Coalitional game theory as a promising approach to identify candidate autism genes

Anika Gupta[1,2,*], Min Woo Sun[1,3,*], Kelley M. Paskov[1], Nate T. Stockham[1], Jae Y. Jung[1], and Dennis P. Wall[1]

[1]Stanford University, Palo Alto, CA; [2]Massachusetts Institute of Technology, Cambridge, MA;
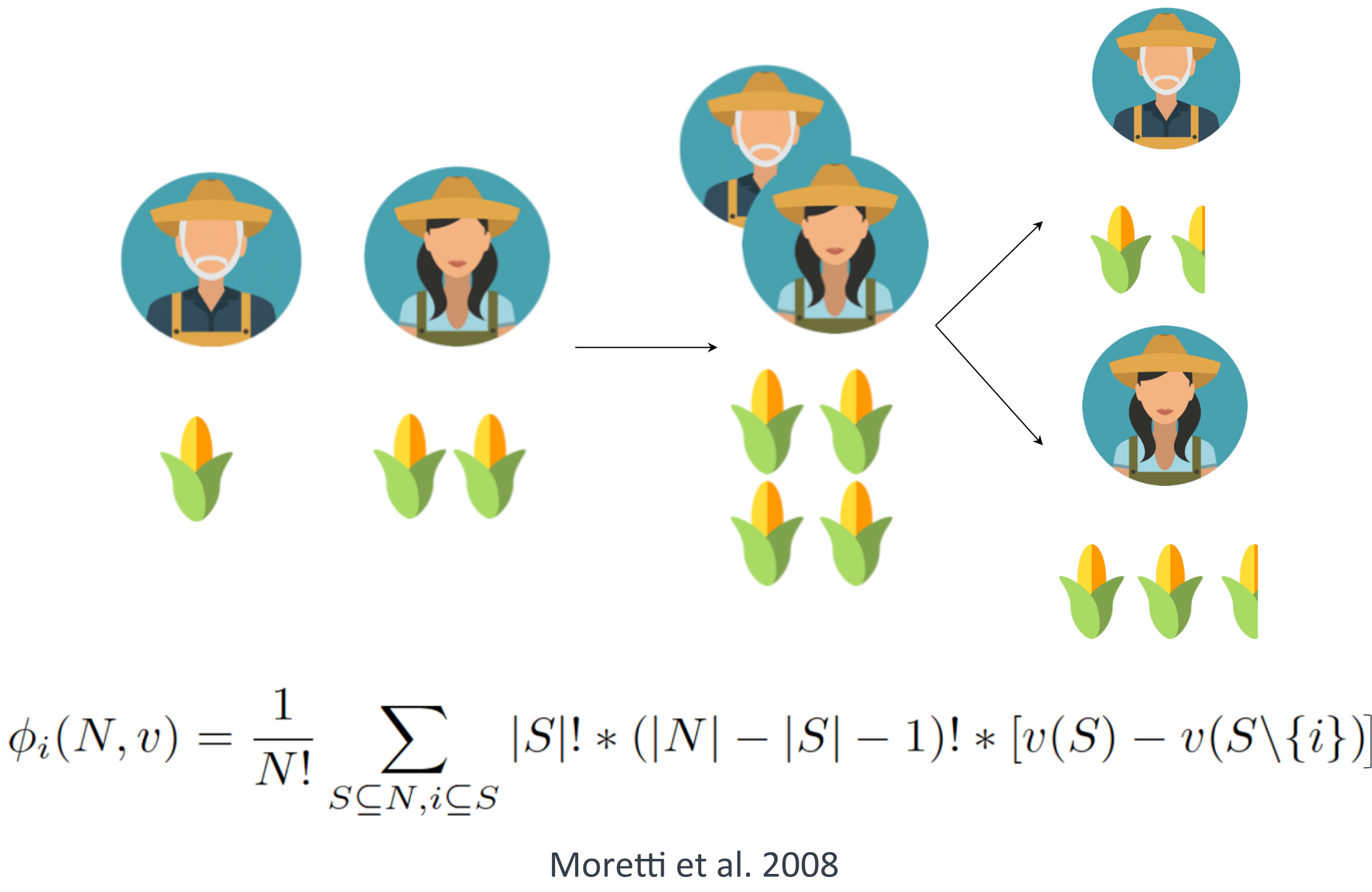[3]University of California Los Angeles, Los Angeles, CA

Abstract

Introduction: Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disease with strong genetic etiology. Despite an estimated 90% heritability of the disorder, only a small fraction of cases can be explained by known molecular causes[1], and typical approaches do not incorporate intergenic associations. Coalitional game theory (CGT) has been suggested as an enhanced signal detection method that takes into account the combinatorial role of groups of genes in a given condition and identifies genes with the greatest marginal contribution across all genes[2,3]. We propose that CGT can be effectively applied to gene alteration data to highlight candidate genes in ASD and other complex diseases with combinatorial molecular underpinnings.
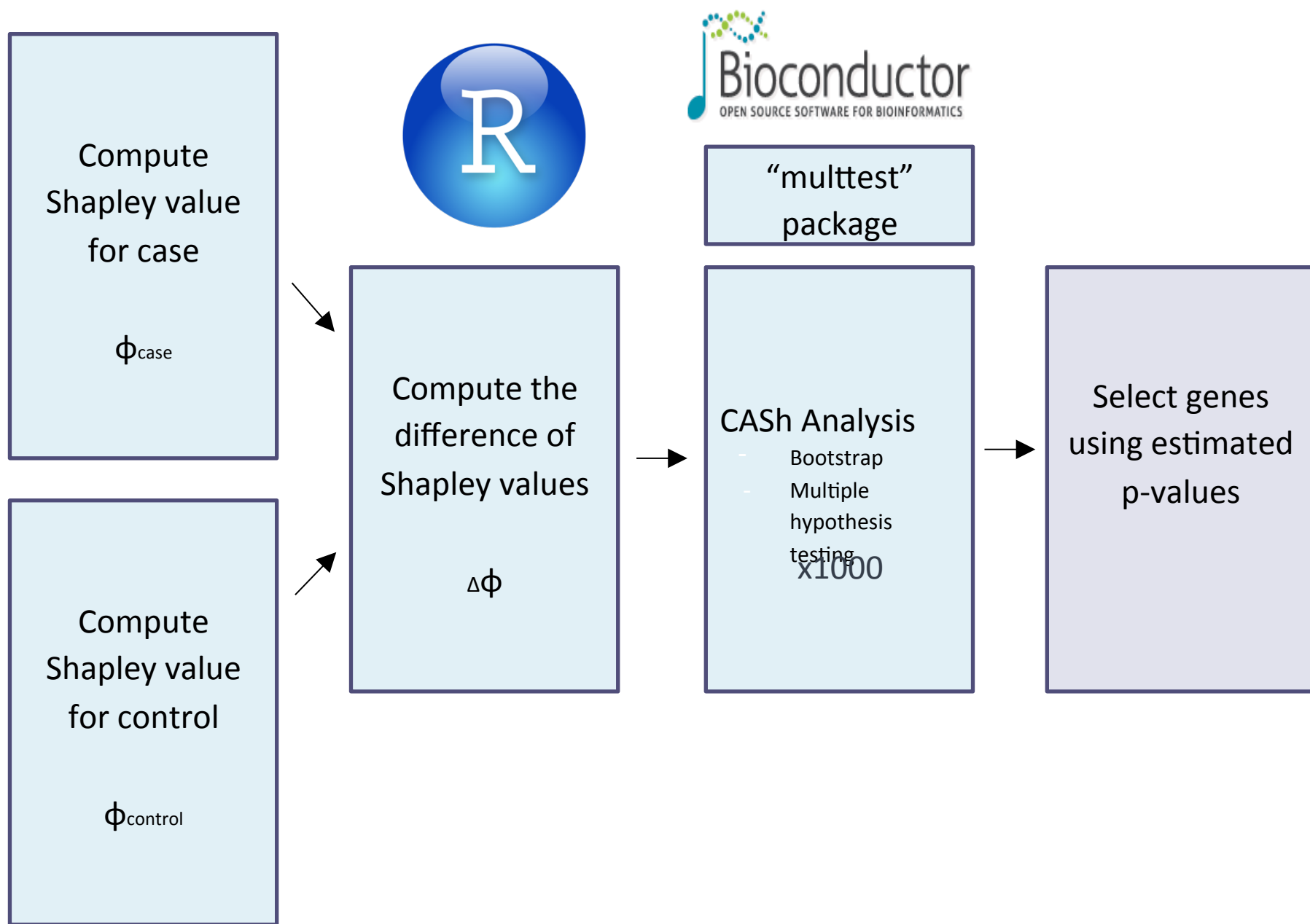
Methods: We analyzed 30x-coverage whole genome sequencing data for 2,710 fully sequenced genomes from the Hartwell Foundation's Autism Research and Technology Initiative (iHART). We encoded alterations into binary matrices for ASD (case) and unaffected (control) samples, indicating likely gene-disrupting, inherited mutations in altered genes. To determine individual gene contributions given an ASD phenotype, we calculated the Shapley value (a player metric, Φ), for each gene in the case and control cohorts (Figure 1). We performed Comparative Analysis of Shapley value (CASh), a resampling-based multiple hypothesis testing procedure and filtered for significant genes (p > 0.05).

Results: We identified genes (CGT genes) as key contributors in the genetic coordination of ASD using coalitional game theory, as determined by the difference in Shapley value between cases and controls. 208 genes showed statistical significance at the 0.05 significance level (unadj. p < 0.05), 136 genes at the 0.01 significance level (unadj. p < 0.01), and 8 genes passed the threshold for significance after correcting for multiple hypothesis testing (FDR<0.01). Rerunning CGT on randomly sampled cases and controls from each of the families confirmed that the family structure of the dataset did not confound the results. Cross-referencing CGT genes with high confidence ASD genes extracted the known biological functions represented by these candidate ASD genes. Five of the 8 highest confidence (FDR<0.01) CGT genes have protein products that directly interact with protein products of genes in the SFARI and Root 66 gene lists, as determined by the functional protein association networks tool STRING.

Methods

Data Preprocessing

GENOMES (965 genes)

*1,616 cases
349 controls

INDIVIDUALS* (1,965 people)

Coalitional game theory identifies key players

$$\phi_i(N,v) = \frac{1}{N!} \sum_{S \subseteq N, i \subseteq S} |S|! * (|N| - |S| - 1)! * [v(S) - v(S \setminus \{i\})]$$

Moretti et al. 2008

Coalitional game theory implementation

Results

Coalitional Game Theory Genes

*Protein interaction partners of known ASD genes
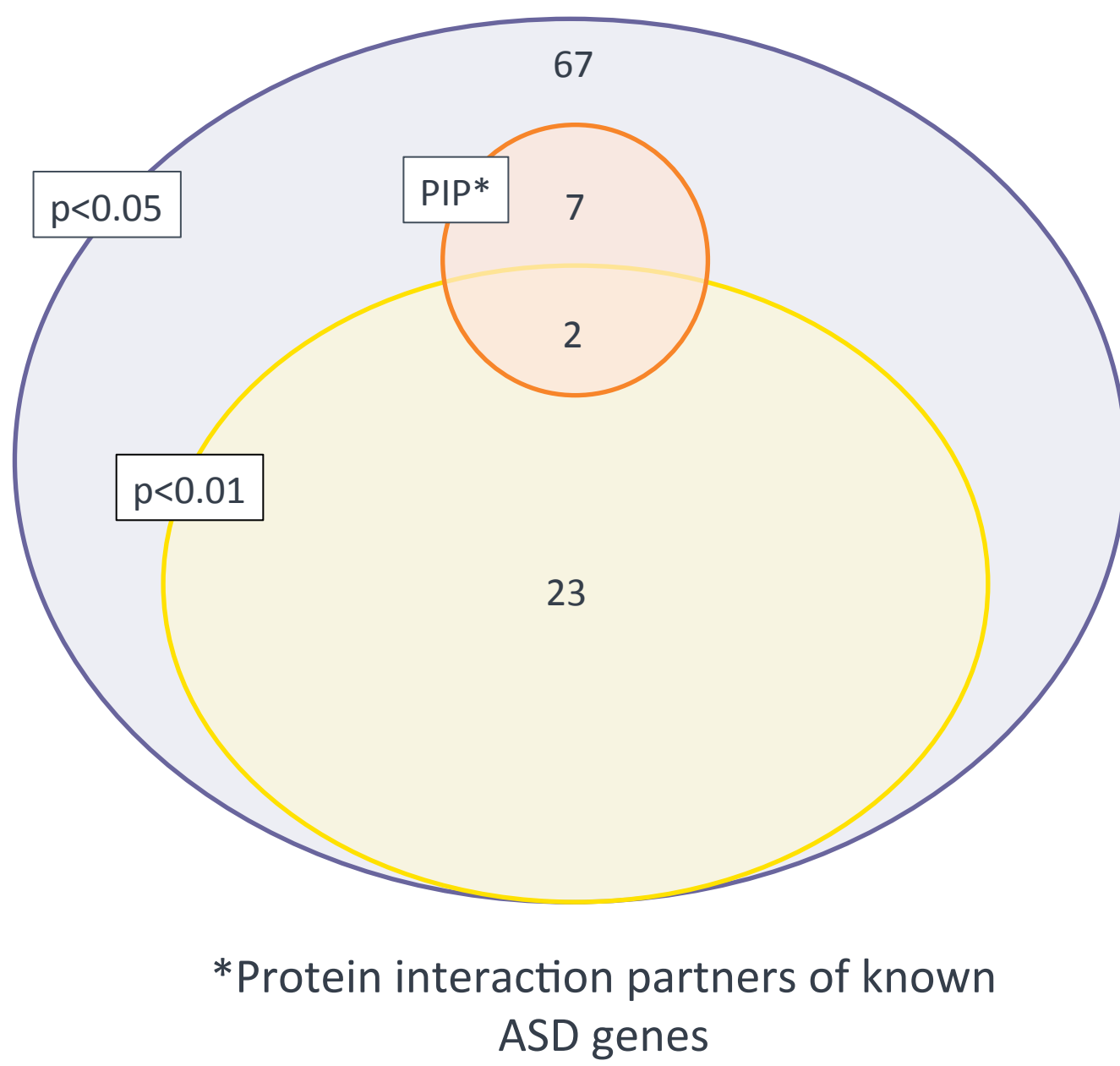
ASD-related Biological Functions Represented

*Pertaining to axon development
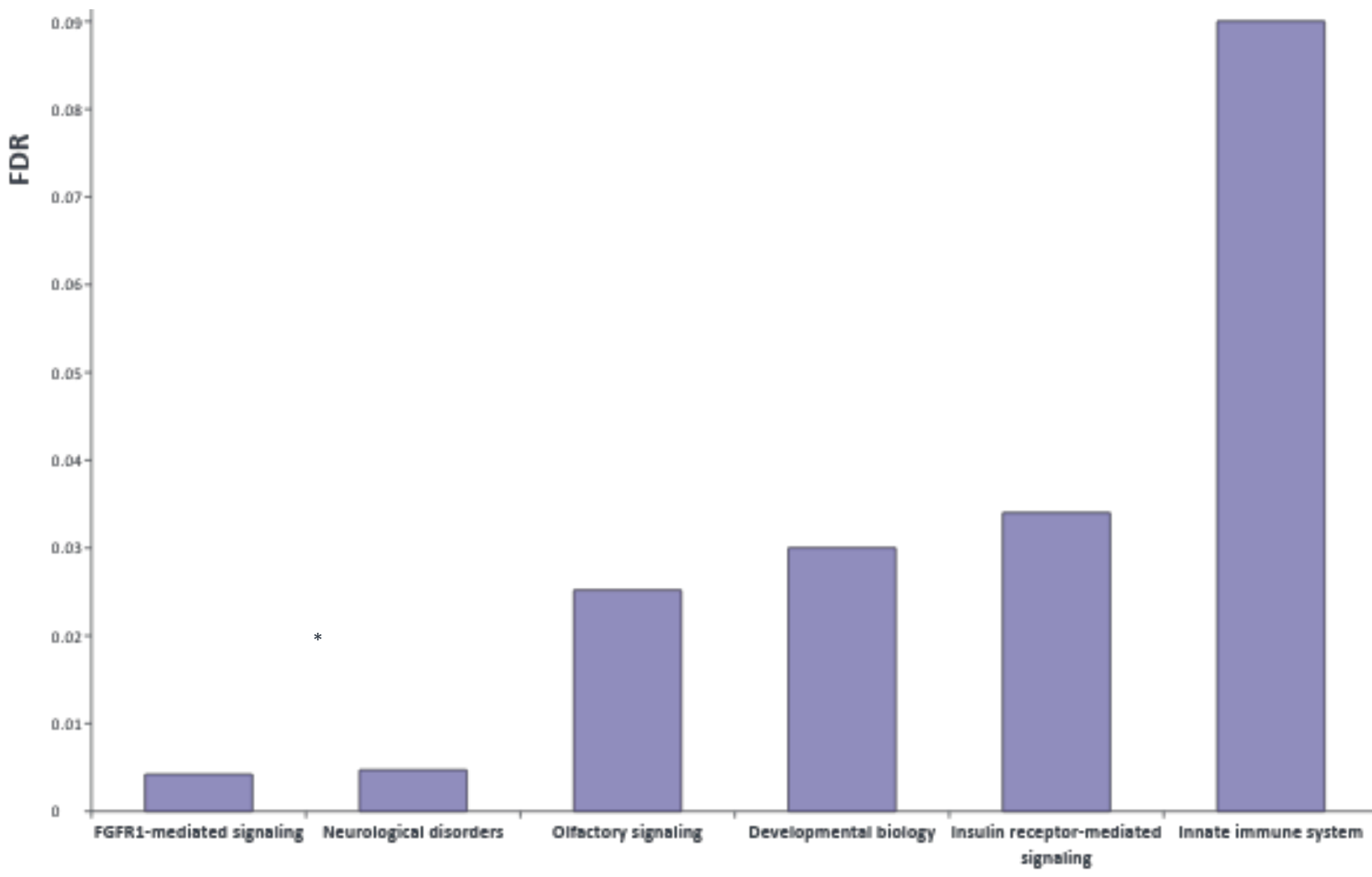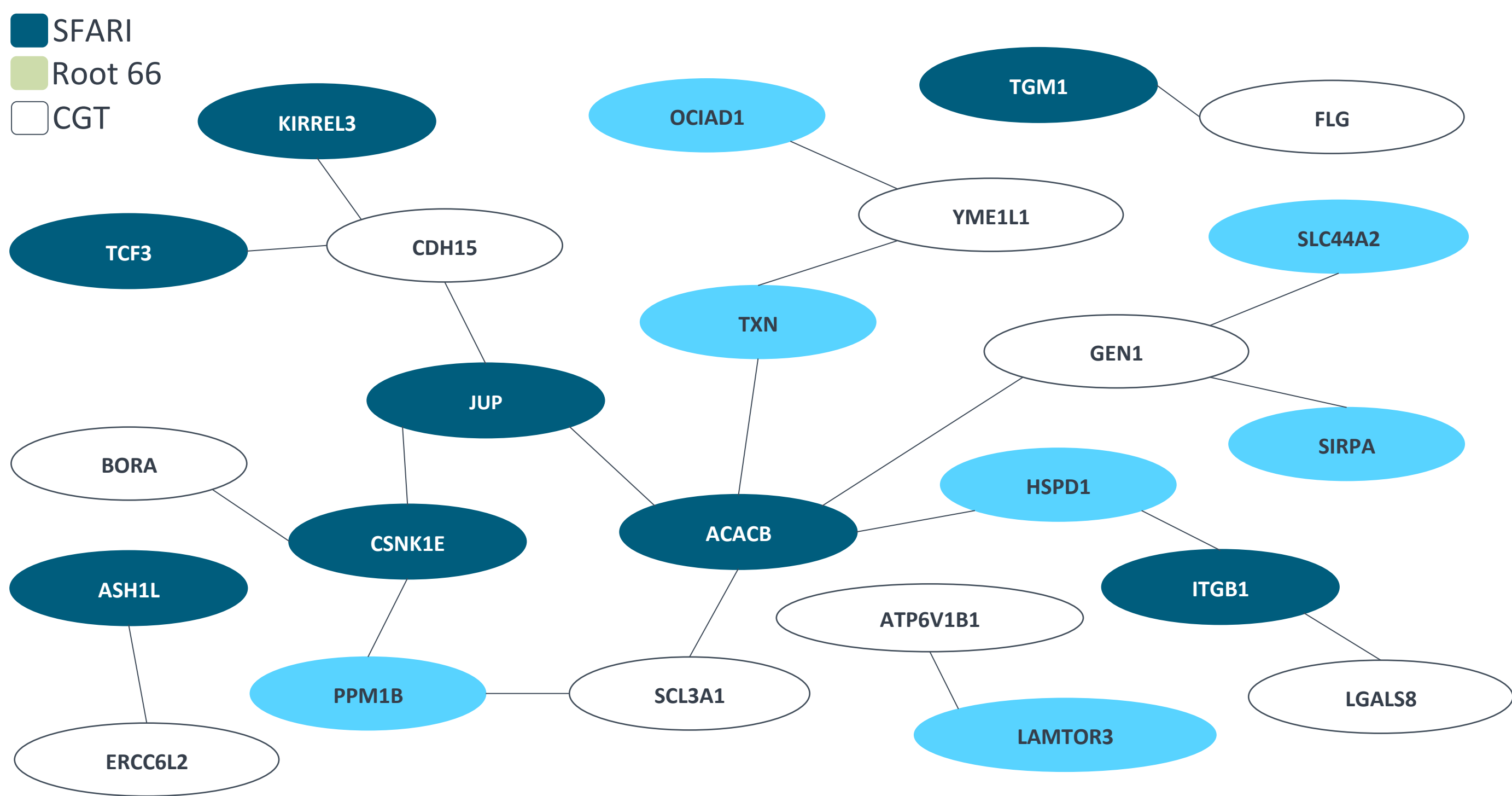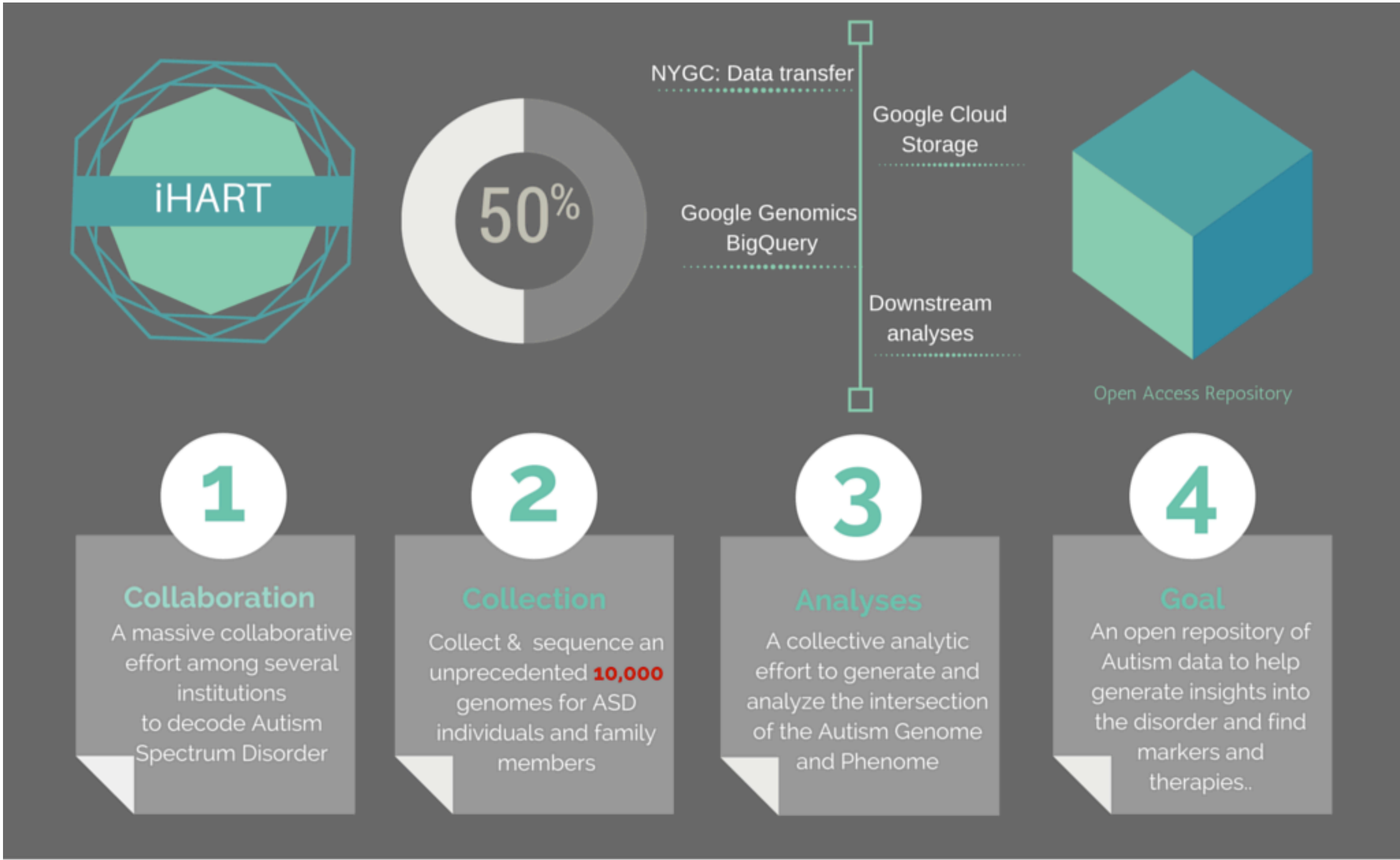
Ashwin et al. 2014, Park et al. 2016, Lee et al. 2010, Peltier et al. 2007, and Goines and Van de Water 2010

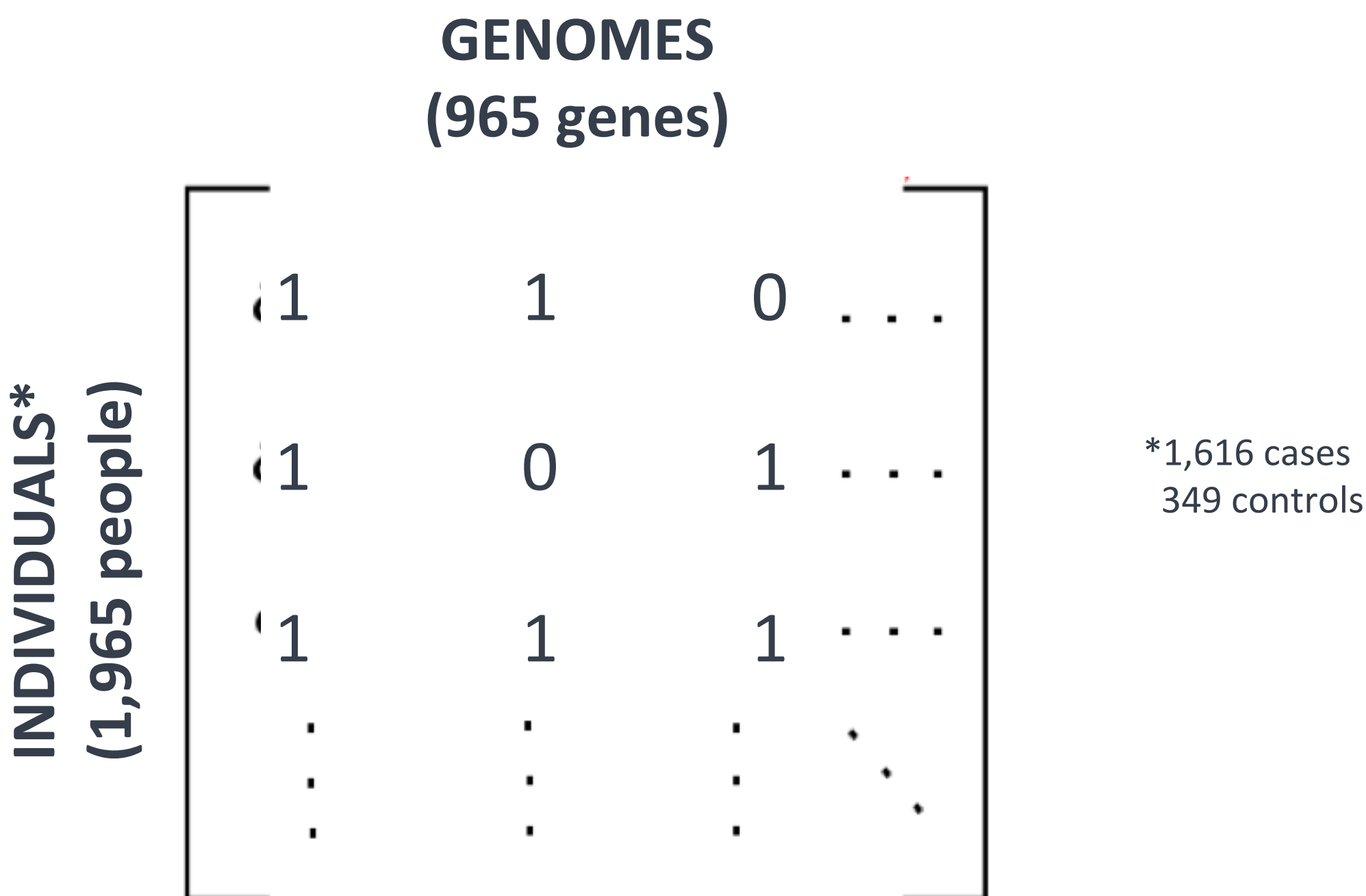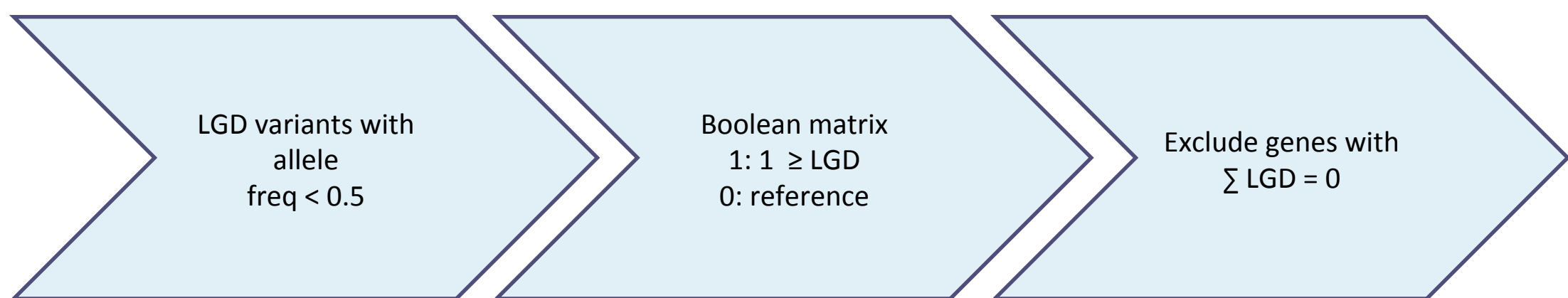### CGT genes interact with known ASD genes

STRING: functional protein association networks
Abrahams et al. 2013, Diaz-Beltran et al. 2016

Discussion

Whereas classical genome-wide association studies to pinpoint genes relevant to a biological condition focus on individual genes, coalitional game theory presents a cooperative view of the alteration landscape, more comprehensively accounting for polygenic complexity. Calculating the Shapley value and filtering for genes with the highest average marginal contribution over all possible alteration combinations can boost biologically informative signal. Stratifying patients through PCA according to their landscape of co-alterations could improve the precision of diagnosis, and knocking out groups of genes identified in functional assays could reveal potent combinations in therapeutically targeting the molecular underpinnings of ASD.

Coalitional game theory thus serves as a powerful approach to elucidate epistatic interactions that may only emerge in a multi-gene model. Capitalizing on the unparalleled rate of genomes being sequenced in increasingly diverse demographic and disease populations, unconventional yet statistically sound tools such as CGT may accelerate the search for biomarkers, particularly in polygenic conditions of mental health.

References

Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. Nature Reviews. 2008;9:341–355.
Esteban EJ, Wall DP. Using game theory to detect genes involved in Autism Spectrum Disorder. https://doi.org/10.1007/s11750-009-0111-6. 2011;19:121-129.
Moretti S, van Leeuwen D, Gmuender H, Bonassi S, van Delft J, Kleinjans J, Patrone F, Merlo DF. Combining Shapley value and statistics to the analysis of gene expression data in children exposed to air pollution. BMC Bioinformatics 2008;9:361.