bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Deep multitask learning of gene risk for comorbid neurodevelopmental disorders

Ilayda Beyreli[1,#], Oguzhan Karakahya[1,#], and A. Ercument Cicek[1,2,*]

[1] Department of Computer Engineering, Bilkent University, Ankara, Turkey 06800
[2] Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA 15213

Abstract. Autism Spectrum Disorder (ASD) and Intellectual Disability (ID) are comorbid neurodevelopmental disorders with complex genetic architectures. Despite large-scale sequencing studies only a fraction of the risk genes were identified for both. Here, we present a novel network-based gene risk prioritization algorithm named DeepND that performs cross-disorder analysis to improve prediction power by exploiting the comorbidity of ASD and ID via multitask learning. Our model leverages information from gene co-expression networks that model human brain development using graph convolutional neural networks and learns which spatio-temporal neurovelopmental windows are important for disorder etiologies. We show that our approach substantially improves the state-of-the-art prediction power in both single-disorder and cross-disorder settings. DeepND identifies prefrontal and primary motor-somatosensory cortex brain region, and periods from early fetal to mid fetal periods and from early childhood to young adulthood as the highest neurodevelopmental risk windows for both ASD and ID. Finally, we investigate frequent ASD and ID associated copy number variation regions and report our findings for several susceptibility gene candidates. DeepND can be generalized to analyze any combinations of comorbid disorders and is released at `http://github.com/ciceklab/deepnd`.
# Equal contribution.*Correspondance: cicek@cs.bilkent.edu.tr

1 Introduction

Autism Spectrum Disorder (ASD) is a common neurodevelopmental disorder with a complex genetic architecture in which around a thousand risk genes have a role [36]. Large consortia efforts have been paving the way for understanding the genetic, functional and cellular aspects of this complex disorder via large scale exome [18, 41, 73, 67, 66, 62, 40] and genome [33, 23, 20, 57, 1, 3] sequencing studies. Latest and also the most comprehensive study to date analyzed $\sim 36k$ samples (6,430 trios) to pinpoint 102 risk genes ($FDR \leq 0.1$) [75]. Overwhelming evidence suggests that genetic architectures of neuropsychiatric disorders overlap [60, 71, 53]. For instance, out of the twenty five SFARI Cat I ASD risk genes (i.e., highest risk), only five are solely associated with ASD. Genes like *CHD2*, *SCN2A* and *ARID1B* are associated with six neurodevelopmental disorders. Intellectual Disability (ID) is one of such comorbid disorders which manifests itself with impaired mental capabilities. Reminiscent of ASD, ID also has a complex genetic background with hundreds of risk genes involved and identified by rare *de novo* disruptive mutations observed in whole exome and genome sequencing studies [17, 24, 26, 69, 77, 80, 86]. ASD and ID are frequently observed together [61]. In 2018, CDC reported that 31% of children with ASD were also diagnosed with ID and 25% were borderline [5]. They also share a large number of risk genes [54]. Despite these similarities, Robinson et al. also point to differences in genetic architectures and report that intelligence quotient (IQ) positively correlates with family history of psychiatric disease and negatively correlates with *de novo* disruptive mutations [70]. Yet, the shared functional circuitry behind is mostly unknown.

The current lack of understanding on how comorbid neuropsychiatric disorders relate mostly stems from the incomplete profiling of individual genetic architectures. Statistical methods have been used to assess gene risk using excess genetic burden from case-control and family studies [36] which are recently extended to work with multiple traits [64]. Yet, these tools work with genes with observed disruptive mutations (mainly *de novo*). It is often of interest to use these as prior risk and

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

obtain a posterior gene interaction network-adjusted risk which can also assess risk for genes with no prior signal. Network-based computational gene risk prediction methods come handy for (i) imputing the insufficient statistical signal and providing a genome-wide risk ranking, and (ii) finding out the affected cellular circuitries such as pathways and networks of genes [36, 30, 25, 38, 65, 28, 27, 63, 8, 21]. While these methods have helped unraveling the underlying mechanisms, they have several limitations. First, by design, they are limited to work with a single disorder. In order to compare and contrast comorbid disorders such as ASD and ID using these tools, one approach is to bag the mutational burden observed for each disorder assuming two are the same. However, disorder specific features are lost as a consequence [30]. The more common approach is to perform independent analyses per disorder and intersect the results. Unfortunately, this approach ignores valuable source of information coming from the shared genetic architecture and lose prediction power as per-disorder analyses have less input (i.e., samples, mutation counts) and less statistical power [75, 10, 38]. Second, current network-based gene discovery methods can work with one or two integrated gene interaction networks [28, 56, 38, 49, 8]. This means numerous functional interaction networks (e.g., co-expression, protein interaction etc.) are disregarded which limits and biases the predictions. Gene co-expression networks that model brain development are a promising source of diverse information regarding gene risk, but currently cannot be fully utilized, as the signal coming from different networks cannot be deconvoluted. Usually, investigating such risky neurodevelopmental windows is an independent downstream analysis [83, 49]. Should this process be integrated within the risk assessment framework, it has potential to provide valuable biological insights and also to improve the performance of the genome-wide risk assessment task.

Here, we address these challenges with a novel cross-disorder gene discovery algorithm (Deep Neurodevelopmental Disorders algorithm - *DeepND*.) For the first time, DeepND analyzes comorbid neurodevelopmental disorders simultaneously over multiple gene co-expression networks and explicitly learns the shared and disorder-specific features using multitask learning. Thus, the predictions for the disorders *depend* on each other's genetic architecture. The proposed DeepND architecture uses graph convolution to extract associations between genes from gene co-expression networks. This information is processed by a mixture-of-experts model that can self-learn critical neurodevelopmental time windows and brain regions for each disorder etiology which makes the model interpretable. We provide a genome-wide risk ranking for each disorder and show that the prediction power is improved in both singletask (single disorder) and multitask settings. DeepND identifies prefrontal cortex brain region and from early fetal to late mid-fetal and from mid-childhood to young adulthood periods as the highest neurodevelopmental risk windows for both disorders.

Finally, we investigate frequent ASD and ID associated copy number variation regions and pinpoint risk genes that are disregarded by other algorithms. The software is released at `http://github.com/ciceklab/deepnd`. This neural network architecture can easily be generalized to other disorders with a shared genetic component and can be used to prioritize focused functional studies and possible drug targets.

2 Results

Using a deep learning framework which combines graph convolutional neural networks with a mixture of experts model, we perform a genome-wide risk assessment for ASD and ID simultaneously in a multitask learning setting and detect neurodevelopmental windows that are informative for risk assessment for each disorder. Our results point to the shared disrupted functionalities and novel risk

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

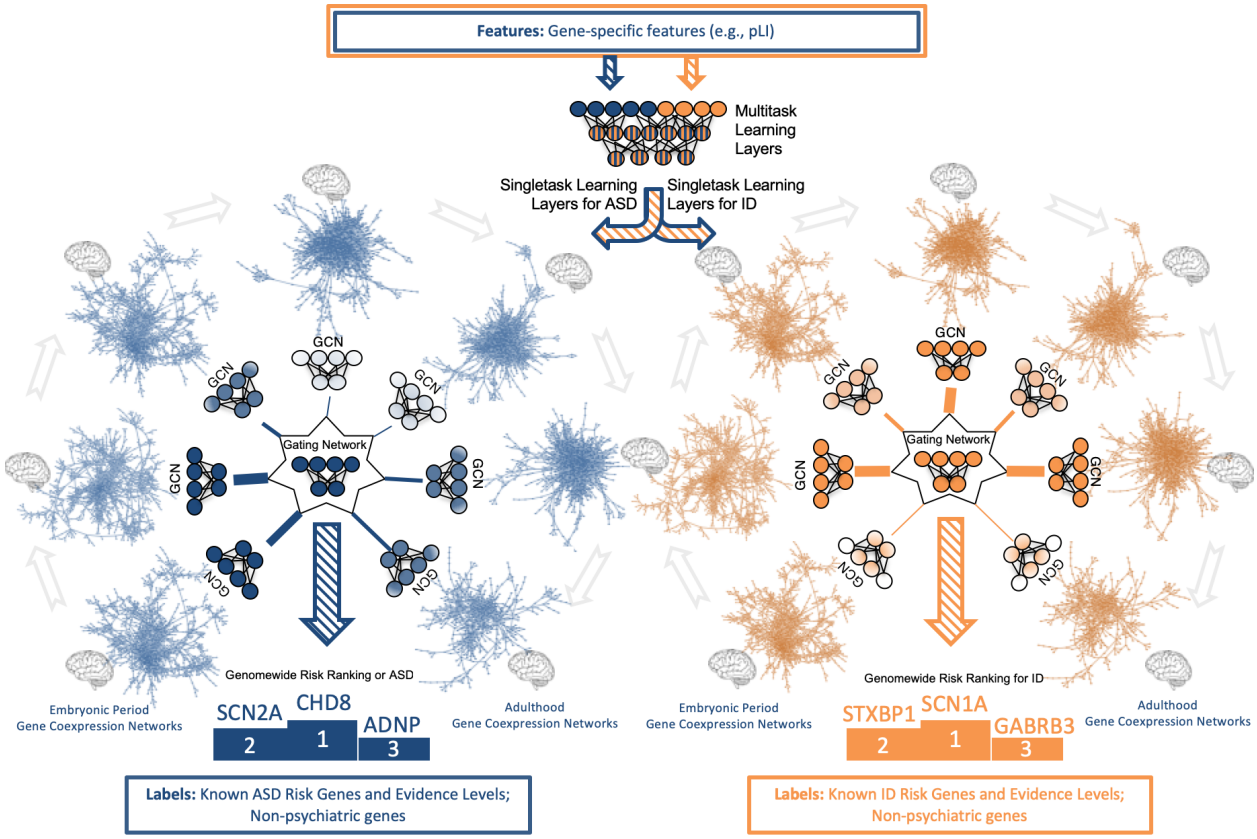

Fig. 1: System model of the proposed deep learning architecture for genome-wide cross-disorder risk assessment (DeepND). The algorithm takes the following information as input: (i) gene specific features such as pLI; (ii) disorder specific ground truth genes that are labeled as positive with varying level of evidence based on a literature search; and (iii) non-psychiatric genes which are labeled as negative. The features are passed through fully-connected multitask layers that learn shared weights for both disorders and produces a new feature representation. This new representation is then input to singletask graph convolutional neural networks (GCNs), each processing one of fifty-two gene co-expression networks that represent different brain regions and neurodevelopmental time windows. The output of GCNs are then weighted by the Gating Network to learn which networks are informative for the gene risk assessment (shade of the network indicates importance). Thus, DeepND learns which neurodevelopmental windows confer more risk for each disorder's etiology. The final output is a genome-wide risk probability ranking per disorder, which are then used for various downstream analyses to understand the underlying functional mechanisms and to compare/contrast both disorders. The singletask layers are exclusively trained with the ground truth genes of the disorder they belong. Thus, they learn only disorder specific parameters and disorder-specific networks that implicate risk.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genes which provides a road map to researchers who would like to understand the ties between these two comorbid disorders.

2.1 Genome-wide Risk Prediction for ASD and ID

To have a model of evolving gene-interactions throughout brain development, we extracted 52 gene co-expression networks from the BrainSpan dataset [78, 44] using hierarchical clustering of brain regions and a sliding window approach [83]. These networks are used to assess the network-adjusted posterior risk for each gene given disease agnostic the prior risk features such as pLI. Note that one can use disease specific features such as mutation burden in case-control or family studies. However, current knowledge on ground truth risk genes mostly rely on this data and using these as features would cause a leak in the training procedure.

The deep neural network architecture (DeepND) performs a semi-supervised learning task, which means a set of positively and negatively labeled genes are required as ground truth per disorder. As also done in [49], we obtained 594 ASD-positive genes (Supplementary Table 1) from SFARI dataset (http://gene.sfari.org) [2] which are categorized into 3 evidence levels based on the strength of evidence in the literature. For ID, we curated a ground truth ID risk gene set of 580 genes using landmark review studies on ID gene risk [26, 34, 79, 39, 12, 45] (Supplementary Table 2). We generated three evidence level sets similar to the ASD counterpart: E1, E2 and E3 sets where each set includes genes which are recurrently indicated in multiple studies. As for the negatively labeled genes, we use 1074 non-mental-health related genes for both disorders which is curated by Krishnan et al., (2016) after discarding 181 of them which are shown to be related to other neurodevelopmental disorders, such as epilepsy, schizophrenia or bipolar disorder (Supplementary Table 1).

DeepND uses the multitask learning paradigm where multiple tasks are solved concurrently (i.e., genome wide risk assessment for ASD and ID). Thus, the network learns a shared set of weights for both disorders and also disorder-specific set of weights (Figure 1). First, the model inputs pLI for each gene and using fully-connected layers produces an embedded feature set per gene. The set of weights learnt in these layers are affected by the ground truth labels for both disorders, and thus, are shared. Then, the architecture branches out to 2 single task layers, one per disorder (blue for ASD and yellow for ID in Figure 1). For each single task branch, these embeddings are input to 52 graph-convolutional neural networks (GCNs). Each GCN processes a co-expression network that represents a neurodevelopmental window and extracts network-adjusted gene risk signatures (i.e., embeddings) [75, 63] (Methods). Finally, these embeddings are fed into a fully-connected gating network along with the prior risk features. The gating network assigns a weight to each GCN which is proportional to the informativeness of the embedding coming from each neurodevelopmental window. Thus, the model also learns which windows are important for prediction of and ASD/ID risk genes. This also means they are important in the etiology of the disorder. In the end, each disorder-specific subnetwork produces a genome-wide ranking of genes being associated with that disorder along with risk probabilities. To quantify the contribution of the co-analyzing comorbid disorders (i.e., multitask) as opposed to individual analysis (i.e., singletask), we also present our results of DeepND when it is run on a single task mode (DeepND-ST). In this mode, the fully connected layers that generates embeddings are removed and the feature (i.e., pLI) is directly fed into the GCNs (Methods). We show that the genome-wide risk assessment of DeepND is robust, precise and sensitive and substantially improves the state of the art both in terms of performance and interpretability of the predictions.

We benchmark the performances of DeepND and other gene discovery algorithms for neurodevelopmental disorders. First, we compare DeepND and the state of the art algorithm by Krishnan

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

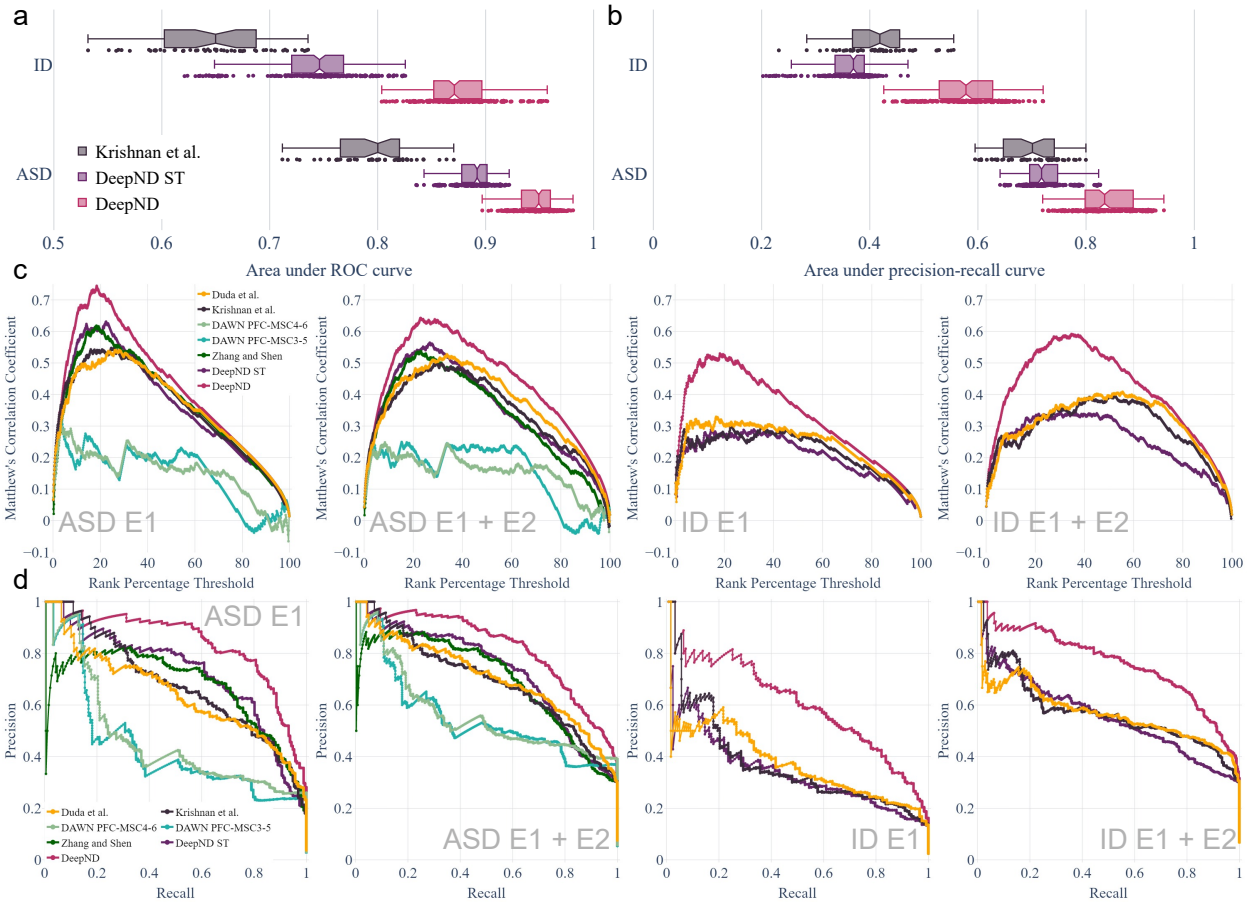

Fig. 2: Evaluation of DeepND genome-wide risk assessment for ASD and ID. (**a**) The area under ROC curve distributions of Krishnan et al., DeepND and DeepND-ST for ASD and ID genome-wide risk assessments. Every point corresponds to the performance on a test fold in the repeated cross validation setting (**b**) The area under precision-recall curve distributions for the same methods as in (**a**). Center line: median; box limits: upper and lower quartiles; whiskers: 1.5x interquartile range; points: outliers. (**c**) MCC between the rankings of each method and the ground truth genes are shown for varying rank percentage threshold values (x axis). This ranking threshold p sets the top $p\%$ of a ranking as positive predictions. Methods compared are DeepND, DeepND - ST, Evidence weighted SVM of Krishnan et al., DAWN (only compared for ASD, due to low intersection ratio with ground truth gene set of ID), DAMAGES score of Zhang and Shen, and Random Forest based method of Duda et al.. Results for ASD and ID shown when (i) E1 genes are used as the true risk genes and (ii) E1+E2 genes are considered as the true risk genes. (**d**) Precision - Recall curves to compare the same set of methods in panel c).

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

et al. using their experimental settings. Evaluating performances of the algorithms using E1 genes through 5-fold cross shows that DeepND achieves a median AUC of 94% and a median AUPR of 83% for ASD, which correspond to 4% and 47% improvements over the evidence weighted SVM algorithm of Krishnan et al., respectively (Figures 2a and 2b). Similarly for ID, DeepND achieves a median AUC of 87% and a median AUPR of 57%, which corresponds to an improvement of 15% and 41%, respectively. We observe that even the singletask mode DeepND-ST performs better than Krishnan et al.'s algorithm in all settings other than ID AUPR. As expected, the multitask setting of DeepND performs better than DeepND-ST and leads to improved AUC (up to 13%) and AUPR (up to 20%) for both disorders.

We also compare the ranking performance DeepND with the following methods from the literature using a rank percentage threshold based method: Krishnan et al.[49], DAWN [56], DAMAGES score [84], and Random Forest based method of Duda et al. [21]. That is, we obtain the final ranking of each method and mark $p\%$ of the top genes predicted as risk genes. We calculate the Matthews Correlation Coefficient of these marked lists with the ground truth. We vary p by 1% and obtain the plots in Figure 2c. Results show that for varying threshold levels, DeepND curve always dominates all others indicating that it achieves the highest correlation with the ground truth both when we use E1 and E1 + E2 genes as positive ground truth. Comparing precision-recall performance of the same set of methods show that the PR curve of DeepND domiates all other in all settings and for both disorders (Figure 2d).

Finally, we compare DeepND with an ensemble learner, forecASD [8]. This method uses the outputs of other methods including the ones we compared above and learns to predict ASD/ID risk genes in a semi-supervised setting. However, labels used by forecASD are also used by the voter methods whose outputs are used as features, which suggests information leakage during the training process and the performance might be overly optimistic. Comparing DeepND and forecASD shows that despite this advantage of forecASD, DeepND performs slightly worse or on par with respect to rank-percentage threshold curve and PR curve comparisons in ASD (see Supplementary Figure 1). DeepND performs better than forecASD in all settings for ID. We also considered adding DeepND as an additional voter into the ensemble of forecASD to check if it benefits the final performance. For both disorders and in all settings forecASD that uses DeepND as a voter outperforms both DeepND and original forecASD. This shows the informativeness of DeepND predictions.

2.2 Critical neurodevelopmental windows for ASD and ID Risk

We observe that the top percentile genes in the DeepND ASD and ID rankings have a high overlap with a Jaccard index of 65%. Top 3 deciles also have relatively high overlap with Jaccard indices 0.66, 0.71, and 0.69, respectively (Supplementary Figure 2).

To further investigate the shared genetic component, we focus on the spatio-temporal neurodevelopmental windows that are deemed important by DeepND for accurate ranking of risk genes for both disorders. The neural network analyzes co-expression networks that represent 13 neurodevelopmental time windows (from embryonic period to late adulthood; Figure 3a) and 4 brain region clusters (PFC-MSC: Prefrontal and motor-somatosensory cortex; MDCBC: Mediodorsal nucleus of the thalamus and cerebellar cortex; V1C-STC: Primary visual and superior temporal cortex; SHA: Striatum, Hippocampus, Amygdala) generated using Brainspan RNA-Seq dataset [44] in accordance with Willsey et al. [83] (Methods; Figure 3b).

We investigate using which neurodevelopmental windows more confidently distinguish disorder risk genes. Figure 3c shows normalized average probabilities assigned to top percentile risk genes

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

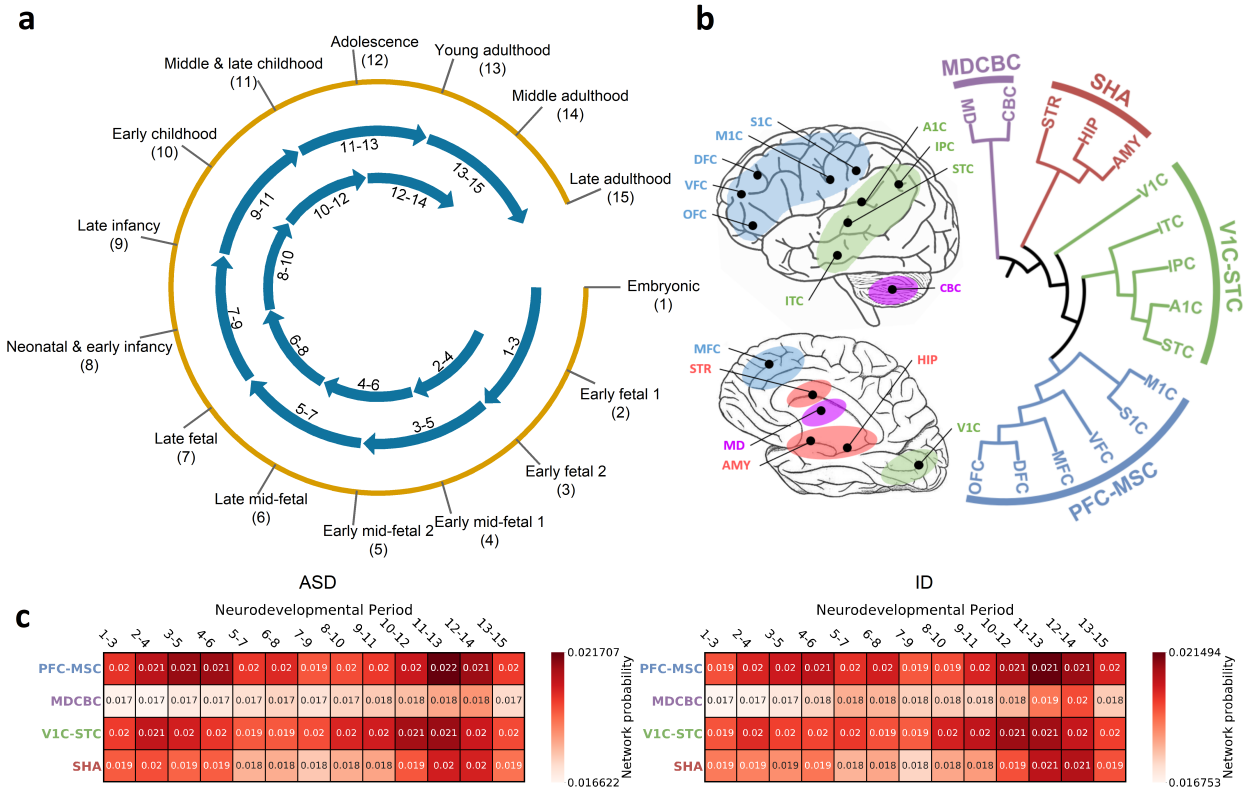

Fig. 3: The BrainSpan dataset which models the spatio-temporal gene expression of human neurodevelopment [44] is used to obtain gene co-expression networks. This dataset contains samples from 12 brain regions and spans 15 time points from early fetal period to late adulthood. **(a)** We generate 13 neurodevelopmental time-windows using a sliding window of length 3 which provides sufficient data in each window as also done by Willsey et al. [83]. **(b)** We obtain 4 brain region clusters based on transcriptional similarity during fetal development which also reflects topographical closeness and functional segregation. The brain regions considered are as follows. HIP: Hippocampal anlage (for (1)–(2)), Hippocampus (for (3)–(15)); OFC: Orbital prefrontal cortex; DFC: Dorsal prefrontal cortex; VFC: Ventral prefrontal cortex; MFC: Medial prefrontal cortex; M1C: Primary motor cortex; S1C: Primary somatosensory cortex; IPC: Posterior inferior parietal cortex; A1C: Primary auditory cortex; STC: Superior temporal cortex; ITC: Inferior temporal cortex; V1C: Primary visual cortex; AMY: Amygdala; STR: Striatum; MD: Mediodorsal nucleus of the thalamus; CBC: cerebellar cortex. **(c)** Heatmaps show which spatio-temporal windows lead to assignment of higher risk probabilities to top percentile genes for ASD (left) and ID (right). The numbers in boxes are softmaxed outputs of each respective GCN, averaged for top percentile genes and then normalized. The weights assigned to each co-expression network by the MoE lets each GCN learn to make a better prediction.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

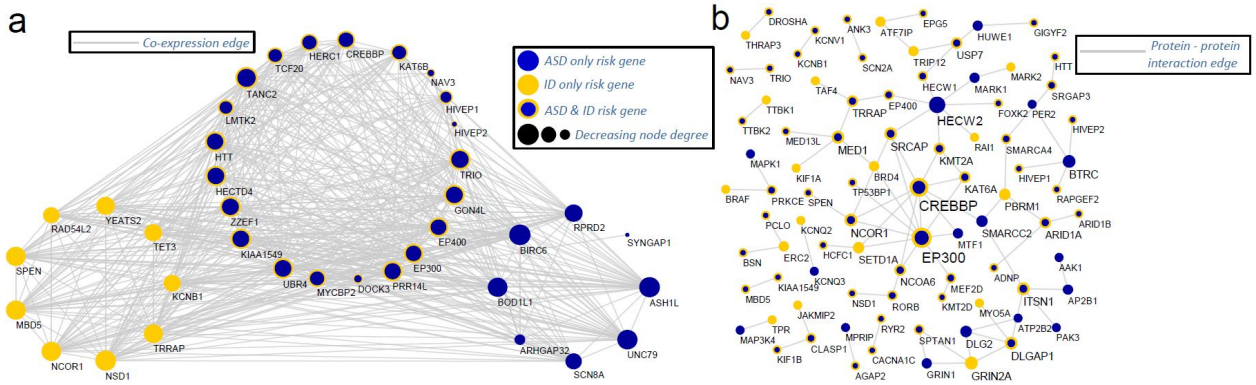

Fig. 4: Network Analyses of the risk genes. (a) The co-expression relationships between top 30 ASD and ID risk genes in PFC-MSC 4-6 region. This is found as the informative co-expression network for DeepND and also indicated in the literature as an important window for ASD [83]. Only links for at least 0.95 absolute correlation and genes with at least one connection are shown. (b) The protein-protein interactions between the top percentile risk genes are shown. The PPIs are obtained from the tissue specific PPI network of frontal cortex in DifferentialNet database [6] Only interactions between the top percentile risk genes are shown.

(Methods). We observe that the networks of the PFC-MSC brain region, spanning several time windows from early fetal to mid fetal periods and from early childhood to young adulthood periods consistently are better predictors for ASD and ID risk. The periods 3-5 and 4-6 were also previously indicated as one of the highest risk regions and were subject to network analyses for ASD gene discovery [83, 56]. DeepND also captures strong signals from these windows, in addition to period 11-13, for both ASD and ID. We see that V1C-STC brain regions are second to PFC-MSC and MD-CBC regions provide the most subtle signal. We assess the empirical of significance of these findings via a permutation test. That is, we run DeepND in the exact same settings but shuffle the labels of the risk genes, keeping the same number of positively and negatively labeled genes. For each of the squares in Figure 3c, we obtain a distribution of average probabilities (background distribution) and check if the actual informativeness of that window is higher than the average background value. See these distributions in Supplementary Figures 4 and 5, for ASD and ID, respectively. We observe that informativeness of PFC-MSC windows are the most significant compared other brain regions and MD-CBC results are consistently insignificant. Overall, we observe that not only earlier time windows but also late time windows are important for both etiologies. Supplementary Figure 3 shows the heatmaps of DeepND-ST for ASD and ID.

The weakest source of information is MDCBC 2-4 network. We observe roughly 12.5k links between top percentile ASD and ID genes. The network contains close to 37.5m links. On the other hand, PFC-MSC 4-6 contains close to 785k edges. Yet, there exists close to 13k links between top percentile ASD and ID genes which is the reason behind DeepND focusing on this window as its top predictor. Visualization of the top 30 genes for each disorder in PFC-MSC 4-6 network with only very high co-expression links ($r^2 > 0.80$) is provided in Figure 4a.

2.3 Enrichment Analysis of the Predicted Risk Genes

In addition to the prediction performance benchmark above, we also evaluate the enrichment of our ASD and ID gene risk rankings in gene lists which are shown to be related to these disorders. That is, while these gene sets are not *ground truth* sets, they have been implicated as being associated

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

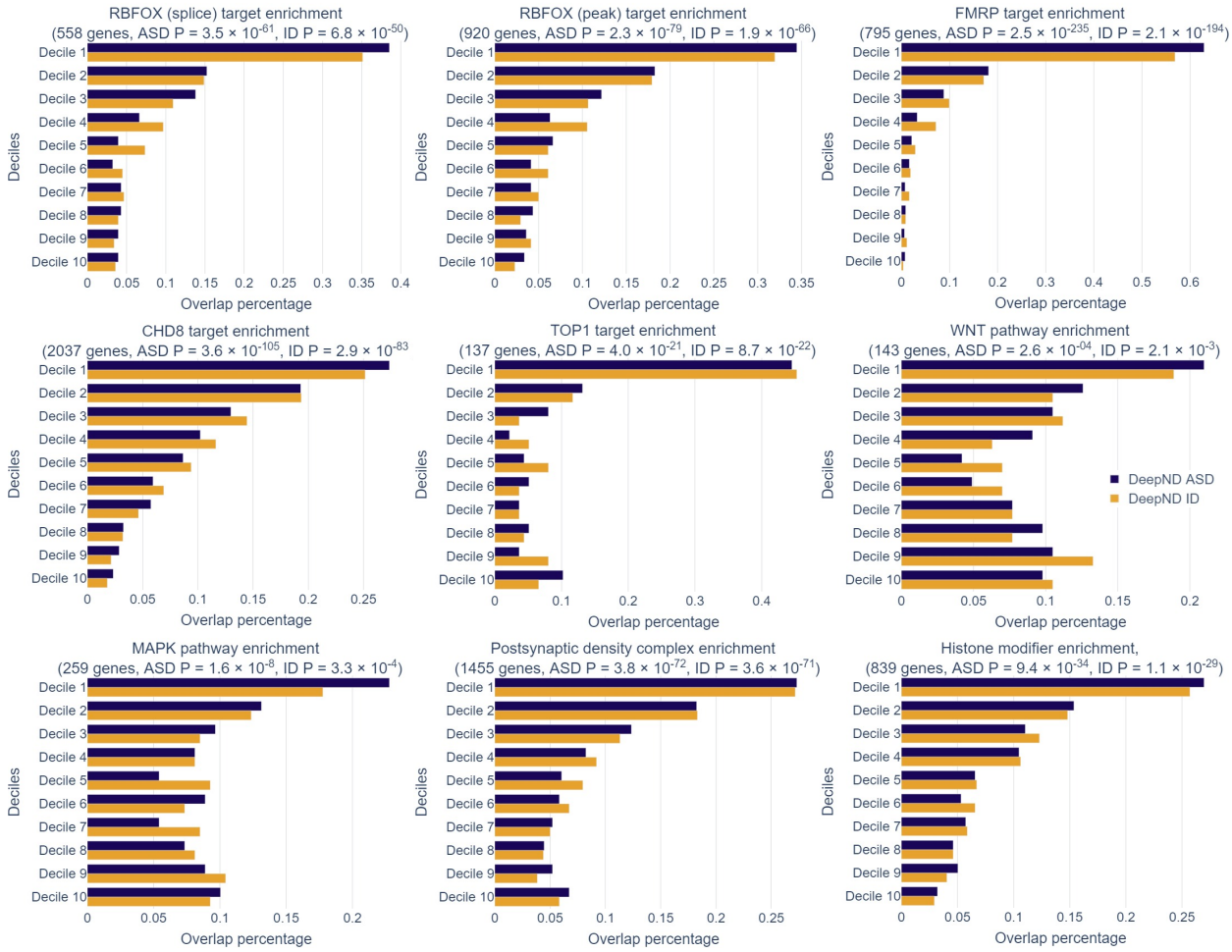

Fig. 5: The enrichment of our ASD and ID gene risk rankings in various disease-related gene lists (i.e., each panel) are shown: (i) ASD and/or ID-related transcription regulators, (ii) ASD and/or ID-related pathways, and (iii) ASD and/or ID-related biological functions or protein complexes. While these do not fully contain ground truth genes, they have been indicated in the literature as being enriched with risk genes for either disorder. Percentage of genes in the corresponding gene set (x axis) that occurred within each decile of the genome-wide risk ranking per ASD (blue) and ID (yellow) are shown. The gene sets used are as follows: (i) Targets of RBFOX (splice), (ii) Targets of RBFOX (splice target), (iii) Targets of FMRP (all peak), (iv) Targets of CHD8, (v) Targets of TOP1, (vi) WNT Pathway, (vii) MAPK Signaling Pathway, (viii) GTPase regulator activity, (x) Postsynaptic density complex genes, (xi) Synaptic genes (xii) Histone modifier genes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with the etiology of either disorder. Thus, enrichment of members of these sets in the higher deciles of the genome-wide risk ranking of DeepND is an indication of the wellness of the ranking and also provides a means of comparing and contrasting the disrupted circuitries affected by ASD and ID. These lists are (i) targets of transcription regulators like CHD8 [15], FMRP [16, 4, 18], RBFOX [81, 82, 18] and TOP1 [46]; (ii) Susceptibility pathways like WNT signaling [43] and MAPK signaling [68]; and (iii) affected biological processes and molecular functions like, post-synaptic density complex [7, 87], histone modification [18, 42]).

The first decile of ASD-risk genes has the highest enrichment in all categories. All enrichments are significant with respect to Binomial test (Figure 5; Methods). We observe the same trend for ID that the top decile of the genome-wide risk ranking is the most enriched in most categories but the P values are more subtle. $CHD8$ (Chromodomain Helicase DNA Binding Protein 8) is the highest ASD-risk gene known to date with the highest mutation burden in large ASD cohorts [18, 75] and with downstream functional analysis [15]. Accordingly, DeepND ranks $CHD8$ as the 37th top risk gene whereas Krishnan et al. ranks it 1943th genome-wide. While it has a solid association to ASD etiology, the ties to ID is not well-established. Bernier et al. reports that 9 out of 15 ASD probands with mutations in $CHD8$ also has ID. In accordance, DeepND places CHD8 as the 53rd highest risk gene for ID.

We also perform an untargeted enrichment analysis of the top percentile predictions using the Enrichr system [11, 51]. Unsurprisingly, we find that for both disorders *nervous system development* is the top enriched Biological Process GO term (Fisher's exact test, $P = 1.48 \times 10^{-12}$ for ASD and $P = 3.95 \times 10^{-14}$ for ID) which indicates that it is a shared function that is disrupted in the etiology of both disorders (Supplementary Table 3). As for the GO Molecular Function enrichment, two disorders overall have similar terms but top terms are different. The top function affected for ASD is *histone-lysine N-methyltransferase activity* indicating a disruption in the transcription factor activity whereas for ID the top affected function is *voltage-gated cation channel activity* which indicates a disruprion in synaptic activity.

As transcription factor activity is highly affected in both disorders, we further investigate if there are any master transcription regulators upstream that regulate the high risk genes for ASD and ID in the ChEA Database [52]. We find that SMARCD1 is the top transcription factor with 82 of its targets coincide with 258 the top percentile DeepND predicted ASD risk genes (Fisher's exact test, $P = 1.14 \times 10^{-20}$) and 99 of its targets coincide with the top percentile DeepND predicted ID risk genes (Fisher's exact test, $P = 7.18 \times 10^{-18}$). 62 targets are shared among ASD and ID. Note that 204 genes overlap in top percentile ASD and ID risk genes. This means 30% of the shared set of genes among two disorders are targeted by SMARCD1 (Supplementary Table 3). When the union set of the top percentile genes for ASD and ID are considered (312 genes), we find that the top transcription factor targeting these is again SMARCD1 (82 out of 2071 targets; Fisher's exact test, $P = 1.24 \times 10^{-21}$). This results indicates that SMARCD1 might be playing role in the convergent etiologies of these comorbid disorders as a shared upstream chromatin modeller.

2.4 Other Interactions between ASD and ID Genes

We further investigate the protein-protein interactions (PPI) between the top percentile risk genes in tissue specific PPI network for frontal cortex obtained from DifferentialNet database [6] using the NetworkAnalyst system [85]. This analysis reveals several hub proteins such as HECW2, EP300, and CREBBP1 as shown in Figure 4b. In this list, $HECW2$ gene stands out as it has the highest degree in this network and has very low prior risk for both ASD (TADA $P = 0.95$) and ID (extTADA

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Q = 0.97). Yet, it is in the top percentile for ID risk. Note that DeepND did not use any PPI network information in its reasoning, and yet, was able to identify this hub protein which has been linked with ASD [41] and ID [35] via de novo disruptive mutations in simplex families.

2.5 Evaluation of Novel Predictions and Identification of Candidates within ASD and ID Associated CNV Regions

Recurrent copy number variations in several regions of the genome are associated with ASD and ID etiology. However, these are large regions and it is not clear which genes in particular are driver genes. We investigate the DeepND risk ranking of genes within (i) six regions which frequently harbor ASD-related CNVs (16p11.2, 15q11-13, 15q13.3, 1q21.1 and 22q11) [49], and (ii) six regions which were reported to harbor mental retardation related CNVs (16p11.2, 1q21.1, 22q11.21, 22q11.22, 16p13 and 17p11.2) [59]. Note that these CNVs in these regions might confer risk for both disorders (Supplementary Table 4).

DeepND highly ranks several candidate genes for ASD and ID, which (i) are within these CNV regions, (ii) have low prior risk (e.g., E2 or lower, low TADA P value etc.), and (iii) low posterior risk assigned by other algorithms (Supplementary Table 4). We discuss some of them below.

NIPA2 is an ASD E3-E4 gene which encodes a magnesium transporter and located on 15q11-13. It does not participate in any of the relevant gene sets (e.g., risk pathways) and it has a low TADA $P = 0.768$. Although it is ranked in the 5th decile for ASD, it is listed in the top decile by DeepND for ID. Its linkage to Prader-Willi Syndrome by [31] also suggest that *NIPA2* might is an important candidate for ID.

DeepND links *MICAL3* to ASD and ID which is in the 22q11.21-22 CNV region. It is a gene related to actin and Rab GTPase binding, ranked in as the 191st for ASD risk and 147th for ID risk despite having ASD-TADA $Q = 0.77$ and not being in relevant risk gene groups. Krishnan et al. rank it $11,665^{th}$ and DAMAGES score ranks it 1916th. DAWN provides no ranking as it is not co-expressed with other is networks of interest for DAWN. While disruption in Rab GTPase cycle was shown to cause intellectual disability [76], there are no established ties between *MICAL3* and ASD/ID. Moreover, MICAL family genes are related to cytoskeletal organization which recently has been deemed as an important molecular function in ASD etiology [75, 65].

It is also an important task to pinpoint genes that might seem related to a neurodevelopmental disorder but actually is not. This also corresponds to untangling the genetic architectures of comorbid disorders. One example that DeepND pinpoints is *ZBTB20*. Although *ZBTB20* is not located within investigated CNV regions, it is regarded as an important risk gene for ASD. It is a CHD8 target and is listed as an E1 gene for ASD with TADA $Q = 0.17$. DAWN and Krishnan et al. rank it as $\sim 500^{th}$ ASD risk gene. However, DeepND consistently ranks it in the last decile with 0.252 probability of being an ASD risk gene. DeepND predicts that there is somewhat a higher chance for *ZBTB20* to be a candidate risk gene for ID while it is still in the 5th decile. This gene acts as a transcriptional repressor, it plays a role in various processes such as postnatal growth. Its shown relation to Primrose Syndrome which is specifically characterized by intellectual disability rather than other symptoms of ASD [14, 13] also suggests that *ZBTB20* has a closer relation to ID. Yet, more mutational evidence is required for a more concrete assessment.

Finally, DeepND ranks *LMTK2* as the 2nd and 7th highest risk gene for ASD and ID, respectively, while other algorithms do not even list it in the top 1000 (Supplementary Table 1). It has very low TADA Q values for both disorders. However, it is a target of CHD8 and FMRP, and also is a post synaptic gene. It is a gene playing role in nerve growth factor (NGF)-TrkA signalling and plays a

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

role in spermatogenesis. With this backgorund in the literature, we think this could be a novel new candidate for both disorders identified by DeepND.

3 Discussion

Neurodevelopmental disorders have been challenging geneticist and neuroscientists for decades with complex genetic architectures harboring hundreds of risk genes. Tracing inherited rare and *de novo* variation burden has been the main driver of risk gene discovery. However, overlapping genetic components and confounding clinical phenotypes make it hard to pinpoint disorder-specific susceptible genes and to understand differences. For instance, Satterstrom et al. pinpoint 102 ASD risk genes with FDR < 10% with the largest ASD cohort to date covering nearly 6.5k trios [75]. Yet, they still need to manually segregate these risk genes into two groups as (i) 53 ASD predominant risk genes which are distributed across a spectrum of ASD phenotypes, and (ii) 49 neurodevelopmental delay risk genes causing impaired cognitive, social, and motor skills. Thus, comorbidity is a further obstacle to be reckoned with in addition to identifying individual susceptible genes. Nevertheless, the shared risk genes and biological pathways offer opportunities for computational risk assessment methods which were not explored before. So far, only disorder specific analyses were possible by design of the network-based gene discovery algorithms. These are limited in power due to distinct datasets which lead to limited cohort sizes. Here, we proposed a novel approach which can co-analyze comorbid neurodevelopmental disorders for gene risk assessment. The method is able to leverage the shared information using multitask learning paradigm for the first time for this task. DeepND learns both a shared and a disorder-specific set of weights to calculate the genome-wide risk for each disorder. DeepND is a multitask deep learner which uses state-of-the-art techniques underneath such as graph convolution and mixture of experts to learn non-linear relationships of genes on 52 brain co-expression networks. The model is also interpretable as it is able to learn which neurodevelopmental windows (i.e., networks) provide more information for distinguishing high risk genes, and thus, are important for understanding disease etiology. Our benchmarks show these techniques enable DeepND to outperform existing gene discovery methods.

We focus on ASD and ID in this study and identify similarities such as shared affected pathways and neurodevelopmental windows, and differences such as regulatory relationships and novel risk genes. We think the findings in this paper will help guiding neuroscientists researching ASD and ID in prioritizing downstream functional studies. DeepND is not an algorithm specific to these disorders though. It can easily be extended to consider other comorbid disorders such as schizophrenia and epilepsy by adding similar singletask layers for each disorder. It can also be used for other comorbid disorders that are not related to neurodevelopment at all.

We demonstrated the advantage of being able to employ multiple co-expression networks and pointed to cases where, for instance, DAWN was not able to capture relationships between genes as they are limited with a single co-expression network. For the clarity of discussion and interpretability, we focused on only networks produced from BrainSpan dataset. However, DeepND can also employ any combination of other types of gene interaction networks such as protein interaction networks.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4 Methods

4.1 Ground Truth Risk Gene Sets

Our genome-wide gene-risk prediction algorithm is based on a semi-supervised learning framework in which some of the samples (i.e., genes) are labeled (as ASD/ID risk gene or not) and they are used to learn a ranking for the unlabeled samples.

We obtain the labels for ASD from SFARI dataset [2]. These genes are classified into three evidence levels indicating the quality of the evidence (E1 - E3, E1 indicating the highest risk). The list contains 185 E1 genes (from SFARI Cat I), 200 E2 genes (from SFARI Cat II) and 470 E3 genes (from SFARI Cat III)as positively labeled ASD-risk genes. We also obtain 893 non-mental health related genes from OMIM and Krishnan et al. as negatively labeled genes. The list and corresponding evidence levels are listed in Supplementary Table 1. In the loss calculations during training of the model, genes in these categories are assigned the following weights: E1 genes (1.0), E2 genes (0.50), E3/E4 genes (0.25) and negative genes (1.0). The performance is evaluated only on E1 genes for the sake of compatibility with other methods.

For ID, we rely on review studies from the literature which provide in depth analyses and lists of ID risk genes. We considered the *known gene* lists from 2 landmark review studies as our base ground truth ID risk gene list [26, 34]. We divide this set into 2 parts (i.e., E1 and E2) with respect to evidence obtained from 4 other studies [79, 39, 12, 45]. The genes which are indicated by 5 or more studies are assigned to the highest risk class, E1. Remaining genes which are indicated by 4 studies are assigned to the second highest risk class, E2. See Supplementary Table 2 for a detailed breakdown of evidence for each gene. We use the same set of negative genes as ASD which are non-mental health related genes. Consequently, we have 123 E1, 224 E2 and 232 E3 genes and 893 negative genes for ID. See Supplementary Table 1 for a complete list of ground truth genes for both disorders. The weights we use per gene class are as follows: E1 genes (1.0), E2 genes (0.50), E3 genes (0.25), and negative genes (1.0). For the sake of consistency, we also report the performance of the model on E1 genes for ID.

4.2 Gene Risk Features

The only gene risk feature we use the gene pLI for both disorders.See Supplementary Table 1 for pLI of each gene obtained from [75]. Note that neurodevelopmental risk genes are known to be conserved. Yet, the ground truth labels described in Section 4.1 are assigned independently and has no ties to this information.

4.3 Gene Co-expression Networks

We used the BrainSpan dataset of the Allen Brain Atlas [78, 44] in order to model gene interactions through neurodevelopment and generated a spatio-temporal system of gene co-expression networks. This dataset contains 57 postmortem brains (16 regions) that span 15 consecutive neurodevelopmental periods from 8 postconception weeks to 40 years. To partition the dataset into developmental periods and clusters of brain regions, we follow the practice in Willsey et al. (2013) [83]. Brain regions were hierarchically clustered according to their transcriptional similarity and four clusters were obtained (Figure 3b): (i) V1C-STC (primary visual cortex and superior temporal cortex), (ii) PFC-MSC (prefrontal cortex and primary motor-somatosensory cortex), (iii) SHA (striatum, hippocampal anlage/hippocampus, and amygladia), and (iv) MDCBC (mediodorsal nucleus of the thalamus and cerebellar cortex). In the temporal dimension, 13 neurodevelopmental windows (Figure 3a) were obtained using a sliding window approach (i.e., [1–3], [3–5], $\cdots$, [13–15]). A spatio-temporal window

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of neurodevelopment and its corresponding co-expression network is denoted by the abbreviation for its brain region cluster followed by the time window of interest, e.g. "PFC-MSC(1-3)" represents interactions among genes in the region PFC-MSC during the time interval [1–3].

Using the above-mentioned partitioning, we obtained 52 spatio-temporal gene co-expression networks, each of which contain 25,825 nodes representing genes. An undirected edge between two nodes is created if their absolute Pearson correlation coefficient $|r|$ is greater than or equal to 0.8 in the related partition of BrainSpan data.

4.4 DeepND Model

Problem Formulation Each 52 co-expression network j described in Section 4.3 is represented as a graph $G_j = (V, E_j)$, where the vertex set $V = (v_1, \cdots, v_n)$ contains genes in the human genome and $E_j \in \{0,1\}^{n \times n}$ denotes the binary adjacency matrix. Note that $n = 25,825$. Let $X_D \in \mathbb{R}^{n \times d}$ be the feature matrix for disorder D where each row $X_D[i]$ is a list of d features associated with gene (and node) $i, \forall i \in [1,n]$. Let y_{ASD} be an l dimensional vector, where y_{ASD}[i] = 1 if the the node i is a risk gene for ASD, and y_{ASD}[i] = 0 if the gene is non-mental health related, $\forall i \in [1,l]$, $l < n$. Note that, in this semi-supervised learning task, only first l genes out of n have labels. y_{ID} is defined similarly for ID, using its ground truth risk and non-risk gene sets. The goal of the algorithm is to learn a function $f(X_{ASD}, X_{ID}, y_{ASD}, y_{ID}, G_1, \cdots, G_{52}) \rightarrow P \in \mathbb{R}^{n \times 2}$. $P[i][ASD]$ denotes $p(y_{ASD}[i] = 1)$, and $P[i][ID]$ denotes $p(y_{ID}[i] = 1), \forall i \in [1,n]$.

Graph Convolutional Network Model Convolutional Neural Networks (CNNs) have revolutionized the computer vision field by significantly improving the state-of-the art by extracting local patterns on grid-structured data [50]. Applying the same principle on arbitrarily structured graph data have also been a success [9, 22, 37]. While all these spectral approaches have proven useful, they are computationally expensive. Kipf and Welling have proposed an approach (graph convolutional network - GCN) to approximate the filters as a Chebyshev expansion of the graph Laplacian [19] and let them operate on the 1-hop neighborhood of each node [48]. This fast and scalable approach extracts a network-adjusted feature vector for each node (i.e., embedding) which incorporates high-dimensional information about each node's neighborhood in the network. The convolution operation of DeepND is based on this method. Given a gene co-expression network G_j, GCN inputs the normalized adjacency matrix $\hat{E}_j$ with self loops (i.e., $\hat{E}_j[i,i] = 1, \forall i \in [1,n]$) and the feature vector $X_D[i] \in \mathbb{R}^{n \times d}$, for gene i and for disorder D. $d = 1$ in this application and shared among ASD and ID. Then, the first layer embedding $H^1[i] \in \mathbb{R}^{d_1}$ produced by GCN is computed using the following propagation rule $H_1[i] = ReLU(\hat{D}^{-0.5}\hat{E}\hat{D}^{-0.5}X_D[i]W_0)$ where W_0 is the weight matrix at the input layer to be learnt, $ReLU$ is the rectified linear unit function and $\hat{D}$ is the normalized version of a diagonal matrix where $\hat{D}_{ii} = \sum_j \hat{E}_{ij}$. We pick $d_1 = 4$ in this application. Each subsequent layer k is defined similarly as follows: $H_k[i] = ReLU(\hat{D}^{-0.5}\hat{E}\hat{D}^{-0.5}H_{k-1}[i]W_{k-1})$. Thus, the output of a k-layered GCN, for gene i on co-expression network j is denoted as follows: $GCN(G_j, X_D[i]) = H_k[i] \in \mathbb{R}^{d_k}$. In this application, we use 2 GCN layers and $d_2 = 1$. The final layer is softmax to produce probabilities for the positive class. That is, the output of a GCN model j, for a gene i is $GCN(G_j, X_D[i]) = softmax(H_2[i]) = v_i^j \in \mathbb{R}$.

Mixture of Experts Model Mixture of experts model (MoE) is an ensemble machine learning approach which aims to find out informative models (experts) among a collection [58]. Specifically, MoE inputs the features and assigns weights to the outputs of the experts. In DeepND architecture, individual experts are the 52 GCNs which operate on 52 gene co-expression networks as explained

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

above $(G_1, \cdots, G_{52})$. For every gene i, MoE inputs $X_D[i]$ and produces a weight $\vec{w}$ of length 52 which is passed through a softwax layer (i.e., $\sum_{h=1}^{52} \vec{w_i}[h] = 1$). The output of the GCNs are weighted by this network and the weighted sum is used to produce a risk probability for every gene i using softmax. That is, the $MoE(X_D[i]) = \vec{w}_i \in \mathbb{R}^{52}$. The GCNs produce 52 v_i^j values (one per co-expression network) which are concatenated to produce $\vec{v}_i \in \mathbb{R}^{52}$. Finally, the following dot product is used to produce the risk probability of gene i for disorder D: $\vec{w}_i \cdot \vec{v}_i = P[i][D]$.

Multitask Learning Model Above-mentioned GCN and MoE cascade inputs $X_D[i]$ and co-expression networks $G_1, .., G_{52}$ along with labels for a single disorder to predict the risk probability for every gene. Thus, it corresponds to the single-task version of DeepND (i.e., DeepND-ST.) On the other hand, DeepND is designed to work concurrently with multiple disorders (i.e., ASD and ID.) DeepND employs one DeepND-ST cascade per disorder and puts a multi-layer perceptron (MLP) as a precursor to two DeepND-ST subnetworks. The weights learnt on these subnetworks are only affected by the back-propagated loss of the corresponding disorder, and hence, these are single task parts of the architecture. On the contrary, the weights learnt on the MLP part are affected by the loss back-propagated from both subnetworks. Thus, this part corresponds to the multitask component of the DeepND architecture. The MLP layer inputs the union of X_{ASD} and X_{ID} and passes it through a fully connected layer followed by Leaky ReLU activation (negative slope = -1.5) to learn a weight matrix W_{MLP} and output a d' dimensional embedding to be input DeepND-ST instead of $X_{ASD}[i]$ and $X_{ID}[i]$. That is, $MLP(X_{ASD}[i] \cup X_{ID}[i]) = LRELU(W_{MLP}^{\intercal} \cdot (X_{ASD}[i] \cup X_{ID}[i]) + \vec{b}_{MLP})$.

Learning and the Cross Validation Setting of DeepND To evaluate both DeepND-ST and DeepND approaches, we use a five-fold cross-validation scheme. All labeled genes are uniformly and randomly distributed to the folds. At each training iteration, we leave one fold for validation and one fold for testing. We train the model on the remaining three folds of the labeled genes. For all the genes in the left-out fold, their feature vectors are nullified when input to training in order to prevent information leakage.

The model is trained up to 1000 epochs with early stop which is determined with respect to the loss calculated on the validation fold using only E1 genes as the positives and all negative genes. Once the model converges, the test performance is reported on the test fold in the same manner as in the validation fold. The model uses cross entropy loss (evidence weighted) and ADAM optimizer [47] for updating the weights which are initialized using Xavier initialization [29]. DeepND-ST uses a fixed learning rate of 7×10^{-4} for both disorders. DeepND uses the learning rates of 7×10^{-4}, 7×10^{-4} and 7×10^{-3}, for the shared layer, ASD single task layer, and ID single task layer, respectively. To ensure proper convergence of DeepND, should one of the singletask subnetworks converge, the learning rate of that subnetwork and the shared layer are cut down twenty folds. This lets the yet underfit subnetwork to keep learning till early stop or the epoch limit.

Above-mentioned procedure produces 4 results as for every left-out test fold as all remaining 4 folds are used as validation. We repeat this process 10 times with random initialization to obtain a total of 200 results for performance comparison (Figure 2). For ASD, positively labeled genes' weights equal to 1.00 (E1), 0.50 (E2), 0.25 (E3) and 0.25 (E4). For ID, positively labeled genes' weights equal to 1.00 (E1) and 0.50 (E2). For both disorders, negatively labeled genes have weight 1.00. Note that this procedure is in line with the setting of Krishnan et al. for fairness of comparison.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.5 Genome-wide risk prediction for ASD and ID and Comparison with Other Methods

We compare the performance of DeepND-ST and DeepND with other state-of-the-art network based neurodevelopmental disorder gene risk assessment methods from the literature which output a genome-wide risk ranking: Krishnan et al. [49], DAWN [55], DAMAGES score [84], Random Forest based method of Duda et al. [21] and forecASD [8].

We run Krishnan et al.'s approach as described in their manuscript [49]. This is an evidence-weighted Support Vector Machine (SVM) classifier which identifies risk genes based on similarity of network features. They use a human brain-specific functional interaction network to generate features as input [32]. Note that, the ground truth gene set is the same as ours as well as the evidence weights for ASD. We perform a 5-fold cross validation. That is, for each iteration, we train their SVM model on 80% of the labeled genes and evaluate the model on E1 genes and all negative genes in the left-out 20% of the labeled genes as suggested. We repeat this procedure 10 times. We post-process SVM outputs to produce risk probabilities using isotonic regression which ensures that the gene ranking is preserved. We use the pLI value of each gene as the dependent variable. In a 10-fold cross validation setting, we detect *knots* on the left-out fold, and fit another isotonic regression line to interpolate the knots. We use SVM output for all genes to produce a gene risk probability values for the corresponding disorder and produce the genome-wide risk ranking. We compare this method and DeepND-ST/DeepND with respect to (i) the area under receiver operating curve (AUC) and area under precision recall curve (AUPR) distributions calculated on the left-out fold at each cross validation iteration (Figure 2a-b), (ii) Matthews Correlation Coefficient of the final predictions with the ground truth (using a rank percentage threshold - Figure 2c), and (iii) precision-recall curves (Figure 2d).

DAWN is a hidden Markov random field based approach that assigns a posterior, network-adjusted disorder risk score to every gene based on guilt by association principle. It inputs TADA p-values as prior features along with a partial co-expression network to assess connectivity. We input the TADA p-values to DAWN [75, 63]. The method also uses partial co-expression networks. We use two networks which the authors suggest as the most useful for this task in their manuscript [55]. These represent prefrontal cortex/mid-fetal period (i.e., PFC-MSC 3-5 and PFC-MSC 4-6). We generate these networks using the RNA-Seq data in the BrainSpan dataset. Note that DeepND utilizes the same dataset and uses these networks and 50 others. Instead of partial co-expression networks, DeepND uses co-expression networks. The DAMAGES score is a principal component analysis based technique that assess the risk of genes based on (i) the similarity of their expression profiles in 24 specific mouse central nervous system cell types, (ii) the enrichment of mutations in cases as opposed to controls, and (iii) the pLI score of the gene. We directly obtain the risk ranking from Shen and Zhang, 2017 [84]. We also benchmark the Random Forest based method presented by Duda et al. as described in [21]. The method inputs a combination of (i) gene expression data from several microarray studies, (ii) PPI interactions from multiple databases, (iii) a quantitative physical interaction (PI) score for all protein isoform pairs in mouse, and (iv) phenotype annotations from the Mouse Genome Informatics (MGI) database. They combine these modalities by a Bayesian network and the final probability of functional interaction between gene pairs are represented with a matrix. Since this method does not support evidence weights for the ground truth set, in the experimental setting, we used our ground truth gene set for ASD after discarding the evidence information to obtain results. We compare all above mentioned methods with respect to (i) Matthews Correlation Coefficient of the final predictions with the ground truth (using a rank percentage threshold - Figure 2c), and (ii) precision-recall curves (Figure 2d).

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Apart from the previously mentioned methods, we also consider forecASD which is a random forest based ensemble method. It inputs features derived from (i) BrainSpan transcriptome, (ii) the STRING PPI network and the outputs of several previously published ASD gene prediction methods: Krishnan et al. [49], DAWN [55], DAMAGES score [84] and TADA [74]. These features are combined in a two-layer architecture. In the first layer, each derived feature set is used to train a random forest branch. Then, in the second step, the results from the first layer are combined and are used to train the final ensemble classifier. We trained the original forecASD architecture with our ground truth sets for ASD and ID and obtained their results. Then, in a separate run, we also added the final ranking of DeepND as an extra feature and retrained the model. We compared these two results with our final ranking using (i) Matthews Correlation Coefficient of the final predictions with the ground truth (using a rank percentage threshold), and (ii) precision-recall curves (Supplementary Figure 1).

All algorithms are trained and tested on a SuperMicro SuperServer 4029GP-TRT with 2 Intel Xeon Gold 6140 Processors (2.3GHz, 24.75M cache), 251GB RAM, 6 NVIDIA GeForce RTX 2080 Ti (11GB, 352Bit) and 2 NVIDIA TITAN RTX GPUs (24GB, 384Bit). For DeepND-ST, we used 3 2080 RTX and 1 TITAN RTX cards and the cross validation setup took approximately 5 hours. For DeepND, we used 5 RTX 2080 and 1 TITAN RTX cards and the cross validation setup took approximately 16 hours.

4.6 Enrichment Analyses

We evaluate the enrichment of DeepND's ASD and ID gene risk rankings in gene lists which are known to be enriched in disorder-risk genes in the literature. These lists are (i) targets of transcription regulators like CHD8 [15], FMRP [16, 72], RBFOX1 [81, 82] and TOP1 [46]; (ii) Susceptibility pathways like WNT signaling [43] and MAPK signaling [68]; and (iii) affected biological processes and molecular functions like, post-synaptic density complex [7, 87], histone modification [18, 42]). We use the binomial test to determine whether the top decile in the corresponding ranking significantly deviate from uniform enrichment (Figure 5).

We perform a Gene Ontology term enrichment analysis of the top percentile predictions using the Enrichr system [11, 51] with respect to Biological Process terms and Molecular Function terms. As both analyses point to *transcription factor regulation*, we investigate the connectivity of the high risk genes in the ChEA Database [52] which is a large repository that lists experimentally validated transcriptional regulation relationships in various organisms.

4.7 Spatio-Temporal Network Analyses

We investigate which GCNs (i.e., neurodevelopmental windows) are better predictors of gene risk for each disorder. For each gene i, the average risk probability assigned by the corresponding GCN is calculated over all iterations such that i is in the test fold. The average values for top percentile genes for each disorder are shown in Figure 3.

To check whether obtained average risk probabilities per neurodevelopmental window that are assigned by that GCN are empirically significant, we performed a permutation test. First, we randomly redistribute the labels of the ground truth genes, while keeping the counts same. We obtain 100 randomized ground truth sets. Keeping all other settings same, we train 100 DeepND models and obtain 100 ASD and 100 ID risk rankings. Using these rankings, we calculate the the average risk probability assigned by each GCN and obtain two background distributions for each GCN; one for ASD and one for ID (Supplementary Figures 4 and 5). We then compare the original average

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

probability assigned by that GCN with the corresponding background distribution to assess significance.

Code Availability. DeepND is implemented and released at http://github.com/ciceklab/deepnd. We provide the environment which contains all dependencies for an easy setup. We give a small example to train and test both DeepND-ST/ DeepND models. Finally, we provide the code and links to the full set of datasets to reproduce the results (genomewide risk rankings and data for heatmaps) presented in this manuscript for ASD and ID.

Data Availability. All datasets used in this study are publicly available, which are referenced in the relevant methods subsections. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The GTEx data used for the analyses described in this manuscript were obtained from the GTEx Portal on Feb 2020. All data supporting the key findings such as gene risk evidence levels and gene risk predictions are available within the article and corresponding supplementary tables.

Acknowledgements: We thank all families who participated in the studies which enable this study with data. Authors would like to acknowledge Autism Sequencing Consortium PIs and contributors. This work was supported by a pilot grant from the Simons Foundation (SFARI 640935, AEC). We also acknowledge the support by TUBA GEBIP 2017 and Bilim Akademisi BAGEP 2020 Awards to AEC.

Author Contributions. AEC conceived, designed, and supervised the study. IB and OK implemented the DeepND software. IB, OK and AEC performed the computational experiments and wrote the manuscript.

Competing Financial Interests. Authors declare no competing financial interests.

Supplementary Figure and Table Legends.

Supplementary Figure 1. Figure compares forecASD, DeepND and forecASD + DeepND performances. (a) MCC between the rankings of each method and the ground truth genes are shown for varying rank percentage threshold values (x axis). This ranking threshold p sets the top $p\%$ of a ranking as positive predictions. and (b) precision-recall curves.

Supplementary Figure 2. (a) The percentage overlap between corresponding deciles of the ASD and ID genome-wide risk rankings are shown. (b) The percentage overlap between corresponding percentiles of the ASD and ID genome-wide risk rankings are shown.

Supplementary Figure 3. Heatmaps show which spatio-temporal windows lead to assignment of higher risk probabilities to the top percentile genes for ASD (left) and ID (right). The numbers in boxes for network probabilities are softmaxed outputs of each respective GCN, averaged for top percentile genes and then normalized. The categorization of brain regions and time windows are provided in Figure 3.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figure 4. Results for the permutation test to assess the significance of the informativeness of each neurodevelopmental window for ASD. The boxplots show the distribution of the average probabilities assigned by the corresponding GCN to top percentile genes, for each of the 100 randomized ground truth sets. The red cross indicates the result obtained with actual ground truth labels and match with the results presented in Figure 3b.

Supplementary Figure 5. Results for the permutation test to assess the significance of the informativeness of each neurodevelopmental window for ID. The boxplots show the distribution of the average probabilities assigned by the corresponding GCN to top percentile genes, for each of the 100 randomized ground truth sets. The red cross indicates the result obtained with actual ground truth labels and match with the results presented in Figure 3b.

Supplementary Table 1. Genome-wide risk probability predictions and rankings of DeepND for ASD and ID. The table marks the gold standard genes: 855 positively and 893 negatively labeled genes for ASD; and 579 positively and 893 negatively labeled genes for ID are given along with their evidence levels (E1 - E4, E1 indicating the highest risk). The table also provides the rankings from other gene discovery algorithms from the literature. For each gene, participation in disorder-related gene sets are provided (e.g., WNT pathway, *CHD8* targets etc.). We also provide 29 gene specific details compiled from the literature.

Supplementary Table 2. The table provides information about the studies used to generate the ground truth labels for ID. For each gene, the base studies which indicates it as a risk gene are provided.

Supplementary Table 3. The table lists (i) the GO enrichment analysis for the top percentile risk genes for both disorders (Biological Process and Molecular Function). It also provides lists of top transcription factor (TF) regulators whose targets are enriched with the top percentile ASD and/or ID risk genes, respectively. The TF enrichment results are based on the experimentally validated TF-gene relationships in ChEA 2016 database. All results are obtained using the EnrichR system.

Supplementary Table 4. DeepND risk rankings for the genes within 6 ASD related (16p11.2, 15q11-13, 15q13.3, 1q21.1 and 22q11) (Krishnan et al., 2016) and 6 mental retardation related CNV regions (16p11.2, 1q21.1, 22q11.21, 22q11.22, 16p13 and 17p11.2). Frequency of these CNV regions within ASD and ID individuals are provided when available.

References

1. Meta-analysis of gwas of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24. 32 and a significant overlap with schizophrenia. Molecular autism **8**, 1–17 (2017)
2. Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., Packer, A.: Sfari gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (asds). Molecular autism **4**(1), 36 (2013)
3. Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., Bolshakova, N., Bolte, S., Bolton, P.F., Bourgeron, T., Brennan, S., Brian, J., Casey, J., Conroy, J., Correia, C., Corsello, C., Crawford, E.L., de Jonge, M., Delorme, R., Duketis, E., Duque, F., Estes, A., Farrar, P., Fernandez, B.A., Folstein, S.E., Fombonne, E., Gilbert, J., Gillberg, C., Glessner, J.T., Green, A.: Individual common variants exert weak effects on risk for autism spectrum disorders. Hum Mol Genet **21** (2012). https://doi.org/10.1093/hmg/dds301, `https://doi.org/10.1093/hmg/dds301`

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. Ascano, M., Mukherjee, N., Bandaru, P., Miller, J.B., Nusbaum, J.D., Corcoran, D.L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M., et al.: Fmrp targets distinct mrna sequence elements to regulate protein expression. Nature **492**(7429), 382–386 (2012)
5. Baio, J., Wiggins, L., Christensen, D.L., Maenner, M.J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Rosenberg, C.R., White, T., et al.: Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2014. MMWR Surveillance Summaries **67**(6), 1 (2018)
6. Basha, O., Shpringer, R., Argov, C.M., Yeger-Lotem, E.: The differentialnet database of differential protein–protein interactions in human tissues. Nucleic acids research **46**(D1), D522–D526 (2018)
7. Bayés, À., Van De Lagemaat, L.N., Collins, M.O., Croning, M.D., Whittle, I.R., Choudhary, J.S., Grant, S.G.: Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nature neuroscience **14**(1), 19–21 (2011)
8. Brueggeman, L., Koomar, T., Michaelson, J.J.: Forecasting risk gene discovery in autism with machine learning and genome-scale data. Scientific Reports **10**(1), 1–11 (2020)
9. Bruna, J., Zaremba, W., Szlam, A., LeCun, Y.: Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203 (2013)
10. Buxbaum, J., Cicek, E., Devlin, B., Klei, L., Roeder, K., De Rubeis, S.: Combining autism and intellectual disability exome data implicates disruption of neocortical development in both disorders. European Neuropsychopharmacology **27**, S437 (2017)
11. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., Ma'ayan, A.: Enrichr: interactive and collaborative html5 gene list enrichment analysis tool. BMC bioinformatics **14**(1), 128 (2013)
12. Chiurazzi, P., Pirozzi, F.: Advances in understanding–genetic basis of intellectual disability. F1000Research **5** (2016)
13. Cleaver, R., Berg, J., Craft, E., Foster, A., Gibbons, R.J., Hobson, E., Lachlan, K., Naik, S., Sampson, J.R., Sharif, S., et al.: Refining the primrose syndrome phenotype: a study of five patients with zbtb20 de novo variants and a review of the literature. American Journal of Medical Genetics Part A **179**(3), 344–349 (2019)
14. Cordeddu, V., Redeker, B., Stellacci, E., Jongejan, A., Fragale, A., Bradley, T.E., Anselmi, M., Ciolfi, A., Cecchetti, S., Muto, V., et al.: Mutations in zbtb20 cause primrose syndrome. Nature genetics **46**(8), 815–817 (2014)
15. Cotney, J., Muhle, R.A., Sanders, S.J., Liu, L., Willsey, A.J., Niu, W., Liu, W., Klei, L., Lei, J., Yin, J., et al.: The autism-associated chromatin modifier chd8 regulates other autism risk genes during human neurodevelopment. Nature communications **6**, 6404 (2015)
16. Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y.S., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al.: Fmrp stalls ribosomal translocation on mrnas linked to synaptic function and autism. Cell **146**(2), 247–261 (2011)
17. De Ligt, J., Willemsen, M.H., Van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., De Vries, P., Gilissen, C., et al.: Diagnostic exome sequencing in persons with severe intellectual disability. New England Journal of Medicine **367**(20), 1921–1929 (2012)
18. De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al.: Synaptic, transcriptional and chromatin genes disrupted in autism. Nature **515**(7526), 209–215 (2014)
19. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: Advances in neural information processing systems. pp. 3844–3852 (2016)
20. Devlin, B., Melhem, N., Roeder, K.: Do common variants play a role in risk for autism? evidence and theoretical musings. Brain research **1380**, 78–84 (2011)
21. Duda, M., Zhang, H., Li, H.D., Wall, D.P., Burmeister, M., Guan, Y.: Brain-specific functional relationship networks inform autism spectrum disorder gene prediction. Translational psychiatry **8**(1), 1–9 (2018)
22. Duvenaud, D.K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., Adams, R.P.: Convolutional networks on graphs for learning molecular fingerprints. In: Advances in neural information processing systems. pp. 2224–2232 (2015)
23. Feliciano, P., Daniels, A.M., Snyder, L.G., Beaumont, A., Camba, A., Esler, A., Gulsrud, A.G., Mason, A., Gutierrez, A., Nicholson, A., et al.: Spark: a us cohort of 50,000 families to accelerate autism research. Neuron **97**(3), 488–493 (2018)
24. Firth, H.V., Wright, C.F., study, D.: The deciphering developmental disorders (ddd) study. Developmental Medicine & Child Neurology **53**(8), 702–703 (2011)
25. Gandal, M.J., Haney, J.R., Parikshak, N.N., Leppa, V., Ramaswami, G., Hartl, C., Schork, A.J., Appadurai, V., Buil, A., Werge, T.M., et al.: Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. Science **359**(6376), 693–697 (2018)
26. Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., et al.: Genome sequencing identifies major causes of severe intellectual disability. Nature **511**(7509), 344 (2014)

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

27. Gilman, S.R., Chang, J., Xu, B., Bawa, T.S., Gogos, J.A., Karayiorgou, M., Vitkup, D.: Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. Nature neuroscience **15**(12), 1723–1728 (2012)
28. Gilman, S.R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., Vitkup, D.: Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron **70**(5), 898–907 (2011)
29. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256 (2010)
30. Gonzalez-Mantilla, A.J., Moreno-De-Luca, A., Ledbetter, D.H., Martin, C.L.: A cross-disorder method to identify novel candidate genes for developmental brain disorders. JAMA psychiatry **73**(3), 275–283 (2016)
31. Goytain, A., Hines, R.M., Quamme, G.A.: Functional characterization of nipa2, a selective mg2+ transporter. American Journal of Physiology-Cell Physiology (2008)
32. Greene, C.S., Krishnan, A., Wong, A.K., Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C., et al.: Understanding multicellular function and disease with human tissue-specific networks. Nature genetics **47**(6), 569 (2015)
33. Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al.: Identification of common genetic risk variants for autism spectrum disorder. Nature genetics **51**(3), 431 (2019)
34. Grozeva, D., Carss, K., Spasic-Boskovic, O., Tejada, M.I., Gecz, J., Shaw, M., Corbett, M., Haan, E., Thompson, E., Friend, K., et al.: Targeted next-generation sequencing analysis of 1,000 individuals with intellectual disability. Human mutation **36**(12), 1197–1204 (2015)
35. Halvardson, J., Zhao, J.J., Zaghlool, A., Wentzel, C., Georgii-Hemming, P., Månsson, E., Ederth Sävmarker, H., Brandberg, G., Soussi Zander, C., Thuresson, A.C., Feuk, L.: Mutations in hecw2 are associated with intellectual disability and epilepsy. Journal of Medical Genetics **53**(10), 697–704 (2016). https://doi.org/10.1136/jmedgenet-2016-103814, `https://jmg.bmj.com/content/53/10/697`
36. He, X., Sanders, S.J., Liu, L., Rubeis, S.D., Lim, E.T., Sutcliffe, J.S., Schellenberg, G.D., Gibbs, R.A., Daly, M.J., Buxbaum, J.D., et al.: Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. PLoS Genetics **9**(8) (2013). https://doi.org/10.1371/journal.pgen.1003671
37. Henaff, M., Bruna, J., LeCun, Y.: Deep convolutional networks on graph-structured data. arXiv preprint arXiv:1506.05163 (2015)
38. Hormozdiari, F., Penn, O., Borenstein, E., Eichler, E.E.: The discovery of integrated gene networks for autism and related disorders. Genome Research **25**(1), 142–154 (May 2014). https://doi.org/10.1101/gr.178855.114
39. Inlow, J.K., Restifo, L.L.: Molecular and comparative genetics of mental retardation. Genetics **166**(2), 835–881 (2004)
40. Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A., Kendall, J., Grabowska, E., Ma, B., Marks, S., Rodgers, L., Stepansky, A., Troge, J., Andrews, P., Bekritsky, M., Pradhan, K., Ghiban, E., Kramer, M., Parla, J., Demeter, R., Fulton, L.L., Fulton, R.S., Magrini, V.J., Ye, K., Darnell, J.C., Darnell, R.B.: De novogene disruptions in children on the autistic spectrum. Neuron **74** (2012). https://doi.org/10.1016/j.neuron.2012.04.009, `https://doi.org/10.1016/j.neuron.2012.04.009`
41. Iossifov, I., O?Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al.: The contribution of de novo coding mutations to autism spectrum disorder. Nature **515**(7526), 216–221 (2014)
42. Iwase, S., Bérubé, N.G., Zhou, Z., Kasri, N.N., Battaglioli, E., Scandaglia, M., Barco, A.: Epigenetic etiology of intellectual disability. Journal of Neuroscience **37**(45), 10773–10782 (2017)
43. Kalkman, H.O.: A review of the evidence for the canonical wnt pathway in autism spectrum disorders. Molecular autism **3**(1), 10 (2012)
44. Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., Guennel, T., Shin, Y., Johnson, M.B., Krsnik, Z., Mayer, S., Fertuzinhos, S., Umlauf, S., Lisgo, S.N., Vortmeyer, A., Weinberger, D.R., Mane, S., Hyde, T.M., Huttner, A., Reimers, M., Kleinman, J.E., Sestan, N.: Spatio-temporal transcriptome of the human brain. Nature **478** (2011). https://doi.org/10.1038/nature10523, `https://doi.org/10.1038/nature10523`
45. Kaplanis, J., Samocha, K.E., Wiel, L., Zhang, Z., Arvai, K.J., Eberhardt, R.Y., Gallone, G., Lelieveld, S.H., Martin, H.C., McRae, J.F., et al.: Evidence for 28 genetic disorders discovered by combining healthcare and research data. Nature **586**(7831), 757–762 (2020)
46. King, I.F., Yandava, C.N., Mabb, A.M., Hsiao, J.S., Huang, H.S., Pearson, B.L., Calabrese, J.M., Starmer, J., Parker, J.S., Magnuson, T., et al.: Topoisomerases facilitate transcription of long genes linked to autism. Nature **501**(7465), 58–62 (2013)
47. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

48. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
49. Krishnan, A., Zhang, R., Yao, V., Theesfeld, C.L., Wong, A.K., Tadych, A., Volfovsky, N., Packer, A., Lash, A., Troyanskaya, O.G.: Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. Nature neuroscience **19**(11), 1454–1462 (2016)
50. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
51. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., McDermott, M.G., Monteiro, C.D., Gundersen, G.W., Ma'ayan, A.: Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research **44**(W1), W90–W97 (2016). https://doi.org/10.1093/nar/gkw377, `+http://dx.doi.org/10.1093/nar/gkw377`
52. Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., Ma'ayan, A.: Chea: transcription factor regulation inferred from integrating genome-wide chip-x experiments. Bioinformatics **26**(19), 2438–2444 (2010)
53. Lee, P.H., Anttila, V., Won, H., Feng, Y.C.A., Rosenthal, J., Zhu, Z., Tucker-Drob, E.M., Nivard, M.G., Grotzinger, A.D., Posthuma, D., et al.: Genome wide meta-analysis identifies genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. bioRxiv p. 528117 (2019)
54. Li, J., Cai, T., Jiang, Y., Chen, H., He, X., Chen, C., Li, X., Shao, Q., Ran, X., Li, Z., et al.: Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from npdenovo database. Molecular psychiatry **21**(2), 290 (2016)
55. Liu, L., Lei, J., Roeder, K.: Network assisted analysis to reveal the genetic basis of autism. The annals of applied statistics **9**(3), 1571 (2015)
56. Liu, L., Lei, J., Sanders, S.J., Willsey, A.J., Kou, Y., Cicek, A.E., Klei, L., Lu, C., He, X., Li, M., Muhle, R.A., Ma'ayan, A., Noonan, J.P., Šestan, N., McFadden, K.A., State, M.W., Buxbaum, J.D., Devlin, B., Roeder, K.: Dawn: a framework to identify autism genes and subnetworks using gene expression and genetics. Molecular Autism **5**(1), 22 (Mar 2014). https://doi.org/10.1186/2040-2392-5-22, `https://doi.org/10.1186/2040-2392-5-22`
57. Ma, D., Salyakina, D., Jaworski, J.M., Konidari, I., Whitehead, P.L., Andersen, A.N., Hoffman, J.D., Slifer, S.H., Hedges, D.J., Cukier, H.N., et al.: A genome-wide association study of autism reveals a common novel risk locus at 5p14. 1. Annals of human genetics **73**(3), 263–273 (2009)
58. Masoudnia, S., Ebrahimpour, R.: Mixture of experts: a literature survey. Artificial Intelligence Review **42**(2), 275–293 (2014)
59. Mefford, H.C., Cooper, G.M., Zerr, T., Smith, J.D., Baker, C., Shafer, N., Thorland, E.C., Skinner, C., Schwartz, C.E., Nickerson, D.A., et al.: A method for rapid, targeted cnv genotyping identifies rare variants associated with neurocognitive disease. Genome research **19**(9), 1579–1585 (2009)
60. Mitchell, K.J.: The genetics of neurodevelopmental disease. Current opinion in neurobiology **21**(1), 197–203 (2011)
61. Moreno-De-Luca, A., Myers, S.M., Challman, T.D., Moreno-De-Luca, D., Evans, D.W., Ledbetter, D.H.: Developmental brain dysfunction: revival and expansion of old concepts based on new genetic evidence. The Lancet Neurology **12**(4), 406–414 (2013)
62. Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V., Polak, P., Yoon, S., Maguire, J., Crawford, E.L., Campbell, N.G., Geller, E.T., Valladares, O., Schafer, C., Liu, H., Zhao, T., Cai, G., Lihm, J., Dannenfelser, R., Jabado, O., Peralta, Z., Nagaswamy, U., Muzny, D., Reid, J.G., Newsham, I., Wu, Y.: Patterns and rates of exonicde novomutations in autism spectrum disorders. Nature **485** (2012). https://doi.org/10.1038/nature11011, `https://doi.org/10.1038/nature11011`
63. Nguyen, H.T., Bryois, J., Kim, A., Dobbyn, A., Huckins, L.M., Munoz-Manchado, A.B., Ruderfer, D.M., Genovese, G., Fromer, M., Xu, X., et al.: Integrated bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and neurodevelopmental disorders. Genome medicine **9**(1), 114 (2017)
64. Nguyen, T.H., Dobbyn, A., Brown, R.C., Riley, B.P., Buxbaum, J.D., Pinto, D., Purcell, S.M., Sullivan, P.F., He, X., Stahl, E.A.: mtada is a framework for identifying risk genes from de novo mutations in multiple traits. Nature Communications **11**(1), 1–12 (2020)
65. Norman, U., Cicek, A.E.: St-steiner: a spatio-temporal gene discovery algorithm. Bioinformatics **35**(18), 3433–3440 (2019)
66. O'Roak, B.J., Vives, L., Fu, W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., Munson, J., Hiatt, J.B., Turner, E.H., Levy, R., O'Day, D.R., Krumm, N., Coe, B.P., Martin, B.K., Borenstein, E., Nickerson, D.A., Mefford, H.C., Doherty, D., Akey, J.M., Bernier, R., Eichler, E.E., Shendure, J.: Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science **338** (2012). https://doi.org/10.1126/science.1227764, `https://doi.org/10.1126/science.1227764`

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

67. O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., Turner, E.H., Stanaway, I.B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J.M., Borenstein, E., Rieder, M.J., Nickerson, D.A., Bernier, R., Shendure, J., Eichler, E.E.: Sporadic autism exomes reveal a highly interconnected protein network ofde novomutations. Nature **485** (2012). https://doi.org/10.1038/nature10989, `https://doi.org/10.1038/nature10989`
68. Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., et al.: Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. The American Journal of Human Genetics **94**(5), 677–694 (2014)
69. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., et al.: Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. The Lancet **380**(9854), 1674–1682 (2012)
70. Robinson, E.B., Samocha, K.E., Kosmicki, J.A., McGrath, L., Neale, B.M., Perlis, R.H., Daly, M.J.: Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. Proceedings of the National Academy of Sciences **111**(42), 15161–15165 (2014)
71. Roeper, J.: Closing gaps in brain disease—from overlapping genetic architecture to common motifs of synapse dysfunction. Current opinion in neurobiology **48**, 45–51 (2018)
72. Rosina, E., Battan, B., Siracusano, M., Di Criscio, L., Hollis, F., Pacini, L., Curatolo, P., Bagni, C.: Disruption of mtor and mapk pathways correlates with severity in idiopathic autism. Translational psychiatry **9**(1), 1–10 (2019)
73. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., Walker, M.F., Ober, G.T., Teran, N.A., Song, Y., El-Fishawy, P., Murtha, R.C., Choi, M., Overton, J.D., Bjornson, R.D., Carriero, N.J., Meyer, K.A., Bilguvar, K., Mane, S.M., Sestan, N., Lifton, R.P., Gunel, M., Roeder, K., Geschwind, D.H., Devlin, B., State, M.W.: De novomutations revealed by whole-exome sequencing are strongly associated with autism. Nature **485** (2012)
74. Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al.: Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron **87**(6), 1215–1233 (2015)
75. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y., Peng, M., Collins, R.L., Grove, J., Klei, L., et al.: Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism (2019)
76. Seabra, M.C., Mules, E.H., Hume, A.N.: Rab gtpases, intracellular traffic and disease. Trends in molecular medicine **8**(1), 23–30 (2002)
77. Srivastava, S., Cohen, J.S., Vernon, H., Barañano, K., McClellan, R., Jamal, L., Naidu, S., Fatemi, A.: Clinical whole exome sequencing in child neurology practice. Annals of neurology **76**(4), 473–483 (2014)
78. Sunkin, S.M., Ng, L., Lau, C., Dolbeare, T., Gilbert, T.L., Thompson, C.L., Hawrylycz, M., Dang, C.: Allen brain atlas: an integrated spatio-temporal portal for exploring the central nervous system. Nucleic acids research **41**(D1), D996–D1008 (2012)
79. Vissers, L.E., Gilissen, C., Veltman, J.A.: Genetic studies in intellectual disability and related disorders. Nature Reviews Genetics **17**(1), 9 (2016)
80. Vissers, L.E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., et al.: A de novo paradigm for mental retardation. Nature genetics **42**(12), 1109 (2010)
81. Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., Geschwind, D.H.: Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature **474**(7351), 380–384 (2011)
82. Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q., et al.: Hits-clip and integrative modeling define the rbfox splicing-regulatory network linked to brain development and autism. Cell reports **6**(6), 1139–1152 (2014)
83. Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., Murtha, M., Bichsel, C., Niu, W., Cotney, J., Ercan-Sencicek, A.G., Gockley, J., Gupta, A.R., Han, W., He, X., Hoffman, E.J., Klei, L., Lei, J., Liu, W., Liu, L., Lu, C., Xu, X., Zhu, Y., Mane, S.M., Lein, E.S., Wei, L.: Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell **155** (2013). https://doi.org/10.1016/j.cell.2013.10.020, `https://doi.org/10.1016/j.cell.2013.10.020`
84. Zhang, C., Shen, Y.: A cell type-specific expression signature predicts haploinsufficient autism-susceptibility genes. Human Mutation **38**(2), 204–215 (2 2017). https://doi.org/10.1002/humu.23147, `http:https://doi.org/10.1002/humu.23147`
85. Zhou, G., Soufan, O., Ewald, J., Hancock, R.E., Basu, N., Xia, J.: Networkanalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic acids research **47**(W1), W234–W241 (2019)

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.13.150201; this version posted May 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

86. Zhu, X., Petrovski, S., Xie, P., Ruzzo, E.K., Lu, Y.F., McSweeney, K.M., Ben-Zeev, B., Nissenkorn, A., Anikster, Y., Oz-Levi, D., et al.: Whole-exome sequencing in undiagnosed genetic diseases: interpreting 119 trios. Genetics in Medicine **17**(10), 774 (2015)
87. Zoghbi, H.Y., Bear, M.F.: Synaptic dysfunction in neurodevelopmental disorders associated with autism and intellectual disabilities. Cold Spring Harbor perspectives in biology **4**(3), a009886 (2012)