



Research paper

Chronic myelomonocytic leukaemia stem cell transcriptomes anticipate disease morphology and outcome

Daniel H Wiseman^{a,*}, Syed M Baker^{b,1}, Arundhati V Dongre^a, Kristian Gurashi^a, Joanna A Storer^a, Tim CP Somervaille^c, Kiran Batta^{a,*}

^a Epigenetics of Haematopoiesis Laboratory, Division of Cancer Sciences, The University of Manchester, Manchester M20 4GJ, UK

^b Division of Informatics, Imaging & Data Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester M13 9PL, UK

^c Leukaemia Biology Laboratory, Cancer Research UK Manchester Institute, The University of Manchester, Manchester M20 4GJ, UK



ARTICLE INFO

Article History:

Received 28 May 2020

Revised 30 June 2020

Accepted 7 July 2020

Available online xxx

Keywords:

CMML

Stem cells

sc-RNA Seq

Leukaemia

ABSTRACT

Background: Chronic myelomonocytic leukaemia (CMML) is a clinically heterogeneous stem cell malignancy with overlapping features of myelodysplasia and myeloproliferation. Over 90% of patients carry mutations in epigenetic and/or splicing genes, typically detectable in the Lin⁻CD34⁺CD38⁻ immunophenotypic stem cell compartment in which the leukaemia-initiating cells reside. Transcriptional dysregulation at the stem cell level is likely fundamental to disease onset and progression.

Methods: We performed single-cell RNA sequencing on 6826 Lin⁻CD34⁺CD38⁻ stem cells from CMML patients and healthy controls using the droplet-based, ultra-high-throughput 10x platform.

Findings: We found substantial inter- and intra-patient heterogeneity, with CMML stem cells displaying distinctive transcriptional programs. Compared with normal controls, CMML stem cells exhibited transcriptomes characterized by increased expression of myeloid-lineage and cell cycle genes, and lower expression of genes selectively expressed by normal haematopoietic stem cells. Neutrophil-primed progenitor genes and a MYC transcription factor regulome were prominent in stem cells from CMML-1 patients, whereas CMML-2 stem cells exhibited strong expression of interferon-regulatory factor regulomes, including those associated with IRF1, IRF7 and IRF8. CMML-1 and CMML-2 stem cells (stages distinguished by proportion of downstream blasts and promonocytes) differed substantially in both transcriptome and pseudotime, indicating fundamentally different biology underpinning these disease states. Gene expression and pathway analyses highlighted potentially tractable therapeutic vulnerabilities for downstream investigation. Importantly, CMML patients harboured variably-sized subpopulations of transcriptionally normal stem cells, indicating a potential reservoir to restore functional haematopoiesis.

Interpretation: Our findings provide novel insights into the CMML stem cell compartment, revealing an unexpected degree of heterogeneity and demonstrating that CMML stem cell transcriptomes anticipate disease morphology, and therefore outcome.

Funding: Project funding was supported by Oglesby Charitable Trust, Cancer Research UK, Blood Cancer UK, and UK Medical Research Council.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Chronic myelomonocytic leukaemia (CMML) is a myelodysplastic/myeloproliferative overlap syndrome typically affecting older people, characterized by bone marrow (BM) failure, leukocytosis and monocytosis [1,2]. CMML is clinically heterogeneous and categorized by both percentage accumulation of blasts and promonocytes (stages

CMML-0/-1/-2), and by circulating white blood cell (WBC) count into dysplastic ($<13 \times 10^9/L$) and proliferative ($>13 \times 10^9/L$) forms [2,3]. Advanced stage and proliferative disease are both associated with shorter survival and higher risk of leukaemic transformation [4]. The genomic landscape is dominated by mutations in epigenetic modifier (e.g. *TET2*; *ASXL1*) and splicing (e.g. *SRSF2*) genes [5,6]. Despite its clinico-pathological heterogeneity, most patients have dismal prognosis with few treatment options [7,8]. Haematopoietic stem cell (HSC) transplantation is potentially curative but only an option for the minority of younger patients [9]. For the majority, hydroxycarbamide remains standard of care, affording cytoreduction without substantially altering disease biology, clonal composition or

* Corresponding authors.

E-mail addresses: daniel.wiseman@manchester.ac.uk (D.H. Wiseman), kiran.batta@manchester.ac.uk (K. Batta).

¹ These authors contributed equally.

Research in context

Evidence before this study

Chronic myelomonocytic leukaemia (CMML) is a rare haematological malignancy with dismal prognosis. Currently there are no treatment options available for this disease, largely due to inadequate mechanistic understanding of disease initiation and progression. The mutational landscape in CMML and insights from other myeloid malignancies implicate transcriptional dysregulation at the level of the disease initiating haematopoietic stem cells in CMML leukaemogenesis. However, to date no studies have directly investigated the transcriptome of these cells in detail. Such understanding is critical for the rational design of novel targeted therapeutic strategies to address this major area of unmet clinical need.

Added value of this study

This is the first study to evaluate the CMML transcriptome at single cell level, and the first to directly evaluate transcriptional dysregulation specifically in the immunophenotypic stem cell compartment. It reveals marked heterogeneity, both within and between individual patients, highlighting CMML as a common phenotypic endpoint of quite different transcriptionally dysregulated stem cell programs. It highlights some candidate genes and pathways as potential novel therapeutic targets and, importantly, identified a proportion of transcriptionally near-normal residual stem cells in most cases, raising the potential for an important therapeutic window and sparing of normal haematopoietic output.

Implications of all the available evidence

Our study reveals new insights into CMML biology and provides a useful platform for future hypothesis-driven exploration of potentially tractable therapeutic vulnerabilities.

explain, in part, the inability to cure these diseases without allogeneic stem cell transplantation.

Single-cell RNA-sequencing (scRNA-seq) is a powerful tool for elaborating transcriptional heterogeneity in cell populations [22–24]. We hypothesised that the remarkable clinical heterogeneity observed in CMML might reflect transcriptomic heterogeneity within the disease-initiating stem cell compartment. ScRNA-seq could reveal genes and pathways selectively dysregulated in CMML stem cells, highlighting potentially tractable vulnerabilities for novel therapeutic strategies. It could also shed light on the origin and pathobiology of CMML disease progression: defined by the accumulation of blasts and promonocytes into clinical stages (CMML-0; –1; –2; AML) tightly linked to overall prognosis [1].

We performed scRNA-seq on 6826 sorted Lin[–]CD34⁺CD38[–] immunophenotypic stem cells from seven treatment-naïve CMML patients and three healthy controls using the droplet-based, ultra-high-throughput 10x platform [24]. We identified distinctive CMML stem cell signatures but with a remarkable degree of inter- and intra-patient transcriptomic heterogeneity, distinct patterns of transcription factor network dysregulation, and substantial differences between CMML-1 and CMML-2 evident even at the primitive stem cell level.

2. Materials and methods

Samples: BM mononuclear cell (MNC) samples from CMML patients were obtained from the Manchester Cancer Research Centre Tissue Biobank (initiated with the approval of South Manchester Research Ethics Committee). Samples were donated, with informed consent, from patients presenting to The Christie NHS Foundation Trust (Manchester, UK), and MNC preparations cryopreserved as previously described [25]. All CMML samples were taken at time of diagnosis and before any definitive therapy. Healthy control BM MNC samples were purchased from Stem Cell Technologies (Cambridge, UK, Cat #70,001). **Further details are provided in Table S1**

Data and code availability: The data generated in the study are available at Array Express (E-MTAB-8884) <https://www.ebi.ac.uk/arrayexpress/>.

Fluorescence-activated cell sorting (FACS): BM-MNC samples were thawed, washed and incubated for 30 min on ice with Hoechst 33-258 (Cat #94,403; Sigma Aldrich, Gillingham, UK) and the following antibodies (all at 1:100): anti-CD34-APC (clone 581; #555,824; BD Biosciences, Oxford, UK); anti-CD38-PE (HIT2; #25–0389–42; eBioscience, Altrincham, UK); anti-Lin-FITC (pooled antibodies targeting CD2/3/4/7/8/10/11b/14/19/20/56/235a; #25–0389–71; BD Biosciences, Oxford, UK). Viable Lin[–]CD34⁺CD38[–] cells were sorted using a FACS Aria III (BD Biosciences). To ensure maximum purity with minimal contamination of the targeted population, sorted samples were immediately subjected to a second sort (when cell numbers permitted) using identical gating. Up to 2000 Lin[–]CD34⁺CD38[–] cells were targeted per sample, based on the empirically-determined 10x capture efficiency (40–50%).

Single-cell RNA-sequencing: Each sample was prepared for single-cell capture immediately after sorting to ensure cell viability. Single cell capture was performed as per manufacturer's protocol (Chromium Single Cell v2 chemistry; 10x Genomics; Leiden, The Netherlands), with 12 cycles of whole transcriptome amplification. Barcoded cDNA libraries were pooled for sequencing on an Illumina NextSeq 500 (Illumina, San Diego, CA, USA) across two high-output runs.

Data processing and filtering: FASTQ files were processed using 10x Genomics' custom pipeline Cell Ranger (v2.2.0) using default parameters (E-MTAB-8884). This identifies valid barcodes as cells and counts Unique Molecular Identifiers (UMIs) mapped to each cell, reported by cell count matrix in sparse matrix format. Sequences were mapped to the prebuild hg38 genome. Cell Ranger's aggr

natural history. Hypomethylating agents (HMA) have an important role for a subset of non-proliferative patients, but their utility is mostly restricted to transiently rebalancing haematopoiesis, with inevitable progression and only modest influence on survival [10]. Thus, novel therapeutic targets and strategies are required to improve patient outcome; this will require deeper understanding of the biology driving the malignant evolution of CMML.

While the transcriptome of CMML monocytes and bulk BM cells has been previously evaluated, the malignant stem cells in CMML have not been directly studied at the genome-wide level [11–15]. The Lin[–]CD34⁺CD38[–] BM compartment represents the apex of the haematopoietic hierarchy, typically representing <5% of total CD34⁺ cells and <0.25% of total BM MNCs [16]. Imbued with properties of self-renewal and multilineage differentiation potential, normal HSCs sustain and replenish haematopoiesis throughout life. In most myeloid malignancies there is substantial enrichment of leukaemia-initiating cells (LICs) within this compartment [17–19]. Whereas CD34⁺CD38[–] and even CD34[–] LICs are recognized in some acute myeloid leukaemias (AML), disease-propagating cells in myelodysplasia (MDS) are thought to reside exclusively in the CD34⁺CD38[–] fraction [20]. This may also apply in CMML, with the CD34⁺CD38[–] compartment consistently harbouring initiating and driver mutations [13] and able to initiate disease in murine xenotransplantation experiments [21]. LICs are distinct from downstream blasts and other clonal progeny in their biology and response to therapy. Their persistence following chemotherapy is a major contributor to the high relapse rate in AML, and the primitive nature of MDS/CMML LICs may

command was used to aggregate samples, with default down-sampling normalization enabled. We evaluated three metrics to remove low quality cells: number of UMIs identified per cell barcode (i.e. library size); number of genes called per barcode; and proportion of UMIs mapping to mitochondrial genes. For each a threshold of three Median Absolute Deviations was applied to reject poor quality cells. Violin plots for each metric were visualized to exclude outlier distributions (no potential doublets/multiplets were observed). Genes with average counts below 0.01 were excluded, being minimally informative and unreliable for statistical inference [26]. A total of 12,695 genes were retained for downstream analysis. To consider the effect of variable library size for each cell, raw counts were normalized using a deconvolution-based method and then log-transformed [27]. Scater (v1.14.6) and its dependent packages were downloaded from Bioconductor [28].

Cell cycle analysis: We used the Seurat package (version 3.1.4) *CellCycleScoring* function to calculate the cell cycle phase score for each cell using canonical marker genes [29]. For this calculation, we took counts for all cells and log normalized them. Next, we performed cell cycle scoring analysis that gives a score for S and G2/M phase of cell cycle. The cell cycle phase is then determined based on a highest positive score given for S or G2/M phase of the cell cycle. Any cell not scoring positive for either of these phases is assigned to G1/G0 phase. Canonical marker genes used for scoring were loaded from Seurat package [29]. No corrections for cell cycle were made, in view of the possibility that cell cycle differences were an important biological variable in comparing cells from different samples in this study.

Visualization and clustering: The variance of expression of each gene was decomposed to technical and biological components, and highly variable genes identified where biological components were significantly >0.5 . This gave a list of genes for which the difference between average expression in any two cells would be at least 2-log fold. These were used for dimensional reduction using Principle Component Analysis (PCA). T-distributed Stochastic neighbour Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) plots were generated using 1–14 components of the PCA. No batch effects were observed for sample BC572 (sequenced on both runs), indicating that batch corrections were not required. To cluster cells we used the hierarchical iterative clustering from the *scrattch*. *hicat* package (<https://github.com/AllenInstitute/scrattch.hicat>) [30]. This starts with coarse-level clustering and iteratively splits into increasingly fine clusters using the phonograph algorithm, which creates a graph with phenotypic similarities of cells by calculating Jaccard distance between their nearest neighbours [31].

Differential gene/pathway analysis: Marker genes for each cluster were identified as those displaying differential expression on comparing each cluster against all others and reporting the genes that are differentially expressed, using *edgeR* [32]. Pairwise differential expression (DE) analysis was performed between patients or between clusters, with each cell considered as a sample in *edgeR* convention. All comparisons used the DE analysis from *sSeq* package [33]. Cluster 17 (derived from sample BC278) returned a prominent signature of highly expressed erythroid progenitor genes; since low cell numbers had precluded double sorting on this sample we could not exclude contamination from CD38⁺ or CD34⁻ downstream cells, so excluded this cluster from all subsequent DE analyses (CD34 mRNA expression was relatively lower in cells from this cluster). Gene set enrichment analysis (GSEA) was performed using GSEA software (<http://software.broadinstitute.org/gsea>) with default parameters, 1000 permutations on gene sets, and gene sets downloaded from MSigDB or other relevant studies [23,34,35] (Table S3).

Pseudotime analysis: We ordered single cells along their developmental trajectory using the *Monocle* (v2.0) R package (<http://cole-trapnell-lab.github.io/monocle-release/>) and default workflow [36]. Size factors and dispersions were first estimated and genes with a global minimum expression detection threshold of 0.1 were selected

for reordering, using *dpFeature*. We then used tSNE for dimension reduction, and pseudotime trajectories were generated using the *plot_cell_trajectory* function.

SCENIC analysis: We used SCENIC (<https://github.com/aertslab/SCENIC>) to construct gene regulatory networks and identify stable cell states [37]. We first applied a two-stage soft filter to remove genes with low expression (UMI count lower than the value of UMI counts of 3 in 1% of cells) or expressed only in few ($<1\%$) cells. For further downstream analysis we identified genes available on RcisTarget database and ran GENIE3 [38] to identify potential targets for each transcription factor based on co-expression. We then identified potential direct targets based on DNA-motif analysis using RcisTarget. Finally, we identified cell states of individual cells by analysing the network activity in the cell and scoring them using AUCell.

3. Results

To compare the transcriptional landscape of immunophenotypic CMML stem cells with their normal counterparts we performed scRNA-seq on double-sorted Lin⁻CD34⁺CD38⁻ cells from seven untreated CMML patients at presentation (CMML-1, $n=4$; CMML-2, $n=3$; six males, one female; age range, 60–75) and three normal healthy volunteers (two males, one female; age range, 36–52) (Figs. 1A, S1A–B, Table S1). In total we sequenced 5870 patient and 1913 normal stem cells, of which 5031 and 1795 respectively passed quality controls for downstream analysis (Fig. S2A). For each sample we detected a median of 10,520 unique transcripts (range, 2411–14,493) mapping to median 2162 genes per cell (range, 917–2524) (Figs. S2B–C).

3.1. CMML stem cells display a myeloid skewed, proliferative transcriptional signature

Two dimensional visualization by t-SNE based on 526 highly variable genes, (defined as those with biological component of variance significantly greater than zero, i.e. false discovery rate (FDR) 5%) revealed a substantial and clear separation between CMML and control Lin⁻CD34⁺CD38⁻ cells with, in addition, striking transcriptomic heterogeneity between different CMML patients (Fig. 1B). By contrast, spatial separation of the three control samples was much more limited. We focused initially on common differences between CMML and control stem cells by analysing differentially expressed genes. We performed differential gene expression analyses between control and CMML Lin⁻CD34⁺CD38⁻ cells, with each cell considered as a separate sample. We found 943 genes up-regulated and 296 genes down-regulated in CMML by comparison with normal cells (FDR $<5\%$, fold change >1.5 ; Table S2). KEGG and Gene Ontology (GO) analysis of up-regulated genes highlighted enrichment of pathways related to cell cycle and also those characteristic of myeloid lineage cells, including “cytokine-cytokine receptor interaction”, “JAK-STAT signalling” and “chemokine signalling” (Fig. 1C and D). There were no similar enrichments among down-regulated genes. Chemokine genes up-regulated included *CXCL8*, *CXCL12*, *CCL3L3*, and *CXCL10* (Fig. S3A, Table S2). In keeping with the significant level of inter-patient heterogeneity, we did not find any genes consistently up-regulated across every CMML patient sample. By contrast, many genes were consistently down-regulated, including *FOS* and *JUN* proto-oncogenes, both coding for components of the AP-1 transcription factor complex and important regulators of myeloid differentiation (Fig. S3B). Silencing of *FOS* has been reported to promote abnormalities of myeloid differentiation [39], while both are among the most prominently up-regulated genes in (unsorted) BM upon clinical response to HMAs in CMML [10].

We next computed mean normalized expression values for each gene across the population of sorted single Lin⁻CD34⁺CD38⁻ cells, on a sample-by-sample basis for each CMML and control specimen (i.e.

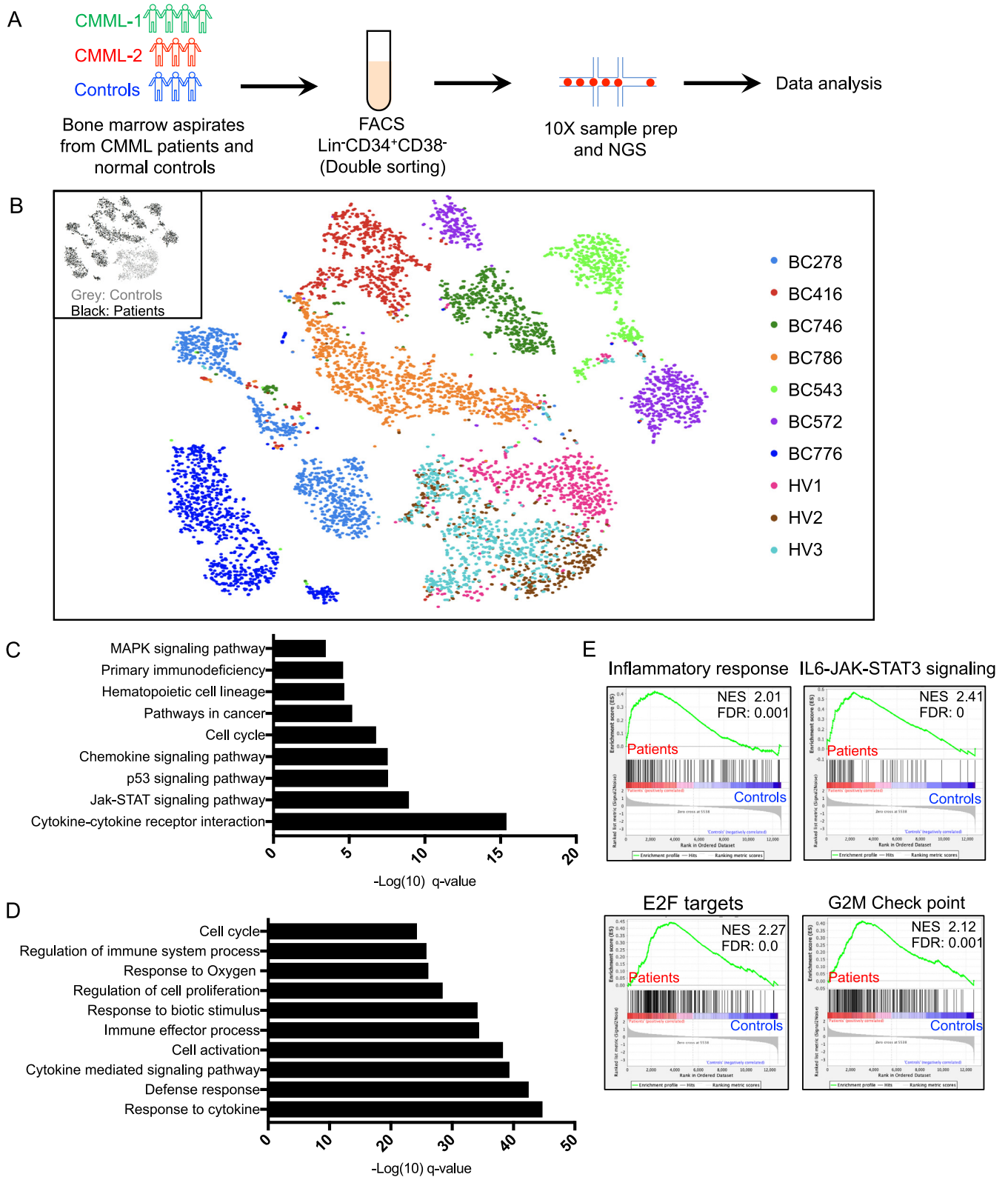


Fig. 1. CMML stem cells display distinct transcriptional signatures: (A) Schematic overview of the experimental workflow used in the study. (B) T-distributed stochastic neighbour embedding (t-SNE) plot based on highly variable genes for all cells passing filtering thresholds, coloured by sample type. In the inset, cells derived from CMML patients are coloured black and from healthy controls are coloured grey. BC: Biobank case number; HV: healthy volunteer; (C-D) KEGG (C) and GO biological processes (D) pathway analysis of up-regulated genes in stem cells from CMML samples as compared with control samples. The top pathways differentially active in CMML stem cells are shown; horizontal bars indicate the $-\log_{10}$ q-value for that term. (E) Gene set enrichment analysis plots showing enrichment of the indicated gene signatures in CMML stem cells as compared with control samples. See also Figs. S1–S3.

seven versus three samples), and performed gene set enrichment analysis (GSEA) using a signal-to-noise ranking metric. We interrogated the ranked list with the Molecular Signatures Database Hallmark Gene Set collection, each of which conveys a specific biological state or process and displays coherent expression [40]. The most significant enrichments were for gene sets characteristic of differentiated myeloid cells such as “Inflammatory response” and “IL6-JAK-STAT signalling” (strongly enriched in CMML versus normal Lin⁻CD34⁺CD38⁻ cells) and cell cycle gene sets “E2F targets” and “G2M checkpoint” (Fig. 1E, Tables S3 and S4). No Hallmark gene sets exhibited enriched expression in control versus CMML Lin⁻CD34⁺CD38⁻ cells. To investigate whether changes in gene expression between CMML and control stem cells are merely reflective of age differences between these groups, we performed GSEA analysis for genes differentially expressed between aged vs. young HSCs [35]. No gene sets associated with normal HSC ageing were differentially enriched in CMML or normal stem cells (Fig. S3C). Further, out of 943 up-regulated genes in CMML stem cells only 8 were associated with ageing (Fig. S3D).

Since stem cell division and differentiation are intimately linked we performed additional GSEA using (i) the gene set up-regulated in human Lin⁻CD34⁺CD38⁻ HSCs by comparison with their downstream Lin⁺CD34⁺CD38⁺ progenitor cells [34] and (ii) gene sets up-regulated in Lin⁻CD34⁺CD38⁺ oligopotent progenitors committed to one particular lineage versus all other cells with the same immunophenotype (including B-cell progenitors, megakaryocyte/erythrocyte (MegE) committed progenitors, neutrophil-primed progenitors, monocyte/dendritic cell progenitors and eosinophil/basophil/mast cell (EBM) progenitors) (Table S3). We observed a highly significant down-regulation of a normal Lin⁻CD34⁺CD38⁻ HSC gene set in CMML versus normal Lin⁻CD34⁺CD38⁻ HSCs (Fig. 2A) and a highly significant up-regulation of genes selectively expressed in oligopotent Lin⁻CD34⁺CD38⁺neutrophil-primed or monocyte/dendritic cell-primed progenitors (Fig. 2B). There was by contrast minimal or absent enrichment of expression of B-cell, MegE or EBM progenitor gene sets (data not shown). These findings are further supported by violin plots shown (Fig. 2C) for single-cell expression of transcription factor genes which have been associated with HSC self-renewal, including *MEIS1*, *HLF*, *KLF4* and *EGR1* (all down-regulated in CMML versus normal stem cells); and transcription factor and other genes associated with myeloid commitment and differentiation such as *RUNX1*, *CEBPD*, *LGALS1* and *AZU* (all up-regulated in CMML versus normal HSCs (Fig. 2C). We also found up-regulation of several genes unique to subsets of patients e.g. *CDKN1A* was up-regulated in a subset of patients' stem cells, as observed previously by targeted q-RT-PCR analysis [11].

Taken together, these data demonstrate that by comparison with normal Lin⁻CD34⁺CD38⁻ HSCs, CMML stem cells exhibit a transcriptome characterized by increased expression of myeloid lineage and cell cycle genes, and lower expression of genes selectively expressed by normal HSCs.

3.2. Detection of normal Lin⁻CD34⁺CD38⁻ single cell transcriptional signature in CMML patients

We next performed unsupervised iterative clustering and identified 17 transcriptionally separable subpopulations across all sequenced cells (Fig. 3A). For each comparison of a subpopulation with all remaining populations, the 20 most up-regulated genes were identified (Table S5). Expression of these marker genes for each of the 17 clusters is shown in Fig. 3B, along with the top three KEGG terms exhibiting significant enrichment in each cluster. As expected for CMML stem cell clusters, and as before, terms include those associated with myeloid lineage differentiation and the cell cycle (Fig. 3B). Normal Lin⁻CD34⁺CD38⁻ HSCs almost exclusively clustered in Cluster 15 (Fig. 3C). Notably, most CMML samples contained a

variable proportion of cells also assigned to Cluster 15 (median 3.1%; range, 0.1–52.3%; Fig. 3C), suggesting the residual presence of cells akin to normal HSCs within each CMML patient.

Leukemic stem cells could influence normal stem cell function via paracrine signalling or through alterations to the microenvironment. We found samples BC786 and BC746 included unusually high (52.4% and 46.5%, respectively) proportions of cells assigned to the “normal” Cluster 15 (Fig. 3C), alongside distinctive patient-specific leukemic clusters. Interestingly, when Cluster 15 CMML stem cells were compared with Cluster 15 stem cells from healthy controls, significant and consistent gene expression differences were evident. We again observed a relative down-regulation of normal HSC genes in Cluster 15 stem cells recovered from CMML patients BC786 and BC746 by comparison with normal individuals; and in the case of Cluster 15 stem cells from patient BC746 there was a significant relative increase in genes selectively expressed in both neutrophil and monocyte/dendritic progenitor cells (Figs. S4A–C). It remains unclear whether these transcriptional differences within Cluster 15 are driven by CMML-associated genetic lesions or alternatively reflect residual genetically normal HSCs in the BM of CMML patients whose transcriptome has been modified by an altered disease-associated BM microenvironment.

3.3. Stem cells from CMML-1 and CMML-2 patients harbour distinct transcriptional programs

An increased BM blast percentage is strongly associated with inferior outcome in CMML [1]. To determine whether this clinicopathological distinction is reflected by transcriptomic differences within the CMML stem cell compartment, we compared single stem cell transcriptomes from patients with WHO-defined CMML-0 or CMML-1 (<10% BM blasts and promonocytes; hereafter “CMML-1”) with those from patients with “CMML-2” (≥10% BM blasts and promonocytes). Again each cell was considered as a separate sample. We observed substantial differences: 782 genes were up-regulated and 802 genes down-regulated in CMML-1 versus CMML-2 stem cell populations (FDR<5%, fold change >1.5; Table S6). KEGG and GO Biological Process analysis showed enriched expression in CMML-1 of gene sets with annotations “DNA replication”, “Cell cycle”, “Translational elongation”, among others. By contrast, gene sets enriched in CMML-2 stem cells by comparison with CMML-1 included those related to inflammatory and cytokine response, JAK-STAT signalling and apoptosis (Table S7). Given this observation of highly distinct categories of KEGG and GO terms enriched in CMML-1 versus CMML-2 stem cells, we made use of the alternate visualization method UMAP, which better represents cellular origins and trajectories during development than the more widely used t-SNE method [41]. This revealed remarkable segregation of CMML-2 stem cell transcriptomes away from CMML-1 and control stem cell transcriptomes, but for a subfraction of stem cells from CMML-2 patient BC572 (Fig. 4A).

To further interrogate the relatedness of normal and CMML stem cell populations we performed pseudotemporal ordering analysis, a method that constructs putative cellular trajectories by statistically ordering cells based on transcriptomic similarity [36,41]. In a two-dimensional linear representation clear demarcation of CMML-1, CMML-2 and normal stem cells was observed (Fig. 4B). This separation was further emphasized by a branched representation where pseudotime ordered stem cells into distinct branches without *a priori* assumption of normal HSC differentiation (Fig. 4C and D). In this projection control and CMML-1 stem cells are largely located above branching point 2, whereas those before branching point 2 are predominantly CMML-2 stem cells (Fig. S5A).

Given the distinct transcriptomes of CMML-1 versus CMML-2 stem cells, we next evaluated levels of transcription factor expression, since these are core regulators of cellular state and differentiation. Consistent with more extensive monocytic differentiation

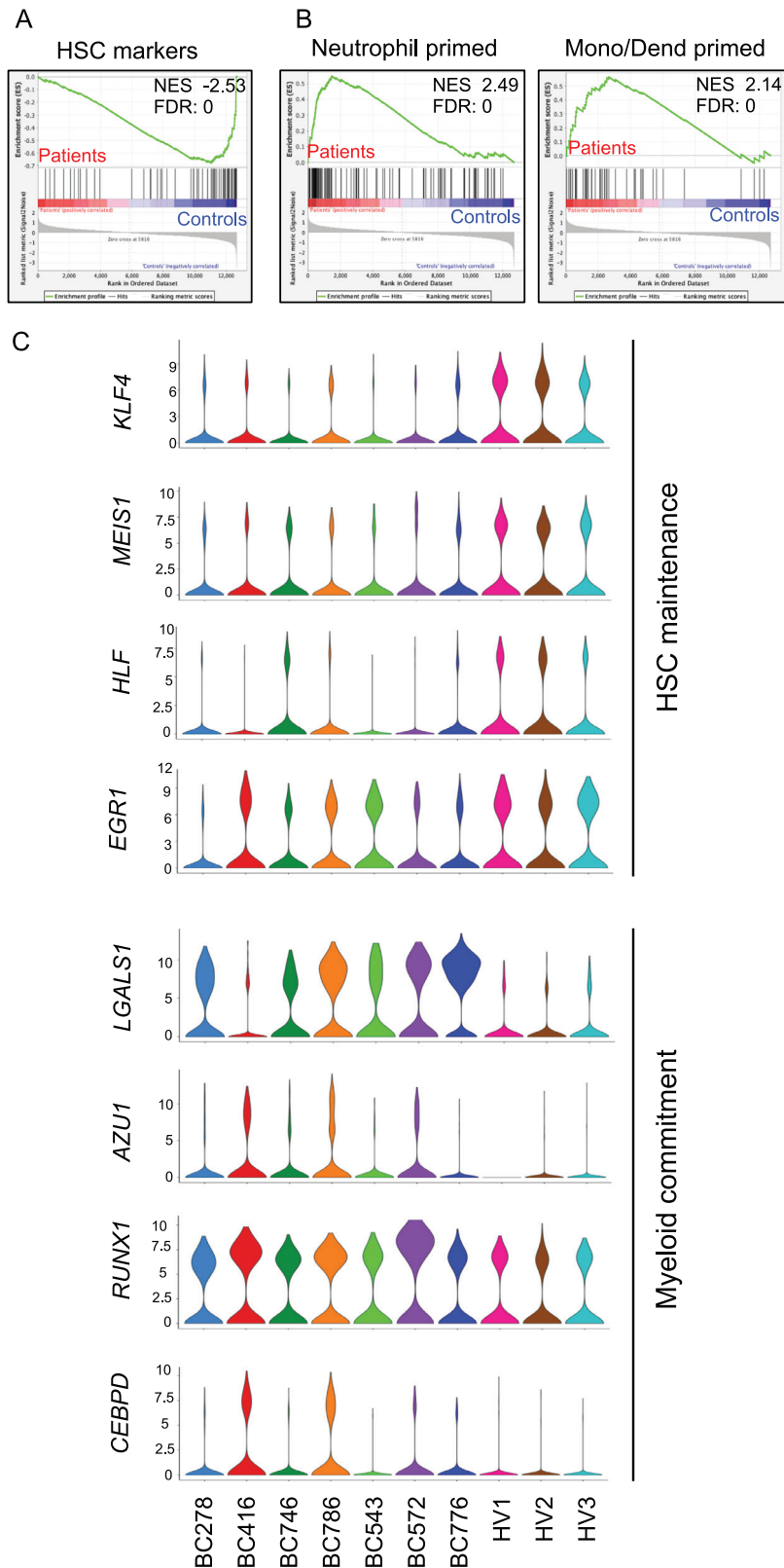


Fig. 2. CMML stem cells lose features of stem cell quiescence and exhibit myeloid bias: (A and B) Gene set enrichment analysis plots showing enrichment of the indicated gene signatures in CMML stem cells as compared with control samples. (C) Violin plots showing the expression of selected genes relating to HSC maintenance and myeloid commitment in CMML and control stem cells. Expression levels are shown as $\log_{10}(\text{normalized counts} + 1)$.

observed in CMML-2 cases, we noted significantly higher expression of transcription factor genes associated with monocytic differentiation such as *IRF1*, *IRF7*, *IRF9*, *MAFF* and *CEBPB* in CMML-2 versus CMML-1 stem cells (Fig. 5A, Fig. S5B). In contrast, the key oncogenic

transcription factor *MYC* and its downstream targets were up-regulated in CMML-1 versus either CMML-2 or normal stem cells, in keeping with the observed enrichment of *MYC*-associated KEGG and GO terms (Fig. 5B). Furthermore, mature granulocytic genes including

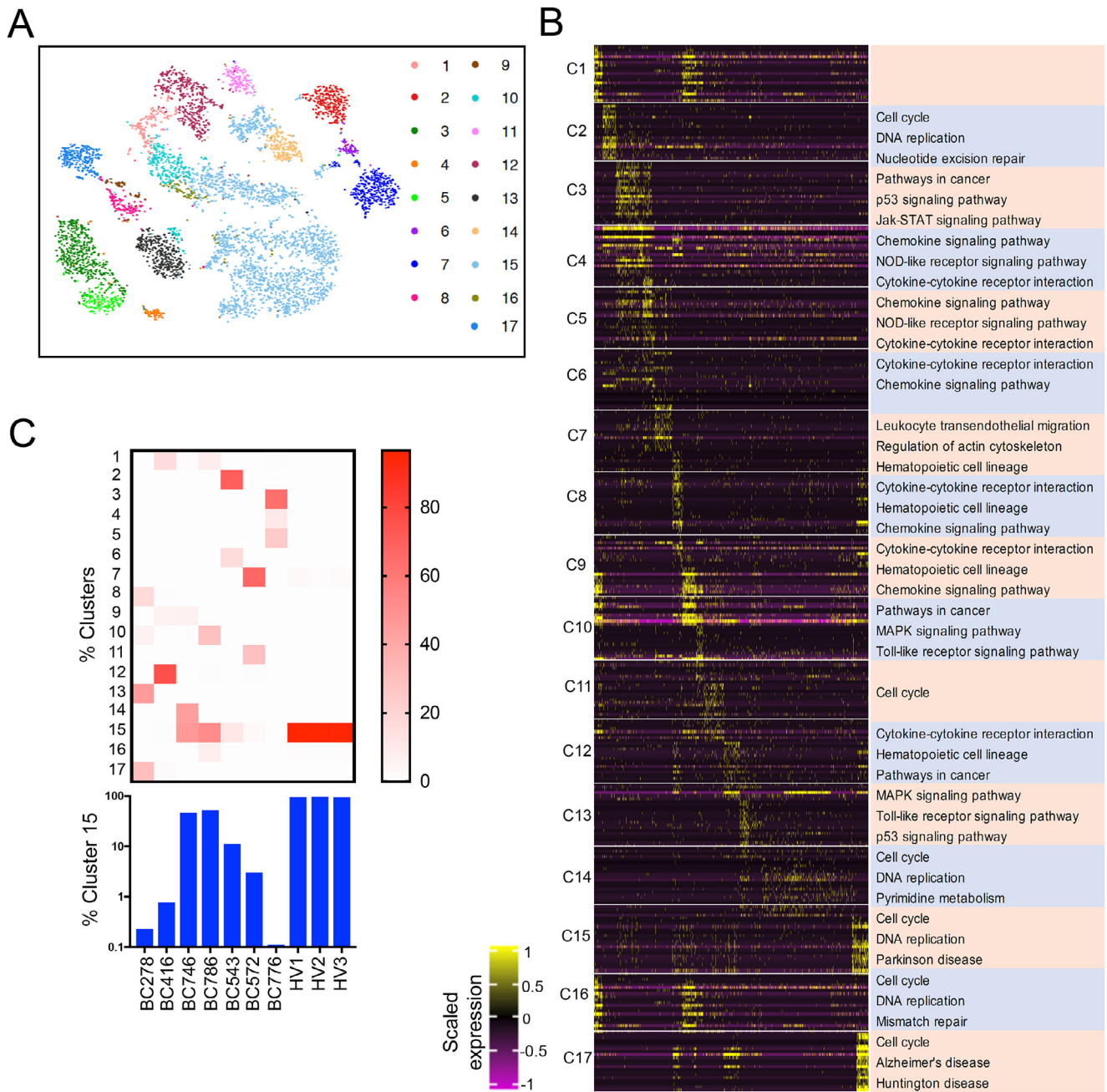


Fig. 3. A variable fraction of CMML stem cells share a transcriptional signature with normal stem cells: (A) t-SNE plot based on highly variable genes for all cells passing filtering thresholds, indicating 17 transcriptionally separable subpopulations (distinguished by colour) as determined by unsupervised iterative clustering. Normal stem cells (from all three healthy controls) segregate together in Cluster 15. (B) Heat map displaying scaled expression of the top 20 differentially expressed genes in each cluster, with respect to cells from all other clusters. Top 3 KEGG pathways (lowest q-value) active in each cluster are shown next to the corresponding cluster (where identified). (C) Heat map indicating the distribution of cells for each sample (columns) across the 17 defined clusters (rows); shading indicates scaled percentage of cells within each cluster for that patient/control sample. Lower panel indicates percentage of cells within Cluster 15 (on a logarithmic scale). See also Fig. S4.

ELANE, *MPO* and *LYZ* were significantly up-regulated in CMML-1 stem cells (Fig. S5C). Transcription factors specifying granulomonocytic commitment and differentiation, such as *SPI1* and *CEBPA*, were also consistently more highly expressed in CMML-1 versus CMML-2 stem cells, but did not reach statistical significance possibly due to small sample size (Fig. 5C). Taken together, our data reveal that stem cell transcriptomes from patients with CMML-2 are quite distinct from those from patients with CMML-1, apparently anticipating disease morphology and thus outcome.

Another source of clinicopathological heterogeneity in CMML is the prognostically-significant distinction between dysplastic and proliferative disease. Differential gene expression analysis found 195

genes up-regulated and 596 down-regulated in dysplastic stem cells with respect to proliferative stem cells (Table S8). Terms including “negative regulation of cell proliferation” and “negative regulation of developmental process” were enriched in dysplastic stem cells (Table S9), but no hallmark gene sets were enriched in stem cells from dysplastic CMML cases as compared with proliferative cases (data not shown). No genes were uniquely expressed in CMML stem cells from cases of proliferative versus dysplastic disease manifestation. Of note, gene expression profiles of whole unsorted BM mononuclear cells and of sorted CD34⁺ cells from CMML patients similarly did not segregate samples based on proliferative versus dysplastic phenotype [15]. Together these observations suggest that whereas

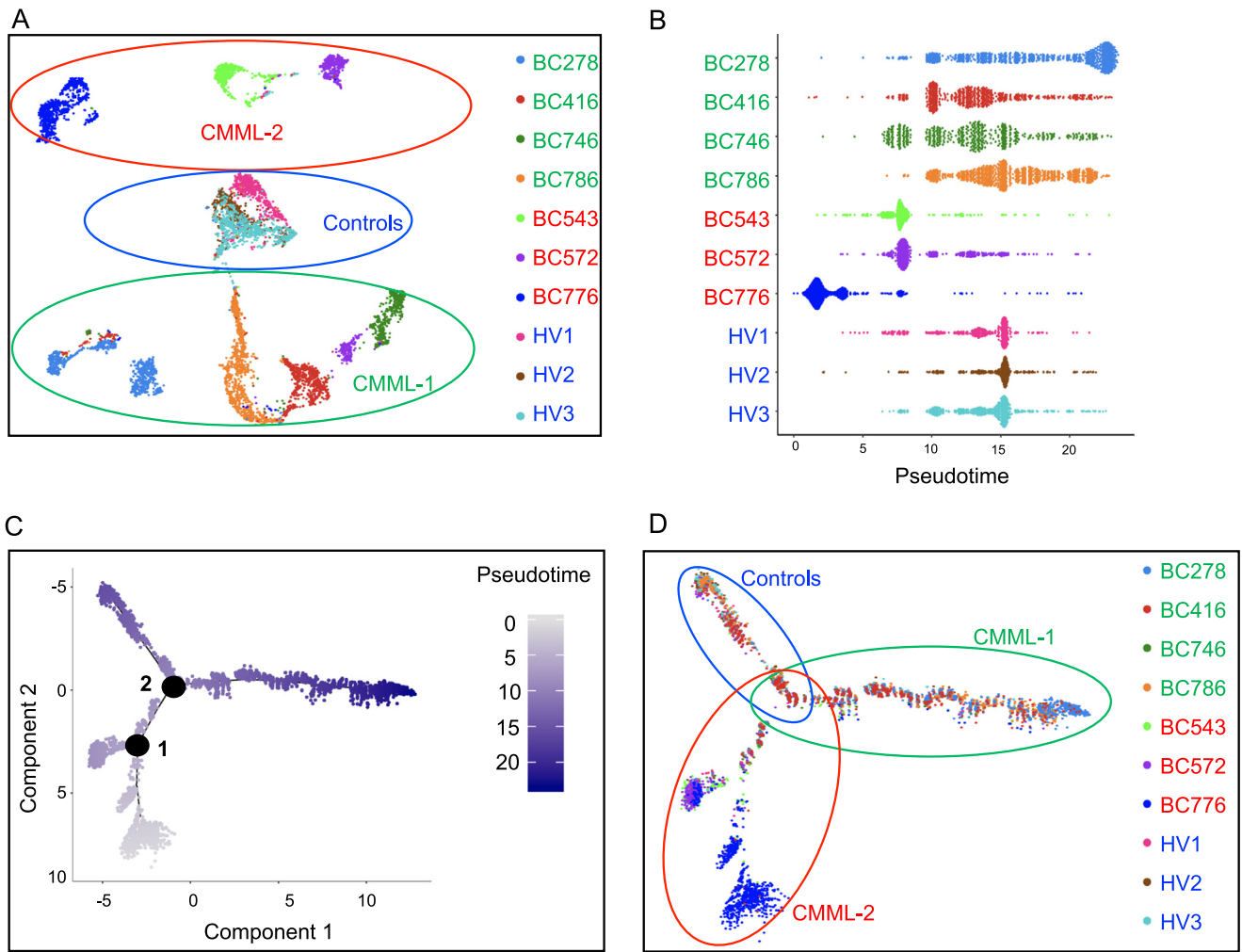


Fig. 4. CMML-1 and CMML-2 stem cells show distinct developmental trajectories: (A) Uniform Manifold Approximation and Projection (UMAP) plot based on highly variable genes for all cells. (B) Cellular trajectories displayed as linear stream plots for each sample, illustrating the spread and density of individual cells in pseudotime; colour coding indicates (C) scaled pseudotime; and (D) the patient/control sample from which each cell derived. Branching points 1 and 2 are indicated in (C) by bold black dots. CMML-1, CMML-2 and HV sample IDs are labelled in green, red and blue respectively. See also Fig. S5.

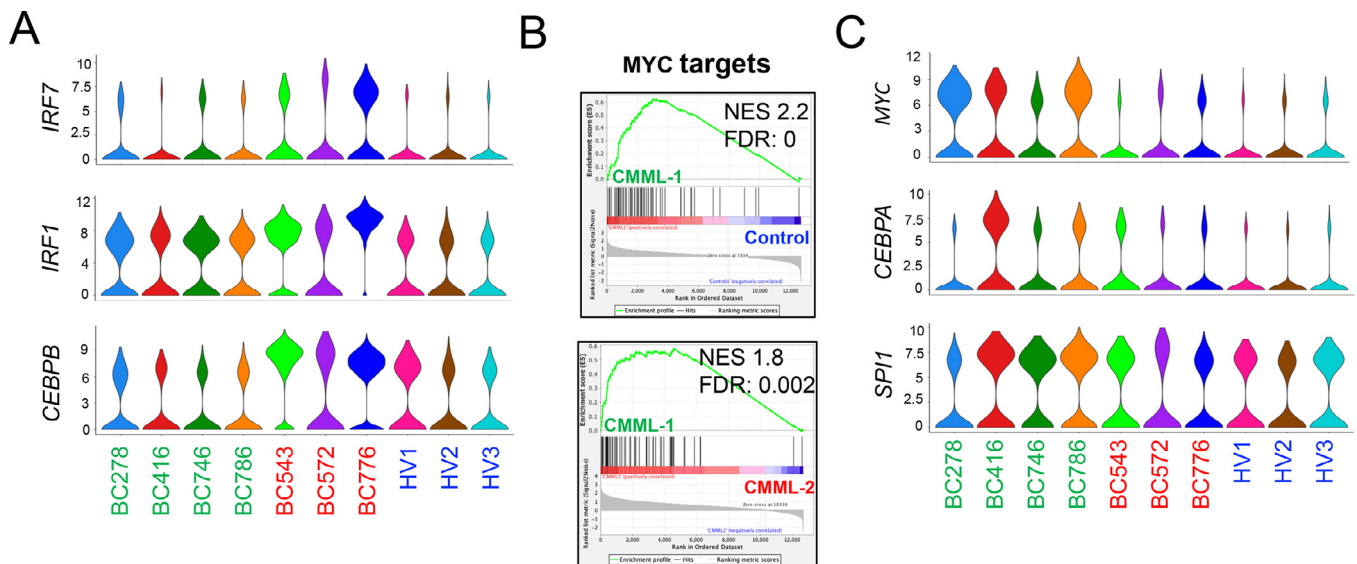


Fig. 5. Stem cells from CMML-1 and CMML-2 patients harbour distinct transcriptional signatures: (A and C) Violin plots showing the relative expression level of selected genes involved in myeloid differentiation. BC: Biobank case number; HV: healthy volunteer; CMML-1, CMML-2 and HV sample IDs are labelled in green, red and blue respectively. (B) Gene set enrichment analysis plots of MYC target genes comparing CMML-1 stem cells as compared with control (upper pane) or CMML-2 (lower pane).

the distinction of CMML-1 from CMML-2 stage disease is primed by upstream transcriptional programs within the primitive stem cell compartment, this does not appear to be the case for dysplastic versus proliferative phenotype; presumably the degree of proliferative potential is instead programmed by genetic or epigenetic events downstream of the stem cell compartment.

3.4. Transcription factor networks involved in CMML stem cells

To more deeply evaluate transcription factor dysregulation in CMML stem cells, we performed a transcriptional regulatory network analysis. A regulon is a gene regulatory network comprising a transcription factor and all of its putative downstream targets. This approach is particularly suited to scRNA-seq, with opportunity to evaluate directly the coordinated expression of a regulating transcription factor and its dependent targets within the same cell. Moreover, considering a regulon as a whole, it is robust against gene dropouts which are common in scRNA-seq datasets. For this we employed SCENIC (Single-Cell rEgulatory Network Inference and Clustering) [37], a computational algorithm that identifies regulon modules by integrating co-expression of transcription factor and genes containing the corresponding transcription factor -binding motif in their cis-regulatory elements.

Clustering based on a total of 296 regulons again showed that normal and CMML $\text{Lin}^- \text{CD34}^+ \text{CD38}^-$ stem cells were widely separable in transcriptional space (Fig. 6A). Whilst some regulons were equally active across all cells (e.g. ELF1, HMGA1, ETS2), certain transcription factor networks displayed unique activity in subsets of samples (Fig. 6B). Consistent with our earlier analyses, regulons associated with the proto-oncogene *JUN* and with transcription factors linked to HSC self-renewal (e.g. *KLF4*, *HLF*) were relatively inactive in CMML (Fig. 6C). CMML-1 stem cells showed a relative increase in activity of regulons associated with known myeloid transcription factors *CEBPD*, *CEBPA* and *MYC*. Networks related to monocytic differentiation were preferentially active in CMML-2 stem cells (Fig. 6C and Fig. S6), consistent with our earlier single gene and GSEA observations. These included regulons centred on *NFKB1*, *IRF8*, *RELA*, *RELB* and *IRF7*, and other transcription factors strongly associated with monocytic lineage differentiation. Thus, the transcriptome of CMML stem cells and their lineage potential are characterized by expression of distinctive transcriptional circuits according to whether they are from patients with CMML-1 (*MYC*) or CMML-2 (*IRF* transcription factors).

3.5. Inter- and intra-patient heterogeneity in CMML HSPCs

Understanding clonal heterogeneity in the stem cell compartment within and between patients is critical for optimal design of novel therapies, which to be effective must target all tumour sub-populations within that patient. Cell-to-cell transcriptional correlation analysis found no consistent differences in the magnitude of overall transcriptomic heterogeneity between samples, or between CMML and controls (Fig. S7A). This suggests that overall CMML stem cells are no more heterogeneous than normal stem cells. However, unlike control stem cells, each patient's stem cells were distributed across 2–3 distinct clusters (Fig. 7A and S7B).

Next we performed pairwise pathway analyses comparing transcriptional profiles of each CMML-patient specific cluster with that of Cluster 15, excluding cells from CMML patients that mapped to this cluster (Fig. 7B). There was highly heterogeneous differential expression of genes mapping to distinct KEGG pathways. Some (e.g. “cytokine-cytokine receptor interaction”, “chemokine signalling”) were generally active across most disease clusters, whilst others were only active in certain clusters in some patients. Moreover, we observed striking heterogeneity within certain CMML patients. For example, patient BC776 displayed three major clusters: 3, 4 and 5. While Clusters 3 and 5 shared many features, “Cell cycle” and “DNA replication”

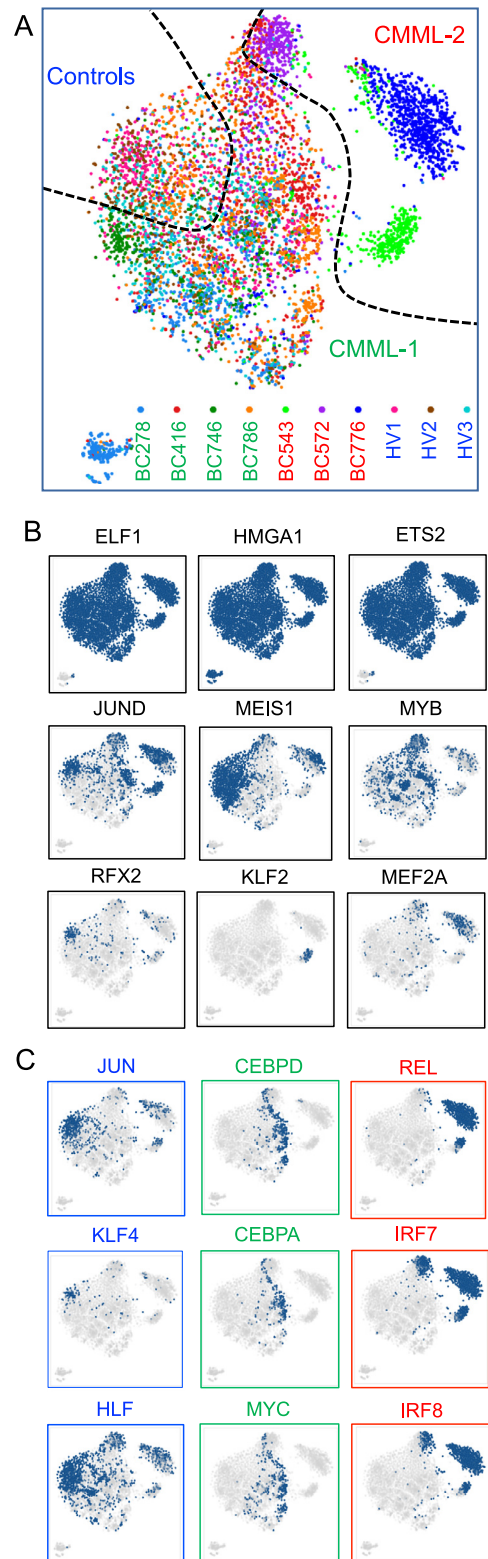


Fig. 6. Distinct transcription factor networks operational in CMML-1 and CMML-2 stem cells:- (A) t-SNE plot based on regulon activity in all CMML and control stem cells included in the study, as determined by SCENIC analysis. Segregation of cells derived from CMML-1, CMML-2 and normal controls is indicated by the dashed lines. BC: Bio-bank case number; HV: healthy volunteer; CMML-1, CMML-2 and HV sample IDs are labelled in green, red and blue respectively. (B & C) t-SNE plots based on activity of regulons centred on the indicated transcription factors, displayed as a binary measure (grey: inactive; blue: active). Regulons that are uniquely active in HV, CMML-1 and CMML-2 cells are highlighted in blue, green and red boxes, respectively (C). Regulons that are either active in all cells or in cells that do not fall under any specific category are highlighted black boxes (B). See also Fig. S6.

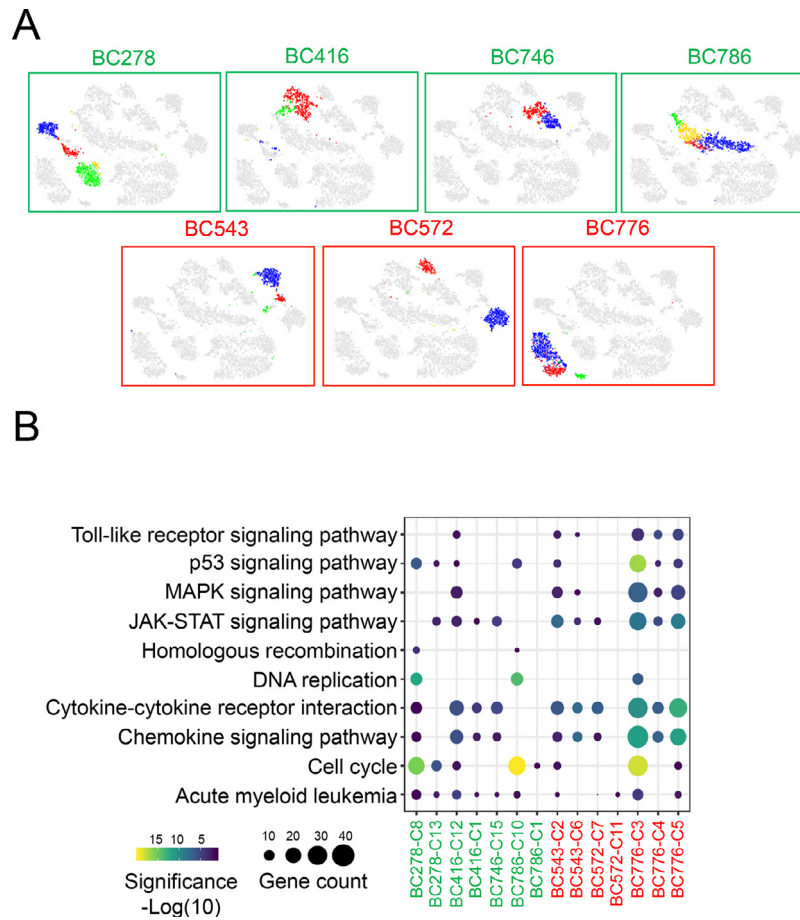


Fig. 7. Inter and intra patient heterogeneity in CMML stem cells: (A) t-SNE plots based on highly variable genes for each CMML patient separately, highlighting the distribution of their stem cells across the defined clusters; distinct clusters represented within a sample are assigned different colours for emphasis. t-SNE plots presenting CMML-1 and CMML-2 stem cells are highlighted in green and red boxes respectively (B) Bubble chart showing relative enrichment of the indicated pathways in each distinct cluster from the indicated CMML patient samples, all with respect to Cluster 15 from all merged control samples. Bubble colour represents scaled $-\log_{10}$ FDR significance; bubble size indicates the number of up-regulated genes enriched in the indicated pathways. See also Fig. S7.

gene sets were more active in Cluster 3 versus Cluster 5. Similarly, Cluster 4 showed minimal activation of JAK-STAT and MAPK signaling pathways in comparison with Clusters 3 and 5. These data further highlight the transcriptional heterogeneity not just between, but within CMML patient stem cell populations.

4. Discussion

Novel inhibitors targeting pathogenic gain-of-function mutations (e.g. IDH1, IDH2, FLT3) are changing the treatment landscape in myeloid malignancies, but have a limited role in CMML in which these mutations are comparatively rare [5]. The typical mutations driving CMML (TET2, SRSF2, ASXL1, RUNX1, N/KRAS) have proved difficult to target directly, and no tractable vulnerabilities specific to the CMML disease-initiating compartment have been identified to rationally guide targeted therapy design. Our data reveal that the defining myelomonocytic expansion of CMML is primed from the Lin⁻CD34⁺CD38⁻ stem cell compartment, with fundamental myelomonocytic transcriptomic perturbation and dysregulation of key differentiation pathways. These results are consistent with the previously observed up-regulation of the myelomonocytic transcription factor *SPI1* in sorted common myeloid progenitor cells from CMML patients in comparison with the healthy controls [13]. We highlight candidate genes and pathways consistently and exclusively dysregulated in CMML stem cells, apparently across genotypes: for example, down-regulation of the AP-1 subunits *FOS* and *JUN*, implicating NF- κ B pathway activation as in other cancers [42] and overexpression of *LGALS1*,

a positive regulator of both RAS signalling and the anti-apoptotic proto-oncogenes BCL-2 and MCL-1 [43]. More broadly, we identify pathways aberrantly operational in stem cells from subsets of patients, including JAK-STAT signalling, inflammatory response, translation elongation and cell cycle progression, each of which are increasingly amenable to therapeutic modulation with novel agents. Aberrant expression of genes involved in inflammation and cell cycle was also observed in whole unsorted BM mononuclear cells and in monocytes from CMML patients, indicating that these altered regulatory networks are programmed within primitive stem cells and propagated throughout differentiation [11,13]. Activation of JAK-STAT pathway, without necessarily a demonstrable JAK2 mutation and particularly as a hypersensitive response to native GM-CSF, has been reported as a key feature of CMML, and drugs targeting this pathway have shown promise in both preclinical models and early phase clinical trials [44–48].

Our data reveal a substantial degree of both inter- and intra-patient heterogeneity within the Lin⁻CD34⁺CD38⁻ compartment of CMML. Strikingly, every patient displayed a private set of differentially expressed genes and aberrantly operational transcription factor networks. This extent of transcriptomic heterogeneity between patients mirrors the observed clinical heterogeneity, which far exceeds the mutational diversity of this disease and indicates that effective targeted therapy approaches in CMML may need to be highly personalized [7].

Strikingly, we observed wide segregation and distinct transcriptomic profiles between CMML-1 and CMML-2 stem cells; disease stages clinically distinguished solely by the percentage of blasts and

promonocytes. CMML-1 and CMML-2 stem cells were segregated widely by UMAP. CMML-1 stem cells exhibited a program defined by high expression of *MYC*, and myeloid differentiation and cell cycle genes. CMML-2 stem cells were characterized by high expression of regulomes associated with *IRF1*, *IRF7* and *IRF8*, factors that are highly expressed in monocytic lineage differentiation. These data highlight the distinctive priming of CMML-1 versus CMML-2 malignant stem cells and suggest against a simplistic, linear succession model of CMML disease progression. This implies that CMML-1 and CMML-2 might represent quite different disease states rather than necessarily steps on a continuum, although our sample size was too small for definitive confirmation. Notably, many CMML-1 patients do not transform to AML despite sometimes exhibiting highly proliferative, aggressive disease. CMML-2 patients, by contrast, frequently presents with a disease closely resembling, or rapidly evolving into AML.

We identified variably sized populations of CMML stem cells transcriptionally resembling normal HSCs. It remains unclear whether these are part of the malignant clone, partially-transformed pre-leukemic stem cells or genuinely normal residual stem cells. That they might be uninvolved bystanders is supported by a previous report that founding mutations were detectable in most, but not all, HSC/MPP-derived colonies from a majority of CMML patients [13]. Nevertheless, the relationship between genotype and functional LIC status is uncertain, particularly since CMML founding mutations are also common markers of age-related clonal haematopoiesis of indeterminate potential [49]. Indeed, our scRNA-seq, at much higher throughput, is also consistent with a hypothesis that the BM of CMML patients contains residual genetically uninvolved stem cells, with transcriptomes subtly altered due to a non-cell autonomous effect. These could represent transcriptional perturbations induced by paracrine signalling from neighbouring malignant stem cells or from the leukemic microenvironment, as has been observed in chronic myeloid leukaemia [50]. Irrespective, this raises the prospect of a therapeutic window, with a potential reservoir of cells to potentially survive a targeted therapy exploiting more pervasive transcriptional vulnerabilities and restore functional haematopoiesis.

Whether the transcriptomic heterogeneity of CMML stem cells is mediated by genetic, epigenetic or microenvironmental influences remains uncertain. We found no convincing correlation of transcriptional clusters with patient mutation status as genotyped at bulk MNC level (data not shown), but these analyses were indirect and limited by small sample size. Our limited sample size also precluded us from a detailed investigation into the transcriptional consequences of TET2/SRSF2 mutant combination in patient HSPCs. Studies leveraging emerging technologies to correlate transcriptome with genotype at the single-cell level, and in larger patient cohorts, are required to comprehensively address the link between transcriptional and mutational heterogeneity. Taken together, though, our data demonstrate CMML to be a heterogeneous stem cell malignancy, with multiple transcriptional routes converging on the common downstream clinicopathological phenotype.

Author contributions

KB, TS and DW designed the study and performed the experiments. KB, SM, TS and DW analysed the data and wrote the manuscript, with contribution from all other authors.

Declaration of Competing Interest

Authors declare no competing interests.

Acknowledgements

We thank Jeff Barry, Toni Banyard, Wolfgang Breitwieser and Gillian Newton at Cancer Research UK Manchester Institute core

facilities for technical support. We also thank Deepti Wilks and Michelle Greenhalgh at MCRC Biobank for assisting with patient sample procurement and processing.

This work was supported by the Oglesby Charitable Trust. DW is additionally supported by a Bloodwise Clinician Scientist Fellowship [grant number 15030]. SM is supported by a UK Medical Research Council single-cell centre award [MR/M008908/1]. JS is supported by Cancer Research UK grant numbers C5759/A27445 and C147/A25254. TS is supported by Cancer Research UK grant number C5759/A20971. Targeted mutation sequencing was supported by a grant from The Christie Charity. Funders had no role in study design, data collection, data analysis, interpretation or writing of the report.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2020.102904.

References

- Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 2016;127(20):2391–405.
- Itzykson R, Solary E. An evolutionary perspective on chronic myelomonocytic leukemia. *Leukemia* 2013;27(7):1441–50.
- McCullough KB, Patnaik MM. Chronic Myelomonocytic Leukemia: a Genetic and Clinical Update. *Curr Hematol Malig Rep* 2015;10(3):292–302.
- Patnaik MM, Tefferi A. Chronic myelomonocytic leukemia: 2018 update on diagnosis, risk stratification and management. *Am J Hematol*. 2018;93(6):824–40.
- Patnaik MM, Itzykson R, Lasho TL, Kosmider O, Finke CM, Hanson CA, et al. ASXL1 and SETBP1 mutations and their prognostic contribution in chronic myelomonocytic leukemia: a two-center study of 466 patients. *Leukemia* 2014;28(11):2206–12.
- Mason CC, Khorashad JS, Tantravahi SK, Kelley TW, Zabriskie MS, Yan D, et al. Age-related mutations and chronic myelomonocytic leukemia. *Leukemia* 2016;30(4):906–13.
- Ball M, List AF, Padron E. When clinical heterogeneity exceeds genetic heterogeneity: thinking outside the genomic box in chronic myelomonocytic leukemia. *Blood* 2016;128(20):2381–7.
- Solary E, Itzykson R. How I treat chronic myelomonocytic leukemia. *Blood* 2017;130(2):126–36.
- Eissa H, Gooley TA, Sorror ML, Nguyen F, Scott BL, Doney K, et al. Allogeneic hematopoietic cell transplantation for chronic myelomonocytic leukemia: relapse-free survival is determined by karyotype and comorbidities. *Biol Blood Marrow Transpl* 2011;17(6):908–15.
- Merlevede J, Droin N, Qin T, Meldi K, Yoshida K, Morabito M, et al. Mutation allele burden remains unchanged in chronic myelomonocytic leukaemia responding to hypomethylating agents. *Nat Commun* 2016;7:10767.
- Hurtado AM, Luengo-Gil G, Chen-Liang TH, Amaral F, Batta K, Palomo L, et al. Transcriptomic rationale for synthetic lethality-targeting ERCC1 and CDKN1A in chronic myelomonocytic leukaemia. *Br J Haematol* 2018.
- Franzini A, Pomietti AD, Yan D, Khorashad JS, Tantravahi SK, Than H, et al. The transcriptome of CMML monocytes is highly inflammatory and reflects leukemia-specific and age-related alterations. *Blood Adv* 2019;3(20):2949–61.
- Itzykson R, Kosmider O, Renneville A, Morabito M, Preudhomme C, Berthon C, et al. Clonal architecture of chronic myelomonocytic leukemias. *Blood* 2013;121(12):2186–98.
- Gelsi-Boyer V, Cervera N, Bertucci F, Trouplin V, Remy V, Olschwang S, et al. Gene expression profiling separates chronic myelomonocytic leukemia in two molecular subtypes. *Leukemia* 2007;21(11):2359–62.
- Gelsi-Boyer V, Cervera N, Bertucci F, Brecqueville M, Finetti P, Murati A, et al. Molecular similarity between myelodysplastic form of chronic myelomonocytic leukemia and refractory anemia with ring sideroblasts. *Haematologica* 2013;98(4):576–83.
- Fritsch G, Stimpfl M, Kurz M, Printz D, Buchinger P, Fischmeister G, et al. The composition of CD34 subpopulations differs between bone marrow, blood and cord blood. *Bone Marrow Transpl* 1996;17(2):169–78.
- Wiseman DH, Greystoke BF, Somerville TC. The variety of leukemic stem cells in myeloid malignancy. *Oncogene* 2014;33(24):3091–8.
- Pang WW, Pluvinau JV, Price EA, Sridhar K, Arber DA, Greenberg PL, et al. Hematopoietic stem cell and progenitor cell mechanisms in myelodysplastic syndromes. *Proc Natl Acad Sci USA* 2013;110(8):3011–6.
- Petzer AL, Eaves CJ, Lansdorp PM, Ponchio L, Barnett MJ, Eaves AC. Characterization of primitive subpopulations of normal and leukemic cells present in the blood of patients with newly diagnosed as well as established chronic myeloid leukemia. *Blood* 1996;88(6):2162–71.
- Woll PS, Kjallquist U, Chowdhury O, Doolittle H, Wedge DC, Thongjuea S, et al. Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. *Cancer Cell* 2014;25(6):794–808.

- [21] Gregor Eisenwort MP, Keller Alexandra, Willmann Michael, Sadovnik Irina, Greiner Georg MP, Karoline V, Gleixner M, et al. Phenotypic Characterization of Leukemia-Initiating Stem Cells in Chronic Myelomonocytic Leukemia (CMML). *Blood* 2019;134:4223.
- [22] van Galen P, Hovestadt V, Wadsworth Li MH, Hughes TK, Griffin GK, Battaglia S, et al. Single-Cell RNA-Seq reveals AML hierarchies relevant to disease progression and immunity. *Cell* 2019;176(6):1265–81 e24.
- [23] Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* 2017;19(4):271–81.
- [24] Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 2017;8:14049.
- [25] Greystoke BF, Huang X, Wilks DP, Atkinson S, Somerville TC. Very high frequencies of leukaemia-initiating cells in precursor T-acute lymphoblastic leukaemia may be obscured by cryopreservation. *Br J Haematol* 2013;163(4):538–41.
- [26] Bourgon R, Gentleman R, Huber W. Independent filtering increases detection power for high-throughput experiments. *Proc Natl Acad Sci USA* 2010;107(21):9546–51.
- [27] Lun AT, Bach K, Marioni JC. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol* 2016;17:75.
- [28] McCarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 2017;33(8):1179–86.
- [29] Tirosh O, Cohen Y, Shitrit A, Shani O, Le-Trilling VT, Trilling M, et al. The transcription and translation landscapes during human cytomegalovirus infection reveal novel host-pathogen interactions. *PLoS Pathog* 2015;11(11):e1005288.
- [30] Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* 2016;19(2):335–46.
- [31] Levine JH, Simonds EF, Bendall SC, Davis KL, Amir el AD, Tadmor MD, et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* 2015;162(1):184–97.
- [32] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26(1):139–40.
- [33] Yu D, Huber W, Vitek O. Shrinkage estimation of dispersion in Negative Binomial models for RNA-seq experiments with small sample size. *Bioinformatics* 2013;29(10):1275–82.
- [34] Georgantas 3rd RW, Tanadve V, Malehorn M, Heimfeld S, Chen C, Carr L, et al. Microarray and serial analysis of gene expression analyses identify known and novel transcripts overexpressed in hematopoietic stem cells. *Cancer Res* 2004;64(13):4434–41.
- [35] Adelman ER, Huang HT, Roisman A, Olsson A, Colaprico A, Qin T, et al. Aging Human Hematopoietic Stem Cells Manifest Profound Epigenetic Reprogramming of Enhancers That May Predispose to Leukemia. *Cancer Discov* 2019;9(8):1080–101.
- [36] Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;32(4):381–6.
- [37] Aibar S, Gonzalez-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 2017;14(11):1083–6.
- [38] Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 2010;5(9).
- [39] Shafarenko M, Amanullah A, Gregory B, Liebermann DA, Hoffman B. Fos modulates myeloid cell survival and differentiation and partially abrogates the c-Myc block in terminal myeloid differentiation. *Blood* 2004;103(11):4259–67.
- [40] Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27(12):1739–40.
- [41] Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 2018.
- [42] Ji Z, He L, Regev A, Struhl K. Inflammatory regulatory network mediated by the joint action of NF- κ B, STAT3, and AP-1 factors is involved in many human cancers. *Proc Natl Acad Sci USA* 2019;116(19):9453–62.
- [43] Ruvoilo PP. Galectins as regulators of cell survival in the leukemia niche. *Adv Biol Regul* 2019;71:41–54.
- [44] Geissler K, Ohler L, Fodinger M, Virgolini I, Leimer M, Kabrna E, et al. Interleukin 10 inhibits growth and granulocyte/macrophage colony-stimulating factor production in chronic myelomonocytic leukemia cells. *J Exp Med* 1996;184(4):1377–84.
- [45] Padron E, Painter JS, Kunigal S, Mailloux AW, McGraw K, McDaniel JM, et al. GM-CSF-dependent pSTAT5 sensitivity is a feature with therapeutic potential in chronic myelomonocytic leukemia. *Blood* 2013;121(25):5068–77.
- [46] Patnaik MM, Sallman DA, Mangaonkar A, Heuer R, Hirvela J, Zblewski D, et al. Phase 1 study of lenzilumab, a recombinant anti-human GM-CSF antibody, for chronic myelomonocytic leukemia (CMML). *Blood* 2020.
- [47] Padron E, Dezern A, Andrade-Campos M, Vaddi K, Scherle P, Zhang Q, et al. A multi-institution phase I trial of ruxolitinib in patients with chronic myelomonocytic leukemia (CMML). *Clin Cancer Res* 2016;22(15):3746–54.
- [48] Assi R, Kantarjian HM, Garcia-Manero G, Cortes JE, Pemmaraju N, Wang X, et al. A phase II trial of ruxolitinib in combination with azacytidine in myelodysplastic syndrome/myeloproliferative neoplasms. *Am J Hematol* 2018;93(2):277–85.
- [49] Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* 2015;126(1):9–16.
- [50] Giustacchini A, Thongjuea S, Barkas N, Woll PS, Povinelli BJ, Booth CAG, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat Med* 2017;23(6):692–702.