METHODS

Detailed protocols and computational methods are available in Supplementary Material.

Single-cell ATAC-Seq

Single-cell ATAC-Seq datasets were provided by 10xGenomics. We used Cell Ranger ATAC v 1.1.0 to process them. At first, we applied count procedure to process individual samples (old and young mice), later these were aggregated with dedicated aggregate Cell Ranger ATAC procedure with library depth normalization.

Count and aggregate command used:

cellranger-atac count --id=<id> --fastqs=<folder> --sample <sample> --reference refdata-cellranger-atac-mm10-1.1.0

cellranger-atac aggr --id=<id> --csv merged.csv --normalize=depth --reference=refdata-cellranger-atac-mm10-1.1.0

Raw reads profiles were visualized by deeptools v 3.1.3 as a part of a technical pipeline for NGS data using Snakemake workflow tool (https://github.com/JetBrains-Research/chipseq-smk-pipeline).

Both bulk reads, as well as individual cell populations, can be explored online in JBR Genome Browser viewer (https://research.jetbrains.org/groups/biolabs/tools/jbr-genome-browser) at http://artyomovlab.wustl.edu/jbr/2019\_gxfer1\_DT1634\_Denis.

Original ATAC-Seq analysis includes bulk peak calling with their algorithm, without any additional parameters, some meaningful biological peaks were missing. Bulk libraries were reprocessed with SPAN v 0.11.0.4882 and JBR v 1.0.beta.4882 (https://research.jetbrains.org/groups/biolabs/tools/span-peak-analyzer) to ensure the presence of these locations in downstream analysis.

Pipeline consisted of the following steps: cell calling, normalization, feature selection, dimensionality reduction, t-SNE projection, clustering, and differential markers analysis. It is written in Python v3.7 and is available at https://github.com/JetBrains-Research/sc-atacseq-explorer. Cell calling was done with reads in barcode thresholds. 200 was used to filter out the noise and 8000 to get rid of possible “barcode multiplets” (<https://www.biorxiv.org/content/early/2019/10/30/824003.full.pdf>).

Blacklist regions by Kundaje lab were downloaded from http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/mm10-mouse/mm10.blacklist.bed.gz.
All the reads intersecting blacklist regions were ignored.

We applied standard Cell Ranger normalization to median count per UMI with peak length normalization to make DNA accessible sites of different widths to make the same impact on analysis. Sites with extremely high coverage (>99%) were ignored as they likely represent sequencing errors or housekeeping genes. We noticed that filtering out sites with low variation (<1%) improves the robustness of real biological signal. (Like SnapATAC and Seurat). Standard Cell Ranger ATAC methods were applied to perform further steps: Inverse Document Frequency normalization, IRLBA Singular Vector Decomposition, graph clustering, and t-SNE projection and differential analysis. Also, we used a novel UMAP method for data visualization.

We annotated peaks with corresponding genes with bedtools command:

bedtools closest -D b to associate arbitrary genome location to closest gene reporting distance to Transcription Start Site upstream or downstream. Genes markup v M22 was downloaded from the Gencode portal (ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode\_mouse/release\_M22/gencode.vM22.annotation.gtf.gz).

All the plots were created with matplotlib v XXX and seaborn v XXXX.

Ready-to-explore dataset for Single Cell Explorer is available at <https://artyomovlab.wustl.edu/sce/?token=2019_scatacseq_Denis>.