SUBGROUP DISCOVERY SHEDS LIGHT ON REGULATORY CROSSTALK IN PARKINSON'S DISEASE DOPAMINERGIC NEURONS

Dr Samuel Neaves
Dr Viola Volpato
Professor Caleb Webber

WEBBER LAB
93

UK Dementia Research Institute

neavess@cardiff.ac.uk

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.

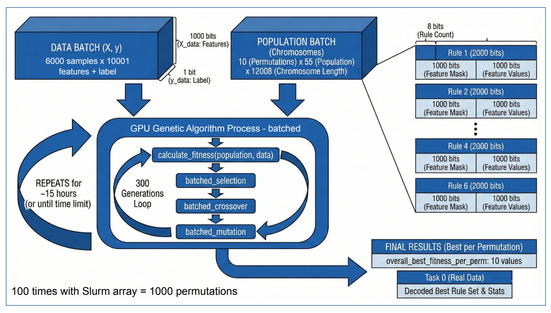

Figure 1: The Discovery Engine. Parallel evolution of rule sets.

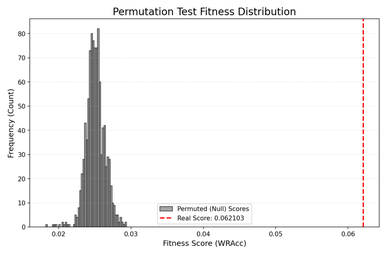

Figure 2: Validation. Signal vs Noise ($p < 0.001$).

3. Discovery: Three Distinct Failure Modes

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

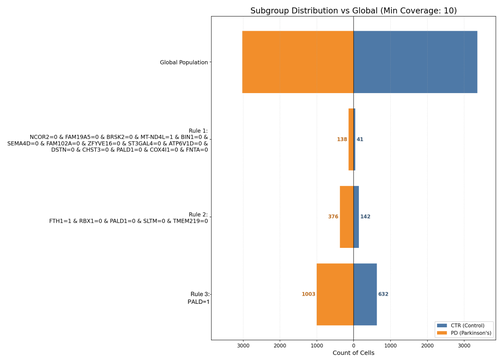

Figure 3: Discovered Rules.

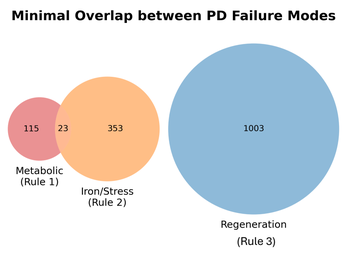

Figure 4: Distinct States. Minimal overlap.

4. Biological Characterization

Functional enrichment analysis (Enrichr) reveals that each subgroup is charaterized by distinct pathological mechanism, supporting the "Orthogonal Failure Mode" hypothesis.

- **Rule 1 (Metabolic):** Characterized by a **bioenergetic collapse**, indicated by extreme enrichment for *Mitochondrial Myopathy* and *Leigh Disease* signatures ($P < 10^{-18}$).
- **Rule 2 (Iron/Stress):** Defined by the dysregulation of **Iron Homeostasis** and *Secretory Granules*, pointing to specific lysosomal and vesicular toxicity.
- **Rule 3 (Regeneration):** Displays a **"Frustrated Regeneration"** phenotype. The upregulation of acute injury factors (*EGR1*) and **Axon Guidance** machinery suggests an active attempt at structural repair.

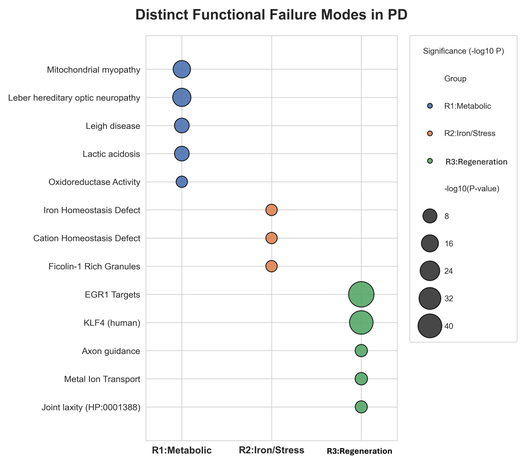

Figure 5: Deconvolution of Pathology. Analysis against GO, Jensen Diseases, and TF databases reveals three non-overlapping mechanisms.

5. Impact: Functional States vs. Identity

Standard clustering groups cells by global transcriptomic identity. Our framework reveals states that are **orthogonal** to this identity.

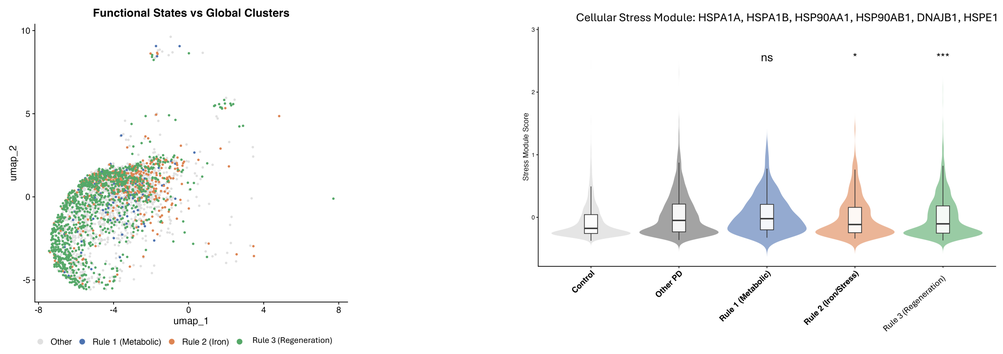

Figure 6: States are Orthogonal to Identity. Left: Functional failure modes are scattered across diverse transcriptomic clusters. Right: These states represent a distinct, higher tier of cellular stress compared to the general PD background.

6. Molecular & Regulatory Architecture

We identified the specific signaling pathways and protein machinery assoicated with these states.

The "Regeneration" Machinery (Rule 3)

STRING interaction analysis reveals that Rule 3 markers form a **coordinated structural remodeling network**, indicating active reorganization of the cell architecture.

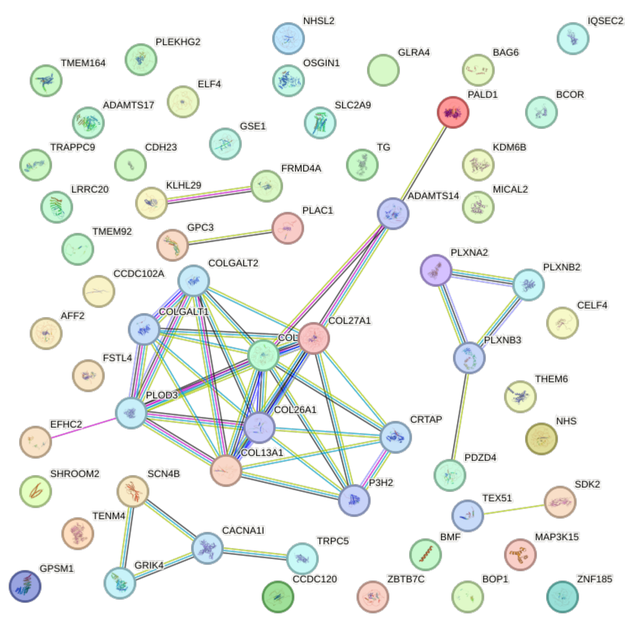

Figure 7: Coordinated Repair Program. The cytoskeletal regulator *PALD1* is functionally coupled to the matrix remodeler *ADAMTS14*, linking internal actin dynamics with external scaffolding.

Distinct Regulatory Architectures

We explored specific signaling pathways of interest by calculating single-cell module scores. This analysis characterizes the distinct regulatory profile associated with each failure mode.

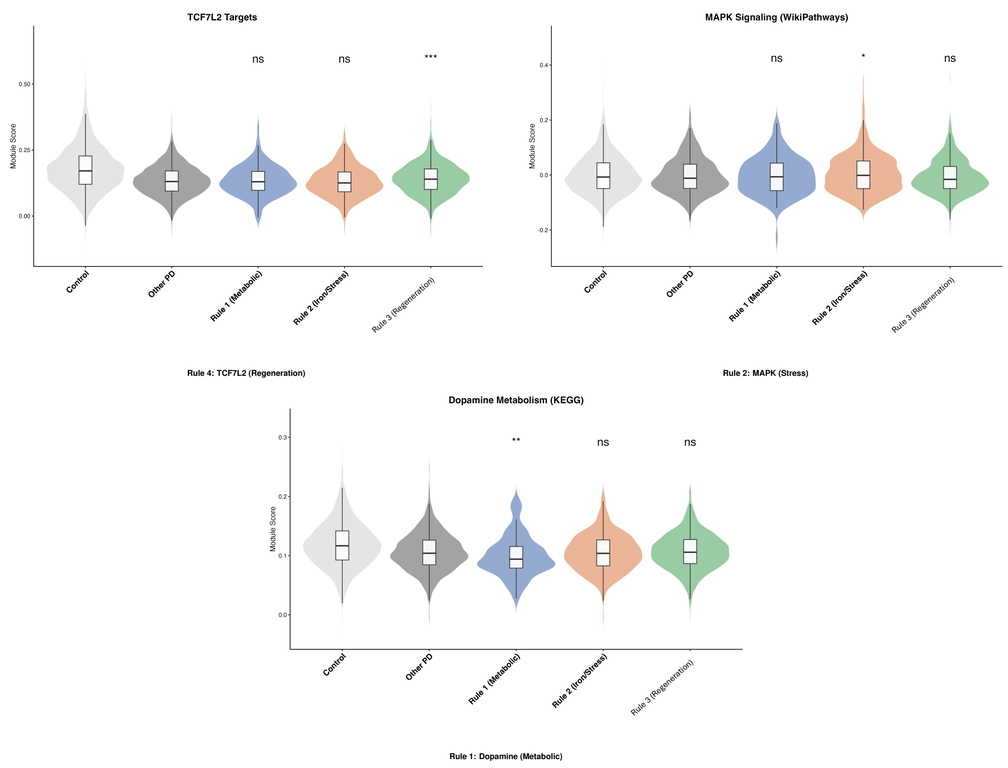

Figure 8: Distinct Signaling Mechanisms. *TCF7L2* (Wnt) is upregulated in Rule 4. ($P < 0.001$). *MAPK* drives Rule 2 ($P < 0.05$). *Dopamine Metabolism* is downregulated in Rule 1 ($P < 0.01$).

7. Pseudotime Dynamics

We mapped the subgroups onto a pseudotime disease trajectory.

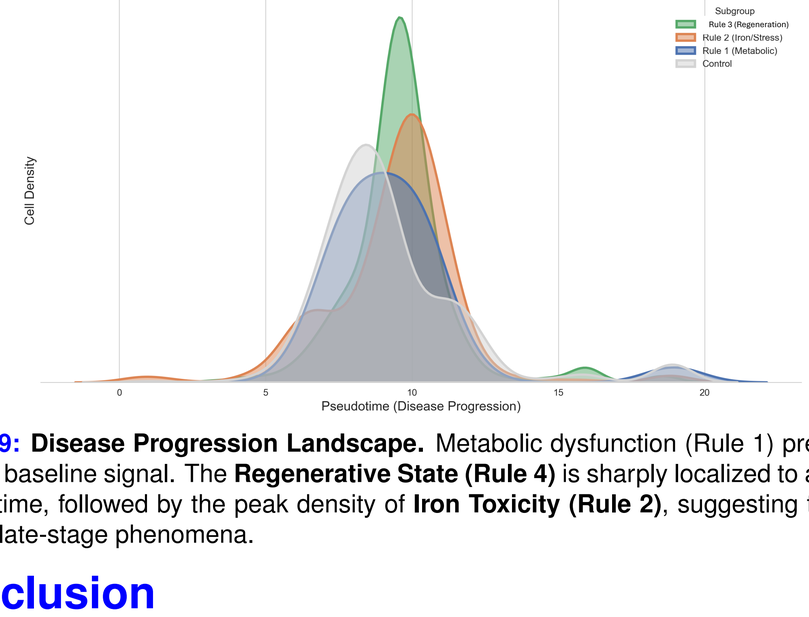

Figure 9: Disease Progression Landscape. Metabolic dysfunction (Rule 1) presents as a broad baseline signal. The **Regenerative State (Rule 4)** is sharply localized to advanced pseudotime, followed by the peak density of **Iron Toxicity (Rule 2)**, suggesting these are distinct late-stage phenomena.

Conclusion

We have mapped the regulatory crosstalk of PD into three potentially actionable states:

1. **Metabolic (Baseline):** Mitochondrial support.
2. **Iron-Toxic (Acute):** Marked by MAPK and functional demand.
3. **Regenerative (Structural):** A coordinated but frustrated attempt at repair linked to TCF7L2.

This work used compute time awarded through the UKRI AIRR Early Access Programme on the Dawn supercomputer (project ANON-BYYG-VX4J-7), supported by UKRI.

Principal funder:
Cardiff University / Prifysgol Caerdydd
UKRI Medical Research Council