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]

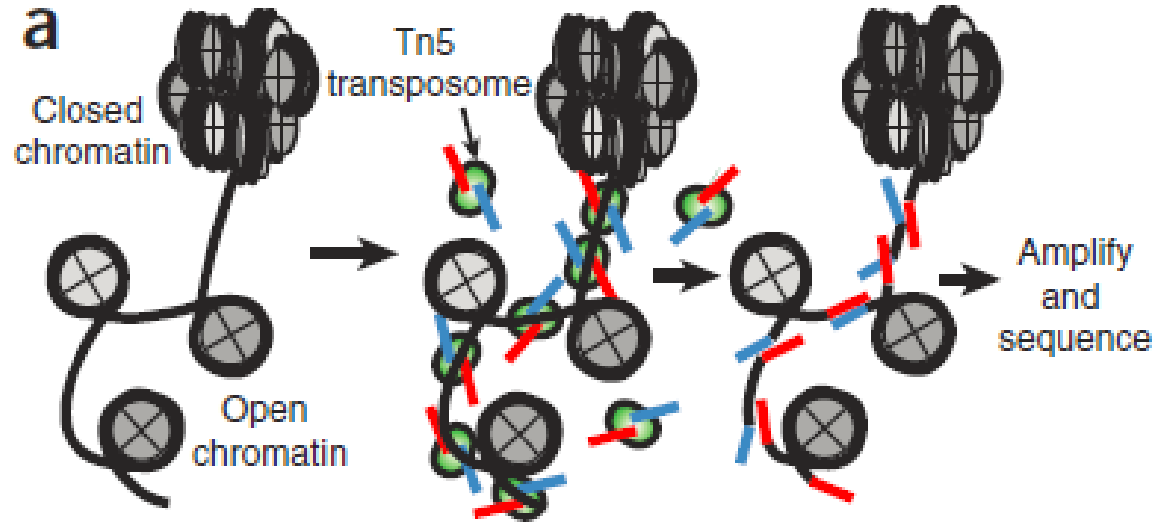
AH

JC

Introduction

- Heterogeneity within cellular populations has been evident since the first microscopic observations of individual cells.
- Single cell method provide deep insight into characteristics underlying developmental plasticity, cancer heterogeneity, and drug resistance
- Recent studies showed the broad diversity of activity within regulatory elements when comparing phenotypically distinct cell populations
- Heterogeneity at the single-cell level extends to accessibility variability within cell types at regulatory elements.

ATAC-seq: ***A***ssay for ***T***ransposase ***A***ccessible ***C***hromosome followed by sequencing (Buenrostro, 2013)

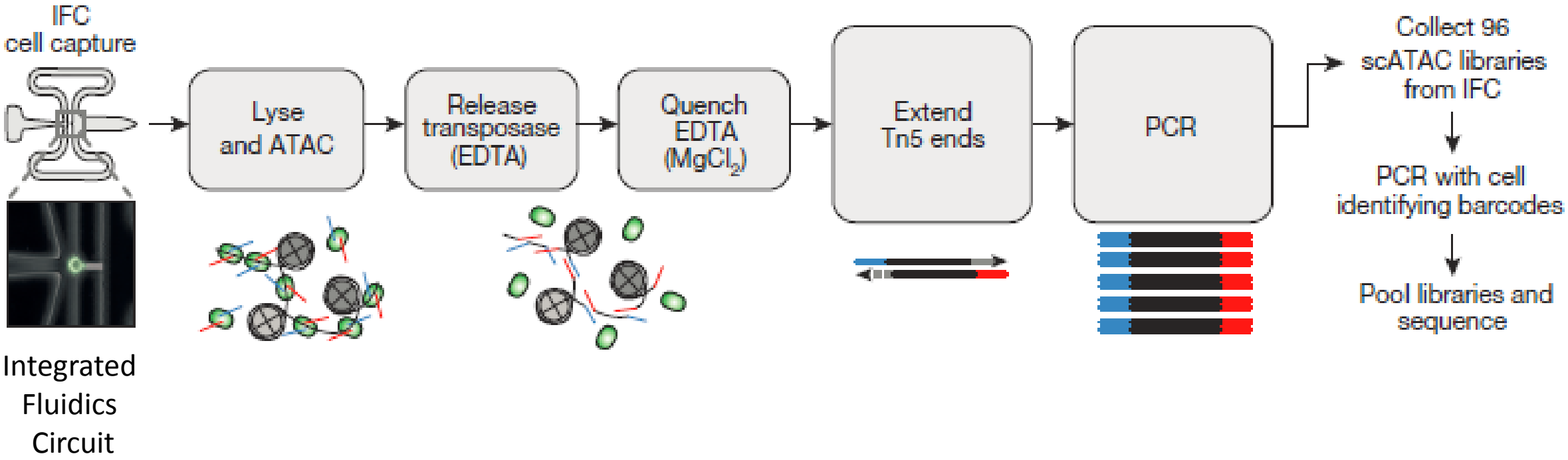


scATAC-seq: Single Cell ATAC-seq

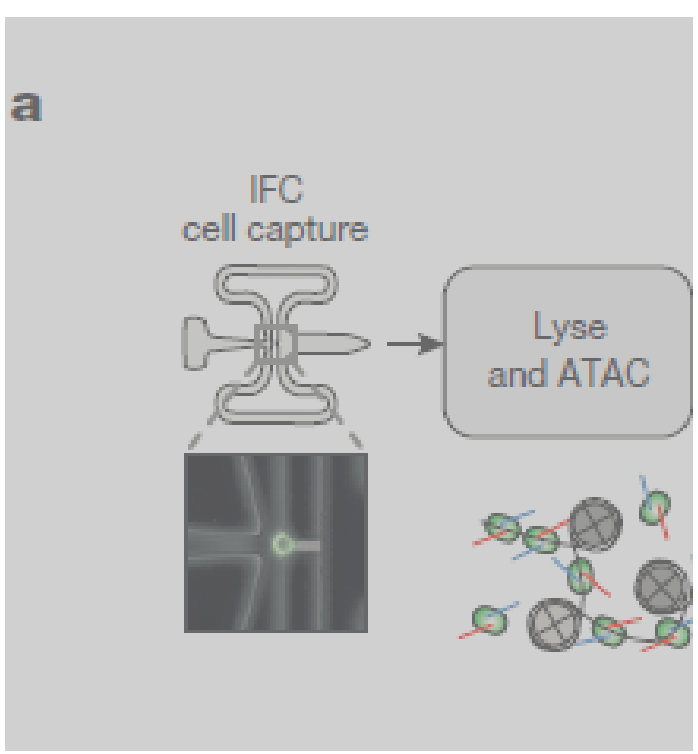
ATAC-seq: ***A***ssay for ***T***ransposase ***A***ccessible ***C***hromosome followed by sequencing

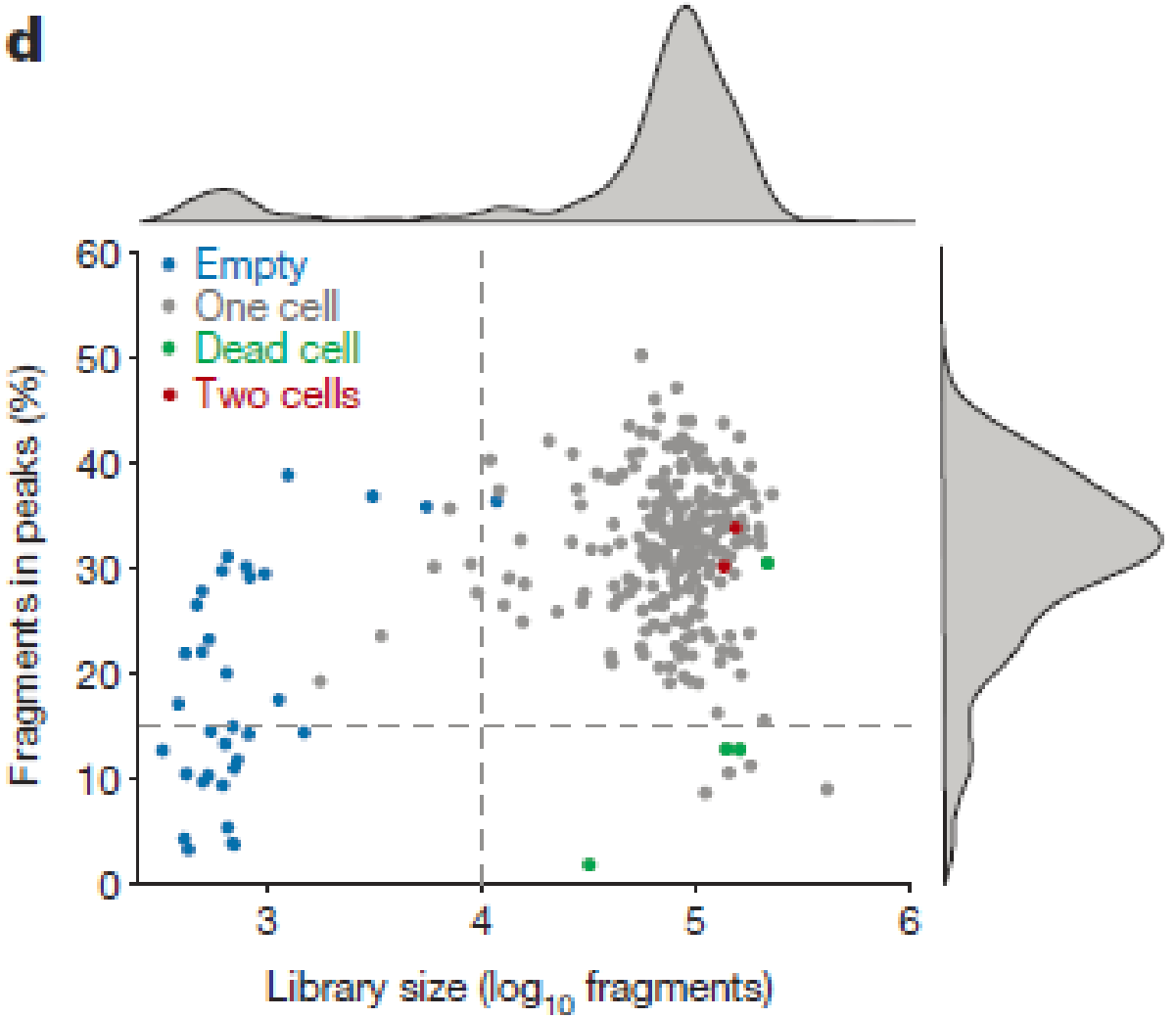
Identifies the parts of the genome which are transposase accessible. Transposase inserts sequence adapters to the accessible regions.

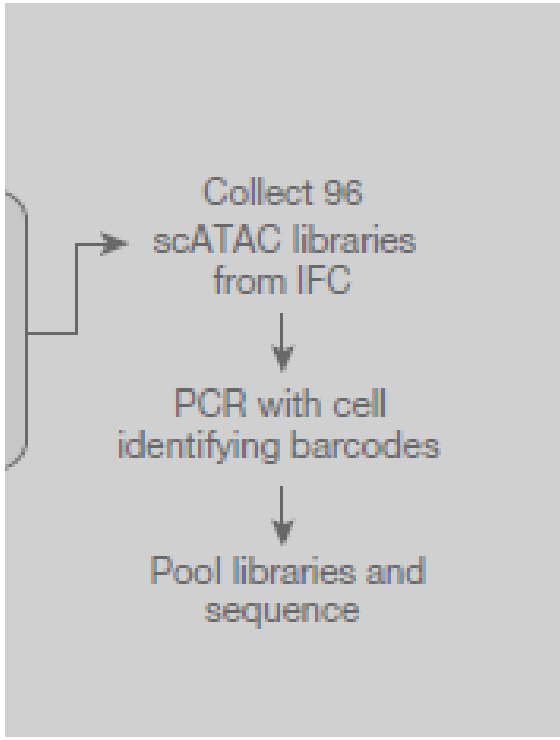
a



scATAC-seq: Single Cell ATAC-seq



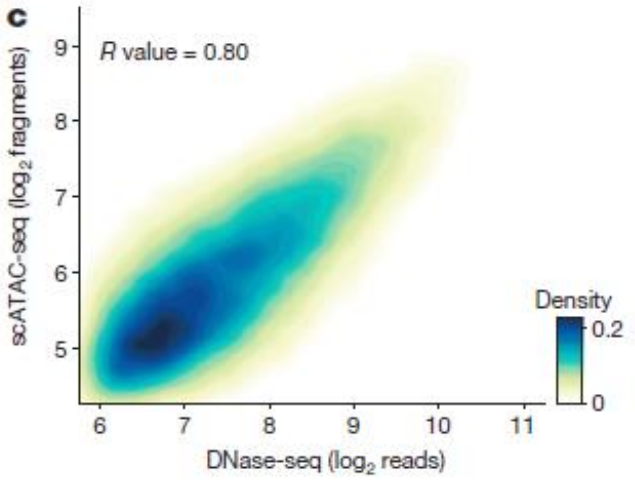


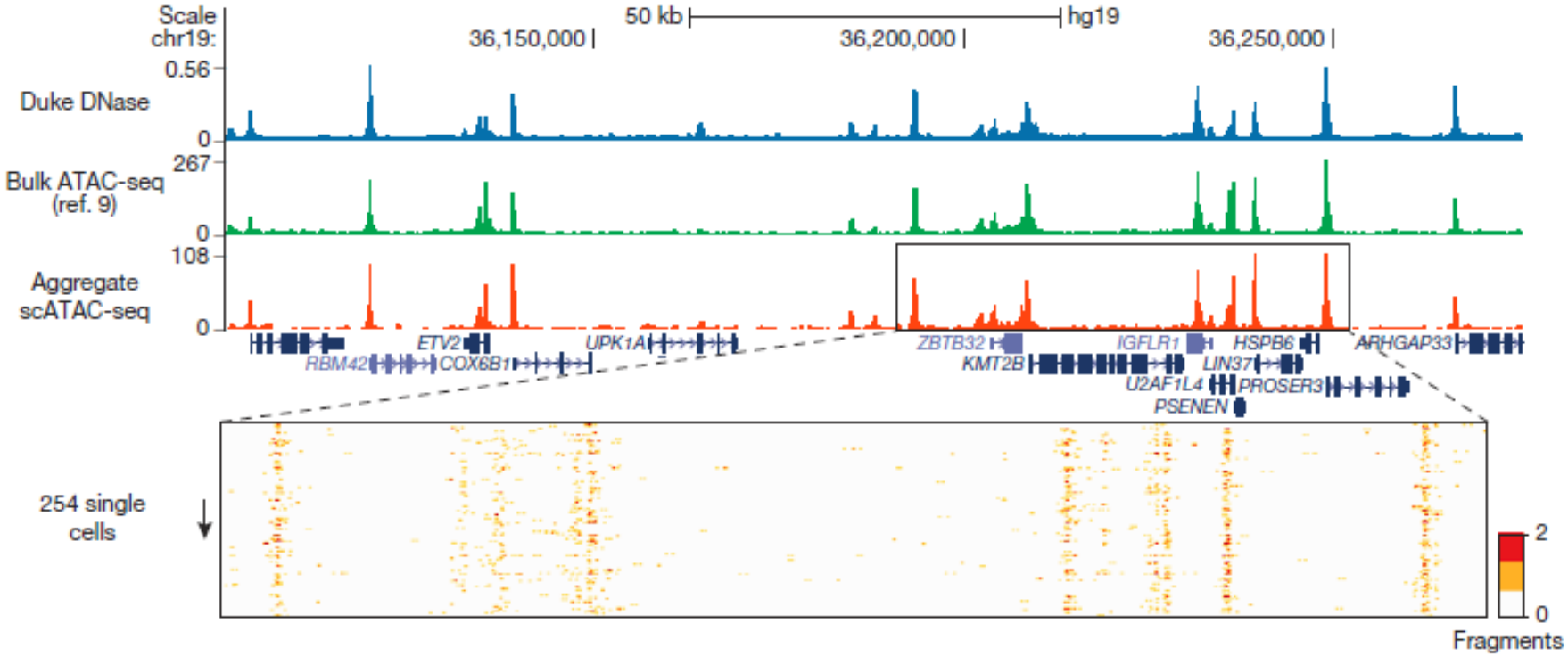


scATAC-seq: Single Cell ATAC-seq

- 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.
- Data from single cells recapitulate several characteristics of bulk ATAC-seq data

Fragment-size periodicity corresponding to integer multiples of nucleosomes, and a strong enrichment of fragments within regions of accessible chromatin.





Studied Cell Lines

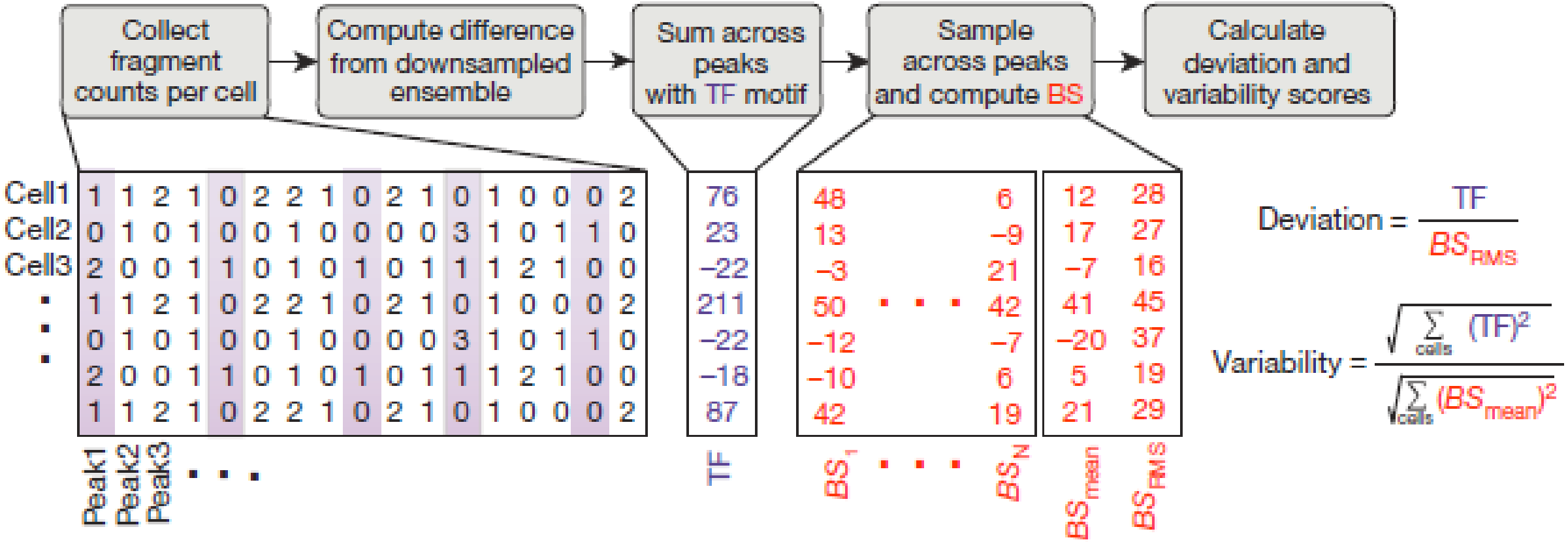
- We further validated the approach by measuring chromatin accessibility from a total of 1,632 IFC chambers representing three tier 1 ENCODE cell lines:
  - H1 human embryonic stem cells (ES cells),
  - K562 chronic myelogenous leukaemia and
  - GM12878 lymphoblastoid cells
- And V6.5 mouse ES cells,
- EML$_1$ cells (mouse haematopoietic progenitors),
- TF-1 cells (human erythroblast),
- HL-60 cells (human promyeloblasts)
- BJ fibroblasts (human foreskin fibroblasts).

(Data is available for download)

Analysis Strategy: Deviation and Variability

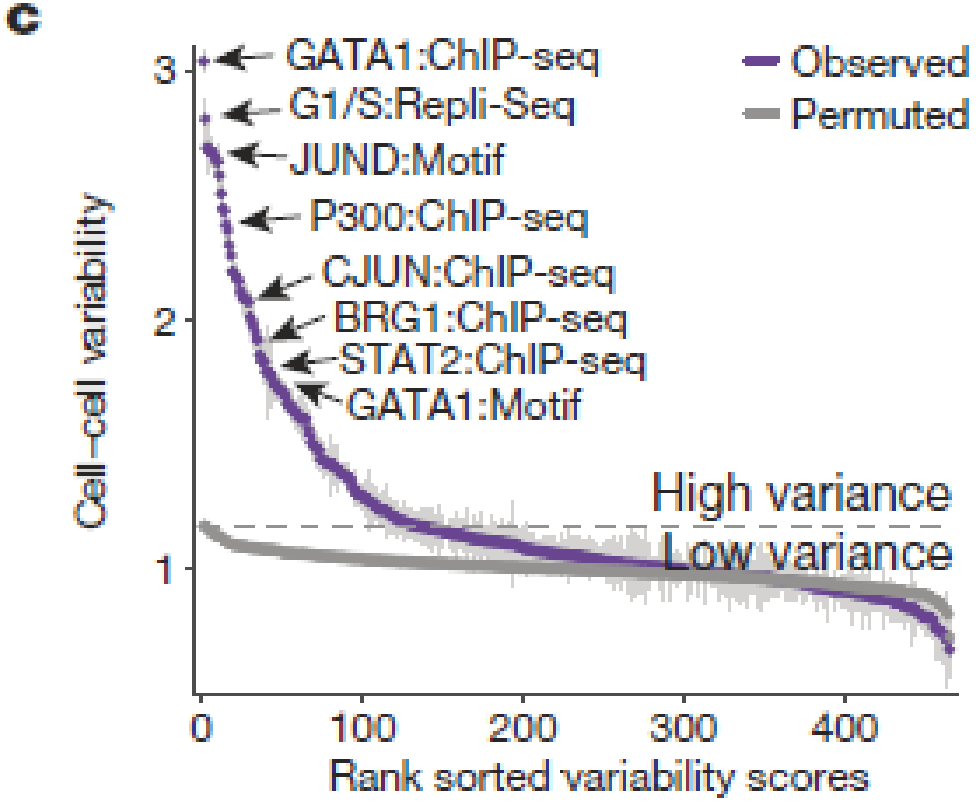
- The sparse nature of scATAC-seq data makes analysis of cellular variation at individual regulatory elements impractical.
- Authors develop an analysis infrastructure to measure regulatory variation using changes of accessibility across sets of genomic features.
  - Deviation: How high/low is the accessibility compared to a random background over all the cells?
  - Variability: How variable is the observed accessibility over the cell population?
- Choose a set of open chromatin peaks 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).
- Calculate the observed fragments in these regions minus the expected fragments, downsampled from the aggregate profile, within individual cells.
- To correct for bias, 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

Deviation and Variability



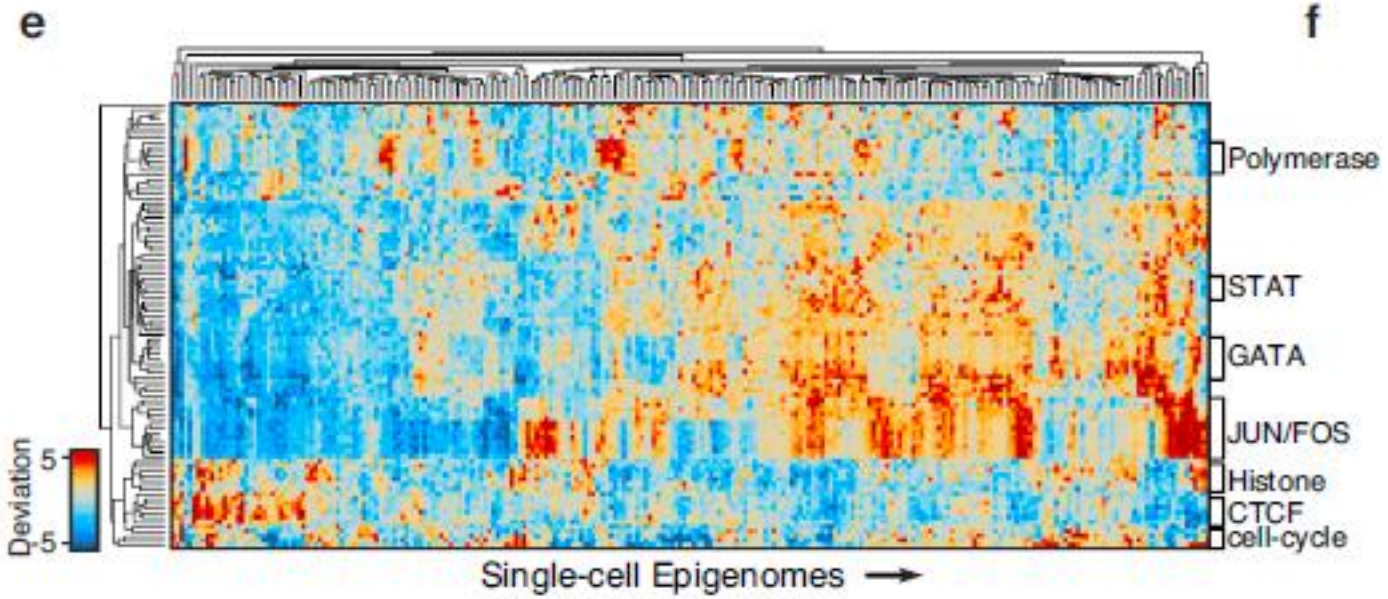
K562 Results

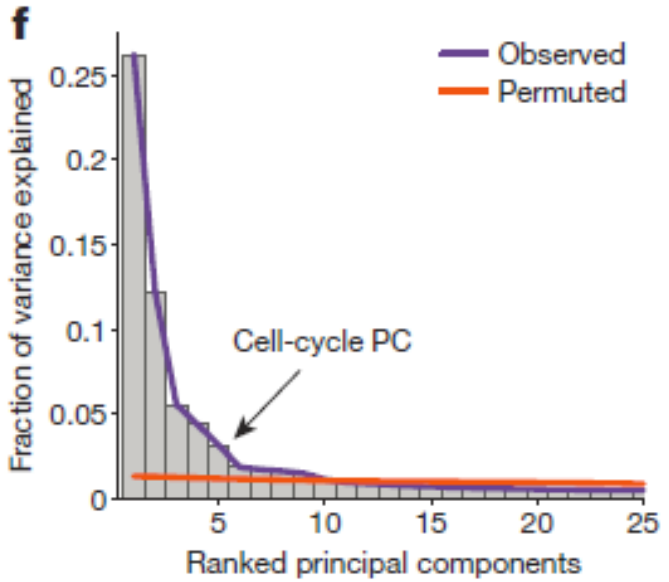
- To comprehensively characterize variability associated with trans-factors within individual K562 cells, they computed variability across:
  - All available ENCODE ChIP-seq,
  - Transcription factor motifs
  - Regions that differed in replication timing
- They found measures of cell-to-cell variability were highly reproducible across biological replicates
- Increased variability within different replication timing domains
  - Variable ATAC-seq signal associated with changes in DNA content across the cell cycle.
  - Trans-factors associated with high variability. GATA1/2, JUN and STAT2, and chromatin effectors, such as BRG1 (also known as SMARCA4) and P300 (also known as EP300).



K562 Results (cont)

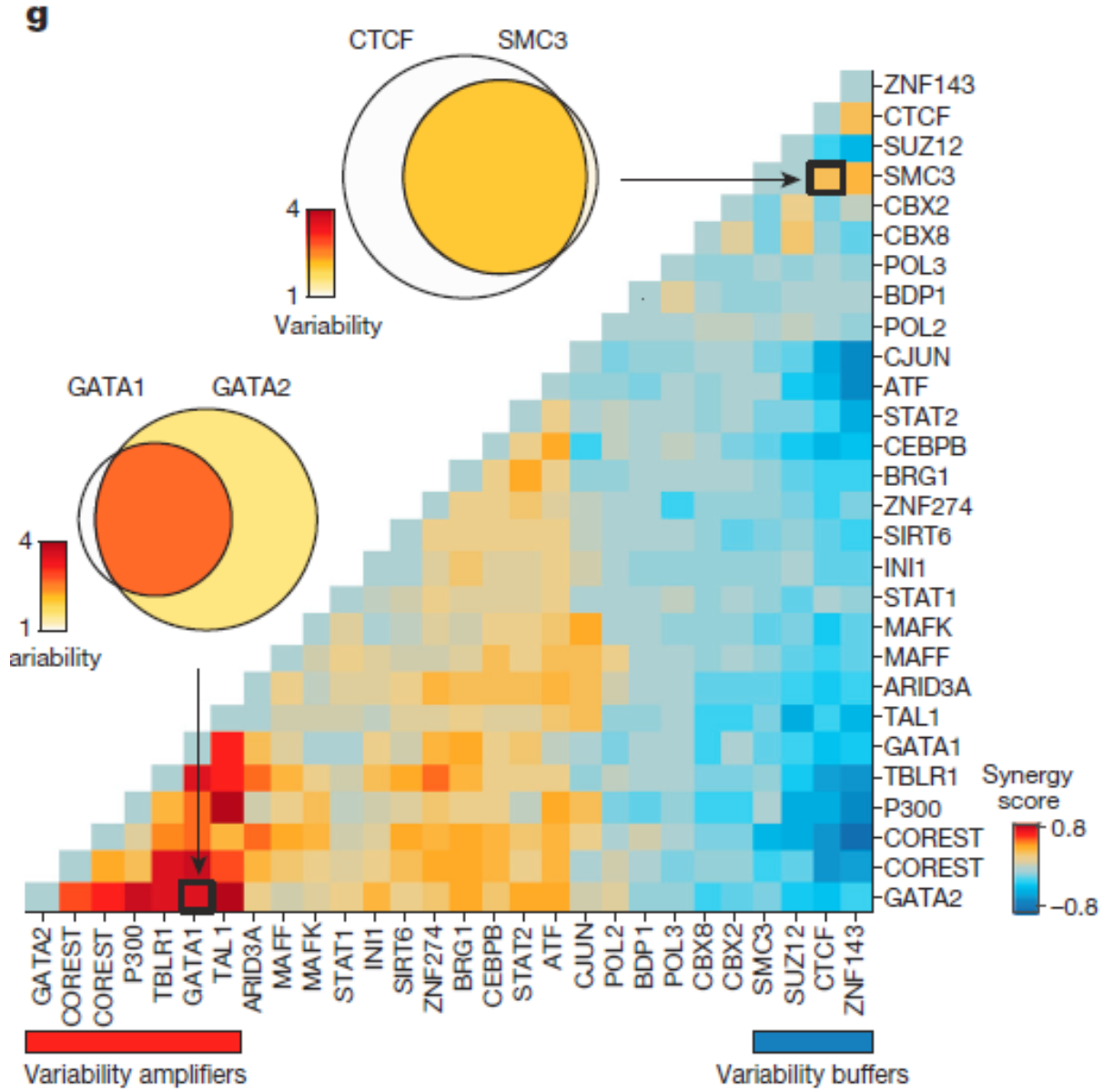
- Immunostaining followed by microscopy or flow cytometry confirmed heterogeneous expression of GATA1 and GATA2.
- Principal component (PC) analysis of single-cell deviations across all trans-factors show seven significant PCs,
  - PC 5 describing changes in DNA abundance throughout the cell cycle.
  - This analysis suggests that high-variance trans-factors are variable independent of the cell cycle.
  - The remaining PCs show contributions from several transcription factors, suggesting that variance across sets of trans-factors represent distinct regulatory states in individual cells.





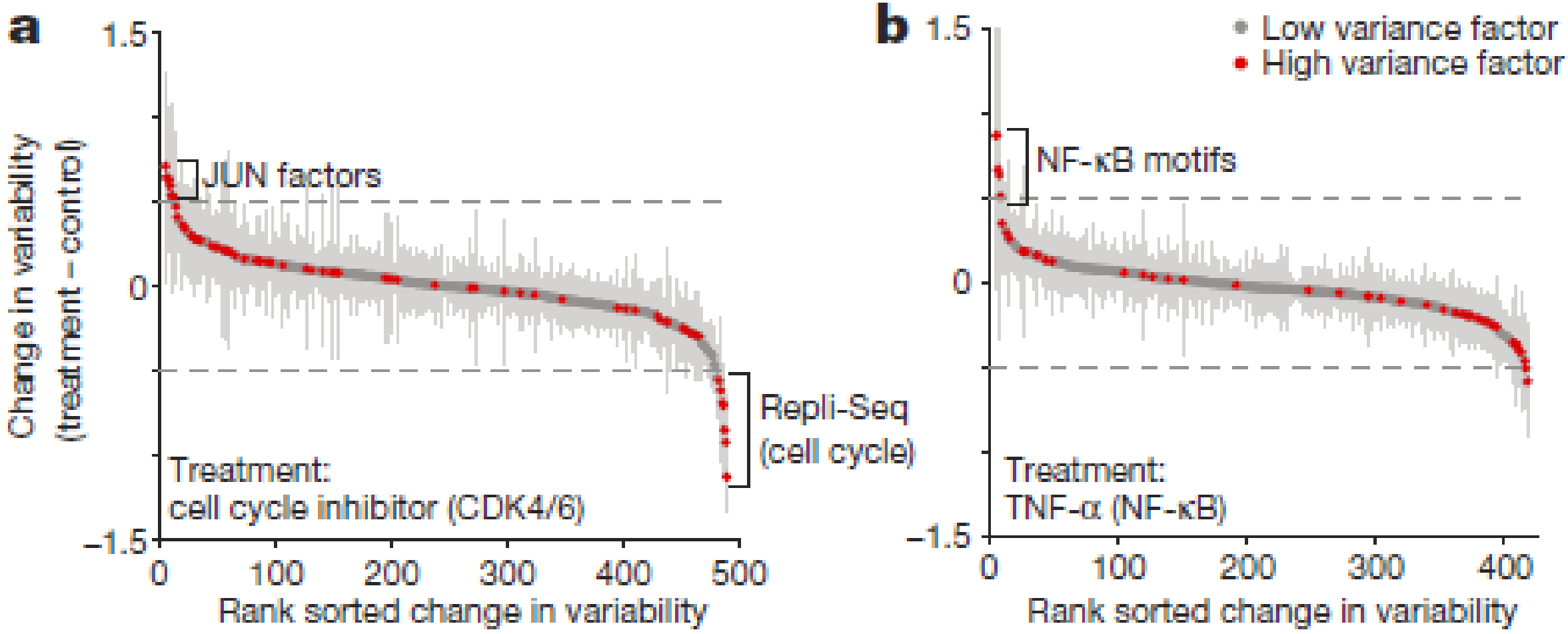
K562 – Synergistic Effects

- 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.
- 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
- Thus, single cell accessibility profiles nominate distinct trans-factors that, in combination, induce or suppress cell-to-cell regulatory variation.



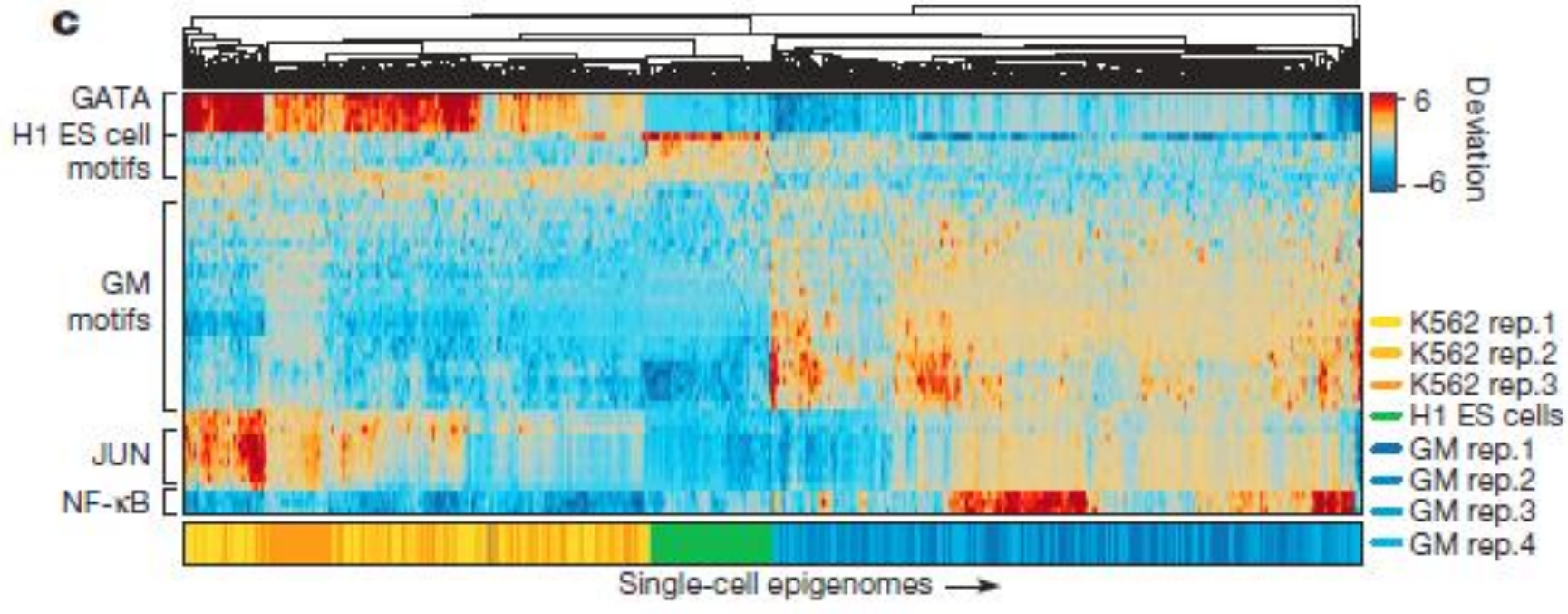
Accessibility can be Modulated Experimentally

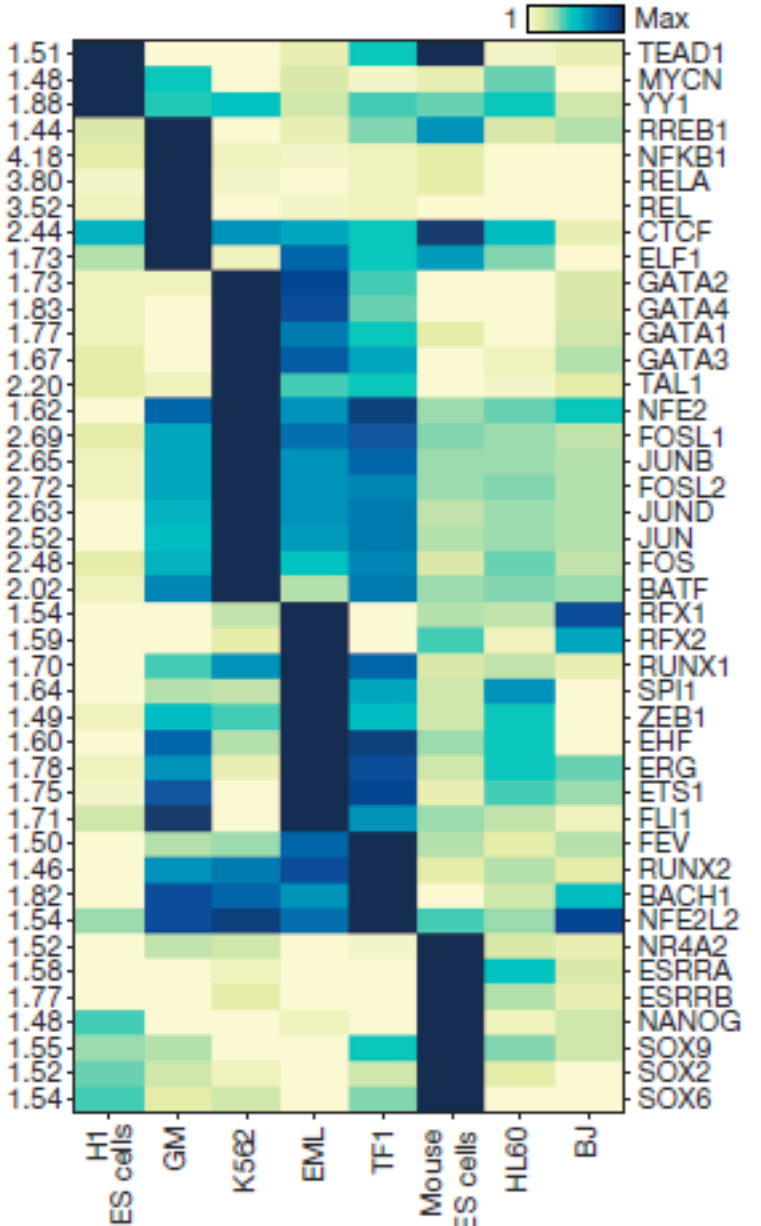
- Variability can be experimentally modulated and further demonstrates that variability is not solely dependent on the cell cycle.



Cell Type Specific Regulators of Deviations in Accessibility

- Cells from different biological replicates cluster with their cell type of origin
- scATAC-seq can also be used to deconvolve heterogeneous cellular mixtures.
- High variance trans-factor motifs that are generally unique to specific cell types
  - Regions associated with GATA transcription factors are most variant in K562 cells, whereas regions associated with master pluripotency
  - Nanog and Sox2 are most variant in mouse ES cells





Higher Order Chromosome Folding and Deviation of Accessibility Overlaps Well

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.