***FANCONI Proyect folder organization***

All the data is in /datos\_2/FANCONI, navarrabiomed server

**AUCell analysis**: all the AUCell analysis carried out. Each of the folders is the name of the pathway that is used to do the analysis. The important one here is the one named INTERSECTION\_our\_data. Here we do the AUCell analysis with our signature.

**DATA**: where all the .rds files are

* Fanconi\_02002:
  + Fanconi\_02002.rds: fanconi data
  + Integration.RDS: integration of fanconi\_02002 and healthy
* Fanconi\_02004:
  + CD34: all the data from bone marrow CD34 cells
    - Fanconi\_02004.RDS: fanconi sample data
    - Integration\_02004.RDS: integration of fanconi 02004 and healthy
  + SP: peripheral blood sample
* Fanconi\_02006:
  + CD34: all the data from bone marrow CD34 cells
    - Fanconi\_02006\_updated.RDS: fanconi sample data
    - Integration.RDS: integration fanconi 02006 and healthy
  + SP: peripheral blood sample
* Fanconi\_02006\_followup: followup data, but NOT USE, we have the same data resequenced
* fanconi\_02006\_followup\_resequenced: the follow up sample resequenced, .rds is the good data
* Fanconi\_02008:
  + CD34: all the data from bone marrow CD34 cells
    - Fanconi.RDS: fanconi sample data
    - Integration.RDS: fanconi 02008 and healthy samples integrated
  + SP: peripheral blood data
    - Sp\_2008.rds
* Healthy\_sample:
  + Mo268\_names.RDS: the healthy donor used in the analysis
  + Seurat\_young: the healthy donors (3 inside the sample). This data was used to do the label transfer (annotation)
* Integration\_3\_samples: the integrated data of the first 3 patients (02004,02006,02008) and the healthy
  + All\_integrated: 3 fanconi and healthy integrated
  + Fanconis\_integrated: only the 3 fanconi samples integrated
  + SP\_integrated: peripheral blood sample of the 3 fanconi patients
* Integration\_4\_samples: integration of all fanconi samples and the healthy one
  + 4\_fanconi\_healthy.RDS: 4 fanconis and healthy integrated
  + 4\_integrated.RDS: 4 fanconis integrated

**FANCONI\_02002:** the results of the analysis of fanconi 02002 sample

* Plots
* DEA (Corrected\_vs\_corrected) per cell type
* GSEA:
  + GSEA analysis per celltype, all the results stored in a rda file
  + Integration: the analysis of fanconi 02002 and healthy integrated
    - Plots
    - DEA: healthy vs corrected and healthy vs uncorrected
    - GSEA: all the results stored in a rdata

**FANCONI\_02004**: the results of the analysis of fanconi 02004 sample

* Plots
* DEA (corrected\_vs\_uncorrected)
* NEW\_GSEA: all the gsea analyses results and stored in a rda file
* MYC\_p53: GSEA analysis of these two pathways
* SP: results of peripheral blood analysis (heatmaps, percentages, UMAPS)

**FANCONI\_02006:** the results of the analysis of fanconi 02006 sample

* Annotation featureplots: feature plots of cell type markers to make sure the annotation was working well
* DEA\_all: differential expression of the sample (corrected\_vs\_uncorrected)
* GSEA\_public\_paper: GSEA analysis with the MYC pathway they describe in the public paper
* HD\_figures: high definition figures to paper
* MYC\_p53: GSEa of these pathways; MYC and p53
* NEW\_GSEA: GSEA analysis, all data stored in a rda
* PLOTS
* SP: the analysis done with peripheral blood sample (integration with public healthy dataset)

**Fanconi\_02006\_followup:** plots, dea and GSEA analysis. NOT TAKE INTO ACCOUNT the good data is the resequenced one.

**Fanconi\_02008:** the results of the analysis of fanconi 02008 sample

* Plots
* DEA corrected vs uncorrected
* Integration:
  + All the analysis with the integration of fanconi 02008 and healthy
* MYC\_p53: GSEA analysis of these two pathawys
* NEW\_GSEA: GSEA analysis of fanconi sample
* SP: analysis of the peripheral blood sample

**Fanconi\_2006\_followup\_resequenced:**

* Seurat\_modulescores: a small try with this function, very similar to AUCell. One pdf for each one of the samples, density plot of the score in corrected and uncorrected cells
* Fanconi\_2006\_resequenced: all the standart analysis of this sample, plots and DEA
* Integration: analysis of both 02006 samples integrated, the original and the follow up.

**Fanconi\_3\_integrated:** the final analysis of three fanconis (02004,02006,02008) and the healthy one integrated. Most of the pdfs here where use for the first draft of the paper.

**Gene lists and heatmaps:** all this data is obtained using only 3 fanconi patients, not usefull for the paper

**GSE\_microarray:** a small analysis with microarrays to check that our results were robusts

**GSEA\_ridgeplots:** all the plots of GSEA analysis of the 3 fanconi patients. .RDA with all the info the contrast healthy\_vs\_uncorrected

**Integration 02004 healthy:** the same analysis with the integration of 02004 and healthy donor

**Integration 02006 healthy:** the same analysis with the integration of 02006 and healthy donor

**MERGE\_4:** all the final pdf and files need to create the figures with 4 samples

**Pathview:** pathway visualization. The first list is with the contrast corrected vs uncorrected. Then there are two more folders: healthy vs uncorrected and mean values. In the mean values folder, the color of the graph is calculated with the mean of the 3 fanconi samples

**Public data:**  the data of the public paper fastq, counts of each one of the samples and some small analysis

**Public\_SP:** the analysis of some public peripheral blood data (healthy) to compare the expression values with our fanconi samples

**SCRIPTS**

**Statistical analysis**: all the results of the statistical test, binomial, correlation etc for each one of the samples and corrected by multiple testing