

# 1. The Challenge: The "Mixed Bag"

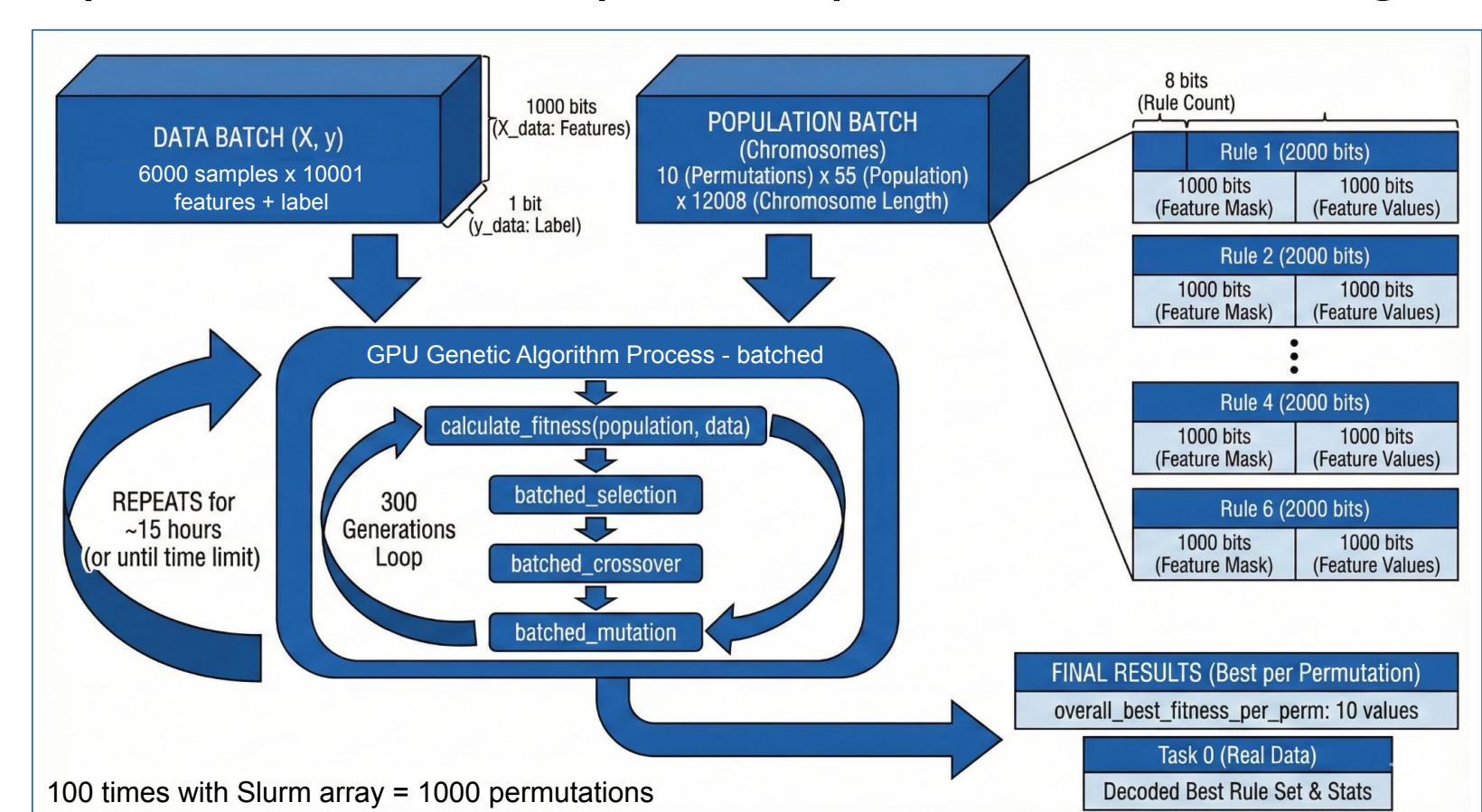
Standard clustering groups cells by Global Transcriptomic Identity, forcing cells into mutually exclusive groups. By ignoring the disease label of individual cells, this approach obscures specific functional failure modes that cut across neuronal subtypes.

**Objective:** To use GPU-accelerated Subgroup Discovery to find **robust combinatorial rules** that define these hidden states without prior biological assumptions.

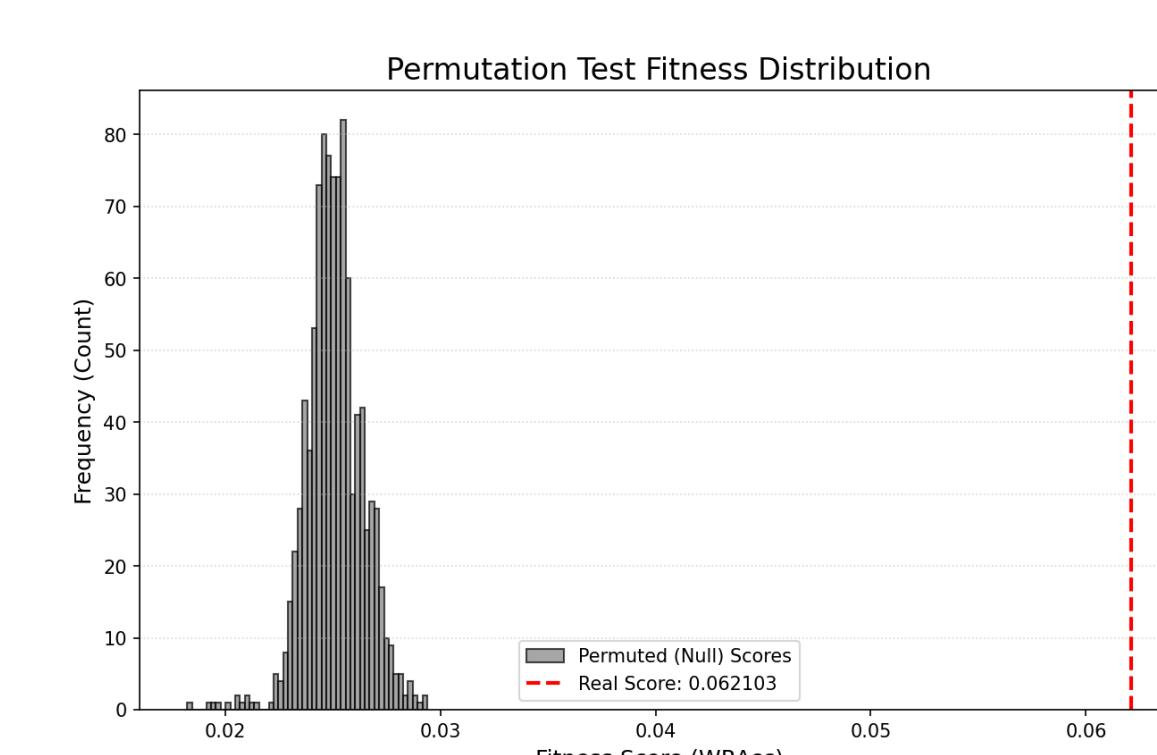
# 2. The Solution: GPU-Accelerated Discovery

To ensure robust biological signals, single-cell expression data was **Z-scored**, **donor-regressed**, and **discretized**. To tackle complexity, we developed a custom **Parallel Genetic Algorithm** to perform Diverse Subgroup Set Discovery (DSSD) on GPU architecture:

- **Glass-Box AI:** Evolves human-readable logical rules (e.g., *IF Gene A AND NOT Gene B*).
- **Set Optimization:** Optimizes a **portfolio** of rules simultaneously, penalizing overlap.
- **High-Throughput:** Leveraged ~10,000 GPU hours on the **DAWN AI Supercomputer** to perform massive parallel permutation testing, ensuring rigorous statistical validation.



**Figure 1: The Discovery Engine.** Parallel evolution of rule sets.



**Figure 2: Validation.** Signal vs Noise ( $p < 0.001$ ).

# 3. Discovery: Three Distinct Failure Modes

The DSSD framework identified three robust, distinct "**Functional States**" within the heterogeneous PD population.

