

Convergent Cerebrospinal Fluid Proteomes and Metabolic Ontologies in Humans and Animal Models of Rett Syndrome.

Stephanie A. Zlatic^{*1#}, Duc Duong^{*2}, Kamal K.E. Gadalla³, Brenda Murage³, Lingyan Ping², Ruth Shah⁶, Omar Khwaja⁴, Lindsay C. Swanson⁴, Mustafa Sahin⁴, Sruti Rayaprolu⁵, Prateek Kumar⁵, Srikant Rangaraju⁵, Adrian Bird⁶, Daniel Tarquinio⁷, Stuart Cobb³, Victor Faundez^{1#}

Departments of Cell Biology¹, Biochemistry², and Neurology⁵, Emory University, Atlanta, GA, USA, 30322. Simons Initiative for the Developing Brain, Centre for Discovery Brain Sciences, University of Edinburgh, Edinburgh, EH8 9XD, UK³. Department of Neurology, Rosamund Stone Zander Translational Neuroscience Center, Boston Children's Hospital, Boston, MA, 02115, USA⁴. The Wellcome Centre for Cell Biology, University of Edinburgh, Michael Swann Building, King's Buildings, Max Born Crescent, Edinburgh EH9 3BF, UK⁶. Center for Rare Neurological Diseases. Norcross, GA 30093, USA⁷.

#Address Correspondence to

vfaunde@emory.edu

szlatic@gmail.com

* These authors contributed equally

Acknowledgements

VF was funded by the Rett Syndrome Research Trust, the Loulou Foundation, and 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

Conflict 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.

Abstract

MECP2 loss-of-function mutations cause Rett syndrome, a disorder that results from a disrupted brain transcriptome. How these transcriptional defects are decoded into a disease proteome remains unknown. We studied the proteome in Rett syndrome cerebrospinal fluid (CSF) across vertebrates. We identified a consensus proteome and ontological categories shared across Rett syndrome cerebrospinal fluid (CSF) from three species, including humans. Rett CSF proteomes enriched proteins annotated to HDL lipoproteins, complement, mitochondria, citrate/pyruvate metabolism, as well as synapse compartments. We used these prioritized and shared ontologies to select analytes for orthogonal quantification. These studies independently validated our proteome and ontologies. Ontologically selected CSF hits had genotypic discriminatory capacity as determined by Receiver Operating Characteristic (ROC) analysis and distinguished Rett from a related neurodevelopmental disorder, CDKL5 deficiency disorder. We propose that *Mecp2* mutant CSF proteomes and ontologies inform novel putative mechanisms and biomarkers of disease. We suggest that Rett syndrome is a metabolic disorder impacting synapse function.

Introduction

The cellular and molecular understanding of neurodevelopmental disorders has been greatly advanced by the study of single gene defects ^{1,2}. 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 ³, affecting the expression of a vast number of RNAs in the brain ³⁻⁶. 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 ⁷⁻¹¹. 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 ¹². Moreover, known Rett syndrome phenotypes such as synaptic, circuit, and behavioral alterations ultimately originate in alterations of protein expression and function ^{4,13,14}. 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 ^{15,16}.

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 ¹⁷⁻¹⁹. CSF carries neurodevelopmental instructive signals and nutrients and accrues secretions and metabolites that reflect functional states of diverse cell types in brain parenchyma and the choroid ²⁰. For example, composition of the CSF proteome in individuals with Alzheimer's predicts key diagnostic molecular pathology in the Alzheimer's brain ²¹⁻²³. 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 ²⁴. 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 ²⁵⁻³⁰. 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 ³¹⁻³³. 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 ³⁴. We collectively refer to these proteins as the brain secreted proteome. We sought to identify secreted proteins sensitive to *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 from *Mecp2* null male rat and mouse models, and the CSF from female individuals with Rett syndrome collected before and after recombinant IGF1 treatment (Fig. 1A) ^{35,36}. 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 experimental sites, quantification platforms, and species to inform biomarker selection (Fig. S1). We chose to quantify proteomes with Tandem Mass Tagging (TMT) mass spectrometry as a high precision method ^{37,38}. TMT datasets were analyzed by fold of change/p value volcano thresholding plus machine learning approaches (Fig. 1A).

The Secreted Proteome of a MeCP2 Deficient Neuronal Cell Line.

We began using differentiated post-mitotic human neurons, LUHMES cells, where the *MECP2* gene was edited by CRISPR-Cas9 ³⁹. We reasoned the secreted proteome of a single cell type would define cell autonomous protein candidates for cell-type annotation of hits obtained in cerebrospinal fluids from individuals with Rett syndrome and rodent *Mecp2*-mutant models. We characterized the culture media before and after cell conditioning (Fig. 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 a N2 supplement starting from high purity reagents. Our customized N2 allowed us to distinguish proteins contributed by either wild type or mutant neurons (Fig. 1C, compare lanes 1 with 2-3). TMT mass spectrometry of media alone identified 704 proteins (Fig. 1D and G). In contrast, cell-conditioned media revealed 958 additional proteins contributed by wild type cells (Fig. 1D and G). 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 (Fig. 1E and H). 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 ^{40,41}. To these volcano 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 (Fig. 1J, asterisks for common volcano and machine learning hits) ⁴². 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 (Fig. 1K) and confusion matrix analysis where predicted and actual genotypic classes closely matched (Fig. 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 (Fig. 1F and I). The overlap between these two MECP2 sensitive proteomes was minor and barely significant (Fig. 1M). We found just six common hits, among them SST and IGFBPL1 (Fig. 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 (GO:0006090 $p = 3.96E-05$ Bonferroni corrected, Fig. 1N). These data suggest that ontologies may be better positioned than proteomic hits to identify convergence between the secreted and cellular proteome in MECP2 gene mutations within a simple cellular system.

Secreted Proteomes of 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 cerebrospinal fluid from wild type and *Mecp2* mutants in two rodent species and Rett syndrome individuals. We performed studies in rats (Fig. 2A-C) as well as in a large cohort of wild type and *Mecp2* null mice (Fig. 2D-F). We identified 70 and 64 CSF proteins whose expression was downregulated in mutant rat and mouse CSF, respectively (Fig. 2A, B, D, and E). 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 (Fig. 2A, B, D, and E). 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)³⁷. This analysis identified 44 proteins whose expression was affected in mutant CSF (Fig. S2A) and confirmed the downregulation of apolipoproteins and complement factors in *Mecp2*-null CSF (Fig. S2A-C, 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 (Fig. 2B and E, S2A-B, Table S2). Synaptic proteins were prominent among factors upregulated in the mouse *Mecp2* mutant CSF (Fig. 2D-E). These synaptic proteins include, but were not limited to, Snap25, Stx1b, Stxbp1, Syn1, and Syn2 (Fig. 2D-E)⁴³. Finally, we also identified mitochondria proteins within both the significant up- and downregulated proteomes in mutant rat and mouse CSF (Fig. 2D-E).

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, Fig. 2F and S2C-E). 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 (Fig. S2D). We identified Myg1, a mitochondrial protein⁴⁴, and *Mecp2* as the top-two most important classifiers to discern between wild type and *Mecp2* mutant brain tissue (Fig. S2D). All these machine learning analyses performed satisfactorily as determined by area under the curve in ROC analysis (0.72-0.92, Fig. 2F, and S2) and/or confusion matrices (Fig. 2F, and S2). 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 protein whose expression was modified in *Mecp2* mutant cortex among 6752 proteins quantified by TMT (Fig. 2C). None of the *Mecp2*-sensitive cortical proteins overlapped with rat CSF hits (Fig. 2G). We reasoned that a lack of overlap could emerge from a *Mecp2*-sensitive cortex proteome with small expression differences as compared to CSF due to the complexity of the former. Thus, we relaxed thresholding to include proteins whose expression was significantly different ($p < 0.05$) without applying a fold-of-change cutoff. This resulted in a relaxed dataset of 543 cortical proteins in whose expression may be modified in *Mecp2* mutants (Fig. 2C). However, this enlarged and relaxed cortical dataset produced only one overlapping hit with CSF (Fig. 2G, Ptpn2). An alternative model to account for the lack of overlap between *Mecp2* cortex and CSF proteomes could come from sparse hits identified in mutant rat CSF. Yet comparison of a 405 proteins pooled rat and mouse *Mecp2*-sensitive CSF proteome, obtained by thresholding and machine learning, yielded a non-significant overlap of just four hits (Fig. 2G). We expanded our analysis searching for CSF hits in *Mecp2* mutant rats that were also quantified in rat cortices. We 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 (Fig. 2H). Finally, we evaluated whether the *Mecp2*-sensitive CSF and cortical proteomes could overlap at the

ontology level. To evaluate this hypothesis, we compared whether the rat *Mecp2*-sensitive cortex and CSF proteomes could converge on similar ontologies. We identified significant ontological categories only if the *Mecp2*-sensitive CSF proteome was analyzed in conjunction with the relaxed *Mecp2* cortical proteome of 543 proteins. Under these conditions, we found overlapping ontologies including the synapse, which was represented by 75% of the annotated proteins contributed by the cortical and 25% by the CSF *Mecp2*-sensitive proteomes (Fig. 2I, GO:0045202, Bonferroni corrected group $p=2.467E-22$). These results argue that ontologies predicting putative brain disease mechanisms can be inferred from changes in the CSF proteome. However, 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 homogeneously deficient in *Mecp2* protein. In contrast, the human CSF proteome reports a genetically mosaic female brain³⁻⁵. We studied a cohort of individuals with Rett syndrome where CSF was collected as part of a phase I clinical trial⁴⁵. These participants were subjected to extended treatment with recombinant IGF-1 during the trial. CSFs were collected from 12 individuals. In nine 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 due to ethical constraints⁴⁵. Even though IGF-1 treatment did not improve clinical outcomes in Rett subjects³⁵, 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 (Fig. 2J-L). We identified by volcano thresholding 7 proteins whose expression was decreased after IGF-1 treatment and 17 proteins whose expression was increased (Fig. 2J-K). Importantly, 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 (Fig. 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 (Fig. 2L). 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.

Composite Ontologies of *Mecp2* Mutant Cerebrospinal Fluids and Conditioned Media

The secreted proteomes from human cultured neurons, rat, mouse, and human mutant CSFs revealed common proteins across these diverse experimental systems. These proteins belong to apolipoproteins, complement, or mitochondrial pathways (Fig. 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 distributions⁴⁶. 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 cells⁴⁷. We collated proteins from each of the four *Mecp2*- and *MECP2*-deficient experimental systems, selected by volcano thresholding plus machine learning (Fig. 1A), to identify a space of shared ontologies (Fig. 3A). We simultaneously 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 contributions to each ontology. We identified a space of 87 ontologies significantly represented in all the *Mecp2*- and *MECP2*-sensitive proteome datasets (Bonferroni corrected p values <10E-3, Fig. 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 (Fig. 3B). Importantly, 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 (Fig. 3C-D). 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 (Fig. 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 (Fig. 3E). This outcome was similar if we performed HumanBase analyses either with astrocyte- or neuron-centric queries (Fig. 3E). Finally, we confirmed that *Mecp2*- and *MECP2*-sensitive 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 (Fig. 3F-G)^{41,43,48}. The collated *Mecp2*- and *MECP2*-sensitive proteome contained 80 proteins in common with the HDL proteome (Fig. 3F). This represents a significant 7.9-fold enrichment above what is expected by chance (Fig. 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 (Fig. 3F)⁴⁹. Similarly, the collated *Mecp2*- and *MECP2*-sensitive 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. These ontological findings demonstrate a varied and novel set of molecular phenotypes associated to Rett syndrome. These Rett syndrome molecular phenotypes offer a distinctive path for disease biomarkers.

Ontologies Inform Putative Biomarkers of Rett Syndrome.

We used convergent Rett molecular 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 the mutant secreted proteomes (Fig. 3B, E and F). We performed absolute quantification (AQUA) of proteins by mass spectrometry to confirm expression changes in HDL apolipoproteins in independent cohorts of rat (Fig. 4A) and mouse CSF from wild type and mutant animals (Fig. 4B-C)⁵⁰. We also confirmed CSF hits in Rett syndrome individuals before and after IGF1 treatment by electrochemiluminescent ELISA assays (Fig. 4C)⁵¹.

AQUA quantification of the HDL lipoprotein particle components ApoA1, Apob, Pon1, C3, and C9 revealed decreased levels in mutant as compared to wild type rat CSF (Fig. 4A). Levels of loading controls, App and A2m, were not affected by genotype in rats (Fig. 4A). These findings were extended to Rett mouse models (Fig. 4B-C). We used isotopolog peptide standards to measure proteins associated to HDL particles in mouse CSF. Diverse apolipoproteins (ApoA1, Apoc1, Apoc2, Apoe), complement subunits (C6, C7, C8a, C8b, C8g, C9), and antiproteases (A1at5, Serpina3k) were reduced in *Mecp2* deficient animals compared to controls (Fig. 4B-C). 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 mutant mice (Fig. 4C). These results independently validate our TMT findings in the CSF of two rodent models of Rett syndrome.

We next focused on the nine participants where CSF samples were obtained before and after IGF1 treatment. We deployed principal component analysis to assess individual and group responses to IGF1. The main variable segregating participant CSF proteomes was their identity rather than treatment itself (Fig. 4D). Second, treatment modified a participant's CSF composition discretely, mostly driven by changes in the levels of apolipoproteins (Fig. 4B). Importantly, the levels of blood proteins such as hemoglobins (HBB and HBD), albumin (ALB), and immunoglobulins could not account for the changes in apolipoprotein protein levels (Fig. 4B). We confirmed that IGF1 treatment increased the content of APOA1 in Rett participant's CSF by ELISA (Fig. 4C). In contrast, two proteins whose expression was not modified by treatment in TMT mass spectrometry quantifications, APOE and B2M, did not change their levels after treatment in ELISA

assays (Fig. 4C). These results confirm protein hits and ontologies in human and rodent Rett syndrome models.

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 models³³ (Fig. 4F-G). We found that the CSF proteome of *Cdkl5* mutants was different from *Mecp2* mutant CSF, as indicated by the 16 different proteins selected from *Mecp2* CSF whose levels remained unchanged in *Cdkl5* null CSF. The only exception was Apoe, which was decreased in both *Mecp2* and *Cdkl5* mutant CSF (Fig. 4G). These findings demonstrate that *Mecp2* CSF hits discriminate between phenotypically related forms of neurodevelopmental disorders. Second, we interrogated whether selected *Mecp2* CSF proteins were expressed in disease-relevant cell types (Fig. S3A-B). HDL apolipoproteins such as Apoa1, Apoe, Pon1, and Clu were expressed in neurons and astrocytes (Fig. 4A). Single cell RNAseq data 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 (Fig. S3B). Apoc1 mRNA or its protein were undetectable in neurons and glia but present in plasma (compare Fig. S3A and B). 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 (Fig. S3B). For example, C8b and C9 mRNAs were expressed in a discrete neuronal population whereas C7 was broadly expressed in cortex and hippocampus (Fig. S3B). Thus, ontologically selected protein hits are expressed in disease-relevant neuronal and glial cell types. Finally, we determined if selected proteins could distinguish subjects by genotype. We performed ROC analysis focusing on the *Mecp2* mutant mouse CSF hits as the animal cohort of the biggest size, the mouse TMT dataset (Fig. 4H). Each ontologically selected analyte efficiently distinguished genotypes irrespective of if they belonged to either the HDL lipoprotein, synapse, or mitochondrion category. In fact, all analytes ROC area under the curve were between 0.77 to 0.9 with significant p values (Fig. 4H). 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 datasets were analyzed (Fig. 4H, compare grey and purple symbols). None of these validated analytes experienced modifications in their mRNA expression across diverse brain regions in *Mecp2* mutant mice, indicating that the utility of these analytes as putative biomarkers of *Mecp2* gene defect is restricted to their protein levels in CSF (Fig. S4). 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 a X-linked neurodevelopmental disorder, Rett syndrome, reproducibly and distinctively impacts the composition of the CSF. We defined a consensus proteome and ontological 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, as well as synapse compartments. 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 two sites for CSF collection and mass spectrometry analysis (Fig. S1). Although, the different mutant CSF proteomes produced discrete overlap across individual analytes, they all shared significant overlap at the ontology level. 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 a proof of principle to the use of convergent ontologies as a strategy to select analytes predicted to report 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. We propose that *Mecp2* mutant CSF ontologies inform robust CSF analytes to act as Rett syndrome biomarkers in humans.

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^{52,53}. 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. Importantly, Apoe brain levels are locally controlled without contributions from plasma⁵⁴. 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 postmitotic 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^{55,56}. Thus, the *Mecp2* secreted proteome offers multiple analytes to assess phenotypes in multiple brain cell types lacking *Mecp2*.

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^{4,57}. 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^{58,59}. 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 synthesis⁶⁰. 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. Pyruvate and lactate are increased in Rett syndrome individual's CSF, 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^{29,30}. However, there are not enough studies to tie together our observations in CSF with potential models of mitochondrial dysfunction in *Mecp2* mutant cells⁶¹⁻⁶⁵. 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.

Materials and Methods

Rat and Sample Collection

All 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*^{-/-} rats were maintained by crossing *Mecp2*^{+/-} females with wild type Sprague Dawley males. Animals were maintained on 12-hour light/dark cycles with free access to normal rat food. WT and *Mecp2*^{-/-} rats at 25 days of postnatal age were weighed and assessed for the development of the RTT-like phenotypes prior to surgery. Rats were anaesthetized using intraperitoneal administration of an injectable cocktail of medetomidine (0.5mg/kg) and ketamine (75mg/kg). Once the animal was deeply anaesthetized, 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 2.5ml 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.

Mice and CSF Collection

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

Our terminal CSF collection method was adapted from a previously published protocol⁶⁶. Mice were deeply anaesthetized 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

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 μ l or clear CSF was drawn into the syringe by slow and smooth aspiration. The CSF was immediately spun down for 30 seconds and clear CSF was inspected with the naked eye and frozen immediately on dry ice. Frankly blood contaminated samples discarded.

Human Subjects

Clinical features of the cohort used in these studies are described by Khwaja et al.⁴⁵. The referred 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.

Cell Culture and Conditioned Media Preparation.

LUHMES wild-type control, and MECP2 knock-out 2_7 cell line were differentiated and conditioned media was collected. Nunclon flasks and plates were treated with a 44 μ g/ml Poly-L-Ornithine (Sigma P3655) and 1 μ g/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), 2mM L-glutamine (Sigma G7513), and 40ng/ml beta-FGF (R&D Systems 4114-TC-01M). After 24 hours, media was changed to differentiation media (Advanced DMEM/F12 with N2, 2mM L-glutamine, 1mM DbcAMP (Sigma D0627), 1 μ g /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. 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 μ g/ml progesterone (Sigma P6149), 1.61 mg/ml putrescine dihydrochloride (Sigma P5780), 0.52 μ g/ml sodium selenite (Sigma S5261) and DMEM/F12 (Thermo Fisher 21331020). 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 μ g/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 minutes. 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 minutes to precipitate out proteins. The solution was centrifuged at 16,000 x g for 15min 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 10minutes. Acetone was aspirated and the pellet was lightly air-dried, dissolved in 200 µl of 8M Urea, and flash frozen on dry ice.

Mass Spectrometry Emory

Sample Processing. All CSF (5µl) samples were diluted with 50 µl of 50 mM NH₄HCO₃ and treated with TCEP and CAA and heated at 90°C for 10 minutes. 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 a HLB microelution plate (Waters) and dried under vacuum.

Tandem Mass Tag (TMT) Labeling. For each sample, labeling was performed as previously described^{22,67}. Briefly, each was re-suspended in 100 mM TEAB buffer (100 µL). The TMT and TMTPro labeling reagents were equilibrated to room temperature, and anhydrous ACN (256 µL) 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 1 h at room temperature. The reaction was quenched with 5% (vol/vol) 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% (vol/vol) TFA, then acidified to a final concentration of 1% (vol/vol) FA and 0.1% (vol/vol) TFA. Peptides were desalted with a 30 mg C18 Sep-Pak column (Waters). Each Sep-Pak column was activated with 1 mL of methanol, washed with 1 mL of 50% (vol/vol) ACN, and equilibrated with 2x1 mL of 0.1% TFA. The samples were then loaded and each column was washed with 2x1 mL 0.1% (vol/vol) TFA, followed by 1 mL of 1% (vol/vol) FA. Elution was performed with 2 volumes of 0.5 mL 50% (vol/vol) ACN. The eluates were then dried to completeness.

High pH Fractionation. High pH fractionation was performed essentially as described^{68,69} with slight modification. Dried samples were re-suspended in high pH loading buffer (0.07% vol/vol NH₄OH, 0.045% vol/vol FA, 2% vol/vol ACN) and loaded onto Water's BEH C18 column (2.1mm x 150 mm with 1.7 µm beads). An Thermo Vanquish system was used to carry out the fractionation. Solvent A consisted of 0.0175% (vol/vol) NH₄OH, 0.01125% (vol/vol) FA, and 2% (vol/vol) ACN; solvent B consisted of 0.0175% (vol/vol) NH₄OH, 0.01125% (vol/vol) FA, and 90% (vol/vol) 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 TMT batches. For the Cdk5 batch, 192 fractions were collected and combined into 96 fractions. All fractions were dried to completeness using a vacuum centrifugation.

Liquid Chromatography Tandem Mass Spectrometry for TMT. Each of the peptide fractions was resuspended in loading buffer (0.1% FA, 0.03% TFA, 1% ACN). Peptide eluents were either separated on a self-packed C18 (1.9 μ m Dr. Maisch, Germany) fused silica column (15 cm \times 100 μ m internal diameter (ID), New Objective, Woburn, MA) or a Water's 1.7 μ m CSH C18 column (15 cm \times 150 μ m internal diameter). An Easy nLC 1200 (ThermoFisher 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 Spectrometry for Parallel Reaction Monitoring (PRM). AQUA standard peptides (ThermoFisher Scientific) were spiked into digested mouse CSF samples. For each sample and equivalent of 1 μ l of CSF was loaded onto a Water's 1.7 μ m CSH C18 column (15 cm \times 150 μ m internal diameter). Peptides were eluted using a 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 20ppm precursor mass window, 0.05 Da product mass window, dynamic modifications methionine (+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 suite⁷⁰.

Mass Spectrometry Ann Arbor

Sample Preparation. For tissues, samples were washed twice in 1X PBS and lysed in 8M urea, 50mM 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 1h at RT in 12mM DTT followed by alkylation for 1h at RT in 15mM 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 each sample was acidified in

formic acid and subjected to SPE on an Empore SD C18 plate (3M catalogue# 6015 SD). Each sample was lyophilized and reconstituted in 140mM HEPES, pH 8.0, 30% acetonitrile.

Label Free Quantification Mass Spectrometry. A 2µg aliquot was analyzed by nano LC/MS/MS with a Waters NanoAcquity HPLC system interfaced to a ThermoFisher Fusion Lumos. Peptides were loaded on a trapping column and eluted over a 75µm analytical column at 350nL/min; both columns were packed with Luna C18 resin (Phenomenex). A 4h 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 3s 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: 20ppm 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. 40µL of acetonitrile was added to each TMT tag tube and mixed aggressively. Tags were incubated at RT for 15min. 30µL of label was added to each peptide sample and mixed aggressively. Samples were incubated in an Eppendorf Thermomixer at 300rpm 25°C for 1.5h. Reactions were terminated with the addition of 8µL of fresh 5% hydroxylamine solution and 15min incubation at room temperature. Samples were subjected to high pH reverse phase fractionation as follows; Buffers: Buffer A: 10mM NaOH, pH 10.5, in water Buffer B: 10mM NaOH, pH 10.5, in acetonitrile. We used XBridge C18 columns, 2.1mm ID x 150mm length, 3.5µm particle size (Waters, part #186003023) attached to a Agilent 1100 HPLC system equipped with a 150µL sample loop operating at 0.3mL/min, detector set at 214 nm wavelength. Dried peptides were resolubilized in 150µL of Buffer A and injected manually. Fractions were collected every 30s from 1-49min (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 75µm analytical column at 350nL/min; both columns were packed with Luna C18 resin (Phenomenex). Each high pH RP fraction was separated over a 2h gradient (24h 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 3s 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), Phospho

(STY; PO4 data only). Fragment Mass Tolerance: 20ppm. 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 AGVLAHLEEEER; IAQLEEQLDNETK. Pon1 IFFYDSENPPGSEVLR; LLIGTVFHR. App TEEISEVK; THTHIVIPYR. A2m AIAYLNTGYQR; LPSDVVEESAR. Apoa1 DYVSQFESSTLGK; WNEEVEAYR. Apob TEVIPPLIENR; GFPTLEALFGK. C3 GLEVSITAR; SSVAVPYVIVPLK. C9 SIEVFGQFQGK; TTSFNANLALK. Thbs1 FVFGTTPEDILR; IENANLIPPVPDDK. A 3-4 µg aliquot of each CSF tryptic peptide digests was spiked with isotopolog peptide 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 ThermoFisher Fusion Lumos mass spectrometer. 1.5µg per sample was loaded on a trapping column and eluted over a 75µm analytical column at 350nL/min; both columns were packed with Luna C18 resin (Phenomenex). A 1h 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 5e4, and maximum IT of 22ms. 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 charts⁴². 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 ExtraTrees (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-Classifer (random_state = 23 learning_rate = 0.3 min_split_loss = 0 max_depth = 6 min_child_weight = 1), a AdaBoost-Classifer (random_state = 23 n_estimators = 100 learning_rate = 1.0), or RandomForest-Classifer (random_state = 23 n_estimators = 100 criterion = gini max_features = auto). Classifiers 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

RNAseq data analysis was described in Wynne et al.⁷¹. Single cell RNA seq data were described in⁷². Gene expression data matrix (matrix.csv) and cell metadata (metadata.csv) data were downloaded from [the Allen Institute Portal](https://portal.gdc.cancer.gov/) and processed as described⁷¹ 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.

Bioinformatic Analyses

Gene ontology analyses were performed with Cluego and HumanBase⁴⁷. ClueGo v2.58 run on Cytoscape v3.8.2^{46,73}. 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 webased parameters⁴⁷. *In silico* interactome data were downloaded from Genemania predicted and physical interactions and processed in Cytoscape v3.8.2⁷³. Interactome connectivity graph parameters were generated in Cytoscape.

Statistical Analyses

Volcano plot p values were calculated using Qlucore Omics Explorer Version 3.6(33) without multiple corrections. Experiments in Figure 4A-B and G statistical analyses were performed with the engine <https://www.estimationstats.com/#/> with a two-sided permutation t-test and alpha of 0.05⁷⁴. ROC analysis and paired t-test were performed with Prism v9.2.0(283).

Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE⁷⁵ partner repository with dataset identifiers: PXD029808, PXD029809, PXD029811, PXD029835

Figure Legends

Figure 1. The Secreted Proteome of Post-Mitotic MECP2 Mutant Human Neurons

A. Experimental strategy to define a consensus Rett syndrome secreted proteome and infer conserved ontologies. **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 ⁷⁶. Lane 1 in B and C represent naïve 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 to 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 to I** show clustered heat maps of hits selected in D to F. Arrows mark some cardinal hits. Rows are depicted as minimum and maximum intensities (blue-yellow scale) and annotated by log2 fold of change (rainbow scale, see table S1). **J to L** analysis of TMT data in panel E using an XGBoost machine learning algorithm. **J** presents main hits discriminating wild type and *MECP2* conditioned media in the decision tree. Asterisks mark proteins identified both by volcano thresholding and machine learning. **K-L,** performance of the machine learning protocol estimated by ROC analysis, **J**, and confusion matrix in **I**. Area under the curve in **J** = 0.97±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 (see table S1). P value estimated with exact hypergeometric probability Bonferroni corrected.

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 see table S2 for hit lists at p<0.01. n=5. **D and E** show mouse CSF TMT volcano plot and heat map 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 ± 0.11 . **G**, Top Venn diagram shows the overlap between *Mecp2*-sensitive rat CSF and rat cortex hits using two thresholding criteria $p < 0.05$ and $p < 0.01$. Bottom Venn diagram compares *Mecp2*-sensitive hits in rat and mouse CSF pooled together with *Mecp2*-sensitive rat cortex hits. **H** depict the correlation in expression of *Mecp2*-sensitive hits in rat CSF and rat cortex. **I**, shows the % overlap between the cellular proteome ontologies inferred from the datasets shown in A and C calculated with the ClueGo application (see table S2). **J** and **K** show Rett syndrome female individual CSF TMT volcano plot and heat map 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. **L** analysis of TMT data presented in panel J 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 ± 0.26 .

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 of contribution of each experimental system to each ontology. Grey 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-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. **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 overlap beyond what is expected by chance. See Table S3 for all ontological data.

Figure 4. Ontologically Selected and Confirmed Rett Syndrome CSF Proteome Hits Perform as Putative Biomarkers of Disease.

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 mouse CSF (**B**). In **A**, femtomoles of the endogenous CSF peptide were normalized to a randomly chosen control sample. In **B**, the ratio of the CSF endogenous peptide to the isotopologue peptide were used to quantify relative analyte abundance. Gray bars correspond to wild type CSF and blue bars *Mecp2* null CSF. All analytes were measured independently with 2-3 isotopologue peptides as standards. In **A** two batches of 5 rats of each genotype were used while in **B**, one batch of 10 mice of each genotype were analyzed. **C** depicts a heat map of all the modified AQUA determinations performed in mouse CSF samples that showed significant differences between genotypes. Every isotopologue peptide correspond to a row. Data are depicted as row median divided by the row median absolute deviation both as heat map and by symbol size. **D**, principal component analysis of Rett syndrome participants before and after treatment (gray and blue symbols, respectively), every subject is color coded and the after-treatment PCI coordinates are indicated by a callout triangle. Analyte expression levels from a TMT mass spectrometry quantification were mapped to the PCI coordinates as a heat map. Apolipoproteins increased expression after IGF-1 treatment. Note controls do not experience changes (B2M, HBB, HBD, ALB). **E**, MesoScale ELISA confirmation of APOA1 levels in Rett individuals CSF before and after IGF-1 treatment. **F** and **G**, a heat map of ontology selected and confirmed analytes in B-C tested in the CSF of *Cdkl5*-null mice. Data correspond to the normalized abundance measured by TMT mass spectrometry. **G** shows APOE, the only analyte whose expression was modified in *Cdkl5*-null CSF, shown by the gray box in **F**. **H**, ROC analysis of selected mouse CSF *Mecp2*-sensitive hits either quantified by TMT mass spectrometry (grey symbols) or by modified AQUA (purple symbols). Number represents p value that tests the null hypothesis that the area under the curve=0.50 (non-discriminatory). P values in **A**, **B**, **E** and **G** were calculated with a two-sided permutation t-test. See table S4 for data to panels B-C and F-G.

Fig. S1. Design of a Robust Strategy to Identify Consensus Rett Syndrome Proteomes for Ontology Inference and Biomarker Selection.

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.

Fig. S2. Orthogonal Approaches to Identify Rett Syndrome CSF Proteomes.

A volcano plot of LFQ mass spectrometry determinations in rat cerebrospinal fluid. Cutoffs at log2 0.5 fold of change in protein abundance and a p value <0.05, n=5. See table SS2. **B** shows

clustered heat maps of hits selected in A. For A and B, see legend to Figure 1 for additional details. **C to E**, depict analysis of the indicated mass spectrometry datasets using a Random Forest machine learning algorithm. Bar graph showing main hits discriminating wild type and *Mecp2* mutant CSF in the decision tree. Inserts show performance of the machine learning protocol estimated by ROC analysis and confusion matrix.

Fig. S3. 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 ^{76,77} or abundance in the plasma proteome according to Geyer et al ⁷⁸. 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 ⁷⁹. Color codes denote neuronal subclasses described by Yao et al. ⁷⁹. 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 ⁸⁰.

Fig. S4. Brain Parenchyma Transcripts Levels of Analytes Selected by Ontology and with Biomarker Potential.

RNAseq analysis of transcript expression of indicated analytes in three microdissected brain regions from wild type and *Mecp2* mutant brains. *Mecp2* mRNA are presented as a control. N=5 animals per each phenotype. Only *Mecp2* mRNA is differentially expressed p for three brain regions p< 1.40E-236, Benjamini and Hochberg corrected Wald test. See table SS4.

References

1. Sztainberg, Y. & Zoghbi, H.Y. Lessons learned from studying syndromic autism spectrum disorders. *Nature neuroscience* **19**, 1408-1417 (2016).
2. Lee, C.E., Singleton, K.S., Wallin, M. & Faundez, V. Rare Genetic Diseases: Nature's Experiments on Human Development. *iScience* **23**, 101123 (2020).
3. Lyst, M.J. & Bird, A. Rett syndrome: a complex disorder with simple roots. *Nature reviews. Genetics* **16**, 261-275 (2015).
4. Banerjee, A., Miller, M.T., Li, K., Sur, M. & Kaufmann, W.E. Towards a better diagnosis and treatment of Rett syndrome: a model synaptic disorder. *Brain : a journal of neurology* **142**, 239-248 (2019).
5. Sandweiss, A.J., Brandt, V.L. & Zoghbi, H.Y. Advances in understanding of Rett syndrome and MECP2 duplication syndrome: prospects for future therapies. *Lancet neurology* **19**, 689-698 (2020).
6. Johnson, B.S., *et al.* Biotin tagging of MeCP2 in mice reveals contextual insights into the Rett syndrome transcriptome. *Nature medicine* **23**, 1203-1214 (2017).
7. Chahrour, M., *et al.* MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320**, 1224-1229 (2008).
8. Raman, A.T., *et al.* Apparent bias toward long gene misregulation in MeCP2 syndromes disappears after controlling for baseline variations. *Nature communications* **9**, 3225 (2018).
9. Cholewa-Waclaw, J., *et al.* Quantitative modelling predicts the impact of DNA methylation on RNA polymerase II traffic. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 14995-15000 (2019).
10. Wu, H., *et al.* Genome-wide analysis reveals methyl-CpG-binding protein 2-dependent regulation of microRNAs in a mouse model of Rett syndrome. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 18161-18166 (2010).
11. Cheng, T.L., *et al.* MeCP2 suppresses nuclear microRNA processing and dendritic growth by regulating the DGCR8/Drosha complex. *Developmental cell* **28**, 547-560 (2014).
12. Mullin, A.P., *et al.* Neurodevelopmental disorders: mechanisms and boundary definitions from genomes, interactomes and proteomes. *Translational psychiatry* **3**, e329 (2013).
13. Chao, H.T., *et al.* Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **468**, 263-269 (2010).
14. Chao, H.T., Zoghbi, H.Y. & Rosenmund, C. MeCP2 controls excitatory synaptic strength by regulating glutamatergic synapse number. *Neuron* **56**, 58-65 (2007).
15. Matarazzo, V. & Ronnett, G.V. Temporal and regional differences in the olfactory proteome as a consequence of MeCP2 deficiency. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7763-7768 (2004).
16. Pacheco, N.L., *et al.* RNA sequencing and proteomics approaches reveal novel deficits in the cortex of Mecp2-deficient mice, a model for Rett syndrome. *Molecular autism* **8**, 56 (2017).
17. Chau, K.F., *et al.* Progressive Differentiation and Instructive Capacities of Amniotic Fluid and Cerebrospinal Fluid Proteomes following Neural Tube Closure. *Developmental cell* **35**, 789-802 (2015).

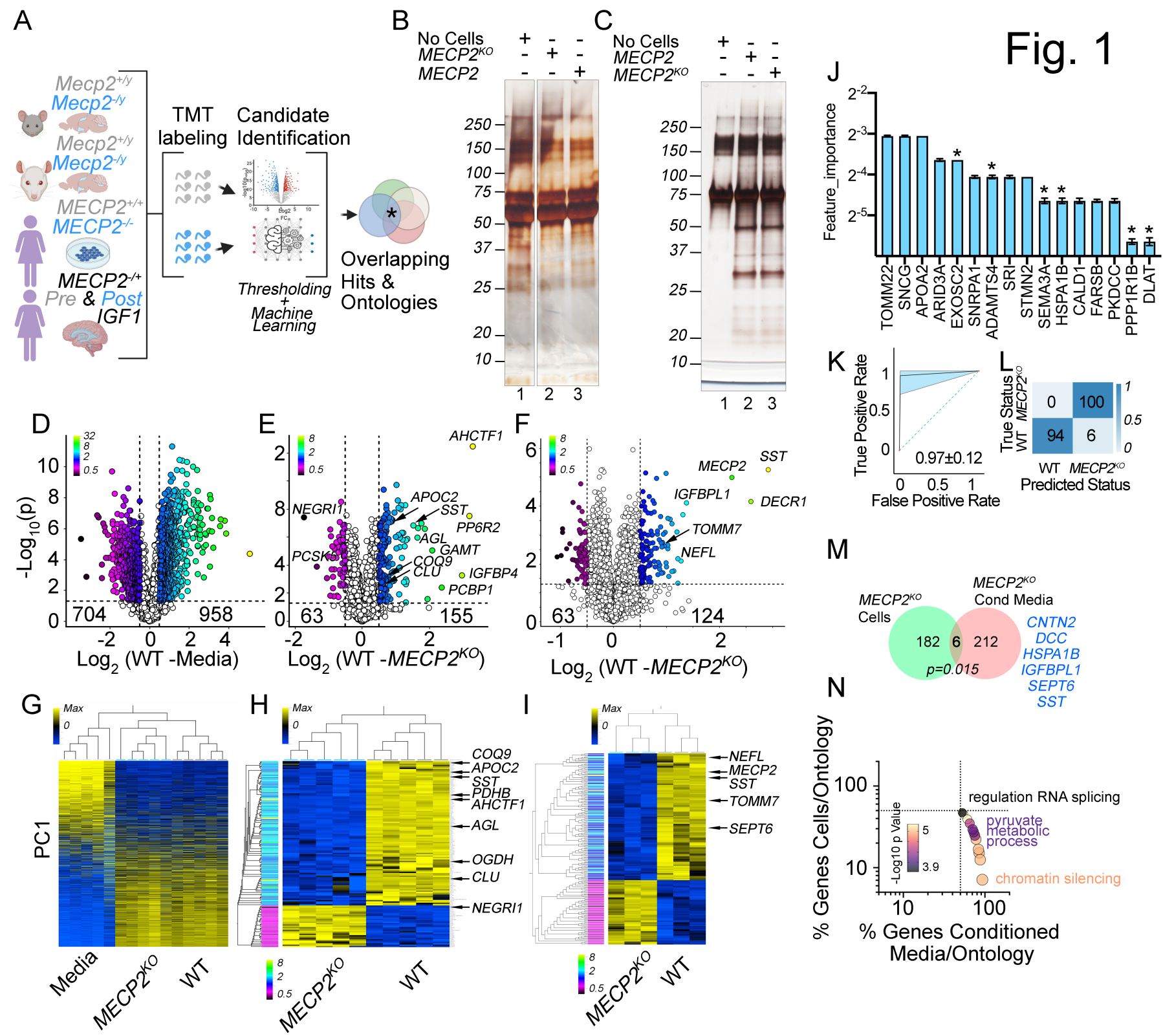
18. Kaiser, K., *et al.* WNT5A is transported via lipoprotein particles in the cerebrospinal fluid to regulate hindbrain morphogenesis. *Nature communications* **10**, 1498 (2019).
19. Lehtinen, M.K. & Walsh, C.A. Neurogenesis at the brain-cerebrospinal fluid interface. *Annual review of cell and developmental biology* **27**, 653-679 (2011).
20. Johanson, C.E., *et al.* Multiplicity of cerebrospinal fluid functions: New challenges in health and disease. *Cerebrospinal Fluid Res* **5**, 10 (2008).
21. Olsson, B., *et al.* CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet neurology* **15**, 673-684 (2016).
22. Higginbotham, L., *et al.* Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic Alzheimer's disease. *Sci Adv* **6**(2020).
23. Johnson, E.C.B., *et al.* Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nature medicine* **26**, 769-780 (2020).
24. Abbasi, D.A., *et al.* Characterization of the Cerebrospinal Fluid Proteome in Patients with Fragile X-Associated Tremor/Ataxia Syndrome. *Cerebellum* (2021).
25. Zimmerman, A.W., *et al.* Cerebrospinal fluid and serum markers of inflammation in autism. *Pediatr Neurol* **33**, 195-201 (2005).
26. Riikonen, R., *et al.* Cerebrospinal fluid insulin-like growth factors IGF-1 and IGF-2 in infantile autism. *Dev Med Child Neurol* **48**, 751-755 (2006).
27. Oztan, O., Garner, J.P., Constantino, J.N. & Parker, K.J. Neonatal CSF vasopressin concentration predicts later medical record diagnoses of autism spectrum disorder. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 10609-10613 (2020).
28. Oztan, O., *et al.* Autism-associated biomarkers: test-retest reliability and relationship to quantitative social trait variation in rhesus monkeys. *Molecular autism* **12**, 50 (2021).
29. Budden, S.S., Myer, E.C. & Butler, I.J. Cerebrospinal fluid studies in the Rett syndrome: biogenic amines and beta-endorphins. *Brain Dev* **12**, 81-84 (1990).
30. Matsuishi, T., *et al.* Abnormal carbohydrate metabolism in cerebrospinal fluid in Rett syndrome. *J Child Neurol* **9**, 26-30 (1994).
31. Evans, J.C., *et al.* Early onset seizures and Rett-like features associated with mutations in CDKL5. *European journal of human genetics : EJHG* **13**, 1113-1120 (2005).
32. Weaving, L.S., *et al.* Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. *American journal of human genetics* **75**, 1079-1093 (2004).
33. Zerbi, V., *et al.* Brain mapping across 16 autism mouse models reveals a spectrum of functional connectivity subtypes. *Molecular psychiatry* (2021).
34. Kalluri, R. & LeBleu, V.S. The biology, function, and biomedical applications of exosomes. *Science* **367**(2020).
35. O'Leary, H.M., *et al.* Placebo-controlled crossover assessment of mecamermin for the treatment of Rett syndrome. *Ann Clin Transl Neurol* **5**, 323-332 (2018).
36. Tropea, D., *et al.* Partial reversal of Rett Syndrome-like symptoms in MeCP2 mutant mice. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2029-2034 (2009).

37. O'Connell, J.D., Paulo, J.A., O'Brien, J.J. & Gygi, S.P. Proteome-Wide Evaluation of Two Common Protein Quantification Methods. *Journal of proteome research* **17**, 1934-1942 (2018).
38. Li, Z., *et al.* Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *Journal of proteome research* **11**, 1582-1590 (2012).
39. Shah, R.R., *et al.* Efficient and versatile CRISPR engineering of human neurons in culture to model neurological disorders. *Wellcome Open Res* **1**, 13 (2016).
40. Rasala, B.A., Orjalo, A.V., Shen, Z., Briggs, S. & Forbes, D.J. ELYS is a dual nucleoporin/kinetochore protein required for nuclear pore assembly and proper cell division. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 17801-17806 (2006).
41. Rath, S., *et al.* MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic acids research* **49**, D1541-D1547 (2021).
42. Torun, F.M., *et al.* Transparent exploration of machine learning for biomarker discovery from proteomics and omics data. *bioRxiv* (2021).
43. Koopmans, F., *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217-234 e214 (2019).
44. Grover, R., *et al.* Myg1 exonuclease couples the nuclear and mitochondrial translational programs through RNA processing. *Nucleic acids research* **47**, 5852-5866 (2019).
45. Khwaja, O.S., *et al.* Safety, pharmacokinetics, and preliminary assessment of efficacy of mecasermin (recombinant human IGF-1) for the treatment of Rett syndrome. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 4596-4601 (2014).
46. Bindea, G., *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091-1093 (2009).
47. Greene, C.S., *et al.* Understanding multicellular function and disease with human tissue-specific networks. *Nature genetics* **47**, 569-576 (2015).
48. Davidson, S. & Shah, A.S. HDL Proteome Watch. (2021).
49. Montojo, J., *et al.* GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics* **26**, 2927-2928 (2010).
50. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. & Gygi, S.P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 6940-6945 (2003).
51. Chowdhury, F., Williams, A. & Johnson, P. Validation and comparison of two multiplex technologies, Luminex and Mesoscale Discovery, for human cytokine profiling. *J Immunol Methods* **340**, 55-64 (2009).
52. Tsujita, M., *et al.* 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 (2021).
53. Stukas, S., *et al.* Intravenously injected human apolipoprotein A-I rapidly enters the central nervous system via the choroid plexus. *J Am Heart Assoc* **3**, e001156 (2014).
54. Linton, M.F., *et al.* Phenotypes of apolipoprotein B and apolipoprotein E after liver transplantation. *J Clin Invest* **88**, 270-281 (1991).

55. Zhang, Y., *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **34**, 11929-11947 (2014).
56. Xu, H., *et al.* Choroid plexus NKCC1 mediates cerebrospinal fluid clearance during mouse early postnatal development. *Nature communications* **12**, 447 (2021).
57. Na, E.S., Nelson, E.D., Kavalali, E.T. & Monteggia, L.M. The Impact of MeCP2 Loss- or Gain-of-Function on Synaptic Plasticity. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* (2012).
58. Wellington, C.L. & Frikke-Schmidt, R. Relation between plasma and brain lipids. *Curr Opin Lipidol* **27**, 225-232 (2016).
59. Liu, Q., *et al.* Neuronal LRP1 knockout in adult mice leads to impaired brain lipid metabolism and progressive, age-dependent synapse loss and neurodegeneration. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **30**, 17068-17078 (2010).
60. Buchovecky, C.M., *et al.* A suppressor screen in Mecp2 mutant mice implicates cholesterol metabolism in Rett syndrome. *Nature genetics* **45**, 1013-1020 (2013).
61. Kriaucionis, S., *et al.* Gene expression analysis exposes mitochondrial abnormalities in a mouse model of Rett syndrome. *Mol Cell Biol* **26**, 5033-5042 (2006).
62. Shulyakova, N., Andreazza, A.C., Mills, L.R. & Eubanks, J.H. Mitochondrial Dysfunction in the Pathogenesis of Rett Syndrome: Implications for Mitochondria-Targeted Therapies. *Frontiers in cellular neuroscience* **11**, 58 (2017).
63. Jagtap, S., *et al.* Aberrant mitochondrial function in patient-derived neural cells from CDKL5 deficiency disorder and Rett syndrome. *Human molecular genetics* **28**, 3625-3636 (2019).
64. Hirofujii, S., *et al.* Mitochondrial dysfunction in dopaminergic neurons differentiated from exfoliated deciduous tooth-derived pulp stem cells of a child with Rett syndrome. *Biochemical and biophysical research communications* **498**, 898-904 (2018).
65. Grosser, E., *et al.* Oxidative burden and mitochondrial dysfunction in a mouse model of Rett syndrome. *Neurobiology of disease* **48**, 102-114 (2012).
66. Boire, A., *et al.* Complement Component 3 Adapts the Cerebrospinal Fluid for Leptomenigeal Metastasis. *Cell* **168**, 1101-1113 e1113 (2017).
67. Ping, L., *et al.* Global quantitative analysis of the human brain proteome in Alzheimer's and Parkinson's Disease. *Sci Data* **5**, 180036 (2018).
68. Ping, L., *et al.* Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. *Sci Data* **7**, 315 (2020).
69. Gokhale, A., *et al.* Mitochondrial Proteostasis Requires Genes Encoded in a Neurodevelopmental Syndrome Locus that are Necessary for Synapse Function. *bioRxiv* (2020).
70. MacLean, B., *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010).
71. Wynne, M.E., *et al.* Heterogeneous Expression of Nuclear Encoded Mitochondrial Genes Distinguishes Inhibitory and Excitatory Neurons. *eNeuro* (2021).
72. Yao, Z., *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* <https://doi.org/10.1016/j.cell.2021.04.021> (2020).

73. Shannon, P., *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504 (2003).
74. Ho, J., Tumkaya, T., Aryal, S., Choi, H. & Claridge-Chang, A. Moving beyond P values: data analysis with estimation graphics. *Nature methods* **16**, 565-566 (2019).
75. Deutsch, E.W., *et al.* The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic acids research* **48**, D1145-D1152 (2020).
76. Scholz, D., *et al.* Rapid, complete and large-scale generation of post-mitotic neurons from the human LUHMES cell line. *J Neurochem* **119**, 957-971 (2011).
77. Zhang, Y., *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89**, 37-53 (2016).
78. Geyer, P.E., *et al.* Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst* **2**, 185-195 (2016).
79. Yao, Z., *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222-3241 e3226 (2021).
80. Wynne, M.E., *et al.* Heterogeneous Expression of Nuclear Encoded Mitochondrial Genes Distinguishes Inhibitory and Excitatory Neurons. *eNeuro* **8**(2021).

Fig. 1



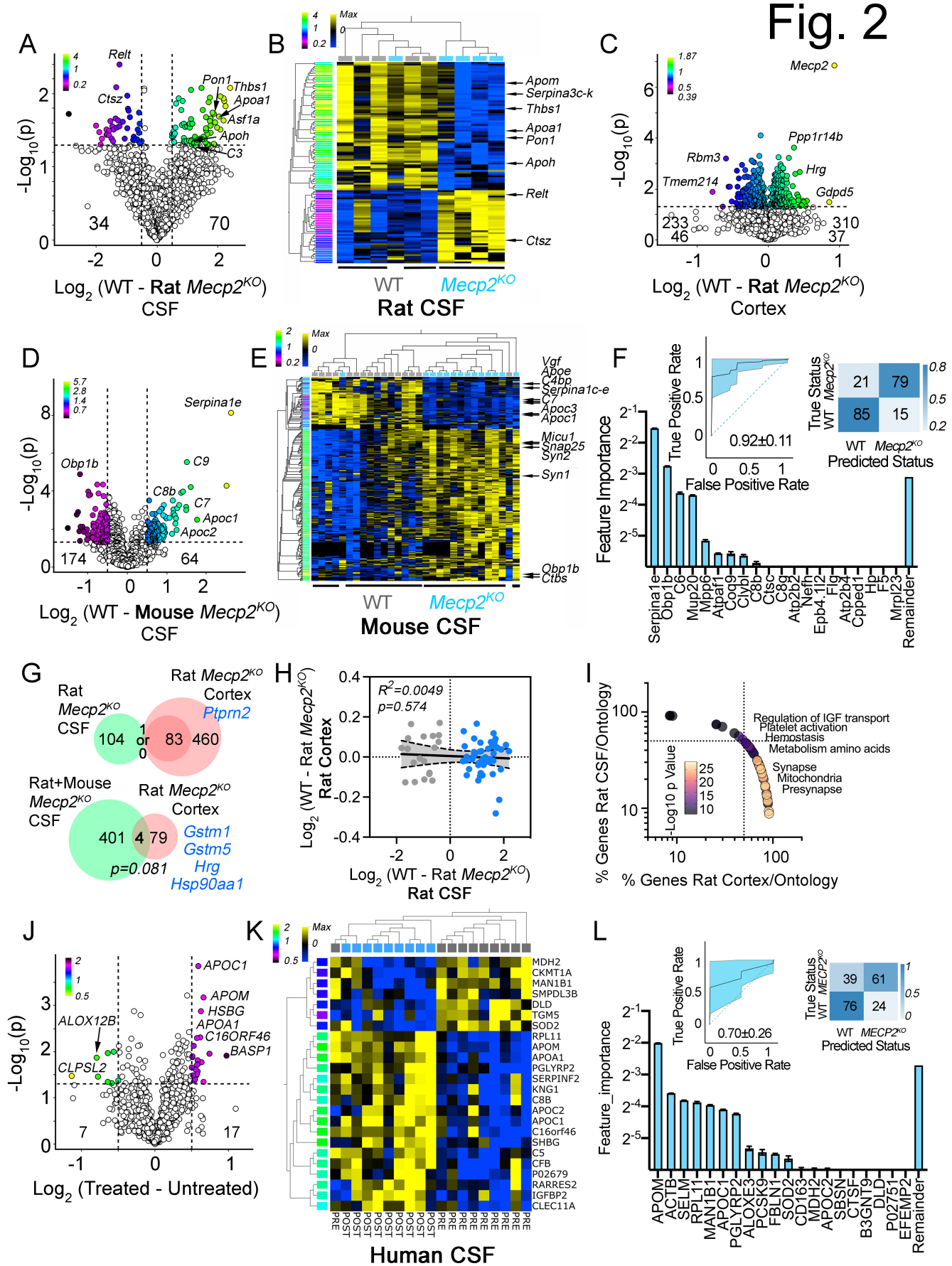


Fig. 3

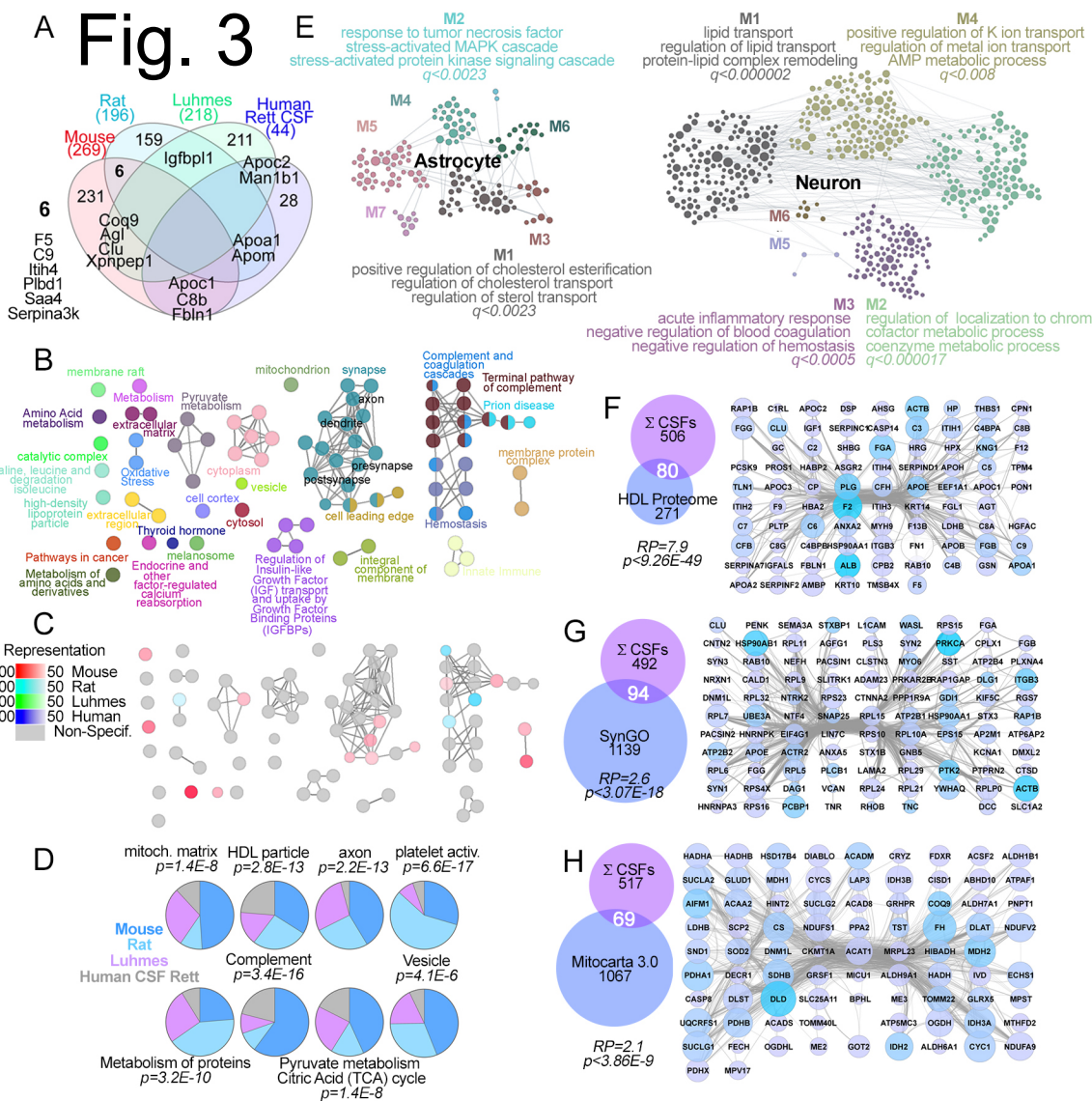


Fig. 4

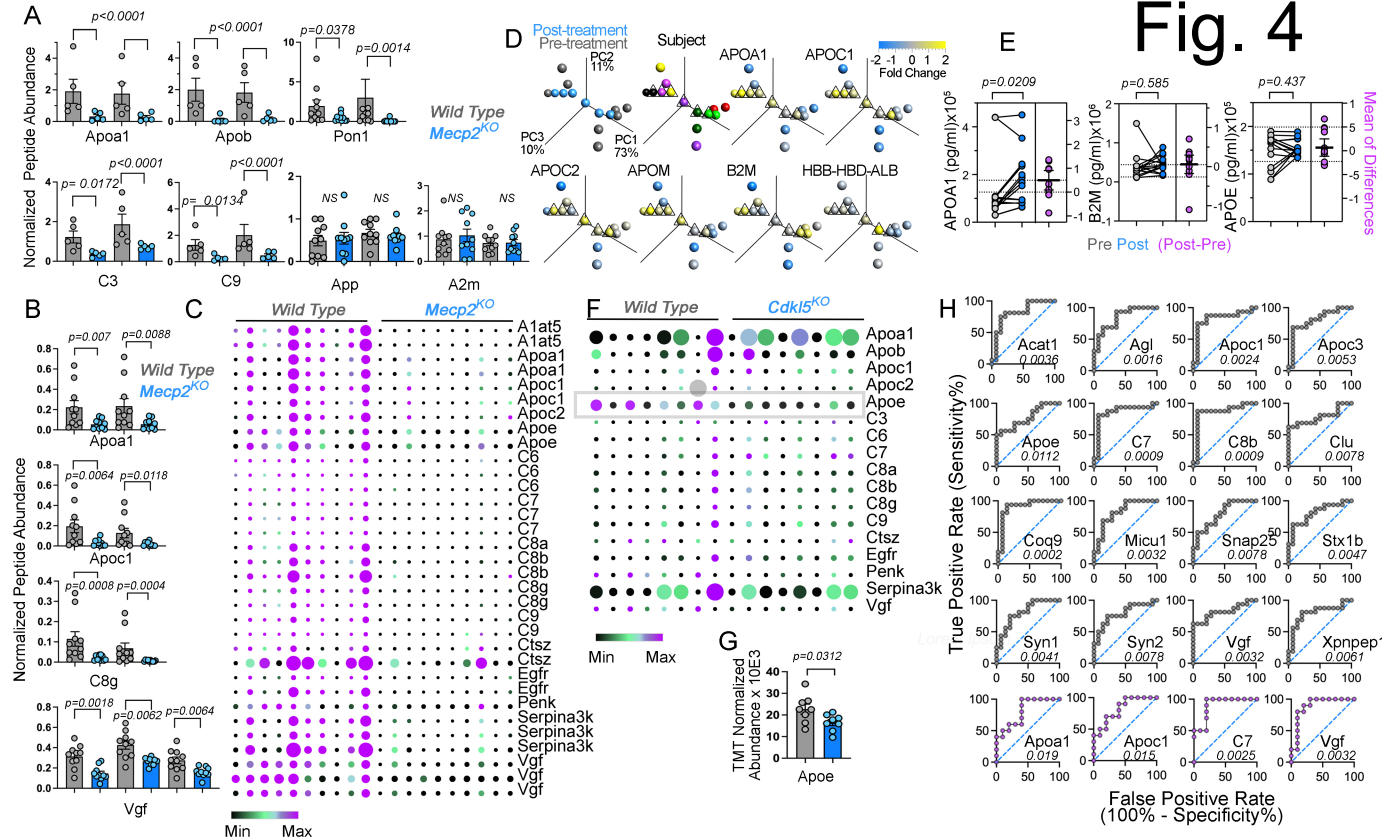


Fig. S1

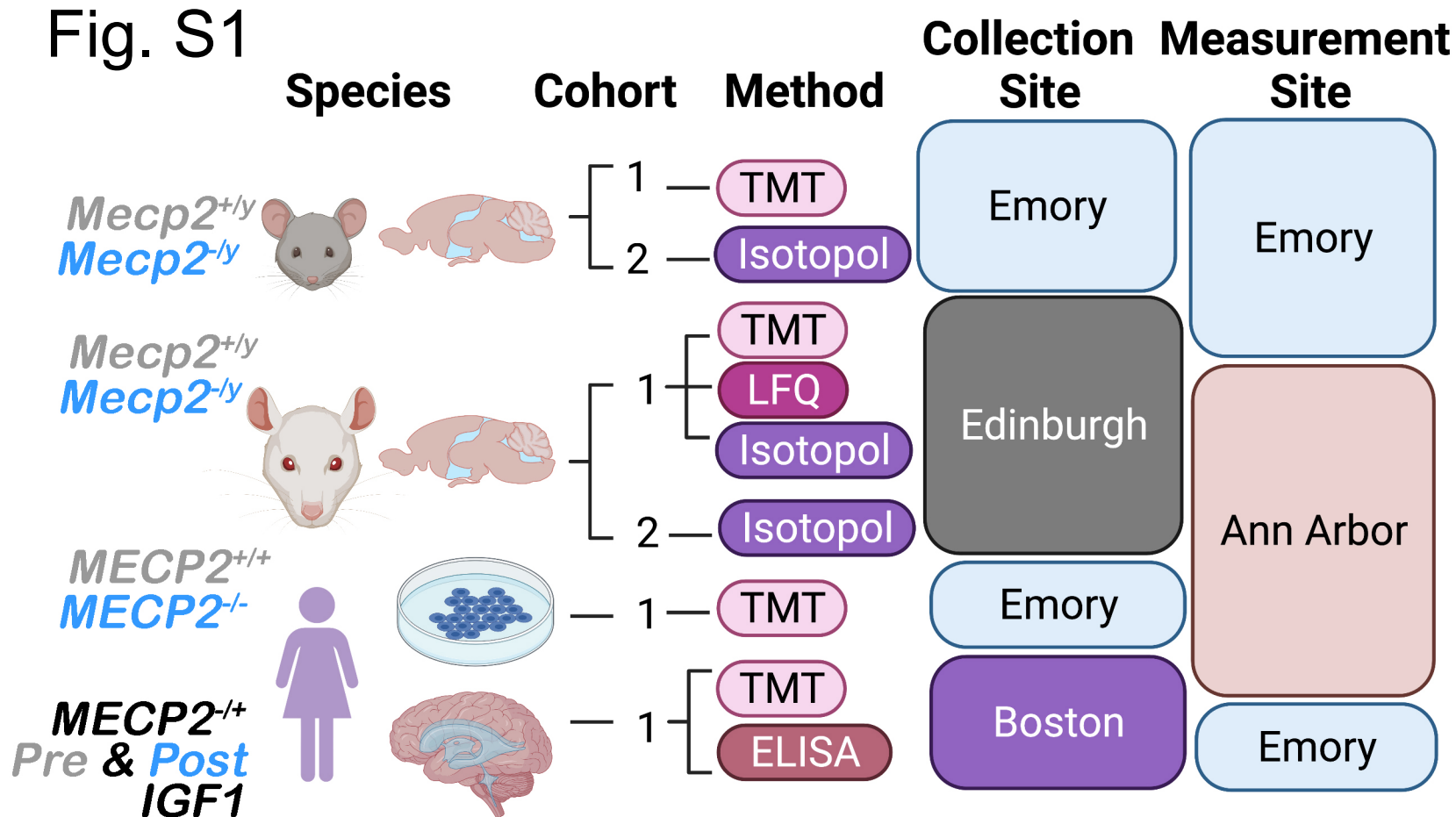


Fig. S2

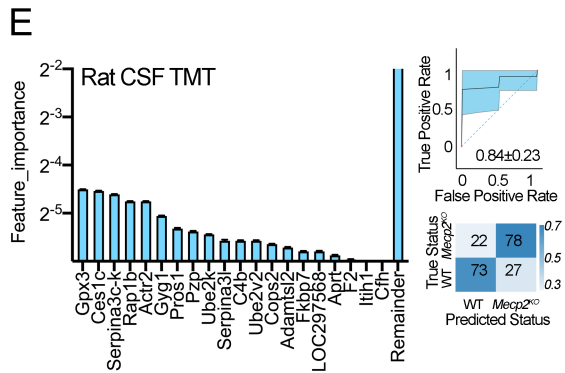
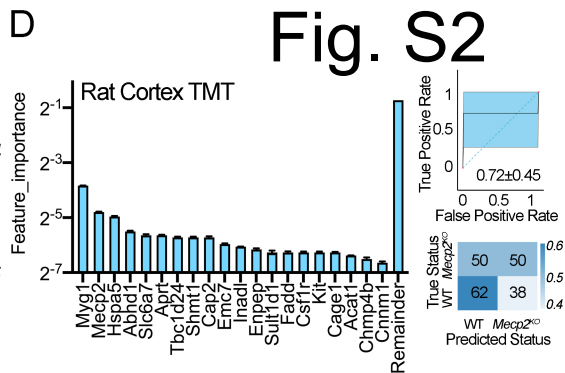
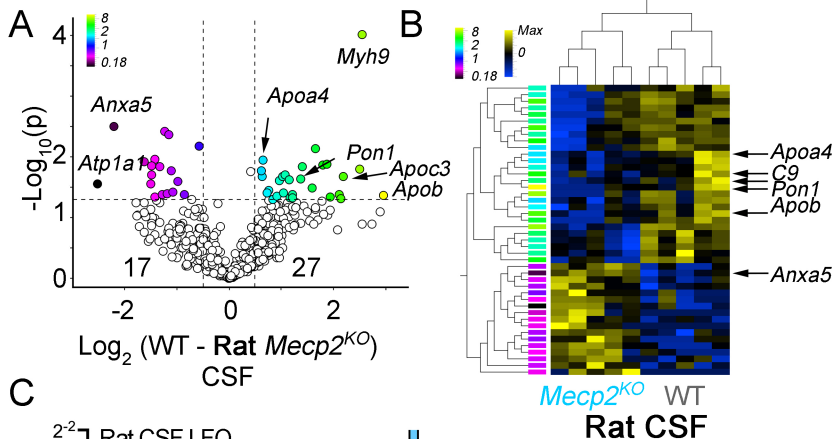


Fig. S3

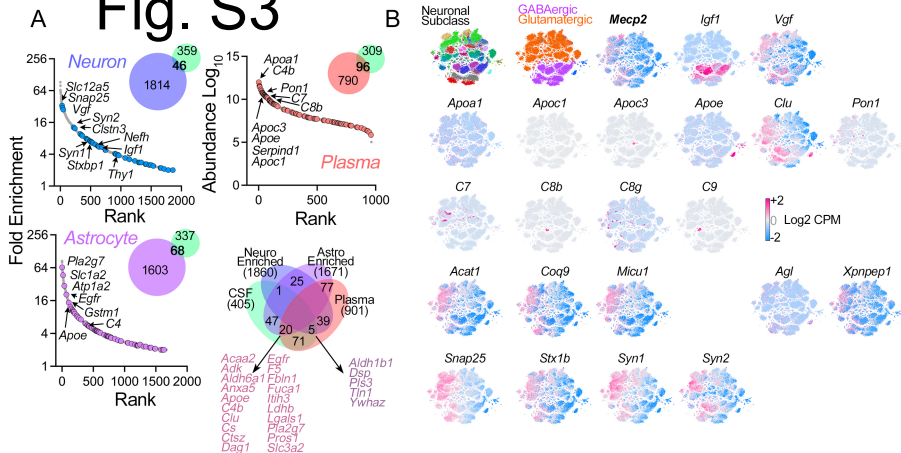


Fig. S4

Cortex
Hippocampus
Striatum

