signaling. Notch signaling has shown to synergize with B cell receptor signalling before [Thomas et al., 2007]. Besides this, regulation of integrins in trisomy12 samples is modulated by notch signalling [Riches et al., 2014]. This suggests decreased B cell receptor activity in M-CLL sample effect trisomy12 depended integrin expression via notch signalling. In

contrast genes in other cluster as those showing a strong buffering effect were enriched in G2M checkpoint pathway. This points towards changes in cell cycle mechanisms as an alternative tumour driver (see figure 5).

4.3 The interaction between IGHV status and trisomy12 affect ex vivo drug response in CLL

Ex vivo sensitivity to drugs is an informative cellular phenotype that reflects pathway dependencies of the tumor cells. We therefore asked whether the epistatic interaction between IGHV status and trisomy12 on expression level also affects the drug response phenotype. In our previous PACE project, we measured the the ex vivo sensitivity of 184 CLL samples towards 63 compounds [Dietrich et al., 2017]. Using the same epistasis model on the drug screening dataset from PACE, we identified 6 drugs whose effect on cell viability was significantly affected by the interaction between IGHV and trisomy12 (see figure 6) ($p_{adj} < 0.1$). For four drugs, namely, vorinostat, NU7441, fludarabine and AZD7762, we observed a suppression effect of the trisomy12 phenotype by IGHV status: compared with non-trisomy12 samples, the trisomy12-only samples showed increased sensitivity to those four drugs, indicated by the lower viability of cells after drug treatment. Additional presence of IGHV mutation reduced sensitivity to those four drugs. While for the other two drugs, chaetoglobosin A and BIX02188, the samples with combined IGHV mutation and trisomy12 showed increased resistance compared with either wild type samples, IGHV mutated samples or trisomy12 samples.
Overall, the interaction of IGHV mutation and trisomy12 on drug response phenotype level further validates the presence of the epistasis effect of IGHV and trisomy12 in CLL and implicates this effect is functional relevant and maybe further explored for clinical therapeutics.

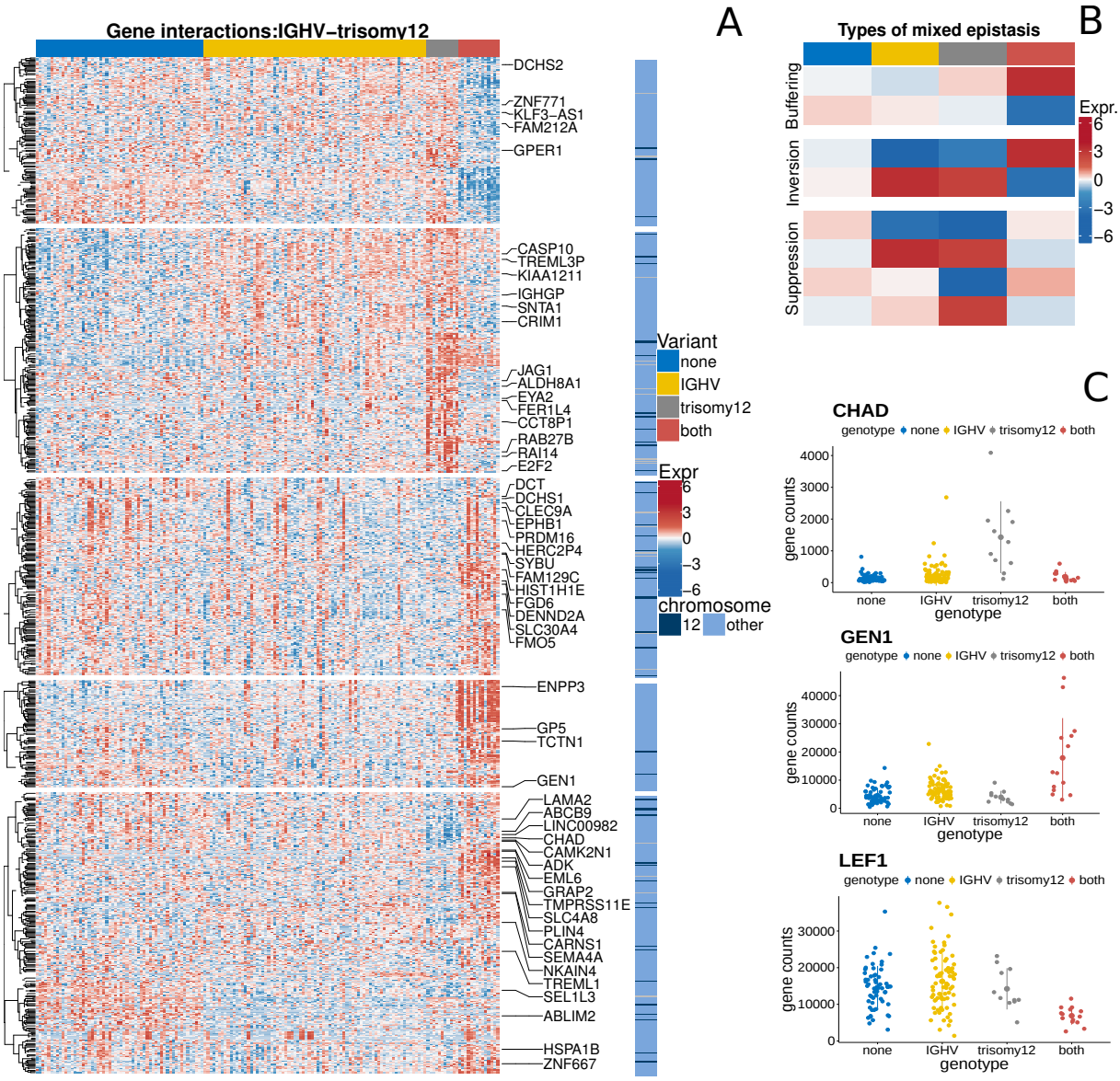

Figure 4: **Mixed epistasis in Trisomy12 and IGHV mutated samples:** A) Expression of epistatic gene interactions between Trisomy12 and M-CLL ($p_{adj} < 0.1$). Genes are clustered by different ways of epistatic interaction. B) Scheme of mixed epistasis in CLL samples: buffering, inversion, suppression. C) Epistatic expression of single genes. Raw gene counts of CHAD ($p_{adj} = 6.17 \times 10^{-08}$), GEN1 ($p_{adj} = 1.47 \times 10^{-04}$) and LEF1 ($p_{adj} = 6.21 \times 10^{-03}$) show different ways of epistasis as supression, buffering up and buffering down.

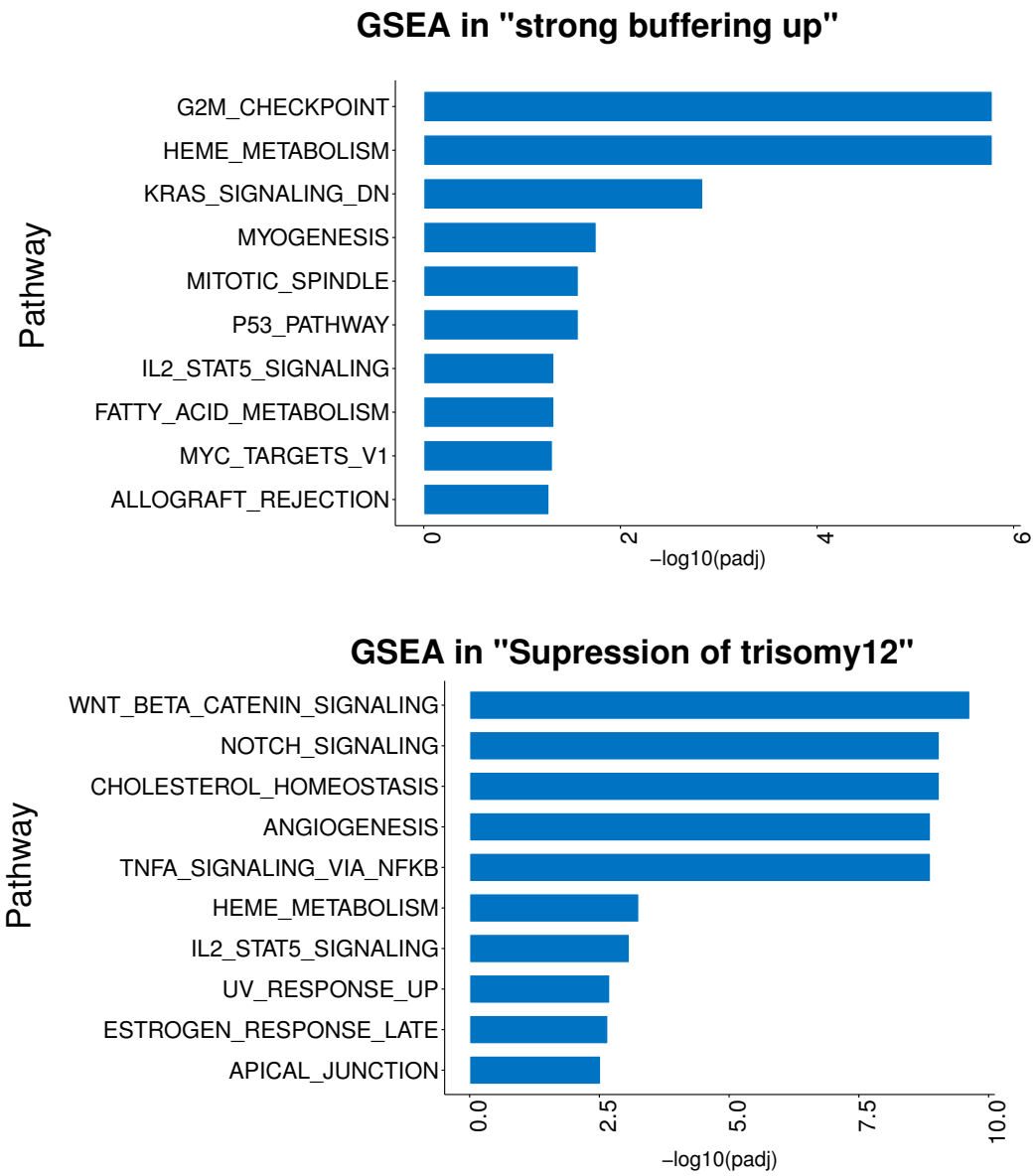

Figure 5: **Gene set enrichment analysis of mixed epistasis cluster:** A) Genes showing a strong positive buffering effect are upregulated in pathways related to G2M checkpoint genes and heme metabolism B) Genes which show an suppression compared to Trisomy12 only samples are enriched in pathways related to Notch signaling pathway and Wnt-beta-Catenin signaling

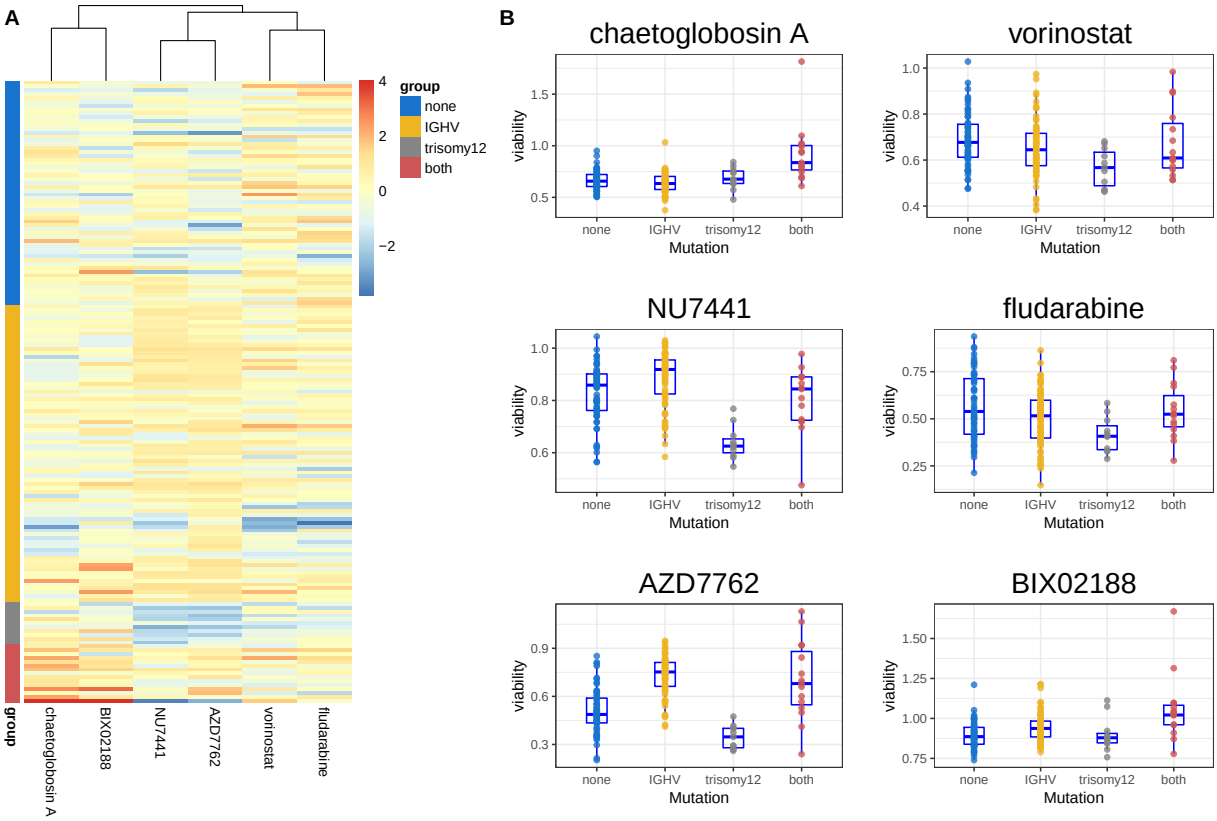

Figure 6: **Gene set enrichment analysis of mixed epistasis cluster:** A) Genes showing a strong positive buffering effect are upregulated in pathways related to G2M checkpoint genes and heme metabolism B) Genes which show an suppression compared to Trisomy12 only samples are enriched in pathways related to Notch signaling pathway and Wnt-beta-Catenin signaling

5 Discussion

Gene expression profiling has yet provided little insight into molecular mechanism underlying genomic variation in CLL. The main hurdle is the high phenotypic and genotypic heterogeneity. To overcome these challenges we collected and analysed 184 CLL transcriptomes, representing the largest CLL transcriptome dataset so far. Indeed we were able to identify gene expression signatures for the most prevalent genetic variants and could establish the hypermutation status of IGHV and trisomy12 as main driver of gene expression variability. In contrast to previous findings we found 9.5 % of gene expression variance associated with the IGHV status, revealing a higher impact on transcriptional changes than previously assumed [Ferreira et al., 2014]. We find clear associations between genotype data and gene expression and show that these can be used to investigate underlying mechanisms. So far global expression pattern in CLL have not been related to major genetic variation [Ferreira et al., 2014]. We could further validate our findings by the expression of some marker genes as CD38 and other previously reported associations. Using unsupervised clustering on the most variable genes, we could also distinguish samples by their associations to methylation groups, which support for the importance and function of this refined sample classification. Altogether we show how this dataset can provide an important resource to complement current understanding of CLLs by supporting previous findings as well as providing novel hypothesis on molecular mechanisms underlying the pathogenesis of CLL.

As CLL is often characterized by multiple co-occurring genetic lesions we further investigated the effect of genetic interaction on expression changes. Thereby we focused on the main driver of gene expression variability Trisomy12 and the IGHV status. We could not only identify numerous genes with specific expression pattern due to this interaction, but were also able to cluster them by different types of epistatic interactions like buffering, suppression and inversion. This concept of mixed epistasis between genetic lesions reflected on gene expression has been described in yeast before [Sameith et al., 2015]. Here we show the functional role and mechanistic importance of interaction between variants. We apply the concept of mixed epistasis and show how this can be used to unravel another layer of complexity in the diversity of tumorigenesis. This systematic approach was used to identify potential pathways as G2M checkpoint pathway, that collaboratively connect the IGHV status and trisomy12 and can help us to understand subtype specific differences. We further find epistatic interaction pattern in ex-vivo drug response data. Sample with hypermutated IGHV and trisomy12 are less sensitive toward a group of drugs indirectly targeting DNA. Our findings from expression analysis suggest more functional DNA damage machinery in these sample and thus provide a mechanistic explanation for this phenotype.

References

Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome biology*, 11(10):R106, 2010. ISSN 1474-760X. doi: 10.1186/gb-2010-11-10-r106. URL http://www.ncbi.nlm.nih.gov/pubmed/20979621http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3218662.

Simon Andrews. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data, 2010. URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):115–117, 1995. ISSN 00359246. doi: 10.2307/2346101. URL http://www.citeulike.org/user/ccr/article/1042553.

S. Decker, K. Zirlik, L. Djebatchie, D. Hartmann, G. Ihorst, A. Schmitt-Graeff, D. Herchenbach, H. Jumaa, M. Warmuth, H. Veelken, and C. Dierks. Trisomy 12 and elevated GLI1 and PTCH1 transcript levels are biomarkers for Hedgehog-inhibitor responsiveness in CLL. *Blood*, 119(4):997–1007, jan 2012. ISSN 0006-4971. doi: 10.1182/blood-2011-06-359075. URL http://www.ncbi.nlm.nih.gov/pubmed/22130798http://www.bloodjournal.org/cgi/doi/10.1182/blood-2011-06-359075.

S. Dietrich, M. Oleś, L. Sellner, S. Anders, J. Lu, B. Velten, A. Mock, C. Oakes, L. Sutton, E. Young, R. Rosenquist, D. Rossi, K. Zirlik, M. Herling, F. Nguyen-Khac, C. Plass, C. von Kalle, J. Dürig, I. Ringshausen, W. Huber, and T. Zenz. DRUG PERTURBATION BASED STRATIFICATION OF LYMPHOPROLIFERATIVE DISORDERS. *Hematological Oncology*, 35(S2):56–56, jun 2017. ISSN 02780232. doi: 10.1002/hon.2437_41. URL http://doi.wiley.com/10.1002/hon.2437{_}41.

Alexander Dobin and Thomas R Gingeras. Mapping RNA-seq Reads with STAR. *Current protocols in bioinformatics*, 51(3):11.14.1–19, sep 2015. ISSN 1934-340X. doi: 10.1002/0471250953.bi1114s51. URL http://www.ncbi.nlm.nih.gov/pubmed/26334920http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4631051.

Heidi Dvinge, Rhonda E Ries, Janine O Ilagan, Derek L Stirewalt, Soheil Meshinchi, and Robert K Bradley. Sample processing obscures cancer-specific alterations in leukemic transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(47):16802–7, nov 2014. ISSN 1091-6490. doi: 10.1073/pnas.1413374111. URL http://www.ncbi.nlm.nih.gov/pubmed/25385641http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4250124.

Simone Ecker, Vera Pancaldi, Daniel Rico, and Alfonso Valencia. Higher gene expression variability in the more aggressive subtype of chronic lymphocytic leukemia. *Genome medicine*, 7(1):8, 2015. ISSN 1756-994X. doi: 10.1186/s13073-014-0125-z. URL http://www.ncbi.nlm.nih.gov/pubmed/25632304http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4308895.

Giulia Fabbri and Riccardo Dalla-Favera. The molecular pathogenesis of chronic lymphocytic leukaemia. *Nature Reviews Cancer*, 16(3):145–162, feb 2016. ISSN 1474-175X. doi: 10.1038/nrc.2016.8. URL http://www.nature.com/doifinder/10.1038/nrc.2016.8.

Pedro G Ferreira, Pedro Jares, Daniel Rico, Gonzalo Gómez-López, Alejandra Martínez-Trillos, Neus Villamor, Simone Ecker, Abel González-Pérez, David G Knowles, Jean Monlong, Rory Johnson, Victor Quesada, Sarah Djebali, Panagiotis Papasaikas, Mónica López-Guerra, Dolors Colomer, Cristina Royo, Maite Cazorla, Magda Pinyol, Guillem Clot, Marta Aymerich, Maria Rozman, Marta Kulis, David Tamborero, Anaïs Gouin, Julie Blanc, Marta Gut, Ivo Gut, Xose S Puente, David G Pisano, José Ignacio Martin-Subero, Nuria López-Bigas, Armando López-Guillermo, Alfonso Valencia, Carlos López-Otín, Elías Campo, and Roderic Guigó. Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome research*, 24(2):212–26, feb 2014. ISSN 1549-5469. doi: 10.1101/gr.152132.112. URL http://www.ncbi.nlm.nih.gov/pubmed/24265505http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3912412.

R. A. Fisher. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, 52(02):399–433, jul 1919. ISSN 0080-4568. doi: 10.1017/S0080456800012163. URL http://www.journals.cambridge.org/abstract{_}S0080456800012163.

Ronald Fisher. *Statistical methods for research workers,*. Oliver and Boyd, Edinburgh, 1932. URL http://www.worldcat.org/title/statistical-methods-for-research-workers/oclc/4971991.

Paul Flicek, M. Ridwan Amode, Daniel Barrell, Kathryn Beal, Konstantinos Billis, Simon Brent, Denise Carvalho-Silva, Peter Clapham, Guy Coates, Stephen Fitzgerald, Laurent Gil, Carlos García Girón, Leo Gordon, Thibaut Hourlier, Sarah Hunt, Nathan Johnson, Thomas Juettemann, Andreas K. Kähäri, Stephen Keenan, Eugene Kulesha, Fergal J. Martin, Thomas Maurel, William M. McLaren, Daniel N. Murphy, Rishi Nag, Bert Overduin, Miguel Pignatelli, Bethan Pritchard, Emily Pritchard, Harpreet S. Riat, Magali Ruffier, Daniel Sheppard, Kieron Taylor, Anja Thormann, Stephen J. Trevanion, Alessandro Vullo, Steven P. Wilder, Mark Wilson, Amonida Zadissa, Bronwen L. Aken, Ewan Birney, Fiona Cunningham, Jennifer Harrow, Javier Herrero, Tim J.P.

Hubbard, Rhoda Kinsella, Matthieu Muffato, Anne Parker, Giulietta Spudich, Andy Yates, Daniel R. Zerbino, and Stephen M.J. Searle. Ensembl 2014. *Nucleic Acids Research*, 42(D1):D749–D755, jan 2014. ISSN 0305-1048. doi: 10.1093/nar/gkt1196. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt1196.

Gianluca Gaidano and Davide Rossi. The mutational landscape of chronic lymphocytic leukemia and its impact on prognosis and treatment. *Hematology. American Society of Hematology. Education Program*, 2017(1):329–337, 2017. ISSN 1520-4383. doi: 10.1182/asheducation-2017.1.329. URL http://www.ncbi.nlm.nih.gov/pubmed/29222275.

Sylvia Ganghammer, Evelyn Hutterer, Elisabeth Hinterseer, Gabriele Brachtl, Daniela Asslaber, Peter William Krenn, Tamara Girbl, Petra Berghammer, Roland Geisberger, Alexander Egle, Antonella Zucchetto, Anna Kruschinski, Valter Gattei, Alexandre Chigaev, Richard Greil, Tanja Nicole Hartmann, Sylvia Ganghammer, Evelyn Hutterer, Elisabeth Hinterseer, Gabriele Brachtl, Daniela Asslaber, Peter William Krenn, Tamara Girbl, Petra Berghammer, Roland Geisberger, Alexander Egle, Antonella Zucchetto, Anna Kruschinski, Valter Gattei, Alexandre Chigaev, Richard Greil, and Tanja Nicole Hartmann. CXCL12-induced VLA-4 activation is impaired in trisomy 12 chronic lymphocytic leukemia cells: a role for CCL21. *Oncotarget*, 6(14): 12048–12060, may 2015. ISSN 1949-2553. doi: 10.18632/oncotarget.3660. URL http://www.oncotarget.com/fulltext/3660.

Zuguang Gu, Roland Eils, and Matthias Schlesner. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18):2847–2849, sep 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btw313. URL http://www.ncbi.nlm.nih.gov/pubmed/27207943https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw313.

Minoru Kanehisa, Miho Furumichi, Mao Tanabe, Yoko Sato, and Kanae Morishima. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1):D353–D361, jan 2017. ISSN 0305-1048. doi: 10.1093/nar/gkw1092. URL http://www.ncbi.nlm.nih.gov/pubmed/27899662http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5210567https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw1092.

Dirk Kienle, Axel Benner, Carolin Läufle, Dirk Winkler, Christof Schneider, Andreas Bühler, Thorsten Zenz, Annett Habermann, Ulrich Jäger, Peter Lichter, Riccardo Dalla-Favera, Hartmut Döhner, and Stephan Stilgenbauer. Gene expression factors as predictors of genetic risk and survival in chronic lymphocytic leukemia. *Haematologica*, 95(1):102–9, jan 2010. ISSN 1592-8721. doi: 10.3324/haematol.2009.010298. URL http://www.ncbi.nlm.nih.gov/pubmed/19951976http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2805743.

Jesse Krijthe. *Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation*, 2017. URL https://cran.r-project.org/package=Rtsne.

Dan A Landau, Scott L Carter, Petar Stojanov, Aaron McKenna, Kristen Stevenson, Michael S Lawrence, Carrie Sougnez, Chip Stewart, Andrey Sivachenko, Lili Wang, Youzhong Wan, Wandi Zhang, Sachet A Shukla, Alexander Vartanov, Stacey M Fernandes, Gordon Saksena, Kristian Cibulskis, Bethany Tesar, Stacey Gabriel, Nir Hacohen, Matthew Meyerson, Eric S Lander, Donna Neuberg, Jennifer R Brown, Gad Getz, and Catherine J Wu. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*, 152(4):714–26, feb 2013. ISSN 1097-4172. doi: 10.1016/j.cell.2013.01.019. URL http://www.ncbi.nlm.nih.gov/pubmed/23415222http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3575604.

Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*, 1(6):417–425, dec 2015. ISSN 24054712. doi: 10.1016/j.cels.2015.12.004. URL http://www.ncbi.nlm.nih.gov/pubmed/26771021http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4707969http://linkinghub.elsevier.com/retrieve/pii/S2405471215002185.

Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, dec 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0550-8. URL http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8.

Amit K Mittal, Nagendra K Chaturvedi, Rae A Rohlfsen, Payal Gupta, Avadhut D Joshi, Ganapati V Hegde, R Gregory Bociek, and Shantaram S Joshi. Role of CTLA4 in the proliferation and survival of chronic lymphocytic leukemia. *PloS one*, 8(8):e70352, 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0070352. URL http://www.ncbi.nlm.nih.gov/pubmed/23936412http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3731360.

Fionn Murtagh and Pierre Legendre. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, 31(1):274–295, 2014. doi: 10.1007/s00357-014-9161-z. URL http://adn.biol.umontreal.ca/{~}numericalecology/Reprints/Murtagh{_}Legendre{_}J{_}Class{_}2014.pdf.

Christopher C Oakes, Marc Seifert, Yassen Assenov, Lei Gu, Martina Przekopowitz, Amy S Ruppert, Qi Wang, Charles D Imbusch, Andrius Serva, Sandra D Koser, David Brocks, Daniel B Lipka, Olga Bogatyrova, Dieter Weichenhan, Benedikt Brors, Laura Rassenti, Thomas J Kipps, Daniel Mertens, Marc Zapatka, Peter

Lichter, Hartmut Döhner, Ralf Küppers, Thorsten Zenz, Stephan Stilgenbauer, John C Byrd, and Christoph Plass. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nature genetics*, 48(3):253–64, mar 2016. ISSN 1546-1718. doi: 10.1038/ng.3488. URL http://www.ncbi.nlm.nih.gov/pubmed/26780610http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4963005.

Young Mi Oh, Young Eun Kwon, Joo Mi Kim, Sung Jun Bae, Bo Keun Lee, Soon Ji Yoo, Chin Ha Chung, Raymond J. Deshaies, and Jae Hong Seol. Chfr is linked to tumour metastasis through the downregulation of HDAC1. *Nature Cell Biology*, 11(3):295–302, mar 2009. ISSN 1465-7392. doi: 10.1038/ncb1837. URL http://www.ncbi.nlm.nih.gov/pubmed/19182791http://www.nature.com/articles/ncb1837.

Xose S. Puente, Magda Pinyol, Víctor Quesada, Laura Conde, Gonzalo R. Ordóñez, Neus Villamor, Georgia Escaramis, Pedro Jares, Sílvia Beà, Marcos González-Díaz, Laia Bassaganyas, Tycho Baumann, Manel Juan, Mónica López-Guerra, Dolors Colomer, José M. C. Tubío, Cristina López, Alba Navarro, Cristian Tornador, Marta Aymerich, María Rozman, Jesús M. Hernández, Diana A. Puente, José M. P. Freije, Gloria Velasco, Ana Gutiérrez-Fernández, Dolors Costa, Anna Carrió, Sara Guijarro, Anna Enjuanes, Lluís Hernández, Jordi Yagüe, Pilar Nicolás, Carlos M. Romeo-Casabona, Heinz Himmelbauer, Ester Castillo, Juliane C. Dohm, Silvia de Sanjosé, Miguel A. Piris, Enrique de Alava, Jesús San Miguel, Romina Royo, Josep L. Gelpí, David Torrents, Modesto Orozco, David G. Pisano, Alfonso Valencia, Roderic Guigó, Mónica Bayés, Simon Heath, Marta Gut, Peter Klatt, John Marshall, Keiran Raine, Lucy A. Stebbings, P. Andrew Futreal, Michael R. Stratton, Peter J. Campbell, Ivo Gut, Armando López-Guillermo, Xavier Estivill, Emili Montserrat, Carlos López-Otín, and Elías Campo. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*, 475(7354):101–105, jul 2011. ISSN 0028-0836. doi: 10.1038/nature10113. URL http://www.ncbi.nlm.nih.gov/pubmed/21642962http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3322590http://www.nature.com/articles/nature10113.

Alejandro Reyes, Carolin Blume, Vincent Pelechano, Petra Jakob, Lars M Steinmetz, Thorsten Zenz, and Wolfgang Huber. Mutated SF3B1 is associated with transcript isoform changes of the genes UQCC and RPL31 both in CLLs and uveal melanomas. *bioRxiv*, 1(1):1, 2013. doi: 10.1101/000992. URL https://www.biorxiv.org/content/biorxiv/early/2014/07/13/000992.full.pdf.

John C Riches, Conor J O'Donovan, Sarah J Kingdon, Fabienne McClanahan, Andrew J Clear, Donna S Neuberg, Lillian Werner, Carlo M Croce, Alan G Ramsay, Laura Z Rassenti, Thomas J Kipps, and John G Gribben. Trisomy 12 chronic lym-

phocytic leukemia cells exhibit upregulation of integrin signaling that is modulated by NOTCH1 mutations. *Blood*, 123(26):4101–10, jun 2014. ISSN 1528-0020. doi: 10.1182/blood-2014-01-552307. URL http://www.ncbi.nlm.nih.gov/pubmed/24829201http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4073326.

A Rosenwald, A A Alizadeh, G Widhopf, R Simon, R E Davis, X Yu, L Yang, O K Pickeral, L Z Rassenti, J Powell, D Botstein, J C Byrd, M R Grever, B D Cheson, N Chiorazzi, W H Wilson, T J Kipps, P O Brown, and L M Staudt. Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *The Journal of experimental medicine*, 194(11):1639–47, dec 2001. ISSN 0022-1007. URL http://www.ncbi.nlm.nih.gov/pubmed/11733578http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2193523.

Davide Rossi and Gianluca Gaidano. The clinical implications of gene mutations in chronic lymphocytic leukaemia. *British journal of cancer*, 114 (8):849–54, apr 2016. ISSN 1532-1827. doi: 10.1038/bjc.2016.78. URL http://www.ncbi.nlm.nih.gov/pubmed/27031852http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4984803.

Katrin Sameith, Saman Amini, Marian J. A. Groot Koerkamp, Dik van Leenen, Mariel Brok, Nathalie Brabers, Philip Lijnzaad, Sander R. van Hooff, Joris J. Benschop, Tineke L. Lenstra, Eva Apweiler, Sake van Wageningen, Berend Snel, Frank C. P. Holstege, and Patrick Kemmeren. A high-resolution gene expression atlas of epistasis between gene-specific transcription factors exposes potential mechanisms for genetic interactions. *BMC Biology*, 13(1):112, dec 2015. ISSN 1741-7007. doi: 10.1186/s12915-015-0222-5. URL http://bmcbiol.biomedcentral.com/articles/10.1186/s12915-015-0222-5.

Matthew Thomas, Marco Calamito, Bhaskar Srivastava, Ivan Maillard, Warren S Pear, and David Allman. Notch activity synergizes with B-cell-receptor and CD40 signaling to enhance B-cell activation. *Blood*, 109(8):3342–50, apr 2007. ISSN 0006-4971. doi: 10.1182/blood-2006-09-046698. URL http://www.ncbi.nlm.nih.gov/pubmed/17179224.

Leif Väremo, Jens Nielsen, and Intawat Nookaew. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research*, 41(8):4378–4391, apr 2013. ISSN 1362-4962. doi: 10.1093/nar/gkt111. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt111.

Thorsten Zenz, Daniel Mertens, Ralf Küppers, Hartmut Döhner, and Stephan Stilgenbauer. From pathogenesis to treatment of chronic lymphocytic leukaemia.

Nature Reviews Cancer, 10(1):37–50, jan 2010. ISSN 1474-175X. doi: 10.1038/nrc2764. URL http://www.ncbi.nlm.nih.gov/pubmed/19956173http://www.nature.com/articles/nrc2764.

Wen-Ya Zhou, Hong Zheng, Xiao-Ling Du, and Ji-Long Yang. Characterization of FGFR signaling pathway as therapeutic targets for sarcoma patients. *Cancer biology & medicine*, 13(2):260–8, jun 2016. ISSN 2095-3941. doi: 10.20892/j.issn.2095-3941.2015.0102. URL http://www.ncbi.nlm.nih.gov/pubmed/27458533http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4944539.

6 Supplements

6.1 Data set

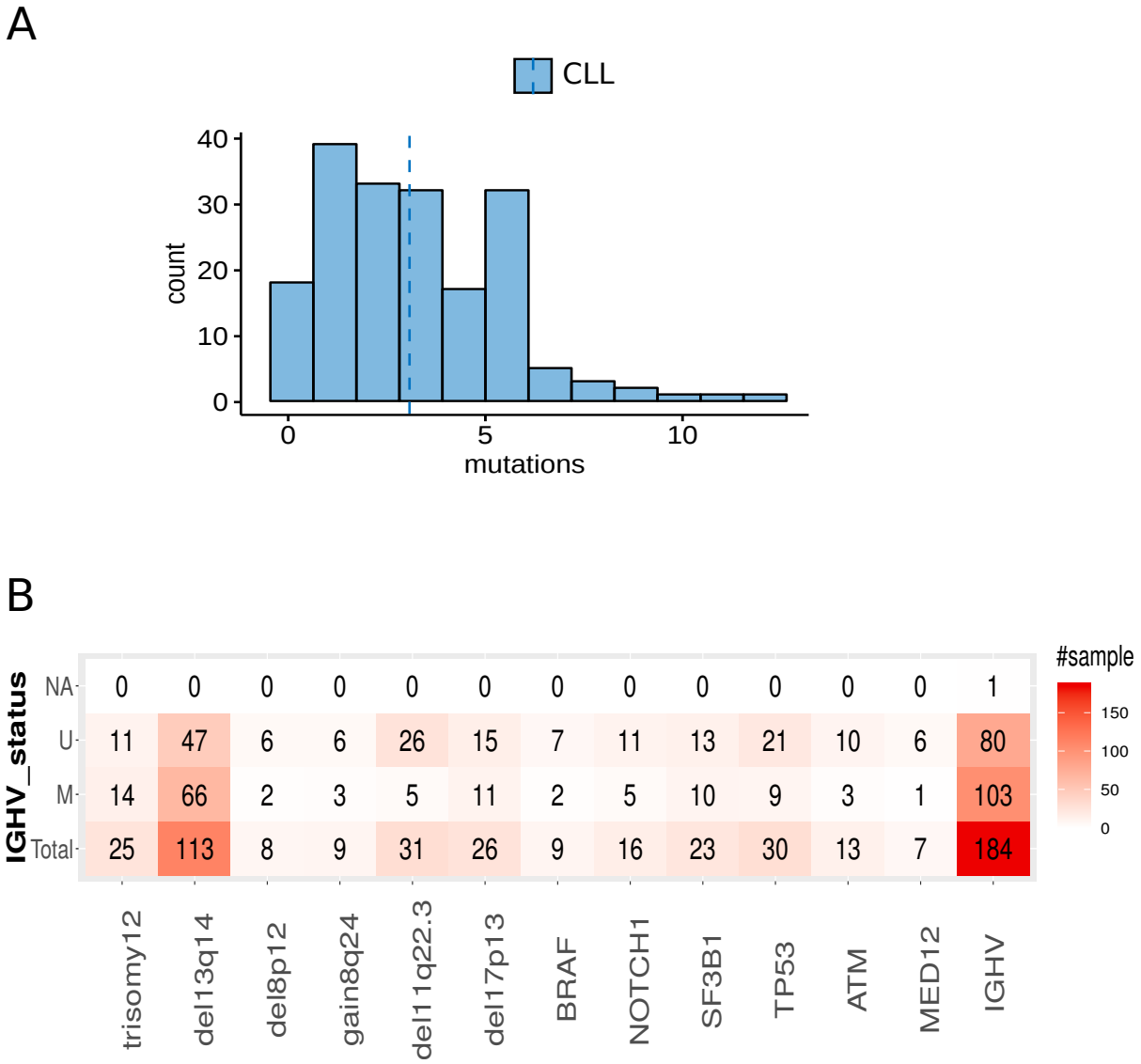

Figure 7: **Occurence of mutations:** A) Histogramm of mutational load per sample. On avarage each sample carries 3 of the 88 tested recurrent mutations and CNVs. B) Occurence of mutations and CNVs by IGHV status.