Spatial Clustering of de Novo Missense Mutations Identifies Candidate Neurodevelopmental Disorder-Associated Genes

Stefan H. Lelieveld,[1,6] Laurens Wiel,[1,2,6] Hanka Venselaar,[2] Rolph Pfundt,[3] Gerrit Vriend,[2] Joris A. Veltman,[3,4] Han G. Brunner,[3,5] Lisenka E.L.M. Vissers,[3,6] and Christian Gilissen[3,6,*]

Summary

Haploinsufficiency (HI) is the best characterized mechanism through which dominant mutations exert their effect and cause disease. Non-haploinsufficiency (NHI) mechanisms, such as gain-of-function and dominant-negative mechanisms, are often characterized by the spatial clustering of mutations, thereby affecting only particular regions or base pairs of a gene. Variants leading to haploinsufficency might occasionally cluster as well, for example in critical domains, but such clustering is on the whole less pronounced with mutations often spread throughout the gene. Here we exploit this property and develop a method to specifically identify genes with significant spatial clustering patterns of de novo mutations in large cohorts. We apply our method to a dataset of 4,061 de novo missense mutations from published exome studies of trios with intellectual disability and developmental disorders (ID/DD) and successfully identify 15 genes with clustering mutations, including 12 genes for which mutations are known to cause neurodevelopmental disorders. For 11 out of these 12, NHI mutation mechanisms have been reported. Additionally, we identify three candidate ID/DD-associated genes of which two have an established role in neuronal processes. We further observe a higher intolerance to normal genetic variation of the identified genes compared to known genes for which mutations lead to HI. Finally, 3D modeling of these mutations on their protein structures shows that 81% of the observed mutations are unlikely to affect the overall structural integrity and that they therefore most likely act through a mechanism other than HI.

De novo mutations affecting protein-coding genes are a major cause of intellectual disability (ID) and other developmental disorders (DDs).[1,2] Several whole exome sequencing (WES) studies have identified ID syndromes molecularly characterized by very specific spatial clustering of de novo missense mutations.[3–6] Similarly, large-scale WES studies of individuals affected by ID/DD have recently leveraged this phenomenon as supporting evidence of the involvement of a gene in disease.[7,8] This spatial clustering of de novo mutations (DNMs) is typical for missense mutations in genes without clear, or limited numbers of, truncating mutations subsequently degraded by nonsense-mediated mRNA decay, suggesting that these clustered mutations act through a different mechanism than haploinsufficiency (HI).[9] Alternative pathophysiological mechanisms that might underlie (de novo) mutation clustering are gain-of-function or dominant-negative effects, resulting in the alteration or impairment of specific protein function.[10,11] We note that while spatial clustering is commonly taken to indicate a mechanism different from loss-of-function,[12] this is not an absolute rule, and a loss-of-function mechanism cannot be excluded without functional evidence.[13] Here, we developed a method to identify genes with spatially clustered DNMs and applied this to DNMs identified in a large cohort of individuals with ID/DD.[14]

We downloaded all DNMs occurring in individuals with ID/DD from de novo-db version 1.3[14] identified through WES and whole genome sequencing which were then re-annotated with our in-house variant annotation pipeline. The de novo mutations included in the analysis were previously validated by a second independent method or showed a high validation rate for a subset of de novo mutations. In addition, we added 1,183 de novo variants identified in the exomes of an in-house ID cohort that was previously published.[8] To further reduce the risk of including sequencing artifacts and/or genotyping errors, we excluded all de novo variants that were present more than once in the ExAC dataset (Table S1).[15] These efforts resulted in 6,495 protein coding DNMs, including 4,061 missense mutations, in 5,302 individuals with ID/DD (Table S2).

We set out to determine for any gene whether the observed de novo missense mutations cluster more than expected compared to random permutations. Hereto, we selected for each the longest representative transcript (i.e., part of the GENCODE basic set)[16] and calculated the geometric mean distance δ_g over all missense DNMs on

[1]Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, 6525 GA, the Netherlands; [2]Centre for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, 6525 GA, the Netherlands; [3]Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, 6525 GA, the Netherlands; [4]Institute of Genetic Medicine, International Centre for Life, Newcastle University, Newcastle upon Tyne, NE1 3BZ, UK; [5]Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, 6202 AZ, the Netherlands
[6]These authors contributed equally to this work
*Correspondence: christian.gilissen@radboudumc.nl
http://dx.doi.org/10.1016/j.ajhg.2017.08.004.

Table 1. List of Identified Genes with Clustering de Novo Missense Mutations

Gene name	Transcript ID	# de novo missense	Median distance (bp)	p value	Adj. p value
ACTL6B[a]	ENST00000160382	3	0	5.70E-07	1.10E-02
ALG13	ENST00000394780	3	0	1.50E-07	2.89E-03
CDK13	ENST00000181839	12	273	<1.00E-08	<1.93E-04
COL4A3BP	ENST00000380494	6	18	2.60E-07	5.01E-03
GABBR2[a]	ENST00000259455	3	0	9.00E-08	1.74E-03
GRIN2B	ENST00000609686	11	354	1.57E-06	3.03E-02
KCNH1	ENST00000271751	7	65	1.00E-07	1.93E-03
KCNQ2	ENST00000354587	20	301	5.00E-08	9.64E-04
KIF5C	ENST00000435030	3	0	1.40E-07	2.70E-03
PACS1	ENST00000320580	9	0	<1.00E-08	<1.93E-04
PACS2[a]	ENST00000458164	3	0	1.50E-07	2.89E-03
PCGF2	ENST00000360797	3	0	1.11E-06	2.14E-02
PPP2R1A	ENST00000322088	4	5	4.60E-07	8.87E-03
PPP2R5D	ENST00000485511	16	10	<1.00E-08	<1.93E-04
SMAD4	ENST00000398417	4	6	1.60E-07	3.08E-03

P values are based on a permutation test (N = 1.00E+08). Adj. p values are corrected by Bonferroni correction. The three identified genes that have not yet been implicated in ID/DD are indicated by an [a].

cDNA. δ_g was calculated by taking the mean distance normalized for transcript length l over all (M) combinations of x_i and x_j of the missense DNMs (Equation 1), where x represents the position for mutation i and j respectively. Statistical significance was determined by performing 1.00E+08 (or N) permutations and calculating for each permuted geometric mean distance (δ'_g) how many times this resulted in the same or smaller geometric mean distance as observed (Equation 2) Permutation p values were corrected for multiple testing via Bonferroni procedure based on the 19,280 genes of the Agilent SureSelect v5 exome enrichment kit.

$$\delta_g = \left(\prod_{\substack{i,j=1 \\ i<j}}^{M} \frac{|\, x_i - x_j \,| + 1}{l+1} \right)^{\frac{1}{M}} \quad (1)$$

$$p = \frac{\sum_{i=1}^{N} \left[\delta'_g \leq \delta_g \right] + 1}{N+1} \quad (2)$$

We first validated our method on a dataset of DNMs identified in 2,448 unaffected siblings and healthy control studies[14,17–22] (Table S3). In this cohort, we failed to identify genes for which clustering of de novo missense mutations reached statistical significance (Table S4). However, application of our method to the dataset of 4,061 DNMs, containing 583 genes with more than one de novo missense mutation, revealed 15 genes with significant clustering[7,8,23–25] (Table 1, Figure 1, Figures S1–S15). In these genes, a total of 107 de novo missense mutations contributed to mutation clustering, ranging from three to 20 mutations per gene with an average distance ranging from 0 to 354 bp. To exclude a correlation between the extent of clustering and the total number of de novo missense mutations analyzed, we applied our method to a cohort of 6,154 de novo missense variants present in de novo-db excluding the five studies incorporated in the ID/DD cohort, and found no such correlation (Figure S16). To examine whether this set of 15 genes is relevant in the context of ID/DD, we compared these genes to a list of 1,541 genes for which mutations are known to cause ID/DD (Table S5). This list of genes was a compilation of two manually curated lists of disease-associated genes including "confirmed" unique genes from DDG2P (n = 1,098; see Web Resources) and 1,034 genes offered for diagnostic testing in individuals with ID/DD by our in-house diagnostic facility (see Web Resources). Among the 15 identified genes with mutation clustering, we find 12 genes for which mutations have previously been implicated in ID/DD, constituting a significant enrichment (p = 3.09e-03; Fisher's exact test; Tables S6 and S7), and confirming that our method is valid for its purpose. The inclusion of exome data of two large DDD-studies in both the DDG2P gene list and the ID/DD cohort of this study could introduce a potential bias.[1,7] To exclude such bias, we repeated this analysis while excluding the DDD-specific genes identified in the two exome studies yielding a significant enrichment (p = 3.68E-02; Table S7A–S7C).

We also identified three genes with clustered de novo missense mutations that have not yet been implicated in ID/DD: *ACTL6B* (MIM:612458), *GABBR2* (MIM:607340),

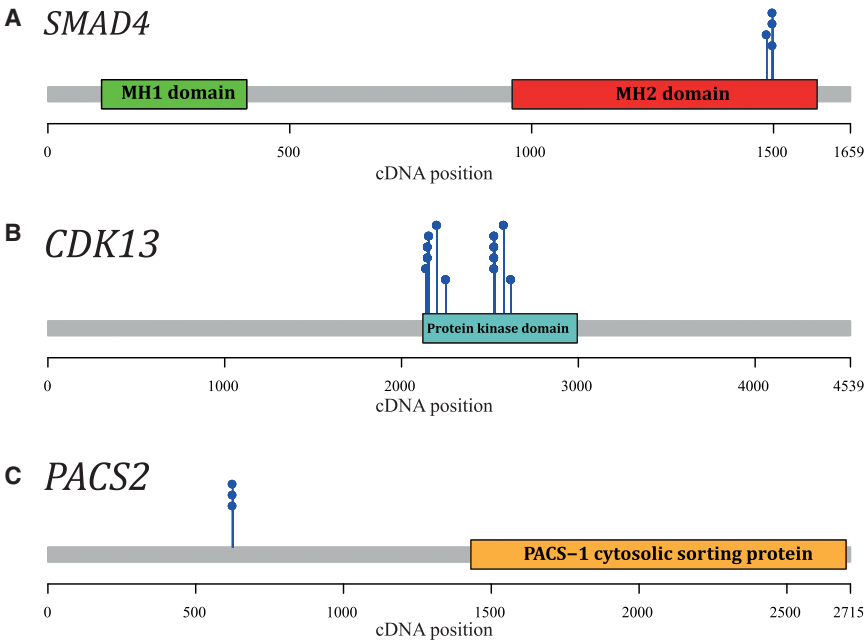

Figure 1. Examples of Identified Genes with Clustering Mutations
Protein domains are annotated based on Pfam HMM search.[26] cDNA locations of de novo missense mutations are depicted by blue pins. Genes shown here are as follows: *SMAD4* (A), *CDK13* (B), *PACS2* (C). Figures visualizing the clustering of de novo missense mutations in the other 12 genes are provided in Figures S1–S15.

and *PACS2* (MIM:610423). None of these genes would have been identified based on an enrichment for de novo mutations in this cohort (Table S8). Further systematic evaluation of gene function supports a role in (neuro)development for two of these genes (Table 2 and Table S9). *ACTL6B*, encoding Actin-like 6B (also known as *BAF53B*), is a pivotal co-factor for the SWI/SNF neuron-specific chromatin remodeling complex nBAF, which is required for neural development and dendritic outgrowth.[27,28] Also, *GABBR2*, which is a component of the G protein-coupled GABA receptor, plays a critical role in the fine-tuning of inhibitory synaptic transmission,[29–31] and other members of the GABA receptor family have already been conclusively linked to neurodevelopmental disorders.[32,33] *GABBR2* was very recently also reported by others to show significant de novo mutation clustering in a neurodevelopmental cohort.[6]

Our method might potentially identify clustering based on identical mutations in multiple individuals only as a result of issues in the underlying cohort. It could for instance be that the same individual was included in multiple studies and therefore occurs twice in the cohort. For 99 out of 107 de novo missense mutations (92.5%) occurring in the 15 genes with clustering mutations, we could decisively conclude that they occurred as unique events in separate individuals based on a combination of the gender of the affected individual and the presence of additional de novo mutations (Table S10). Nevertheless, it might be possible that siblings of affected individuals were included who share a DNM due to parental gonadal mosaicism.[34] Alternatively, DNMs might occur multiple times in disease cohorts as a consequence of a locally increased mutation rate. Examples of the latter might for instance incur a selective growth advantage (i.e., selfish mutations[35]) and thereby result in a pattern of mutational clustering such as known for *FGFR2* (MIM 176943) mutations in Apert syndrome (MIM 101200).[35] However, biological relevance for the mutations in the identified genes in the context of ID/DD is suggested by the fact that in our control cohort genes with significant clusters were absent, and that for the majority of our identified genes experimental evidence in literature supports a NHI mutational mechanism (Table S11).

We hypothesized that the clustering de novo missense mutations of the 15 genes might exert their effects through mechanisms other than haploinsufficiency. To validate this hypothesis, we compiled a set of 116 genes known for mutations that exert disease through non-haploinsufficient (NHI) mechanisms. Hereto, we selected for genes that have a "confirmed" status in the DDG2P list, or are present on both the Radboudumc ID/DD diagnostic testing and DDG2P lists (irrespective of the DDG2P status). Furthermore, genes were selected to be dominant (mono-allelic), with the pathophysiological mechanism being either "activating," "all missense/in frame," and/or "dominant negative" (Table S12). In addition, we generated a set of 183 haploinsufficient genes for which mutations are associated with ID/DD from the DDG2P gene list by selecting "loss-of-function" as the "mutation consequence" and "mono-allelic" for the "allelic requirement" in the DDG2P gene list (Table S13).

Interestingly, for eight of the 12 genes for which mutations are known to cause ID/DD and for which we identified mutation clustering, the disease mechanism on the constructed gene list was reported to be NHI. For these eight genes, it is either gain-of-function or dominant negative, thereby showing statistical enrichment for NHI mechanisms ($p = 2.66E-03$, Fisher's exact test; Table S14 and S15). For two of the three remaining genes (*GRIN2B* [MIM 138252] and *SMAD4* [MIM 600993]) both HI and NHI consequences have been reported,[36–39] suggesting that for mutations in these genes more complex genotype-phenotype relations might exist, where HI and NHI mechanisms cause clinically distinct ID/DD-related disorders. For *KCNQ2* (MIM 602235), the reported mutational mechanism is HI although a literature search also revealed cases with dominant-negative effects.[40] We also investigated the extent of the evidence for NHI mechanisms

Table 2. Gene Function for Candidate Genes with Clustered Mutations

Gene Name	Summary of Gene Function	Interactions
ACTL6B	Belongs to the neuron-specific chromatin remodeling complex (nBAF complex) and is required for postmitotic neural development and dendritic outgrowth.	Complex formation with ACTB, ARID1A, SMARCA2, SMARCA4, SMARCE1, SMARCC1, SMARCC2, SMARCD2, SMARCB1
GABBR2	Postsynaptic GABAB receptor activity regulates excitatory neuronal architecture and spatial memory.	Heterodimerization is required for the formation of a functional GABA-B receptor.
PACS2	Multifunctional sorting protein, controling endoplasmic reticulum-mitochondria communication and Bid-mediated apoptosis.	N/A

Third column indicates whether the encoded protein has physical interactions with other proteins. See Table S9 for extended information.

and found that extensive functional work of mutations supporting NHI mechanisms has been previously published for eight of the 12 known genes (Table S11).

Further we hypothesized that NHI genes should be depleted for truncating mutations in individuals with ID/DD, i.e., mutations resulting in premature translation termination, whereby the mRNA is targeted for nonsense mediated decay. In our initial analyses focusing on de novo missense mutations only, we excluded truncating mutations from our dataset. Retrospectively, we searched for truncating DNMs in the 15 identified genes with clustering de novo missense mutations. We found only three predicted truncating mutations in two of 15 genes, which is significantly less than expected based on the total number of DNMs found in the total cohort for all HI genes ($p<1.00e-05$; Permutation test).

We have previously hypothesized that genes with mutations acting through NHI mechanisms might be more intolerant to normal variation than genes with mutations acting though a HI mechanism for ID/DD.[8] To test for tolerance to variation, existing scores like pLI[15] are not useful because these capture tolerance to mRNA truncating variation rather than tolerance to variation in general. Therefore, we measured tolerance to variation as the ratio of missense over synonymous variation "d_N/d_S," which has been used by us and others previously for predicting disease genes.[2,41,42] We downloaded all PASS-filtered single nucleotide variants (SNVs) from ExAC (n = 9,035,134) and constructed a "d_N/d_S" measure by counting the unique missense SNVs $missense_{obs}$, and the unique synonymous SNVs $synonymous_{obs}$, while correcting for sequence composition using the total possible unique missense and synonymous SNVs ($missense_{bg}$ and $synonymous_{bg}$ respectively): $d_N/d_S = ((missense_{obs}/missense_{bg})/(synonymous_{obs}/synonymous_{bg}))$ (Table S16).

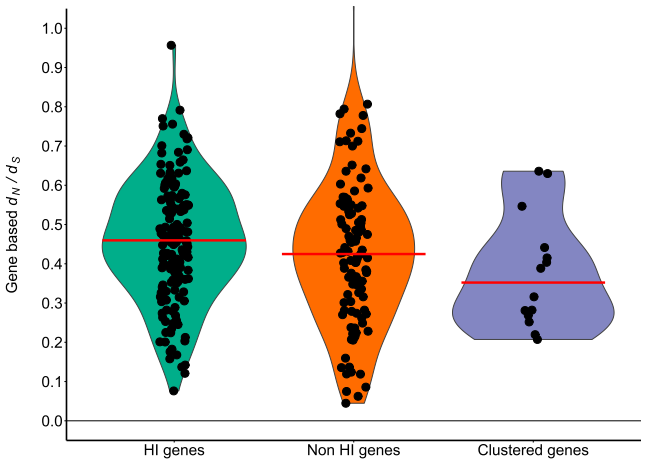

Figure 2. Intolerance to Missense Variation
Violin plots show the distribution of the gene-based d_N/d_S (y axis) per gene set (x axis). The median d_N/d_S is indicated by a red horizontal line. The NHI genes are more intolerant to missense variation than HI genes (HI genes median: 0.460; NHI genes median: 0.428; p = 2.24e-03). In addition, the identified genes with clustering mutations are more intolerant to missense variation than HI genes (genes with clustering mutations median: 0.352; p = 8.45e-03).

Based on calculations of these scores for the sets of 116 NHI, and 183 HI genes, we indeed find that genes with mutations acting through a NHI mechanism are significantly more intolerant to missense variation than genes with mutations acting though a HI mechanism (p = 2.24e-03; permutation test, Figure 2). In line with our hypothesis, also our set of 15 genes with clustered DNMs was significantly less tolerant to missense variation compared to the set of 183 genes with mutations acting through a HI mechanism (p = 8.45e-03; permutation test, Figure 2).

Modeling of missense mutations in a 3D protein structure is helpful to gain more insight into the possible structural and functional effects.[43] Conceptually, mutations in the core of the protein structure are more likely to prevent proper folding than mutations on the protein surface.[44] The impact of a surface change, however, depends entirely on the spatial context and is therefore less likely to result in misfolding and subsequent protein degradation.[45] Consequently, de novo disease-causing missense mutations preventing proper folding cause protein degradation, and thus indirectly lead to HI, similar to protein truncating mutations in such genes. To test the hypothesis that our clustered de novo missense mutations do not generally result in HI due to protein misfolding, we modeled mutations onto the 3D protein structure using YASARA & WHAT IF Twinset.[46,47] A (partial) protein 3D structure was available

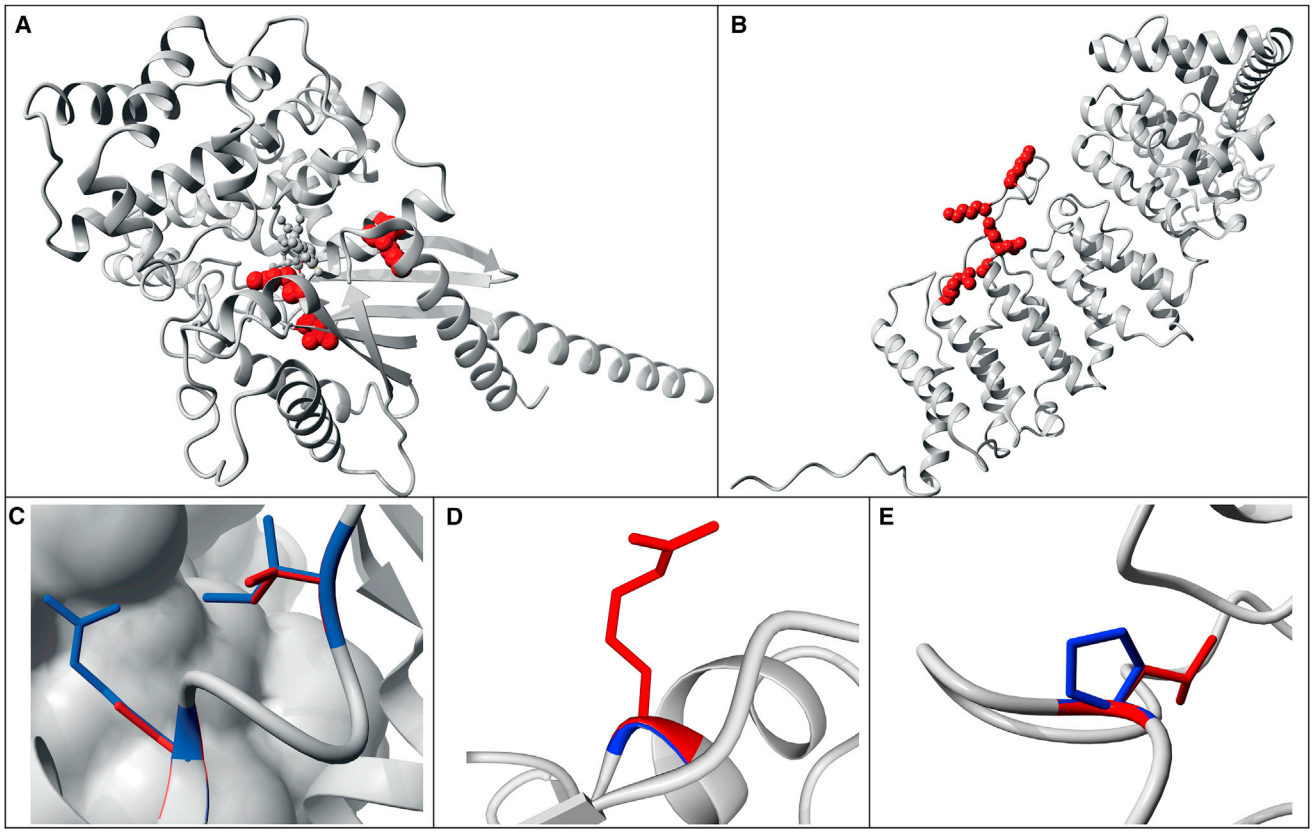

Figure 3. Examples of Modeling of Missense Mutations on 3D Protein Structures
Wild-type residues are marked in blue; de novo mutations are indicated as red globes or lines (Tables S17).
(A) 3D structure of GNA1, acting through HI, showing that the modeled missense mutations are buried and likely to disrupt protein folding.
(B) Structure of PPP2R5D, acting through NHI, where the modeled missense mutations affect mostly surface residues and are expected to have no or only local structural effects.
(C) Zoom-in of known missense variants p.Arg496Cys and p.Ile500Val in SMAD4 known to act through a gain-of-function mechanism. These variants are located on the surface of the monomer and in contact with another SMAD4 monomer.[38]
(D) Zoom-in of the missense variant p.Gly343Arg in ACTL6B which is located at the surface. The side-chain points toward the solvent, therefore the larger Arginine will fit.
(E) Zoom-in of the missense variant p.Pro65Leu in PCGF2 close to the interaction site with other molecules.

or could be created via homology modeling[7] for 10 of the 15 identified genes. We assessed 48 missense mutations on the 3D structure (i.e., buried, at the surface, or semi-buried) and whether the mutation was likely to affect protein folding (no effect, local effect, or large effect; Figure 3, Table S17). To compare the results of 3D modeling of clustered mutations, we also modeled 75 de novo disease-causing missense mutations in 25 genes with mutations acting though HI (Table S13) for which a structure was available (Table S17). For the HI genes, 42% of missense mutations were buried and 34% of mutations were located at the protein surface. In the 10 genes for which a mutational NHI effect is proposed, only 11% of mutations was buried whereas 61% was located at the protein surface ($p = 1.26E-03$, chi-square test; Table S17). Even more strikingly, only 19% of the clustering de novo missense mutations were likely to result in a large structural change that would affect protein function whereas this was observed for 63% of de novo missense mutations in HI genes ($p = 8.43E-06$, chi-square test). These results support the notion that the majority of clustered de novo disease-causing missense mutations do not result in haploinsufficiency at the protein structure level, but exert their effect through other mechanisms. Possibly this could be through the functional impairment of protein-protein interactions, as we noted that two of the three candidate ID/DD-associated genes require complex formation or joining of protein subunits (e.g., multimerization) to be functional (Table 2).

In conclusion, we developed a method for the identification of disease-associated genes based on the significance of spatial mutation clustering within a gene. We show that our method successfully identifies genes previously implicated in ID/DD. Moreover, we identified three genes with similar clustering patterns that we propose as candidate ID/DD-associated genes. Our findings support the concept that these mutations mostly exert their pathogenic effect through disease mechanisms other than haploinsufficiency. Thus, our findings might indicate a larger contribution of non-haploinsufficient mechanisms to ID/DD than previously thought.

Supplemental Data

Supplemental Data include 16 figures and 17 tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2017.08.004.

Acknowledgments

This work was in part financially supported by grants from the Netherlands Organization for Scientific Research (916-14-043 to C.G. and 918-15-667 to J.A.V.), the European Research Council (ERC Starting grant DENOVO 281964 to J.A.V.) and from the Radboud Institute for Molecular Life Sciences, Radboud university medical center (R0002793 to G.V.). We thank Stephan Boersma for help with writing the software for analysis of cluster mutations.

Received: February 14, 2017
Accepted: August 4, 2017
Published: August 31, 2017

Web Resources

gene2phenotype, http://www.ebi.ac.uk/gene2phenotype/downloads
Genome Diagnostics Gene List: https://www2.radboudumc.nl/Informatievoorverwijzers/Genoomdiagnostiek/en/Pages/Intellectualdisability.aspx
OMIM, http://www.omim.org/
YASARA: http://www.yasara.org/

References

1. The Deciphering Developmental Disorders Study (2015). Large-scale discovery of novel genetic causes of developmental disorders. Nature *519*, 223–228.
2. Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., et al. (2014). Genome sequencing identifies major causes of severe intellectual disability. Nature *511*, 344–347.
3. Srour, M., Caron, V., Pearson, T., Nielsen, S.B., Lévesque, S., Delrue, M.A., Becker, T.A., Hamdan, F.F., Kibar, Z., Sattler, S.G., et al. (2016). Gain-of-Function Mutations in RARB Cause Intellectual Disability with Progressive Motor Impairment. Hum. Mutat. *37*, 786–793.
4. Hoischen, A., van Bon, B.W., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M., de Vries, P., de Reuver, R., Wieskamp, N., Mortier, G., et al. (2010). De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat. Genet. *42*, 483–485.
5. Schuurs-Hoeijmakers, J.H., Oh, E.C., Vissers, L.E., Swinkels, M.E., Gilissen, C., Willemsen, M.A., Holvoet, M., Steehouwer, M., Veltman, J.A., de Vries, B.B., et al. (2012). Recurrent de novo mutations in PACS1 cause defective cranial-neural-crest migration and define a recognizable intellectual-disability syndrome. Am. J. Hum. Genet. *91*, 1122–1127.
6. Geisheker, M.R., Heymann, G., Wang, T., Coe, B.P., Turner, T.N., Stessman, H.A.F., Hoekzema, K., Kvarnung, M., Shaw, M., Friend, K., et al. (2017). Hotspots of missense mutation identify neurodevelopmental disorder genes and functional domains. Nat. Neurosci. *20*, 1043–1051.
7. Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature *542*, 433–438.
8. Lelieveld, S.H., Reijnders, M.R., Pfundt, R., Yntema, H.G., Kamsteeg, E.J., de Vries, P., de Vries, B.B., Willemsen, M.H., Kleefstra, T., Löhner, K., et al. (2016). Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. Nat. Neurosci. *19*, 1194–1196.
9. Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. PLoS Genet. *6*, e1001154.
10. Turner, T.N., Douville, C., Kim, D., Stenson, P.D., Cooper, D.N., Chakravarti, A., and Karchin, R. (2015). Proteins linked to autosomal dominant and autosomal recessive disorders harbor characteristic rare missense mutation distribution patterns. Hum. Mol. Genet. *24*, 5995–6002.
11. Wilkie, A.O. (1994). The molecular basis of genetic dominance. J. Med. Genet. *31*, 89–98.
12. Stehr, H., Jang, S.H., Duarte, J.M., Wierling, C., Lehrach, H., Lappe, M., and Lange, B.M. (2011). The structural impact of cancer-associated missense mutations in oncogenes and tumor suppressors. Mol. Cancer *10*, 54.
13. Kamburov, A., Lawrence, M.S., Polak, P., Leshchiner, I., Lage, K., Golub, T.R., Lander, E.S., and Getz, G. (2015). Comprehensive assessment of cancer missense mutation clustering in protein structures. Proc. Natl. Acad. Sci. USA *112*, E5486–E5495.
14. Turner, T.N., Yi, Q., Krumm, N., Huddleston, J., Hoekzema, K., HA, F.S., Doebley, A.L., Bernier, R.A., Nickerson, D.A., and Eichler, E.E. (2017). denovo-db: a compendium of human de novo variants. Nucleic Acids Res *45*, D804–D811.
15. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.
16. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.
17. Besenbacher, S., Sulem, P., Helgason, A., Helgason, H., Kristjansson, H., Jonasdottir, A., Jonasdottir, A., Magnusson, O.T., Thorsteinsdottir, U., Masson, G., et al. (2016). Multi-nucleotide de novo Mutations in Humans. PLoS Genet. *12*, e1006315.
18. Genome of the Netherlands Consortium (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet. *46*, 818–825.
19. Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216–221.
20. Krumm, N., Turner, T.N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., Raja, A., Coe, B.P., Stessman, H.A., He, Z.X., et al. (2015). Excess of rare, inherited truncating mutations in autism. Nat. Genet. *47*, 582–588.
21. Turner, T.N., Hormozdiari, F., Duyzend, M.H., McClymont, S.A., Hook, P.W., Iossifov, I., Raja, A., Baker, C., Hoekzema, K., Stessman, H.A., et al. (2016). Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. Am. J. Hum. Genet. *98*, 58–74.

22. Conrad, D.F., Keebler, J.E., DePristo, M.A., Lindsay, S.J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C.L., Torroja, C., Garimella, K.V., et al.; 1000 Genomes Project (2011). Variation in genome-wide mutation rates within and between human families. Nat. Genet. *43*, 712–714.
23. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., et al. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet *380*, 1674–1682.
24. de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. *367*, 1921–1929.
25. Halvardson, J., Zhao, J.J., Zaghlool, A., Wentzel, C., Georgii-Hemming, P., Månsson, E., Ederth Sävmarker, H., Brandberg, G., Soussi Zander, C., Thuresson, A.C., and Feuk, L. (2016). Mutations in HECW2 are associated with intellectual disability and epilepsy. J. Med. Genet. *53*, 697–704.
26. Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. *44* (D1), D279–D285.
27. Choi, K.Y., Yoo, M., and Han, J.H. (2015). Toward understanding the role of the neuron-specific BAF chromatin remodeling complex in memory formation. Exp. Mol. Med. *47*, e155.
28. Kuroda, Y., Oma, Y., Nishimori, K., Ohta, T., and Harata, M. (2002). Brain-specific expression of the nuclear actin-related protein ArpNalpha and its involvement in mammalian SWI/SNF chromatin remodeling complex. Biochem. Biophys. Res. Commun. *299*, 328–334.
29. Ramírez, O.A., Vidal, R.L., Tello, J.A., Vargas, K.J., Kindler, S., Härtel, S., and Couve, A. (2009). Dendritic assembly of heteromeric gamma-aminobutyric acid type B receptor subunits in hippocampal neurons. J. Biol. Chem. *284*, 13077–13085.
30. Robbins, M.J., Calver, A.R., Filippov, A.K., Hirst, W.D., Russell, R.B., Wood, M.D., Nasir, S., Couve, A., Brown, D.A., Moss, S.J., and Pangalos, M.N. (2001). GABA(B2) is essential for g-protein coupling of the GABA(B) receptor heterodimer. J. Neurosci. *21*, 8043–8052.
31. Jones, K.A., Borowsky, B., Tamm, J.A., Craig, D.A., Durkin, M.M., Dai, M., Yao, W.J., Johnson, M., Gunwaldsen, C., Huang, L.Y., et al. (1998). GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2. Nature *396*, 674–679.
32. Møller, R.S., Wuttke, T.V., Helbig, I., Marini, C., Johannesen, K.M., Brilstra, E.H., Vaher, U., Borggraefe, I., Talvik, I., Talvik, T., et al. (2017). Mutations in GABRB3: From febrile seizures to epileptic encephalopathies. Neurology *88*, 483–492.
33. Johannesen, K., Marini, C., Pfeffer, S., Møller, R.S., Dorn, T., Niturad, C.E., Gardella, E., Weber, Y., Søndergård, M., Hjalgrim, H., et al. (2016). Phenotypic spectrum of GABRA1: From generalized epilepsies to severe epileptic encephalopathies. Neurology *87*, 1140–1151.
34. Acuna-Hidalgo, R., Bo, T., Kwint, M.P., van de Vorst, M., Pinelli, M., Veltman, J.A., Hoischen, A., Vissers, L.E., and Gilissen, C. (2015). Post-zygotic Point Mutations Are an Underrecognized Source of De Novo Genomic Variation. Am. J. Hum. Genet. *97*, 67–74.
35. Goriely, A., and Wilkie, A.O. (2010). Missing heritability: paternal age effect mutations and selfish spermatogonia. Nat. Rev. Genet. *11*, 589.
36. Endele, S., Rosenberger, G., Geider, K., Popp, B., Tamer, C., Stefanova, I., Milh, M., Kortüm, F., Fritsch, A., Pientka, F.K., et al. (2010). Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat. Genet. *42*, 1021–1026.
37. Lemke, J.R., Hendrickx, R., Geider, K., Laube, B., Schwake, M., Harvey, R.J., James, V.M., Pepler, A., Steiner, I., Hörtnagel, K., et al. (2014). GRIN2B mutations in West syndrome and intellectual disability with focal epilepsy. Ann. Neurol. *75*, 147–154.
38. Le Goff, C., Mahaut, C., Abhyankar, A., Le Goff, W., Serre, V., Afenjar, A., Destrée, A., di Rocco, M., Héron, D., Jacquemont, S., et al. (2011). Mutations at a single codon in Mad homology 2 domain of SMAD4 cause Myhre syndrome. Nat. Genet. *44*, 85–88.
39. Gallione, C., Aylsworth, A.S., Beis, J., Berk, T., Bernhardt, B., Clark, R.D., Clericuzio, C., Danesino, C., Drautz, J., Fahl, J., et al. (2010). Overlapping spectra of SMAD4 mutations in juvenile polyposis (JP) and JP-HHT syndrome. Am. J. Med. Genet. A. *152A*, 333–339.
40. Wuttke, T.V., Jurkat-Rott, K., Paulus, W., Garncarek, M., Lehmann-Horn, F., and Lerche, H. (2007). Peripheral nerve hyperexcitability due to dominant-negative KCNQ2 mutations. Neurology *69*, 2045–2053.
41. Ge, X., Kwok, P.Y., and Shieh, J.T. (2015). Prioritizing genes for X-linked diseases using population exome data. Hum. Mol. Genet. *24*, 599–608.
42. Ge, X., Gong, H., Dumas, K., Litwin, J., Phillips, J.J., Waisfisz, Q., Weiss, M.M., Hendriks, Y., Stuurman, K.E., Nelson, S.F., et al. (2016). Missense-depleted regions in population exomes implicate ras superfamily nucleotide-binding protein alteration in patients with brain malformation. Npj Genomic Medicine *1*, 16036.
43. Lee, C., and Levitt, M. (1991). Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core. Nature *352*, 448–451.
44. Yue, P., Li, Z., and Moult, J. (2005). Loss of protein structure stability as a major causative factor in monogenic disease. J. Mol. Biol. *353*, 459–473.
45. Venselaar, H., Camilli, F., Gholizadeh, S., Snelleman, M., Brunner, H.G., and Vriend, G. (2013). Status quo of annotation of human disease variants. BMC Bioinformatics *14*, 352.
46. Krieger, E., Koraimann, G., and Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA–a self-parameterizing force field. Proteins *47*, 393–402.
47. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. J Mol Graph *8*, 52–56, 29.

The American Journal of Human Genetics, Volume *101*

Supplemental Data

Spatial Clustering of de Novo Missense Mutations Identifies Candidate Neurodevelopmental Disorder-Associated Genes

Stefan H. Lelieveld, Laurens Wiel, Hanka Venselaar, Rolph Pfundt, Gerrit Vriend, Joris A. Veltman, Han G. Brunner, Lisenka E.L.M. Vissers, and Christian Gilissen

Supplemental Figures

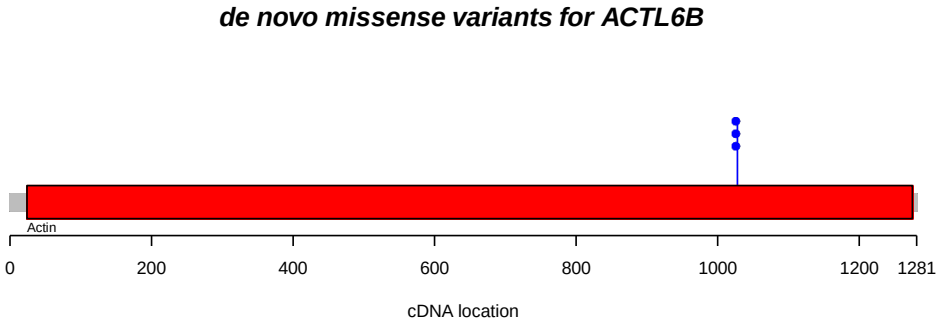

Figure S1 Schematic representation of *ACTL6B* (ENST00000160382). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

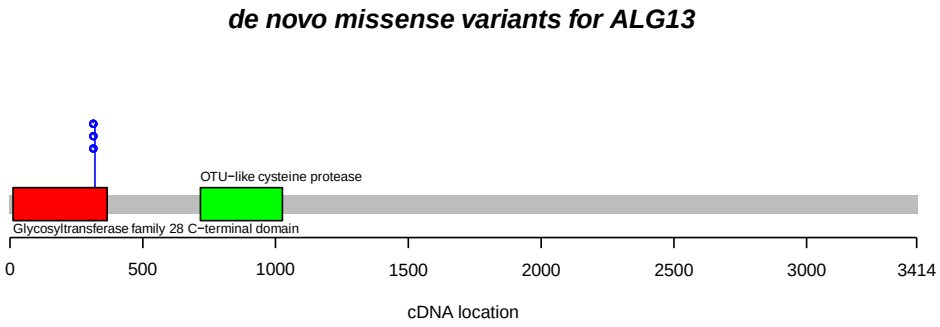

Figure S2 Schematic representation of *ALG13* (ENST00000394780). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

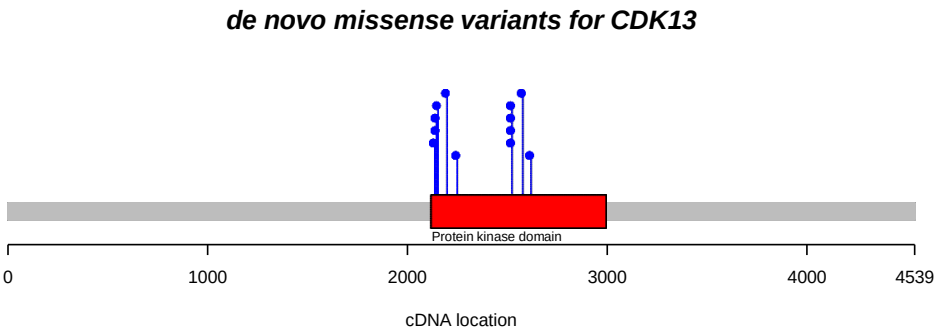

Figure S3 Schematic representation of *CDK13* (ENST00000181839). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

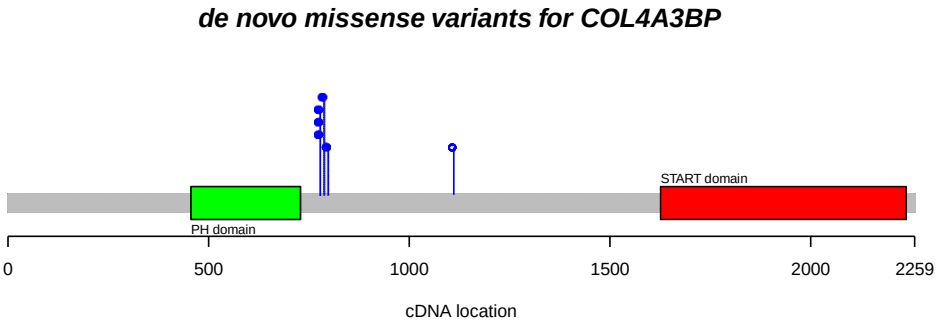

Figure S4 Schematic representation of *COL4A3BP* (ENST00000380494). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

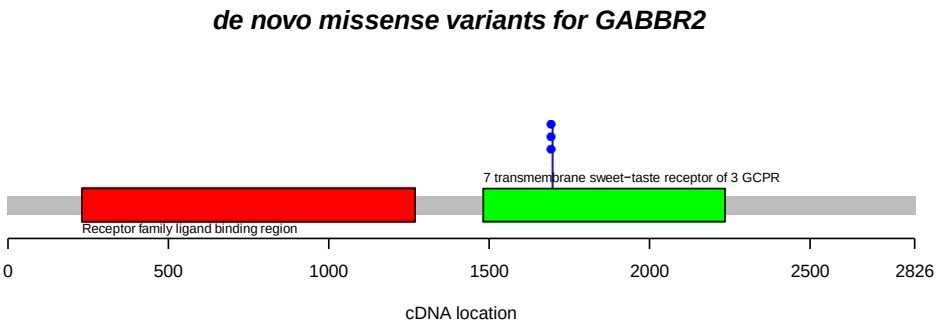

Figure S5 Schematic representation of *GABBR2* (ENST00000259455). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

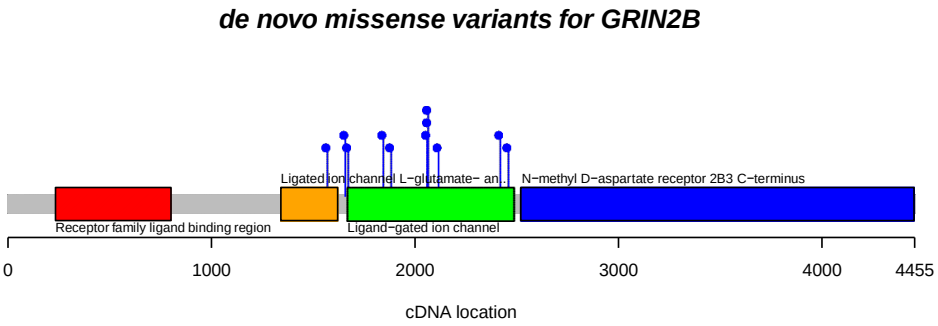

Figure S6 Schematic representation of *GRIN2B* (ENST00000609686). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

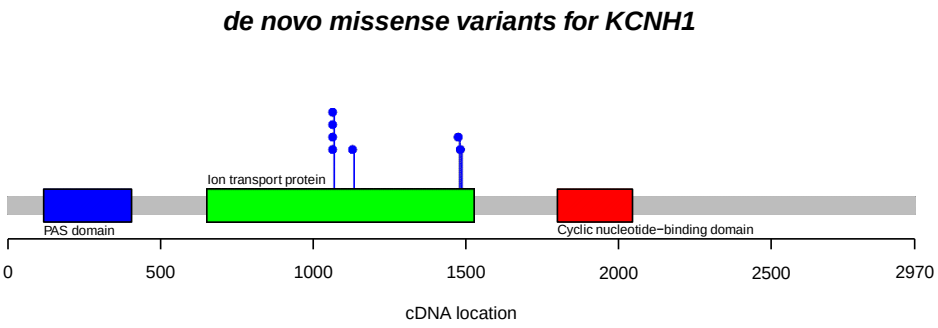

Figure S7 Schematic representation of *KCNH1* (ENST00000271751). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

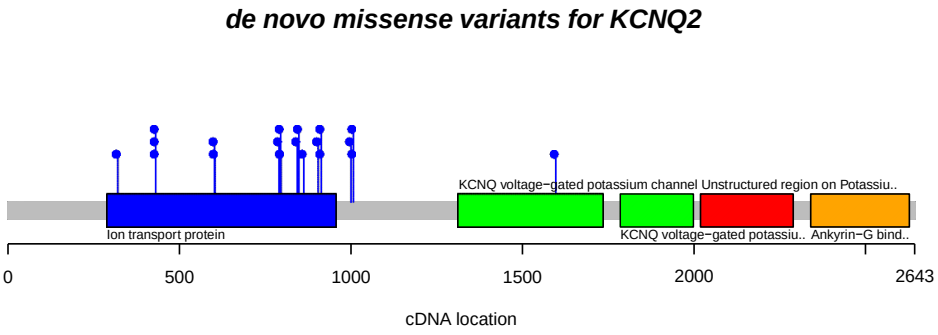

Figure S8 Schematic representation of *KCNQ2* (ENST00000354587). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

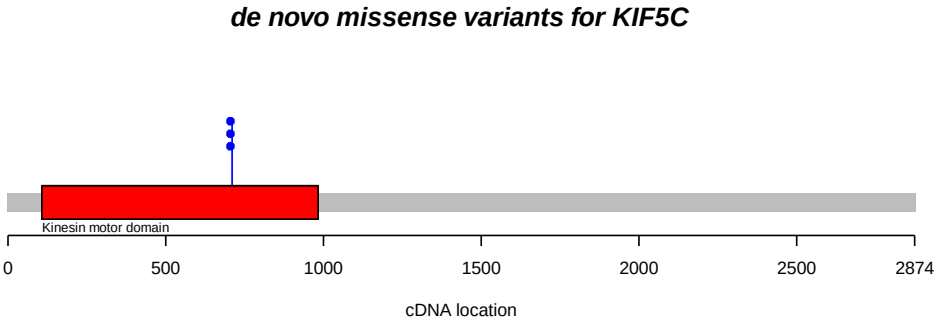

Figure S9 Schematic representation of *KIF5C* (ENST00000435030). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

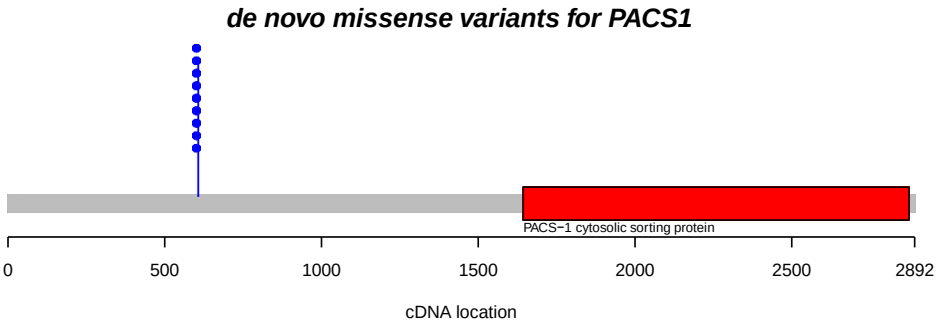

Figure S10 Schematic representation of *PACS1* (ENST00000320580). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

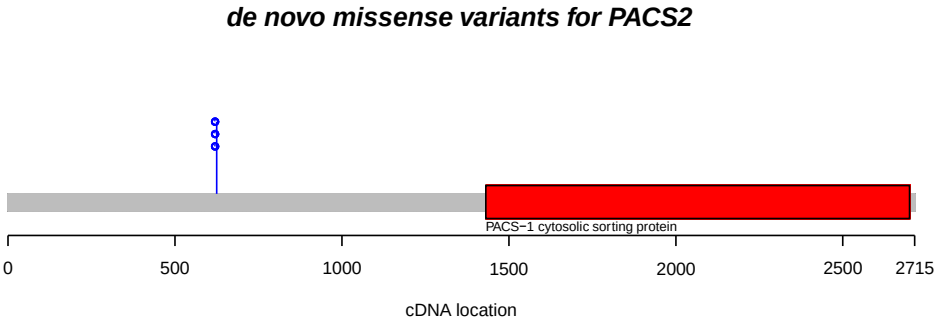

Figure S11 Schematic representation of *PACS2* (ENST00000458164). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

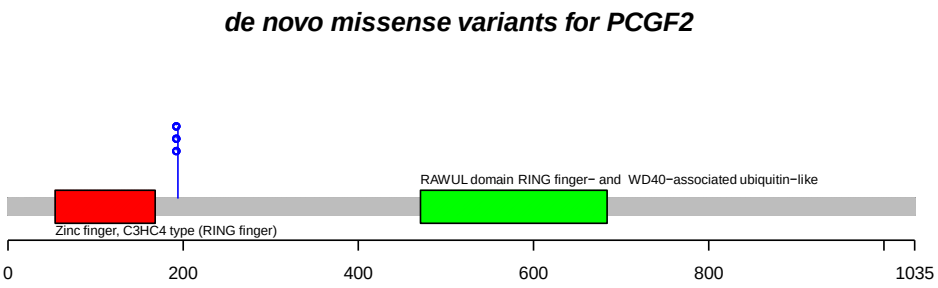

Figure S12 Schematic representation of *PCGF2* (ENST00000360797). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

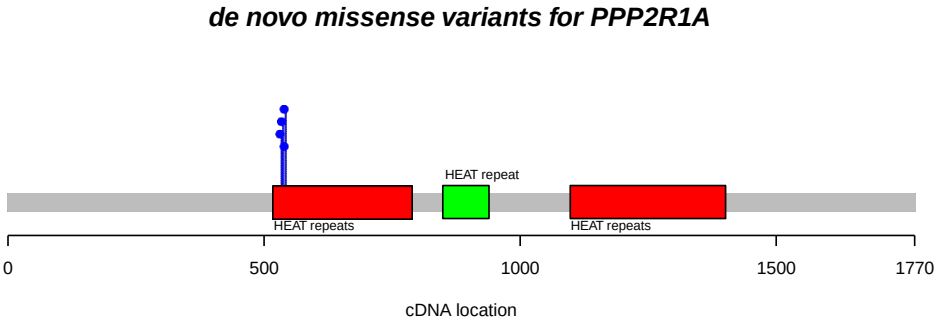

Figure S13 Schematic representation of *PPP2R1A* (ENST00000322088). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

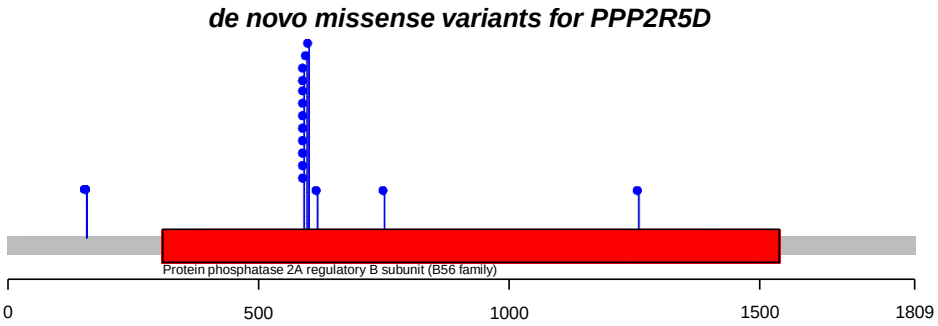

Figure S14 Schematic representation of *PPP2R5D* (ENST00000485511). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

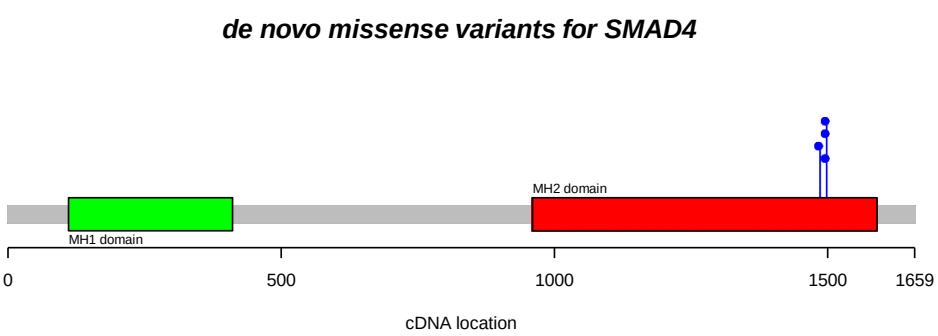

Figure S15 Schematic representation of *SMAD4* (ENST00000398417). The locations of the missense de novo variants are indicated by blue pins. Recurrent de novo missense mutations are indicated by stacked blue pins. Protein domains are annotated based on Pfam HMM search[1] and are illustrated by colored squares.

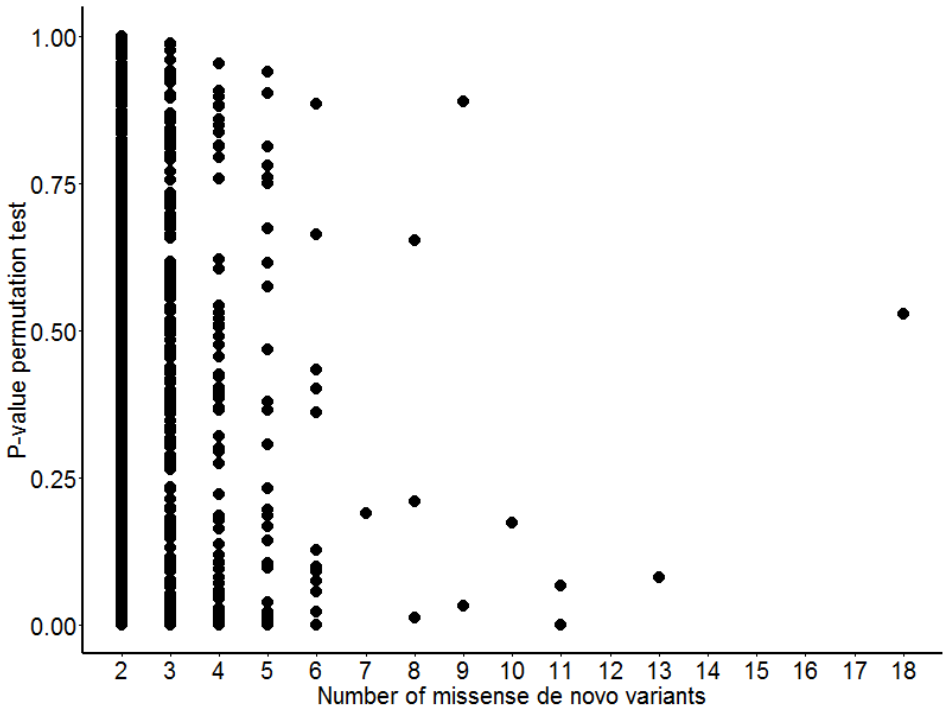

Figure S16. number of missense variants and the extend of clustering. We have studied the possibility of a correlation between the number of missense variants and the extend of clustering. Hereto, we have analysed a much larger set of 6,154 de novo missense variants from various patient cohorts present in the denovo-db version 1.3[2] (other than the five selected ID/DD cohorts presented in this manuscript). The variants were filtered as described in Table S1. Correlation analysis (Kendall rank correlation coefficient) between the number of de novo missense variants per gene (x-axis) and the corresponding p-values based on the spatial clustering signal (y-axis) yielded a correlation coefficient of -0.04 indicating there is no correlation.

Supplemental Tables

Column name	Number of variants remaining	
	ID and DD set	Control set
Unfiltered	9,770	35,632
Variants in coding regions and canonical splice-sites (±2 bp)	9,202	2,749
ExAC number of heterozygous variants less than 2 and no homozygous variants.	6,495	1,961
Variant identified in exome of genome sequencing study	6,495	1,948

Table S1. Filtering of de novo mutations used in study. Overview of filters that were applied to specific columns of the annotated de novo variants used in the analysis (left column). The middle and right columns indicate the number of variants left after the filtering for the ID and DD set and control set respectively.

Study	Trios	Disorder	Coding	Missense	LoF
McRae *et al.*[3]	4,293	DD	5,375	3,375	961
Lelieveld *et al.* [4a]	820	ID	948	579	177
de Ligt *et al.*[5]	100	ID	56	35	10
Rauch *et al.*[6]	51	ID	90	54	31
Halvardson *et al.*[7]	38	ID/EE	26	18	4
Total	5,302		6,495	4,061	1,183

Table S2. Dataset composition of intellectual disability and developmental disorders cohort (ID + DD). Columns indicate (from left to right), the study reference, number of trios included from the study, the disorder that was studied (DD= Developmental disorders, ID = Intellectual disability and EE = Epileptic Encephalopathies), and the number of coding, missense and loss-of-function de novo mutations after filtering. [a]Including 100 de novo mutations (median GATK quality score of 241) not included in the original publication.

Study	Trios	Disorder	Coding	Missense	LoF
Iossifov *et al.*[8]	1.786	Sibs. (ASD)	1,267	805	139
Krumm *et al.*[9a]			323	220	30
GONL[10]	250	Control	102	67	5
Turner *et al.*[11]	43	Sibs. (ASD)	28	18	1
Gulsuner *et al.*[12]	84	Sibs. (SCZ)	50	28	9
Besenbacher *et al.*[13b]	283	Control	178	135	15
Total	2,448		1,948	1,273	199

Table S3. Dataset composition of cohort of healthy controls (Control). Columns (from left to right) indicate the reference to the studies, number of trios that were included from the study, the disorder that was studied (Sibs. = siblings, ASD = Autism spectrum disorder and SCZ = Schizophrenia), the number of coding, missense and loss-of-function de novo mutations that were included from the study. [a]The study by Krumm *et al*[9]*.* performed a re-analysis on existing data of Iossifov *et al.*[8] [b]Variants in the study of Besenbacher *et al.*[13] are annotated to reference genome HG18. We used LiftOver to convert the HG18 coordinates to HG19 coordinates (See .web resources).

Gene name	Gene ID	Miss. DNMs	Avg. distance	p-value	Adj. p-value
SYNE1	ENST00000367255.5	2	36	0.0027	1
TMPRSS15	ENST00000284885.3	2	9	0.0061	1
MAPK8	ENST00000374189.1	2	6	0.0101	1
FAT4	ENST00000394329.3	2	86	0.0116	1
SPOCK3	ENST00000357154.3	2	14	0.0220	1
PRPS2	ENST00000398491.2	2	15	0.0299	1

Table S4. Results from clustering analysis on the control cohort. Columns from left to right indicate the gene name, used GENCODE gene identifier, number of de novo missense variants, average distance between de novo variants, permutations test based p-value and Bonferroni corrected p-value. None of the genes reach statistical significance after multiple testing correction.

Table S5. List of known genes involved in intellectual disability and developmental disorders

External file: table_S5_known_DD_genes.xls

Table S6. List of identified clustering mutations

External file: table_S6_dnm_clustering_genes.xls

	Genes with sig. cluster	Other genes	**total**
Known ID/DD genes	12	187	199
Other genes	3	381	384
total	15	568	583

Table S7A. Enrichment of genes have previously been implicated in ID/DD.

Data used to calculate statistical enrichment of genes associated to ID/DD based on the associated ID/DD genes on the combined list of the DDG2P and RUMC. The dataset contained in total 583 genes that were recurrently mutated of which 15 genes contained clustering de novo variants. 13 of the 15 genes were associated to ID/DD. Enrichment was tested with a two-sided Fisher's Exact test and yielded a significant p-value of 3.09E-04.

	Genes with sig. cluster	Other genes	**total**
Known ID/DD genes	11	162	173
Other genes	4	406	410
total	15	568	583

Table S7B. Enrichment of genes have previously been implicated in ID/DD solely on RUMC diagnostis list.

Data used to calculate statistical enrichment of genes associated to ID/DD based on the associated ID/DD genes solely on the RUMC list (see web resources). Enrichment was tested with a two-sided Fisher's Exact test and yielded a significant p-value of 5.13E-04.

	Genes with sig. cluster	Other genes	**total**
Known ID/DD genes	8	153	161
Other genes	7	415	422
total	15	568	583

Table S7C. Enrichment of genes have previously been implicated in ID/DD excluding two large DDD exome studies.

To completely remove the weight of the DDD exomes we have extended the enrichment test by excluding a set of genes based on the following criteria:

1. Novel genes enriched for de novo mutations found in the two large scale exome data studies of the DDD[3; 14]
2. Complement to the genes found enriched for de novo variants in Lelieveld *et al.* [4]

Analysis via a two-sided Fisher's Exact test yielded a significant P-value of 3.68E-02

Gene name	De novo missense	p-value	Adj. p-value
ACTL6B	3	8.48E-04	1
ALG13	3	3.66E-03	1
CDK13	12	4.34E-14	7.93E-10
COL4A3BP	6	1.28E-07	2.33E-03
GABBR2	3	6.82E-03	1
GRIN2B	11	4.22E-11	7.71E-07
KCNH1	7	1.36E-07	2.49E-03
KCNQ2	20	3.60E-28	6.59E-24
KIF5C	3	3.22E-03	1
PACS1	9	1.95E-10	3.57E-06
PACS2	3	5.88E-03	1
PCGF2	3	2.99E-04	1
PPP2R1A	4	9.88E-05	1
PPP2R5D	16	2.89E-24	5.28E-20
SMAD4	4	4.40E-05	8.04E-01

Table S8. Statistical significance of enrichment for de novo mutations for all identified genes. P-values are calculated based on Gene Specific Mutation Rates from Samocha *et al.*[15] and corrected for multiple testing by the Bonferroni correction (for 19,280 genes). Only some of the previously known genes reached statistical significance. Genes identified by our clustering analysis would not have been identified as candidate genes based on a statistical approach for the enrichment of de novo mutations.

Table S9. Known functions of identified genes

External file: table_S9_known_function_of_identified_genes.xlsx

Table S10: Overview of de novo variants found per individual

External file: table_S10_overview_all_DNM_per_sample_clustering_genes.xlsx

Table S11. Overview of literature supporting a NHI mechanism

External file: table_S11_overview_literature_study_known_NHI_genes.xlsx

Table S12. List of non-haploinsufficient genes involved in intellectual disability and developmental disorders

External file: table_S12_List_of_NHI_genes_involved_in_id_dd.xlsx

Table S13. List of haploinsufficient genes involved in intellectual disability and developmental disorders

External file: table_S13_List_of_HI_genes_involved_in_id_dd.xlsx

Gene name	Allelic requirement	mutation consequence
ALG13[a]	x-linked dominant	all missense/in frame
CDK13	monoallelic	all missense/in frame
COL4A3BP	monoallelic	activating
GRIN2B[a]	monoallelic	loss of function / all missense/in frame
KCNH1	monoallelic	activating
KCNQ2	monoallelic	loss of function
KIF5C	monoallelic	all missense/in frame
PACS1	monoallelic	activating
PCGF2	monoallelic	activating
PPP2R1A	monoallelic	dominant negative
PPP2R5D	monoallelic	dominant negative
SMAD4[a]	monoallelic	loss of function / all missense/in frame

Table S14. Identified known genes and their mutational consequence. Allelic requirement and mutational consequence according to DDG2P. [a]In the statistical analysis we excluded genes for reasons mentioned in the main text .

	Hi genes	NHI genes	Total
Genes with significant cluster	1	8	9
Other genes	182	108	290
Total	183	116	299

Table S15. Enrichment of genes based on their mutational consequence. Statistical enrichment testing of disease mechanisms of the genes identified in this study. Of the 15 identified genes, 9 are annotated with a disease mechanism after filtering based on the DDG2P and RUMC lists. Analysis via Fisher's Exact test yielded a significant P-value of 2.66E-03.

Table S16. Tolerance scores based on dn/ds ratios for all human genes

External file: table_S16_Dn_ds_ratios_for_all_human_genes.xls

Table S17. HI and NHI DNMs analysed to protein structures

External file: table_S17_DNMs_in_protein_structures.xlsx

Supplemental web resources

DDG2P: https://www.ebi.ac.uk/gene2phenotype/downloads

RUMC Genome diagnostics gene list:

https://www2.radboudumc.nl/Informatievoorverwijzers/Genoomdiagnostiek/en/Pages/Intellectualdisability.aspx

Lift Over tool: https://genome.ucsc.edu/cgi-bin/hgLiftOver

Supplemental References

1. Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44, D279-285.
2. Turner, T.N., Yi, Q., Krumm, N., Huddleston, J., Hoekzema, K., HA, F.S., Doebley, A.L., Bernier, R.A., Nickerson, D.A., and Eichler, E.E. (2017). denovo-db: a compendium of human de novo variants. Nucleic Acids Res 45, D804-D811.
3. Deciphering Developmental Disorders Study. (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature.
4. Lelieveld, S.H., Reijnders, M.R., Pfundt, R., Yntema, H.G., Kamsteeg, E.J., de Vries, P., de Vries, B.B., Willemsen, M.H., Kleefstra, T., Lohner, K., et al. (2016). Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. Nature neuroscience 19, 1194-1196.
5. de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. The New England journal of medicine 367, 1921-1929.
6. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., et al. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 380, 1674-1682.
7. Halvardson, J., Zhao, J.J., Zaghlool, A., Wentzel, C., Georgii-Hemming, P., Mansson, E., Ederth Savmarker, H., Brandberg, G., Soussi Zander, C., Thuresson, A.C., et al. (2016). Mutations in HECW2 are associated with intellectual disability and epilepsy. Journal of medical genetics 53, 697-704.
8. Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014).

The contribution of de novo coding mutations to autism spectrum disorder. Nature 515, 216-221.
9. Krumm, N., Turner, T.N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., Raja, A., Coe, B.P., Stessman, H.A., He, Z.X., et al. (2015). Excess of rare, inherited truncating mutations in autism. Nature genetics 47, 582-588.
10. Genome of the Netherlands, C. (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nature genetics 46, 818-825.
11. Turner, T.N., Hormozdiari, F., Duyzend, M.H., McClymont, S.A., Hook, P.W., Iossifov, I., Raja, A., Baker, C., Hoekzema, K., Stessman, H.A., et al. (2016). Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. American journal of human genetics 98, 58-74.
12. Gulsuner, S., Walsh, T., Watts, A.C., Lee, M.K., Thornton, A.M., Casadei, S., Rippey, C., Shahin, H., Consortium on the Genetics of, S., Group, P.S., et al. (2013). Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell 154, 518-529.
13. Besenbacher, S., Sulem, P., Helgason, A., Helgason, H., Kristjansson, H., Jonasdottir, A., Jonasdottir, A., Magnusson, O.T., Thorsteinsdottir, U., Masson, G., et al. (2016). Multi-nucleotide de novo Mutations in Humans. PLoS genetics 12, e1006315.
14. Deciphering Developmental Disorders Study. (2015). Large-scale discovery of novel genetic causes of developmental disorders. Nature 519, 223-228.
15. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnstrom, K., Mallick, S., Kirby, A., et al. (2014). A framework for the interpretation of de novo mutation in human disease. Nature genetics 46, 944-950.