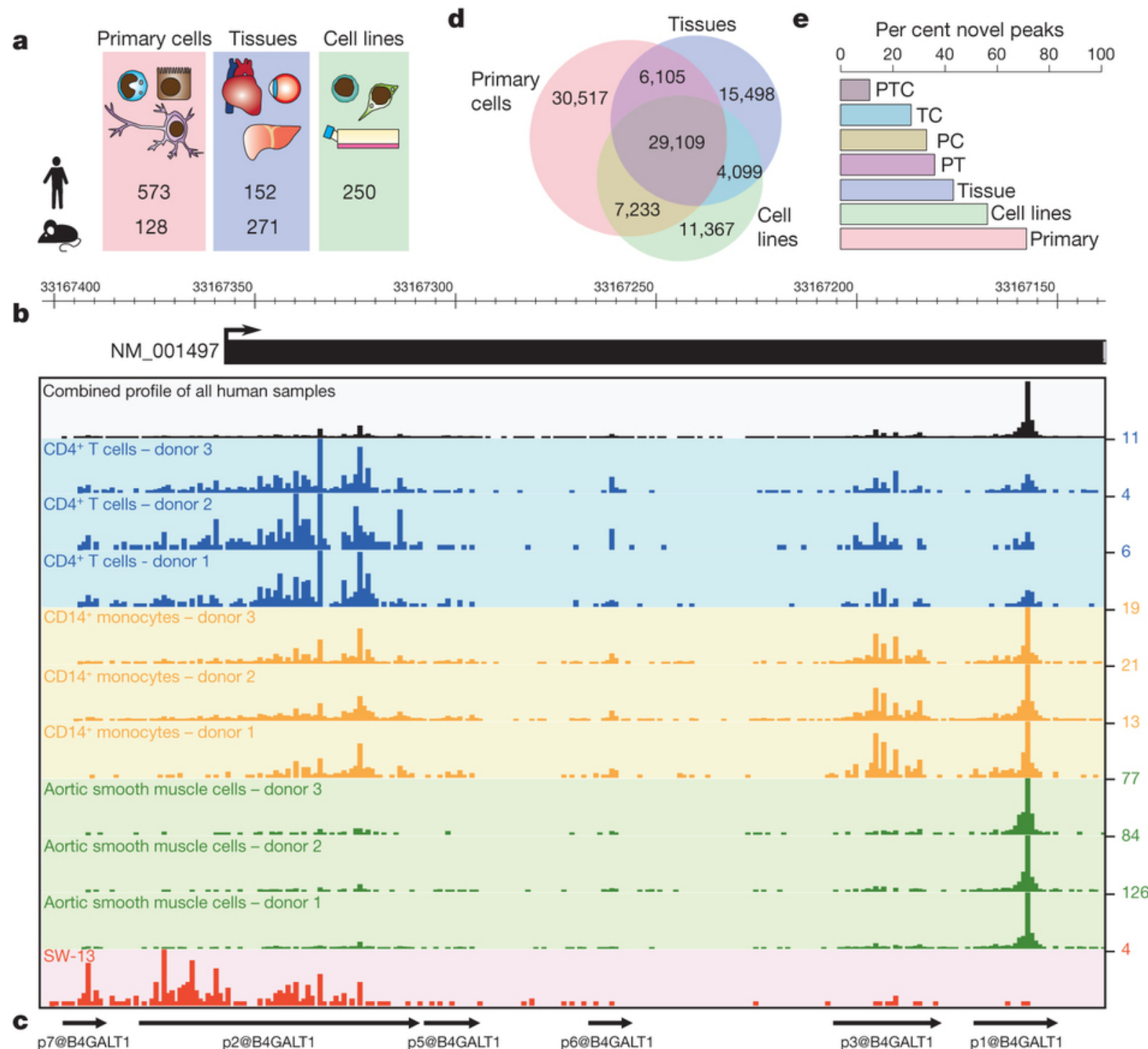# FANTOM 5 GO exercise

This exercise is best done in groups of 2-3 persons.

Dataset: FANTOM5

FANTOM 5 is a consortium that ran CAGE on almost all human tissues

From that, it is possible to define tissue-specific genes (or, really, TSSs)

I have given you 7 sets of tissue specific genes based on this data set, and 3 sets of random genes

All in the zip folder

gProfiler: the online tool of choice

- Go to the gProfiler website: http://biit.cs.ut.ee/gprofiler/

- Click on "Example query" for an example dataset

- What is included in the output?

  - What part is the GO-terms?

  - Where are the p-values? Is it corrected for multiple testing? (We will get back to this later…). Can you change to the FDR multiple testing correction?

- Try and copy-paste the FANTOM5 **setA.txt** into the query field and press "g:Profile!" to analyse.

  - Did you get any warnings? (HINT: Look at the bottom of the page)

- Now try supplying your own background set from **setExpressed.xt** (*HINT: look under "Show advanced options"*)

  - Did the results change?

- Discuss with your group what the remaining settings are

Task

- There is one set of all expressed genes, to be used as background.

- There are 10 different gene sets:

  - 7 sets of tissue-specific genes (Obtained from FANTOM5)

  - 3 random sets

- It is your task to guess what each set corresponds to using g:Profiler enrichment:

  - Which are the 3 random gene sets?

  - Which are the 7 tissue-specific gene sets?

  - What tissue are each of the 7 tissue specific sets actually from?

- When done, write down your best guesses in the absalon quiz called Gene Ontology quiz/contest (password is GoGoGo)

- Possible tissues are:

  - *Adipose*

  - *Aorta*

  - *Brain*

  - *Cerebellum*

  - *Colon*

  - *Heart*

  - *Kidney*

  - *Liver*

  - *Lung*

  - *Pituitary Gland*

  - *Skin*

  - *Smooth muscle*

  - *Spinal cord*

  - *Spleen*

  - *Submaxillary gland*

  - *Testis*

  - *Thyroid*

  - *Uterus*

  - *Vein*

gProfiler R-package

- **IF YOU ARE QUICK:**

  - Download the accompanying R-package gProfiler from CRAN.

  - Read the documentation for the included gprofiler() function

  - Figure out how to solve the assignment using only R!

  - Is there a way you can analyse all sets in a single function call?