Replication of Figure 1B: Correlation of Fetal Cortex Single Cells to Bulk reference Cortical Zones and Purified Cell Types

The aim is to:

1. Compute **Spearman correlation** between individual fetal cortex single cells (12–13 wpc) and:
    a. Four cortical zones (VZ, iSVZ, oSVZ, CP)
    b. Three purified cell types (aRG, bRG, neuron)
2. Perform **per-cell Z-score normalization**
3. Cluster cells using **Pearson correlation distance**
4. Visualize results as two aligned heatmaps
5. Assign each cell to its maximum-correlation reference

🗁 Input Datasets

1 Single-Cell Dataset (sc_raw)

File: GSE75140_hOrg.fetal.master.data.frame.txt

- Rows: cells
- Columns: gene symbols
- Contains both fetal + organoid cells

Filtered this data using R to:

- Fetal only
- 12–13 wpc only
- Removal of:
    - 1 PECAM1+ endothelial cell
    - 5 interneurons (GAD1,ERBB4,DLX1,DLX2,DLX5,DLX6)

Final retained cells: **220**

2 Bulk Cortical Zones

File: GSE38805_human_FPKM.txt

Obtained from the article : pnas.1209647109

- Rows: Ensembl gene IDs
- Columns: cortical zones (extract only 13 wpc)
    - VZ
    - iSVZ
    - oSVZ

- CP

3 Purified Cell Types

File: GSE65000_hsa_fpkm_matrix.txt

Obtained from the article : science.aaa1975

- Rows: Ensembl gene IDs
- Columns: four replicate samples
  - aRG
  - bRG
  - Neuron

Replicates are averaged (mean) per cell type.

Data Processing Workflow

Step 1 — Filter Single Cells

Usinhg sc_raw converted to sc_fetal (filtered) then transposed it to sc_matrix

- Retain fetal cells only
- Retain 12–13 wpc only
- Remove:
  - Highest PECAM1-expressing cell (endothelial contaminant)
  - Top 5 cells ranked by cumulative interneuron marker score

interneuron Markers used:

GAD1, ERBB4, DLX1, DLX2, DLX5, DLX6

Step 2 — Harmonize Gene Identifiers

- Reference datasets use Ensembl IDs
- Single-cell dataset uses gene symbols
- Ensembl IDs mapped to gene symbols using provided mapping
- Duplicated symbols averaged

- Genes intersected across all datasets

Final gene set used for correlation: **9699 genes**

Step 3 — Correlation Analysis

For each single cell:

Compute Spearman correlation with:

- 4 cortical zones
- 3 purified cell types

Produces:

- zone_cor (220 × 4)
- type_cor (220 × 3)

Step 4 — Per-Cell Z-score Scaling

Each cell's correlations are scaled across references,

This enables relative comparison within each cell.

Produces:

- zone_z
- type_z

Step 5 — Hierarchical Clustering

Cells clustered using:

- Pearson correlation distance
- Complete linkage

Clustering based on zone_z.

Row order applied to both heatmaps.

Step 6 — Visualization

Two aligned heatmaps generated:

Left panel:

- Z-scored correlations to cortical zones

Right panel:

- Z-scored correlations to purified cell types

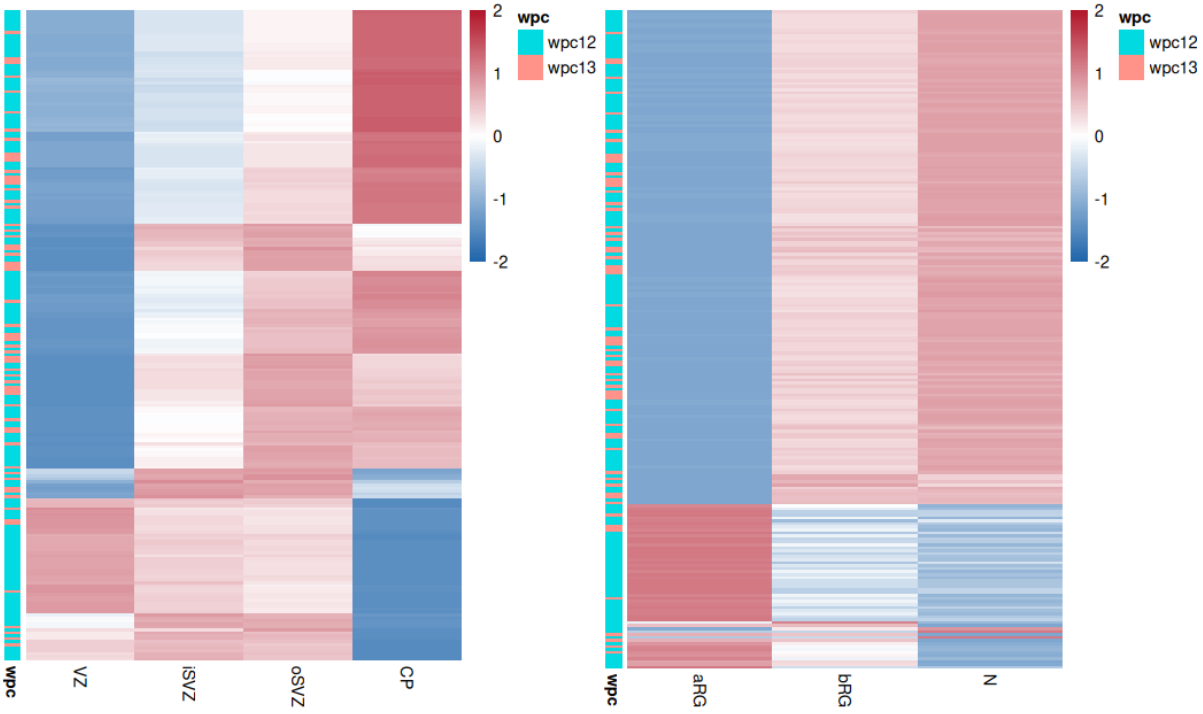
Color scale:

- Blue → white → red
- Z-score range: −2 to +2

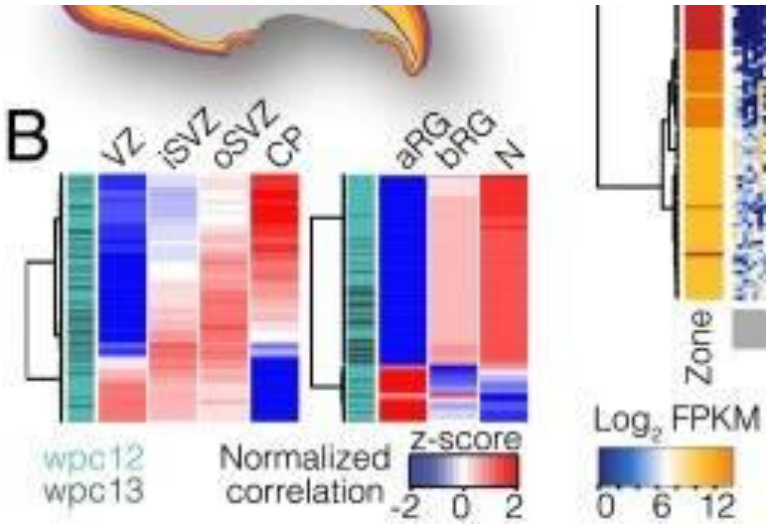
Key Objects in Workspace

Object	Description
sc_raw	Original single-cell dataset
sc_fetal	Filtered fetal 12–13 wpc cells
sc_matrix	Genes × cells matrix for correlation
cells_mean	Averaged purified cell-type reference
zones_symbol	Zone reference (symbol-mapped)
cells_symbol	Purified reference (symbol-mapped)
common_genes	Genes shared across all datasets
zone_cor	Raw Spearman correlations (zones)
type_cor	Raw Spearman correlations (cell types)
zone_z	Z-scored correlations (zones)
type_z	Z-scored correlations (cell types)
hc	Hierarchical clustering object
annotation_row	Developmental stage annotation
hm_zone, hm_type	Final heatmap plot objects

created 1B :



Original 1B


B
VZ
iSVZ
oSVZ
CP
aRG
bRG
N
wpc12
wpc13
Normalized correlation
z-score
-2 0 2
Zone
Log2 FPKM
0 6 12