BodyMap: Reference Gene Expression Data Collection

March 24, 2026

sean.yao@insilicio-biosystems.com

Contents

I. Normal adult tissue BodyMaps of human and nonclinical species

II. Developmental tissue BodyMaps of human and nonclinical species

III. Single-Cell BodyMaps (10X Genomics and SMART-Seq2) of Human and Mouse

IV. Bulk-Cell BodyMaps: CCLE Cell Line, Cancer Cell, and Immune Cell BodyMaps

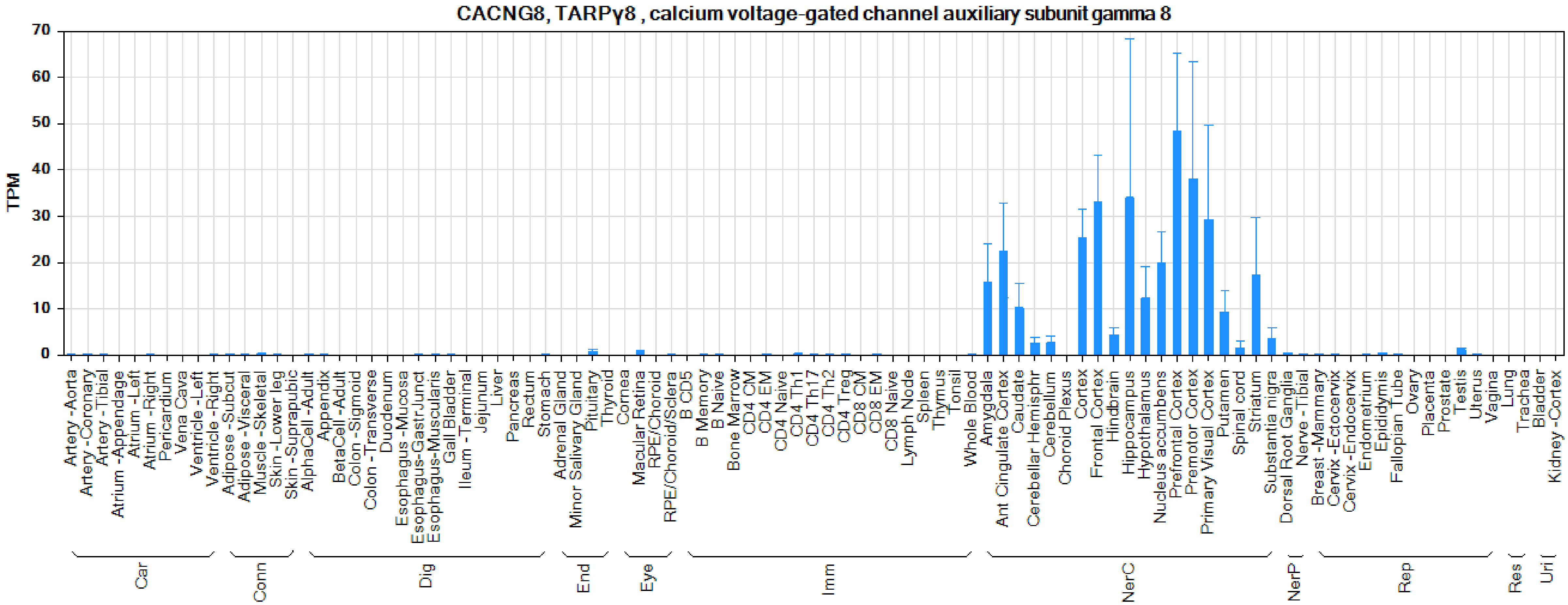
V. BodyMap data collection and R-Shily app for data visualization

I. Normal Adult tissue BodyMaps of human and nonclinical species

- Species included: Human, Mouse, Rat, Dog, and Cyno
- Data collection, processing, QC, curation and integration:
  - Survey published large tissue atlas studies using compatible RNA-Seq protocols. Select 2-5 studies having high quality and coverage for each species
  - Re-analyze raw fastq data by applying the same data processing and QC protocols → Expression levels (TPM) are directly comparable across genes or species
  - Apply QC criteria more stringent that original publication → Better data quality
  - Curation: Use common ontology for tissue and cell type name across studies
  - Integration: Integrate all data of the same species into a single dataset

Normal adult tissue BodyMap: Human

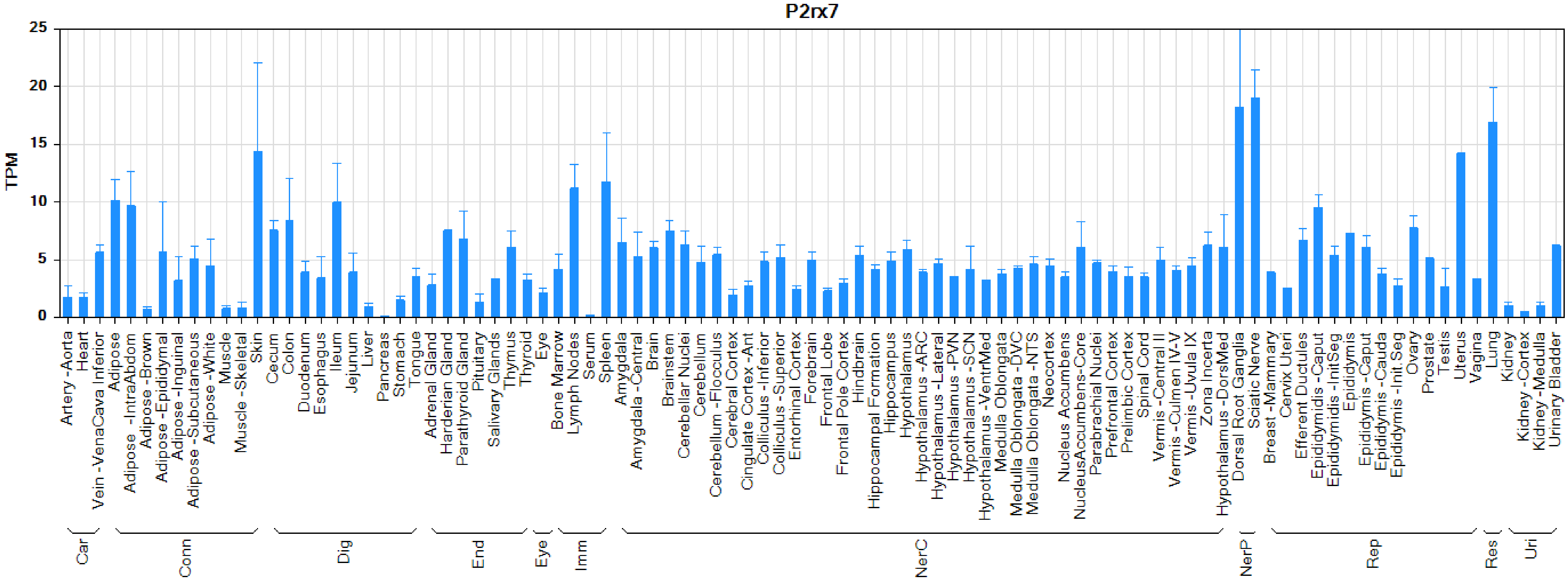
Example: TARPγ8 specifical expression in brain regions



- Reprocessed all GTEx RNA-Seq data and kept 12490 of 16700 samples / 51 tissues types
- Added matching RNA-Seq data of 49 more tissue types from other sources

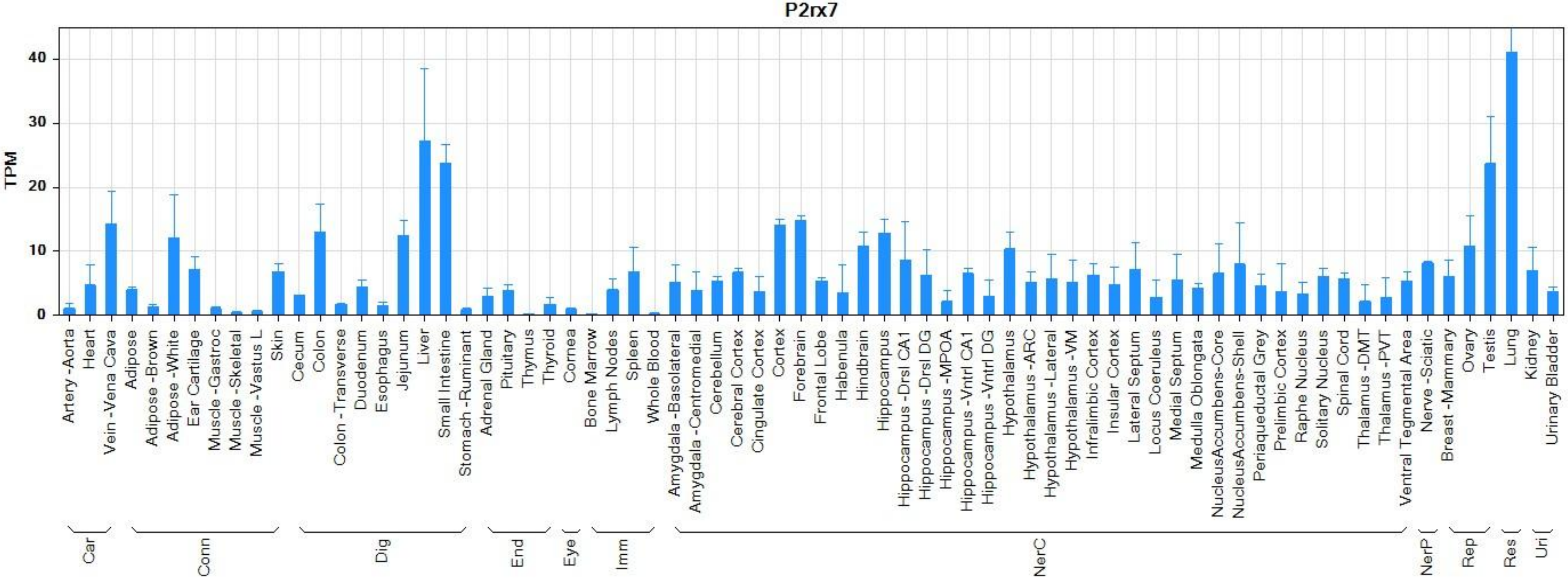
➢ **A total of 95 unique tissue types passed QC**

Normal adult tissue BodyMap: Mouse



- Reprocessed all AbbVie RNA-Seq data of 47 tissue types;
- Processed matching RNA-Seq data of 69 tissue types from other sources
- ➢ **A total of 95 unique tissue types passed QC**

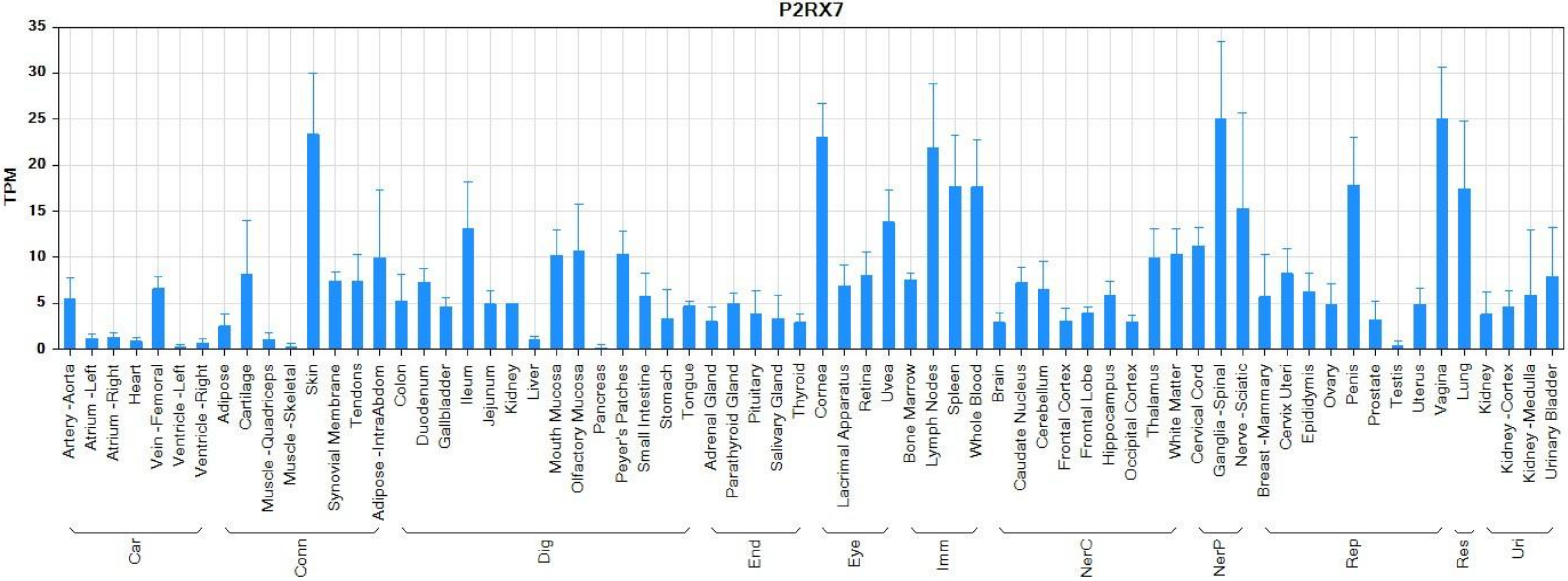
Normal adult tissue BodyMap: Rat



- Reprocessed all AbbVie RNA-Seq data of 46 tissue types
- Processed matching RNA-Seq data of 60 tissue types from other sources

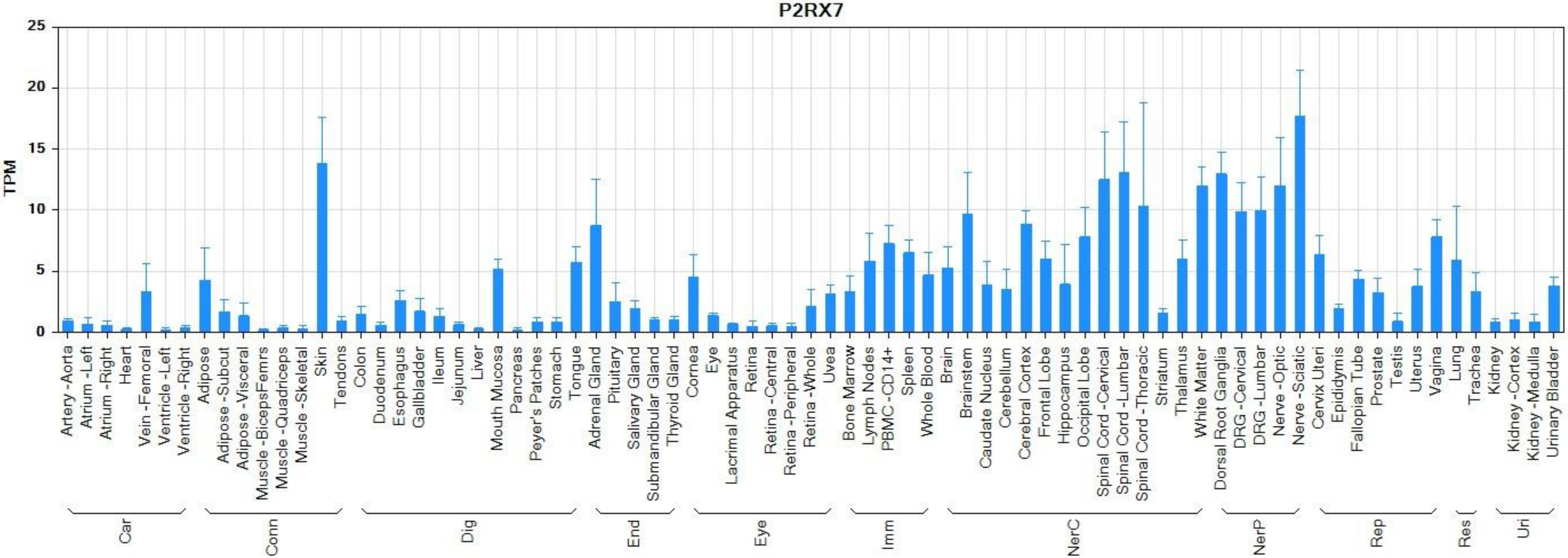
➢ **A total of 73 unique tissue types passed QC**

Normal adult tissue BodyMap: Dog



- Reprocessed all AbbVie RNA-Seq data of 53 tissue types
- Processed matching RNA-Seq data of 54 tissue types from other sources
- ➢ **A total of 67 unique tissue types passed QC**

Normal adult tissue BodyMap: Cyno

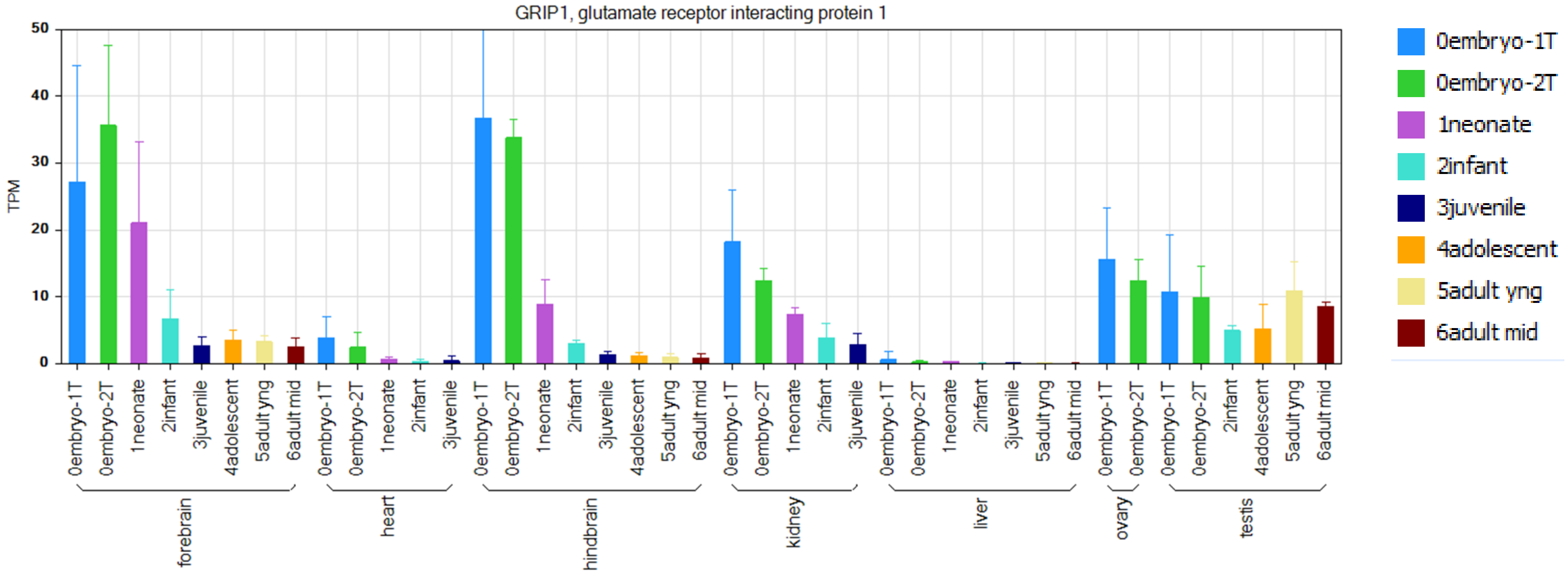


- Reprocessed all AbbVie RNA-Seq data of 42 tissue types
- Processed matching RNA-Seq data of 45 tissue types from other sources
- **A total of 78 unique tissue types passed QC**

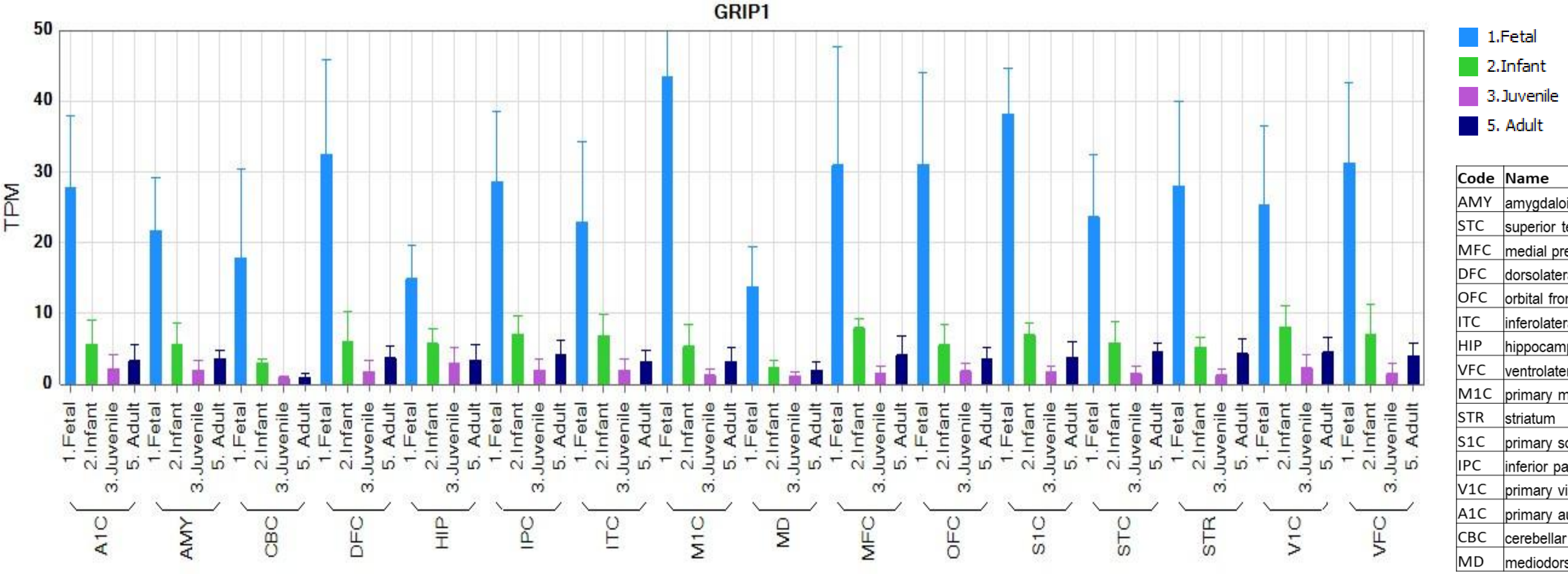
II. Developmental tissue BodyMaps

- Species included: Human, Mouse, Rat, Rabbit, and Rhesus
- Data collection, processing, QC, curation and integration: The same as the normal adult tissue BodyMaps

Developmental tissue BodyMap: Human

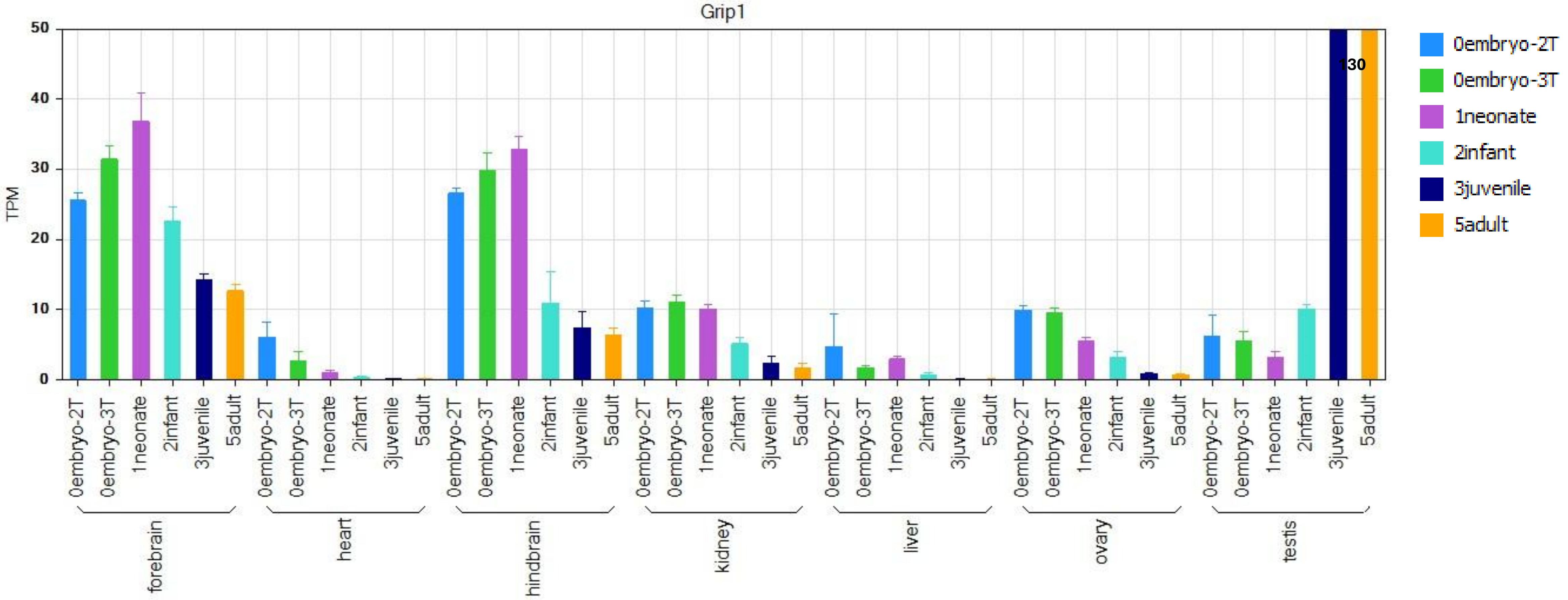


Developmental tissue BodyMap: Human brain regions

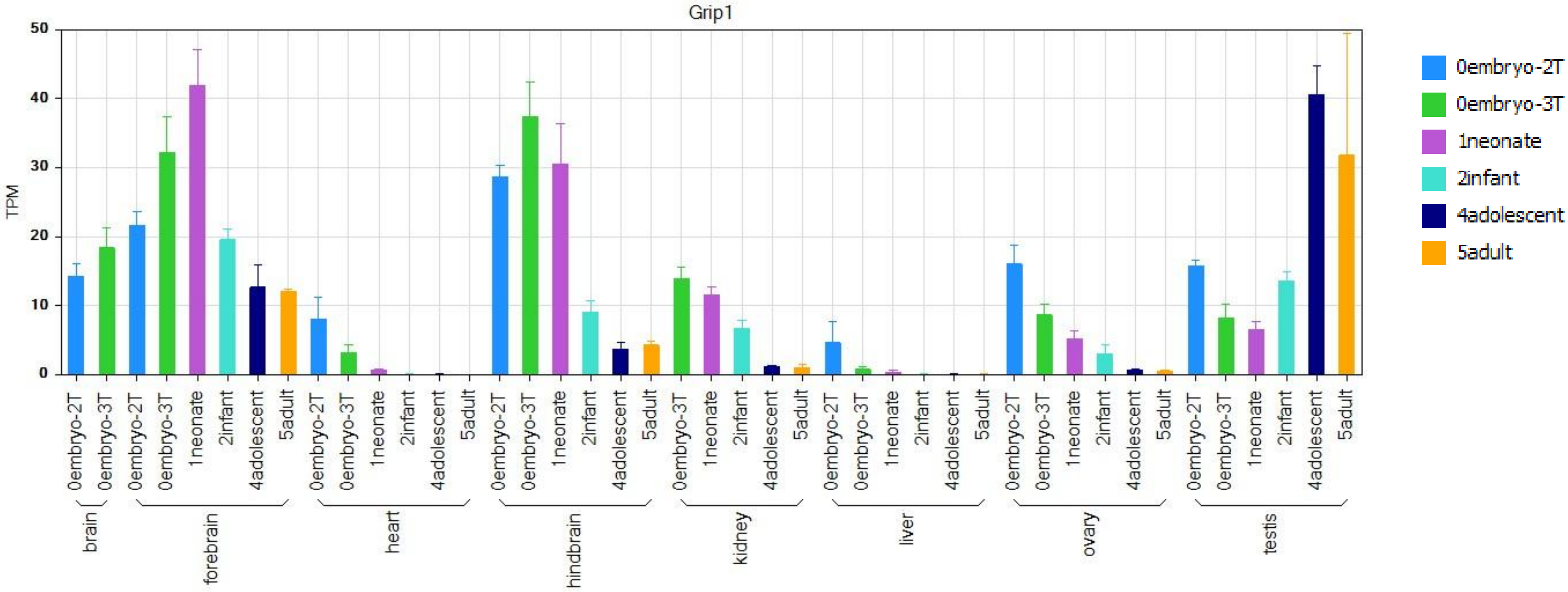


Code	Name
AMY	amygdaloid complex
STC	superior temporal cortex
MFC	medial prefrontal cortex
DFC	dorsolateral prefrontal cortex
OFC	orbital frontal cortex
ITC	inferolateral temporal cortex
HIP	hippocampus
VFC	ventrolateral prefrontal cortex
M1C	primary motor cortex (area M1)
STR	striatum
S1C	primary somatosensory cortex (area S1)
IPC	inferior parietal cortex
V1C	primary visual cortex (area V1/17)
A1C	primary auditory cortex (core)
CBC	cerebellar cortex
MD	mediodorsal nucleus of thalamus

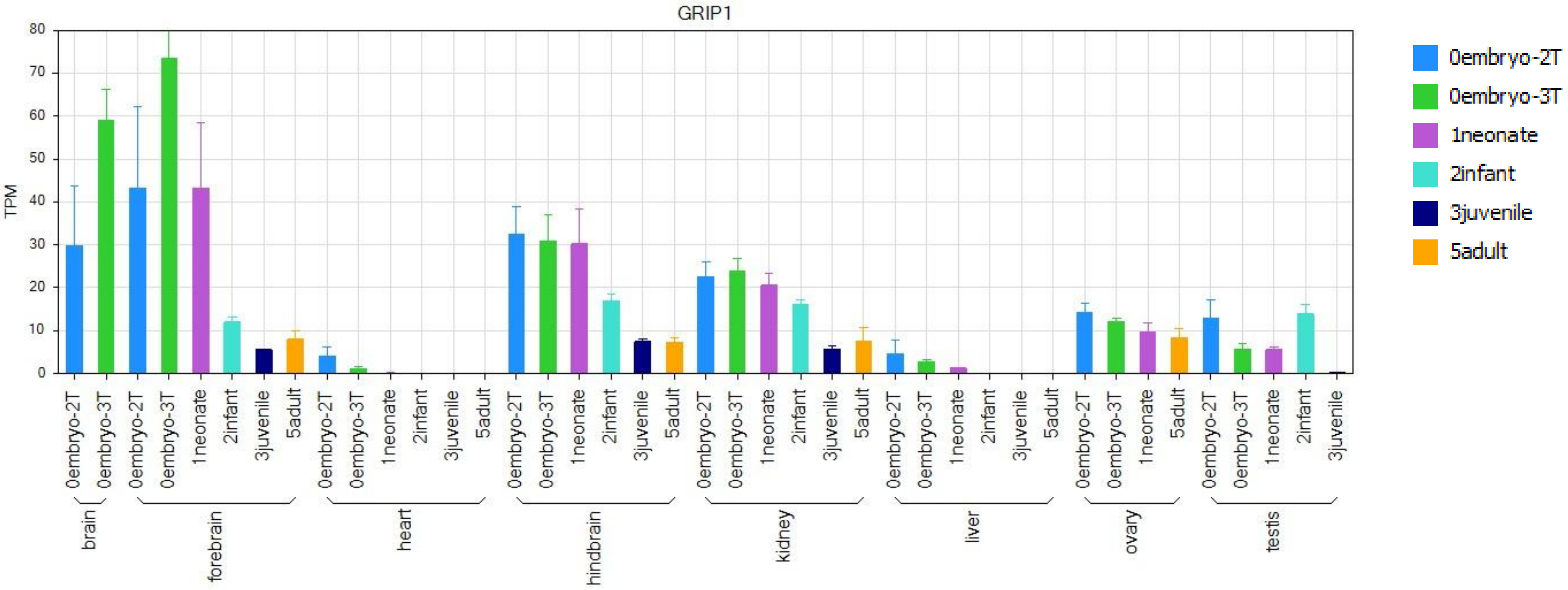
Developmental tissue BodyMap: Mouse



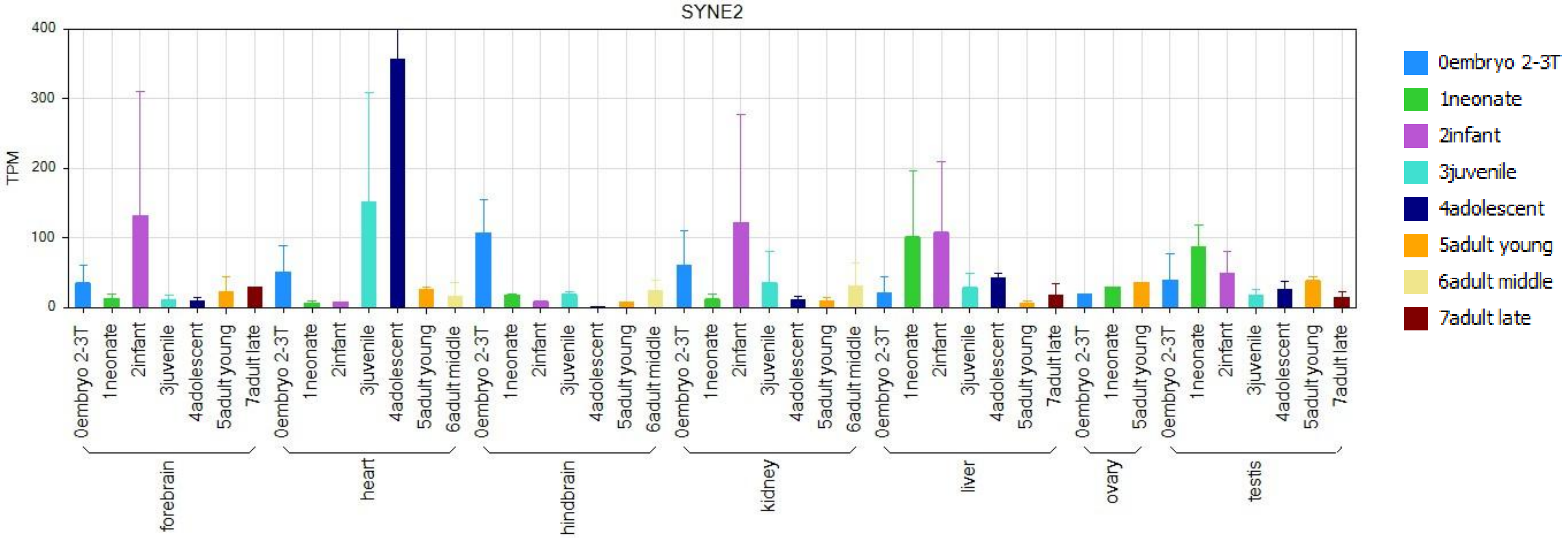
Developmental tissue BodyMap: Rat



Developmental tissue BodyMap: Rabbit



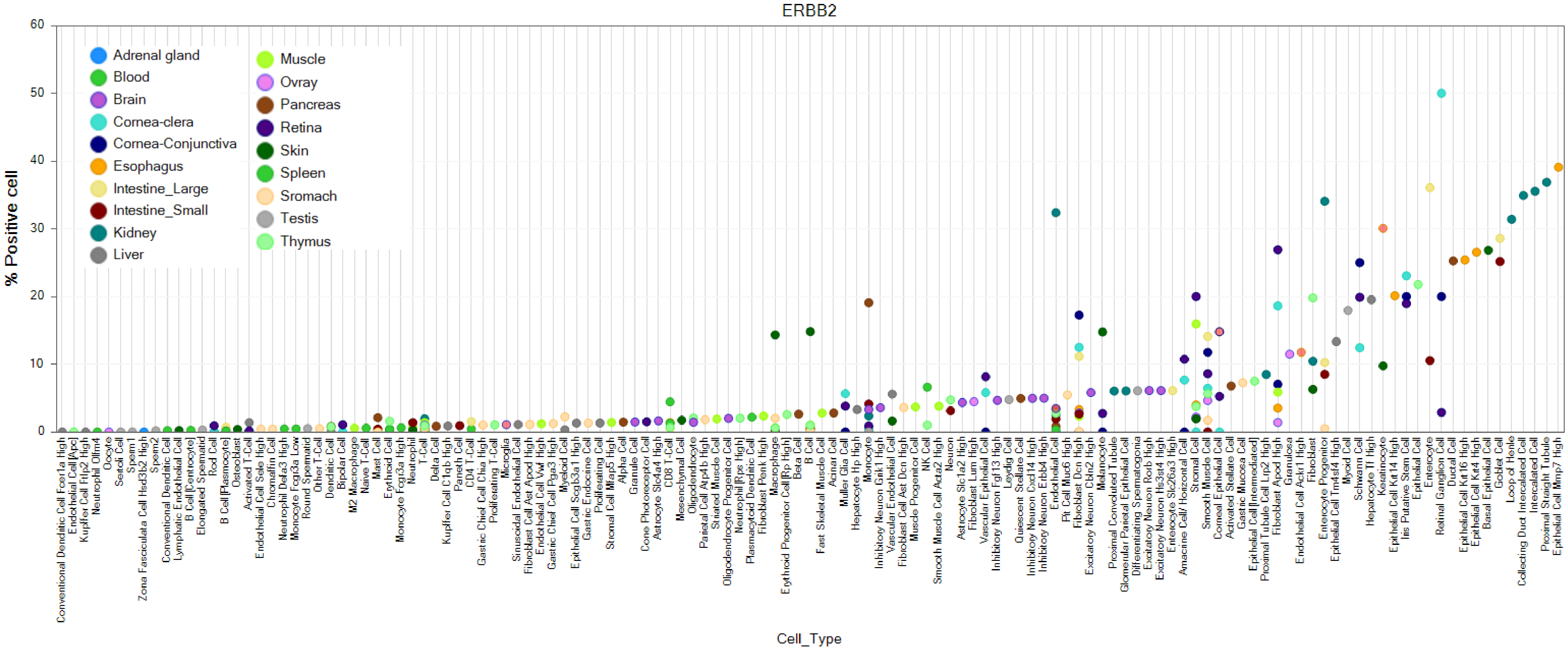
Developmental tissue BodyMap: Rhesus



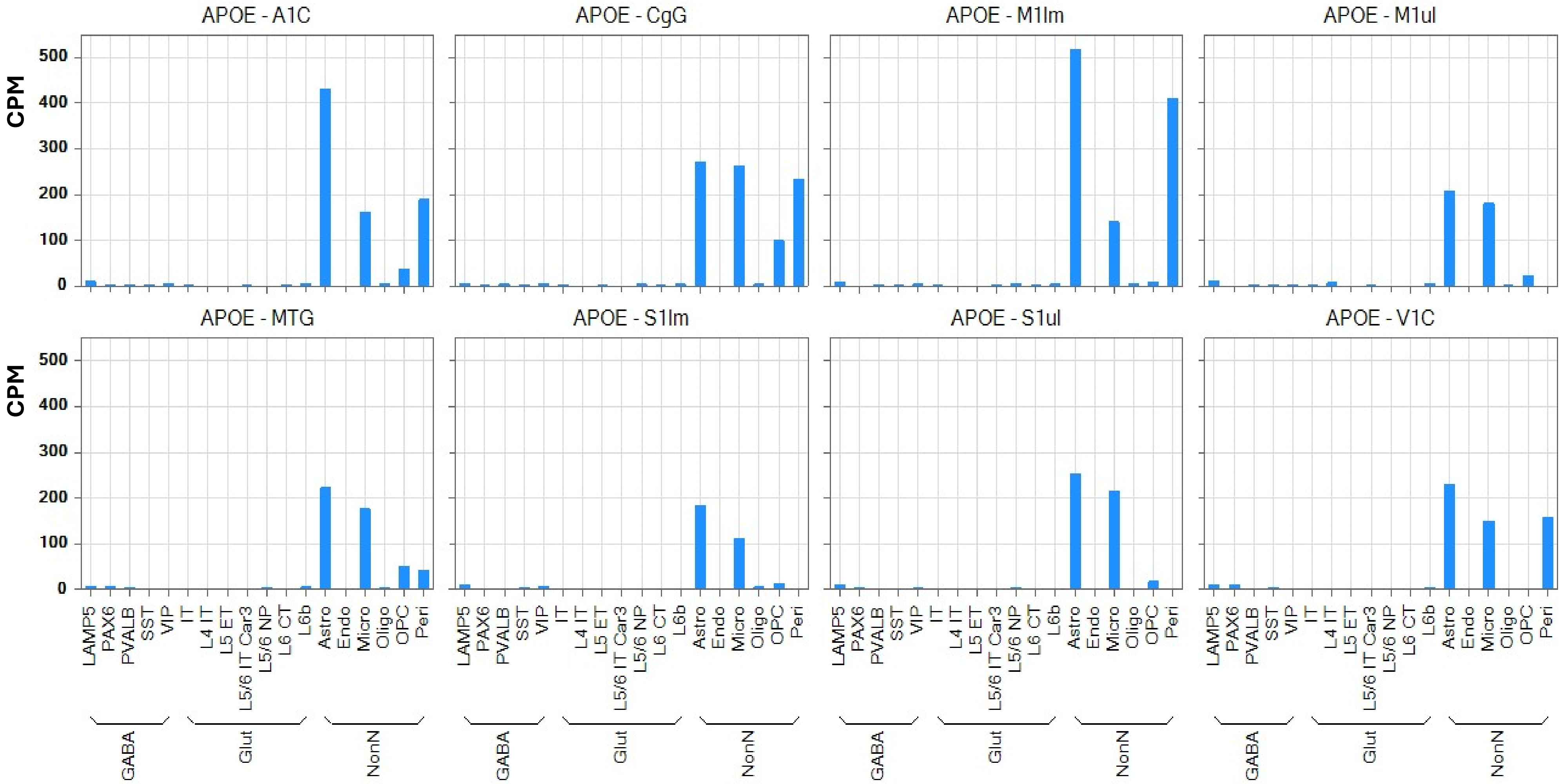
III. Single cell BodyMaps of human and mouse

- Species included: Human and Mouse
- Data collection, processing, QC, curation and integration:
  - Survey published large single cell atlas studies and select the ones with high data quality and sample coverage
  - Re-analyze data by applying additional QC and curation:
    - QC: Remove cells with low sequence or gene coverage, weak association with assigned cell type → better data quality than original paper
    - Curation: Use common ontology for tissue and cell type name across studies
    - Calculate a gene's percentage of detection among cells in each cell type → more accurate assessment of genes at lower expression level

Expression in SingleCell BodyMap - Human (10X Genomics)

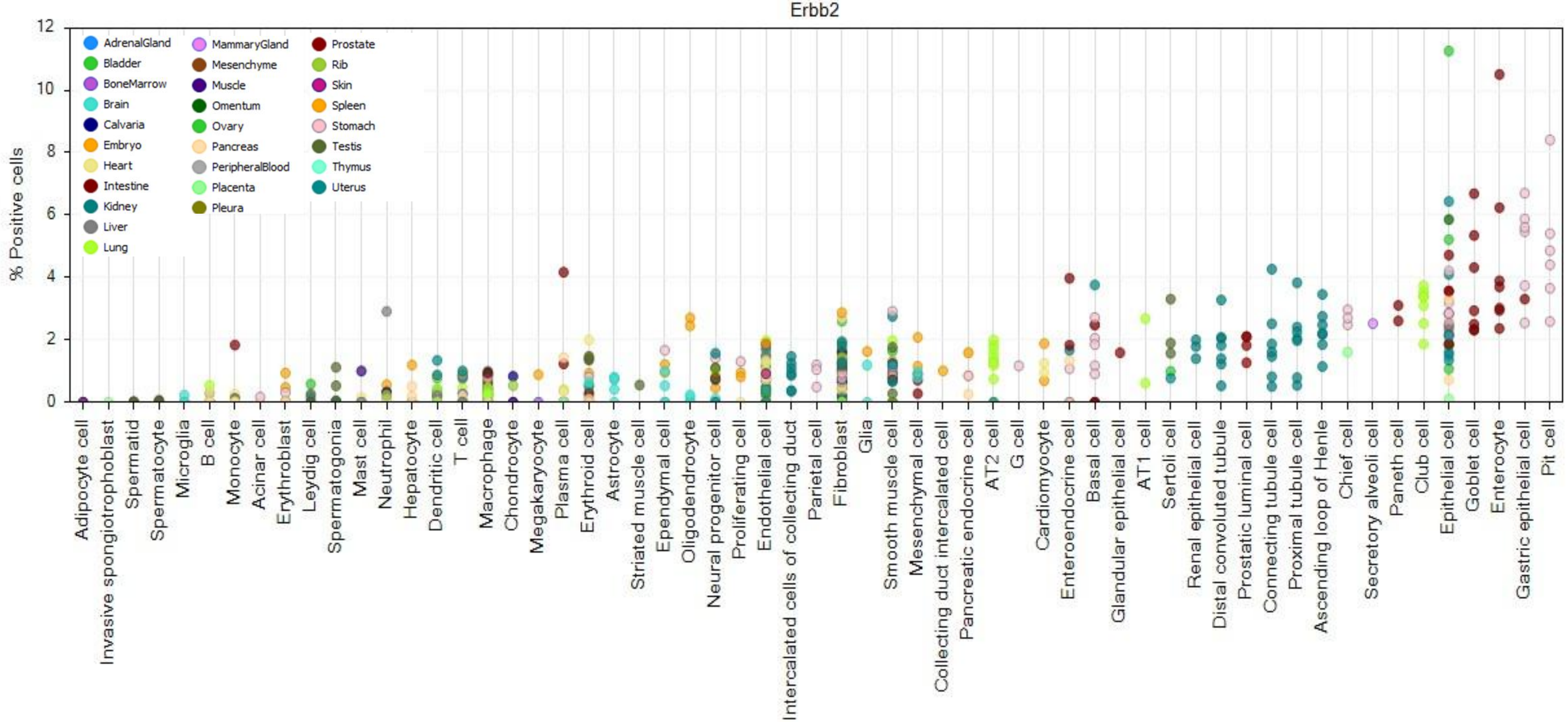


Expression in single cell of normal human regions (SMART-Seq2)



Code	Brain region name
A1C	A1C primary auditory cortex
CgG	CgG anterior cingulate gyrus
M1lm	M1 lower limb, primary motor
M1ul	M1 upper limb, primary motor
MTG	MTG middle temporal gyrus
S1lm	S1 lower limb, primary somatosensory
S1ul	S1 upper limb, primary somatosensory
V1C	V1C primary visual cortex

Expression in SingleCell BodyMap – Mouse (10X Genomics)

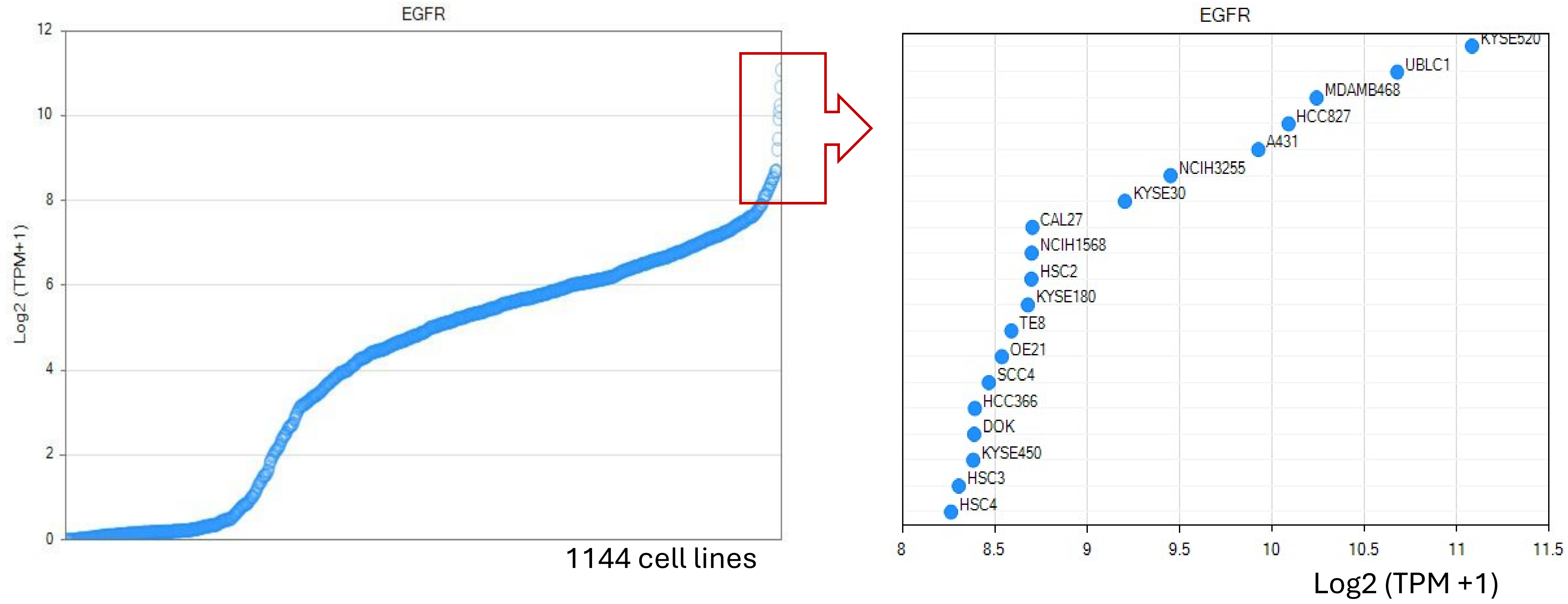


IV. Bulk-Cell BodyMap Collections

CCLEplus cell lines

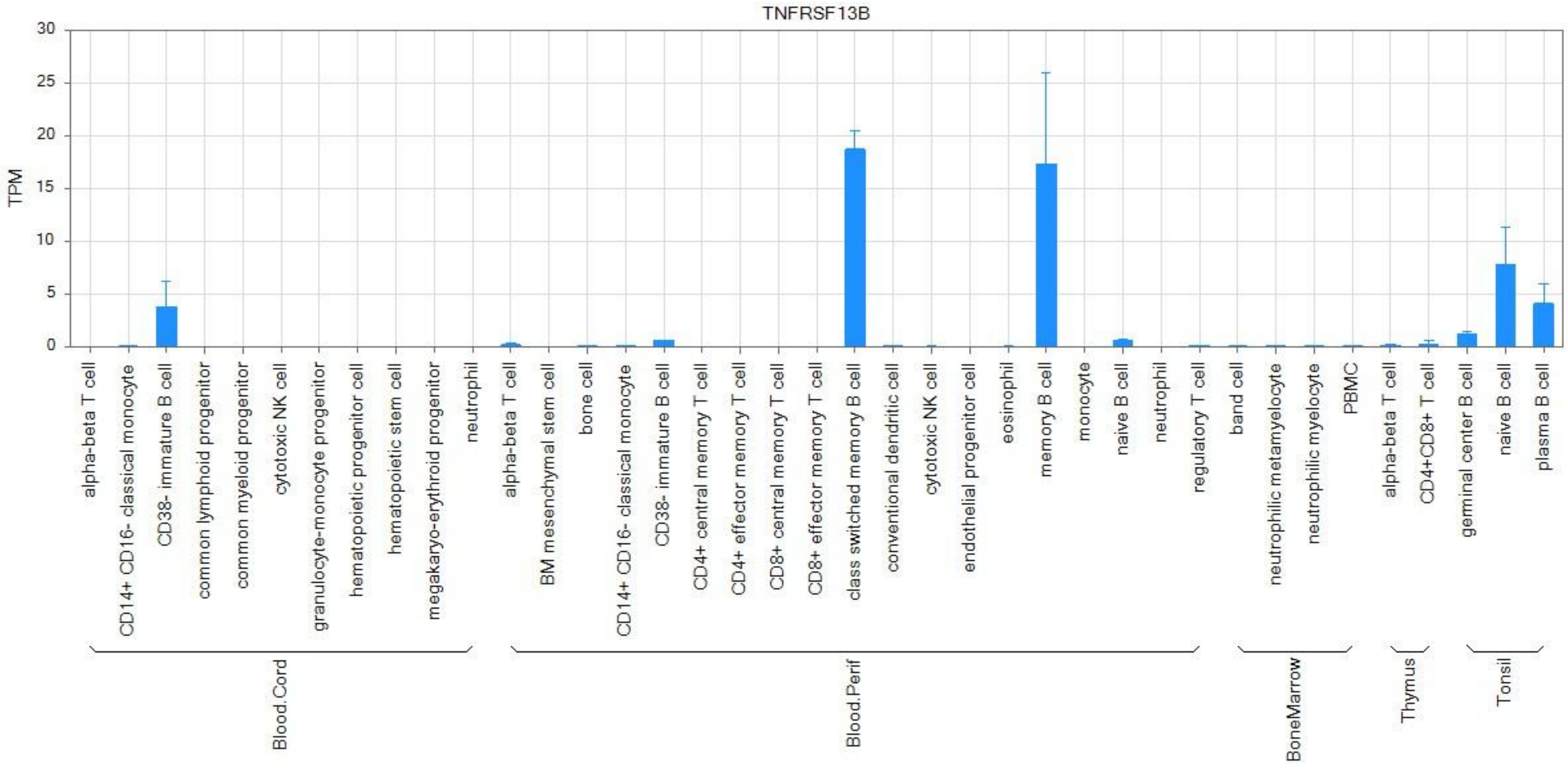
Bulk RNA-Seq of CCLE cell line collections (1144 unique cell lines).

Example: EGFR gene expression across all cell lines



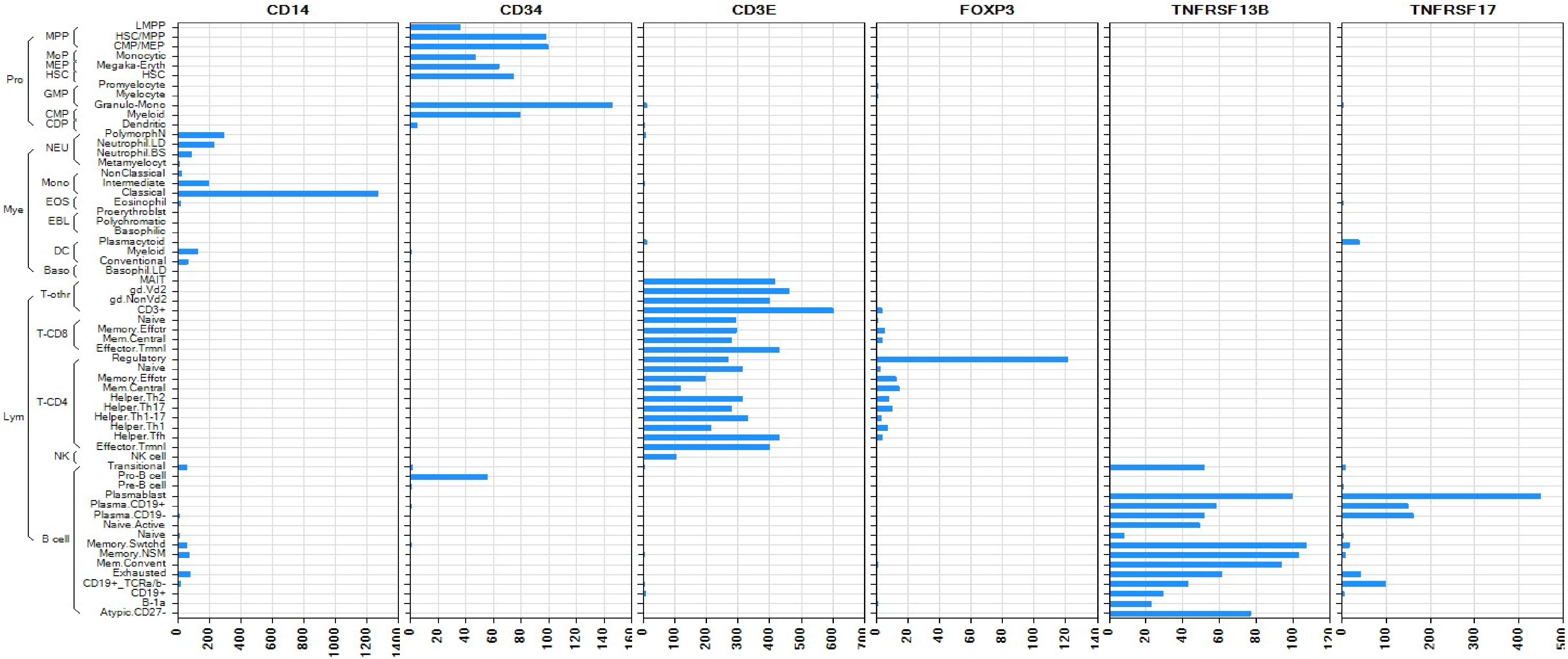
Blueprint

Bulk RNA-Seq of isolated immune cells



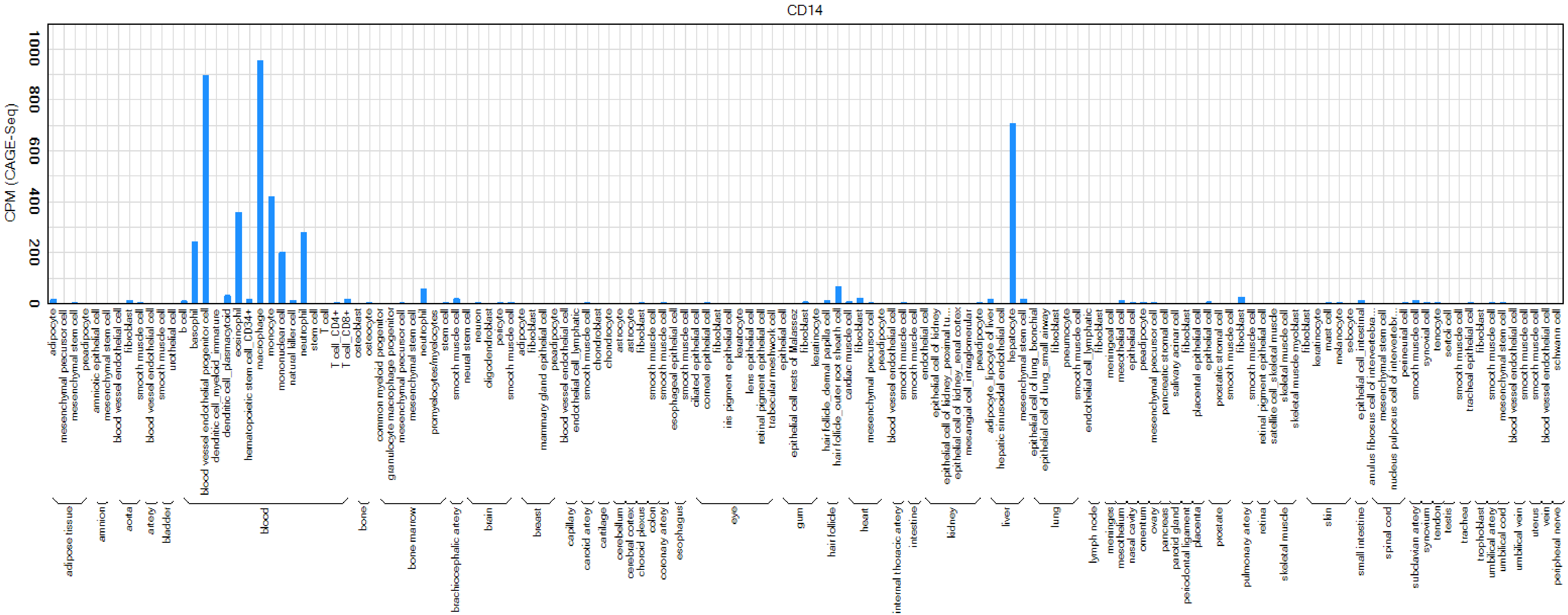
Immune Cell Atlas

Bulk RNA-Seq of isolated immune cells



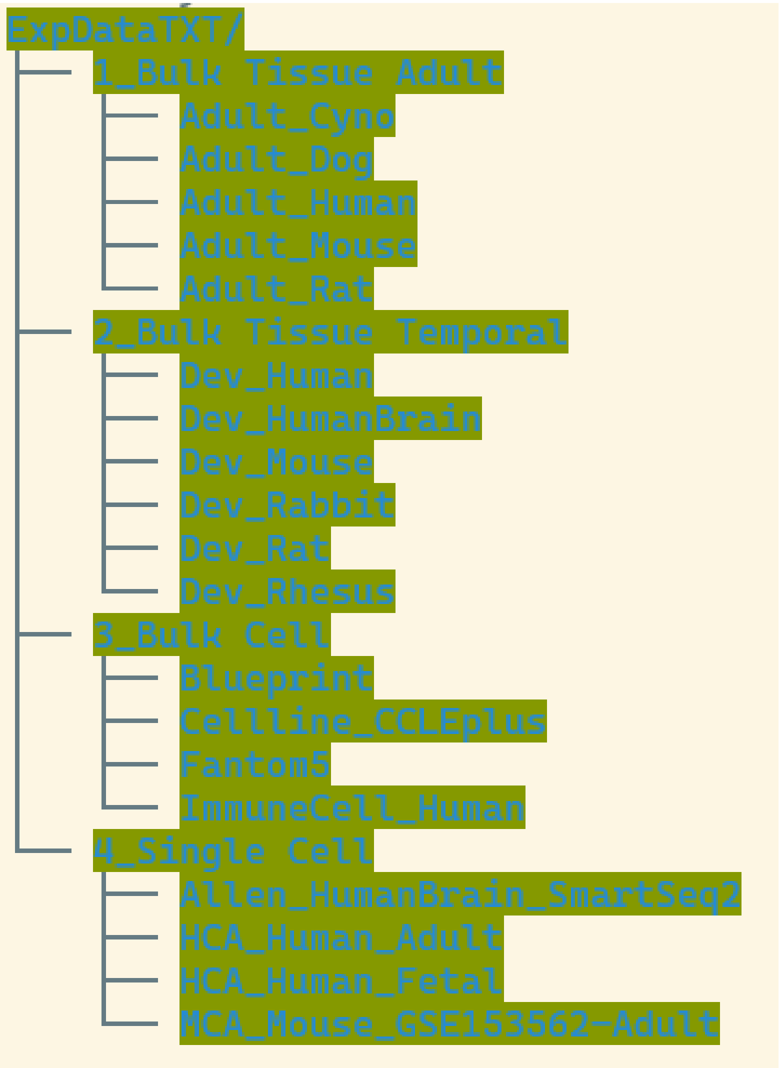
Fantom5

CAGE-Seq of isolated cells



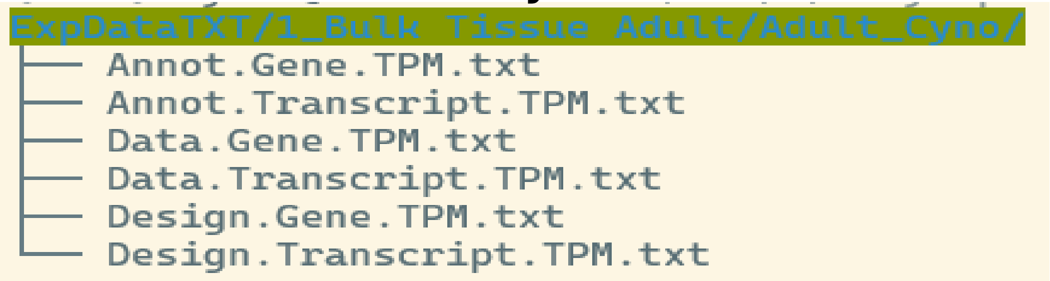
V. BodyMap data collection

- Data categories and Studies (TXT)



```
ExpDataTXT/
├── 1_Bulk Tissue Adult
│   ├── Adult_Cyno
│   ├── Adult_Dog
│   ├── Adult_Human
│   ├── Adult_Mouse
│   └── Adult_Rat
├── 2_Bulk Tissue Temporal
│   ├── Dev_Human
│   ├── Dev_HumanBrain
│   ├── Dev_Mouse
│   ├── Dev_Rabbit
│   ├── Dev_Rat
│   └── Dev_Rhesus
├── 3_Bulk Cell
│   ├── Blueprint
│   ├── Cellline_CCLEplus
│   ├── Fantom5
│   └── ImmuneCell_Human
└── 4_Single Cell
    ├── Allen_HumanBrain_SmartSeq2
    ├── HCA_Human_Adult
    ├── HCA_Human_Fetal
    └── MCA_Mouse_GSE153562-Adult
```

- Data sets in a study



```
ExpDataTXT/1_Bulk Tissue Adult/Adult_Cyno/
├── Annot.Gene.TPM.txt
├── Annot.Transcript.TPM.txt
├── Data.Gene.TPM.txt
├── Data.Transcript.TPM.txt
├── Design.Gene.TPM.txt
└── Design.Transcript.TPM.txt
```

  - Annot.xxx: Annotation of features (genes or transcripts)
  - Data.xxx: Expression measurements
  - Design.xxx: Sample information

- Package (data and app) folders

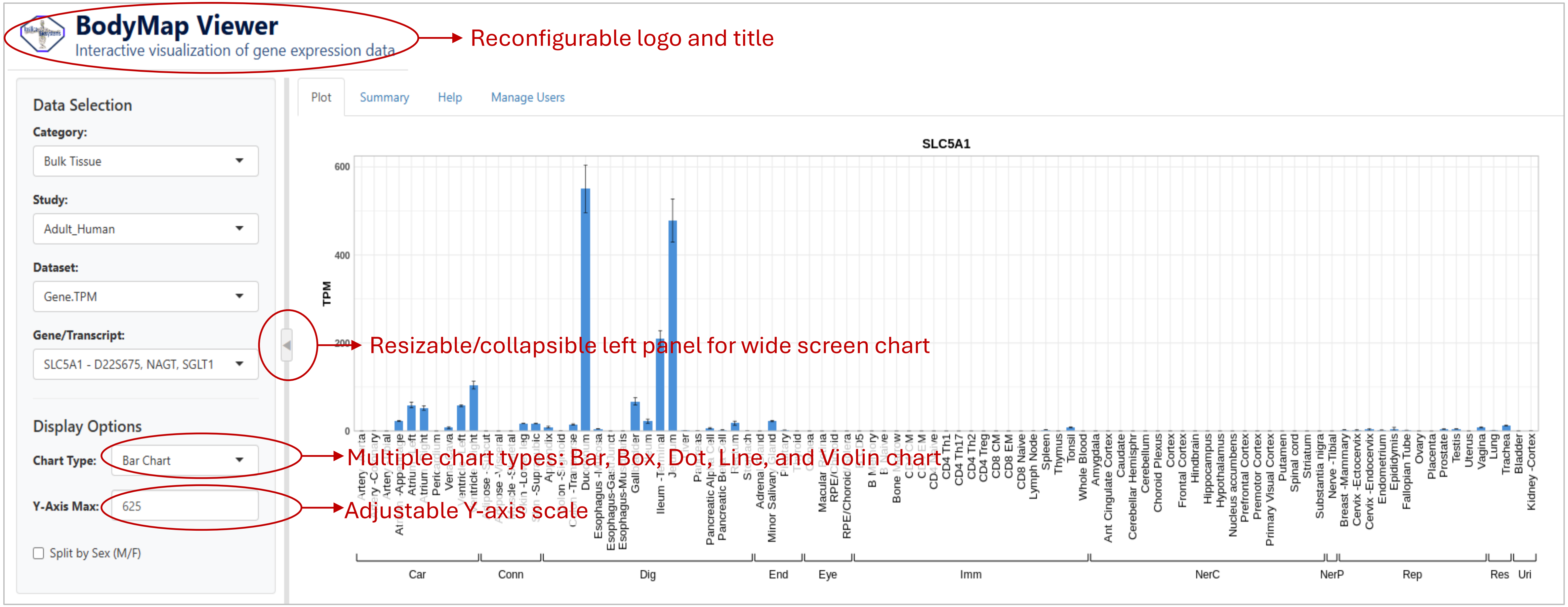
```
BodyMapShiny/ExpDataFST
BodyMapShiny/ExpDataTXT
BodyMapShiny/ExpressionViewer
```

- Code repository:

  https://github.com/isbs-xyao/BodyMapShiny/

V. BodyMap R-Shily app for data visualization

R-Shiny server with indexed data files for fast retrieval and easy configuration, interactive interface for flexible data presentation



BodyMap Viewer

Interactive visualization of gene expression data

Data Selection

Category: Bulk Tissue

Study: Adult_Human

Dataset: Gene.TPM

Gene/Transcript: SLC5A1 - D22S675, NAGT, SGLT1

Display Options

Chart Type: Bar Chart

Y-Axis Max: 625

☐ Split by Sex (M/F)

Reconfigurable logo and title

Resizable/collapsible left panel for wide screen chart

Multiple chart types: Bar, Box, Dot, Line, and Violin chart

Adjustable Y-axis scale

Note: Additional configurations in config.R

Company Website:

https://insilico-biosystems.com/