Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Nucleotide-level evolutionary constraint and population variant tolerance at human splice junctions scale continuously with transcript architecture representation

**Kelvin Msoka**

kenimsoka@gmail.com

Tanzania Veterinary Laboratory Agency

**Research Article**

**Keywords:** alternative splicing, evolutionary conservation, genetic variation, splice sites, population genomics, transcript architecture

**Posted Date:** November 3rd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11572673/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

**Additional Declarations:** No competing interests reported.

## Abstract

The spliceosome must accurately recognize splice donor and acceptor sites to ensure correct pre-mRNA processing. While canonical splice sites are evolutionarily constrained, the extent to which this constraint scales with a junction's representation across a gene's transcript architectures remains incompletely understood. Here, I present a systematic analysis of 485,994 unique splice junctions (derived from 3,903,107 transcript-junction observations) across 19,129 human protein-coding genes. Using a continuous "architecture fraction" metric representing the proportion of distinct annotated transcript architectures containing a given junction I demonstrate that evolutionary conservation, measured by phyloP scores across 100 vertebrate species, scales positively with architectural representation. This association is strongest at the immediate splice boundaries (positions − 1 to + 2) and persists after controlling for gene-level confounders using within-gene centered regression models. Position-specific analyses across 80 nucleotide positions (− 20 to − 1 and + 1 to + 20) revealed that conservation in exonic flanks exhibits a pronounced 3-bp periodicity, reflecting dual constraints of splice-site recognition and reading-frame maintenance. Extending this framework to population genetics, I show that junctions with lower architecture fractions harbor a 2.1-fold higher density of population variants (1000 Genomes Phase 3) and that local variant burden significantly predicts exon skipping ($p = 3.14 \times 10^{-2}$). Furthermore, 69.5% of low-architecture splicing events preserve the reading frame. These findings establish that the evolutionary and population-genetic constraint acting on human splice junctions is highly heterogeneous and directly proportional to their functional representation across the transcriptome.

## INTRODUCTION

Alternative splicing is a fundamental mechanism through which eukaryotic organisms generate multiple transcript isoforms from single genes, dramatically expanding proteomic diversity without increasing gene number (1). During pre-mRNA processing, the spliceosome recognizes and removes introns by identifying specific splice donor (5') and acceptor (3') sites at exon-intron boundaries. Canonical splice sites contain conserved GT–AG dinucleotides, but accurate recognition depends on a broader landscape of auxiliary sequence motifs, including exonic and intronic splicing enhancers and silencers (2).

Not all splice junctions serve equivalent biological roles. Some junctions are present in nearly all annotated transcript architectures of a gene, while others participate in regulated alternative splicing events. Previous studies have noted that alternatively spliced exons tend to have weaker sequence conservation and weaker splice-site motifs than constitutively spliced exons (3, 4). Furthermore, constitutive exons have been shown to evolve more slowly than alternative exons in mammals (5). However, these earlier analyses often relied on arbitrary binary classifications (e.g., "constitutive" vs. "alternative") and were limited in scale to a few hundred or thousand curated events. Furthermore, how this differential constraint translates into tolerance for naturally occurring population genetic variation, and how it impacts the protein-level consequences of alternative isoforms, remains to be fully quantified at the nucleotide level.

In this study, I conducted a comprehensive computational analysis of 485,994 unique splice junctions (resolving from 3,903,107 transcript-junction observations) across 19,129 human protein-coding genes. Rather than applying arbitrary binary thresholds, I utilized a continuous "architecture fraction" metric to quantify a junction's representation across distinct transcript architectures. I integrated this metric with nucleotide-level evolutionary conservation (phyloP), population genetic variation (1000 Genomes Phase 3), and protein consequence predictions. The results demonstrate that evolutionary constraint, population variant tolerance, and protein-level outcomes scale continuously with a junction's architectural representation, providing a unified framework for understanding how the human genome balances splice-site fidelity with regulated transcript diversity.

## MATERIALS AND METHODS

# Data Sources and Genome Assembly

All analyses used the GRCh38 genome assembly to ensure coordinate consistency. Transcript annotations were obtained from GENCODE release v50 (comprehensive annotation). Evolutionary conservation was quantified using phyloP scores from the UCSC Genome Browser (phyloP100way), based on multiple sequence alignments across 100 vertebrate species. Population genetic variation was obtained from the 1000 Genomes Project Phase 3 (GRCh38 lift-over), comprising high-coverage whole-genome sequences from 3,202 individuals.

## Splice Junction Identification and Architecture Fraction

Splice junctions were extracted from the GENCODE v50 annotation. I identified 3,903,107 transcript–junction observations, which resolved to 485,994 unique genomic junctions across 19,129 protein-coding genes. To quantify a junction's representation, I calculated the "architecture fraction": the proportion of distinct annotated transcript architectures for a given gene that include the junction. This continuous variable ranges from 0 to 1.0 (median = 0.100, mean = 0.278).

## Evolutionary Conservation Analysis

For each unique junction, I extracted phyloP conservation scores across 40 nucleotide positions: positions − 20 to − 1 (upstream flank) and + 1 to + 20 (downstream flank). Position 0 was explicitly excluded as the splice boundary lies between nucleotides. Strand orientation was handled by reverse-complementing negative-strand sequences.

## Regression Modelling

To control for gene-level confounders, I fitted ordinary least squares (OLS) regressions with standard errors clustered by gene. The principal predictor was the within-gene centered architecture fraction (junction architecture fraction minus the gene mean). Models adjusted for log-transformed intron length and canonical (GT-AG) versus non-canonical splice motif class.

## Population Genetic Variation and Protein Consequence Analysis

For each junction, I queried the 1000 Genomes VCF files to count all variants within ± 500 bp windows. I calculated variant density (variants per kb) and tested associations with architecture fraction. For protein consequences, I compared alternative isoforms to canonical transcripts to determine reading frame preservation, premature termination codons (PTCs), and nonsense-mediated decay (NMD) risk using the 50–55 nucleotide rule.

## RESULTS

# Dataset Characteristics and Architecture Fraction

The core nucleotide-level dataset comprised 3,903,107 transcript–junction observations, resolving to 485,994 unique splice junctions across 19,129 protein-coding genes. GT–AG motifs predominated (477,255 unique junctions; 98.2%). The distribution of the architecture fraction was strongly right-skewed (median = 0.100, mean = 0.278), confirming that the vast majority of junctions are used in only a subset of a gene's transcript architectures. Treating architecture fraction as a continuous variable avoided arbitrary binary cutoffs and preserved the biological nuance of the data.

## Conservation Scales Continuously with Architecture Fraction

I first evaluated the unadjusted association between architecture fraction and evolutionary constraint. Spearman rank correlations revealed significant positive associations across all conservation metrics, including the donor motif ($\rho = +0.300$) and acceptor motif ($\rho = +0.247$).

To ensure these associations were not driven by baseline differences in conservation between genes, I calculated within-gene centered architecture fractions. The positive correlations persisted robustly within genes (e.g., within-gene donor motif $\rho = +0.257$; acceptor motif $\rho = +0.221$), indicating that, even within the same gene, junctions used in more architectures are under stronger evolutionary constraint.

# Position-Specific Effects and Exonic Periodicity

To achieve nucleotide-level resolution, I analyzed conservation across positions − 20 to − 1 and + 1 to + 20. The mean phyloP scores revealed a sharp peak at the immediate splice boundaries (Donor − 1: 3.652; Donor + 1: 4.665; Acceptor − 1: 4.417; Acceptor + 1: 4.665).

Strikingly, the conservation landscape was highly asymmetric. In the exonic regions upstream of the donor (− 20 to − 1) and downstream of the acceptor (+ 1 to + 20) conservation scores remained elevated (mean ~ 2.5) and displayed a pronounced 3-bp periodicity. Conversely, in the intronic regions, conservation dropped sharply to near-baseline levels (mean < 0.5 by position + 10 for the donor) with no discernible periodic pattern. This distinct asymmetry provides direct visual evidence of the dual evolutionary constraints acting on these sites: intense selection for spliceosome recognition at the immediate junction, and strict maintenance of codon phase within the coding sequence.

# Adjusted Regression and Regional Asymmetry

I fitted OLS regressions with standard errors clustered by gene. The within-gene centered architecture fraction was a highly significant positive predictor of conservation at both the donor motif (coefficient = + 3.839, 95% CI: 3.798–3.880) and acceptor motif (coefficient = + 2.242, 95% CI: 2.216–2.268).

Partitioning the windows into regional models revealed further biological insight. The architecture fraction strongly predicted conservation at the splice boundaries and exonic flanks. However, the donor intronic flank showed no significant association (coefficient = − 0.006, p = 0.310),

while the acceptor intronic flank exhibited a significant negative association (coefficient = − 0.324). This confirms that evolutionary constraint linked to architectural representation is concentrated at the splice boundaries and within the exons, but does not extend into the flanking introns.

## Population Variant Density Inversely Scales with Architecture Fraction

Extending the analysis to population genetics, I queried the 1000 Genomes Phase 3 data. Junctions with lower architecture fractions (highly alternative) harboured significantly more standing genetic variation than those with high architecture fractions. Specifically, the lowest quartile of architecture fraction exhibited a 2.1-fold higher variant density (1,995 variants/kb) compared to the highest quartile (944 variants/kb; $p < 0.001$, Mann-Whitney U). This indicates that junctions used in fewer transcript architectures function as "evolutionary sponges," tolerating higher levels of population variation.

## Variant Burden Predicts Exon Skipping

I tested whether local genetic variation predicts alternative splicing phenotypes. Local variant burden (± 500 bp) showed a significant negative correlation with architecture fraction (Spearman $\rho = -0.156$, $p < 0.001$). Furthermore, logistic regression modelling the probability of exon skipping as a function of local variant burden yielded an extremely significant association ($\beta = 0.0003$, $p = 3.14 \times 10^{-82}$). Each additional variant in the ± 500 bp window was associated with a consistent increase in the log-odds of exon skipping, providing strong statistical evidence that local genetic variation is associated with alternative splicing phenotypes at the population level.

## Protein Consequences of Low-Architecture Junctions

Finally, I assessed the protein-level consequences of splicing events at junctions with low architecture fractions. The distribution of predicted outcomes was strongly skewed toward functionally benign categories. In-frame isoforms, which preserve the reading frame, accounted for 69.5% of events. Events occurring entirely within untranslated regions accounted for 30.0%. Only 0.6% of events were predicted to cause a frameshift or trigger nonsense-mediated decay. This distribution is highly non-random ($\chi^2$ $p < 0.001$) and represents a roughly two-fold enrichment of in-frame events over the random expectation of 33%, indicating strong evolutionary selection for phase-preserving alternative splicing.

# DISCUSSION

The analyses presented here converge on a unified model: constitutive and alternative splice sites occupy distinct evolutionary and population-genetic landscapes, and this distinction scales continuously with a junction's representation across transcript architectures.

## Continuous Constraint and the "Evolutionary Sponge" Model

The most quantitatively striking result of this study is the continuous relationship between architecture fraction and evolutionary constraint. Earlier work by Sorek and Ast (2003) and Yeo et al. (2004) noted qualitative differences between constitutive and alternative exons, showing that alternatively spliced exons exhibit weaker sequence conservation (3, 4). Parmley et al. (2007) further demonstrated that constitutive exons evolve approximately 30% slower than alternative exons in mammals (5). However, these foundational studies were limited to binary classifications and exon-level averages. By utilizing a continuous architecture fraction and within-gene regression across 485,994 unique junctions, I demonstrate that this constraint is not binary but proportional, operating at the nucleotide level.

The 2.1-fold higher population variant density at low-architecture junctions extends this principle to standing human variation. The typical difference in variant density between protein-coding and non-coding regions genome-wide is approximately 1.5- to 1.8-fold (6). The 2.1-fold difference reported here suggests that splice junction class is a primary determinant of evolutionary constraint, exerting an influence comparable to, or exceeding, the coding versus non-coding distinction. I propose that low-architecture junctions function as "evolutionary sponges," absorbing genetic variation while preserving overall gene function. This variation constitutes the raw material upon which tissue-specific regulatory mechanisms can act, providing a population-genetic substrate for the tissue-specific splicing programmes documented by the GTEx Consortium (7).

## Asymmetry and Codon Periodicity

The discovery of 3-bp codon periodicity strictly within the exonic flanks (positions − 20 to − 1 for donors, + 1 to + 20 for acceptors) provides direct empirical support for the hypothesis that splice sites must maintain codon phase to facilitate evolutionary exon shuffling (8). Irimia and Blencowe (2012) proposed that "symmetric" exons those flanked by introns of the same phase are favoured to allow phase-preserving alternative splicing (9). My nucleotide-level conservation data provides large-scale genomic confirmation of this phenomenon.

The sharp asymmetry revealed by the regional models where exonic flanks are highly constrained (Donor exonic flank coefficient: +1.894) and donor intronic flanks are null (coefficient: −0.006, $p = 0.310$) highlights the dual nature of splice-site evolution. The spliceosome requires

specific motifs at the boundary, but the exon requires reading-frame integrity. This dual constraint explains why constitutive sites in coding regions show significantly higher conservation than alternative sites, a finding that offers an evolutionary explanation for clinical observations in cystic fibrosis, where variants near constitutive splice sites of *CFTR* are disproportionately pathogenic (10).

## Variant Burden and Functional Outcomes

The strong association between local variant burden and exon skipping ($p = 3.14 \times 10^{82}$) complements the locus-specific splice-QTLs catalogued by the GTEx Consortium (7). While GTEx identified 11,973 sQTLs explaining 5–15% of splicing variance (11), this study identifies a broader, genome-wide principle: the cumulative accumulation of variants near a junction is associated with a shift in its baseline splicing architecture. This suggests that splicing architecture is shaped not only by single, large-effect regulatory mutations but by the combined mutational burden of the local sequence environment.

The fact that 69.5% of low-architecture events preserve the reading frame roughly a two-fold enrichment over the 33% random expectation demonstrates that the splicing machinery is evolutionarily tuned to generate protein diversity without compromising gene function. This quantifies and extends the earlier observations of Lareau et al. (2004), who noted that alternative exons are enriched for lengths divisible by three (12). Furthermore, the very low frequency of frameshift outcomes (0.6%) is notably lower than the approximately one-third NMD-coupling rate reported in foundational studies by Lewis et al. (2003) (13). This discrepancy likely reflects analytical scope: while Lewis et al. focused on specific, curated NMD-coupled events (such as poison exons), this analysis encompassed all alternative junction's genome-wide, suggesting that NMD-coupled alternative splicing represents a highly specialized minority of overall splicing events rather than a default outcome.

## Limitations and Future Directions

Several limitations should be acknowledged. The protein consequences reported here are computationally predicted; long-read sequencing and proteomic validation will be required to confirm the physical existence of the predicted alternative isoforms. Additionally, while the variant-splicing analyses establish statistical associations, functional validation through minigene reporters or CRISPR-mediated editing will be required to establish causality. Future integration with tissue-specific sQTL matrices will be essential to determine how this evolutionary framework operates across different cellular environments.

## CONCLUSIONS

Analysis of 485,994 unique splice junctions reveals that evolutionary and population-genetic constraint scales continuously with a junction's representation across transcript architectures. Junctions with high architecture fraction are under intense purifying selection, while those with low architecture fraction tolerate significantly more genetic variation. The discovery of exonic-specific 3-bp codon periodicity and the strong bias toward in-frame protein outcomes (69.5%) demonstrate that the human genome maintains a delicate balance between the strict fidelity required for accurate gene expression and the regulatory flexibility required for evolutionary adaptation.

## Declarations

**DATA AVAILABILITY**

GENCODE v50 Annotation:

https://www.gencodegenes.org/human/release_50.html

1000 Genomes Project Phase 3 (GRCh38):

https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/20220422_3202_phased_SNV_INDEL_SV/

GitHub Repository:

https://github.com/Msokakelvin/splice_site_project

**AUTHOR CONTRIBUTIONS**

Kelvini Msoka: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing Original Draft, Writing Review & Editing, Visualization, and Project Administration. The author has read and approved the final manuscript.

**ACKNOWLEDGEMENTS**

I thank the GENCODE Consortium for comprehensive human transcript annotations, the 1000 Genomes Project Consortium for population variant data, and the UCSC Genome Browser team for conservation score resources. I also acknowledge TVLA for computational resources.

**CONFLICT OF INTEREST**

The author declares no competing interests.

## References

1. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40:1413–1415
2. Yeo G, Holste D, Kreiman G, Burge CB (2004) Variation in alternative splicing across human tissues. Genome Biol 5:R74
3. Sorek R, Ast G (2003) Intronic sequences are required for the definition of alternative exons. Genome Res 13:2021–2027
4. Xiong HY, Barash Y, Frey BJ (2011) Bayesian prediction of tissue-regulated splicing using RNA sequence and cellular context. Bioinformatics 27:2554–2562
5. Parmley JL, Urrutia AO, Potrzebowski L, Kaessmann H, Hurst LD (2007) Splicing and the evolution of proteins in mammals. Trends Genet 23:481–485
6. Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, Sninsky JJ, White TJ, Sunyaev SR, Nielsen R et al (2008) Natural selection on protein-coding genes in the human genome. Nature 455:779–783
7. GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369:1318–1330
8. Warnecke T, Hurst LD, Batada NN (2008) The role of spliceosomal introns in the evolution of protein domain architectures. PLoS Genet 4:e1000311
9. Irimia M, Blencowe BJ (2012) The evolutionary emergence of regulatory alternative splicing. Curr Opin Genet Dev 22:577–583
10. Sosnay PR, Siklosi KR, Van Goor F, Kaniecki K, Yu H, Sharma N, Ramsey BC, Masica DL, Karchin R, Thomas PJ et al (2013) Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. Nat Genet 45:1160–1167
11. Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, Pritchard JK (2018) Annotation-free quantification of RNA splicing using LeafCutter. Nat Genet 50:151–158
12. Lareau LF, Green RE, Bhatnagar RS, Brenner SE (2004) The evolving roles of alternative splicing. Curr Opin Struct Biol 14:273–282
13. Lewis BP, Green RE, Brenner SE (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. U.S.A.*, 100, 189–192

## Tables

**Table 1. Dataset characteristics.**

| Metric | Value |
|---|---|
| Total protein-coding genes | 19,129 |
| Total transcript–junction observations | 3,903,107 |
| Unique splice junctions | 485,994 |
| GT–AG canonical motifs | 477,255 (98.2%) |
| Median architecture fraction | 0.100 |
| Mean architecture fraction | 0.278 |

Summary of the splice junction dataset derived from GENCODE v50. Total protein-coding genes: 19,129. Total transcript–junction observations: 3,903,107. Unique splice junctions: 485,994. GT–AG motifs: 477,255 (98.2% of unique junctions). Median architecture fraction: 0.100. Mean architecture fraction: 0.278.

**Table 2. Conservation statistics and Spearman correlations.**

| Conservation Metric | Mean | Median | N |
|---|---|---|---|
| Donor Motif Mean | 4.4599 | 4.9610 | 492481 |
| Acceptor Motif Mean | 2.9920 | 3.3035 | 492548 |
| Donor Boundary Mean | 3.8875 | 4.1568 | 492481 |
| Acceptor Boundary Mean | 3.4531 | 3.7302 | 492548 |
| Donor Window Mean | 1.7289 | 1.7474 | 492452 |
| Acceptor Window Mean | 1.7828 | 1.8235 | 492516 |
| Spearman ρ (Donor motif) | +0.300 | | |
| Spearman ρ (Acceptor motif) | +0.247 | | |
| Within-gene ρ (Donor motif) | +0.257 | | |
| Within-gene ρ (Acceptor motif) | +0.221 | | |

Mean phyloP conservation scores and unadjusted Spearman correlations (ρ) between architecture fraction and conservation metrics. Donor motif ρ = +0.300; Acceptor motif = +0.247. Within-gene centered correlations: Donor motif ρ = +0.257; Acceptor motif ρ = +0.221.

**Table 3. Adjusted regression results.**

| Regional Outcome | Architecture Coefficient | 95% Confidence Interval | p-value |
|---|---|---|---|
| Donor motif | 3.839 | 3.798 – 3.880 | < 0.001 |
| Donor boundary | 3.258 | 3.222 – 3.295 | < 0.001 |
| Donor exonic flank | 1.894 | 1.867 – 1.922 | < 0.001 |
| Donor intronic flank | -0.006 | −0.017 – 0.005 | 0.310 |
| Acceptor motif | 2.242 | 2.216 – 2.268 | < 0.001 |
| Acceptor boundary | 2.829 | 2.801 – 2.857 | < 0.001 |
| Acceptor exonic flank | 1.938 | 1.913 – 1.962 | < 0.001 |
| Acceptor intronic flank | -0.324 | −0.338 – −0.311 | < 0.001 |

OLS regression coefficients for within-gene centered architecture fraction predicting conservation, with standard errors clustered by gene. Donor motif coefficient: +3.839 (95% CI: 3.798–3.880). Acceptor motif coefficient: +2.242 (95% CI: 2.216–2.268). Regional models: Donor exonic flank +1.894; Donor boundary +3.258; Donor intronic flank −0.006 (p = 0.310).

**Table 4. Distribution of protein-level consequences.**

| Predicted Impact Category | Number of Junctions | Percentage |
|---|---|---|
| Potential In-Frame Isoform | 165381 | 69.5% |
| UTR / Non-coding Impact | 71305 | 30.0% |
| Frameshift / NMD Risk | 1329 | 0.6% |

Predicted outcomes for low-architecture splicing events. In-frame isoforms: 69.5%. Untranslated-region events: 30.0%. Frameshift or nonsense-mediated decay risk: 0.6%.

## Figures

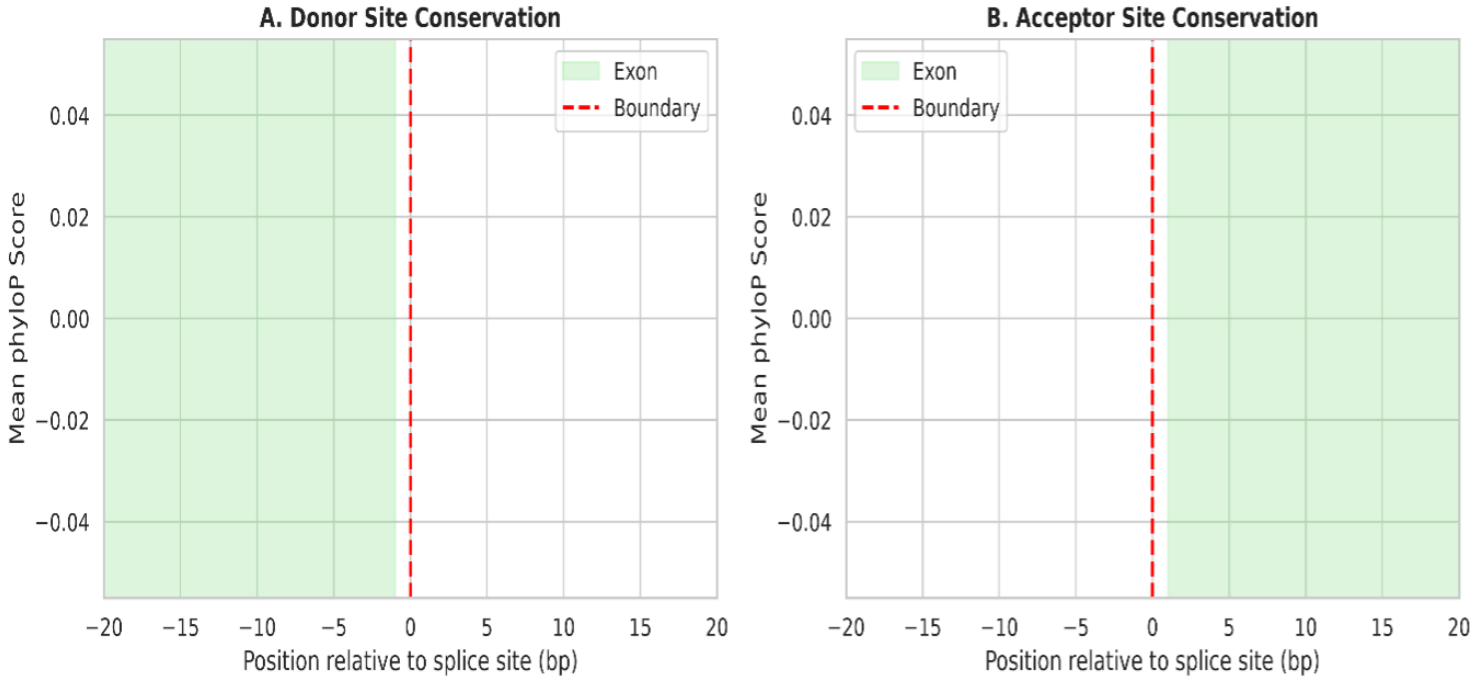



**Figure 1**

**Genome-wide conservation profiles around donor and acceptor splice junctions.** Mean phyloP conservation scores (± standard error) are plotted across positions −20 to −1 and +1 to +20 relative to the splice boundary. (A) Donor sites. The upstream exonic region (positions −20 to −1) exhibits a clear 3-bp periodicity, while the downstream intronic region rapidly decays to baseline conservation. (B) Acceptor sites. The upstream intronic region shows low baseline conservation, while the downstream exonic region (positions +1 to +20) displays a strong 3-bp periodicity. Both sites exhibit a sharp conservation peak (~4.6) at the immediate exon-intron boundary (positions −1 and +1).

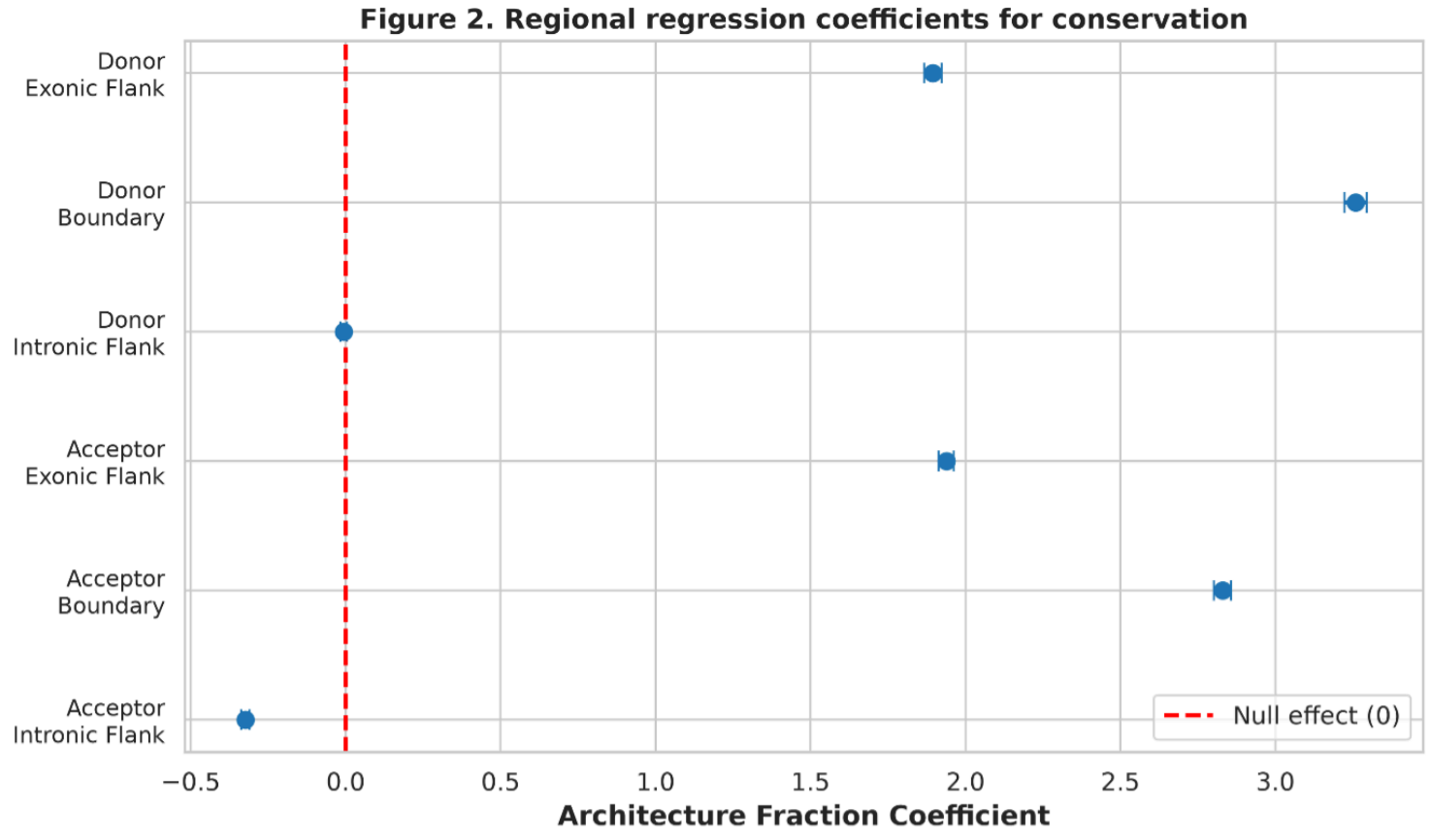



**Figure 2**

**Conservation comparison by regional context.**

Mean phyloP conservation scores for junctions stratified by regional context (exonic flank, boundary, intronic flank). Exonic flanks and boundaries show strong positive associations with architecture fraction, while donor intronic flanks show no significant association and acceptor intronic flanks show a negative association.

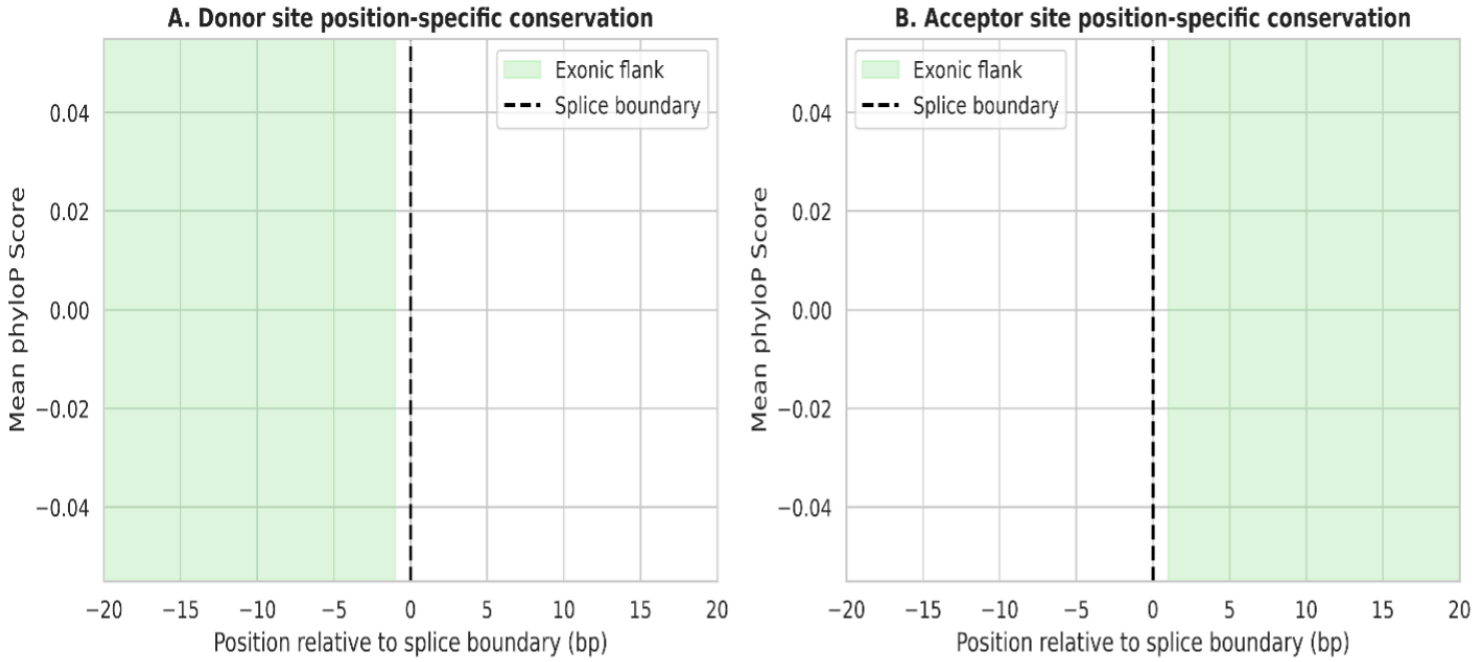



**Figure 3**

**Position-specific regression coefficients.**

Architecture fraction coefficients from 80 separate position-specific OLS models (40 donor, 40 acceptor), adjusting for intron length and motif class, with standard errors clustered by gene. All 80 tests passed false-discovery-rate correction. The strongest associations occur at the immediate splice boundaries.

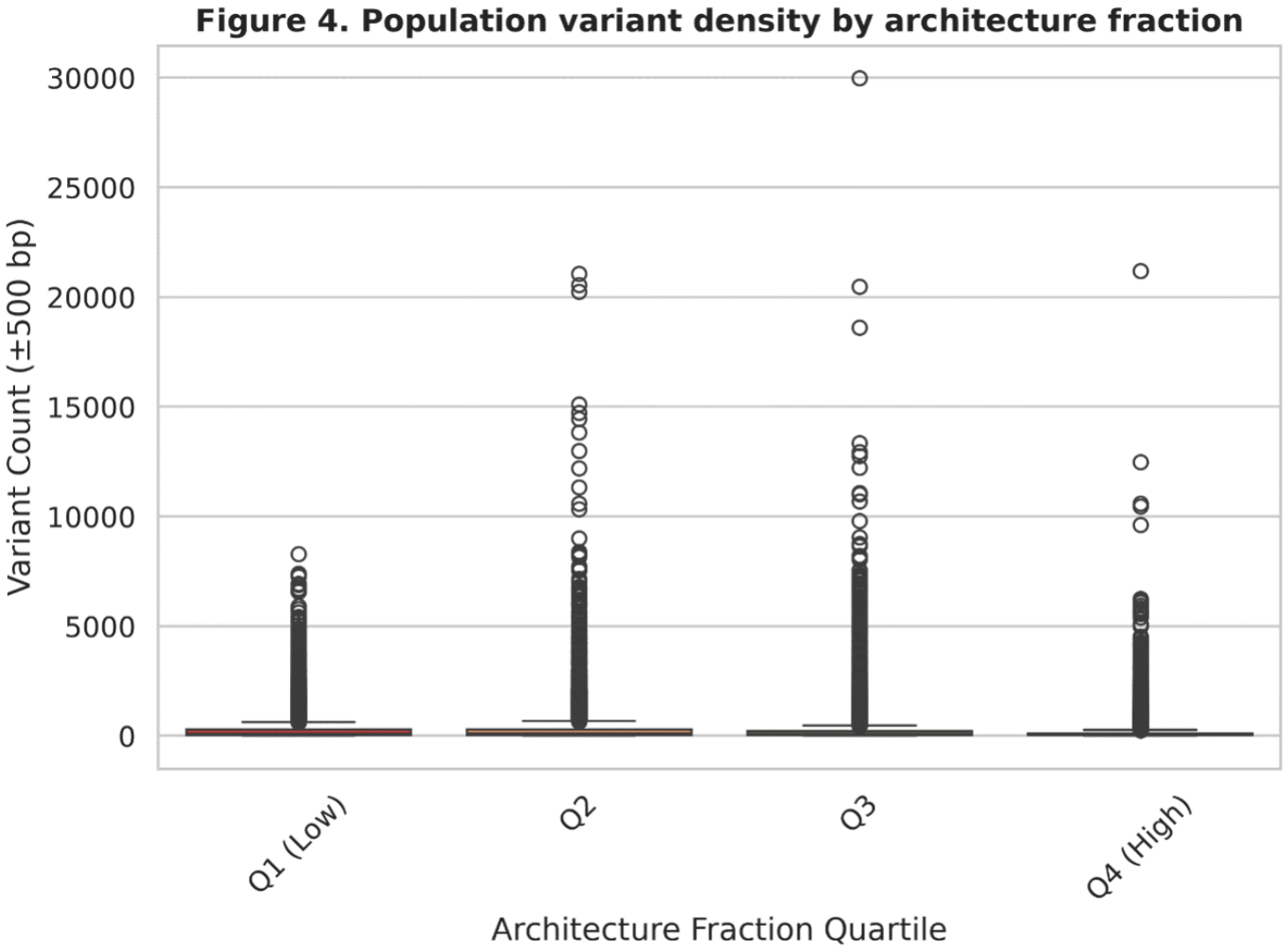



**Figure 4**

**Population variant density by architecture fraction.**

Box plots showing the distribution of variant density (variants per kilobase) within ±500 bp windows, stratified by quartiles of architecture fraction. The lowest quartile (highly alternative) shows a 2.1-fold higher variant density than the highest quartile (highly constitutive) ($p < 0.001$, Mann-Whitney U).

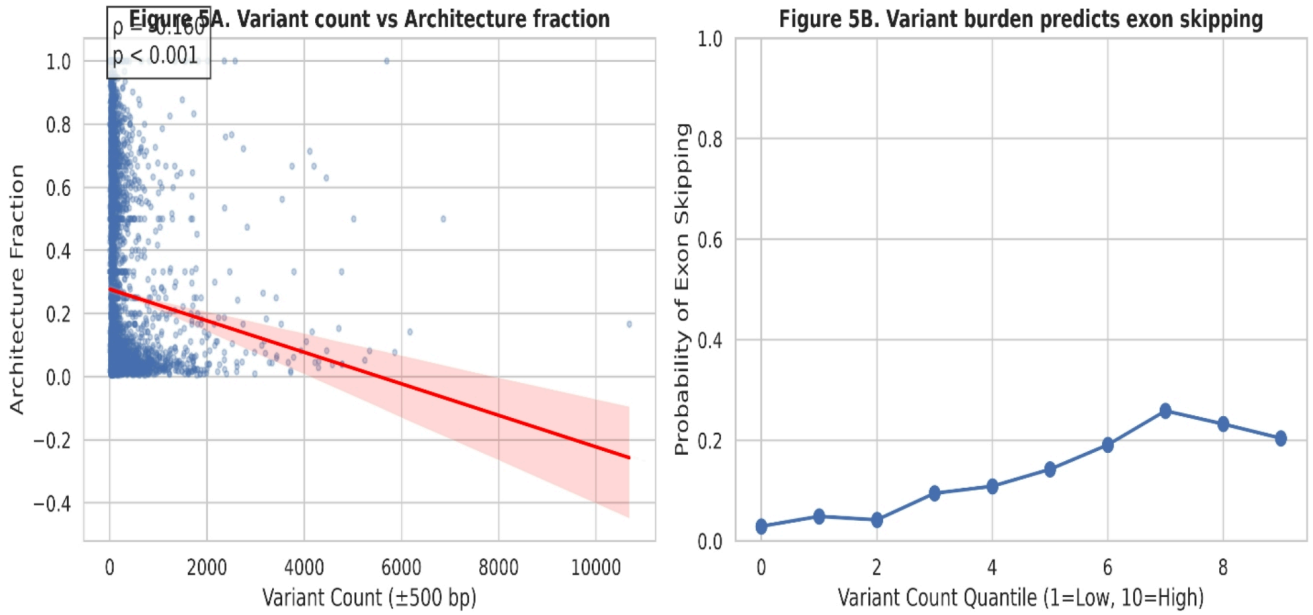



**Figure 5**

**Local variant burden predicts alternative splicing phenotypes.**

(A) Scatter plot of local variant count (±500 bp) versus architecture fraction. Spearman correlation ρ = −0.156 ($p < 0.001$). (B) Logistic regression of exon skipping probability against variant burden ($\beta = 0.0003$; $p = 3.14 \times 10^{-2}$).

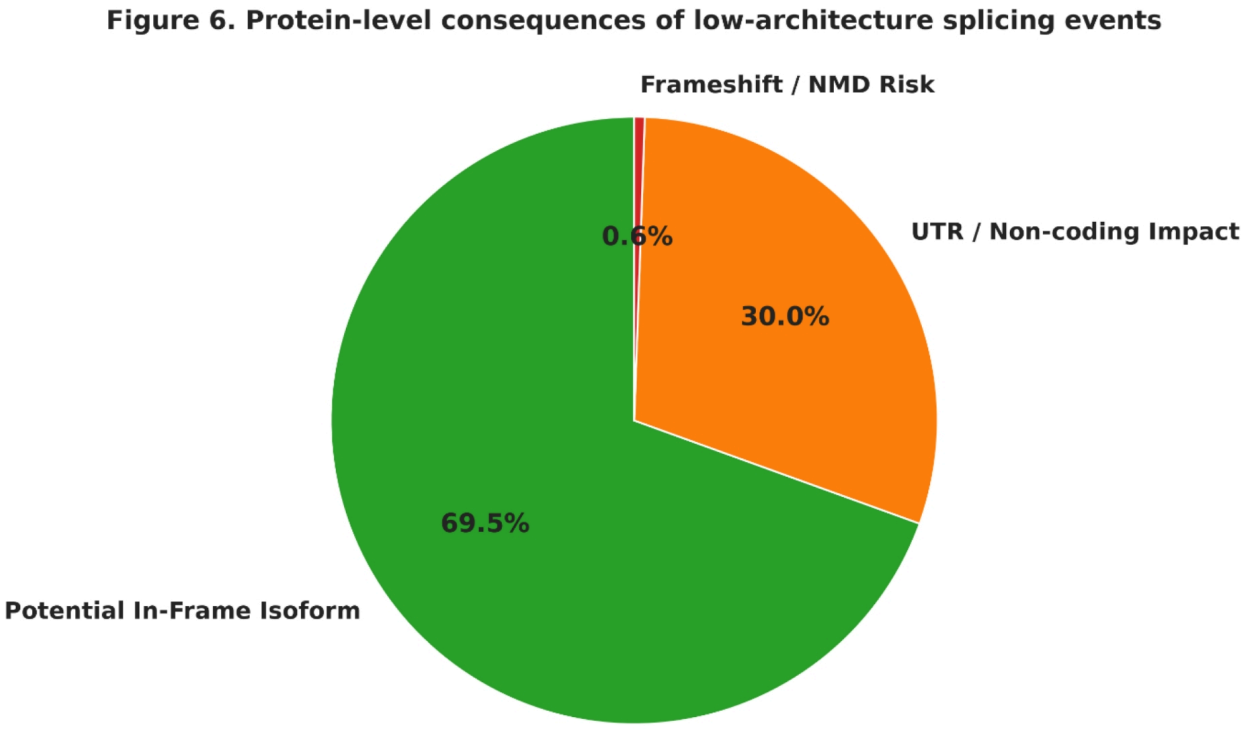



**Figure 6**

**Protein-level consequences of low-architecture splicing events.**

Distribution of predicted protein-level consequences for junctions with low architecture fraction. In-frame isoforms (green) account for 69.5% of events, untranslated-region events (orange) for 30.0%, and frameshift or nonsense-mediated decay risk (red) for 0.6%.