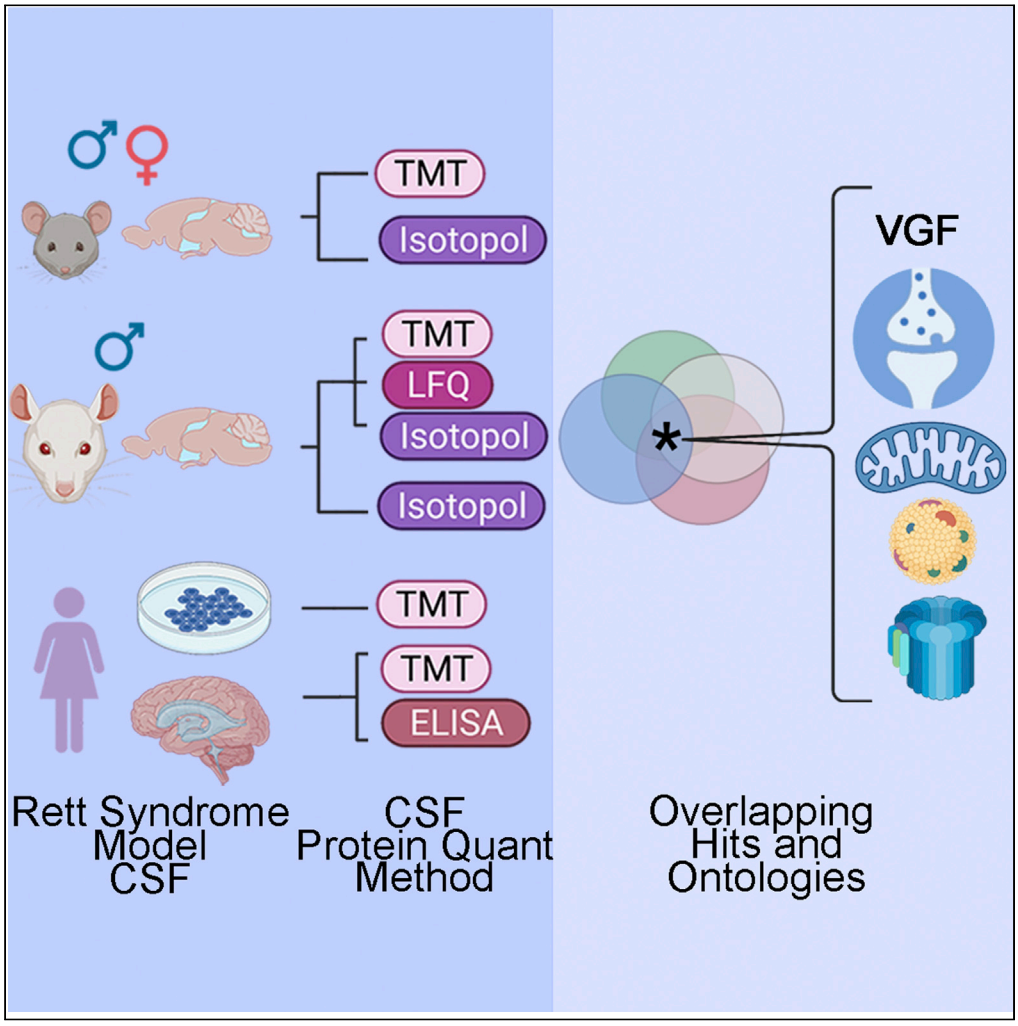
iScience

CellPress
OPEN ACCESS

Article

Convergent cerebrospinal fluid proteomes and metabolic ontologies in humans and animal models of Rett syndrome



Stephanie A. Zlatic, Duc Duong, Kamal K.E. Gadalla, ..., Randall Carpenter, Stuart Cobb, Victor Faundez

szlatic@gmail.com (S.A.Z.)
vfaunde@emory.edu (V.F.)

Highlights

Shared Rett CSF proteome ontologies across species and experimental approaches

Rett CSF proteome enriches HDL lipoprotein, mitochondria, and synapse ontologies

VGF levels discriminate *Mecp2*$^{-/y}$, *Mecp2*$^{-/+}$, from control CSF

Rett CSF proteome and ontologies are distinct from those in *Cdkl5*$^{-/y}$ CSF

Zlatic et al., iScience *25*, 104966
September 16, 2022
https://doi.org/10.1016/j.isci.2022.104966

iScience

CellPress
OPEN ACCESS

Article

Convergent cerebrospinal fluid proteomes and metabolic ontologies in humans and animal models of Rett syndrome

Stephanie A. Zlatic,[1,10,*] Duc Duong,[2,10] Kamal K.E. Gadalla,[3] Brenda Murage,[3] Lingyan Ping,[2] Ruth Shah,[6] James J. Fink,[9] Omar Khwaja,[4] Lindsay C. Swanson,[4] Mustafa Sahin,[4] Sruti Rayaprolu,[5] Prateek Kumar,[5] Srikant Rangaraju,[5] Adrian Bird,[6] Daniel Tarquinio,[7] Randall Carpenter,[8] Stuart Cobb,[3] and Victor Faundez[1,11,*]

SUMMARY

MECP2 loss-of-function mutations cause Rett syndrome, a neurodevelopmental disorder resulting from a disrupted brain transcriptome. How these transcriptional defects are decoded into a disease proteome remains unknown. We studied the proteome of Rett cerebrospinal fluid (CSF) to identify consensus Rett proteome and ontologies shared across three species. Rett CSF proteomes enriched proteins annotated to HDL lipoproteins, complement, mitochondria, citrate/pyruvate metabolism, synapse compartments, and the neurosecretory protein VGF. We used shared Rett ontologies to select analytes for orthogonal quantification and functional validation. VGF and ontologically selected CSF proteins had genotypic discriminatory capacity as determined by receiver operating characteristic analysis in *Mecp2*$^{-/y}$ and *Mecp2*$^{-/+}$. Differentially expressed CSF proteins distinguished Rett from a related neurodevelopmental disorder, CDKL5 deficiency disorder. We propose that *Mecp2* mutant CSF proteomes and ontologies inform putative mechanisms and biomarkers of disease. We suggest that Rett syndrome results from synapse and metabolism dysfunction.

INTRODUCTION

The cellular and molecular understanding of neurodevelopmental disorders has been greatly advanced by the study of single gene defects (Sztainberg and Zoghbi, 2016; Lee et al., 2020). Among these monogenic neurodevelopmental disorders, Rett syndrome, caused by mutations in *MECP2*, stands out because of its severity and developmental regression. The molecular function of MECP2 as an epigenetic chromatin regulator is well defined (Lyst and Bird, 2015), affecting the expression of a vast number of RNAs in the brain (Lyst and Bird, 2015; Banerjee et al., 2019; Sandweiss et al., 2020; Johnson et al., 2017). The molecular complexity of Rett syndrome is compounded by the extensive and varied modifications of coding and non-coding transcriptomes across brain cell types and regions (Chahrour et al., 2008; Raman et al., 2018; Cholewa-Waclaw et al., 2019; Wu et al., 2010; Cheng et al., 2014). This fact makes transcriptional prediction of Rett syndrome proteomes a complex and uncertain endeavor. Thus, we focus on the proteome to identify biochemical and cellular alterations in Rett syndrome brains (Mullin et al., 2013). Moreover, known Rett syndrome phenotypes such as synaptic, circuit, and behavioral alterations ultimately originate in alterations of protein expression and function (Chao et al., 2007, 2010; Banerjee et al., 2019). Thus, the proteome decodes *Mecp2*-dependent transcriptional changes and executes a diseased phenome. Despite these advantages of the proteome to illuminate pathogenic mechanisms and to identify disease biomarkers, few studies examine how the proteome is modified in Rett syndrome brains (Matarazzo and Ronnett, 2004; Pacheco et al., 2017).

Our goal was to identify a Rett proteome capable of distinguishing normal and disease brain states across neurodevelopment that is clinically accessible for sample collection, diagnostic testing, and assessing treatment outcomes. We focused on CSF as a clinically accessible sample whose composition is dictated by the brain and neurodevelopment (Chau et al., 2015; Kaiser et al., 2019; Lehtinen and Walsh, 2011). CSF carries neurodevelopmental instructive signals and nutrients and accrues secretions and metabolites that reflect functional states of diverse cell types in brain parenchyma, the choroid, and the ependymal cells

[1]Departments of Cell Biology, Emory University, Atlanta, GA 30322, USA

[2]Departments of Biochemistry, Emory University, Atlanta, GA 30322, USA

[3]Simons Initiative for the Developing Brain, Centre for Discovery Brain Sciences, University of Edinburgh, Edinburgh EH8 9XD, UK

[4]Department of Neurology, Rosamund Stone Zander Translational Neuroscience Center, Boston Children's Hospital, Boston, MA 02115, USA

[5]Departments of Neurology, Emory University, Atlanta, GA 30322, USA

[6]The Wellcome Centre for Cell Biology, University of Edinburgh, Michael Swann Building, King's Buildings, Max Born Crescent, Edinburgh EH9 3BF, UK

[7]Center for Rare Neurological Diseases, Norcross, GA 30093, USA

[8]Rett Syndrome Research Trust, Trumbull, CT 06611, USA

[9]Q-State Biosciences, Cambridge, MA 02139, USA

[10]These authors contributed equally

[11]Lead contact

*Correspondence: szlatic@gmail.com (S.A.Z.), vfaunde@emory.edu (V.F.)

https://doi.org/10.1016/j.isci.2022.104966

(Johanson et al., 2008). For example, composition of the CSF proteome in individuals with Alzheimer's predicts key diagnostic molecular pathology in the Alzheimer's brain (Olsson et al., 2016; Higginbotham et al., 2020; Johnson et al., 2020). Thus, the CSF proteome has the potential to inform us about brain-wide normal and pathological states.

The study of the CSF in genetic or sporadic forms of neurodevelopmental disorders lags behind similar studies in neurological diseases. Less than a handful of studies analyze the proteome of this biofluid (Abbasi et al., 2021). The study of the CSF in neurodevelopmental disorders has mostly been focused to targeted studies of few analytes such as cytokines, growth factors, neuropeptides, or metabolites (Zimmerman et al., 2005; Riikonen et al., 2006; Oztan et al., 2020, 2021; Budden et al., 1990; Matsuishi et al., 1994). Thus, we do not know if a neurodevelopmental disorder, such as Rett syndrome, reproducibly and distinctively modifies the CSF proteome to predict disease mechanisms and biomarkers of disease. Here, we address this fundamental question by comprehensively and unbiasedly exploring the CSF proteome of three species carrying mutations in *MECP2*/*Mecp2*. We defined a consensus proteome and ontologies predictive of Rett syndrome disease mechanisms. Analytes found in the *Mecp2*-sensitive proteome behaved with sufficient sensitivity and specificity to act as Rett syndrome biomarkers and to discriminate Rett syndrome CSF from a CDKL5 deficiency disorder CSF, a phenotypically related syndrome (Evans et al., 2005; Weaving et al., 2004; Zerbi et al., 2021). This is the first multispecies study of the CSF proteome in a monogenic disorder of neurodevelopment. Based on our proteomic analysis, we propose that Rett syndrome is a synaptic and metabolic neurodevelopmental disorder. Further, our experimental strategy offers a platform for the identification of proteomes and biomarkers in the CSF of any childhood genetic neurological disorder and to infer putative mechanisms of disease.

RESULTS

The CSF proteome is composed of proteins secreted by conventional and non-conventional secretory pathways, exosomes, and ectosomes from brain (Kalluri and LeBleu, 2020). We collectively refer to these proteins as the brain secreted proteome. We sought to identify secreted proteins sensitive to *MECP2*/*Mecp2* gene defects. To achieve this goal, we designed a multipronged strategy to quantify secreted proteomes from wild type and *MECP2*-null neuron conditioned media, cerebrospinal fluid (CSF) from *Mecp2* null male rat and mouse models, and the CSF from female individuals with Rett syndrome collected before and after recombinant IGF1 treatment (Figure 1A) (O'Leary et al., 2018; Tropea et al., 2009). We reasoned that overlapping proteins and ontologies across diverse experimental systems would identify robust proteins and ontologies to inform putative disease mechanisms and Rett syndrome biomarkers. Furthermore, we designed our studies with an emphasis in replicability across sample collection and measurement experimental sites, quantification platforms, and species in order to inform biomarker selection (Figure 1A). We chose to quantify proteomes with tandem mass tagging (TMT) mass spectrometry as a high precision method (O'Connell et al., 2018; Li et al., 2012). TMT datasets were analyzed by fold of change/p value volcano thresholding plus machine learning approaches (Figure 1A).

The secreted proteome of a MECP2 deficient neuronal cell line

We began using cultured human neurons that were differentiated from the immortalized mesencephalic neuronal cell line LUHMES (Scholz et al., 2011) in which the *MECP2* gene was edited by CRISPR/Cas9 (Shah et al., 2016). We reasoned the secreted proteome of a single cell type would define cell autonomous protein candidates for cell-type annotation of hits obtained in CSFs from individuals with Rett syndrome and rodent *Mecp2*-mutant models. We characterized the culture media before and after cell conditioning (Figure 1B, compare lanes 1 with 2–3). The protein complexity of media alone prevented the identification of proteins contributed by differentiated neurons. The source of protein contaminants was commercial N2 supplements; thus, we customized an N2 supplement starting from high purity reagents. Our customized N2 allowed us to distinguish proteins contributed by either wild type or mutant neurons (Figure 1C, compare lanes 1 with 2–3). TMT mass spectrometry of media alone identified 704 proteins (Figures 1D and 1G). In contrast, cell-conditioned media revealed 958 additional proteins contributed by wild type cells (Figures 1D and 1G). Next, we used this custom media formulation to compare the proteome of wild type and *MECP2* null cells. We identified 63 upregulated and 155 downregulated proteins in *MECP2* mutant cells by p value and fold-of-change volcano thresholding (Figures 1E and 1H). Prominent downregulated proteins in *MECP2* mutant cells included two apolipoproteins (APOC2 and Clusterin, CLU or APOJ), the nucleoporin component AHCTF1, the mitochondrial protein COQ9, as well as factors implicated in citrate cycle and glycolysis such as PDHB and OGDH (Rasala et al., 2006; Rath et al., 2021). To these volcano

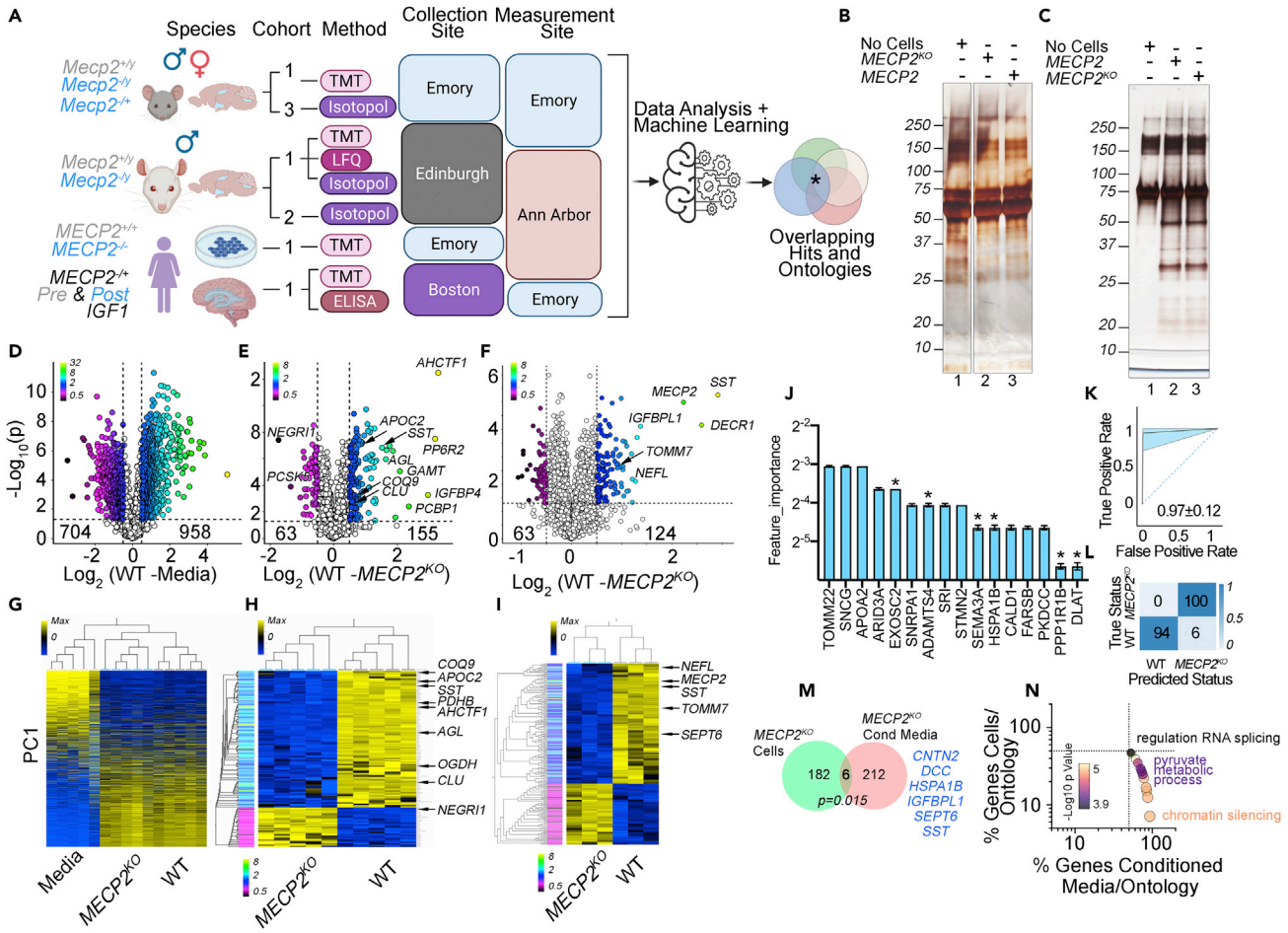


Figure 1. The secreted proteome of post-mitotic *MECP2* mutant human neurons

(A) Diagram shows three species and experimental systems. Cohorts represent independent collections of samples with the strategy used for analyte quantification, the place of sample collection, and sample measurement location. Isotopol refers to AQUA and modified AQUA strategies, LFQ corresponds to label free quantification, and TMT denotes tandem mass tagging.

(B and C) Silver stain of two formulations of media conditioned by wild type and *MECP2*-null differentiated LUHMES cells, a post-mitotic human neuron line (Scholz et al., 2011). Lane 1 in B and C represent naive non-conditioned media. Lanes 2 and 3 depict conditioned media by wild type and mutant cells. B presents experiments performed with commercial N2 supplement. C shows experiments where the N2 supplement was custom generated from high grade purity reagents.

(D–F) Present volcano plots of TMT mass spectrometry experiments with thresholds at log2 of 0.5-fold of change in protein abundance and a p value <0.05. Symbol color represents fold of change in linear scale (see insert). (D) presents a comparison between protein hits obtained by comparing media conditioned by wild type neurons and non-conditioned media. All hits to the right of the X axis correspond to proteins secreted by neurons (n = 5). (E) shows a comparison of the wild type and *MECP2* mutant secreted proteome. All hits to the right correspond to proteins whose expression is higher in wild type than in mutant cells (n = 5). (F) shows the total cellular proteome of wild type and *MECP2*-null cells used in (E), n = 3.

(G–I) Show clustered heat maps of hits selected in D to F. Arrows mark some cardinal differentially expressed proteins. Rows are depicted as minimum and maximum intensities (blue-yellow scale) and annotated by log2 fold of change (rainbow scale, see Table S1).

(J–L) Analysis of TMT data in panel E using an XGBoost machine learning algorithm. J presents main hits discriminating wild type and *MECP2* mutant conditioned media in the decision tree. Asterisks mark proteins identified both by volcano thresholding and machine learning. (K and L) performance of the machine learning protocol estimated by ROC analysis, (J) and confusion matrix in (I). Confusion matrix refers to the percentage of samples assigned to a genotype. Area under the curve in J =0.97 ± 0.12.

(M) Venn diagram of the overlap between hits found in conditioned media in panel E and cellular hits in F, p value calculated with exact hypergeometric probability.

(N) Shows the % overlap between the cellular proteome ontologies inferred from the datasets shown in (E and F) calculated with the ClueGo application. p value estimated with exact hypergeometric probability Bonferroni corrected.

See extended data in Table S1.

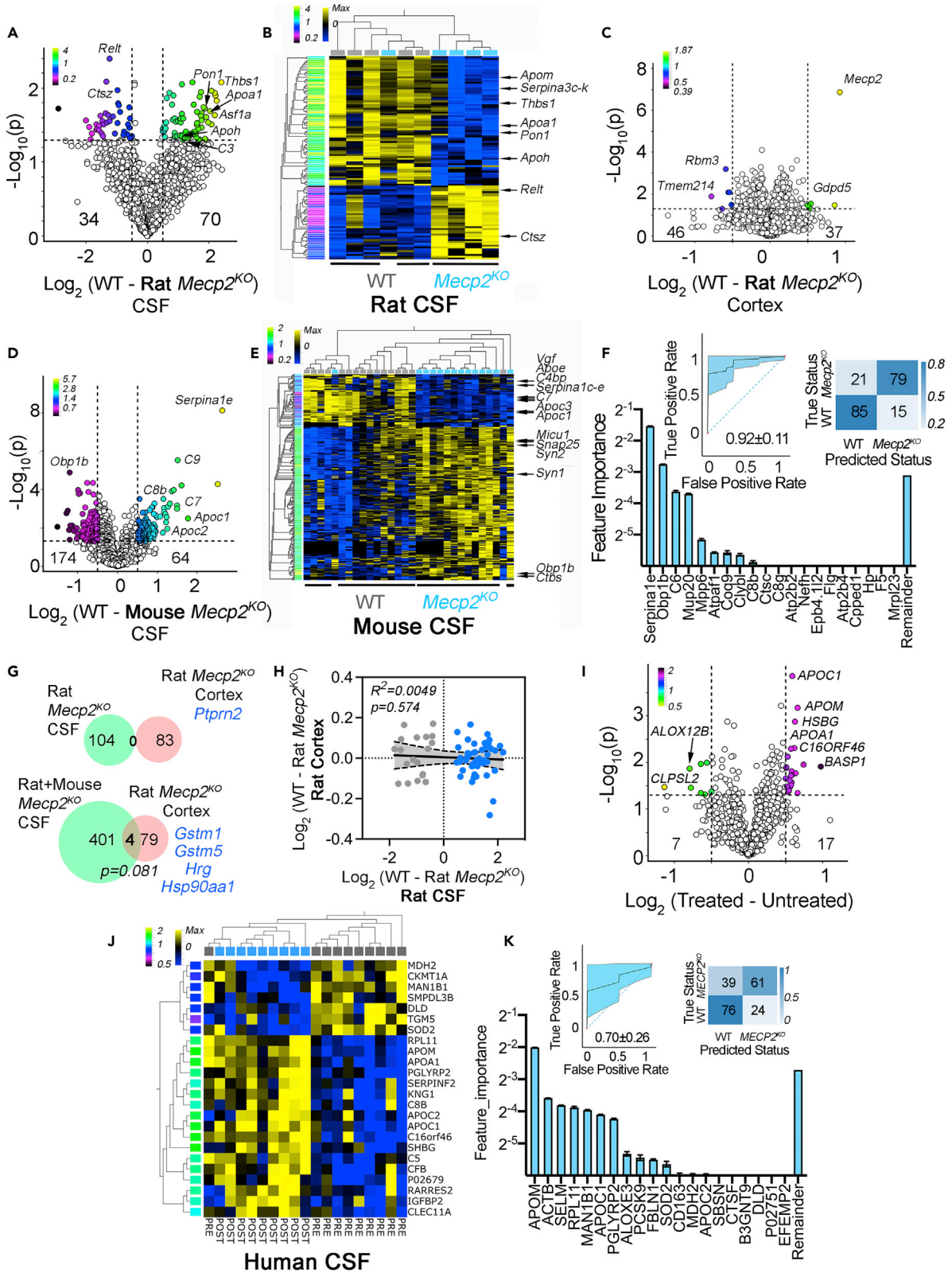

A
Relt
Pon1
Thbs1
Apoa1
Asf1a
Apoh
C3
Ctsz
34
70
-Log10(p)
Log2 (WT - Rat Mecp2KO)
CSF
B
Max
Apom
Serpina3c-k
Thbs1
Apoa1
Pon1
Apoh
Relt
Ctsz
WT
Mecp2KO
Rat CSF
C
Mecp2
Rbm3
Tmem214
Gdpd5
46
37
Log2 (WT - Rat Mecp2KO)
Cortex
D
Serpina1e
Obp1b
C9
C8b
C7
Apoc1
Apoc2
174
64
Log2 (WT - Mouse Mecp2KO)
CSF
E
Vgf
Apoe
C4bp
Serpina1c-e
C7
Apoc3
Apoc1
Micu1
Snap25
Syn2
Syn1
Obp1b
Ctbs
WT
Mecp2KO
Mouse CSF
F
True Positive Rate
False Positive Rate
0.92±0.11
True Status
WT Mecp2KO
21
79
85
15
WT Mecp2KO
Predicted Status
Feature Importance
Serpina1e
Obp1b
C6
Mup20
Mpp6
Atpaf1
Coq9
Clyol
C8b
Ctsc
C8g
Atp2b2
Nefl
Epb41l2
Flg
Atp2b4
Cpped1
Hp
F5
Mrpl23
Remainder
G
Rat Mecp2KO CSF
Rat Mecp2KO Cortex
Ptprn2
104
0
83
Rat+Mouse Mecp2KO CSF
Rat Mecp2KO Cortex
401
4
79
Gstm1
Gstm5
Hrg
Hsp90aa1
p=0.081
H
R2=0.0049
p=0.574
Log2 (WT - Rat Mecp2KO)
Rat Cortex
Log2 (WT - Rat Mecp2KO)
Rat CSF
I
APOC1
APOM
HSBG
APOA1
C16ORF46
BASP1
ALOX12B
CLPSL2
7
17
-Log10(p)
Log2 (Treated - Untreated)
J
Max
MDH2
CKMT1A
MAN1B1
SMPDL3B
DLD
TGM5
SOD2
RPL11
APOM
APOA1
PGLYRP2
SERPINF2
KNG1
C8B
APOC2
APOC1
C16orf46
SHBG
C5
CFB
P02679
RARRES2
IGFBP2
CLEC11A
PRE
POST
Human CSF
K
True Positive Rate
False Positive Rate
0.70±0.26
True Status
WT MECP2KO
39
61
76
24
WT MECP2KO
Predicted Status
Feature_importance
APOM
ACTB
SELM
RPL11
MAN1B1
APOC1
PGLYRP2
ALOX5E3
PCSK9
FBLN1
SOD2
CD163
MDH2
APOC2
SBSN
CTSF
B3GNT9
DLD
P02751
EFEMP2
Remainder

Figure 2. The Rett syndrome CSF proteome across three species

(A) Volcano plot of TMT mass spectrometry determinations in rat cerebrospinal fluid. Cutoffs at log2 of 0.5-fold of change in protein abundance and a p value <0.05 n = 5.

(B) Shows clustered heat maps of hits selected in A. For A and B, see legend to Figure 1 for additional details.

(C) Depicts volcano plot of rat cortices analyzed by TMT mass spectrometry. Shown are hits selected by p value <0.05 and log2 change of 0.5 n = 5.

(D and E) Show mouse CSF TMT volcano plot and heatmap of selected hits at cutoffs of log2 0.5-fold of change in protein abundance and a p<0.05. n = 16 wild type and 14 *Mecp2*-null mice.

(F) Analysis of TMT data presented in panel D using an XGBoost machine learning algorithm. Main hits discriminating wild type and *Mecp2* mutant CSF proteomes in decision tree are shown. Inserts show performance of the machine learning protocol estimated by ROC analysis and confusion matrix. Area under the curve in J =0.92 $\pm$ 0.11. Confusion matrix refers to the percentage of samples assigned to a genotype.

(G) Top Venn diagram shows the overlap between *Mecp2*-sensitive rat CSF and rat cortex hits using thresholding criteria p<0.05 and log2 fold of change of 0.5. Bottom Venn diagram compares *Mecp2*-sensitive hits in rat and mouse CSF pooled together with *Mecp2*-sensitive rat cortex hits.

(H) Depicts the correlation in expression of *Mecp2*-sensitive hits in rat CSF and rat cortex.

(I and J) Show Rett syndrome female individual CSF TMT volcano plot and heatmap of selected hits at log2 of 0.5-fold of change in protein abundance and a p<0.05 comparing before and after IGF-1 treatment. n = 10 before treatment and 9 after treatment.

(K) Analysis of TMT data presented in panel I using an AdaBoost machine learning algorithm. Main hits discriminating CSF before and after treatment in decision tree are shown. Inserts show performance of the machine learning protocol estimated by ROC analysis and confusion matrix. Area under the curve in insert ROC analysis 0.70 $\pm$ 0.26. Confusion matrix refers to the percentage of samples assigned to a genotype.

See Tables S2 and S3.

selected hits, we added 10 additional proteins from a total of 16 proteins whose expression was sensitive to the *MECP2* mutation as defined by machine learning (Figure 1J, asterisks for common volcano and machine learning hits) (Torun et al., 2021). Among these 10 new proteins were the mitochondrial proteins TOMM22 and the E2 component of the pyruvate dehydrogenase complex (DLAT) as well as the apolipoprotein APOA2. The performance of the machine learning algorithm was evaluated by receiver operating characteristics (ROC) analysis with an area under the curve of 0.97 (Figure 1K) and confusion matrix analysis where predicted and actual genotypic classes closely matched (Figure 1L).

We asked whether the *MECP2* secreted proteome was a reflection of the *MECP2* cellular proteome. The *MECP2* cellular proteome was represented by 187 proteins (Figures 1F and 1I). The overlap between these two *MECP2* sensitive proteomes was minor and barely significant (Figure 1M). We found just six common hits, among them SST and IGFBPL1 (Figure 1M). Convergence between these two datasets became evident in few significant ontological categories shared between the *MECP2* secreted and cellular proteomes, one of them pyruvate metabolic process, a mitochondrial ontology (GO:0006090 p = 3.96×10^{-5} Bonferroni corrected, Figure 1N). These data suggest that ontologies may be better positioned than isolated proteomic hits to identify convergence between the secreted and cellular proteome in MECP2 gene mutations within a simple cellular system.

Secreted proteomes of *MECP2/Mecp2* deficient cerebrospinal fluids in three species

In order to identify CSF proteomes and/or ontologies that are robust and convergent at the intra and inter-species level, we analyzed by TMT mass spectrometry the CSF from wild type and *Mecp2* mutants in two rodent species and from Rett syndrome individuals. We performed studies in rats aged 25 days (Figures 2A–2C) and in a large cohort of wild type and *Mecp2* null mice aged 6 weeks (Figures 2D–2F). We identified 70 and 64 CSF proteins whose expression was downregulated in mutant rat and mouse CSF, respectively (Figures 2A, 2B, 2D, and 2E). These mutant CSF downregulated proteins prominently converged on subunits of high-density lipoprotein particles such as Apom, Apoa1, Apoh, and Pon1. Similarly, we identified Apoa1, Apoc1, Apoc2, and Apoe as downregulated hits in mouse *Mecp2* mutant CSF. An additional category of proteins downregulated in both species were proteins belonging to the complement and coagulation cascades (Figures 2A, 2B, 2D, and 2E). To assess intraspecies robustness, we confirmed these rat CSF hits in an independent cohort of wild type and mutant rats using an orthogonal label-free mass spectrometry quantification procedure (LFQ) (O'Connell et al., 2018). This analysis identified 44 proteins whose expression was affected in mutant CSF (Figure S1A) and confirmed the downregulation of apolipoproteins and complement factors in *Mecp2*-null CSF (Figures S1A–S1C, Apoa4, Apob, Apoc3, Pon1, and C9). Rodent apolipoprotein and complement cascade hits could not be attributed to blood contamination of the CSF as evidenced by albumin, immunoglobulins, or hemoglobin species, which failed to co-cluster with apolipoproteins and complement (Figures 2B,2E, S1A, and S1B). Synaptic proteins were prominent among factors upregulated in the mouse Mecp2 mutant CSF (Figures 2D and 2E). These synaptic proteins include, but were not limited to, Snap25, Stx1b, Stxbp1, Syn1, and Syn2 (Figures 2D and 2E) (Koopmans et al., 2019). Finally, we also identified mitochondria

proteins within both the significant up- and downregulated proteomes in mutant rat and mouse CSF (Figures 2D and 2E).

We further scrutinized mouse and rat CSF datasets using machine learning algorithms. We sought to identify additional proteins sensitive to *Mecp2* deficiency that could otherwise escape detection by volcano thresholding. In addition, we reasoned proteins categorized as apolipoproteins, synaptic, complement-related, or mitochondria-annotated should emerge as priority hits in non-linear decision trees segregating wild type and *Mecp2* mutant CSFs. Both mouse and rat CSF machine learning analyses identified complement factors, mitochondrial proteins (Atpaf1, Clybl, Coq9, and Mrpl23), as well as synaptic proteins as priority hits (Actr2, Atp2b2 and Nefh, Figure 2F and S1C–S1E). We validated the performance of machine learning approaches asking their capacity to identify MECP2 as a priority protein hit in a proteome dataset of wild type and Mecp2 mutant rat cortex (Figure S1D). We identified Myg1, a mitochondrial protein (Grover et al., 2019), and MECP2 as the top two most important classifiers to discern between wild type and *Mecp2* mutant brain tissue (Figure S1D). These machine learning analyses performed to different degrees as determined by area under the curve in ROC analysis (0.72–0.92, Figures 2F and S1) and/or confusion matrices (Figures 2F and S1). Thus, interrogation of CSF proteome datasets with boosting mathematical algorithms provides similar answers as to protein families enriched in rat and mouse mutant CSFs.

The above-described changes to the *Mecp2* mutant CSF could closely parallel brain proteome modifications. Alternatively, changes to the CSF proteome could be in proteins different from those in the *Mecp2* brain proteome yet both *Mecp2* proteomes representing alterations in the same compartment or pathway, a converging ontology. To address this question, we compared the rat *Mecp2* mutant cortex proteome of animals where we simultaneously collected CSF. Volcano thresholding by p value and fold of change identified 83 proteins whose expression was modified in *Mecp2* mutant cortex among 6,752 proteins quantified by TMT (Figure 2C). None of the *Mecp2*-sensitive cortical proteins overlapped with rat CSF hits (Figure 2G). Correlation analysis searching for CSF hits in *Mecp2* mutant rats that were also quantified in rat cortices found 67 proteins shared between the rat cortical dataset and the rat CSF *Mecp2*-sensitive proteome. These 67 proteins showed no correlation in their expression levels (Figure 2H). These results show that there is a limited capacity to predict specific protein candidates in *Mecp2* mutant CSF from brain proteomes and vice versa.

The *Mecp2* mutant rodent CSF differs from the human Rett CSF in that the former represents brain tissue homogenously deficient in Mecp2 protein. In contrast, the human CSF proteome reports a genetically mosaic female brain (Lyst and Bird, 2015; Banerjee et al., 2019; Sandweiss et al., 2020). We studied a cohort of individuals with Rett syndrome where CSF was collected as part of a phase I clinical trial (Khwaja et al., 2014). These participants were subjected to extended treatment with recombinant IGF-1 during the trial. CSFs were collected from 12 female individuals (5.7 $\pm$ 2.5 years, range 2–10 years, average $\pm$ SD (Khwaja et al., 2014)). In seven participants, fluids were collected before and after treatment; in two participants, collections occurred only before treatment; and in three participants, CSF was sampled only after IGF-1 treatment. The phase I clinical trial did not include typically developing control participants because of ethical constraints (Khwaja et al., 2014). Even though IGF-1 treatment did not improve clinical outcomes in Rett subjects (O'Leary et al., 2018), we reasoned that if IGF-1 treatment were to modify some aspects of CSF proteome, it should do so by changing CSF proteins whose expression was *Mecp2*-sensitive in rodents. In addition, we hypothesized that any IGF-1-induced proteome modifications in individuals with Rett syndrome should be in the opposite direction of what we observed in *Mecp2*-null rodents and conditioned media from *MECP2* null neurons. The Rett CSF proteome revealed a discrete number of proteins sensitive to IGF-1 (Figures 2I–2K). We identified by volcano thresholding 7 proteins whose expression was decreased after IGF-1 treatment and 17 proteins whose expression was increased (Figures 2I–2K). Of importance, the most prominent among the IGF1-upregulated proteins were the high-density lipoprotein proteins APOA1, APOC1 and APOM; a change precisely in the opposite direction of what we found in mouse and rat *Mecp2* mutants. Apolipoproteins APOC1, APOC2, and APOM as well as the apolipoprotein regulatory factor PCSK9 were also identified by machine learning (Figure 2L). In fact, APOM was assigned the top priority as a discriminatory factor in a decision tree segregating Rett participants by their IGF-1 treatment (Figure 2K). Our data suggest that discrete CSF proteome changes correlate with IGF-1 treatment in Rett female individuals. In the case of apolipoproteins, protein identity and the direction of change in humans can be informed from the study of *Mecp2* mutant CSF proteomes in preclinical animal models.

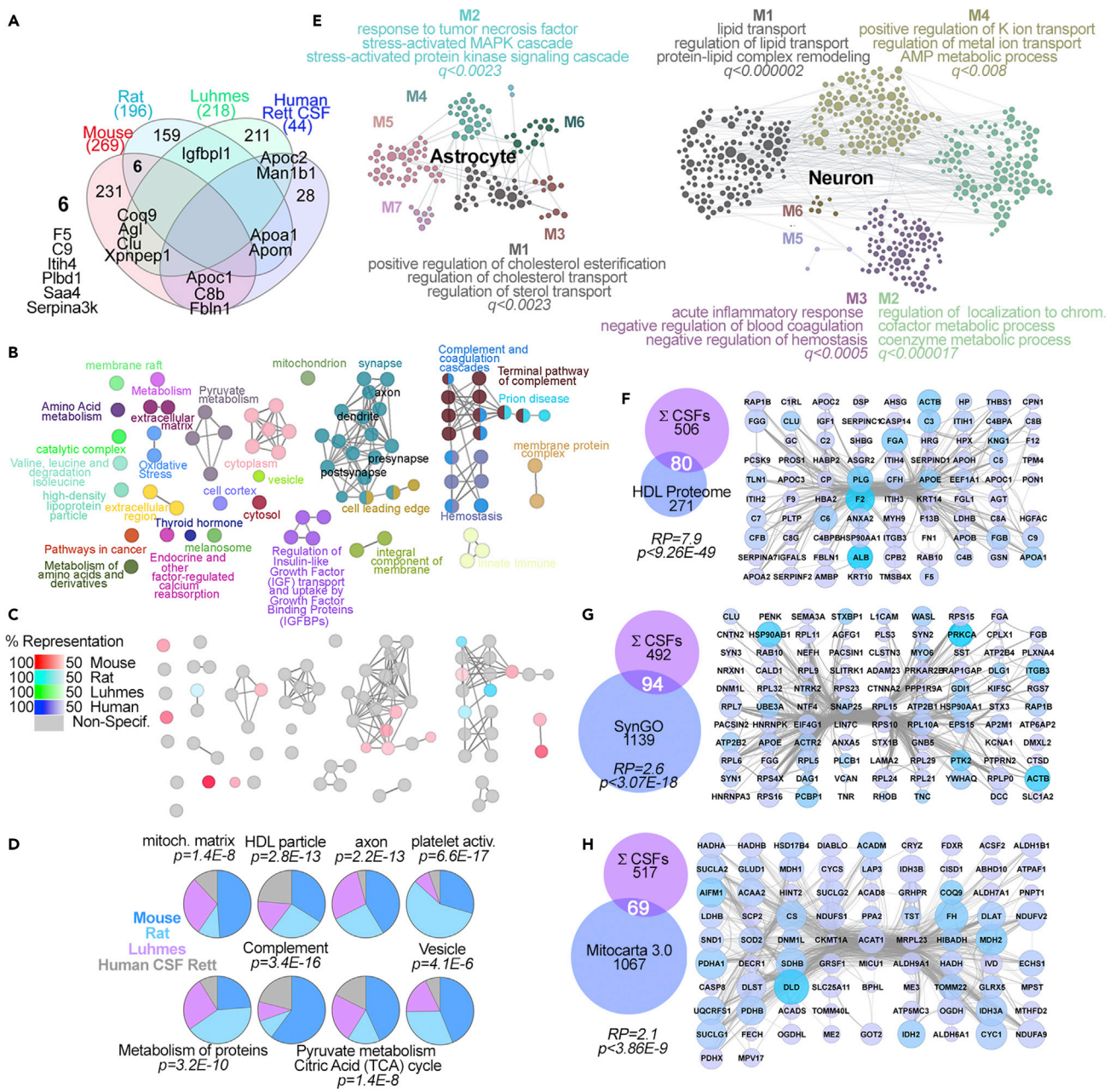


Figure 3. Convergent ontologies inferred from Rett syndrome CSF and CSF-mimic proteomes

(A) Venn diagram depicted hit overlaps among the four experimental systems studied. Hits represent the sum of volcano thresholding- and machine learning-selected hits.

(B and C) Integrated gene ontology analysis of the four datasets in A annotated with the experimental system that originated the dataset. Nodes represent individual ontologies. GO_CellularComponent, KEGG, Reactome, and WikiPathways were queried with the ClueGo application. All ontologies have p<0.001. Exact hypergeometric probability Bonferroni corrected. C, shows the percent contribution of each experimental system to each ontology. Gray denotes ontologies represented by all four experimental systems.

(D) Pie charts of the percent contribution of each experimental system to the top ontologies identified in (B and C), p values hypergeometric probability Bonferroni corrected for the ontology.

(E) Shows ontology analysis using the same datasets in B-C but using the Bayesian engine HumanBase. Analyses were performed either with astrocyte- or neuron-centric queries. Nodes correspond to genes and their functional interrelationships. Nodes are grouped into clusters M(n), prioritized by q value calculated using one-sided Fisher's exact tests and Benjamini–Hochberg corrections. Most significant cluster is M1.

Figure 3. *Continued*

(F–H) Show Venn diagrams and protein-protein interaction networks of the sum of datasets presented in A overlapping with the curated HDL proteome in F, the Synapse knowledge database of annotated genes in (G), and the Mitocarta 3.0 annotated mitochondrial proteins in (H). Venn diagram p values calculated with exact hypergeometric probability and the representation factor (RP) estimates enrichment beyond what is expected by chance.
See extended ontological data in Table S4.

Composite ontologies of *MECP2*/*Mecp2* mutant cerebrospinal fluids and conditioned media

The secreted proteomes from human cultured neurons, rat, mouse, and human mutant CSFs revealed some common proteins across these diverse experimental systems. These proteins belong to apolipoproteins, complement, or mitochondrial pathways (Figure 3A). We asked if these *Mecp2*-sensitive hits overlapped just as isolated occurrences or, instead, secreted proteomes obtained from each experimental system sampled a common ontological space. We tested this hypothesis using ClueGO, an application that performs composite and comparative enrichment tests based on hypergeometric distributions (Bindea et al., 2009). To test the robustness of our ontological predictions, we used HumanBase, a genomics data-driven Bayesian machine learning algorithm that identifies functional modules in tissues and cells (Greene et al., 2015). We collated proteins from each of the four *Mecp2*-and *MECP2*-deficient experimental systems, selected by volcano thresholding plus machine learning (Figure 1A), to identify a space of shared ontologies (Figure 3A). We simultaneous queried an ontological database composite with ClueGo (GO CC, REACTOME, KEGG and WikiPathways). Each of the four mutant proteome datasets was tagged in ClueGo to discern their individual contribution to each ontology. We identified a space of 87 ontologies significantly represented in all the*Mecp2*-, *MECP2*-sensitive, and human Rett CSF proteome datasets (Bonferroni corrected p values $< 10 \times 10^{-3}$, Figure 3B). These ontologies revealed a significant enrichment of hits in mitochondrial compartments and pathways, pyruvate and amino acid metabolism, complement subunits, HDL lipoproteins, and synapse-related ontologies (Figure 3B). Of importance, these ontologies were qualified as non-dataset specific, as each mutant proteome dataset contributed less than 50% of hits to each one of these ontologies (Figures 3C and 3D). For example, the HDL particle ontology was made up by 34, 26, 16 and 25% of hits derived from mouse, rat, neuron conditioned media, and human Rett CSF, respectively (Figure 3D, GO:0034364, Bonferroni corrected group p value=2.81E-13). We confirmed these ontological findings using HumanBase, where we identified lipid and cholesterol transport ontologies as the most significantly enriched functional modules (Figure 3E). This outcome was similar if we performed HumanBase analyses either with astrocyte- or neuron-centric queries (Figure 3E). Finally, we confirmed that the *Mecp2*-, *MECP2*-sensitive, and human Rett CSF proteomes were enriched in HDL, synaptic, and mitochondria annotated proteins interrogating different databases. We used the curated HDL proteomes database, the SynGo knowledge base of annotated synaptic proteins, and the Mitocarta 3.0 database of annotated mitochondrial proteins (Figures 3F–3H) (Davidson and Shah, 2021; Koopmans et al., 2019; Rath et al., 2021). The collated *Mecp2*-and *MECP2*-sensitive proteome contained 80 proteins in common with the HDL proteome (Figure 3F). This represents a significant 7.9-fold enrichment above what is expected by chance (Figure 3F, $p < 9.26E-49$). Among these overlapping HDL proteins, we found diverse apolipoproteins, complement subunits, antiproteases of the Serpin family, and factors such as clusterin (Clu), and PCSK9. All these overlapping HDL components formed an interconnected network of protein-protein interactions as determined with the Genemania application (Figure 3F) (Montojo et al., 2010). Similarly, the collated *Mecp2*-, *MECP2*-sensitive, and human Rett CSF proteome was also significantly either enriched in interconnected synaptic or mitochondrial proteins >2-fold above what is predicted by chance. These findings demonstrate that Rett syndrome CSF proteomes from diverse species converge on a common set of ontologies.

Functional validation of Rett syndrome CSF ontologies

We tested the predictive value of the ontologies derived from the collated *Mecp2*-and *MECP2*-sensitive CSF proteomes by measuring functional outcomes inferred from these ontologies. Mitochondrial ontologies were significantly enriched across the collated *Mecp2*-and *MECP2*-sensitive CSF proteomes (Figures 3B–3E and 3H) suggesting mitochondrial dysfunction in *MECP2*/*Mecp2*-deficient cells. Therefore, we measured mitochondrial respiration with Seahorse flow oximetry. We used as a model the post-mitotic human neurons differentiated from the LUHMES cell line (Figure 1), with their *MECP2* gene edited by CRISPR-Cas9 (Shah et al., 2016). These *MECP2*-null cells were infected with a lentivirus expressing the human *MECP2* gene from a vector also encoding a Red Fluorescent Protein (TagRFP) under the control of a *MECP2* minimal promoter. Controls were performed with *MECP2* null cells infected with a virus carrying just the TagRFP reporter (Figure 4). Viral expression of MECP2-RFP restored MECP2 expression in most null

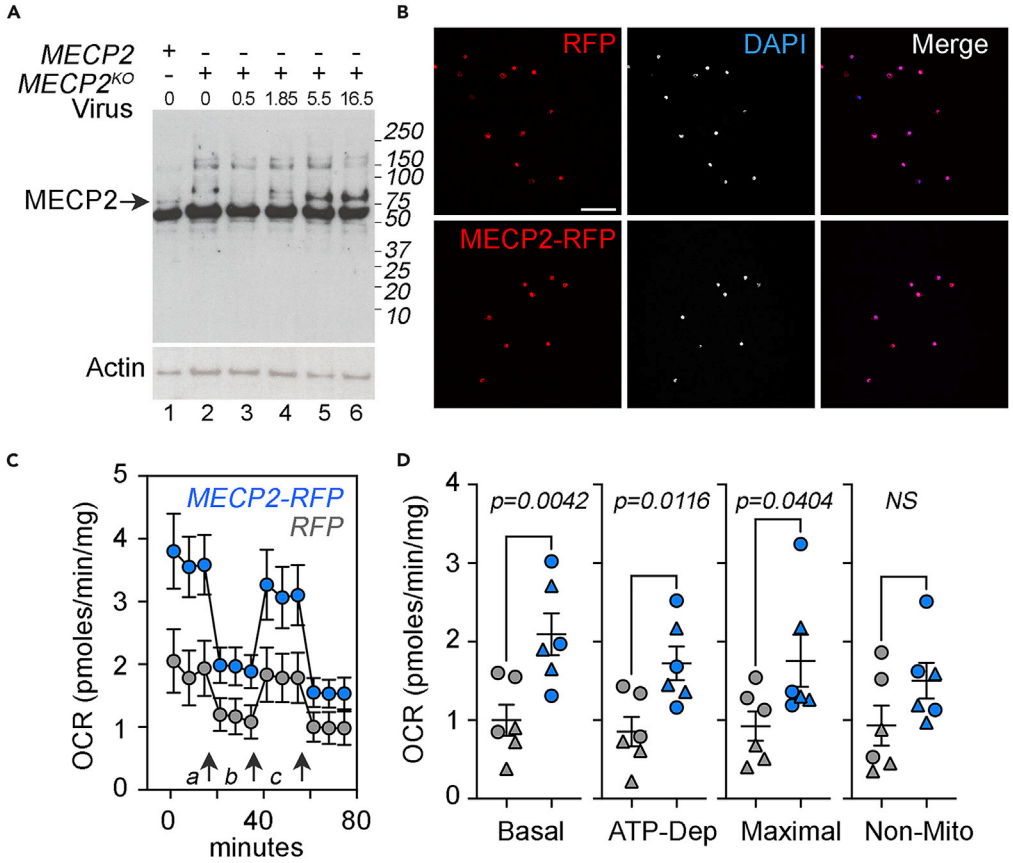


Figure 4. Expression of MECP2 increases mitochondrial respiration in *MECP2* null post-mitotic human neurons

Wild type and *MECP2* null neurons differentiated from the LUHMES cell line were infected with lentiviruses encoding RFP or MECP2-RFP for 72 h.

(A) presents immunoblot of wild type and *MECP2* null cells (lanes 1 and 2) mock infected or infected with increasing amounts of a virus encoding MECP2-RFP (lanes 3-6).

(B) depicts indirect immunofluorescence with RFP antibodies and DAPI for DNA of *MECP2* null cells infected with a virus encoding either RFP or MECP2-RFP. Calibration bar 50 μm.

(C and D) Seahorse stress test of cells as in A and B. Arrows a to c indicate the moment of addition of oligomycin to derive ATP dependent respiration (D), FCCP to drive maximal respiration (D), and rotenone plus antimycin to estimate non-mitochondrial respiration (D), respectively. p values were calculated with a two-sided permutation t test. Circles and triangles in (D), represent independent mutant LUHMES clones. (C) and (D) Mean ± Standard Error of the Mean.

cells (>80% infection) as determined by immunoblotting and immunofluorescent microscopy with antibodies against MECP2 or RFP (Figure 4A, compare lanes 2 with 4–6, and Figure 4B). Similarly, most cells infected with a control RFP virus expressed this fluorescent protein (Figure 4B, >80% infection). We used these conditions to measure mitochondrial oxygen consumption (Figure 4C) 72 h after viral infection in two *MECP2* null clones (Figure 4D, triangles and circle symbols). Basal, oligomycin-sensitive, and maximal mitochondrial respiration were significantly increased by MECP2-RFP expression as compared to RFP controls. Non-mitochondrial oxygen consumption rates, measured after treatment with rotenone and antimycin to abolish mitochondrial respiration, were not altered by MECP2-RFP expression. These results demonstrate that re-expression of MECP2 increases mitochondrial oxygen consumption in *MECP2*-null post-mitotic neurons. These mitochondrial respiration phenotypes strengthen the argument that ontologies derived from *Mecp2*-and *MECP2*-sensitive CSF proteomes could predict disease biomarkers.

Biochemical and genetic validation of Rett syndrome CSF ontologies

We used convergent Rett ontologies to inform a selection of proteins for confirmatory studies and to assess their potential as disease biomarkers. The HDL lipoprotein proteome was the most significantly enriched ontology among all mutant secreted proteomes (Figures 3B, 3E, and 3F). We performed absolute

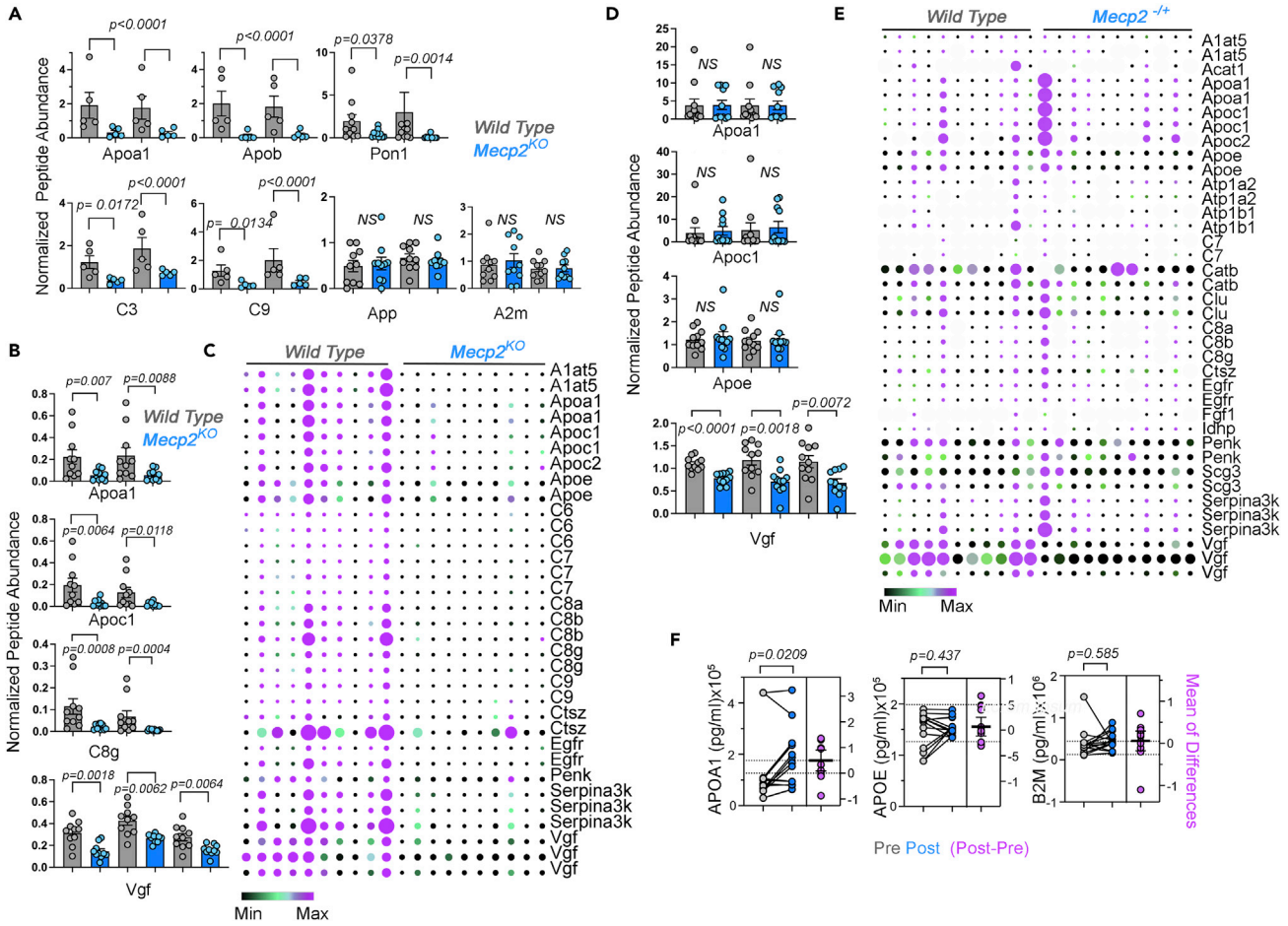


Figure 5. Ontologically selected and confirmed Rett syndrome CSF proteome hits

(A and B) shown independent confirmatory analyses using isotopologue peptides mapping the primary sequence of the indicated proteins using AQUA mass spectrometry in rat CSF (A) or a modified AQUA approach in *Mecp2* null male mouse CSF (B and C) and in *Mecp2*$^{-/+}$heterozygotic female mouse CSF (D and E). In A, femtomoles of the endogenous CSF peptide were normalized to a randomly chosen control sample. In (B–E), the ratio of the CSF endogenous peptide to the isotopologue peptide was used to quantify relative analyte abundance. Gray bars correspond to wild type CSF and blue bars *Mecp2* mutant CSF either from male null (B and C) or female heterozygotes (D and E). All analytes were measured independently with 2–3 isotopologue peptides as standards. In A two batches of 5 rats of each genotype were used whereas in (B and C) one batch of 10 mice of each genotype was analyzed, and in (C and D) one batch of 11 females per genotype was used.

(C and E) depict heatmap of all modified AQUA determinations performed in mouse CSF samples selected because they showed significant differences between genotypes in males. Every isotopologue peptide corresponds to a row. Data are depicted as row median divided by the row median absolute deviation both as heatmap and by symbol size.

(F) MesoScale ELISA confirmation of APOA1 levels in Rett individuals CSF before and after IGF-1 treatment.

p values in (A, B, D and F) were calculated with a two-sided permutation t test. See Table S5 for raw data.

quantification (AQUA) of proteins by mass spectrometry to confirm expression changes in HDL apolipoproteins in independent cohorts of male *Mecp2*-null rats (Figure 5A) and mouse CSF from *Mecp2*-null male mice (Figures 5B and 5C) (Gerber et al., 2003). We extended these AQUA confirmatory studies to CSF from female *Mecp2*$^{-/+}$ mice (Figures 5D–5E), a model with genetic validity for the human disease that preponderantly affects females. We selected female *Mecp2*$^{-/+}$ mice at six weeks of age to compare to male *Mecp2*$^{-/y}$ mice of the same age. These female mice offered a way to measure how gene dosage plus a presymptomatic status could affect the differential expression of an analyte in CSF. *Mecp2* deficient males have overt symptoms at three weeks of age whereas female animals develop symptoms at ~3 months (Guy et al., 2001; Schaevitz et al., 2013; Ribeiro and MacDonald, 2020). Finally, we tested CSF proteins differentially expressed in Rett syndrome female human subjects before and after IGF1 treatment by ELISA (Figure 5F) using the experimental design and subject cohort studied by Khwaja et al.Figures 2J–2L) (Khwaja et al., 2014).

AQUA quantification of the HDL lipoprotein particle components Apoa1, Apob, Pon1, C3, and C9 revealed decreased levels in *Mecp2* null males as compared to wild type rat male CSF (Figure 5A). Levels of loading controls, App and A2m, were not affected by genotype in male rats (Figure 5A). These findings were extended to *Mecp2*-null male mouse models (Figures 5B and 5C). We used isotopolog peptide standards to measure proteins associated with HDL particles in mouse male CSF. Diverse apoliproteins (Apoa1, Apoc1, Apoc2, Apoe), complement subunits (C6, C7, C8a, C8b, C8g, C9), and antiproteases (A1at5, Serpina3k) were reduced in *Mecp2* null males compared to controls (Figures 5B and 5C). Similarly, the levels of other secreted proteins such as the neurosecretory protein VGF, proenkephalin-A (Penk), cathepsin Z (Ctsz), or the transmembrane epidermal growth factor receptor (Egfr) were robustly and significantly reduced in the CSF of *Mecp2*-null male mice (Figure 5C). To address if these analytes are sensitive to the dosage of a *Mecp2*-null allele and the presence of symptoms, we used AQUA to measure these analytes in CSF from heterozygote *Mecp2*$^{-/+}$ female mice at 6 weeks of age. HDL-annotated proteins; such as Apoa1, Apoc1-2, Apoe, C7, C8, and Clu; did not detectably change their abundance in the CSF from heterozygote *Mecp2*$^{-/+}$ female mice as measured by AQUA. This was despite their robust change in the CSF of two male *Mecp2*$^{-/y}$ rodent models of the same age (Figures 5A–5C). However, the levels of Vgf were significantly reduced in heterozygote *Mecp2*$^{-/+}$ female mice CSF (Figures 5D and 5E) to a similar extent found in *Mecp2*-null male mice.

Finally, we focused on the nine human subjects where CSF samples were obtained before and after IGF1 treatment. Treatment modified a Rett CSF composition discretely, mostly driven by changes in the levels of some apolipoproteins such as APOA1 (Figures 2J–2L). We confirmed that IGF1 treatment increased the content of APOA1 in Rett participant's CSF by ELISA (Figure 5F). In contrast, two proteins whose expression was not modified by treatment in TMT mass spectrometry quantifications, APOE and B2M (Figures 2J–2L), did not change their levels after treatment in ELISA assays (Figure 5C). These results confirm protein hits and ontologies in human and rodent Rett syndrome models.

Ontology selected analytes performance as putative biomarkers of Rett syndrome

To evaluate the potential of ontologically selected *Mecp2*-sensitive hits to serve as disease biomarkers, we addressed the following questions. First, do CSF *Mecp2* hits discriminate between genetic forms of autism spectrum disorder? We selected a null mutation in the *Cdkl5* gene, which is causative of the CDKL5 deficiency disorder, an X-linked neurodevelopmental disorder. The behavioral and brain anatomy phenotypes of a mouse model of this syndrome closely mimic those in Rett syndrome mouse models (Zerbi et al., 2021). We found that the CSF proteome of *Cdkl5* male mutants was different from *Mecp2* male mutant CSF, as indicated by the 15 AQUA confirmed proteins selected from *Mecp2* CSF whose levels remained unchanged in *Cdkl5* null male CSF. (Figure 6A). There is no significant overlap between *Mecp2* and *Cdkl5* male mutant CSFs when considering all differentially expressed proteins in CSF from *Mecp2*-null male rodents as compared to all proteins differentially expressed in the CSF from *Cdkl5*-null male mice. The later, we identified by TMT mass spectrometry in two cohorts of *Cdkl5*-null males totaling 20 animals per genotype (Figure 6B, hypergeometric test, and unpublished data). Similarly, there was minimal ontological overlap between *Mecp2*-null and *Cdkl5*-null male mice as assessed with the Metascape tool (Figure 6C) (Zhou et al., 2019). These findings demonstrate that *Mecp2* CSF differentially expressed proteins discriminate between phenotypically related forms of neurodevelopmental disorders.

Second, we interrogated whether selected *Mecp2* CSF proteins were expressed in disease-relevant cell types using publicly available datasets (Figures 7A and 7B) (Scholz et al., 2011; Zhang et al., 2016; Yao et al., 2021). HDL apolipoproteins such as Apoa1, Apoe, Pon1, and Clu are expressed in neurons and astrocytes (Figure 7A). Single cell RNAseq datasets showed that transcripts encoding these HDL proteins were expressed in diverse populations of glutamatergic and GABAergic neurons across the multiple layers of the cortex and hippocampus (Figure 7B). *Apoc1* mRNA or its protein were undetectable in neurons and glia but present in plasma and expressed in the choroid (compare Figures 7A and 7B and(Xu et al., 2021)). In contrast, complement components (C7, C8b, C8g, and C9), growth factors (Igf1 and Vgf), mitochondrial proteins (Acat1, Coq9, and Micu1), and synaptic annotated proteins (Snap25, Stx1b, Syn1, and Syn2) were expressed in diverse glial and neuronal cell populations to a different degree (Figure 7B). For example, C8b and C9 mRNAs were expressed in a discrete neuronal population whereas C7 was broadly expressed in cortex and hippocampus (Figure 7B). Thus, ontologically selected protein hits are expressed in disease-relevant neuronal and glial cell types.

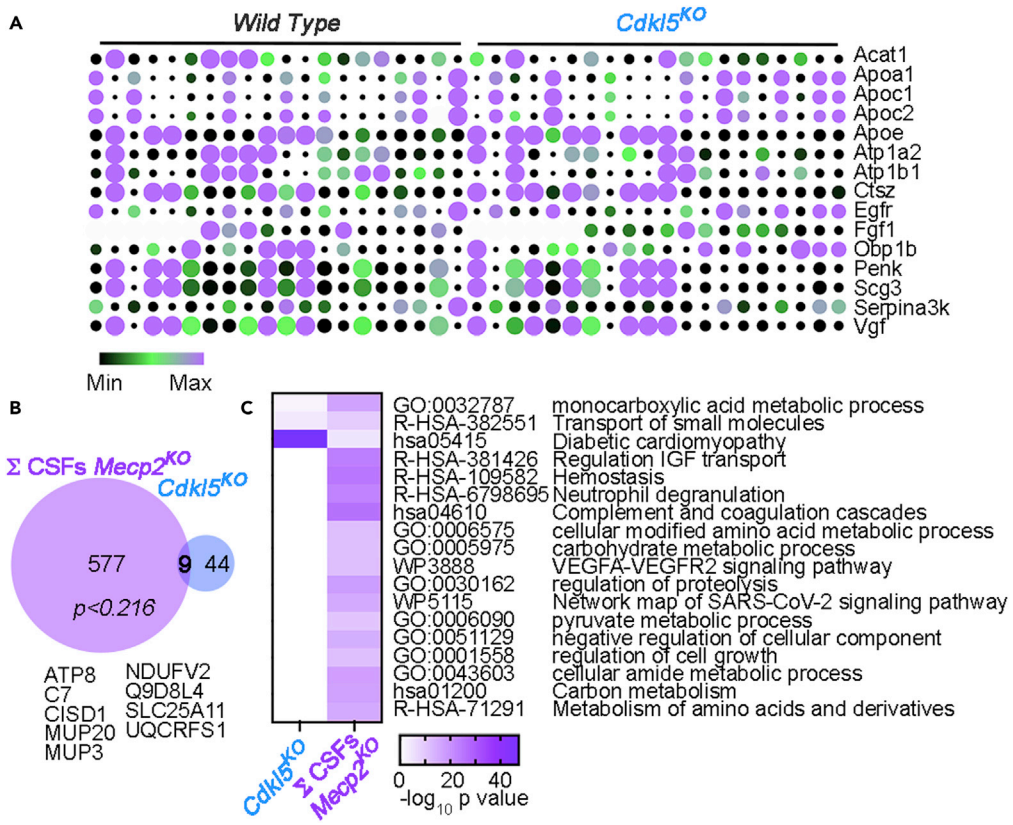


Figure 6. *Mecp2* and *Cdkl5* mutant CSFs are distinct

(A) Heatmap of ontology selected and confirmed analytes in from Figures 5B and 5C tested in the CSF of *Cdkl5*-null mice. Data correspond to the normalized abundance measured by TMT mass spectrometry in two independent cohorts of wild type and *Cdkl5*$^{-/y}$mutant males (total n = 20 per genotype). Data are depicted as row median divided by the row median absolute deviation both as heatmap and by symbol size.

(B) Venn diagram of total overlap between the collated *Mecp2*-and *MECP2*-sensitive CSF proteomes (Σ CSFs *Mecp2*^{KO}) and all differentially expressed proteins in *Cdkl5*$^{-/y}$mutant males as determined by TMT mass spectrometry.

(C) Shared ontologies as defined by the Metascape tool using the datasets in (B).

See extended data in Table S5.

Finally, we determined if selected and confirmed proteins could distinguish subjects by genotype. We performed ROC analysis focusing on the *Mecp2* mutant male mouse CSF hits as the animal cohort of the biggest size, the mouse TMT dataset (Figure 7C). Each ontologically selected analyte efficiently distinguished genotypes irrespective of if they belonged to either HDL lipoprotein, synapse, or mitochondrion category. In fact, all analytes ROC area under the curve were between 0.77 and 0.9 with significant p values (Figure 7C). In other words, these analytes have a 77 to 90% chance to distinguish wild type and mutant CSF. The ROC performance of an analyte was similar whether TMT or AQUA male datasets were analyzed (Figure 7C, compare gray and blue symbols). A similar ROC performance was obtained with Vgf AQUA measurements in heterozygote *Mecp2*$^{-/+}$ female mouse CSF samples (Figure 7C, pink symbols). None of these validated analytes experienced modifications in their mRNA expression across diverse brain regions in *Mecp2* mutant male mice, indicating that the utility of these analytes as putative biomarkers of *Mecp2* gene defect is restricted to their protein levels in CSF (Figure S2, Table S6). These findings demonstrate that CSF analytes identified in rodent models of Rett syndrome are sensitive, conserved, and represent specific candidates for Rett biomarkers with potential for human applications.

DISCUSSION

Here we demonstrate that an X-linked neurodevelopmental disorder, Rett syndrome, reproducibly and distinctively impacts the composition of the CSF. We defined a consensus proteome and ontological

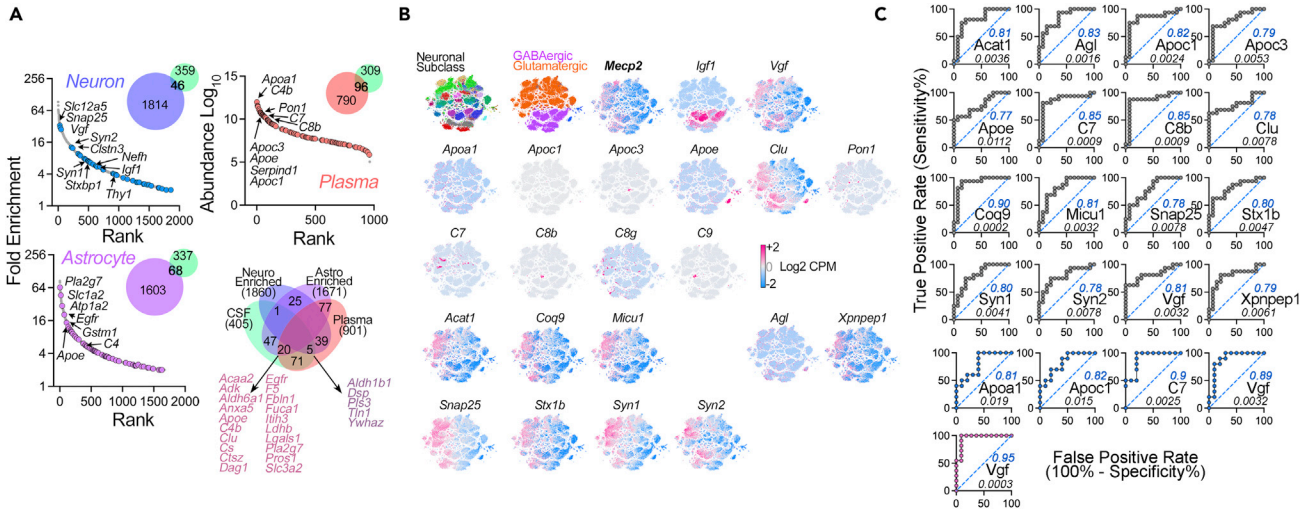


Figure 7. Expression patterns of ontology selected analytes in brain cells and plasma

(A) Fold enrichment and rank order of mRNAs most expressed in neurons and astrocytes according to Zhang (Scholz et al., 2011; Zhang et al., 2016) or abundance in the plasma proteome according to Geyer et al. (Geyer et al., 2016). Superimposed are CSF hits. Venn diagrams present overlaps with each cell type gene expression category or plasma proteome.

(B) Depicts a t-SNE cell atlas generated with the expression levels of all transcripts encoding selected hits from the mouse CSF Mecp2-sensitive proteome. The t-SNE atlas encompasses >20 areas of mouse cortex and hippocampus, totaling 76,307 cells (Yao et al., 2021). Color codes denote neuronal subclasses described by Yao et al. (Yao et al., 2021). Neurotransmitter annotation is depicted as well as the expression levels of Mecp2 mRNA across brain regions and cell types. Each atlas depicts the mRNA expression of the indicated analyte. Note analytes such as Apoc1 whose mRNA is not detectable in this dataset. t-SNE cell atlases were assembled using the Allen single-cell RNAseq dataset as describe by Wynne et al. (Wynne et al., 2021).

(C) ROC analysis of selected mouse CSF *Mepc2*-sensitive hits either quantified by TMT mass spectrometry (gray symbols) or by modified AQUA (blue symbols) in males. Pink symbols depict *Mecp2*$^{-/+}$ female mice. Number represents p value that tests the null hypothesis that the area under the curve = 0.50 (non-discriminatory). Blue figures depict area under the curve per each analyte.

See extended data in Table S5.

categories shared by four experimental systems across three species deficient in *Mecp2/MECP2*. These proteomes converged on proteins annotated to HDL lipoproteins, complement cascade, mitochondrial compartments, citrate cycle/pyruvate metabolism, synapse compartments, and secreted factors such as Vgf. The robustness of our findings is founded on the multipronged nature of our experimental design, which includes diversified *in vitro* and *in vivo* systems, multiple species studied, distinct mathematical processing of datasets capturing similarly annotated proteins, replication across different proteomic platforms (LFQ and TMT), and replicability across three sites for CSF collection and two sites of mass spectrometry analysis (Figure 1C). Although, the different mutant CSF proteomes produced discrete overlap across individual analytes, they all shared significant overlap at the ontology level. We functionally confirmed mitochondrial compartment-related ontologies, which were bioinformatically predicted, by discovering mitochondrial respiration sensitive to MECP2 expression in human neurons (Figure 4). We used these convergent ontologies to inform the selection of analytes for orthogonal confirmatory efforts. These confirmatory approaches independently validated our LFQ and TMT findings. Confirmed analytes provided additional proof of principle to the use of convergent ontologies as a strategy to select analytes suitable to discriminate a mutant genotype across species and quantification platforms. For example, even though Apoa1 was not identified as a significant hit in the mouse CSF TMT proteome, selection of Apoa1, based on the ontology to which it belongs, predicted and resulted in robust confirmation across all species studied and platforms used. Ontologically selected hits performed well as putative biomarkers as determined by ROC analysis and the capacity of multiple *Mecp2*-sensitive hits to discriminate *Mecp2* mutant CSF from another phenotypically related neurodevelopmental disorder, the CDKL5 deficiency disorder in male preclinical disease models. Of the analytes identified in our studies, we think the expression of APOA1 and VGF are likely the best candidates for CSF biomarkers of disease. These two proteins were modified in female Rett cases or female heterozygotic mouse either by genotype or IGF-1 treatment. We propose that *Mecp2* mutant CSF ontologies inform robust CSF analytes to act as Rett syndrome biomarkers in humans. This contention awaits confirmation in human CSF from controls and Rett cases that, until recently, were not available (Zandl-Lang et al., 2022).

The mechanisms that account for the changes in the secreted proteome described here have not been explored yet. However, we reasoned that if the *Mecp2* secreted proteome were to be caused directly by a *Mecp2*-dependent transcriptional defect, there should be parallel modifications in the cellular and secreted proteomes. We found that this is not the case. The CSF and brain cortex proteomes did not correlate. In fact, transcriptomic analysis of several ontologically selected CSF protein hits showed that none of these proteins exhibited correlated modifications in their mRNA levels in brain. Similarly, the neuronal cell and conditioned media proteomes poorly overlapped. These results argue that indirect mechanisms downstream of *Mecp2*-dependent transcription, such as network activity, likely drive the secreted proteome phenotypes.

We have minimized the possibility that accidental plasma contamination of the CSF is a driving factor for some of the CSF expression differences observed. However, CSF protein composition is defined by factors that normally transcytose from the plasma to the CSF plus contributions from neuronal and non-neuronal cells in the brain parenchyma, the choroid plexus, and ependymal cells (Tsujita et al., 2021; Stukas et al., 2014). Therefore, our findings likely represent contributions of diverse cell types in brain to the *Mecp2*-sensitive CSF proteome. With few exceptions, many of our CSF *Mecp2*-sensitive proteins could be attributed to multiple cell types. For example, Apoe and Clu (Apoj) could be ascribed to secretions from astrocytes or the choroid plexus, where Apoe and Clu (Apoj) rank among the most expressed mRNAs; they could be ascribed to neurons, where these mRNAs are also expressed yet at lower levels. Of importance, Apoe brain levels are locally controlled without contributions from plasma (Linton et al., 1991). We directly tested the hypothesis that the expression of apolipoproteins is cell-autonomously controlled in neurons as demonstrated by the reduced levels of Clu (Apoj) in the conditioned media of human post-mitotic neurons. On the other extreme, Vgf and Igf1 mRNAs are expressed in neurons with preferences for neuronal cell types. Such is the case of Igf1, which is mostly expressed in GABAergic interneurons and is minimally or not expressed in glia, endothelial cells, and the choroid plexus (Zhang et al., 2014; Xu et al., 2021). Thus, the *Mecp2* secreted proteome offers multiple analytes to assess phenotypes in multiple brain cell types lacking *Mecp2*.

Vgf is uniquely reduced both in *Mecp2*$^{-/y}$ male and *Mecp2*$^{-/+}$ female mouse models among the analytes identified in CSF as responsive to the *Mecp2* gene defects. In contrast, proteins annotated to HDL ontologies did not change their levels in *Mecp2*$^{-/+}$ female CSF whereas they changed in *Mecp2*$^{-/y}$ male CSF. A possible interpretation for these findings is that the levels of diverse analytes differ in their dependency on gene dosage and whether mice are symptomatic or not. However, an alternative interpretation is that technical limitations measuring Vgf and HDL-annotated protein preclude determinations of these analytes that capture genotype- and symptomology-dependency. For example, the coefficient of variation for Vgf measurements was the lowest, both in male and female of all genotypes (13-38%). However, coefficient of variation for HDL-annotated protein measurements ranged between 93 and 201%. This variability would prevent the detection of subtle changes in analyte levels driven by gene dosage modifications. Thus, a more conservative interpretation of our results is that the female *Mecp2*$^{-/+}$ genotype reduces the magnitude of the differences of some CSF analytes below the threshold of detection.

All *Mecp2* secreted proteomes converged on robust ontologies. Proteins annotated most significantly to HDL lipoprotein, complement, synapse, mitochondria, and mitochondrial pathways such as citrate cycle/pyruvate metabolism ontologies. These consensus ontologies likely point to pathogenic mechanisms in Rett syndrome. For example, the effects of *Mecp2* mutations on synaptic morphology, function, and plasticity have been extensively documented (Na et al., 2012; Banerjee et al., 2019). However, HDL and mitochondrial ontologies have received less attention. HDL particles are assembled by astrocytes and microglia. These lipoproteins transport cholesterol between glial cells and neurons. Thus, a possible mechanism to account for the decreased levels of HDL apolipoproteins in CSF is either a decreased production/secretion by *Mecp2* deficient glial cells or an increased clearance by cells that express HDL receptors in brain, such as neurons (Wellington and Frikke-Schmidt, 2016; Liu et al., 2010). We favor the decreased HDL production model as it can explain the observed increased cholesterol content in brain at postnatal day 56, despite decreased expression of cholesterol synthesis enzymes and decreased *de novo* cholesterol synthesis, as well as the decreased cholesterol content in the CSF of Rett syndrome cases (Buchovecky et al., 2013; Zandl-Lang et al., 2022). We postulate that decreased HDL lipoproteins levels in CSF may be a factor contributing to the accumulation of cholesterol in brain and the concurrent inhibition by product of cholesterol synthesis. A second ontology strongly represented in our datasets is mitochondria compartments and pathways, which we functionally confirmed by measuring mitochondrial respiration (Figure 4). Pyruvate and lactate are increased in the CSF of individuals with Rett syndrome, and Krebs cycle

metabolites are increased in the brain of *Mecp2* mutant mice. This suggests connections between CSF glycolysis and Krebs cycle ontologies and proteins and these metabolites (Budden et al., 1990; Matsuishi et al., 1994). However, there are not enough studies to tie together our observations in CSF with potential models of mitochondrial dysfunction in *Mecp2* mutant cells (Kriaucionis et al., 2006; Shulyakova et al., 2017; Jagtap et al., 2019; Hirofuji et al., 2018; Grosser et al., 2012). Our findings support the idea that *Mecp2* mutant CSF ontologies predict putative brain mechanisms disrupted by mutations in *Mecp2*. We propose that Rett syndrome is a synaptic and metabolic disorder of neurodevelopment.

Limitations of the study

The invasive nature of CSF collection in humans, in particular children, is a severe limitation precluding the acquisition of CSF samples from neurotypic and diseased subjects. Thus, our human studies rely on an already studied cohort of only Rett subjects, before and after a treatment was applied to them. We did not compare to a neurotypic control group. To circumvent this limitation, we devised a strategy involving CSF-mimic fluid from cultured human neurons and CSF from two rodent models of the disease to build a robust portfolio of CSF analytes and ontologies strongly and consistently associated with Rett syndrome. The potential of these analytes and ontologies to act as biomarkers of disease awaits careful testing in humans and non-human models of the disease undergoing genetic or pharmacological therapies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - Human samples
  - Animal models
  - Cell lines
- METHOD DETAILS
  - Rat CSF sample collection
  - Mouse CSF collection
  - LUHMES conditioned media preparation
  - LUHMES cell viral transduction
  - Lysis
  - Immunofluorescence
  - Seahorse flux oximetry
  - Mass spectrometry Emory
  - *Tandem mass tag (TMT) labeling*
  - *High pH fractionation*
  - *Liquid chromatography tandem mass spec for TMT*
  - *Liquid chromatography tandem mass spec for PRM*
  - *Data processing protocol*
  - Mass spectrometry Ann Arbor
  - *Label-free quantification mass spectrometry*
  - *TMT quantification mass spectrometry*
  - *AQUA mass spectrometry*
  - *Data processing*
  - RNAseq and single cell RNAseq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Bioinformatic analyses
  - Statistical analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104966.

ACKNOWLEDGMENTS

VF was funded by the Rett Syndrome Research Trust, the Loulou Foundation, and NIH 1RF1AG060285. S Rangaraju was partly funded by the NIH (5R01NS114130). S Rayaprolu was partly supported by the NIH (F32 AG064862).SC was funded by the Rett Syndrome Research Trust and Simons Initiative for the Developing Brain. JJF was funded by the Rett Syndrome Research Trust. VF is grateful for mitochondria provided by Maria Olga Gonzalez.

AUTHOR CONTRIBUTIONS

Conceptualization: S.A.Z., R.C., D.T., and V.F.; Methodology: S.A.Z., D.D., L.P., K.K.E.G., B.M., S.R., P.K., and S.R; Investigation: S.A.Z., D.D., L.P., K.K.E.G., and B.M; Visualization: V.F., Writing – original draft: V.F. and S.A.Z.; Writing – review and editing: V.F., S.A.Z., R.C., A.B., and D.T.; Reagents: R.S., L.K., L.C.S., M.H., A.B., and J.J.F.

DECLARATION OF INTERESTS

Mustafa Sahin reports grant support from Novartis, Biogen, Astellas, Aeovian, Bridgebio, and Aucta. He has served on Scientific Advisory Boards for Novartis, Roche, Regenxbio, SpringWorks Therapeutics, Jaguar Therapeutics and Alkermes. James J. Fink is an employee of Q-State Biosciences. VF is a member of the iScience editorial board.

Received: June 17, 2022
Revised: July 30, 2022
Accepted: August 12, 2022
Published: September 16, 2022

REFERENCES

Abbasi, D.A., Nguyen, T.T.A., Hall, D.A., Robertson-Dick, E., Berry-Kravis, E., and Cologna, S.M. (2021). Characterization of the cerebrospinal fluid proteome in patients with fragile X-associated tremor/ataxia syndrome. Cerebellum *21*, 99–100.

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. *46*, W537–W544.

Andrews, S. (2012). FASTQC [Online]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Banerjee, A., Miller, M.T., Li, K., Sur, M., and Kaufmann, W.E. (2019). Towards a better diagnosis and treatment of Rett syndrome: a model synaptic disorder. Brain *142*, 239–248.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics *25*, 1091–1093.

Boire, A., Zou, Y., Shieh, J., Macalinao, D.G., Pentsova, E., and Massagué, J. (2017). Complement component 3 adapts the cerebrospinal fluid for leptomeningeal metastasis. Cell *168*, 1101–1113.e13.

Buchovecky, C.M., Turley, S.D., Brown, H.M., Kyle, S.M., Mcdonald, J.G., Liu, B., Pieper, A.A., Huang, W., Katz, D.M., Russell, D.W., et al. (2013). A suppressor screen in Mecp2 mutant mice implicates cholesterol metabolism in Rett syndrome. Nat. Genet. *45*, 1013–1020.

Budden, S.S., Myer, E.C., and Butler, I.J. (1990). Cerebrospinal fluid studies in the Rett syndrome: biogenic amines and beta-endorphins. Brain Dev. *12*, 81–84.

Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T.C., Qin, J., and Zoghbi, H.Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. Science *320*, 1224–1229.

Chao, H.T., Chen, H., Samaco, R.C., Xue, M., Chahrour, M., Yoo, J., Neul, J.L., Gong, S., Lu, H.C., Heintz, N., et al. (2010). Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. Nature *468*, 263–269.

Chao, H.T., Zoghbi, H.Y., and Rosenmund, C. (2007). MeCP2 controls excitatory synaptic strength by regulating glutamatergic synapse number. Neuron *56*, 58–65.

Chau, K.F., Springel, M.W., Broadbelt, K.G., Park, H.Y., Topal, S., Lun, M.P., Mullan, H., Maynard, T., Steen, H., Lamantia, A.S., and Lehtinen, M.K. (2015). Progressive differentiation and instructive capacities of amniotic fluid and cerebrospinal fluid proteomes following neural tube closure. Dev. Cell *35*, 789–802.

Cheng, T.L., Wang, Z., Liao, Q., Zhu, Y., Zhou, W.H., Xu, W., and Qiu, Z. (2014). MeCP2 suppresses nuclear microRNA processing and dendritic growth by regulating the DGCR8/Drosha complex. Dev. Cell *28*, 547–560.

Cholewa-Waclaw, J., Shah, R., Webb, S., Chhatbar, K., Ramsahoye, B., Pusch, O., Yu, M., Greulich, P., Waclaw, B., and Bird, A.P. (2019). Quantitative modelling predicts the impact of DNA methylation on RNA polymerase II traffic. Proc. Natl. Acad. Sci. USA *116*, 14995–15000.

Davidson, S., and Shah, A.S. (2021). HDL Proteome Watch [Online]. https://homepages.uc.edu/~davidswm/HDLproteome.html.

Deutsch, E.W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J.J., Kundu, D.J., García-Seisdedos, D., Jarnuczak, A.F., Hewapathirana, S., Pullman, B.S., et al. (2020). The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. Nucleic Acids Res. *48*, D1145–D1152.

Evans, J.C., Archer, H.L., Colley, J.P., Ravn, K., Nielsen, J.B., Kerr, A., Williams, E., Christodoulou, J., Gécz, J., Jardine, P.E., et al. (2005). Early onset seizures and Rett-like features associated with mutations in CDKL5. Eur. J. Hum. Genet. *13*, 1113–1120.

Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., and Gygi, S.P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc. Natl. Acad. Sci. USA *100*, 6940–6945.

Geyer, P.E., Kulak, N.A., Pichler, G., Holdt, L.M., Teupser, D., and Mann, M. (2016). Plasma proteome profiling to assess human health and disease. Cell Syst. *2*, 185–195.

Gokhale, A., Lee, C.E., Zlatic, S.A., Freeman, A.A.H., Shearing, N., Hartwig, C., Ogunbona, O., Bassell, J.L., Wynne, M.E., Werner, E., et al. (2020). Mitochondrial proteostasis requires genes encoded in a neurodevelopmental syndrome locus that are necessary for synapse function.

Journal of Neuroscience *41*, 6596–6616. https://doi.org/10.1523/JNEUROSCI.2197-20.2021.

Greene, C.S., Krishnan, A., Wong, A.K., Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C., et al. (2015). Understanding multicellular function and disease with human tissue-specific networks. Nat. Genet. *47*, 569–576.

Grosser, E., Hirt, U., Janc, O.A., Menzfeld, C., Fischer, M., Kempkes, B., Vogelgesang, S., Manzke, T.U., Opitz, L., Salinas-Riester, G., and Müller, M. (2012). Oxidative burden and mitochondrial dysfunction in a mouse model of Rett syndrome. Neurobiol. Dis. *48*, 102–114.

Grover, R., Burse, S.A., Shankrit, S., Aggarwal, A., Kirty, K., Narta, K., Srivastav, R., Ray, A.K., Malik, G., Vats, A., et al. (2019). Myg1 exonuclease couples the nuclear and mitochondrial translational programs through RNA processing. Nucleic Acids Res. *47*, 5852–5866.

Guy, J., Hendrich, B., Holmes, M., Martin, J.E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat. Genet. *27*, 322–326.

Higginbotham, L., Ping, L., Dammer, E.B., Duong, D.M., Zhou, M., Gearing, M., Hurst, C., Glass, J.D., Factor, S.A., Johnson, E.C.B., et al. (2020). Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic Alzheimer's disease. Sci. Adv. *6*, eaaz9360.

Hirofuji, S., Hirofuji, Y., Kato, H., Masuda, K., Yamaza, H., Sato, H., Takayama, F., Torio, M., Sakai, Y., Ohga, S., et al. (2018). Mitochondrial dysfunction in dopaminergic neurons differentiated from exfoliated deciduous tooth-derived pulp stem cells of a child with Rett syndrome. Biochem. Biophys. Res. Commun. *498*, 898–904.

Ho, J., Tumkaya, T., Aryal, S., Choi, H., and Claridge-Chang, A. (2019). Moving beyond P values: data analysis with estimation graphics. Nat. Methods *16*, 565–566.

Jagtap, S., Thanos, J.M., Fu, T., Wang, J., Lalonde, J., Dial, T.O., Feiglin, A., Chen, J., Kohane, I., Lee, J.T., et al. (2019). Aberrant mitochondrial function in patient-derived neural cells from CDKL5 deficiency disorder and Rett syndrome. Hum. Mol. Genet. *28*, 3625–3636.

Johanson, C.E., Duncan, J.A., 3rd, Klinge, P.M., Brinker, T., Stopa, E.G., and Silverberg, G.D. (2008). Multiplicity of cerebrospinal fluid functions: new challenges in health and disease. Cerebrospinal Fluid Res. *5*, 10.

Johnson, B.S., Zhao, Y.T., Fasolino, M., Lamonica, J.M., Kim, Y.J., Georgakilas, G., Wood, K.H., Bu, D., Cui, Y., Goffin, D., et al. (2017). Biotin tagging of MeCP2 in mice reveals contextual insights into the Rett syndrome transcriptome. Nat. Med. *23*, 1203–1214.

Johnson, E.C.B., Dammer, E.B., Duong, D.M., Ping, L., Zhou, M., Yin, L., Higginbotham, L.A., Guajardo, A., White, B., Troncoso, J.C., et al. (2020). Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. Nat. Med. *26*, 769–780.

Kaiser, K., Gyllborg, D., Procházka, J., Salašová, A., Kompaníková, P., Molina, F.L., Laguna-Goya, R., Radaszkiewicz, T., Harnoš, J., Procházková, M., et al. (2019). WNT5A is transported via lipoprotein particles in the cerebrospinal fluid to regulate hindbrain morphogenesis. Nat. Commun. *10*, 1498.

Kalluri, R., and Lebleu, V.S. (2020). The biology, function, and biomedical applications of exosomes. Science *367*, eaau6977.

Khwaja, O.S., Ho, E., Barnes, K.V., O'leary, H.M., Pereira, L.M., Finkelstein, Y., Nelson, C.A., Vogel-Farley, V., Degregorio, G., Holm, I.A., et al. (2014). Safety, pharmacokinetics, and preliminary assessment of efficacy of mecasermin (recombinant human IGF-1) for the treatment of Rett syndrome. Proc. Natl. Acad. Sci. USA *111*, 4596–4601.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Koopmans, F., Van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijsouw, T., Coba, M.P., Cornelisse, L.N., Farrell, R.J., Goldschmidt, H.L., Howrigan, D.P., et al. (2019). SynGO: an evidence-based, expert-curated knowledge base for the synapse. Neuron *103*, 217–234.e4.

Kriaucionis, S., Paterson, A., Curtis, J., Guy, J., Macleod, N., and Bird, A. (2006). Gene expression analysis exposes mitochondrial abnormalities in a mouse model of Rett syndrome. Mol. Cell Biol. *26*, 5033–5042.

Lee, C.E., Singleton, K.S., Wallin, M., and Faundez, V. (2020). Rare genetic diseases: nature's experiments on human development. iScience *23*, 101123.

Lehtinen, M.K., and Walsh, C.A. (2011). Neurogenesis at the brain-cerebrospinal fluid interface. Annu. Rev. Cell Dev. Biol. *27*, 653–679.

Li, Z., Adams, R.M., Chourey, K., Hurst, G.B., Hettich, R.L., and Pan, C. (2012). Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. J. Proteome Res. *11*, 1582–1590.

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.

Linton, M.F., Gish, R., Hubl, S.T., Bütler, E., Esquivel, C., Bry, W.I., Boyles, J.K., Wardell, M.R., and Young, S.G. (1991). Phenotypes of apolipoprotein B and apolipoprotein E after liver transplantation. J. Clin. Invest. *88*, 270–281.

Liu, Q., Trotter, J., Zhang, J., Peters, M.M., Cheng, H., Bao, J., Han, X., Weeber, E.J., and Bu, G. (2010). Neuronal LRP1 knockout in adult mice leads to impaired brain lipid metabolism and progressive, age-dependent synapse loss and neurodegeneration. J. Neurosci. *30*, 17068–17078.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lyst, M.J., and Bird, A. (2015). Rett syndrome: a complex disorder with simple roots. Nat. Rev. Genet. *16*, 261–275.

Maclean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and Maccoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966–968.

Matarazzo, V., and Ronnett, G.V. (2004). Temporal and regional differences in the olfactory proteome as a consequence of MeCP2 deficiency. Proc. Natl. Acad. Sci. USA *101*, 7763–7768.

Matsuishi, T., Urabe, F., Percy, A.K., Komori, H., Yamashita, Y., Schultz, R.S., Ohtani, Y., Kuriya, N., and Kato, H. (1994). Abnormal carbohydrate metabolism in cerebrospinal fluid in Rett syndrome. J. Child Neurol. *9*, 26–30.

Montojo, J., Zuberi, K., Rodriguez, H., Kazi, F., Wright, G., Donaldson, S.L., Morris, Q., and Bader, G.D. (2010). GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinformatics *26*, 2927–2928.

Mullin, A.P., Gokhale, A., Moreno-De-Luca, A., Sanyal, S., Waddington, J.L., and Faundez, V. (2013). Neurodevelopmental disorders: mechanisms and boundary definitions from genomes, interactomes and proteomes. Transl. Psychiatry *3*, e329.

Na, E.S., Nelson, E.D., Kavalali, E.T., and Monteggia, L.M. (2012). The impact of MeCP2 loss- or gain-of-function on synaptic plasticity. Neuropsychopharmacology *38*, 212–219.

O'connell, J.D., Paulo, J.A., O'brien, J.J., and Gygi, S.P. (2018). Proteome-wide evaluation of two common protein quantification methods. J. Proteome Res. *17*, 1934–1942.

O'leary, H.M., Kaufmann, W.E., Barnes, K.V., Rakesh, K., Kapur, K., Tarquinio, D.C., Cantwell, N.G., Roche, K.J., Rose, S.A., Walco, A.C., et al. (2018). Placebo-controlled crossover assessment of mecasermin for the treatment of Rett syndrome. Ann. Clin. Transl. Neurol. *5*, 323–332.

Olsson, B., Lautner, R., Andreasson, U., Öhrfelt, A., Portelius, E., Bjerke, M., Hölttä, M., Rosén, C., Olsson, C., Strobel, G., et al. (2016). CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. Lancet Neurol. *15*, 673–684.

Oztan, O., Garner, J.P., Constantino, J.N., and Parker, K.J. (2020). Neonatal CSF vasopressin concentration predicts later medical record diagnoses of autism spectrum disorder. Proc. Natl. Acad. Sci. USA *117*, 10609–10613.

Oztan, O., Talbot, C.F., Argilli, E., Maness, A.C., Simmons, S.M., Mohsin, N., Del Rosso, L.A., Garner, J.P., Sherr, E.H., Capitanio, J.P., and Parker, K.J. (2021). Autism-associated biomarkers: test-retest reliability and relationship to quantitative social trait variation in rhesus monkeys. Mol. Autism. *12*, 50.

Pacheco, N.L., Heaven, M.R., Holt, L.M., Crossman, D.K., Boggio, K.J., Shaffer, S.A., Flint, D.L., and Olsen, M.L. (2017). RNA sequencing and proteomics approaches reveal novel deficits in

the cortex of Mecp2-deficient mice, a model for Rett syndrome. Mol. Autism. *8*, 56.

Ping, L., Duong, D.M., Yin, L., Gearing, M., Lah, J.J., Levey, A.I., and Seyfried, N.T. (2018). Global quantitative analysis of the human brain proteome in Alzheimer's and Parkinson's Disease. Sci. Data *5*, 180036.

Ping, L., Kundinger, S.R., Duong, D.M., Yin, L., Gearing, M., Lah, J.J., Levey, A.I., and Seyfried, N.T. (2020). Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. Sci. Data *7*, 315.

Raman, A.T., Pohodich, A.E., Wan, Y.W., Yalamanchili, H.K., Lowry, W.E., Zoghbi, H.Y., and Liu, Z. (2018). Apparent bias toward long gene misregulation in MeCP2 syndromes disappears after controlling for baseline variations. Nat. Commun. *9*, 3225.

Rasala, B.A., Orjalo, A.V., Shen, Z., Briggs, S., and Forbes, D.J. (2006). ELYS is a dual nucleoporin/kinetochore protein required for nuclear pore assembly and proper cell division. Proc. Natl. Acad. Sci. USA *103*, 17801–17806.

Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T.J., Goodman, R.P., Grabarek, Z., Haas, M.E., Hung, W.H.W., et al. (2021). MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. Nucleic Acids Res. *49*, D1541–D1547.

Ribeiro, M.C., and Macdonald, J.L. (2020). Sex differences in Mecp2-mutant Rett syndrome model mice and the impact of cellular mosaicism in phenotype development. Brain Res. *1729*, 146644. https://doi.org/10.1016/j.brainres.2019.146644.

Riikonen, R., Makkonen, I., Vanhala, R., Turpeinen, U., Kuikka, J., and Kokki, H. (2006). Cerebrospinal fluid insulin-like growth factors IGF-1 and IGF-2 in infantile autism. Dev. Med. Child Neurol. *48*, 751–755.

Sandweiss, A.J., Brandt, V.L., and Zoghbi, H.Y. (2020). Advances in understanding of Rett syndrome and MECP2 duplication syndrome: prospects for future therapies. Lancet Neurol. *19*, 689–698.

Schaevitz, L.R., Gómez, N.B., Zhen, D.P., and Berger-Sweeney, J.E. (2013). MeCP2 R168X male and female mutant mice exhibit Rett-like behavioral deficits. Genes Brain Behav. *12*, 732–740.

Scholz, D., Pöltl, D., Genewsky, A., Weng, M., Waldmann, T., Schildknecht, S., and Leist, M. (2011). Rapid, complete and large-scale generation of post-mitotic neurons from the human LUHMES cell line. J. Neurochem. *119*, 957–971.

Shah, R.R., Cholewa-Waclaw, J., Davies, F.C.J., Paton, K.M., Chaligne, R., Heard, E., Abbott, C.M., and Bird, A.P. (2016). Efficient and versatile CRISPR engineering of human neurons in culture to model neurological disorders. Wellcome Open Res. *1*, 13.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shulyakova, N., Andreazza, A.C., Mills, L.R., and Eubanks, J.H. (2017). Mitochondrial dysfunction in the pathogenesis of Rett syndrome: implications for mitochondria-targeted therapies. Front. Cell. Neurosci. *11*, 58.

Stukas, S., Robert, J., Lee, M., Kulic, I., Carr, M., Tourigny, K., Fan, J., Namjoshi, D., Lemke, K., Devalle, N., et al. (2014). Intravenously injected human apolipoprotein A-I rapidly enters the central nervous system via the choroid plexus. J. Am. Heart Assoc. *3*, e001156.

Sztainberg, Y., and Zoghbi, H.Y. (2016). Lessons learned from studying syndromic autism spectrum disorders. Nat. Neurosci. *19*, 1408–1417.

Torun, F.M., Winter, S.V., Doll, S., Riese, F.M., Vorobyev, A., Mueller-Reif, J.B., Geyer, P.E., and Strauss, M.T. (2021). Transparent exploration of machine learning for biomarker discovery from proteomics and omics data. Preprint at bioRxiv. https://doi.org/10.1101/2021.03.05.434053.

Tropea, D., Giacometti, E., Wilson, N.R., Beard, C., Mccurry, C., Fu, D.D., Flannery, R., Jaenisch, R., and Sur, M. (2009). Partial reversal of Rett Syndrome-like symptoms in MeCP2 mutant mice. Proc. Natl. Acad. Sci. USA *106*, 2029–2034.

Tsujita, M., Vaisman, B., Chengyu, L., Vickers, K.C., Okuhira, K.I., Braesch-Andersen, S., and Remaley, A.T. (2021). Apolipoprotein A-I in mouse cerebrospinal fluid derives from the liver and intestine via plasma high-density lipoproteins assembled by ABCA1 and LCAT. FEBS Lett. *595*, 773–788.

Weaving, L.S., Christodoulou, J., Williamson, S.L., Friend, K.L., McKenzie, O.L.D., Archer, H., Evans, J., Clarke, A., Pelka, G.J., Tam, P.P.L., et al. (2004). Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. Am. J. Hum. Genet. *75*, 1079–1093.

Wellington, C.L., and Frikke-Schmidt, R. (2016). Relation between plasma and brain lipids. Curr. Opin. Lipidol. *27*, 225–232.

Wu, H., Tao, J., Chen, P.J., Shahab, A., Ge, W., Hart, R.P., Ruan, X., Ruan, Y., and Sun, Y.E. (2010). Genome-wide analysis reveals methyl-CpG-binding protein 2-dependent regulation of microRNAs in a mouse model of Rett syndrome. Proc. Natl. Acad. Sci. USA *107*, 18161–18166.

Wynne, M.E., Lane, A.R., Singleton, K.S., Zlatic, S.A., Gokhale, A., Werner, E., Duong, D., Kwong, J.Q., Crocker, A.J., and Faundez, V. (2021). Heterogeneous expression of nuclear encoded mitochondrial genes distinguishes inhibitory and excitatory neurons. eNeuro *8*. ENEURO.0232-21.2021.

Xu, H., Fame, R.M., Sadegh, C., Sutin, J., Naranjo, C., Della, S., Cui, J., Shipley, F.B., Vernon, A., Gao, F., et al. (2021). Choroid plexus NKCC1 mediates cerebrospinal fluid clearance during mouse early postnatal development. Nat. Commun. *12*, 447.

Yao, Z., Van Velthoven, C.T.J., Nguyen, T.N., Goldy, J., Sedeno-Cortes, A.E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., et al. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell *184*, 3222–3241.e26.

Zandl-Lang, M., Züllig, T., Trötzmüller, M., Naegelin, Y., Abela, L., Wilken, B., Scholl-Buergi, S., Karall, D., Kappos, L., Köfeler, H., and Plecko, B. (2022). Changes in the cerebrospinal fluid and plasma lipidome in patients with Rett syndrome. Metabolites *12*, 291.

Zerbi, V., Pagani, M., Markicevic, M., Matteoli, M., Pozzi, D., Fagiolini, M., Bozzi, Y., Galbusera, A., Scattoni, M.L., Provenzano, G., et al. (2021). Brain mapping across 16 autism mouse models reveals a spectrum of functional connectivity subtypes. Mol. Psychiatry *26*, 7610–7620.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. *34*, 11929–11947.

Zhang, Y., Sloan, S.A., Clarke, L.E., Caneda, C., Plaza, C.A., Blumenthal, P.D., Vogel, H., Steinberg, G.K., Edwards, M.S.B., Li, G., et al. (2016). Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. Neuron *89*, 37–53.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *10*, 1523.

Zimmerman, A.W., Jyonouchi, H., Comi, A.M., Connors, S.L., Milstien, S., Varsou, A., and Heyes, M.P. (2005). Cerebrospinal fluid and serum markers of inflammation in autism. Pediatr. Neurol. *33*, 195–201.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
MECP2 polyclonal antibody	Invitrogen	Cat# PA-1-888; RRID:AB_2281966
Beta-Actin monoclonal antibody, Clone AC-15	Sigma	Cat# A5441; RRID:AB_476744
Living Colors® dsRed Polyclonal Antibody	Takara Bio	Cat# 632,496; RRID:AB_10013483
Goat anti-Rabbit IgG (H + L) Cross-Adsorbed Secondary Antibody, HRP	ThermoFisher	Cat# G21234; RRID:AB_2536530
Goat anti-Mouse IgG, IgM, IgA (H + L) Secondary Antibody, HRP	ThermoFisher	Cat# A10668; RRID:AB_2534058
Goat anti-Rabbit IgG (H + L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 555	ThermoFisher	Cat# A21429; RRID:AB_2535850
Bacterial and virus strains		
pQS136-03(fMECP2-MECP2_P2A-nucTagRFP)	Q-State Biosciences	NA
pQS140-04(fhSyn-H2B(nls)-TagRFP)	Q-State Biosciences	NA
Biological samples		
Human Rett cases, CSF pre/post recombinant IGF1 treatment	Boston Children's Hospital; PMID: 24623853	NA
Chemicals, peptides, and recombinant proteins		
Myh9: AGVLAHLEEER	New England Peptide	NA
Myh9: IAQLEEQLDNETK	New England Peptide	NA
Pon1: IFFYDSENPPGSEVLR	New England Peptide	NA
Pon1: LLIGTVFHR	New England Peptide	NA
App: TEEISEVK	New England Peptide	NA
App: THTHIVIPYR	New England Peptide	NA
A2m: AIAYLNTGYQR	New England Peptide	NA
A2m: LPSDVVEESAR	New England Peptide	NA
Apoa1: DYVSQFESSTLGK	New England Peptide	NA
Apoa1: WNEEVEAYR	New England Peptide	NA
Apob: TEVIPPLIENR	New England Peptide	NA
Apob: GFEPTLEALFGK	New England Peptide	NA
C3: GLEVSITAR	New England Peptide	NA
C3: SSVAVPYVIVPLK	New England Peptide	NA
C9: SIEVFGQFQGK	New England Peptide	NA
C9: TTSFNANLALK	New England Peptide	NA
Thbs1: FVFGTTPEDILR	New England Peptide	NA
Thbs1: IENANLIPPVPDDK	New England Peptide	NA
Critical commercial assays		
TMT Labeling Reagent	ThermoFisher	Cat# A34808
TMTPro labeling Reagent	ThermoFisher	Cat# A44520
RNA 6000 Nano Kit	Agilent Technologies	Cat# 5067-1511

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
TMT mass spectrometry analysis of wild type and Mecp2null rat CSF of animals aged 23-26 days	This Paper	PXD029808 ProteomeXchange Consortium
TMT mass spectrometry analysis of wild type and Mecp2 null mouse CSF of animals aged 46 days	This Paper	PXD029809 ProteomeXchange Consortium
TMT mass spectrometry analysis of wild type and Cdkl5 null mouse CSF of animals aged 46 days	This Paper	PXD029811 ProteomeXchange Consortium
MOUSE CSF Parallel Reaction Monitoring	This Paper	PXD029835 ProteomeXchange Consortium
Convergent Cerebrospinal Fluid Proteomes and MetabolicOntologies in Humans and Animal Models of Rett Syndrome	This Paper	PXD030426 ProteomeXchange Consortium
Whole cortex & hippocampus - smart-seq (2019)	Allen Institute Mouse Whole Cortex and Hippocampus SMART-seq	RRID:SCR_019013; https://portal.brain-map.org/atlases-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-smart-seq
RNAseq signatures of cortex, hippocampus, striatum, liver, muscle and kidney from wild type and Mecp2 null male mice	PMID: 34312306	GEO: GSE140054
Experimental models: Cell lines		
LUHMES	ATCC	Cat# CRL-2927; RRID:CVCL_B056
LUHMES MECP2 KO clone 2_7	PMID: 27976757	NA
Experimental models: Organisms/strains		
WildType Sprague-Dawley Rat	Charles River Laboratory	Cat # CRL:SD; RRID:RGD_737891
SD-Mecp2^{em1Sage}Sprague-Dawley Rat	Envigo	Strain: HsdSage:SD-Mecp2^{em1Sage}; RRID:RGD_11567272
C57BL/6J Mouse	The Jackson Laboratory	Cat# 000664; RRID:IMSR_JAX:000664
B$_6$.129P2(C)-Mecp2$^{tm1.1Bird}$/J Mouse	The Jackson Laboratory	Cat# 003890; RRID:IMSR_JAX:003890
B$_6$.129(FVB)-Cdkl5$^{tm1.1Joez}$/J Mouse	The Jackson Laboratory	Cat# 021967; RRID:IMSR_JAX:021967
Software and algorithms		
Seahorse Wave v2.6.1	Agilent Technologies	RRID:SCR_014526; https://www.agilent.com/en/product/software-informatics/cell-analysis-software
Proteome Discoverer suite (version 2.4.1)	ThermoFisher	RRID:SCR_014477; https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/proteome-discoverer-software.html
Galaxy v.21.01	PMID:20738864, PMID:20069535, PMID:16169926	RRID:SCR_006281; http://galaxyproject.org/
iGenome	Illumina	https://ewels.github.io/AWS-iGenomes/
STAR	PMID:23104886	RRID:SCR_004463; http://code.google.com/p/rna-star/
Delimit Pro for Windows 10/8.1/7	Delimit	http://delimitware.com/download.html
Skyline v4.2	MacLean et al. (2010); PMID:20147306	RRID:SCR_014080; https://skyline.ms/project/home/software/Skyline/begin.view

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MaxQuant v1.6.2.3	PMID:19029910, PMID:25059623	RRID:SCR_014485; https://www.maxquant.org/maxquant/
Perseus v1.5.5.3	PMID:27348712	RRID:SCR_015753; https://www.maxquant.org/perseus/
Qlucore Omics Explorer Version 3.6(33)	Qlucore	https://qlucore.com/omics-explorer
OmicLearn (v1.0.0)	Torun et al. (2021); bioRxiv	https://github.com/OmicEra/OmicLearn
Percolator (v2.8.1)	PMID:17952086	RRID:SCR_005040; http://percolator.ms/
Pandas package (1.0.1)	Pandas	RRID:SCR_018214; https://pandas.pydata.org/
NumPy package (1.18.1)	PMID: 32939066	RRID:SCR_008633; https://numpy.org/
scikit-learn package (0.22.1)	Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011. No PMID	RRID:SCR_002577; http://scikit-learn.org/
Plotly (4.9.0)	Plotly	RRID:SCR_013991; https://plotly.com/
OmicLearn	https://doi.org/10.1101/2021.03.05.434053	https://omicera-omiclearn-omic-learn-7nhy3b.streamlitapp.com/
Cluego v2.58	PMID:19237447	RRID:SCR_005748; https://apps.cytoscape.org/apps/cluego
Cytoscape v3.8.2	PMID:21149340, PMID:14597658	RRID:SCR_003032; https://cytoscape.org/
Metascape	PMID: 30944313	RRID:SCR_016620; https://metascape.org/gp/index.html#/main/step1
Estimation Stats	PMID:31217592	RRID:SCR_018321; https://www.estimationstats.com/#/
Prism v9.2.0(283)	GraphPad	RRID:SCR_002798; https://www.graphpad.com/scientific-software/prism/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Victor Faundez (vfaunde@emory.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Deutsch et al., 2020) partner repository with dataset identifiers: PXD029808, PXD029809, PXD029811, PXD029835, and PXD030426. Dataset Identifiers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contacton request.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Human samples

Clinical features of the cohort used in these studies are described by Khwaja et al. (2014). The referred to study was approved by the Institutional Review Board of Boston Children's Hospital and informed consent was obtained from the parent of each participant. CSF samples were received and remained deidentified for these studies.

Animal models

All rat experiments were carried out in accordance with the European Communities Council Directive (86/609/EEC) and with the terms of a project license under the UK Scientific Procedures Act (1986). The *Mecp2*$^{-/y}$ rats were maintained by crossing *Mecp2*$^{+/-}$ females (SD-*Mecp2*^{em1Sage}) with wild type Sprague Dawley males. Animals were maintained on 12-h light/dark cycles with free access to normal rat food and water. WT and *Mecp2*$^{-/y}$ rats at 25 days of postnatal age were weighed and assessed for the development of the RTT-like phenotypes prior to surgery.

Mouse animal husbandry and euthanasia was carried out as approved by the Emory University Institutional Animal Care and Use Committees. Male and female mice of C57BL/6J, *Mecp2* deficient (*Mecp2*$^{tm1.1Bird}$), and *Cdkl5*-deficient (*Cdkl5*$^{tm1.1Joez}$) were obtained from the The Jackson Laboratory stocks (The Jackson Laboratory #000664, #003890 and #021967, respectively). All animals were of 6 weeks of age. Animals were maintained on 12-h light/dark cycles with free access to mouse chow and water.

Cell lines

Female LUHMES wild-type control clone, and *MECP2* knock-out clone 2_7 cell lines were cultured and differentiated on Nunclon flasks and plates treated with a 44 μg/mL Poly-L-Ornithine (Sigma P3655) and 1 μg/mL fibronectin (Sigma F1141) solution overnight in a 37°C incubator. LUHMES cells were differentiated as follows: three million cells were plated in a T75 flask with proliferation media (Advanced DMEM/F12 (Gibco 12634-010) with N2 (Gibco 17502048), 2 mM L-glutamine (Sigma G7513), and 40 ng/mL beta-FGF (R&D Systems 4114-TC-01M). After 24 h, media was changed to differentiation media (Advanced DMEM/F12 with N2, 2 mM L-glutamine, 1 mM DbcAMP (Sigma D0627), 1 μg/mL tetracycline (Sigma T7660), and 2 ng/mL GDNF (R&D Systems 212-GD-050) for a pre-differentiation phase of two days. Pre-differentiated cells were lifted with trypsin method. Trypsin activity was blocked with aprotinin after lifting the cells.

METHOD DETAILS

Rat CSF sample collection

Rats were anesthetized using intraperitoneal administration of an injectable cocktail of medetomidine (0.5 mg/kg) and ketamine (75 mg/kg). Once the animal was deeply anesthetized, as indicated by the absence of withdrawal reflexes (tail and limbs) and the eye positioning reflex, the surgical area was shaved and the animal was secured in the stereotaxic frame with the head tilted at roughly 45°. The surgical area was then cleaned with Hibiscrub and a surgical drape was placed around the operating area with a hole to expose only the surgical area. A skin incision along the midline of the skull extending from between the eyes to 3-4 cm caudally to make sure the back of the neck is fully exposed. The fascia and the superficial and deep layers of the neck muscles were then dissected to expose the membrane of the dura mater at the atlanto-occipital joint between the occipital condyles and the rostral facets of atlas. The cisterna magna was then carefully pierced by a pulled glass pipette (1 cm long) connected to a 2.5 mL syringe through 30 cm of PE-50 tubing. A small volume of CSF entered the glass pipette through the capillary action and the flow was maintained by gently pulling the plunger. The CSF was collected into cryoprotective tubes and snap-frozen immediately in liquid nitrogen. Animals were then given a lethal dose of anesthesia, decapitated, and the brain was exposed and the areas of interest were dissected and snap-frozen in liquid nitrogen.

Mouse CSF collection

Our terminal CSF collection method was adapted from a previously published protocol (Boire et al., 2017). Mice were deeply anesthetized by intraperitoneal injection of a mixture of ketamine (73.5 mg/kg; Akron, USA), xylazine (9.2 mg/kg; Bayer Pharma, Germany), and acepromazine maleate (2.75 mg/kg; Boehringer Ingelheim, USA) in 0.9% (v/w) NaCl. The back of the neck overlying the occiput were first shaved then cleaned and disinfected with 70% ethanol. Using the thumb and index finger, the mouse was placed prone with the neck in flexion on a 15 mL conical tube at an approximately 45-degree angle to access the cisterna magna using landmarks between occipital protuberances and the spine of the atlas. A Hamilton syringe containing 30 G needle was inserted through the skin at a 45-degree angle with the horizontal, to reach a depth of approximately 4 mm into the cisterna magna for CSF collection without need for an incision. The syringe was kept stable without any lateral movement and 4–12 μL of clear CSF was drawn into the syringe by slow and smooth aspiration. The CSF was immediately spun down for 30 s and clear CSF was

inspected with the naked eye and frozen immediately on dry ice. Frankly blood contaminated samples discarded.

LUHMES conditioned media preparation

LUHMES wild-type control, and MECP2 knock-out clone 2_7 cell lines were differentiated and conditioned media was collected. To reduce background signal in mass spectrometry, the last phase of differentiation utilized high purity and BSA-free components including high purity N2 components. A 100x high purity N2 solution was made with 10 mg/mL human holo-transferrin (Sigma T4132), 0.5mg/mL human recombinant insulin solution (Sigma I9278), 0.63 μg/mL progesterone (Sigma P6149), 1.61 mg/mL putrescine dihydrochloride (Sigma P5780), 0.52 μg/mL sodium selenite (Sigma S5261) and DMEM/F12 (Thermo Fisher 21,331,020). One million pre-differentiated cells were plated to each well of a Nunclon 6-well dish with 2 mL of the high purity differentiation media: DMEM/F12 (Thermo Fisher 21331020) containing high purity N2 (above), 2 mM L-glutamine (Sigma G7513), 1 mM DbcAMP (D0627), 1 μg/mL tetracycline (T7660), and 2 ng/mL GDNF (R&D Systems 212-GD-050). Cells conditioned the media for 3 days at 5% CO2 in a 37°C incubator. On the third day, the conditioned media was collected and Complete antiprotease (Roche 11697498001) was added. Cellular debris was pelleted at 16,000 x g in an Eppendorf microcentrifuge at 4°C for 20 min. The supernatant was collected and flash frozen on dry ice. A trichloroacetic acid (TCA) precipitation was done on 750 μL of the conditioned cell media by adding 9.8 μg sodium deoxycholate per 100 μL of conditioned media followed by trichloroacetic acid to 10%. The solution was incubated on ice for 20 min to precipitate out proteins. The solution was centrifuged at 16,000 x g for 15 min at 4°C. TCA supernatant was aspirated out and the pellet was washed in an equal volume of ice-cold acetone and vortexed. Precipitate was repelleted by centrifugation at 16,000 x g at 4°C for 10 min. Acetone was aspirated and the pellet was lightly air-dried, dissolved in 200 μL of 8 M Urea, and flash frozen on dry ice.

LUHMES cell viral transduction

LUHMES cells were cultured as above and infected with 0–16.5 μl/mL titer with pQS136-03(fMECP2-MECP2_P2A-nucTagRFP) and pQS140-04(fhSyn-H2B(nls)-TagRFP). Virus was added during media change to pre-differentiation media. Cells were incubated in virus containing media for 48 (immunofluorescence and flux oximetry) or 72 h (lysis) and then plated onto coated coverslips for immunofluorescence, to Seahorse Flux Oximetry 96 cellular well plates for differentiation, or to culture flasks for lysis.

Lysis

For cellular lysis, cells were washed in PBS containing MgCl2 1 mM, CaCl2 0.1 mM and lysed in 8 M Urea containing complete antiprotease. Lysis was incubated for 30 min and then sonicated with 10 quick bursts to shear DNA. Following Bradford Assay for protein concentration, lysis was prepped for running an SDS-Page electrophoresis gel and transferred to PVDF membrane. Membranes were blocked in 5% non-fat milk in TBS containing 0.05% Triton X-100. Primary antibodies were incubated at 4°C overnight (Anti-MECP2, Invitrogen PA-1-888, 1:500 rabbit polyclonal and anti-beta actin Sigma A5451 1:1000 mouse monoclonal). Secondary antibodies were Thermo HRP conjugated anti-rabbit and anti-mouse (Thermo G21234 and A10668).

Immunofluorescence

For immunofluorescence, cells were washed in PBS containing MgCl2 1 mM, CaCl2 0.1 mM, fixed in 4% paraformaldehyde for 20 min on ice, and rinsed again in PBS containing MgCl2 1 mM, CaCl2 0.1 mM. Cells were permeabilized in 0.2% Triton X-100 in PBS for 5 min and blocked in 2% BSA, 1% fish gelatin, and 15% horse serum all in PBS for 30 min at room temperature. After blocking, coverslips were incubated for 30 min at 37°C in primary antibody diluted in block solution (anti-dsRed, 1:500, Clontech catalog #632496), washed 3 times quickly in blocking solution and incubated for 30 min at 37°C with secondary antibody diluted in block solution (anti-rabbit AlexaFluor 555, 1:1000, Thermo A21429). Coverslips were washed 3 times in blocking solution, and once in PBS containing MgCl 1 mM, CaCl 0.1 mM followed by mounting with DAPI Fluoromount-G (Southern Biotech 0100-20). Images were captured on a widefield fluorescence laser scan confocal Ti2 microscope with Nikon A1 LFOV camera, objective Apo 60× oil λS DIC N2/NA 1.4. with Galvano scanner. DAPI acquisition was with PMT detector (EmW: 450.0, ExW: 405.0) and Alexa Fluor 555 was acquired with GaAsP detection (EmW:595.0, ExW: 561.0).

Seahorse flux oximetry

For Flux Oximetry, assays were performed using manufacturer's guidelines on the Seahorse XFe96 analyzer. Each well contained approximately 20,000 differentiated LUHMES cells. Assay media consisted of DMEM based assay media supplemented with 2 mM L-glutamine, 1 mM sodium pyruvate, 10 mM glucose. Stress test drug injection strategies resulted in final well concentrations of 1mM oligomycin, 0.25 μM FCCP, and 0.5 μM each of rotenone and antimycin A. Assay oxygen and pH recording parameters were based on the preset Agilent Seahorse Stress Test protocol in Wave 2.6.1.

Mass spectrometry Emory

Sample processing

All CSF (5 μL) samples were diluted with 50 μL of 50 mM NH4HCO3 and treated with TCEP and CAA and heated at 90°C for 10 min. The samples were digested with 1:20 (w/w) lysyl endopeptidase (Wako) at 25°C overnight. Further overnight digestion was carried out with 1:20 (w/w) trypsin (Promega) at 25°C. Resulting peptides were desalted with an HLB microelution plate (Waters, Cat# 186001828BA) and dried under vacuum.

Tandem mass tag (TMT) labeling

For each sample, labeling was performed as previously described (Higginbotham et al., 2020; Ping et al., 2018). Briefly, each was re-suspended in 100 mM TEAB buffer (100 μL). The TMT and TMTPro labeling reagents (ThermoFisher Cat# A34808, A44520) were equilibrated to room temperature, and anhydrous ACN (256μL) was added to each reagent channel. Each channel was gently vortexed for 5 min, and then 41 μL from each TMT channel was transferred to the peptide solutions and allowed to incubate for 1h at room temperature. The reaction was quenched with 5% (v/v) hydroxylamine (8 μL) (Pierce). All channels were then combined and dried by SpeedVac (LabConco) to approximately 150 μL and diluted with 1 mL of 0.1% (v/v) TFA, then acidified to a final concentration of 1% (v/v) FA and 0.1% (v/v) TFA. Peptides were desalted with a 30 mg C18 Sep-Pak column (Waters, Cat# 186008055). Each Sep-Pak column was activated with 1mL of methanol, washed with 1 mL of 50% (v/v) ACN, and equilibrated with 2 × 1 mL of 0.1% TFA. The samples were then loaded and each column was washed with 2 × 1 mL 0.1% (v/v) TFA, followed by 1 mL of 1% (v/v) FA. Elution was performed with 2 volumes of 0.5 mL 50% (v/v) ACN. The eluates were then dried to completeness.

High pH fractionation

High pH fractionation was performed essentially as described (Ping et al., 2020; Gokhale et al., 2020) with slight modification. Dried samples were re-suspended in high pH loading buffer (0.07% v/v NH4OH, 0.045% v/v FA, 2% v/v ACN) and loaded onto BEH C18 column (2.1× 150 mm with 1.7 μm beads) (Waters, Cat# 186002353). A Thermo Vanquish system was used to carry out the fractionation. Solvent A consisted of 0.0175% (v/v) NH4OH, 0.01125% (v/v) FA, and 2% (v/v) ACN; solvent B consisted of 0.0175% (v/v) NH4OH, 0.01125% (v/v) FA, and 90% (v/v) ACN. The sample elution was performed over a 22 min gradient with a flow rate of 0.6 mL/min from 0 to 50% solvent B. A total of 96 individual equal volume fractions were collected across the gradient and subsequently pooled by concatenation into 48 fractions for the *Mecp2* null TMT batches. For the *Cdlk5* null batch, 192 fractions were collected and combined into 96 fractions. All fractions were dried to completeness using a vacuum centrifugation.

Liquid chromatography tandem mass spec for TMT

Each of the peptide fractions was re-suspended in loading buffer (0.1% FA, 0.03% TFA, 1% ACN). Peptide eluents were either separated on a self-packed C18 (1.9 μm) (Dr. Maisch, Germany, Cat# ReproSil-Pur: 120C18-AQ 1.9 um 1 g), fused silica column (15 cm × 100 μM internal diameter) (New Objective, Cat# FS360-100-15-N-5-C25) or a 1.7 μm CSH C18 column (15 cm × 150 μM internal diameter) (Waters, Cat# 186008814). An Easy nLC 1200 (Thermo Fisher Scientific) or Ultimate U300 RSLCnano (Thermo Scientific) was used to elute the peptide ion. Mass spectra were collected either on a Fusion Lumos or Fusion Eclipse mass spectrometer. Both mass spectrometers were outfitted with the FAIMS Pro ion mobility source.

Liquid chromatography tandem mass spec for PRM

AQUA standard peptides (Thermo Fisher Scientific) were spiked into digested mouse CSF samples. For each sample and equivalent of 1 μl of CSF was loaded onto a Water's 1.7 μm CSH C18 column

(15 cm × 150 μM internal diameter). Peptides were eluted using an Ultimate 3000 RSLCnano and PRM spectra were collected using an Orbitrap HFX mass spectrometer.

Data processing protocol

All TMT raw files were searched using Thermo's Proteome Discoverer suite (version 2.4.1) with Sequest HT. The spectra were searched against rat or mouse uniprot database. Search parameters included 20 ppm precursor mass window, 0.05 Da product mass window, dynamic modifications methione (+15.995 Da), deamidated asparagine and glutamine (+0.984 Da), phosphorylated serine, threonine and tyrosine (+79.966 Da), and static modifications for carbamidomethyl cysteines (+57.021 Da) and N-terminal and Lysine-tagged TMT (+229.26340 Da or +304.207 Da). Percolator was used filter PSMs to 0.1%. Peptides were grouped using strict parsimony and only razor and unique peptides were used for protein level quantitation. Reporter ions were quantified from MS2 scans using an integration tolerance of 20 ppm with the most confident centroid setting. Only unique and razor (i.e., parsimonious) peptides were considered for quantification. PRM spectra were processed using the Skyline quantitation suite (MacLean et al., 2010).

Mass spectrometry Ann Arbor

Sample preparation

For tissues, samples were washed twice in 1X PBS and lysed in 8 M urea, 50 mM Tris HCl, pH 8.0, 1X Roche Complete Protease Inhibitor and 1X Roche PhosStop. Other samples were processed directly for protein quantification using Qubit fluorometry following by digestion overnight with trypsin. Briefly, samples were reduced for 1 h at RT in 12 mM DTT followed by alkylation for 1 h at RT in 15 mM iodoacetamide. Trypsin was added to an enzyme:substrate ratio of 1:20. Each sample was acidified in formic acid and subjected to SPE on an Empore SD C18 plate. For TMT labeling, after trypsin digestion ach sample was acidified in formic acid and subjected to SPE on an Empore SD C18 plate (3M catalog# 6015 SD). Each sample was lyophilized and reconstituted in 140 mM HEPES, pH 8.0, 30% acetonitrile.

Label-free quantification mass spectrometry

A 2 μg aliquot was analyzed by nano LC/MS/MS with a Waters NanoAcquity HPLC system interfaced to a Thermo Fisher Fusion Lumos. Peptides were loaded on a trapping column and eluted over a 75 μm analytical column at 350 nL/min; both columns were packed with Luna C18 resin (Phenomenex). A 4-h gradient was employed. The mass spectrometer was operated in data-dependent mode, with MS and MS/MS performed in the Orbitrap at 60,000 FWHM resolution and 15,000 FWHM resolution, respectively. APD was turned on. The instrument was run with a 3-s cycle for MS and MS/MS. The acquisition order was randomized. Data Processing Data were processed through the MaxQuant software v1.6.2.3 (www.maxquant.org). Data were searched using Andromeda with the following parameters: Enzyme: Trypsin, Database: Uniprot Rat, Fixed modification: Carbamidomethyl (C), Variable modifications: Oxidation (M), Acetyl (Protein N-term), Fragment Mass Tolerance: 20 ppm Pertinent. MaxQuant settings were: Peptide FDR 0.01 Protein FDR 0.01 Min. peptide Length 7 Min. razor + unique peptides 1 Min. unique peptides 0 Min. ratio count for LFQ 1 Second Peptides^ TRUE Match Between Runs* TRUE.

TMT quantification mass spectrometry

40 μL of acetonitrile was added to each TMT tag tube and mixed aggressively. Tags were incubated at RT for 15 min 30 μL of label was added to each peptide sample and mixed aggressively. Samples were incubated in an Eppendorf Thermomixer at 300 rpm 25°C for 1.5h. Reactions were terminated with the addition of 8 μL of fresh 5% hydroxylamine solution and 15-min incubation at room temperature. Samples were subjected to high pH reverse phase fractionation as follows; Buffers: Buffer A: 10 mM NaOH, pH 10.5, in water Buffer B: 10 mM NaOH, pH 10.5, in acetonitrile. We used XBridge C18 columns, 2.1 mm ID x 150 mm length, 3.5 μm particle size (Waters, part #186003023) attached to a Agilent 1100 HPLC system equipped with a 150 μL sample loop operating at 0.3 mL/min, detector set at 214 nm wavelength. Dried peptides were resolubilized in 150 μL of Buffer A and injected manually. Fractions were collected every 30 s from 1 to 49 min (96 fractions total, 150μL/fraction). We analyzed by mass spectrometry 10% per pool for the full proteome in a nano LC/MS/MS with a Waters NanoAcquity HPLC system interfaced to a ThermoFisher Fusion Lumos mass spectrometer. Peptides were loaded on a trapping column and eluted over a 75 μm analytical column at 350 nL/min; both columns were packed with Luna C18 resin (Phenomenex). Each high pH RP fraction was separated over a 2-h gradient (24 h instrument time total). The mass spectrometer was operated in data-dependent mode, with MS and MS/MS performed in the Orbitrap at 60,000 FWHM resolution and

50,000 FWHM resolution, respectively. A 3-s cycle time was employed for all steps. Data Processing Data were processed through the MaxQuant software v1.6.2.3 (www.maxquant.org). Data were searched using Andromeda with the following parameters: Enzyme: Trypsin Database: Uniprot Rat, Fixed modification: Carbamidomethyl (C) Variable modifications: Oxidation (M), Acetyl (Protein N-term), Phopho (STY; PO4 data only). Fragment Mass Tolerance: 20 ppm. Pertinent MaxQuant settings were: Peptide FDR 0.01 Protein FDR 0.01 Min. peptide Length 7 Min. razor + unique peptides 1 Min. unique peptides 0 Second Peptides FALSE Match Between Runs FALSE The protein Groups.txt files were uploaded to Perseus v1.5.5.3 for data processing and analysis.

AQUA mass spectrometry

Synthetic peptides labeled with Arginine (13C6,15N4) or Lysine (13C6,15N2) at >95% purity were made by New England Peptide MA 01440 USA. The following peptides were used: Myh9 AGVLAHLEEER; IAQLEEQLDNETK. Pon1 IFFYDSENPPGSEVLR; LLIGTVFHR. App TEEISEVK; THTHIVIPYR. A2m AIAYLNTG YQR; LPSDVVEESAR. Apoa1 DYVSQFESSTLGK; WNEEVEAYR. Apob TEVIPPLIENR; GFEPTLEALFGK. C3 GLEVSITAR; SSVAVPYVIVPLK. C9 SIEVFGQFQGK; TTSFNANLALK. Thbs1 FVFGTTPEDILR; IENANLIP PVPDDK. A 3-4 μg aliquot of each CSF tryptic peptide digests was spiked with isotopologe peptides at a concentration of 100 or 133 fmol/μg peptide digest. Peptides mixes were analyzed in analytical duplicate by nano LC/PRM using a Waters NanoAcquity HPLC system interfaced to a Thermo Fisher Fusion Lumos mass spectrometer. 1.5 μg per sample was loaded on a trapping column and eluted over a 75 μm analytical column at 350 nL/min; both columns were packed with Luna C18 resin (Phenomenex). A 1-h gradient was employed. The mass spectrometer was operated in PRM mode without scheduling; instrument settings included 15,000 FWHM resolution, NCE 30, AGC target value 5×10^4, and maximum IT of 22 ms. Data were processed using Skyline v4.2.

Data processing

Proteomics data were log2 converted. Data analysis was performed with two methods. We used Qlucore Omics Explorer Version 3.6(33) normalizing data to a mean of 0 and a variance of 1. No filtering by standard deviation was applied. All data were thresholded by a log2 fold of change of 0.5 and a non-corrected p value of 0.05.

A second method used was OmicLearn (v1.0.0) for performing the data analysis, model execution, and generating the plots and charts (Torun et al., 2021). Machine learning was done in Python (3.8.8). Feature tables were imported via the Pandas package (1.0.1) and manipulated using the Numpy package (1.18.1). The machine learning pipeline was employed using the scikit-learn package (0.22.1). For generating the plots and charts, Plotly (4.9.0) library was used. No normalization on the data was performed. To impute missing values, a mean-imputation strategy is used. Features were selected using a Extra-Trees (n_trees = 100) strategy with the maximum number of 20 features. Normalization and feature selection was individually performed using the training data of each split. For classification, we used either a XGBoost-Classifier (random_state = 23 learning_rate = 0.3 min_split_loss = 0 max_depth = 6 min_child_weight = 1), a AdaBoost-Classifier (random_state = 23 n_estimators = 100 learning_rate = 1.0), or RandomForest-Classifier (random_state = 23 n_estimators = 100 criterion = gini max_features = auto). Clasifiers were chosen based on the proximity of ROC curves to a value of 1. When using (RepeatedStratifiedKFold) a repeated (n_repeats = 10), stratified cross-validation (n_splits = 5) approach to classify datasets based on their genotype.

RNAseq and single cell RNAseq

These procedures were reported before (Wynne et al., 2021). RNA, library construction, and sequencing were performed by BGI. Total RNA was isolated with Trizol and quality control was done with the Agilent 2100 Bio analyzer (Agilent RNA 6000 Nano Kit, Cat# 5067-1511) to do total RNA sample QC: RNA concentration, RIN value 28S/18S and the fragment length distribution.

For library generation, poly-A containing mRNAs were isolated by poly-T oligo-attached magnetic beads. mRNA was fragmented into pieces using divalent cations under elevated temperature. RNA fragments were copied into first strand cDNA with reverse transcriptase plus random primers. Second strand cDNA synthesis was done using DNA Polymerase I and RNase H. cDNA fragments underwent addition of a single 'A' base followed by ligation of the adapter. The products were isolated and enriched by PCR amplification. We quantified PCR yield by Qubit and pooled samples together to make a single-strand DNA circle (ssDNA

circle), which gave the final library. DNA nanoballs were generated with the ssDNA circle by rolling circle replication to enlarge fluorescent signals at the sequencing. DNA nanoballs were loaded into the patterned nanoarrays and pair-end reads of 100 bp were read through on the BGISEQ-500 platform for data analysis. The BGISEQ-500 platform combined the DNA nanoball-based nanoarrays and stepwise sequencing using Combinational Probe-Anchor Synthesis Sequencing Method. We generated about 5.64 Gb bases per sample on average. Average mapping ratio with reference genome was 93.47%, average mapping ratio with gene was 67.04%; 19,972 genes were identified in which 2,659 of them are novel genes. 29,781 transcripts were identified.

The sequencing reads were uploaded to the Galaxy platform, and we used the usegalaxy.org server for analysis (Afgan et al., 2018). FastQC was performed to remove samples of poor quality (Andrews, 2012). All mapping was executed with the Galaxy server (v. 21.01) running Hisat2 (Galaxy Version 2.1.0 + galaxy7), FeatureCounts (Galaxy Version 2.0.1), and Deseq2 (Galaxy Version 2.11.40.6 + galaxy1) (Kim et al., 2015; Liao et al., 2014; Love et al., 2014). We employed the GRCm38 build of the reference sequence and GTF files (Ensembl) were used, iGenome (Illumina). Hisat2 was run under the following conditions: Paired-end, unstranded, default settings were used except for a GTF file was used for transcript assembly. The aligned SAM/BAM files were processed using Featurecounts (Default settings except used Ensembl GRCm38 GTF file and output for DESeq2 and gene length file). FeatureCounts files and raw files are available at GEO with accession GSE140054. The FeatureCounts compiled file is GSE140054_AllTissueFeatureCounts.txt.gz. Gene counts were normalized using DESeq2 (Love et al., 2014) followed by a regularized log transformation. Differential Expression was determined by DESeq2 using settings: Factors were tissue type, pairwise comparisons across tissues was done, output all normalized tables, size estimation was the standard median ratio, fit type was parametric, outliers were filtered using a Cook's distance cutoff. See Table S6.

Single cell RNA seq data are from Yao et al. (2021). Gene expression data matrix (matrix.csv) and cell metadata (metadata.csv) were obtained at Whole cortex & hippocampus - smart-seq (2019) with 10x-smart-seq taxonomy (2020). Data were downloaded from the Allen Institute Portal. Dataset contains RNA sequencing data of single cells from >20 areas of mouse cortex and hippocampus. The dataset includes 76,307 single cells. The sequencing results were aligned to exons and introns in the GRCm38.p3 reference genome using the STAR algorithm, and aggregated intron and exon counts at the gene level were calculated.

Matrix files were processed with Delimit Pro for Windows 10/8.1/7. Data were exported as tab delimited text file and analyzed with the Qlucore Omics Explorer Version 3.6(33). Data were log2 converted and normalized to a mean of 0 and a variance of 1. 2D t-SNE plots were generated using a perplexity of 40 and default settings.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatic analyses

Gene ontology analyses were performed with Cluego and HumanBase (Greene et al., 2015). ClueGo v2.58 run on Cytoscape v3.8.2 (Bindea et al., 2009; Shannon et al., 2003). ClueGo was run querying GO CC, REACTOME, KEGG and WikiPathways considering all evidence at a Medium Level of Network Specificity and selecting pathways with a Bonferroni corrected p value <10E-3. ClueGo was run with Go Term Fusion. HumanBase was run using default web-based parameters (Greene et al., 2015). *In silico* interactome data were downloaded from Genemania predicted and physical interactions and processed in Cytoscape v3.8.2 (Shannon et al., 2003). Interactome connectivity graph parameters were generated in Cytoscape. To compare ontologies between *Mecp2* and *Cdkl5* mutant males we used the online tool Metascape using multiple gene list and express analysis settings (Zhou et al., 2019).

Statistical analyses

Volcano plot p values were calculated using Qlucore Omics Explorer Version 3.6(33) without multiple corrections. Experiments in Figure 4D and 5A, 5B, 5D and 5F statistical analyses were performed with the engine https://www.estimationstats.com/#/ with a two-sided permutation t-test and alpha of 0.05 (Ho et al., 2019). ROC analysis and paired t-test were performed with Prism v9.2.0(283).