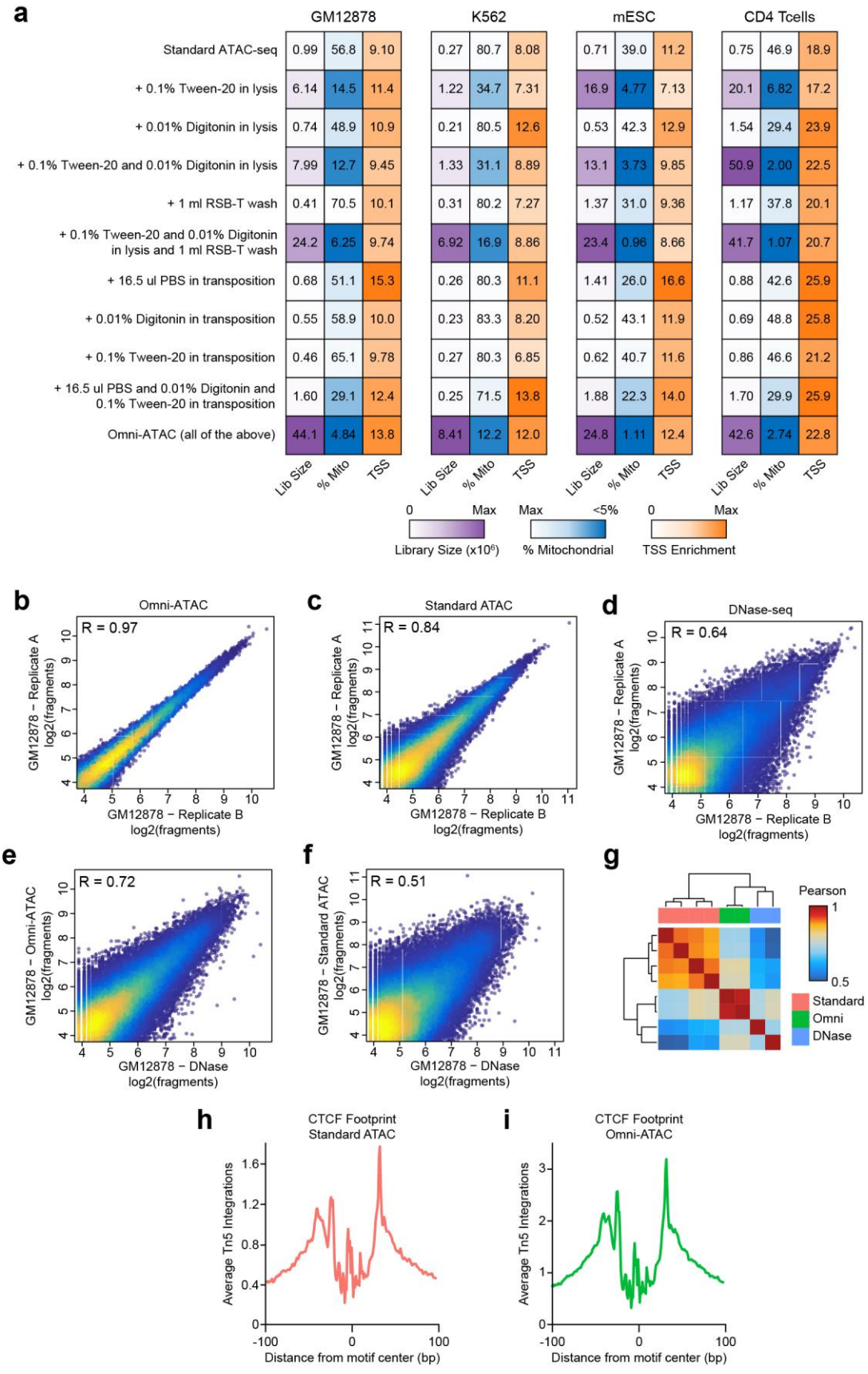
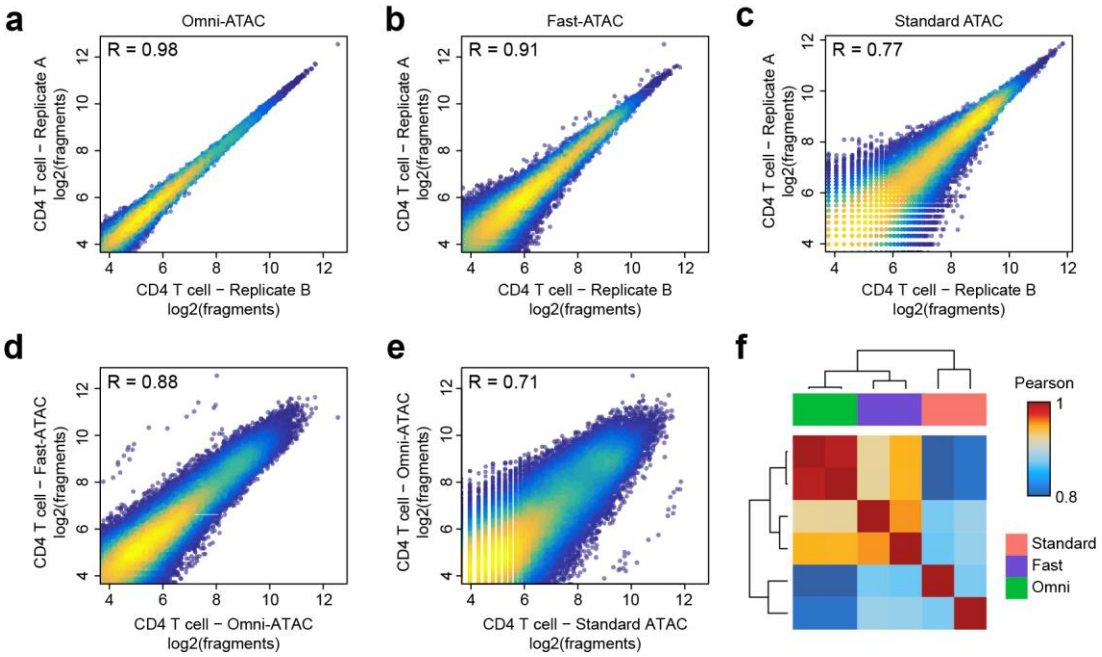


Nature Methods: doi:10.1038/nmeth.4396

Supplementary Figure 1

Comparison of technical replicate consistency between and across the standard ATAC-seq method, DNase-seq, and Omni-ATAC.

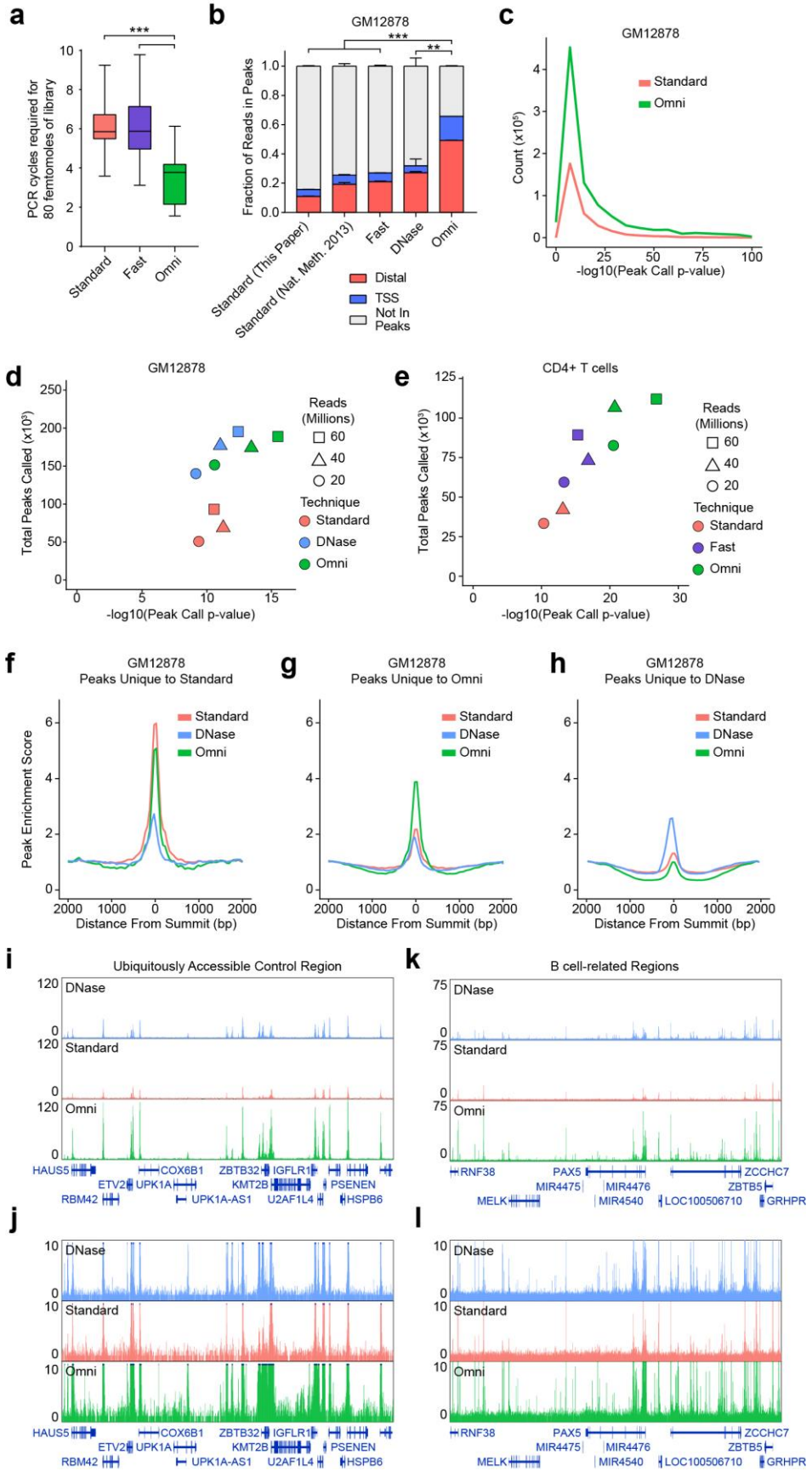
(**a**) Heatmap-based representation of ATAC-seq quality control metrics including library size (purple), percent of reads mapping to mitochondrial DNA (blue), and enrichment of signal at TSSs (orange) for each optimization made to produce the Omni-ATAC protocol. Deeper color is used to depict the most desirable value of each statistic. All values were determined from 100,000 random aligned reads. Data represents the average of two technical replicates per square. (**b-d**) Correlation of technical replicates from GM12878 ATAC-seq data generated using (b) the Omni-ATAC protocol, (c) the standard ATAC-seq protocol, and (d) DNase-seq. Data for standard ATAC-seq and DNase-seq were obtained from publicly available data. Each dot represents an individual peak. Peaks analyzed were derived from a union peak set using data from all three methods. Only 50,000 random peaks with more than 5 reads are shown. R value indicated at the top of each plot. (**e,f**) Correlation of GM12878 ATAC-seq data generated using (e) Omni-ATAC protocol and (f) the standard ATAC-seq protocol with DNase-seq as in (b-d). (**g**) Pearson correlation heatmap showing sample by sample unsupervised clustering on all peaks identified across all technical replicates from each method. Each sample represents an individual technical replicate. (**h,i**) CTCF meta-footprint across the top 20,000 ATAC-seq peaks containing a CTCF motif in data from GM12878 cells derived using (h) the standard ATAC-seq protocol and (i) the Omni-ATAC protocol. Data for standard ATAC-seq method derived from published data[1].



Supplementary Figure 2

Comparison of Omni-ATAC to the Fast-ATAC method in primary human T cells.

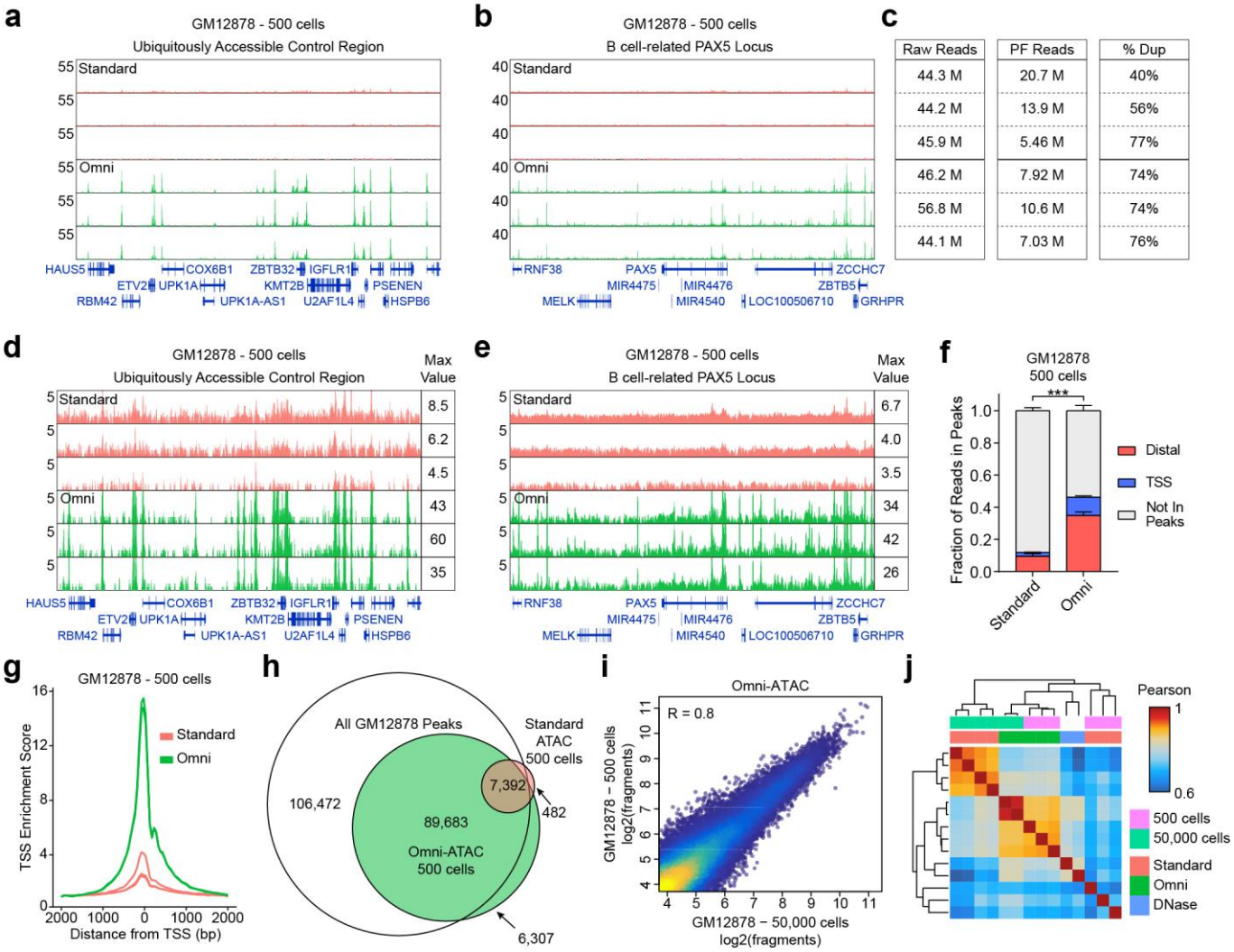
(**a-c**) Correlation of technical replicates from primary human CD4+ T cells cells using (a) the Omni-ATAC protocol, (b) the Fast-ATAC protocol, and (c) the standard ATAC-seq protocol. All data shown was generated in this study from the same sample of T cells. Each dot represents an individual peak. Peaks analyzed were derived from a union peak set using data from all three methods. Only 50,000 random peaks with more than 5 reads are shown. R value indicated at the top of each plot. (**d,e**) Correlation of ATAC-seq data generated using the Omni-ATAC protocol with data generated using (d) the Fast-ATAC protocol or (e) the standard ATAC-seq protocol in CD4+ T cells as in (a-c). (**f**) Pearson correlation heatmap showing sample by sample unsupervised clustering on all peaks identified across all technical replicates of CD4+ T cells from each method.


a
PCR cycles required for 80 femtomoles of library
Standard
Fast
Omni
b
GM12878
Fraction of Reads in Peaks
Standard (This Paper)
Standard (Nat. Meth. 2013)
Fast
DNase
Omni
Distal
TSS
Not In Peaks
c
GM12878
Standard
Omni
Count (x10⁶)
-log10(Peak Call p-value)
d
GM12878
Total Peaks Called (x10³)
Reads (Millions)
60
40
20
Technique
Standard
DNase
Omni
-log10(Peak Call p-value)
e
CD4+ T cells
Total Peaks Called (x10³)
Reads (Millions)
60
40
20
Technique
Standard
Fast
Omni
-log10(Peak Call p-value)
f
GM12878
Peaks Unique to Standard
Peak Enrichment Score
Standard
DNase
Omni
Distance From Summit (bp)
g
GM12878
Peaks Unique to Omni
Standard
DNase
Omni
Distance From Summit (bp)
h
GM12878
Peaks Unique to DNase
Standard
DNase
Omni
Distance From Summit (bp)
i
Ubiquitously Accessible Control Region
DNase
Standard
Omni
HAUS5
COX6B1
ZBTB32
IGFLR1
ETV2
UPK1A
KMT2B
PSENEN
RBM42
UPK1A-AS1
U2AF1L4
HSPB6
k
B cell-related Regions
DNase
Standard
Omni
RNF38
PAX5
ZCCHC7
MIR4475
MIR4476
ZBTB5
MELK
MIR4540
LOC100506710
GRHPR
j
DNase
Standard
Omni
HAUS5
COX6B1
ZBTB32
IGFLR1
ETV2
UPK1A
KMT2B
PSENEN
RBM42
UPK1A-AS1
U2AF1L4
HSPB6
l
DNase
Standard
Omni
RNF38
PAX5
ZCCHC7
MIR4475
MIR4476
ZBTB5
MELK
MIR4540
LOC100506710
GRHPR

Supplementary Figure 3

Comparison of Omni-ATAC to standard ATAC-seq and DNase-seq at the level of peaks.

(**a**) The number of PCR cycles required to amplify ATAC-seq libraries to an absolute yield of 80 femtomoles (20 ul at 4 nM) using the standard ATAC-seq protocol, the Fast-ATAC protocol, and the Omni-ATAC protocol. N=28 for each box chart, representing 2 technical replicates from 14 distinct cell types and cell contexts (Supplementary Table 1). (**b**) Fraction of reads in peaks mapping to TSSs (+/- 500 bp of TSS) and distal elements (>500 bp from TSS) from GM12878 libraries generated in this study and from public repositories using the standard ATAC-seq protocol ("Standard"), the Fast-ATAC-seq protocol ("Fast"), the Omni-ATAC-seq protocol ("Omni-ATAC"), and DNase-seq[11] ("DNase"). Each bar represents the mean of 2 technical replicates with the exception of the previously published ATAC-seq data which had 4 technical replicates. Error bars represent standard deviation. **$p < 0.01$, ***$p < 0.001$ by two-tailed unpaired students t-test comparing the fraction of reads in peaks to reads outside of peaks. All values were determined from 5 million random aligned de-duplicated reads. (**c**) Histogram comparing the distribution of macs2 peak scores (-log10(pvalue of peak call)) from peaks called from 20 million randomly selected aligned de-duplicated reads in GM12878 cells as in (d). (**d,e**) In silico simulation of peak calling from variable read depth using data derived from (d) GM12878 cells with the standard ATAC-seq protocol, DNase-seq, and the Omni-ATAC protocol or (e) CD4+ T cells with the standard ATAC-seq protocol, the Fast-ATAC protocol, and the Omni-ATAC protocol. Input reads were randomly selected from a pool of properly aligned, non-duplicate reads. All input reads were trimmed to equal length (36 bp) prior to alignment to match DNase-seq data. GM12878 data from the standard ATAC-seq protocol and DNase-seq were obtained from publicly available sources. All data for CD4+ T cells was generated from a single donor. Data points represent the mean of three individual down-sampled technical replicates. (**f-h**) Signal at peaks called as "unique" in the analysis presented in Figure 1c for (f) the standard ATAC-seq protocol, (g) the Omni-ATAC protocol, and (h) DNase-seq. Peak score is calculated in the same way as the TSS enrichment score but for a different set of specified peaks. (**i-l**) Input read normalized sequencing tracks at a control locus (i,j) and a B cell-related locus (k,l) from GM12878 cells derived using standard ATAC-seq, DNase-seq, and Omni-ATAC. In each case, 60 million properly aligned, non-duplicate reads were used to allow for direct comparison. All tracks in each panel are shown on the same y-axis scale. Tracks shown in (j,l) depict the same regions shown in (i,k) but with a reduced y-axis scale to show the difference in signal-to-background in data generated using the Omni-ATAC protocol compared to the standard ATAC-seq protocol and DNase-seq. Peaks that pass above the 10 read y-axis limit are truncated at y=10. All input reads were trimmed to equal length (36 bp) prior to alignment to match DNase-seq. Regions shown represent (i,j) chr19:36,097,939-36,272,939, and (k,l) chr9:36,373,966-37,494,500.

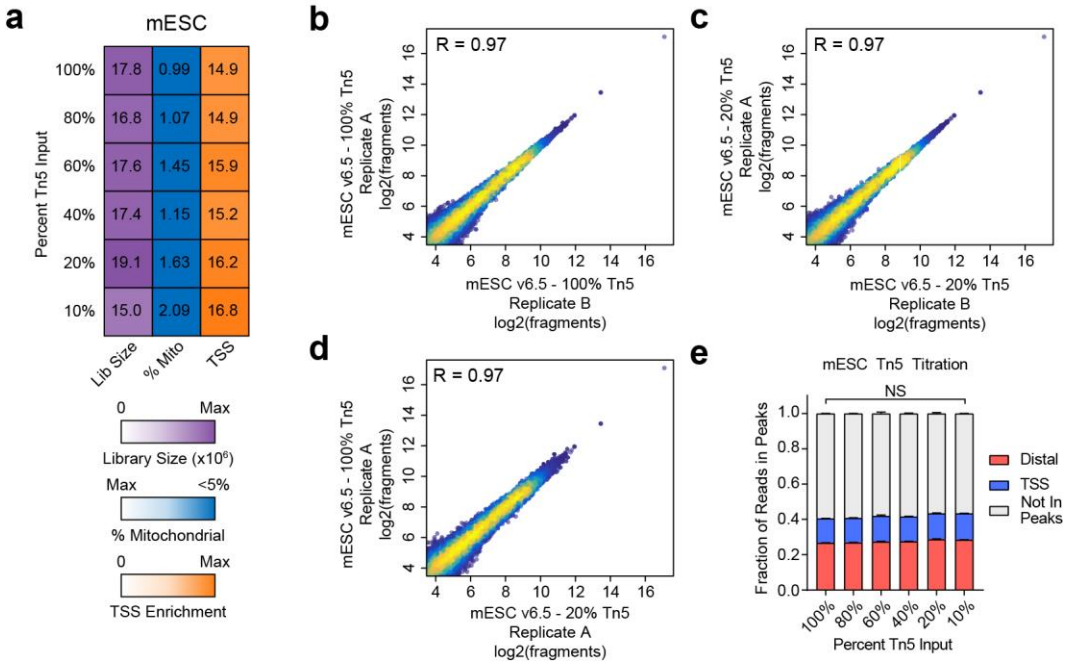


Supplementary Figure 4

Omni-ATAC outperforms the standard ATAC-seq method in situations requiring low cell numbers.

(**a,b**) Sequencing tracks of ATAC-seq data at (a) a control locus and (b) the PAX5 locus derived from the standard ATAC-seq protocol and the Omni-ATAC protocol using 500 GM12878 cells. No normalization was performed. All available filtered reads are shown under the assumption that the libraries have been sequenced to relatively complete depth. Each track represents an individual technical replicate. Data for standard ATAC-seq from Buenrostro et al. 2013[1]. Region shown represents (a) chr19:36,097,939-36,272,939 or (b) chr9:36,373,966-37,494,500. (**c**) The total raw reads, filtered reads, and the percent duplication (how completely the libraries have been sequenced) are shown for each library to the right. (**d,e**) Sequencing tracks shown in (a,b) but with a y-axis scale of 5 to show the difference in signal-to-background in data derived using the Omni-ATAC protocol compared to the standard ATAC-seq protocol. Peaks that pass above the 5 read y-axis limit are truncated at y=5. Each track represents an individual technical replicate. Region shown represents (d) chr19:36,097,939-36,272,939 or (e) chr9:36,373,966-37,494,500. (**f**) Fraction of reads in peaks mapping to TSSs (+/- 500 bp of TSS) and distal elements (>500 bp from TSS) from the libraries shown in (a-e). Each bar represents the mean of 3 technical replicates. Error bars represent standard deviation. ***p<0.001 by two-tailed unpaired students t-test. Data for standard ATAC-seq from Buenrostro et al. 2013[1]. (**g**) Metaplot showing

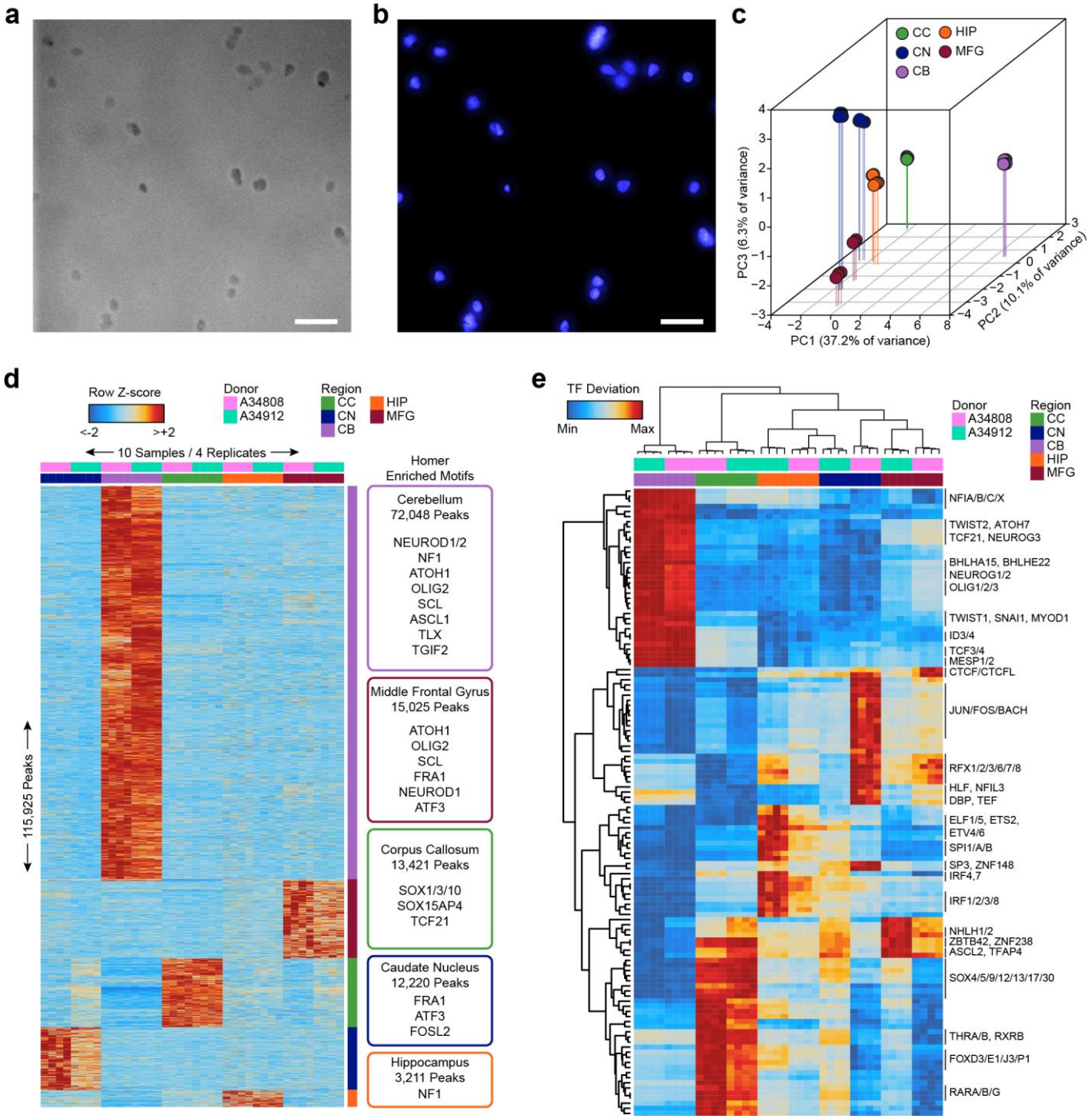
enrichment of reads at TSSs in data derived from 500 GM12878 cells using Omni-ATAC (N=3) compared to published data using standard ATAC-seq[1] (N=3). (**h**) Overlap of 500 cell GM12878 peaks called from all available reads using the standard ATAC-seq protocol and the Omni-ATAC protocol. Numbers represent the mean of three individual technical replicates in each case. Overlap was performed with a union set of GM12878 peaks identified by DNase-seq, standard ATAC-seq, and Omni-ATAC. Data for standard ATAC-seq from Buenrostro et al. 2013[1]. (**i**) Correlation of Omni-ATAC data from 500 GM12878 cells with Omni-ATAC data from 50,000 GM12878 cells. Each dot represents an individual peak. Peaks analyzed were derived from a union peak set using data from the standard ATAC-seq method, DNase-seq, and Omni-ATAC. Only 50,000 random peaks with more than 5 reads are shown. R value indicated at the top of the plot. (**j**) Pearson correlation heatmap showing sample by sample unsupervised clustering on all peaks identified across all technical replicates of 500 and 50,000 GM12878 cells. Data for standard ATAC-seq for both 500 cell and 50,000 cell GM12878 data from Buenrostro et al. 2013[1]. DNase-seq was performed on an unknown number of cells.



Supplementary Figure 5

Omni-ATAC enables the use of less amounts of transposase without decreasing data quality.

(**a**) Heatmap-based representation of ATAC-seq quality control metrics including library size, percent of reads mapping to mitochondrial DNA, and enrichment of signal at TSSs. Deeper color is used to depict the most desirable value of each statistic. All values were determined from 5 million random aligned reads. Data represents two technical replicates per square. (**b-d**) Correlation of Omni-ATAC data generated with varying amounts of transposase enzyme. Technical replicates using (b) 100% Tn5 transposase input and (c) 20% transposase input are shown alongside (d) the correlation of 100% Tn5 transposase input with 20% Tn5 transposase input. Peaks analyzed were derived from DNase-seq data. Only 50,000 random peaks with more than 5 reads are shown. R value indicated at the top of each plot. (**e**) Fraction of reads in peaks mapping to TSSs (+/- 500 bp of TSS) and distal elements (>500 bp from TSS) for Omni-ATAC data performed with various amounts of Tn5 transposase input. Peak set used was derived from DNase-seq data. Each bar represents the mean of 2 technical replicates. Error bars represent standard deviation. NS = not significant.

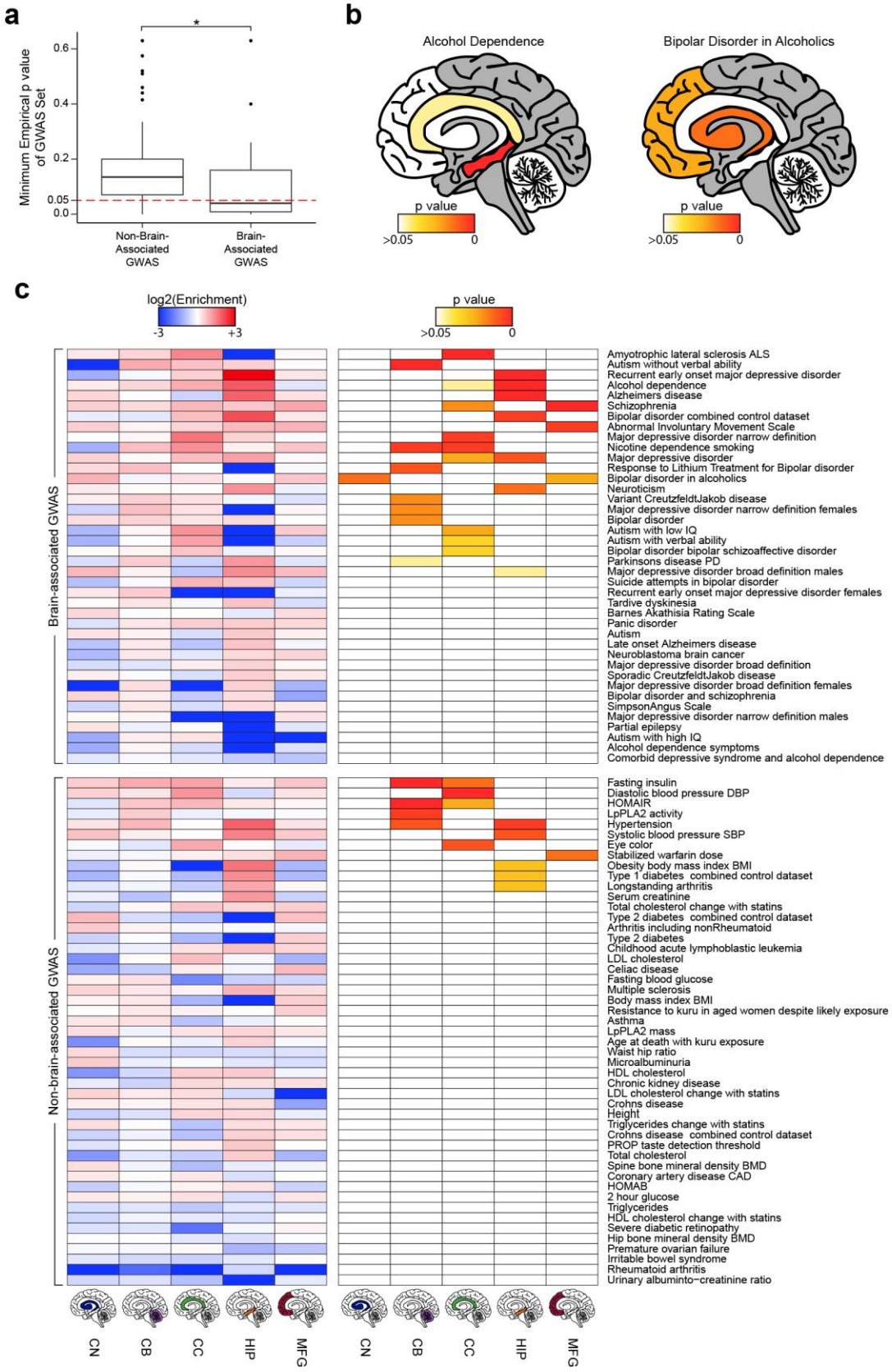


Supplementary Figure 6

Application of Omni-ATAC to frozen post-mortem human brain samples yields robust data.

(**a**) Representative phase contrast image of nuclei isolated from frozen post-mortem human brain. Scale bar represents 10 µm. Experiment was repeated 5 times with 5 total images collected. (**b**) Representative DAPI staining of the nuclei shown in (a). Scale bar represents 10 µm. (**c**) Three-dimensional principal component analysis showing the first 3 principal components from ATAC-seq data generated with the Omni-ATAC protocol in 5 different brain regions from 2 different biological donors, each with 2 technical replicates.

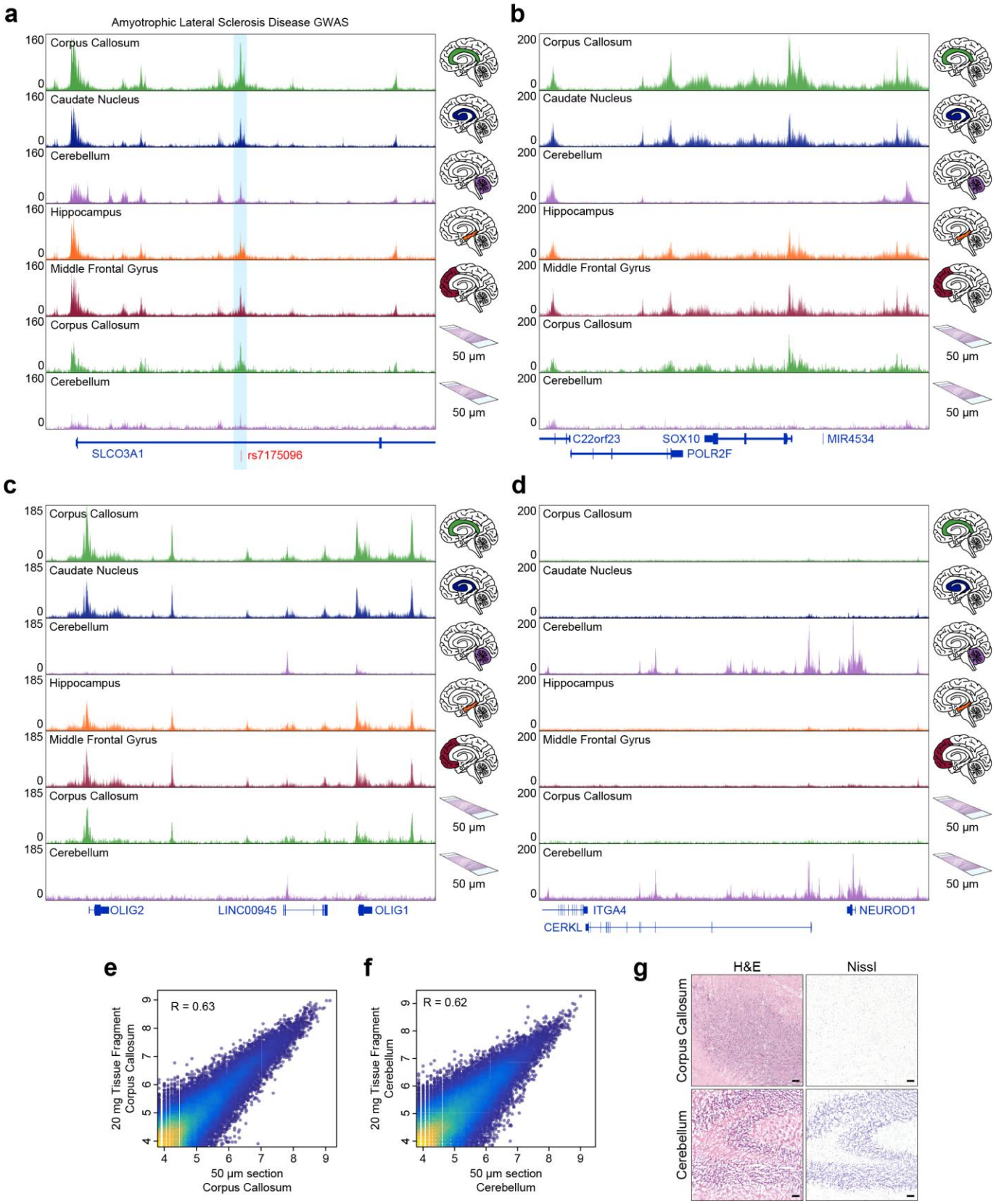
Different brain regions are indicated by color. CC = corpus callosum; CN = caudate nucleus; CB = cerebellum; HIP = hippocampus; MFG = middle frontal gyrus. (**d**) Differentially-accessible peaks defined as being at least 2 standard deviations away from the mean of all other regions. Transcription factors whose binding motifs are identified as enriched using the Hypergeometric Optimization of Motif Enrichment (Homer) algorithm in the accessible chromatin regions of each brain region are shown to the right. (**e**) Transcription factor deviation of all technical replicates and biological donors. TFs enriched in each cluster are shown to the right. Color scale represents the minimum and maximum of each row. Each row represents an individual TF motif. Each column represents an individual replicate of ATAC-seq using the Omni-ATAC protocol.



Supplementary Figure 7

GWAS analysis of ATAC-seq data from different brain regions provides key insights into the ontogeny of various brain diseases.

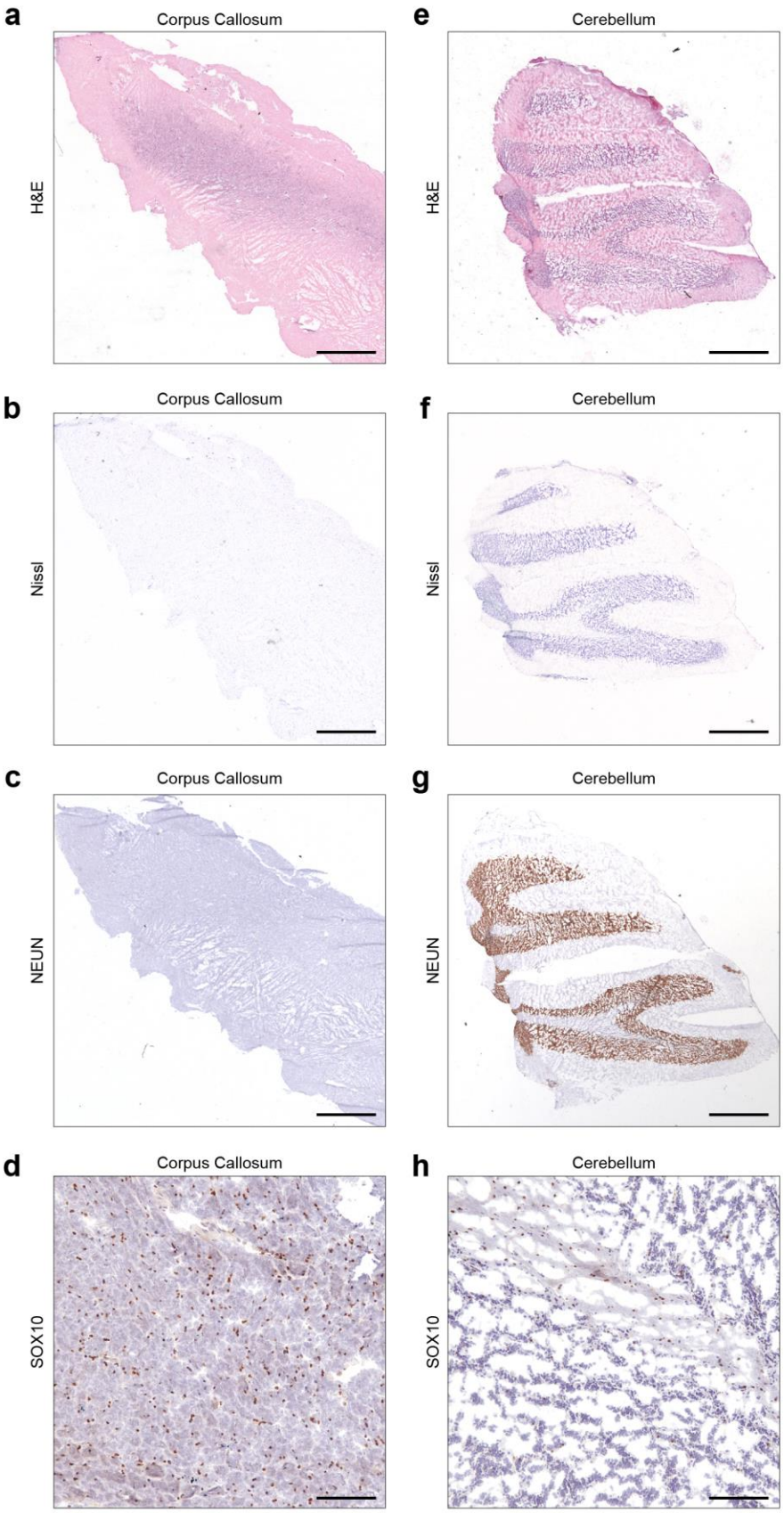
(**a**) Box plot of the minimum empirical p value from all brain regions for each GWAS SNP set assayed. *p < 0.05 by two-tailed unpaired students t-test. (**b**) Significance of enrichment of disease-specific GWAS polymorphisms in the uniquely-accessible regions of the 5 different brain regions shown in Fig. 2b. The empirical p value is depicted colorimetrically with reference to association-based permutations of the GWAS SNPs. (**c**) Heatmap representation of the enrichment (left) and significance (right) of the enrichment of GWAS SNPs in the uniquely-accessible regions of each brain region. For enrichment, 0 values are represented colorimetrically the same as a log2(enrichment) of -3. Regions without significant associations are shown as white in the p value heatmap.



Supplementary Figure 8

The Omni-ATAC protocol enables generation of robust ATAC-seq data from 50-μm-thick frozen tissue sections.

(**a**) Depth-normalized sequencing tracks surrounding amyotrophic lateral sclerosis disease-associated SNP rs7175096. Tracks were normalized according to the total number of reads in peaks. Each individual replicate (N=4 technical replicates for each of 2 donors per tissue fragment, N=1 technical replicate for each of 2 donors for corpus callosum sections, and N=2 technical replicates for each of 2 donors for cerebellum sections) is displayed with 25% transparency to show consistency among technical replicates and biological donors. Region shown represents chr15:92,390,775-92,470,760. Data derived from 20 mg tissue fragments and 50 μm tissue sections are represented by a cartoon image to the right. (**b-d**) Depth normalized sequencing tracks from all replicates from all regions comparing ATAC-seq data generated using the Omni-ATAC protocol in 50 μm tissue sections and larger 20 mg tissue fragments. Each individual replicate (N=4 technical replicates for each of 2 donors per tissue fragment, N=1 technical replicate for each of 2 donors for corpus callosum sections, and N=2 technical replicates for each of 2 donors for cerebellum sections) is displayed with 25% transparency to show consistency among technical replicates and biological donors. Regions shown represent (c) chr2:182,378,211-182,585,768, (d) chr21:34,391,306-34,455,298, and (e) chr22:38,347,790-38,398,955. (**e-f**) Correlation of ATAC-seq data from 50 μm sections of (f) corpus callosum or (g) cerebellum with ATAC-seq data generated from a nearby 20 mg fragment of the same brain region from the same donor. Each dot represents an individual peak from a union set of peaks called from all technical replicates and biological donors of data derived from tissue fragments from the same brain region. Only 50,000 random peaks with more than 5 reads are shown. (**g**) Histology and immunohistochemistry of 5 μm frozen sections from cerebellum and corpus callosum immediately adjacent to the 50 μm section used for ATAC-seq shown in Figure 2d-e. Stains from left to right: hematoxylin and eosin (H&E), Nissl. Scale bar represents 100 μm in each image.


a
Corpus Callosum
H&E
e
Cerebellum
H&E
b
Corpus Callosum
Nissl
f
Cerebellum
Nissl
c
Corpus Callosum
NEUN
g
Cerebellum
NEUN
d
Corpus Callosum
SOX10
h
Cerebellum
SOX10

Supplementary Figure 9

Histology of tissue immediately adjacent to a 50-µm section used for ATAC-seq.

(**a,e**) Hematoxylin and eosin staining of corpus callosum and cerebellum. Scale bar represents 1 mm. (**b,f**) Nissl staining of corpus callosum and cerebellum. Scale bar represents 1 mm. (**c,g**) Anti-NEUN immunohistochemistry of corpus callosum and cerebellum. Scale bar represents 1 mm. (**d,h**) Anti-SOX10 immunohistochemistry of corpus callosum and cerebellum. Scale bar represents 100 µm. All images represent same tissue shown in Figure 2d.

SUPPLEMENTARY NOTE 1

ATAC-seq Quality Assessment

Cumulative experience has demonstrated that the original ATAC-seq method performs sub-optimally on certain cell types and cell contexts. The primary metrics used to determine the quality of ATAC-seq data are (i) enrichment of signal at transcriptional start sites (TSSs), (ii) a low percentage of reads mapping to non-chromatinized mitochondrial DNA, (iii) high library complexity, and (iv) an insert size distribution that follows a nucleosomal spacing periodicity. In this work, we also use the fraction of reads in peaks as a measure of data quality; however, we note that this requires a priori knowledge of peak locations and is correlated with the TSS enrichment score. We also note that the value of the TSS enrichment score is dependent on the set of TSSs that is used. We have chosen to use RefSeq-defined TSSs as downloaded from the UCSC Genome Browser.

Rationale and Efficacy of Chosen Optimizations

The individual optimizations made to the original ATAC-seq protocol that comprise the Omni-ATAC protocol are as follows:
(1) The addition of 0.1% Tween-20 and 0.01% digitonin to the NP40-based lysis buffer
(2) Washing lysed cells in a resuspension buffer containing 0.1% Tween-20
(3) The addition of PBS, 0.1% Tween-20, and 0.01% digitonin to the transposition mix
These optimizations were rationally designed to improve the overall quality (as described above) of the ATAC-seq data and improve the breadth of applications to which ATAC-seq can be applied without individualized protocol alterations.

The addition of Tween-20 and digitonin to the NP40-based lysis buffer (i) leads to a reduction in mitochondrial reads and an increase in library complexity and (ii) enables lysis in a wider array of cell types. Washing of lysed cells in resuspension buffer containing Tween-20 on its own does not improve data quality, but in the context of lysis with Tween-20, digitonin, and NP40, further increases the library complexity and further reduces reads from mitochondrial DNA. The addition of PBS to the transposition reaction on its own increases the enrichment of signal at TSSs by 48% on average (Supplementary Fig. 1a).

Collectively, these optimizations (i) improve ATAC-seq data quality in cell types and cell contexts that were already tractable with the standard ATAC-seq protocol and (ii) make ATAC-seq possible in cell types and cell contexts that were not possible with the standard ATAC-seq protocol.

Decreasing Transposase Concentration in the ATAC-seq Reaction

We have found that the amount of Tn5 transposase added per ATAC-seq reaction can be reduced without changes in data quality. Titration of Tn5 input from 100 nM to 10 nM in mouse embryonic stem cells (mESCs) showed no appreciable change in fragment enrichment at open chromatin or total fragment yield, except in the lowest Tn5 concentration of 10 nM, which shows a reduction in library size (Supplementary Fig. 5a-e and Supplementary Table 1). Thus, the Omni-ATAC protocol can provide high-quality ATAC-seq data with a reduced per-reaction cost. However, we note that this was performed using homemade Tn5 transposase that may have different properties than Tn5 transposase purchased from Illumina. Additionally, transposase requirements may vary between different cell types and we have only tested mouse embryonic stem cells here. Aside from

Supplementary Fig. 5, all ATAC-seq reactions performed in this manuscript used the standard 2.5 µl Tn5 transposase per 50 µl ATAC-seq reaction. We encourage users to test the result of reducing Tn5 transposase concentration in their system of interest.