LETTER

doi:10.1038/nature14590

Single-cell chromatin accessibility reveals principles of regulatory variation

Jason D. Buenrostro[1,2], Beijing Wu[1]*, Ulrike M. Litzenburger[2]*, Dave Ruff[3], Michael L. Gonzales[3], Michael P. Snyder[1], Howard Y. Chang[2] & William J. Greenleaf[1,4]

Cell-to-cell variation is a universal feature of life that affects a wide range of biological phenomena, from developmental plasticity[1,2] to tumour heterogeneity[3]. Although recent advances have improved our ability to document cellular phenotypic variation[4–8], the fundamental mechanisms that generate variability from identical DNA sequences remain elusive. Here we reveal the landscape and principles of mammalian DNA regulatory variation by developing a robust method for mapping the accessible genome of individual cells by assay for transposase-accessible chromatin using sequencing (ATAC-seq)[9] integrated into a programmable microfluidics platform. Single-cell ATAC-seq (scATAC-seq) maps from hundreds of single cells in aggregate closely resemble accessibility profiles from tens of millions of cells and provide insights into cell-to-cell variation. Accessibility variance is systematically associated with specific *trans*-factors and *cis*-elements, and we discover combinations of *trans*-factors associated with either induction or suppression of cell-to-cell variability. We further identify sets of *trans*-factors associated with cell-type-specific accessibility variance across eight cell types. Targeted perturbations of cell cycle or transcription factor signalling evoke stimulus-specific changes in this observed variability. The pattern of accessibility variation in *cis* across the genome recapitulates chromosome compartments[10] *de novo*, linking single-cell accessibility variation to three-dimensional genome organization. Single-cell analysis of DNA accessibility provides new insight into cellular variation of the 'regulome'.

Heterogeneity within cellular populations has been evident since the first microscopic observations of individual cells. Recent proliferation of powerful methods for interrogating single cells[4–8] has allowed detailed characterization of this molecular variation, and provided deep insight into characteristics underlying developmental plasticity[1,2], cancer heterogeneity[3], and drug resistance[11]. In parallel, genome-wide mapping of regulatory elements in large ensembles of cells have unveiled substantial variation in chromatin structure across cell types, particularly at distal regulatory regions[12]. In particular, methods for probing genome-wide DNA accessibility have proven extremely effective in identifying regulatory elements across a variety of cell types[13] and quantifying changes that lead to both activation or repression of gene expression. Given this broad diversity of activity within regulatory elements when comparing phenotypically distinct cell populations, it is reasonable to hypothesize that heterogeneity at the single-cell level extends to accessibility variability within cell types at regulatory elements. However, the lack of methods to probe DNA accessibility within individual cells has prevented quantitative dissection of this hypothesized regulatory variation.

We have developed a single-cell assay for transposase-accessible chromatin (scATAC-seq). ATAC-seq is an ensemble measure of open chromatin that uses the prokaryotic Tn5 transposase[14,15] to tag regulatory regions by inserting sequencing adapters into accessible regions of the genome. In scATAC-seq, individual cells are captured and assayed using a programmable microfluidics platform (Fluidigm) with methods optimized for this task (Fig. 1a, Extended Data Fig. 1 and Supplementary Discussion). After transposition and PCR on the integrated fluidics circuit (IFC), libraries were collected and PCR amplified with cell-identifying barcoded primers. Single-cell libraries were then pooled and sequenced on a high-throughput sequencing instrument. Using single-cell ATAC-seq, we generated DNA accessibility maps from 254 individual GM12878 lymphoblastoid cells. Aggregate profiles of scATAC-seq data closely reproduce ensemble measures of accessibility profiled by DNase-seq and ATAC-seq generated from $\sim 10^7$ or $\sim 10^4$ cells, respectively (Fig. 1b, c and Extended Data Fig. 2a). Data from single cells recapitulate several characteristics of bulk ATAC-seq data, including fragment-size periodicity corresponding to integer multiples of nucleosomes, and a strong enrichment of fragments within regions of accessible chromatin (Extended Data Fig. 2b, c). Microfluidic chambers generating low library diversity or poor measures of accessibility, which correlate with empty chambers or dead cells, were excluded from further analysis (Fig. 1d and Extended Data Fig. 2d–l). Chambers passing filter yielded an average of 7.3×10^4 fragments mapping to the nuclear genome. We further validated the approach by measuring chromatin accessibility from a total of 1,632 IFC chambers representing three tier 1 ENCODE cell lines[16] (H1 human embryonic stem cells (ES cells), K562 chronic myelogenous leukaemia and GM12878 lymphoblastoid cells), as well as from V6.5 mouse ES cells, EML[1] cells (mouse haematopoietic progenitors), TF-1 cells (human erythroblast), HL-60 cells (human promyeloblasts) and BJ fibroblasts (human foreskin fibroblasts).

Because regulatory elements are generally present at two copies in a diploid genome, we observe a near digital (0 or 1) measurement of accessibility at individual elements within individual cells (Extended Data Fig. 3a). For example, within a typical single cell we estimate a total of 9.4% of promoters are represented in a typical scATAC-seq library (Extended Data Fig. 3b–d). The sparse nature of scATAC-seq data makes analysis of cellular variation at individual regulatory elements impractical. We therefore developed an analysis infrastructure to measure regulatory variation using changes of accessibility across sets of genomic features (Fig. 2a, b). To quantify this variation we first choose a set of open chromatin peaks, identified using the aggregate accessibility track, which share a common characteristic (such as transcription factor binding motif, ChIP-seq peaks or cell cycle replication timing domains). We then calculate the observed fragments in these regions minus the expected fragments, downsampled from the aggregate profile, within individual cells. To correct for bias, we divide this by the root mean square of fragments expected from a background signal constructed to estimate technical and sampling error within single-cell data sets (Methods and Extended Data Fig. 4). Hereafter, we refer to this metric as 'deviation'. Finally, for any set of features, we also calculate an overall 'variability' score across

[1]Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. [2]Program in Epithelial Biology and the Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA. [3]Fluidigm Corporation, South San Francisco, California 94080, USA. [4]Department of Applied Physics, Stanford University, Stanford, California 94025, USA.
*These authors contributed equally to this work.

©2015 Macmillan Publishers Limited. All rights reserved

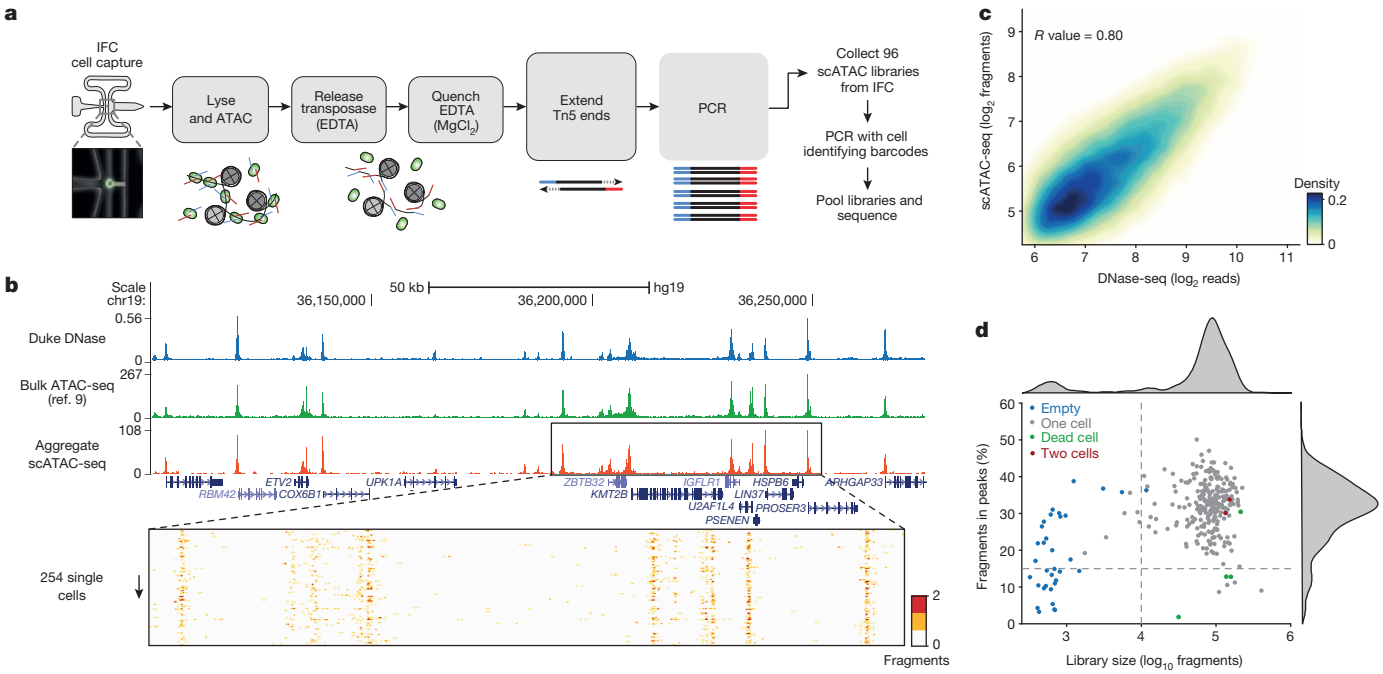


Figure 1 | Single-cell ATAC-seq provides an accurate measure of chromatin accessibility genome-wide. **a**, Workflow for measuring single epigenomes using scATAC-seq on a microfluidic device (Fluidigm). **b**, Aggregate single-cell accessibility profiles closely recapitulate profiles of DNase-seq and ATAC-seq in GM12878 cells. **c**, Genome-wide accessibility patterns observed by scATAC-seq are correlated with DNase-seq data ($R = 0.80$). **d**, Library size versus percentage of fragments in open chromatin peaks (filtered as described in Methods) within K562 cells ($n = 288$). Dotted lines (15% and 10,000) represent cutoffs used for downstream analysis.

all cells (Fig. 2b), a metric of excess variance over the background signal.

We first focused our analysis on K562 myeloid leukaemia cells, a cell type with extensive epigenomic data sets[17,18]. To comprehensively characterize variability associated with *trans*-factors within individual K562 cells, we computed variability across all available ENCODE ChIP-seq, transcription factor motifs and regions that differed in replication timing (as determined from Repli-Seq data sets[19]) (Fig. 2c, d). We found measures of cell-to-cell variability were highly reproducible across biological replicates (Extended Data Fig. 5). 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. In addition, we discover a set of *trans*-factors associated with high variability. These factors include sequence-specific transcription factors, such as GATA1/2, JUN and STAT2, and chromatin effectors, such as BRG1 (also known as SMARCA4) and P300 (also known as EP300). Immunostaining followed by microscopy or flow cytometry (Fig. 2e and Extended Data Fig. 6a–d) confirmed heterogeneous expression of GATA1 and GATA2. Principal component (PC) analysis of single-cell deviations across all *trans*-factors show seven significant PCs, with PC 5 describing changes in DNA abundance throughout the cell cycle. This analysis suggests that high-variance *trans*-factors are variable independent of the cell cycle (Fig. 2f and Extended Data Fig. 6e–g). The remaining PCs show contributions from several transcription factors, suggesting that variance across sets of *trans*-factors represent distinct regulatory states in individual cells.

We hypothesized that variation associated with different *trans*-factors can synergize, either through cooperative or competitive binding, to induce or suppress site-to-site variability in chromatin accessibility. For example, the most variant factors in K562 cells, GATA1 and GATA2, display expression heterogeneity and also bind an identical consensus sequence GATA, suggesting these factors may compete for access to DNA sequences. In support of this hypothesis, we find regulatory elements with both GATA1 and GATA2 ChIP-seq signals show increased variability in accessibility, whereas sites with only GATA1 or GATA2 show substantially less variability (Fig. 2g and Extended Data Fig. 6h). In contrast, we find no substantial change in variability of GATA1 binding sites that co-occur with JUN or CEBPB (Extended Data Fig. 6i). We also find peaks unique to GATA1 binding are significantly more accessible than peaks unique to GATA2 (Extended Data Fig. 6k–l) supporting the hypothesis that GATA1, an activator of accessibility, competes with GATA2 to induce single-cell variability. Extending this analysis to all transcription factor ChIP-seq data sets revealed a *trans*-factor synergy landscape for accessibility variation (Fig. 2g and Extended Data Fig. 6j). For example, chromatin accessibility variance associated with GATA2 binding is significantly enhanced when the same region could also be bound by GATA1, TAL1 or P300. In contrast, CTCF, SUZ12, and ZNF143 appear to act as general suppressors of accessibility variance, unless associated with proximal binding of ZNF143 or SMC3, the latter a cohesin subunit involved in chromosome looping[18,20]. Thus, single cell accessibility profiles nominate distinct *trans*-factors that, in combination, induce or suppress cell-to-cell regulatory variation.

To validate our ability to detect changes in accessibility variance, we used chemical inhibitors to modulate potential sources of cell-cell variability. Inhibition of cyclin-dependent kinases 4 and 6 (CDK4/6), essential components of the cell cycle, caused a marked reduction of variability within peaks associated with DNA replication timing domains (Repli-Seq) (Fig. 3a). The addition of inhibitors of JUN or BCR–ABL kinases (JNKi and imatinib, respectively) increased G1/S-associated variability suggesting an increase in the subpopulation of G1/S cells, which was validated with flow cytometry (Extended Data Fig. 7). JUN variability was significantly gained in response to JNKi but not imatinib treatment, suggesting that high-variance *trans*-factors can also be specifically and pharmacologically modulated. Tumour necrosis factor (TNF) treatment of GM12878 cells specifically modulated accessibility variability at NF-κB sites (Fig. 3b), consistent with the known stochastic and oscillatory property of nuclear shuttling in

©2015 Macmillan Publishers Limited. All rights reserved

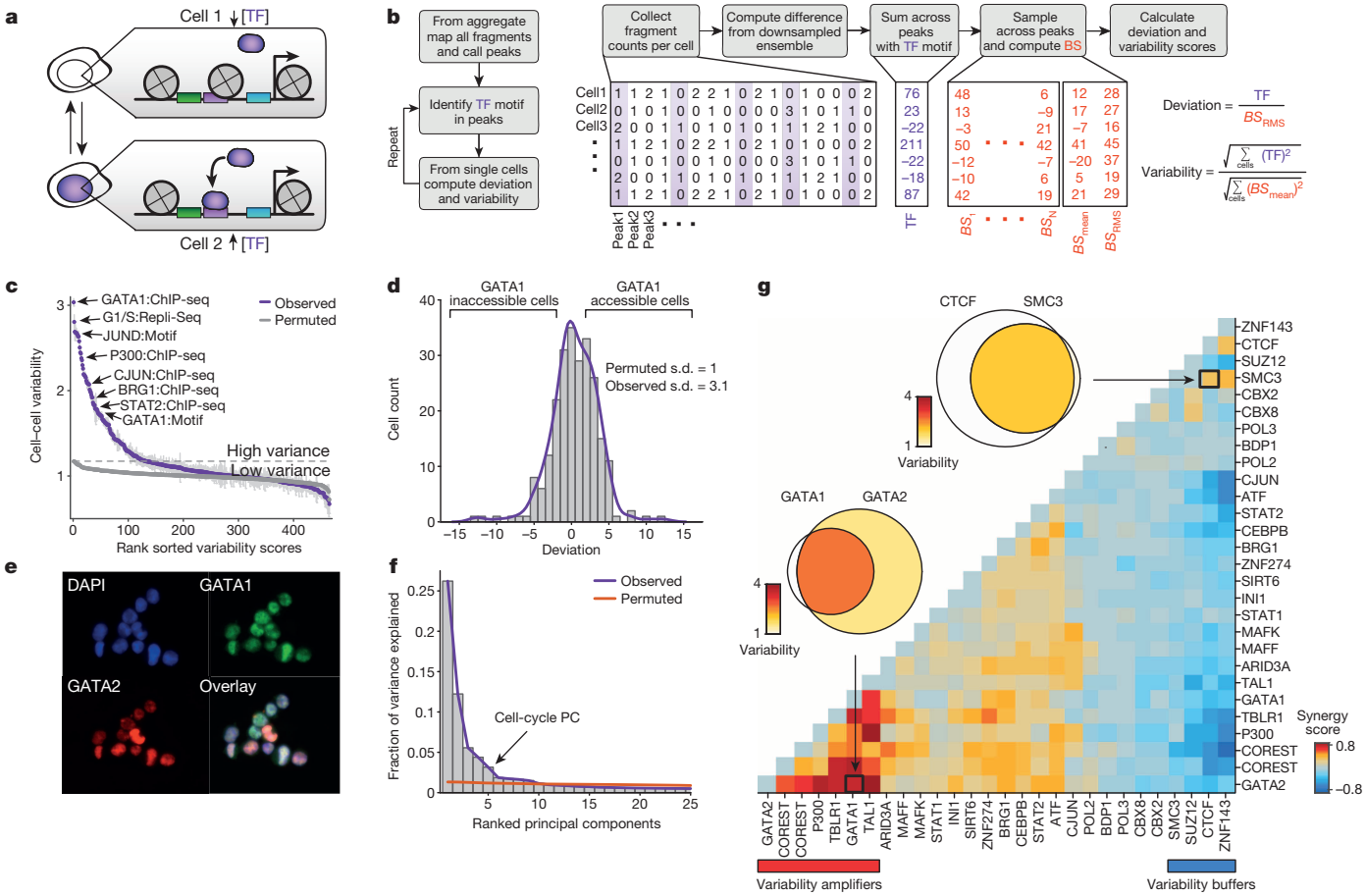


Figure 2 | *Trans*-factors are associated with single-cell epigenomic variability. **a,** Schematic showing two cellular states (transcription factor high and transcription factor low) leading to differential chromatin accessibility. TF, transcription factor. **b**, Analysis infrastructure, which uses a calculated background signal (BS; see Supplementary Methods, section 3.2) to calculate transcription factor deviations and variability from scATAC-seq data. The transcription factor value is calculated by subtracting the number of expected fragments from the observed fragments per cell (see Supplementary Methods, section 3.1). **c**, Observed cell-to-cell variability within sets of genomic features associated with ChIP-seq peaks, transcription factor motifs, and replication timing (error estimates shown in grey, see Methods for details). Variability measured from permuted background (see Methods) is shown in grey dots. **d**, Distribution of normalized deviations from expected accessibility signal for GATA1 sites in individual cells, histogram of cells shown in grey, density profile shown in purple (see Methods). **e**, Immunostaining of GATA1 (green) and GATA2 (red) shows protein expression in K562 cells. **f**, Principal components ranked by fraction of variance explained from observed deviation data (purple) and permuted data (orange). Bar plot of observed data shown in grey. **g**, Calculated changes in associated variability of factors when present together versus independently, depicting a context-specific *trans*-factor variability landscape (see Methods). Venn-diagrams show variability associated with GATA1 and/or GATA2 and CTCF and/or SMC3 (co-)occurring ChIP-seq

this system[21]. Together, these results show that variability can be experimentally modulated and further demonstrates that variability is not solely dependent on the cell cycle.

We observe that *trans*-factors associated with high variability are generally cell-type specific. Hierarchical bi-clustering of single-cell deviations generated from three cell lines reveals cell-type specific sets of transcription factor motifs associated with high variability (Fig. 3c). This analysis also shows cells from different biological replicates cluster with their cell type of origin (with a single exception), suggesting scATAC-seq can also be used to deconvolve heterogeneous cellular mixtures. Systematic analysis of all assayed cell types identified high-variance *trans*-factor motifs that are generally unique to specific cell types (Fig. 3d and Extended Data Fig. 8a). For example, regions associated with GATA transcription factors are most variant in K562 cells, whereas regions associated with master pluripotency transcription factors Nanog and Sox2 are most variant in mouse ES cells, consistent with previous observations of expression variation of these factors[22,23]. We also find high variability of GATA1 and PU.1 (SPI1) binding accessibility in EML cells, a cell type previously shown to have >200-fold GATA1 and >15-fold PU.1 expression differences within clonal cellular subpopulations[1]. The complete set of identified high-variance *trans*-factors contains a number of transcription factors previously reported to dynamically localize into the nucleus, including NF-κB, JUN and ETS/ERG[21,24,25], suggesting that temporal fluctuations in transcription factor concentration may be driving observed chromatin accessibility heterogeneity. Finally, we find BJ fibroblasts and HL-60 cells exhibit less variance among this set of annotated *trans*-factor motifs, suggesting differences in the global levels of *trans*-factor variability across cell lines. Specific chromatin states and histone modifications[26] are also sometimes associated with accessibility variation in single cells (Extended Data Fig. 8b, c). Overall these findings suggest that *trans*-factors promote cell-type specific chromatin accessibility variation genome-wide.

Patterns of variation in accessibility along the linear genome in individual cells reveal an unexpected connection to higher-order chromosome folding. We calculated single-cell deviations within sliding windows across the genome, each encompassing a fixed number of peaks ($n = 25$) (Fig. 4a). We determined which windows co-varied within individual cells by calculating the co-correlation of each window across all others within the same chromosome within individual cells (Extended Data Fig. 9a, b). We further enhanced this co-correlation matrix using a secondary correlation analysis using methods similar to those used in chromosome conformation studies[10] (Methods). The resulting matrix, which identifies pairs of positions in the genome

©2015 Macmillan Publishers Limited. All rights reserved

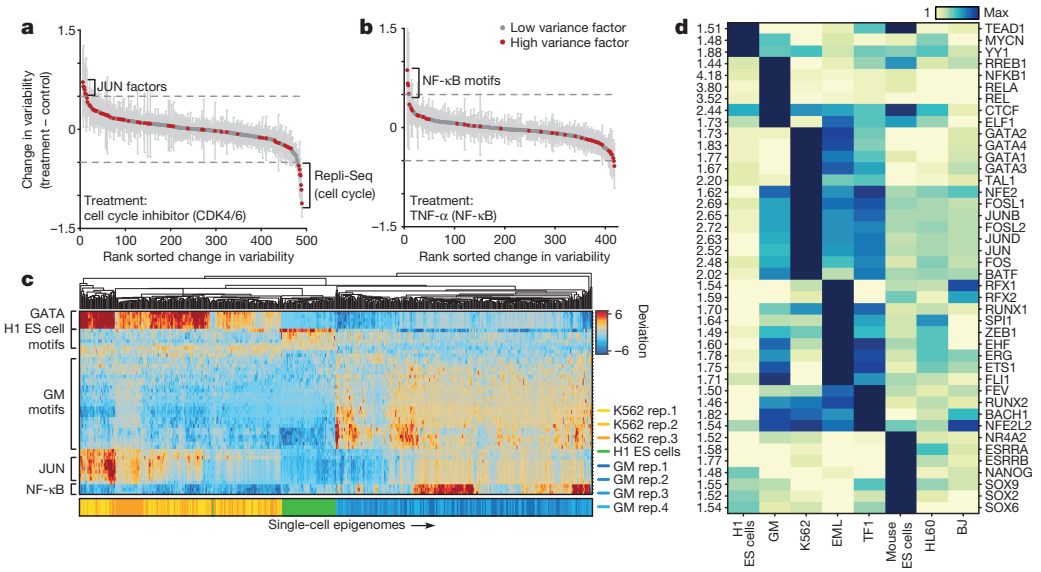


Figure 3 | Cell-type-specific epigenomic variability. **a**, **b**, Change of cellular variability due to chemical perturbations using CDK4/6 cell-cycle inhibitor (K562) (**a**) or TNF-α stimulation (GM12878) (**b**). Error bars (shown in grey) represent one standard deviation of bootstrapped cells across the two conditions. **c**, Heat map of deviations from expected accessibility signal across *trans*-factors (rows) and of single cells (columns) from 3 cell types. Bottom colour map represents assignment classification from hierarchical clustering. **d**, Variability associated with *trans*-factor motifs across 7 cell types. Each row is normalized to the maximum variability for that motif across cell types (left).

where accessibility co-varies within individual cells, yields megabase-scale correlation domains highly concordant with previously observed chromosome compartments[27] (Fig. 4b–d and Extended Data Fig. 9c–i) ($R = 0.61$ for chromosome 1). These data provide independent biological validation of large-scale compartmentalization of higher-order chromatin structure[10,27]. Moreover, these results suggest that higher-order chromatin interactions may drive regulatory variability in *cis* (elements that are proximal together tend to be accessible together). Thus, ensemble chromosome conformation data may arise in part from the statistical properties of single cell variation in co-regulated accessibility, a hypothesis also supported by single-cell fluorescent *in situ* hybridization (FISH) measurements of interactions between DNA loci[28].

Using scATAC-seq, we dissected single-cell epigenomic heterogeneity and linked *cis*- and *trans*-effectors to variability in accessibility profiles within individual epigenomes. We identify *trans*-factors associated with increased accessibility variance, which we call high-variance *trans*-factors. Additionally, other *trans*-factors such as CTCF appear to buffer variability, perhaps by providing a stable anchor of chromatin accessibility or insulator function that dampens potential fluctuations. Conversely, co-occurance with other factors such as P300 appears to amplify variability, perhaps due to synergistic interactions.

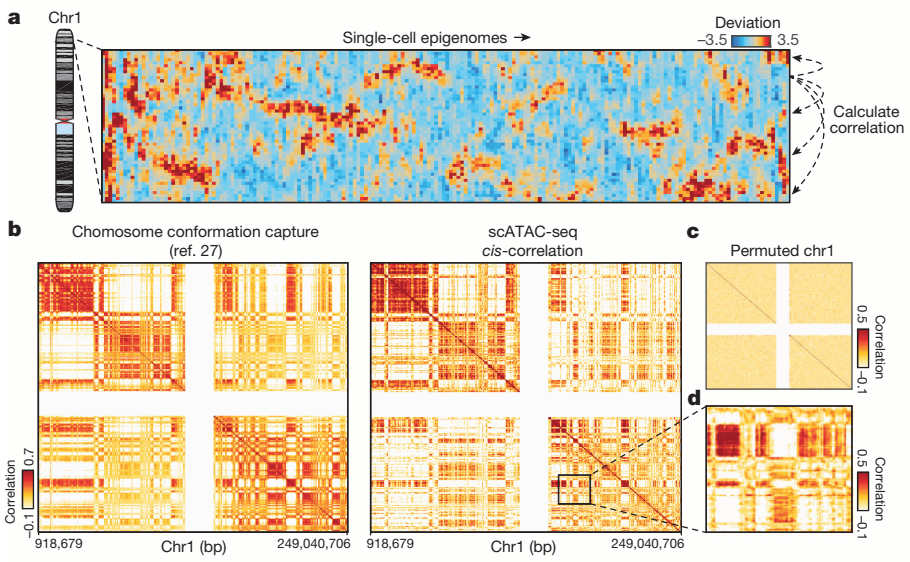


Figure 4 | Structured *cis*-variability across single epigenomes. **a**, Per-cell deviations of expected fragments across a region within chromosome 1 (see Methods). For display, only large deviation cells are shown ($n = 186$ cells). **b**, Pearson correlation coefficient representing chromosome compartment signal (see Methods) of interaction frequency from a chromatin conformation capture assay (left, analysis carried out of data from ref. 27) or doubly correlated normalized deviations of scATAC-seq (right) from chromosome 1 (see Methods). Data in white represents masked regions due to highly repetitive regions. **c**, Permuted *cis*-correlation map for chromosome 1 (analysed identically to **b**). **d**, Box highlights a representative region depicting long-range covariability.

©2015 Macmillan Publishers Limited. All rights reserved

Lineage-specific master regulators are associated with cell-type specific single-cell epigenomic variability across several cell types, suggesting that control of single-cell variance is a fundamental characteristic of different biological states. Finally, variation of chromatin accessibility in *cis* is highly correlated with previously reported chromosome compartments, opening the intriguing possibility that this component of epigenomic noise has its roots in higher-order chromatin organization. Together these data provide a new hypothesis of regulatory mechanisms that give rise to single-cell heterogeneity.

We envision that future studies will enhance the utility of scATAC-seq by further improving the recovery of DNA fragments, increasing throughput, and refining methods of data analysis (Supplementary Discussion). Improvements to throughput and new statistical tools will enable single-cells to be partitioned by cell-state and analysed in aggregate to find the individual peaks that drive variability (Extended Data Fig. 10). In addition, we anticipate scATAC-seq may be paired with existing approaches in microscopy and single-cell RNA-seq to provide opportunities for systems analysis of individual cells. Such an approach will link regulatory variation to details of phenotypic variation, providing new insights into the molecular underpinnings of cellular heterogeneity. We believe scATAC-seq will also enable the interrogation of the epigenomic landscape of small or rare biological samples allowing for detailed, and potentially *de novo*, reconstruction of cellular differentiation or disease at the fundamental unit of investigation—the single cell.

Received 12 January; accepted 26 May 2015.
Published online 17 June 2015.

1. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453,** 544–547 (2008).
2. Imayoshi, I. *et al.* Oscillatory control of factors determining multipotency and fate in mouse neural progenitors. *Science* **342,** 1203–1208 (2013).
3. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344,** 1396–1401 (2014).
4. Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332,** 687–696 (2011).
5. Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463,** 913–918 (2010).
6. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343,** 776–779 (2014).
7. Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature Methods* **11,** 817–820 (2014).
8. Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338,** 1622–1626 (2012).
9. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10,** 1213–1218 (2013).
10. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
11. Michor, F. *et al.* Dynamics of chronic myeloid leukaemia. *Nature* **435,** 1267–1270 (2005).
12. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
13. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 (2012).
14. Goryshin, I. Y. & Reznikoff, W. S. Tn5 *in vitro* transposition. *J. Biol. Chem.* **273,** 7367–7374 (1998).
15. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol.* **11,** R119 (2010).
16. ENCODE Project Consortium. User's guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol.* **9,** e1001046 (2011).
17. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489,** 91–100 (2012).
18. Xie, D. *et al.* Dynamic *trans*-acting factor colocalization in human cells. *Cell* **155,** 713–724 (2013).
19. Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107,** 139–144 (2010).
20. Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132,** 422–433 (2008).
21. Tay, S. *et al.* Single-cell NF-κB dynamics reveal digital activation and analogue information processing. *Nature* **466,** 267–271 (2010).
22. Grün, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nature Methods* **11,** 637–640 (2014).
23. Singer, Z. S. *et al.* Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* **55,** 319–331 (2014).
24. Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455,** 485–490 (2008).
25. Levine, J. H., Lin, Y. & Elowitz, M. B. Functional roles of pulsing in genetic circuits. *Science* **342,** 1193–1200 (2013).
26. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).
27. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature Biotechnol.* **30,** 90–98 (2012).
28. Giorgetti, L. *et al.* Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157,** 950–963 (2014).

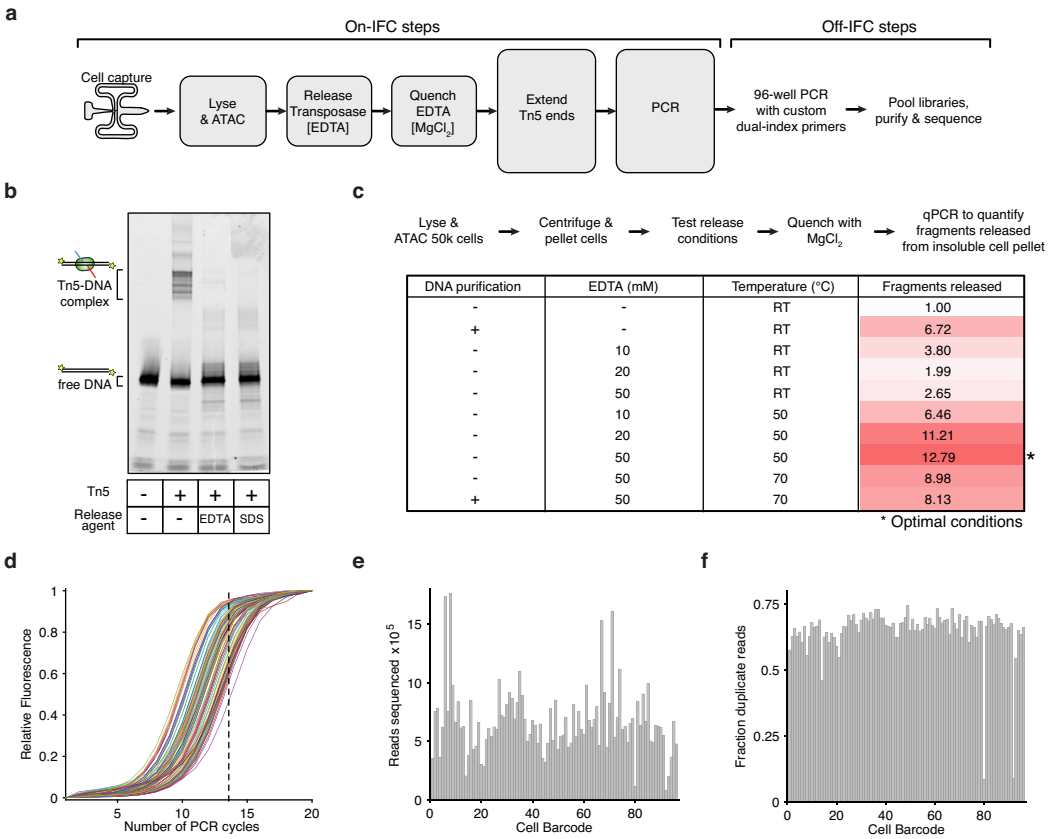
Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by National Institutes of Health (NIH) P50HG007735 (to H.Y.C. and W.J.G.), UH2 AR067676 and Lifespan Extension Foundation (H.Y.C.), U19AI057266 (to W.J.G.) and the Rita Allen Foundation (to W.J.G.) and the Baxter Foundation Faculty Scholar Grant (to W.J.G); H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute. J.D.B. acknowledges support from the National Science Foundation Graduate Research Fellowships and NIH training grant T32HG000044 for support. M.P.S. acknowledges the NIH and the National Human Genome Research Institute (NHGRI) for funding through 5U54HG00455805. We thank members of Greenleaf and Chang laboratories, as well as the Fluidigm team, including L. Xi for discussions. We acknowledge the S. Kim laboratory for assistance with FACS sorting and the C. Bustamante laboratory for help with sequencing. We also thank R. Nichols, C. Mazumdar, V. Sebastiano and V. Risca for cells.

Author Contributions J.D.B., H.Y.C. and W.J.G. conceived of the method. J.D.B., B.W., M.G. and D.R. developed the Fluidigm C1 microfluidic protocols. B.W. performed all scATAC-seq experiments with supervision from J.D.B. U.M.L. conducted the flow analysis, immunostains and drug treatments. J.D.B. developed and implemented the analysis infrastructure with input from W.J.G. All authors interpreted the data and wrote the manuscript. W.J.G. and H.Y.C. supervised all aspects of this work.

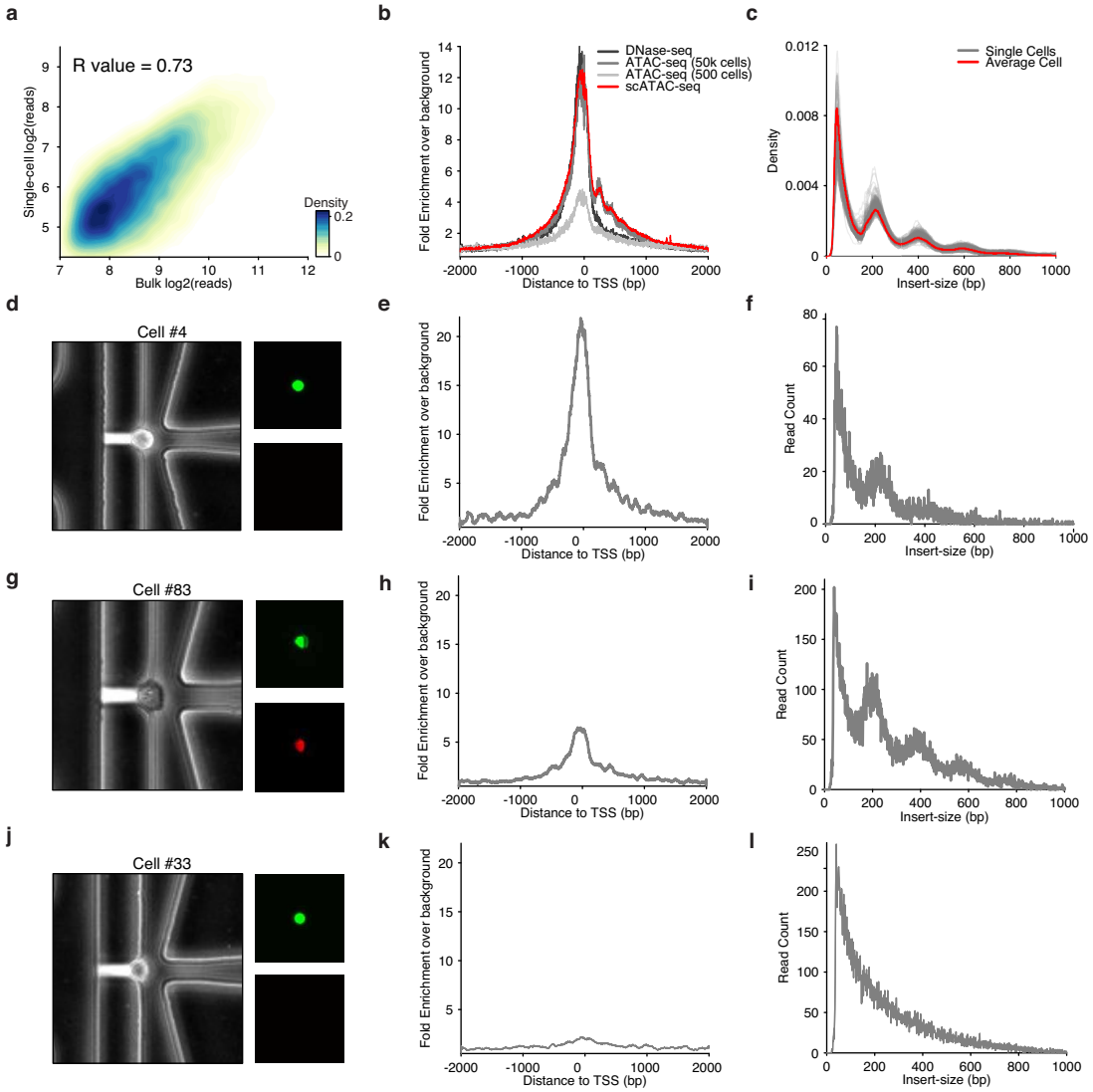
Author Information All data has been deposited in GEO under the accession number GSE65360. Fluidigm C1 scripts for performing scATAC-seq are available at https://www.fluidigm.com/c1openapp/scripthub/script/2015-06/single-cell-chromatin-accessib-1433443631246-1. Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details are available in the online version of the paper. Correspondence and requests for materials should be addressed to W.J.G. (wjg@stanford.edu) or H.Y.C. (howchang@stanford.edu).

©2015 Macmillan Publishers Limited. All rights reserved



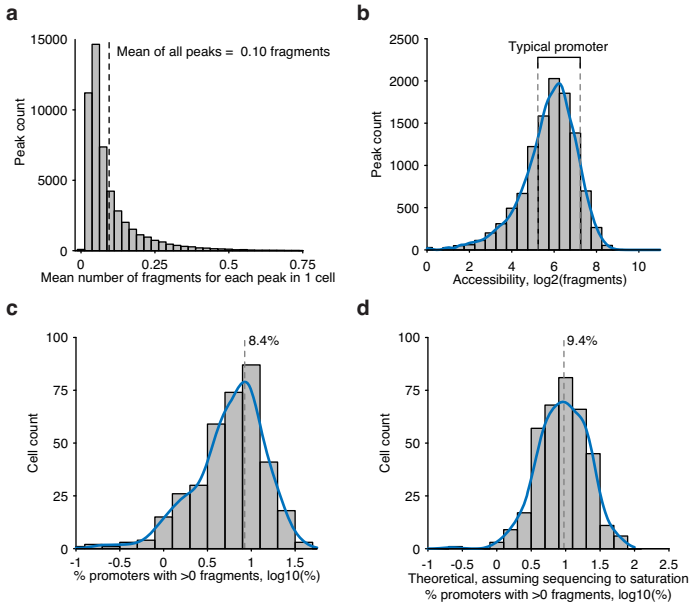
Extended Data Figure 1 | Methods development for assaying single epigenomes. **a**, scATAC-seq workflow for steps performed both on and off the integrated fluidic chip (IFC). **b**, **c**, The development of an efficient Tn5 release protocol designed to permit downstream enzymatic reactions without DNA purification. **b**, An *in vitro* electrophoretic mobility gel shift assay using a fluorescently labelled PCR product (lane 1), showing a stable Tn5-DNA complex (lane 2) dissociated with 50 mM EDTA (lane 3) or 0.1% SDS (lane 4). **c**, Workflow and associated table of conditions used to optimize release protocol, showing conditions that markedly improve fragment yield over no release conditions or purifying DNA. Fragments released represents the fold gain in library diversity, as measured by quantitative PCR (qPCR). **d**, qPCR fluorescence traces of 96 libraries generated using scATAC-seq. For all subsequent libraries we used a total of 14 PCR cycles (dotted line). **e**, **f**, A bar plot of per-cell library sequencing depth (**e**) and fraction of duplicate reads (**f**), showing each library was sequenced to varying depths to a similar fraction of duplicate reads.

©2015 Macmillan Publishers Limited. All rights reserved



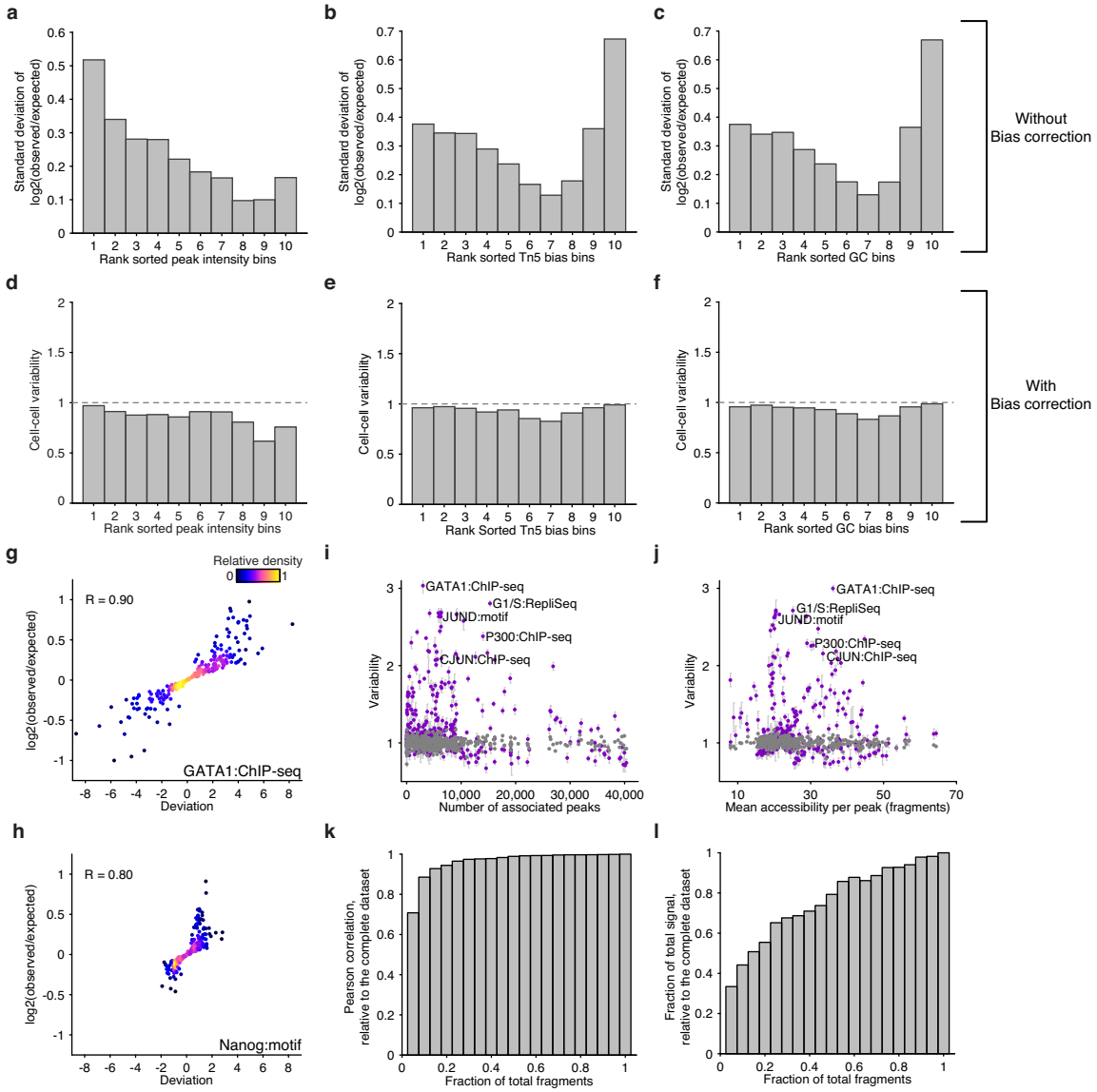
Extended Data Figure 2 | scATAC-seq data recapitulate bulk ATAC-seq characteristics. **a**, Fragments observed in open chromatin peaks identified from aggregate scATAC-seq data ($n = 384$ libraries) are highly correlated with reads observed from bulk ATAC-seq in GM12878 cells. **b**, Histogram of aggregated read starts around all transcription start sites (TSS) (in K562 cells) comparing ensemble approaches, including 500 cell ATAC-seq reported in a previous publication, to scATAC-seq shows high enrichment above background level of reads. **c**, DNA fragment size distribution of ATAC-seq fragments from single cells (grey) and the average of all single cells (red) display characteristic nucleosome-associated periodicity. **d**, Phase-contrast (left) and epifluorescence images (right) of captured cell no. 4 displaying characteristic live cell stain (Calcein) and exclusion of ethidium bromide. **e**, Histogram of read starts around TSSs for cell no. 4 shows high enrichment. **f**, DNA fragment size distribution for cell no. 4 showing nucleosomal periodicity. **g**, Images similar to **d** showing staining of cell no. 83, suggesting low viability due to ethidium bromide staining. **h**, Histogram of read starts around transcription start sites shows lower enrichment than cell no. 4. **i**, DNA fragment size distribution for cell no. 83. **j**, Images similar to **d** showing staining of cell no. 33 suggesting viability. **k**, Histogram of read starts around transcription start sites of this cell shows low levels of enrichment. **l**, DNA fragment size distribution showing no nucleosome-associated periodicity.

©2015 Macmillan Publishers Limited. All rights reserved



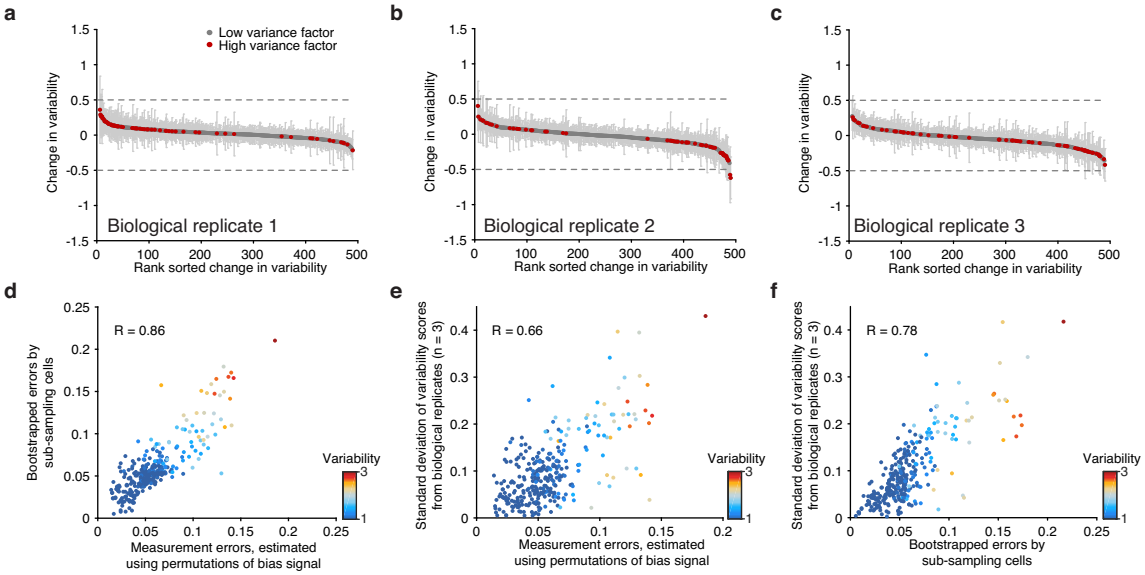
Extended Data Figure 3 | Fragment recovery metrics within scATAC-seq libraries. **a**, Accessibility across all peaks (n = 50,000) in GM12878 cells. **b**, Accessibility across all annotated promoters in GM12878 cells. Typical promoters used for subsequent analysis are boxed with dotted lines. **c**, **d**, Recovery of typical promoters shown in **a** within single cells within observed (**c**) and extrapolated (**d**) data using measures of predicted library complexity.

©2015 Macmillan Publishers Limited. All rights reserved



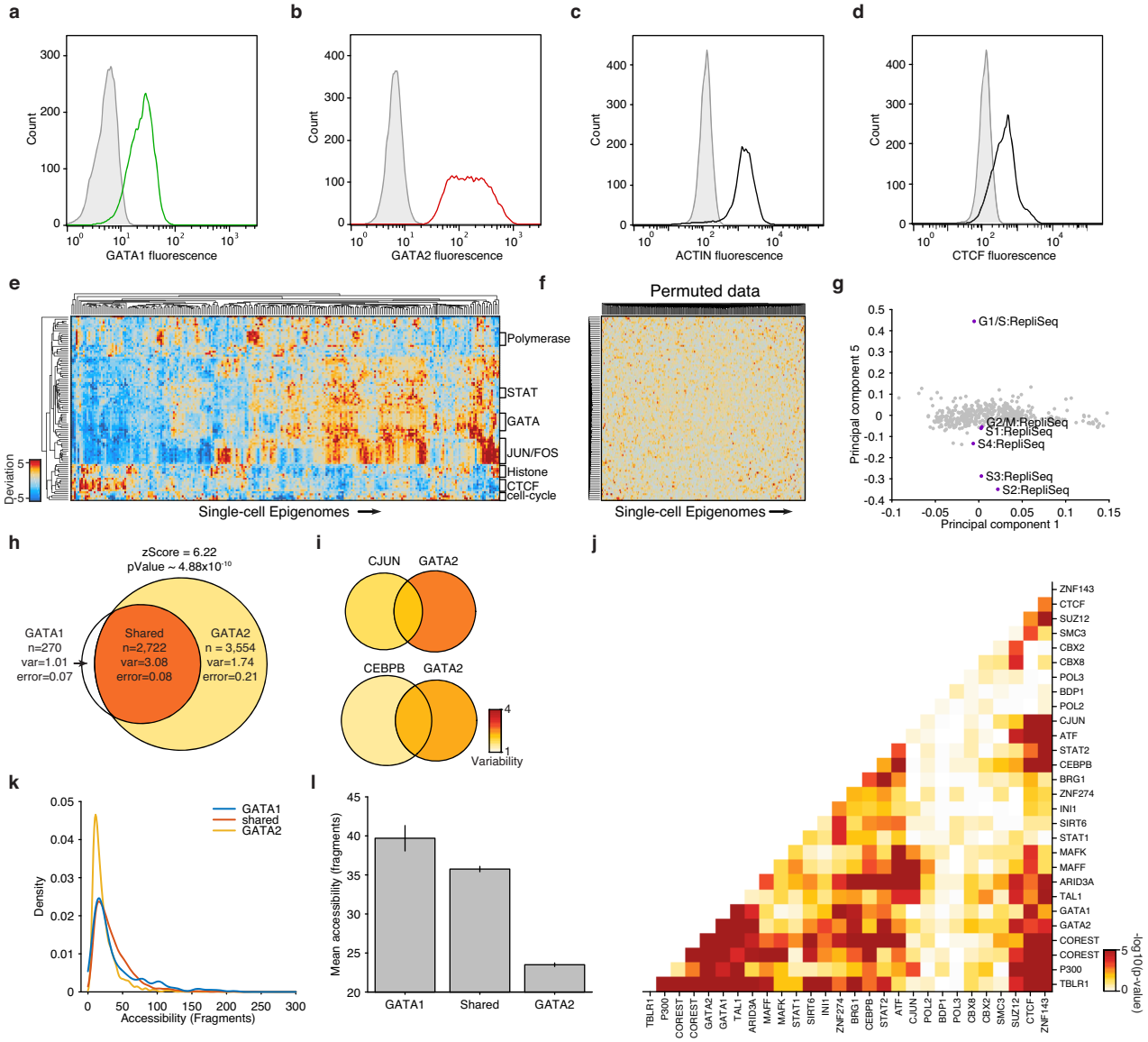
Extended Data Figure 4 | scATAC-seq data analysis pipeline and validation of bias normalization. **a**–**c**, Standard deviation of log-fold change in reads across cells within peaks binned by deciles of peak intensity (**a**), Tn5 bias (**b**) and GC bias (**c**). **d**–**f**, Variability scores (incorporating bias normalization) within the same peaks shown in **a**–**c**, peaks are binned by deciles of peak intensity (**d**), Tn5 bias (**e**) and GC bias (**f**). **g**, **h**, Log-fold change versus deviation scores across single K562 cells for GATA1 ChIP-seq target sites (**g**) and peaks (**h**) containing a Nanog motif. **i**, **j**, Variability scores for factors (purple) and the permuted background (grey) ranked by number of peak associations (**i**) and the mean accessibility per annotated peak (**j**). **k**, **l**, K562 single-cell data sets showing the effect on variability scores as a function of downsampling fragments. Fidelity after downsampling is measured with correlation (**k**) and dynamic range (**l**) relative to the complete data set.

©2015 Macmillan Publishers Limited. All rights reserved



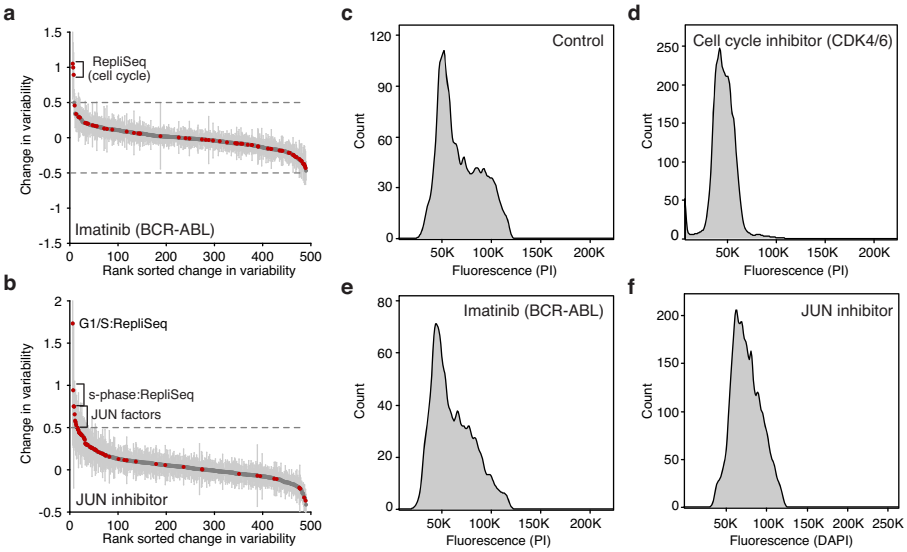
Extended Data Figure 5 | Biological replicates and measurement error analysis. **a**–**c**, Observed changes in variability comparing the merged set of replicates (K562) to each individual biological replicate. Error bars represent one standard deviation of the variability scores after bootstrapping cells from each replicate. **d**–**f,** Correlation of errors computed using three distinct approaches.

©2015 Macmillan Publishers Limited. All rights reserved



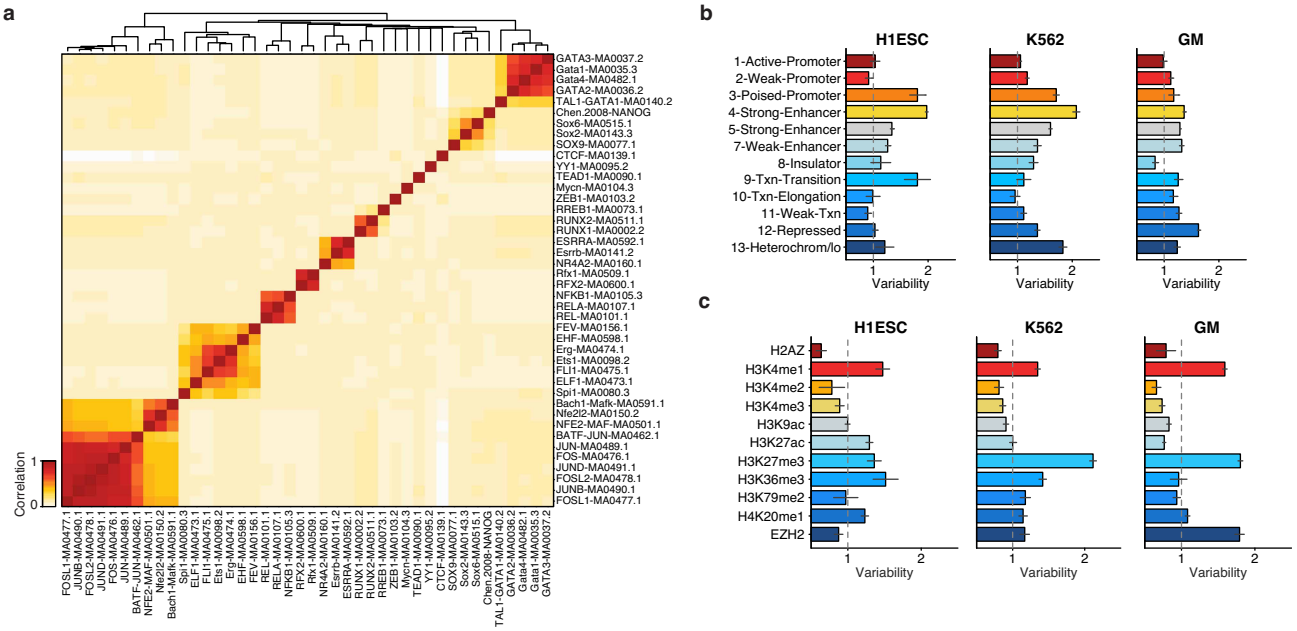
Extended Data Figure 6 | Characterization of high-variance *trans*-factors in K562 cells. **a**–**d**, Distribution of GATA1 (**a**), GATA2 (**b**), actin (**c**) and CTCF (**d**) fluorescence observed by flow cytometry. Distributions in grey depict isotype controls. **e**, Bi-clustered heat map of single-cell deviations as observed within K562 cells ($n = 239$). Labels on right identify co-clustering of related factors. **f**, Bi-clustered heat map of single-cell deviations observed from permuted data. **g**, Projection of factor loadings onto principal component 1 versus 5 from principal component (PC) analysis of heatmap shown in **e**. Factor loadings do not vary along PC5, although peaks associated with regions with different replication timings (Repli-Seq) have strong variation along this axis. **h**, **i**, Venn diagrams showing variability of GATA1 and/or GATA2 (**h**), cJUN and/or GATA2 and CEBPB and/or GATA2 (co-)occurring ChIP-seq sites (**i**). **j**, The $-\log_{10}$ (P values) of calculated changes in co-occurring ChIP-seq sites shown in Fig. 2g. **k**, Distribution of accessibility among GATA1 only, GATA2 only, and shared sites. **l**, Mean accessibility from GATA1 only, GATA2 only, and shared sites in **k**, error bars represent one standard deviation generated by bootstrapping ChIP-seq peaks.

©2015 Macmillan Publishers Limited. All rights reserved



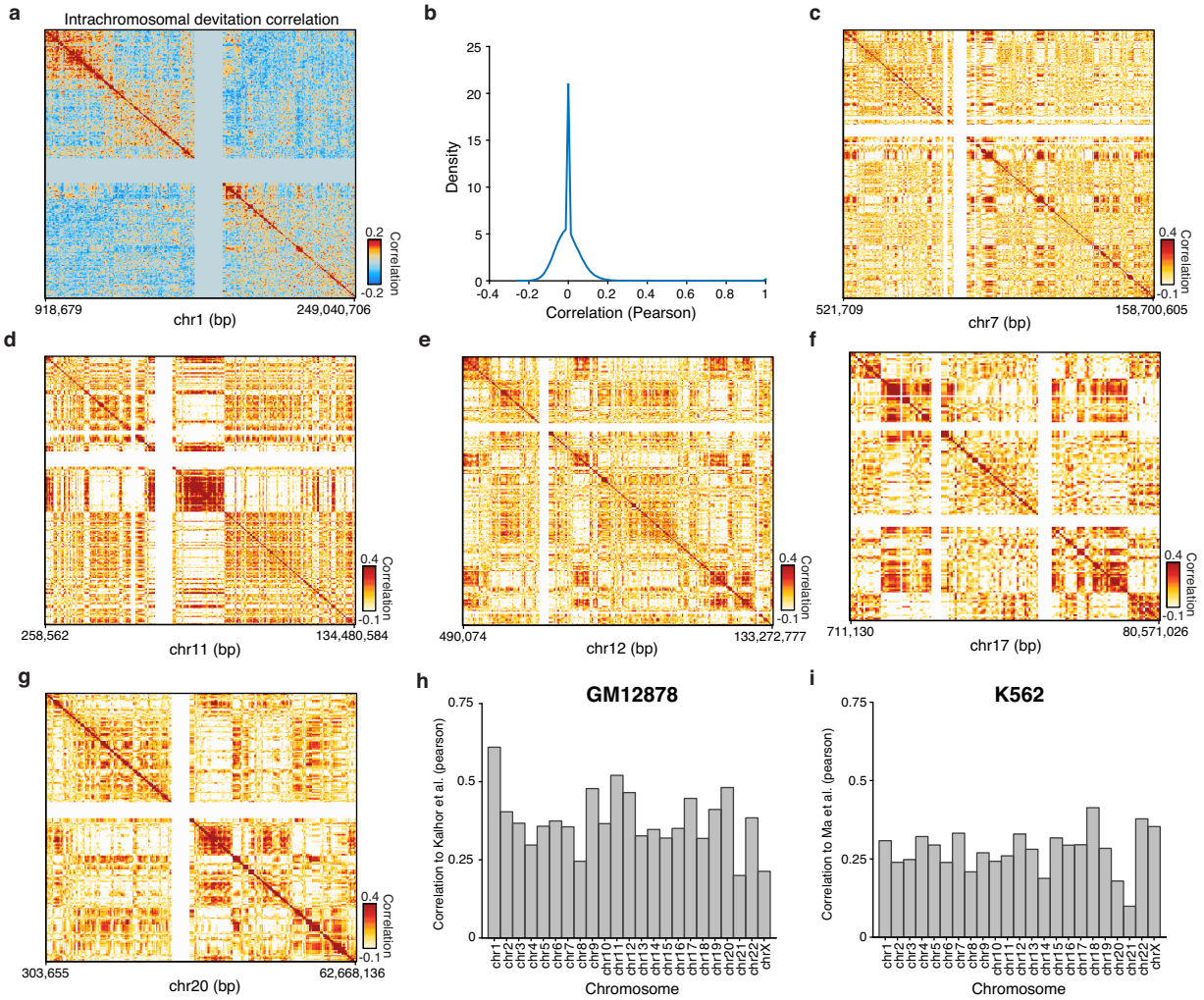
Extended Data Figure 7 | Drug treatments modulate factor variability. **a**, **b**, Change in variability of untreated K562 cells versus cells treated with imatinib (**a**) and JUN inhibitor (**b**) show increase of variability in factors associated with the cell cycle or S phase and JUN factors, respectively. **c**–**f**, Flow cytometry data depicting DNA content, using DAPI or propidium iodide, in control K562 cells (**c**) or cells showing altered cell-cycle status after treatment with cell-cycle inhibitor (**d**), imatinib (**e**) or JUN inhibitor (**f**).

©2015 Macmillan Publishers Limited. All rights reserved



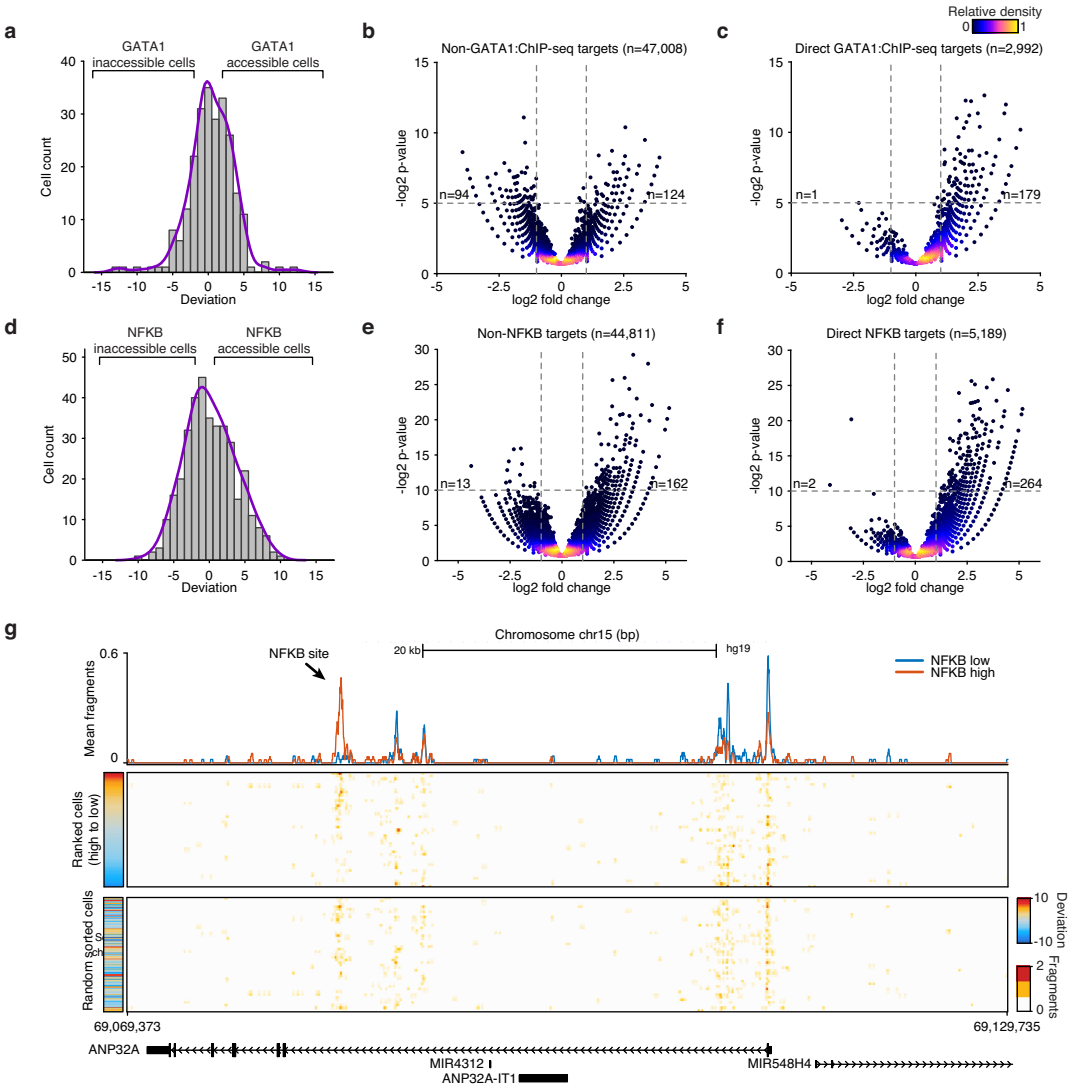
Extended Data Figure 8 | Transcription factor motif correlation and variability across chromatin state. **a**, Hierarchical bi-clustering of high-variance transcription factor motif annotations using the Pearson correlation. **b**, **c**, Variability of regions associated with chromatin states (**b**), as identified in ref. 26, and histone modifications (**c**).

©2015 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 9 | *Cis*-variability analysis within single cells. **a**, Interchromosomal chromosome 1 co-correlations of deviation scores within single cells calculated for bins of 25 peaks within GM12878 cells. **b**, Distribution, using density estimation, of correlation values shown in **a**. **c**–**g**, Analysis of *cis*-correlation (identical to Fig. 4) for representative chromosomes 7, 11, 12, 17 and 20. Correlation between scATAC-seq *cis*-correlation and chromosome conformation capture methods for each chromosome in GM12878 (**h**) and K562 (**i**) cells.

©2015 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 10 | Measurements of individual peaks within single cells. **a**, The distribution of GATA1 deviation scores for single K562 cells. **b**, **c**, Volcano plots of non-GATA1 (**b**) and GATA1 (**c**) peaks in K562 cells, *P* values were calculated using a binomial test. **d**, The distribution of NF-κB deviation scores for single GM12878 cells. **e**, **f**, Volcano plots of non-NF-κB (**e**) and NF-κB (**f**) peaks in GM12878 cells, *P* values were calculated using a binomial test. Inset numbers show the number of points in upper left or upper right quadrants of the panel. **g**, Accessibility at a genomic locus, showing (top) aggregate NF-κB low (blue) and NF-κB high (red) profiles, (middle) single GM12878 cells ranked by NF-κB deviations scores and (bottom) unranked single cells.

©2015 Macmillan Publishers Limited. All rights reserved