Decoding the Intronic Signature of Cancer

IDSC Fellows

Dec. 3. 2025

Decoding the Intronic Signature of Cancer

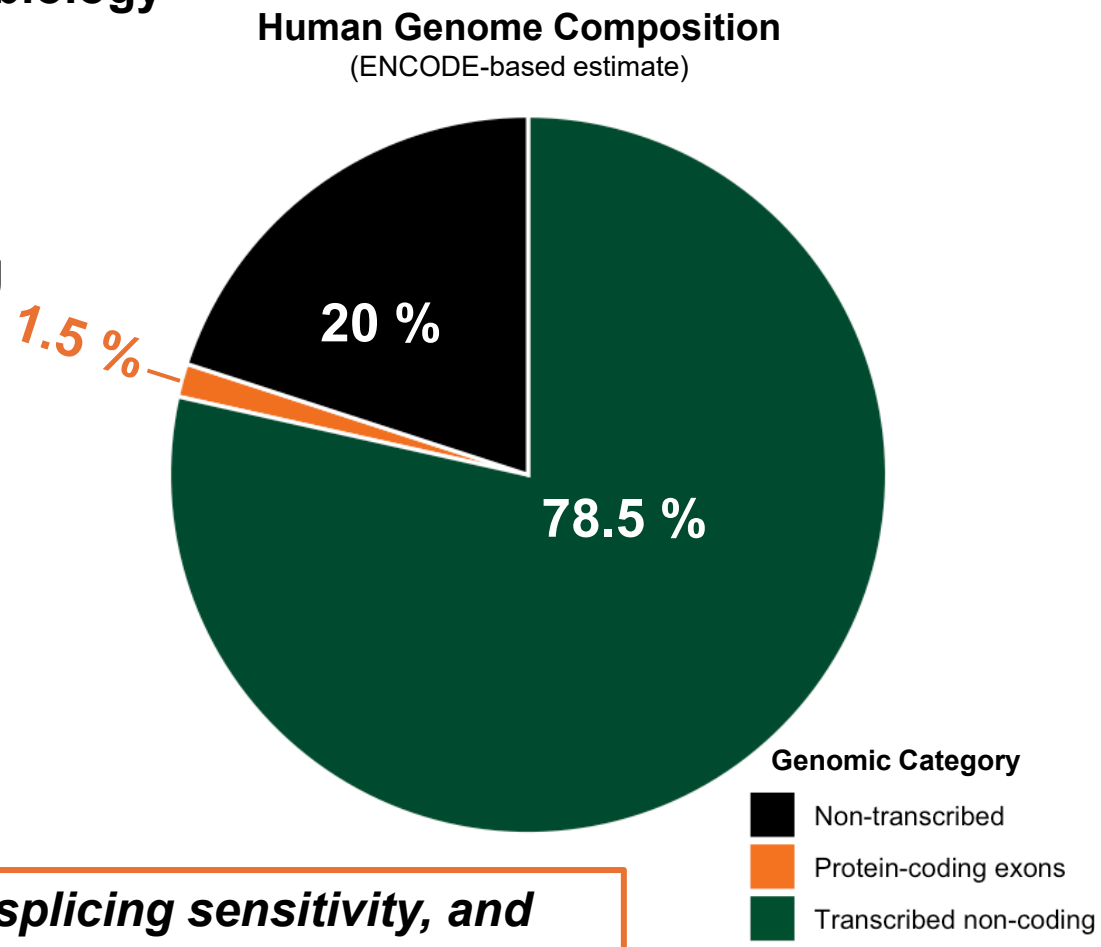
Linking intron subtypes, splicing dysregulation, and tumor biology

Only **~1-2%** of the human genome encodes proteins

~78.5% of the human genome is transcribed non-coding regions, including introns.

Because most transcribed sequences are non-coding, these regions play major regulatory roles in transcription and RNA processing.

Disruptions in the non-coding genome are increasingly recognized as a **major drivers of human disease**, including **cancer.**



How do different intron subtypes shape mutation burden, splicing sensitivity, and pathway-level vulnerabilities in cancer genomes?

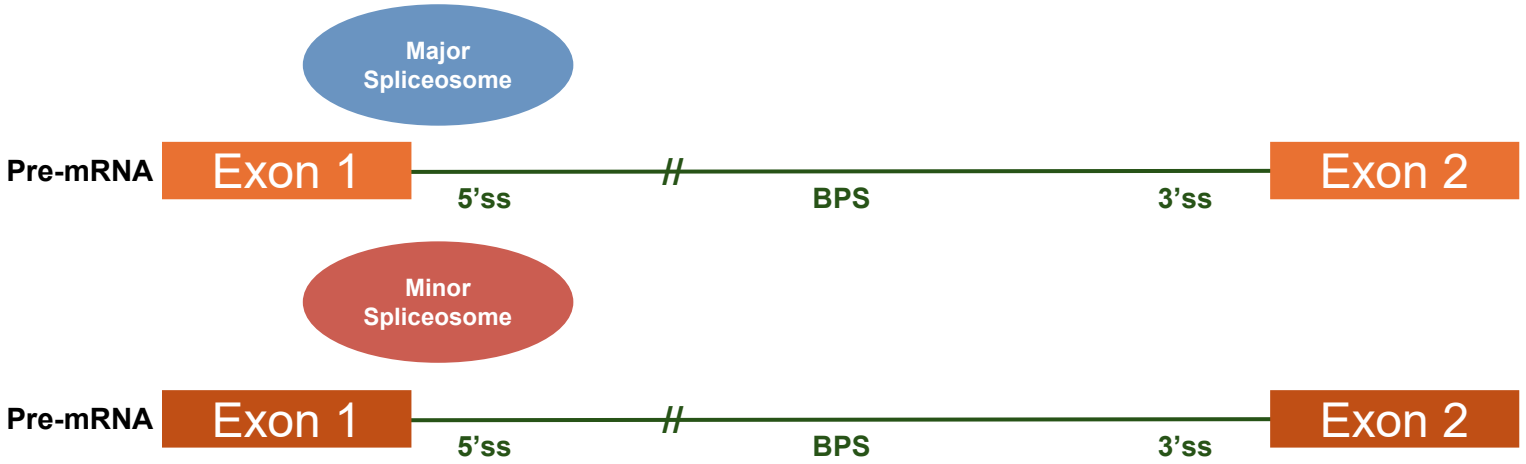
[1] Evans et al., 2016

What Are Introns?

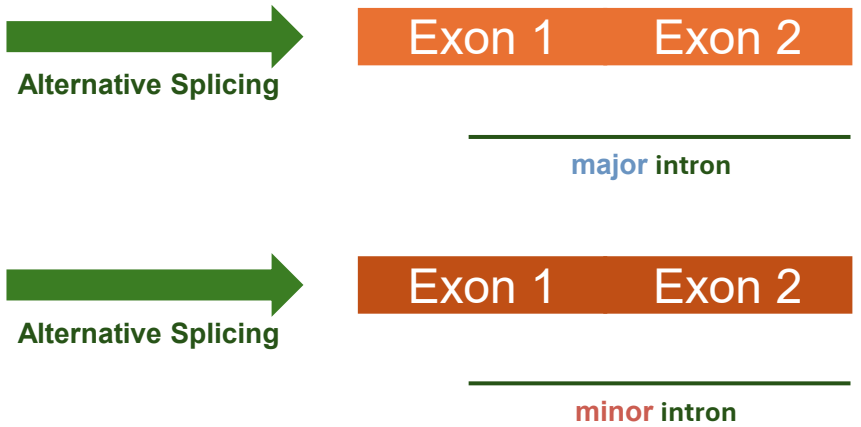
UNIVERSITY OF MIAMI

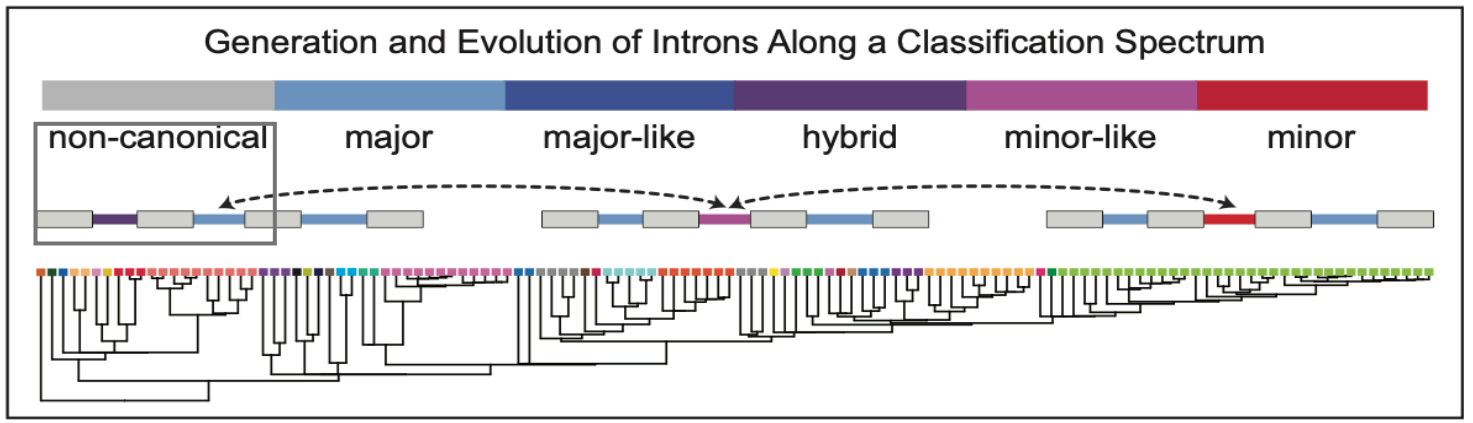
Non-coding sequences within genes, removed during RNA processing

Not all introns are equal – Processed by **distinct spliceosomes**



Distinct spliceosomes recognize and process major **vs.** minor **introns**





Olthof *et al.* (2024) reclassified introns into **six distinct types** based on splice site features and evolutionary signature: **minor**, **minor-like**, **hybrid**, **major-like**, **major** and **non-canonical**

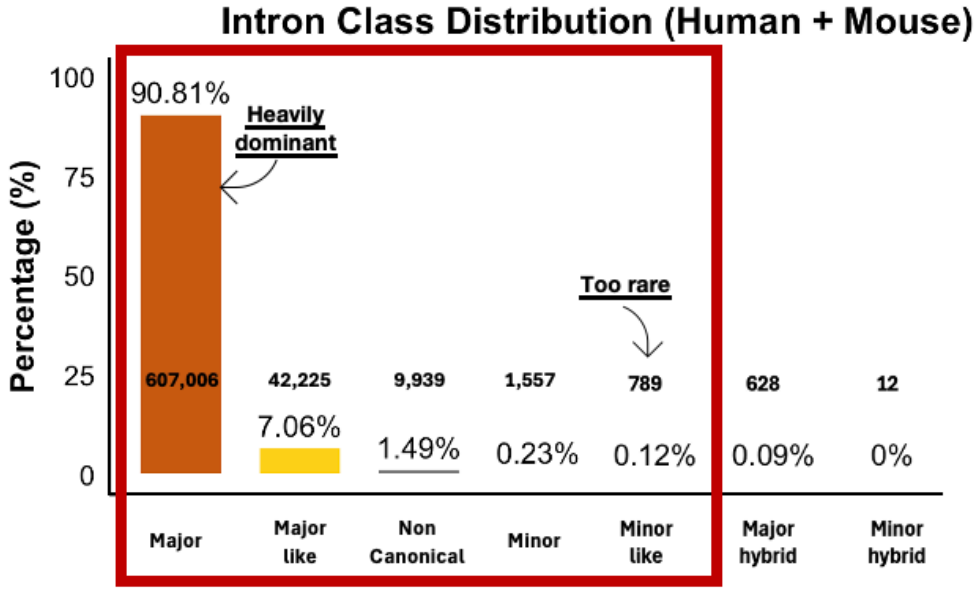
Research Overview

Integrating intron subtypes with pan-cancer mutations

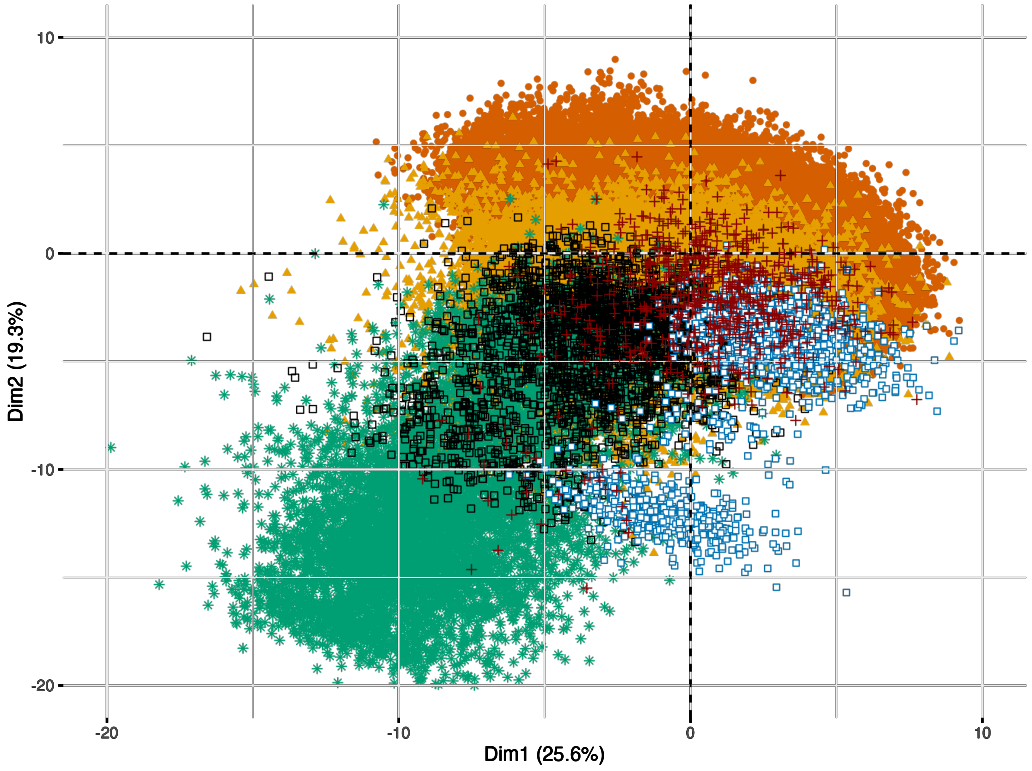
1. **Annotate refined intron subtypes across TCGA + POC matched tumor/normal**
2. **Identify subtype-enriched mutations:**
   - Splice-site disruptions
   - Branchpoint mutation
   - Loss of Function variants / NMD-triggering variants
   - Enrichment analysis vs. background mutation rate
3. **Connect intron subtypes to oncogenic pathways**
4. **Quantify subtype-specific splicing sensitivity**

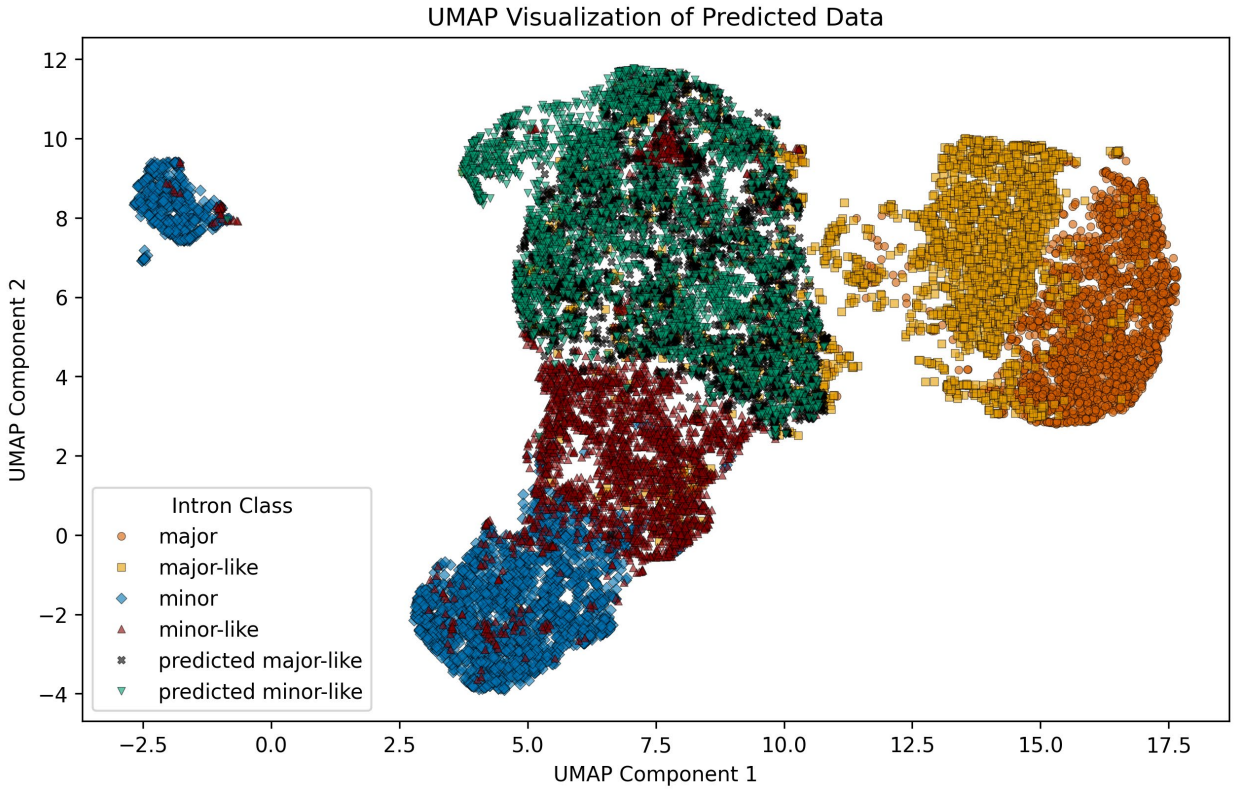
Computational Challenges

- **Severe Intron Class imbalance**
- **Massive scale integration**: millions of mutations × intron subtypes
- **Subtype sensitivity detection** under noisy cancer genomes
- **Extracting real signals** from mutational background
- **Linking mutation → subtype → pathway** in a unified model



Preliminary Result: Non-Canonical Introns Reveal Unresolved Structure



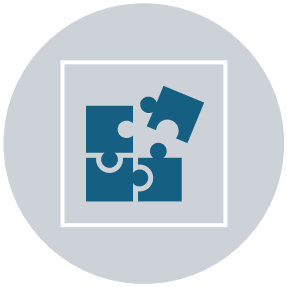


Suggests ambiguous boundaries within major/minor groups

→ **Requires high-resolution modeling with larger datasets + advanced sequence models.**

UNIVERSITY OF MIAMI

What This Research Enables



Reveal hidden intron subtype structure

Detect aberrant splicing events with high sensitivity

Uncover novel isoforms uniquely captured by long-read sequencing

Identify subtype-specific splicing vulnerabilities in cancer

Ultimately, my goal is to bridge the gap between large-scale genomic data and biological understanding by using computational tools to reveal hidden patterns, vulnerabilities, and novel isoforms within the non-coding genome.