Practical Guidelines for High-Resolution Epigenomic Profiling of Nucleosomal Histones in Postmortem Human Brain Tissue

Supplemental Information

Glossary

BAM: sequencing data (typically derived from FASTQ files) are stored in compressed indexed binary forms

Bigwig: A file that allows presentation of genomics data in a genome browser, to ‘zoom out’ to single chromosome screenview or ‘zoom’ in at base pair resolution at any genomic locus of interest

BWA: Burrows-Wheeler Aligner (to align short reads of sequenced DNA to the genome)

ChIP-qPCR: chromatin immunoprecipitation followed by quantification of DNA in the immunoprecipitate via quantitative real-time polymerase chain reaction

ChIP-seq: chromatin immunoprecipitation followed by deep sequencing of DNA in the immunoprecipitate

CommonMind Consortium: An industry/NIH funded collaborative project to sequence the transcriptome from postmortem brain from hundreds of cases with schizophrenia and a similar number of controls (www.synapse.org/cmc)

DLPFC: dorsolateral prefrontal cortex

FASTQ: An ASCII text file informing about the nucleotide sequence plus some measures on sequence call confidence

FASTQC: A modular set of analyses used to provide a simple and fast initial check of the quality of high throughput sequencing data

Flow cell: A device used in deep sequencing technology, enabling massively parallel sequencing of many DNA molecules. Each molecule is spatially separated from other molecules, amplified and sequenced.

GATK: Genome Analysis Toolkit, in order to assess sequence variants, including single nucleotide polymorphisms and small insertions or deletions in deep sequencing datasets (www.broadinstitute.org)

Integrative Genomics Viewer (IGV): A tool to visualize and integrate a range of genomics data in browser-like fashion

MNAse: micrococcal nuclease, an enzyme that digests non-nucleosomal DNA

Mono-nucleosome: the elementary unit of chromatin, comprised of the core histone octamer (histones H3, H4, H2A, H2B) and 147 base pairs of DNA wrapped around it in 2.5 loops

NRF (nonredundant fraction): a measure of sequencing library complexity, representing the fraction of reads that remain after removing duplicates

Paired end sequencing: DNA fragments are sequenced from both ends (sense and antisense strand, in 5' to 3' direction on each strand)

PBC (PCR bottleneck coefficient): The PCR bottleneck coefficient is the fraction of regions that have been sequenced and contain no duplicates

PsychENCODE: A National Institutes of Mental Health-funded consortium to map epigenomes in ~1000 postmortem brain samples, including subjects diagnosed with schizophrenia, bipolar disorder and autism spectrum disorder, and additional cell lines and tissues from disease and control cases (www.synapse.org/pec)

RNA-seq: RNA extracted from tissue is reversed transcribed into cDNA followed by deep sequencing

Roadmap Epigenomics Project: A National Institutes of Health-funded consortium to provide a public resource and database to chart the transcriptome and epigenome (including DNA methylation and histone modifications) in various human tissues across different stages of development (www.roadmapepigenomics.org)

RSC (relative strand correlation): A quality metric used to assess clustering of immunoprecipitated DNA at specific locations in the genome

SAM: sequencing data (typically derived from FASTQ) are stored in tab delimited ASCII columns

Single end sequencing: DNA fragments are sequenced from one (of the two 5' strand) ends only, thus from a single strand

Superenhancer: A block of sequences, on average 8 kb wide, with broad stretches of open chromatin and clustered transcription factor occupancies, functioning as regulatory element to drive expression of a target gene potentially many kilobases further up- or downstream of the superenhancer

VCF: variant call format. A standardized text file used by the GATK for variant call (gatkforums.broadinstitute.org)

Histone H3 Array

	1	2	3	4	5	6	7	8	9	10	11	12	
A	H3 1-19 unmod	H3 1-19 R2me1	H3 1-19 R2me2a	H3 1-19 R2me2s	H3 1-19 T3P	H3 1-19 K4ac	H3 1-19 K4me1	H3 1-19 K4me2	H3 1-19 K4me3	H3 1-19 R8me1	H3 1-19 R8me2a	H3 1-19 R8me2s	**100 ng**
B	H3 1-19 unmod	H3 1-19 R2me1	H3 1-19 R2me2a	H3 1-19 R2me2s	H3 1-19 T3P	H3 1-19 K4ac	H3 1-19 K4me1	H3 1-19 K4me2	H3 1-19 K4me3	H3 1-19 R8me1	H3 1-19 R8me2a	H3 1-19 R8me2s	**10 ng**
C	H3 1-19 K9ac	H3 1-19 K9me1	H3 1-19 K9me2	H3 1-19 K9me3	H3 1-19 S10P	H3 1-19 T11P	H3 7-26 unmod	H3 7-26 K14ac	H3 7-26 R17me1	H3 7-26 R17me2a	H3 7-26 R17me2s	H3 7-26 K18ac	**100 ng**
D	H3 1-19 K9ac	H3 1-19 K9me1	H3 1-19 K9me2	H3 1-19 K9me3	H3 1-19 S10P	H3 1-19 T11P	H3 7-26 unmod	H3 7-26 K14ac	H3 7-26 R17me1	H3 7-26 R17me2a	H3 7-26 R17me2s	H3 7-26 K18ac	**10 ng**
E	H3 16-34 unmod	H3 16-34 K23ac	H3 16-34 K27ac	H3 16-34 K27me1	H3 16-34 K27me2	H3 16-34 K27me3	H3 16-34 R26me1	H3 16-34 R26me2a	H3 16-34 R26me2s	H3 16-34 S28P	H3 26-44 unmod	H3 26-44 K36ac	**100 ng**
F	H3 16-34 unmod	H3 16-34 K23ac	H3 16-34 K27ac	H3 16-34 K27me1	H3 16-34 K27me2	H3 16-34 K27me3	H3 16-34 R26me1	H3 16-34 R26me2a	H3 16-34 R26me2s	H3 16-34 S28P	H3 26-44 unmod	H3 26-44 K36ac	**10 ng**
G	H3 26-44 K36me1	H3 26-44 K36me2	H3 26-44 K36me3	H3 26-44 Y41P	H3 47-65 unmod	H3 47-65 K56ac	H3 71-89 unmod	H3 71-89 K79me1	H3 71-89 K79me2	H3 71-89 K79me3	Rat IgG **10 ng**	Sheep IgG **10 ng**	**100 ng**
H	H3 26-44 K36me1	H3 26-44 K36me2	H3 26-44 K36me3	H3 26-44 Y41P	H3 47-65 unmod	H3 47-65 K56ac	H3 71-89 unmod	H3 71-89 K79me1	H3 71-89 K79me2	H3 71-89 K79me3	Mouse IgG **10 ng**	Rabbit IgG **10 ng**	**10 ng**

Figure S1. Histone peptide array

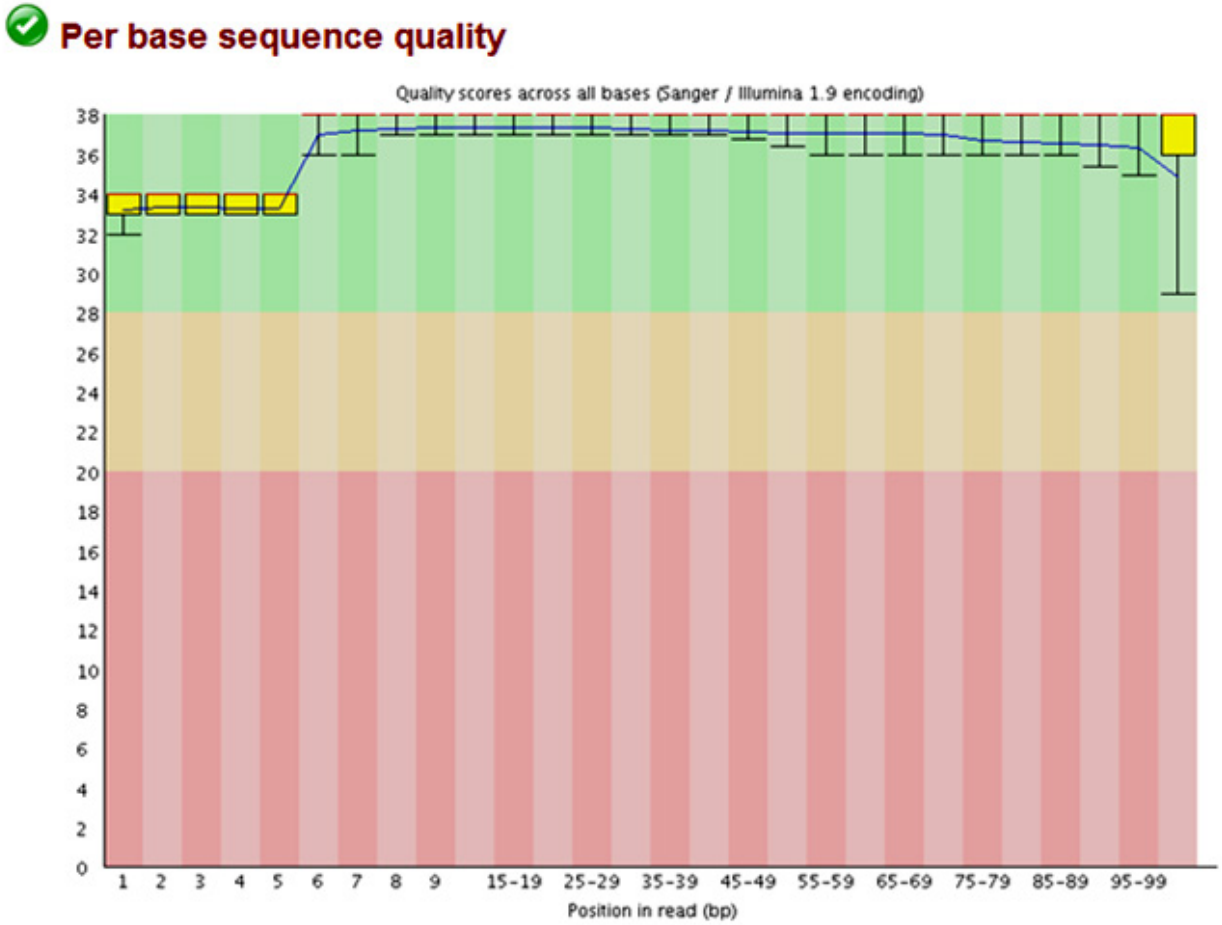

Figure S2. FASTQC analysis of the ChIP-seq data. FastQC analysis provides a simple way to do some quality control checks on raw sequencing data coming straight from the sequencer, before performing more specific data processing. Here we present one example of the FASTQC data output, which shows an overview of the range of quality values (Y axis) across all bases at each position in the FastQ file (X axis). The plot has the following elements: The blue line represents the mean quality. The red line is the median value. The yellow box represents the inter-quartile range (25-75%). The upper and lower whiskers represent the 10% and 90% points.

The y-axis shows the quality scores and the higher the score the better the base call. More specifically, the graph background colors separate the y axis into very good quality calls (green, score > 28), calls of reasonable quality (orange, score = 20-28), and calls of poor quality (red, score < 20). Note that our representative ChIP-seq data shows remarkably high quality throughout the entire sequence, including the sequence towards the end of the read (up to 100 bp).

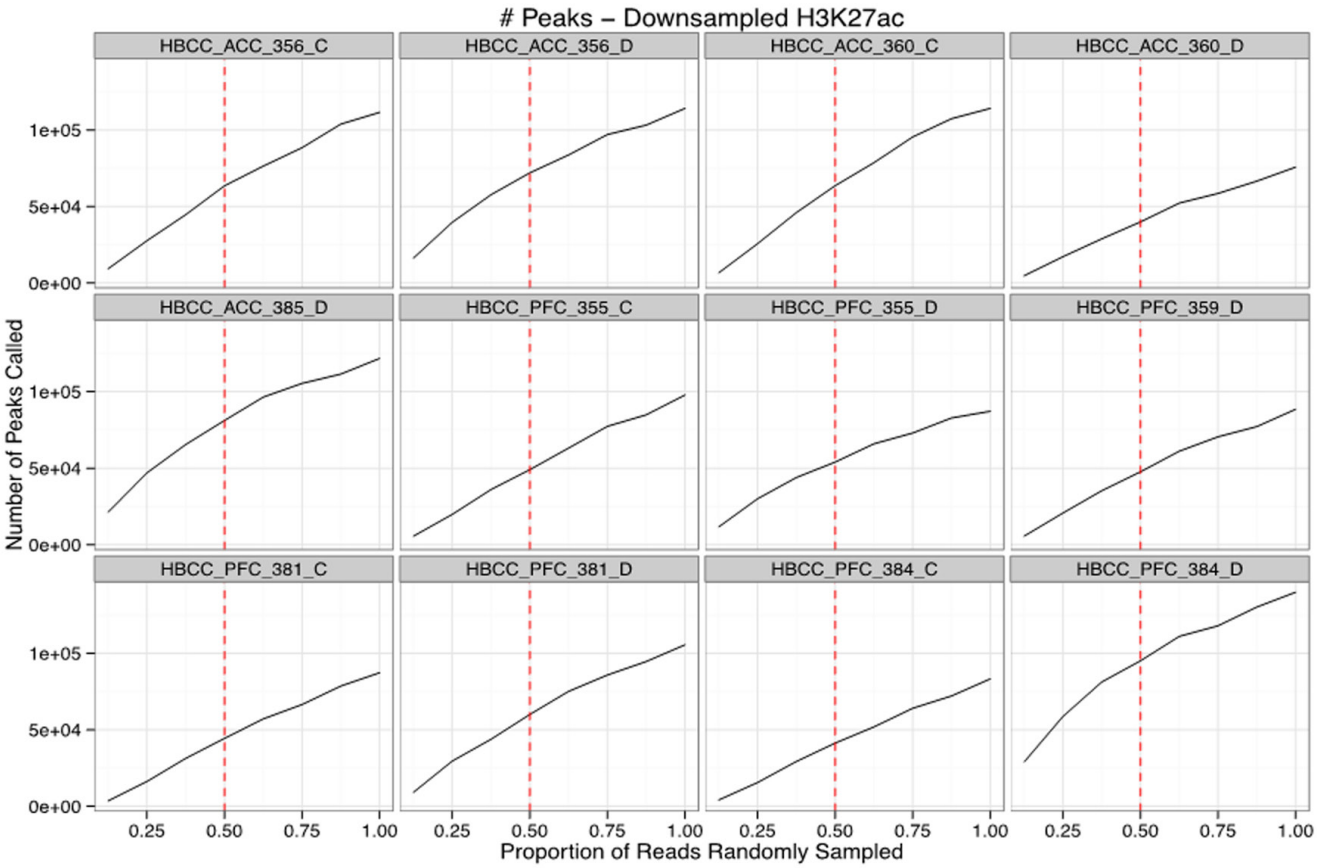

Figure S3. In silico downsampling of the filtered BAM files were performed at various proportions of the number of reads in the BAM file (12.5%, 25%, 37.5%, 50%, 62.5%, 75% and 87.5%) and peaks were then called using MACS2 identically to how they were called for main BAM file. Shown above are the number of peaks called in each of the downsamples, separately for each sample and each histone mark.

Table S1. Specimen information and ChIP-seq data

Brain	Institution ID	Gender	Ethnicity	Age of Death	PMI (in hours)	pH	Dx
A	HBCC-PFC-361	Male	Caucasian	59	22	6.35	Control
B	HBCC-PFC-369	Male	Caucasian	40	44.5	6.51	Control

Brain	BMI (Traditional)	Height (Inches)	Weight (pounds)	Brodmann Area	Hemisphere	Estimated gray matter (mg)
A	29.66465981	68.5	198	9_10	right	492
B	33.5701308	72.5	251	9_10	right	545.175

Brain	NeuN	# of nuclei	Input DNA (ng, 10%)	Histone mark	Histone Antibody catalog #
				H3K4me3	Cell Signaling Cat # 9751
	Pos	3.71E+06	334.8	H3K27ac	Active Motif Cat # 39133
				H3K4me3	Cell Signaling Cat # 9751
A	Neg	2.46E+06	214.2	H3K27ac	Active Motif Cat # 39133
				H3K4me3	Cell Signaling Cat # 9751
	Pos	3.99E+06	296.1	H3K27ac	Active Motif Cat # 39133
				H3K4me3	Cell Signaling Cat # 9751
B	Neg	4.43E+06	191.7	H3K27ac	Active Motif Cat # 39133

Brain	ChIP DNA(ng)	Barcode sequence	Library DNA (ng)	Sequencing Submission #	Sample ID	Yield (Mbases)
	14.85	ACTTGA	123.6	B749	HBCC_PFC_361_A	4509
	14.625	CTTGTA	74.4	B753	HBCC_PFC_361_C	5471
	7.245	TTAGGC	101.6	B752	HBCC_PFC_361_B	5049
A	9.585	TAGCTT	122	B754	HBCC_PFC_361_D	7580
	9.135	GATCAG	37.6	B748	HBCC_PFC_369_A	6142
	20.385	TTAGGC	77.2	B754	HBCC_PFC_369_C	8077
	2.3085	ATCACG	89.2	B749	HBCC_PFC_369_B	3357
B	9.765	ACTTGA	80.8	B754	HBCC_PFC_369_D	9010

Brain	% PF	# Reads	% of raw clusters per lane	% Perfect Index Reads	% One Mismatch Reads (Index)	% of >= Q30 Bases (PF)	Mean Quality Score (PF)
	100	4.51E+07	13.67	98.9	1.1	94.01	36.16
	100	5.47E+07	20.04	99.34	0.66	90.97	34.96
	100	5.05E+07	15.27	99.05	0.95	93.45	35.95
A	100	7.58E+07	21.85	98.71	1.29	94.45	36.31
	100	6.14E+07	18.21	98.78	1.22	85.02	32.94
	100	8.08E+07	23.28	98.94	1.06	94.18	36.21
	100	3.36E+07	10.18	98.86	1.14	92.1	35.62
B	100	9.01E+07	25.96	99.27	0.73	94.17	36.22

Table S2. ChIP-seq quality metrics

NSC		RSC	
< 1.05	0	< 0.8	18
≥ 1.05	127	≥ 0.8	109

Supplemental Methods

Postmortem Tissue

Samples for this study were collected at the National Institute of Mental Health Human Brain Collection Core (HBCC; http://www.nimh.nih.gov/hbcc) through the Offices of the Chief Medical Examiner of the District of Columbia and of Northern and Central Virginia, after autopsy. Informed consent was obtained from the surviving next-of-kin, according to Protocol #90-M-0142 approved by the Combined Neuroscience Institutional Review Board and only with the permission of next-of-kin. A telephone interview with the next-of-kin to gather basic demographic information and medical, substance use, and psychiatric history was conducted within 1 week of tissue donation. Clinical characterization, neuropathological screening, toxicological analyses, and dissection of the dorsolateral prefrontal cortex (DLPFC) were performed as previously described (e.g., (1)). Samples for the study were dissected at the HBCC from either right or left hemisphere of fresh frozen coronal slabs cut at autopsy from the DLPFC within cortical areas Brodmann 9/10. Immediately after dissection, tissue blocks were cooled to -78.5°C and transferred on dry ice to Icahn School of Medicine at Mount Sinai for nuclei isolation and ChIP-seq. The entire pipeline, including bioinformatical analyses, has so far been completed for more than 100 ChIP-seq samples. Here we show a representative example of n = 8 ChIP-seq datasets from two PFC specimens, each processed for $NeuN^{+}$H3K4me3, $NeuN^{-}$H3K4me3, $NeuN^{+}$H3K27ac, and $NeuN^{-}$H3K27ac (Figures 3-4, Table S1).

Library Preparation Procedure

Libraries are usually prepared in batches of 8 samples, using KAPA Hyper Prep Kit (KK8502) and BIOO Scientific adapters (Catalog #514102). For H3K4me3 and H3K27ac, the starting amount of ChIP DNA varies from <2 ng to 20 ng. Although this is considered a low-input library preparation, our protocol performs excellently in this range with more than 99% of libraries passing the final QC and yielding data of good quality. The amount of DNA used for library prep from the input DNA for ChIP is 4 ng. After each step of the procedure, DNA is purified using AMPure beads (SPRI select) and final library size selection (200-350 bp) is performed using Pippin Prep (Sage Science). In general, 0.8X or 1.0X AMPure bead clean-up can also be used for lower size selection, particularly for libraries derived from sonicated DNA samples. The presence of adapter-dimers (125 bp) in the library is a concern since the dimers amplify and are sequenced more efficiently than the larger major library products, resulting in uninformative sequencing reads. Although AMPure beads are a very convenient way to remove adapter dimers, we have noticed that in the 0.8-1.2X bead clean-up process, there can be significant loss of the main library fragment (275 bp) as well. Therefore, we regularly use 1.8X SPRI clean-up (retaining all DNA fragments above 100 bp) and perform the final size selection using automated gel size selection apparatus (Pippin prep) which removes all the adapter dimers and retains the majority of the 275 bp library. Using Pippin prep also ensures a consistency in the library preparation.

In the ligation step, the addition of the adapters containing different barcodes allows pooling and sequencing of the multiple libraries in the same lane of the Illumina flow cell. The sequencing pooling strategy should be decided before the library preparation (based on the sequencing depth needed for adequate coverage and will be specific for each mark and chemistry version) as each pooled library should have a unique barcode and the barcode pools should be chosen based on the manufacturer's recommendations to allow for most efficient sequencing. In the majority of cases, 12 PCR cycles produce sufficient amount of library DNA

for sequencing (at least 2 ng/µl in 20 µl of the final library volume). If the input ChIP DNA is very low (<2 ng DNA or undetectable on Qubit), the number of cycles should be increased to 15.

Next Generation (Deep) Sequencing

Our samples are sequenced on HiSeq 2500 Illumina instruments using V3 or V4 Illumina chemistry. The following options should be considered when designing a deep sequencing experiment: sequencing depth and multiplexing, read length, and whether to perform single or paired-end sequencing. In addition, the development of an appropriate analytic pipeline is essential for the efficient processing and QC of the sequencing data. Whenever possible, batch effects should be controlled for at both the experimental and computational level.

Sequencing Depth and Multiplexing

The number of ChIP-seq peaks (i.e., the histone modification enrichment regions) will be dependent on the abundance of a histone mark and number of sequencing reads. Ideally, sequencing depth should allow for the discovery of the majority of enrichment sites, while at the same time the number of reads should be kept to a minimum to save costs. For sequencing depth, according to the ENCODE guidelines, histone marks enriched at specific locations in the genome and defined by sharp, narrow peaks (such as H3K4me3, mainly associated with active promoters), 20 million single end (SE) or 40 million paired end (PE) reads should be sufficient. For more broadly distributed marks, a minimum of 40 million SE or 80 million PE reads is required to identify the majority of the enrichment sites across the genome. In our study, we aimed for 40 million uniquely mapped 100 base pair PE reads for H3K4me3 and 80 million uniquely mapped PE reads for H3K27ac. To achieve this, samples were sequenced in batches of eight (for H3K4me3) or four (for H3K27ac) samples per flow cell lane. As shown in Figure 3B, using this strategy, we achieved the expected sequencing depth for the majority of the samples. Importantly, ENCODE criteria consider H3K27ac a narrow mark and recommends the same

number of sequencing reads as for H3K4me3 (40 million PE). We performed a random downsampling of our H3K27ac ChIP-seq samples and compared the number of peaks that we retrieved with different sequencing depths (10-100% of the original depth, Figure S3). This analysis showed that, at least for the adult brain samples, the number of peaks gained for H3K27ac in this sequencing range is still linear (though, as expected, the additional peaks have moderately lower fold changes and statistical significance). Therefore, we do not rule out that, at least for some of the histone modifications, current sequencing depth guidelines may underestimate the actual number of peaks, at least in brain. Sequencing options also include different read length and SE vs. PE sequencing. Currently, most frequently used read lengths are 50 bp, 75 bp, and 100 bp, while mappability (including of potential sequence variants due to single nucleotide polymorphisms or small insertions and deletions) has been shown to improve with both increased read length and with PE sequencing (2).

Furthermore, it is noteworthy but unsurprising that in our pipeline, the average number of reads from $NeuN^+$ ChIP-seq only showed minimal differences from $NeuN^-$ (with N = 136 libraries sequenced to date, we observed the $N_{(average)}$ reads for H3K27ac, $NeuN^+$ 112,533,170, $NeuN^-$ 120,169,901; H3K4me3 $NeuN^+$, 58,101,963; NeuN- 57,383,033. H3K27ac may show more variability in numbers of reads because the H3K27ac samples were sequenced to a higher depth.

Sequencing Data Processing and QC

The initial processing and QC of the sequencing data includes the following steps: FASTQC analysis, alignment to the reference genome, peak calling, annotation analysis, and data visualization (Figure 2G) (3). FASTQC analysis is a fast and simple way to initially assess the quality of raw sequencing data (usually using FASTQ files) coming straight from the sequencer before doing any specific data processing (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) (Figure 2H, Figure S2). FASTQC

provides a modular set of analyses which include: the analysis of Per Base Sequence Quality, Per Sequence Quality Scores, Per Base Sequence Content, Per Sequence GC Content, Per Base N Content, Sequence Length Distribution, Duplicate Sequences, Overrepresented Sequences, Adapter Content, Kmer Content, and Per Tile Sequence Quality. The FASTQC data outputs can be used to get a quick impression of whether the data have any problems of which one should be aware before pursuing further analysis (an example of the FASTQC plot for per Base Sequence Quality is given in Figure 2H and Figure S2). Following the initial quality control, our data processing pipeline (Figure 3H) begins with paired FASTQ files aligned to the Human Genome (version HG19) with BWA mem (4) (version 0.7.8) generating SAM files, which are then converted to indexed and sorted BAM files. Picard (version 1.112) (http://broadinstitute.github.io/picard) MarkDups is used to mark duplicates in the BAM files and multi-mapped reads and improperly paired reads were filtered out with 'samtools (5) view -f 2 -F 2828 -q 1' (version 1.1). Marking duplicates has been shown to be beneficial in removing noise in non-enriched regions, while not impacting the sensitivity of peak detection (6). MACS2(7) is used to call peaks from the duplicate marked, filtered BAM with a q-value threshold of 0.01. MACS2's advantages, including speed and ease of use, have made it into one of the most extensively used benchmarked peaks callers. We called narrow and broad peaks for both H3K4me3 and H3K427ac, as these are considered to be some of the more punctate chromatin marks. We then took gapped peaks as broad regions containing a narrow peak. This matches the ENCODE recommended gapped peak calling pipeline.

Bigwig files are generated from the bedgraph files output by MACS2 and used for visualization in the Integrative Genomics Viewer (8). We perform, whenever possible, a genotype check to verify sample identity across data files by comparing genetic variants called on the ChIP-Seq data to RNA-seq or genomic DNA (genotyping or sequencing), if available for that specimen. For each ChIP-Seq sample, variants are called using GATK best practices on the aligned reads. VCF files for each sample are compared to every other sample using the

gtcheck function of htslib (5). The distribution of discordant values between samples is analyzed to identify cases of non-self matching and lack of expected self-matching.

Batch Effects, Sample Randomization and Data Tracking

Epigenomic studies, similar to other studies that are based on high-throughput technologies, are complicated by batch effects including but not limited to changes in reagent lots and personnel (9). These will affect large-scale projects disproportionately, but will also affect smaller studies. Experimental and computational approaches should be implemented to address this. In order to avoid misinterpreting technical artifacts as biologically relevant, all experiments should be designed so that experimental conditions (e.g., disease status) are evenly distributed throughout batches and across the duration of the experimental processing timeline. If brains from multiple brain banks are being used in the study, the randomization should also take into account the brain bank as a possible confounding factor. In our experiments, we use multiple rounds of randomization in order to avoid correlation between disease status, brain bank, and processing order. Paired disease samples are randomly ordered for FACS sorting and ChIP. Samples are once again randomized before library preparation and are randomly assigned to flow cell and lane for sequencing. These randomization steps are intended to account for the variability in FACS sorting, ChIP assay, library preparation, and sequencing run (different flow cell or sequencing lane), which we have found can display batch effects. In addition to randomization, removal of technical artifacts by statistical modeling (e.g., logistic regression), to the degree possible, can increase power to detect biological differences. Therefore, a tracking system that captures information about variables throughout the study that may have an effect on the data such as date of sample processing, reagents and lots, personnel, instruments, facilities, etc., should be implemented so that relevant information is available for downstream data analysis. Because it is not feasible to predict or record all sources of technical variation in a study, additional statistical modeling — including principal component analysis or hierarchical

clustering, followed by adjustment for these unknown batch effects in downstream analyses — are often required to minimize their impact (9).

Detailed Protocol

Introductory Remarks:

The protocol described below should be applicable to ChIP-seq for any histone modification or histone variant for which an antibody with sufficient sensitivity and specificity is available. Previous work, using a relative small cohort with 32 brains, suggests that postmortem confounds, including tissue pH and postmortem interval, do not show strong correlations with numbers and proportions of specific reads in ChIP-seq libraries (10) at least with postmortem intervals ranging from 3-40 hours or tissue pH ranging from 6.1 to 6.7 (10). However, a more definitive answer on the role of postmortem confounds for the outcome and quality of ChIP-seq experiments will have to await studies in larger cohorts. Furthermore, the protocol has been tested for H3K4me3, H3K27ac, and H3K9me3 ChIP-seq but other types of ChIP-seq experiments may benefit from modifications of this protocol, including the addition of specific kinase or histone deacetylase inhibitors. Finally, the protocol below avoids formaldehyde-based fixation to maximize yields of sorted nuclei and specificity and efficacy of antibody pull-down (both of which could be diminished in input prepared by crosslinking) (11). However, sorting of fixed nuclei with cell type-specific markers for subsequent chromatin assays is feasible (12-14).

MATERIALS REQUIRED:

***ChIP**:*

1. Micrococcal Nuclease (MNase) (Sigma, Cat# N3755)
2. Anti-H3K4me3 antibody (Cell Signaling, Cat# 9751BC, lot 7)
3. Anti-H3K27ac antibody (Active Motif, Cat# 39133, Lot # 01613007)
4. A/G Magnetic beads (Thermo Scientific, Cat# 88803)
5. GlycoBlue (Ambion, Cat# AM9515)
6. Phenol:Chloroform:Isoamyl Alcohol (25:24:1) (Invitrogen, Cat# 15593-031)

Library Prep:

1. KAPA Hyper Prep Kit – KK8502 – KAPA Biosystems
   - KAPA End Repair & A-Tailing Buffer
   - KAPA End Repair & A-Tailing Enzyme
   - KAPA Ligation Buffer
   - KAPA DNA Ligase
   - KAPA HiFi HotStart ReadyMix
2. AMPure Beads – SPRIselect Part #B23317 – Beckman Coulter
3. Pippin Prep – CDF2010 – Sage Science
   - 2% DF Marker L Cassette
   - Marker L solution/marker mix
   - Electrophoresis buffer
4. BIOO Scientific - Catalog #514102
   - BIOO PCR Primer Mix (12.5 µM)
   - BIOO Adapters 1-12 (25 µM)
5. Elution Buffer (EB)
6. 80% Ethanol

EQUIPMENT REQUIRED:

1. Ultracentrifuge (Thermo Scientific Sorvall WX Floor Ultra Centrifuge 100,000 rpm) and rotor (Thermo Scientific SureSpin 630 (17 mL) Swinging Bucket Rotor)

2. BD™ FACSAria II (SOP) Cell Sorter, enclosed in a Baker BioPROtect® III Hood, equipped with either 355nm and/or 405nm laser to excite DAPI (Thermoscientific, Cat# 62248), and 488nm laser to excite Neuronal Nuclei NeuN-Alexa-488 conjugated antibody (Millipore, Cat# MAB377X)
3. 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA)
4. Qubit Fluorometer
5. Pippin Prep
6. Magnet for 1.5 ml tubes
7. Magnet for 0.2 ml tubes
8. Microfuge for 1.5 ml tubes - room temp
9. Microfuge for 1.5 ml tubes - chilled
10. Mini centrifuge with rotor for 8-well strips
11. Quantitative real-time PCR machine

BUFFERS:

Lysis Buffer

0.32M	Sucrose	5.47 g
5 mM	CaCl2	250 ul
3 mM	Mg(Ace)2	150 ul
0.1 mM	EDTA	10 ul
10mM	Tris-HCl, pH8	500 ul
1 mM	DTT	17 ul
0.1%	Triton X-100	50 ul

--- Adjust with ddH2O to 50 mL

Sucrose Solution

1.8 M	Sucrose	30.78 g
3 mM	Mg(Ace)2	150 ul
1 mM	DTT	17 ul
10 mM	Tris-HCl, pH8	500 ul

--- Adjust with ddH2O to 50 mL

Douncing Buffer (10mM Tris, 4mM MgCl2, 1mM CaCl2, pH 7.5)

10 mM Tris	0.242 g
4 mM MgCl2	0.163 g
1 mM CaCl2	0.03 g

--- Adjust pH to 7.5 and volume to 200 ml

10X FSB

50 mM EDTA	20 ml 500 mM EDTA, pH 8
200 mM Tris,	4.85 g
500 mM NaCl	5.84 g

--- Adjust pH to 7.5 and volume to 200 ml

Low Salt Buffer

0.1% SDS	2 ml 10% SDS
1 % Triton X-100	2 ml Triton X-100
2 mM EDTA	0.8 ml 500 mM EDTA, pH 8
20 mM Tris	4 ml 1M Tris-HCl, pH 8
150 mM NaCl	10 ml 3M NaCl

Adjust volume to 200 ml

High Salt Buffer

0.1% SDS	2 ml 10% SDS
1% Triton X-100	2 ml Triton X-100

2 mM EDTA 0.8 ml 500 mM EDTA, pH 8
20 mM Tris 4 ml 1M Tris-HCl, pH 8
500 mM NaCl 33 ml 3M NaCl
Adjust volume to 200 ml

TE Buffer
10 mM Tris 2 ml 1M Tris, pH 8
1 mM EDTA 0.4 ml 500 mM EDTA, pH 8
Adjust volume to 200 ml

Elution Buffer
0.1M NaHCO3 0.21 g NaHCO3
1% SDS 2.5 ml 10% SDS
Adjust volume to 25 ml

3M Sodium Acetate
3M sodium acetate 19.69 g anhydrous sodium acetate (M.W.=82.03)
Adjust pH to 5.2 with concentrated acetic acid and volume to 80 ml

0.1M PMSF (phenylmethanesulfonyl fluoride) (M.W. = 174.19)
0.017g/1ml isopropanol

3M 1,4-Dithio-DL-threitol (DTT) (M.W. = 154.3)
1g/2ml autoclaved ddH2O

I. Nuclei Isolation and ChIP Protocol

The protocol below yields, on average, 600,000 NeuN+ and 600,000 NeuN- nuclei per 100 mg cortical gray matter from a postmortem brain. For each ChIP, we used the minimum of 400,000 NeuN+ or NeuN- nuclei as starting material. On average, 2-4 µg of chromatin DNA was used for each ChIP, and the ChIP efficiency for H3K4me3 and H3K27ac was 0.3% and 0.5%, respectively.

DAY 1 : Nuclei Extraction, FACS, and ChIP-day 1

Douncing and extracting nuclei

1. Turn on ultracentrifuge and set temperature to 4°C and turn on the vacuum
2. Add 5 ml of lysis buffer to human tissue douncer
3. Add dissected human postmortem brain tissue (gray matter) into lysis buffer and dounce 50 times
4. Transfer tissue homogenate to ultracentrifuge tube
5. Add 9 ml of sucrose solution to the bottom of the tube, carefully not to disrupt layers in between
6. Process another sample and balance 2 tubes by weight
7. Ultracentrifuge at 24,000 rpm (106803.1x g RCF) for 1 hour at 4°C
8. While waiting put away solutions and clean douncer with bleach, rinse well with distilled water (4-5x) and final time with milliQ water

15 mins before ultracentrifuge is complete:

9. Label tubes: sample, DAPI-only control (stained with DAPI only, but no NeuN-Alexa 488)
10. Make two mixes, one for the sample and the other for DAPI-only control

    Sample mix (per sample): 500 µL of 1xDPBS
    10 µL of 10% BSA
    1 µL of anti-NeuN-Alexa 488 (Millipore, Cat# MAB377X)

    DAPI-only mix (per sample): 1000 µL of 1xDPBS
    10 µL of 10%BSA
11. Place all tubes on ice

After the sample is taken out of the ultracentrifuge

12. Vacuum away the supernatant including the interface between the sucrose solution and lysis buffer solution gradients
13. Add 500 µL of 1xDPBS in tube and let it sit on ice for 10 mins
14. Pipette up and down 50 times to re-suspend nuclei pellet on ice
15. Pipette 500 µL of nuclei solution into the sample mix from Step 10
16. Pipette 20 µL nuclei solution into DAPI-only control mix from Step 10
17. Incubate at 4°C, 45-60 mins, rotate and keep in dark
18. Add 1 µL of DAPI into all the samples before bringing samples to FACS
19. Bring samples to FACS facility

Turn on 28°C water bath, thaw MNase, PMSF, DTT, Benzamidin and place on ice, put 0.2mM EDTA on ice

FACS

20. Filter each sample through a 35um cell strainer
21. Lightly vortex sample to make mixture homogenous (not clumped)
22. Collection tubes should each contain 1mL DPBS prior to sorting
23. Load sample tube into chamber
24. Analyze sample and set appropriate gates to capture maximum nuclei while avoiding unnecessary debris.
    a. Gating Parameters (Y:X):

i. SSC-A:FSC-A (Size, Complexity, and Granularity Gates)
ii. FSC-W:FSC-A
iii. SSC-W:SSC-A
iv. Pacific Blue-W:Pacific Blue-A (Selecting for nuclei from non-dividing cells)
v. Pacific Blue-A:FITC-A (Alexa488 Expression Gate)

25. Sort entire volume of sample

After FACS

26. For 10 mL of FACS sample, add:
    2 mL of Sucrose solution
    50 µL of 1M CaCl2
    30 µL of 1M Mg(Ace)2
27. Invert sample several times and leave on ice for 15 minutes
28. Centrifuge at 4000 rpm for 5 mins at 4°C
29. Vacuum away supernatant
30. Add 300 µL of Douncing buffer, pipette up and down to re-suspend the pellet, and transfer it to o-ring tubes
31. Pre-warm the sample at 28°C for 5 minutes
32. Add MNase (Amount relative to number of nuclei present in sample - to be optimized for each lot of MNase, e.g. 0.5 µL of 0.2U/ µL MNase per 1 million nuclei)
33. Vortex and incubate at 28°C for 5 min
34. Add 30 µL of 0.5M EDTA (1:10) to stop reaction and place on ice
35. Add 1200 ul of pre-chilled 0.2mM EDTA containing 1.5 µL of 0.1M PMSF (1:1000), 0.5 µL of 3M DTT (1:3000), 1.5 µL of 0.1 M Benzamidin (1:1000)
36. Add 15 µL of 10% NP40, pipette up and down 20 times
37. Spin down at 4000 rpm for 3 minutes
38. Save 150 µL (or 10%) of the supernatant for Input at -80 °C
39. Take 1350 µL for ChIP, transfer to a 2-ml rube, and add 150 µL of 10XFSB
40. Add 10 µg/10 µL of anti-H3K4me3 antibody or 4 µL of anti-H3K27ac antibody (or, instead of adding antibody, save chromatin at -80°C for future use)
41. Incubate overnight at 4°C on rotator

Day 2: ChIP

Preparing magnetic beads and beads incubation

42. Allow 30 minutes for magnetic beads to warm to room temperature
43. Mix well and pipette 20 µL of A/G Magnetic beads per sample and re-suspend with 1 mL of 1x FSB with 0.1%NP40
44. Place tube with magnetic beads on rack for five minutes allowing beads to settle on wall of tube
45. Remove solution and re-suspend again with 1 mL of 1x FSB with 0.1%NP40
46. Place tube on magnetic rack for five minutes, remove solution
47. Estimate and add amount of 1xFSB with 0.1%NP40 to magnetic beads to give final 20 µL volume for each sample
48. Pipette 20 µL of magnetic beads solution to ChIP samples
49. Incubate on rotator in 4°C for two hours

During incubation

50. Make fresh elution buffer (final volume 25ml)
    0.21g NaHCO3
    2.5ml 10%SDS
    22.5ml ddH2O

Washing beads and chromatin complex (30 min)

51. Place ChIP samples on magnetic rack for five minutes
52. Carefully remove supernatant
53. Wash beads with 1 ml of each washing solution
54. Discard the supernatant using vacuum
    * Each washing solution
    -- Low salt washing buffer
    -- High salt washing buffer
    -- TE buffer (10 mM Tris, 1 mM EDTA pH=8)

55. Transfer magnetic beads in TE buffer to newly labeled tubes
56. Repeat TE buffer wash and carefully remove the supernatant by pipetting

Elution of chromatin from beads (40 min)

57. Add 250 µL of elution buffer from Step 44 to sample
58. Rotate 15 minutes at room temperature
59. Transfer 250 µL of supernatant into 2 ml o-ring tube, place at room temperature
60. Add another 250 µL of elution buffer to sample
61. Rotate 15 minutes
62. Transfer 250 µL of supernatant into 2ml o-ring tube from step 53, final elution volume is 500 µL

RNA and Protein digestion (4 hr)

63. For each ChIP sample (total 500 µL), add:
    20 µL of 5M NaCl
    50 µL of 1M Tris, PH 8
    5 µL of 0.5M EDTA

64. Take input samples from Step 33 out of freezer, for each input sample (total 150 µL), add:
    277.5 µL of water
    2.5 µL of 20% SDS
    20 µL of 5M NaCl
    50 µL of 1M Tris, PH 8

65. Add 1 µL of 20 mg/ml RNaseA, incubate at 37 °C for 15 min
66. Add 2.5 µL of 10 mg/mL proteinase K (1:200), incubate at 52°C for at least 2 hours

Phenol/Chloroform extraction (30 min)

67. Add 500 µL of phenol chloroform into each sample (under fume hood)
68. Vortex 1 min
69. Centrifuge at 12,000 rpm for 5 minutes at room temperature
70. Transfer aqueous (top) into a 2 mL silicone tube
    Add: 2 µL of GlycoBlue
    50 µL of 3M sodium acetate

71. Add 1.375 ml of pre-chilled 100% ethanol
72. Invert several times and store at -80°C overnight to precipitate DNA

Day 3: ChIP DNA precipitation

73. Take sample from Step 66 out of freezer, thaw on ice
74. Centrifuge at 13,000 rpm for 30 minutes at 4°C
75. Discard supernatant
76. Add 1 mL of pre-chilled 75% ethanol
77. Centrifuge at 13,000 rpm for 10 minutes at 4°C
78. Carefully remove all the supernatant
79. Air dry the pellet
80. Re-suspend pellet in 45 µL of EB Buffer, pH 8
81. Store at -80°C

ChIP Assay Quality Control

The efficiency of each ChIP assay was validated using:
1) Agilent Bioanalyzer (to confirm Input DNA product of around 150 bp);
2) Qubit concentration measurement;
3) qPCR for positive (*GRIN2B*, *DARPP32*) and negative (*HBB*) control genomic regions.

Only ChIP assays that passed quality control were further processed for library preparation and sequencing; this included ChIP DNA that was not detectable on Qubit but showed a good signal and expected enrichment patterns in qPCR (note that *GRIN2B* promoter, although enriched for H3K4me3 and H3K27ac in both NeuN+ and NeuN- nuclei, consistently showed stronger enrichment in NeuN+ nuclei fraction).

II. ChIP-Seq Library Preparation Protocol

Libraries are prepared using **KAPA Hyper Prep Kit (KK8502)**. After MNase digestion, the main DNA product is around 150 bp and, after ligation of the adapters, the size of the library is around 275 bp. The size selection (200-350 bp) is performed using Pippin Prep.

The starting amount of ChIP DNA varies from <2 ng to 20 ng and this protocol works well for this range. The amount of Input DNA that we use for Input Libraries is 4 ng.

Step 1: End Repair and A-Tailing

1a. Combine the following components in a PCR tube:

ChIP DNA or 4 ng Input DNA (+EB buffer)	50 µL
End Repair and A-Tailing Buffer	7 µL
End Repair and A-Tailing Enzyme Mix	3 µL
Total volume	**60 µL**

1b. Mix thoroughly and centrifuge briefly.

1c. Incubate in a thermocycler with the following profile:

20˚C	30 min
65˚C	30 min
10˚C	∞

1d. Proceed immediately to the next step.

Step 2: Adapter Ligation

2a.Combine and mix the following components in PCR tube:

Product of End Repair/A-Tailing	60 µL
EB	9 µL
Ligation Buffer	30 µL
DNA Ligase	10 µL
*BIOO Adapter 1-12 (1.5 µM)	1 µL
Total volume	**110 µL**

*Please note that the original BIOO Scientific Adapters have concentration of 25 µM. Make 1:16.7 dilution of the original adapters to make a final concentration of 1.5 µM

* Libraries are barcoded (1-12) to allow for multiplexing. The following four-plex barcode pools are routinely used: **Pool 1:** 1, 2, 5, 10
Pool 2: 3, 4, 6, 12
Pool 3: 7, 8, 9, 11

2b. Mix thoroughly and centrifuge briefly.
2c. Incubate in thermocycler for 15 min at 20˚C.
2d. Proceed immediately to the next step.

Step 3: Post-ligation AMPure Beads Cleanup

3a. Perform a 1.8x Ampure cleanup by mixing the following:

Adapter Ligation reaction product	110 µL
AMPure beads	198 µL
Total Volume	308 µL

3b. Mix thoroughly (pipet 10 times up and down) and incubate for 15 min at room temperature to allow DNA to bind to beads.

3c. Place tube on a magnet to capture the beads. Incubate until liquid appears clear (can be up to several min).

3d. Remove and discard supernatant without drying out the beads.

3e. On the magnet, wash twice with 200 µL of freshly prepared 80% ethanol each time.

3f. Spin down and pipette off any residual ethanol.

3g. Remove from magnet and resuspend the beads with 26 µL of EB. Incubate for 5 min at room temperature to elute DNA off the beads.

3h. Place tube on the magnet and incubate until liquid is clear. Transfer 24 µL of the clear solution to a new tube.

Step 4: PCR Amplification

4a. Combine and mix the following components in PCR tubes:

Adapter-ligated product	24 µL
2x KAPA HiFi HotStart Ready Mix	25 µL
*BIOO PCR primer mix (6.25 µM)	1 µL
Total volume	**50 µL**

*Note that the BIOO PCR primer mix comes in a 12.5 µM concentration so make a 2-fold dilution. Also, make sure that the PCR primer mix comes from the same kit in which BIOO adapters are supplied (Cat #514102); do NOT use the PCR primers from the KAPA kit.

4b. Mix thoroughly and centrifuge briefly.

4c. Amplify using the following PCR protocol:

98˚C	2 min	1 cycle
98˚C	30 sec	12-15 cycles
65˚C	30 sec	
72˚C	30 sec	
72˚C	4 min	1 cycle
10˚C	∞	1 cycle

In the majority of cases the number of cycles is 12. If the input ChIP DNA is very low (undetectable on Qubit), the number of cycles should be increased to 15.

Step 5: Post-PCR Ampure Cleanup

5a. Perform a 1.8x AMPure cleanup by mixing the following:

Adapter Ligation reaction product	50 µL
Ampure beads	90 µL
Total Volume	140 µL

5b. Mix thoroughly and incubate for 15 min at room temperature to allow DNA to bind to beads.

5c. Place tube on a magnet to capture the beads. Incubate until liquid appears clear.

5d. Remove and discard supernatant without beads drying.

5e. On the magnet, wash twice with 200 µL of freshly prepared 80% ethanol each time.

5f. Spin down and pipette off any residual ethanol on the magnet.

5g. Remove from magnet and resuspend the beads with 22 µL of EB. Incubate for 4 min at room temperature to elute DNA off the beads.

5h. Place tube on the magnet and incubate until liquid is clear. Transfer 20 µL of the clear solution to a new tube.

Step 6: Pippin Prep

6a. Add EB to DNA samples for a total volume of 30 µL while loading solution comes to room temperature.

6b. For each sample, combine 30 µL of DNA sample with 10 µL of Marker L and mix thoroughly.

6c. In Pippin Prep Software, select **2% DF Marker L** and Use Internal Standards. Save the protocol.

6b. To calibrate, use calibration fixture and enter **0.80** in the "Target I ph, mA" field. Press Calibrate and Exit.

6c. Remove gel cassette from packaging and check buffer levels, gel column breakage, and bubbles.

6d. Release any bubbles trapped behind the elution wells and place cassette into the optical nest.

6e. Remove the white adhesive from the cassette.

6f. Remove buffer from elution wells and replace with 40 µL of fresh electrophoresis buffer.

6g. Seal the elution wells with tape provided in cassette package.

6h. Remove ~70 µL buffer from samples wells and replace with 70 µL of fresh electrophoresis buffer.

6i. Perform continuity test.

6j. For each sample, remove 40 µL of buffer from the sample well and carefully add 40 µL of the sample. Close the lid.

6k. Ensure the proper protocol is loaded and press "Start".

6l. Remove samples with a 100-200 µL pipette followed by washes with buffer (40 µL, 35 µL, 35 µL for total volume 110 µL) and dispose of cassette. Repeat as necessary.

Step 7: Post-Pippin AMPure Cleanup

7a. Perform a 1.8x AMPure cleanup by mixing the following:

Adapter Ligation reaction product	110 µL
AMPure beads	198 µL
Total Volume	308 µL

7b. Mix thoroughly and incubate for 15 min at room temperature to allow DNA to bind to beads.

7c. Place tube on a magnet to capture the beads. Incubate until liquid appears clear.

7d. Remove and discard supernatant without beads drying.

7e. On the magnet, wash twice with 200 µL of freshly prepared 80% ethanol each time.

7f. Spin down and pipette off any residual ethanol on the magnet.

7g. Remove from magnet and resuspend the beads with 22 µL of EB. Incubate for 5 min at room temperature to elute DNA off the beads.

7h. Place tube on the magnet and incubate until liquid is clear. Transfer the 20 µL of clear solution to a new tube – this is the final library.

Step 8: Final Library Quality Control

8a. Check concentrations with High Sensitivity DNA Qubit.

8b. Check size/quality of samples using Agilent BioAnalyzer.

The presence of the main library product (275 bp) and the absence of adapter dimer (125 bp) are confirmed using Agilent Bioanalyzer. **Only libraries that pass Bioanalyzer quality control are submitted for sequencing.**

Alternative Library Preparation Protocol

Alternatively, libraries were prepared using TruSeq ChIP Sample Prep Kit (set A) (Illumina, IP-202-1012) with some modifications. ChIP or Input DNA (<2 ng) was used for library preparation in batches of eight, end-repaired with the End-It DNA Repair Kit from Epicentre (ER0720) followed by A-tailing using Exo-Minus Klenow DNA Polymerase (Epicentre, KL11101K), followed by overnight ligation at 16°C overnight using Fast-Link DNA Ligation Kit (Epicentre, KL11101K) with 1 ul of adaptor from TruSeq ChIP Sample Prep Kit (set A) (Illumina, IP-202-1012), followed by PCR amplification using TruSeq v2 LT Sample Prep Kit-PCR box (Illumina, 15027084). A default 12 PCR cycle setting yields 300 ng of library DNA on average. In case of low starting material of ChIP DNA (undetectable by Qubit), 15 PCR cycles may be necessary in order to yield 200-300 mg of DNA without compromising on library quality. 1.8X SPRI is used for each clean-up step, other than the final clean-up after PCR amplification, for which 1.0X SPRI is used to remove any adapter-dimers (125 bp). The main library products are expected around 275 bp (mononucleosomal DNA (~150 bp) plus adapters. Pippin selection with the setting of 200 to 350 bp is used in case further separation of high molecular library product (~425 bp, or 575 bp) from di or tri-nucleosome DNA (~300 bp, or 450 bp) is required.

Supplemental References

1. Kunii Y, Zhang W, Xu Q, Hyde TM, McFadden W, Shin JH, et al. (2015): CHRNA7 and CHRFAM7A mRNAs: Co-Localized and Their Expression Levels Altered in the Postmortem Dorsolateral Prefrontal Cortex in Major Psychiatric Disorders. *Am J Psychiatry*.appiajp201514080978.
2. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP (2014): Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet*. 15:121-132.
3. Bailey T, Krajewski P, Ladunga I, Lefebvre C, Li Q, Liu T, et al. (2013): Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLoS computational biology*. 9:e1003326.
4. Li H, Durbin R (2009): Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25:1754-1760.
5. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. (2009): The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:2078-2079.
6. Chen Y, Negre N, Li Q, Mieczkowska JO, Slattery M, Liu T, et al. (2012): Systematic evaluation of factors influencing ChIP-seq fidelity. *Nat Methods*. 9:609-614.
7. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. (2008): Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 9:R137.
8. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. (2011): Integrative genomics viewer. *Nat Biotechnol*. 29:24-26.
9. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. (2010): Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature reviews Genetics*. 11:733-739.
10. Shulha HP, Cheung I, Whittle C, Wang J, Virgil D, Lin CL, et al. (2012): Epigenetic signatures of autism: trimethylated H3K4 landscapes in prefrontal neurons. *Arch Gen Psychiatry*. 69:314-324.
11. O'Neill LP, Turner BM (2003): Immunoprecipitation of native chromatin: NChIP. *Methods*. 31:76-82.
12. Jiang Y, Matevossian A, Huang HS, Straubhaar J, Akbarian S (2008): Isolation of neuronal chromatin from brain tissue. *BMC Neurosci*. 9:42.
13. Marion-Poll L, Montalban E, Munier A, Herve D, Girault JA (2014): Fluorescence-activated sorting of fixed nuclei: a general method for studying nuclei from specific cell populations that preserves post-translational modifications. *Eur J Neurosci*. 39:1234-1244.
14. Bonn S, Zinzen RP, Perez-Gonzalez A, Riddell A, Gavin AC, Furlong EE (2012): Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. *Nat Protoc*. 7:978-994.