Index table v. 2.0

Matt Maciejewski

20 November 2019

Inflammation and Immunology Research Unit

Pfizer WORLDWIDE RESEARCH & DEVELOPMENT

SCIENCE FOR LIFE-CHANGING IMPACT

UC colon

- 22 datasets available with some combination of inflamed and non-inflamed samples
- Developing 2 contrasts here:
  - UC (inflamed) vs healthy,
  - and UC inflamed vs UC non-inflamed

UC blood (👍) & ileum (👎)!

For UC blood only a modest amount of data was available:

```
uc_blood_data:
['GSE86434', 'GSE3365', 'GSE94648']
  dataset_name                                                disease  \
0     GSE86434  ulcerative colitis (UC);crohn's disease (CD);n...
1      GSE3365  ulcerative colitis (UC);crohn's disease (CD);n...
2     GSE94648  ulcerative colitis (UC);normal control;crohn's...

            tissues sample_pathology  \
0  peripheral blood          No Info
1  peripheral blood          No Info
2  peripheral blood          No Info

                                        cell_type treatment
0  ; monocyte; helper T cell; naive CD4+ T cell; ...   No Info
1           peripheral blood mononuclear cell (PBMC)   No Info
2                                          No Info   No Info
```

- 2/3 contained “weird” celltypes rather than bulk
- GSE86434 was salvageable as it contains a number of samples (21 UC, 13 healthy) w/o a specific celltype designation – so I have 2 datasets for the UC vs healthy contrast here

Similarly, only a couple of ileal datasets available:

```
uc_ileum_data:
['GSE16879', 'GSE1152', 'GSE1141', 'GSE1142', 'GSE75214']
  dataset_name                                                disease  \
0     GSE16879  normal control;ulcerative colitis (UC);crohn's...
1      GSE1152  normal control;crohn's disease (CD);ulcerative...
2      GSE1141  normal control;crohn's disease (CD);ulcerative...
3      GSE1142  normal control;ulcerative colitis (UC);crohn's...
4     GSE75214  crohn's disease (CD);normal control;ulcerative...

                    tissues                sample_pathology cell_type
0  colonic mucosa;ileal mucosa                       No Info   No Info
1  ileal mucosa;colonic mucosa                   non-inflamed   No Info
2  ileal mucosa;colonic mucosa                   non-inflamed   No Info
3  ileal mucosa;colonic mucosa                   non-inflamed   No Info
4  colonic mucosa;ileal mucosa  non-inflamed; normal; inflamed   No Info
```

- None of these were salvageable as if you specify “ileum” & “inflamed UC or healthy” & “no weird celltypes” you get only healthy samples from GSE16879 and GSE75214

CD

- 12 datasets available where a combination of inflamed & non-inflamed samples is present
- 4 ileum datasets (CD vs healthy)
- 2 blood datasets (CD vs healthy)

RA

- 5 synovial datasets: healthy vs RA
- 2 blood datasets: healthy vs RA
  - Out of 10! But the majority were notoriously focused on celltypes…

AD

- 8 AD vs healthy datasets
- 11 AD involved vs non-involved
- No blood data!

Psoriasis

- ~30 skin datasets – pso vs healthy & pso involved vs non-involved
- Again, no blood samples (5/6 available were revealed to be singular cell - focused)

Systemic Lupus Erythematosus (SLE)

- 8 blood datasets – SLE vs healthy (more available for singular celltypes, e.g. easily 8 with PBMCs)

NASH / NAFLD

- ~10 liver datasets for the NASH / NAFLD / healthy trio (all pair permutations calculated)

Obesity

- Comparison of obesity vs healthy in:
- 13 adipose tissue samples, and
- 7 blood datasets

Asthma

- Here's the celltype situation for the asthma datasets:
- Taking all of these still, as tissue was already specified as airway epithelial etc. (lavage included, e.g. bronchoalveolar lavage)
- So here we have 16 tissue datasets for asthma vs healthy
- Only 2 potential blood datasets (out of 11, where the majority are T-cell, or PBL or some other singular cell types).

	cell_type
0	No Info
1	No Info
2	airway epithelial cell
3	airway epithelial cell
4	airway epithelial cell
5	No Info
6	bronchial epithelial cell
7	airway epithelial cell
8	airway epithelial cell
9	bronchial epithelial cell
10	bronchial epithelial cell
11	nasal epithelial cell
12	bronchoalveolar lavage cell
13	No Info
14	alveolar macrophage
15	epithelial cell

COPD

- Similar celltype situation as in asthma, with lots of epithelial celltype datasets
- So here we have 14 tissue datasets for COPD vs healthy
- ~5 blood datasets

Psoriatic arthritis

- This was abandoned since the only 2 datasets that had both pso arthritis and normal control were blood datasets with singular celltypes (rather than cell cocktail)

Allergic rhinitis

- 3 nasal datasets, as shown opposite
- No blood datasets

	tissues
0	bronchial airway epithelium;nasal epithelium
1	nasal mucosa
2	nasosinus mucosa;nasal polyps

	cell_type	treatment
0	airway epithelial cell	No Info
1	airway epithelial cell; No Info	No Info
2	No Info	No Info

Diabetes type 1

- Only 1 pancreas dataset – ignoring
- 4 blood datasets (out of 11 where the rest are singular cell)

Diabetes type 2

- Looking at diabetes type 2 vs healthy
- 3 adipose datasets
- 9 muscle datasets
- 1 pancreas dataset – ignoring
- 3 blood datasets

Multiple sclerosis

- 3 brain datasets (1 of 4 was removed as there was only 1 control sample)

Liver cirrhosis/fibrosis

- ~5 datasets on the list that mention liver cirrhosis/fibrosis
- However, once we request datasets with liver samples from both cirrhosis/fibrosis AND healthy, only 1 dataset remains
- So, disregarding

Skin sclerosis

- 5 seemingly eligible datasets with skin sclerosis and healthy samples…
- Only 1 with a cell cocktail, 4 with singular cell types
- Disregarding!

Idiopathic pulmonary fibrosis

- Checking for IPF vs healthy datasets in lung, blood, and lavage
- Only one dataset in lavage – disregarding
- Only 2 datasets in blood, one of which is PBMCs – disregarding
- 7 datasets in lung that have cell cocktail samples – deriving metaanalysis for these