Supplementary data:

Whole exome sequencing identifies an important role of loss of function de novo mutations in severe human neural tube defects

Philippe Lemay[1], Marie-Claude Guyot[1], Élizabeth Tremblay[1], Alexandre Dionne-Laporte[2], Dan Spiegelman[2], Édouard Henrion[2], Ousmane Diallo[2], Patrizia De Marco[3], Elisa Merello[3], Christine Massicotte[1], Valérie Désilets[1], Jacques L. Michaud[1], Guy A. Rouleau[2], Valeria Capra[3], Zoha Kibar[1,*]

[1] CHU Ste-Justine Research Center, Université de Montréal, Montréal, Québec H3T 1C5 Canada ; [2] Montreal Neurological Institute, McGill University, Montréal, Québec H3A 2B4,Canada [3] Istituto Giannina Gaslini, 16148, Genoa, Italy

*Corresponding Author at University of Montréal, Department of Neurosciences, CHU Sainte Justine Research Center, 3175 Cote-Ste-Catherine, Room 5999C, Montreal, QC H3T 1C5 Canada; Tel: (514) 345 4931 X 3984; Fax: (514) 345 4801; E-mail : zoha.kibar@ recherche-ste-justine.qc.ca

Material and Methods

Ethics statement

All samples were collected with the approval of the Local Ethics Committees: CHU Sainte Justine Hospital (Protocols' numbers: 2598 and 2899) and Istituto Giannina Gaslini, Genoa, Italy (Protocol number: 213/2013). Written informed consent was obtained from all participating individuals.

Cohort

Forty-three families each composed of one affected child and two unaffected parents with no family history of NTDs were recruited through the Montreal Ste-Justine Hospital Spina Bifida Center, the 3D study of the Integrated Research Network in Perinatology of Quebec and Eastern Ontario and the Istituto Giannina Gaslini in Genoa, Italy. Detailed information including folate status, tissue of origin and type of NTDs of this cohort is summarized in Table S1. Further description of the families with discussed findings can be found in Table S2. Briefly, all 43 cases were affected with NTDs including 35 myelomeningocele and 8 anencephaly cases. A total of 21 cases were fetuses and 55.6 % took folate peri-conceptionally. Tissues from fetuses were all obtained following induced abortions. The average maternal and paternal ages were 30.0 ± 4.8 years and 30.7 ± 5.9 years respectively. Saliva or blood samples were obtained from affected patients and their parents. Liver biopsies or umbilical cord tissues were obtained from aborted fetuses.

Alignment and SNV annotation

Whole exome sequencing (WES) on genomic DNA was performed according to the manufacturer's procedures at the McGill University and Genome Quebec Innovation Center (MUGQIC). Capture was done using SureSelectXT Human All Exon V4 or the SureSelect Human All Exon 50Mb. Sequencing was done on Hiseq 2000 as pair-end with 3 samples per lane. Raw data were aligned to the GRCh37 build using Burrows-Wheeler Aligner (BWA)[1]. Duplicated reads were removed and the rest were locally realigned followed by variant calling using GATK 2.6.4.[2] [3] [4] Variants were annotated using ANNOVAR (http://www.openbioinformatics.org/annovar/). Coding regions defined by refseq (33 499 779bp) were covered in average by 132 reads with 90% of target covered 20 times or more. Identified DNMs were annotated with the following bioinformatics tools: CADD score [5], Polyphen HDIV [6] [7] [8], RVIS rank [9] and HI score [10].

Statistical Analysis

To allow an accurate comparison of the LoF *de novo* rate in our sample to the control in the published data, the number of LoF *de novo* were divided by the number of coding bases (33 499 779bp) covered at least with 20 or more reads. This per-base LoF *de novo* rate was then compared to the published control rate for two LoF DNM rates of 0.05×10^{-8} and 0.17×10^{-8}.[11] [12] The rate of LoF DNMs in *SHROOM3* was obtained by multiplying the two per-base *de novo* rate with the length of *SHROOM3* long isoform (5991bp). The gene specific per-trio LoF *de novo* rate from Samocha KE et al., Nat genet, 2014[13] was obtained by combining the rate of splicing, nonsense and frameshift mutations in *SHROOM3*. Significance for differences between those rates and the observed number of mutations was assessed using a two tailed binomial exact test performed using R V3.0.1.[14] Bonferroni correction was defined using the Ensembl gene

annotation (http://aug2014.archive.ensembl.org/Homo_sapiens/Info/Annotation #assembly). A total of 20 389 genes has been annotated in the GRCh38 suggesting an acceptation threshold of $0.05 / 20\ 389 = 2.45 \times 10^{-6}$.

References:

1 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl* 2009;**25**:1754–60.

2 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;**20**:1297–303.

3 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;**43**:491–8.

4 Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In: *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc. 2002. http://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi1110s43/abstract (accessed 23 May2014).

5 Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;**46**:310–5.

6 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248–9.

7 Liu X, Jian X, Boerwinkle E. dbNSFP: A Lightweight Database of Human Nonsynonymous SNPs and Their Functional Predictions. *Hum Mutat* 2011;**32**:894–9.

8 Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: A Database of Human Non-synonymous SNVs and Their Functional Predictions and Annotations. *Hum Mutat* 2013;**34**:E2393–402.

9 Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* 2013;**9**:e1003709.

10 Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and Predicting Haploinsufficiency in the Human Genome. *PLoS Genet* 2010;**6**:e1001154.

11 Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Günel M, Roeder K, Geschwind DH, Devlin B, State MW. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012;**485**:237–41.

12 Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, Smith JD, Paeper B, Nickerson DA, Dea J, Dong S, Gonzalez LE, Mandell JD, Mane SM, Murtha MT, Sullivan CA, Walker MF, Waqar Z, Wei L, Willsey AJ, Yamrom B, Lee Y-H, Grabowska E, Dalkic E, Wang Z, Marks S, Andrews P, Leotta A, Kendall J, Hakker I, Rosenbaum J, Ma B, Rodgers L, Troge J, Narzisi G, Yoon S, Schatz MC, Ye K, McCombie WR, Shendure J, Eichler EE, State MW, Wigler M. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* Published Online First: 29 October 2014. doi:10.1038/nature13908

13 Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki JA, Rehnström K, Mallick S, Kirby A, Wall DP, MacArthur DG, Gabriel SB, DePristo M, Purcell SM, Palotie A, Boerwinkle E, Buxbaum JD, Cook Jr EH, Gibbs RA, Schellenberg GD, Sutcliffe JS, Devlin B, Roeder K, Neale BM, Daly MJ. A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014;**46**:944–50.

14 R Development Core Team. The R Project for Statistical Computing. http://www.r-project.org./ (accessed 14 Aug2013).