



LETTER TO THE EDITOR OPEN

Threshold for neural tube defect risk by accumulated singleton loss-of-function variants

Cell Research (2018) 0:1–3; <https://doi.org/10.1038/s41422-018-0061-3>

Dear Editor,

Neural tube defects (NTDs) are a class of major structural malformations affecting the brain and spinal cord. They are among the most common congenital anomalies with a worldwide prevalence of 0.1%.^{1,2} Elucidating the genetic basis of their complex etiology has eluded our best efforts to date. Although there are more than 400 genes capable of producing an NTD phenotype when mutated in the mouse,^{3,4} studies of human candidate genes based on mouse NTD genes have not been informative, except for genes in the planar cell polarity pathway.⁵ Recently, an omnigenic model of inheritance was proposed for complex traits, suggesting that the associated signals tend to spread across almost the entire genome.⁶ In light of this new perspective on the genomic architecture of complex traits, we re-evaluated whole-genome sequencing (WGS) data from three different NTD cohorts (Han Chinese, Caucasian USA, and Middle Eastern/Qatar).

We initially evaluated our Chinese NTD cohort (100 cases, primarily anencephaly, Supplementary information, Table S1) and used the 1000 Genomes Project (1KGP)⁷ as controls (208 Chinese Han). Due to the limited sample size of the NTD cohort and uneven coverage of 1KGP sequences (higher coverage in coding regions), only rare (MAF < 0.01) protein-coding variants in NTDs and controls were selected for functional prediction. The selected deleterious missense (D-mis) and loss-of-function (LoF) variants were further compared with 1KGP and ExAC databases (MAF_{1KGP} < 0.001 and MAF_{ExAC} < 0.001). Although D-mis and LoF variants are considered more likely to be causative variants, we failed to observe more rare D-mis variants in our Chinese NTD cases compared to the 1KGP controls (one-sided Wilcoxon test, $P = 1$; Fig. 1a).

We also did not observe significant enrichment of damaging variants in human orthologs of 249 mouse NTD-associated genes⁴ (χ^2 test, $P = 0.48$; Supplementary information, Table S2). By contrast, when only the rare LoF variants were examined, there are significantly more such variants in NTD cases (median = 17) than in controls (median = 12; two-sided Wilcoxon test, $P = 1.2 \times 10^{-14}$; Fig. 1b). This suggested that rare LoF variants statistically correlate well with NTDs.

We wanted to validate our observations in different cohorts; however, when we examined WGS data from the 1KGP, we found that the number of rare LoF variants varies among different populations. However, the number of singleton LoF variants (SLoFVs), which are those LoF variants that appeared only once when compared to entire 1KGP data, is very similar among different populations.⁷ This suggests that the number of SLoFVs is a stable and reliable genomic indicator of NTD risk in humans. We re-examined the SLoFVs in our Chinese cohort and found that the median number of SLoFVs per NTD case is 15, whereas the medium number of SLoFVs is 6 in controls (two-sided Wilcoxon test, $P = 1.7 \times 10^{-31}$; Fig. 1c). We went on to examine the SLoFVs in our US and Middle Eastern NTD cohorts. Seventy-four US spina

bifida samples were compared to 99 Caucasians from the 1KGP. Again, a statistically significant difference was detected, with 11.5 SLoFVs in NTDs and 5 SLoFVs in controls (two-sided Wilcoxon test, $P = 6.4 \times 10^{-14}$; Fig. 1d). The WGS data from a Middle Eastern cohort consisting of 69 spina bifida samples and 108 matched controls (no matched controls in 1KGP) were also examined, and a statistically significant difference was detected, with 7 SLoFVs in NTDs and 5 SLoFVs in controls (two-sided Wilcoxon test, $P = 2.7 \times 10^{-6}$; Fig. 1e). When the three NTD cohorts (243 cases) were combined and compared to 2,504 controls from the 1KGP, the median number of SLoFVs was 11.5 in NTDs and 5 in controls (two-sided Wilcoxon test, $P = 3.3 \times 10^{-61}$, Fig. 1f; Binomial test, $P = 5.97 \times 10^{-237}$, Supplementary information, Table S3). Both results reached a Bonferroni-corrected threshold based on correction for testing of $\sim 20,389$ genes ($P < 2.5 \times 10^{-6}$).

Further comparison between the two major subtypes of NTDs, anencephaly (from Chinese cohort) vs. spina bifida (from US cohort) demonstrated that anencephalic cases carried more SLoFVs than spina bifida cases (15 vs. 9, two-sided Wilcoxon test, $P = 1.7 \times 10^{-13}$; Fig. 1g). This suggested that the more severe the subtype of NTDs, the more SLoFVs it likely contains. Therefore, we calculated the odds ratios (ORs) of NTDs with different numbers of SLoFVs. Our results demonstrated that there is a threshold SLoFV number for NTD risk. When the number of SLoFVs reaches 9, the OR for NTD is >1 [OR = 3.3 (1.8–6.0)] (Fig. 1h). When an individual carries more SLoFVs, the risk for NTD increases exponentially, e.g., OR = 4.8 (2.7–8.3) for 10 SLoFVs, OR = 6.6 (2.9–13.8) for 12 SLoFVs, and OR = 43.1 (18.8–107.6) for 14 SLoFVs.

Both gene and pathway distributions of those SLoFVs identified in NTD cases from Chinese Han and US CEU cohorts were analyzed to determine whether there is any specific enrichment. Each SLoFV was mapped to KEGG pathways; the frequency of pathways with mutations, based on the percentage of individual NTD cases carrying mutations in each pathway, was calculated. There are 14 pathways containing SLoFVs in > 10% of the NTD samples (Supplementary information, Fig. S1), including the PI3K-AKT signaling pathway, tight junction, and focal adhesion pathways. Some of these pathways, such as tight junctions, were also found to differ significantly at the transcriptome level between NTD and control fetal cells isolated from amniotic fluid⁸ (Supplementary information, Fig. S2). However, there is no pathway in NTDs that is significantly enriched with SLoFVs after Bonferroni correction (Supplementary information, Table S4). In fact, the expression of over 20,000 genes was detected in NTD fetal cells from amniotic fluid,⁸ which suggests that most of the genome is transcriptionally active during neural tube development.

We have now demonstrated that the genetic basis for NTD risk can be measured by the accumulation of SLoFVs in one's genome. Approximately nine SLoFVs is a genomic threshold for NTD risk, regardless of their genetic background or ethnicity. Together with

Received: 26 January 2018 Revised: 3 May 2018 Accepted: 11 June 2018
Published online: 05 July 2018

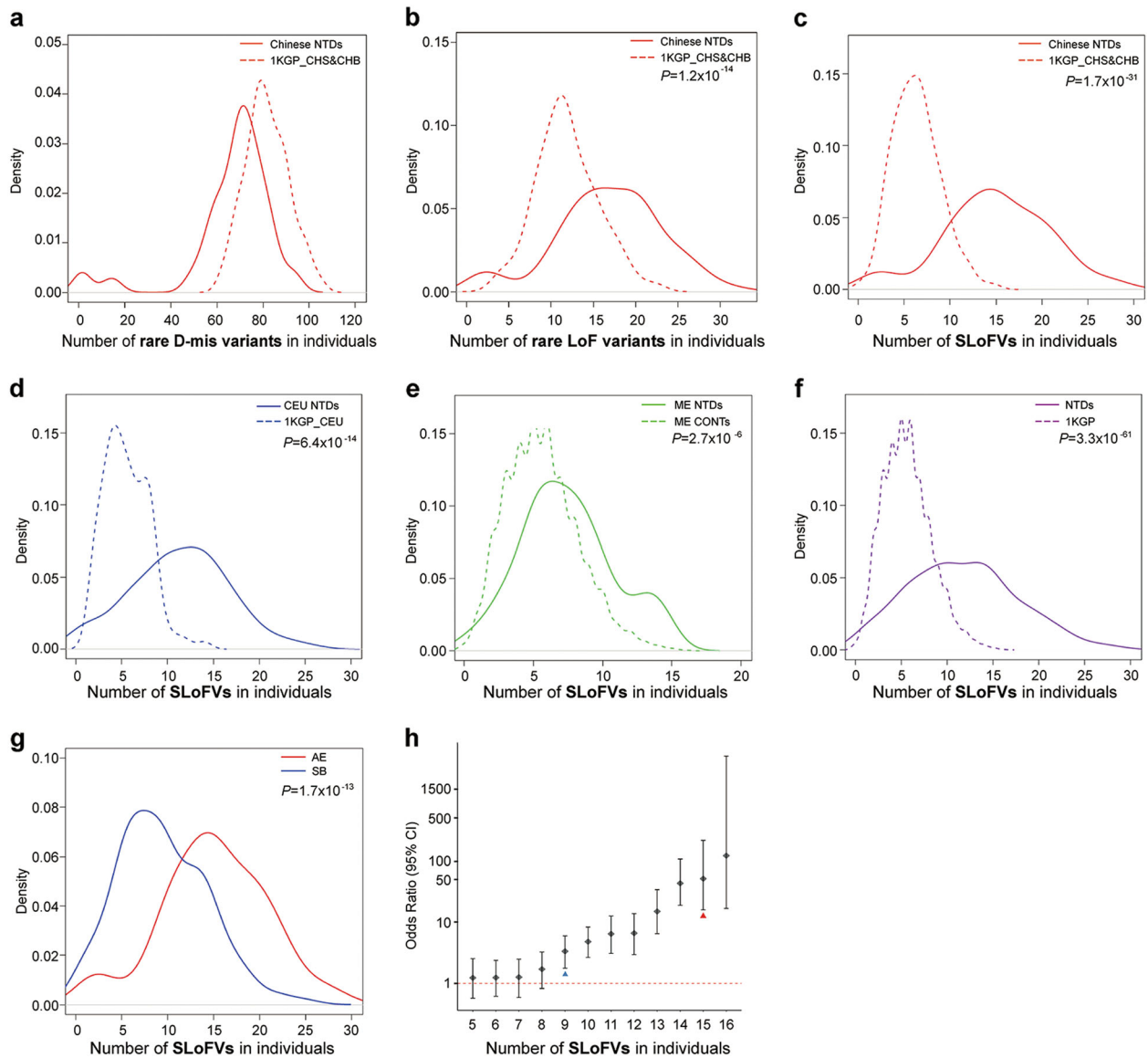


Fig. 1 **a** Chinese NTD cases do not carry more rare D-mis variants than controls from the 1KGP. **b** The distribution of rare LoF variants shows significantly more LoF variants for Chinese NTDs than controls. **c–e** Significantly more SLoFVs in NTD samples than their matched controls for Chinese (**c**), Caucasian in USA (**d**), and Middle Eastern (ME) cohorts (**e**). **f** Combined data shows significantly more SLoFVs for NTDs than controls from 1KGP. **g** Significantly more SLoFVs were found in anencephalic (AE) than spina bifida (SB) samples. **h** Odds ratios of SLoFVs in NTDs. The blue arrowhead represents the median number of SLoFVs in SB and the red arrowhead represents the median number of SLoFVs in AE. The density (y-axis) at each point is the average contribution from each of the kernels at that point

the finding that no single pathway was enriched for SLoFVs in NTD cases, our study suggests that almost all genes have some minor impact on the etiology of NTDs. In summary, our data demonstrated how a complex reproductive disadvantage disease like NTDs can benefit from the omnigenic model.⁶ This enhances our understanding of the genomic architecture underlying susceptibility to these severe congenital malformations.

Materials and Methods are available in Supplementary information, Data S1.

ACKNOWLEDGEMENTS

We thank the families for their participation in this study. Data analysis and computing resource were supported by Center for Big Data Research in Health, Institute of Biophysics, Chinese Academy of Sciences. This project was supported by the National Key Basic Research Program of China (2016YFC1000502), the National Natural Science Foundation of China (81430005, 31521003, 31771669, 81741048), and the

Commission for Science and Technology of Shanghai Municipality (17JC1400902) to H.W. and Y.Z.; and NIH (HD081216, HD083809, and HD067244) to R.H.F. and M.E.R.

ADDITIONAL INFORMATION

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41422-018-0061-3>.

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Zhongzhong Chen^{1,2}, Yunping Lei^{3,4}, Yufang Zheng^{1,2,5}, Vanessa Aguiar-Pulido⁶, M. Elizabeth Ross⁶, Rui Peng¹, Li Jin^{1,2}, Ting Zhang⁷, Richard H. Finnell^{3,4} and Hongyan Wang^{1,2,8,9}
¹Obstetrics and Gynecology Hospital, State Key Laboratory of Genetic Engineering at School of Life Sciences, Institute of Reproduction and Development, Fudan University, Shanghai 200011, China; ²Key

Laboratory of Reproduction Regulation of NPFPC, Collaborative Innovation Center of Genetics and Development, Fudan University, Shanghai 200032, China; ³Departments of Molecular and Cellular Biology and Medicine, Baylor College of Medicine, Houston, TX 77030, USA; ⁴Collaborative Innovation Center for Genetics & Development, School of Life Sciences, Fudan University, Shanghai 200438, China; ⁵Institute of Developmental Biology & Molecular Medicine, Fudan University, Shanghai 200433, China; ⁶Center for Neurogenetics, Weill Cornell Medicine, New York, NY 10021, USA; ⁷Capital Institute of Pediatrics, Beijing 100020, China; ⁸Children's Hospital, Fudan University, 399 Wanyuan Road, Shanghai 201102, China and ⁹Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China

These authors contributed equally: Zhongzhong Chen, Yunping Lei, and Yufang Zheng

Correspondence: Ting Zhang (zhangtingcv@126.com) or Richard H. Finnell (Richard.Finnell@bcm.edu) or Hongyan Wang (wanghy@fudan.edu.cn)

REFERENCES

1. Copp, A. J. et al. *Nat. Rev. Dis. Prim.* **1**, 15007 (2015).
2. Wallingford, J. B., Niswander, L. A., Shaw, G. M. & Finnell, R. H. *Science* **339**, 1222002 (2013).

3. Eppig, J. T. et al. *Nucleic Acids Res.* **33**, D471–D475 (2005).
4. Harris, M. J. & Juriloff, D. M. *Birth Defects Res. A Clin. Mol. Teratol.* **88**, 653–669 (2010).
5. Juriloff, D. M. & Harris, M. J. *Birth Defects Res. A Clin. Mol. Teratol.* **94**, 824–840 (2012).
6. Boyle, E. A., Li, Y. I. & Pritchard, J. K. *Cell* **169**, 1177–1186 (2017).
7. Genomes Project, C. et al. *Nature* **526**, 68–74 (2015).
8. Nagy, G. R. et al. *Clin. Chem.* **52**, 2013–2020 (2006).



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

© The Author(s) 2018