**JMMLC\_Outline\_MHartmann\_03**

Title suggestions

**Multi omics profiling determines CD52 as a prognostic factor and therapeutic target of high-risk leukemia stem cells in JMML**

**Epigenetic reprogramming induces CD52 as a subgroup-specific prognostic biomarker and therapeutic target in JMML leukemia stem cells**

**Epigenetic reprogramming characterizes high-risk leukemia stem cells in JMML with CD52 as a subgroup-specific prognostic biomarker and therapeutic target**

\_\_

Aspect 1:

**Multi-modal analysis of epigenetic subgroups (risk groups) in JMML**

A. Schematic of concept and experimental overview.

> Comprehensive molecular characterization to identify novel prognostic markers and therapeutic targets in JMML.

B. Genetic and epigenomic annotation of cohort.

> Cohort comprises entire spectrum of epigenotypes / risk-groups

ToDo:

- New schematic including more information about the approaches and specific aims.

\_\_

Aspect 2:

**JMML-related aberrations affect total hematopoiesis starting from HSPCs**

A-E. scRNA-seq of total hematopoiesis of JMML and healthy references.

A. UMAP colored by dataset.

> scRNAseq reveals that global structures of the hematopoietic system are conserved in JMML.

B. Densityplots of lineage marker expression in adult BM and JMML samples. Marker genes were selected from Hay et al. 2018 Experimental Hematology).

> All major hematopoietic lineages can be identified in JMML.

C. UMAP colored by cell types.

> Label transfer from adult hematopoiesis reveals that virtually all major hematopoietic cell types are expressed in JMML.

D. Stacked barplot of cell type frequencies across JMML patients for CD34+ enriched cells and MNCs.

> Spleens of low-risk patients are biased towards B cells, as expected.

> High risk patients reveal expanded extramedullary hematopoiesis.

E. Integrated pseudotime analysis per lineage of JMML in the context of normal hematopoiesis.

> Pronounced differences occur at the beginning of the differentiation trajectories in the HSPC compartment.

\_\_

Aspect 3:

**Subgroup-specific heterogeneity reveals HSPCs as the compartment of origin in JMML**

A-B. FACS analysis of JMML HSPCs.

A. Contourplots of FACS analysis of JMML HSPCs (Lin-CD34+CD38-).

> Large immunophenotypic heterogeneity across patients.

> Appearance of aberrant CD45RA+CD90+ HSPCs in HM patients only.

B. Quantification of Lin-CD34+CD38-CD45RA+CD90+ in validation cohort (n = 29).

> Aberrant CD45RA+CD90+ HSPCs are HM-specific.

C-I. Ultra-low input WGBS (µWGBS) of flow-cytometrically enriched JMML CD34+ HSPCs.

C. Boxplots of genome-wide CpG island (CGI) methylation

> Genome-wide methylation differences in JMML HSPCs across patients.

> Subgroup-related CpG island methylation phenotype (CIMP).

D. Principal component analysis (PCA) of µWGBS data of JMML HSPCs.

> CD45RA/CD90 quadrants of JMML HSPCs show patient-specific methylation patterns and can be used as replicates.

> Epigenetic subgroups of JMML are conserved at the top of the hematopoietic differentiation hierarchy in HSPCs.

> Epigenetics subgroups are most dominant factor beyond the genotype.

E. Methylation-based cell type classification (Farlik et al. 2016 Cell Stem Cell).

> Global methylation differences are probably predominantly disease-driven and not related to differences in cell type composition, as assumed based on the intra-patient heterogeneity of immunophenotypes from JMML HSPCs.

> The CD45RA/CD90 quadrants do not determine epigenetic cell identity.

> WGBS samples have HSC signature and can be used as replicates for DMR calling between high-risk (HM) and low-risk (LM+IM) JMML.

F. Heatmap of differentially methylated regions (DMRs) between HM and nonHM patients.

> Genome-wide differences (>30,000 DMRs) separate high- and low-risk HSPCs.

> Neither the genotype nor the immunophenotypic CD45RA/CD90 quadrant determine subclustering of samples.

> Patient-specific subclustering confirms suitability of samples as replicates.

G. Bubbleplot of genome-wide enrichment analysis of DMRs in chromatin features (ChromHMM) using locus overlap analysis (LOLA).

> Specific chromatin features are differentially methylated, which suggests functional consequences in JMML HSPCs.

> Developmentally regulated regions (bivalent enhancers and promoters (TSSBiv, BicFlnk, EnhBiv) and polycomb-repressed (ReprPC) domains) are differentially methylated.

H. Bubbleplot of the genome-wide analysis of enriched motifs across DMRs using Hypergeometric Optimization of Motif EnRichment (HOMER).

> Strongest enrichment of central hematopoietic and developmental transcription factors across DMRs.

I. Genome browser tracks of µWGBS of JMML HSPCs (P1-P8) and healthy normal controls from cordblood (neo) and adult bonemarrow (adu) HSCs, and aggregated bulk methylation array data of XXX JMML patients (LM, IM, HM).

> Well known hematopoietic factors such as FLI1 are affected by differential methylation.

> Conservation of the epigenotypes from bulk JMML samples in highly purified HSPCs.

J-K. scRNA-seq of flow-cytometrically enriched JMML CD34+ HSPCs.

J. UMAP colored by patient.

> Patient-specific patterns dominate global structure of JMML HSPC compartment.

K. Heatmap of differentially expressed genes (DEGs) between HSCs of HM and non-HM patients.

> DEG analysis between HSCs of high- and low-risk patients reveal risk group-specific programs.

> Markers of healthy and malignant hematopoiesis, differentiation, inflammation, and development (HOXA9, HOXA10, HBG2).

> Different myeloid markers are enriched across JMML HSCs (AREG, HES4, BEST1, EREG, TCF4, SELL, CD74, CLEC7A, LST1)

> Non-HM HSCs are enriched for B cell markers (IGLL1, DNTT), whereas HM HSCs are enriched for NK/T markers (CD69, CD96, CITED2, HCST, HOPX, LTB).

> Specific HSC (AVP, CRHBP, HOXA9, HOXA10, CD34, CD164) and non-HSC (CD52, CD69, CD96) markers are differentially expressed across subgroups.

> Potentially novel prognostic surface markers.

L-M. Integrative analysis of scRNA-seq and µWGBS data of JMML HSPCs.

L. Juxtaposition of heatmaps from differentially expressed genes correlating with differentially methylated regions. Of note, multiple DMRs can be associated with single genes.

> Association between DNA methylation and gene expression reveal subgroup-specific epigenetically regulated programs in JMML HSCs.

M. Scatterplots of genes showing correlation between gene expression and DNA methylation. A subset of examples is shown in the major figure.

> Impact of DNA methylation on HSPC homeostasis associated to oncogenic RAS signaling (RAB12, RALA, RALB), tumorigenesis (MAF, MYCN, VGLL4), epigenetic gene regulation (DOT1L, DROSHA, ERG, IDH3B, KCNQ1OT1, SETBP1, SETD7, TET3, TOX2), hematopoietic differentiation (AVP, CRHBP, CLEC9A, RUNX1/2/3, THY1/CD90), function (CD8A, CD96, CXCR4, HCST, HLA-DQA1, HLA-DQA2, IL12RB1, IL17RE, LAT), and development (HMGA2, HOXA10-AS, HOXB-AS3, HOXB3, LHX3, NPM2, PBX1).

> The quint-neg LM patient P1 shows both typical pan-JMML and LM-specific signatures on both layers – transcriptomes and methylomes. It is clearly distinct from healthy HSCs and should therefore be categorized as JMML.

ToDo:

- Get epi-/genotype information of unannotated patients from FACS cohort (C. Flotho)

\_\_

Aspect 4:

**Distinct developmental programs in HSPCs across JMML subgroups**

A. Schematic summarizing the systematic comparison of HSPCs from JMML subgroups and four different developmental stages, namely fetal liver (fet), cord blood (neo), juvenile (juv) and adult (adu) bone marrow.

> JMML hematopoiesis in the context of different developmental stages.

B. UMAP of scRNA-seq data from JMML and four healthy hematopoietic references colored by dataset. References for normal hematopoiesis encompass fetal liver (fet, Popescu et al.), cord blood (neo, Human Cell Atlas), as well as juvenile (juv) and adult (adu) bone marrow. JMML samples are colored per patient.

C. Dotplot showing the percentage of cells (dot size) and the corresponding average expression (color) of selected marker genes of postnatal (grey bar) or prenatal (black bar) HSCs, and the corresponding values per JMML patient (blue, yellow, and red bars).

> Well known hematopoietic and cancer genes, as well as developmental factors are differentially regulated across JMML patients.

> LM HSCs are enriched for postnatal marker genes, HM HSCs are enriched for fetal marker genes, whereas there is a large degree of heterogeneity across patients.

D. Violinplots showing the anticorrelated expression of developmental HSC markers in healthy and malignant HSCs. AVP and CRHBP are the most significant adult HSC markers (Hay et al.). HMGA2 is a fetal HSC marker. SPINK2 is both, a fetal (Popescu et al.) as well as a postnatal HSC marker.

> HM and LM HSCs show reciprocal developmental HSC signatures.

> HM HSCs show upregulated fetal HSC signature genes, whereas LM HSCs upregulate postnatal HSC markers.

> Intra-patient heterogeneity for every marker, small subsets of HSCs behaving different than majority.

E-F. Heatmaps showing the top 50 ‘unique’ differentially expressed genes (DEGs) per reference. DEGs between healthy hematopoietic references representing the differences in normal HSC development.

E. Heatmap ordered by samples.

> Specific transcription programs distinguish healthy HSCs across development stages from fetal to adult hematopoiesis.

> JMML HSCs show a complex transcription signature of markers form different stages across patients.

F. Heatmap clustered

> Clustering reveals subgroup-specific programs of developmental signature genes.

> Upregulation of specific healthy HSC signature genes across patients.

G. Gene ontology analysis of transcription programs from DEG clustering in G.

> …

H-I. Logistic regression to determine the similarity of JMML HSCs to all hematopoietic cell types (n = 116) of all four developmental stages examined here. A value above 0.5 (pink) and below 0.5 (grey) refers to similarity and dissimilarity, respectively, relative to the reference cell type. 0.5 is random (white).

H. Heatmap showing the complete analysis including all reference cell types. grey

I. Enlargement of the top part of the heatmap in H showing all cell types with a positive

> Subgroup-specific similarities to different developmental stages

> Fetal signature in HMs, postnatal signature in LMs, neonatal signature pan-JMML.

\_\_

Aspect 5:

**Epigenomic reprogramming of high-risk HSPCs results in accelerated aging and onco-fetal signatures**

A-F. Integrated analysis of whole-genome bisulfite sequencing (WGBS) data from JMML and healthy HSCs from four developmental stages (as described in figure 4.A). FEL corresponds to fetal liver, FES to fetal spleen, NEO to neonatal, JUV to juvenile, and ADU to adult HSCs.

A-C. Identification of DNA methylation changes in normal HSC development for the identification of the phylogenetic relationships between JMML and normal HSCs.

A. Genomic regions whose methylation status changes during development from fetal to adult HSCs. FET corresponds to both FES and FEL samples.

> Epigenomic reprogramming in the course of healthy HSC development.

B. Phylogenetic tree based on genomic regions defined in A.

> Selected methylation regions recapitulate the normal HSC development.

C. Integration of JMML samples into the phylogenetic tree of the normal HSC development.

> All patients have a postnatal signature, suggesting that they passed through age-appropriate development.

> LM HSCs follow the normal HSC trajectory and appearing in the postnatal space around the juvenile reference.

> IM and HM HSCs branch off the normal trajectory, suggesting amplified epigenetic aberrations.

D. Principal component analysis (PCA) of healthy and malignant HSCs using the 10,000 most variable CpGs (mvCpGs).

> LM HSCs are epigenetically very similar to postnatal HSCs.

> PC1 separates HM HSCs from all other samples.

> PC2 shows a convergence of HM HSCs to healthy fetal HSCs.

E-F. Average methylation of epigenetic scars across normal and tumor samples. Hypo and hyper correspond to scars whose methylation de- or increase during developmental state transition. Epigenetic scars are per definition DNA methylation changes that occur at a single step during development but do not change downstream of this event.

E. Epigenetic scars from the transition between fetal and neonatal stages.

> As expected, fetal samples cluster away from the other healthy reference samples of neo- to postnatal HSCs.

F. Epigenetic scars from the transition between neonatal and postnatal stages.

> All JMML HSCs cluster with the healthy neo- to postnatal HSCs, suggesting that the HSCs from all patients passed the age-appropriate development beyond birth.

> Additionally, HM patients partially gain fetal signature, across patients in a progressive manner towards higher methylated states.

> This result suggests an onco-fetal reprogramming rather than a fetal origin of those patients.

G. Boxplots showing the average methylation of regions that have been described to be epigenetically regulated upon ageing of HSCs (Adelman et al.).

> Promoters reveal accelerated ageing in HM HSCs.

> Enhancers confirm Onco-fetal reprogramming of HM HSCs.

H. Boxplots showing the average methylation of JMML HSPCs of the signature CpGs which were associated with proliferation activity (so called ‘epiTOC’, Young et al. 2016 Genome Biology).

> Increased proliferative history towards higher methylated HSCs.

I. Boxplots epigenetic clock on methylation array data of XX patients…

> Reveals a conservation of the increased proliferation signature in downstream hematopoietic bulk samples in a larger cohort with methylation array data.

J. Horvath’s epigenetic clock applied to methylation array data of XX patients…

> Reveals extremely increased ‘epigenetic age’ of JMML patients, which might be a consequence of the increased proliferation history of hematopoietic cells in JMML.

>> High-risk HM HSCs reveal both an aging phenotype on the level of methylation and signatures pointing towards progressive onco-fetal reprogramming in the course of the disease.

ToDo:

- Exclude scIGMT-seq data from tree and scars analysis!

- Epigenetic clock on HSPCs?

\_\_

Aspect 6:

**Subgroup-specific aberrations reveal CD52 as a prognostic biomarker for high-risk HSPCs.**

A. Heatmap showing the distances of samples in the phylogenetic tree. Manhatten distances were calculated based on the branch length per subgroup.

> Fetal HSCs are the closest to neonatal and the furthest to adult HSCs, as expected.

> All JMML subgroups reveal proximity to all postnatal HSCs including neonatal, juvenile, and adult samples.

> All JMML HSCs examined here reveal a postnatal epigenetic signature.

B. Upsetplot of all differentially methylated regions (DMRs) called per subgroup against single postnatal reference methylomes, which were used as a union per JMML subgroup for the overlap analysis.

> Subgroup- and pan-JMML-specific epigenomic aberrations.

C. Dotplot showing the percentage of cells (sot size) and the corresponding average expression (color) of selected surface marker genes, which show subgroup-specific expression patterns across subgroups and relative to all healthy references. The annotation ‘DMR’ marks all genes that are differentially methylated. ‘Drugs’ marks factors for which therapeutics exist at least in clinical trial. ‘HM’ marks genes which significantly correlate between transcription in scRNA-seq data of HSCs and DMR methylation exclusively in HM patients.

> 9 surface markers appear to be upregulated in higher-risk patients and relative to healthy HSCs of all stages.

> The methylation and expression of CD52, CD69, and CD164 are significantly correlating, which suggests that those genes are promising candidates for prognosis and therapy.

> CD52 is the only gene which gene expression correlates exclusively in HM HSCs.

> And only for CD52 an approved drug exists that can be used to validate the functional relevance of CD52.

> Furthermore, CD52 is a promising candidate because

- prognostic factor in CLL + AML

- diff. meth in CN-AML

- conditioning of allo HSCT

D. Locusplot showing the methylation beta values in the CD52 locus in HSCs across JMML patients and heathy references.

> Diff. methylation of CD52 across JMML subgroups and relative to healthy references.

> Intermediate beta values reveal intra-patient heterogeneity of CD52 methylation.

E. Violinplots showing the expression of CD52 in healthy and malignant HSCs.

> CD52 is upregulated in HM-HSCs.

> Intra-patient heterogeneity of CD52 expression.

F-G. Analysis of CD52 surface protein expression using FACS.

F. Boxpot showing the percent of CD52+ Lin-CD34+CD38- cells.

G. Violinplot showing the MFI of CD52 across subgroups.

> Cell surface expression of CD52 is upregulated in HM HSCs.

> Confirmation of the intra-patient heterogeneity of CD52 surface expression, as seen at the transcriptional and methylation levels.

\_\_

Aspect 7:

**An anti-CD52 treatment efficiently depletes HM LSCs and improves survival in a JMML PDX model**

A. Schematic of patient-derived xenograft (PDX) mouse experiments. The analysis of the primary transplants focuses on the immunophenotypic analysis of the composition of engrafted human cells. The analysis of the secondary transplants focuses on the survival analysis of secondary recipients following treatment of primary transplanted mice.

B. Quantification of engrafted human CD45+ cells 4 weeks after treatment with anti-CD52 (light grey) or control (dark grey) after primary transplantation in four tissues (bone marro, spleen, liver, and lung).

> anti-CD52 treatment efficiently depletes JMML cells.

C. FACS analysis of primary transplanted PDX mice with an antibody panel specific for HSPCs, including CD52.

> All kind of HSPCs including HSCs, MPPs, LMPPs, CMPs, and GMPs are efficiently depleted upon treatment.

> CD52+ cells are efficiently depleted, including HSPCs.

D. FACS analysis of primary transplanted PDX mice with an antibody panel specific for neutrophils, erythrocytes, and lymphocytes.

> Hematopoietic cells of all major lineages are depleted upon treatment, including neutrophil granulocytes, erythrocytes, B and T cells.

> This affects all tissues, here only shown for spleen.

E. FACS analysis of primary transplanted PDX mice with an antibody panel specific for HSPCs and myeloid cells.

> Various kinds of myeloid subpopulations are depleted upon treatment, including maturing neutrophils, monocytes, and blasts.

> HSPCs are depleted upon treatment.

> This affects all tissues, here only shown for bone marrow.

>> High specificity to deplete CD52+ cells.

>> High efficacy to virtually deplete the entire malignant hematopoietic system.

F. Representative FACs plot showing CD52+ values of human CD45+ JMML cells with (Alemtuzumab) and without (Control) anti-CD52 treatment.

> Efficient depletion of CD52 + cells.

> Heterogeneity of CD52 expression per mouse, in line with the observed patterns on the transcriptional and methylation level of patient-specific HSPCs.

G. Kaplan-Meyer curve summarizing the survival rate of secondary transplanted PDX.

> Anti-CD52 treatment results in an improved survival for recipient mice.

H. Engraftment of human CD45+ cells with and without Alemtuzumab treatment…

> …

>> CD52 is a functionally relevant factor in JMML pathogenesis

ToDo:

- JMML LSC panel FACS data of Xenos (Erlacher)

- Repeat primary transplantations for overall 2 patients and up to 10 mice each (Erlacher)

- Repeat secondary transplantations for overall 2 patients and up to 10 mice each (Erlacher)

\_\_

Aspect 8:

**A native single-cell multi-omics analysis reveals CD52 as a marker of high-risk leukemia stem cells**

A-F. Multi-omics analysis applying our scIGMT-seq framework to integrate immunophenotypes, genotypes, methylomes, and transcriptomes per single cell.

A. tSNE of the multi-omics factor analysis (Mofa) of scIGMT-seq data from JMML HSPCs.

> The HSPC compartments of HM and IM patients contain two different subpopulations – HML or IML cells, rspectively, and LML cells.

B. tSNE of the multi-omics factor analysis (Mofa) of scIGMT-seq data from JMML HSPCs with projected DNA methylation programs.

> LML cells mimic LM cells on the single-cell methylation layer.

C. Heatmap to compare bulk with single-cell methylation data as pseudobulks. Epigenetic programs (EPs) are based on DMRs from HM vs non-HM patients by hierarchical clustering with manhattan distance.

> Pseudobulks of scIGMT data recapitulate established methylome signatures of HM vs non-HM cells.

> Average methylation of pseudobulks from HML and LML cells mimic methylation patterns of bulk HSPC methylomes, suggesting an association with risk in JMML.

D. Densityplot comparing the 1% of the most discriminating variants between HML and LML pseudobulks with a permutation of revealing the most discriminating (red, 95% quantile of random distribution) variants between HML and LML pseudobulk cells in transcriptome, methylome, and merged data.

> Non-random discriminating sequence variants exist that distinguish HML and LML cells.

E. Heatmap showing the cosine similarity of each cell to the LML and HML SNV profile determined before.

> HML and LML cells constitute distinct genetic clones.

F. Heatmap integrating all layers of information from scIGMT data. Depicted are per cell: the expression values of the most significant HML and LML marker genes, the average methylation values for all epigenetic programs from figure 8.C, and the HSPC transcription signatures of the 4 different developmental stages (logistic regression).

> HML- and LML-specific transcription signatures further refine the high- and low-risk signatures from 10X data and include CD52 as a marker gene.

> The epigenetic programs strongly correlate with HSPC subpopulations.

> Post- and prenatal HSPC signatures are differentially enriched in HSPC subpopulations.

> Immunophenotypes of classical HSPC markers do not correlate with HSPC subpopulations.

G. Violinplot showing the expression of CD52 per single cell in HML and LML cells.

> CD52 is a high-confidence prognostic candidate gene of high-risk HML cells.

H. UMAPs showing the calculated HML and LML scores on high-throughput 10X scRNA-seq data

> Transcriptional signatures of HML and LML HSPCs are a pan-JMML phenomenon.

I. Barplot of the logarithmic ratios of HML to LML cells.

> High- (HML) and low-risk (LML) HSPCs exist across all 8 patients analysed in this project.

> The only difference is the ratio of HML to LML cells, which might explain intermediate methylation levels in bulk methylome analyses.

J. Multi-step model of pathogenesis in JMML. HSPCs acquire a first hit, most likely a somatic mutation, which leads to the expansion of a less aggressive clone and still allows remission. A second hit generates an even more proliferative and more aggressive clone. The second hit can be a secondary mutation or another non-genetic event. The second clone expands and overgrows the first clone, leading to intermediate to increased methylation values at diagnosis. It is not clear what finally leads to overt leukemia.

> Novel pathogenetic mechanism explaining methylation, risk, and anti-CD52 efficacy.

ToDo:

- Sort bulk LML and HML cells for genotyping – new samples required.

- Generate scIGMT-seq data for all 8 patients – new samples required.

**Discussion:**

> First demonstration of the conservation of the epigenetic subgroups in HSPCs.

> Novel prognostic biomarkers (subgroup-specific surface markers), which can be exploited for fast, accurate, and cheap clinical risk stratification.

> Novel therapeutic target, which efficiently depletes malignant JMML cells from HSPCs until differentiated immune cells.

> Anti-CD52 can be used as an immune therapeutic and for conditioning of allo HSCT in parallel.

> De novo identification of high- and low-risk HSPCs present in individual patients, which is possible by use of a sophisticated native single-cell multi-omics approach.

> Novel hypothesis for cellular and molecular pathogenesis pathways in JMML.

> …