scientific reports

OPEN

Screening autism-associated environmental factors in differentiating human neural progenitors with fractional factorial design-based transcriptomics

Abishek Arora[1,2], Martin Becker[1,2], Cátia Marques[3], Marika Oksanen[1,2], Danyang Li[1,2], Francesca Mastropasqua[1,2], Michelle Evelyn Watts[1,2], Manish Arora[4], Anna Falk[5,6], Carsten Oliver Daub[7,8], Ingela Lanekoff[3] & Kristiina Tammimies[1,2]✉

Research continues to identify genetic variation, environmental exposures, and their mixtures underlying different diseases and conditions. There is a need for screening methods to understand the molecular outcomes of such factors. Here, we investigate a highly efficient and multiplexable, fractional factorial experimental design (FFED) to study six environmental factors (lead, valproic acid, bisphenol A, ethanol, fluoxetine hydrochloride and zinc deficiency) and four human induced pluripotent stem cell line derived differentiating human neural progenitors. We showcase the FFED coupled with RNA-sequencing to identify the effects of low-grade exposures to these environmental factors and analyse the results in the context of autism spectrum disorder (ASD). We performed this after 5-day exposures on differentiating human neural progenitors accompanied by a layered analytical approach and detected several convergent and divergent, gene and pathway level responses. We revealed significant upregulation of pathways related to synaptic function and lipid metabolism following lead and fluoxetine exposure, respectively. Moreover, fluoxetine exposure elevated several fatty acids when validated using mass spectrometry-based metabolomics. Our study demonstrates that the FFED can be used for multiplexed transcriptomic analyses to detect relevant pathway-level changes in human neural development caused by low-grade environmental risk factors. Future studies will require multiple cell lines with different genetic backgrounds for characterising the effects of environmental exposures in ASD.

Intensified research continues to identify genetic variation, environmental exposures and their mixtures underlying different diseases and conditions. As several factors can be associated with the same outcome, there is a need for better methods to investigate the molecular, cellular, and developmental effects of these factors, both in parallel and as mixtures. At present, there is a growing need to develop in-vitro experimental models of environmental factors that recapitulate real-life exposures and clinical outcomes[1–3].

Fractional factorial experimental designs (FFED)[4], adapted from methods used in managerial and industrial processes, drastically reduce the number of samples required to observe effects with high statistical power. It is a subset of a full factorial design that increases experimental efficiency and makes it possible to study the

[1]Center of Neurodevelopmental Disorders (KIND), Centre for Psychiatry Research, Department of Women's and Children's Health, Karolinska Institutet, BioClinicum J9:30, Visionsgatan 4, 171 56 Solna, Stockholm, Sweden. [2]Astrid Lindgren Children's Hospital, Karolinska University Hospital, Region Stockholm, Stockholm, Sweden. [3]Department of Chemistry - BMC, Uppsala University, Uppsala, Sweden. [4]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, USA. [5]Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden. [6]Lund Stem Cell Center, Division of Neurobiology, Department of Experimental Medical Science, Lund University, Lund, Sweden. [7]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden. [8]Science for Life Laboratory, Stockholm, Sweden. ✉email: kristiina.tammimies@ki.se

interactions between different experimental conditions using statistical modelling. Earlier, it has not been coupled with omic readouts such as RNA sequencing to investigate any changes in the transcriptional landscape.

Autism spectrum disorder (ASD) is a neurodevelopmental disorder (NDD) diagnosed in nearly 1–2% of the population[5]. More than 100 genes are associated with ASD affected by both, common variants with small effect sizes and rare variants with larger effect sizes[5–7]. Heritability studies show that genetic factors account for 50–86.8% of ASD risk, suggesting that the remaining risk arises from environmental factors[8]. Early exposure to chemical stressors may induce or exacerbate neurodevelopmental trajectories underlying ASD. Environmental factors associated with ASD include lead (Pb)[9], valproic acid (VPA)[10,11], bisphenol A (BPA)[12,13], ethanol (EtOH)[14], fluoxetine hydrochloride (FH)[15] and zinc dysregulation[16,17]. The majority of genes and environmental factors associated with ASD have also been implicated in other NDDs, suggesting more generalised effects on neurodevelopment rather than being ASD specific.

Despite the growing interest in environmental factors, the molecular mechanisms leading to the increased liability of behavioural and cognitive difficulties remain largely unknown, especially in genetically vulnerable conditions. The use of induced pluripotent stem cell (iPSC) derived neural progenitors, either from typically developed individuals or individuals with ASD, can enable the study of the effects of environmental factors during neurodevelopment[18]. As a large number of both genetic and environmental factors have been indicated in ASD, there is a need for efficient methods to estimate their effects during early development.

Here, we present a multiplexed experimental design based on the FFED coupled with RNA-sequencing to perform simultaneous analyses of six environmental factors associated with ASD and four iPSC lines, two neurotypical controls and two with a known ASD genetic variant. The main outcomes were determined using transcriptional and pathway level analysis in differentiating neural progenitors from human iPSCs. The analyses were done at day 5 post induction of differentiation, to capture early but stable changes induced by the low-grade exposures. Furthermore, we validated the specific lipid pathways identified after exposure to FH using mass spectrometry-based metabolome detection. We show that FFED-RNA-seq can help pinpoint relevant mechanisms related to ASD-associated environmental exposures, and that there is a need to model these effects in multiple genetic backgrounds.

Results

Cellular effects of environmental factors.
We evaluated the molecular effects of Pb, VPA, BPA, EtOH, FH and zinc deficiency (Zn-) in cell lines with different clinical backgrounds for ASD during differentiation of human neural progenitors (Fig. 1A). The Non-ASD cell lines included two neurotypical controls (CTRL$_{Male}$, CTRL$_{Female}$), while the ASD cell lines included two males with a known ASD genetic variant (ASD$_{CASK}$, ASD$_{HNRNPU}$) as described earlier[19–22]. For studying the effects of the six exposures, we set up an FFED framework (Table S1A).

First, we performed a cytotoxicity assay to test the exposure ranges of the six selected environmental factors in differentiating human iPSC-derived neuroepithelial stem (NES) cells. NES cells are neural progenitors that are predominantly neurogenic in nature but also have the potential to be gliogenic, under defined differentiation conditions[20,23]. Exposure ranges were selected based on an extensive literature review as described in Table S1B–S1F and tested in the CTRL$_{Male}$ cell line. The highest concentration with no significant (Tukey post hoc $p > 0.05$) or trending decrease in cell viability and no gross changes in cell morphology at 120 h of exposure was selected for each environmental factor (Figure S1A); 3 µM for Pb, BPA and FH, and 3 mM for VPA and EtOH. This was done to avoid the toxicity of a single environmental factor overshadowing the combined effects of the other factors. Based on the six factors, we fitted the L8 Orthogonal Array based FFED for our study (Table S1A). The selected concentrations were then tested for cytotoxicity using the MTS assay in the FFED. At 120 h of exposure, a significant difference (Tukey post hoc $p < 0.05$) in cell viability was only observed for Pb, VPA and EtOH (Figure S1B). Based on the observed cytotoxicity effects of VPA on gross cellular morphology during NES cell differentiation in the FFED, the exposure concentration was adjusted to 3.0 µM in line with previous in-vitro exposures reported in the literature (Table S1D).

Next, we tested if the exposures affected the population of neural progenitors in two of the four cell lines, CTRL$_{Male}$ and ASD$_{CASK}$, by analysing the percentage of SOX2 positive cells at day 5 of exposure using the FFED. No significant changes (Tukey post hoc $p > 0.05$) in the number of SOX2 positive cells were detected for any of the exposures (Figure S1C). Additionally, we tested for any proliferation changes using the BrdU assay and FFED at day 5 of exposure in all the cell lines. A significant decrease in proliferation was detected after exposure to Pb in the ASD$_{CASK}$ cell line (Tukey post hoc $p < 0.001$, Figure S1D).

Herein, the cell viability testing enabled us to select the optimal exposure concentrations of the chosen environmental factors to be analysed in our model system (3 µM for Pb, VPA, BPA and FH, and 3 mM for EtOH). No changes in SOX2 positivity were detected following the exposures and only Pb reduced cell proliferation in ASD$_{CASK}$.

Approaching multiplexed exposures and gene expression.
We investigated if FFED coupled with RNA-seq could detect relevant transcriptomic changes after low-grade environmental exposures on differentiating human neural progenitors for 5 days. We performed RNA-seq using the L8 orthogonal array for the four cell lines (n = 32). We included a biological replicate of the full set for CTRL$_{Male}$ (n = 8) and a partial set for CTRL$_{Female}$ (n = 4), for variability and sample size testing. Principal component analysis (PCA) of the RNA-seq data (N = 43) showed that the genetic background of the samples was the main driver of the differences across PC1 (46.12%) and PC2 (33.57%), and the exposure effects were minimal for the total variation in the transcriptomic profiles (Figure S2A and S2B).
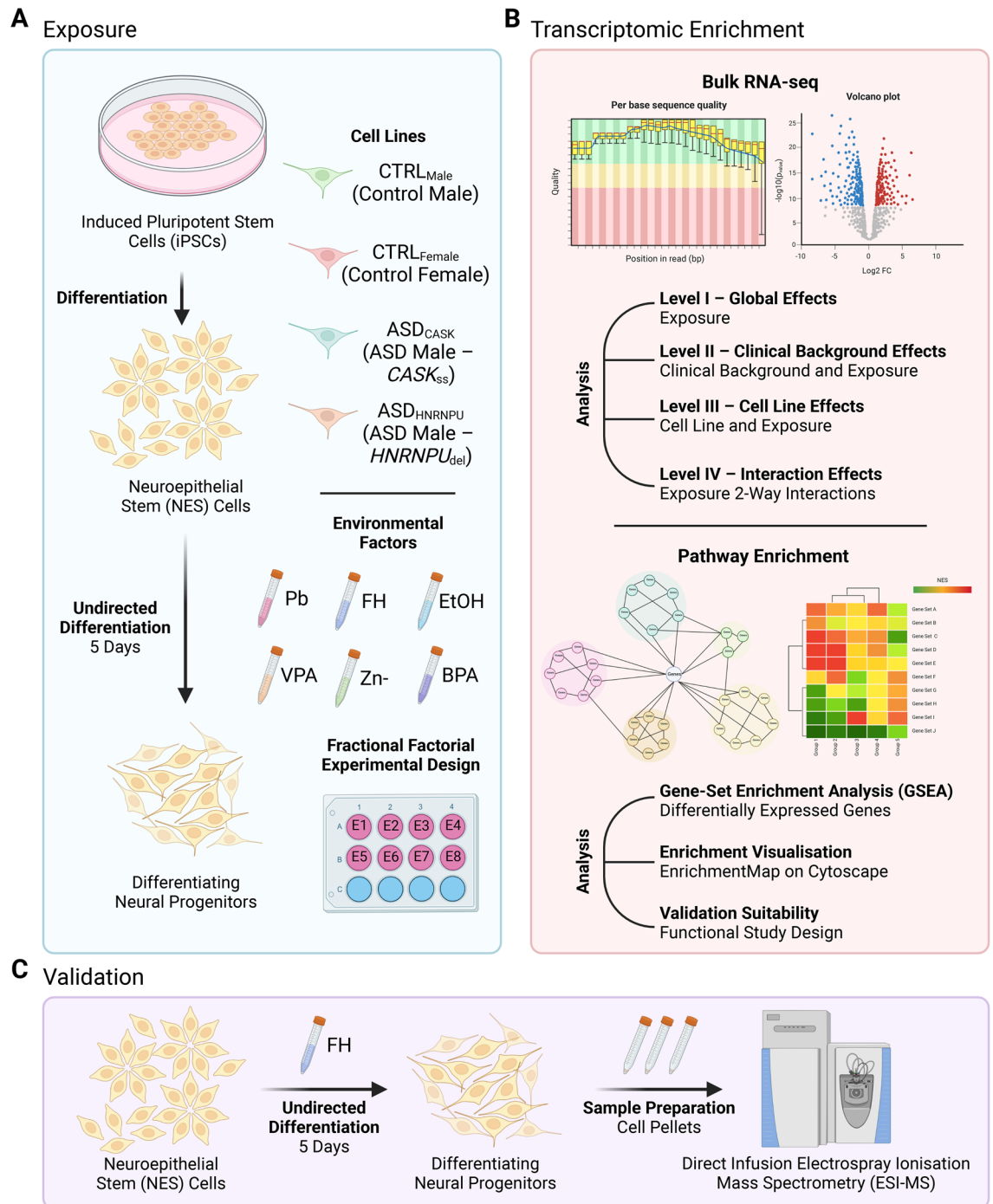
Figure 1. Overview of study plan using fractional factorial experimental design (FFED). (**A**) Exposure of neuroepithelial stem (NES) cells derived from human induced pluripotent stem cells (iPSCs) during neural progenitor differentiation for 5 days. Four cell lines, neurotypical controls (CTRL_{Male}, CTRL_{Female}) and males with autism spectrum disorder (ASD) diagnoses (ASD_{CASK}: *CASK* splice site variant, ASD_{HNRNPU}: *HNRNPU* deletion), were exposed to six environmental factors during differentiation, namely lead (Pb), fluoxetine hydrochloride (FH), ethanol (EtOH), valproic acid (VPA), bisphenol A (BPA) and zinc deficiency (Zn-). (**B**) Using RNA samples from (**A**), bulk RNA-sequencing was performed followed by differential gene expression analysis for global effects, clinical background effects, cell line effects and interaction effects. Using this, pathway enrichment and network visualisation was done. (**C**) Enriched pathways identified from (**B**) following FH exposure were detected using direct infusion electrospray ionisation mass spectrometry (ESI–MS) and quantified across the different levels of analyses. (Created with BioRender.com).

We then performed identification of differentially expressed genes (DEGs) and gene-set enrichment analysis (GSEA) for enriched pathways at three levels. Briefly, at *Level I—Global Effects*, exposure effects were analysed across all cell lines; *Level—II, Clinical Background Effects*, exposure comparisons were made separately for the non-ASD and ASD cell lines; and *Level III—Cell Line Effects*, exposure effects were analysed for every individual cell line (Fig. 1B).

In short, the genetic background of the cell lines included in this study had a larger effect than the environmental factor exposures themselves on the detected transcriptomic responses.

Early exposure effects of environmental factors. In the global analyses (*Level I*), no genes responding to BPA, VPA, EtOH and Zn- exposures (adj. $p > 0.05$) were identified. Nevertheless, pathway analysis highlighted biological processes following these exposures (Table S3A and S3B). For BPA, eleven pathways were significantly upregulated, including protein lipid complex assembly ($q = 0.0027$) and mitochondrial membrane organisation ($q = 0.032$), and one was significantly downregulated, detection of chemical stimulation involved in sensory perception ($q = 0.035$). For VPA, thirty processes were significantly downregulated, including neuron projection guidance ($q = 0.0023$) and axon development ($q = 0.0089$). For EtOH and Zn-, negative regulation of endothelial cell apoptotic process ($q = 0.0017$) and positive chemotaxis ($q = 0.048$) were the only significantly upregulated pathways, respectively. Additionally, for Zn- three were significantly downregulated: plasminogen activation ($q = 0.0006$), protein activation cascade ($q = 0.0048$) and fibrinolysis ($q = 0.0066$).

We detected significant changes in genes responding to Pb (69 genes, adj. $p < 0.05$, Table S2A, Fig. 2A) and FH (50 genes, adj. $p < 0.05$, Table S2H, Fig. 2C). Furthermore, pathway analysis revealed several biological processes affected by the exposures (Fig. 2B and 2D). Amongst the significantly upregulated pathways after Pb exposure
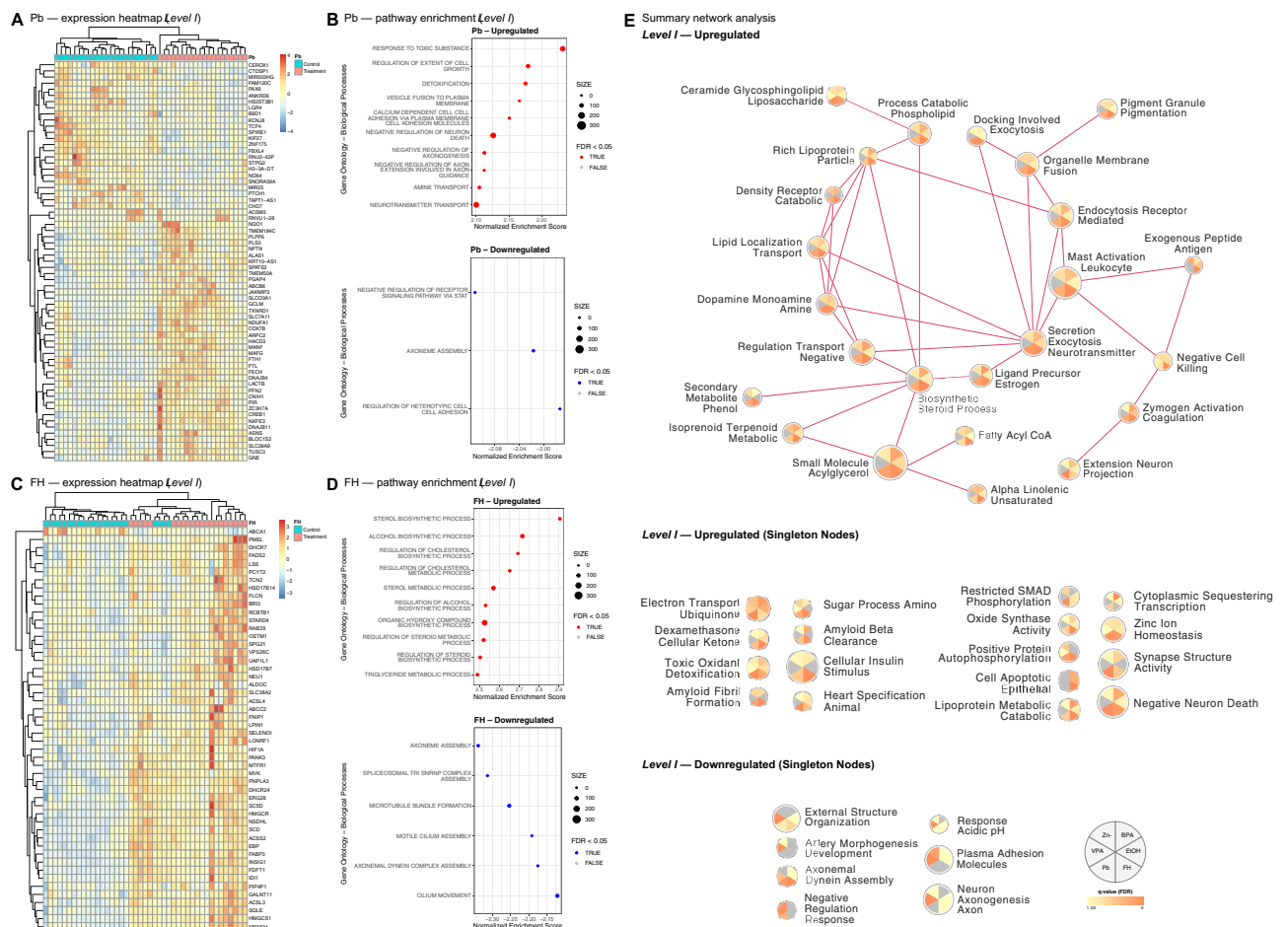
Figure 2. Differential gene expression and pathway enrichment—*Level I* analysis of global effects following environmental factor exposure. (**A**) Heatmap of significant (adj. $p < 0.05$) differentially expressed genes following lead (Pb) exposure for 5 days. (**B**) Gene Ontology (GO) enrichment analysis with top 10 significantly (FDR < 0.05) upregulated and downregulated biological processes following Pb exposure for 5 days. (**C**) Heatmap of significant (adj. $p < 0.05$) differentially expressed genes following fluoxetine (FH) exposure for 5 days. (**D**) GO enrichment analysis with top 10 significantly (FDR < 0.05) upregulated and downregulated biological processes following FH exposure for 5 days. (**E**) Summary network of upregulated and downregulated clusters enriched for day 5 exposure of Pb, FH, ethanol (EtOH), valproic acid (VPA), bisphenol A (BPA) and zinc deficiency (Zn-).

(Table S3A) were cholinergic synaptic transmission (q = 0.0099), axon extension (q = 0.012), and synapse assembly (q = 0.012). Those significantly downregulated (Table S3B) were negative regulation of receptor signalling pathway via STAT (q = 0.0091), regulation of heterotypic cell adhesion (q = 0.043) and axoneme assembly (q = 0.046). For FH (Table S3A), the significantly upregulated pathways included alcohol biosynthetic process (q < 0.0001), sterol biosynthetic process (q < 0.0001), and regulation of steroid metabolic process (q < 0.0001). There were six that were significantly downregulated (Table S3B), including motile cilium assembly (q = 0.015), axonemal dynein complex assembly (q = 0.018) and cilium movement (q = 0.035).

To assess the involvement of the genes responding to Pb and FH exposures across spatio-temporal trajectories during human neurodevelopment, we analysed their expression using available RNA-seq data from the BrainSpan Consortium[24]. The genes responding to Pb (Figure S2C) and FH (Figure S2E) both formed two groups for early and late neurodevelopment. Both Pb (Figure S2D) and FH (Figure S2F) gene groups were significantly (adj. $p < 0.05$) detected in the frontal cortex, temporo-parietal cortex, sensorimotor cortex, and subcortical regions. When considering their role across developmental periods, Pb genes (early) were significantly identified (adj. $p < 0.05$) in all brain regions from the late mid-foetal period (19–24 weeks), other than in the subcortical region that was only significant (adj. $p < 0.05$) from the early foetal period (0–12 weeks). For FH (late), the genes were significantly found (adj. $p < 0.05$) in all studied brain regions, however, were limited to postnatal periods of neurodevelopment.

Additionally, we tested for enrichment of the responding genes in neurodevelopmental disorder gene lists, including SFARI for ASD[25], Genetic epilepsy syndromes (EPI) panel[26], intellectual disability (ID) panel[26] and a general developmental gene list[27]. No significant findings were observed (adj. $p > 0.05$).

To summarise, gene level responses were only detected following Pb and FH exposures, while pathway level responses could be detected for all environmental factors.

Convergence of affected molecular pathways for all exposures.

Several shared biological processes were revealed, when we analysed the molecular pathways across environmental factor exposures (*Level I* analysis). There was significant upregulation detected (Table S3A) in glycosphingolipid metabolic processes (FH q = 0.034, Pb q = 0.023), steroid catabolic processes (FH q = 0.0003, Pb q = 0.016), terpenoid metabolic processes (BPA q = 0.046, FH q = 0.031), negative regulation of endothelial cell apoptosis (EtOH q = 0.0017, FH q = 0.003), synaptic vesicle priming (FH q = 0.014, Pb q = 0.018) and cholinergic synaptic transmission (FH q = 0.015, Pb q = 0.01). Those that were significantly downregulated (Table S3B) were axoneme assembly (FH q = 0.0008, Pb q = 0.046) and microtubule bundle formation (FH q = 0.0065, Pb q = 0.05). The summary network of the six exposures, indicated connected nodes across the exposures for several of the upregulated clusters and only singleton nodes for the downregulation (Fig. 2E, Table S4A and S4B).

Based on our pathway level analyses, we demonstrated that the environmental exposures resulted in systemic responses of which many were shared physiological pathways and/or pathological effects.
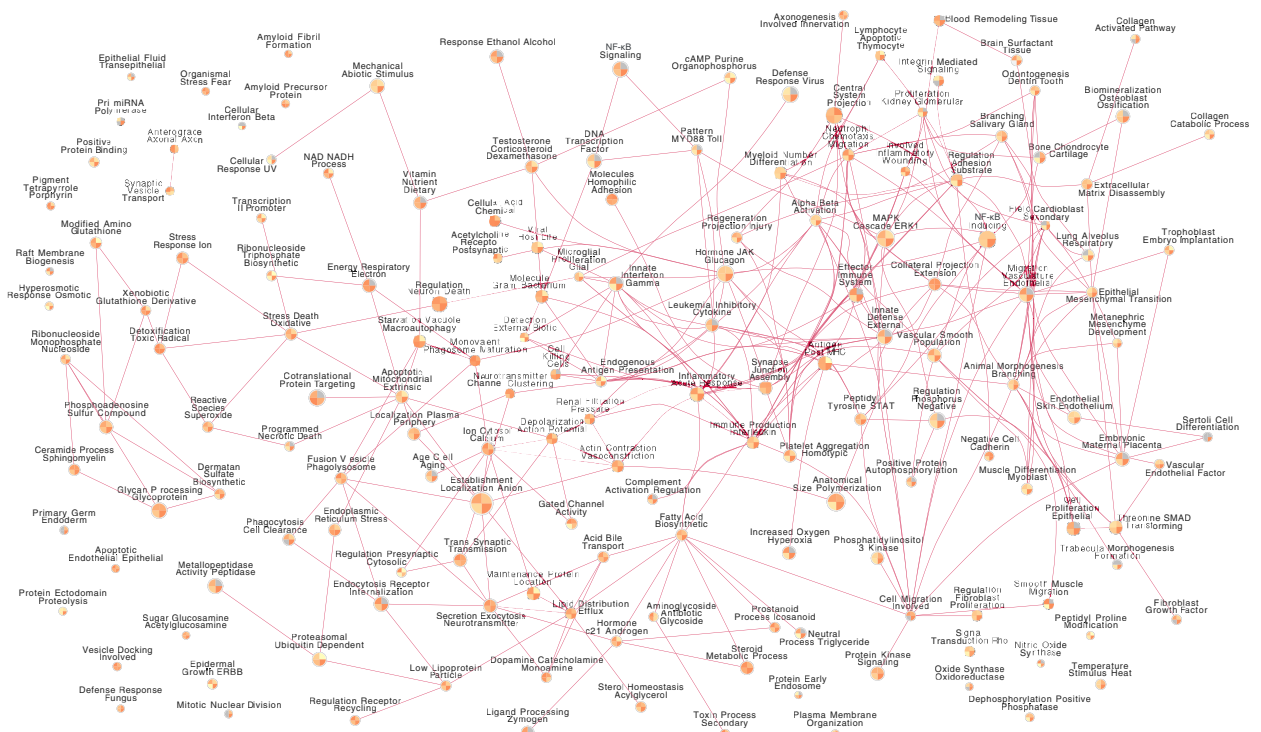
Clinical background modulates exposure effects.

Similar pathway analyses were performed for *Level—II, Clinical Background Effects,* by analysing the Non-ASD cells (CTRL$_{Male}$ and CTRL$_{Female}$; Table S2B and S2I) and ASD cells (ASD$_{CASK}$ and ASD$_{HNRNPU}$; Table S2C and S2J) separately. The environmental factor exposures uniquely and commonly modulated several pathways. A total of 460 biological processes were upregulated (Table S3C) and 79 were downregulated (Table S3D) in the Non-ASD cells. In the ASD cells, 861 pathways were upregulated (Table S3E), while 17 were downregulated (Table S3F). The summary networks were used to visualize hub nodes (Table S4C and S4D, Figure S3A and S3C for the Non-ASD cell lines; Table S4E and S4F, Figure S3F and S3D for the ASD cell lines).

Furthermore, we analysed the global network properties of the Non-ASD and ASD cell responses to the environmental factor exposures to represent the differential responses. In the upregulation network for Pb (Table S3G), the diameter measuring the network size, was greater for the ASD cell lines (17.0) than the Non-ASD cell lines (7.0). In contrast, the clustering coefficient indicating the nodal neighbourhood connectivity was higher of the Non-ASD cell lines (0.68) than the ASD cell lines (0.54). The average degree, representing the number of edges per node in a network, was conversely higher in the ASD cell lines (8.06) than the Non-ASD cell lines (6.67). Interestingly, no downregulation network could be generated for the ASD cell lines based the selected thresholds (q < 0.05). For the upregulation network of FH (Table S3G), the diameter of the Non-ASD cell lines (18.0) was larger than the ASD cell lines (4.0). The clustering coefficient was greater for the ASD cell lines (0.95) than for the Non-ASD cell lines (0.59), while the average degree was higher for the Non-ASD cell lines (7.53) when compared to the ASD cell lines (5.72). Similar to Pb, no downregulation network could be generated for FH exposure in the ASD cell lines.
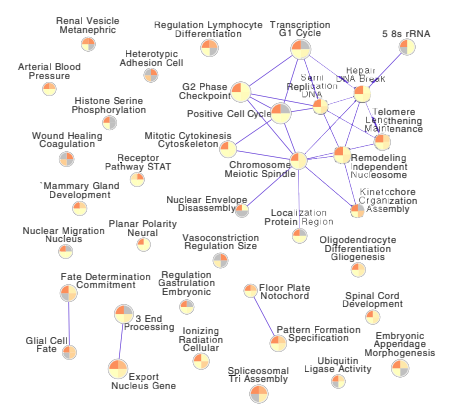
Lastly, we performed the *Level III* analyses for each cell line (Table S2D–S2G and S2K–S2N). Several pathways were found to be significantly modulated (Table S5). Interestingly, the exposures per cell line regulated several unique biological processes. For Pb, significantly upregulated (q < 0.05) unique processes (Table S5A), were indicated for all cell lines, however ASD$_{HNRNPU}$ showed the largest effects with 1105 of 1274 that were significantly upregulated. While, CTRL$_{Male}$ had the largest effects that were Pb-driven and significantly downregulated (q < 0.05) (Table S5B) with 266 of 290 pathways. FH induced significant upregulation (q < 0.05) was largely detected in CTRL$_{Male}$ (250 of 378) and none in ASD$_{CASK}$ (Table S5C). FH-driven significant downregulation (q < 0.05) of pathways was identified mostly in CTRL$_{Male}$ (75 of 85) with a few in CTRL$_{Female}$ (13 of 71), such as spliceosomal snRNP assembly (CTRL$_{Male}$ q < 0.0001) and cilium movement (CTRL$_{Female}$ q = 0.0002) (Table S5D).

Pb and FH exposure effects on the cell lines were visualised using summary network analysis for upregulated (Fig. 3A and 3C, Table S4G and S4I, respectively) and downregulated (Fig. 3B and 3D, Table S4H and S4J, respectively) common clusters. The upregulated clusters (Fig. 3A, Table S4G), significantly altered after
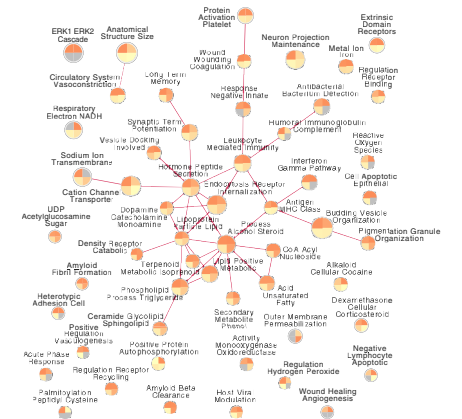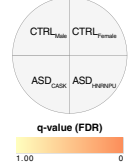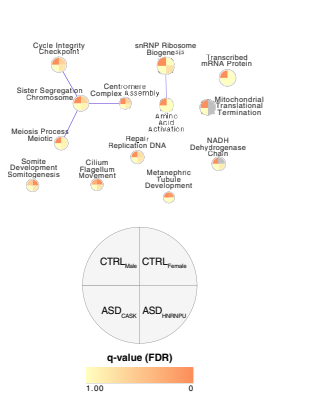
Figure 3. Summary networks from pathway enrichment—*Level III* analysis of cell line effects following lead (Pb) and fluoxetine (FH) exposure. (**A**) Upregulated and (**B**) downregulated clusters enriched for day 5 exposure of Pb in CTRL$_{Male}$, CTRL$_{Female}$, ASD$_{HNRNPU}$ and ASD$_{CASK}$. (**C**) Upregulated and (**D**) downregulated clusters enriched for day 5 exposure of FH in CTRL$_{Male}$, CTRL$_{Female}$, ASD$_{HNRNPU}$ and ASD$_{CASK}$.

Pb exposure and common between at least two cell lines included detoxification toxic radical (ASD$_{HNRNPU}$ q = 0.0099, ASD$_{CASK}$ q = 0.0472, CTRL$_{Male}$ q = 0.0325), complement activation regulation (ASD$_{HNRNPU}$ q = 0.0183, CTRL$_{Male}$ q = 0.0192) and xenobiotic glutathione derivative (ASD$_{HNRNPU}$ q = 0.0296, CTRL$_{Male}$ q = 0.0467). Some that were downregulated (Fig. 3B, Table S4H) and unique to CTRL$_{Male}$ or CTRL$_{Female}$ were also detected, several of which were related to cell cycle regulation. For FH, the upregulated clusters (Fig. 3C, Table S4I), significant and common amongst at least two cell lines included alcohol steroid process (CTRL$_{Male}$ q = 0.0312, CTRL$_{Female}$ q = 0.0083), receptor recycling regulation (CTRL$_{Male}$ q = 0.0240, CTRL$_{Female}$ q = 0.0403), and ERK1 ERK2 cascade (CTRL$_{Male}$ q = 0.0238, CTRL$_{Female}$ q = 0.0114). Those that were unique and downregulated (Fig. 3D, Table S4J) were also identified, like cilium flagellum movement (CTRL$_{Female}$ q = 0.0197), DNA replication repair (CTRL$_{Male}$ q = 0.0146), and sister chromosome segregation (CTRL$_{Male}$ q = 0.0184).

BPA predominantly modulated pathways with significance (q < 0.05) in CTRL$_{Male}$ (upregulated 138 of 139, downregulated 1 of 1) (Table S5E). On the other hand, EtOH and VPA predominantly upregulated (Table S5I and S5G, respectively) those with significance (q < 0.05) in CTRL$_{Female}$ (EtOH: 67 of 72, VPA: 5 of 5). VPA-driven significant downregulation (q < 0.05) (Table S5H) was detected mainly in CTRL$_{Female}$ (37 of 84) and ASD$_{CASK}$

(51 of 84). Zn- significantly upregulated pathways (Table S5J) mainly in CTRL$_{Male}$ (21 of 34), such as ribosome biogenesis (CTRL$_{Male}$ q = 0.0133) and translation initiation (CTRL$_{Male}$ q = 0.0332). Those that were significantly downregulated (Table S5K) like protein activation cascade (CTRL$_{Female}$ q < 0.0001) and sterol import (CTRL$_{Female}$ q = 0.0248), were primarily detected in CTRL$_{Female}$ (34 of 40).

In summary, several shared as well as unique biological processes were detected when selecting based on clinical background or individual cell lines. While these were predominantly mediated by Pb and FH, small effects were also detected for BPA and VPA. Differences were also observed based on the properties of the generated networks.

Detecting interaction effects with FFED.

In addition to single exposures, the FFED design enabled exposure interaction analyses. In our design, we were able to analyse three two-way interactions (*Level IV* analyses): Pb–Zn-, VPA–FH, and BPA–EtOH. However, interaction modulated genes were only detected in CTRL$_{Male}$ when analysing for interaction effects. This could be attributed to the presence of two full biological replicates of the FFED design for CTRL$_{Male}$ (Table S1A). Several biological processes were modulated by the two-way interactions (Table S6A and S6B) and in the summary network several clusters were identified. Significant upregulation (Table S6C, Figure S3B) was detected in nucleosome organization assembly (BPA–EtOH q = 0.035), mitochondrial electron respiration (Pb–Zn- q = 0.015), membrane raft organization (Pb–Zn- q = 0.040), lipoprotein density particle (VPA–FH q = 0.023) and steroid biosynthetic process (VPA–FH q = 0.034). Significant downregulation (Table S6D, Figure S3E) was detected in clusters such as G1 phase cell cycle (Pb–Zn- q = 0.015), chromatin remodelling (Pb–Zn- q = 0.0023), kinetochore organization assembly (Pb–Zn- q = 0.0034, VPA–FH q = 0.015), and snRNP biogenesis (VPA–FH q = 0.0071).

We were able to detect interaction effects in the CTRL$_{Male}$ for the specific combinations of exposures. For future interaction studies, at least two biological replicates in an L8 orthogonal array based FFED are needed to detect interaction effects.

Lead and Bisphenol A induce alternative splicing events.

After investigating gene response differences, we performed differential exon usage (DEU) analysis on the *Level I* to *III* comparisons (Table S7A). In line with the gene response analysis, Pb exposure presented specific effects on alternative splicing. Pb exposure induced more alternative splicing events in the ASD cell lines (45 genes) compared to the Non-ASD cell lines (4 genes). In addition to Pb, BPA also affected alternative splicing. BPA exposure caused extensive alternative splicing events on the global level (223 genes) and in the Non-ASD cell lines (346 genes). Furthermore, in the female control cell line, CTRL$_{Female}$, alternative splicing events were detected following analysis in both the full set of samples (381 genes) and excluded partial set (868 genes). No alternative splicing events were detected after exposure to the other environmental factors.

The genes with alternative splicing events in CTRL$_{Female}$ after BPA exposure (partial sample set, 868 genes) were over-represented in several biological processes (Table S7B, Figure S3G). These included cytoskeleton organisation (adj. *p* = 0.0091), neuron development (adj. *p* = 0.036), regulation of RNA splicing (adj. *p* = 0.0359), neuron differentiation (adj. *p* = 0.048) and intracellular protein transport (adj. *p* = 0.044). No significant enrichment (adj. *p* < 0.05) in biological processes were detected for the genes driving alternative splicing events following Pb exposure at all levels of analyses.

Furthermore, genes with significant alternative splicing events following BPA exposure in CTRL$_{Female}$ were tested for enrichment in publicly available gene lists, as previously done for significantly responding genes. Significant enrichment was observed in the SFARI gene list (adj. *p* = 0.045), SFARI high confidence gene list (adj. *p* = 0.022), ID panel (adj. *p* = 0.0045) and developmental gene list (adj. *p* = 0.0013). No significant enrichment (adj. *p* > 0.05) was detected in the EPI panel.

In our DEU level analyses, we indicated that both Pb and BPA mediated alternative splicing events in the study samples, however these were not enriched in any biological processes. At the cell line level, BPA predominantly drove the effects in CTRL$_{Female}$ with enrichment in processes that regulate neuronal physiology and in publicly available gene lists relevant to ASD.

Fluoxetine alters lipid metabolism during neural progenitor differentiation.

As we revealed an extensive dysregulation of lipid metabolism related pathways in the transcriptomic profiles after FH exposures, we set out to validate the findings using direct infusion electrospray ionisation mass spectrometry (ESI–MS) for metabolomics (Fig. 1C). We initially evaluated over 2500 metabolite signals, including adducts, for FH exposure effects in all the cell lines, out of which 79 were selected and reported based on their signal threshold level. At the *Level I* analysis for global effects, we could confirm several lipid-based metabolites to be significantly elevated after exposure to FH, including fatty acids (FA 18:1, adj. *p* = 0.0015; FA 20:2, adj. *p* = 0.0056; FA 20:3, adj. *p* = 0.0040), plasmanylcholines (PC e 32:0, adj. *p* = 0.0006; PC e 32:1, adj. *p* = 0.0003) and the phosphatidylethanolamine, PE 36:4 (adj. *p* = 0.0062) (Fig. 4A, Table S8A).

When separating for the clinical background (*Level* II) effects, several shared metabolites showed a significant elevation following FH exposure, however, with a difference in magnitude of concentrations, including fatty acids FA 20:2 (Non-ASD q = 0.0217; ASD q = 0.0320), FA 20:3 (Non-ASD q = 0.0134, ASD q = 0.0198); phosphatidylcholines PC 34:3 (Non-ASD q = 0.0245, ASD q = 0.0313), PC 36:3 (Non-ASD q = 0.0246, ASD q = 0.0432) and phosphatidylethanolamines PE 34:2 (Non-ASD q = 0.0032, ASD q = 0.0007), and plasmanylethanolamine PE e 38:4 (non-ASD q = 0.0189, ASD q = 0.0022) (Fig. 4D, Table S8B and S8C).

Furthermore, we detected several metabolites that were elevated significantly only in either of the clinical backgrounds. For the Non-ASD cell lines, additional 11 metabolites showed significant changes, including aspartic acid (q = 0.0436), PC e 32:0 (adj. *p* = 0.0003), PE 34:2 (adj. *p* = 0.0033) and PE e 40:5 (adj. *p* = 0.0072) (Fig. 4B,

7
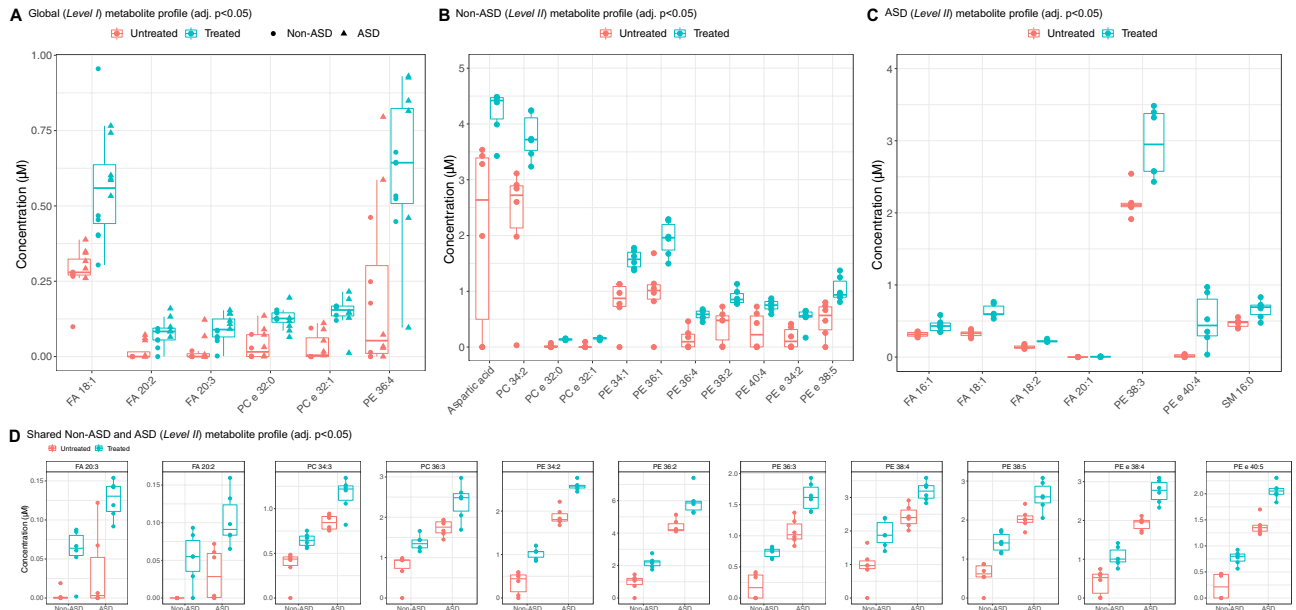
Figure 4. Metabolic changes induced by fluoxetine (FH) exposure. (**A**) Significantly enriched metabolites (adj. $p < 0.05$) from day 5 FH exposure for global effects (*Level I* analysis). (**B**) Significantly enriched (adj. $p < 0.05$) and unique metabolites from day 5 FH exposure, in the non-ASD cell lines (*Level II* analysis). (**C**) Significantly enriched (adj. $p < 0.05$) and unique metabolites from day 5 FH exposure, in the ASD cell lines (*Level II* analysis). (**D**) Significantly enriched (adj. $p < 0.05$) and common metabolites from day 5 FH exposure, in the non-ASD and ASD cell lines (*Level II* analysis).

Table S8B). For the ASD cell lines, seven unique metabolites showed significant changes, including FA 18:2 (adj. $p = 0.0009$), PE 38:3 (adj. $p = 0.0280$), PE e 40:4 (adj. $p = 0.0426$) and SM 16:0 (adj. $p = 0.0409$) (Fig. 4C, Table S8C).

At the *Level III* analysis for cell line effects, in addition to some of the earlier significant metabolites, few more metabolites were also significantly elevated. In CTRL$_{Male}$ (Table S8D), PC e 32:0 (adj. $p = 0.020$) and PC e 32.1 (adj. $p = 0.0068$) were elevated upon FH exposure, while no metabolites were significantly elevated in CTRL$_{Female}$ (Table S8E). In ASD$_{HNRNPU}$ (Table S8F), significant elevation was detected in six metabolites including FA 20:3 (adj. $p = 0.033$), PE 34:2 (adj. $p = 0.033$) and PC 34:4 (adj. $p = 0.048$). Five additional metabolites were significantly elevated in ASD$_{CASK}$ (Table S8G) including PE e 38:4 (adj. $p = 0.012$), PE 38:3 (adj. $p = 0.0010$) and PE 36:4 (adj. $p = 0.0015$).

Based on the above-mentioned results, we can summarise that FH exposure elevated the levels of several lipid-based metabolites in relation to their baseline levels detected in all comparisons in this study. This is turn could have several effects on cellular physiology.

Discussion

As the identification of both genetic and environmental factors in different conditions such as ASD and other NDDs is increasing, there is a need for efficient, multiplexable platforms to model their effects and interactions during differentiation of neural progenitors. Here, we describe our study using FFED coupled with RNA-seq to identify affected biological pathways in a human iPSC derived model of early neurodevelopment, followed by a layered analytical workflow. We find modest but significant changes at the pathway level after low-grade exposure with six selected environmental factors and showed the feasibility of separating the effects at different levels of complexity, vis-à-vis global, clinical background, and cell line effects, independently and as two-way interactions.

A means to highlight the efficiency and utility of the FFED would be to compare the input sample count in similar transcriptomic studies. While we performed RNA-seq on 32 samples and in-depth analysis to study the effects of six environmental factors, a theoretical full design study would require 120 samples to study 6 factors in 4 cell lines using 5 technical replicates. Similarly, a conventional experimental design to investigate the effects of Pb on human neural progenitors at two different concentrations and multiple time-points used 60 samples[28]. A recent report analysing the effects of a mixture of endocrine disrupting chemicals on neurodevelopment and language delay in human cerebral organoids and in-vivo models, generated vast insights on perturbed regulatory networks[2]. The FFED, would not only make it possible to analyse such interaction effects, but simultaneously decipher independent effects too, creating an in-depth transcriptomic profile with the same or even lower levels of sampling. As in the future, we need to better model complex mixtures of environmental factors across different doses and time points in-vitro[1–3], the FFED approach will prove advantageous for such research.

In addition to showing the usefulness of the FFED method to increase the experimental efficiency for transcriptomics or other high-throughput techniques, we have also gained insight into the biological mechanisms of the studied environmental factors. Several of the pathways identified to be changed have been linked to ASD through genetic studies[7,29]. These pathways were found to be differentially modulated by the six environmental

factors, however, such effects were predominantly driven by the exposures of Pb and FH. It is important to note here that these effects represent sustained, steady state transcriptomic responses following low-grade environmental exposure for 5 days and therefore provide stable insights with greater translational value.

We detect several relevant transcriptomic changes in the neural progenitors after low-grade exposure to Pb, primarily related to synaptic function. Our findings support previous reports on Pb exposure effects, a well-known neurotoxicant linked to cognitive impairments, ASD, and other NDDs[9,30]. In-vitro studies looking at the mechanism of Pb action with the help of primary neuronal cultures showed that Pb can affect synaptic plasticity[30] and the expression of multiple ASD susceptibility genes[28]. These reports suggest that Pb induces modulation of selected genes implicated in vesicle release, neurite growth, synaptogenesis, and transcription factors involved in neuronal differentiation[31]. Our analyses align with the earlier studies, as we detect upregulated pathways related to axonogenesis and synaptogenesis. Interestingly, we also observed a large cluster of downregulated pathways related to DNA repair, cell cycle/division, and nucleosome remodelling. While such effects are known cytotoxic changes induced by Pb exposure[32–35], these changes have also been previously identified in ASD[7,36–38].

Another exposure resulting in large pathway level changes was FH, a commonly prescribed SSRI for depression and anxiety disorders. We show that FH modulates transcriptomic pathways related to lipid metabolism, which we also functionally validated using ESI–MS. We indicate an increase in lipid-based metabolites following FH exposure in differentiating neural progenitors. The importance of lipids in the human brain is well known, which has the highest lipid content after adipose tissue[39]. Neural progenitors are rich in lipid droplets and their abundance influences the proliferative state of such progenitors, which may also contribute to cell fate determination[40]. In neural embryonic stem cells, FH exposure at different time points of neural stem cell differentiation was shown to alter cell type markers[41]. Mechanistically, FH binds to the transmembrane domain of the dimerised neurotropic tyrosine kinase receptor 2 (TKRB), otherwise responsible for long-term potentiation thorough brain derived neurotrophic factor (BDNF) signalling, which we also detect as altered after FH exposure[42].

How the changes in lipid metabolism during differentiation of neural progenitors would affect outcomes such as ASD later in life, requires further investigation. Nevertheless, it is still intriguing to note that deviations in lipid profiles have been previously observed in individuals with ASD. In a systematic review by Esposito and colleagues, 37 studies linking ASD to abnormal lipid profiles were identified[43], showing an association between ASD and hypocholesterolaemia but unclear association for fatty acids. Furthermore, recently, hypolipidemia was identified in 367 individuals with an ASD diagnosis[44]. We also noted that the baseline concentrations for many of the affected lipids were lower in the Non-ASD cell lines, and these get elevated following short-term FH exposure in both clinical backgrounds. While transcriptomic enrichment of lipid-based pathways has been noted in ASD[45] with an increasing interest in defining an ASD subtype based on dyslipidaemia[46], we provide evidence that FH exposure in early development could alter metabolomic pathways, thereby suggesting the investigation of long-term exposure effects as well as comparisons with other SSRIs should be made.

Interestingly, evidence of FH effects on lipid metabolism has started gaining traction. For instance, long term exposure to FH for 2 years in male rhesus macaques decreased polyunsaturated fatty acids in the medial prefrontal cortex[47]. An increase in serum lipid levels was observed in patients of clinical depression following treatment with FH for 8 weeks[48]. Even though there has been uncertainty regarding the direction of such changes, a report on FH effects in a murine model of depression revealed that the nature of the biological sample used in a metabolomic investigation is a key determinant of the observed lipid patterns[49]. When focussing on the specific metabolites that were reported in our study, the detected fatty acids, FA 18:1, FA 20:2 and FA 20:3 have been reported to be elevated in plasma samples of children with ASD[50]. FA 18:1 and FA 20:3 have also been recently associated with ASD diagnosis based on plasma lipidomics[51]. There are no direct reports of the other detected metabolites, following FH exposure, to be associated with ASD.

We also provide proof that when investigating effects of environmental exposures using in-vitro cellular models, diverse genetic backgrounds need to be taken into consideration. When analysing the effects separately for the clinical background and cell lines, several significantly enriched pathways were found to be predominantly affected by the exposure of Pb and FH, with both independent and combined effects. Remarkably, with a higher network diameter, Pb had a larger upregulatory response in the ASD cell lines (ASD$_{CASK}$, ASD$_{HNRNPU}$), while FH had such an impact on the Non-ASD cell lines (CTRL$_{Male}$, CTRL$_{Female}$). This could signify that the ASD cell lines are more vulnerable to Pb, and the Non-ASD cell lines are to FH, thereby showing a larger transcriptomic response to the factors respectively. Additionally, few pathways modulated by VPA, BPA, EtOH and Zn- were also detected.

What was particularly interesting was the ability of BPA to induce differential exon usage events in differentiating human neural progenitors. To find such changes by an already established endocrine disruptor[2] with previously reported effects on alternative splicing[52], emphasises the necessity to carry out further research in the context of NDDs. Moreover, we are aware that alternative splicing events can generally effect health and disease status[53]. A recent study has highlighted the contributions of alternative splicing to gene regulation in the developing brain and ASD, while providing an expansive resource for future research[54]. Based on human in-vitro and post-mortem samples, there was widespread dysregulation of neural microexons in ASD that are generated by differential exon usage events and are known to modulate axonogenesis and synaptogenesis[55]. Similar findings in related pathways, along with differences in the expression levels of splicing factors have also been reported in a murine model of ASD[56].

Our study shows that the FFED is a promising design to generate multiplexed data, when evaluating the effects of several factors on biological outcomes. We have demonstrated that it is possible to translate from exploratory analysis to functional validation with the FFED. We saw robust effects after the Pb and FH exposures in the four cell lines included in the study, but only modest pathway level changes for the other four exposures. A majority of such pathway level changes were also unique to one cell line, which indicated a need for larger studies with several cell lines, selected based on genetic and clinical profiles. While the FFED coupled with RNA-seq is an

efficient approach to explore in-vitro exposure effects in parallel, there may be limitations in detecting smaller effects or those that polarise and end-up cancelling each other out. In conclusion, we have provided evidence for an efficient and multiplexable resource that can be used for better understanding the functional ramifications of environmental factors and gene-environment interactions in ASD and other clinical conditions with neurodevelopmental links. A molecular level understanding of the biological underpinnings of both genetic and environmental factors is essential for identifying plausible aetiologies and emergent future preventive interventions in ASD.

Methods

Cell culture. For the purpose of this study, previously generated and described human iPSC lines from a neurotypical male: CTRL9II–CTRL[Male][22] and neurotypical female: AF22–CTRL[Female][20], and two males with known genetic variants for NDDs including ASD: ASD12BI–ASD[HNRNPU][21] and ASD17AII–ASD[CASK][19], were used (Fig. 1A). Information pertaining to the generation and quality control of the iPSC lines, as well as relevant clinical background are provided in the cited publications. The iPS cells were differentiated into neural progenitor state and maintained as neuroepithelial stem (NES) cells, as previously described[20,23]. While NES cells are neural progenitors with both neurogenic and gliogenic potential, upon differentiation they are primarily neurogenic unless they are specifically differentiated to glia[20,23].

The NES cells were seeded on tissue culture treated plates (Sarstedt), coated with 20 µg/mL poly-ornithine (Sigma-Aldrich) and 1 µg/mL Laminin2020 (Sigma-Aldrich). These were grown in DMEM/F-12 Glutamax basal medium (Gibco) supplemented with $0.05 \times B27$ (Gibco), $1 \times N$-2 (Gibco), 10 U/ml Penicillin–Streptomycin (Gibco), 10 ng/mL recombinant human bFGF (Gibco) and 10 ng/mL recombinant human EGF (Peprotech). Undirected differentiation of NES cells was achieved by the withdrawal of growth factors from the culture medium[20,23]. The NES cells were seeded on tissue culture treated plates (Sarstedt), coated with 20 µg/mL poly-ornithine (Sigma-Aldrich) and 1 µg/mL Laminin2020 (Sigma-Aldrich). These were grown in DMEM/F-12 Glutamax basal medium (Gibco), that was made zinc free for use in the FFED approach, supplemented with $0.5 \times B27$ (Gibco), $1 \times N$-2 (Gibco) and 10 U/ml Penicillin–Streptomycin (Gibco). When using zinc free basal medium, 1.5 µM zinc sulphate ($ZnSO_4$, Sigma) was added to maintain normal zinc concentration in the basal medium (Table S1A). The cells were differentiated for 5 days post induction of differentiation, followed by sampling based on the downstream application. To control for any variance introduced by experimentation, the experiments were performed by the same cell culture scientist in a single experimental batch with passage matched NES cells and all samples were collected at the timepoint stated in the methods section.

Zinc free DMEM/F-12 Glutamax basal medium was prepared following treatment of 500 mL DMEM/F-12 Glutamax basal medium (Gibco) with 25 g of Chelex 100 (Biorad). Following gentle mixing on an orbital shaker for 1 h at room temperature, the pH of the medium was adjusted to 7.0–7.4 using hydrochloric acid (HCl, Sigma) and sterile filtered using a 0.02 µ filtration unit (Sarstedt). As Chelex 100 is not a specific chelator for zinc ions but rather chelates all positively charged metal ions, other metal ions had to be restored in the medium by addition of a self-formulated metal supplement. To prepare 1 mL of metal supplement for 500 mL of medium, add 525.2 µL of 1 M calcium chloride ($CaCl_2$, Sigma), 0.65 µL of 1 mg/mL copper(II) sulphate ($CuSO_4$, Sigma), 4.17 µL of 100 mg/mL iron(II) sulphate ($FeSO_4$, Sigma) and 301.5 µL of 1 M magnesium chloride (MgCl2, Sigma) to 168.5 µL of Milli-Q water. The metal supplement was sterile filtered using 0.02 µ syringe filter prior to thoroughly mixing with the basal medium.

Fractional factorial experimental design (FFED). Based on our aim to study six exposures with two levels each (untreated control and treated), we selected the L8 Orthogonal Array based FFED[4]. The cell culture media was spiked with the environmental factors as per the FFED and determined treatment concentration, to mimic sustained low-grade exposures of the environmental factors in-vitro (Table S1A).

Cytotoxicity assays for environmental factors. We evaluated the cytotoxicity of lead (Pb), valproic acid (VPA), bisphenol A (BPA), ethanol (EtOH), fluoxetine hydrochloride (FH) and zinc deficiency (Zn-) in our cell lines. The treatment concentration ranges (0.00–30.00 µM or 0.00–200.00 mM) for the selected environmental factors were determined based on those previously reported in different cell types and tissue samples from the literature. (Table S1), followed by testing in our cell lines using the MTS cytotoxicity assay (Promega).

The NES cells were exposed to a concentration range of the environmental factor being studied. Growth factors were withdrawn from the cell culture, to initiate differentiation. The exposure media was changed every second day. At 24- and 120-h (5 days) post exposure, the cells were treated with MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) reagent, based on instructions from the manufacturer. After 3 h of incubation at 37 °C and 5% CO_2, absorbance was recorded at 490 nm using a spectrophotometric microplate reader. The same experimental approach was repeated when checking the selected concentrations for the environmental factors using the FFED (Table S1A). The response was represented as a regression line generated using the Loess method for local regression fitting and the shaded area denoted the 95% confidence interval. Two-way ANOVA followed by Tukey's post-hoc correction for multiple comparisons, was used to test for significance (adj. $p < 0.05$) of the reported findings.

Immunocytochemistry. The stem cell specific marker, SOX2 was visualised using immunocytochemistry. Cells were cultured on glass coverslips with exposure to the environmental factors in the FFED for 120 h (5 days) and fixed for 20 min in 4% paraformaldehyde. The primary antibody used was SOX2-AB5603, 1:1000 (Merck-Millipore). All images were taken with LSM 700 Zeiss Confocal Microscope (Zeiss Plan-Apochromat $63 \times /1.40$na Oil DIC Objective M27), with $63 \times$ magnification at 1024- $\times$ 1024- pixel (pxl) resolution, resulting

in an aspect ratio of 0.099233 µm per pixel. Fiji—ImageJ (v2.3.0/1.53f.)[57] was used to estimate SOX2 positivity, normalised against total cellular nuclei. Quantification of percentage positivity was done in R (v4.1.2)[58].

Cell proliferation. To analyse differences in cellular proliferation, when exposed to the environmental factors in the FFED, the BrdU assay (Abcam) was performed at 24 h and 120 h (5 days) post induction of differentiation. 24 h prior to the read-out, 1X-BrdU reagent was added to the cell culture vessels for incorporation by incubation at 37 °C and 5% CO_2. For the read-out, the culture media was aspirated, and cells were fixed with the supplied fixing solution. This was followed by exposure to the anti-BrdU antibody (primary antibody), followed by incubation at room temperature for 1 h and washing with the supplied plate wash buffer. The cells were incubated with the HRP-tagged secondary antibody at room temperature for 30 min followed by TMB exposure and recording absorbance at 450 nm. Quantification of percentage proliferation was done in R (v4.1.2)[58].

Bulk RNA sequencing (RNA-seq). Cell lysates were collected on day 5 of exposure (N = 43). At least one biological replicate was included for each of the four cell lines (n = 32). An additional full biological replicate was collected for $CTRL_{Male}$ (n = 8) and a partial for $CTRL_{Female}$ (n = 4) to account for cell culture related effects. RNA samples were extracted using a spin-column based kit (Promega) and following manufacturer's instructions. RNA integrity was determined using a fluorescence based micro-capillary detection system (Qsep100, Bioptic) using an RNA specific kit (Bioptic) and following manufacturer's guidelines. RNA samples with an RNA quality number (RQN) greater than 9 were selected for library preparation and sequencing.

Library preparation and sequencing were performed at the National Genomic Infrastructure (NGI), Stockholm. The sequencing library was prepared using the Illumina TruSeq RNA RiboZero GOLD kit. Pooling and sequencing were done on NovaSeq6000 (NovaSeq Control Software 1.6.0/RTA v3.4.4) with a 2 × 151 setup using 'NovaSeqXp' workflow in 'S4' mode flowcell. We obtained, on average, 46.5 million reads per sample. The Bcl to FastQ conversion was performed using bcl2fastq (v2.20.0.422) from the CASAVA software suite. Results from the best practice bioinformatics nf- core/RNAseq pipeline of the NGI, were used for further analysis[59]. In short, quality control of read sequences was performed with FastQC, followed by preparation for alignment using UMI-tools (extraction of unique molecular identifiers), Trim Galore! (adapter and quality trimming), BBSplit (removal of genomic contaminants) and SortMeRNA (removal of ribosomal RNA). Read alignment was completed using STAR aligner. Differential gene expression (DGE) analysis was performed using the DESeq2 package (v3.14)[60] in R (v4.1.2)[58].

The analysis for the detection of differentially expressed genes (DEGs) was stratified depending on several levels of complexity. In the *Level I—Global Effects*, comparisons across all cell lines, independent of genetic background were made and DEGs and globally effected pathways were identified. *Level II—Clinical Background Effects*, comparisons based on the clinical background of the iPSC lines were made to identify DEGs. Here, the neurotypical male and female iPSC lines belonged to the non-ASD clinical background, while the male ASD iPSC lines belong to the ASD clinical background. At *Level III—Cell Line* Effects, comparisons were made between the control and treatment groups of each individual cell line included in the study. Here, the linear model used was *design = ~ exposure*. The data was filtered based on the level of analysis, prior to the application of the linear model. Lastly, *Level IV—Interaction Effects,* were analysed where comparisons were made across all cell lines, both independent as well as dependent on clinical background, for the three possible interactions as per the FFED. This included interactions between Pb and Zn, VPA and FH, and BPA and EtOH. The linear model used was *design = ~ cell line + replicate + interaction + cell line * interaction*. Based on instructions provided by the package developer, the PCA was performed and were exported for easy visualisation using the pcaExplorer package (v2.20.1)[61] in R (v4.1.2)[58].

Pathway enrichment. Pathway analysis was done according to previously described protocols[62]. Following DGE analysis, RNK files were generated, and Gene-Set Enrichment Analysis (GSEA, v4.2.1)[63] was performed for biological processes, molecular function, and cellular components, with a significance threshold of FDR < 0.05. The scored gene ontology information (biological processes, MSigDB v7.4) was imported into Cytoscape (v3.9.0)[64] and visualised using the EnrichmentMap (v3.3.3)[65] app. The generated clusters with FDR adjusted p values (q values) were subsequently labelled using the AutoAnnotate (v1.3.5)[66] app, to identify enriched pathways, that were both differentially upregulated and downregulated. Summary networks were also created following the annotation. The generated network(s) were exported as PDFs and aesthetically organised using Affinity Designer (v1.10.4).

Network parameters for the generated gene ontology biological processes (GOBP) networks were analysed using the online tool NetConfer[67]. From the networks generated using the EnrichmentMap (v3.3.3)[65] app in Cytoscape (v3.9.0)[64], a list of connected nodes and their edge similarity coefficient was compiled. These lists were then used in the "WF2: Identify and compare key nodes" workflow from NetConfer[67] to compute global network properties including total nodes, total edges, diameter, density, clustering coefficient and average degree.

Gene enrichment. We used the publicly available BrainSpan[24] RNA-sequencing dataset that contains spatio-temporal gene expression data pertaining to typical neurodevelopment to analyse the enrichment of DEGs in certain time and regions. After removing low quality samples (RIN < 9.0) from the dataset, and genes with low expression (< 1 FPKM in at least 2 samples) and with low variable expression between samples (> 0 FPKM less than 50% of samples and coefficient of variance < 0.25)[68], a total of 344 samples and 24,434 genes were screened for further analysis. Four brain regions and eight developmental periods from 8 PCW (weeks post conception) to 40 years of age, were defined[69]. Hierarchical clustering was performed to group significant gene lists to 2 clusters separately, based on their standardized mean expression pattern across all brain regions and timepoints.

To identify if there were any significantly enriched gene clusters in any brain regions or time periods, the mean expression of random genes with similar numbers of selected cluster at each region and time period were calculated. The p value was based on the proportion of selected gene clusters' mean expression that were higher than the random genes' mean expression after 10,000 times of permutation. FDR[70] was used to adjust p values for multiple comparisons and were reported.

To examine for enrichment of the significant DEGs following exposure to the environment risk factors in publicly available ASD and NDD related gene lists, genetic data was obtained from SFARI gene (v2021 Q3)[25], Genetic epilepsy syndromes (EPI) (v2.489) panel[26], Intellectual disability (v3.1500) (ID) panel[26] and a developmental gene list[27]. Enrichment analysis based on hypergeometric testing was performed using the phyper function in R[58]. The background gene list for the analysis was generated from the RNA-seq expression data, for expressed genes with a base mean > 20.

Differential exon usage. Differential exon usage (DEU) analyses were performed using DEXSeq package (v1.36.0)[71]. Flattened annotation file was created using provided python script excluding the aggregate exon bins and exon counts were calculated using provided python script. For the global effects (*Level I*) analysis, the cell line factor was used as a blocking factor using the linear models $formulaFullModel = \sim sample + exon + cell\ line{:}exon + exposure{:}exon$ and $formulaReducedModel = \sim sample + exon + cell\ line{:}exon$. Therefore, DEU exon bins were called based on differences after treatments and differences between cell line gene expression backgrounds was blocked from the analysis. The analysis was repeated for ASD (ASD_{CASK}, ASD_{HNRNPU}) and non-ASD cell lines ($CTRL_{Male}$, $CTRL_{Female}$) again blocking gene expression backgrounds between cell lines. Finally, the analysis was performed for all the cell lines separately per treatment by $\sim sample + exon + exposure{:}exon$. A gene was called to have evidence for DEU, if it had at least one exon bin differentially used between conditions. The difference was considered significant with FDR[70] adjusted $p < 0.05$, exon base mean ≥ 10 and an absolute log fold change ≥ 1.5. Over-representation analysis (ORA) was performed using the online tool WebGestalt[72].

Direct infusion electrospray ionisation mass spectrometry (ESI–MS). Metabolomic profiling was performed using the direct infusion probe (DIP) for ESI-MS[73] on snap-frozen cell lysates (N = 24) at day 5 of exposure to 3.0 μM FH during undirected differentiation of NES cells, with 3 technical replicates each for the treated and untreated groups in all four cell lines: $CTRL_{Male}$, $CTRL_{Female}$, ASD_{HNRNPU} and ASD_{CASK}. The cells were stored as pellets in -80 freezer prior to lysing with the electrospray solvent containing 9:1 methanol:water and 0.1% formic acid. To enable comparison, all cell samples had the similar number of cells per volume. The samples were analysed on a QExactive Orbitrap instrument at 140,000 mass resolving power using full scanning between 70 and 1000 Da in untargeted mode. The resulting data was extracted using 1 min data per sample and sorted in MZmine2[74]. Quantification was performed using a one-point calibration and the metabolite concentration in the sample was calculated from the concentration of the corresponding internal standard multiplied with the intensity ratio of the endogenous metabolite and the internal standard.

To test for significance of the resulting metabolite concentrations (μM) following the mass spectrometry analysis pipeline, a linear model was applied to first the complete dataset, then to subset datasets based on the clinical background and individual cell lines. Here, the linear model used was *lm(concentration ~ exposure)*. The data was filtered prior to the application of the linear model, based on the level of analyses: global effects, clinical background effects, and cell line effects. The obtained p values were adjusted for multiple comparisons using the FDR method in R (v4.1.2)[58]. Significantly modulated metabolites (Untreated vs. Treated) were selected based on a significance threshold of adjusted $p < 0.05$ and were visualised using the ggplot2 package (v3.3.5)[75] in R (v4.1.2)[58].

Statistical analyses and plots. All statistical analyses, unless otherwise stated, were performed in R (v4.1.2). The statistical models and tests used for the analyses are described in the methodology relevant to the experimental technique, in the sections above. All plots, unless otherwise stated, were created using the ggplot2 package (v3.3.5)[75] or the pheatmap package (v1.0.12, https://CRAN.R-project.org/package=pheatmap) in R (v4.1.2)[58].

Data availability
The generated data is available on GitHub (https://github.com/Tammimies-Lab/FFED-RNASeq) or available from the corresponding author (kristiina.tammimies@ki.se). The RNA-sequencing read counts are additionally available on the NCBI GEO repository (GSE229546).

Code availability
The source code is available on GitHub (https://github.com/Tammimies-Lab/FFED-RNASeq) or available from the corresponding author (kristiina.tammimies@ki.se).

References
1. Austin, C. Dynamical properties of elemental metabolism distinguish attention deficit hyperactivity disorder from autism spectrum disorder. *Transl. Psychiatry* https://doi.org/10.1038/s41398-019-0567-6 (2019).
2. Caporale, N. *et al.* From cohorts to molecules: Adverse impacts of endocrine disrupting mixtures. *Science* **375**, eabe8244 (2022).
3. Curtin, P. *et al.* Dynamical features in fetal and postnatal zinc-copper metabolic cycles predict the emergence of autism spectrum disorder. *Sci. Adv.* **4**, eaat1293 (2018).
4. Mee, R. *A Comprehensive Guide to Factorial Two-Level Experimentation* (Springer, Berlin, 2009).

5. Lord, C. *et al.* Autism spectrum disorder. *Nat. Rev. Dis. Primer* **6**, 5 (2020).
6. Iakoucheva, L. M., Muotri, A. R. & Sebat, J. Getting to the cores of autism. *Cell* **178**, 1287–1298 (2019).
7. Satterstrom, F. K. *et al.* Large-scale exome sequencing study implicates both developmental and functional changes in the neuro-biology of autism. *Cell* https://doi.org/10.1016/j.cell.2019.12.036 (2020).
8. Bai, D. *et al.* Association of genetic and environmental factors with autism in a 5-country cohort. *JAMA Psychiat.* **76**, 1035 (2019).
9. Heyer, D. B. & Meredith, R. M. Environmental toxicology: Sensitive periods of development and neurodevelopmental disorders. *Neurotoxicology* **58**, 23–41 (2017).
10. Christensen, J. *et al.* Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *JAMA* **309**, 1696–1703 (2013).
11. Moore, S. J. A clinical study of 57 children with fetal anticonvulsant syndromes. *J. Med. Genet.* **37**, 489–497 (2000).
12. Thongkorn, S. *et al.* Sex differences in the effects of prenatal bisphenol a exposure on genes associated with autism spectrum disorder in the hippocampus. *Sci. Rep.* **9**, 1–14 (2019).
13. Thongkorn, S. *et al.* Sex differences in the effects of prenatal bisphenol A exposure on autism-related genes and their relationships with the hippocampus functions. *Sci. Rep.* **11**, 1241 (2021).
14. Stevens, S. A., Nash, K., Koren, G. & Rovet, J. Autism characteristics in children with fetal alcohol spectrum disorders. *Child Neuropsychol.* **19**, 579–587 (2013).
15. Boukhris, T., Sheehy, O., Mottron, L. & Berard, A. Antidepressant use during pregnancy and the risk of autism spectrum disorder in children. *JAMA Pediatr.* **170**, 117–124 (2016).
16. Arora, M. & Austin, C. Teeth as a biomarker of past chemical exposure. *Curr. Opin. Pediatr.* **25**, 261–267 (2013).
17. Arora, M. *et al.* Fetal and postnatal metal dysregulation in autism. *Nat. Commun.* **8**, 15493 (2017).
18. Pintacuda, G., Martín, J. M. & Eggan, K. C. Mind the translational gap: using iPS cell models to bridge from genetic discoveries to perturbed pathways and therapeutic targets. *Mol. Autism* **12**, 10 (2021).
19. Becker, M. *et al.* Presynaptic dysfunction in CASK-related neurodevelopmental disorders. *Transl. Psychiatry* **10**, 312 (2020).
20. Falk, A. *et al.* Capture of neuroepithelial-like stem cells from pluripotent stem cells provides a versatile system for in vitro production of human neurons. *PLoS ONE* **7**, e29597 (2012).
21. Mastropasqua, F. *et al.* Deficiency of heterogeneous nuclear ribonucleoprotein U leads to delayed neurogenesis. *bioRxiv* https://doi.org/10.1101/2022.09.14.507275 (2022).
22. Uhlin, E. *et al.* Derivation of human iPS cell lines from monozygotic twins in defined and xeno free conditions. *Stem Cell Res.* **18**, 22–25 (2017).
23. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280 (2009).
24. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
25. Banerjee-Basu, S. & Packer, A. SFARI gene: An evolving database for the autism research community. *Dis. Model. Mech.* **3**, 133–135 (2010).
26. Stark, Z. *et al.* Scaling national and international improvement in virtual gene panel curation via a collaborative approach to discordance resolution. *Am. J. Hum. Genet.* **108**, 1551–1557 (2021).
27. Li, D. *et al.* Rare variants in the outcome of social skills group training for autism. *Autism Res.* https://doi.org/10.1002/aur.2666 (2021).
28. Jiang, P., Hou, Z., Bolin, J. M., Thomson, J. A. & Stewart, R. RNA-seq of human neural progenitor cells exposed to lead (Pb) reveals transcriptome dynamics, splicing alterations and disease risk associations. *Toxicol. Sci.* **159**, 251–265 (2017).
29. Eyring, K. W. & Geschwind, D. H. Three decades of ASD genetics: Building a foundation for neurobiological understanding and treatment. *Hum. Mol. Genet.* **30**, R236–R244 (2021).
30. Neal, A. P. & Guilarte, T. R. Molecular neurobiology of lead (Pb(2+)): Effects on synaptic function. *Mol. Neurobiol.* **42**, 151–160 (2010).
31. Sanchez-Martin, F. J., Fan, Y., Lindquist, D. M., Xia, Y. & Puga, A. Lead induces similar gene expression changes in brains of gestationally exposed adult mice and in neurons differentiated from mouse embryonic stem cells. *PLoS ONE* **8**, e80558 (2013).
32. Gidlow, D. A. Lead toxicity. *Occup. Med.* **65**, 348–356 (2015).
33. Hartwig, A. Role of DNA repair inhibition in lead- and cadmium-induced genotoxicity: A review. *Environ. Health Perspect.* **102**, 6 (1994).
34. Jannuzzi, A. T. & Alpertunga, B. Evaluation of DNA damage and DNA repair capacity in occupationally lead-exposed workers. *Toxicol. Ind. Health* **32**, 1859–1865 (2016).
35. Senut, M.-C. *et al.* Epigenetics of early-life lead exposure and effects on brain development. *Epigenomics* **4**, 665–674 (2012).
36. Courchesne, E. *et al.* The ASD living biology: From cell proliferation to clinical phenotype. *Mol. Psychiatry* **24**, 88–107 (2019).
37. Markkanen, E., Meyer, U. & Dianov, G. DNA damage and repair in schizophrenia and autism: Implications for cancer comorbidity and beyond. *Int. J. Mol. Sci.* **17**, 856 (2016).
38. Mossink, B., Negwer, M., Schubert, D. & Nadif Kasri, N. The emerging role of chromatin remodelers in neurodevelopmental disorders: A developmental perspective. *Cell. Mol. Life Sci.* **78**, 2517–2563 (2021).
39. Hamilton, J. A., Hillard, C. J., Spector, A. A. & Watkins, P. A. Brain uptake and utilization of fatty acids, lipids and lipoproteins: Application to neurological disorders. *J. Mol. Neurosci.* **33**, 2–11 (2007).
40. Ramosaj, M. *et al.* Lipid droplet availability affects neural stem/progenitor cell metabolism and proliferation. *Nat. Commun.* **12**, 7362 (2021).
41. de Leeuw, V. C. *et al.* Differential effects of fluoxetine and venlafaxine in the neural embryonic stem cell test (ESTn) revealed by a cell lineage map. *Neurotoxicology* **76**, 1–9 (2020).
42. Casarotto, P. C. *et al.* Antidepressant drugs act by directly binding to TRKB neurotrophin receptors. *Cell* **184**, 1299-1313.e19 (2021).
43. Esposito, C. M., Buoli, M., Ciappolino, V., Agostoni, C. & Brambilla, P. The role of cholesterol and fatty acids in the etiology and diagnosis of autism spectrum disorders. *Int. J. Mol. Sci.* **22**, 3550 (2021).
44. Tierney, E. *et al.* Sterol and lipid analyses identifies hypolipidemia and apolipoprotein disorders in autism associated with adaptive functioning deficits. *Transl. Psychiatry* **11**, 471 (2021).
45. David, M. M. *et al.* Comorbid analysis of genes associated with autism spectrum disorders reveals differential evolutionary constraints. *PLoS ONE* **11**, e0157937 (2016).
46. Luo, Y. *et al.* A multidimensional precision medicine approach identifies an autism subtype characterized by dyslipidemia. *Nat. Med.* **26**, 1375–1379 (2020).
47. Tkachev, A. *et al.* Long-term fluoxetine administration causes substantial lipidome alteration of the juvenile macaque brain. *Int. J. Mol. Sci.* **22**, 8089 (2021).
48. Pan, S. *et al.* Fluoxetine induces lipid metabolism abnormalities by acting on the liver in patients and mice with depression. *Acta Pharmacol. Sin.* **39**, 1463–1472 (2018).
49. Zhao, J. *et al.* A comprehensive metabolomics investigation of hippocampus, serum, and feces affected by chronic fluoxetine treatment using the chronic unpredictable mild stress mouse model of depression. *Sci. Rep.* **9**, 7566 (2019).
50. Usui, N. *et al.* VLDL-specific increases of fatty acids in autism spectrum disorder correlate with social interaction. *EBioMedicine* **58**, 102917 (2020).
51. Yap, C. X. *et al.* Interactions between the lipidome and genetic and environmental factors in autism. *Nat. Med.* **29**, 936–949 (2023).

52. Kim, C.-H. *et al.* Bisphenol A exposure changes the transcriptomic and proteomic dynamics of human retinoblastoma Y79 cells. *Genes* **12**, 264 (2021).
53. Li, Y. I. *et al.* RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016).
54. Leung, S. K. *et al.* Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *Cell Rep.* **37**, 110022 (2021).
55. Irimia, M. *et al.* A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511–1523 (2014).
56. Thacker, S., Sefyi, M. & Eng, C. Alternative splicing landscape of the neural transcriptome in a cytoplasmic-predominant Pten expression murine model of autism-like Behavior. *Transl. Psychiatry* **10**, 380 (2020).
57. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
58. R Core Team. R: A language and environment for statistical computing. (2020).
59. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
61. Marini, F. & Binder, H. pcaExplorer: An R/Bioconductor package for interacting with RNA-seq principal components. *BMC Bioinform.* **20**, 331 (2019).
62. Reimand, J. *et al.* Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and Enrichment-Map. *Nat. Protoc.* **14**, 482–517 (2019).
63. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
64. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
65. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **5**, e13984 (2010).
66. Kucera, M., Isserlin, R., Arkhangorodsky, A. & Bader, G. D. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. https://doi.org/10.12688/f1000research.9090.1 (2016).
67. Nagpal, S., Baksi, K. D., Kuntal, B. K. & Mande, S. S. NetConfer: A web application for comparative analysis of multiple biological networks. *BMC Biol.* **18**, 53 (2020).
68. Eising, E. *et al.* A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development. *Mol. Psychiatry* **24**, 1065–1078 (2019).
69. Lin, G. N. *et al.* Spatiotemporal 16p11.2 protein network implicates cortical late mid-fetal brain development and KCTD13-Cul3-RhoA pathway in psychiatric diseases. *Neuron* **85**, 742–754 (2015).
70. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
71. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–2017 (2012).
72. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).
73. Marques, C., Liu, L., Duncan, K. D. & Lanekoff, I. A direct infusion probe for rapid metabolomics of low-volume samples. *Anal. Chem.* https://doi.org/10.1021/acs.analchem.2c02918 (2022).
74. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* **11**, 395 (2010).
75. Wickham, H. *ggplot2* (Springer, New York, 2009). https://doi.org/10.1007/978-0-387-98141-3.

Acknowledgements

Author contributions

Conceptualization: A.A., M.B., and K.T.; methodology: A.A., M.B., and K.T.; software development: A.A., M.B., and K.T.; validation: A.A., C.M., I.L. and K.T.; formal analysis: A.A., M.B, C.M., M.O., D.L, M.W. and F.M., ; investigation: A.A., M.B. and C.M.; resources: I.L. and K.T.; data curation: A.A., M.B., C.M., M.O., D.L., M.W. and F.M.; writing—original draft: A.A. and K.T.; writing—review and editing: A.A. and K.T.; visualization: A.A., M.O., D.L. and K.T.; supervision: M.A., A.F., C.O.D., I.L. and K.T.; project administration: A.A. and K.T.; funding acquisition: K.T.

Funding

Competing interests

A.A., C.M., M.O., D.L., F.M., M.E.W, M.A., A.F., C.O.D., I.L., and K.T. declare no competing interests. M.B. is a full-time employee of Bayer AG, Germany.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-37488-0.

Correspondence and requests for materials should be addressed to K.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.