# Transcriptional Profiling Reveals Strong Impact of Major Molecular Disease Subgroups and Mixed Epistasis in Chronic Lymphocytic Leukemia

Thorsten Zenz, MD,[1] Almut Luetge,[*,2] Junyan Lu, PhD,[*,3] Huellein Jennifer, PhD,[*,2] Sascha Dietrich, MD,[4] Leopold Sellner, MD,[*,5] Wolfgang Huber, PhD[*,3]

[1]Department of Medical Oncology and Hematology, University Hospital and University of Zürich, Zürich, Switzerland

[2]EMBL, Heidelberg, Germany

[3]European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

[4]Department of Medicine V, University of Heidelberg, Heidelberg, Germany

[5]Department of Internal Medicine V, University Hospital Heidelberg, Heidelberg, Germany

While recurrent mutations in CLL have been extensively catalogued, how driver mutations affect disease phenotypes remains incompletely understood. To address this, we performed RNA sequencing on 184 CLL patient samples and linked gene expression changes to molecular subgroups, gene mutations and copy number variants.

Library preparation was performed according to the Illumina TruSeq RNA sample preparation v2 protocol. Samples were paired-end sequenced and 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. Internal trimming with STAR version 2.5.2a was used to remove adapters before mapping. Mapping was performed using STAR version 2.5.2a against the Ensembl human reference genome release 75 (Homo sapiens GRCh37.75). 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 with default parameters and union mode. Thus, only fragments unambiguously overlapping with one gene were counted. The count data were then imported into R (version 3.4) for subsequent analysis.

We identified robust and previously unknown gene expression signatures associated with recurrent copy number variants (including trisomy 12, del11q22.3, del17p13, del18p12 and gain8q24), gene mutations

(TP53, BRAF and SF3B1) and the mutation status of the immunoglobulin heavy-chain variable region (IGHV). The most profound gene expression changes were associated with IGHV, methylation groups and trisomy 12. We found evidence for a significant influence of CNVs beyond the gene dosage effect. In line with these observations, unsupervised clustering showed that these major biological subgroups form distinct clusters and are discernible by unsupervised clustering (IGHV, methylation groups and trisomy 12).

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% . In total 9.5 % of variance within gene expression was associated with the IGHV status. These data suggest a much larger impact on transcriptional changes than previously detected (Ferreira et al. 2014), a finding much more in line with the key impact of IGHV on clinical course and biology of disease.

We found distinct expression pattern of up- and downregulated genes for trisomy 12 samples. Even though many upregulated genes are located on chromosome 12, the majority of differentially expressed genes are indeed distributed among the other chromosomes and cannot be therefore not be ascribed to a simple gene dosage effect.

To investigate the role of genetic interactions, we tested the collaborative effect on gene expression phenotypes. We investigated epistatic gene expression changes for IGHV status and trisomy 12. Epistasis was defined as a non-linear effect on gene expression between sample with both variants co-occuring and the single variants alone. In total 893 genes showed specific expression pattern in a combined genotype (padj<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. 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. 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.

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 in a previously underappreciated fashion. This includes epistatic interaction between trisomy 12 and IGHV. Using a novel way to describe coordinated changes we can group genes into sets related to buffering, inversion and suppression.

**Disclosures**
**Sellner:** *Takeda:* Employment.

# Author notes

*Asterisk with author names denotes non-ASH members.