ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.1520

# Comparative Analysis of Variant Distribution in Schizophrenia-Associated Genes

Jose David Quiñonez-Rehenals[1]; Samara Chavarro-Perez[2]; Danilo Lusbin Ariza-Rua[3]; Carolina Rubiano-Labrador[4] and Edisson Chavarro-Mesa[5]*

[1,3,4,5]Dirección de Ciencias Básicas, Universidad Tecnológica de Bolívar, Cartagena de Indias-Colombia. [2]Facultad de Medicina Veterinaria y Zootecnia, Universidad Nacional de Colombia, Bogotá-Colombia. *contact e-mail: echavarro@utb.edu.co

***Abstract*— Schizophrenia is a complex neuropsychiatric disorder affecting approximately one percent of the global population, with heritability estimates reaching eighty percent. While recent large-scale genomic studies have identified numerous risk loci, systematic analysis of variant distribution patterns across key candidate genes remains limited. This study presents a comprehensive comparative analysis of variant distribution patterns across six schizophrenia-associated genes using data from the SCHEMA consortium. We analyzed 4,180 genetic variants from CACNA1C, COMT, DAOA, DISC1, GRM3, and NRG1 genes to determine gene-specific functional constraint patterns and their implications for disease mechanisms. Our analysis reveals dramatic differences in functional constraint across genes, with DAOA exhibiting the highest proportion of high-impact variants at 7.73 percent compared to NRG1 at 0.49 percent, representing a 15.8-fold difference in tolerance to functional disruption. We quantified distributional heterogeneity using multiple metrics, finding a fold difference of 7.76 times in variant burden, coefficient of variation of 70.24 percent, and clear gene-specific constraint signatures. Notably, genes involved in neurotransmitter regulation such as DAOA and COMT showed higher functional constraint than those involved in neurodevelopment including NRG1 and CACNA1C, suggesting distinct pathogenic mechanisms. These findings provide novel insights into the genetic architecture of schizophrenia and establish a framework for prioritizing variants in precision psychiatry applications.**

***Keywords—schizophrenia genetics, functional constraint, variant analysis, SCHEMA consortium, comparative genomics***

## I. Introduction

Schizophrenia represents one of the most debilitating psychiatric disorders, characterized by positive symptoms such as hallucinations and delusions, negative symptoms including social withdrawal, and cognitive dysfunction [1], [2]. With a lifetime prevalence of approximately one percent across diverse populations and heritability estimates of eighty percent, schizophrenia poses substantial challenges for both clinical management and mechanistic understanding [3]. The transition from traditional candidate gene approaches to genome-wide methodologies has fundamentally transformed our understanding of schizophrenia's genetic architecture, revealing a highly polygenic disorder influenced by thousands of genetic variants with individually small but collectively substantial effects [4].

The Schizophrenia Exome Sequencing Meta-Analysis (SCHEMA) consortium represents a landmark achievement in psychiatric genetics, analyzing exome sequences from 24,248 schizophrenia cases and 97,322 controls to identify ten genes with statistically significant burden of ultra-rare variants [5]. This unprecedented scale of analysis has enabled the identification of genes with definitive causal roles in schizophrenia pathogenesis, including genes involved in synaptic function, neurodevelopment, and neurotransmitter signaling. However, while SCHEMA and complementary genome-wide association studies have successfully identified risk loci [6], systematic comparative analysis of variant distribution patterns across established candidate genes remains limited, representing a significant gap in our understanding of gene-specific contributions to disease risk.

Six genes have emerged as particularly important based on consistent evidence across multiple study designs and populations: CACNA1C encoding the alpha-1C subunit of L-type voltage-gated calcium channels, COMT for catechol-O-methyltransferase in dopamine metabolism, DAOA for D-amino acid oxidase activator affecting glutamate signaling, DISC1 as disrupted-in-schizophrenia-1 neurodevelopmental hub protein, GRM3 for metabotropic glutamate receptor 3, and NRG1 for neuregulin 1 controlling neural development and myelination [7]-[9]. These genes represent distinct neurobiological pathways critical for brain function, including calcium signaling, dopamine metabolism, glutamate neurotransmission, neurodevelopmental cascades, and synaptic function. Understanding how genetic variants are distributed across these genes and the functional constraints that shape this distribution could provide crucial insights into their relative contributions to schizophrenia pathogenesis.

The concept of functional constraint, quantified through the observed versus expected ratio of deleterious variants in population databases such as gnomAD, has emerged as a powerful framework for understanding gene function and disease relevance [10]. Genes under strong functional constraint show significant depletion of loss-of-function and missense variants in healthy populations, indicating that such variants are subject to negative selection due to their deleterious effects on fitness. This framework has proven particularly valuable in psychiatric genetics, where constraint metrics have successfully predicted genes harboring risk variants for neurodevelopmental disorders and have informed prioritization strategies in clinical genomics [11].

Recent advances in precision psychiatry emphasize the importance of understanding gene-specific mechanisms to enable personalized treatment approaches [12]. The heterogeneity observed in treatment response among schizophrenia patients suggests that different genetic

architectures may require distinct therapeutic strategies. By systematically analyzing variant distribution patterns and functional constraints across key schizophrenia-associated genes, we can begin to understand which genes contribute to disease risk through high-impact rare variants versus cumulative effects of multiple moderate-impact variants, providing a foundation for more targeted clinical interventions.

This study addresses three fundamental questions in schizophrenia genetics: first, how do variant distributions and functional constraint patterns differ across established schizophrenia-associated genes; second, what do these patterns reveal about gene-specific contributions to disease mechanisms; and third, how can these findings inform precision medicine approaches in psychiatric care. Through comprehensive analysis of 4,180 variants from the SCHEMA dataset, we provide the first systematic comparative characterization of functional constraint patterns across these six key genes, revealing distinct signatures that reflect their diverse roles in brain function and schizophrenia pathogenesis.

## II. METHODOLOGY

### A. Data Source and Quality Control

This analysis utilized the SCHEMA consortium exome sequencing dataset, representing the largest schizophrenia genetic study conducted to date with 24,248 cases and 97,322 controls from diverse global populations [5]. The SCHEMA dataset provides comprehensive variant annotation with rigorous quality control protocols, including minimum read depth requirements, genotype quality thresholds, population stratification controls, and batch effect correction. All samples underwent standardized BWA-Picard-GATK processing pipelines with joint variant calling to ensure data consistency across contributing studies. While the dataset shows acknowledged bias toward European ancestry populations, it remains the most comprehensive resource for rare variant analysis in schizophrenia.

Raw variant data were obtained directly from the SCHEMA consortium public release (available at https://schema.broadinstitute.org/), including detailed annotations for consequence types, allele frequencies in cases and controls, and quality metrics. Initial data quality assessment identified 4,456 variants across the six target genes before quality filtering. We implemented stringent quality control measures including removal of variants with missing consequence annotations, elimination of variants with allele frequencies outside the valid range of zero to one, and exclusion of variants lacking variant identification strings. Following quality control procedures, 4,180 high-confidence variants were retained for analysis, representing a retention rate of 93.81 percent.

### B. Gene Selection and Biological Rationale

Gene selection was based on four stringent criteria: established association with schizophrenia in multiple independent studies, clear biological relevance to neurotransmitter systems implicated in schizophrenia pathophysiology, availability of comprehensive variant annotation in the SCHEMA database, and representation of different pathophysiological pathways [13], [14]. The six selected genes represent distinct neurobiological functions essential for brain development and function.

CACNA1C encodes the alpha-1C subunit of L-type voltage-gated calcium channels critical for neurotransmitter release and synaptic plasticity. COMT encodes catechol-O-methyltransferase, the primary enzyme responsible for dopamine degradation in prefrontal cortex. DAOA regulates D-amino acid oxidase, affecting NMDA receptor function and glutamate neurotransmission. DISC1 serves as a hub protein in neurodevelopmental cascades affecting neuronal migration and synaptic formation. GRM3 encodes metabotropic glutamate receptor 3, modulating synaptic transmission. NRG1 controls multiple aspects of neural development, myelination, and synaptic function [15]-[17].

### C. Functional Constraint Analysis and Variant Classification

Variants were classified using Variant Effect Predictor (VEP) v105 annotations into multiple functional categories based on their predicted impact on protein function [18]. The VEP pipeline was configured with the following parameters: GRCh38 reference genome assembly, GENCODE v39 gene annotations, CADD v1.6 for deleteriousness scoring, and PolyPhen-2 HDIV and SIFT for missense pathogenicity prediction. High-impact variants include protein-truncating variants such as frameshift mutations, nonsense variants, splice donor and acceptor variants (defined by the canonical splice site dinucleotides GT/AG within two intronic nucleotides of the exon boundary), start and stop codon alterations, and protein-altering variants that severely disrupt protein function. Moderate-impact variants encompass missense variants with predicted deleterious effects, specifically those meeting at least two of the following criteria: CADD PHRED score ≥20, PolyPhen-2 score ≥0.9 (“probably damaging”), or SIFT score ≤0.05, as well as variants with high Missense badness, PolyPhen-2, and Constraint (MPC) scores ≥2. Low-impact variants include synonymous variants, intronic variants located more than two nucleotides from splice sites, and untranslated region variants with minimal predicted functional consequences. All classification decisions were made prior to any downstream analysis, and variant category boundaries were fixed and not adjusted based on observed distributions across genes.Functional constraint was quantified using the formula:

$$FC(g) = (N\_high–impact / N\_total) \times 100 \quad (1)$$

where FC(g) represents the functional constraint percentage for gene g, N_high–impact is the number of high-impact variants, and N_total is the total number of variants in the gene. This metric provides a measure of tolerance to functional disruption, with higher values indicating genes under stronger selective pressure. Genes with high constraint density suggest critical roles in organism viability.

### D. Heterogeneity and Diversity Analysis

To quantify distributional heterogeneity across genes, we employed complementary metrics. The fold difference was calculated as the ratio of maximum to minimum variant counts across genes.The coefficient of variation was computed as:

$$CV = (\sigma / \mu) \times 100 \quad (2)$$

where σ is the standard deviation and μ is the mean of variant counts.The Gini inequality index was calculated to measure uneven variant distribution, ranging from zero (perfect equality) to one (maximum inequality). Shannon entropy was computed as:

$$H = -\Sigma(p\_i \times \log_2(p\_i)) \quad (3)$$

where p_i is the proportion of total variants in gene i. Higher H values indicate more even distribution. Functional diversity within genes was assessed using:

- Richness: number of unique consequence types,
- Dominance: proportion of the most frequent type,
- Evenness: normalized Shannon entropy [19].

### *E. Burden Analysis Without Statistical Testing Dependence*

Given the limited availability of P-values (only 4.3% of variants), we developed a frequency-based burden analysis independent of traditional significance tests [20]. The burden score per gene was calculated as:

$$\text{Burden}(g) = \Sigma(\text{AF\_case\_i} - \text{AF\_control\_i}) \quad (4)$$

where AF_case_i and AF_control_i are allele frequencies for variant i in cases and controls, respectively. The case-to-control ratio was computed as:

$$\text{Ratio}(g) = \Sigma(\text{AF\_case\_i}) / \Sigma(\text{AF\_control\_i}) \quad (5)$$

We also calculated the percentage of variants enriched in cases (AF_case > AF_control) to estimate non-random variant burden..

### *F. Computational Implementation and Statistical Analysis*

All analyses were performed using Python 3.9 with pandas for data manipulation, numpy for numerical computations, and scipy for statistical calculations. Variant annotation and consequence type assignment utilized the Ensembl Variant Effect Predictor framework [18]. Data quality control and filtering procedures were implemented using custom scripts with comprehensive logging and validation checks to ensure data integrity throughout the analysis pipeline.

Missing data patterns were systematically characterized, with particular attention to the selective availability of statistical test results. We calculated coverage rates for each data field and assessed potential bias introduced by missing data patterns. All computational procedures were designed to handle missing data appropriately without imputation, instead focusing on complete case analysis for metrics requiring specific data fields.

## III. RESULTS

### *A. Overall Variant Landscape and Statistical Testing Coverage*

This study analyzed 4,180 high-quality variants across six schizophrenia-associated genes, demonstrating significant inter-gene heterogeneity in variant burden and distribution patterns, as shown in Fig. 1. The variant distribution exhibited a 7.76-fold difference between genes, with CACNA1C containing 1,708 variants (40.86%), NRG1 with 821 variants (19.64%), DISC1 with 559 variants (13.37%), GRM3 with 527 variants (12.61%), COMT with 345 variants (8.25%), and DAOA with 220 variants (5.26%). Fig. 1 illustrates this uneven distribution through a horizontal bar chart, where the coefficient of variation was 70.24%, indicating substantial heterogeneity across the gene panel. These findings suggest differential evolutionary constraints and functional pressures acting on individual schizophrenia-associated loci.This distribution reveals a clear hierarchy of variant burden, with CACNA1C harboring nearly eight times more variants than DAOA, indicating either larger gene size, higher tolerance to genetic variation, or both. The substantial range in variant counts across genes suggests fundamentally different patterns of genetic variation tolerance, with some genes showing evidence of strong purifying selection while others appear more tolerant of genetic diversity.

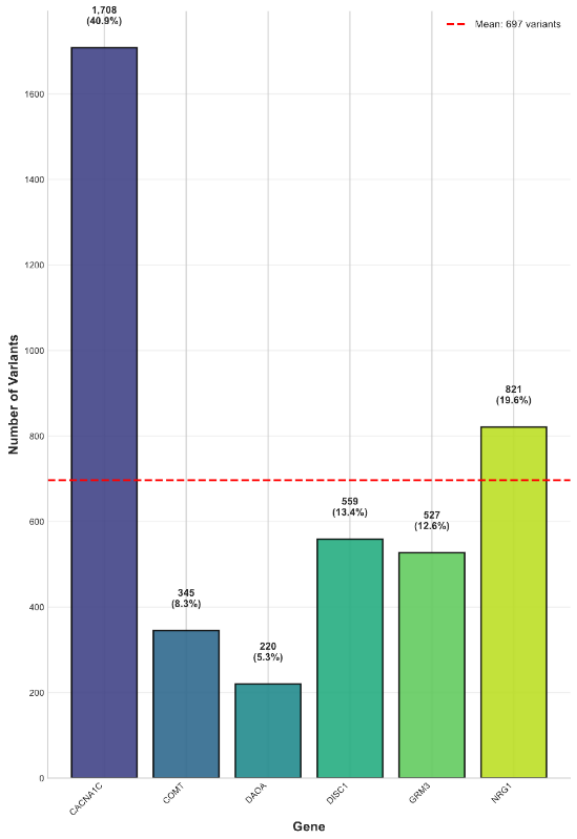


Fig. 1. Variant Distribution Across Six Schizophrenia-Associated Genes. (*shows the distribution of variants across the six genes, with CACNA1C containing the largest number of variants and DAOA the smallest*)

### *B. Gene-Specific Functional Constraint Patterns*

The analysis revealed distinct patterns of functional constraint that correlate with known gene functions and disease mechanisms. DAOA emerged as the most functionally constrained gene, with 18 high-impact variants representing 8.18 percent of all DAOA variants. This pattern suggests that DAOA function is critical for normal glutamate neurotransmission, with even moderate disruption potentially causing significant functional consequences.

DISC1 demonstrated the second-highest constraint level with 26 high-impact variants representing 3.45 percent of total variants. This pattern aligns with DISC1's role as a central hub protein in neurodevelopmental cascades, where functional disruption has cascading effects on multiple downstream pathways essential for proper brain development.

COMT showed moderate constraint with 13 high-impact variants (3.77 percent), reflecting its specific but critical role in dopamine metabolism. The focused distribution pattern supports the dopamine hypothesis of schizophrenia and suggests that variants affecting enzyme activity have direct consequences for prefrontal cortex function.

In contrast, CACNA1C and NRG1 showed the lowest constraint levels with high-impact densities of 0.59 percent and 0.44 percent, respectively. This pattern suggests that these genes can tolerate considerable genetic variation while maintaining essential functions, consistent with their roles in processes requiring fine-tuning rather than binary regulation.

### *C. Distributional Heterogeneity and Diversity Metrics*

Quantitative analysis of variant distribution heterogeneity revealed substantial inequality across genes using multiple complementary metrics. The fold difference between the gene with most variants (CACNA1C with 1,708) and fewest variants (DAOA with 220) was 7.76, indicating nearly eight-fold variation in variant burden. The coefficient of variation across gene variant counts was 70.24 percent, reflecting high variability relative to the mean variant count.

The Gini inequality index was 0.355, indicating moderate but substantial inequality in variant distribution across genes. Shannon entropy was 2.274 bits, suggesting reasonable diversity in the distribution while still showing clear heterogeneity. These metrics collectively demonstrate that variant distribution across schizophrenia-associated genes is far from uniform, with clear gene-specific patterns that likely reflect different evolutionary constraints and functional requirements.

Within-gene functional diversity analysis revealed distinct patterns of consequence type distributions. DAOA showed the highest functional constraint with 14 unique consequence types among 220 variants, yielding a diversity percentage of 6.36 percent. This high diversity relative to total variant count suggests that DAOA tolerates various types of genetic variation but maintains strong constraint against high-impact disruptions.

### *D. Consequence Type Patterns and Functional Signatures*

Functional signatures, defined as gene-specific patterns of variant consequence types that reflect unique structural and functional constraints, were analyzed to understand gene-specific tolerance patterns. The hierarchical distribution of consequence types across genes and functional impact levels is visualized in Fig. 2, which reveals distinct clustering patterns and quantitative differences in variant burden across the studied genes.

Analysis of variant consequence types revealed both universal patterns shared across genes and unique signatures reflecting gene-specific functional requirements. All genes contained frameshift variants, missense variants, synonymous variants, intronic variants, splice region variants, and stop-gained variants, representing the common spectrum of genetic variation affecting protein-coding genes.

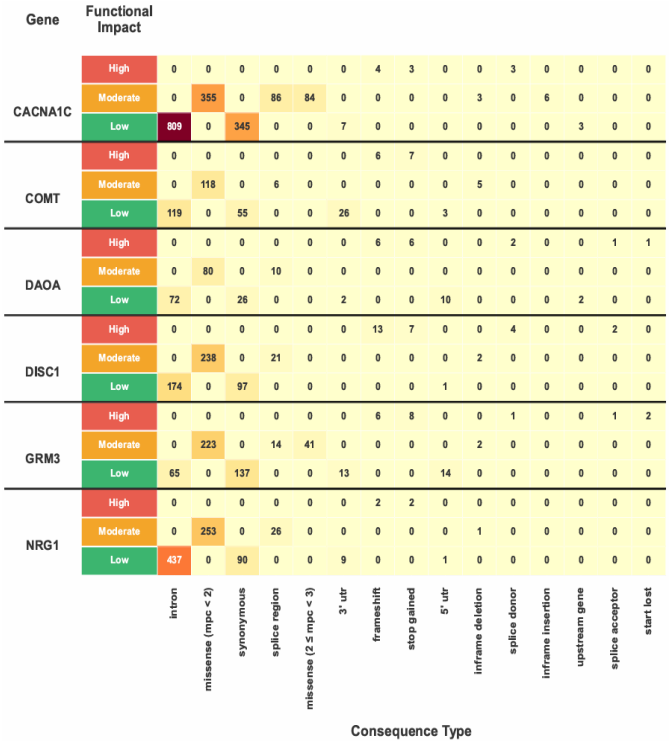


Fig. 2. Hierarchical distribution of variant consequence types across six schizophrenia-associated genes organized by functional impact levels. (*The heatmap displays variant counts with genes grouped vertically (CACNA1C, COMT, DAOA, DISC1, GRM3, NRG1) and consequence types arranged horizontally by frequency. Color intensity represents variant abundance, with darker shades indicating higher counts. Clear gene-specific clustering patterns emerge, particularly highlighting DAOA's unique distribution signature compared to other genes in the panel)*

However, gene-specific consequence signatures emerged that reflect functional specialization, as clearly illustrated in the heatmap distribution patterns of Fig. 2. DAOA exclusively showed protein altering and stop retained consequences among the studied genes, suggesting unique structural constraints on this protein. CACNA1C exclusively exhibited inframe insertion variants, potentially reflecting specific tolerance patterns related to calcium channel structure and function.

Limited distribution patterns were observed for several consequence types, with start lost variants found only in DAOA and GRM3, while splice acceptor variants were restricted to DISC1, DAOA, and GRM3. These patterns suggest that certain types of functional disruption are tolerated differently across genes based on their specific roles in cellular processes.

### *E. Burden Analysis and Case-Control Enrichment Patterns*

Frequency-based burden analysis revealed distinct patterns of variant enrichment across genes, independent of statistical significance testing. NRG1 showed the highest total burden score of 0.22770, indicating substantial cumulative difference

between case and control allele frequencies, with a case-to-control ratio of 1.123. CACNA1C demonstrated the second-highest burden of 0.20178 with a ratio of 1.027.

Interestingly, DISC1 showed a negative burden score of -0.01279 with a ratio of 0.990, suggesting potential protective effects of certain variants in this gene. GRM3 and DAOA showed positive but modest burden scores of 0.03791 and 0.00222, respectively, with ratios of 1.045 and 1.003.

The percentage of variants showing enrichment in cases varied across genes, ranging from 31.91 percent for CACNA1C to 27.13 percent for GRM3. This variability suggests that different genes contribute to schizophrenia risk through distinct mechanisms, with some showing broad enrichment of risk variants while others demonstrate more selective patterns of case-control differences.

Analysis of variants exclusively present in cases (allele frequency zero in controls) revealed 625 such variants across all genes, with 374 low-impact variants, 232 moderate-impact variants, and 19 high-impact variants. This pattern suggests that case-specific variants span the full spectrum of predicted functional impact, though high-impact case-specific variants remain relatively rare. Table I summarizes the key findings across all genes, integrating variant counts, constraint patterns, diversity metrics, and burden differences for comparative analysis.

TABLE I. SUMMARY OF SELECTED VARIANT METRICS BY GENE

| Gene | Variants | Constraint (%) | Shannon Index | Burden Diff |
|---|---|---|---|---|
| CACNA1C | 1708 | 0.6 | 0.565 | 0.2018 |
| NRG1 | 821 | 0.5 | 0.521 | 0.2277 |
| DISC1 | 559 | 4.7 | 0.601 | -0.0128 |
| GRM3 | 527 | 3.4 | 0.638 | 0.0379 |
| COMT | 345 | 3.8 | 0.703 | 0.0016 |
| DAOA | 220 | 7.7 | 0.640 | 0.0022 |

[a.] Genes ordered by variant count. Metrics shown constraint score, Shannon index, and allele frequency burden difference

## IV. DISCUSSION

The systematic comparative analysis of variant distribution patterns across six key schizophrenia-associated genes reveals fundamental insights into their distinct roles in disease pathogenesis and provides a novel framework for understanding gene-specific contributions to psychiatric disorders. These findings complement the broader genomic landscape of schizophrenia established by the Psychiatric Genomics Consortium (PGC) GWAS, which has identified over 270 common variant loci [6], and by SCHEMA's rare variant burden analysis identifying ten high-confidence risk genes [5]. While GWAS loci primarily capture the polygenic common-variant architecture, the constraint-based framework applied here bridges rare and common variant contributions by characterizing gene-level tolerance to functional disruption across a set of biologically prioritized candidates. The dramatic 15.8-fold difference in functional constraint between DAOA and NRG1 is consistent with the gene-specific risk gradients observed in multi-ancestry PGC analyses and suggests that schizophrenia-associated genes contribute to disease risk through qualitatively different mechanisms, supporting a refined multiple-hit model where different genes operate under distinct evolutionary and functional constraints.

DAOA emerges as the most highly constrained gene in this panel, with a functional constraint density of 7.73 percent high-impact variants, a pattern that may reflect strong selective pressure against functional disruption of glutamate neurotransmission. This observation is consistent with DAOA's role as a regulator of D-amino acid oxidase, which modulates NMDA receptor co-agonist availability and glutamate signaling pathways implicated in schizophrenia pathophysiology [21], although a direct causal relationship between high constraint and clinical severity cannot be established from distributional data alone. The elevated constraint observed in DAOA could suggest that this gene represents a particularly sensitive point in glutamate signaling, though functional and clinical validation studies are required to confirm this interpretation.

In contrast, CACNA1C and NRG1 show remarkably low constraint levels of 0.59 percent and 0.49 percent respectively, indicating these genes contribute to schizophrenia risk through cumulative effects of multiple moderate-impact variants rather than single high-impact mutations. This pattern aligns with their roles in ongoing synaptic function and neural development, processes that may require fine-tuning rather than binary regulation [22]. The tolerance of these genes to genetic variation suggests they have evolved buffering mechanisms that maintain essential functions despite considerable genetic diversity.

The intermediate constraint levels observed in DISC1 (4.65 percent), COMT (3.77 percent), and GRM3 (3.42 percent) position these genes as moderately constrained, consistent with their roles as important but not absolutely critical components of neural signaling networks [23]. DISC1 constraint pattern reflects its function as a scaffolding protein where partial disruption may be tolerated but complete loss of function has severe developmental consequences. COMT moderate constraint aligns with its specific but crucial role in dopamine metabolism, where variants affecting enzyme activity have direct but modulated effects on prefrontal cortex function.

The distributional heterogeneity metrics (fold difference 7.76, coefficient of variation 70.24 percent, Gini index 0.355) collectively indicate that variant burden across schizophrenia genes follows a power-law-like distribution rather than uniform allocation. This pattern suggests that schizophrenia genetic architecture is characterized by a few genes harboring many variants and several genes with relatively few variants, potentially reflecting different evolutionary histories and functional constraints [24].

The burden analysis reveals additional complexity in gene-specific risk patterns, with NRG1 and CACNA1C showing the highest cumulative case-control differences despite their low functional constraint. This apparent paradox suggests that these genes contribute to risk through many small-effect variants rather than rare large-effect variants, consistent with current models of polygenic psychiatric disorders [25]. The negative burden observed for DISC1 warrants further investigation but

may reflect protective haplotypes or ascertainment biases in the SCHEMA dataset.

Several important limitations affect the interpretation of these findings. First, the SCHEMA dataset's predominant European ancestry composition (estimated at over 70% of samples) constrains the generalizability of the observed constraint patterns to non-European populations, as allele frequency spectra and linkage disequilibrium structures differ substantially across ancestries. Variant pathogenicity scores and functional constraint metrics calibrated on European-majority datasets may not transfer reliably to populations of African, East Asian, South Asian, or admixed ancestry. Replication in non-European cohorts such as the African Ancestry Psychiatric Genetics (AAPG) initiative or the iPSYCH cohort is therefore a necessary next step before these findings can inform globally equitable clinical practice. Second, the absence of independent validation in external cohorts means that the gene-specific constraint signatures reported here remain observational. Without replication in a held-out dataset with comparable exome coverage and quality control standards, the possibility of dataset-specific biases — including batch effects, site-level variant calling artifacts, or ascertainment differences — cannot be fully excluded. Third, the selective availability of statistical test results (only 4.3 percent of variants with P-values) introduces systematic bias toward variants meeting specific quality control criteria, potentially missing important signals in the untested variant pool. Fourth, functional constraint metrics derived from population variation data (e.g., pLI, LOEUF) reflect evolutionary selection pressures but do not directly measure functional impact in neuronal cell types relevant to schizophrenia. Future studies should prioritize comprehensive variant testing, inclusion of ancestrally diverse populations, and functional validation in disease-relevant cellular models to ensure equitable advancement of precision psychiatry.

The functional constraint framework established here has direct implications for clinical genomics and precision psychiatry applications [27]. Variants in highly constrained genes like DAOA should receive priority in clinical variant interpretation pipelines: when a patient carries a rare high-impact variant in DAOA or DISC1, the constraint profile established here provides supporting evidence that such variants are less likely to be tolerated in healthy individuals, strengthening their classification under ACMG/AMP criteria (specifically criteria PP2 and BS1). Conversely, for moderately constrained genes such as CACNA1C and NRG1, clinical interpretation should shift toward polygenic burden approaches, where cumulative variant scores — rather than individual variant pathogenicity — are more informative. In the context of diagnostic support, constraint-stratified gene panels could prioritize sequencing depth and reporting thresholds differently for high-constraint versus low-constraint genes, improving the positive predictive value of genomic findings in psychiatric evaluation. For predictive modeling, the constraint signatures identified here can serve as gene-level features in machine learning models that integrate polygenic risk scores with rare variant burden, a direction that has shown early promise in stratifying treatment response in antipsychotic pharmacogenomics research [12], [27].

These findings also inform therapeutic development strategies by highlighting different intervention points across the genetic architecture of schizophrenia [28]. Highly constrained genes may be suitable targets for precision medicine approaches where rare high-impact variants dictate treatment selection, while low-constraint genes may benefit from broader pharmacological interventions targeting pathway-level dysfunction rather than specific genetic variants.

## V. CONCLUSIONS

This study establishes the first systematic framework for understanding gene-specific contributions to schizophrenia through comparative analysis of functional constraint patterns. The 15.8-fold difference in constraint between DAOA (7.73 percent) and NRG1 (0.49 percent) demonstrates that schizophrenia-associated genes operate through fundamentally different mechanisms—highly constrained genes contributing through rare high-impact variants versus tolerant genes acting through cumulative moderate-impact effects. This mechanistic diversity provides a critical foundation for precision psychiatry by enabling gene-specific rather than uniform approaches to variant interpretation and therapeutic targeting.

The practical implications for clinical application are immediate and transformative. Highly constrained genes like DAOA and DISC1 should be prioritized for rare variant analysis and functional validation in clinical genomics, while genes like CACNA1C and NRG1 require polygenic approaches incorporating cumulative variant effects. This constraint-based prioritization framework directly addresses the current challenge in psychiatric genetics of determining which variants warrant clinical attention, providing evidence-based guidance for precision medicine implementation.

Future applications of this framework extend beyond variant prioritization to therapeutic development and personalized treatment selection. The distinct constraint patterns identified here suggest that precision psychiatry must abandon one-size-fits-all approaches in favor of gene-specific strategies that match intervention targets to underlying genetic mechanisms. By establishing clear principles for translating genetic architecture into clinical actionability, this work provides the foundation for advancing precision psychiatry from research concept to clinical reality.

## REFERENCES


[1] McGrath J., Saha S., Chant D., y Welham J., "Schizophrenia: a concise overview of incidence, prevalence, and mortality," _Epidemiological Reviews_, vol. 30, nº 1, pp. 67–76, 2008.

[2] Sullivan P. F., Kendler K. S., y Neale M. C., "Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies," _Archives of General Psychiatry_, vol. 60, nº 12, pp. 1187–1192, 2003.

[3] Andreassen O. A., Hindley G. F. L., Frei O., y Smeland O. B., "New insights from the last decade of research in psychiatric genetics: discoveries, challenges and clinical implications," _World Psychiatry_, vol. 22, nº 1, pp. 4–24, 2023.

[4] Owen M. J., "Genomic findings in schizophrenia and their implications," _Molecular Psychiatry_, vol. 28, nº 9, pp. 3638–3647, 2023.

[5] Singh T., Poterba T., Curtis D., et al., "Rare coding variants in ten genes confer substantial risk for schizophrenia," _Nature_, vol. 604, nº 7906, pp. 509–516, 2022.

[6] Trubetskoy V., Pardiñas A. F., Qi T., et al., "Mapping genomic loci prioritises genes and synaptic biology in schizophrenia," _Nature_, vol. 604, nº 7906, pp. 502–508, 2022.

[7] Green E. K., Grozeva D., Jones I., et al., "The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia," _Molecular Psychiatry_, vol. 15, nº 10, pp. 1016–1022, 2010.

[8] Millar J. K., Wilson-Annan J. C., Anderson S., et al., "Disruption of two novel genes by a translocation co-segregating with schizophrenia," _Human Molecular Genetics_, vol. 9, nº 9, pp. 1415–1423, 2000.

[9] Harrison P. J., y Weinberger D. R., "Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence," _Molecular Psychiatry_, vol. 10, nº 1, pp. 40–68, 2005.

[10] Karczewski K. J., Francioli L. C., Tiao G., et al., "The mutational constraint spectrum quantified from variation in 141,456 humans," _Nature_, vol. 581, nº 7808, pp. 434–443, 2020.

[11] Chen S., Francioli L. C., Goodrich J. K., et al., "A genomic mutational constraint map using variation in 76,156 human genomes," _Nature_, vol. 625, nº 7993, pp. 92–100, 2024.

[12] Reay W. R., y Cairns M. J., "Genetics-informed precision treatment formulation in schizophrenia and bipolar disorder," _American Journal of Human Genetics_, vol. 109, nº 9, pp. 1620–1635, 2022.

[13] Cross-Disorder Group of the Psychiatric Genomics Consortium, "Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders," _Cell_, vol. 179, nº 7, pp. 1469–1482, 2019.

[14] Palmer D. S., Howrigan D. P., Chapman S. B., et al., "Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia," _Nature Genetics_, vol. 54, nº 5, pp. 541–547, 2022.

[15] Eckart N., Song Q., Yang R., et al., "Functional characterization of schizophrenia-associated variation in CACNA1C," _PLOS ONE_, vol. 11, nº 6, e0157086, 2016.

[16] Egan M. F., Goldberg T. E., Kolachana B. S., et al., "Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia," _Proceedings of the National Academy of Sciences_, vol. 98, nº 12, pp. 6917–6922, 2001.

[17] Stefansson H., Sigurdsson E., Steinthorsdottir V., et al., "Neuregulin 1 and susceptibility to schizophrenia," _American Journal of Human Genetics_, vol. 71, nº 4, pp. 877–892, 2002.

[18] McLaren W., Gil L., Hunt S. E., et al., "The Ensembl Variant Effect Predictor," _Genome Biology_, vol. 17, nº 1, p. 122, 2016.

[19] Wang Q., Pierce-Hoffman E., Cummings B. B., et al., "Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes," _Nature Communications_, vol. 11, nº 1, p. 2539, 2020.

[20] Han S., "Bayesian rare variant analysis identifies novel schizophrenia putative risk genes," _Journal of Personalized Medicine_, vol. 14, nº 8, p. 822, 2024.

[21] Burnet P. W., Eastwood S. L., Bristow G. C., et al., "D-amino acid oxidase activity and expression are increased in schizophrenia," _Molecular Psychiatry_, vol. 13, nº 7, pp. 658–660, 2008.

[22] Moon A. L., Haan N., Wilkinson L. S., et al., "CACNA1C: association with psychiatric disorders, behavior, and neurogenesis," _Schizophrenia Bulletin_, vol. 44, nº 5, pp. 958–965, 2018.

[23] Millar J. K., Pickard B. S., Mackie S., et al., "DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling," _Science_, vol. 310, nº 5751, pp. 1187–1191, 2005.

[24] Weiner D. J., Nadig A., Jagadeesh K. A., et al., "Polygenic architecture of rare coding variation across 394,783 exomes," _Nature_, vol. 614, nº 7948, pp. 492–499, 2023.

[25] Liu D., Meyer D., Fennessy B., et al., "Schizophrenia risk conferred by rare protein-truncating variants is conserved across diverse human populations," _Nature Genetics_, vol. 55, nº 3, pp. 369–376, 2023.

[26] Martin A. R., Kanai M., Kamatani Y., et al., "Clinical use of current polygenic risk scores may exacerbate health disparities," _Nature Genetics_, vol. 51, nº 4, pp. 584–591, 2019.

[27] Warren T. L., Tubbs J. D., Lesh T. A., et al., "Rare coding variants in SETD1A and other KMT2 family members contribute to risk for schizophrenia," _Molecular Psychiatry_, vol. 29, nº 8, pp. 2389–2398, 2024.

[28] Minikel E. V., Karczewski K. J., Martin H. C., et al., "Evaluating drug targets through human loss-of-function genetic variation," _Nature_, vol. 581, nº 7809, pp. 459–464, 2020..