Genes with high penetrance for syndromic and non-syndromic autism typically function within the nucleus and regulate gene expression

Emily L. Casanova¹²*, Julia L. Sharp³, Hrishikesh Chakraborty⁴, Nahid Sultana Sumi⁴ and Manuel F. Casanova¹²

Abstract

Background: Intellectual disability (ID), autism, and epilepsy share frequent yet variable comorbidities with one another. In order to better understand potential genetic divergence underlying this variable risk, we studied genes responsible for monogenic IDs, grouped according to their autism and epilepsy comorbidities.

Methods: Utilizing 465 different forms of ID with known molecular origins, we accessed available genetic databases in conjunction with gene ontology (GO) to determine whether the genetics underlying ID diverge according to its comorbidities with autism and epilepsy and if genes highly penetrant for autism or epilepsy share distinctive features that set them apart from genes that confer comparatively variable or no apparent risk.

Results: The genetics of ID with autism are relatively enriched in terms associated with nervous system-specific processes and structural morphogenesis. In contrast, we find that ID with highly comorbid epilepsy (HCE) is modestly associated with lipid metabolic processes while ID without autism or epilepsy comorbidity (ID only) is enriched at the Golgi membrane. Highly comorbid autism (HCA) genes, on the other hand, are strongly enriched within the nucleus, are typically involved in regulation of gene expression, and, along with IDs with more variable autism, share strong ties with a core protein-protein interaction (PPI) network integral to basic patterning of the CNS.

Conclusions: According to GO terminology, autism-related gene products are integral to neural development. While it is difficult to draw firm conclusions regarding IDs unassociated with autism, it is clear that the majority of HCA genes are tightly linked with general dysregulation of gene expression, suggesting that disturbances to the chronology of neural maturation and patterning may be key in conferring susceptibility to autism spectrum conditions.

Keywords: Mental retardation, Epilepsy, Epigenomics, Body patterning, Regulation of gene expression, Chromatin assembly and disassembly

Background

Intellectual disability (ID), epilepsy, and autism are highly comorbid with one another, suggesting shared etiologies in at least some forms of these conditions. In both autism and ID, epilepsy occurs in approximately one-third of cases, respectively [1, 2]. In those individuals with epilepsy eligible for surgery who have seizure onset prior to 24 months of age, approximately 46 % also have comorbid ID [3]. Meanwhile, about 31 % of autistic children aged 8 have IQs within the ID range, and an additional 23 % fall within the borderline region [4]. These high comorbidity rates stress an etiological commonality among some forms of the conditions, though it leaves unanswered the question of which forms are more susceptible to co-occurrence.

In recent years, there has been increased interest in both the genetic and phenotypic overlap of ID, epilepsy, and autism. Popular foci of study include monogenic syndromes such as tuberous sclerosis (TSC), fragile X...
(FXS), and Angelman syndromes (AS), whose respective gene mutations can lead to disturbed neurogenesis and various perturbations in neuroblast and neuronal maturation. This can be inferred by the different yet often overlapping malformations of cortical development (MCD) found in these syndromes. TSC, for instance, is defined in part by the characteristic tubers for which the condition is so named, a form of multifocal cortical and subcortical dysgenesis [5]. In FXS, features of macrocephaly and abnormalities of neuronal migration are also sometimes noted [6, 7]. In addition, mouse models of the syndrome have revealed alterations in neurogenesis and early neuroblast differentiation, particularly affecting the glutamatergic population [8]. Secondary microcephaly is likewise a feature in the majority of those with AS, potentially resulting from disturbances in early neuroblast maturation and subsequent downstream effects on neuronal differentiation and overall cortical structure [9, 10]. In short, commonalities exist between these conditions not necessarily in the precise malformations encountered but in the general presence of MCD, though they may sometimes require microscopic investigation in order to identify. And in fact, numerous types of MCD are commonly found across many forms of ID, epilepsy, and autism, suggesting that these malformations may be indicative of similar physiologies, e.g., excitatory-inhibitory imbalance [11–14].

Because these three conditions overlap so frequently and share phenotypic features such as MCD, we questioned whether the genetics of different forms of ID might segregate according to their comorbidities with either autism or epilepsy, indicating differences in their etiological underpinnings. In particular, we find that ID with high rates of autism comorbidity present with a particularly homogenous genetic profile that has been previously unreported.

**Methods**

**Curation**

A comprehensive list of forms of ID were accessed from the Mendelian Inheritance in Man (MIM) database [15]. Only conditions with known molecular basis were collected. Keywords for initial search comprised “intellectual disability”, “mental retardation”, “mentally retarded”, “global developmental delay”, “severe developmental delay”, and “profound developmental delay” (for a full listing of OMIM numbers, gene/locus numbers, and associated data, see Additional file 1). Conditions were then curated and removed if: (1) the ID was not a primary feature but was variably expressed; (2) the ID displayed onset later than three years of age; (3) the condition tended to be lethal in infancy or early childhood; (4) the condition had a known yet complex genetic etiology, e.g., large recombination events that include two or more genes (with the exception of chromosome 2p16.3 deletion syndrome which has been directly linked to NRXN1 mutations); (5) autism was a defining symptom for diagnosis, as in the cases of certain “susceptibility” genes; (6) only one or two cases were noted in the literature; (7) mutations occurred in only a single family; (8) the condition was a chromosome instability syndrome, leading to variable features due to the accumulation of different mutations; or (9) a condition contained an unconfirmed or potentially spurious mapping as indicated by a “?” before the disease name. This led to a final list of 465 different forms of ID and 434 unique genes (some genes whose functions are unknown were also removed from analyses although are still included in the main list. In addition, a small selection of genes was not recognized by gene ontology and therefore was not included within those analyses. Therefore, gene lists sometimes varied minimally between analyses).

Associated genes were assigned to one of five categories according to the information available regarding their comorbidities with autism and epilepsy. This information was initially derived from the MIM database and was subsequently confirmed through thorough literature review. Genes were additionally cross-referenced with Pinto et al. [16], whose supplemental material includes a large list of conditions associated with autism and ID. The categories ultimately were: (1) ID with highly comorbid autism (HCA) \((N = 72\) conditions, 71 genes), (2) ID with variable autism (VarAut) \((N = 139\) conditions, 124 genes), (3) ID with highly comorbid epilepsy (HCE) \((N = 88\) conditions, 86 genes), (4) ID with variable epilepsy (VarEp) \((N = 84\) conditions, 78 genes), and (5) ID without autism or epilepsy (ID only) \((N = 82\) conditions, 75 genes) (Fig. 1a).

If a gene was redundant across two or more IDs, those with occurrences of autism were preferentially given preference and placed within HCA or VarAut, followed by epilepsy (HCE, VarEp), and so forth. Conditions were assigned the above comorbidity groups if: (1) there was even minor evidence of overlapping comorbidity reported on MIM or in the broader literature; (2) if, in the case of autism, the associated gene was included under “spondromic”, “high confidence”, “strong candidate”, or “suggestive evidence” headings in the SFARI Gene database or rated 9+ within the AutismKB database; (3) there were no reported instances of epilepsy/seizures but abnormal epileptiform activity had been noted; or (4) the gene was included in the epilepsy gene database, CarpeDB [17–19]. In instances in which a single MIM condition had multiple associated genes, all genes were placed in the same category provided any of those were not already in a superseding category.

For the purposes of dividing all autism-associated IDs into either the HCA or VarAut groups, any condition that had ≥20 % autism comorbidity rating according to intensive literature review was placed within the HCA category;
meanwhile, all other IDs with at least two unrelated case studies involving autism or autistic symptomology were placed within VarAut (see Additional file 1 and Additional file 2 for comorbidity references). For the purposes of studying genetic penetrance, HCA and HCE conditions were annotated according to their inheritance patterns (dominant, recessive) and broken into their respective pattern groups for further analysis.

Epileptic comorbidity was rated upon whether epilepsy was reported in any cases within MIM or the larger literature and which were not apparently due to trauma or another identifiable illness; in conditions in which the reported $N$ was low for the entire ID ($\leq 3$) and only one or a few instances of epilepsy/seizure were noted, these disorders were removed from the analysis entirely. Similar to autism, if epilepsy occurred only within a single family, this condition was not included. For assessing frequency of epilepsy, conditions were included in the HCE group if seizures were listed as a common neurological feature for a given condition within the clinical synopses portion of the MIM database, and this was subsequently corroborated via literature search. The same percentage cut-off for inclusion within HCE (20%) was used as in HCA.

**Gene ontology**

In order to assess associated gene product functions, we analyzed differences in sample frequency for gene ontology (GO) terms according to group across categories of biological process, cellular component, and molecular function [20]. A listing of significantly enriched terms were initially accessed for each gene group, followed by the removal of redundant parent terms and those unusable for direct annotation. Absolute frequencies were then compared for each of these significant terms across all groups. The `prop.test()` function in the statistical computing software, R, was used for most statistical analyses. All pairs of proportions were compared using a Chi-square test of two proportions with one degree of freedom. A false discovery rate adjustment was applied to account for multiple comparisons.

The same approach was used to study GO term enrichment across HCA dominant x HCA recessive and HCE dominant x HCE recessive conditions. Meanwhile,
ratios of dominant:recessive in HCA vs. HCE were assessed using a two-tailed heteroscedastic *t* test.

For more intensive examination, UniProt/Swiss-Prot and Entrez Gene summaries of gene products’ molecular functions were used as the basis to determine whether a given gene was considered a nuclear epigenetic regulator (transcription factor/repressor, methylator, ubiquitinase, chromatin remodeler, etc.) [21, 22]. The same statistical analyses as used in the main GO experiment were also used here. For the purposes of studying rates of nuclear epigenetic regulators across dominant x recessive subgroups, a proportions comparison was used, without need for correction for multiple comparisons.

In addition, because a large minority of the HCA genes are not currently included or rated within the SFARI Gene Database, we compared the SFARI-only HCA group to the HCA non-SFARI group in terms of number of nuclear epigenetic regulators to ensure that no significant differences existed. This was performed using a two-tailed proportions comparison.

Finally, a thorough literature search was performed to determine which conditions within HCE and VarEp were considered neurodegenerative. HCE and VarEp rates of neurodegeneration were compared using a two-tailed proportions comparison.

### Protein-protein interaction networks

For the protein-protein interaction experiment, each gene group was loaded individually into String 10 alongside a selection of proteins representative of the core PPI network (WNT = CTNNB1, SHH = PTCHD1, NCOR = NCOR1, SWI/SNF = SMARCA1, NOTCH = NOTCH1, ERK1/2 = FGFR8, TGF-β/BMP = SMAD4), and confidence data were analyzed [23]. Both the percentage of experimental genes connected with the core PPI and the number of intermediary nodes that lay between the genes of interest and their nearest core PPI neighbors were assessed according to group. For the former analysis, a proportions test with correction for multiple comparisons was used; meanwhile, for the latter, an ANOVA was utilized.

### Results

#### Comorbidity data

We studied the comorbidities of a substantial list of IDs with known molecular origins derived from the MIM database. Of the 465 forms of IDs collected, we found that 45 % (N = 211) were comorbid with autism in at least a minority of cases. Meanwhile, 15 % (N = 72) were highly comorbid with autism, co-occurring in ≥20 % of reported cases. Some of these include conditions well known for autism association, such as FXS and TSC, however, also included conditions less well known, such as non-photosensitive trichothiodystrophy and Mowat-Wilson Syndrome. Meanwhile, 44 % (N = 204) of IDs were highly comorbid with epilepsy, while an additional 32 % (N = 151) exhibited variable rates of epilepsy. In fact, 55 % (N = 116) of autism-related conditions listed epilepsy as a common feature, which was consistent across both HCA and VarAut groups, reinforcing ideas of their shared etiologies (Table 1).

### Trends in gene product function

We went on to study the genetic etiologies of our conditions of interest, investigating GO term associations in the areas of biological process, molecular function, and cellular component (see Table 2 for gene list by category). Both autism groups, HCA and VarAut, exhibited enrichment in nervous system development compared to HCE and ID only (p = 0.0035–0.052, see Additional file 3 for full statistical results). However, despite differences in this overarching parent term, our comorbidity groups did not differ significantly in the child terms neurogenesis (p = 0.1218–0.8594), neuron differentiation (p = 0.3118–1.00), neuron projection development (p = 0.2608–1.00), and synaptic transmission (p = 0.5075–0.7988). HCA was mildly enriched in the regulation of synaptic structure or activity compared to HCE (p = 0.0477) and ID only (p = 0.0453), but not compared to the variable groups. In summary, both autism groups exhibit stronger nervous system enrichment than either HCE or ID only, suggesting that gene product involvement in nervous system development may characterize a significant subset of autism risk genes (Fig. 1b).

HCA was also particularly enriched in anatomical structure development compared to all groups (p = 0.0021–0.0419) except VarAut (p = 0.2084), indicating the genes’ probable roles in structural morphogenesis (Fig. 1b). However, above all else, HCA was typified by regulation of gene expression (p = 0.000–0.0404) and was involved in regulation of DNA-templated transcription (p = 0.000–0.0406) and chromatin binding (p = 0.0018–0.0326). Matching its functional enrichment, HCA was strongly enriched within the nucleus (p = 0.0002–0.0009) in contrast to all other groups and was also modestly enriched at the chromosome compared to HCE (p = 0.0392) and ID only (p = 0.0416) (Fig. 1c).

UniProt/Swiss-Prot and Entrez Gene analysis further revealed that HCA gene expression regulation was largely carried out through nuclear epigenetic means, such as transcription factors and repressors, methylation regulators, ubiquitin ligases, and other chromatin remodelers, which comprised over half of that gene group, a substantial increase compared to all other comorbidity groups (p = 0.000–0.0004) (Fig. 2a). In addition, 45 % of the genes in HCA are not currently included or rated within the SFARI database, yet even with their removal, the SFARI-only HCA group did not differ from those not included within
| Intellectual disability (OMIM #) | Gene symbol | SFARI rating | Autism comorbidity | References | Epilepsy frequency |
|---------------------------------|-------------|--------------|--------------------|------------|-------------------|
| Adenylosuccinase Deficiency (#103050) | ADSL | 5 | 38 % (N = 8) | Jaeken et al. 1988 [36] | Common |
| Alpha-thalassemia/Mental Retardation Syndrome (#301040) | ATRX | 6 | 27 %, males (N = 73) | Wada and Gibbons 2003 [37] | Common |
| Angelman Syndrome (#105830) | UBE3A | 5 | 79 % (N = 39) | Bonati et al. 2007 [38] | Common |
| Autosomal Dominant Mental Retardation 1 (#156200) | MBDS | 3S | 100 % (N = 14) | Talkowski et al. 2011 [39] | Common |
| Autosomal Dominant Mental Retardation 5 (#612621) | SYNGAP1 | 1S | 60 % (N = 10) | Hamdan et al. 2011 [40] | Common |
| Autosomal Dominant Mental Retardation 12 (#614562) | ARID1B | S6 | 3 % (N = 8) | Halgren et al. 2012 [41] | Variable |
| Autosomal Dominant Mental Retardation 23 (#615761) | SETD5 | 1S | ≤ 71 % (N = 7) | Grozeva et al. 2014 [42] | No evidence |
| Autosomal Dominant Mental Retardation 24 (#615828) | DEAF1 | 2S | 75 % (N = 4) | Vulto-van Silfhout et al. 2014 [43] | Variable |
| Autosomal Dominant Mental Retardation 26 (#615834) | AUTS2 | 3 | 41 % (N = 17) | Beunders et al. 2013 [44] | Variable |
| Autosomal Dominant Mental Retardation 30 (#616083) | ZMYND11 | 3 | 29 % (N = 7) | Coe et al. 2014 [45] | Variable |
| Autosomal Recessive Mental Retardation 3 (#608443) | CC2D1A | 3 | 31 % (N = 16) | Manzini et al. 2014 [46] | Variable |
| Autosomal Recessive Mental Retardation 38 (#615516) | HERC2 | NS | 86 % (N = 7) | Puffenberger et al. 2012 [47] | Common |
| Athabaskan Brainstem Dysgenesis Syndrome (#601536) | HOXA1 | 5 | 22 % (N = 9) | Tischfield et al. 2005 [48] | Variable |
| Branched-chain Ketoacid Dehydrogenase Kinase Deficiency (#614923) | BCKDK | 3 | 100 % (N = 6) | Novarino et al. 2012 [49] | Variable |
| Brunner Syndrome (#300615) | MAOA | 4 | 71 %, males (N = 7) | Piton et al. 2014 [50] | Uncommon |
| Cardiofaciocutaneous Syndrome 1 (#115150) | BRAF | NS | 20 % (N = 15) | Nava et al. 2007 [51] | Common |
| Cardiofaciocutaneous Syndrome 3 (#615279) | MAP2K1 | Nil | 63 % (N = 8) | Nava et al. 2007 [52] | Common |
| Cerebral Creatine Deficiency Syndrome 1 (#300352) | SLC6A8 | 4 | 68 % (N = 28) | Dunbar et al. 2014 [53] | Common |
| Cerebral Creatine Deficiency Syndrome 2 (#612736) | GAMT | Nil | 43 % (N = 7) | Cheillan et al. 2012 [54] | Common |
| Cerebral Creatine Deficiency Syndrome 3 (#612718) | GATM | NS | 35 % (N = 20) | Mercimek-Mahmutoglu et al. 2014 [55] | Variable |
| CHARGE Syndrome (#214800) | CHD7 | S | 60 % (N = 10) | Smith et al. 2005 [56] | Variable |
| Childhood-onset Epileptic Encephalopathy (#615369) | CHD2 | 2S | 50 % (N = 6) | Chénier et al. 2014 [57] | Common |
| Christian-type of X-linked Syndromic Mental Retardation (#300243) | SLC9A6 | 3 | 89 %, males (N = 9) | Pescosolido et al. 2014 [58] | Common |
| Chromosome 2p16.3 Deletion Syndrome (#614332) | NRXN1 | 2 | 50 % (N = 40) | Dabell et al. 2013 [59] | Variable |
| Cohen Syndrome (#216550) | VPS13B | S | 49 % (N = 45) | Howlin et al. 2005 [60] | Common |
| Congenital Rett Syndrome (#613454) | FOXG1 | 5 | 100 % (N = 26) | Kortüm et al. 2011 [61] | Common |
| Cornelia de Lange Syndrome 1-5 (#122470, 300590, 610759, 614701, 300882) | NIPBL | Nil | 62 % (N = 34) | Moss et al. 2008 [62] | Variable |
| | SMC1A | | | | |
| | SMC3 | | | | |
| | RAD21 | | | | |
| | HDAC8 | | | | |
| Early Infantile Epileptic Encephalopathy 4 (#612164) | STXBP1 | NS | 29 % (N = 7) | Barcia et al. 2014 [63] | Common |
| Early Infantile Epileptic Encephalopathy 6 (#607208) | SCN1A | S | 24 % (N = 37) | Li et al. 2011 [64] | Common |
### Table 1  List of intellectual disabilities highly comorbid with autism, including gene symbols, SFARI ratings, estimates of autism comorbidity, and indications of epilepsy comorbidity (Continued)

| Disorder                                      | Gene Symbol | SFARI Rating | Comorbidity Estimate | epilepsy comorbidity |
|-----------------------------------------------|-------------|--------------|-----------------------|----------------------|
| Early Infantile Epileptic Encephalopathy 9 (#300088) | PCDH19      | S            | 22 %, females (N = 27) | Scheffer et al. 2008 [69] Common |
| Early Infantile Epileptic Encephalopathy 24 (#615871) | HCN1        | NS           | 66 % (N = 6)          | Nava et al. 2014 [70] Common |
| Fragile X Mental Retardation Syndrome (#300624) | FMR1        | S            | 45 % (N = 64)         | Clifford et al. 2007 [71] Common |
| Glass Syndrome (#612313)                      | SATB2       | 4            | 29 % (N = 7)          | Balasubramanian et al. 2011 [72] Common |
| Helmoortel-Van der AA Syndrome (#615873)      | ADNP        | 1            | 100 % (N = 11)        | Helmoortel et al. 2014 [73] Common |
| KEBG Syndrome (#148050)                       | ANKRD11     | 33 % (N = 9) |                       | Ockelen et al. 2014 [74] Uncommon |
| Kleefstra Syndrome (#610253)                  | EHMT1       | 3S           | most, % unknown (N = 20) | Willemsen et al. 2012 [75] Common |
| Lowe Oculocerebrorenal Syndrome (#305000)    | OCRL        | NS           | 71 %, males (N = 52)  | Oliver et al. 2011 [76] Common |
| Lubs X-linked Mental Retardation Syndrome (#300260) | MECP2    | S            | 100 %, males (N = 18) | Ramocki et al. 2009 [77] Common |
| Lujan-Fryns Syndrome (#309520)                | MED12       | 6            | ≤62 %, males (N = 32) | Williams 2006 [78] Common |
| Marshall-Smith Syndrome (#602535)            | NFIX        | Nil          | 83 % (N = 6)          | van Balkom et al. 2011 [79] Variable |
| Mental Retardation with Language Impairment and Autistic Features (#613670) | FOXP1 | 3           | 75 % (N = 4)         | Le Fereve et al. 2013 [80] Variable |
| Mowat-Wilson Syndrome (#235730)               | ZEB2        | Nil          | 40 % (N = 6)          | Evans et al. 2012 [81] Common |
| Mucopolysaccharidosis, Type IIIA (#252900)    | SGSH        | Nil          | 29 % (N = 73)         | Heron et al. 2011 [82] Common |
| Muscular Dystrophy-dystroglycanopath (Congenital with Mental Retardation), Type B3 (#613151) | POMGNT1 | NS           | 22 % (N = 9)          | Hehr et al. 2007 [83] Variable |
| Myhre Syndrome (#139210)                     | SMAD4       | NS           | 25 % (N = 8)          | Caputo et al. 2012 [84] Variable |
| Myotonic Dystrophy 1 (#160900)               | DMPK        | S            | 49 % (N = 57)         | Ekstrom et al. 2008 [85] Variable |
| Neurodegeneration due to Cerebral Folate Transport Deficiency (#613068) | FOLR1    | Nil          | 35 % (N = 20)         | Ramaekers and Blau 2004 [86] Common |
| Nicolaides-Baraitser Syndrome (#601358)       | SMARCA2     | Nil          | 28 % (N = 18)         | Sousa et al. 2009 [88] Common |
| Noonan Syndrome 3 (#600942)                  | KRAS        | Nil          | 33 % (N = 6)          | Nava et al. 2007 [55] Uncommon |
| Nonphotosensitive Trichoiodothyropath #234050 | MPLKIP      | Nil          | 60 % (N = 5)          | Heller et al. 2015 [89] Variable |
| Noonan Syndrome 3 (#600942)                  | KRAS        | Nil          | 33 % (N = 6)          | Nava et al. 2007 [55] Variable |
| Norrie Disease (#310600)                     | NDP         | Nil          | 27 %, males (N = 56)  | Smith et al. 2012 [90] Uncommon |
| Phelan-McDermid Syndrome (#606232)           | SHANK3      | 1S           | 52 % (N = 130)        | Phelan et al. 2001 [91] Common |
| Pitt-Hopkins Syndrome (#610954)              | TCF4        | NS           | 75 % (N = 8)          | van Balkom et al. 2011 [79] Uncommon |
| Renpenning Syndrome 1 (#309500)              | PQBP1       | Nil          | 38 %, males (N = 13)  | Germainaud et al. 2011 [94] Variable |
| Rett Syndrome (#312750)                      | MECP2       | S            | 100 %, females (N = 35) | Hagberg et al. 1983 [95] Common |
| Schaan-Yang Syndrome (#615547)               | MAGEL2      | NS           | 100 % (N = 6)         | Schaaf et al. 2013 [96] Common |
| Smith-Lemli-Opitz Syndrome (#270400)         | DHCR7       | S            | 75 % (N = 14)         | Sikora et al. 2006 [98] Common |
| Smith-Magenis Syndrome (#182290)             | RAI1        | S            | 90 % (N = 26)         | Laje et al. 2010 [99] Common |
| Terntamy Syndrome (#218340)                  | C12orf57    | NS           | 100 % (N = 10)        | Akizu et al. 2013 [100] Common |
| Tuberous Sclerosis 2 (#613254)               | TSC2        | S            | 40 % (N = 103)        | Numis et al. 2011 [101] Common |
| Warburg Micro Syndrome 4 (#615663)           | TBC1D20     | Nil          | 100 % (N = 7)         | Liegel et al. 2013 [102] Common |
| Wiedemann-Steiner Syndrome (#605130)         | KMT2A       | 2S           | 33 % (N = 6)          | Jones et al. 2012 [103] Variable |
the database in terms of their functional enrichment ($p = 0.4501$, $Z = 0.8$, Diff $= -0.1436$, 0.3236).

VarAut exhibited similar though more modest trends in functional enrichment as seen within HCA, such as regulation of gene expression ($p = 0.0101$) and regulation of DNA-templated transcription ($p = 0.0034$), although this was only apparent compared to HCE, the latter which tended to house particularly low enrichment in all of these terms.

HCE, VarEp, and ID Only did not show consistent differences in enrichments in biological processes, with the exception of HCE functional enrichment in lipid metabolic processes compared to all groups ($p = 0.0092$–0.0486) except VarEp ($p = 0.1135$). Compartmental enrichments for the non-autism groups were also minimal, with the exception of ID only enrichment within the Golgi membrane compared to all groups ($p = 0.0298$) except VarEp ($p = 0.3132$).

For more in-depth analysis, the HCA and HCE comorbidity groups were each divided in two according to their patterns of inheritance (dominant vs. recessive) and were compared against one another for significant GO term enrichments. Significant functional enrichments differentiated both sets of dominant and recessive groups. HCA dominant genes, for instance, were comparatively more enriched than HCA recessive genes in anatomical structure development ($p = 0.0340$), nervous system development ($p = 0.0126$), cell differentiation ($p = 0.0395$), regulation of gene expression ($p = 0.0033$), regulation of DNA-templated transcription ($p = 0.0048$), and chromosome organization ($p = 0.0232$) (Fig. 2b). This suggests that many of the significant GO terms associated with the larger HCA group are primarily driven by this dominantly inherited subgroup.

Meanwhile, the HCE recessive gene group neared significant enrichment in the term lipid metabolic process compared to HCE recessive ($p = 0.0561$). In addition, though they likewise did not reach significance, the recessive group was also comparatively enriched in the endoplasmic reticulum as well as the endoplasmic reticulum membrane ($p = 0.0913$). This suggests that disturbances to protein trafficking through the cell, particularly the endoplasmic reticulum, may be a risk factor for recessive forms of epilepsy.

Dominant HCE genes, in contrast, were enriched in terms related to structural constituent of cytoskeleton ($p = 0.0018$), transmembrane transporter complex ($p = 0.0003$), potassium ion transmembrane transporter ($p = 0.0018$), protein complex ($p < 0.0001$), and myelin sheath ($p = 0.007$) (Fig. 2c). The functional significance of these associations is not currently well understood.

These results together suggest that dominant and recessive patterns of inheritance may diverge according to gene function, though the reasons for this are currently unknown. It is possible that haploinsufficiency may be more or less detrimental according to broader groups of protein function, leading to variations in penetrance. In addition, we also found that the HCA group had a higher ratio of dominant:recessive disorders than HCE, though the relevance of this also cannot currently be determined and may simply be a reflection of the divergent classes of functional enrichment ($p < 0.0001$).

### Protein-protein interaction network data

Upon further study of our gene groups, we find that neither of our autism groups differ significantly from one another in their connectivity to the core PPI network, either in the number of proteins that connect to the core PPI ($p = 0.1053$) nor in the number of intermediary nodes that lay between our proteins and their nearest core PPI neighbor (network “tightness”) ($p = 0.6098$) (Fig. 3c, d). In addition, HCA does not differ from HCE in terms of the number of proteins that connect with the network ($p = 0.9151$), but they do vary according to

---

**Table 1** List of intellectual disabilities highly comorbid with autism, including gene symbols, SFARI ratings, estimates of autism comorbidity, and indications of epilepsy comorbidity (Continued)

| Disorder | Gene Symbol | SFARI Rating | Autism Comorbidity | Epilepsy Comorbidity |
|----------|-------------|--------------|--------------------|----------------------|
| ARX      | GRIA3       | NI           | 32%, males (N=6)   | Philips et al. 2014 [104] Common |
| X-linked Mental Retardation 1 (#309530) | IQSEC2      | NI           | 55%, males (N=9)   | Tran Mau-Them et al. 2013 [105] Uncommon |
| X-linked Mental Retardation 2 (#300271) | RAB39B      | 4            | 33%, males (N=9)   | Russo et al. 2000 [107] Common |
| X-linked Mental Retardation 98 (#300912) | KIAA2022    | NS           | 43%, males (N=7)   | van Maldergem et al. 2013 [109] Common |
| X-linked Mental Retardation with or without Seizures (#300419) | ARX         | S            | 50%, males (N=6)   | Turner et al. 2002 [110] Variable |
| X-linked Syndromic Mental Retardation 14 (#300676) | UPF3B       | S            | 50%, males (N=8)   | Tarpey et al. 2007 [111] Variable |

Thirty-two of these genes are either not included in the SFARI Gene Database, are unscored, or are scored as a “6”. SFARI rating system: S = syndromic; 1 = high confidence; 2 = strong evidence; 3 = suggestive evidence; 4 = minimal evidence; 5 = hypothesized; 6 = not supported; NI = not included; NS = included but not scored (see Additional file 3 for references used to estimate comorbidities)
### Table 2: Full gene list by category

| ID with highly comorbid autism (HCA) N=71 | ID with variable autism (VarAut) N=124 | ID with highly comorbid epilepsy (HCE) N=86 | ID with variable epilepsy (VarEp) N=78 | ID without autism or epilepsy (ID only) N=75 |
|------------------------------------------|--------------------------------------|------------------------------------------|--------------------------------------|------------------------------------------|
| ADNP                                    | ACSL4                                | AASS                                    | ADAR                                 | AAAS                                    |
| ADSL                                    | ACY1                                 | ACTB                                    | ADC1K                                | ADAT3                                   |
| ANKRD11                                 | ADGRG1                               | ACTG1                                   | ALDH3A2                              | AHDC1                                   |
| ARID1B                                  | AFF2                                 | ADK                                     | AP4B1                                | AIFM1                                   |
| ARX                                     | AHI1                                 | AGA                                     | AP4E1                                | ALX4                                     |
| ATRX                                    | ALDH5A1                              | ALDH18A1                                | AP4SI1                               | AP1S1                                   |
| AUTS2                                   | ALDH7A1                              | ALDH4A1                                | ARID1A                                | AP4M1                                   |
| BCKD1K                                  | AP1S2                                | ALG13                                   | ATIC                                 | ARSE                                     |
| BRAF                                    | ARHGEP6                              | ALG6                                   | BSCL2                                | B4GALT1                                 |
| C12orf57                                | ASL                                  | ARG1                                   | CACNG2                                | B4GALT7                                 |
| CC2D1A                                  | BBS10                                | ARHGEF9                                | CDKSRAP2                              | BBS7                                     |
| CHD2                                    | BCKDHA                               | ASPM                                    | CENPJ                                | BRWD3                                   |
| CHD7                                    | BCKDHB                               | ASXL1                                   | CHMP1A                               | C12orf65                                 |
| DEAF1                                   | BTD                                  | ATP1A2                                 | COG8                                  | CSorH2                                   |
| DHR87                                   | CACNA1D                              | ATP1A3                                 | COL4A2                                | CA8                                     |
| DMPK                                    | CAMTA1                               | ATP60A2                                | CSPP1                                 | CASC5                                   |
| EHMT1                                   | CASK                                 | ATR                                     | CYP2F1A                               | CDH15                                   |
| FM1R                                    | CBS                                  | CACD8B6C                               | DCAF17                                | CEP152                                   |
| FOLR1                                   | CC2D2A                               | CLN5                                   | DHTKD1                               | CLCNK8                                   |
| FOXG1                                   | CDKL5                                | CLN8                                   | DP2B                                 | COG1                                     |
| FOXP1                                   | CDON                                 | CLP1                                   | DNMT3A                                | COG6                                     |
| GAMT                                    | CEP290                               | CTSA                                   | EF2AK3                                | CRADD                                    |
| GATM                                    | CEP41                                | CUL4B                                   | ESCO2                                 | CRBN                                     |
| GRIA3                                   | CHK8                                 | D2HGDH                                 | FAT4                                  | CTDP1                                    |
| HCN1                                    | CNTNAP2                              | DHC2R24                                | FGF14                                 | DDHD2                                   |
| HDAC8                                   | COGS                                 | DHFR                                   | FGRF1                                 | DDX59                                   |
| HERC2                                   | CREBBP                               | DPM1                                   | GALE                                  | DLAT                                     |
| HOXA1                                   | CTCF                                 | EFTUD2                                 | GALT                                  | EMD                                      |
| IQSEC2                                  | CTNNB1                               | ELOVL4                                 | GATAD28                               | ENTPD1                                   |
| KIAA2022                                | DAG1                                 | EPG5                                   | GJC2                                  | FGFR3                                    |
| KRAS                                    | DBT                                  | ERLIN2                                 | GRIN1                                 | FTC3                                     |
| MAGEL2                                  | DCHS1                                | FAM126A                                | HFE                                   | GIF                                      |
| MADD                                    | DCX                                 | FBXL4                                   | KSF1                                  | GMP4A                                    |
| MAP2K1                                  | DDC                                 | GABRA1                                 | KAT6B                                 | GNPTAB                                    |
| MBDS                                    | DEPDSC5                              | GMP8                                   | KIAA1279                              | HPRT1                                    |
| MECP2                                   | DLG3                                 | GRM1                                   | KRAS                                  | IDUA                                     |
| MED12                                   | DMD                                 | KANS1                                  | LBR                                   | IGBP1                                    |
| MLL                                     | DOC8                                 | KCNQ2                                  | LINS1                                 | INPPSE                                   |
| MPLK1P                                  | DPYD                                 | KCNT1                                  | MANBA                                 | IRX5                                     |
| NDP                                    | DPYS                                 | KCTD7                                  | MTRR                                  | KCNK9                                    |
| NEF                                | DYM                                 | KIF5C                                  | MYCN                                  | KIFSA                                    |
| NIPBL                                   | DYNIC1H1                            | KPTN                                   | NIN                                  | LARGE                                    |
| NRXN1                                   | DYSR1A                              | MGAT2                                  | NTRK1                                 | LARP7                                    |
| Gene  | Gene  | Gene  | Gene  |
|-------|-------|-------|-------|
| OCLR  | EP300 | MLYCD | OFD1  |
| PCDH19| FGFR2 | MTR   | PEPD  |
| POMGNT1| FH   | NAGA  | PEX1  |
| PQBP1 | FTSJ1 | NDE1  | PIGO  |
| RAB39B| GABRB3| NSDHL | PORCN |
| RAD21 | OD1   | OCLN  | PPXO  |
| RA1   | GLYCTK| PAK3  | PTCH1 |
| SATB2 | GNS   | PDX1  | PYCR1 |
| SCN1A | GRK2  | PDX2  | POLR3B|
| SETD5 | GRIN2A| PGAP2 | RAB23 |
| SGSH  | GRIN2B| PGAP3 | RAB3GAP1|
| SHANK3| GSS   | PHGDH | RAB3GAP2|
| SLC6A8| HCFC1 | PIGA  | RBBP8 |
| SLC9A6| HDAC4 | PIKL  | RIT1  |
| SMAD4 | HEPACAM| PIK3R2| SERAC1 |
| SMARCA2| HGSNAT| PLP1  | SLC4A4|
| SMCA1 | HPD   | PNNK  | SMT1  |
| SMC3  | HRAS  | PPT1  | TAF2  |
| STXB1P| HUWE1 | QARS  | SMARCA4|
| SYNGAP1| IL1RAPL1| RAB18 | SMARCB1|
| TBC1D20| KANK1 | RANBP2| SOBP  |
| TCF4  | KCNJ10| RTTN  | SOX3  |
| TSC2  | KDM4C | SEPECS| STIL  |
| UBE3A | KDM6A | SLC19A3| SYP   |
| UFP3B | KIAA0196| SLC2A1| TAT   |
| VPS13B| KIF7  | SMS   | UQCRQ |
| ZEB2  | KIF7  | SMS   | TBC1D7|
| ZMYND11| KIF7 | SMS   | TBC1D7|
|        | KMT2D | SPTAN1| TBC1D7|
|        | KMT2D | SPTAN1| TBC1D7|
|        | L1CAM | ST3GAL5| TRAPPC9|
|        | L2HG DH| SFT2  | UBE3B |
|        | LAMB1 | TBC3  | UBR1  |
|        | MAN1B1| TECPR2| UBR1  |
|        | MCPH1 | TRMT10A| VLDLR |
|        | MEF2C | TUBA8  | ZDHHC9|
|        | METTL23| TUBB2A| ZIC2   |
|        | MKKS  | TUBB3  |       |
|        | MTHFR | UBE2A  |       |
|        | NAGLU | VPSS3  |       |
|        | NDN   | WDR45  |       |
|        | NR2F1 | WDR62  |       |
|        | NSD1  | WWOX   |       |
|        | NSUN2 | ZSWIM6 |       |
|        | OPHN1 | PAH    |       |
Meanwhile, HCA and VarAut exhibit a larger core network than both VarEp ($p = 0.0001$–$0.0011$) and ID only ($p = 0.0001$–$0.0003$), but the level of network tightness does not differ significantly. Overall, our autism groups exhibit a larger, tighter protein network surrounding the core PPI compared to all other groups.

The core PPI network is a single extensive protein network integral for patterning of the central nervous system (e.g., dorsoventral patterning) as well as later processes of neural maturation and ongoing plasticity. This network includes morphogens such as Wingless Integration Site (WNT), NOTCH, nuclear receptor co-repressor (NCOR), SWItch/Sucrose Non-Fermentable (SWI/SNF), Sonic Hedgehog (SHH), transforming growth factor-β (TGF-β), bone morphogenetic proteins (BMP), and extracellular signal-related kinase 1 and 2 (ERK1/2)—a functional module with considerable overlap.

Table 2 Full gene list by category (Continued)

| Gene        |
|-------------|
| PAX6        |
| PGM3        |
| PHF6        |
| PHF8        |
| PIGV        |
| POMT1       |
| PRSS12      |
| PTEN        |
| PTPN11      |
| RELN        |
| RODGI       |
| RPGRIP1L    |
| RPS6KA3     |
| SCN2A       |
| SCN8A       |
| SETBP1      |
| SHH         |
| SLC16A2     |
| SLC17A5     |
| SLC35C1     |
| SOX10       |
| SOX11       |
| SRCAP       |
| STAMBP      |
| TCN2        |
| TCTN3       |
| TSC1        |
| TSPAN7      |
| TUBA1A      |
| TUBG1       |
| TUSC3       |
| UPB1        |
| USP9X       |
| ZBIB20      |
| ZNF711      |
| ZNF81       |
| Casanova et al. Molecular Autism (2016) 7:18 |
with recent reports by Hormozdiari et al. [24] pinpointing a particularly enriched protein network in autism that centers around WNT, NOTCH, SWI/SNF, and NCOR [25].

Why the VarEp group diverged from HCE somewhat in the String analysis, as well as various GO term enrichments, is currently unknown. However, the HCE group \( (N = 88) \) contained a significantly greater number of neurodegenerative conditions compared to VarEp \( (N = 84) \) \( (z = 2.5, p = 0.0119, \text{Diff} = 0.0375, 0.3025) \), suggesting that the prevalence of neurodegeneration could underlie divergence in their respective etiologies.

**Discussion**

**Research implications**

Despite that the rare conditions reported here within the HCA group display very high rates of autism comorbidity, genes associated with close to half of these conditions are not currently included within the SFARI Gene Database, are included but unscored, or are scored as “6” indicating that the “evidence does not support a role in [autism spectrum disorders]” (Table 1) [19]. In addition, 69% of HCA genes that are scored in SFARI but not considered “syndromic” do in fact present with multiple physical dysmorphia, such as complex craniofacial malformations,
indicating that with further investigation, these genes should and are likely to be subsumed under this umbrella category.

Some of the non-SFARI conditions that are used in this study are based on small numbers of patients and likely require further in depth diagnostics to confirm the presence of autism symptoms in these conditions at a higher-than-expected rate. In addition, a number of the studies is also hampered by poor diagnostic methodology, an issue that has plagued many of the earlier genetics studies, whereas today’s gold standards include the use of the ADOS and ADI-R as well as ID comparison groups for the purposes of research. Most tellingly, however, the removal of these conditions from the HCA group did not change our results; genes with high penetrance for syndromic and nonsyndromic autism are typically localized to the nucleus and are involved in transcription regulation. These results in particular stress the importance of grouping risk genes according to penetrance when possible, because this information was already present within the SFARI database, though is currently lost amidst the other genetic data.

Because genes associated with HCA are overrepresented within the nucleus and tend to directly regulate transcription, this suggests that mutation penetrance for autism may be strongly linked with regulatory, as opposed to enzymatic, transduction, and structural, cellular networks [26]. Transcription factors and regulators are the most common examples within this group; however, other epigenetic regulators, such as heterochromatin remodelers, ubiquitin ligases, and methylation regulators, were also overrepresented. It may be for their phenotypic penetrance that dominantly inherited conditions were so common in the HCA comorbidity group.

In addition, the presence of gene subgroups within HCA that share considerable overlap with a module for autism risk reported by Hormozdiari et al. [24] suggests that a significant portion of these cases, as well as those in VarAut, are rooted in disturbances to patterning of the CNS and ongoing deviations in neural maturation and

---

**Fig. 3 Central nervous system patterning.**

- **A** String 10 results for general interaction of the core PPI within itself.
- **B** General locations of the major embryonic organizing centers of the central nervous system that underlie variations in its dorsoventral and rostrocaudal patterning.
- **C** Results across groups for absolute connectivity of experimental proteins to the core PPI network.
- **D** Network tightness of proteins surrounding the core PPI network, based upon the average number of intermediary nodes between a target protein and its nearest core PPI neighbor (Figure B adapted from Wurst and Bally-Cuif. Nat Rev Neurosci 2001;2(2):99-108)
plasticity. All of these morphogenetic pathways share considerable crosstalk that is foundational to dorsoventral and rostrocaudal patterning, planar cell polarity, locomotion, neuritogenesis, and finally synaptogenesis and plasticity [25, 27–29]. If disturbed, they are likely to affect all stages of neuronal development, spanning from the most foundational to the most nuanced. Further work at the cellular and tissue levels will be required to investigate whether disturbances to patterning may play roles in these conditions and how such patterning defects, alongside later impairments to neuronal development and plasticity, underlie the behavioral and neurological phenotypes.

In support of this, previous work investigating high-risk autism-related genes has suggested that disturbances to neural maturation may be a common theme to autism [30]. Our present results indicate that epigenetic dysregulation could inappropriately suppress or prematurely promote the expression of gene products, leading to chronological changes in the typical developmental process and ultimately to gross structural, microstructural, and physiological perturbations.

For example, the fragile X mental retardation protein (FMRP) associated with FXS normally aids in suppression of translation, thereby controlling timing of neural differentiation. Instead, when the FMR1 gene is mutated leading to decreased production of FMRP, neurogenesis occurs prematurely [31]. These neurons also fail to express mature markers in a timely fashion (e.g., GAD67), a disparity likely resulting in poor maturation and circuit integration of adult neurons, and absolute or relative macrocephaly, periventricular heterotopias, and volumetric increase in periventricular white matter, further evidence of a pathological heterochrony and disturbances to patterning [6, 7, 32].

In contrast to our autism, variable epilepsy, and ID only groups, ID with highly comorbid epilepsy exhibited particularly low enrichment in nervous system-specific processes, was more often involved in lipid metabolism, and, compared to conditions with variable epilepsy, had higher rates of neurodegeneration. While this evidence is tantalizing, further work is needed to determine whether ID and epilepsy related to neurodegenerative processes follow a different etiological path than those related to general nervous system development as may be seen in ID with autism or variable epilepsies.

In contrast to the other groups, it was clear that the HCA group is surprisingly homogeneous, suggesting that risk for autism lies within a specific and very definable set of molecular events that confer greater risk the further downstream these elements are affected, i.e., at the level of the gene and its product. This likewise suggests that the further upstream a particular risk factor or environmental effector, the more variable the penetrance for autism due to the number of elements that may intercede and alter events, e.g., feedback inhibition. This is strongly suggested by the divergent compartmental enrichments seen in HCA vs. VarAut, in which the former is highly enriched for the nucleus while the latter is mildly enriched throughout numerous cellular compartments and within cell projections in particular. Ultimately, risk is a threshold effect and a risk factor must be closely upstream of its target (e.g., in the case of epigenetic regulators) or, if further upstream, then it must be capable of avoiding feedback inhibition in order to reach threshold in a consistent highly penetrant fashion (e.g., in the case of select sodium channel mutations).

On a similar note, factors that are comparatively less penetrant yet still confer measurable risk suggest the presence of additional variables, e.g., polygenic effects, environmental agencies, etc., in the determination of their etiologies. Given the nature of genetic selection, common gene variants that provide variable autism risk (i.e., common disease-common variant) are more likely to explain a wider breadth of cases than the rare, often de novo, mutations that confer higher penetrance for the phenotype. Although interestingly, a recent study by Alvarez-Mora et al. [33] reported that in a subset of high-functioning cases they studied, over 50 % (6/11) of the identified rare potentially deleterious single nucleotide variants (SNV) occurred within the HCA genes reported here, suggesting that these genes may be targets with variable penetrance dependent upon the specific type of mutation. Sanders et al. [34] have found that highly penetrant deleterious SNVs tend to affect the same genes that are also targets of small copy number variants (CNV) in autism, such as occurs in the monoogenic conditions studied here. Meanwhile, individual genes that comprise larger CNVs each confer comparatively lesser risk. It is possible that if rare SNVs tend to overlap HCA genes, less penetrant SNVs (e.g., common variants) may overlap genes typically comprising larger CNVs and reflect polygenic risk.

In the future, we may find that the genetics of autism tends to diverge according to levels of severity, with rare mutations (e.g., small and large CNVs, highly deleterious rare SNVs) responsible for a significant portion of low-functioning individuals with intellectual disability while other rare SNVs and common variants, perhaps even polygenic and/or environmentally driven, play important roles in a larger portion of the moderate-to-high-functioning ranges. This hypothesis is not entirely unlike that proposed by Folstein [35] in which she suggests that autistic individuals with profound ID, complex dysmorphic features, or specific genetic conditions represent phenotypes that are clinically unique compared to the idio-pathic autism reported by Kanner. In this case, however, we are suggesting that the genetics, though not necessarily the overall biology, diverges between the two.
Additional limitations
Aside from the limitations mentioned above concerning questions of the diagnostic reliability of some of the HCA conditions, additional shortcomings of this study involve the availability of information regarding what are typically rare conditions and potential underreporting regarding comorbidities such as autism and epilepsy. There are, for instance, a number of VarAut conditions in which case studies or small group studies reporting autism incidence are available but no larger studies have been performed in order to provide better estimates of co-occurrence. Examples include conditions such as Succinic Semialdehyde Dehydrogenase Deficiency (#271980), Autosomal Dominant Mental Retardation 21 (#615502), and Dihydropyrimidine Dehydrogenase Deficiency (#274270) to name just a few that are likely worthy of more intensive study in relation to autism. Therefore, it is highly likely that some of the conditions presented within this study have been mis-categorized. In order to limit that occurrence to an absolute minimum, various genetic databases were used in conjunction with phenotypic data.

In addition, the fact that we limited our study of autism and epilepsy to forms of ID subsequently limits the potential scope of applicability of our results, although in doing so we were able to estimate comorbidity rates. We therefore hope that future research may elucidate which of the results presented here are applicable to the broader autism spectrum or whether these data solely define a subgroup of autism.

Because genetic mutations are infrequently identical across different individuals with a single form of ID, it is possible that some cases of autism or epilepsy within our variable groups were not due to mutations involving the primary gene associated with the monogenic condition but were instead due to confounding effects of other genes, such as may be seen in larger chromosomal rearrangements. However, most of the results presented here exhibit strong functional patterns and therefore while individual IDs may ultimately be mis-categorized, we are confident that the conclusions regarding the larger groups are relatively sound.

Conclusions
While there were distinctive genetic differences between groups, particularly between ID with autism vs. ID without, the strongest findings within this study were overwhelmingly those regarding the HCA group. In particular, we find that the majority of genes that confer high risk for autism are located within the nucleus and function as nuclear epigenetic regulators.

Our results also suggest that both autism groups represents a collection of disabilities that share not only the autism and ID phenotypes, but also likely share developmental similarities in disruption to patterning of the central nervous system. Further work by way of molecular and animal studies is still needed to address this hypothesis.

Aside from novel conclusions derived from the genetic data presented here, we also hope that this curated list may be useful for others and can be updated as new information becomes available. In addition, we hope that this study can be used to inform further clinical research in order to better update databases such as SFARI, affecting research foci in future.

Availability of data and materials
Additional information on the list of monogenic intellectual disabilities used in this study is available through the Online Mendelian Inheritance in Man (OMIM) database accessible through http://www.omim.org/. OMIM numbers are included within the table in Additional file 1.

Additional files

| Additional file 1: Full list of monogenic intellectual disabilities used in this study, gene information, relevant phenotypic data, and short references for categorization. (XLSX 138 kb) | Additional file 2: References for Additional file 1. (DOCX 164 kb) |
| Additional file 3: Full statistical results. (DOCX 157 kb) |

Abbreviations
HCA: intellectual disability with highly comorbid autism; HCE: intellectual disability with highly comorbid epilepsy and without autism; ID: intellectual disability; ID only: intellectual disability without autism or epilepsy; VarAut: intellectual disability with variable autism; VarEp: intellectual disability with variable epilepsy.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
ELC conceived the study, curated the data, provided expertise in genetics, and performed a minimal amount of the statistical analyses. JLS, HC, and NSS performed the bulk of the statistical analyses. MFC provided expertise on autism and neuropathology and was integral in helping to design the study. All authors contributed to the drafts and have read and approved the final manuscript.

Acknowledgements
Financial support for this work came from NIH grant R01 HD-65279.

Author details
1Department of Biomedical Sciences, University of South Carolina, South Carolina, USA. 2Department of Pediatrics, Greenville Health System, Patewood Medical Campus, 200A Patewood Dr, Greenville, SC 29615, USA. 3Department of Mathematical Sciences, Clemson University, Clemson, USA. 4Department of Biostatistics and Epidemiology, University of South Carolina, South Carolina, USA.

Received: 24 November 2015 Accepted: 1 March 2016
Published online: 15 March 2016

References
1. Viscidi EW, Triche EW, Pescosolido MF, McLean RL, Joseph RM, Spence SJ, et al. Clinical characteristics of children with autism spectrum disorder and co-occurring epilepsy. PLoS One. 2013;8(7):e67797.
52. Novarino G, EFishawy P, Kayserili H, Meguid NA, Scott EM, Schroth J, et al. Mutations in BCKD-kinase lead to a potentially treatable form of autism with epilepsy. Science. 2012;338:394–7.

53. Piton A, Poquet H, Redin C, Masuel A, Lauer J, Muller J, et al. 2014. 20 ans après: a second mutation in MADA identified by targeted high-throughput sequencing in a family with altered behavior and cognition. Eur J Hum Genet. 2014;22:776–83.

54. Palmer EE, Leffer M, Rogers C, Shaw M, Carroll R, Earl J, et al. New insights into Brunner syndrome and potential for targeted therapy. Clin Genet. 2016;89:120–7.

55. Nava C, Hanna N, Michot C, Pereira S, Pouvreau N, Niihori T, et al. Cardio-facio-cutaneous and Noonan syndromes due to mutations in the RAS/MAPK signaling pathway: genotypo-phenotype relationships and overlap with Costello syndrome. J Med Genet. 2007;44:763–71.

56. Dunbar M, Jaggi-Munarti S, Sargent M, Stockler-Ipsiroglu S, van Karnebeek CD. Treatment of X-linked creatine transporter (SLC6A8) deficiency: systematic review of the literature and three new cases. Mol Genet Metab. 2014;112:259–74.

57. Cheillan D, Jonczak-Chevallier Curt M, Briand G, Salomons GS, Mention-Muluye K, Dobbelare D, et al. Screening for primary creatine deficiencies in French patients with unexplained neurological symptoms. Orphanet J Rare Dis. 2012;7:96.

58. Mercimek-Mahmutoglu S, Ndika J, Kanhai W, de Villemeur TB, Cheillan D, Christensen E, et al. Thirteen new patients with guanidinoacetate methyltransferase deficiency and functional characterization of nineteen novel missense variants in the GAMT gene. Hum Mutat. 2014;35:462–9.

59. Smith IM, Nichols A, Iseukutz K, Blake K. Canadian Paediatric Surveillance Program. Behavioral profiles and symptoms of autism in CHARGE syndrome: preliminary Canadian epidemiological data. Am J Med Genet. 2005;133A:248–56.

60. Chénier S, Yoon G, Argiroloups B, Lazor J, Lafamboise R, Ahn JW, et al. CHD2 haploinsufficiency is associated with developmental delay, intellectual disability, epilepsy and neurobehavioral problems. J Neurodev Disord. 2014;6:69.

61. Pescosolido MF, Schwede M, Johnson Harrison A, Schmidt M, Gamsz ED, Chen WS, et al. Expansion of the clinical phenotype associated with mutations in activity-dependent neuroprotective protein. J Med Genet. 2014;51:587–9.

62. Diabell MP, Rosenfeld JA, Bader P, Escobar LF, El-Hsheh D, Valle SE, et al. Investigation of NRXN1 deletions: clinical and molecular characterization. Am J Med Genet A. 2013;161A:717–31.

63. Schaaf CP, Boone PM, Sampath S, Williams C, Bader PI, Mueller JM, et al. Phenotypic spectrum and genotypo-phenotype correlations of NRXN1 exon deletions. Eur J Hum Genet. 2012;20:1240–8.

64. Howlin P, Karpf J, Turk. J. Behavioural characteristics and autistic features in individuals with Cohen syndrome. Eur Child Adolesc Psychiatry. 2005;14:557–64.

65. Korunt F, Das S, Flindt M, Morris-Rosendahl DJ, Stefanova I, Goldstein A, et al. The core FOXG1 syndrome phenotype consists of postnatal microencephaly, severe mental retardation, absent language, dyskinesia, and corpus callosum hypogenesis. J Med Genet. 2011;48:396–406.

66. Moss JF, Oliver C, Berg K, Kaur G, Jephcott L, Cornish K. Prevalence of autism spectrum phenomenology in Comella de Lange and Cru du Chat syndromes. Am J Ment Retard. 2008;113:278–91.

67. Barcia G, Chemaly N, Gobin S, Mili M, Van Bogaert P, Barnerias C, et al. Early epileptic encephalopathies associated with STXBP1 mutations: Could we better delineate the phenotype? Eur J Hum Genet. 2014;22:15–20.

68. Li BM, Liu XR, Yi YH, Deng YH, Su T, Zou X, et al. Autism in Dravet syndrome: prevalence, features, and relationship to the clinical characteristics of epilepsy and mental retardation. Epilepsy Behav. 2011;21:291–5.

69. Scheffer IE, Turner SJ, Dibbens LM, Bayly MA, Friend K, Hodgson B, et al. Epilepsy and mental retardation limited to females with fragile X full mutation and Epilepsy and mental retardation limited to females with fragile X full mutation and a second mutation in MADA identified by targeted high-throughput sequencing in a family with altered behavior and cognition. Eur J Hum Genet. 2014;22:776–83.

70. Nava C, Hanna N, Michot C, Pereira S, Pouvreau N, Niihori T, et al. Cardio-facio-cutaneous and Noonan syndromes due to mutations in the RAS/MAPK signaling pathway: genotypo-phenotype relationships and overlap with Costello syndrome. J Med Genet. 2007;44:763–71.

71. Dunbar M, Jaggi-Munarti S, Sargent M, Stockler-Ipsiroglu S, van Karnebeek CD. Treatment of X-linked creatine transporter (SLC6A8) deficiency: systematic review of the literature and three new cases. Mol Genet Metab. 2014;112:259–74.

72. Cheillan D, Jonczak-Chevallier Curt M, Briand G, Salomons GS, Mention-Muluye K, Dobbelare D, et al. Screening for primary creatine deficiencies in French patients with unexplained neurological symptoms. Orphanet J Rare Dis. 2012;7:96.

73. Mercimek-Mahmutoglu S, Ndika J, Kanhai W, de Villemeur TB, Cheillan D, Christensen E, et al. Thirteen new patients with guanidinoacetate methyltransferase deficiency and functional characterization of nineteen novel missense variants in the GAMT gene. Hum Mutat. 2014;35:462–9.

74. Smith IM, Nichols A, Iseukutz K, Blake K. Canadian Paediatric Surveillance Program. Behavioral profiles and symptoms of autism in CHARGE syndrome: preliminary Canadian epidemiological data. Am J Med Genet. 2005;133A:248–56.

75. Chénier S, Yoon G, Argiroloups B, Lazor J, Lafamboise R, Ahn JW, et al. CHD2 haploinsufficiency is associated with developmental delay, intellectual disability, epilepsy and neurobehavioral problems. J Neurodev Disord. 2014;6:69.

76. Pescosolido MF, Schwede M, Johnson Harrison A, Schmidt M, Gamsz ED, Chen WS, et al. Expansion of the clinical phenotype associated with mutations in activity-dependent neuroprotective protein. J Med Genet. 2014;51:587–9.

77. Diabell MP, Rosenfeld JA, Bader P, Escobar LF, El-Hsheh D, Valle SE, et al. Investigation of NRXN1 deletions: clinical and molecular characterization. Am J Med Genet A. 2013;161A:717–31.

78. Schaaf CP, Boone PM, Sampath S, Williams C, Bader PI, Mueller JM, et al. Phenotypic spectrum and genotypo-phenotype correlations of NRXN1 exon deletions. Eur J Hum Genet. 2012;20:1240–7.

79. Howlin P, Karpf J, Turk. J. Behavioural characteristics and autistic features in individuals with Cohen syndrome. Eur Child Adolesc Psychiatry. 2005;14:557–64.

80. Korunt F, Das S, Flindt M, Morris-Rosendahl DJ, Stefanova I, Goldstein A, et al. The core FOXG1 syndrome phenotype consists of postnatal microencephaly, severe mental retardation, absent language, dyskinesia, and corpus callosum hypogenesis. J Med Genet. 2011;48:396–406.

81. Moss JF, Oliver C, Berg K, Kaur G, Jephcott L, Cornish K. Prevalence of autism spectrum phenomenology in Comella de Lange and Cru du Chat syndromes. Am J Ment Retard. 2008;113:278–91.

82. Barcia G, Chemaly N, Gobin S, Mili M, Van Bogaert P, Barnerias C, et al. Early epileptic encephalopathies associated with STXBP1 mutations: Could we better delineate the phenotype? Eur J Hum Genet. 2014;22:15–20.

83. Li BM, Liu XR, Yi YH, Deng YH, Su T, Zou X, et al. Autism in Dravet syndrome: prevalence, features, and relationship to the clinical characteristics of epilepsy and mental retardation. Epilepsy Behav. 2011;21:291–5.

84. Scheffer IE, Turner SJ, Dibbens LM, Bayly MA, Friend K, Hodgson B, et al. Epilepsy and mental retardation limited to females: an under-recognized disorder. Brain. 2008;131:918–27.

85. Nava C, Dalle C, Rastetter A, Stixiana P, de Kovel CG, Nabbout R, et al. De novo mutations in HCN1 cause early infantile epileptic encephalopathy. Nat Genet. 2014;46:640–5.

86. Clifford S, Disnayake C, Bui QM, Huggins R, Taylor AK, Loesch DZ. Autism spectrum phenotype in males and females with fragile X full mutation and permuation. J Autism Dev Disorder. 2007;37:378–47.

87. Babakrubosmanian M, Smith K, Basel-Vanagiate L, Feingold MF, Brock S, Gowans GC, et al. Case series: 2q31.3 microdeletion syndrome—further delineation of the phenotype. J Med Genet. 2011;48:290–8.

88. Helsmoortel C, Vulo-van Silfhout AT, Cole BP, Vandeneyere G, Rooms L, van den Ende J, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. Nat Genet. 2014;46:380–4.
98. Sikora DM, Pettit-Kekel K, Penfield J, Merkens LS, Steiner RD. The near universal presence of autism spectrum disorders in children with Smith-Lemli-Opitz syndrome. Am J Med Genet A. 2006;140:1511–8.

99. Laje G, Morse R, Richter W, Ball J, Pao M, Smith AC. Autism spectrum features in Smith-Magenis syndrome. Am J Med Genet C Semin Med Genet. 2010;154C:456–62.

100. Akizu N, Shembesh NM, Ben-Omran T, Bastaki L, Al-Tawari A, Zaki MS, et al. Whole-exome sequencing identifies mutated c12orf57 in recessive corpus callosum hypoplasia. Am J Hum Genet. 2013;92:392–400.

101. Numis AL, Major P, Montenegro MA, Muzykewicz DA, Pulsifer MB, Thiele EA. Identification of risk factors for autism spectrum disorders in tuberous sclerosis complex. Neurology. 2011;76:981–7.

102. Liegel RP, Handley MT, Ronchetti A, Brown S, Langemeeyer L, Linford A, et al. Loss-of-function mutations in the TBC1D20 cause cataracts and male infertility in blind sterile mice and Warburg micro syndrome in humans. Am J Hum Genet. 2013;93:1001–14.

103. Jones WD, Dafou D, McEntagart M, Woolard WJ, Elmslie IV, Halder-Espinasse M, et al. De novo mutations in MLL cause Wiedemann-Steiner syndrome. Am J Hum Genet. 2011;98:1–7.

104. Liegel RP, Handley MT, Ronchetti A, Brown S, Langemeeyer L, Linford A, et al. Loss-of-function mutations in the TBC1D20 cause cataracts and male infertility in blind sterile mice and Warburg micro syndrome in humans. Am J Hum Genet. 2013;93:1001–14.

105. Tran Mau-Them F, Willems M, Albrecht B, Sanchez E, Puechberty J, Endele S, et al. Expanding the phenotype of IQSEC2 mutations: truncating mutations in severe intellectual disability. Eur J Hum Genet. 2014;22:89–92.

106. Shoubridge C, Tarpey PS, Abidi F, Ramsden SL, Rujirabanjerd S, Murphy JA, et al. Mutations in the guanine nucleotide exchange factor gene IQSEC2 cause nonsyndromic intellectual disability. Nat Genet. 2010;42:486–8.

107. Russo S, Cogliati F, Cavalleri F, Casagrande MG, Giglioli R, Toniolo D, et al. Mapping to distal Xq28 of nonspecific X-linked mental retardation MRX72: linkage analysis and clinical findings in a three-generation Sicilian family. Am J Med Genet. 2000;94:376–82.

108. Giannandrea M, Bianchi V, Mignogna ML, Siri A, Carabinio S, Delia E, et al. Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. Am J Hum Genet. 2010;86:185–95.

109. van Maldergem L, Hou Q, Kalscheuer V, Rio M, Doco-Fenay M, Medeira A, et al. Loss of function of KIAA2022 causes mild to severe intellectual disability with autism spectrum disorder and impaired neurite outgrowth. Hum Mol Genet. 2013;22:3306–14.

110. Turner G, Partington M, Kerr B, Mangelsdorf M, Geicz J. 2002. Variable expression of mental retardation, autism, seizures, and dystonic hand movements in two families with an identical ARX gene mutation. Am J Med Genet. 2002;112:405–11.

111. Tarpey PS, Raymond FL, Nguyen LS, Rodriguez J, Hackett A, VandeIe L, et al. 2007. Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. Nat Genet. 2007;39:1127–33.