



HAL
open science

Rare deleterious mutations of HNRNP genes result in shared neurodevelopmental disorders.

M. A. Gillentine, T. Wang, K. Hoekzema, J. Rosenfeld, P. Liu, H. Guo, C. N. Kim, B. B. A. de Vries, L. E. L. M. Vissers, M. Nordenskjold, et al.

► To cite this version:

M. A. Gillentine, T. Wang, K. Hoekzema, J. Rosenfeld, P. Liu, et al.. Rare deleterious mutations of HNRNP genes result in shared neurodevelopmental disorders.. *Genome Medicine*, 2021, *Genome Medicine*, 13 (63), 10.1186/s13073-021-00870-6 . hal-04219731

HAL Id: hal-04219731

<https://hal.univ-lille.fr/hal-04219731>

Submitted on 27 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH

Open Access

Rare deleterious mutations of *HNRNP* genes result in shared neurodevelopmental disorders



Madelyn A. Gillentine¹, Tianyun Wang¹, Kendra Hoekzema¹, Jill Rosenfeld^{2,3}, Pengfei Liu², Hui Guo^{1,4}, Chang N. Kim^{5,6,7,8}, Bert B. A. De Vries⁹, Lisenka E. L. M. Vissers⁹, Magnus Nordenskjold^{10,11}, Malin Kvarnung^{10,11}, Anna Lindstrand^{10,11}, Ann Nordgren^{10,11}, Jozef Gecz^{12,13,14}, Maria Iacone¹⁵, Anna Cereda¹⁶, Agnese Scatigno¹⁶, Silvia Maitz¹⁷, Ginevra Zanni¹⁸, Enrico Bertini¹⁸, Christiane Zweier¹⁹, Sarah Schuhmann¹⁹, Antje Wiesener¹⁹, Micah Pepper^{20,21}, Heena Panjwani^{20,21}, Erin Torti²², Farida Abid^{23,24}, Irina Anselm²⁵, Siddharth Srivastava²⁵, Paldeep Atwal²⁶, Carlos A. Bacino³, Gifty Bhat²⁷, Katherine Cobian²⁷, Lynne M. Bird^{28,29}, Jennifer Friedman^{28,30,31}, Meredith S. Wright^{28,30}, Bert Callewaert³², Florence Petit³³, Sophie Mathieu³⁴, Alexandra Afenjar³⁴, Celenie K. Christensen³⁵, Kerry M. White³⁶, Orly Elpeleg³⁷, Itai Berger^{38,39}, Edward J. Espineli^{23,24}, Christina Fagerberg⁴⁰, Charlotte Brasch-Andersen⁴⁰, Lars Kjærsgaard Hansen⁴¹, Timothy Feyma⁴², Susan Hughes^{43,44}, Isabelle Thiffault^{44,45}, Bonnie Sullivan⁴³, Shuang Yan⁴³, Kory Keller⁴⁶, Boris Keren⁴⁷, Cyril Mignot⁴⁷, Frank Kooy⁴⁸, Marije Meuwissen⁴⁸, Alice Basinger⁴⁹, Mary Kukulich⁴⁹, Meredith Philips⁴⁹, Lucia Ortega⁴⁹, Margaret Drummond-Borg⁴⁹, Mathilde Lauridsen⁴⁰, Kristina Sorensen⁴⁰, Anna Lehman^{50,51}, CAUSES Study⁵¹, Elena Lopez-Rangel^{50,52,53}, Paul Levy⁵⁴, Davor Lessel⁵⁵, Timothy Lotze²³, Suneeta Madan-Khetarpal^{56,57}, Jessica Sebastian⁵⁶, Jodie Vento⁵⁶, Divya Vats⁵⁸, L. Manace Benman⁵⁹, Shane Mckee⁶⁰, Ghayda M. Mirzaa^{61,62,63}, Candace Muss⁶⁴, John Pappas⁶⁵, Hilde Peeters⁶⁶, Corrado Romano⁶⁷, Maurizio Elia⁶⁷, Ornella Galesi⁶⁷, Marleen E. H. Simon⁶⁸, Koen L. I. van Gassen⁶⁸, Kara Simpson⁶⁹, Robert Stratton⁷⁰, Sabeen Syed⁷¹, Julien Thevenon⁷², Irene Valenzuela Palafoll⁷³, Antonio Vitobello^{74,75}, Marie Bournez^{76,77}, Laurence Faivre^{75,77}, Kun Xia⁴, SPARK Consortium, Rachel K. Earl^{20,21,78}, Tomasz Nowakowski^{5,6,7,8}, Raphael A. Bernier^{20,21,78} and Evan E. Eichler^{1,79*} 

Abstract

Background: With the increasing number of genomic sequencing studies, hundreds of genes have been implicated in neurodevelopmental disorders (NDDs). The rate of gene discovery far outpaces our understanding of genotype–phenotype correlations, with clinical characterization remaining a bottleneck for understanding NDDs. Most disease-associated Mendelian genes are members of gene families, and we hypothesize that those with
(Continued on next page)

* Correspondence: eee@gs.washington.edu

¹Department of Genome Sciences, University of Washington School of Medicine, 3720 15th Ave NE S413A, Box 355065, Seattle, WA 981095-5065, USA

⁷⁹Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

related molecular function share clinical presentations.

Methods: We tested our hypothesis by considering gene families that have multiple members with an enrichment of de novo variants among NDDs, as determined by previous meta-analyses. One of these gene families is the heterogeneous nuclear ribonucleoproteins (hnRNPs), which has 33 members, five of which have been recently identified as NDD genes (*HNRNPK*, *HNRNPU*, *HNRNPH1*, *HNRNPH2*, and *HNRNPR*) and two of which have significant enrichment in our previous meta-analysis of probands with NDDs (*HNRNPU* and *SYNCRIP*). Utilizing protein homology, mutation analyses, gene expression analyses, and phenotypic characterization, we provide evidence for variation in 12 *HNRNP* genes as candidates for NDDs. Seven are potentially novel while the remaining genes in the family likely do not significantly contribute to NDD risk.

Results: We report 119 new NDD cases (64 de novo variants) through sequencing and international collaborations and combined with published clinical case reports. We consider 235 cases with gene-disruptive single-nucleotide variants or indels and 15 cases with small copy number variants. Three hnRNP-encoding genes reach nominal or exome-wide significance for de novo variant enrichment, while nine are candidates for pathogenic mutations. Comparison of *HNRNP* gene expression shows a pattern consistent with a role in cerebral cortical development with enriched expression among radial glial progenitors. Clinical assessment of probands ($n = 188$ – 221) expands the phenotypes associated with *HNRNP* rare variants, and phenotypes associated with variation in the *HNRNP* genes distinguishes them as a subgroup of NDDs.

Conclusions: Overall, our novel approach of exploiting gene families in NDDs identifies new *HNRNP*-related disorders, expands the phenotypes of known *HNRNP*-related disorders, strongly implicates disruption of the hnRNPs as a whole in NDDs, and supports that NDD subtypes likely have shared molecular pathogenesis. To date, this is the first study to identify novel genetic disorders based on the presence of disorders in related genes. We also perform the first phenotypic analyses focusing on related genes. Finally, we show that radial glial expression of these genes is likely critical during neurodevelopment. This is important for diagnostics, as well as developing strategies to best study these genes for the development of therapeutics.

Keywords: Neurodevelopmental disorders, hnRNPs, Cortex development, Gene families

Background

Among the hundreds of candidate genes proposed for neurodevelopmental disorders (NDDs), genes involved in RNA metabolic processing and regulation of gene expression have been shown to be enriched for de novo variants (DNVs) among probands with NDDs [1]. RNA processing (splicing, transport, localization, translation, and degradation) is critically important for brain development and function, as neurons are post-mitotic cells dependent on RNA expression, as well as spatiotemporal isoform specificity, for individual growth and functionality [2]. To successfully regulate RNA processing and protein synthesis, over 500 RNA-binding proteins (RBPs) in humans are abundantly and ubiquitously expressed, found primarily in the nucleus [3]. Although ubiquitous, there are tissue-specific changes in alternative splicing from RBPs resulting in cell-specific phenotypes [4]. As RBPs are necessary for many steps of neuronal RNA metabolism, there are multiple opportunities for dysfunction, which is highlighted by the range of neurological phenotypes resulting from variation in RBP-encoding genes, including neurodegenerative diseases, muscular atrophies, and various cancers. RBPs have also been

implicated in NDDs, most notably *FMRI* in fragile X syndrome.

Here, we hypothesize that variation in gene families with related structure and function in the brain will result in subtypes of NDDs with related pathology. With over 80% of Mendelian disease-associated genes being part of gene families and/or having functionally redundant paralogs, this provides an opportunity to divide many NDD candidate genes into subgroups [5, 6]. In fact, it has recently been shown that DNVs are enriched among a subset of gene families in probands with NDDs [7]. This approach aids in understanding the biological impacts of variation in groups of genes, provides an opportunity for gene discovery, and allows for development of gene/protein family-specific therapeutics impacting a larger number of individuals than when targeting a single gene.

Focusing on RNA processing, we applied this gene family approach to identify gene families implicated in NDDs. Of the four gene families identified from previous meta-analyses of NDD exomes, the heterogeneous nuclear ribonucleoproteins (hnRNPs for proteins, *HNRNPs* for genes) stand out as strong NDD

candidate genes that have yet to be fully investigated, as the fewest members of the family are associated with known disorders (Table 1). The hnRNPs are a large family of RBPs consisting of 33 core and minor members implicated in many steps of RNA processing. Several, primarily through changes in expression or localization, have been associated with neurodegenerative disorders, and, more recently, five (*HNRNPH1*, *HNRNPH2*, *HNRNPK*, *HNRNPR*, and *HNRNPU*) have been described as having DNVs among probands with NDDs (Table 2). We have also previously shown that two *HNRNPs* (*HNRNPU* and *SYNCRIP*) have a significant enrichment of DNVs among probands with NDDs, making them very strong candidate genes [1]. Supporting their role in NDDs as a gene family, it is known that hnRNPs function cooperatively and compensatorily, suggesting that disruption among them may result in similar phenotypic consequences. Thus, we hypothesize that there may be additional hnRNPs with shared structure and function that impact neurodevelopment, resulting in shared phenotypes. Overall, the hnRNPs are just one example of how multiple members of a gene family can be involved in related NDDs. Recent data from large-scale sequencing efforts suggest that other gene families (e.g., chromodomain DNA-binding helicase gene family) may benefit from similar coordinated investigations.

Here, through international collaborations, protein, mutation and expression analyses, and an exhaustive literature search, we identify additional *HNRNPs* with pathogenic likely gene-disrupting (LGD) and severe missense single-nucleotide variants (SNVs), indels, or small copy number variants (CNVs) among probands with NDDs while utilizing novel methods of grouping genetic disorders by gene family instead of broader gene function. We compare findings from a total of 250 probands with NDDs with de novo or likely de novo SNVs/CNVs: 122 novel cases (118 SNVs or small indels and four chromosomal deletions) and 128 published probands (117 SNVs or small indels and 11 CNVs). Of these, 240 probands have variants occurring in 12 *HNRNPs* where three or more probands have been identified (termed NDD *HNRNPs*:

HNRNPAB, *HNRNPD*, *HNRNPF*, *HNRNPH1*, *HNRNPH2*, *HNRNPH3*, *HNRNPK*, *HNRNPR*, *SYNCRIP*, *HNRNPU*, *HNRNPUL1*, and *HNRNPUL2*). We establish variation in a subset of the *HNRNPs* as a risk factor for NDDs and highlight the utility of a gene family-based approach to identify NDD-related disorders.

Methods

Identification of candidate NDD *HNRNPs*

Literature review to determine NDD candidate *HNRNPs*

First, *HNRNPs* with published cases were identified in large exome or genome sequencing studies, as identified by denovo-db v1.6 or case reports found by searching PubMed with all known gene aliases ($n = 128$; Fig. 1, Additional File 1: Table S1) [11–18, 20–61]. Cohorts in denovo-db v1.6 include ASD samples ($n = 5886$: including the Simons Simplex collection [SSC, $n = 2508$] and MSSNG [$n = 1625$]), DD/ID samples ($n = 1010$), and epilepsy samples ($n = 532$, including the Epi4K Consortium [$n = 264$]). Samples were also identified from the Autism Sequencing Consortium (ASC, $n = 8157$) and the Epilepsy Genetics Initiative ($n = 166$). We also identified samples in the Deciphering Developmental Disorders (DDD) cohort, described below as we recalled the variants using our own analyses [39, 49, 62, 63]. In total, 51,616 published individuals were screened. Cases had to have a variant in an *HNRNP* to be included (i.e., not solely changes in gene expression as seen in some neurodegenerative phenotypes). Publications had to include at least one of the disorders under the HPO terms “abnormality of the nervous system physiology” (HP:0012638) and/or “abnormality of the nervous system morphology” (HP:0012639). This includes, but is not limited to, developmental delay/intellectual disability (DD/ID), seizures, and behavioral abnormalities such as autism spectrum disorder (ASD). Note that neurodegenerative phenotypes, such as ALS, were not included as part of this analysis. Clinical information was requested from authors when possible for larger sequencing studies. Duplicates were removed based on sample ID, sharing of multiple variants, and disclosure of duplicates in publications. When possible, samples were excluded if

Table 1 Gene families involved in RNA processing reaching FDR significance in Coe et al. [1] and their role in disease

Gene family	Genes reaching FDR significance	# of NDD candidate genes in gene family
Chromodomain DNA-binding proteins	<i>CHD2</i> , <i>CHD3</i> , <i>CHD4</i> , <i>CHD7</i> , <i>CHD8</i>	6/10
BAF complex	<i>ARID1B</i> , <i>BCL11A</i> , <i>SMARCA2</i> , <i>SMARCA4</i> , <i>SMARCD1</i>	16/24
Heterogeneous nuclear ribonuclear proteins	<i>HNRNPU</i> , <i>SYNCRIP</i> , <i>HNRNPD</i> , <i>HNRNPK</i>	7/33
Lysine acetyltransferases	<i>CREBBP</i> , <i>DLG4</i> , <i>EP300</i> , <i>TAF1</i>	5/17

Gene families were defined by the HUGO Gene Nomenclature Consortium. At least three members of the gene family must reach FDR significance to be included. NDD candidate genes were determined by OMIM and a literature search. Coe et al. [1] $n = 11,722$ exomes

Table 2 Previous disease associations for genetic variation in *HNRNP* genes

Gene	Disorder	Type of variation
<i>HNRNPA1</i>	ALS/FTL [8] Multisystem proteinopathy [8, 9]	MIS MIS
<i>HNRNPA2/B1</i>	ALS/FTLD [8, 9] Multisystem proteinopathy [8, 9]	MIS MIS
<i>HNRNPA3</i>	ALS/FTLD [8, 9] Multisystem proteinopathy [8, 9]	MIS MIS
<i>HNRNPD/AUF1</i>	4q21 microdeletion/duplication/triplication syndrome [10]	CNVs
<i>HNRNPH1</i>	<i>HNRNPH1</i> -related syndromic ID [11]	MIS20/small CNVs
<i>HNRNPH2</i>	Bain-type ID [12]	MIS20
<i>HNRNPK</i>	AKS/Okamoto syndrome [13, 14] Kabuki syndrome [15–17]	LGD/MIS20/chromosomal deletions MIS
<i>HNRNPR</i>	<i>HNRNPR</i> -related disorder [18]	LGD/MIS20
<i>SYNCRIP/HNRNPQ</i>	6q proximal deletions [19]	Chromosomal deletion
<i>HNRNPP/FUS/TLS</i>	ALS/FTLD [3]	LGD/MIS
<i>HNRNPU</i>	1q43q44 microdeletion syndrome [20] <i>HNRNPU</i> -related disorder [20–23]	Chromosomal deletion LGD/MIS20/chromosomal duplication

MIS missense, MIS20 missense variants with CADD scores ≥ 20 , indicating they are in the top 1% of likely pathogenic variants, CNVs copy number variants, LGD likely gene disrupting, ALS amyotrophic lateral sclerosis, FTLD frontotemporal lobar degeneration, ID intellectual disability, AKS Au-Kline syndrome

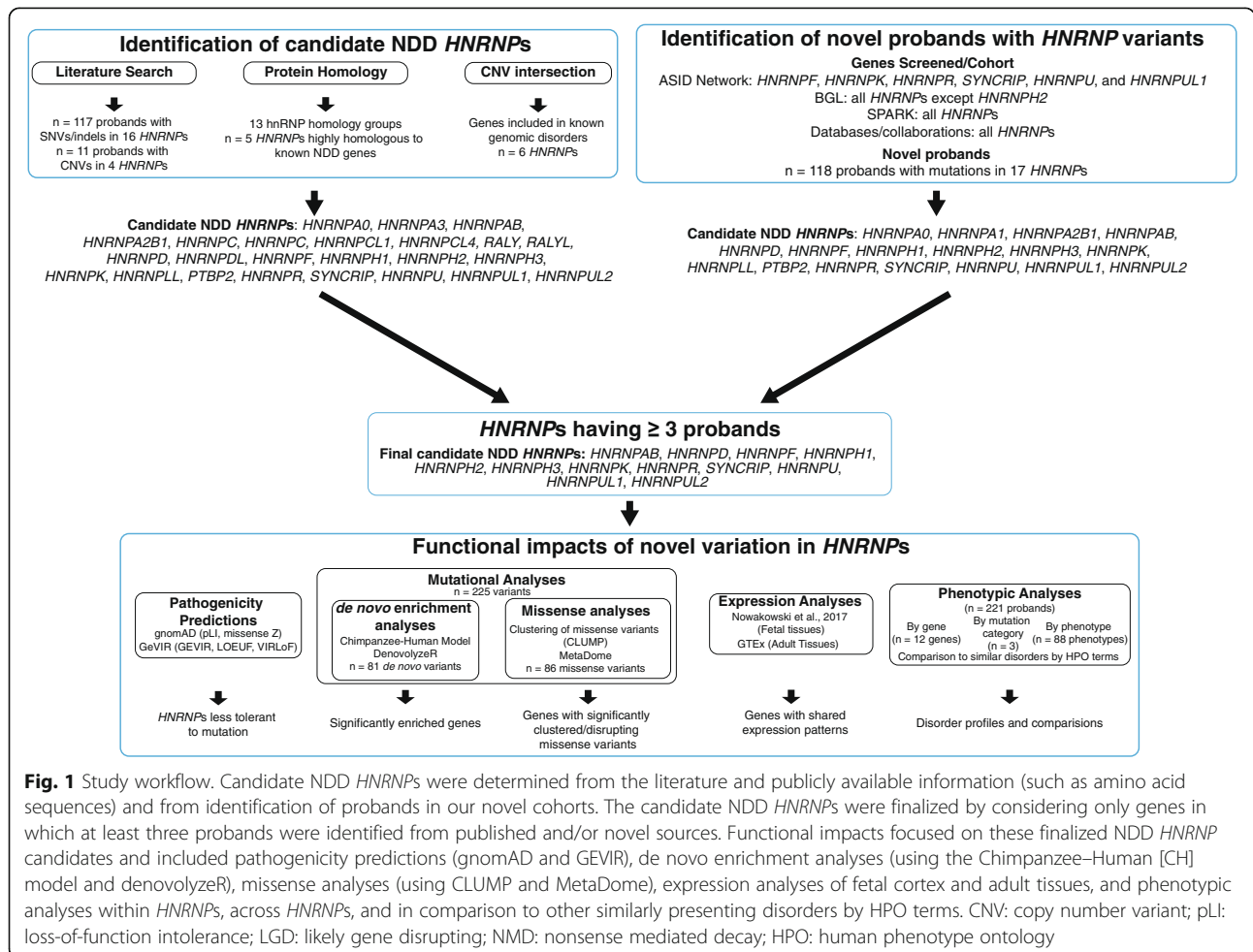


Fig. 1 Study workflow. Candidate NDD *HNRNP*s were determined from the literature and publicly available information (such as amino acid sequences) and from identification of probands in our novel cohorts. The candidate NDD *HNRNP*s were finalized by considering only genes in which at least three probands were identified from published and/or novel sources. Functional impacts focused on these finalized NDD *HNRNP* candidates and included pathogenicity predictions (gnomAD and GeVIR), de novo enrichment analyses (using the Chimpanzee-Human [CH] model and denovolyzeR), missense analyses (using CLUMP and MetaDome), expression analyses of fetal cortex and adult tissues, and phenotypic analyses within *HNRNP*s, across *HNRNP*s, and in comparison to other similarly presenting disorders by HPO terms. CNV: copy number variant; pLI: loss-of-function intolerance; LGD: likely gene disrupting; NMD: nonsense mediated decay; HPO: human phenotype ontology

they had a likely causative variant in an NDD candidate gene, as determined by multiple publications [1, 58].

Protein homology analysis to identify additional candidate NDD HNRNPs

To determine if there were a subset of hnRNPs with sequence similarity to known NDD hnRNPs, and thus potential candidate NDD genes, we utilized protein homology analysis (Fig. 2 and Additional File 2: Fig. S1). Protein homology was determined using Clustal Omega with canonical transcript sequences obtained from UniProt and clustering was performed in R (v.3.6.1) using the corrplot package (v.0.84) to identify shared protein homology.

CNV intersection

In order to determine if any *HNRNPs* are dosage sensitive, we intersected the *HNRNPs* with a list of 58 genomic disorders based on previous CNV morbidity maps and the DECIPHER database (Table 3) [10, 13, 19, 20, 64–67]. From this, we identified 15 probands with CNVs spanning only an NDD *HNRNP* gene or neighboring genes not predicted to be haploinsufficient.

Identification of novel probands with *HNRNP* variants

Probands were collected from multiple cohorts ($n = 32,359$ probands) (Additional File 1: Table S1). We also queried ClinVar and DECIPHER databases. We identified 122 probands with SNVs, indels, or CNVs impacting 14 *HNRNPs* (Fig. 3, Additional File 1: Tables S7, S10 and S11).

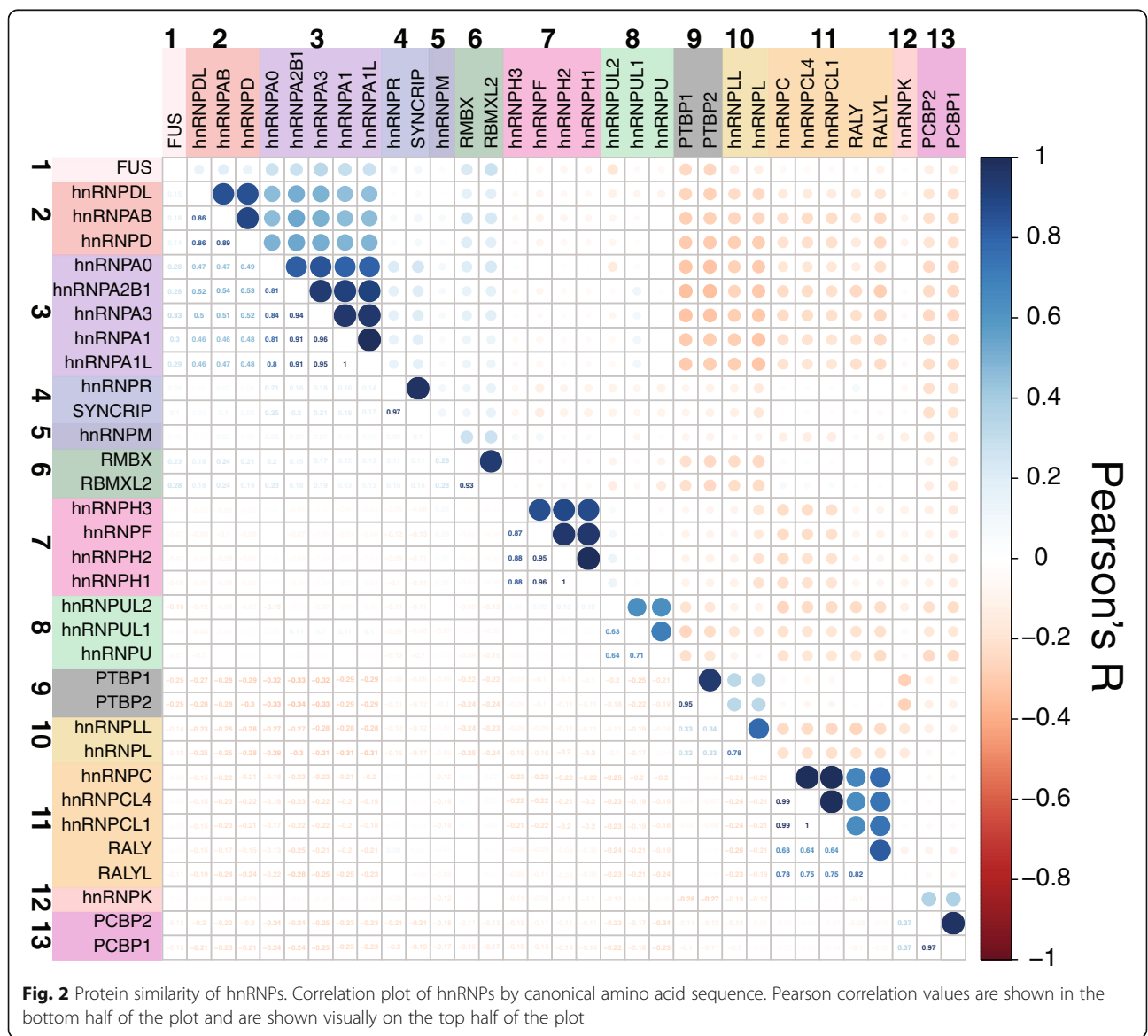
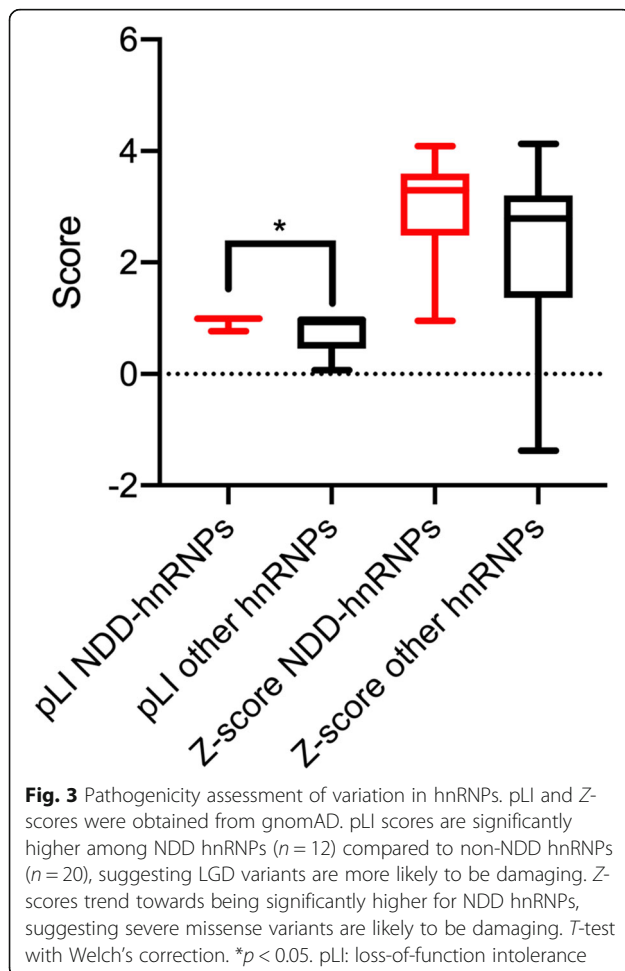


Table 3 Genomic disorders spanning NDD *HNRNPs*

Genomic disorder	<i>HNRNP</i>	Gene previously considered candidate for CNVs?	Shared phenotypes
5q35 deletions	<i>HNRN PH1</i>	No	DD/ID, characteristic facial features, overgrowth, microcephaly
1q43q44 deletions	<i>HNRN PU</i>	Yes	DD/ID, seizures, structural brain abnormalities, speech delay
9q21.32 deletions	<i>HNRNPK</i>	Yes	DD/ID, motor delay, speech delay, structural brain abnormalities, hypotonia, skeletal abnormalities, hand/feet abnormalities, cardiac abnormalities, genitourinary issues, dysmorphic features
6q proximal deletions	<i>SYNCRIP</i>	Yes	DD/ID, ASD, structural brain abnormalities, behavioral issues
4q21 microdeletion syndrome	<i>HNRN PD</i>	No	DD/ID, emotional/behavioral issues, speech delay
1p36 monosomy	<i>HNRN PR</i>	No	DD/ID, skeletal abnormalities, genitourinary issues, seizures, structural brain abnormalities

CNV copy number variant, DD/ID developmental delay/intellectual disability, ASD autism spectrum disorder



Families where recontact was possible were invited to participate in a remote comprehensive clinical workup that included diagnostic evaluation, medical history, and neuropsychological assessment. All experiments carried out on these individuals were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national), and proper informed consent was obtained for sequencing, recontact for inheritance testing, and phenotypic workup. When physicians or families could not be contacted, deidentified clinical information was used. Proband were excluded if they reportedly had (1) a causative variant in another NDD-related gene, (2) a known inherited variant in an *HNRNP* (other than mosaic), (3) a variant observed in gnomAD, (4) a missense variant with a CADD (v1.6, GRCh38) score < 20 , (5) consanguineous parents (which was based on clinical history), and/or were (6) born prematurely (prior to 37 weeks), or (7) too young to assess for DD (< 1 year). However, individuals excluded for our *HNRNPs* cohort were still considered in the total number of individuals for statistical analyses. For probands from large sequencing studies, these criteria were not always available; therefore, a small number of those not meeting our criteria may be included due to limited information.

Cohorts

ASID Network

For our Autism Spectrum/Intellectual Disability (ASID) network ($n = 16,294$ individuals), six *HNRNPs* were previously assessed using targeted sequencing with single-molecule molecular inversion probes: *HNRNPF*, *HNRNPK*, *HNRNPR*, *SYNCRIP*, *HNRNPU*, and *HNRN PULL1*, as described by Stessman et al., identifying 14 probands (seven with DNVs) included in the current study [34, 68]. All variants were validated using Sanger

sequencing, although not all had parental DNA available for inheritance testing. Information for a subset of these probands are available in denovo-db v1.6.

BGL

From Baylor Genetics Laboratory (BGL; $n = 9536$, including 1744 trios), 34 novel probands are included in the current study (four with DNVs) and four previously published probands with DNVs (for a total of 38 probands) [23]. These probands are all from clinical exome sequencing samples broken down into “neurologic” phenotypes ($n = 2364$, including 429 trios; developmental delay [DD]/intellectual disability [ID], speech delay, and autism spectrum disorder [ASD]) and “neurologic plus other organ systems” ($n = 7172$, including 1315 trios). Exome sequencing was performed as described previously, and DNVs were validated by Sanger sequencing [69]. All *HNRNPs* were queried in the BGL data except *HNRNPH2*. Phenotypic information was obtained from BGL clinical indications for sequencing and/or from reporting physicians when available.

SPARK Consortium

From the SPARK Consortium, with individuals ascertained for ASD diagnoses, nine probands with DNVs are included in the current study, as well as an additional published proband (for a total of 10 probands) [42]. Probands identified by the SPARK Consortium were sequenced as previously described [42]. SPARK Consortium probands were used in all statistical analyses, although complete phenotypic information was not always available.

Deciphering Developmental Disorders (DDD13K)

Data for 9860 trios from the DDD cohort were obtained and reanalyzed [62]. We applied FreeBayes (version v1.1.0–3-g961e5f3) and GATK (version 3.7) for SNV/indel calling independently in 9308 DDD families where both parents were available (with 28,476 samples including 9860 probands) [70, 71]. BCFtools (version 1.3.1) was used for left-align and normalization post calling [72]. For quality control, we filtered for read depth ($DP > 9$) for all family members and filtered allele balance ($AB > 0.25$) and PHRED genotype quality ($GQ > 20$) filters in probands. Candidate de novo SNVs/indels were only observed in probands but not in parents and were required to be identified by both FreeBayes and GATK. Variants were then annotated using CADD (v1.6, GRCh38).

International collaborations

GeneMatcher and GeneDx were utilized to identify 24 individuals, 21 of whom have DNVs [73]. GeneMatcher requests were made for each *HNRNP* identified in an NDD cluster (Fig. 2). Probands were sequenced at

varying locations, including several by GeneDx, whose protocol was described previously. Deidentified clinical information was obtained from physicians or families were invited for remote clinical workup. These probands were not used in de novo statistical calculations but were considered among phenotypic analyses.

Four probands with DNVs were obtained from the Radboud University Medical Center (RUMC) in the Netherlands. These probands were not used in de novo statistical calculations but were considered among phenotypic analyses.

Databases

Large databases were also used to identify novel probands. Each *hnRNP*-encoding gene was queried in MyGene2 (<https://mygene2.org/MyGene2/>), a database generated by the University of Washington Center for Mendelian Genomics, identifying two probands (one proband was identified via both MyGene2 and GeneMatcher) [74]. MyGene2 probands were not used for de novo statistical calculations but were considered for phenotypic analyses.

DECIPHER ($n > 33,000$ individuals; <https://decipher.sanger.ac.uk/>) and ClinVar ($n = 874,088$; <https://www.ncbi.nlm.nih.gov/ncbi-nlm-nih.gov.offcampus.lib.washington.edu/clinvar/>) were queried and only those with phenotypic information available were included in this study, adding 12 (10 with DNVs) and 19 (seven with DNVs) new probands, respectively. Probands were considered novel if they had not been previously published. DECIPHER and ClinVar probands were included in phenotypic analyses but not de novo statistics.

Functional impacts of novel variation in *HNRNP* genes Pathogenicity predictions, de novo enrichment, and missense analyses

To further characterize the types of variation among NDD *HNRNPs*, we compared the probability of pathogenicity of LGD and missense variants using gnomAD (loss-of-function intolerance [pLI] and missense Z-scores; <https://gnomad.broadinstitute.org/>) and Gene Variation Intolerance Ranking (GeVIR; <http://www.gevirank.org/>) scores and percentiles, which include loss-of-function observed/expected upper bound fraction (LOEUF) metrics, of NDD *HNRNPs* to non-NDD *HNRNPs* using a one-way *t*-test in GraphPad Prism (v.8.3.0) (Fig. 3 and Additional File 2: Fig. S3, Additional File 1: Table S2) [75].

For de novo enrichment analyses, 33,688 probands with trio sequencing ($n = 36,460$ for *HNRNPH2* and $n = 36,814$ for *HNRNPL1*) were used (Additional File 1: Table S6). Only studies with a clear sample size were used to estimate de novo burden (i.e., $n > 1$, cohort total sizes included in publication). Only genes determined as NDD

HNRNPs were analyzed. In total, 78 *HNRNP* variants were analyzed (Additional File 1: Table S7). Two models, the chimpanzee–human divergence (CH) model and denovolyzeR, were used to assess the excess of de novo LGD and missense variants as previously described (Fig. 4a) [1, 34, 76, 77]. Briefly, the CH model utilizes locus-specific transition/transversion/indel rates and chimpanzee–human coding sequence divergences to estimate the expected number of de novo mutations, while the denovolyzeR utilizes trinucleotide context, exome depth, and divergence between macaque and human over a ± 1 megabase-pair window and includes known mutational biases like CpG hotspots. Each model was corrected for the number of genes included in the model ($n = 18,946$ for the CH model and $n = 19,618$ for denovolyzeR) as well as three mutation-type tests per gene and two tests per mutation type (FWER for CH model: $q < 4.4 \times 10^{-7}$, FWER for denovolyzeR $q < 4.24 \times 10^{-7}$).

Missense variants were analyzed using two methods: CLUMP (CLUstering by Mutation Position) analysis and MetaDome assessment [78–80]. Only genes determined

NDD *HNRNPs* were analyzed. CLUMP assesses the clustering of missense mutations and was performed as previously described, but here using the gnomAD non-neuropsychiatric group as controls (gnomAD $n = 114,704$; 1958 missense variants; current cohort missense $n = 80$ variants) (Fig. 4b, Additional File 1: Table S6) [80]. All calculations were performed using R (v.4.0.0) or python (v.3.7.7). Missense variants were also assessed for clustering by phenotype. Mutations were queried in MetaDome to assess for protein domain disruptions (Fig. 5) [78]. LGD variants were also assessed for susceptibility to nonsense mediated decay (NMD) using NMDescPredictor (<https://nmdprediction.shinyapps.io/nmdescpredictor/>) and splicing impact using MaxEntScan (Fig. 5, Additional File 1: Table S7) [81, 82].

GTEx and single-cell expression analyses

Transcript-level expression from adult tissues obtained from GTEx (<https://gtexportal.org/>) for each gene was compared using *t*-tests between NDD and non-NDD *HNRNPs* as well as by two-way ANOVA for comparing

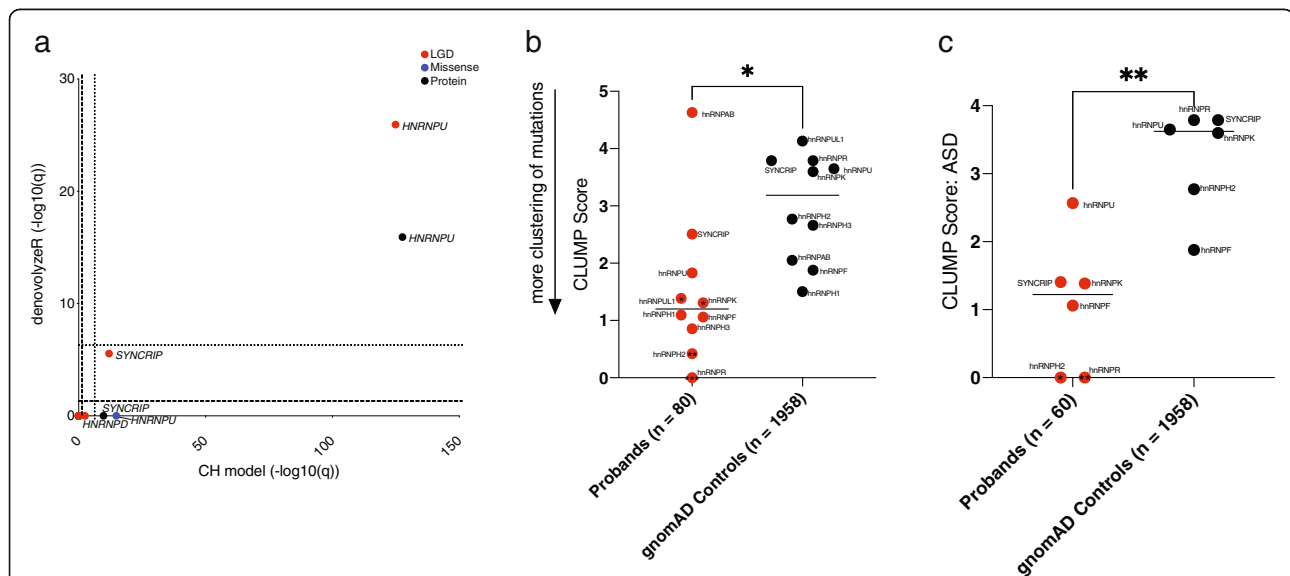
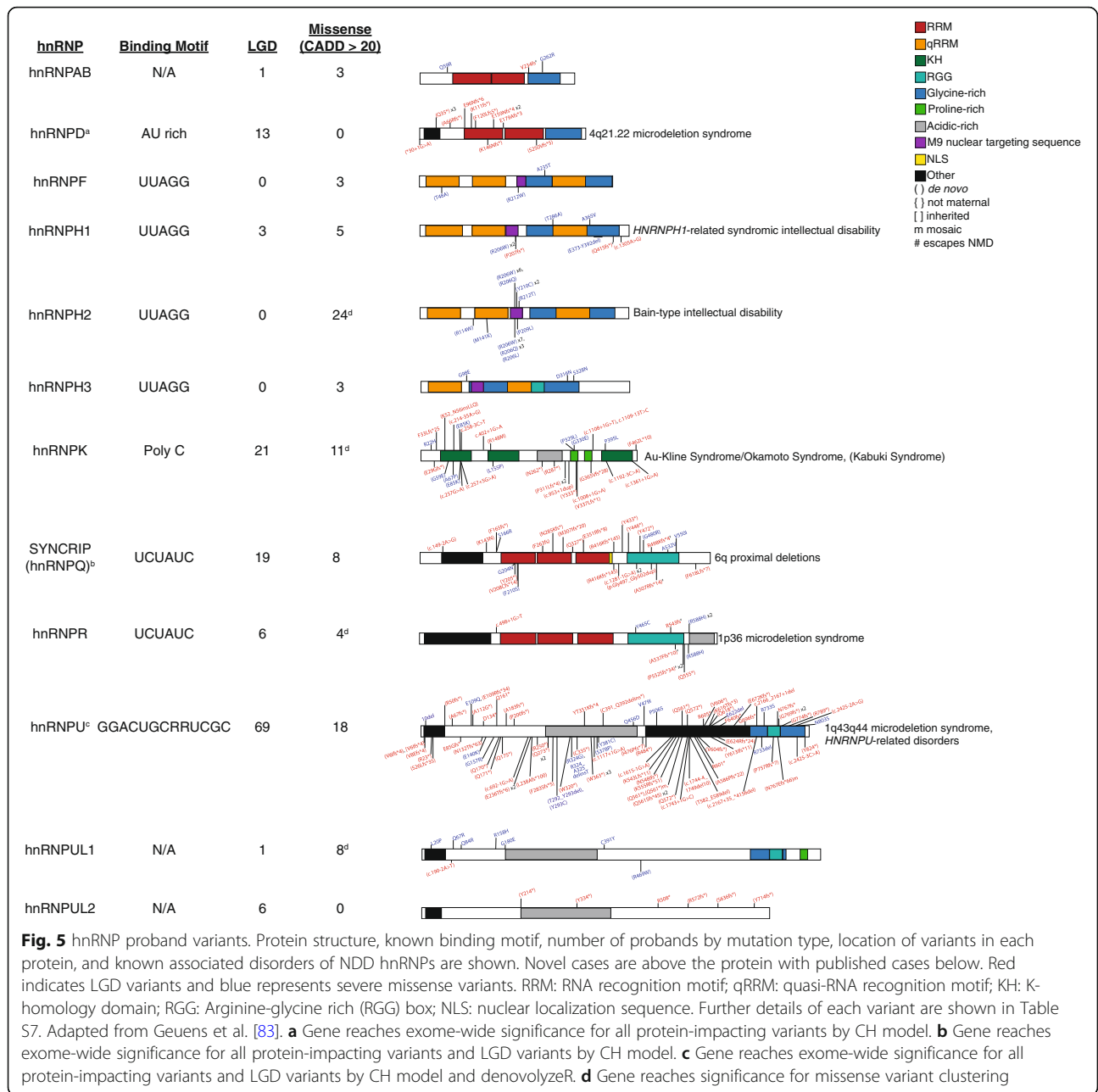


Fig. 4 De novo enrichment and clustering of missense variation analyses of NDD *hnRNPs*. **a** De novo variation was assessed for NDD *HNRNPs* using two statistical models: the CH model and denovolyzeR. Right/above the dotted line indicates the gene achieves exome-wide significance ($q < 4.24 \times 10^{-7}$) while right/above the dashed line indicates the gene reaches nominal significance ($q < 0.05$). *HNRNPU* reaches exome-wide significance for all protein-impacting variants (Protein) and LGD variants, with severe missense variants reaching significance by only the CH model. *SYNCRIP* reaches exome-wide significance for LGD variants and all protein-impacting variants by the CH model alone. *HNRNPD* reaches nominal significance by the CH model. *P* values are FDR corrected with the number of genes ($n = 18,946$ for CH model and $n = 19,618$ for denovolyzeR) with three tests per gene (LGD, missense, and all protein changes) and two tests (CH model and denovolyzeR) per mutation type. Only cohorts with known de novo status were included, as listed in Tables S1 and S7. Statistics can be seen in Table S6. **b** Analysis of clustered missense variants. Clustering of missense variants was analyzed using CLUMP; scores are shown in Table S6 (paired *t*-test). Compared to the non-neuropsychiatric subset of gnomAD ($n = 114,704$, 1958 missense variants), the CLUMP score for NDD *hnRNPs* (red) among probands is significantly lower than controls in gnomAD (black), indicating more clustering of mutations (shown by arrow). Note that only genes with variants in the current cohort could undergo this analysis. *hnRNPH2*, *hnRNPK*, *hnRNPR*, and *hnRNPUL1* each have independent significant clumping compared to gnomAD controls. **c** CLUMP scores for missense variants in probands with ASD ($n = 60$). *HnRNPH2*, *hnRNPK*, *hnRNPR*, and *hnRNPUL1* reach significance independently. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. LGD: likely gene disrupting; Missense: severe missense (CADD ≥ 20); Protein: all protein-affecting variants



between individual genes (Fig. 6a, Additional File 2: Figs. S3C,D, Additional File 1: Table S3). We used previously described single-cell RNA (scRNA) sequencing data generated from 48 individuals to assess gene expression during fetal human cortical development, which is available from the UCSC cell browser: <https://cells.ucsc.edu/> [84]. To determine cell enrichment, data were tested using a Wilcoxon ranked sum test followed by Bonferroni correction in R, as previously described [84]. Student *t* tests and two-way ANOVA among all *HNRNPs*, between NDD and non-NDD *HNRNPs*, or *HNRNPs* and genes

associated with similarly presenting disorders determined the difference in fold-enrichment using GraphPad Prism (Figs. 6b,c and 8b,c, Additional File 2: Figs. S3A,B, Additional File 1: Tables S4–5). Correlations among scRNA sequencing data were performed in R (v.3.6.1) using the corplot package (v.0.84). One-way *t*-tests were performed between each NDD *HNRNP* group (as in Fig. 2) using GraphPad Prism. *HNRNPs* were also entered into Specific Expression Analyses (SEA) to determine enrichment of expression across brain regions (Fig. 6d) [85]. Expression among these tissues was

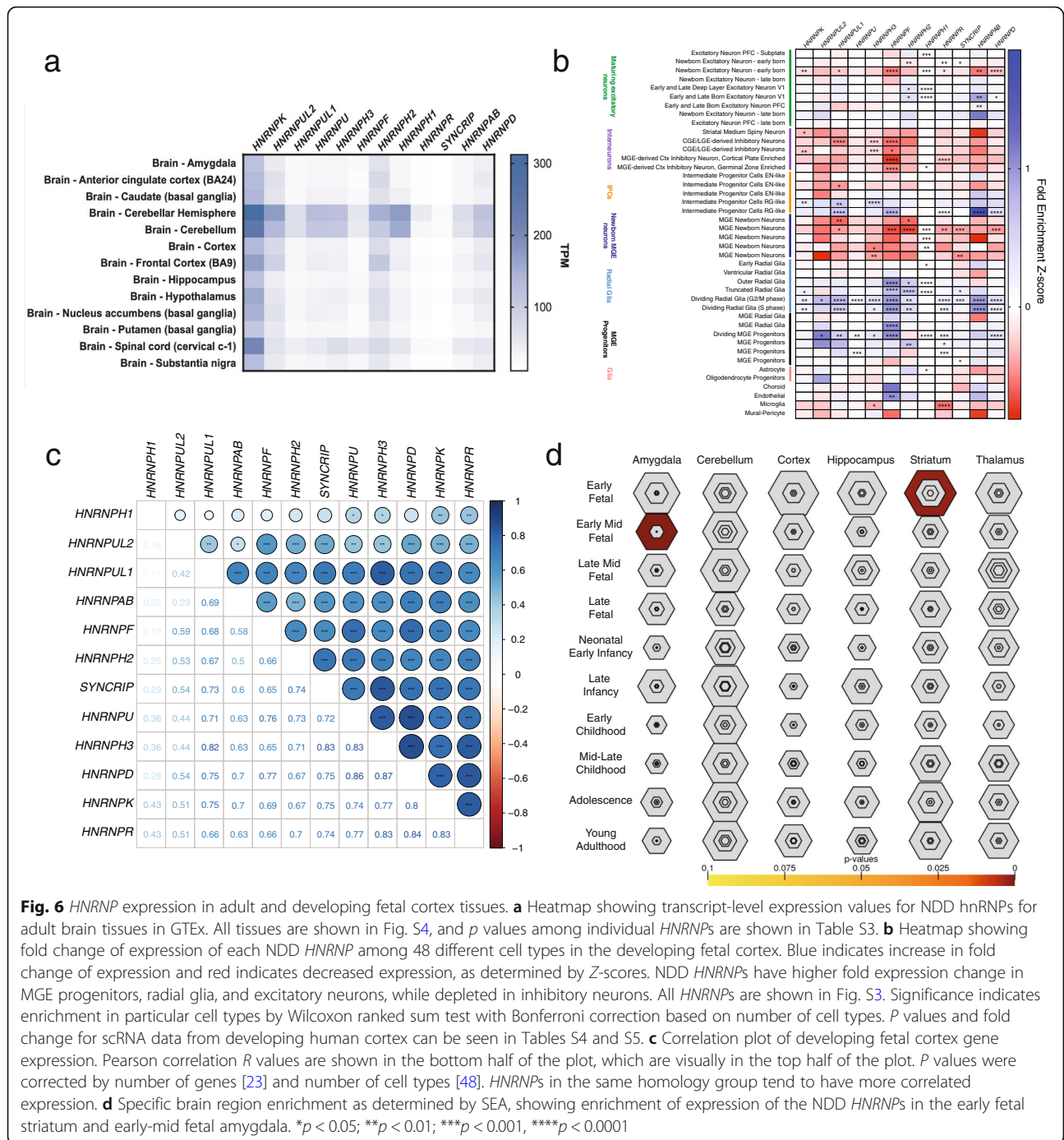


Fig. 6 *HNRNP* expression in adult and developing fetal cortex tissues. **a** Heatmap showing transcript-level expression values for NDD *hnrnps* for adult brain tissues in GTEx. All tissues are shown in Fig. S4, and *p* values among individual *HNRNPs* are shown in Table S3. **b** Heatmap showing fold change of expression of each NDD *HNRNP* among 48 different cell types in the developing fetal cortex. Blue indicates increase in fold change of expression and red indicates decreased expression, as determined by Z-scores. NDD *HNRNPs* have higher fold expression change in MGE progenitors, radial glia, and excitatory neurons, while depleted in inhibitory neurons. All *HNRNPs* are shown in Fig. S3. Significance indicates enrichment in particular cell types by Wilcoxon ranked sum test with Bonferroni correction based on number of cell types. *P* values and fold change for scRNA data from developing human cortex can be seen in Tables S4 and S5. **c** Correlation plot of developing fetal cortex gene expression. Pearson correlation *R* values are shown in the bottom half of the plot, which are visually in the top half of the plot. *P* values were corrected by number of genes [23] and number of cell types [48]. *HNRNPs* in the same homology group tend to have more correlated expression. **d** Specific brain region enrichment as determined by SEA, showing enrichment of expression of the NDD *HNRNPs* in the early fetal striatum and early-mid fetal amygdala. **p* < 0.05; ***p* < 0.01; ****p* < 0.001, *****p* < 0.0001

compared using Fisher’s exact tests and followed by Benjamini-Hochberg correction.

Phenotypic analyses

Phenotypes were defined by Human Phenotype Ontology (HPO) terms when possible. Those that occurred in ≥ 20% of probands in at least one *HNRNP* cohort were considered characteristic for the disorder and used for statistical analyses. Traits occurring less than that were

considered not likely to be syndromic. This resulted in a total of 88 phenotypes. For quantitative measures (e.g., head circumference), qualitative terms were used if provided by a clinician (e.g., microcephalic) or determined based on top/bottom 3rd percentile.

Phenotypic information was compared across *HNRNP* cohorts using RVAideMemoire (version 0.9–73) pairwise Fisher’s exact tests with Bonferroni correction considering number of genes (*n* = 12 for all

NDD *HNRNPs*, $n = 9$ for LGD NDD *HNRNPs*, $n = 10$ for missense NDD *HNRNPs*, number of phenotypes ($n = 88$), and number of mutation categories ($n = 3$: all mutations, LGD, and missense) tested, as well as across mutation types (LGD vs. missense) within *HNRNP* cohorts using Fisher's exact tests (Fig. 7,

Additional File 1: Tables S8–9). All calculations were performed using GraphPad Prism or R. Phenotypes were compared to other known disorders by characteristic HPO terms using PhenPath (<http://phenpath.biocomp.unibo.it/phenpath/>), which returns disorders and genes with shared HPO terms [86].

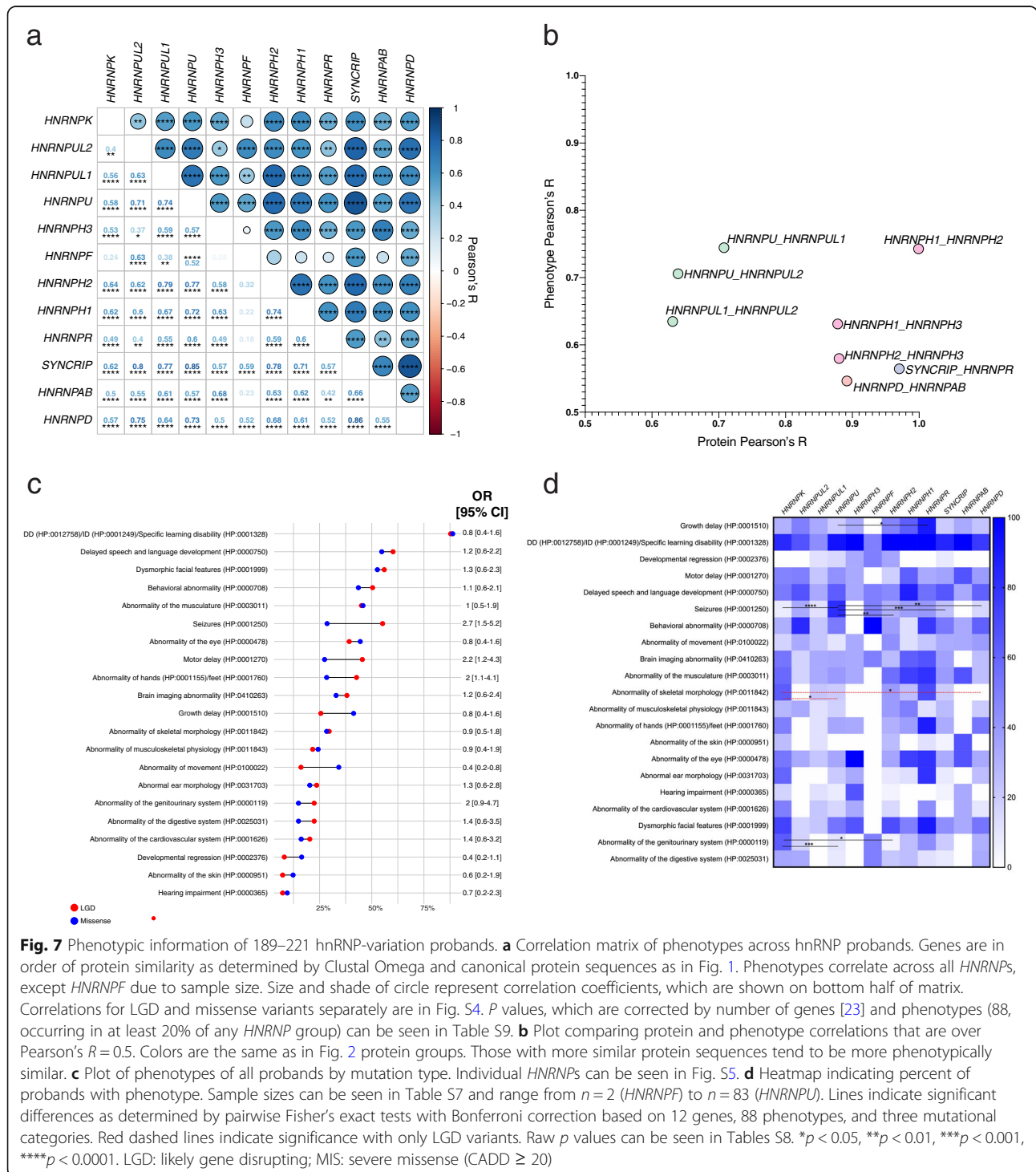


Fig. 7 Phenotypic information of 189–221 hnRNP-variation probands. **a** Correlation matrix of phenotypes across hnRNP probands. Genes are in order of protein similarity as determined by Clustal Omega and canonical protein sequences as in Fig. 1. Phenotypes correlate across all *HNRNPs*, except *HNRNPF* due to sample size. Size and shade of circle represent correlation coefficients, which are shown on bottom half of matrix. Correlations for LGD and missense variants separately are in Fig. S4. *P* values, which are corrected by number of genes [23] and phenotypes (88, occurring in at least 20% of any *HNRNP* group) can be seen in Table S9. **b** Plot comparing protein and phenotype correlations that are over Pearson's $R = 0.5$. Colors are the same as in Fig. 2 protein groups. Those with more similar protein sequences tend to be more phenotypically similar. **c** Plot of phenotypes of all probands by mutation type. Individual *HNRNPs* can be seen in Fig. S5. **d** Heatmap indicating percent of probands with phenotype. Sample sizes can be seen in Table S7 and range from $n = 2$ (*HNRNPF*) to $n = 83$ (*HNRNPU*). Lines indicate significant differences as determined by pairwise Fisher's exact tests with Bonferroni correction based on 12 genes, 88 phenotypes, and three mutational categories. Red dashed lines indicate significance with only LGD variants. Raw *p* values can be seen in Tables S8. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. LGD: likely gene disrupting; MIS: severe missense (CADD ≥ 20)

Results

Twelve *HNRNPs* are NDD candidates based on literature search, protein homology, CNV intersection, and identification of new probands

In order to identify NDD candidate *HNRNPs*, we initially considered cases with DNVs in the literature. This resulted in 16 genes with at least one published case as potential NDD candidate genes: *HNRNPA0*, *HNRNPA3*, *HNRNPAB*, *HNRNPA2B1*, *HNRNPC*, *HNRNPD*, *HNRNPF*, *HNRNPH1*, *HNRNPH2*, *PTBP2*, *HNRNPK*, *HNRNPLL*, *SYNCRIP*, *HNRNPR*, *HNRNPU*, and *HNRNPUL1* ($n = 117$ probands; Fig. 1, Additional File 1: Tables S1, S7, S8, and S11).

To identify other genes in the family that may not have published cases, we considered the extent of shared protein homology among hnRNPs (Fig. 2 and Additional File 2: Fig. S1). Among 23 core hnRNPs and 10 minor hnRNPs, we distinguished 13 protein groups based on protein homology clustering: (1) FUS (Fused in sarcoma, also known as hnRNPP), (2) hnRNPD-related proteins, (3) hnRNPA-related proteins, (4) hnRNPR/SYNCRIP, (5) hnRNPM, (6) RMBX (RNA-binding motif protein X-linked, also known as hnRNPG)-related proteins, (7) hnRNPH/F-related proteins, (8) hnrnpu-related proteins, (9) PTBP (also known as hnRNPI)-related proteins, (10) hnRNPL-related proteins, (11) hnRNPC-related proteins, (12) hnRNPK, and (13) PCBP (also known as hnRNPE)-related proteins. From this, we identified seven additional hnRNPs that, while they do not have published cases with NDDs, cluster with our other candidate hnRNPs: hnRNPC1, hnRNPC4, RALY, RALYL, hnRNPH3, hnRNPD1, and hnRNPU2.

We determined if any *HNRNPs* are dosage sensitive by intersecting the *HNRNPs* with a list of 58 genomic disorders based on previous CNV morbidity maps and the DECIPHER database (Table 3) [10, 13, 19, 20, 64–67]. Of the 33 *HNRNPs*, six are in regions of known genomic CNV disorders: *HNRNPD*, *HNRNPH1*, *HNRNPK*, *HNRNPR*, *SYNCRIP*, and *HNRNPU*, further supporting their likely role in NDDs. This analysis did not identify any new candidate NDD *HNRNPs* but did provide additional evidence for their roles in NDD pathogenicity.

Genes were also considered as candidate NDD *HNRNPs* if we identified variants in our novel cohorts. This resulted in 17 genes with at least one novel case as potential NDD candidate genes: *HNRNPA0*, *HNRNPA1*, *HNRNPA2B1*, *HNRNPAB*, *HNRNPD*, *HNRNPF*, *HNRNPH1*, *HNRNPH2*, *HNRNPH3*, *HNRNPK*, *HNRNPLL*, *PTBP2*, *SYNCRIP*, *HNRNPR*, *HNRNPU*, *HNRNPUL1*, and *HNRNPUL2* ($n = 118$ probands; Fig. 1, Additional File 1: Tables S1, S7, S8, and S11).

Finally, NDD *HNRNPs* were finalizing by identification of at least three probands in our combined literature and novel cohorts. This resulted in the following being considered NDD *HNRNPs*: *HNRNPAB*, *HNRNPD*, *HNRNPF*,

HNRNPH1, *HNRNPH2*, *HNRNPH3*, *HNRNPK*, *HNRNPR*, *SYNCRIP*, *HNRNPU*, *HNRNPUL1*, and *HNRNPUL2*.

NDD *HNRNPs* are more sensitive to mutation than non-NDD *HNRNPs*

We considered each gene's tolerance to mutation by assessing gnomAD metrics pLI and missense Z -scores as well as GeVIR scores (Fig. 3 and Additional File 2: Fig. S2, Additional File 1: Table S2) [75, 87]. As a group, pLI scores are significantly higher among NDD *HNRNPs* compared to non-NDD *HNRNPs* ($p = 0.03$, one-way Student's t test). Similarly, NDD *HNRNPs*' missense Z -scores trend higher than non-NDD *HNRNPs*; GeVIR scores show the same pattern. Specifically, NDD *HNRNPs* have significantly higher LOEUF autosomal dominant scores ($p = 0.03$, one-way Student's t test) and Variation Intolerant Region Loss-of-Function (VIRLoF) scores, which takes both missense and LGD variants into account ($p = 0.04$, one-way Student's t test). GeVIR autosomal recessive scores show a similar pattern for LOEUF scores ($p = 0.04$, one-way Student's t test), but as all of our variants are heterozygous, this information was not utilized in our study. This suggests that LGD and severe missense variants are more likely to be damaging among NDD *HNRNPs* compared to the non-NDD *HNRNP* family members.

Enrichment of de novo *HNRNP* variants

In total, 225 probands with SNVs or indels in NDD *HNRNPs* were identified from 83,975 individuals. Inheritance was determined for 73.8% of NDD *HNRNP* SNV and indel variants (Table 4, Fig. 3, and Additional File 1: Table S7). Of variants with known inheritance, DNVs account for 98.8% of variants, with one mosaic variant inherited from the unaffected father, who also carries the variant mosaically (Proband 124, *HNRNPU* proband 1) [26]. LGD variants represent 61.8% ($n = 139/225$ variants, inheritance known for 82.7%; 99.1% de novo) of the NDD *HNRNP* cohort, while severe missense variants (MIS20: CADD score ≥ 20 , being in the top 1% of predicted pathogenic mutations and MIS30: CADD ≥ 30 top 0.1% of predicted pathogenic variants) account for 38.2% of variants ($n = 86/225$ variants, inheritance known for 59.3%; 98% de novo; MIS20, MIS30, and those without CADD scores available were included). Twelve of the 16 small CNVs have known inheritance status, and all are de novo (Additional File 1: Table S12).

In order to identify genes with a significant excess of DNVs, we examined 41,779 parent–child trios from 13,437 families with DD/ID, 20,542 families with ASD, 1421 families with epilepsy, and 6379 families from clinical exome testing. We applied two statistical models, denovolzeR and the CH model, to genes that had DNVs from cohorts with known sample size ($n = 78$

Table 4 Inheritance of 225 SNVs and indels in NDD *HNRNPs*

Variant type	De novo % of variant type (n)	Inherited % of variant type (n)	Unknown % of variant type (n)	Total % of all variants (n)
LGD	82 (114/139)	0.7 (1/139)	17.3 (24/139)	61.8 (139/225)
All MIS	58.1 (50/86)	1.2 (1/86)	40.7 (35/86)	38.2 (86/225)
MIS—no CADD score	50 (4/8)	0 (0/8)	50 (4/8)	3.6 (8/225)
MIS20	57.5 (42/73)	1.4 (1/73)	41.1 (30/73)	32.4 (73/225)
MIS30	80 (4/5)	0 (0/5)	28.6 (2/5)	2.7 (6/225)
All variants	73.2 (164/225)	0.9 (2/225)	25.3 (56/225)	100 (225/225)

LGD likely gene disruptive, MIS missense, MIS20 CADD score ≥ 20 , MIS30 CADD score ≥ 30 . NDD *HNRNPs* include *HNRNPAB*, *HNRNPD*, *HNRNPF*, *HNRNPH1*, *HNRNPH2*, *HNRNPH3*, *HNRNPK*, *SYNCRIP*, *HNRNPR*, *HNRNPU*, *HNRNPUL1*, and *HNRNPUL2*. Variants identified in non-NDD *HNRNPs* are in Table S7

variants, Fig. 4a, Additional File 1: Table S6) [76, 77]. We find that only *HNRNPU* achieves exome-wide significance ($p < 4.24 \times 10^{-7}$) for de novo LGD and all protein-impacting variants by both models. *SYNCRIP*, for all protein-impacting variants and LGD variants, reaches exome-wide significance only by the CH model, while all protein-impacting variants of *HNRNPD* reach nominal significance ($p < 0.05$) by the CH model. Only *HNRNPU* reaches exome-wide significance for an excess of severe de novo missense variation after multiple testing correction (FWER, $p < 4.24 \times 10^{-7}$), and only by the CH model. Based on these DNV models, only *HNRNPU*, *SYNCRIP*, and *HNRNPD* reach significance for increased number of variants than one would expect.

Missense mutation analyses

As pathogenic missense variants are known to cluster for some proteins, such as hnRNPH2 in the M9 nuclear targeting signal, we assessed whether there was any additional evidence of missense variant clustering across hnRNPs [12]. We examined all MIS20 and MIS30 variants in our cohort ($n = 80$) and compared them to the non-neuropsychiatric subset in gnomAD using CLUMP ($n = 1958$) (Fig. 4b, Additional File 1: Table S6) [80]. In order to control for platform differences, variants were only included if exon coverage was at least 20-fold in gnomAD (the first two exons of *HNRNPUL1* were excluded). The CLUMP scores for NDD hnRNP missense variants are significantly lower than those of controls from gnomAD ($p = 0.03$, paired t -test) suggesting clustering of variants. hnRNPH2, hnRNPK, hnRNPR, and hnRNPU independently have significantly more clustering of missense variants than controls ($p < 0.05$ for hnRNPK and hnRNPU, $p < 0.01$ for hnRNPH2 and hnRNPR). Thus, for multiple hnRNPs, clustering of missense variants likely contributes to pathogenicity, potentially impacting particular protein domains. Notably, this clustering was more significant among particular phenotypes, including ASD ($p = 0.005$, *HNRNPR* alone $p = 0.007$). In hnRNPH2, missense variants cluster significantly among probands with growth delay ($p = 0.01$),

motor delay ($p = 0.001$), speech delay ($p = 0.006$), microcephaly ($p = 0.01$), hypotonia ($p = 0.003$), seizures ($p = 0.01$), and cardiac abnormalities ($p = 0.02$) compared to controls. Cardiac abnormalities also have a clustering of missense variants among *HNRNPK* probands ($p = 0.004$). Variants cluster among probands with hypospadias, joint hypermobility, and scoliosis among *HNRNPH1* probands ($p = 0.04$ for each). In addition to ASD, variants in probands with speech delay or microcephaly and *HNRNPR* variants are clustered ($p = 0.005$ and $p = 0.001$, respectively). We also applied MetaDome to identify such domains and as another predictor of pathogenicity. There was no significant difference between variants in known domains and unknown domains. Other than the M9 nuclear localization signal, there was clustering of missense variants in the RNA-binding domain of hnRNPH2. In hnRNPK, 70% of missense variants ($n = 7/10$) cluster in the KH domains, which are important in RNA binding and recognition. Almost half (42.9%, $n = 3/7$) of missense variants in *SYNCRIP* cluster in its RNA recognition motif while de novo missense mutations in hnRNPU cluster in the SPRY domain and AAA domain ($n = 3/16$ each). Twenty-five percent (2/8) of missense variants in hnRNPU occur in the SAP domain, important for RNA and DNA binding. Thus, RNA recognition and binding are likely responsible for many of the missense variants observed in the hnRNPs.

HNRNP brain expression analyses

In order to understand the pathogenesis of variation in *HNRNPs*, we explored the expression of the genes in both adult tissues and the developing human cerebral cortex [84]. GTEx analysis of adult tissues shows that the *HNRNPs* as a whole ($n = 33$) are expressed across all brain regions, as well as ubiquitously across tissues, with the exception of *HNRNPCL1*, *RALYL* (*HNRNPCL3*), *HNRNPCL4*, and *RBMXL2*, which broadly show low levels of expression (Fig. 6a and Additional File 2: Figs. S3C,D). Significant changes in expression are seen based on individual genes and by tissue type ($p < 0.0001$ for both by two-way ANOVA), with highest expression

levels in the brain seen typically in the cerebellum. Between NDD and non-NDD *HNRNPs*, no significant differences are observed (Additional File 2: Fig. S3). The general high expression of these genes may contribute to phenotypes observed.

scRNA sequencing from the developing cortex confirms widespread expression across brain cells, especially neuronal and neuronal progenitor cell types, suggesting that these genes could play a role in neuronal differentiation (Fig. 5b and Additional File 2: Figs. S4C,D) [84]. We find that *HNRNP* expression is enriched among radial glia, which act as neural stem cells, but decreased among immature inhibitory and excitatory neurons, although it does vary by *HNRNP* ($p < 0.05$ to $p < 0.0001$, Wilcoxon ranked sum test, Bonferroni correction; Fig. 6a and Additional File 2: S3C, Additional File 1: Table S4). While significant differences are seen by two-way ANOVA between individual *HNRNPs*, particularly for *PCBP1* (*HNRNPI*) and *HNRNPCL3/RALYL*, no significant differences are observed between all NDD and non-NDD *HNRNPs* (Additional File 2: Figs. S3A,B, Additional File 1: Table S5). It is plausible that there are differences between NDD and non-NDD *HNRNP* expression that contribute to the development of NDD phenotypes when perturbed.

Specifically, NDD *HNRNPs* are consistently significantly enriched among dividing radial glia in G2/M phase (all but *HNRNPH1*; $p < 0.05$ to $p < 0.0001$, Wilcoxon ranked sum test, Bonferroni correction), dividing radial glial cells during S phase (*HNRNPK*, *HNRNPUL1*, *HNRNPH3*, *HNRNPF*, *HNRNPH2*, *HNRNPR*, *HNRNPAB*, and *HNRNPD*; $p < 0.05$ to $p < 0.0001$, Wilcoxon ranked sum test, Bonferroni correction), and in the medial ganglionic eminences (MGE), which give rise to major populations of inhibitory neurons of the cortex (*HNRNPUL2*, *HNRNPUL1*, *HNRNPU*, *HNRNPH3*, *HNRNPF*, *HNRNPH1*, *HNRNPR*, and *HNRNPD*, $p < 0.05$ to $p < 0.0001$, Wilcoxon ranked sum test, Bonferroni correction), while depleted in newborn neurons in the MGE (*HNRNPUL1*, *HNRNPH3*, *HNRNPF*, *HNRNPH2*, *HNRNPH1*, *HNRNPR*, *SYNCRIP*, and *HNRNPD*; $p < 0.05$ to $p < 0.0001$, Wilcoxon ranked sum test, Bonferroni correction) and newborn excitatory neurons (*HNRNPK*, *HNRNPUL1*, *HNRNPF*, *HNRNPH1*, *HNRNPR*, *HNRNPAB*, and *HNRNPD*; $p < 0.05$ to $p < 0.0001$, Wilcoxon ranked sum test, Bonferroni correction). Among individual NDD *HNRNPs*, no significant differences are observed by two-way ANOVA for the contribution of each gene, highlighting their shared expression.

Hierarchical clustering shows that gene expression in the developing cortex is highly correlated among specific *HNRNPs*. Strong positive correlations are seen between *HNRNPD* and *HNRNPH3* (Pearson's $R = 0.87$, $p = 1.9e-15$), *HNRNPU* (Pearson's $R = 0.86$, $p = 3e-15$), *HNRN*

PR (Pearson's $R = 0.83$, $p = 1.4e-13$), and *HNRNPK* expression patterns (Pearson's $R = 0.8$, $p = 6.1e-12$), *HNRNPR* and *HNRNPK* expression patterns (Pearson's $R = 0.83$, $p = 2.7e-08$), *HNRNPUL1* and *HNRNPH3* expression patterns (Pearson's $R = 0.82$, $p = 9.5e-13$), *HNRNPU* and *HNRNPH3* expression patterns (Pearson's $R = 0.83$, $p = 4e-13$), and *HNRNPH3* and *HNRNPR* (Pearson's $R = 0.83$, $p = 4.1e-13$) and *SYNCRIP* expression patterns (Pearson's $R = 0.83$, $p = 1.8e-13$). While there are positive correlations between almost all NDD *HNRNP* gene expression patterns, the differences in expression along with protein homology likely play a role in the phenotypes observed.

Finally, we assessed the NDD *HNRNP* expression among specific brain regions during development (Fig. 5d). Expression is enriched in the striatum in early fetal development ($p = 0.002$, Fisher's exact test, Benjamini-Hochberg correction) and the amygdala ($p = 2.6e-4$, Fisher's exact test, Benjamini-Hochberg correction) during early-mid fetal development. *HNRNPK*, *HNRNPR*, *SYNCRIP*, *HNRNPF*, and *HNRNPU* are expressed in the striatum while *HNRNPH1*, *HNRNPAB*, *HNRNPUL1*, *HNRNPH3*, and *HNRNPU* are expressed in the amygdala. Expression among the striatum and amygdala is likely impacted by variation in these genes, contributing to probands' phenotypes.

Phenotypic assessment of probands with *HNRNP* variation

The pathogenesis of variation in the *HNRNPs*, and most genes in NDDs, has previously been discussed as independent syndromes, as opposed to a spectrum of related syndromes. While there are distinctions between the 12 *HNRNPs* and their related phenotypes, there is also considerable overlap, which is expected due to the known shared targets and functionality of hnRNPs [88]. Here, we examine the phenotypic similarities and differences among probands with variation in NDD *HNRNPs* (Fig. 7 and Additional File 2: Figs. S4 and S5, Additional File 1: Table S7).

Our findings support the five reported disorders associated with *HNRNPs*—Au-Kline syndrome (AKS, *HNRNPK*), Bain-type ID (*HNRNPH2*), *HNRNPH1*-related syndromic ID, *HNRNPR*-related syndrome, and *HNRNPU*-related disorder—and we propose seven new disorders [12–18, 21–23, 25–27, 30].

Phenotypic comparisons and description

Overall, the phenotypes across the *HNRNP*-related disorders are significantly highly correlated ($p > 0.05$ for all pairs of *HNRNPs* except for *HNRNPF* likely due to small sample size; Fig. 7a). These correlations are observed when considering mutation type as well (Additional File 2: Fig. S4), which is more critical for missense variation as phenotypes could be due to a variety of molecular

changes. Furthermore, this correlation is in line with the degree of protein homology, with those being more similar at a protein level also being more alike at the phenotypic level (Fig. 7b). We caution that ascertainment bias likely contributes to some of this, as the majority of our cohorts have DD/ID and/or ASD, and the availability of phenotypic information varies by gene and clinical referring center.

Neurobehavioral phenotypes have the most overlap among *HNRNP* genetic disorders (Fig. 7c, d, Additional File 1: Table S7). As expected, the most common phenotype among probands in our cohort is DD/ID (88.9%, $n = 192/216$, 12/12 disorders). DD/ID varies among disorders, with diagnoses ranging from 44.4 to 100% of probands but occurs at similar rates among LGD and missense probands (Additional File 2: Fig. S5). Delayed speech and language development (57.9%, $n = 114/197$, 11/12 disorders) is common among all *HNRNP* genetic disorders except *HNRNPF*-related disorder, again with similar frequency among both mutation types. Probands with Bain-type ID (*HNRNPH2*) tend to have regression more often than other *HNRNP*-related disorders (34.8%, $n = 8/23$). Seizures are reported in 45.6% of probands ($n = 98/215$, 10/12 disorders) with variation in *HNRNPs*, primarily driven by probands with *HNRNPU*-related disorder (83.1%, $n = 69/83$), who also have a significantly higher prevalence of seizures compared to other *HNRNP*-related disorders ($p = 0.009$ compared to *HNRNPD* probands, $p = 0.009$ compared to *HNRNPH2* probands, $p = 1.89 \times 10^{-10}$ compared to *HNRNPK* probands, $p = 0.002$ compared to *SYNCRIP* probands; Fig. 7d, Additional File 1: Table S8). Seizures overall are more common among LGD probands, but only with the inclusion of *HNRNPU* and *HNRNPR* probands (OR = 2.7, 95% CI 1.5–5.2, Fisher's exact test).

A range of brain imaging abnormalities are observed among 37% ($n = 70/189$) of probands, with abnormalities of the corpus callosum (particularly for *HNRNPR* probands) and cerebellar vermis hypoplasia (particularly among *HNRNPH1*, *HNRNPH2*, and *HNRNPR* probands, although not significantly) being most common. Behavioral diagnoses, including ASD (35.6%, $n = 73/205$, 10/12 disorders), ADHD (8.7%, $n = 17/195$, 6/12 disorders), among others, are shared across disorders. While not significantly different, ASD is more common among probands with variation in *HNRNPD* (46.2%, $n = 6/13$), *HNRNPF* (100%, $n = 3$), *HNRNPR* (44.4%, $n = 4/9$), *SYNCRIP* (57.7%, $n = 15/26$), and *HNRNPUL2* (66.7%, $n = 4/6$), while ADHD is prevalent among *HNRNPR* probands (44.4%, $n = 4/9$) and aggression is common among *HNRNPD* probands (33.3%, $n = 4/12$). *HNRNPU*-related disorder probands with LGD variants also exhibit stereotypy (21.7%, $n = 15/69$, mostly recurrent hand flapping), as do *HNRNPR*-related disorder probands (33.3%, $n = 3/9$). Abnormal movements (22.2%, $n = 44/198$) are

reported in a subset of *HNRNP* genetic disorders, including abnormal gait among probands with *SYNCRIP*-related disorder (20%, $n = 5/25$), dystonia for probands with *HNRNPD*-related disorder (25%, $n = 3/12$), and poor gross motor coordination among probands with *HNRNPUL2*-related disorder (33.3%, $n = 2/6$).

Physical abnormalities appear to be specific to certain disorders. The exception to this are abnormalities of the hands and feet (36.7%, $n = 73/199$, 9/12 disorders), including small hands and feet, abnormalities of palmar creases, clinodactyly and brachydactyly (the latter two prevalent among probands with *HNRNPR*-related syndrome, 77.8%, $n = 7/9$ and 44.4%, $n = 4/9$, respectively, clinodactyly significantly more prevalent among *HNRNPR* probands compared to *SYNCRIP* [$p = 0.02$] and *HNRNPU* probands [$p = 0.0004$]), among others with less consistency. A subset of probands have growth delay (30.5%, $n = 61/200$), particularly probands with variation in *HNRNPH1* (62.5%, $n = 5/8$), *HNRNPH2* (47.8%, $n = 11/23$), *HNRNPR* (90%, $n = 9/10$, significantly more so than *HNRNPK* probands [$p = 0.03$]), and *HNRNPUL2* (50%, $n = 3/6$). Muscular abnormalities are common (46.1%, $n = 89/193$), with hypotonia (44%) occurring more often than hypertonia (6.2%). Hypertonia is relatively specific to *HNRNPUL1* variation (22.2%, $n = 2/9$). Of eye abnormalities (41.2%, $n = 82/199$), strabismus is the most common, observed frequently in *HNRNPAB*-related disorder (66.7%, $n = 2/3$), *HNRNPH1*-related syndromic ID (62.5%, $n = 5/8$), and *HNRNPR*-related syndrome (55.6%, $n = 5/9$). Hearing impairment is observed primarily among probands with *HNRNPH3* variation (66.7%, $n = 2/3$), although with only three probands it's unclear if this is characteristic of a majority of *HNRNPH3* variants. Skeletal morphological abnormalities (29.9%, $n = 59/197$) are primarily observed among probands with AKS (*HNRNPK*, 63%, $n = 17/27$ and *HNRNPR*-related syndrome (70%, $n = 7/10$). The most common among these probands are abnormalities of curvature of the vertebral column, including scoliosis and kyphosis (*HNRNPK*: 39.1%, $n = 9/23$; *HNRNPR*: 33.3%, $n = 3/9$). Skeletal physiological abnormalities, including joint hyper-extensibility or laxity, only occur in 27.6% ($n = 43/190$) of probands over all, but are particularly common among probands with *HNRNPD*-related disorder (20%, $n = 2/10$), *HNRNPH1*-related syndromic ID (50%, $n = 4/8$), Bain-type ID (39.1%, $n = 9/23$), AKS (27.6%, $n = 8/29$), and *SYNCRIP*-related disorder (26.1%, $n = 6/23$). Cardiac abnormalities are another known AKS phenotype, occurring in half (50%, $n = 15/30$) of probands with AKS, but only 19.4% ($n = 38/196$) of all *HNRNP* genetic disorder probands [13]. Genitourinary abnormalities (21.5%, $n = 43/200$) are a known AKS phenotype (63.3%, $n = 19/30$) and occur significantly more often among probands with AKS ($p = 0.02$ compared to Bain-type ID, $p = 0.0001$ compared to *HNRNPU*-related

disorder). The most common genitourinary phenotypes observed in AKS are vesicoureteral reflux (26.7%, $n = 8/30$) and hydronephrosis (16.7%, $n = 5/30$). Enuresis is specific to *HNRNPD*-related disorder (23.1%, $n = 3/13$), while hypospadias is mostly specific to *HNRNPHI*-related syndromic ID (40% of missense variants, $n = 2/5$).

Dysmorphic facies (56.8%, $n = 113/199$) tend to not be consistent, although a few are seen among multiple disorders, including ear abnormalities (AKS, *HNRNPR*-related syndrome, *HNRNPD*-related disorder, and *HNRNPU*-related disorders), small nasal alae (*HNRNPHI*-related syndromic ID and AKS), microcephaly (AKS, *HNRNPHI*-related disorder, *HNRNPR*-related syndrome, and *HNRNPU*-related disorder), and round faces (*HNRNPD*-related disorder, Bain-type ID, and *HNRNPR*-related syndrome) (Additional File 1: Table S7). *HNRNPD*-related disorder most consistently results in a round face (23.1%, $n = 3/13$) and large ears (25%, $n = 3/12$). *HNRNPHI*-related syndromic ID probands tend to have hypoplastic alae (25%, $n = 2/8$). Bain-type ID probands tend to have abnormal palpebral fissures (22.7%, $n = 5/22$), consistent with the almond-shaped eyes previously reported for this syndrome. AKS dysmorphic features have been previously described and also consist of cleft palate (31%, $n = 9/29$), downturned corners of the mouth (20.7%, $n = 6/29$), an open bite (20.7%, $n = 6/29$), teeth abnormalities (17.2%, $n = 5/29$), a wide nasal bridge (31%, $n = 9/29$), shallow orbits/prominent eyes (20%, $n = 6/30$ each), long palpebral fissures (41.4%, $n = 11/26$), ptosis (30%, $n = 9/30$), abnormal eyebrows (20.7%, $n = 6/29$), metopic ridging (24.1%, $n = 7/29$), and abnormal nipples (20%, $n = 6/30$), all of which are more prevalent among probands with LGD variants [13]. *HNRNPR* probands have a large range of facies as well, including micro/retrognathia (50%, $n = 4/8$), teeth abnormalities (50%, $n = 4/8$), anteverted nares (37.5%, $n = 3/8$), a low hanging columella (25%, $n = 2/8$), a narrow forehead (37.5%, $n = 3/8$), up-slanted palpebral fissures (37.5%, $n = 3/8$), unusual hair (37.5%, $n = 3/8$), a broad and short neck (22.2%, $n = 2/9$), and brachycephaly (22.2%, $n = 2/9$).

Other *HNRNP* genes

In addition to *HNRNPs* with multiple cases of disruptive DNVs, we identified seven with less than three cases: *HNRNPA0* ($n = 2$, missense), *HNRNPA1* ($n = 1$, missense), *HNRNPA3* ($n = 1$, missense), *HNRNPA2B1* ($n = 2$, LGD), *HNRNPC* ($n = 2$, 1 LGD and 1 missense), *HNRNPLL* ($n = 1$, missense), and *PTBP2* ($n = 1$, missense) (Additional File 1: Table S10). Four of these seven (*HNRNPA0*, *HNRNPA3*, *HNRNPA1*, and *HNRNPA2B1*) are in a group closely related to an NDD *hNRNP* group (group 2; Fig. 1), suggesting that they may be implicated in NDDs, but mutations occur much less frequently. These were not included in the de novo or phenotypic

statistical analyses but may be of importance for future screening of patients for candidate NDD-related genes.

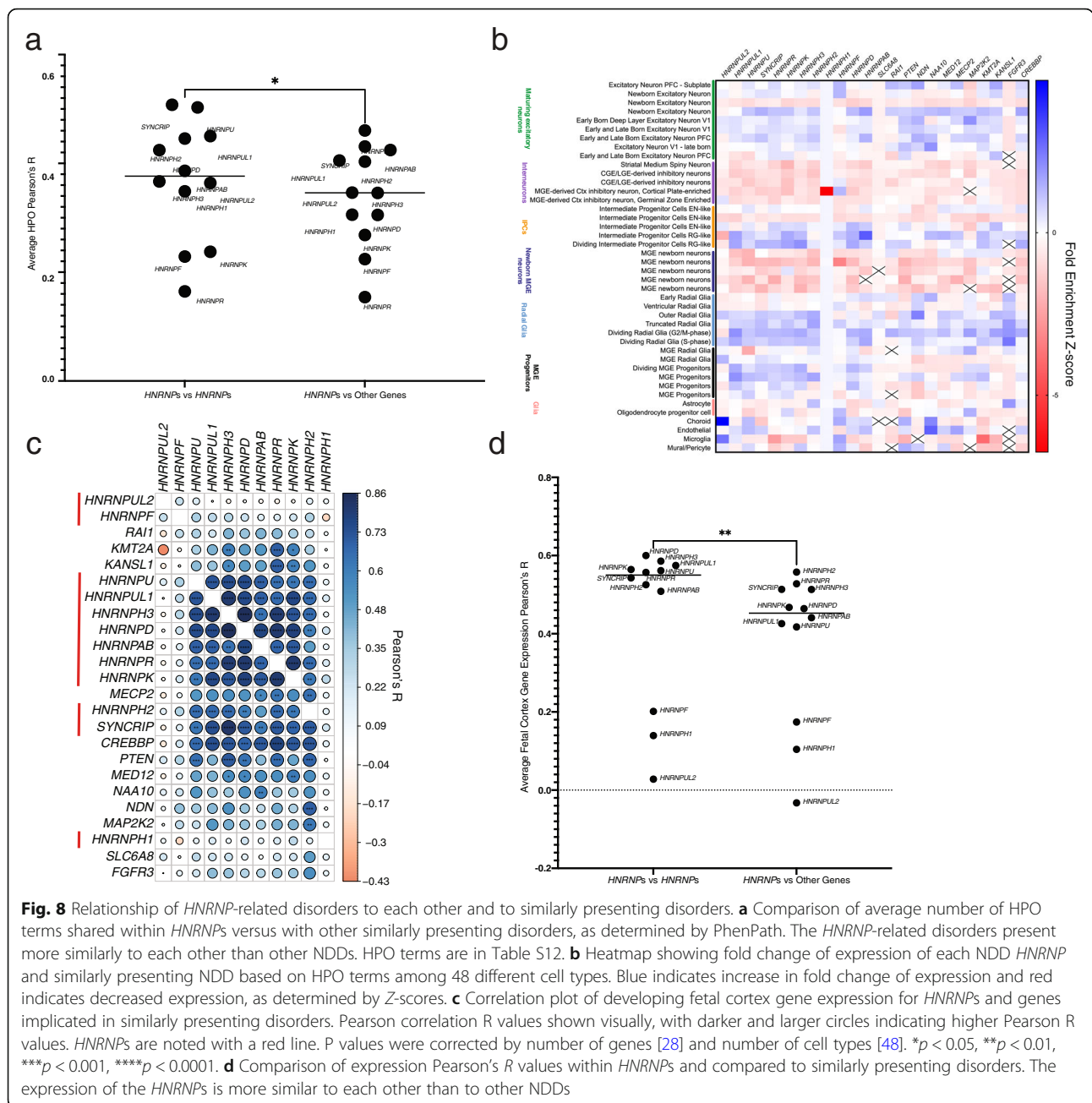
Comparison to similar disorders

It is plausible that similarities observed between *HNRNP*-related disorders may be due to ascertainment and that our *HNRNP*-related disorders may actually have more similarities to other known genetic disorders. To address this, we compared each disorder's characteristic features to those of known NDDs (Fig. 7). HPO terms that occur in 20% or more probands were entered into the PhenPathTOOL and genes sharing the most HPO terms were compared to the expression of the *HNRNP* (Additional File 1: Table S12) [86]. HPO terms are shared with *SLC6A8* (cerebral creatine deficiency syndrome 1), *CREBBP* (Rubenstein-Taybi syndrome), *MECP2* (Rett syndrome, mental retardation, X-linked syndromic, Lubs type), *KANSL1* (Koolen-de Vries syndrome), *RAI1* (Smith-Magenis syndrome), *PTEN*, *SNRN*, *MAP2K2*, *FGFR3*, *KMT2A*, *MED12*, *NDN*, and *NAA10*. Notably, AKS has been suggested to be similar to Kabuki syndrome, although they do not share the most HPO terms. Additionally, methylation analyses have shown that these disorders have distinct profiles, suggesting that their shared clinical features are likely not simply due to shared genetic causes (Au et al., ASHG poster 3117). The number of genes with similar phenotypes that overlap between the *HNRNPs* highlights their shared phenotypic spectra. While there are similarities with other NDDs, correlation of HPO terms present for each disorder is higher among the *HNRNP*-related disorders, suggesting they may have shared molecular pathogenesis ($p = 0.0013$, paired *t*-test, Fig. 8a).

We assessed if shared gene expression is seen among similarly presenting disorders in developing cortical tissues (Fig. 8b, Additional File 2: Fig. S6). While there is a positive correlation (Pearson's $R > 0.5$) among almost all genes in our analyses, correlations between genes with shared HPO terms are lower than those among the *HNRNPs* (Fig. 7c). This suggests that, while there are some shared expression patterns that are likely involved in neurodevelopment broadly, the *HNRNP*-related genetic disorders are indeed a unique family based on gene expression.

Discussion

The goal of this study was to investigate genotype-phenotype correlations in the context of a molecularly related gene family. We selected the *HNRNPs* because they have been implicated in NDDs due to rare missense and LGD variation, as well as neurodegenerative disease and cancers primarily through changes in function, expression, and/or localization (Table 1) [8, 9, 12–14, 18, 20–23, 25, 27, 28, 89–103]. In addition, *HNRNPs* are



candidates for haploinsufficiency for multiple contiguous gene deletion syndromes resulting in NDDs (*HNRNPK*, *SYNCRIP*, and *HNRNPU*, and likely *HNRNPD* and *HNRNPR*). More recently, SNVs and indels in *HNRNPU*, *HNRNPK*, *HNRNPH1/2*, and *HNRNPR* have been reported in probands with NDDs, and *HNRNPU* and *SYNCRIP* show an excess of DNVs in individuals with NDDs from recent large-scale sequencing studies [11–18, 20–23, 25–27, 35–39, 41, 58, 60, 104–106]. Thus, we hypothesized that there are (1) similar phenotypic (clinical and molecular) spectra across these disorders due to

shared structure and function and (2) additional *HNRNP*-related disorders yet to be observed. From our analyses, we identified 12 *HNRNPs* as particularly relevant to NDDs, which we term NDD *HNRNPs*: *HNRNPAB*, *HNRNPD*, *HNRNPF*, *HNRNPH1*, *HNRNPH2*, *HNRNPH3*, *HNRNPK*, *HNRNPR*, *SYNCRIP*, *HNRNPU*, *HNRNPUL1*, and *HNRNPUL2*. Multiple lines of evidence, including protein homology, expression analysis, and mutation intolerance, indicate that genes/proteins within this group are more molecularly related and more sensitive to variation than the other *HNRNP* family

members. In addition to the five *HNRNPs* previously implicated in NDDs, we report seven new *HNRNPs* as candidates for emerging disorders and have integrated and compared their phenotypic features (Fig. 7, Additional File 1: Tables S7) [19, 38, 66, 104].

From over 80,000 individuals, we identify 225 probands (115 novel) with likely pathogenic SNVs and indels among the NDD *HNRNPs*—almost all of which with known inheritance are de novo (Table 2, Fig. 5). This study, thus, increases the sample size for de novo analyses by almost eightfold when compared to previous surveys, identifying a third gene family member, *HNRNPD*, that achieves nominal significance for an excess of de novo LGD or missense mutations in comparison to chimpanzee- and macaque-human mutation models (Fig. 4a). With a larger sample size, we predict that most of these NDD *HNRNPs* will ultimately become significantly associated with an excess of DNVs. Due to the limitations of these models and modes of ascertainment, many variants were not considered in de novo analyses ($n = 147/225$ not in analyses); therefore, we are underestimating the significance of variation in these genes. Overall, in large cohorts evolutionary mutational modeling provides a valuable resource for gene discovery and statistical support for NDD genes but still may miss rare disorders that are observed in the clinic. Additionally, we identified multiple disorders (*HNRNPF*-, *HNRNPH3*-, *HNRNPUL1*-, and *HNRNPUL2*-related disorders) that were not identified in the clinic or by previous statistical analyses, but by comparison to known NDDs. This highlights the need for intersection between statistical modeling, clinical identification, and novel approaches to identifying NDD candidate genes to uncover the rarest of disorders.

Our results support emerging syndromes associated with seven genes and expand upon known *HNRNP*-related disorders [10, 19, 38, 66, 104]. We are able to explain some of the phenotypic similarities and differences based on protein homology, gene expression—which is highly correlated among almost all *HNRNPs*—and known shared function. At cell type resolution, NDD *HNRNP* expression is most strongly enriched among progenitor populations of the human cerebral cortex and MGE, including mitotic radial glia, and their expression is depleted in both early-born excitatory neurons and newborn MGE neurons (Fig. 6b, Additional File 2: S3A,B). Radial glia, one of the earliest cells found in brain development, act as neural stem cells of the cerebral cortex and are important in the differentiation and migration of many neuronal cell types. Enriched expression of NDD *HNRNPs* in actively dividing radial glia suggests a possibility that disruption of hnRNPs may lead to aberrant neuronal development. Disruption of radial glia can happen at many steps in early development,

including changes in proliferation leading to differing number of neurons, changes in scaffolding leading to impaired neuronal migration, and subtler defects in neurite extension, synaptogenesis, and neuronal connectivity [107–109]. These types of changes have been detected in ASD, DD/ID, epilepsy, and schizophrenia. Interestingly, some NDD hnRNPs have already been implicated in cell division through cancer studies, including hnRNPD, hnRNPK, SYNCRIP, and hnRNPU [110–114]. Thus, it is plausible that altered cell division, differentiation, and neuronal migration may be playing a role in the pathogenesis of *HNRNP* variation. At a brain region level, we see enrichment in the early fetal striatum and the early-mid fetal amygdala, both of which have been implicated in NDDs [1, 115–117]. At the adult tissue level, it is clear that *HNRNPK* has the highest and most ubiquitous expression, which may account for some of the more physical phenotypes observed in probands with AKS. hnRNPR is also known to regulate *HOX* genes, which likely plays a role in the physical manifestations seen in *HNRNPR*-related disorder, and dysregulation of *HOX* genes may be important to other *HNRNP*-related disorders as well [18]. Overall, we observe that missense variants cluster for all hnRNPs more so among probands than controls, and especially for hnRNPH2, hnRNPK, hnRNPR, and hnRNPU1, and that particular protein domains appear to be impacted, including RNA- and DNA-binding domains.

Consistent with our ascertainment, DD/ID are the most prevalently shared features among all probands with *HNRNP*-related disorders, which may be due to ascertainment bias (Fig. 7). Over half have speech delay and over 40% have seizures and/or hypotonia, while over one third each have structural brain abnormalities or ASD. Structural brain abnormalities include abnormalities of the corpus callosum, cerebellar vermis hypoplasia, and dysmorphic ventricles, which may be explained by the high radial glia expression of *HNRNPs*. Most have some dysmorphic features, and several have a range of physical abnormalities impacting the cardiac, skeletal, and genitourinary systems, although generally these are inconsistent, except for AKS and *HNRNPR*-related syndrome, which have been previously described (Additional File 1: Table S7) [13, 18].

While there is subjectively a pattern of phenotypes, we also observe a strong correlation between almost all of the *HNRNP*-related disorders, as hypothesized given they have similar molecular structure and function and shared gene expression (Fig. 7a). Notably, we observe that phenotypic correlations follow a similar pattern to those of protein homology, suggesting that more similar proteins molecularly do have more similar consequences when disrupted (Fig. 7b). *HNRNPF*-related disorder has the weakest correlations, although this is likely due to

(1) lower sample size ($n = 3$) with limited phenotypic information and (2) a potential association with autism without intellectual disability ($n = 1/2$ with ASD and no DD/ID) [34]. There are similar patterns in correlations based on LGD and severe missense variation, highlighting the shared impacts of the two types of variants on each gene (Additional File 2: Fig. S4). While we have far fewer probands with missense variation, they typically have a less severe phenotype, and their pathogenic significance remains to be determined given the small sample size.

Overall, our phenotypic analyses of the previously reported *HNRNPU*-related disorder, AKS, Bain-type ID, *HNRNPH1*-related syndromic ID, and *HNRNPR*-related syndrome are consistent with what has been published [12, 13, 18, 20–23, 25–27, 60]. We generally find that the phenotypic spectrum of newly recruited patients matches what has previously been reported for LGD mutations, although the phenotypic consequences of missense variation are less well established. Only a few differences were observed. One is that our data do not support the previous suggestion that probands with *HNRNPU*-related disorder have renal and cardiac abnormalities, or these phenotypes have much lower penetrance. Another is that ASD has a higher prevalence than previously reported for AKS, particularly for missense variants. We also provide evidence phenotypically and based on gene expression that *HNRNPH1*-related syndromic ID and Bain-type ID (*HNRNPH2*) are indeed independent disorders, which has been under debate. Also supporting this are the variant types, which cluster in the M9 nuclear targeting region of *hnRNPH2*, likely impacting its localization and global translation based on in vitro work, but are seen across the entirety of *hnRNPH1* [118, 119].

We report seven novel disorders: *SYNCRIP*-, *HNRNPD*-, *HNRNPUL1*-, *HNRNPUL2*-, *HNRNPAB*-, *HNRNPF*-, and *HNRNPAB*-related disorders. Both *SYNCRIP* and *HNRNPD* have been NDD candidates due to being in the critical region of genomic disorders (Table 3). *SYNCRIP*, having high protein homology to *hnRNPR*, has also been observed to have recurrent de novo LGD variants in large NDD sequencing studies [39, 69, 74]. Indeed, we see support of this both phenotypically and statistically, with *SYNCRIP* LGD variants reaching exome-wide significance by the CH model (Fig. 5). We propose that *SYNCRIP*-related disorder primarily manifests as DD/ID, ASD, ADHD, hypotonia, speech delay, and structural brain abnormalities (including abnormal/dysmorphic ventricles, abnormalities of the corpus callosum, and Chiari malformations), with some probands having abnormal gait or ataxia, hyper-extensible joints, and hand and feet abnormalities (Additional File 1: Table S7). Many *SYNCRIP* RNA and protein interactors

are critical in neuronal RNA and membrane trafficking, synaptic plasticity, axonogenesis, neuronal morphology, and dendritic translation [120–125].

Probands with *HNRNPD*-related syndrome show a high prevalence of DD/ID, speech delay, and ASD and/or other behavioral phenotypes (Additional File 1: Table S7) [55]. These probands also have consistent facial features of a round face, epicanthus, and large ears. Other emotional disturbances are also noted, including aggression, intermittent explosive disorder, and depression. It has been suggested that the gene may play a role in the development of enkephalergic neurons, known to be important in neurodevelopment, particularly for emotional responses and ASD, which aligns with more emotional disturbances in *HNRND* probands [126, 127]. *HNRNPD* expression is also altered following NMDA receptor stimulation, which may impact the wide range of intracellular cascades, including those involved in cell survival, differentiation, and neuroplasticity.

hnRNPU and *hnRNPU* have high protein homology to *hnRNPU*, although they appear to have milder phenotypes when disrupted. *HNRNPUL1*-related disorder consists of DD/ID, although about half as often as with *HNRNPU* variation, short stature, motor and speech delay, structural brain, skeletal, and cardiac abnormalities, abnormal gait, ASD, failure to thrive, and hypertonia (Additional File 1: Table S7). Previously, it was suggested that missense variation in *HNRNPUL1* contributes to “high-functioning” autism, meaning an autism diagnosis and IQ in the normal range, which our data support [34]. Missense variants cluster significantly among probands compared to controls, although not in a clear protein domain, suggesting that disruption within that part of the protein may be responsible for some phenotypes (Fig. 4). Probands with *HNRNPUL2*-related disorder exhibit DD/ID, ASD, and motor and speech delay. *HNRNPUL2* gene expression has also been shown to be significantly decreased in the blood of first episode psychosis patients, suggesting a neuronal functional role and that we may have under-ascertained individuals with neuropsychiatric phenotypes [128]. Although understudied, both *HNRNPUL1* and *HNRNPUL2* are known to be critical in the DNA damage response, which has been associated with neurodevelopment [129–132].

Three of our novel disorders only have three probands each: *HNRNPAB*-, *HNRNPF*-, and *HNRNPH3*-related disorders, making it difficult to draw conclusions. For probands with *HNRNPAB* variants, DD/ID is common, as are delayed speech and language, ASD, hypotonia, and strabismus. Little is known about *hnRNAB* outside of cancer, but it has been shown to regulate expression of neurodevelopmental genes and those involved in glutamate signaling, as well as neural cell motility [133, 134]. Probands with *HNRNPH3*-related disorder have a

high prevalence of DD/ID, like the other *HNRNPH* disorders, although they manifest characteristic hearing loss as well. However, the gene expression is shared with both other *HNRNPH* genes, suggesting that their shared phenotypes may have similar molecular causes and that additional probands may reveal a stronger phenotypic relationship. hnRNPF is highly homologous to hnRNPH1/2/3 (Fig. 2, group 7), and like those genes, we identified only reports of severe missense variation. Thus, *HNRNPF*-related disorder may have more molecular similarities to *HNRNPH*-related disorders, although with only three probands this is difficult to conclude. All probands have ASD, and half have DD/ID, suggesting that autism with a normal IQ may be an associated phenotype as previously suggested (Table S7) [34]. While little is known about the specific function of hnRNPF, it has been shown to interact with other hnRNPH proteins and regulate neuronal-specific splicing, as well as myelin basic protein synthesis in oligodendrocytes [135, 136]. The lack of LGD variants in our cohort as well as gnomAD for *HNRNPH2*, *HNRNPH3*, and *HNRNPF* may suggest that loss-of-function mutations are embryonic lethal.

As there are some shared features among NDDs overall, we wanted to ensure that the common phenotypes and expression were specific among our *HNRNP*-related disorders compared to similarly presenting disorders. While several disorders share multiple HPO terms with our *HNRNP*-related disorders, the *HNRNP*-related disorders are more similar to each other than to other NDDs. It is plausible that similarly presenting disorders share disrupted gene regulation with the *HNRNP*-related disorders, explaining some shared phenotypes. Based on gene expression in the developing fetal cortex, the expression of the *HNRNPs* are more similar to each other than similarly presenting disorders (Fig. 8c,d). Thus, we conclude that, while of course there are many disorders that may share phenotypes, the *HNRNP*-related disorders are unique in their presentation and gene expression, as well as protein function, supporting a gene family approach.

Our study highlights the value and challenge of integrating data across multiple centers and the published literature for a group of highly heterogeneous NDDs. A strength is our large sample size of > 80,000 individuals, pulling from published work, our own unpublished resources, and international collaborations, but for rare disorders this is still insufficient. As shown by our de novo mutation analyses, the sample sizes used are not high enough to detect evolutionary evidence that variation in many of these genes is pathogenic, but based on clinical patterns, protein homology, and gene expression, there is reason to conclude pathogenicity. Multiple *HNRNPs* have been implicated in autism with normal IQ and/or neuropsychiatric disorders that have little

effect on IQ, which are severely under-ascertained not only in this study but across the board in NDD research. Additionally, the average age of patients in our study is 10.5 years, suggesting that any phenotypes associated with adults are likely missed. Thus, while much of our data shows that there are not many significant differences among *HNRNP*-related disorders, this is likely not the full story of these disorders. Much work is needed to establish the molecular phenotypes to fully understand the pathophysiology of these disorders, which will determine if there are shared pathways affected by these disorders and inform potential treatment strategies.

Conclusions

In summary, the hnRNPs are a family of proteins that have been shown repeatedly to impact neurological disease. Here, we establish both a statistical enrichment of DNVs and phenotypic evidence that many of the *HNRNPs* are involved in NDDs. The conservation, shared expression, and phenotypic consequences clearly support a subset of these as high likelihood candidates of NDD when mutated. More work is necessary to understand the molecular underpinnings of NDD-related *HNRNP* variation, especially since compounds utilized in diseases such as addiction and cancer are available to target, both up- and downregulating, a subset of hnRNPs [83, 137]. Future work may identify hnRNP-targeting compounds and/or downstream-targeting compounds that could prove to be efficacious in modifying patient outcomes. With the shared function, localization, and binding targets of many of the NDD hnRNPs, it is likely that one compound may benefit probands with changes in related genes. Using the hnRNPs as an example, we have provided an evidence-based foundation supporting that, by identifying relevant NDD gene families and distinguishing molecularly related subgroups, characterization of de novo mutated genes among NDDs can be performed more effectively. This facilitates improved diagnosis and prospective assessment of NDDs as well as potential future development of more impactful therapeutics, while also providing information on neurodevelopment as a whole.

Abbreviations

ADHD: Attention deficit hyperactivity disorder; AKS: Au-Kline syndrome; ALS: Amyotrophic lateral sclerosis; ANOVA: Analysis of variance; ASD: Autism spectrum disorder; BGL: Baylor Genetics Laboratory; CH model: Chimpanzee-human divergence model; CLUMP: CLUstering by Mutation Position; CNV: Copy number variant; DD: Developmental delay; DDD: Deciphering Developmental Disorders; DNA: Deoxyribonucleic acid; DNV: De novo variant; FTLD: Frontotemporal lobar degeneration; GERD: Gastroesophageal reflux disease; GeVIR: Gene Variation Intolerance Rank; hnRNP/*HNRNP*: Heterogeneous nuclear ribonucleoprotein (hnRNP for protein, *HNRNP* for gene); HPO: Human Phenotype Ontology; ID: Intellectual disability; LGD: Likely gene disrupting; MIS: Missense; MIS20: Missense variant with CADD score ≥ 20 ; MIS30: Missense variant with CADD score ≥ 30 ; NDD: Neurodevelopmental disorder; NMDA: *N*-methyl-D-aspartate receptor;

pLI: Probability of loss of function; RNA: Ribonucleic acid; SEA: Specific Expression Analyses; SNV: Single-nucleotide variant

Supplementary Information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13073-021-00870-6>.

Additional file 1: Supplementary Tables S1-S12. Table S1. Cohorts utilized in current study. **Table S2.** Pathogenicity predictions by gnomAD and GeVIR for each hnRNP. **Table S3.** Significant *p* values of two-way ANOVA between cell type and transcripts per million (TPM) for each NDD *HNRNP*'s GTEx data. **Table S4.** *P* values and fold change for scRNA data from developing human cortex. **Table S5.** Significant *p* values of two-way ANOVA between cell type and fold expression for each *HNRNP*. **Table S6.** De novo enrichment and CLUMP analyses of current cohort. **Table S7.** Phenotypes among *HNRNP*-related disorders. **Table S8.** Uncorrected *p* values for pairwise Fisher's exact tests for each phenotype occurring in 20% of an *HNRNP* group. **Table S9.** Uncorrected *p* values for phenotype correlations between each *HNRNP*-related disorder. **Table S10.** Variants and phenotypes for *HNRNPs* with < 3 probands. **Table S11.** Probands with copy number variants. **Table S12.** Shared HPO terms with similarly presenting genetic disorders.

Additional file 2: Supplementary Figures S1-S6. Fig. S1.

Dendrogram of hnRNPs based on multiple sequence alignment of canonical amino acid sequences. Colors match those seen in Fig. 2. NDD hnRNPs are shown in black boxes. **Fig. S2.** Pathogenicity assessment of variation in hnRNPs. A) Gene Variation Intolerance Ranking (GeVIR), loss-of-function observed/expected upper bound fraction (LOEUF), and Variation Intolerant Region Loss-of-Function (VIRLoF) percentiles. Average LOEUF percentile is significantly higher for NDD *HNRNPs* ($n = 13$) compared to other *HNRNPs* ($n = 15$). B) Average fold change for GeVIR, LOEUF, and VIRLoF for autosomal dominant (AD) and autosomal recessive (AR) variants. Average LOEUF fold change for AD mutations is significantly higher for NDD *HNRNPs* compared to other *HNRNPs*, with the AD VIRLoF fold change trending in the same direction. The AR LOEUF fold change is trending towards being significantly higher among other *HNRNPs* compared to NDD *HNRNPs*. One-way t-test. * $p < 0.05$. **Fig. S3.** Expression of *HNRNP* among adult tissues and the developing human cortex. A) Heatmap of all *HNRNP* expression in developing cortex tissues. B) Comparison of fold expression of NDD *HNRNPs* to non-NDD *HNRNPs*. C) Heatmap of all *HNRNP* expression (transcript level expression) in adult brain tissues from GTEx. D) Heatmap of NDD *HNRNP* expression (transcript level expression) in all tissues from GTEx. **Fig. S4.** Phenotypic correlations for LGD and missense variant probands. A) Correlation matrix of phenotypes across *HNRNP* probands with LGD variation (genes with only missense variation excluded) and B) severe missense variation (genes with only LGD variation excluded). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. **Fig. S5.** Phenotypic comparisons between LGD and missense variants by *HNRNP*. **Fig. S6.** GTEx expression of NDD *HNRNPs* and genes associated with similarly presenting disorders.

Additional file 3: Table S13. GenBank accession numbers.

Acknowledgements

We thank all of the families participating in this study. We are grateful to all of the families at the participating SSC sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, E. Wijsman). We appreciate obtaining access to phenotypic data on SFARI Base for both SSC and SPARK samples, as well as SPARK exome data from the SPARK Consortium. Approved researchers can obtain the SSC population dataset described in this study (<https://www.sfari.org/resource/resources/simons-simplex-collection/>) by applying at <https://base.sfari.org>. We thank the DDD study, which presents independent research commissioned by the Health Innovation Challenge Fund (grant number HICF-1009-003), a parallel funding partnership between the Wellcome Trust and the Department of Health, and the Wellcome Trust Sanger Institute (grant number WT098051). The views expressed in this publication are those of the authors and not necessarily

those of the Wellcome Trust or the Department of Health. The study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC). The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network. The authors would like to thank the University of Washington Center for Mendelian Genomics and all contributors to MyGene2 for use of data included in MyGene2. 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 content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We thank Emily Guilbert for aid in recontacting a family included in this study. We thank Tonia Brown for assistance in editing this manuscript and Amy Wilfert for advice on statistics.

Consortia

SPARK Consortium: John Acampado¹, Andrea J. Ace¹, Alpha Amatya¹, Irina Astrovskaya¹, Asif Bashar¹, Elizabeth Brooks¹, Martin E. Butler¹, Lindsey A. Cartner¹, Wubin Chin¹, Wendy K. Chung^{1,2}, Amy M. Daniels¹, Pamela Feliciano¹, Chris Fleisch¹, Swami Ganesan¹, William Jensen¹, Alex E. Lash¹, Richard Marini¹, Vincent J. Myers¹, Eirene O'Connor¹, Chris Rigby¹, Beverly E. Robertson¹, Neelay Shah¹, Swapnil Shah¹, Emily Singer¹, LeeAnne G. Snyder¹, Alexandra N. Stephens¹, Jennifer Tjernagel¹, Brianna M. Vernoia¹, Natalia Volfovsky¹, Lorán Casey White¹, Alexander Hsieh², Yufeng Shen², Xueya Zhou², Tychele N. Turner³, Ethan Bahl⁴, Taylor R. Thomas⁴, Leo Brueggeman⁴, Tanner Koomar⁴, Jacob J. Michaelson⁴, Brian J. O'Roak⁵, Rebecca A. Barnard⁵, Richard A. Gibbs⁶, Donna Muzny⁶, Aniko Sabo⁶, Kelli L. Baalman Ahmed⁶, Evan E. Eichler⁷, Matthew Siegel⁸, Leonard Abbeduto⁹, David G. Amaral⁹, Brittani A. Hilscher⁹, Deana Li⁹, Kaitlin Smith⁹, Samantha Thompson⁹, Charles Albright¹⁰, Eric M. Butter¹⁰, Sara Eldred¹⁰, Nathan Hanna¹⁰, Mark Jones¹⁰, Daniel Lee Coury¹⁰, Jessica Scherr¹⁰, Taylor Pifher¹⁰, Erin Roby¹⁰, Brandy Dennis¹⁰, Lorrin Higgins¹⁰, Melissa Brown¹⁰, Michael Alessandri¹¹, Anibal Gutierrez¹¹, Melissa N. Hale¹¹, Lynette M. Herbert¹¹, Hoa Lam Schneider¹¹, Giancarla David¹¹, Robert D. Annett¹², Dustin E. Sarver¹², Ivette Arriaga¹³, Alexia Camba¹³, Amanda C. Gulsrud¹³, Monica Haley¹³, James T. McCracken¹³, Sophia Sandhu¹³, Maira Tafolla¹³, Wha S. Yang¹³, Laura A. Carpenter¹⁴, Catherine C. Bradley¹⁴, Frampton Gwynette¹⁴, Patricia Manning¹⁵, Rebecca Shaffer¹⁵, Carrie Thomas¹⁵, Raphael A. Bernier¹⁶, Emily A. Fox¹⁶, Jennifer A. Gerds¹⁶, Micah Pepper¹⁶, Theodore Ho¹⁶, Daniel Cho¹⁶, Joseph Piven¹⁷, Holly Lechniak¹⁸, Latha V. Soorya¹⁸, Rachel Gordon¹⁸, Allison Wainer¹⁸, Lisa Yeh¹⁸, Cesar Ochoa-Lubinoff¹⁹, Nicole Russo¹⁹, Elizabeth Berry-Kravis²⁰, Stephanie Booker²¹, Craig A. Erickson²¹, Lisa M. Prock²², Katherine G. Pawlowski²², Emily T. Matthews²², Stephanie J. Brewster²², Margaret A. Hojlo²², Evi Abada²², Elena Lamarche²³, Tianyun Wang²⁴, Shwetha C. Murali²⁴, William T. Harvey²⁴, Hannah E. Kaplan²⁵, Karen L. Pierce²⁵, Lindsey DeMarco²⁶, Susannah Horner²⁶, Juhi Pandey²⁶, Samantha Plate²⁶, Mustafa Sahin²⁷, Katherine D. Riley²⁷, Erin Carmody²⁷, Julia Constantini⁷, Amy Esler²⁸, Ali Fatemi²⁹, Hanna Hutter²⁹, Rebecca J. Landa²⁹, Alexander P. McKenzie²⁹, Jason Neely²⁹, Vini Singh²⁹, Bonnie Van Metre²⁹, Ericka L. Wodka²⁹, Eric J. Fombonne³⁰, Lark Y. Huang-Storms³⁰, Lillian D. Pacheco³⁰, Sarah A. Mastel³⁰, Leigh A. Coppola³⁰, Sunday Francis³¹, Andrea Jarrett³¹, Suma Jacob³¹, Natasha Lillie³¹, Jaclyn Gunderson³¹, Dalia Istephanous³¹, Laura Simon³¹, Ori Wasserberg³¹, Angela L. Rachubinski³², Cordelia R. Rosenberg³², Stephen M. Kanne^{33,34}, Amanda D. Shocklee³⁴, Nicole Takahashi³⁴, Shelby L. Bridwell³⁴, Rebecca L. Klimczak³⁴, Melissa A. Mahurin³⁴, Hannah E. Cottrell³⁴, Cortaiga A. Grant³⁴, Samantha G. Hunter³⁴, Christa Lese Martin³⁵, Cora M. Taylor³⁵, Lauren K. Walsh³⁵, Katherine A. Dent³⁵, Andrew Mason³⁶, Anthony Sziklay³⁶, Christopher J. Smith³⁶.

¹ Simons Foundation, New York, USA

² Columbia University, New York, USA

³ Washington University School of Medicine, St. Louis, USA

⁴ University of Iowa Carver College of Medicine, Iowa City, USA

⁵ Oregon Health & Science University, Portland, USA

⁶ Baylor College of Medicine, Houston, USA

⁷ University of Washington School of Medicine, Howard Hughes Medical Institute, Seattle, USA

⁸ Maine Medical Center Research Institute, Portland, USA

⁹ University of California, Davis, Sacramento, USA

¹⁰ Nationwide Children's Hospital, Columbus, USA

¹¹ University of Miami, Coral Gables, USA

¹² University of Mississippi Medical Center, Jackson, USA

- ¹³ University of California, Los Angeles, Los Angeles, USA
¹⁴ Medical University of Southern Carolina (MUSC), Portland, USA
¹⁵ Cincinnati Children's Hospital Medical Center – Research Foundation, Cincinnati, USA
¹⁶ Seattle Children's Autism Center/UW, Seattle, USA
¹⁷ University of North Carolina at Chapel Hill, Chapel Hill, USA
¹⁸ Department of Child & Adolescent Psychiatry, Rush University Medical Center, Chicago, USA
¹⁹ Department of Developmental & Behavioral Pediatrics, Rush University Medical Center, Chicago, USA
²⁰ Department of Neurological Sciences, Department of Pediatrics, Department of Biochemistry, Rush University Medical Center, Chicago, USA
²¹ Cincinnati Children's Hospital Medical Center – Research Foundation, Cincinnati, USA
²² Boston Children's Hospital (BCH), Boston, USA
²³ University of North Carolina at Chapel Hill, Chapel Hill, USA
²⁴ University of Washington School of Medicine, Seattle, USA
²⁵ University of California, San Diego, School of Medicine, La Jolla, USA
²⁶ Children's Hospital of Philadelphia, Philadelphia, USA
²⁷ Boston Children's Hospital (BCH), Boston, USA
²⁸ University of Minnesota, Minneapolis, USA
²⁹ Kennedy Krieger Institute, Baltimore, USA
³⁰ Oregon Health & Science University, Portland, USA
³¹ University of Minnesota, Minneapolis, USA
³² University of Colorado School of Medicine, Aurora, USA
³³ Department of Health Psychology, University of Missouri, Columbia, USA
³⁴ Thompson Center for Autism and Neurodevelopmental Disorders, University of Missouri, Columbia, USA
³⁵ Geisinger Autism & Developmental Medicine Institute, Lewisburg, USA
³⁶ Southwest Autism Research and Resource Center, Phoenix, USA

Authors' contributions

M.A.G. and E.E.E. led the conception and design of the work and manuscript drafting. T.W. and K.H. performed sequencing on ASID samples, and T.W. participated in analysis of SPARK data. J.R. and P.L. analyzed and contributed BGL data. R.A.B., R.K.E., M.P., and H.P. collected and provided data from probands at the UW Autism Center, as well as information from probands in the SSC and other cohorts. E.T. provided information on probands from GeneDx. CNK and T.N. analyzed and contributed the scRNA-seq data. B.B.A.d.V and L.E.L.M.V. provided genomic and phenotypic data on RUMC probands. All other authors provided phenotypic information on probands. All authors read and approved the final manuscript.

Funding

This work was supported, in part, by the U.S. National Institutes of Health (R01MH101221) to E.E.E. Research reported in this publication was supported, in part, by the National Institute of Neurological Disorders and Stroke (NINDS) under award number K08NS092898, Jordan's Guardian Angels, and the Brotman Baty Institute (to G.M.M.). M.I., A.C., and A.S. were supported by the GENE (Genomic analysis Evaluation Network) Research Project founded by Progetti di Innovazione in Ambito Sanitario e Socio Sanitario (Bando EX decreto n.2713 28.02.2018) Regione Lombardia. D. L. was supported by the German Research Foundation (DFG; LE 4223/1). B.B.A.d.V. and L.E.L.M.V. were supported by grants from the Dutch Organization for Health Research and Development (ZON-MW grants 917–86–319 and 912–12–109). M.E., O.G., and C.R. received funding from the Italian Ministry of Health (Project RC n. 2757328). I.T. is supported by generous donors to the Children's Mercy Research Institute and the Genomic Answers for Kids program. K.X. is supported by the National Natural Science Foundation of China (NSFC: 8173000779) and the Science and Technology Major Project of Hunan Provincial Science and Technology Department (2018SK1030). M.A.G. was supported by the U.S. National Institutes of Health (T32HG000035). E.E.E. is an investigator of the Howard Hughes Medical Institute.

Availability of data and materials

All data from probands generated or analyzed during this study are included in this published article and its supplementary information files. Accession numbers can be found in Additional File 3: Table S13 [138]. All other data is available from previously published resources.

Web Resources

ape: <http://ape-package.ird.fr/>
 CADD: <https://cadd.gs.washington.edu/>
 ClinVar: <https://www.ncbi.nlm.nih.gov/clinvar/>
 Clustal Omega: <http://clustal.org/omega>

corrplot: <https://github.com/taiyun/corrplot>
 corr: <https://github.com/tidymodels/corr>
 CLUMP: <https://github.com/tycheleturner/clump>
 CSEA: <http://genetics.wustl.edu/jdlab/csea-tool-2/>
 DECIPHER: <http://depher.sanger.ac.uk/>
 denovo-db: <http://denovo-db.gs.washington.edu>
 Ensembl VEP (hg38): http://uswest.ensembl.org/Homo_sapiens/Tools/VEP?db=core
 GeneMatcher: <http://genematcher.org>
 GenBank: <https://www.ncbi.nlm.nih.gov/genbank/>
 GeViR: <http://www.gevirank.org/>
 gnomAD: <https://gnomad.broadinstitute.org/>
 GTeX: <https://gtexportal.org/>
 MyGene2: <https://mygene2.org/MyGene2/>
 MetaDome: <https://stuart.radboudumc.nl/metadome/>
 NMDescPredictor: <https://nmdprediction.shinyapps.io/nmdescpredictor/>
 PhenPath: <http://phenpath.biocomp.unibo.it/phenpath/>
 RVAideMemoire: <https://CRAN.R-project.org/package=RVAideMemoire>
 SFARI Base: <https://base.sfari.org>
 SPARK: <https://sparkforautism.org/>
 UCSC Cell Browser: <https://cells.ucsc.edu>
 Uniprot: <http://uniprot.org>

Declarations

Ethics approval and consent to participate

Probands or their families were contacted directly and consented by the University of Washington Autism Center under the TIGER study (IRB#: STUDY00000813) to participate in the study. Patient samples were obtained from the ASID network with informed consent, and all clinical recontact performed on the individuals in this study were approved by the University of Washington Institutional Review Board (IRB), in accordance with the ethical standards of the responsible local institutional and national committees.

Consent for publication

Deidentified clinical data (case histories and clinical information) and variant data (exempt from IRB review) for probands or families who could not be contacted was used, and research conformed to the Helsinki Declaration. Probands or their families were contacted directly and consented by the University of Washington Autism Center under the TIGER study (IRB#: STUDY00000813), which includes consent for age and sex to be published.

Competing interests

The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing conducted at Baylor Genetics Laboratory. E.T. is an employee of GeneDx. The remaining authors declare no competing interests.

Author details

¹Department of Genome Sciences, University of Washington School of Medicine, 3720 15th Ave NE S413A, Box 355065, Seattle, WA 981095-5065, USA. ²Baylor Genetics Laboratories, Houston, TX, USA. ³Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX, USA. ⁴Center for Medical Genetics and Human Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China. ⁵Department of Anatomy, University of California, San Francisco, CA, USA. ⁶Department of Psychiatry, University of California, San Francisco, CA, USA. ⁷Weill Institute for Neurosciences, University of California at San Francisco, San Francisco, CA, USA. ⁸The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, CA, USA. ⁹Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands. ¹⁰Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden. ¹¹Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden. ¹²School of Medicine and the Robinson Research Institute, the University of Adelaide at the Women's and Children's Hospital, Adelaide, South Australia, Australia. ¹³Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia. ¹⁴South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia. ¹⁵Laboratorio di Genetica Medica - ASST Papa Giovanni XXIII, Bergamo, Italy. ¹⁶Department of Pediatrics, ASST Papa Giovanni XXIII, Bergamo, Italy. ¹⁷Genetic Unit, Department of Pediatrics, Fondazione MBM S. Gerardo Hospital, Monza, Italy. ¹⁸Unit of Neuromuscular and Neurodegenerative

Disorders, Department Neurosciences, Bambino Gesù Children's Hospital, IRCCS, 00146 Rome, Italy. ¹⁹Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany. ²⁰Center on Human Development and Disability, University of Washington, Seattle, WA, USA. ²¹Seattle Children's Autism Center, Seattle, WA, USA. ²²GeneDX, Gaithersburg, MD, USA. ²³Department of Pediatrics-Neurology, Baylor College of Medicine, Houston, TX, USA. ²⁴Texas Children's Hospital, Houston, TX, USA. ²⁵Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. ²⁶The Atwal Clinic: Genomic & Personalized Medicine, Jacksonville, FL, USA. ²⁷Department of Pediatrics, Section of Genetics, University of Illinois at Chicago, Chicago, IL, USA. ²⁸Department of Pediatrics, University of California San Diego, San Diego, CA, USA. ²⁹Genetics/Dysmorphology, Rady Children's Hospital San Diego, San Diego, CA, USA. ³⁰Rady Children's Institute for Genomic Medicine, San Diego, CA, USA. ³¹Department of Neurosciences, University of California San Diego, San Diego, CA, USA. ³²Department of Biomolecular Medicine, Ghent University Hospital, Ghent, Belgium. ³³Clinique de Génétique, Hôpital Jeanne de Flandre, Bâtiment Modulaire, CHU, 59037 Lille Cedex, France. ³⁴Sorbonne Universités, Centre de Référence déficiences intellectuelles de causes rares, département de génétique et embryologie médicale, Hôpital Trousseau, AP-HP, Paris, France. ³⁵Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN, USA. ³⁶Department of Medical and Molecular Genetics, IU Health, Indianapolis, IN, USA. ³⁷Department of Genetics, Hadassah, Hebrew University Medical Center, Jerusalem, Israel. ³⁸Pediatric Neurology, Assuta-Ashdod University Hospital, Ashdod, Israel. ³⁹Health Sciences, Ben-Gurion University of the Negev, Beersheba, Israel. ⁴⁰Department of Clinical Genetics, Odense University Hospital, Odense, Denmark. ⁴¹H C Andersen Childrens Hospital, Odense University Hospital, Odense, Denmark. ⁴²Gillette Children's Specialty Healthcare, Saint Paul, MN, USA. ⁴³Division of Clinical Genetics, Children's Mercy Kansas City, Kansas City, MO, USA. ⁴⁴The University of Missouri-Kansas City, School of Medicine, Kansas City, MO, USA. ⁴⁵Children's Mercy Kansas City, Center for Pediatric Genomic Medicine, Kansas City, MO, USA. ⁴⁶Oregon Health & Science University, Corvallis, OR, USA. ⁴⁷Department of Genetics, Hôpital Pitié-Salpêtrière, Paris, France. ⁴⁸Department of Medical Genetics, University of Antwerp, Antwerp, Belgium. ⁴⁹Genetics Department, Cook Children's Hospital, Fort Worth, TX, USA. ⁵⁰Department of Medical Genetics, University of British Columbia, Vancouver, Canada. ⁵¹BC Children's Hospital and BC Women's Hospital, Vancouver, BC, Canada. ⁵²Division of Developmental Pediatrics, Department of Pediatrics, BC Children's Hospital, University of British Columbia, Vancouver, BC, Canada. ⁵³Sunny Hill Health Centre for Children, Vancouver, BC, Canada. ⁵⁴Department of Pediatrics, The Children's Hospital at Montefiore, Bronx, NY, USA. ⁵⁵Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ⁵⁶Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, USA. ⁵⁷UPMC Children's Hospital of Pittsburgh, Pittsburgh, PA, USA. ⁵⁸Kaiser Permanente Southern California, Los Angeles, CA, USA. ⁵⁹The Permanente Medical Group, Oakland, CA, USA. ⁶⁰Northern Ireland Regional Genetics Service, Belfast City Hospital, Belfast, UK. ⁶¹Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA, USA. ⁶²Department of Pediatrics, University of Washington, Seattle, WA, USA. ⁶³Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. ⁶⁴AI Dupont Hospital for Children, Wilmington, DE, USA. ⁶⁵NYU Grossman School of Medicine, Department of Pediatrics, Clinical Genetic Services, New York, NY, USA. ⁶⁶Center for Human Genetics, KU Leuven and Leuven Autism Research (LAuRes), Leuven, Belgium. ⁶⁷Oasi Research Institute-IRCCS, Troina, Italy. ⁶⁸Department of Genetics, University Medical Center, Utrecht University, Utrecht, The Netherlands. ⁶⁹Rare Disease Institute, Children's National Health System, Washington, DC, USA. ⁷⁰Department of Genetics, Driscoll Children's Hospital, Corpus Christi, TX, USA. ⁷¹Department of Pediatric Gastroenterology, Driscoll Children's Hospital, Corpus Christi, TX, USA. ⁷²Àrea de Genètica Clínica i Molecular, Hospital Vall d'Hebrón, Barcelona, Spain. ⁷³Centre de référence Anomalies du développement, CHU Grenoble-Alpes, Grenoble, France. ⁷⁴UF Innovation en Diagnostic Génomique des Maladies Rares, FHU-TRANSLAD, CHU Dijon Bourgogne and INSERM UMR1231 GAD, Université de Bourgogne Franche-Comté, F-21000 Dijon, France. ⁷⁵INSERM UMR 1231 Génétique des Anomalies du Développement, Université Bourgogne Franche-Comté, Dijon, France. ⁷⁶Centre de Référence Maladies Rares « déficience intellectuelle », Centre de Génétique, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon, France. ⁷⁷Centre de Référence Maladies Rares « Anomalies du Développement et Syndromes malformatifs » Université Bourgogne Franche-Comté, Dijon, France. ⁷⁸Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA. ⁷⁹Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA.

Received: 4 September 2020 Accepted: 16 March 2021
Published online: 19 April 2021

References

- Coe BP, Stessman HAF, Sulovari A, Geisheker MR, Bakken TE, Lake AM, et al. Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nat Genet.* 2019;51(1):106–16.
- Khalil B, Morderer D, Price PL, Liu F, Rossoll W. mRNP assembly, axonal transport, and local translation in neurodegenerative diseases. *Brain Res.* 2018;1693(Pt A):75–91.
- Purice MD, Taylor JP. Linking hnRNP function to ALS and FTD pathology. *Front Neurosci.* 2018;12.
- Fu X-D, Ares M. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet.* 2014;15(10):689–701.
- Dickerson JE, Robertson DL. On the origins of Mendelian disease genes in man: the impact of gene duplication. *Mol Biol Evol.* 2012;29(1):61–9.
- Chen W-H, Zhao X-M, van Noort V, Bork P. Human monogenic disease genes have frequently functionally redundant paralogs. *PLoS Comput Biol.* 2013;16:9(5).
- Lal D, May P, Perez-Palma E, Samocha KE, Kosmicki JA, Robinson EB, et al. Gene family information facilitates variant interpretation and identification of disease-associated genes in neurodevelopmental disorders. *Genome Med.* 2020;12(1):1–12.
- Kim HJ, Kim NC, Wang Y-D, Scarborough EA, Moore J, Diaz Z, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature.* 2013;495(7442):467–73.
- Le Ber I, Van Bortel I, Nicolas G, Bouya-Ahmed K, Camuzat A, Wallon D, et al. hnRNPA2B1 and hnRNPA1 mutations are rare in patients with “multisystem proteinopathy” and frontotemporal lobar degeneration phenotypes. *Neurobiol Aging.* 2014;35(4):934.e5–934.e6.
- Bhoj E, Halbach S, McDonald-McGinn D, Tan C, Lande R, Waggoner D, et al. Expanding the spectrum of microdeletion 4q21 syndrome: a partial phenotype with incomplete deletion of the minimal critical region and a new association with cleft palate and Pierre Robin sequence. *Am J Med Genet A.* 2013;161(9):2327–33.
- Reichert SC, Li R, Turner SA, van Jaarsveld RH, Massink MPG, van den Boogaard M-JH, et al. HNRNPH1-related syndromic intellectual disability: seven additional cases suggestive of a distinct syndromic neurodevelopmental syndrome. *Clin Genet.* 2020;98(1):91–8.
- Bain JM, Cho MT, Telegrafi A, Wilson A, Brooks S, Botti C, et al. Variants in HNRNPH2 on the X chromosome are associated with a neurodevelopmental disorder in females. *Am J Hum Genet.* 2016;99(3):728–34.
- Au PYB, Goedhart C, Ferguson M, Breckpot J, Devriendt K, Wierenga K, et al. Phenotypic spectrum of Au-Kline syndrome: a report of six new cases and review of the literature. *Eur J Hum Genet EHG.* 2018;26(9):1272–81.
- Okamoto N. Okamoto syndrome has features overlapping with Au-Kline syndrome and is caused by HNRNPK mutation. *Am J Med Genet A.* 2019; 179(5):822–6.
- Dentici ML, Barresi S, Niceta M, Pantaleoni F, Pizzi S, Dallapiccola B, et al. Clinical spectrum of Kabuki-like syndrome caused by HNRNPK haploinsufficiency. *Clin Genet.* 2018;93(2):401–7.
- Lange L, Pagnamenta AT, Lise S, Clasper S, Stewart H, Akha ES, et al. A de novo frameshift in HNRNPK causing a Kabuki-like syndrome with nodular heterotopia. *Clin Genet.* 2016;90(3):258–62.
- Miyake N, Inaba M, Mizuno S, Shiina M, Imagawa E, Miyatake S, Nakashima M, Mizuguchi T, Takata A, Ogata K, Matsumoto N A case of atypical kabuki syndrome arising from a novel missense variant in HNRNPK. *Clin Genet.* 2017;92(5):554–5.
- Duijkers FA, McDonald A, Janssens GE, Lezzerini M, Jongejan A, van Koningsbruggen S, et al. HNRNPR variants that impair Homeobox gene expression drive developmental disorders in humans. *Am J Hum Genet.* 2019;104(6):1040–59.
- Engwerda A, Frenzt B, den Ouden AL, Flapper BC, Swertz MA, Gerkes EH, et al. The phenotypic spectrum of proximal 6q deletions based on a large cohort derived from social media and literature reports. *Eur J Hum Genet.* 2018;26(10):1478–89.
- Depienne C, Nava C, Keren B, Heide S, Rastetter A, Passerard S, et al. Genetic and phenotypic dissection of 1q43q44 microdeletion syndrome and neurodevelopmental phenotypes associated with mutations in ZBTB18 and HNRNPU. *Hum Genet.* 2017;136(4):463–79.
- Yates TM, Vasudevan PC, Chandler KE, Donnelly DE, Stark Z, Sadedin S, Willoughby J, Broad Center for Mendelian Genomics, DDD study,

- Balasubramanian M. De novo mutations in HNRNPU result in a neurodevelopmental syndrome. *Am J Med Genet A*. 2017;173(11):3003–12.
22. Bramswig NC, Lüdecke H-J, Hamdan FF, Altmüller J, Beleggia F, Elcioglu NH, et al. Heterozygous HNRNPU variants cause early onset epilepsy and severe intellectual disability. *Hum Genet*. 2017;136(7):821–34.
 23. Leduc MS, Chao H-T, Qu C, Walkiewicz M, Xiao R, Magoulas P, et al. Clinical and molecular characterization of de novo loss of function variants in HNRNPU. *Am J Med Genet A*. 2017;173(10):2680–9.
 24. Turner TN, Yi Q, Krumm N, Huddleston J, Hoekzema K, F Stessman HA, et al. denovo-db: a compendium of human de novo variants. *Nucleic Acids Res*. 2017;45(D1):D804–11.
 25. Harmsen S, Buchert R, Mayatepek E, Haack TB, Distelmaier F. Bain type of X-linked syndromic mental retardation in boys. *Clin Genet*. 2019;95(6):734–5.
 26. Jepsen WM, Ramsey K, Szelinger S, Llaci L, Balak C, Belnap N, Bilagody C, de Both M, Gupta R, Naymik M, Pandey R, Piras IS, Sanchez-Castillo M, Rangasamy S, Narayanan V, Huentelman MJ. Two additional males with X-linked, syndromic mental retardation carry de novo mutations in HNRNPH2. *Clin Genet*. 2019;96(2):183–5.
 27. Pilch J, Koppolu AA, Walczak A, Murcia Pienkowski VA, Biernacka A, Skiba P, et al. Evidence for HNRNPH1 being another gene for Bain type syndromic mental retardation. *Clin Genet*. 2018;94(3–4):381–5.
 28. Somashekar PH, Narayanan DL, Jagadeesh S, Suresh B, Vaishnavi RD, Bielas S, et al. Bain type of X-linked syndromic mental retardation in a male with a pathogenic variant in HNRNPH2. *Am J Med Genet A*. 2020;182(1):183–8.
 29. denovo-db. Available from: <http://denovo-db.gs.washington.edu>
 30. Durkin A, Albaba S, Fry AE, Morton JE, Douglas A, Beleza A, Williams D, Volker-Touw CML, Lynch SA, Canham N, Clowes V, Straub V, Lachlan K, Gibbon F, el Gamal M, Varghese V, Parker MJ, Newbury-Ecob R, Turmpenny PD, Gardham A, Ghali N, Balasubramanian M. Clinical findings of 21 previously unreported probands with HNRNPU-related syndrome and comprehensive literature review. *Am J Med Genet A*. 2020;182(7):1637–54.
 31. Caliebe A, Kroes HY, van der Smagt JJ, Martin-Subero JI, Tönnies H, Van 't Slot R, et al. Four patients with speech delay, seizures and variable corpus callosum thickness sharing a 0.440 Mb deletion in region 1q44 containing the HNRNPU gene. *Eur J Med Genet*. 2010;53(4):179–85.
 32. de Kovel CGF, Brilstra EH, van Kempen MJA, Van't Slot R, Nijman IJ, Afawi Z, et al. Targeted sequencing of 351 candidate genes for epileptic encephalopathy in a large cohort of patients. *Mol Genet Genomic Med*. 2016;4(5):568–80.
 33. Demos M, Guella I, DeGuzman C, McKenzie MB, Buerki SE, Evans DM, et al. Diagnostic yield and treatment impact of targeted exome sequencing in early-onset epilepsy. *Front Neurol*. 2019;10:434.
 34. Stessman HAF, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M, Kvarnung M, Gerdtz J, Trinh S, Cosemans N, Vives L, Lin J, Turner TN, Santen G, Ruivenkamp C, Kriek M, van Haeringen A, Aten E, Friend K, Liebelt J, Barnett C, Haan E, Shaw M, Gez C, Anderlid BM, Nordgren A, Lindstrand A, Schwartz C, Kooy RF, Vandeweyer G, Helsmoortel C, Romano C, Alberti A, Vinci M, Avola E, Giusto S, Courchesne E, Pramparo T, Pierce K, Nalabolu S, Amaral DG, Scheffer IE, Delatycki MB, Lockhart PJ, Hormozdiari F, Harich B, Castells-Nobau A, Xia K, Peeters H, Nordenskjöld M, Schenck A, Bernier RA, Eichler EE. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat Genet*. 2017;49(4):515–26.
 35. De Rubeis S, He X, Goldberg AP, Poultnery CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*. 2014;515(7526):209–15.
 36. Yuen RK, Merico D, Bookman M, L Howe J, Thiruvahindrapuram B, Patel RV, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci* 2017;20(4):602–611.
 37. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, Raja A, Coe BP, Stessman HA, He ZX, Leal SM, Bernier R, Eichler EE. Excess of rare, inherited truncating mutations in autism. *Nat Genet*. 2015;47(6):582–8.
 38. Lelieveld SH, Reijnders MRF, Pfundt R, Yntema HG, Kamsteeg E-J, de Vries P, et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat Neurosci*. 2016;19(9):1194–6.
 39. Epi4K Consortium, Epilepsy Phenome/Genome Project, Allen AS, Berkovic SF, Cossette P, Delanty N, et al. De novo mutations in epileptic encephalopathies. *Nature*. 2013;501(7466):217–21.
 40. Hinokuma N, Nakashima M, Asai H, Nakamura K, Akaboshi S, Fukuoka M, et al. Clinical and genetic characteristics of patients with Doose syndrome. *Epilepsia Open*.
 41. Helbig KL, Farwell Hagman KD, Shinde DN, Mroske C, Powis Z, Li S, et al. Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. *Genet Med Off J Am Coll Med Genet*. 2016;18(9):898–905.
 42. Feliciano P, Zhou X, Astrovskaia I, Turner TN, Wang T, Brueggeman L, et al. Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes. *NPJ Genomic Med*. 2019;4(1):19.
 43. Yamada M, Shiraishi Y, Uehara T, Suzuki H, Takenouchi T, Abe-Hatano C, et al. Diagnostic utility of integrated analysis of exome and transcriptome: successful diagnosis of Au-Kline syndrome in a patient with submucous cleft palate, scaphocephaly, and intellectual disabilities. *Mol Genet Genomic Med*. 2020;8(9):e1364.
 44. Peron A, Novara F, Briola FL, Merati E, Giannusa E, Segalini E, et al. Missense variants in the Arg206 residue of HNRNPH2: further evidence of causality and expansion of the phenotype. *Am J Med Genet A*. 2020;182(4):823–8.
 45. Borlot F, de Almeida BI, Combe SL, Andrade DM, Filloux FM, Myers KA. Clinical utility of multigene panel testing in adults with epilepsy and intellectual disability. *Epilepsia*. 2019;60(8):1661–9.
 46. Carvill GL, Heavin SB, Yendle SC, McMahon JM, O'Roak BJ, Cook J, et al. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. *Nat Genet*. 2013;45(7):825–30.
 47. Chérot E, Keren B, Dubourg C, Carré W, Fradin M, Lavillaureix A, et al. Using medical exome sequencing to identify the causes of neurodevelopmental disorders: experience of 2 clinical units and 216 patients. *Clin Genet*. 2018;93(3):567–76.
 48. Coppola A, Cellini E, Stamberger H, Saarentaus E, Cetica V, Lal D, et al. Diagnostic implications of genetic copy number variation in epilepsy plus. *Epilepsia*. 2019;60(4):689–706.
 49. The Epilepsy Genetics Initiative. Systematic reanalysis of diagnostic exomes increases yield. *Epilepsia*. 2019;60(5):797–806.
 50. Hamdan FF, Srour M, Capo-Chichi J-M, Daoud H, Nassif C, Patry L, et al. De novo mutations in moderate or severe intellectual disability. *PLoS Genet*. 2014;10(10):e1004772.
 51. Johannessen KM, Nikanorova N, Marjanovic D, Pavbro A, Larsen LHG, Rubboli G, et al. Utility of genetic testing for therapeutic decision-making in adults with epilepsy. *Epilepsia*. 2020;61(6):1234–9.
 52. Mignot C, von Stülpnagel C, Nava C, Ville D, Sanlaville D, Lesca G, et al. Genetic and neurodevelopmental spectrum of SYNGAP1-associated intellectual disability and epilepsy. *J Med Genet*. 2016;53(8):511–22.
 53. Monroe GR, Frederix GW, Savelberg SMC, de Vries TI, Duran KJ, van der Smagt JJ, et al. Effectiveness of whole-exome sequencing and costs of the traditional diagnostic trajectory in children with intellectual disability. *Genet Med Off J Am Coll Med Genet*. 2016;18(9):949–56.
 54. Oates S, Tang S, Rosch R, Lear R, Hughes EF, Williams RE, et al. Incorporating epilepsy genetics into clinical practice: a 360° evaluation. *Npj Genomic Med*. 2018;3(1):1–11.
 55. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet*. 2014;94(5):677–94.
 56. Stamouli S, Anderlid B-M, Willfors C, Thiruvahindrapuram B, Wei J, Berggren S, et al. Copy number variation analysis of 100 twin pairs enriched for neurodevelopmental disorders. *Twin Res Hum Genet*. 2018;21(1):1–11.
 57. Zhu X, Petrovski S, Xie P, Ruzzo EK, Lu Y-F, McSweeney KM, et al. Whole-exome sequencing in undiagnosed genetic diseases: interpreting 119 trios. *Genet Med Off J Am Coll Med Genet*. 2015;17(10):774–81.
 58. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature*. 2017;542(7642):433–8.
 59. Maystadt I, Deprez M, Moortgat S, Benoit V, Karadurmus D. A second case of Okamoto syndrome caused by HNRNPK mutation. *Am J Med Genet A*. 2020;182(6):1537–9.
 60. Shimada S, Oguni H, Otani Y, Nishikawa A, Ito S, Eto K, Nakazawa T, Yamamoto-Shimajima K, Takanashi JI, Nagata S, Yamamoto T. An episode of acute encephalopathy with biphasic seizures and late reduced diffusion followed by hemiplegia and intractable epilepsy observed in a patient with a novel frameshift mutation in HNRNPU. *Brain Dev*. 2018;40(9):813–8.
 61. Wang T, Hoekzema K, Vecchio D, Wu H, Sulovari A, Coe BP, et al. Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. *Nat Commun*. 2020;11(1):4932.
 62. Kaplanis J, Samocha KE, Wiel L, Zhang Z, Arvai KJ, Eberhardt RY, et al. Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature*. 2020;586(7831):757–62.

63. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An J-Y, et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*. 2020;180(3):568–84. e23
64. Iourov IY, Zelenova MA, Vorsanova SG, Voinova VV, Yurov YB. 4q21.2q21.3 duplication: molecular and neuropsychological aspects. *Curr Genomics*. 2018;19(3):173–8.
65. Lebedev IN, Nazarenko LP, Skryabin NA, Babushkina NP, Kashevarova AA. A de novo microtriplication at 4q21.21-q21.22 in a patient with a vascular malignant hemangioma, elongated sigmoid colon, developmental delay, and absence of speech. *Am J Med Genet A*. 2016;170(8):2089–96.
66. Zarrei M, Merico D, Kellam B, Engchuan W, Scriver T, Jokhan R, Wilson MD, Parr J, Lemire EG, Stavropoulos DJ, Scherer SW. A de novo deletion in a boy with cerebral palsy suggests a refined critical region for the 4q21.22 microdeletion syndrome. *Am J Med Genet A*. 2017;173(5):1287–93.
67. Mohamed AM, El-Bassyouni HT, El-Gerzawy AM, Hammad SA, Helmy NA, Kamel AK, et al. Cytogenomic characterization of 1q43q44 deletion associated with 4q32.1q35.2 duplication and phenotype correlation. *Mol Cytogenet*. 2018;11:57.
68. Stessman HA, Bernier R, Eichler EE. A genotype-first approach to defining the subtypes of a complex disease. *Cell*. 2014;156(5):872–7.
69. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014;312(18):1870–9.
70. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *ArXiv:12073907 [q-bio.GN]* 2012. <https://arxiv.org/abs/1207.3907>.
71. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
72. Danecek P, McCarthy SA. BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics*. 2017;33(13):2037–9.
73. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat*. 2015;36(10):928–30.
74. NHGRI/NHLBI University of Washington-Center for Mendelian Genomics (UW-CMG), Seattle, WA. MyGene2. Available from: <http://www.mygene2.org>
75. Abramovs N, Brass A, Tassabehji M. GeVIR is a continuous gene-level metric that uses variant distribution patterns to prioritize disease candidate genes. *Nat Genet*. 2020;52(1):35–9.
76. O’Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*. 2012;485(7397):246–50.
77. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet*. 2014;46(9):944–50.
78. Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C. MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. *Hum Mutat*. 2019;40(8):1030–8.
79. Meyer MJ, Beltrán JF, Liang S, Fragoza R, Rumack A, Liang J, et al. Interactome INSIDER: a structural interactome browser for genomic studies. *Nat Methods*. 2018;15(2):107–14.
80. Turner TN, Douville C, Kim D, Stenson PD, Cooper DN, Chakravarti A, Karchin R. Proteins linked to autosomal dominant and autosomal recessive disorders harbor characteristic rare missense mutation distribution patterns. *Hum Mol Genet*. 2015;24(21):5995–6002.
81. Coban-Akdemir Z, White JJ, Song X, Jhangiani SN, Fatih JM, Gambin T, et al. Identifying genes whose mutant transcripts cause dominant disease traits by potential gain-of-function alleles. *Am J Hum Genet*. 2018;103(2):171–87.
82. Shamsani J, Kazakoff SH, Armean IM, McLaren W, Parsons MT, Thompson BA, O’Mara TA, Hunt SE, Waddell N, Spurdle AB. A plugin for the Ensembl variant effect predictor that uses MaxEntScan to predict variant spliceogenicity. *Bioinformatics*. 2019;35(13):2315–7.
83. Geuens T, Bouhy D, Timmerman V. The hnRNP family: insights into their role in health and disease. *Hum Genet*. 2016;135(8):851–67.
84. Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E, et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science*. 2017;358(6368):1318–23.
85. CSEA tool. Available from: <http://genetics.wustl.edu/jdlab/csea-tool-2/>
86. Babbi G, Martelli PL, Casadio R. PhenPath: a tool for characterizing biological functions underlying different phenotypes. *BMC Genomics*. 2019;20(8):548.
87. gnomAD. Available from: <https://gnomad.broadinstitute.org/>
88. Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, et al. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep*. 2012;1(2):167–78.
89. Cooper-Knock J, Higginbottom A, Stopford MJ, Highley JR, Ince PG, Wharton SB, Pickering-Brown S, Kirby J, Hautbergue GM, Shaw PJ. Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. *Acta Neuropathol (Berl)*. 2015;130(1):63–75.
90. Mohagheghi F, Prudencio M, Stuaní C, Cook C, Jansen-West K, Dickson DW, Petrucelli L, Buratti E. TDP-43 functions within a network of hnRNP proteins to inhibit the production of a truncated human SORT1 receptor. *Hum Mol Genet*. 2016;25(3):534–45.
91. Kashima T, Rao N, David CJ, Manley JL. hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. *Hum Mol Genet*. 2007;16(24):3149–59.
92. Berson A, Barbash S, Shaltiel G, Goll Y, Hanin G, Greenberg DS, et al. Cholinergic-associated loss of hnRNP-A/B in Alzheimer’s disease impairs cortical splicing and cognitive function in mice. *EMBO Mol Med*. 2012;4(8):730–42.
93. Borreca A, Gironi K, Amadoro G, Ammassari-Teule M. Opposite dysregulation of fragile-X mental retardation protein and heteronuclear ribonucleoprotein C protein associates with enhanced APP translation in Alzheimer disease. *Mol Neurobiol*. 2016;53(5):3227–34.
94. Sun Y, Chen H, Lu Y, Duo J, Lei L, OuYang Y, et al. Limb girdle muscular dystrophy D3 HNRNPDL related in a Chinese family with distal muscle weakness caused by a mutation in the prion-like domain. *J Neurol*. 2019;266(2):498–506.
95. Mori K, Lammich S, Mackenzie IRA, Forné I, Zilow S, Kretzschmar H, et al. hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathol (Berl)*. 2013;125(3):413–23.
96. Lee EK, Kim HH, Kuwano Y, Abdelmohsen K, Srikantan S, Subaran SS, et al. hnRNP C promotes APP translation by competing with FMRP for APP mRNA recruitment to P bodies. *Nat Struct Mol Biol*. 2010;17(6):732–9.
97. Moursy A, Allain FH-T, Cléry A. Characterization of the RNA recognition mode of hnRNP G extends its role in SMN2 splicing regulation. *Nucleic Acids Res*. 2014;42(10):6659–72.
98. Barbash S, Garfinkel BP, Maoz R, Simchovitz A, Nadorp B, Guffanti A, et al. Alzheimer’s brains show inter-related changes in RNA and lipid metabolism. *Neurobiol Dis*. 2017;106:1–13.
99. Cho S, Moon H, Loh TJ, Oh HK, Cho S, Choy HE, et al. hnRNP M facilitates exon 7 inclusion of SMN2 pre-mRNA in spinal muscular atrophy by targeting an enhancer on exon 7. *Biochim Biophys Acta*. 2014;1839(4):306–15.
100. Waibel S, Neumann M, Rabe M, Meyer T, Ludolph AC. Novel missense and truncating mutations in FUS/TLS in familial ALS. *Neurology*. 2010;75(9):815–7.
101. Vance C, Rogelj B, Hortobágyi T, De Vos KJ, Nishimura AL, Sreedharan J, et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science*. 2009;323(5918):1208–11.
102. Chen H-H, Chang J-G, Lu R-M, Peng T-Y, Tam W-Y. The RNA binding protein hnRNP Q modulates the utilization of exon 7 in the survival motor neuron 2 (SMN2) gene. *Mol Cell Biol*. 2008;28(22):6929–38.
103. Dombert B, Sivadasan R, Simon CM, Jablonka S, Sendtner M. Presynaptic localization of Smn and hnRNP R in axon terminals of embryonic and postnatal mouse motoneurons. *PLoS One*. 2014;9(10):e110846.
104. Guo H, Duyzend MH, Coe BP, Baker C, Hoekzema K, Gerdtts J, et al. Genome sequencing identifies multiple deleterious variants in autism patients with more severe phenotypes. *Genet Med*. 2019;21(7):1611–1620. A.
105. Iossifov I, O’Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*. 2014;515(7526):216–21.
106. Rauch A, Wieczorek D, Graf E, Wieland T, Ende S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet Lond Engl*. 2012;380(9854):1674–82.
107. Ernst C. Proliferation and differentiation deficits are a major convergence point for neurodevelopmental disorders. *Trends Neurosci*. 2016 [cited 2020 Feb 12];39(5):290–9.
108. Donovan APA, Basson MA. The neuroanatomy of autism – a developmental perspective. *J Anat*. 2017;230(1):4–15.
109. Guarnieri FC, de Chevigny A, Falace A, Cardoso C. Disorders of neurogenesis and cortical development. *Dialogues Clin Neurosci*. 2018;20(4):255–66.

110. Wang Y-C, Chang K-C, Lin B-W, Lee J-C, Lai C-H, Lin L-J, et al. The EGF/hnRNP Q1 axis is involved in tumorigenesis via the regulation of cell cycle-related genes. *Exp Mol Med*. 2018;50(6):70.
111. Chun Y, Kim R, Lee S. Centromere protein (CENP)-W interacts with heterogeneous nuclear ribonucleoprotein (hnRNP) U and may contribute to kinetochore-microtubule attachment in mitotic cells. *PLoS One*. 2016;11(2):e0149127.
112. Douglas P, Ye R, Morrice N, Britton S, Trinkle-Mulcahy L, Lees-Miller SP. Phosphorylation of SAF-A/hnRNP-U serine 59 by Polo-like kinase 1 is required for mitosis. *Mol Cell Biol*. 2015;35(15):2699–713.
113. Sugimasa H, Taniue K, Kurimoto A, Takeda Y, Kawasaki Y, Akiyama T. Heterogeneous nuclear ribonucleoprotein K upregulates the kinetochore complex component NUF2 and promotes the tumorigenicity of colon cancer cells. *Biochem Biophys Res Commun*. 2015;459(1):29–35.
114. Al-Khalaf HH, Colak D, Al-Saif M, Al-Bakheet A, Hendrayani S-F, Al-Yousef N, et al. p16INK4A positively regulates Cyclin D1 and E2F1 through negative control of AUF1. *PLoS One*. 2011;6(7):e21111.
115. Schuetze M, Park MTM, Cho IY, MacMaster FP, Chakravarty MM, Bray SL. Morphological alterations in the thalamus, striatum, and pallidum in autism spectrum disorder. *Neuropsychopharmacology*. 2016;41(11):2627–37.
116. Peixoto RT, Chantranupong L, Hakim R, Lévassieur J, Wang W, Merchant T, et al. Abnormal striatal development underlies the early onset of behavioral deficits in Shank3B^{-/-} mice. *Cell Rep*. 2019;29(7):2016–27. e4
117. Schumann CM, Bauman MD, Amaral DG. Abnormal structure or function of the amygdala is a common component of neurodevelopmental disorders. *Neuropsychologia*. 2011;49(4):745–59.
118. Pérez-Palma E, May P, Iqbal S, Niestroj L-M, Du J, Heyne HO, et al. Identification of pathogenic variant enriched regions across genes and gene families. *Genome Res*. 2020;30(1):62–71.
119. Dusen CMV, Yee L, McNally LM, McNally MT. A glycine-rich domain of hnRNP H/F promotes nucleocytoplasmic shuttling and nuclear import through an interaction with transportin 1. *Mol Cell Biol*. 2010;30(10):2552–62.
120. Choi J-H, Kim S-H, Jeong Y-H, Kim SW, Min K-T, Kim K-T. hnRNP Q regulates internal ribosome entry site-mediated *fmr1* translation in neurons. *Mol Cell Biol*. 2019;39(4):e00371–18.
121. Tratnjek L, Živin M, Glavan G. Synaptotagmin 7 and SYNCRIP proteins are ubiquitously expressed in the rat brain and co-localize in Purkinje neurons. *J Chem Neuroanat*. 2017;79:12–21.
122. Williams KR, McAninch DS, Stefanovic S, Xing L, Allen M, Li W, Feng Y, Mihailescu MR, Bassell GJ. hnRNP-Q1 represses nascent axon growth in cortical neurons by inhibiting Gap-43 mRNA translation. *Mol Biol Cell*. 2015;27(3):518–34.
123. McDermott SM, Yang L, Halstead JM, Hamilton RS, Meignin C, Davis I. *Drosophila* Syncrip modulates the expression of mRNAs encoding key synaptic proteins required for morphology at the neuromuscular junction. *RNA N Y N*. 2014;20(10):1593–606.
124. Chen H-H, Yu H-H, Chiang W-C, Lin Y-D, Shia B-C, Tarn W-Y. hnRNP Q regulates Cdc42-mediated neuronal morphogenesis. *Mol Cell Biol*. 2012;32(12):2224–38.
125. Duning K, Buck F, Barnekow A, Kremerskothen J. SYNCRIP, a component of dendritically localized mRNPs, binds to the translation regulator BC200 RNA. *J Neurochem*. 2008;105(2):351–9.
126. Schneider T, Ziolkowska B, Gienyk A, Tyminska A, Przewlocki R. Prenatal exposure to valproic acid disturbs the enkephalinergic system functioning, basal hedonic tone, and emotional responses in an animal model of autism. *Psychopharmacology*. 2007;193(4):547–55.
127. Dobi A, Szemes M, Lee C, Palkovits M, Lim F, Gyorgy A, et al. AUF1 is expressed in the developing brain, binds to AT-rich double-stranded DNA, and regulates enkephalin gene expression. *J Biol Chem*. 2006;281(39):28889–900.
128. Leirer DJ, Iyegbe CO, Di Forti M, Patel H, Carra E, Fraietta S, et al. Differential gene expression analysis in blood of first episode psychosis patients. *Schizophr Res*. 2019;209:88–97.
129. Servadio M, Manduca A, Melancia F, Leboffe L, Schiavi S, Campolongo P, Palmery M, Ascenzi P, di Masi A, Trezza V. Impaired repair of DNA damage is associated with autistic-like traits in rats prenatally exposed to valproic acid. *Eur Neuropsychopharmacol J Eur Coll Neuropsychopharmacol*. 2018;28(1):85–96.
130. Hong Z, Jiang J, Ma J, Dai S, Xu T, Li H, Yasui A. The role of hnRNPUL1 involved in DNA damage response is related to PARP1. *PLoS One*. 2013;8(4):e60208.
131. Polo SE, Blackford AN, Chapman JR, Baskcomb L, Gravel S, Rusch A, Thomas A, Blundred R, Smith P, Kzhyshkowska J, Dobner T, Taylor AMR, Turnell AS, Stewart GS, Grand RJ, Jackson SP. Regulation of DNA-end resection by hnRNP-like proteins promotes DNA double-strand break signaling and repair. *Mol Cell*. 2012;45(4):505–16.
132. Markkanen E, Meyer U, Dianov GL. DNA damage and repair in schizophrenia and autism: implications for cancer comorbidity and beyond. *Int J Mol Sci*. 2016;17(6).
133. Lampasona AA, Czaplinski K. Hnrnpab regulates neural cell motility through transcription of *Eps8*. *RNA N Y N*. 2019;25(1):45–59.
134. Sinnamon JR, Waddell CB, Nik S, Chen EI, Czaplinski K. Hnrnpab regulates neural development and neuron cell survival after glutamate stimulation. *RNA N Y N*. 2012;18(4):704–19.
135. White R, Gonsior C, Bauer NM, Krämer-Albers E-M, Luhmann HJ, Trotter J. Heterogeneous nuclear ribonucleoprotein (hnRNP) F is a novel component of oligodendroglial RNA transport granules contributing to regulation of myelin basic protein (MBP) synthesis. *J Biol Chem*. 2012;287(3):1742–54.
136. Chou M-Y, Rooke N, Turck CW, Black DL. hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol Cell Biol*. 1999;19(1):69–77.
137. Latorre E, Torregrossa R, Wood ME, Whiteman M, Harries LW. Mitochondria-targeted hydrogen sulfide attenuates endothelial senescence by selective induction of splicing factors HNRNP D and SRSF2. *Aging*. 2018;10(7):1666–81.
138. GenBank Overview. Available from: <https://www.ncbi.nlm.nih.gov/genbank/>

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

