

# UC Office of the President

## Research Grants Program Office (RGPO) Funded Publications

### Title

Single-cell epigenomic variability reveals functional cancer heterogeneity

### Permalink

<https://escholarship.org/uc/item/8q62h5n2>

### Journal

Genome Biology, 18(1)

### ISSN

1474-760X

### Authors

Litzenburger, Ulrike M

Buenrostro, Jason D

Wu, Beijing

et al.

### Publication Date

2017-12-01

### DOI

10.1186/s13059-016-1133-7

Peer reviewed

RESEARCH

Open Access



# Single-cell epigenomic variability reveals functional cancer heterogeneity

Ulrike M. Litzenburger<sup>1</sup>, Jason D. Buenrostro<sup>4,5</sup>, Beijing Wu<sup>2</sup>, Ying Shen<sup>1</sup>, Nathan C. Sheffield<sup>1</sup>, Arwa Kathiria<sup>1,2</sup>, William J. Greenleaf<sup>1,2,3\*</sup> and Howard Y. Chang<sup>1\*</sup>

## Abstract

**Background:** Cell-to-cell heterogeneity is a major driver of cancer evolution, progression, and emergence of drug resistance. Epigenomic variation at the single-cell level can rapidly create cancer heterogeneity but is difficult to detect and assess functionally.

**Results:** We develop a strategy to bridge the gap between measurement and function in single-cell epigenomics. Using single-cell chromatin accessibility and RNA-seq data in K562 leukemic cells, we identify the cell surface marker CD24 as co-varying with chromatin accessibility changes linked to GATA transcription factors in single cells. Fluorescence-activated cell sorting of CD24 high versus low cells prospectively isolated GATA1 and GATA2 high versus low cells. GATA high versus low cells express differential gene regulatory networks, differential sensitivity to the drug imatinib mesylate, and differential self-renewal capacity. Lineage tracing experiments show that GATA/CD24hi cells have the capability to rapidly reconstitute the heterogeneity within the entire starting population, suggesting that GATA expression levels drive a phenotypically relevant source of epigenomic plasticity.

**Conclusion:** Single-cell chromatin accessibility can guide prospective characterization of cancer heterogeneity. Epigenomic subpopulations in cancer impact drug sensitivity and the clonal dynamics of cancer evolution.

**Keywords:** Open chromatin, Gene expression noise, Cancer stem cells

## Background

Epigenetic aberrations are a key driver of cancer pathogenesis. Altered chromatin states can activate oncogenes and silence tumor suppressor genes, leading to uncontrolled growth and metastasis. In contrast to genetic mutations, epigenetic changes are dynamic and potentially reversible, leading to heterogeneity during development, within tumors, or in response to environmental stimuli, drugs, or diseases [1–4]. Epigenomic variability can arise as cell-to-cell differences in the patterning of DNA methylation, histone modifications, or expression of protein coding genes or noncoding RNAs. This epigenomic variation at the single-cell level can create heterogeneity in cancer. However, the functional relevance of this variation is difficult to assess, often due to a lack of methods capable of quantifying it.

Methods for profiling the epigenomic landscape include bisulfite sequencing for analyzing DNA methylation, DNase-seq and MNase-seq [5–7] for accessibility or nucleosome positioning information, and chromatin immunoprecipitation followed by sequencing (ChIP-seq) for binding sites of individual factors or modified nucleosomes [8, 9]. These methods have proven invaluable for identifying the epigenomic features dictating cell states within large cellular populations but are generally unable to detect single-cell epigenomic cell-to-cell variability. Methods for measuring single-cell gene expression have begun to provide genome-wide measures of cell-to-cell differences; however, these methods provide only an indirect readout of genome-wide epigenomic variance [10, 11]. Recently, single-cell methods for measuring DNA methylation [12, 13], histone modifications [14], and chromatin accessibility have been developed to directly quantify epigenomic variation within cellular populations [15–17]; nevertheless, the functional relevance of this observed epigenomic variability remains to be elucidated.

\* Correspondence: wjg@stanford.edu; howchang@stanford.edu

<sup>1</sup>Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA 94305, USA

Full list of author information is available at the end of the article

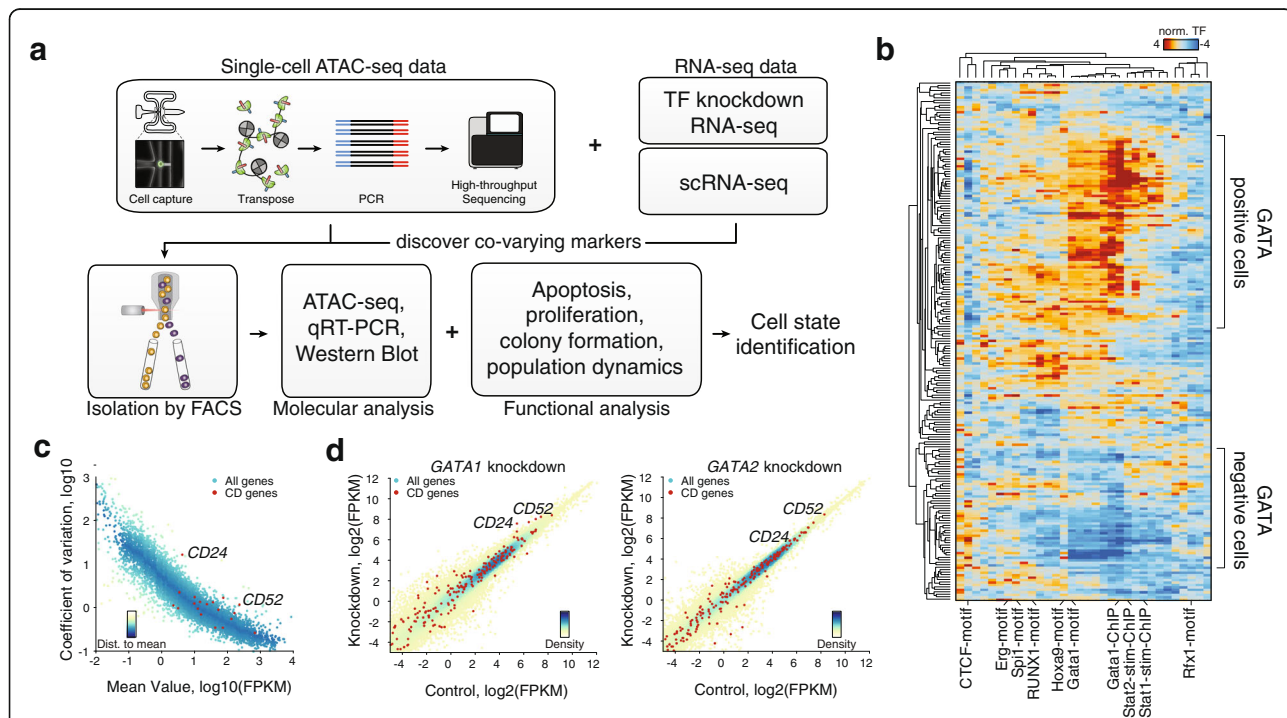
ATAC-seq measures regions of open chromatin using the Tn5-transposase, which preferentially inserts sequencing adapters into accessible chromatin [16]. As applied to single cells [18, 19], this method quantifies cell-to-cell variation in regions of chromatin accessibility. Single cell (sc)ATAC-seq has been used to identify specific transcription factors associated with cell-to-cell regulatory variability, such as GATA1 and GATA2 in K562 cells [19]. While this signal of increased regulatory variation provides a rich platform for hypotheses regarding a potential functional role of GATA factor variation, further experiments are required to identify the phenotypic consequences of this epigenomic variability. Data generated from single-cell techniques like scRNA-seq, scDNA-seq, and scATAC-seq are purely descriptive and require downstream functional validation to link observed heterogeneity to functional subpopulations, such as those with metastatic capabilities or stem cell-like properties that might inform possible treatment strategies. Because most techniques for genomic analysis destroy the cell, it is difficult to combine single-cell approaches with functional cellular assays unless single cells can be identified and

sorted using cell surface markers. However, cell surface markers for partitioning cellular populations based on epigenomic state are often unknown. Here we combine scATAC-seq and RNA-seq to identify a potential co-varying surrogate for cell surface markers (Fig. 1a) that enable prospective isolation of relevant subpopulations, allowing downstream functional dissection of the importance of these single-cell observations.

**Results and Discussion**

**Selection of cell surface marker co-varying with highly variable motifs identified by scATAC-seq**

In previous work, scATAC-seq measurements of K562 chronic myeloid leukemia (CML) cells identified high cell-to-cell variability in the accessibility of the GATA motif (Fig. 1b) [20]. As expected from proliferating cells, we find increased variability within different replication timing domains, representing variable ATAC-seq signal associated with changes in DNA content across the cell cycle. Importantly, the variability in GATA motif accessibility is not influenced by the cell cycle variation [19]. Interestingly, in addition to epigenomic variability



**Fig. 1** Strategy for identifying a cell surface marker co-varying with identified varying transcription factors. **a** Cartoon illustrating the strategy: single-cell ATAC-seq is followed by sequencing and analysis of cell-to-cell variation, focusing on transcription factor (TF) motifs. RNA-seq and single-cell RNA-seq data are used to correlate cell surface expression with expression of the transcription factor with highest identified variability. The expression of the cell surface protein is subsequently used to isolate subpopulations, which can then be analyzed for molecular and functional characteristics. **b** Hierarchical clustering of cells (rows) and high-variance transcription factors (columns). Scores represent relative accessibility and are reproduced from Buenrostro et al. [19]. **c** Single-cell RNA-seq data of K562 cells. Coefficient of variation is plotted against the mean FPKM, data points are colored by distance to running mean. Red dots indicate CD expression markers. **d** Re-analysis of RNA-seq data of GATA1 and GATA2 knockdown in K562 cells. Control FPKM is plotted against knockdown FPKM; data points are colored by density. Red dots indicate CD expression markers. FACS fluorescence-activated cell sorting, qRT-PCR quantitative reverse transcription PCR

associated with GATA binding, we also find high epigenomic variability within transcription factors that are expressed in hematopoietic progenitors, like ERG, HOXA9, SPI1 (PU.1), and RUNX1 [21–24]. We also observe variability associated with STAT1 and STAT2 binding, further reflecting hematopoietic differentiation, as the JAK-STAT pathway is an important regulator enabling cells to respond to interferons and cytokines. In particular, K562 cells contain a BCR-ABL fusion resulting in constitutive STAT activity and ultimately defective erythropoiesis. Furthermore, STAT transcription factors can promote oncogenesis by inducing anti-apoptotic gene expression [25, 26]. These observations suggest that multiple transcription factors involved in regulating the progenitor state significantly vary among K562 cells, pointing to a possible difference in the phenotype of these subpopulations.

Here, we focus on variation in GATA motif accessibility because GATA1 and GATA2 play pivotal roles during erythropoiesis and leukemogenesis [27–30]. Notably, GATA factors have a highly similar binding consensus sequence, WGATAA. Recent genome-wide ChIP-seq analysis using K562 human leukemia cells revealed that 35% of GATA1-binding sites are not occupied by GATA2, while the remaining 65% overlap with GATA2-binding sites [31]. The fact that GATA1 and GATA2 often bind the same subset of genomic locations suggests an underlying mechanism for molecular competition via association and disassociation at the transcription factor binding site. Interestingly, it has also been previously shown that transcription factor crowding on the DNA may increase transcriptional noise through increased variability of the occupancy time of the target sites, leading to cell-to-cell variation [32].

GATA factor interplay is thought to be a common mechanism for controlling developmental processes [33, 34]. During erythropoiesis, GATA2 is expressed prior to GATA1, which suggests that GATA2 binding may promote GATA1 accessibility to GATA motifs. GATA1 occupancy on chromatin has been shown to activate transcription of a differentiation program leading to committed erythroid cells. Here, we test whether the observed variation of DNA accessibility at GATA binding sites resembles functionally distinct developmental cell states. We hypothesize that the accessibility variation results mainly from differential expression levels of GATA in K562 cells (Additional file 1: Figure S1a). To analyze the functional impact of GATA expression and motif accessibility variability, we set out to find a cell surface marker that co-varied with GATA expression levels to allow sorting of live cells from a mixed population for subsequent functional experiments.

Our strategy (Fig. 1a) to identify co-varying transcription factor–cell surface marker pairs starts with analysis

of scATAC-seq data, in which we focus on transcription factor motif variability, identifying a transcription factor of interest with variable binding between cells (Fig. 1b). Second, we investigate existing RNA-seq data for cell surface marker expression. scRNA-seq data helps to focus on highly abundant and variably expressed genes. The addition of transcription factor knockdown RNA-seq data allows us to further narrow down candidates. The third phase is the confirmation of co-variation of the transcription factor with the cell surface marker.

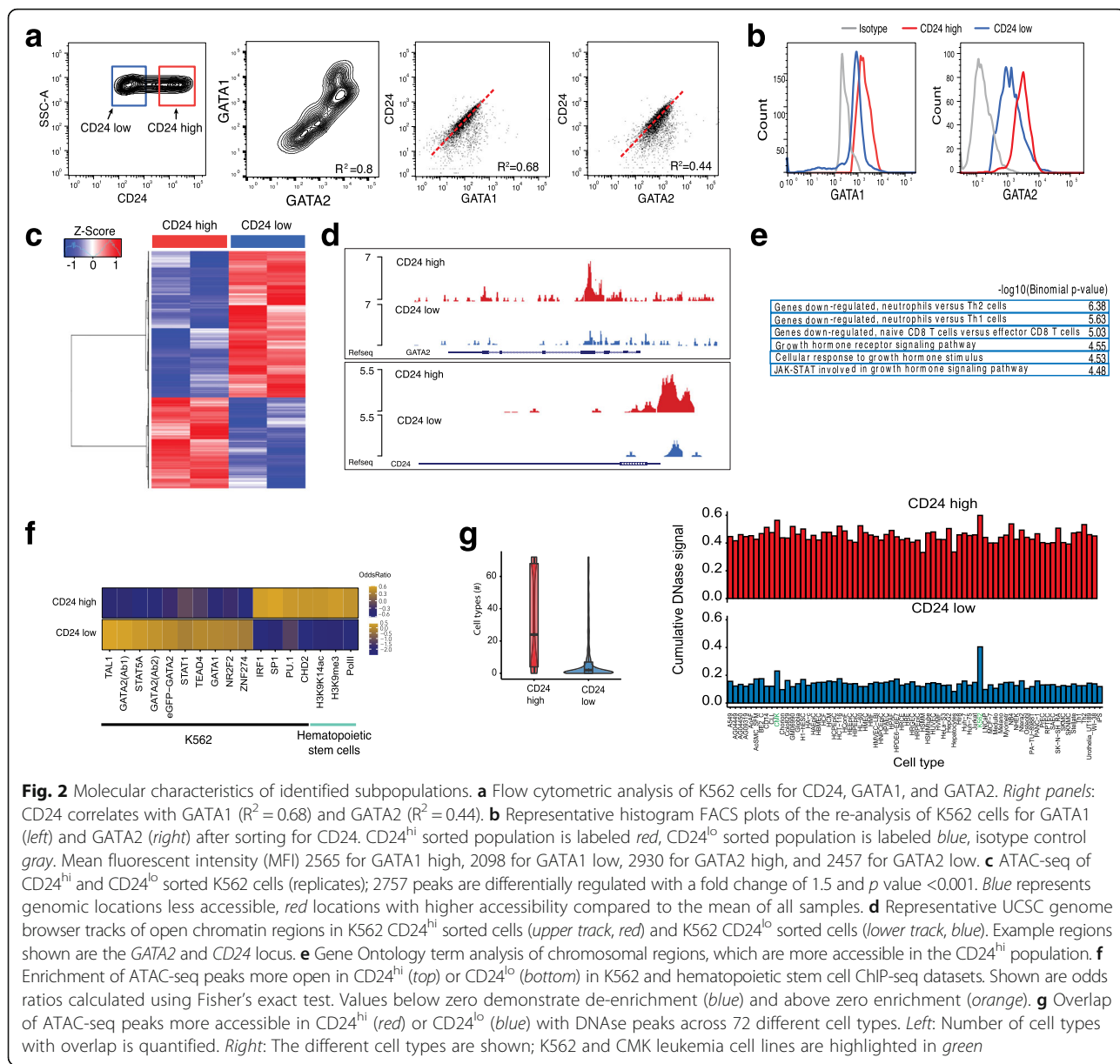
Here, K562 scRNA-seq data [35] were analyzed focusing on highly expressed, yet highly variable, cluster of differentiation (“CD”) cell surface genes (red dots in Fig. 1c). In addition, we re-analyze published *GATA1* and *GATA2* knockdown RNA-seq data [36], identifying CD-annotated genes which were both highly expressed and changed expression following *GATA* knockdown in K562 cells (Fig. 1d). Combining both datasets, we identified *CD24*, *CD44*, and *CD52* mRNAs as encoding candidate cell surface genes that were highly variable.

#### Validation of a co-varying “surrogate” marker for GATA motif variation

To test CD24, CD44, and CD52 as surrogate cell surface markers for GATA variation, we sorted cells with fluorescence-activated cell sorting (FACS). CD44 was only weakly expressed and CD52 did only partially correlate with GATA expression (Additional file 1: Figure S1b). CD24 is expressed and is highly variable in K562 cells (Fig. 2a, left panel); in addition we found two populations, CD24<sup>hi</sup> (red square) and CD24<sup>lo</sup> (blue square) (Additional file 1: Figure S1c). GATA1 and GATA2 are also heterogeneously expressed in K562 cells (Fig. 2a, middle panel), with cells expressing low levels of GATA1 also tending to express low levels of GATA2. In a cell with high CD24 expression, GATA1 and GATA2 tend also to be more highly expressed (Fig. 2a, right panels). To further link high expression of CD24 with GATA high cells, cells sorted for CD24 high and low expression were stained and analyzed for GATA. The result shows that in CD24<sup>hi</sup> cells, protein as well as mRNA levels of GATA1 and GATA2 are higher compared to CD24<sup>lo</sup> sorted cells (Fig. 2b; Additional file 1: Figure S1d). Notably, expression of phospho-JUN, another transcription factor which displayed high variation in motif accessibility in K562 scATAC-seq experiments [20], does not differ between sorted populations (Additional file 1: Figure S1e). In summary, our data show that CD24 cells are GATA positive and CD24 is thus a surrogate marker for GATA factor expression level in K562 cells.

#### Molecular analysis of the identified subpopulations

Focusing on molecular and functional differences of CD24 high versus low K562 subpopulations, we used



our CD24 surrogate marker to identify epigenomic differences of the two subpopulations with ATAC-seq. In contrast to other cell lines, the mitochondria are particularly highly represented in K562 cells, resulting in high mitochondrial DNA representation in ATAC-seq libraries. Therefore, we developed an optimized ATAC-seq protocol for K562, which includes an optimized cell lysis and additional nuclei washes prior to transposition, reducing mitochondrial representation from approximately 75 to 35% (see “Methods” for details). Differential peak analysis showed 2757 differentially accessible peaks (fold change (FC) of 1.5,  $p$  value 0.001; Fig. 2c; Additional file 2: Figure S2a), of which 1698 were more accessible in CD24<sup>lo</sup> and 1059 more accessible in CD24<sup>hi</sup> sorted K562

cells. Representative UCSC genome browser tracks of open chromatin regions of CD24<sup>hi</sup> and CD24<sup>lo</sup> sorted K562 cells are displayed in Fig. 2d and Additional file 2: Figure S2b. Interestingly, open chromatin regions cluster around transcription start sites in CD24<sup>hi</sup> (26% in high versus 4% low), whereas in CD24<sup>lo</sup> K562 cells distal chromatin regions are more accessible (Additional file 2: Figure S2c), suggesting general differential chromatin regulation in these subpopulations. Next we set out to confirm that the differentially accessible sites between CD24<sup>hi</sup> and CD24<sup>lo</sup> are functionally relevant. First, we performed Gene Ontology (GO) analysis [37] with all regions more accessible in the CD24<sup>hi</sup> population, using total accessible locations of K562 cells as

background set. These regions are associated with genes involved in neutrophil versus T-cell differentiation, as well as in growth hormone signaling. In particular, STAT signaling is enriched, a signaling pathway involved in CML and BCR-ABL signaling (Fig. 2e) [38, 39]. The resulting gene list was further analyzed with the PANTHER database (<http://pantherdb.org>), showing the highest biological process GO term enrichment for “regulation of hematopoiesis” (GO:1903706). In contrast, the GO terms resulting from chromatin regions more accessible in CD24<sup>lo</sup> cells are associated with promoters bound by FOXP3, maturation of monocytes in response to inflammation, MYC overexpression, and genes up-regulated in response to BCR-ABL (Additional file 2: Figure S2d). In addition, we correlated the ATAC-seq peaks more open in CD24<sup>lo</sup> (1698 genomic regions) as well as those more open in CD24<sup>hi</sup> (1059 genomic regions) to all available K562 ChIP-seq datasets using LOLA (Locus Overlap Analysis: Enrichment of Genomic Ranges), using total accessible locations of K562 CD24<sup>hi</sup> and CD24<sup>lo</sup> cells as background set [40]. Interestingly, ChIP-seq signals of TAL-1, GATA1, and GATA2, factors involved in hematopoietic differentiation [41, 42], are preferentially enriched in accessible locations in CD24<sup>lo</sup> K562 cells. In CD24<sup>hi</sup> K562 cells on the other hand, binding sites of the ubiquitous transcription factors SP1, SP2, and CHD2 are enriched, as well as PU.1 sites (Fig. 2f). In addition to the intersection of our ATAC-seq data with ChIP-seq data, we intersected our differential ATAC-seq regions with the regulatory elements database DNase hypersensitivity data [43]. In line with the previous results, we found high overlap of CD24<sup>lo</sup> K562 accessible sites with K562 enriched DNase hypersensitivity clusters, but no enrichment for any specific cell/tissue type for the CD24<sup>hi</sup> accessible genomic regions (Fig. 2g; Additional file 2: Figure S2e).

These molecular analyses of K562 subpopulations show significantly higher *GATA2* expression in CD24<sup>hi</sup> cells compared to CD24<sup>lo</sup> K562 cells (Additional file 1: Figure S1d). However, the CD24<sup>lo</sup> population exhibits more accessibility at GATA and TAL1 binding sites (Fig. 2f, g; Additional file 2: Figure S2f), transcription factors regulating differentiation into erythrocytes, suggesting that these cells might be more differentiated erythro-leukemic cells. In contrast, the CD24<sup>hi</sup> K562 population exhibits less erythropoietic-specific transcription factor binding and more accessibility at hematopoietic progenitor maintenance factors, like PU.1 (Fig. 2f, g). PU.1 is a key regulator of hematopoietic differentiation, which is tightly regulated transcriptionally and not expressed in differentiated erythroid or myeloid cells [44] and thus implicates CD24<sup>hi</sup> as a less differentiated “stem-like” subpopulation. Importantly, *GATA2*, and not *GATA1*, is highly expressed in hematopoietic stem cells,

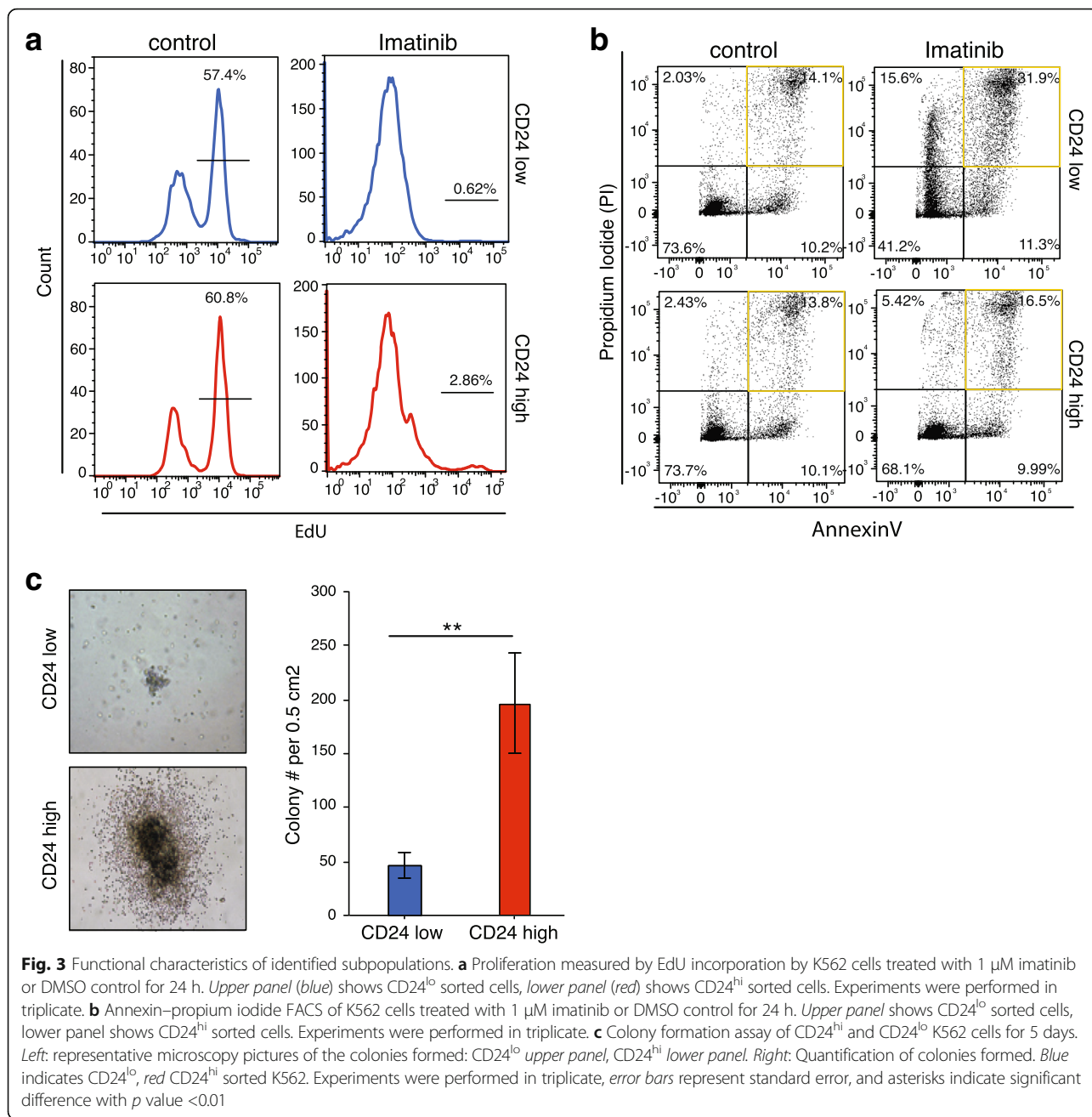
but through erythropoietic differentiation *GATA1* is highly expressed while *GATA2* expression is lost [45]. This “GATA factor switch” is at the center of hematopoietic differentiation and is mediated by GATA factor competition in erythropoietic progenitors, whereby *GATA2* acts as a repressor by inhibiting *GATA1* activation of erythropoietic gene expression [46, 47]. In addition, the overexpression of *GATA2* strongly promotes hematopoietic stem cell self-renewal, altogether implicating *GATA2* as a stem-ness factor [48].

We observe on the one hand higher expression of *GATA1* and *GATA2* in the CD24<sup>hi</sup> population, an expression signature for more differentiated erythroid cells; on the other hand CD24<sup>hi</sup> has more accessible binding sites for stem-ness transcription factors. We assume that the high expression of *GATA* in the CD24<sup>hi</sup> state leads to the overall loss in *GATA* motif accessibility, whereas *GATA* motif chromatin accessibility is higher in the more differentiated CD24<sup>lo</sup> cells, in which *GATA* is also less expressed.

#### Functional analysis of the identified subpopulations

Next, we set out to analyze the functional effects of the observed epigenomic variability. The K562 cell line is derived from female human chronic myelogenous leukemia cells, which are positive for the Philadelphia chromosome and bear characteristics of multipotent progenitors [49, 50]. To further elucidate the phenotypic differences of the two subpopulations we treated the CD24<sup>hi</sup> and CD24<sup>lo</sup> sorted cells with imatinib mesylate (Gleevec) [51], a BCR-ABL tyrosine kinase inhibitor approved for CML treatment, and observed the effects on proliferation and apoptosis (Fig. 3a, b; Additional file 3: Figure S3a, b). We assayed proliferation by monitoring the incorporation of alkyne-containing thymidine analog EdU (5-ethynyl-2'-deoxyuridine), which is incorporated into DNA during active DNA synthesis [52]. EdU incorporation was significantly inhibited in both subpopulations upon treatment, but 2.9% of CD24<sup>hi</sup> sorted cells continued proliferating, in contrast to CD24<sup>lo</sup> sorted cells (Fig. 3a lower right panel; Additional file 3: Figure S3a). To further analyze the differential drug response in more detail, the apoptosis rate of the two cell populations after drug treatment was measured. The percentage of annexin V-propidium iodide (PI)-positive cells increased from 14% in control to 32% in the CD24<sup>lo</sup> population, whereas the number of CD24<sup>hi</sup> cells undergoing apoptosis was similar (13.8 to 16.5%) (Fig. 3b; Additional file 3: Figure S3b). Therefore, we conclude that CD24<sup>hi</sup> cells are more resistant to imatinib mesylate treatment than CD24<sup>lo</sup> cells.

To further support our hypothesis that the CD24<sup>hi</sup> subpopulation might resemble the more stem cell-like population, whereas the CD24<sup>lo</sup> subpopulation might be



**Fig. 3** Functional characteristics of identified subpopulations. **a** Proliferation measured by EdU incorporation by K562 cells treated with 1  $\mu$ M imatinib or DMSO control for 24 h. *Upper panel (blue)* shows CD24<sup>lo</sup> sorted cells, *lower panel (red)* shows CD24<sup>hi</sup> sorted cells. Experiments were performed in triplicate. **b** Annexin–propidium iodide FACS of K562 cells treated with 1  $\mu$ M imatinib or DMSO control for 24 h. *Upper panel* shows CD24<sup>lo</sup> sorted cells, *lower panel* shows CD24<sup>hi</sup> sorted cells. Experiments were performed in triplicate. **c** Colony formation assay of CD24<sup>hi</sup> and CD24<sup>lo</sup> K562 cells for 5 days. *Left*: representative microscopy pictures of the colonies formed: CD24<sup>lo</sup> *upper panel*, CD24<sup>hi</sup> *lower panel*. *Right*: Quantification of colonies formed. Blue indicates CD24<sup>lo</sup>, red CD24<sup>hi</sup> sorted K562. Experiments were performed in triplicate, error bars represent standard error, and asterisks indicate significant difference with *p* value <0.01

more differentiated, we performed a colony forming cell (CFC) assay, which measures the capacity of single cells to replicate in a semisolid medium, with both sorted subpopulations. The CFC assay allows us to assess the amount of leukemic progenitors within these populations. CD24<sup>hi</sup> sorted cells formed over fourfold more colonies CD24<sup>lo</sup> cells (Fig. 3c) and these colonies were generally larger, with a dense core and some outgrowing cells surrounding a ring (Fig. 3c, left panels). These results suggest that the CD24<sup>hi</sup> population has more progenitor capacity than the CD24<sup>lo</sup> subpopulation.

We harvested cells from more than four individual colonies or from the whole plate after the CFC assay to further assess their numbers and differentiation states using FACS. We analyzed the CD24 status of the harvested colonies and were surprised to find that the CD24<sup>hi</sup> subpopulation contained only 30% CD24<sup>hi</sup> expressing cells; thus, the majority lost their CD24 expression (Additional file 3: Figure S3c). In contrast, the majority of the CD24<sup>lo</sup> population stayed in the low state, gaining only 6.68% CD24 positive cells. These results suggest that the differentiation state of cancer

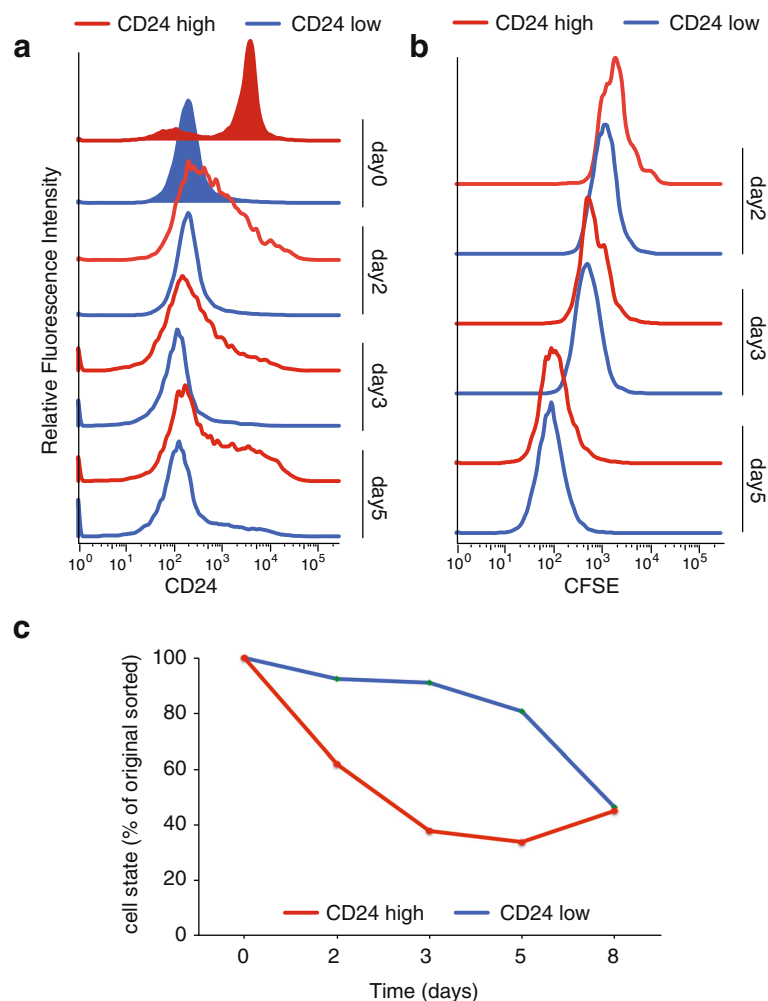
cells is dynamic, consistent with findings in other cancer stem cell systems [53].

#### Epigenomic plasticity of K562 subpopulations

To further investigate these dynamics, K562 cells were sorted for the two subpopulations and immediately stained with the cell tracker 5-(and 6)-carboxyfluorescein diacetate succinimidyl ester (CFSE). CFSE readily crosses intact cell membranes, and after staining cell division can be measured as successive halving of the fluorescence intensity. For five consecutive days CD24 and CFSE signals of the two subpopulations were measured using flow cytometry. Both populations re-established the initial population distribution of CD24<sup>hi</sup> and CD24<sup>lo</sup> cells, suggesting that both correspond to metastable, temporally dynamic epigenomic states. We observed a rapid loss of CD24 high expressing cells of

the CD24<sup>hi</sup> sorted subpopulation, whereas the CD24<sup>lo</sup> subpopulation dynamic changes occurred more slowly (Fig. 4a, c). Both populations proliferated at an equal rate during that time (Fig. 4b). These observations lead to the conclusion that the CD24-GATA-high population is dynamic, and contributes to epigenomic plasticity of K562 cells (Fig. 4c).

To validate the epigenomic plasticity of the identified K562 populations, we cultured the sorted cells (d0) for 5 days (d5) and performed ATAC-seq on CD24 d5 subpopulations. The CD24<sup>hi</sup> population is able to generate both CD24<sup>hi</sup> and CD24<sup>lo</sup> populations within 5 days. We compared the epigenome of the new CD24<sup>hi</sup>-CD24<sup>lo</sup> populations to each other as well as to the initial sorted (parental) population (Additional file 4: Figure S4a, b): 2884 peaks are differentially accessible in the d5 K562 cells started from the CD24<sup>hi</sup> population, 1372 more



**Fig. 4** Epigenomic plasticity of K562 subpopulations. **a** FACS analysis of CD24 sorted K562 cells. Shown are the initial sort (*tinted*) and the flow cytometric re-analysis at days 2, 3, and 5. *Blue* indicates CD24<sup>lo</sup> sorted K562 cells, *red* CD24<sup>hi</sup> sorted population. **b** Proliferation analysis of K562 sorted subpopulations. After the initial sort CD24<sup>hi</sup> and CD24<sup>lo</sup> cells were stained with CFSE and then cultured for 8 days. CFSE fluorescence intensity was measured at days 2, 3, and 5 together with CD24 (**a**). **c** Quantification of the changes in CD24-expressing cells. *Blue*, CD24<sup>lo</sup>; *red*, CD24<sup>hi</sup>



accessible in d5 CD24<sup>hi</sup>, 1512 more accessible in d5 CD24<sup>lo</sup>. The peaks of the parental CD24 sorted K562 cells correlated with the peaks accessible after 5 days with an R of 0.78 and 0.79, respectively (Additional file 4: Figure S4b). Moreover, the new CD24<sup>hi</sup> and CD24<sup>lo</sup> populations show the same molecular and phenotypic features as their respective parental line. We analyzed differentially accessible regions between day 5 CD24<sup>lo</sup> and CD24<sup>hi</sup> originating from CD24<sup>hi</sup> using LoLa. The enrichment of accessibility for the respective hematopoietic or more stem factors is in line with what we found with the parental population (Additional file 4: Figure S4c). In addition, we confirmed the functional difference between day 5 CD24<sup>lo</sup> and CD24<sup>hi</sup> by apoptosis assay after drug treatment. We sorted day 5 CD24<sup>hi</sup> and CD24<sup>lo</sup> K562 cells, treated those with 1  $\mu$ M imatinib and analyzed them for apoptosis by annexin-PI FACS after 24 h (similar to Fig. 3b). The second-generation CD24<sup>hi</sup> population cells were less susceptible to the drug (11.1% (standard deviation = 0.84) annexin- and annexin-PI-positive cells compared to 18.5% (standard deviation = 1.56) annexin and annexin-PI positive cells of the second generation CD24<sup>lo</sup>) (Additional file 4: Figure S4d). These results recapitulate the functional heterogeneity found after the first CD24 sort.

## Conclusions

We demonstrate an integrative strategy to prospectively isolate epigenomic subpopulations of cells defined by single-cell chromatin activity. Data mining of available knockdown as well as scRNA-seq data allow correlation of cell surface marker expression with transcription factor variability. scRNA-seq data are generally sparse, making gene-gene correlations, especially of often lowly expressed transcription factors, a particularly difficult task. Our approach, described above, circumvents these issues by looking at functional co-variation using bulk transcription factor knockdowns. This strategy nominates co-varying cell surface markers, which can then be used to identify functional distinct subgroups in cancer cells. A similar approach has been described to resolve heterogeneity within stem cell populations, combining RNA-seq with flow cytometry data [54]. With new genetic perturbation tools like CRISPR [55, 56] and CRISPRi [57], we anticipate this strategy to become more generally applicable and a common tool for single-cell epigenomics. In addition, we anticipate that new high-throughput single-cell genomics methods will be invaluable for efficiently discovering co-varying cell surface markers. Specifically, high-throughput scRNA-seq profiling has been shown to uncover gene-expression networks [58, 59]. Currently, low throughput epigenomics methods preclude identification of the individual regulatory elements within cell populations; however, we anticipate

that high-throughput epigenomic methods may enable de novo identification of hidden epigenomic states. This strategy should be broadly applicable to many cancer types and disease states to unravel molecular drivers of epigenomic state and to improve therapeutic targeting.

## Methods

### Cell culture and reagents

K562 (ATCC) chronic myeloid leukemia cells were maintained in Iscove's modified Dulbecco's medium (IMDM) containing 10% fetal bovine serum (HyClone, Thermo Scientific) and 1% penicillin streptomycin (Pen/Strep). Cells were maintained at 37 °C and 5% CO<sub>2</sub> at recommended density and were treated and harvested at mid-log phase for all experiments.

### Drug treatments

K562 cells were treated with 1  $\mu$ M imatinib mesylate (Gleevec, Cayman Chemicals, Ann Arbor, MI, USA) or DMSO control for 24 h.

### FACS and flow cytometric analysis

In a 1.5 mL tube, cells were washed with ice cold phosphate-buffered saline (PBS). For (CD) cell surface markers, cells were stained with PE-CD24 (#555428, BD Biosciences), or APC-CD44 (#559942, BD Biosciences) or APC-CD52 (Clone HI186, BioLegend) in PBS containing 2 mM EDTA and 0.5% bovine serum albumin (BSA) on ice in the dark for 30 min. For subsequent intracellular staining, cells were fixed in 1% paraformaldehyde (PFA) for 10 min followed by permeabilization using 0.5% TritonX100 in PBS for 10 min at room temperature. Cells were stained with primary antibodies rabbit anti-GATA1 (1:400, Cell Signaling, D52H6), mouse anti-GATA2 (1:100, Abnova, H00002624-M01), rabbit anti phospho c-JUN II (Ser63, Cell Signaling), or mouse or rabbit IgG as isotype control in PBS containing 0.5% TritonX100, 2 mM EDTA and 0.5% BSA (Sigma) for 1 h at room temperature. After washing with staining buffer, cells were labeled with Alexa-conjugated donkey anti-mouse or anti-rabbit Alexa 488 or Alexa 647 antibodies (life technologies) at a dilution of 1:500 for 30 min at room temperature. Finally, cells were washed and sorted for CD24 or analyzed using the BD FACSAriaII.

Flow cytometric analysis and statistics were performed using FlowJo V.10.0.8.

### ATAC-seq

K562 cells were stained and sorted for CD24 as described above. ATAC of  $5 \times 10^4$  cells was performed as previously described [20], changing the lysis and ATAC conditions slightly. Lysis was performed in 100  $\mu$ l cold buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub> + 0.1% IGEPAL CA-630 + 0.1% Tween 20),

transposition was performed in 50  $\mu$ l buffer containing 25  $\mu$ l 2 $\times$  TD buffer (Illumina #FC-121-1030), 2.5  $\mu$ l Tn5 transposase (Illumina #FC-121-1030), 22.5  $\mu$ l nuclease free H<sub>2</sub>O, 0.5  $\mu$ l Tween-20 (0.1% final), followed by the recommended library preparation protocol. The resulting libraries were quantified and sequencing data were generated on an Illumina HiSeq 4000 that was purchased with funds from NIH under award number S10OD018220.

#### Data processing

All ATAC-seq libraries were sequenced using paired-end, dual-index sequencing using 76  $\times$  8  $\times$  8  $\times$  76 cycle reads on a NextSeq. Adapter sequences were trimmed from FASTQs using custom python scripts to enable mapping fragments smaller than 50 bp. Paired-end reads were aligned to hg19 using BOWTIE2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) with the parameter `-very-sensitive`. Duplicates were removed and library size was estimated using PICARD tools (<http://picard.sourceforge.net>). Reads were subsequently filtered for alignment quality of  $> Q30$  and were required to be properly paired. Reads mapping to the mitochondria or chromosome Y were removed and not considered. We used MACS2 (<http://pypi.python.org/pypi/MACS2>) to call all reported ATAC-seq peaks. MACS2 was used with the following parameters (`-nomodel -shift 0`). Peaks were filtered using the consensus excludable ENCODE blacklist (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/>) and a custom blacklist designed to remove high-signal-causing repeats and mitochondrial homologues. Using the filtered peak set, peak summits were extended  $\pm 250$  bps. The top 50,000 non-overlapping 500-bp summits, which we refer to as accessibility peaks, were used for all downstream analysis.

Peaks from all samples were merged and normalized. For differentially accessible peaks a cutoff of 1.5-fold change and  $p$  value  $< 0.01$  between CD24<sup>hi</sup> and CD24<sup>lo</sup> were used. For ATAC-seq peak–ChIP-seq and DNase-seq correlation analysis we used the LOLA bioconductor package with all K562 peaks from these ATAC-seq experiments as background set. For enrichment of GATA2-bound motifs in ATAC-seq peaks, ChIP-seq dataset GSM935373 was intersected with ATAC-seq peaks.

GO term analysis was performed using GREAT (<http://great.stanford.edu>) [37].

K562 CD24 sorted ATAC-seq data from day 0 and day 5 have been deposited in the Gene Expression Omnibus (GEO) with accession GSE76224.

#### Quantitative RT-PCR

Total RNA was isolated with an RNeasy isolation kit (Qiagen) and cDNA was synthesized using the Super-script III First Strand synthesis kit according to the

manufacturer's instructions (Invitrogen). qRT-PCR reactions were performed in a Roche Lightcycler 480 using 2 $\times$  Brilliant II SYBR QRT-PCR Master Mix from Agilent according to standard protocols. All primers were separated by at least one intron on the genomic DNA to exclude amplification of genomic DNA. PCR reactions were checked by including no-RT controls, by omission of templates, and by examining melting curves. Standard curves were generated for each gene. Relative quantification of gene expression was determined by comparison of threshold values. All samples were analyzed in duplicate in two different dilutions. All results were normalized to actin. All experiments were performed in biological triplicates.

Primer sequences were (5'–3' forward, reverse): actin, CCGGCTTCGCGGGCGACG, TCCCGGCCAGCCAGG TCC; *GATA1*, TGCTCTGGTGTCTCCACAC, TGGGA GAGGAATAGGCTGCT; *GATA2*, AGCGTCTCCAGCC TCATCTTCCGCG, CGAGTCTTGCTGCGCCTGCTT.

#### Proliferation measurements

K562 cells were sorted for CD24 and cultured in the presence of 1  $\mu$ M imatinib mesylate or DMSO for 24 h before proliferation analysis. EdU (10  $\mu$ M) was added directly to the media for 4 h before cells were harvested. After that, cells were fixed and stained according to the manufacturer's protocol (Click-iT EdU kit #C10340, Invitrogen). Briefly, cells were fixed with 3.7% formaldehyde for 15 min and permeabilized using 0.5% Triton X-100 in PBS for 20 min at room temperature. Incorporation of EdU was observed by incubating fixed cells with 2% BSA in PBS for 30 min and Alexa fluor 647 for a further 30 min under Cu(I)-catalyzed click reaction conditions, as described by the manufacturer. Cells were washed with PBS and counterstained with DAPI in PBS right before flow cytometric analysis using the BD FACSAriaII.

Experiments were performed in triplicate; the standard 10,000 cells per gate were recorded and analyzed.

#### Apoptosis measurements

K562 cells were sorted for CD24 and cultured in the presence of 1  $\mu$ M imatinib mesylate or DMSO for 24 h before proliferation analysis. Cells were washed with cold PBS containing 0.5% BSA and then resuspended in Annexin V Binding Buffer (BioLegend, #422201). Cells were then incubated for 15 min with 5  $\mu$ l of FITC annexin V (BioLegend, #640906) and 10  $\mu$ l of 1 mg/ml PI solution (BioLegend, #421301) at room temperature in the dark. Apoptosis was measured by flow cytometry using the BD FACSAriaII.

Experiments were performed in triplicate; the standard 10,000 cells per gate were recorded and analyzed.

### Colony formation assay

K562 cells were sorted for CD24. Immediately after sorting, 500 cells in 0.5 ml medium were added to 3 ml methylcellulose-based media (HSC002, R&D Systems). Using a 10 ml syringe and a 16 gauge needle, 1 ml of this mixture was added to a 35-mm dish, which was then placed in a 15-cm dish filled with water to maintain the humidity necessary for colony formation. After 10 days, colonies were counted on a grid using a light microscope. After that, methylcellulose was dissolved in media to make a single-cell suspension. Cells were washed and stained as described above for flow cytometric analysis of CD24 expression using the BD FACSAriaII. Experiments were performed in triplicate.

### Cell tracing experiments (CFSE staining)

K562 cells were sorted for CD24. Immediately after sorting, 200,000 cells of the high- and low-sorted population were stained with 5  $\mu$ M CFSE (Cell Trace Proliferation Kit, Life Technologies) according to the manufacturer's protocol. Cell proliferation (CFSE dilution) and CD24 surface expression were analyzed every 24 h for 8 days using the BD FACSAriaII.

Experiments were performed in triplicate; the standard 10,000 cells per gate were recorded and analyzed.

### Additional files

**Additional file 1: Figure S1.** Molecular characteristics of identified subpopulations. *a Left:* QRT-PCR of *GATA1* and *GATA2* in K562 cells measured relative to *ACTIN*. *Error bars* represent standard error. *Right:* Representative FACS analysis of K562 cells stained for *GATA1* and *GATA2*. *b* Histograms showing expression (fluorescent intensity) of CD44 (*light blue*) and CD52 (*dark blue*) in K562 cells. *c* Dot plots displaying the gating strategy for sorting CD24<sup>hi</sup> and CD24<sup>lo</sup> expressing K562 cells. *d* Expression analysis of *GATA1* and *GATA2* in CD24 sorted K562 cells measured by qRT-PCR relative to *ACTIN*. \**P* value <0.05 (t-test), *error bars* represent standard error. *e* Representative FACS analysis of CD24<sup>hi</sup> and CD24<sup>lo</sup> sorted K562 cells, stained after sort for pJUN. Mean fluorescent intensity (MFI) was 152 (high), and 137 (low). (PDF 616 kb)

**Additional file 2: Figure S2.** Molecular characteristics of identified subpopulations. *a* Heatmap of the correlation coefficient of K562 CD24 sorted ATAC-seq samples using Spearman correlation. *b* UCSC tracks showing examples of open chromatin in CD24<sup>lo</sup> (*blue*) and CD24<sup>hi</sup> (*red*) K562 cells. *Top:* *JUN* locus, which is more accessible in CD24<sup>lo</sup>. *Bottom:* *SOD1* locus, which is equally accessible in both subpopulations. *c* Bar plot illustrating the distribution of ATAC-seq peaks across genomic locations; promoter proximal (*light blue*) and distal (*dark blue*). The difference in promoter accessibility between CD24<sup>hi</sup> and CD24<sup>lo</sup> K562 cells is significant (using Chi-squared), *p* < 0.001. *d* Gene Ontology terms for accessible chromatin locations in CD24<sup>lo</sup> cells. *e* Q-Q plot illustrating the differences in the overlap of ENCODE DNase-seq peaks and ATAC-seq peaks of CD24<sup>lo</sup> and CD24<sup>hi</sup> populations (shown in Fig. 2g). *P* values for Wilcox (*W*), t-test (*T*), and Komorov-Smirnov (*K*S). *f* Enrichment of *GATA2* ChIP-seq binding sites in CD24<sup>hi</sup> and CD24<sup>lo</sup> K562 ATAC-seq peaks. (PDF 507 kb)

**Additional file 3: Figure S3.** Functional characteristics of identified subpopulations. *a* Quantification of EdU incorporation as a measurement of proliferation. Experiments were done in triplicate; *asterisk* indicates significance, calculated using t-test, *p* value <0.05. *b* Quantification of apoptosis of K562 cells treated with 1  $\mu$ M imatinib or DMSO control for 24 h. AnnexinV-PI negative cells are counted as live, annexin V-positive-

PI-negative cells as early apoptotic and annexin V-PI double positive as late apoptotic. Experiments were done in triplicate; *asterisk* indicates significance, calculated using t-test, *p* value <0.01. *Error bars* represent standard error. *c* Representative FACS analysis of CD24<sup>hi</sup> and CD24<sup>lo</sup> sorted K562 cells for CD24 expression after 5-day colony formation assay. (PDF 138 kb)

**Additional file 4: Figure S4.** Molecular and functional analysis of epigenetic dynamics. *a* Heat map of differentially accessible ATAC-seq peaks of day 5 CD24<sup>hi</sup> and CD24<sup>lo</sup> K562 cells (replicates). CD24<sup>hi</sup> as parental line: 2884 peaks are differentially accessible, 1372 more accessible in day 5 CD24<sup>hi</sup>, 1512 more accessible in day 5 CD24<sup>lo</sup>. Fold change of 1.5 and *p* value <0.001. *Blue* represents genomic locations less accessible, *red* locations with higher accessibility compared to the mean of all samples. *b* Spearman correlation of day 5 K562 CD24<sup>hi</sup> (*top*) and CD24<sup>lo</sup> (*bottom*) ATAC-seq peaks (log counts) and day 0 (parental) ATAC-seq peaks. *R* = 0.78 and 0.79, respectively, *p* value of the correlation <2.2e-16 and 0.0012, respectively. *c* LoLa analysis of differentially accessible peaks of new (day 5) CD24<sup>hi</sup> and CD24<sup>lo</sup>. List of the highest odds ratios in untreated K562 datasets. *d* AnnexinV-PI apoptosis FACS of day 5 CD24<sup>hi</sup> and CD24<sup>lo</sup> 24 h after imatinib treatment. Bar plots demonstrate percentage of total cells in different phases of apoptosis. *Red* bars represent cells originating from CD24<sup>hi</sup>, *blue* bars demonstrate CD24<sup>lo</sup> descendants. *N* = 3. (PDF 7944 kb)

### Abbreviations

ATAC-seq: Assay for transposase-accessible chromatin with high-throughput sequencing; BSA: Bovine serum albumin; CFC: Colony formation assay; CFSE: Carboxyfluorescein succinimidyl ester; ChIP: Chromatin immunoprecipitation; CML: Chronic myeloid leukemia; EdU: 5-ethynyl-2'-deoxyuridine; FACS: Fluorescence-activated cell sorting; GO: Gene Ontology; PBS: Phosphate-buffered saline; PI: Propidium iodide; qRT-PCR: Quantitative reverse transcription polymerase chain reaction; sc: Single-cell; scATAC-seq: Single-cell assay for transposase-accessible chromatin with high-throughput sequencing; scRNA-seq: Single-cell RNA sequencing

### Acknowledgements

We thank the Seung Kim lab for generous assistance with cell sorting. Supported by NIH grants P50-HG007735 (H.Y.C., W.J.G.) and R35-CA209919 (H.Y.C.). N.C.S. was supported by a long-term fellowship of the Human Frontier Science Program (LT000211/2014).

### Funding

H.Y.C. was funded by NIH# P50-HG007735 and R35-CA209919. W.J.G. was funded by NIH#P50-HG007735 and U19AI057266 as well as the Human Frontier Science Program, the Rita Allen Foundation, and the Baxter Foundation. NS was funded by the Human Frontier Science Program #LT000211/2014.

### Availability of data and material

Sequencing data are stored and openly available at the GEO (Gene Expression Omnibus; <https://www.ncbi.nlm.nih.gov/geo/>) under accession GSE76224.

### Authors' contributions

UML, W.J.G., and H.Y.C. designed the experiments and wrote the manuscript. UML designed and performed all experiments. JDB analyzed the RNA-seq and scRNA-seq data and helped write the manuscript. JDB and BW performed the initial K562 scATAC-seq experiments and analysis. AK and JDB optimized the K562 specific ATAC-seq protocol. YS and NS helped with the ATAC-seq analysis. All authors read and approved the final manuscript.

### Competing interests

Stanford University has filed a provisional patent application on the methods described, and J.D.B., H.Y.C. and W.J.G. are named as inventors. The other authors declare that they have no competing interests.

### Ethics approval and consent to participate

Not applicable

### Author details

<sup>1</sup>Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>2</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>3</sup>Department of Applied Physics, Stanford University, Stanford, CA 94035, USA. <sup>4</sup>Broad

Institute of MIT and Harvard, Cambridge, MA 02142, USA. <sup>5</sup>Harvard Society of Fellows, Harvard University, Cambridge, MA 02138, USA.

Received: 8 July 2016 Accepted: 12 December 2016

Published online: 24 January 2017

## References

1. Buffo A, Rite I, Tripathi P, Lepier A, Colak D, Horn AP, et al. Origin and progeny of reactive gliosis: a source of multipotent cells in the injured brain. *Proc Natl Acad Sci U S A*. 2008;105(9):3581–6.
2. Sirko S, Behrendt G, Johansson PA, Tripathi P, Costa Z, Bek S, et al. Reactive glia in the injured brain acquire stem cell properties in response to sonic hedgehog. [corrected]. *Cell Stem Cell*. 2013;12(4):426–39.
3. Friedmann-Morvinski D, Verma IM. Dedifferentiation and reprogramming: origins of cancer stem cells. *EMBO Rep*. 2014;15(3):244–53.
4. Micalizzi DS, Farabaugh SM, Ford HL. Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. *J Mammary Gland Biol Neoplasia*. 2010;15(2):117–34.
5. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007;129(4):823–37.
6. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell*. 2008;132(2):311–22.
7. He HH, Meyer CA, Shin H, Bailey ST, Wei G, Wang Q, et al. Nucleosome dynamics define transcriptional enhancers. *Nat Genet*. 2010;42(4):343–7.
8. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. *Science*. 2007;316(5830):1497–502.
9. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 2007;448(7153):553–60.
10. 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.
11. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*. 2014;509(7500):371–5.
12. Farlik M, Sheffield NC, Nuzzo A, Datlinger P, Schonegger A, Klughammer J, et al. Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep*. 2015;10(8):1386–97.
13. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods*. 2014;11(8):817–20.
14. Rotem A, Ram O, Shoreh N, Sperling RA, Goren A, Weitz DA, et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol*. 2015;33(11):1165–72.
15. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. *Genome Res*. 2007;17(6):877–85.
16. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013;10(12):1213–8.
17. Jin W, Tang Q, Wan M, Cui K, Zhang Y, Ren G, et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature*. 2015;528(7580):142–6.
18. Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*. 2015;348(6237):910–4.
19. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*. 2015;523(7561):486–90.
20. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol*. 2015;109:21.29.1–9.
21. North TE, de Bruijn MF, Stacy T, Talebian L, Lind E, Robin C, et al. Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity*. 2002;16(5):661–72.
22. Burda P, Laslo P, Stopka T. The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. *Leukemia*. 2010;24(7):1249–57.
23. Knudsen KJ, Rehn M, Hasemann MS, Rapin N, Bagger FO, Ohlsson E, et al. ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation. *Genes Dev*. 2015;29(18):1915–29.
24. Ramos-Mejia V, Navarro-Montero O, Ayllon V, Bueno C, Romero T, Real PJ, et al. HOXA9 promotes hematopoietic commitment of human embryonic stem cells. *Blood*. 2014;124(20):3065–75.
25. Steelman LS, Pohnert SC, Shelton JG, Franklin RA, Bertrand FE, McCubrey JA. JAK/STAT, Raf/MEK/ERK, PI3K/Akt and BCR-ABL in cell cycle progression and leukemogenesis. *Leukemia*. 2004;18(2):189–218.
26. Ward AC, Touw I, Yoshimura A. The Jak-Stat pathway in normal and perturbed hematopoiesis. *Blood*. 2000;95(1):19–29.
27. Minegishi N, Morita S, Minegishi M, Tsuchiya S, Konno T, Hayashi N, et al. Expression of GATA transcription factors in myelogenous and lymphoblastic leukemia cells. *Int J Hematol*. 1997;65(3):239–49.
28. Nagai T, Harigae H, Ishihara H, Motohashi H, Minegishi N, Tsuchiya S, et al. Transcription factor GATA-2 is expressed in erythroid, early myeloid, and CD34+ human leukemia-derived cell lines. *Blood*. 1994;84(4):1074–84.
29. Shimamoto T, Ohyashiki JH, Ohyashiki K, Kawakubo K, Kimura N, Nakazawa S, et al. GATA-1, GATA-2, and stem cell leukemia gene expression in acute myeloid leukemia. *Leukemia*. 1994;8(7):1176–80.
30. Shimizu R, Takahashi S, Ohneda K, Engel JD, Yamamoto M. In vivo requirements for GATA-1 functional domains during primitive and definitive erythropoiesis. *EMBO J*. 2001;20(18):5250–60.
31. Fujiwara T, O'Geen H, Keles S, Blahnik K, Linnemann AK, Kang YA, et al. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell*. 2009;36(4):667–81.
32. Zabet NR, Adryan B. The effects of transcription factor competition on gene regulation. *Front Genet*. 2013;4:197.
33. Bresnick EH, Lee HY, Fujiwara T, Johnson KD, Keles S. GATA switches as developmental drivers. *J Biol Chem*. 2010;285(41):31087–93.
34. Bresnick EH, Martowicz ML, Pal S, Johnson KD. Developmental control via GATA factor interplay at chromatin domains. *J Cell Physiol*. 2005;205(1):1–9.
35. Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, et al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol*. 2014;32(10):1053–8.
36. Lan X, Witt H, Katsumura K, Ye Z, Wang Q, Bresnick EH, et al. Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic Acids Res*. 2012;40(16):7690–704.
37. McLean CY, Bristol D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010;28(5):495–501.
38. Lin TS, Mahajan S, Frank DA. STAT signaling in the pathogenesis and treatment of leukemias. *Oncogene*. 2000;19(21):2496–504.
39. Danial NN, Rothman P. JAK-STAT signaling activated by Abl oncogenes. *Oncogene*. 2000;19(21):2523–31.
40. Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics*. 2016;32(4):587–9.
41. Porcher C, Swat W, Rockwell K, Fujiwara Y, Alt FW, Orkin SH. The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell*. 1996;86(1):47–57.
42. Pevny L, Simon MC, Robertson E, Klein WH, Tsai SF, D'Agati V, et al. Erythroid differentiation in chimeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*. 1991;349(6306):257–60.
43. Sheffield NC, Thurman RE, Song L, Safi A, Stamatoyannopoulos JA, Lenhard B, et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res*. 2013;23(5):777–88.
44. Nutt SL, Metcalf D, D'Amico A, Polli M, Wu L. Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors. *J Exp Med*. 2005; 201(2):221–31.
45. Rieger MA, Schroeder T. Hematopoiesis. *Cold Spring Harb Perspect Biol*. 2012;4(12):a008250.
46. Guo Y, Fu X, Jin Y, Sun J, Liu Y, Huo B, et al. Histone demethylase LSD1-mediated repression of GATA-2 is critical for erythroid differentiation. *Drug Design Dev Ther*. 2015;9:3153–62.
47. Graf T, Enver T. Forcing cells to change lineages. *Nature*. 2009;462(7273):587–94.
48. Minegishi N, Suzuki N, Yokomizo T, Pan X, Fujimoto T, Takahashi S, et al. Expression and domain-specific function of GATA-2 during differentiation of the hematopoietic precursor cells in midgestation mouse embryos. *Blood*. 2003;102(3):896–905.
49. Luzzo BB, Luzzo CB, Bamberger EG, Feliu AS. A multipotential leukemia cell line (K-562) of human origin. *Proc Soc Exp Biol Med*. 1981;166(4):546–50.

50. Lozzio CB, Lozzio BB. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood*. 1975;45(3):321–34.
51. Savage DG, Antman KH. Imatinib mesylate—a new oral targeted therapy. *N Engl J Med*. 2002;346(9):683–93.
52. Salic A, Mitchison TJ. A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proc Natl Acad Sci U S A*. 2008;105(7):2415–20.
53. Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*. 2011;146(4):633–44.
54. Wilson NK, Kent DG, Buettner F, Shehata M, Macaulay IC, Calero-Nieto FJ, et al. Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. *Cell Stem Cell*. 2015;16(6):712–24.
55. Parnas O, Jovanovic M, Eisenhaure TM, Herbst RH, Dixit A, Ye CJ, et al. A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. *Cell*. 2015;162(3):675–86.
56. Hart T, Chandrashekar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell*. 2015;163(6):1515–26.
57. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell*. 2014;159(3):647–61.
58. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*. 2015;161(5):1187–201.
59. Macosko EZ, Basu A, Satija R, Nemes J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. 2015;161(5):1202–14.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

