Identification of high and low penetrance genomic variants in Autism Spectrum Disorders from Next Generation Sequencing data

Giovanni Spirito[1], Diego Vozzi[2], Aldamaria Puliti[4,5], Maria Cerminara[5], Martina Servetti[4,5], Margherita Lerone[4], Maria Teresa Divizia[4], Giulia Rosti[5], Livia Pisciotta[5,6], Irene Serio[6], Lino Nobili[5,6], Stefano Gustincich[1,2], Remo Sanges[1,2,3]

1. Area of Neuroscience, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy.
2. Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Napoli, Italy.
3. Department of Neuroscience and Brain Technologies, Italian Institute of Technologies (IIT), Genova, Italy.
4. Medical Genetics Unit, IRCCS Giannina Gaslini, Genova, Italy
5. Dept. of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINOGMI), University of Genoa, Genoa, Italy
6. Child Neuropsychiatry Unit, IRCCS Giannina Gaslini, Genova, Italy

Correspondence to: gspirito@sissa.it

1 Introduction

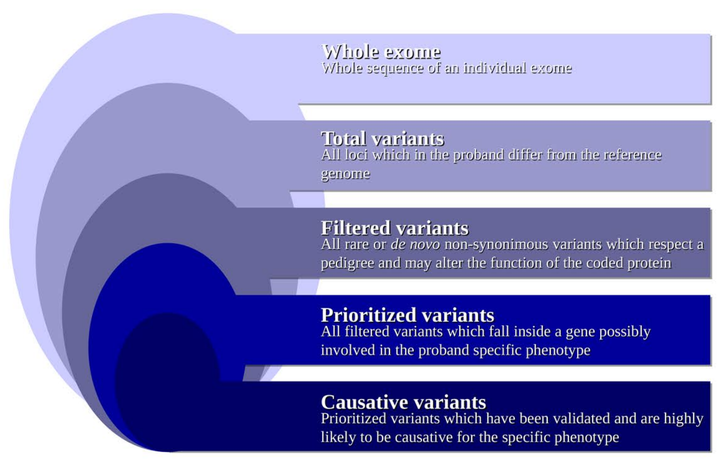
Whole Exome Sequencing in diagnostics, a collaborative effort

Next generation sequencing (NGS) technologies enabled the extensive study of the genomics underlying human diseases. Namely whole exome sequencing (WES) represents a cost-efficient method which can lead to the detection of genomic variants and the discovery of novel disease-associated genes. This is particularly important in the case of common complex diseases such as autism spectrum disorders (ASD), whose genetic etiology is still poorly understood. We built a custom computational framework capable, from raw WES data, to automatically detect genomic variants and prioritize them in regards to their relevance to a specific phenotype. We tested this framework on a small cohort of 10 trios all featuring a proband affected by sporadic ASD.
Our pipeline was built in the context of a joint lab born thanks to a collaboration between IIT and the Gaslini pediatric hospital (GE). The pipeline is systematically upgraded upon counseling and feedback from the hospital.

2 Project aim

Identification of deleterious variants

One of the main drawbacks of the WES approach is the large number of genomic variants detected in each analysis, which is by no means manually manageable. Automated variant prioritization strategies are therefore required. Our method allows to automatically prioritize all detected variants according to multiple factors, and therefore highlight those variants which are more likely to be causative for the specific phenotype. Our results include a number of subsets of variants, which are manually evaluated and in some cases validated *in vitro* by the hospital.



3 Our pipeline

1. **Whole exome sequencing**
Samples were sequenced using the Illumina NovaSeq 6000 System in Genova (IIT)

2. **Variant calling**
All samples underwent the same variant discovery analysis, the pipeline was built following the GATK4 best practices

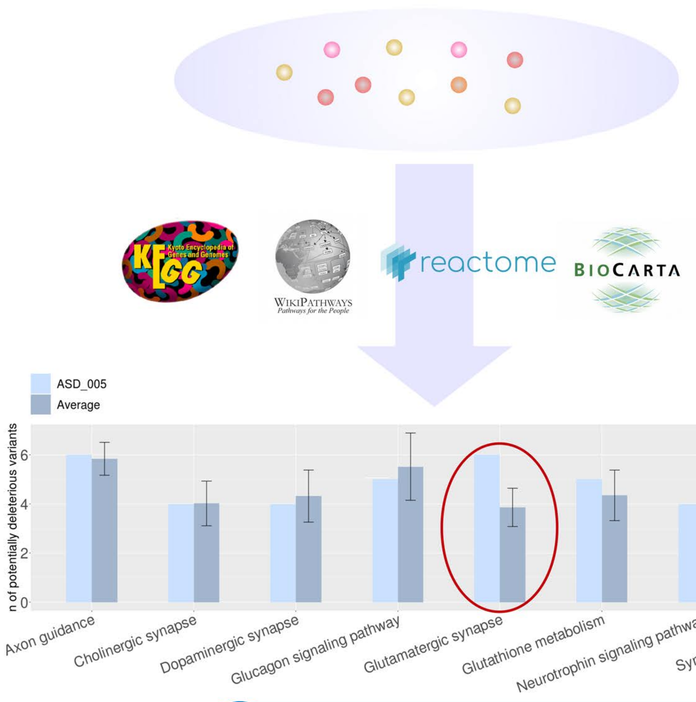
3. **Variant Filtering**
All variants detected underwent a series of filters applied in order to select the Possibly deleterious variants

4. **Gene prioritization**
All candidate variants must also fall within a gene whose function relates to the specific proband phenotype

5. **Variant validation**
Several variants suggested by our pipeline were validated and considered phenotype-causative.

1 Pathway enrichment analysis

All genes affected by possibly deleterious variants are parsed through a number of canonical pathways databases in order to pinpoint ASD-associated pathways with an increased variant burden



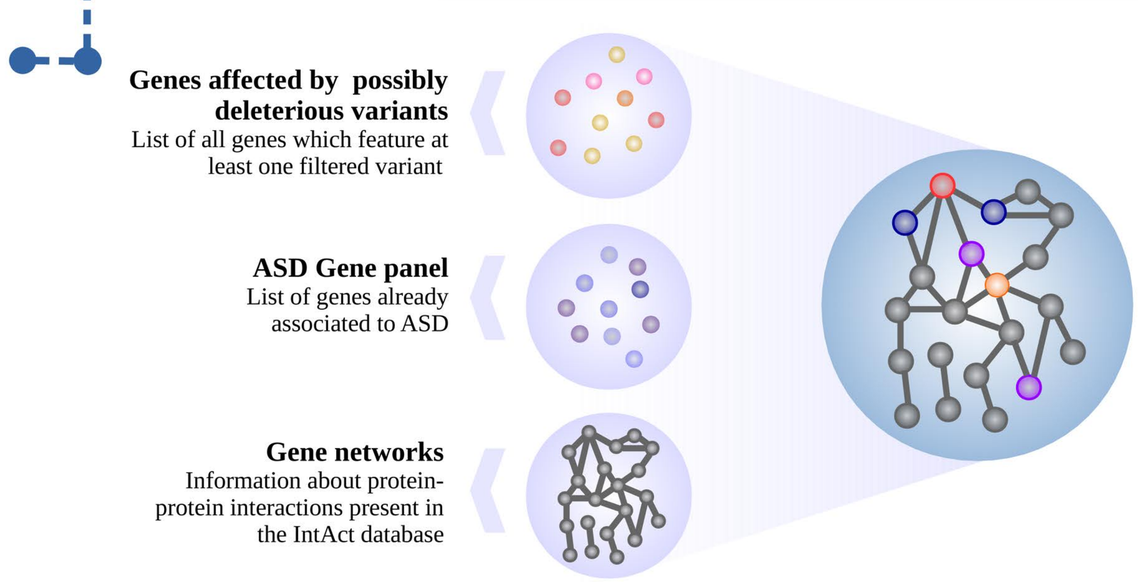
Genes affected by possibly deleterious variants
List of all genes which feature at least one filtered variant

Pathways databases
List of public databases containing information about canonical pathways

Pathways impaired in ASD
Manual selection of pathways commonly impaired in ASD and manual selection of pathways containing an increased burden of possibly deleterious variants compared to an in-house average

2 Gene network analysis

Genes featuring possibly deleterious variants are prioritized based on their putative interaction with genes present in an ASD panel



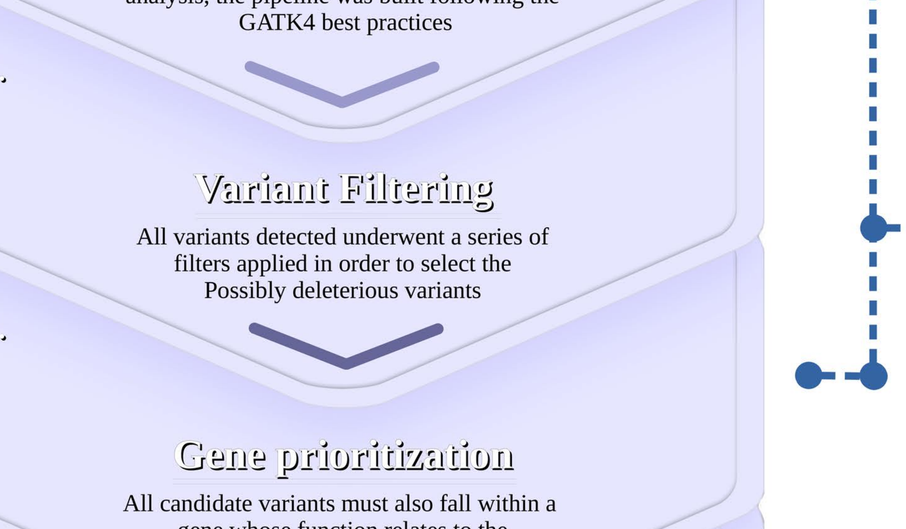
Genes affected by possibly deleterious variants
List of all genes which feature at least one filtered variant

ASD Gene panel
List of genes already associated to ASD

Gene networks
Information about protein-protein interactions present in the IntAct database

Prioritized genes
An annotation related to the likelihood of interaction with an ASD gene is added to each gene

4 Example result 1



chrX:53563401-53563401 C>T

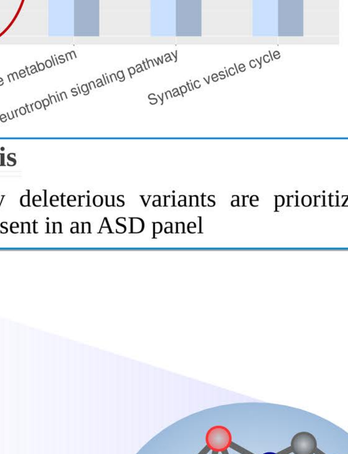
OMIM: Mental retardation, X-linked syndromic, Turner type 309590

Case description

Clinical trio composed by healthy parents and a male proband affected by a sporadic form of ASD. Thanks to our computational framework we were able to pinpoint an unannotated X-linked and possibly deleterious non-synonymous single nucleotide variant (SNV) embedded inside gene HUWE1. The presence of the variant and the predicted inheritance pattern were confirmed through Sanger sequencing.
HUWE1 is a E3 ubiquitin-protein ligase which mediates ubiquitination and subsequent proteasomal degradation of target proteins[1]. Regulates neural differentiation and proliferation by catalyzing the polyubiquitination and degradation of MYCN. Defects in HUWE1 are the cause of intellectual disabilities[2]. Recently, a *de novo* missense variant in HUWE1 was identified in a male ASD proband, but not in the proband's less severely affected brother[2].
It is therefore conceivable that the detected variant could is causative for the proband phenotype.

5 Example result 2

SHANK1	WT
SHANK2	WT/p.I1076T
SHANK3	WT
GRIA1	WT/p.F27L
GRID2	WT
GRIN2A	WT/p.N1076K

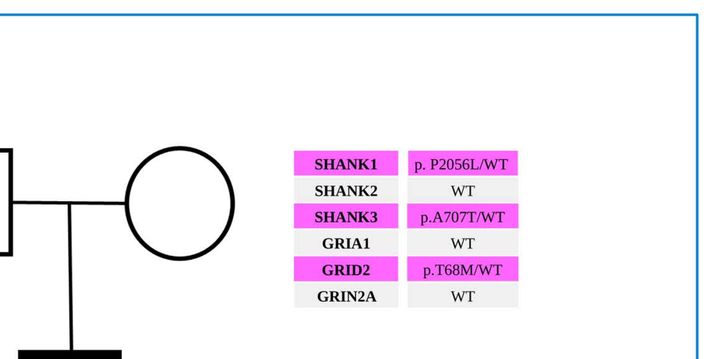


SHANK1	p. P2056L/WT
SHANK2	WT
SHANK3	p.A707T/WT
GRIA1	WT
GRID2	p.T68M/WT
GRIN2A	WT

SHANK1	p. P2056L/WT
SHANK2	WT/p.I1076T
SHANK3	p.A707T/WT
GRIA1	WT/p.F27L
GRID2	p.T68M/WT
GRIN2A	WT/p.N1076K

Case description

Clinical trio composed by healthy parents and a male proband affected by a sporadic form of ASD. In this case we were not able to pinpoint any putatively high-penetrance causative variant. However, thanks to our pathway enrichment analysis, we detected 6 uncommon heterozygous variants all inside genes strongly associated to ASD. While the impact of each of these variants is likely very mild, we believe that an increased burden of variants affecting genes essential for the physiology of the glutamatergic synapse signalling pathway could be detrimental to the point of being causative for the observed ASD phenotype.



Schematic representation of a portion of the glutamatergic synapse signalling pathway, customized from Kegg pathway[3]

6 Conclusions

Detection of low and high penetrance variants putatively involved in ASD

Our workflow helped us to detect and prioritize genomic variants from whole exome sequencing data. The analysis we performed in order to prioritize the detected variants according to their putative relevance to the proband phenotype allowed us to pinpoint both clinical cases were the root of the phenotype is a single high penetrance variant and cases were an increased burden of low penetrance inherited variants could concur to cause the phenotype. By analysing an extensive number of autism spectrum disorders cases, we believe that this workflow could help us to reconstruct the genomics at the basis of complex ASD phenotypes.

7 Future perspectives

Other kinds of variants

Currently the available computational tools allow for a reliable and standardized detection of only small variants such as SNPs and indels. Nonetheless other kinds of genomic variants exist and can be disease-causative. For example, both *de novo* and inherited copy number variants (CNVs) and short tandem repeats (STRs)[5] and transposable elements insertions[6] have been associated with disease[4]. We are working towards implementing the discovery of this three less studied classes of genomic variants into our computational pipeline.

References
1. Moortgat S, Berland S, Aukrust I, Maystadt I, Baker L, Benoit V, Caro-Llopis, A, Cooper NS, Debray FG, Faivre L, Gardeitchik T, Haukanes BI, Houge G, Kivuva E, Martinez F, Mehta SG, Nassogne MC, Powell-Hamilton N, Pfundt R, Rosello M, Prescott T, Vasudevan P, van Loon B, Verellen-Dumoulin C, Verloes A, Lippe CV, Wakeling E, Wilkie AOM, Wilson L, Yuen A, Study D, Low KJ, Newbury-Ecob RA. HUWE1 variants cause dominant X-linked intellectual disability: a clinical study of 21 patients. Eur J Hum Genet. 2018 Jan;26(1):64-74.
2. Nava C, Lamari F, Héron D, Mignot C, Rastetter A, Keren B, Cohen D, Faudet A, Bouteiller D, Gilleron M, Jacquette A, Whalen S, Afenjar A, Périsse D, Laurent C, Dupuits C, Gautier C, Gérard M, Huguet G, Caillet S, Leheup B, Leboyer M, Gillberg C, Delorme R, Bourgeron T, Brice A, Depienne C. Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE. Transl Psychiatry. 2012 Oct 23;2:e179.
3. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016 Jan 4;44(D1):D457-62. doi: 10.1093/nar/gkv1070.
4. Rev, A. & Hum, G. Copy Number Variation in Human Health, Disease, and Evolution. 451–481 (2015). doi:10.1146/annurev.genom.9.081307.164217.
5. Mousavi, N., Shleizer-burko, S. & Yanicky, R. Profiling the genome-wide landscape of tandem repeat expansions. *47*, 1–13 (2019).
6. Payer LM, Burns KH. Transposable elements in human genetic disease. Nat Rev Genet. 2019 Dec;20(12):760-772. doi: 10.1038/s41576-019-0165-8. Epub 2019 Sep 12. Review.