An ATAC-seq atlas of chromatin accessibility in mouse tissues

Liu, Chuanyu; Wang, Mingyue; Wei, Xiaoyu; Wu, Liang; Xu, Jiangshan; Dai, Xi; Xia, Jun; Cheng, Mengnan; Yuan, Yue; Zhang, Pengfan; Li, Jiguang; Feng, Taiqing; Chen, Ao; Zhang, Wenwei; Chen, Fang; Shang, Zhouchun; Zhang, Xiuqing; Peters, Brock A.; Liu, Longqi

Published in:
Scientific Data

DOI:
10.1038/s41597-019-0071-0

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Liu, C., Wang, M., Wei, X., Wu, L., Xu, J., Dai, X., Xia, J., Cheng, M., Yuan, Y., Zhang, P., Li, J., Feng, T., Chen, A., Zhang, W., Chen, F., Shang, Z., Zhang, X., Peters, B. A., & Liu, L. (2019). An ATAC-seq atlas of chromatin accessibility in mouse tissues. Scientific Data, 6, [65]. https://doi.org/10.1038/s41597-019-0071-0
An ATAC-seq atlas of chromatin accessibility in mouse tissues

Chuanyu Liu1,2,3, Mingyue Wang1,2,3, Xiaoyu Wei1,2,3, Liang Wu1,2,3, Jiangshan Xu1,2,3, Xi Dai1,2,3, Jun Xia1,2,4, Mengnan Cheng1,2,3, Yue Yuan1,2,3, Pengfan Zhang1,2,3, Jiguang Li2,4, Taiqing Feng2,4, Ao Chen1,2, Wenwei Zhang5,3, Fang Chen2,3,5,6, Zhouchun Shang1,2,3, Xiqin Zhang1,2,3, Brock A. Peters4,5,6,7 & Longqi Liu2,3

The Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) is a fundamental epigenomics approach and has been widely used in profiling the chromatin accessibility dynamics in multiple species. A comprehensive reference of ATAC-seq datasets for mammalian tissues is important for the understanding of regulatory specificity and developmental abnormality caused by genetic or environmental alterations. Here, we report an adult mouse ATAC-seq atlas by producing a total of 66 ATAC-seq profiles from 20 primary tissues of both male and female mice. The ATAC-seq read enrichment, fragment size distribution, and reproducibility between replicates demonstrated the high quality of the full dataset. We identified a total of 296,574 accessible elements, of which 26,916 showed tissue-specific accessibility. Further, we identified key transcription factors specific to distinct tissues and found that the enrichment of each motif reflects the developmental similarities across tissues. In summary, our study provides an important resource on the mouse epigenome and will be of great importance to various scientific disciplines such as development, cell reprogramming, and genetic disease.

Background & Summary

Although most of the protein-coding genes in human and model animals such as mouse have been extensively annotated, vast regions of the genome are noncoding sequences (e.g., roughly 98% of the human genome) and still poorly understood1,2. During the last decade, the development of next-generation sequencing (NGS) based epigenomics techniques (e.g., ChIP-seq and DNase-seq) have significantly facilitated the identification of functional genomic regions3. For example, by comparing the histone modifications and transcription factor (TF) binding patterns throughout the mouse genome in a wide spectrum of tissues and cell types, Yue et al.4,5 have made significant progress towards a comprehensive catalog of potential functional elements in the mouse genome. So far, the international human epigenome consortium (IHEC), including ENCODE and the NIH Roadmap epigenomics projects, have profiled thousands of epigenomes including DNA methylation, genome-wide binding of TFs, histone modifications, and chromatin accessibility. This has resulted in the discovery of over 5 million cis-regulatory elements (CREs) in the human genome6–8. These data resources have created an important baseline for further study of diverse biological processes, such as development, cell reprogramming, and human disease6–13.

The accessibility of CREs, which is important for switching on and off gene expression14, is strongly associated with transcriptional activity. To date, detection of DNase I hypersensitive sites (DHSs) within chromatin by DNase-seq has been extensively used to map accessible genomic regions in diverse organisms including the laboratory mouse15. In 2013, Buenrostro et al.15 reported an alternative approach, termed ATAC-seq, for fast and sensitive profiling of chromatin accessibility by direct transposition of native chromatin within the nucleus. This method, in comparison to DNase-seq, requires a significantly lower input of cells (only 500–50,000) and a shorter period to process samples16. Moreover, ATAC-seq has been applied to single cells through various methods17–20.

1BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, 518083, China. 2BGI-Shenzhen, Shenzhen, 518083, China. 3China National GeneBank, BGI-Shenzhen, Shenzhen, 518120, China. 4MGI, BGI-Shenzhen, Shenzhen, 518083, China. 5Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, Copenhagen, 2100, Denmark. 6Advanced Genomics Technology Lab, Complete Genomics Inc., 2904 Orchard Parkway, San Jose, California, 95134, USA. These authors contributed equally: Chuanyu Liu, Mingyue Wang and Xiaoyu Wei. Correspondence and requests for materials should be addressed to B.A.P. (email: bpeters@completegenomics.com) or L.L. (email: liulongqi@genomics.cn)

Received: 29 November 2018
Accepted: 10 April 2019
Published online: 20 May 2019
enabling the investigation of regulatory heterogeneity within complex tissues. As such, ATAC-seq has demonstrated great potential to be a leading method in assaying accessible chromatin genome-wide.

The sequence preference of both DNase I and Tn5 enzymes produced distinct but inevitable biases in DNase-seq and ATAC-seq21, making it impractical to directly compare datasets generated from the two methods. Therefore, although the DNase-seq atlas of adult mouse tissues has been published5, a baseline of chromatin accessible regions generated from ATAC-seq is still important for ATAC-seq based studies. Here, we applied Omni-ATAC-seq22, an approach that enables profiling of accessible chromatin from frozen tissues, to the generation of 66 chromatin accessibility datasets from 20 different tissues derived from both adult male and female C57BL/6J mice (Fig. 1a). Systematic analysis of the dataset identified a total of 296,574 accessible elements, of which 26,916 showed highly tissue-specific patterns. We further predicted TFs specific to distinct tissues and importantly, many of these have been validated by previous studies23–27. In this study, we provide a valuable resource, which can be used to elucidate transcriptional regulation and may further help understand diseases caused by regulatory dysfunction.

Methods

Sample collection. All relevant procedures involving animals were approved by the Institutional Review Board on Ethics Committee of BGI (Permit No. BGI-R085-1). C57BL/6J male and female mice were purchased from Beijing Vital River Laboratory Animal Technology Co., Ltd (Beijing, China). 8-week old mice were used for this study. Mice were housed under standard conditions of a specific pathogen-free, temperature-controlled environment with a 12-h day/night cycle28. The mice were sacrificed by cervical dislocation. Whole organs were extracted and cut into 2–3 pieces, respectively (50–200 mg/piece). Each sample was then quickly frozen in liquid nitrogen and stored at −80 °C until nuclei extraction was performed. In this study, we used 20 different organs or tissues, including adrenal gland, bladder, brain (cerebrum and cerebellum), fat (abdominal, brown and mesenteric), heart, intestine (large and small), kidney, liver, ovary, pancreas, skeletal muscle, spleen, stomach, thymus, and uterus (as listed in Table 1).

Library construction and sequencing. Tissues were homogenized in a 2 ml Dounce homogenizer (with a loose and then tight pestle) with 10–20 strokes in 2 ml of 1X homogenization buffer on ice22. 400 μl of this nuclei suspension was transferred to a round-bottom 2 ml Lo-Bind Eppendorf tube for density gradient centrifugation following the protocol by Corces et al.22. After centrifugation, the nuclei band (about 200 μl) was collected, stained with DAPI, and nuclei were counted. Approximately 20,000–100,000 nuclei were transferred into a fresh tube and diluted in 1 ml ATAC-LSB +0.1% Tween-20 (Sigma-Aldrich, Darmstadt, Germany). Nuclei were centrifuged and the supernatant was carefully aspirated. Nuclei were treated in 50 μl transposition reaction mixture containing 10 mM TAPS-NaOH (pH 8.5), 5 mM MgCl2, 10% DMP, 2.5 μl of in-house Tn5 transposase (0.8 U/μl), 0.01% digitonin (Sigma-Aldrich, Darmstadt, Germany), 0.1% Tween-20, 31.5 μl of PBS, and 5 μl of nuclease-free water for 30 mins at 37 °C. Afterward, the DNA was purified with MinElute Purification Kit (Quigen, Venlo, Netherlands) and amplified with primers containing barcodes, as previously described22,29.

All libraries were adapted for sequencing on the BGISEQ-500 platform30. In brief, the DNA concentration was determined by Qubit 3.0 (ThermoFisher, Waltham, MA). Pooled samples were used to make single-strand DNA circles (ssDNA circles). DNA nanoballs (DNBs) were generated from the ssDNA circles by rolling circle replication as previously described30. The DNBs were loaded onto patterned nano-arrays and sequenced on the BGISEQ-500 sequencing platform with paired end 50 base reads.
| Sample ID | Strain  | Serial | Gender | Tissue            | Replicate |
|-----------|---------|--------|--------|-------------------|-----------|
| ATAC-1    | C57BL/6J | 1      | Female | Adrenal gland     | 1         |
| ATAC-2    | C57BL/6J | 1      | Female | Adrenal gland     | 2         |
| ATAC-3    | C57BL/6J | 1      | Female | Cerebrum          | 1         |
| ATAC-4    | C57BL/6J | 1      | Female | Cerebrum          | 2         |
| ATAC-5    | C57BL/6J | 1      | Female | Cerebellum        | 1         |
| ATAC-5    | C57BL/6J | 1      | Female | Cerebellum        | 2         |
| ATAC-7    | C57BL/6J | 1      | Female | Abdominal fat     | 1         |
| ATAC-8    | C57BL/6J | 1      | Female | Abdominal fat     | 2         |
| ATAC-9    | C57BL/6J | 1      | Female | Brown fat         | 1         |
| ATAC-10   | C57BL/6J | 1      | Female | Brown fat         | 2         |
| ATAC-11   | C57BL/6J | 1      | Female | Mesenteric fat    | 1         |
| ATAC-12   | C57BL/6J | 1      | Female | Mesenteric fat    | 2         |
| ATAC-13   | C57BL/6J | 1      | Female | Heart             | 1         |
| ATAC-14   | C57BL/6J | 1      | Female | Heart             | 2         |
| ATAC-15   | C57BL/6J | 1      | Female | Kidney            | 1         |
| ATAC-16   | C57BL/6J | 1      | Female | Kidney            | 2         |
| ATAC-17   | C57BL/6J | 1      | Female | Liver             | 1         |
| ATAC-18   | C57BL/6J | 1      | Female | Liver             | 2         |
| ATAC-19   | C57BL/6J | 1      | Female | Lung              | 1         |
| ATAC-20   | C57BL/6J | 1      | Female | Lung              | 2         |
| ATAC-21   | C57BL/6J | 1      | Female | Ovary             | 1         |
| ATAC-22   | C57BL/6J | 1      | Female | Ovary             | 2         |
| ATAC-23   | C57BL/6J | 1      | Female | Pancreas          | 1         |
| ATAC-24   | C57BL/6J | 1      | Female | Pancreas          | 2         |
| ATAC-25   | C57BL/6J | 1      | Female | Skeletal muscle   | 1         |
| ATAC-26   | C57BL/6J | 1      | Female | Skeletal muscle   | 2         |
| ATAC-27   | C57BL/6J | 1      | Female | Spleen            | 1         |
| ATAC-28   | C57BL/6J | 1      | Female | Spleen            | 2         |
| ATAC-29   | C57BL/6J | 1      | Female | Thymus            | 1         |
| ATAC-30   | C57BL/6J | 1      | Female | Thymus            | 2         |
| ATAC-31   | C57BL/6J | 1      | Female | Uterus            | 1         |
| ATAC-32   | C57BL/6J | 1      | Female | Uterus            | 2         |
| ATAC-33   | C57BL/6J | 2      | Male   | Adrenal gland     | 1         |
| ATAC-34   | C57BL/6J | 2      | Male   | Adrenal gland     | 2         |
| ATAC-35   | C57BL/6J | 2      | Male   | Bladder           | 1         |
| ATAC-36   | C57BL/6J | 2      | Male   | Bladder           | 2         |
| ATAC-37   | C57BL/6J | 2      | Male   | Cerebrum          | 1         |
| ATAC-38   | C57BL/6J | 2      | Male   | Cerebrum          | 2         |
| ATAC-39   | C57BL/6J | 2      | Male   | Cerebellum        | 1         |
| ATAC-40   | C57BL/6J | 2      | Male   | Cerebellum        | 2         |
| ATAC-41   | C57BL/6J | 2      | Male   | Brown fat         | 1         |
| ATAC-42   | C57BL/6J | 2      | Male   | Brown fat         | 2         |
| ATAC-43   | C57BL/6J | 2      | Male   | Mesenteric fat    | 1         |
| ATAC-44   | C57BL/6J | 2      | Male   | Mesenteric fat    | 2         |
| ATAC-45   | C57BL/6J | 2      | Male   | Heart             | 1         |
| ATAC-46   | C57BL/6J | 2      | Male   | Heart             | 2         |
| ATAC-47   | C57BL/6J | 2      | Male   | Large intestine   | 1         |
| ATAC-48   | C57BL/6J | 2      | Male   | Large intestine   | 2         |
| ATAC-49   | C57BL/6J | 2      | Male   | Small intestine   | 1         |
| ATAC-50   | C57BL/6J | 2      | Male   | Small intestine   | 2         |
| ATAC-51   | C57BL/6J | 2      | Male   | Kidney            | 1         |
| ATAC-52   | C57BL/6J | 2      | Male   | Kidney            | 2         |
| ATAC-53   | C57BL/6J | 2      | Male   | Liver             | 1         |
| ATAC-54   | C57BL/6J | 2      | Male   | Liver             | 2         |
| ATAC-55   | C57BL/6J | 2      | Male   | Lung              | 1         |
| ATAC-56   | C57BL/6J | 2      | Male   | Lung              | 2         |
| ATAC-57   | C57BL/6J | 2      | Male   | Pancreas          | 1         |
| ATAC-58   | C57BL/6J | 2      | Male   | Pancreas          | 2         |
| ATAC-59   | C57BL/6J | 2      | Male   | Skeletal muscle   | 1         |
| ATAC-60   | C57BL/6J | 2      | Male   | Skeletal muscle   | 2         |
| ATAC-61   | C57BL/6J | 2      | Male   | Spleen            | 1         |
| ATAC-62   | C57BL/6J | 2      | Male   | Spleen            | 2         |
| ATAC-63   | C57BL/6J | 2      | Male   | Stomach           | 1         |
| ATAC-64   | C57BL/6J | 2      | Male   | Stomach           | 2         |
| ATAC-65   | C57BL/6J | 2      | Male   | Thymus            | 1         |
| ATAC-66   | C57BL/6J | 2      | Male   | Thymus            | 2         |

Table 1. Tissue and corresponding mouse and sample IDs.
### Table 2. ATAC-seq metadata and mapping statistics.

| Sample ID | Total Reads | Mapped Reads | chrM Reads | Usable Reads* | Percentage of Usable Reads | TSS Enrichment | IDR Peaks |
|-----------|-------------|--------------|------------|---------------|---------------------------|----------------|-----------|
| ATAC-1    | 98,303,382  | 96,954,685   | 23,288,521 | 16,098,792    | 16.38                     | 17.21          | 23,296    |
| ATAC-2    | 148,416,230 | 146,416,001  | 25,259,955 | 24,342,678    | 16.34                     | 17.83          | 23,296    |
| ATAC-3    | 150,254,776 | 148,205,698  | 1,773,531  | 116,678,430   | 77.65                     | 18.35          | 93,546    |
| ATAC-4    | 132,171,814 | 130,467,930  | 1,749,710  | 132,048,212   | 92.03                     | 19.11          | 33,546    |
| ATAC-5    | 44,251,998  | 43,312,156   | 254,745    | 35,315,756    | 79.81                     | 13.90          | 32,337    |
| ATAC-6    | 180,555,698 | 178,313,881  | 2,085,323  | 136,439,604   | 75.57                     | 14.41          | 32,337    |
| ATAC-7    | 170,443,536 | 166,598,249  | 1,204,915  | 61,090,112    | 35.80                     | 13.03          | 28,877    |
| ATAC-8    | 165,896,236 | 135,431,464  | 1,175,420  | 61,493,964    | 70.75                     | 12.63          | 28,877    |
| ATAC-9    | 134,223,704 | 129,656,384  | 14,098,291 | 49,193,108    | 61.17                     | 16.75          | 66,620    |
| ATAC-10   | 160,942,400 | 156,016,214  | 23,296     | 59,910,266    | 37.22                     | 17.21          | 66,620    |
| ATAC-11   | 157,502,400 | 154,297,472  | 1,228,888  | 93,223,318    | 91.59                     | 8.10           | 38,451    |
| ATAC-12   | 162,615,494 | 159,625,229  | 2,527,579  | 103,453,218   | 63.62                     | 7.44           | 38,451    |
| ATAC-13   | 238,728,070 | 229,371,653  | 14,280,715 | 116,609,704   | 88.85                     | 7.82           | 38,877    |
| ATAC-14   | 268,182,264 | 259,270,286  | 18,497,575 | 132,829,548   | 49.53                     | 7.11           | 38,877    |

*Unusable reads: number of mapped read minus number of low mapping quality, duplicate and mitochondrial reads.
Preprocessing of the ATAC-seq datasets. The ATAC-seq data were processed (trimmed, aligned, filtered, and quality controlled) using the ATAC-seq pipeline from the Kundaje lab\(^31,32\) (Table 2). The model-based analysis of ChIP-seq (MACS2)\(^33\) version 2.1.2 was used to identify the peak regions with options -B, -q 0.01–nomodel, -f BAM, -g mm. The Irreproducible Discovery Rate (IDR) method\(^34\) was used to identify reproducible peaks between two technical replicates (Fig. 1b). Only peaks reproducible between the two technical replicates were retained for downstream analyses. Peaks for all tissues were then merged together into a standard peak list.

The number of raw reads mapped to each standard peak were counted using the intersect function of BedTools\(^35\) version 2.26.0. The raw count matrix\(^32\) was normalized by Reads Per Million mapped reads (RPM). Pearson correlation coefficients between technical or biological replicates across tissues were calculated based on the Log10 RPM matrix.
Identification of tissue-specific chromatin accessible regions. We used a strategy described previously based on the Shannon entropy to compute a tissue specificity index for each peak \[ i \] as \[ R_i = E_i / \Sigma E \], where \( E_i \) is the RPM value for the peak in the tissue \( i \), \( \Sigma E \) is the sum of RPM values in all tissues, and \( N \) is the total number of tissues. The entropy
score for each peak across tissues can be defined as \( H = -1 \sum (R_i \log_2 R_i) \) (1 < i < N), where the value of \( H \) ranges between 0 to \( \log_2(N) \). An entropy score close to zero indicates the accessibility of this peak is highly tissue-specific, while an entropy score close to \( \log_2(N) \) indicates that this peak is ubiquitously accessible\(^3\). Based on the distribution of entropy scores, peaks with score less than 3.5 were selected as tissue-restricted peaks.

We searched TF motifs in tissue-specific peaks using the *findMotifsGenome.pl* script of the HOMER\(^4\) version 4.9.1 software with default settings. We then generated a motif enrichment matrix\(^3\), where each row represents the P-value of a motif and each column represents a tissue. The 50 motifs with the top CV values and mean values greater than 20 were displayed.

**Data Records**

A complete list of the 66 tissue samples is given in Tables 1 and 2. All raw data have been submitted to the CNGB Nucleotide Sequence Archive\(^5\). The raw data have also been submitted to the NCBI Sequence Read Archive\(^6\). The ATAC-seq QC results and count matrices have been submitted to the Figshare\(^7\).
Technical Validation

Data QC from the pipeline with IDR quality control. We evaluated our ATAC-seq dataset by a series of commonly used statistics, including the number of total read, mapping rate, the proportion of duplicate read, the number of mitochondrial read, and the number of final usable read (Table 2). For each replicate, we obtained an average of 78 million reads, which was previously shown to be enough for the detection of accessible regions\textsuperscript{35,31}. In agreement with published ATAC-seq profiles\textsuperscript{35}, the chromatin accessibility fragments show size periodicity corresponding to integer multiples of nucleosomes\textsuperscript{32} (Fig. 2b). The successful detection of accessible regions is also supported by the observation of strong enrichment of ATAC-seq reads around transcription start sites (TSSs)\textsuperscript{32} (Fig. 2a,d).

To evaluate the reproducibility of accessible element discovery between replicates, we identified accessible regions in both replicates by using the MACS2\textsuperscript{33} algorithm. We then applied the IDR method\textsuperscript{34} to find peaks that were reproducible between replicates in each tissue type (Fig. 2c). We identified an average of 43,421 reproducible peaks (Table 2). For downstream analyses, we filtered out low-quality data where the TSS enrichment scores are less than 10.0 and the number of reproducible peaks are less than 10,000.

Reproducibility of biological samples and comparison with published studies. The Pearson correlation analysis was used to further examine the reproducibility of biological and technical replicates. Heatmap clustering of Pearson correlation coefficients from the comparison of 66 datasets revealed a strong correlation between replicates of the same tissue (Fig. 3b), but a lower correlation between profiles from distinct tissues. This result is also supported by t-distributed stochastic neighbor embedding (t-SNE) analysis with tissue-restricted peaks of all profiles (Fig. 3a,c). Interestingly, correlations between replicates from mice of the same gender are generally higher than those from the opposite gender. This can be seen in the cerebrum where the correlation coefficient between replicates of female mice is 0.99 (Fig. 3d), while the coefficient between male and female is slightly reduced (0.96). We also compared our data to ATAC-seq profiles of postnatal mouse (day 0) tissues downloaded from ENCODE project\textsuperscript{42,43}. Importantly, we found that both heart and lung were comparable with each other (Fig. 3e). Taken together, these analyses strongly suggest that our ATAC-seq profiles can reliably detect accessible chromatin regions in the mouse genome and can be used as a basic reference ATAC-seq dataset for future studies.

Inferring tissue-specific transcription factors. In an effort to validate the tissue-specific TF motifs identified in our dataset, we compared them to results from previous studies. Log2 RPM of the tissue-restricted peaks was shown in the heatmap (Fig. 4a). For example, we observed high enrichment of the NeuroG2 motif in cerebellum and cerebral (Fig. 4b), in agreement with the role of NeuroG2 in controlling the temporal switch from neurogenesis to gliogenesis and regulating laminar fate transitions\textsuperscript{25}. In brown fat, we found the CCAAT-enhancer-binding proteins (CEBP) motif to be highly specific (Fig. 4b). This is supported by a previous study demonstrating that CEBP can cooperate with PRDM16 to induce brown fat determination and differentiation\textsuperscript{25}. In addition, other well-known tissues-specific motifs such as the liver-specific HNF family TF motifs (Hnf1, HNF1b, and HNF4a)\textsuperscript{25,32} and heart or skeletal muscle specific MeF2 family motifs (MeF2a, MeF2b, MeF2c, MeF2d)\textsuperscript{26,27} were validated in our study. To further validate whether the overall motif enrichment in each tissue can reflect tissue specificity we performed hierarchical clustering of tissues using Euclidean distance (Fig. 4c). This provided a result similar to hierarchical clustering of various mouse tissues based on RNA-seq data\textsuperscript{44}. In addition, examination of tissues from the gastrointestinal (GI) tract (i.e., large intestine, small intestine, and stomach) showed tight clustering (Fig. 4c), which is likely due to their common functions such as lipid metabolism and energy hemostasis\textsuperscript{25,46}. Skeletal muscle and heart tissue are found in the same branch, suggesting that patterns of chromatin accessibility in the two tissues are highly influenced by shared TF motifs such as those from the MeF2 family\textsuperscript{35}.

Usage Notes

The ATAC-seq data processing pipeline, including read mapping, peak calling, IDR analysis, and read counting were run on the Linux operating system. The optimized parameters are provided in the main text. All R source codes used for the downstream data analyses and visualization are provided in Supplementary File 1.

Code Availability

The R codes used for correlation analysis, identification of tissue-specific chromatin accessible regions, and tissue-specific TFs are available in the supplementary materials (Supplementary File 1). A repository list containing the chromatin accessibility raw count matrix and the motif enrichment matrix is available online\textsuperscript{46}.

References

1. Consortium, I. H. G. S. Initial sequencing and analysis of the human genome. Nature 409, 860 (2001).
2. Venter, J. C. et al. The sequence of the human genome. Science 291, 1304–1351 (2001).
3. Rivera, C. M. & Ren, B. Mapping human epigenomes. Cell 155, 39–55 (2013).
4. Shen, Y. et al. A map of the cis-regulatory sequences in the mouse genome. Nature 488, 116 (2012).
5. Yue, F. et al. A comparative encyclopedia of DNA elements in the mouse genome. Nature 515, 355–364 (2014).
6. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57 (2012).
7. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317 (2015).
8. Stunnenberg, H. G. et al. The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. Cell 167, 1145–1149 (2016).
9. Xu, Q. & Xie, W. Epigenome in Early Mammalian Development: Inheritance, Reprogramming and Establishment. Trends Cell Biol 28.3, 237–253 (2017).
10. Eckerley-Maslin, M. A., Alda-Catalinas, C. & Reik, W. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat. Rev. Mol. Cell Biol.* **19**, 436–450 (2018).
11. Apostolou, E. & Hochedlinger, K. Chromatin dynamics during cellular reprogramming. *Nature* **502**, 462 (2013).
12. Takahashi, K. & Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat. Rev. Mol. Cell Biol.* **17**, 183 (2016).
13. Karczewski, K. J. & Snyder, M. P. Integrative omics for health and disease. *Nat. Rev. Genet.* **19**, 239 (2018).
14. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
15. 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. *Nat. Methods* **10**, 1213–1218 (2013).
16. Meyer, C. A. & Liu, X. S. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat. Rev. Genet.* **15**, 709 (2014).
17. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486 (2015).
18. Cusanovich, D. A. et al. Multiplex single-cell profiling of chromatin accessibility by combinatory cellular indexing. *Science* **348**, 910–914 (2015).
19. Mezger, A. et al. High-throughput chromatin accessibility profiling at single-cell resolution. *Nat. Commun.* **9**, 3647 (2018).
20. Chen, X., Natarajan, K. N. & Teichmann, S. A. A rapid and robust method for single cell chromatin accessibility profiling. *Nat. Commun.* **9**, 5345 (2018).
21. Karabacak Calviello, A., Hirsekorin, A., Wurnus, R., Yusuf, D. & Ohler, U. Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets using protocol-specific bias modeling. *Genome Biol.* **20**, 42 (2019).
22. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
23. Dennis, D. J. et al. Neurog2 and Ascl1 together regulate a postmitotic derepression circuit to govern laminar fate specification in the murine neocortex. *Proc. Natl Acad. Sci.* **114**, 4934–4943 (2017).
24. Kaijura, S. et al. Initiation of myoblast to brown fat switch by a PRDM16-C/EBP-β transcriptional complex. *Nature* **460**, 1154 (2009).
25. Xanthopoulos, K. G. et al. The different tissue transcription patterns of genes for HNF-1, C/EBP, HNF-3, and HNF-4, protein factors that govern liver-specific transcription. *Proc. Natl Acad. Sci.* **88**, 3807–3811 (1991).
26. Fickett, J. W. Quantitative discrimination of MEF2 sites. *Mol. Cell. Biol.* **16**, 437–441 (1996).
27. Lu, J., McKinsey, T. A., Zhang, C.-L. & Olson, E. N. Regulation of skeletal myogenesis by association of the MEF2 transcription factor with class II histone deacetylases. *Cell Mol. Cell* **6**, 233–244 (2000).
28. Fischer, A. W., Cannon, B. & Nedergraad, J. Optimal housing temperatures for mice to mimic the thermal environment of humans: An experimental study. *Mol. Metab.* **7**, 161–170 (2018).
29. Shang, Z. et al. Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation. *Gigascience* **7**, 1–19 (2018).
30. Huang, J. et al. A reference human genome dataset of the BGISEQ-500 sequencer. *Gigascience* **6**, 1–9 (2017).
31. Koh, P. W. et al. An atlas of transcriptional, chromatin accessibility, and surface marker changes in human mesoderm development. *Sci. Data* **3**, 160109 (2016).
32. Liu C.-Y. et al. An ATAC-seq atlas of chromatin accessibility in mouse tissues. *figshare*, https://doi.org/10.6084/m9.figshare.c.4436264.v1 (2019).
33. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
34. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat* **5**, 1752–1779 (2011).
35. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
36. Barbera, L. O. et al. Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome Res.* **18**, 46–59 (2008).
37. Schug, J. et al. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* **6**, R33 (2005).
38. Xie, W. et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
39. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Cell* **38**, 576–589 (2010).
40. CNGB Nucleotide Sequence Archive. https://db.cnb.org/cnsa/project/CNP0000198/public/ (2018).
41. NCBI Sequence Read Archive. http://identifiers.org/nchb/insdc.sra:SRP167062 (2018).
42. ENCODE https://www.encodeproject.org/experiments/ENCSCR451NAE/ (2017).
43. ENCODE https://www.encodeproject.org/experiments/ENCSCR102NGD/ (2017).
44. Sollner, J. F. et al. An RNA-Seq atlas of gene expression in mouse and rat normal tissues. *Sci. Data* **4**, 170185 (2017).
45. Zhou, Q. et al. A mouse tissue transcription factor atlas. *Nat. Commun.* **8**, 15089 (2017).
46. Sonawane, A. R. et al. Understanding tissue-specific gene regulation. *Cell Rep.* **21**, 1077–1088 (2017).

**Acknowledgements**

We thank all members of the Cell and Development Lab (BGI) for helpful comments and Jian Zhang and Jie Chen from Shenzhen Institutes of Advanced Technology for assistance with sample collection. This work was supported by the National Key R&D Program of China (No. 2016YFC1303902), the Shenzhen Peacock Plan (No. KQTD20150330171505310), and the Science, Technology and Innovation Commission of Shenzhen Municipality (No. JCYJ20160531194327655).

**Author Contributions**

C.L., L.L., M.W. and X.W. conceived the idea. C.L. and M.W. collected samples. M.W. and C.L. generated the data. J.X., J. Xia, J.L., T.F., M.C. and Y.Y. assisted with the experiments. X.W. analyzed the data with the assistance of C.L., X.D., P.Z. and L.W. C.L. wrote the manuscript with the input of X.W. and M.W. L.L. and B.A.P. supervised the study and revised the manuscript. X.Z., F.C., W.Z., A.C. and Z.S. provided helpful comments on this study. All authors reviewed and approved the final manuscript.

**Additional Information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41597-019-0071-0.

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

The Creative Commons Public Domain Dedication waiver http://creativecommons.org/publicdomain/zero/1.0/ applies to the metadata files associated with this article.

© The Author(s) 2019