

Contribution of rare and non-coding genetic variants to Gilles de la Tourette syndrome

Malgorzata Borczyk^{1,*}, Jakub P. Fichna^{2,3,*}, Marcin Piechota¹, Sławomir Gołda¹, Michał Korostyński¹, Piotr Janik⁴, Cezary Żekanowski²

*These authors contributed equally

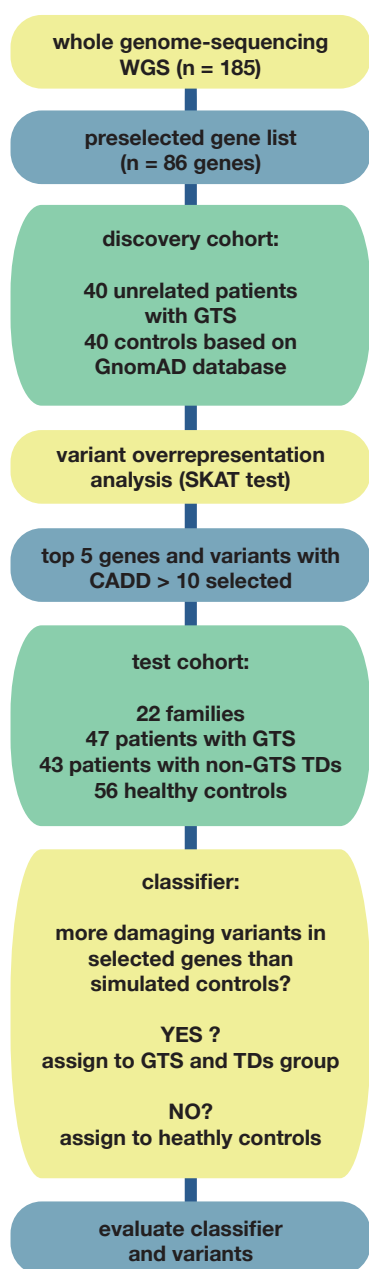
1. Laboratory of Pharmacogenomics, Department of Molecular Neuropharmacology, Maj Institute of Pharmacology, Polish Academy of Sciences, Krakow, Poland
2. Laboratory of Neurogenetics, Department of Neurodegenerative Disorders, Mossakowski Medical Research Institute, Polish Academy of Sciences, Warsaw, Poland
3. Department of Biological Sciences, Purdue University, West Lafayette, Indiana, USA
4. Department of Neurology, Medical University of Warsaw, Warsaw, Poland

presenting author: Malgorzata Borczyk, gosborcz@if-pan.krakow.pl

Background and research hypothesis

Gilles de la Tourette syndrome (GTS) and other Tic Disorders (TDs) have a substantial genetic component with the heritability estimated at between 60 and 80% (Davis et al., 2013; Mataix-Cols et al., 2015). It is now evident that no single gene is responsible for a large fraction of GTS cases, albeit rare variants with large effects have been considered causative in single GTS families (Castellan Baldan et al., 2014). We propose that a substantial part of GTS cases could be explained by an oligogenic model which assumes a compound effect of multiple low- and medium-impact variants with varied population allele frequencies. This hypothesis is supported by exomic data showing that de novo damaging variants in approximately 400 genes contribute to GTS risk in 12% of clinical cases (Wilsey et al., 2017). GWAS results suggest that as much as 21% of GTS heritability can be explained by genotypes with a MAF (minor allele frequency) between 0.001 and 0.05 (Davis et al., 2013). Recently, a whole-exome sequencing study indicated a role of the rare variant burden in 13 families with TD history (Cao et al., 2021). Still, the role of rare, and particularly non-coding variants, particularly with MAF < 0.001 remains largely unexplored for GTS and other TDs.

Methodology and approach



Results

We investigated whether an oligogenic additive model could be used to distinguish healthy individuals from GTS and other TDs subjects. Whole-genome sequencing (WGS) was used to analyze known and identify novel variants in genes previously indicated with varying levels of evidence, to be associated with GTS and other TDs. The search window encompassed also 20,000 bp of both flanks of the gene to identify variants located both in the gene itself and in distant regulatory regions as well. The first phase of analysis was conducted to select candidate genes that best differentiate unrelated GTS patients (discovery group) from the general population. Variants in five genes (HDC, CHADL, MAOA, NAA11, and PCDH10) were then used to build a model assigning individuals from GTS risk families to a group with clinical symptoms (GTS and other TDs) or to a healthy group, with the AUC-ROC (area under the receiver operating characteristic curve) metric of 0.6 ($p < 0.00001$) (Figure 1). The model includes 98 variants putatively associated with GTS and other TDs risks, including 75 non-coding variants with the median allele frequency of ~0.04 (Figure 2).

Conclusions

This is, to our knowledge, the first statistical model using WGS data, including non-coding and rare variants, to predict GTS/TD risk. Overall, the presented approach provides a promising path for further studies of the genomic basis of GTS. The obtained results support the concept that the additive effect of putatively deleterious variants in a small subset of key genes is a substantial risk factor of GTS. We have validated results currently available in the literature and identified a range of rare non-coding variants not previously associated with GTS that could contribute to its etiopathology. The ability of the classifier to distinguish GTS-affected from healthy individuals within families is of particular importance, as the availability of a burden test for affected families could be highly beneficial in genetic counseling. Although the clinical utility of the presented model is limited, it provides an insight into the variant burden associated with familial as well as sporadic GTS. Further WGS studies of substantially larger groups and including an extended panel of genes should provide an even better tool for oligogenic GTS risk prediction.

Figure 1. Comparison of the performance of tested classifiers of GTS risk. Graphs show ROCs of the chosen GTS risk model (orange line) and other tested classifiers. A) Comparison of diverse classifiers based on top genes considered with different CADD thresholds (5,10,15,20) and a different number of genes (2 - 5). The thick orange line represents the best classifier taken as a GTS risk model, with CADD threshold of 10 and five genes. Points A and B have the highest percentage of correct assignments. B) Best classifier from A compared with 478 control classifiers (blue lines) based on sets of random genes. x-axes: fraction of false-positive results; y-axes: fraction of true positive results.

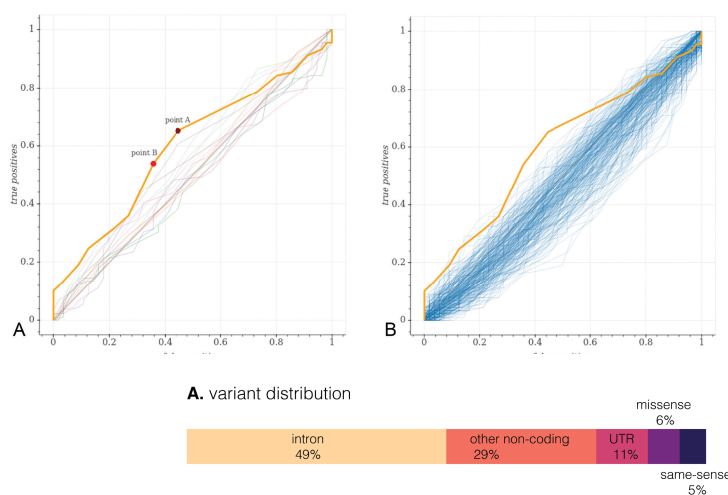
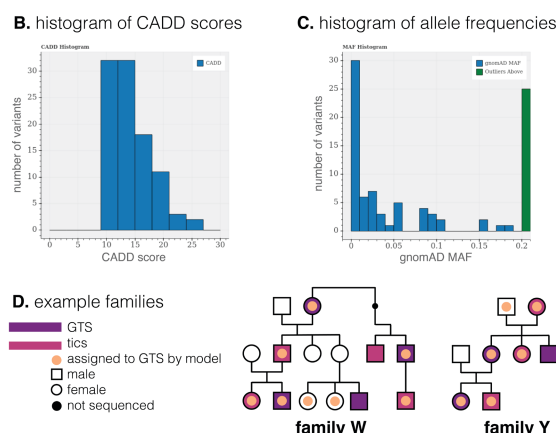


Figure 2. Characteristics of genetic variants included in the GTS risk model and example assignments. A. Localization of variants in genes and flanking regions; B. Histogram of CADD scores of 98 variants included in the risk model; The CADD score is scaled non-linearly in a Phred scale, where a score greater or equal to 10 indicates that a given variant is predicted to be within the top 10% of most deleterious variants substitutions possible in the human genome, whereas a score greater or equal 20 corresponds to the 1% of the most deleterious variants (Rentzsch et al., 2019) C. Distribution allele frequencies (MAF) of 98 chosen variants in non-Finnish Europeans in the gnomAD database; D. An example of classification of members of two families by the GTS risk model. Orange dots indicate individuals assigned to the GTS and other TDs group by the risk model;



Funding

This work was supported by the National Science Center, Poland (NCN) project UMO-2016/23/B/NZ2/03030