1. Simplified Explanations of Method Choices

Why Python/Jupyter?

- Python : Chosen due to its extensive libraries (e.g., Scanpy) for single-cell analysis and ease of use for bioinformatics workflows + my experiences and previous works all were python based
- Jupyter Notebook : Enables step-by-step execution, visualization, and sharing of code, while cloud platforms handle heavy computations better than my own hardware

Steps

1. Data Download & Extraction
    - Tools : **wget** and **tar** for direct GEO dataset access.
    - Why : Avoids manual downloads and ensures reproducibility.
2. Quality Control
    - Cell Filtering : Removed cells with **<200 genes** or **>2500 genes** (common thresholds to exclude low-quality/doublet cells).
    - Mitochondrial Gene Cutoff (**<5%**) : High mitochondrial RNA often indicates dying cells.
3. Cell Cycle Correction
    - Gene Lists : S-phase and G2/M-phase genes sourced from prior literature to score cell cycle phases.
    - Why : Corrects for cell cycle-driven expression noise, ensuring clusters reflect true biological states.
4. Clustering & Dimensionality Reduction
    - Leiden Algorithm : Preferred for its ability to detect fine-grained communities compared to older methods (e.g., Louvain).
    - PCA/UMAP : Reduces data complexity while preserving biological structure.
5. Re-clustering Adrenal Medulla Populations

- Subsetting : Focused on SCPs, Chromaffin, and Sympathoblasts using literature-backed marker genes :
    - SCPs : **SOX10**, **PLP1** (neural crest markers).
    - Chromaffin : **TH**, **PNMT** (catecholamine synthesis).
    - Sympathoblasts : **PHOX2B**, **ASCL1** (sympathetic neuron development).
- Why Re-cluster ?: Initial clusters were too broad, subsetting improves resolution of rare/transitional states.

6. Trajectory Analysis (Task 6a)
    - Tool Choice : Diffusion Pseudotime (DPT) in Scanpy.
        - Why : Efficiently orders cells along developmental paths (e.g., SCP → Chromaffin) and integrates seamlessly with Scanpy.
    - Gene Selection : Union of markers for each trajectory pair (e.g., **TH** for Chromaffin, **PHOX2B** for Sympathoblasts).
    - Binning (20 bins) : Smooths gene expression trends for clear visualization.

Why Task 6a Over 6b?

- Task 6a was prioritized to focus on gene expression dynamics during developmental transitions (e.g., SCP differentiation). Heatmaps directly show how genes like **TH** or **PHOX2B** change, aligning with the goal of understanding neuroblastoma origins.
- SCENIC (6b) was deferred due to its computational complexity and focus on transcription factor networks, which is better suited for advanced follow-up studies.

Key Thresholds & Numbers Explained

- Leiden Resolution : Default (**1.0**) balances cluster granularity.
- Highly Variable Genes : **min_mean=0.0125**, **max_mean=3**, **min_disp=0.5** retain biologically meaningful genes.
- Pseudotime Root Cell : Set to the first SCP cell to anchor differentiation direction.

2. Interpretation of Task 6 Results

The trajectory analysis offers valuable insight into the differentiation pathways from Schwann Cell Precursors (SCPs) into either Chromaffin cells or Sympathoblasts. By examining the expression dynamics of key genes, we can trace distinct transcriptional programs that govern these lineage decisions.

One clear transition observed is from **Chromaffin cells to Sympathoblasts**. This shift is marked by a sharp decline in *PNMT* and *CHGA*, both essential for catecholamine synthesis and Chromaffin identity. In contrast, genes associated with neuronal fate—particularly *ELAVL4* and *PRPH*—show increasing expression. The heatmap reveals that *ELAVL4* rises steadily across pseudotime, highlighting its role in guiding cells toward a sympathetic neuronal fate.

A second transition, from **SCPs to Sympathoblasts**, also centers around *ELAVL4* as a key marker of neuronal commitment. This process is accompanied by a pronounced decline in progenitor markers such as *ISL1*, *PLP1*, and *SOX10*. Notably, the drop in *SOX10* and *ISL1* underscores the cells' exit from the neural crest progenitor state, while the peak of *ELAVL4* in later pseudotime bins signals full commitment to the Sympathoblast lineage.

In parallel, the differentiation path from **SCPs to Chromaffin cells** is characterized by the upregulation of *PNMT* and *CHGA*, indicating a shift toward an endocrine identity and functional catecholamine production. Concurrently, the steady decline of *SOX10* again reflects loss of progenitor identity, suggesting a common early transcriptional exit point before diverging into specific lineages.

These patterns reinforce the idea that **SCPs function as bipotent progenitors**, capable of giving rise to both neuronal (Sympathoblast) and endocrine (Chromaffin) cells through distinct transcriptional trajectories. This flexibility also sheds light on disease mechanisms—particularly **neuroblastoma**, which may arise from arrested or dysregulated transitions. For instance, failure to downregulate *SOX10* or insufficient activation of *ELAVL4* could trap cells in an undifferentiated state, a hallmark of neuroblastoma's origin.

From a developmental standpoint, the upregulation of *PNMT* in the Chromaffin trajectory underlines its role in establishing catecholamine-producing cells, while *ELAVL4* activation appears pivotal for sympathetic neuron maturation.

This analysis describes how dynamic changes in a few key regulators—*SOX10*, *ELAVL4*, and *PNMT*—guide lineage specification. These findings enhance our understanding of adrenal gland development and provide a molecular framework that could inform future research into neuroblastoma and related pathologies.

3. Abstract (Overall results)

Single-cell RNA sequencing of developing human adrenal glands revealed three key cell states: Schwann Cell Precursors (SCPs), Chromaffin cells, and Sympathoblasts. SCPs, marked by SOX10 and PLP1 , differentiated into catecholamine-producing Chromaffin cells (TH , PNMT) or immature sympathetic neurons (Sympathoblasts, PHOX2B , ASCL1). Trajectory analysis showed dynamic gene shifts: ISL1 (SCP marker) declined as TH increased during SCP→Chromaffin transitions, while PHOX2B rose in SCP→Sympathoblast paths. A large "Unknown" cluster suggested unannotated cell types. These findings highlight developmental programs linking neural crest-derived progenitors to adrenal medulla maturation, providing insights into neuroblastoma’s origins in dysregulated precursor cells.