HHS Public Access

Author manuscript

Pac Symp Biocomput. Author manuscript; available in PMC 2018 July 23.

Published in final edited form as:

Pac Symp Biocomput. 2018 ; 23: 436–447.

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

Anika Gupta, **Min Woo Sun**, **Kelley Marie Paskov**, **Nate Tyler Stockham**, **Jae-Yoon Jung**, and **Dennis Paul Wall**

Departments of Pediatrics and Biomedical Data Sciences, Stanford University, 1265 Welch Road, Palo Alto, CA 94305, United States

Abstract

Despite mounting evidence for the strong role of genetics in the phenotypic manifestation of Autism Spectrum Disorder (ASD), the specific genes responsible for the variable forms of ASD remain undefined. ASD may be best explained by a combinatorial genetic model with varying epistatic interactions across many small effect mutations. Coalitional or cooperative game theory is a technique that studies the combined effects of groups of players, known as coalitions, seeking to identify players who tend to improve the performance--the relationship to a specific disease phenotype--of any coalition they join. This method has been previously shown to boost biologically informative signal in gene expression data but to-date has not been applied to the search for cooperative mutations among putative ASD genes. We describe our approach to highlight genes relevant to ASD using coalitional game theory on alteration data of 1,965 fully sequenced genomes from 756 multiplex families. Alterations were encoded 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, a “player” metric, referred to as the Shapley value, was calculated for each gene in the case and control cohorts. Sixty seven genes were found to have significantly elevated player scores and likely represent significant contributors to the genetic coordination underlying ASD. Using network and cross-study analysis, we found that these genes are involved in biological pathways known to be affected in the autism cases and that a subset directly interact with several genes known to have strong associations to autism. These findings suggest that coalitional game theory can be applied to large-scale genomic data to identify hidden yet influential players in complex polygenic disorders such as autism.

Keywords

Coalitional Game Theory; Autism Spectrum Disorder

1 Introduction

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disease with strong genetic etiology. The rapid growth of genome sequencing capabilities has enabled the collection of large datasets and genome-wide association studies in the quest for causative mutations (C Yuen et al. 2017, Hardy and Singleton 2009, Hirschhorn et al. 2002, Iossifov et al. 2014, Leppa et al. 2016, Robinson et al. 2015, Sanders et al. 2015, Weiner et al. 2017,

Manolio et al. 2009). Sequencing and data collection efforts such as those undertaken by the Simons Foundation Autism Research Initiative have thus far identified over 100 high confidence genes correlated to ASD and intellectual disability (Sanders et al. 2015, Abrahams et al. 2013). Expression studies have identified an additional 66 genes with a highly correlated regulatory pattern in both blood and brain samples of individuals with ASD (Diaz-Beltran et al. 2016).

Models explaining the genetic architecture of ASD include inherited and de novo alterations and have evolved to include polygenic, additive alterations (de la Torre-Ubieta et al. 2016). Despite an estimated 90% heritability of the disorder, only a small fraction of cases can be explained by known molecular causes (Abrahams and Geschwind 2008), and typical approaches to detect pertinent genes often do not account for intergenic associations. Epistatic interactions have been proposed to explain a portion of the remaining unknown genetic etiology (Phillips 2008, Coutinho et al. 2007). Due to the rarity of many alterations involved in the polygenic models, studies have largely been unable to achieve the statistical power necessary for pinpointing new genetic causes (Gratten et al. 2014, Eichler et al. 2010).

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. This tactic assumes synergy within coalitions to explain a given phenotype, where individual players are represented by genes. Genes that have the greatest average marginal contribution across all coalitions–the genes that play the most in coalitions associated with the disease phenotype–are selected as significant. Previous work has applied CGT approaches for differential gene expression analysis and has successfully identified groups of genes as contributing to Alzheimer's (Vardarajan et al. 2013) and ASD (Esteban et al. 2011, Moretti et al. 2008).

We propose that CGT can be effectively applied to gene alteration data to highlight candidate ASD genes, supporting the polygenic model. We demonstrate an example on 1,965 genomes of individuals in multiplex ASD families (1616 cases and 349 controls). Identifying inherited alterations in genes not previously linked to ASD, such as those presented here, could lead to more accurate diagnoses of the disorder and targeted, proactive therapeutic development against the combinations of underlying molecular causes. This approach could be similarly applied to other complex diseases with combinatorial molecular underpinnings.

2 Methods

2.1 Data source and preprocessing

We analyzed 30x-coverage whole genome sequencing data from the Hartwell Foundation's Autism Research and Technology Initiative (iHART), which has amassed one of the largest human disease genome sequencing efforts to-date. Specifically, we assessed the genomes and phenotypic measurements from 756 multiplex families containing at least two children affected by ASD, unaffected parents, and zero or more unaffected siblings.

Families grouped by the phenotypes of their children are as follows (number of children with ASD, number of neurotypical children): 380 (2, 0), 243 (2, 1), 62 (3, 0), 29 (3, 1), 23 (2, 2), 5 (3, 2), 4 (4, 0), 2 (3, 3), 2 (3, 4), 2 (4, 1), 2 (5, 0), 1 (4, 4), and 1 (5, 2) families. Quality control of the sequenced data included removal of non-mendelian variants, which removed sequencing error as well as de novo mutations. Removal of all parents led to an imbalance in the number of cases and controls included in the analysis.

To test the hypothesis of the impact of inherited mutations on the ASD phenotype, we restricted our attention to inherited mutations with highest predicted impact (termed likely gene disrupting, or LGD). We filtered the alteration space to include only loss-of-function variants that had high haplotype-aware consequences (CSQ impact = high) from the variant call format files (http://samtools.github.io/bcftools). For each variant, we predicted the inheritance pattern based on the mother's genotype, father's genotype, and the child's genotype. This included autosomal, pseudoautosomal, and sex-linked genes.

We grouped the alterations across all alleles in each gene for each sample, combining homozygous alternate variants and compound heterozygous variants for a given gene. We also included an allele frequency cutoff (keeping only those with a frequency ≤0.5) in order to prevent variants with high allele frequency from masking signal. Collapsing the data from allele to gene level, we encoded these alterations into a binary matrix, with 1 indicating the presence of at least one high impact alteration in a given gene for a given sample (homozygous alternate, compound heterozygous, or both), and 0 indicating that no such alteration existed in that gene/sample combination. Only genes with a high impact alteration in at least one sample were kept in the subsequent analyses.

2.2 Coalitional game theory method

We applied the coalitional game theory method as described by Moretti et al. (2008) to the iHART alteration data. Coalitional game theory studies the synergy among groups of players. Let N be a finite set of players (in our case, genes). We define a coalition T to be a subset of players ($T \subseteq N$), and we select a score function $v: 2^N \rightarrow \mathbf{R}$ which assigns a score to every possible coalition. We assume $v(\emptyset) = 0$. The fundamental concept of coalitional game theory is that players working together in a coalition may produce higher or lower scores than the sum of each of those players working individually. This means we assume there are cases such that

$$v(T) \neq \sum_{t \in T} v(\{t\}) \quad (1)$$

Our goal was to identify players who tend to increase the score of any coalition they join. To do this, we calculated the Shapley value for each player. The Shapley value ϕ_i is the marginal contribution of player i over all possible coalitions.

$$\phi_i = \sum_{T \subseteq N : i \in T} \frac{(|T| - 1)!(|N| - |T|)!}{|N|!}(v(T) - v(T \backslash \{i\})) \quad (2)$$

For our analysis, we used the unanimity score given by Moretti et. al. (2007). Let m be the number of individuals in our dataset. Each individual j has a high impact alteration in a subset M_j of genes. The unanimity score $u(T, j)$ for coalition T and individual j is 1 if all M_j altered genes are contained in the coalition ($M_j \subseteq T$) and 0 otherwise. We then define our characteristic function v to be

$$v(T) = \sum_{j=1}^{m} u(T, j) \quad (3)$$

Moretti et al. (2007) show that this choice of v makes computing the Shapley values ϕ_i tractable for large datasets. The algorithm took 16.793 minutes to run for 965 genes and 1,965 samples on 1 node of a computer cluster, with 4 cores and 32GB of memory. A detailed example of Shapley value calculation can be found in the electronic supplementary material provided by Moretti et al. (2008).

2.3 Game theory analysis

All statistical analyses and applications of coalitional game theory were performed using Bioconductor version 3.5 (http://www.bioconductor.org/) and R version 3.4.0 (http://www.r-project.org/). We removed all genes without any alterations across samples, as the corresponding Shapley values would be 0. This reduced the feature space of genes from 13,853 to 965 genes, resulting in a final boolean matrix of 965 genes and 1,965 samples (1,616 cases and 349 controls). In order to compute the Shapley values, the boolean matrix was split into two matrices corresponding to case and control: B^{case} and $B^{control}$. We adapted the script from Moretti et al. (2008) to compute the Shapley values using the boolean matrices.

We performed Comparative Analysis of Shapley value (CASh), a resampling-based multiple hypothesis testing procedure introduced in Moretti et al. (2008) to filter out genes with Shapley values that could be high due to chance. We used the function MTP from Bioconductor package “multtest” to compute a CASh p-value for each gene. We ran 1,000 nonparametric bootstrap re-samples with replacement on the matrices. The MTP produces unadjusted p-values for each gene, along with a bootstrap p-value, calculated via simulations. We filtered for only those genes that were significant at the 0.05 and 0.01 significance levels.

Given that our cases and controls could be related siblings and share inherited mutations, we reran coalitional game theory on randomly sampled cases and controls from each family (1 of each per eligible family). As the algorithm is run separately on the case and control binary

matrices, none of the samples share familial connection when the Shapley values are computed. The gene identification process via CGT is presented in Figure 1.

2.4 Functional analyses

To elucidate potential associations with previously-correlated ASD gene candidates, we cross-referenced the CGT gene lists with both the high confidence SFARI genes list (Abrahams et al. 2013) and a set of genes found to be significantly dysregulated in both the blood and brain of individuals with autism, the Root 66 gene list (Diaz-Beltran et al. 2016), using the functional protein network analysis tool STRING (string-db.org). We also checked for network representation of the CGT genes using the Reactome Pathway Browser (reactome.org), a free, open-source, curated and peer-reviewed pathway database. We reported pathways enriched for CGT genes with FDR<0.1.

3 Results

The filtration and binary conversion pre-processing steps yielded 1,616 cases and 349 controls with alteration information for 965 genes. 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 (Figure 1). Sixty-seven genes showed statistical significance at the 0.05 significance level ($p < 0.05$), with 23 of those genes significant at the 0.01 level ($p < 0.01$) (Table 1). Rerunning coalitional game theory on randomly sampled cases and controls from each of the eligible families returned the same genes and 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. Nine of the 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 (Figure 2).

Reactome pathway analysis detected 20 significant, non-overlapping functional categories (FDR < 0.1). Of note, pathways representing axon guidance in developmental biology and related neurologic disorders (FDR=0.03 and FDR=0.0047), FGFR1- and insulin receptor-mediated signaling (FDR=0.0042 and FDR=0.034), the innate immune system (FDR=0.09), and olfactory signaling (FDR=0.0252) were enriched for CGT genes. Each of these biological functions has been previously associated with ASD (Ashwin et al. 2014, Park et al. 2016, Lee et al. 2010, Peltier et al. 2007, and Goines and Van de Water 2010).

4 Discussion

Whereas classical genome-wide association studies to pinpoint genes relevant to a biological condition focus on individual genes, coalitional game theory takes into account broader interactions between groups of genes leading to a phenotype. Calculating the Shapley value and filtering for the genes with the highest average marginal contribution over all possible alteration combinations can enable detection of significant gene coaltions and thus boost

biologically informative signal. This cooperative view of the alteration landscape more comprehensively accounts for polygenic complexity and may be essential for understanding ASD's genetic architecture.

In this study, we applied cooperative game theory to a large collection of whole genomes from multiplex autism families in an effort to find cooperative signal among coalitions that relate specifically to the autism phenotype. We focused our analysis on inherited gene disrupting mutations to pursue the hypothesis of ASD being a largely inherited, polygenic disorder. Our sample set consisted of 1,616 cases and 349 family-based controls (due to the exclusion of parents), where a gene was assigned a 1 if it contained at least 1 or more such mutations. By analyzing the cooperative contribution of each gene to the phenotype, we found 67 genes that significantly increased the likelihood of a coalition "winning the game" whenever they joined the coalition, where winning refers to the strength of the association to the autism phenotype. Through random sampling, we found that these genes did not appear to be an artifact of the family structure of our data.

Compellingly, 9 of these genes have published links to autism through either DNA or RNA-based analysis (Brown et al. 2015). Further supporting the role of these genes in autism, we found their protein products to be enriched for interaction with the protein products of known autism candidate genes, often interacting just one node away from hubs of connected ASD gene products. Pathway enrichment analyses revealed that signaling pathways in ASD-associated biological functions such as axon development and innate immunity are enriched for CGT genes.

Potential limitations of this work entail the exclusion of alterations in the non-coding regions, as well as the exclusion of low CSQ impact variants within genes. This limits our ability to find links between the disease phenotype and more subtle genomic and non-coding variation. Exploring such nuanced alterations could shed light on additional high impact, potentially causative molecular states for a disease under consideration. Replication of our analyses and functional characterization will be necessary for comprehensively evaluating the biological implications of our findings. Additionally, many of the genes that were identified via CGT are pseudogenes and have not yet been well studied, making the analysis of such regions even more pressing.

Probing into the specifics of the protein-protein interactions between CGT genes and known ASD candidates, as well as into the groups of co-altered gene coalitions in case subgroups, may provide further mechanistic insights into the underlying molecular causes of ASD. Identifying similarly statistically correlated genes through additional genome-wide ASD gene alteration datasets may elucidate higher degree epistatic connections that can be targeted in both diagnosis and treatment of ASD. 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 characterize 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.

Acknowledgments

This work was supported in part by the Hartwell Foundation award to D.P. Wall and the Hartwell Autism Research and Technology Initiative (iHART).

6 Bibliography

Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, Menashe I, Wadkins T, Banerjee-Basu S, Packer A. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Molecular autism. 2013; 4(1):36. [PubMed: 24090431]

Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. Nature Reviews. Genetics. 2008; 9(5):341–355.

Ashwin C, Chapman E, Howells J, Rhydderch D, Walker I, Baron-Cohen S. Enhanced olfactory sensitivity in autism spectrum conditions. Molecular Autism. 2014; 5:53. [PubMed: 25908951]

Banerjee-Basu S, Packer A. SFARI Gene: an evolving database for the autism research community. Disease Models Mechanisms. 2010; 3(3–4):133–135. [PubMed: 20212079]

Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, Ermolaeva O, Tolstoy I, Tatusova T, Pruitt KD, Maglott DR, Murphy TD. Gene: a gene-centered information resource at NCBI. Nucleic Acids Research. 2015; 43(Database issue):D36–42. [PubMed: 25355515]

Yuen CRK, Merico D, Bookman M, Howe LJ, Thiruvahindrapuram B, Patel RV, Whitney J, Deflaux N, Bingham J, Wang Z, Pellecchia G, Buchanan JA, Walker S, Marshall CR, Uddin M, Zarrei M, Deneault E, D'Abate L, Chan AJS, Koyanagi S, Scherer SW. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nature Neuroscience. 2017; 20(4): 602–611. [PubMed: 28263302]

Coutinho AM, Sousa I, Martins M, Correia C, Morgadinho T, Bento C, Marques C, Ataíde A, Miguel TS, Moore JH, Oliveira G, Vicente AM. Evidence for epistasis between SLC6A4 and ITGB3 in autism etiology and in the determination of platelet serotonin levels. Human Genetics. 2007; 121(2): 243–256. [PubMed: 17203304]

Diaz-Beltran L, Esteban FJ, Wall DP. A common molecular signature in ASD gene expression: following Root 66 to autism. Translational psychiatry. 2016; 6:e705. [PubMed: 26731442]

Dwyer CA, Esko JD. Glycan susceptibility factors in autism spectrum disorders. Molecular aspects of medicine. 2016; 51:104–114. [PubMed: 27418189]

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. Nature Reviews. Genetics. 2010; 11(6):446–450.

Esteban EJ, Wall DP. Using game theory to detect genes involved in Autism Spectrum Disorder. 2011; 19(1):121–129.doi: 10.1007/s11750-009-0111-6

Goines P, Van de Water J. The immune system's role in the biology of autism. Current Opinion in Neurology. 2010; 23(2):111–117. [PubMed: 20160651]

Gratten J, Wray NR, Keller MC, Visscher PM. Large-scale genomics unveils the genetic architecture of psychiatric disorders. Nature Neuroscience. 2014; 17(6):782–790. [PubMed: 24866044]

Hardy J, Singleton A. Genomewide association studies and human disease. The New England Journal of Medicine. 2009; 360(17):1759–1768. [PubMed: 19369657]

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genetics in Medicine. 2002; 4(2):45–61. [PubMed: 11882781]

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, Wigler M. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014; 515(7526):216–221. [PubMed: 25363768]

Lee EH, Kim YH, Hwang JS, Kim SH. Non-type I cystinuria associated with mental retardation and ataxia in a Korean boy with a new missence mutation(G173R) in the SLC7A9 gene. Journal of Korean Medical Science. 2010; 25(1):172–175. [PubMed: 20052367]

Leppa VM, Kravitz SN, Martin CL, Andrieux J, Le Caignec C, Martin-Coignard D, DyBuncio C, Sanders SJ, Lowe JK, Cantor RM, Geschwind DH. Rare inherited and de novo cnvs reveal complex contributions to ASD risk in multiplex families. American Journal of Human Genetics. 2016; 99(3):540–554. [PubMed: 27569545]

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, Mc-Carthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–753. [PubMed: 19812666]

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. [PubMed: 18764936]

Moretti S, Patrone F, Bonassi S. The class of microarray games and the relevance for index genes. 2007; 15(2):256–280.doi: 10.1007/s11750-007-0021-4

Park HJ, Kim SK, Kang WS, Park JK, Kim YJ, Nam M, Kim JW, Chung J-H. Association between IRS1 Gene Polymorphism and Autism Spectrum Disorder: A Pilot Case-Control Study in Korean Males. International Journal of Molecular Sciences. 2016; 17(8)

Peltier J, O'Neill A, Schaffer DV. PI3K/Akt and CREB regulate adult neural hippocampal progenitor proliferation and differentiation. Developmental Neurobiology. 2007; 67(10):1348–1361. [PubMed: 17638387]

Phillips PC. Epistasis–the essential role of gene interactions in the structure and evolution of genetic systems. Nature Reviews. Genetics. 2008; 9(11):855–867.

Robinson EB, Neale BM, Hyman SE. Genetic research in autism spectrum disorders. Current Opinion in Pediatrics. 2015; 27(6):685–691. [PubMed: 26371945]

Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, Murtha MT, Bal VH, Bishop SL, Dong S, Goldberg AP, Jinlu C, Keaney JF, Klei L, Mandell JD, Moreno-De-Luca D, Poultney CS, Robinson EB, Smith L, Solli-Nowlan T, State MW. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015; 87(6):1215–1233. [PubMed: 26402605]

de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH. Advancing the understanding of autism disease mechanisms through genetics. Nature Medicine. 2016; 22(4):345–361.

Vardarajan BN. ProQuest Dissertations and Theses Global 2013 Identification of gene-gene interactions for alzheimer's disease using co-operative game theory. 1179981110

Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, Samocha KE, Goldstein JI, Okbay A, Bybjerg-Grauholm J, Werge T, Hougaard DM, Taylor J, Skuse D, Devlin B, Anney R, Sanders SJ, Bishop S, Robinson EB. iPSYCH-Broad Autism Group Psychiatric Genomics Consortium Autism Group. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nature Genetics. 2017; 49(7): 978–985. [PubMed: 28504703]

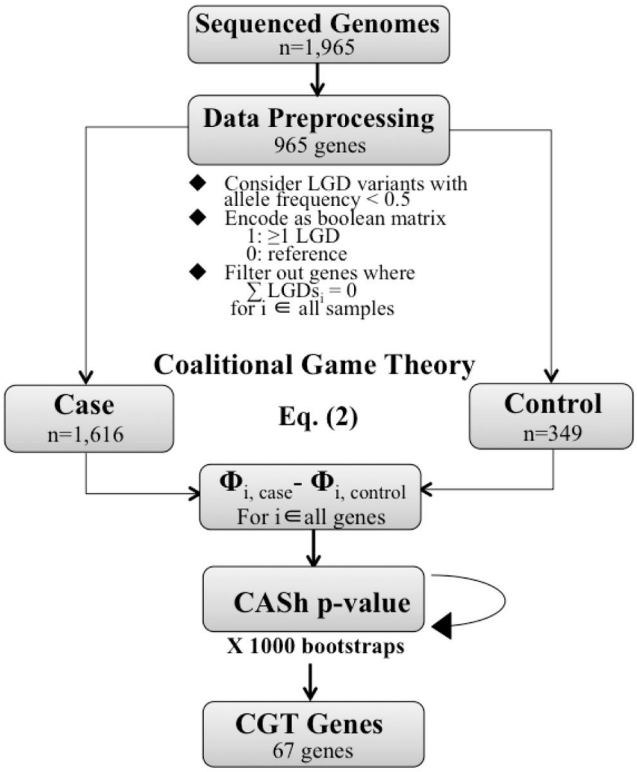

Figure 1.

Data analysis flow diagram, starting from the sequenced genomes to identification of significant genes through coalitional game theory.

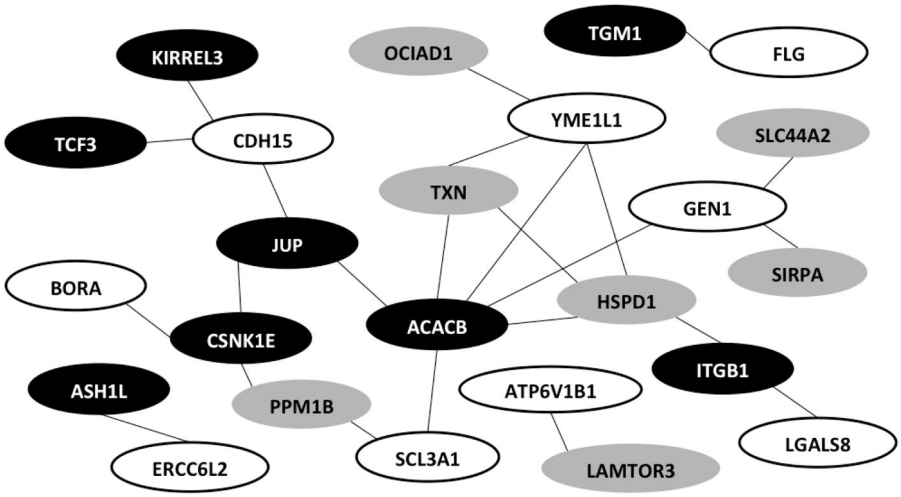

Figure 2.

Functional protein interactions between coalitional game theory genes, SFARI genes, and Root 66 genes. SFARI genes are in black, Root 66 genes are in gray, and CGT genes ($p < 0.05$) are in white. Nine CGT genes have direct links to known candidate ASD genes.

Table 1

Genes selected through coalitional game theory at two levels of significance. The 44 genes listed in $p<0.05$ are the subset of the 67 genes not in $p<0.01$.

Significance	Gene symbol
p-value < 0.05	A2ML1, AC008703.1, AC093911.1, ALOX15P2, ATP13A5, BORA, BPIFB5P, C12orf60, C3orf35, CARD8, CCDC26, CCDC7, CDH15, COQ10A, CTC-525D6.1, DUSP16, ERCC6L2, FAM151A, FAM81B, FLG, GBGT1, HLA-K, LGALS8, MAGEC3, MYCT1, OR2T4, OR4Q2, OR6C1, OR8B3, RBAK-RBAKDN, RP11-104E19.1, RP11-160N1.10, RP11-404K5.2, RP11-56H2.2, RP11-618I10.2, RP11-738O11.13, SLC3A1, SSPO, TCP11, TRBV6-7, TRIM48, UBXN11, YME1L1, ZNF99
p-value < 0.01	AF196972.4, AP002856.6, ATP6V1B1, C10ORF68, CDRT15P1, CTB-23I7.1, CTD-2130O13.1, CTD-2509G16.2, GEN1, KRT43P, MDP1, MPRIP, NT5C1B, OR4P4, OR5M10, OR5M11, OR8I2, PRIM2, RP11-15E18.4, RP11-283G6.4, RP11-705C15.2, SSXP3, VWA7