Check for updates

Single-cell DNA methylation sequencing by combinatorial indexing and enzymatic DNA methylation conversion

Zac Chatterton[1,2]* , Praves Lamichhane[1,2], Diba Ahmadi Rastegar[1,2], Lauren Fitzpatrick[1,2], Hélène Lebhar[3], Christopher Marquis[4], Glenda Halliday[1,2] and John B. Kwok[1,2]

Abstract

Background DNA methylation is a critical molecular mark involved in cellular differentiation and cell-specific processes. Single-cell whole genome DNA methylation profiling methods hold great potential to resolve the DNA methylation profiles of individual cell-types. Here we present a method that couples single-cell combinatorial indexing (sci) with enzymatic conversion (sciEM) of unmethylated cytosines.

Results The sciEM method facilitates DNA methylation profiling of single-cells that is highly correlated with single-cell bisulfite-based workflows ($r^2 > 0.99$) whilst improving sequencing alignment rates, reducing adapter contamination and over-estimation of DNA methylation levels (CpG and non-CpG). As proof-of-concept we perform sciEM analysis of the temporal lobe, motor cortex, hippocampus and cerebellum of the human brain to resolve single-cell DNA methylation of all major cell-types.

Conclusion To our knowledge sciEM represents the first non-bisulfite single-cell DNA methylation sequencing approach with single-base resolution.

Keywords Epigenetics, DNA methylation, Single-cell, Brain, APOBEC

Introduction

The covalent addition of a methyl group to cytosine bases in mammalian DNA (DNA methylation) is one of the most highly studied epigenetic modifications [1]. Primarily occurring in the CpG context, DNA methylation is critical for organism development [2] and plays an essential role in regulating gene expression during cellular differentiation [3]. Cell-types have highly specific DNA methylation patterns [4] necessitating the analysis of DNA methylation in pure cellular populations, however limited cell surface markers or highly interconnected tissue networks prohibit cell isolation from tissues such as the human brain.

Single-cell whole genome bisulfite sequencing techniques have recently been described [5–8] that can produce single-base resolution DNA methylation information from which cell-specific whole genome DNA methylation profiles (methylomes) can be reconstructed bioinformatically. However, these sequencing library preparations are prohibitively expensive for most labs because of high reagent costs associated with single-cell single-well reactions. Recently a single-cell combinatorial indexing (sci-) bisulfite sequencing approach (termed sciMET) was described in which nuclei are sorted and

*Correspondence:
Zac Chatterton
zac.chatterton@sydney.edu.au
[1] Brain and Mind Centre, The University of Sydney, Camperdown, Australia
[2] School of Medical Science, The University of Sydney, Camperdown, Australia
[3] Recombinant Products Facility, University of New South Wales, Kensington, Australia
[4] School of Biotechnology and Biomolecular Science, University of New South Wales, Kensington, Australia

Chatterton *et al. Cell & Bioscience*   (2023) 13:2

Page 2 of 11

tagged with sequencing indexes over multiple rounds, forming unique combinations of indexes per nuclei [9]. The sciMET approach allows multiple-nuclei single-well reactions, thus reducing reagent costs.

Single-cell DNA methylation sequencing is still in its infancy, however the technique's importance cannot be overstated as it enables the DNA methylation profiling of cell-types that lack cell surface markers that can be used for their isolation, such as those in the brain [10]. Further, defining cell-specific DNA methylation patterns have been critical for understanding mechanisms of disease, such as cancer [11], and represent important markers for cell-of-origin molecular diagnostic assays [12]. Notably, targeted single-cell epigenomics is emerging as a promising approach for high throughput functional screening of disease relevant gene regulation [13]. To our knowledge high throughput functional screens using single-cell DNA methylation sequencing have yet to arrive, however, single-cell reduced representation bisulfite sequencing workflows can lower experimental costs by enriching for dense regions of CpG dinucleotides (CpG Islands) that are important in gene regulation at the expense of whole genome coverage [14].

Bisulfite sequencing is the gold standard method for DNA methylation analysis [15, 16] but is not without limitations. It has been estimated that 84–96% of DNA is degraded during the bisulfite conversion reaction [17]. Additionally, methylated DNA is overrepresented in WGBS libraries leading to an over-estimate of DNA methylation levels [18] particularly in CHG and CHH contexts [19]. This is especially relevant in the analysis of DNA methylation in embryonic stem cells and neurons that have been reported to exhibit high levels of CHG and CHH methylation [20–25].

Unmethylated cytosines can also be deaminated by *APOBEC* enzymes, resulting in base changes analogous to bisulfite conversion (sequenced as T) [26]. Notably, enzymatic conversion is less degradative to DNA and can produce high quality single-base resolution DNA methylation data [26]. Such attributes may be particularly beneficial in situations where DNA content is limited such as the single-cell analysis of DNA methylation. However, enzymatic conversion in single cells is challenging due, in part, to multiple reaction cleanup steps required that results in DNA loss. A major advantage of the sci-approach over single-cell/single-well methods is the ability to perform deamination reactions of multiple cells per-well, thus increasing per-well DNA content. Here we combine sci- with enzymatic conversion (termed sciEM, Fig. 1a) and show application by characterizing single-base DNA methylation profiles of human brain cell-types without the need for cell-type markers (e.g. NeuN). The sciEM approach accurately captures CpG methylation dynamics across annotated regulatory features of the human genome. Both CpG and non-CpG (CpH) methylation estimates are lower than bisulfite conversion and we find no evidence of higher global CpH DNA methylation within neurons from the temporal lobe, motor cortex, hippocampus or cerebellum of the human brain. The sciEM approach represents an economical method for single-cell single-base resolution DNA methylation analysis.

Results

Library construction and sequencing read processing

Using frozen post-mortem brain tissue from mouse (NextSeq, n = 1) and human (NextSeq, n = 1 and NovaSeq, n = 4) we were able to construct both sci-MET and sciEM single-cell libraries in parallel (Additional file 1: Fig. S1). Within the sciEM workflow we use a G-depleted (mg) random linear primer that we observed to improve CpH mapping within preliminary experiments (NextSeq, Additional file 1: Fig. S2). Both bisulfite and enzymatic conversion efficiencies were high, 99.99% and 99.94% respectively, however the sci-MET method produced ~10× the amount of library than sciEM (518 nM vs. 53 nM by RT-qPCR). Post-sequencing (NovaSeq), single-nuclei were identified by unique barcode combinations (Tn5, i5 and i7 barcodes). Single-nuclei with >100 unique mapped reads were observed to have significantly higher mapping efficiency, more paired-reads, larger insert sizes and a lower proportion of reads mapped using local alignment (Students t-test p-value's $< 5 \times 10^{-17}$, Additional file 1: Fig. S2a–d), representing high-quality single-nuclei. Following k-means clustering of unique mapped reads, a total of 710 and 64 high quality single-nuclei from sciEM and sciMET workflows were retained for analysis (Additional file 1: Fig. S3e, f), representing 54 and 58% of the nuclei fluorescently sorted for each workflow respectively. The mapping efficiency was $58 \pm 5\%$ and $64.9 \pm 10\%$ for sci-MET and sciEM libraries respectively (Fig. 1b), but we note the sciMET mapping efficiencies were lower than previously reported [9], results that are partially attributed to the removal of 4% of reads that contained substantial linear primer sequences (Fig. 1b). The number of mapped reads were higher in sciMET (mean $254,194 \pm 191,560$) than sciEM $(150,515 \pm 147,388)$ (Students t-test p-value $= 7.25 \times 10^{-5}$, Fig. 1b) leading to higher coverage of mappable cytosine dinucleotides (e.g., CpG p $= 1.91 \times 10^{-10}$, Additional file 1: Fig. S4a). However, the library loading concentration largely influences read counts and, proportional to mapped reads, the sciEM method covered a greater number of all cytosine dinucleotides, particularly CpT and CpC dinucleotides (p $< 6.1 \times 10^{-10}$, Fig. 1c).
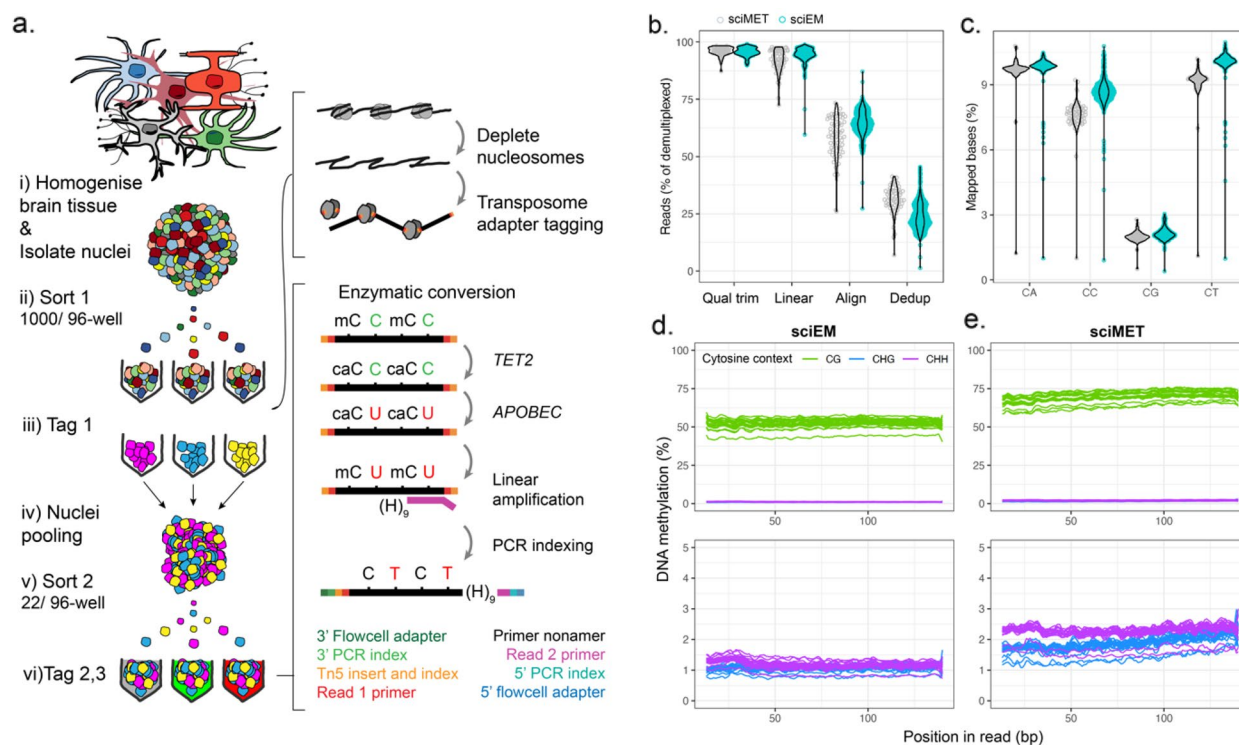
Fig. 1 **a** Single-cell combinatorial indexing and enzymatic conversion (sciEM) workflow in which (i) whole tissue (e.g. brain tissue) is homogenized to dissociates cells. Nuclei from heterogeneous cell-types are isolated and (ii) sorted by Fluorescent Activated Nuclei-Sorting (FANS). (iii) Nuclei membranes are permeabilized, nucleosomes depleted, and molecular tags (tag 1, Tn5 barcode) are attached to genomic DNA via transposome tagmentation. (iv) Nuclei are pooled, (vi) re-sorted by FANS and (vi) unmethylated cytosines are converted to thymine following treatment with *TET2* and *APOBEC* enzymes and Linear amplification. Molecular tags 2, 3 (i5 and i7 barcodes) and sequencing adapters are attached via PCR amplification. (adapted from [9]). **b** Per cell read processing metric's. **c** Cytosine dinucleotides covered as percentage of mapped bases. **d**, **e** DNA methylation bias plots for sciEM and sciMET methods respectively for reads mapping to autosome's and the X-chromosome, with close-up of CHG and CHH DNA methylation (bottom). H = A, C or T

DNA methylation

Global DNA methylation levels were observed to be lower in sciEM than sciMET libraries for CpG (54% vs. 71%), CHG (1.0 vs. 1.2%) and CHH (1.2 vs. 2.2%) cytosine contexts (p-value $< 7.10 \times 10^{-8}$). WGBS library preparation methods have previously been reported to have an over-representation of methylated fragments [18]. In line with these reports, we observed an increase in the 5mC levels from 5′ to 3′ of sciMET reads for CpG, CHG and CHH contexts (student t-test p-value < 0.003, first 50% bases vs. last 50% bases). In contrast, 5mC levels were stable across sciEM reads (Fig. 1d). Of note, the 5mC levels of chromosome 21 were significantly lower than other chromosomes, an effect that was pronounced within sciEM libraries. We observed a very high correlation between the CpG methylation of sciEM and sciMET across annotated genomic features (e.g., $R^2 = 0.996$, ± 5 kb Ensemble genes, Pearson's p-value $= 6.9 \times 10^{-120}$).

CpG hypomethylation is generally associated with chromatin accessibility and gene activity and sciEM successfully captured global DNA methylation dynamics across regulatory regions e.g. hypomethylation of gene promoter regions, CpG Islands and open chromatin (DNase-seq) (Fig. 2a–c). We observed CpG hypomethylation across annotated regions enriched for active histone marks (e.g. H3K4me3, H3K9ac, H3K27ac) and H3K4me1 boundaries (Fig. 2d–g). The H3K27me3 histone modification is a marker of bivalent polycomb regulated promoters in which dynamic crosstalk between DNA methylation controls gene expression [27] and we observed CpG hypomethylation across genomic regions enriched for H3K27me3 (Fig. 2h). Conversely, hypermethylation was observed within annotated regions enriched for repressive histone modifications (H3K36me3 and H3K9me3) (Fig. 3i, j). Of note, sciEM CpG methylation levels are significantly lower than sciMET measurements e.g., 14% lower across annotated genes (± 5 kb) (paired t-test p-value $= 3.46 \times 10^{-65}$). The difference in CpG DNA methylation between sciEM and sciMET was highly
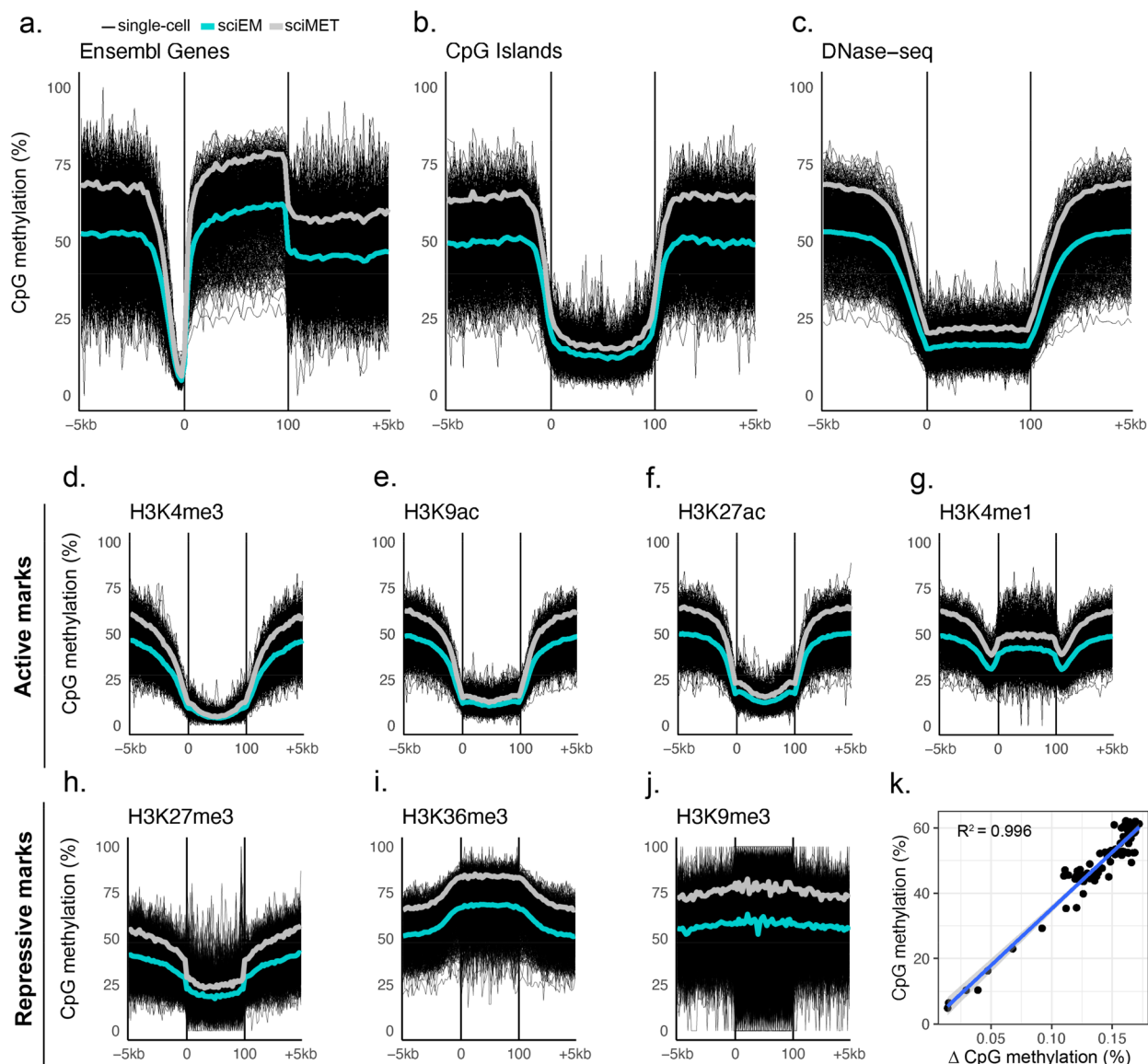
Fig. 2 CpG methylation across genomic features. The mean CpG methylation levels of single-cells (black) from sciEM (cyan) and sciMET (grey) protocols across **a** Ensemble genes **b** CpG Islands and regions of **c** DNase hypersensitivity as well as histone modifications associated with active (**d**–**g**) and repressive (**h**–**j**) chromatin conformations. **k** Scatterplot of CpG methylation levels (sciMET mean) and methylation difference (sciMET-sciEM) across each bin (3%) of annotated Ensemble genes (± 5 kb)

correlated to the underlying CpG methylation levels (Fig. 2k).

Single-cell clustering

To assess the ability of sciEM to discriminate cell types we summarized CpG DNA methylation of regulatory regions (Ensembl Regulatory Build) and CpH methylation across 100 kb genomic bins (used to cluster Neuronal cell-types [7]) across 710 high quality single nuclei (methods). The DNA methylation information of each single nuclei was combined using NMF and projected into 2-dimensional space (tSNE) from which 19 clusters were identified (Fig. 3a). We observed distinct patterns of DNA methylation within established Differentially Methylated Regions (DMRs) distinct to Neurons and Non-Neuronal cell-types from the human brain [22] which enabled the identification of 11 neuron (n = 430) and 8 non-neuronal clusters (n = 280) (Fig. 3b). As gene promoter hypomethylation is associated with gene activity, the promoter DNA methylation status of established Differentially Expressed Genes (DEG's) of non-neuron cell-types [28] enabled the identification of
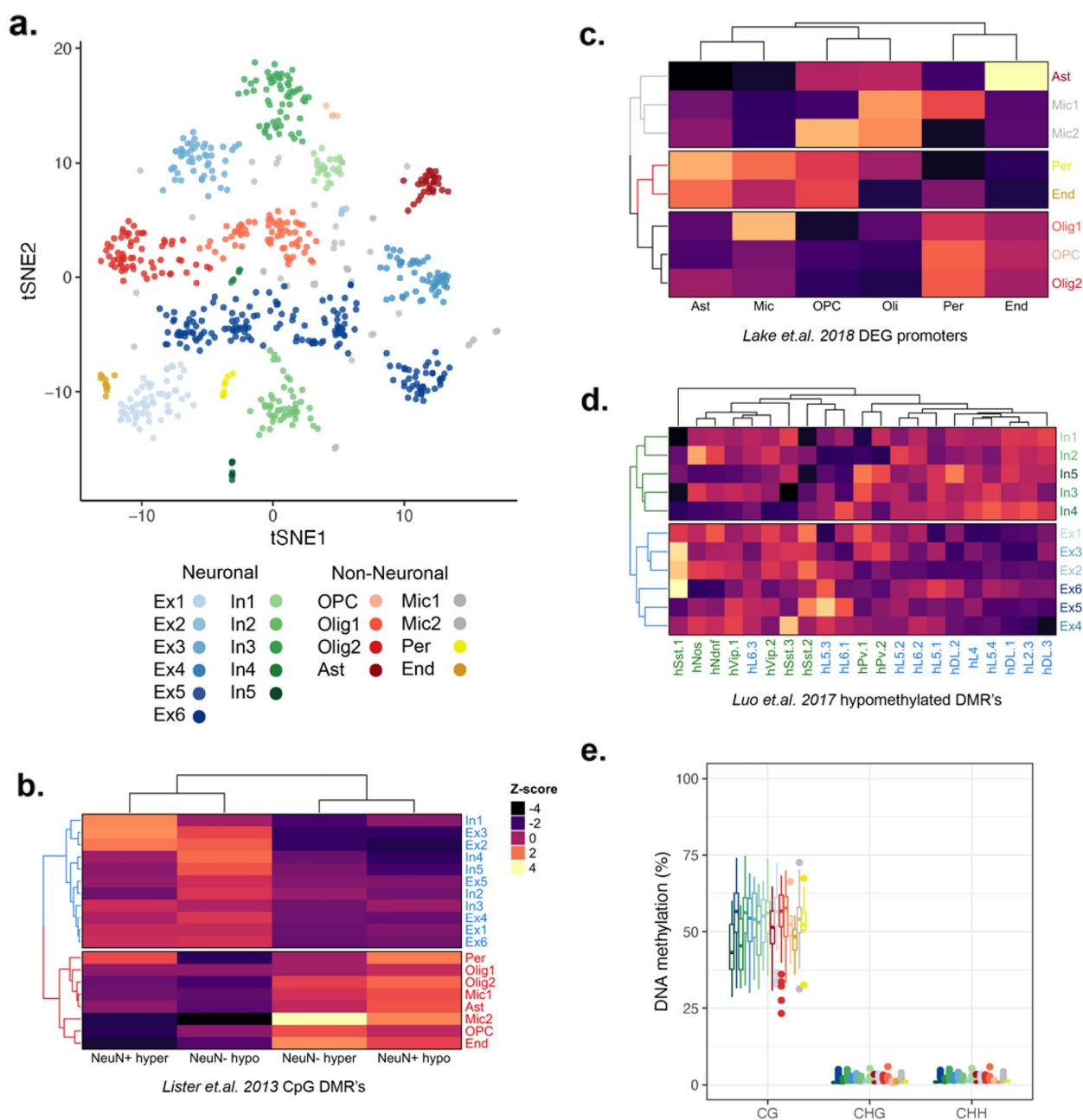
Fig. 3 Cell-type discrimination by sciEM single nuclei DNA methylation. **a** Single nuclei DNA methylation clustering (NMF-tSNE). Clusters (n = 19) are defined by unique colors. **b** Heatmap of summarized CpG methylation z-score's of clusters across annotated Neuron and Non-neuronal DMR's. **c** Heatmap of summarized CpG methylation z-score's of non-neuronal cell clusters across annotated non-neuronal DEG's. **d** Heatmap of summarized CpG methylation z-score's of neuronal clusters across annotated neuronal subtype CpG DMR's. **e** Boxplots of CpG, CHG and CHH DNA methylation of each cell-type. *Ast* astrocyte, *Exc* excitatory neuron, *End* endothelial cells, *Inh* inhibitory neuron, *Mic* microglia, *Olig* oligodendrocyte, *OPC* oligodendrocyte precursor cells, *Per* pericyte

Astrocytes (n = 30), Endothelial cells (n = 12), Microglia (n = 45), Oligodendrocyte precursor cells (OPC's, n = 5), Oligodendrocytes (n = 130) and Pericytes (n = 8) (Fig. 3c). Using the CpG DNA methylation levels at

annotated neuronal cell-type CpG DMRs [7] we identified 6 excitatory (n = 279) and 5 inhibitory (n = 151) neuronal cell-type clusters (Fig. 3d). We found no significant difference in the per-nuclei CpG, CHG or CHH

Chatterton *et al. Cell & Bioscience*   (2023) 13:2

Page 6 of 11

DNA methylation levels between neuron and non-neuronal cell-type clusters (Fig. 3e).

Discussion

To our knowledge we present the first enzyme based single-cell DNA methylation method with single-base resolution. The original sciEM method extends single-cell combinatorial indexing approaches developed using sodium bisulfite (sciMET). Bisulfite sequencing is problematic for single-cell sequencing as it degrades the limited amount of DNA in each cell [29], however enzymatic based conversion of unmethylated cytosines has been shown to be less degradative to the DNA, resulting in more genomic coverage, even at 100 pg amounts [26]. However, the enzymatic conversion of ∼132pg (22 nuclei) produced ∼10× less library than bisulfite conversion, likely due to the increased number of wash steps. We further theorized that a greater reduction in the library would have occurred if the DNA content of each reaction was further reduced via a single-cell single-well system. Hence, our adoption of the combination of the enzymatic conversion process with a multiple-nuclei single-well approach, such as single-cell combinatorial indexing, has likely contributed to the success of creating the first enzymatic DNA methylation single-cell libraries.

To generate sciEM libraries we replaced the 9-N linear amplification primer of the sciMET protocol with a G-depleted random primer which we discovered to have bound more efficiently to genomic fragments devoid of cytosines following conversion (enzymatic or bisulfite converts ∼95% of cytosines to uracils). G-depleted random primers have been previously shown to improve library complexity, coverage uniformity and reduce artefactual reads [8]. Generally, the sciEM approach results in higher library loss during construction, and a lower amount of library input into the second barcoding PCR (i5 and i7), preferential amplification of smaller molecules (reduced insert size), and a higher rate of duplicate sequences compared to sciMET. However, the data quality is high. We show high correlation of DNA methylation between sciEM and sciMET ($r^2 = 0.99$) approaches, accurate recapitulation of DNA methylation dynamics across gene features and the ability to resolve single base resolution of single-cell types.

Since the first whole genome bisulfite sequencing (WGBS) study, a multitude of techniques have been developed to characterize genome-wide methylation [30]. However, as DNA methylation patterns are unique to single-cell-types, it is essential to move towards single-cell DNA methylation profiling [31]. In disease states, DNA methylation patterns are known to be altered, leading to aberrant cascades of molecular changes [3]. Hence it is important to have accurate base level estimates of DNA methylation levels. Whole genome bisulfite sequencing approaches have been reported to over-represent methylated fragments [18]. We observed 5mC bias within sciMET sequencing reads, elevating from 5′ to 3′, that were not observed within sciEM. Further, 5mC levels were lower in sciEM, an effect that was greater within CHH loci (typically unmethylated) relative to CpG loci (typically methylated). Whilst we cannot rule out over conversion of 5mC within sciEM, our observations of a 5mC bias within sciMET sequencing reads (elevating from 5′ to 3′), comparatively higher 5mCHH (typically unmethylated) then 5mCpG (typically methylated) and a slightly higher conversion efficiency indicate a potentially over-representation of methylated fragments in sciMET.

Neurons exhibit distinct DNA methylation patterns, particularly in non-CpG (CpH) loci, compared to non-neuronal brain cells [22, 32] as well as between neuronal subtypes [7, 33]. To our knowledge we present the first whole genome DNA methylation assessment of brain cell-types using enzymatic DNA methylation assessment. We did not observe significantly higher CpH methylation within neurons as previously reported [22]. Further, our single-cell analysis of brain cell types omits the use the NeuN antibody for neuron selection, hence we cannot rule out the possibility that NeuN-positivity (nuclei surface marker) of Neurons relates to CpH DNA methylation.

Conclusion

The brain is a highly heterogeneous environment, comprising multiple neuronal and glial cell types with unique functions and at various stages of differentiation. Previously, the understanding of disease mechanisms progressed by studying cell populations in bulk which revealed only the average features of the population's constituents and can obscure the cell-to-cell variability. Since the first whole genome bisulfite sequencing (WGBS) study, a multitude of techniques have been developed to characterize genome-wide methylation [30] at single-base resolution. Moreover, as DNA methylation patterns are unique to cell-types, it is essential to move towards single-cell DNA methylation profiling [31]. Single-cell strategies have already yielded novel mechanistic insights into brain function [34]. The sciEM approach represents an invaluable tool in assessing CpH DNA methylation function within cell-types with reportedly increased CpH DNA methylation, such as the brain and stem cells [35].

Many DNA methylation signatures have been described for distinct cellular phenotypes including cell-type, pluripotency [36], age [37] and disease state [11]. As sci-based approaches offer a lower cost to entry for single-cell DNA methylation analysis, we anticipate that combining sciEM with high-content screening, such as

Chatterton *et al. Cell & Bioscience*   (2023) 13:2

Page 7 of 11

library of small molecules and CRISPR pooled screens, will create powerful new tools to evaluate mediators and mechanisms of cellular phenotypes in human health and disease and beyond into non-medical fields such as agriculture. The sciEM method represents an economical, high-throughput approach for single-cell DNA methylation at single-base resolution.

Methods

Transposome production

Recombinant transposase enzyme (Tn5) was grown (pTXB1-Tn5 vector) and purified following the protocols described in Picelli et al. [36]. Cytosine depleted sci-MET transposase-loaded oligonucleotides (1–96) were annealed (10 μL each 100 μM) to 10 μL 5′-[Phos]-CTG TCTCTTATACACATCT-3′ oligonucleotide (100 μM) within 80 μL EB buffer (Qiagen), incubating 2 min at 95 °C and cooled to room temperature (0.1 °C/s), following protocols [37]. Annealed oligonucleotides were dilute 2:5 (EB buffer), mixed with glycerol (50% final solution) and loaded (equal volume) into the recombinant Tn5 (15 μm) by incubation for 20 min at room temperature. Annealed oligonucleotide loading was confirmed by gel-shift assay and fragmentation efficiency of each transposome was confirmed (> 50%) by qPCR analysis [38].

Brain sample and nuclei isolation

NextSeq—Post-mortem flash-frozen prefrontal cortex tissue was obtained from a 93-year-old female donor with no diagnosis of neurological disease. Post-mortem flash-frozen cortex was obtained from a genetically modified (*C9orf72*) mouse. Following the protocols of Mulqueen et al., Brain tissue sections were resuspended in 5 mL of ice-cold NIB-HEPES solution (20 mM HEPES, 10 mM NaCl, 3 mM MgCl$_2$). The tissues were equilibrated (5 min) and then dounce homogenized (10 loose strokes and 5 tight strokes) and filtered through 35–40 μm cell strainers (BD Biosciences, 352235). Nuclei were pelleted (600 g) and were transferred to a fresh tube containing 5 mL ice cold NIB-HEPES solution.

NovaSeq—Post-mortem flash-frozen tissue from the Primary Motor Cortex (BA4), Banks of the superior temporal sulcus (BA 22,41/42 [BA22]), Cerebellum (CRB) and Hippocampus (HIP) were obtained from a 47-year-old female donor with no diagnosis of neurological disease. Brain tissue was acquired from the NeuroBioBank (NIH) and approved by the Research Integrity and Ethics Administration of the University of Sydney. A high amount of cellular debris was observed by nuclei isolation protocols described in Mulqueen et al. [9], therefore nuclei were isolated instead following protocols described in Matevossian et al. [39] and resuspended in 5 mL ice cold NIB-HEPES solution.

Nucleosome depletion

Following the protocols of Mulqueen et al., nuclei were cross-linked using 135 μL of 37% formaldehyde, quenched with 400 μL of 2.5 M glycine and resuspended in 5 mL of ice-cold NIB (10 mM Tris HCl pH 7.4, 10 mM NaCl, 3 mM MgCl$_2$, 0.1% Igepal (v/v), 1× protease inhibitors) solution, pelleted (500 g for 5 min), and washed using 900 μL of 1× NEBuffer 2.1 (NEB, B7202). To denature proteins, nuclei were mixed with 800 μL of 1× NEBuffer 2.1 and 12 μL SDS solution (20%) and incubated at 42 °C for 30 min with vigorous shaking. Nuclei were then mixed with 20 μL of 10% Triton-X (Sigma, 9002-93-1) and incubated at 42 °C for 30 min solubilize proteins/ increase nuclei permeabilization.

Fluorescent activated nuclei-sorting (FANS) and tagmentation

The nuclei were stained using 8 μL of 5 mg/mL DAPI dye (Thermo-Fisher, Cat. D1306) and filtered through a 35–40 μm cell strainer. FANS was performed on BD InFlux-7 L (sort 1), separating 1000 single nuclei per well in a 96-well plate containing 5 μL of 2×TB buffer [20 mM Tris(hydroxymethyl)aminomethane, 10 mM MgCl$_2$ and 20% (v/v) dimethylformamide (DMF)] and 5 μL of NIB solution. To each well, 4 μL of 4.56 μM unique transposome (1–96) was added and incubated at 55 °C for 15 min with gentle shaking, adding the "Tn5 index". All wells were then pooled, re-stained with 8 μL of 5 mg/mL DAPI and filtered. FANS was performed again (sort 2), separating 22 or 10 (control wells) single nuclei per well in a 96-well plate containing 2.5 μL of M-digestion buffer (Zymo, Cat. D5020-9), 0.25 μL of Proteinase K (Zymo, D3001-2-5), and 2.5 μL of H$_2$O. Nuclei were then digested at 50 °C for 20 min with gentle shaking and the plate was then spun at 600 g for 5 min at 4 °C.

Bisulfite conversion

Prior to bisulfite conversion, 35 pg of pre-tagmented unmethylated lambda DNA was spiked into wells receiving 10 nuclei (sort 2). NextSeq; Each well was made up to 50 μL with H$_2$O and bisulfite conversion was performed following manufacturer protocols using the EZ-96 DNA Methylation Kit (Zymo, Cat. D5004) and eluted twice (12.5 μL each using elution buffer) for a final volume of 25 μL. NovaSeq; Each well was made up to 20 μL with H$_2$O and bisulfite conversion was performed following manufacturer protocols using the MethylCode BC conversion kit (Applied Biosystems, Cat. MECOV50) and

Chatterton *et al. Cell & Bioscience*   (2023) 13:2

Page 8 of 11

eluted twice (12.5 μL each using elution buffer) for a final volume of 25 μL.

Enzymatic DNA methylation (EM) conversion

Prior to the EM conversion, 35 pg of pre-tagmented unmethylated lambda DNA and 14 pg of the tagmented CpG methylated pUC19 DNA were spiked into wells receiving 10 nuclei (sort 2). Sample volumes were made up to 20 μL with H_2O and cleanup was performed using 1.8× using AMPure XP cleanup beads (Beckman Coulter, Cat. A63881) following manufacturer protocol with the exception that samples were incubated for 10 min at room temperature followed by a single wash step using 60 μL 80% EtOH and eluted 29 μL of elution buffer. Enzymatic conversion was then performed using the NEBNext Enzymatic Methyl-seq Conversion Module (New England Biolabs, Cat. E7125L) following manufacturers protocol (steps 1.5 to 1.9.11) for inserts 370–420 bp. Briefly, 5-Methylcytosines and 5-Hydroxymethylcytosines were oxidized using TET2 enzyme. DNA was cleaned using AMPure XP cleanup beads in place of NEBNext Sample Purification Beads. DNA was denatured in 0.1 N NaOH and cytosines were deaminated by APOBEC3A, cleaned and eluted using 25 μL of Elution Buffer.

Linear amplification

Full elution's from both the bisulfite converted and EM converted libraries were transferred to a plate prepared with the following: 16 μL of PCR-clean H2O, 5 μL of 10×NEBuffer 2.1, 2 μL of 10 mM dNTP mix (New England Biolabs, Cat. N0447), and 2 μL of 10 μM of either the 9-nucleotide random primer (n9) previously described in the sciMET protocols [9] or the G-depleted (mg) random primer [8], containing a partial Illumina Standard Read 2 sequencing primer 5′-GGAGTTCAGACGTGT GCTCTTCCGATCT(H1:33340033)(H1)(H1)(H1)(H1) (H1)(H1)(H1)(H1)-3′. Four rounds of linear amplification were performed using 10 U of Klenow (3′–5′ exo) polymerase (Enzymatics, Cat. P7010-LC-L) followed by AMPure XP cleanup (1.1×) and elution in 21 μL of 10 mM Tris-HCl (pH 8.5) as previously described [9].

Library indexing and quantification

Indexing PCRs were performed in a 96-well plate to incorporate i5 and i7 indexes. The full elution's from the linear amplification reaction were mixed with 2 μL each of the 10 μM forward and reverse indexing primers [9], 25 μL of 2xKAPA NEBNext Q5 Hot Start HiFi PCR Master Mix (New England Biolabs, Cat. M0543L), and 0.5 μL of 100× SYBR Green I dye (FMC BioProducts, Cat. 50513). Real-time PCR was performed on a QuantStudio 6 Flex real-time thermocycler (Applied Biosystems) with

the following thermocycling conditions: 95 °C for 2 min, 20 cycles of 94 °C for 80 s, 65 °C for 30 s and 72 °C for 30 s [image]. The libraries were then pooled, cleaned using AMPure XP beads (0.8×) and eluted in 20 μL of 10 mM Tris-HCl (pH 8.5) as previously described [9]. Quantification of each sciMET(n9), sciMET(mg) and sciEM (combined n9 and mg) were performed using the KAPA qPCR Illumina library quantification kit (Kapa Biosystems Cat. KR0405) and the mean of each sciMET result (n9 = 397 nM, and mg = 638 nM) was calculated.

Library sequencing

NextSeq—sciMET (n9), sciMET (mg) and sciEM (combined n9 and mg libraries) were quantified separately by High Sensitivity D1000 ScreenTape (Agilent, Cat. 5067-5584). Libraries were pooled and sequenced on the Illumina NextSeq 500 (v2 2 × 75 bp cycle Mid-Output Kit) using a 0.9 pM loading concentration, 30% PhiX and custom Read 1 and Index 2 (i5) oligonucleotides matching chemistry temperatures [9]. Sequencing was performed using custom chemistry (Read1: 100 imaged cycles; Read2: 10 imaged cycles; Index1: 10 imaged cycles; Index2: 11 imaged cycles, 9 dark cycles, and 9 imaged cycles).

NovaSeq—sciMET(n9) and sciEM(mg) libraries were pooled and then quantified (as above). Libraries were sequenced on the NovaSeq 6000 (v1.5 2 × 300 bp SP Kit) using 116 pM loading concentration, 10% PhiX and custom Read 1 (as above). Sequencing was performed using custom chemistry (Read1: 142 imaged cycles; Index1: 10 imaged cycles; Index2: 7 dark cycles, 10 imaged cycles, 16 dark cycles, and 11 imaged cycles; Read2: 142 imaged cycles).

Bioinformatics

All scripts used for the processing and analysis of sciMET/sciEM data have been deposited and documented within https://github.com/zchatt/sciem_scripts.

Sequence read demultiplexing

NextSeq—BCL files were converted to fastq format using bcl2fastq "--create-fastq-for-index-reads --with-failed-reads --use-bases-mask Y*,I10,I20,Y*" generating 2 Read files (R1[100 bp] & R2[25 bp]) and 2 Index files (I1[10 bp, i7 index] & I2[20 bp; Tn5 & i5 indexes]) for each sequencing lane. Each R1, R2, I1 and I2 from multiple sequencing lanes were combined by linux cat command and I2 was split into individual Tn5[11 bp] and i5[9 bp] index files using linux awk command. Fastq files were demultiplexed if all 3 indexes (i5, i7 and Tn5) had a Hamming distance < 3 from the reference, as previously described [9].

NovaSeq—BCL files were converted to fastq format using bcl2fastq "--create-fastq-for-index-reads

--use-bases-mask Y*,I10,I21,Y*" generating 2 Read files (R1[142 bp] & R2[142 bp]) and 2 Index files (I1[10 bp, i7] & I2[21 bp; Tn5 & i5]) for each sequencing lane. Reads were processed as described above with the exception I2 was split into individual Tn5[11 bp] and i5[10 bp] index files.

Sequence read trimming, alignment and DNA methylation extraction

Reads were firstly trimmed (trim 1) using TrimGalore! Software (v0.38.0) with options "--illumina --stringency 3". A high number of sequences corresponding to the Linear Primer and read-through of the P7 flow-cell were observed, therefore reads were trimmed again (trim 2) using cutadapt software (v1.8.3) with options "--anywhere = AGATCGGAAGAGCACACGTCTGAACTC CAGTCA --anywhere=GAAGAGCACACGTCTGAA CTC --anywhere=ATCTCGTATGCCGTCTTCTGC TTGAAAAAAAAAAGGGGGGGGGGGGGGGGGGG GGGGGGGGGGG --minimum-length=20 --times=2" [40]. Read 2 sequences from the NovaSeq 6000 instrument displayed increasing "G" content > 60 bp that were largely poly-G sequences (1.2% reads) indicative of low signal intensity. Read 2 sequences were truncated using fastp software (v0.19.6) with options "--max_len1 60". The human (GRCh38) or mouse (GRCm39) reference genomes were each combined with the lambda phage reference genome that is used for bisulfite/ enzymatic conversion control. Alignment of reads were performed using scBS-map software using the options "-l 9 -p 12 -n 10" [41]. Aligned reads were deduplicated using samtools software with options "rmdup" [42]. DNA methylation information was extracted from aligned deduplicated BAM files using cgmaptools with options "convert bam2cgmap" [43].

Single-cell discrimination and quality control

Single nuclei with < 100 unique mapped reads were removed. The unique read counts of single nuclei have previously been used to discriminate high quality single cells [9]. Briefly, k-means clustering (k = 3) of unique aligned reads per barcode (k-means, k = 3) was performed and normal distributions were fitted to each cluster (Additional file 1: Fig. S3). Barcodes with unique read counts passing 95% confidence interval threshold (cluster 1) were retained (64 sciMET and 710 sciEM). Bisulfite and enzymatic conversion efficiencies were, calculated as the 5mC % of reads aligned to unmethylated lambda phage genome. Mapping efficiencies were assessed (reads aligned/reads assigned per barcode). We determined the number of uniquely mappable cytosine dinucleotides by intersecting the within the hg38 reference genome with umap files (k = 100) downloaded https://bismap.hoffm

anlab.org/ [44] using bedtools software with options "getfasta" and umap files (k = 100). NextSeq reads were processed as above with the exception that no unique mapped read thresholds were applied, and single nuclei were assigned as mouse or human based on the greatest read mapping efficiency to gr39 and hg38 genomes independently.

DNA methylation across genomic annotations

CpG and CpH methylation were summarized (3% window) for each single-cell across (± 5 kb) Ensembl gene annotations, CpG Islands (CGI) as well as ChIP-seq and DNase-seq annotations from the middle frontal cortex (ENCFF146VKE, ENCFF225RTW, ENCFF600AYY, ENCFF724XKK, ENCFF727KZF, ENCFF729EZH, ENCFF835ZYG, ENCFF860MVH from https://www.encodeproject.org) using cgmaptools with options "mfg" [43]. Frontal gyrus NeuN+/− CpG Differentially Methylated Regions (DMRs) generated by Lister et al. [22] were downloaded from http://brainome.ucsd.edu/BrainMethylomeData/CG_DMR_lists.tar.gz and converted to hg38 using rtracklayer and hg18ToHg38.over.chain. Neuron CpG DMRs for each of the 21 Neuron cluster described by Luo et al. [7] were converted to hg38 using rtracklayer and hg19ToHg38.over.chain. The hg38 locations of Differentially Expressed Genes (DEG's) across non-neuron cell-types identified by Lake et al. [28] were extracted using R software biomart package (v 2.46.3) and were separated into gene body and promoter (1.5 kb upstream TSS). CpG and CpH methylation were summarized for each DMR and DEG across each single-cell using cgmaptools with options "mtr" [43].

Cell-type clustering analysis

We performed non-negative matrix factorization (NMF) on summarized CpH methylation across 100 kb genomic bins and CpG methylation across the Ensembl Regulatory Build [45] setting k = 12, as previously described [9]. CpH and CpG NMF matrices were weighted, merged by cell, and plotted into two-dimensional space using students t-distributed stochastic neighbor embedding (t-SNE). Cell clustering was performed using DBSCAN, as previously described [9] using an epsilon value of 1.3. Clustering analysis was performed using all sciEM single-cells (n = 710) identifying 16 clusters (Fig. 3A). In addition, we evaluated clustering using both sciMET and sciEM single-cells using summarized CpH and CpG (Additional file 1: Fig. S5), summarized CpG alone (Additional file 1: Fig. S6), as well as summarized CpG for sciEM single-cells alone (Additional file 1: Fig. S7). To identify the cell-type of each cluster, sequencing reads of all cell-types within a cluster were collapsed and CpG methylation

Chatterton *et al. Cell & Bioscience*　　(2023) 13:2

Page 10 of 11

was summarized for NeuN+/− DMR's, as described above. Broad subtypes of non-neuronal cells were further classified by CpG methylation summarization of non-neuron DEG promoters, as described above. Non-neuron cell subtypes were defined by lowest (hypo-methylated promoters) z-score (annotation × cluster matrix). Broad excitatory and inhibitory neuron subtypes were classified by CpG methylation summarization of promoter CpG DMRs of 21 neuron subtypes, as described above, and defined by hierarchical clustering of z-scores. We performed a linear regression analysis between neuron (n = 430) and non-neuronal (n = 280) cell-types using per-nuclei CpG, CHG and CHH DNA methylation levels controlling for read depth using R statistic software.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13578-022-00938-9.

> **Additional file 1.** Supplementary material.

Author contributions
ZC conceived the study and coordinated experiments. GH performed post-mortem brain tissue dissections. ZC, PL, DAR and LF performed nuclei isolation, fluorescent activated nuclei sorting and sequencing library construction. ZC performed bioinformatic and statistical analysis. HL and CM performed recombinant protein and isolation experiments. ZC and JBK wrote the manuscript, with contributions from all authors. All authors read and approved the final manuscript.

Availability of data and materials
The dataset(s) supporting the conclusions of this article are available in the SRA repository (SRR22493491).

Declarations

Ethics approval and consent to participate
The use of human brain tissue was approved by the Research Integrity and Ethics Administration of the University of Sydney (project number: 2018/861). The use of a genetically modified (*C9orf72*) mouse was approved by the Garvan Animal Ethics Committee (project number: 16_14).

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Schubeler D. Function and information content of DNA methylation. Nature. 2015;517(7534):321–6.
2. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99(3):247–57.
3. Suelves M, Carrio E, Nunez-Alvarez Y, Peinado MA. DNA methylation dynamics in cellular commitment and differentiation. Brief Funct Genom. 2016;15(6):443–53.
4. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet. 2009;41(12):1350–3.
5. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods. 2014;11(8):817–20.
6. Clark SJ, Smallwood SA, Lee HJ, Krueger F, Reik W, Kelsey G. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). Nat Protoc. 2017;12(3):534–47.
7. Luo C, Keown CL, Kurihara L, Zhou J, He Y, Li J, et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. Science. 2017;357(6351):600–4.
8. Luo C, Rivkin A, Zhou J, Sandoval JP, Kurihara L, Lucero J, et al. Robust single-cell DNA methylome profiling with snmC-seq2. Nat Commun. 2018;9(1):3824.
9. Mulqueen RM, Pokholok D, Norberg SJ, Torkenczy KA, Fields AJ, Sun D, et al. Highly scalable generation of DNA methylation profiles in single cells. Nat Biotechnol. 2018;36(5):428–31.
10. Liu H, Zhou J, Tian W, Luo C, Bartlett A, Aldridge A, et al. DNA methylation atlas of the mouse brain at single-cell resolution. Nature. 2021;598(7879):120–8.
11. Wong NC, Ashley D, Chatterton Z, Parkinson-Bates M, Ng HK, Halemba MS, et al. A distinct DNA methylation signature defines pediatric pre-B cell acute lymphoblastic leukemia. Epigenetics. 2012;7(6):535–41.
12. Chatterton Z, Mendelev N, Chen S, Carr W, Kamimori GH, Ge Y, et al. Bisulfite amplicon sequencing can detect Glia and neuron cell-free DNA in blood plasma. Front Mol Neurosci. 2021;14:140.
13. Rubin AJ, Parker KR, Satpathy AT, Qi Y, Wu B, Ong AJ, et al. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. Cell. 2019;176(1–2):361-376.e17.
14. Guo H, Zhu P, Guo F, Li X, Wu X, Fan X, et al. Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. Nat Protoc. 2015;10(5):645–59.
15. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell. 2008;133(3):523–36.
16. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature. 2008;452(7184):215–9.
17. Grunau C, Clark SJ, Rosenthal A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res. 2001;29(13):E65.
18. Ji L, Sasaki T, Sun X, Ma P, Lewis ZA, Schmitz RJ. Methylated DNA is over-represented in whole-genome bisulfite sequencing data. Front Genet. 2014;5:341.
19. Feng S, Zhong Z, Wang M, Jacobsen SE. Efficient and accurate determination of genome-wide DNA methylation patterns in *Arabidopsis thaliana* with enzymatic methyl sequencing. Epigenet Chromatin. 2020;13(1):42.
20. Ziller MJ, Muller F, Liao J, Zhang Y, Gu H, Bock C, et al. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS Genet. 2011;7(12):e1002389.
21. Haines TR, Rodenhiser DI, Ainsworth PJ. Allele-specific non-CpG methylation of the Nf1 gene during early mouse development. Dev Biol. 2001;240(2):585–98.

Chatterton *et al. Cell & Bioscience*   (2023) 13:2

Page 11 of 11

22. Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, et al. Global epigenomic reconfiguration during mammalian brain development. Science. 2013;341(6146):1237905.

23. Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci. 2014;17(2):215–22.

24. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature. 2011;471(7336):68–73.

25. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315–22.

26. Vaisvila R, Ponnaluri VKC, Sun Z, Langhorst BW, Saleh L, Guan S, et al. EM-seq: detection of DNA methylation at single base resolution from picograms of DNA. BioRxiv. 2020. https://doi.org/10.1101/2019.12.20. 884692.

27. Reddington JP, Perricone SM, Nestor CE, Reichmann J, Youngson NA, Suzuki M, et al. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of polycomb target genes. Genome Biol. 2013;14(3):R25.

28. Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science. 2016;352(6293):1586–90.

29. Karemaker ID, Vermeulen M, Single-Cell DNA. Methylation profiling: technologies and biological applications. Trends Biotechnol. 2018;36(9):952–65.

30. Plongthongkum N, Diep DH, Zhang K. Advances in the profiling of DNA modifications: cytosine methylation and beyond. Nat Rev Genet. 2014;15(10):647–61.

31. Eberwine J, Sul JY, Bartfai T, Kim J. The promise of single-cell sequencing. Nat Methods. 2014;11(1):25–7.

32. Kozlenkov A, Roussos P, Timashpolsky A, Barbu M, Rudchenko S, Bibikova M, et al. Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. Nucleic Acids Res. 2014;42(1):109–27.

33. Kozlenkov A, Wang M, Roussos P, Rudchenko S, Barbu M, Bibikova M, et al. Substantial DNA methylation differences between two major neuronal subtypes in human brain. Nucleic Acids Res. 2016;44(6):2593–612.

34. Ziffra RS, Kim CN, Ross JM, Wilfert A, Turner TN, Haeussler M, et al. Single-cell epigenomics reveals mechanisms of human cortical development. Nature. 2021;598(7879):205–13.

35. Butcher LM, Ito M, Brimpari M, Morris TJ, Soares FAC, Ahrlund-Richter L, et al. Non-CG DNA methylation is a biomarker for assessing endodermal differentiation capacity in pluripotent stem cells. Nat Commun. 2016;7:10458.

36. Picelli S, Bjorklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 2014;24(12):2033–40.

37. Adey A, Shendure J. Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. Genome Res. 2012;22(6):1139–43.

38. Rykalina V, Shadrin A, Lehrach H, Borodina T. qPCR-based characterization of DNA fragmentation efficiency of Tn5 transposomes. Biol Methods Protoc. 2017;2(1):bpx001.

39. Matevossian A, Akbarian S. Neuronal nuclei isolation from human post-mortem brain tissue. J Vis Exp. 2008. https://doi.org/10.3791/914.

40. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17(1):3.

41. Wu P, Gao Y, Guo W, Zhu P. Using local alignment to enhance single-cell bisulfite sequencing data efficiency. Bioinformatics. 2019;35(18):3273–8.

42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

43. Guo W, Zhu P, Pellegrini M, Zhang MQ, Wang X, Ni Z. CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. Bioinformatics. 2018;34(3):381–7.

44. Karimzadeh M, Ernst C, Kundaje A, Hoffman MM. Umap and Bismap: quantifying genome and methylome mappability. Nucleic Acids Res. 2018;46(20):e120.

45. Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR. The ensembl regulatory build. Genome Biol. 2015;16:56.

Publisher's Note