Developmental disruption to the cortical transcriptome and synaptosome in a model of *SETD1A* loss-of-function

Nicholas E Clifton[1,2,3,*], Matthew L Bosworth[1], Niels Haan[3], Elliott Rees[1], Peter A Holmans[1], Lawrence S Wilkinson[3], Anthony R Isles[1], Mark O Collins[4], Jeremy Hall[1,3]

[1]MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Maindy Road, Cardiff, UK

[2]University of Exeter Medical School, University of Exeter, Exeter, UK

[3]Neuroscience and Mental Health Research Institute, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Maindy Road, Cardiff, UK

[4]School of Biosciences, University of Sheffield, Western Bank, Sheffield, UK

* Corresponding author

Dr Nicholas E Clifton
University of Exeter Medical School
University of Exeter
Exeter
EX1 2HZ
Email: n.clifton@exeter.ac.uk

Abstract

Large-scale genomic studies of schizophrenia implicate genes involved in the epigenetic regulation of transcription by histone methylation and genes encoding components of the synapse. However, the interactions between these pathways in conferring risk to psychiatric illness are unknown. Loss-of-function (LoF) mutations in the gene encoding histone methyltransferase, SETD1A, confer substantial risk to schizophrenia. Among several roles, SETD1A is thought to be involved in the development and function of neuronal circuits. Here, we employed a multi-omics approach to study the effects of heterozygous *Setd1a* LoF on gene expression and synaptic composition in mouse cortex across five developmental timepoints from embryonic day 14 to postnatal day 70. Using RNA sequencing, we observed that *Setd1a* LoF resulted in the consistent downregulation of genes enriched for mitochondrial pathways. This effect extended to the synaptosome, in which we found age-specific disruption to both mitochondrial and synaptic proteins. Using large-scale patient genomics data, we observed no enrichment for genetic association with schizophrenia within differentially expressed transcripts or proteins, suggesting they derive from a distinct mechanism of risk from that implicated by genomic studies. This study highlights biological pathways through which *SETD1A* loss-of-function may confer risk to schizophrenia. Further work is required to determine whether the effects observed in this model reflect human pathology.

Introduction

Schizophrenia is a leading cause of disability in young adults and many patients remain insufficiently treated by current antipsychotics (1). Our understanding of the molecular mechanisms associated with risk for schizophrenia will be crucial for targeting new therapies. A string of recent genomic studies has unearthed hundreds of genomic loci each contributing

small amounts of risk (2–5), improving power for the identification of relevant molecular pathways whilst complicating the recapitulation of their effects in model organisms. However, through advances in exome sequencing, a small number of single genes were identified containing a genome-wide excess of highly penetrant coding mutations in patients (6, 7). This discovery greatly increases the feasibility of studying pathology relevant to schizophrenia in model organisms.

Rare loss-of-function (LoF) variants in the *SETD1A* gene, encoding SET Domain Containing 1A, confer substantial risk to schizophrenia and other neurodevelopmental disorders (6–8). The SETD1A protein catalyses histone H3 (K4) methylation to mediate the expression of target genes. This lends support to the growing evidence that regulation of histone methylation is a point of convergence for genes conferring risk to neuropsychiatric disorders (9). SETD1A is required from very early in development for epigenetic control of the cell cycle and maintaining genome stability (10–13), but remains expressed in brain tissue throughout prenatal and postnatal life and appears to be required for normal neurite outgrowth, neuronal excitability and cognitive function (14–17). These observations suggest that *SETD1A* LoF may impact synaptic structure and function, and expose a mechanism through which risk to schizophrenia might be conferred.

Just as epigenetic control of gene expression is dynamic across development (18–21), the composition of the synapse varies considerably during brain maturation (22–24) as neurons migrate (25, 26), form connections and mature. To explore the biological pathways through which *SETD1A* contributes to risk for schizophrenia, we quantified gene expression and synaptosome composition in the frontal cortex of mice carrying a *Setd1a* LoF allele at multiple prenatal and postnatal stages of development.

Results

Frontal cortex differential gene expression in *Setd1a*$^{+/-}$ mice

Heterozygous knockout of *Setd1a* resulted in loss of approximately 50% Setd1a protein in cortical tissue compared to wildtype controls, reported previously (27). We performed RNA sequencing on 50 high-quality libraries (median RNA integrity = 9.35; Supplementary Figure S1, Supplementary Table S1) from frontal cortex of *Setd1a*$^{+/+}$ and *Setd1a*$^{+/-}$ mice across five developmental timepoints (E14-P70; Figure 1A). We analysed the expression of 16001 protein-coding genes expressed during at least one timepoint. In wildtype frontal cortex samples, *Setd1a* was expressed at all ages, consistent with human expression at matched developmental timepoints (Figure 1B).

We used interaction analyses to identify any genes for which the effect of genotype differed by age. We observed no significant interaction terms after correction for FDR (Supplementary Table S2), despite an overall negative correlation between the differential expression at E14 and E18, as indexed by the log fold change (Supplementary Figure S2). However, in genotype contrasts, controlling for age, we observed 734 genes differentially expressed (FDR < 0.05) between wildtype and *Setd1a*$^{+/-}$ tissue (Figure 1C; Supplementary Table S3). The mutation led to considerably more downregulated genes (N = 616) than upregulated genes (N = 118).

Differentially expressed genes (DEGs) were enriched for seven GO pathways predominantly relating to mitochondrial function (Figure 1D, Supplementary Table S4A). 144 differentially expressed genes intersecting with the GO:0005739 *Mitochondrion* term are listed in Supplementary Table S4B. Mitochondrial pathways were only enriched among downregulated genes (Supplementary Table S5). No GO pathways were significantly enriched in upregulated genes following Bonferroni correction (Supplementary Table S6). Using protein-protein

interactions data, we identified a core network of proteins encoded by the DEGs, consisting of a central group from the mitochondrial NADH:ubiquinone oxidoreductase respiratory complex I and surrounding assembly factors (Figure 1E). The rate of sequencing reads mapping to the mitochondrial genome did not differ by genotype (Supplementary Figure S3). In keeping with the lack of genotype-by-age interaction, the most significant mitochondrial DEGs were consistently downregulated across all the developmental timepoints examined (Figure 1F).

We investigated the consistency of our results with previous studies of *Setd1a* haploinsufficiency. 617 downregulated genes observed in a human neuroblastoma cell line following knockdown of *SETD1A* were also significantly enriched for GO:0005739 *Mitochondrion* genes (28). Of these, 454 genes had unique murine brain-expressed homologs, in which we observed an overlap of 68 genes with our downregulated gene set (Fisher's exact Test: odds ratio = 4.76; $P = 9.6 \times 10^{-22}$). Conversely, 342 DEGs observed following *Setd1a* heterozygous knockout in 6-week-old mouse prefrontal cortex (14) showed proportionally less overlap (21 genes) with our DEGs, and did not exceed the chance level of overlap in Fisher's exact Test (odds ratio = 1.44; $P = 0.11$). The same study also employed chromatin immunoprecipitation and sequencing (ChIP-seq) to identify direct targets of Setd1a on promoter or enhancer regions predicted to mediate gene expression (14). Using this data, we mapped Setd1a target peaks to promoter regions in 4970 genes and enhancer regions in 3738 genes. Notably, the GO term most significantly overrepresented among our DEGs following heterozygous *Setd1a* knockout – GO:0005739 *Mitochondrion* – was also strongly enriched among genes harbouring promoter regions targeted by Setd1a (odds ratio = 1.42; *P*.bonferroni $= 6.1 \times 10^{-4}$). Furthermore, based on this data, 236 of our downregulated genes are targeted by Setd1a at promotor regions (Fisher's exact Test: odds ratio = 1.30; $P = 0.0015$). This lends strength to the possibility that *Setd1a* loss-of-function caused dysregulation of mitochondrial

pathways through direct effects on gene regulation. Genes containing enhancer regions targeted by Setd1a were not enriched for *Mitochondrion* genes (odds ratio = 0.60; *P*.bonferroni = 1.0).

Dysregulation of synaptosomal transcripts in *Setd1a*$^{+/-}$ mice

To examine the effect of the *Setd1a* LoF allele on the regulation of synaptic components, we quantified changes in gene and protein expression relating to the synaptosomal fraction of frontal cortical tissue across the same timepoints (Figure 2A). Using mass spectrometry-based label-free quantitation of isolated synaptosomes, we observed 3653 protein groups present in samples from at least one timepoint, after filtering. Within-sample comparisons of RNA and protein expression revealed good overall correlation (Supplementary Figure S4). Of the 734 DEGs from previous transcriptomic analysis of genotype effects, 127 (106 downregulated, 21 upregulated) encode proteins observed at the synaptosome (Supplementary Table S7). More than half (58) of the downregulated synaptosomal genes were members of GO:0005739 *Mitochondrion*, indicating a strongly significant overrepresentation (odds ratio = 4.99; *P*.bonferroni = 2.9 x 10^{-11}; Supplementary Table S7). No other GO terms were significantly enriched among the downregulated synaptosomal genes. Again, no GO terms were enriched in the upregulated fraction (Supplementary Table S8).

To predict whether *Setd1a* LoF preferentially impacted on mitochondria situated at the synapse, we compared our DEGs to published proteomic data (29) describing the relative abundance of proteins in synaptic vs non-synaptic mitochondrial proteomes. 48 downregulated genes observed in genotype contrasts here encode proteins quantified from mitochondrial proteomes. Of these, 11 were enriched (log fold change > 1) in synaptic mitochondria and 11 were enriched in non-synaptic mitochondria (Figure 2B). The ratio between these is no greater than the overall proportion of proteins enriched in synaptic mitochondria (odds ratio = 0.60; *P*

= 0.92), suggesting the effects of *Setd1a* LoF on mitochondrial components were not specific to synapses but distributed between synaptic and non-synaptic compartments.

Disruption to the synaptosomal proteome in *Setd1a*$^{+/-}$ mice

We tested the effect of the *Setd1a*$^{+/-}$ genotype on the synaptosomal proteome. To identify synaptosomal proteins for which the change in abundance over time was affected by *Setd1a* LoF, we contrasted the difference in protein expression between all pairs of consecutive timepoints in mutant and wildtype samples. The change in developmental expression of two proteins, Kng1 and Ndufa3, differed by genotype (Supplementary Table S9). By examining each contrast, we observed that synaptosomal Kng1 intensity was significantly affected by genotype between E14 and E18 ($t = 2.09$, $P = 0.042$), E18 to P7 ($t = -2.12$, $P = 0.040$) and P35 to P70 ($t = -4.92$, $P = 1.3 \times 10^{-5}$). Ndufa3 was affected by genotype between P7 and P35 ($t = -2.48$, $P = 0.017$) and from P35 to P70 ($t = 5.75$, $P = 8.3 \times 10^{-7}$). Analysing across all timepoints, 6 proteins were significantly altered by genotype (Supplementary Table S10): Synaptotagmin-2 (Syt2), Kininogen-1 (Kng1), NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 3 (Ndufa3), Semaphorin-4C (Sema4c), Transcriptional activator protein Pur-alpha (Pura) and Mitochondrial ribosomal protein L16 (Mrpl16) (Figure 2C). Notably, transcripts encoding Ndufa3 were also differentially expressed in transcriptomic analysis of genotype effects described above (Supplementary Table S2). The functions of these proteins are summarised in Table 1.

To examine the effect of *Setd1a* LoF at different developmental stages more closely, we performed genotype contrasts at each age independently, yielding a set of differentially expressed synaptosomal proteins for each timepoint (Figure 2D; Supplementary Table S11-15). As a whole, differential protein expression, as indexed by the log fold change, was poorly correlated between different stages of development (Supplementary Figure S2), and with

differential gene expression compared at each stage (Supplementary Figure S5). Significantly upregulated or downregulated proteins were further annotated by predictions of their relative abundance in the postsynaptic density (PSD) compared to the total synaptosome, based on previously published data (30). At all timepoints, differentially expressed PSD-enriched proteins were downregulated in *Setd1a*$^{+/-}$ samples compared to wildtype. Presynaptic protein, Syt2, was strongly decreased at E18, such that the developmental upregulation observed in wildtypes was delayed in mutant cortex (Figure 2C). To obtain more biological insight into the types of proteins affected, we performed pathway analysis of those differentially expressed at any timepoint (N = 63), using a background of all synaptosomal proteins. We observed no significantly enriched pathways after multiple testing correction (Supplementary Table S16). It is notable, however, that genes belonging to the top term by significance in DEG pathway analyses (GO:0005739 *Mitochondrion*) represented the highest proportion of differentially expressed proteins (19 proteins), with a nominally significant enrichment (odds ratio: 1.70; *P*.unadjusted = 0.043). 17 of these *Mitochondrion* proteins were differentially expressed at E18 (Supplementary Table S12), 12 of which were downregulated.

Genetic association with schizophrenia of transcripts and proteins disrupted by *Setd1a* loss-of-function

We hypothesised that molecular pathways disrupted by *Setd1a* LoF also contribute risk for schizophrenia through enrichment for genetic association with the disorder. We tested this using case-control data from GWAS (PGC3) and exome sequencing studies. Through gene set association analyses of DEGs and affected proteins across brain development, including subsets defined by membership of the synaptosome or GO:0005739 *Mitochondrion*, we observed no significant enrichment for genetic association with schizophrenia in any gene set, through common or *de novo* rare variants (Table 2).

Due to the small number of significantly differentially expressed proteins, we performed an additional test of genetic association with schizophrenia in gene sets ranked by the probability of differential protein expression in the synaptosome, determined from the main effect of genotype. Gene sets ranked highest for differential protein expression were not enriched for association with schizophrenia through common variation or *de novo* rare variation (Supplementary Figure S6), suggesting that *Setd1a* LoF does not preferentially disrupt synaptosomal proteins that contribute additional genetic risk to schizophrenia.

Discussion

Understanding the biological effects of highly penetrant genetic mutations conferring risk to schizophrenia is crucial for unravelling pathology and improving treatments. We modelled a heterozygous *Setd1a* LoF allele in mice and profiled RNA and protein from frontal cortical tissue across multiple pre- and postnatal developmental timepoints. The mutation caused downregulation of transcripts predominantly enriched for mitochondrial function, irrespective of age. Using mass spectrometry-based protein quantitation, we further examined the effects of the *Setd1a* variant on the constituents of the synaptosome and revealed subsets of proteins disrupted at each timepoint.

Transcriptomic data from *Setd1a*$^{+/-}$ mouse cortex showed evidence of disruption to respiratory chain complex I, mitochondrial assembly proteins and mitochondrial translation. Disruption by *Setd1a* haploinsufficiency of mitochondrial and metabolic functions characterised by a downregulation of associated nuclear transcripts is consistent with previous studies in human neuroblastoma (28). Altered metabolism following *Setd1a* deletion was also observed in hematopoitic stem cells (10). Similarly, loss of other SET domain-containing proteins, including the Set1 ortholog, *Setd1b*, and *Setd5*, induced downregulation of mitochondrial and metabolic pathways (31, 32), together supporting a role of chromatin modifications by this

protein family in regulating mitochondrial function. Whilst further work is needed to establish the nature of this relationship, we report that promoter regions targeted by Setd1a (14) are enriched in genes with functional annotations related to mitochondria, thereby providing evidence of a direct causal relationship.

Oxidative phosphorylation in mitochondria supplies the high metabolic demand of synaptic activity in neurons, and it has been suggested that mitochondrial dysfunction could cause progressive developmental synaptic pathology in schizophrenia (33–37). Fast-spiking parvalbumin interneurons, which have been recurrently implicated in schizophrenia, contain high densities of mitochondria, are highly susceptible to oxidative stress, and may be particularly vulnerable to metabolic disruption (38–40). Mitochondrial dysfunction in schizophrenia is supported by a range of studies presenting transcriptomic, proteomic and metabolomic evidence of reduced mitochondrial activity, predominantly relating to components of respiratory complex I, in post-mortem brain, peripheral tissues and induced pluripotent stem cells from patients (41–49). Genetic studies show some evidence of a burden of patient copy number variants (CNVs) in mitochondrial genes (50), yet the largest schizophrenia GWAS to date reported no enrichment for common genetic association in mitochondrial pathways (5). Consistent with this, we observed no enrichment in differentially expressed mitochondrial genes for schizophrenia association conferred by common variants or *de novo* rare nonsynonymous variants. Hence, through quantifying the biological effects of a highly penetrant schizophrenia risk variant, our study informs potential functional pathways of risk that are not illuminated by primary genetic studies alone.

From previous work, it has been suggested that Setd1a has additional roles in synaptic function and development, and its loss-of-function leads to deficits in working memory (14, 15). However, unlike these previous studies, we observed no enrichment of neuron-specific

functional annotations among DEGs in mutant samples. After restricting our transcriptomic analysis to genes encoding proteins detected in the synaptosome, we found that downregulated transcripts remained enriched for mitochondrial function. Through proteomic analysis, we observed multiple downregulated mitochondrial proteins, principally at E18, coinciding with a critical period of neuronal maturation and synaptogenesis (51). This paralleled a delay in the developmental upregulation of presynaptic neurotransmitter release protein Syt2, suggesting abnormal synaptic maturation in *Setd1a*$^{+/-}$ cortical samples. However, whether the effects on Syt2 and other (non-mitochondrial) synaptic proteins were primary or secondary to *Setd1a* or mitochondrial dysfunction is undetermined.

Despite the apparent disruption to mitochondrial pathways in the synaptosome, we found evidence that *Setd1a* haploinsufficiency impacted synaptic and non-synaptic mitochondria. Therefore, any metabolic consequences of the mutation may be equally likely to influence non-neuronal cell types and other cellular compartments not examined in this study. Poor overall correlation between differential protein expression in the synaptosome and tissue-wide differential gene expression also suggests that many of the transcriptomic effects of the *Setd1a* variant influence non-synaptic compartments. However, due to the high metabolic demands of neurotransmission, synaptic systems may be more sensitive to small changes in mitochondrial activity than other cellular processes (52). Future work using tissue- or cell-specific omics may seek to further characterise cortical metabolic abnormalities caused by *Setd1a* LoF.

DEGs observed here in genotype contrasts exhibited poor overlap with those derived from a previous transcriptomic study of adult *Setd1a*$^{+/-}$ mouse prefrontal cortex (14), which in turn were inconsistent with a third transcriptomic study of *Set1a* haploinsufficiency (15), together with the biological pathways annotated to them. Whilst we extended the investigation to multiple developmental stages, we found no significant effect of age on the differential

expression signature. Critically, each of these three studies were performed using different mouse models, and whilst each resulted in the reduction of Setd1a protein in frontal brain regions by approximately 50% and the induction of schizophrenia-related behavioural phenotypes (14, 15, 27), their effects on particular isoforms or compensatory mechanisms may have differed. Further differences in tissue extraction and library preparation methods could also contribute.

To conclude, our results give evidence of disruption to nuclear-encoded mitochondrial pathways in cortical tissue throughout brain development caused by modelling a *SETD1A* LoF allele that confers substantial risk to schizophrenia. Our findings therefore support the premise of mitochondrial perturbation in psychiatric pathology and expose biological consequences of genetic risk that are not themselves predicted by genetic association studies. We further highlight a subset of synaptic proteins that may be key to understanding neural dysfunction induced by this variant.

Materials and Methods

Subjects and Tissue Preparation

Mice carrying a heterozygous *Setd1a*^{tm1d} loss-of-function allele, with mixed C57BL/6NTac and C57BL/6J background, were generated using a knockout-first design and genotyped as described previously (27). Heterozygous males were paired with wildtype females to generate male experimental subjects at embryonic day 14.5 (E14.5), E18.5, postnatal day 7 (P7), P35 and P70 (N = 5 per genotype per timepoint). Timed matings, determined by plug checks, were used for embryonic timepoints. For P35 and P70 timepoints, offspring were weaned at P28 and housed in single-sex groups. All animals were provided with environmental enrichment, food and water *ad libitum* and maintained at 21°C and 50% humidity with a 12-hour light-dark

cycle. All procedures were conducted in accordance with the United Kingdom Animals (Scientific Procedures) Act 1986 (PPL 30/3375).

At embryonic timepoints, pregnant dams were killed and frontal brain regions immediately dissected from embryos. At postnatal timepoints, littermates were killed and frontal cortex dissected. Brain tissue was snap frozen before storage at -80°C until processing. Bilateral frontal cortices were homogenised using a Dounce homogenizer in Synaptic Protein Extraction Reagent (SynPER, Thermofisher). A fraction of the homogenised sample was taken forward for RNA extraction and the remaining used for synaptosome extraction.

Synaptosome isolation

Synaptosomes were isolated from homogenised cortical tissue using the SynPER protocol, as per the manufacturer's instructions. Briefly, following homogenisation, samples were centrifuged at 1200g for 10 min (4°C) and the pellet discarded. The supernatant was centrifuged again at 15,000g for 20 minutes (4°C) to generate the synaptosome pellet. We resuspended the pellet in 2% SDS, 50 mM Tris pH 7.4 and heated at 70°C for 15 min to extract the protein. Samples were clarified by centrifugation at 20,000g for 10 min.

Transcriptomics

RNA was extracted using an AllPrep DNA/RNA micro kit (QIAGEN) before quantitation and checks for integrity, degradation and contamination. Samples with < 0.5 µg total RNA were replaced. Library preparation and sequencing were performed by Novogene. cDNA libraries with 250-300 bp inserts were prepared using poly-A capture. A single batch of Illumina high-throughput sequencing was performed at 12Gb read depth per sample with 150bp paired-end reads (~40 million paired-end reads).

Raw sequencing reads were trimmed of adapters using Trimmomatic (53) and passed through FastQC quality control (54). Reads were aligned to the mouse genome (GRCm38) with STAR (55) and mapped to genes using featureCounts (56). Processed read counts were filtered for protein-coding genes. EdgeR (57) was used to determine and exclude unexpressed genes, and perform trimmed mean of M values (TMM) normalisation (58). Expressed genes were defined as having at least 10 counts-per-million in at least 5 samples. Differential expression analyses were performed with limma (59). In primary analyses, we tested for genotype effects that varied by age by fitting an age × genotype interaction, coding age as a 5-level factor. In subsequent analysis, gene expression was regressed on genotype, covarying for age. False discovery rate (FDR) was corrected for using the Benjamini-Hochberg method.

Postmortem human prefrontal cortex *Setd1a* expression data across the lifespan was obtained from the BrainSeq Phase I database (http://eqtl.brainseq.org/phase1/) (60). Samples were filtered for individuals with no history of psychiatric condition. Raw gene counts were converted to reads per kilobase of transcript per million mapped reads (RPKM) and averaged across five developmental stages: Late midfetal (17-23 post-conceptual weeks; N = 13), Late fetal (24-37 post-conceptual weeks; N = 3), Childhood (1-12 years; N = 16), Adolescence (13-19 years; N = 47), Adulthood (20-85 years, N = 202).

Quantitative mass spectrometry analysis

50 ug of synaptosome samples were solubilised with 5% SDS, 100 mM TEAB pH 8 and reduced using 10 mM TCEP with heating at 70°C for 15 minutes. Samples were alkylated with 20 mM Iodoacetamide for 30 minutes at 37°C. Protein was precipitated in solution, trapped and washed on S-trap micro spin columns (ProtiFi, LLC) according to the manufacturer's instructions. Protein was digested using 5 µg trypsin sequence grade (Pierce) at 47°C for 1 hour and 37°C for 1 hour. Eluted peptides were dried in a vacuum concentrator and resuspended in

0.5% formic acid for LC-MS/MS analysis. Peptides were analysed using nanoflow LC-MS/MS using an Orbitrap Elite (Thermo Fisher) hybrid mass spectrometer equipped with a nanospray source, coupled to an Ultimate RSLCnano LC System (Dionex). Peptides were desalted online using a nano trap column, 75 μm I.D.X 20mm (Thermo Fisher) and then separated using a 120-min gradient from 5 to 35% buffer B (0.5% formic acid in 80% acetonitrile) on an EASY-Spray column, 50 cm × 50 μm ID, PepMap C18, 2 μm particles, 100 Å pore size (Thermo Fisher). The Orbitrap Elite was operated with a cycle of one MS (in the Orbitrap) acquired at a resolution of 120,000 at m/z 400, with the top 20 most abundant multiply charged (2+ and higher) ions in a given chromatographic window subjected to MS/MS fragmentation in the linear ion trap. An FTMS target value of 1e6 and an ion trap MSn target value of 1e4 were used with the lock mass (445.120025) enabled. Maximum FTMS scan accumulation time of 500 ms and maximum ion trap MSn scan accumulation time of 100 ms were used. Dynamic exclusion was enabled with a repeat duration of 45 s with an exclusion list of 500 and an exclusion duration of 30 s. Raw mass spectrometry data were analysed with MaxQuant version 1.6.10.43 (61). Data were searched against a mouse UniProt reference proteome (downloaded May 2020) using the following search parameters: digestion set to Trypsin/P, methionine oxidation and N-terminal protein acetylation as variable modifications, cysteine carbamidomethylation as a fixed modification, match between runs enabled with a match time window of 0.7 min and a 20-min alignment time window, label-free quantitation (LFQ) was enabled with a minimum ratio count of 2, minimum number of neighbours of 3 and an average number of neighbours of 6. A protein FDR of 0.01 and a peptide FDR of 0.01 were used for identification level cut-offs based on a decoy database searching strategy. This protocol yielded synaptic proteomes with comparable composition to those observed previously in mice (30).

Protein groups were converted to single proteins by prioritising those explaining the most data. These 5142 proteins were filtered to include only those detected in at least 4 of 5 samples from

at least one experimental group, giving 3710 proteins for analysis. Raw LFQ intensity values were log converted and scaled by median intensity normalization. Missing values were imputed from a normal distribution (mean = μ - 1.8; standard deviation = σ × 0.3). Genotype contrasts were performed using limma (59). In primary analyses, all within-age genotype contrasts were tested in the same linear model to determine the mean effect of genotype across development for each protein. In interaction analyses, the difference in protein intensity between all pairs of consecutive timepoints was contrasted between wildtype and *Setd1a*$^{+/-}$ samples, as described previously (62). For any significant interactions following correction for FDR ($P < 0.05$), the interaction terms from each pair of consecutive timepoints were extracted individually to identify specific periods when the protein is affected by genotype. In secondary analyses, genotype contrasts were performed at each age independently.

Data describing the relative abundance of proteins in synaptic vs non-synaptic mitochondria were acquired from a study of neuronal bioenergetic control in adult rat forebrain (29).

The localisation of proteins to presynaptic or postsynaptic fractions of the synaptosome was predicted *in-silico* using a previous report of synaptic protein enrichment or depletion in postsynaptic density compared to synaptosome preparations from mouse brain (30).

Pathway analysis

Functional annotations of genes were compiled from the Gene Ontology (GO) database (June 8, 2021), excluding gene annotations with evidence codes IEA (inferred from electronic annotation), NAS (non-traceable author statement), or RCA (inferred from reviewed computational analysis). GO terms annotated to fewer than 10 genes were excluded, leaving 8557 terms used in pathway analyses. Comparisons between gene or protein sets were made using the mouse Ensembl ID (63). Enrichment of gene sets derived from differential expression analysis for GO annotations, or other functionally-defined gene sets, was determined by

Fisher's exact test, whereby all remaining tissue-expressed genes or proteins were used as the statistical background. Multiple testing was corrected for using the Bonferroni method.

Protein-protein interaction networks were compiled using GeneMANIA (64). Networks were filtered to include only physical interactions, and exclude interactions defined by co-expression, co-localization, shared domains or predictions.

Mapping of Setd1a targets to genes

Data containing predicted Setd1a genomic binding sites were obtained from a recent study (14) of Setd1a targets in 6-week-old mouse prefrontal cortex using chromatin immunoprecipitation and sequencing (ChIP-seq). Setd1a peaks located at promoter or enhancer regions were mapped to genes using the mm10 mouse genome assembly. Peaks mapping to zero or multiple genes were excluded.

Genetic association analysis

Recent schizophrenia case-control genome-wide association study (GWAS) summary statistics were provided by the Psychiatric Genomics Consortium. The primary GWAS consisted of 69,369 cases and 94,015 controls of European or Asian descent (5). Single nucleotide polymorphisms (SNPs) with minor allele frequency greater than 1% were annotated to genes using a 35kb upstream / 10kb downstream window to allow for proximal regulatory regions. SNP association *P*-values were combined in MAGMA v1.08 (65) using the SNP-wise Mean model, controlling for linkage disequilibrium with the 1000 Genomes European reference panel (66). Gene set association analysis was performed using one-tailed competitive tests in MAGMA, conditioning on a background of tissue-expressed genes.

De novo coding variants observed in people diagnosed with schizophrenia were taken from published exome sequencing studies. In total, de novo variant data were derived from 3444

published schizophrenia-proband parent trios (67–76), as described previously (68, 77, 78). Gene set enrichment statistics were generated by a two-sample Poisson rate ratio test comparing the ratio of observed vs expected *de novo* variants in the gene set to a background set of genes. Expected numbers of variants were determined from per-gene mutation rates (79). The background set contained all tissue-expressed genes.

Data Availability

Transcriptomic data from RNA sequencing is available from the Gene Expression Omnibus (GEO) with identifier GSE199428. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (80) partner repository with the dataset identifier PXD032742.

Acknowledgements

This research was supported by Medical Research Council (MRC) grant MR/R011397/1 and Wellcome Trust grant 100202/Z/12/Z. The MRC Harwell Institute generated the C57BL/6NTac-Setd1a$^{tm1c(EUCOMM)Wtsi}$/WtsiCnrm mouse strain from which we derived the strain used in this study. We thank the Psychiatric Genomics Consortium for providing genome-wide association study summary statistics.

Conflict of Interest Statement

The authors report no conflicts of interest.

439 Citations

440 1. Vigo,D., Thornicroft,G. and Atun,R. (2016) Estimating the true global burden of mental
441 illness. *Lancet Psychiatry*, **3**, 171–178.

442 2. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological
443 insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421–427.

444 3. Pardiñas,A.F., Holmans,P., Pocklington,A.J., Escott-Price,V., Ripke,S., Carrera,N.,
445 Legge,S.E., Bishop,S., Cameron,D., Hamshere,M.L., *et al.* (2018) Common
446 schizophrenia alleles are enriched in mutation-intolerant genes and in regions under
447 strong background selection. *Nat. Genet.*, **50**, 381–389.

448 4. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium (2011)
449 Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.*,
450 **43**, 969–976.

451 5. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke,S.,
452 Walters,J.T. and O'Donovan,M.C. (2020) Mapping genomic loci prioritises genes and
453 implicates synaptic biology in schizophrenia. *medRxiv*,
454 10.1101/2020.09.12.20192922.

455 6. Singh,T., Poterba,T., Curtis,D., Akil,H., Al Eissa,M., Barchas,J.D., Bass,N., Bigdeli,T.B.,
456 Breen,G., Bromet,E.J., *et al.* (2020) Exome sequencing identifies rare coding variants
457 in 10 genes which confer substantial risk for schizophrenia. *medRxiv*,
458 10.1101/2020.09.18.20192815.

459 7. Singh,T., Kurki,M.I., Curtis,D., Purcell,S.M., Crooks,L., McRae,J., Suvisaari,J.,
460 Chheda,H., Blackwood,D., Breen,G., *et al.* (2016) Rare loss-of-function variants in
461 SETD1A are associated with schizophrenia and developmental disorders. *Nat.*
462 *Neurosci.*, **19**, 571–577.

463 8. Takata,A., Ionita-Laza,I., Gogos,J.A., Xu,B. and Karayiorgou,M. (2016) De novo
464 synonymous mutations in regulatory elements contribute to the genetic etiology of
465 autism and schizophrenia. *Neuron*, **89**, 940–947.

466 9. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium (2015)

Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.*, **18**, 199–209.

10. Arndt,K., Kranz,A., Fohgrub,J., Jolly,A., Bledau,A.S., Di Virgilio,M., Lesche,M., Dahl,A., Höfer,T., Stewart,A.F., *et al.* (2018) SETD1A protects HSCs from activation-induced functional decline in vivo. *Blood*, **131**, 1311–1324.

11. Tajima,K., Matsuda,S., Yae,T., Drapkin,B.J., Morris,R., Boukhali,M., Niederhoffer,K., Comaills,V., Dubash,T., Nieman,L., *et al.* (2019) SETD1A protects from senescence through regulation of the mitotic gene expression program. *Nat. Commun.*, **10**, 2854.

12. Higgs,M.R., Sato,K., Reynolds,J.J., Begum,S., Bayley,R., Goula,A., Vernet,A., Paquin,K.L., Skalnik,D.G., Kobayashi,W., *et al.* (2018) Histone Methylation by SETD1A Protects Nascent DNA through the Nucleosome Chaperone Activity of FANCD2. *Mol. Cell*, **71**, 25–41.e6.

13. Tajima,K., Yae,T., Javaid,S., Tam,O., Comaills,V., Morris,R., Wittner,B.S., Liu,M., Engstrom,A., Takahashi,F., *et al.* (2015) SETD1A modulates cell cycle progression through a miRNA network that regulates p53 target genes. *Nat. Commun.*, **6**, 8257.

14. Mukai,J., Cannavò,E., Crabtree,G.W., Sun,Z., Diamantopoulou,A., Thakur,P., Chang,C.-Y., Cai,Y., Lomvardas,S., Takata,A., *et al.* (2019) Recapitulation and Reversal of Schizophrenia-Related Phenotypes in Setd1a-Deficient Mice. *Neuron*, **104**, 471–487.e12.

15. Nagahama,K., Sakoori,K., Watanabe,T., Kishi,Y., Kawaji,K., Koebis,M., Nakao,K., Gotoh,Y., Aiba,A., Uesaka,N., *et al.* (2020) Setd1a Insufficiency in Mice Attenuates Excitatory Synaptic Function and Recapitulates Schizophrenia-Related Behavioral Abnormalities. *Cell Rep.*, **32**, 108126.

16. Wang,S., van Rhijn,J.-R., Akkouh,I.A., Kogo,N., Maas,N., bleeck,A., Lewerissa,E., Wu,K.M., Schoenmaker,C., Djurovic,S., *et al.* (2021) Loss-of-function variants in the schizophrenia risk gene *SETD1A* alter neuronal network activity in human neurons through cAMP/PKA pathway. *BioRxiv*, 10.1101/2021.05.25.445613.

17. Kummeling,J., Stremmelaar,D.E., Raun,N., Reijnders,M.R.F., Willemsen,M.H., Ruiterkamp-Versteeg,M., Schepens,M., Man,C.C.O., Gilissen,C., Cho,M.T., *et al.*

(2020) Characterization of SETD1A haploinsufficiency in humans and Drosophila defines a novel neurodevelopmental syndrome. *Mol. Psychiatry*, 10.1038/s41380-020-0725-5.

18. John,R.M. and Rougeulle,C. (2018) Developmental epigenetics: phenotype and the flexible epigenome. *Front. Cell Dev. Biol.*, **6**, 130.

19. Isles,A.R. (2018) Epigenetics, chromatin and brain development and function. *Brain Neurosci. Adv.*, **2**, 239821281881201.

20. Perzel Mandell,K.A., Price,A.J., Wilton,R., Collado-Torres,L., Tao,R., Eagles,N.J., Szalay,A.S., Hyde,T.M., Weinberger,D.R., Kleinman,J.E., *et al.* (2021) Characterizing the dynamic and functional DNA methylation landscape in the developing human cortex. *Epigenetics*, **16**, 1–13.

21. Spiers,H., Hannon,E., Schalkwyk,L.C., Smith,R., Wong,C.C.Y., O'Donovan,M.C., Bray,N.J. and Mill,J. (2015) Methylomic trajectories across human fetal brain development. *Genome Res.*, **25**, 338–352.

22. Moczulska,K.E., Pichler,P., Schutzbier,M., Schleiffer,A., Rumpel,S. and Mechtler,K. (2014) Deep and precise quantification of the mouse synaptosomal proteome reveals substantial remodeling during postnatal maturation. *J. Proteome Res.*, **13**, 4310–4324.

23. Gonzalez-Lozano,M.A., Klemmer,P., Gebuis,T., Hassan,C., van Nierop,P., van Kesteren,R.E., Smit,A.B. and Li,K.W. (2016) Dynamics of the mouse brain cortical synaptic proteome during postnatal brain development. *Sci. Rep.*, **6**, 35456.

24. Li,J., Zhang,W., Yang,H., Howrigan,D.P., Wilkinson,B., Souaiaia,T., Evgrafov,O.V., Genovese,G., Clementel,V.A., Tudor,J.C., *et al.* (2017) Spatiotemporal profile of postsynaptic interactomes integrates components of complex brain disorders. *Nat. Neurosci.*, **20**, 1150–1161.

25. Ohtaka-Maruyama,C., Okamoto,M., Endo,K., Oshima,M., Kaneko,N., Yura,K., Okado,H., Miyata,T. and Maeda,N. (2018) Synaptic transmission from subplate neurons controls radial migration of neocortical neurons. *Science*, **360**, 313–317.

26. Kanold,P.O., Deng,R. and Meng,X. (2019) The integrative function of silent synapses on

524 subplate neurons in cortical development and dysfunction. *Front. Neuroanat.*, **13**, 41.

525 27. Bosworth,M., Isles,A., Wilkinson,L. and Humby,T. (2021) Behavioural consequences of
526 Setd1a haploinsufficiency in mice: evidence for heightened emotional reactivity and
527 impaired sensorimotor gating. *bioRxiv*.

528 28. Cameron,D., Blake,D.J., Bray,N.J. and Hill,M.J. (2019) Transcriptional Changes
529 following Cellular Knockdown of the Schizophrenia Risk Gene SETD1A Are
530 Enriched for Common Variant Association with the Disorder. *Mol. Neuropsychiatry*,
531 **5**, 109–114.

532 29. Graham,L.C., Eaton,S.L., Brunton,P.J., Atrih,A., Smith,C., Lamont,D.J.,
533 Gillingwater,T.H., Pennetta,G., Skehel,P. and Wishart,T.M. (2017) Proteomic
534 profiling of neuronal mitochondria reveals modulators of synaptic architecture. *Mol.*
535 *Neurodegener.*, **12**, 77.

536 30. Bayés,À., Collins,M.O., Reig-Viader,R., Gou,G., Goulding,D., Izquierdo,A.,
537 Choudhary,J.S., Emes,R.D. and Grant,S.G.N. (2017) Evolution of complexity in the
538 zebrafish synapse proteome. *Nat. Commun.*, **8**, 14613.

539 31. Schmidt,K., Zhang,Q., Tasdogan,A., Petzold,A., Dahl,A., Arneth,B.M., Slany,R.,
540 Fehling,H.J., Kranz,A., Stewart,A.F., *et al.* (2018) The H3K4 methyltransferase
541 Setd1b is essential for hematopoietic stem and progenitor cell homeostasis in mice.
542 *Elife*, **7**.

543 32. Sessa,A., Fagnocchi,L., Mastrototaro,G., Massimino,L., Zaghi,M., Indrigo,M.,
544 Cattaneo,S., Martini,D., Gabellini,C., Pucci,C., *et al.* (2019) SETD5 Regulates
545 Chromatin Methylation State and Preserves Global Transcriptional Fidelity during
546 Brain Development and Neuronal Wiring. *Neuron*, **104**, 271–289.e13.

547 33. Rajasekaran,A., Venkatasubramanian,G., Berk,M. and Debnath,M. (2015) Mitochondrial
548 dysfunction in schizophrenia: pathways, mechanisms and implications. *Neurosci.*
549 *Biobehav. Rev.*, **48**, 10–21.

550 34. Hjelm,B.E., Rollins,B., Mamdani,F., Lauterborn,J.C., Kirov,G., Lynch,G., Gall,C.M.,
551 Sequeira,A. and Vawter,M.P. (2015) Evidence of Mitochondrial Dysfunction within
552 the Complex Genetic Etiology of Schizophrenia. *Mol. Neuropsychiatry*, **1**, 201–219.

553 35. Gonçalves,V.F., Andreazza,A.C. and Kennedy,J.L. (2015) Mitochondrial dysfunction in
554 schizophrenia: an evolutionary perspective. *Hum. Genet.*, **134**, 13–21.

555 36. Ben-Shachar,D. (2017) Mitochondrial multifaceted dysfunction in schizophrenia;
556 complex I as a possible pathological target. *Schizophr. Res.*, **187**, 3–10.

557 37. Srivastava,R., Faust,T., Ramos,A., Ishizuka,K. and Sawa,A. (2018) Dynamic changes of
558 the mitochondria in psychiatric illnesses: new mechanistic insights from human
559 neuronal models. *Biol. Psychiatry*, **83**, 751–760.

560 38. Steullet,P., Cabungcal,J.H., Coyle,J., Didriksen,M., Gill,K., Grace,A.A., Hensch,T.K.,
561 LaMantia,A.S., Lindemann,L., Maynard,T.M., *et al.* (2017) Oxidative stress-driven
562 parvalbumin interneuron impairment as a common mechanism in models of
563 schizophrenia. *Mol. Psychiatry*, **22**, 936–943.

564 39. Stedehouder,J. and Kushner,S.A. (2017) Myelination of parvalbumin interneurons: a
565 parsimonious locus of pathophysiological convergence in schizophrenia. *Mol.*
566 *Psychiatry*, **22**, 4–12.

567 40. Gulyás,A.I., Buzsáki,G., Freund,T.F. and Hirase,H. (2006) Populations of hippocampal
568 inhibitory neurons express different levels of cytochrome c. *Eur. J. Neurosci.*, **23**,
569 2581–2594.

570 41. Rosenfeld,M., Brenner-Lavie,H., Ari,S.G.-B., Kavushansky,A. and Ben-Shachar,D.
571 (2011) Perturbation in mitochondrial network dynamics and in complex I dependent
572 cellular respiration in schizophrenia. *Biol. Psychiatry*, **69**, 980–988.

573 42. Ben-Shachar,D. and Karry,R. (2008) Neuroanatomical pattern of mitochondrial complex
574 I pathology varies between schizophrenia, bipolar disorder and major depression.
575 *PLoS One*, **3**, e3676.

576 43. Prabakaran,S., Swatton,J.E., Ryan,M.M., Huffaker,S.J., Huang,J.T.J., Griffin,J.L.,
577 Wayland,M., Freeman,T., Dudbridge,F., Lilley,K.S., *et al.* (2004) Mitochondrial
578 dysfunction in schizophrenia: evidence for compromised brain metabolism and
579 oxidative stress. *Mol. Psychiatry*, **9**, 684–97, 643.

580 44. Karry,R., Klein,E. and Ben Shachar,D. (2004) Mitochondrial complex I subunits

expression is altered in schizophrenia: a postmortem study. *Biol. Psychiatry*, **55**, 676–684.

45. Maurer,I., Zierz,S. and Möller,H. (2001) Evidence for a mitochondrial oxidative phosphorylation defect in brains from patients with schizophrenia. *Schizophr. Res.*, **48**, 125–136.

46. Rollins,B.L., Morgan,L., Hjelm,B.E., Sequeira,A., Schatzberg,A.F., Barchas,J.D., Lee,F.S., Myers,R.M., Watson,S.J., Akil,H., *et al.* (2018) Mitochondrial complex I deficiency in schizophrenia and bipolar disorder and medication influence. *Mol. Neuropsychiatry*, **3**, 157–169.

47. Gandal,M.J., Haney,J.R., Parikshak,N.N., Leppa,V., Ramaswami,G., Hartl,C., Schork,A.J., Appadurai,V., Buil,A., Werge,T.M., *et al.* (2018) Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*, **359**, 693–697.

48. Kathuria,A., Lopez-Lengowski,K., Jagtap,S.S., McPhie,D., Perlis,R.H., Cohen,B.M. and Karmacharya,R. (2020) Transcriptomic Landscape and Functional Characterization of Induced Pluripotent Stem Cell-Derived Cerebral Organoids in Schizophrenia. *JAMA Psychiatry*, **77**, 745–754.

49. Li,J., Ryan,S.K., Deboer,E., Cook,K., Fitzgerald,S., Lachman,H.M., Wallace,D.C., Goldberg,E.M. and Anderson,S.A. (2019) Mitochondrial deficits in human iPSC-derived neurons from patients with 22q11.2 deletion syndrome and schizophrenia. *Transl. Psychiatry*, **9**, 302.

50. Szatkiewicz,J.P., O'Dushlaine,C., Chen,G., Chambert,K., Moran,J.L., Neale,B.M., Fromer,M., Ruderfer,D., Akterin,S., Bergen,S.E., *et al.* (2014) Copy number variation in schizophrenia in Sweden. *Mol. Psychiatry*, **19**, 762–773.

51. Tau,G.Z. and Peterson,B.S. (2010) Normal development of brain circuits. *Neuropsychopharmacology*, **35**, 147–168.

52. Oyarzabal,A. and Marin-Valencia,I. (2019) Synaptic energy metabolism and neuronal excitability, in sickness and health. *J. Inherit. Metab. Dis.*, **42**, 220–236.

53. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

54. Andrews,S. (2010) FastQC: a quality control tool for high throughput sequence data. Babraham Institute, Cambridge, United Kingdom.

55. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

56. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.

57. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

58. Robinson,M.D. and Oshlack,A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*, **11**, R25.

59. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.

60. Jaffe,A.E., Straub,R.E., Shin,J.H., Tao,R., Gao,Y., Collado-Torres,L., Kam-Thong,T., Xi,H.S., Quan,J., Chen,Q., *et al.* (2018) Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat. Neurosci.*, **21**, 1117–1125.

61. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.

62. Peart,M.J., Smyth,G.K., van Laar,R.K., Bowtell,D.D., Richon,V.M., Marks,P.A., Holloway,A.J. and Johnstone,R.W. (2005) Identification and functional significance of genes regulated by structurally different histone deacetylase inhibitors. *Proc. Natl.*

Acad. Sci. USA, **102**, 3697–3702.

63. Howe,K.L., Achuthan,P., Allen,J., Allen,J., Alvarez-Jarreta,J., Amode,M.R., Armean,I.M., Azov,A.G., Bennett,R., Bhai,J., *et al.* (2021) Ensembl 2021. *Nucleic Acids Res.*, **49**, D884–D891.

64. Mostafavi,S., Ray,D., Warde-Farley,D., Grouios,C. and Morris,Q. (2008) GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol.*, **9 Suppl 1**, S4.

65. de Leeuw,C.A., Mooij,J.M., Heskes,T. and Posthuma,D. (2015) MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.*, **11**, e1004219.

66. 1000 Genomes Project Consortium, Auton,A., Brooks,L.D., Durbin,R.M., Garrison,E.P., Kang,H.M., Korbel,J.O., Marchini,J.L., McCarthy,S., McVean,G.A., *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.

67. Fromer,M., Pocklington,A.J., Kavanagh,D.H., Williams,H.J., Dwyer,S., Gormley,P., Georgieva,L., Rees,E., Palta,P., Ruderfer,D.M., *et al.* (2014) De novo mutations in schizophrenia implicate synaptic networks. *Nature*, **506**, 179–184.

68. Rees,E., Han,J., Morgan,J., Carrera,N., Escott-Price,V., Pocklington,A.J., Duffield,M., Hall,L.S., Legge,S.E., Pardiñas,A.F., *et al.* (2020) De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. *Nat. Neurosci.*, **23**, 179–184.

69. Ambalavanan,A., Girard,S.L., Ahn,K., Zhou,S., Dionne-Laporte,A., Spiegelman,D., Bourassa,C.V., Gauthier,J., Hamdan,F.F., Xiong,L., *et al.* (2016) De novo variants in sporadic cases of childhood onset schizophrenia. *Eur. J. Hum. Genet.*, **24**, 944–948.

70. Girard,S.L., Gauthier,J., Noreau,A., Xiong,L., Zhou,S., Jouan,L., Dionne-Laporte,A., Spiegelman,D., Henrion,E., Diallo,O., *et al.* (2011) Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat. Genet.*, **43**, 860–863.

71. Guipponi,M., Santoni,F.A., Setola,V., Gehrig,C., Rotharmel,M., Cuenca,M., Guillin,O., Dikeos,D., Georgantopoulos,G., Papadimitriou,G., *et al.* (2014) Exome sequencing in 53 sporadic cases of schizophrenia identifies 18 putative candidate genes. *PLoS One*,

9, e112745.

72. Gulsuner,S., Walsh,T., Watts,A.C., Lee,M.K., Thornton,A.M., Casadei,S., Rippey,C., Shahin,H., Consortium on the Genetics of Schizophrenia (COGS), PAARTNERS Study Group, *et al.* (2013) Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell*, **154**, 518–529.

73. McCarthy,S.E., Gillis,J., Kramer,M., Lihm,J., Yoon,S., Berstein,Y., Mistry,M., Pavlidis,P., Solomon,R., Ghiban,E., *et al.* (2014) De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol. Psychiatry*, **19**, 652–658.

74. Takata,A., Xu,B., Ionita-Laza,I., Roos,J.L., Gogos,J.A. and Karayiorgou,M. (2014) Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. *Neuron*, **82**, 773–780.

75. Wang,Q., Li,M., Yang,Z., Hu,X., Wu,H.-M., Ni,P., Ren,H., Deng,W., Li,M., Ma,X., *et al.* (2015) Increased co-expression of genes harboring the damaging de novo mutations in Chinese schizophrenic patients during prenatal development. *Sci. Rep.*, **5**, 18209.

76. Xu,B., Ionita-Laza,I., Roos,J.L., Boone,B., Woodrick,S., Sun,Y., Levy,S., Gogos,J.A. and Karayiorgou,M. (2012) De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.*, **44**, 1365–1369.

77. Rees,E., Carrera,N., Morgan,J., Hambridge,K., Escott-Price,V., Pocklington,A.J., Richards,A.L., Pardiñas,A.F., GROUP Investigators, McDonald,C., *et al.* (2019) Targeted Sequencing of 10,198 Samples Confirms Abnormalities in Neuronal Activity and Implicates Voltage-Gated Sodium Channels in Schizophrenia Pathogenesis. *Biol. Psychiatry*, **85**, 554–562.

78. Clifton,N.E., Rees,E., Holmans,P.A., Pardiñas,A.F., Harwood,J.C., Di Florio,A., Kirov,G., Walters,J.T.R., O'Donovan,M.C., Owen,M.J., *et al.* (2020) Genetic association of FMRP targets with psychiatric disorders. *Mol. Psychiatry*, 10.1038/s41380-020-00912-2.

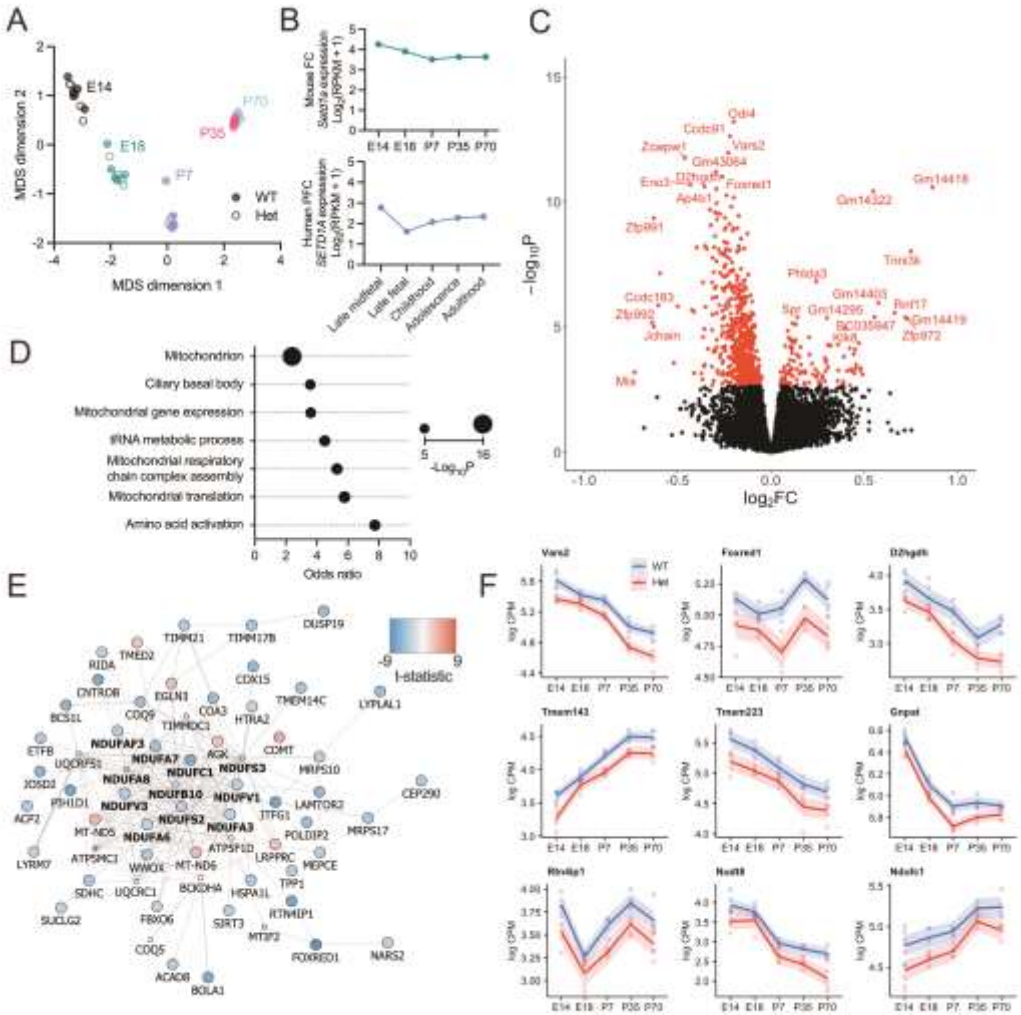
79. Samocha,K.E., Robinson,E.B., Sanders,S.J., Stevens,C., Sabo,A., McGrath,L.M., Kosmicki,J.A., Rehnström,K., Mallick,S., Kirby,A., *et al.* (2014) A framework for the

694 interpretation of de novo mutation in human disease. *Nat. Genet.*, **46**, 944–950.

695 80. Perez-Riverol,Y., Bai,J., Bandla,C., García-Seisdedos,D., Hewapathirana,S.,
696 Kamatchinathan,S., Kundu,D.J., Prakash,A., Frericks-Zipper,A., Eisenacher,M., *et al.*
697 (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based
698 proteomics evidences. *Nucleic Acids Res.*, **50**, D543–D552.

699

700



701

702 **Figure 1** Transcriptomic effects of *Setd1a* loss-of-function determined by RNA sequencing of
703 mouse frontal cortical tissue across prenatal and postnatal development. **A)** Multidimensional
704 scaling (MDS) plot representing the similarity of sequencing libraries and clustering by
705 timepoint. **B)** Mean normalised frontal cortical expression of wildtype mouse *Setd1a* (top) or

human *SETD1A* (bottom) across matched developmental timepoints. **C)** Differential gene expression analysis contrasting *Setd1a*$^{+/-}$ with wildtype samples, covarying for effects of age. Significantly differentially expressed genes (DEGs) are shown in red. **D)** Enrichment of DEGs in genotype contrasts for functionally-defined gene sets from the Gene Ontology (GO) database. The size of the dot relates to the significance in Fisher's exact test. **E)** Protein-protein interactions among human orthologs of DEGs from genotype contrasts annotated by GO:0005739 *Mitochondrion*. Interaction data was obtained from GeneMANIA. Only proteins with interactions in the core network are displayed. Proteins forming part of the mitochondrial NADH:ubiquinone oxidoreductase respiratory complex I are shown in bold. Node colour relates to the t-statistic in differential gene expression analysis. Smaller nodes indicate proteins inserted by GeneMANIA to improve the network but were non-significant in differential gene expression analysis. **F)** Developmental expression of the top 9 DEGs in genotype contrasts, for wildtype and *Setd1a*$^{+/-}$ samples. Shown is log counts per million (logCPM) ± standard error from embryonic day 14 (E14) to postnatal day 70 (P70). Wildtype (WT); Heterozygous (Het); Frontal cortex (FC); Prefrontal cortex (PFC); Reads per kilobase of transcript per million mapped reads (RPKM); Fold change (FC); Counts per million (CPM).

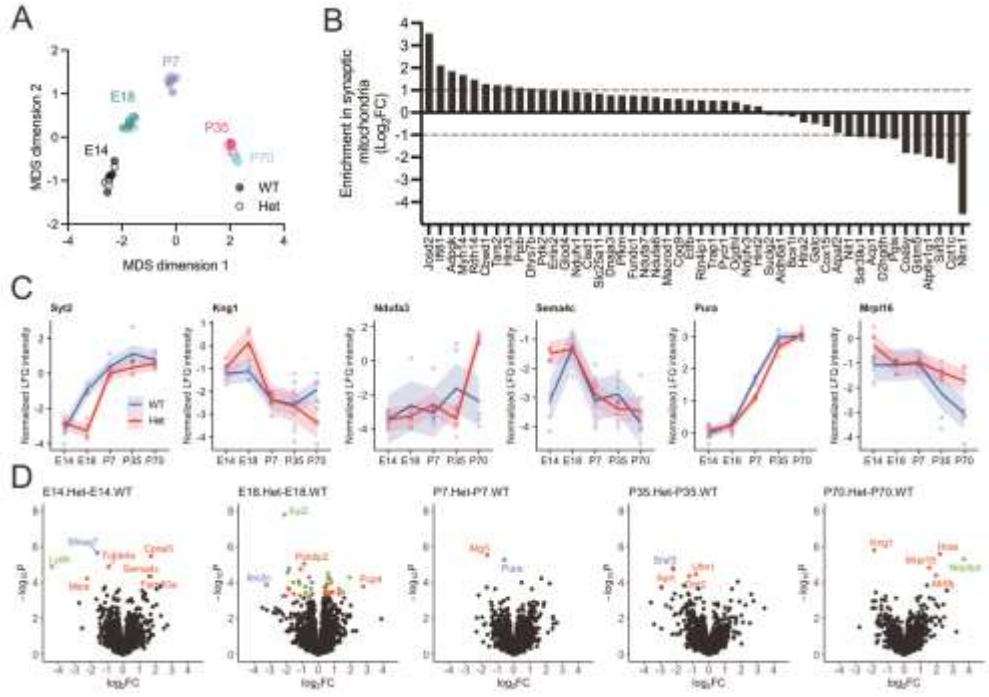


Figure 2 Alterations to the synaptosome caused by *Setd1a* loss-of-function determined by mass spectrometry-based label-free quantitation of isolated mouse synaptosomes across prenatal and postnatal development. **A)** Multidimensional scaling (MDS) plot indicating the clustering of

synaptosome samples based on normalized LFQ intensity of each protein. **B)** Relative abundance in synaptic vs non-synaptic mitochondria of proteins encoded by differentially expressed genes in genotype contrast. Mitochondrial proteomics data obtained from a previous study (29). **C)** Wildtype and *Setd1a*$^{+/-}$ synaptosomal protein abundance across development of differentially expressed proteins in primary genotype contrasts. Displayed is the normalized LFQ intensity ± standard error from embryonic day 14 (E14) to postnatal day 70 (P70). **D)** Differential protein expression analyses contrasting *Setd1a*$^{+/-}$ with wildtype synaptosomes at each timepoint independently. Colours indicate significantly differentially expressed proteins enriched in the postsynaptic density (PSD; blue), depleted in the PSD (green), or of similar abundance in the PSD compared to the total synaptosome / no data (red), based on published data (30).

Protein	Function	Synaptic compartment
Synaptotagmin-2 (Syt2)	Mediates calcium-dependent synaptic vesicle exocytosis and neurotransmitter release.	PSD-depleted
Kininogen-1 (Kng1)	Precursor to the proinflammatory peptides of the kallikrein-kinin system.	No data
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 3 (Ndufa3)	Subunit of the mitochondrial respiratory chain complex I.	PSD-depleted
Semaphorin-4C (Sema4c)	Receptor for Plexin-B2, important for regulation of axon guidance, dendritic morphology and synapse formation.	No data
Transcriptional activator protein Pur-alpha (Pura)	DNA and RNA binding protein involved in transcriptional control and cytoplasmic RNA localization.	PSD-enriched
Mitochondrial ribosomal protein L16 (Mrpl16)	Nuclear-encoded subunit of mitochondrial ribosomes, required for protein synthesis within mitochondria.	No data

Table 1 Differentially expressed proteins in cortical synaptosomes of *Setd1a*$^{+/-}$ mice compared to wildtype, controlling for age. Synaptic compartment localisation was determined from a previous study (30).

Differential expression contrast	N genes / proteins	Schizophrenia common variant association	Schizophrenia *de novo* rare nonsynonymous variant association
RNA: Genotype effect (All timepoints)			
All DEGs	734	β = -0.071 *P* = 0.95	Rate ratio = 1.03 *P* = 0.39
Downregulated	616	β = -0.091 *P* = 0.97	Rate ratio = 0.96 *P* = 0.80
Upregulated	118	β = 0.052 *P* = 0.32	Rate ratio = 1.31 *P* = 0.11
Synaptosomal	127	β = -0.063 *P* = 0.73	Rate ratio = 0.52 *P* = 0.97
Non-synaptosomal	607	β = -0.028 *P* = 0.72	Rate ratio = 1.15 *P* = 0.12
DEGs belonging to GO:0005739 *Mitochondrion*	101	β = -0.15 *P* = 0.91	Rate ratio = 1.08 *P* = 0.75
Protein: Genotype effects (Any timepoint)			
All unique proteins	67	β = 0.049 *P* = 0.37	Rate ratio = 1.49 *P* = 0.11
Downregulated	35	β = 0.098 *P* = 0.30	Rate ratio = 0.88 *P* = 0.67
Upregulated	33	β = -0.025 *P* = 0.55	Rate ratio = 2.45 *P* = 0.019 (*P*.adj = 0.21)
PSD-enriched	8	β = -0.39 *P* = 0.88	Rate ratio = 2.71 *P* = 0.17
PSD-depleted	22	β = 0.23 *P* = 0.15	Rate ratio = 1.18 *P* = 0.44

Table 2 Genetic association with schizophrenia of transcripts and proteins disrupted by *Setd1a* loss-of-function. Differentially expressed genes or proteins observed from the specified contrasts were tested for enrichment for genetic association with schizophrenia through common or rare variation. *P*-values follow conditioning on all expressed genes (RNA analyses) or all synaptosomal genes (protein analyses) and are uncorrected for multiple testing, unless specified. PSD enrichment or depletion was predicted based on published data (30). Differentially expressed gene (DEG); Postsynaptic density (PSD); Bonferroni-adjusted *P*-value (*P.adj*).

Abbreviations

GWAS Genome-wide association study

SNP Single nucleotide polymorphism

DEG Differentially expressed gene

PSD Postsynaptic density

LoF Loss-of-function

SETD1A SET Domain Containing 1A

FDR False discovery rate

GO Gene ontology

ChIP Chromatin immunoprecipitation

RNA Ribonucleic acid

E14 Embryonic day 14

P7 Postnatal day 7

PGC Psychiatric genomics consortium

SynPER Synaptic protein extraction reagent

SDS Sodium dodecyl sulfate

TEAB Tetraethylammonium bromide

TCEP Tris(2-carboxyethyl)phosphine

MS Mass spectrometry

FTMS Fourier transform mass spectrometry

LFQ Label-free quantitation

IEA Inferred from electronic annotation

NAS Non-traceable author statement

RCA Inferred from reviewed computational analysis