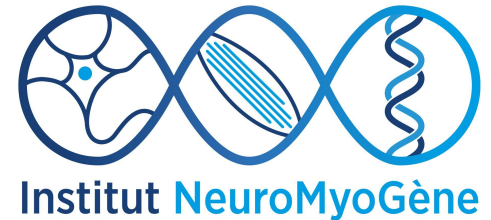# RNA-seq analysis of 6 cell types in regeneration muscle tissue between young and old mice
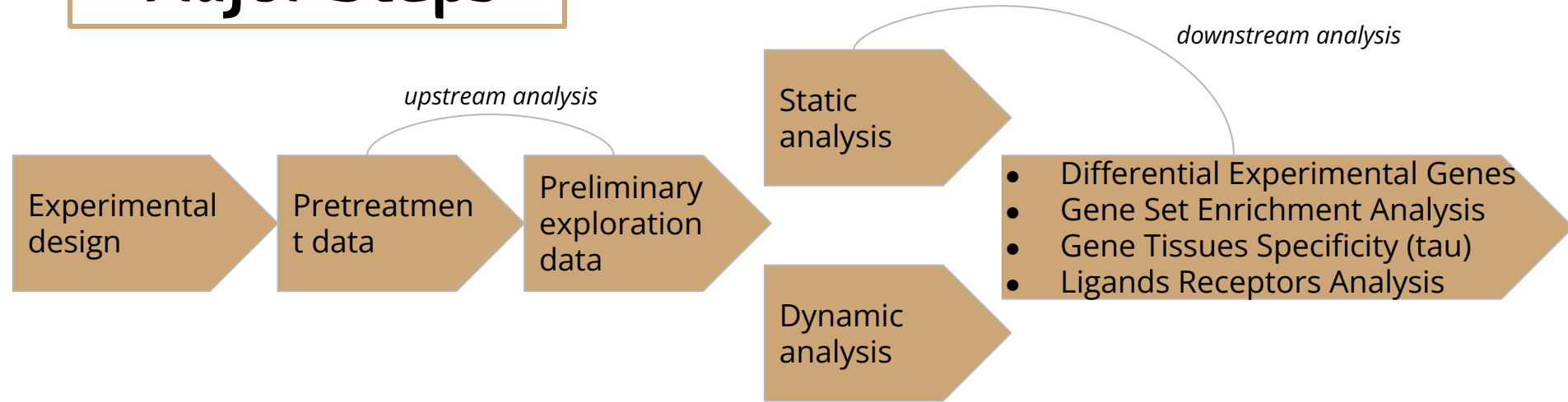
Partenariat des équipes de:

- Dr. Fabien LE GRAND
- Dr. Bénédicte CHAZAUD
- Dr. Jérôme FEIGE

Contributeurs Principaux:

- PhD Hoang DIEU-HUONG
- IE Johanna GALVIS-LACROUX
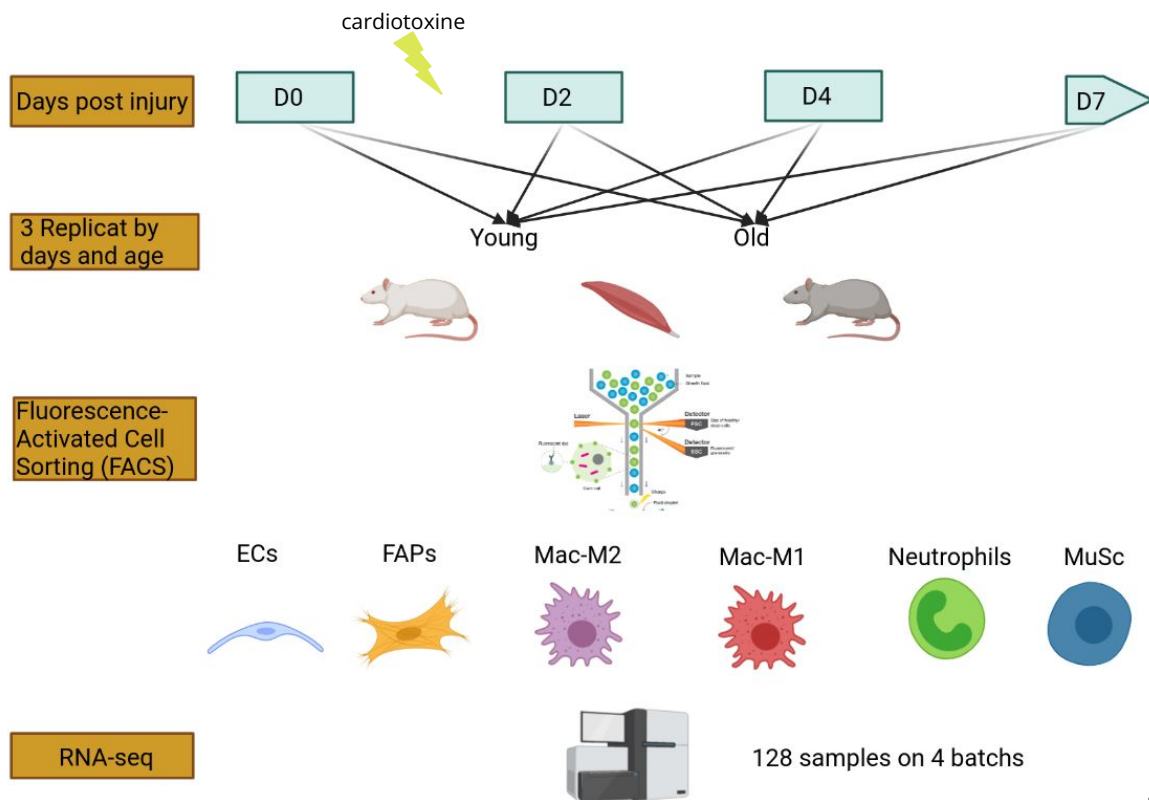- IE William JARASSIER
- IE Pauline MOULLÉ

Institut NeuroMyoGène

Major Steps

Experimental design → Pretreatment data → *upstream analysis* → Preliminary exploration data

Static analysis

Dynamic analysis

downstream analysis

- Differential Experimental Genes
- Gene Set Enrichment Analysis
- Gene Tissues Specificity (tau)
- Ligands Receptors Analysis

Objectives: Find the pathways and genes differentially expressed between young and old

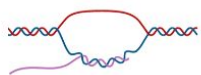Represent the interactome of these changes

Experiment design

18 conditions

Cell sequenced	D0	D2	D4	D7
ECs				
FAPs				
Mac-M1				
Mac-M2				
Neutrophils				
MuSc				

cardiotoxine

Days post injury — D0 — D2 — D4 — D7

3 Replicat by days and age — Young — Old

Fluorescence-Activated Cell Sorting (FACS)

ECs   FAPs   Mac-M2   Mac-M1   Neutrophils   MuSc

RNA-seq — 128 samples on 4 batchs

Pretreatment RNA-seq data

Filter1: Keep only 40% express gene

105 samples and 20 257 genes expressed
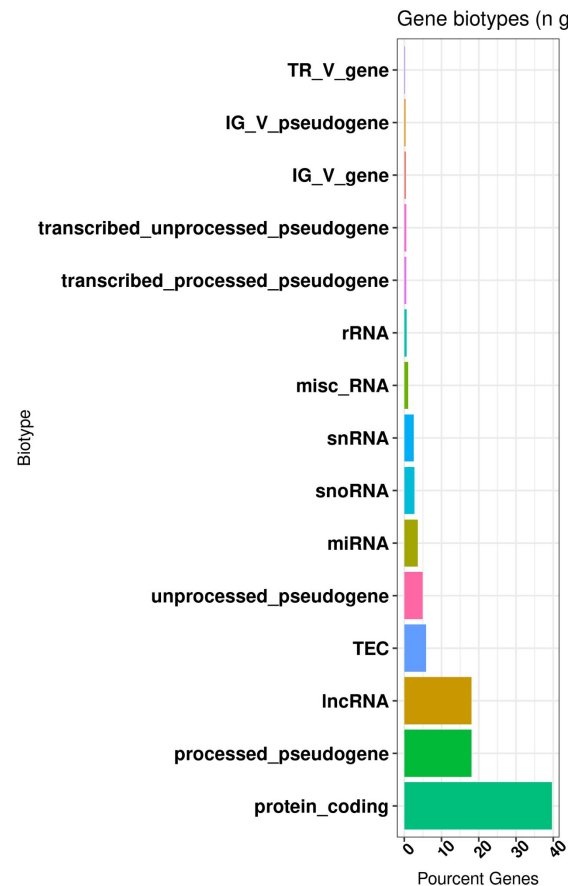
* Filter4: Genes associated with severals transcripts

Preliminary exploration

Principal Component Analysis

PCA for each type of cell

Expression grouping by days post injury

PCA on all data frame

Expression grouping by type cell

PCA for each type of cell and for each day post injury
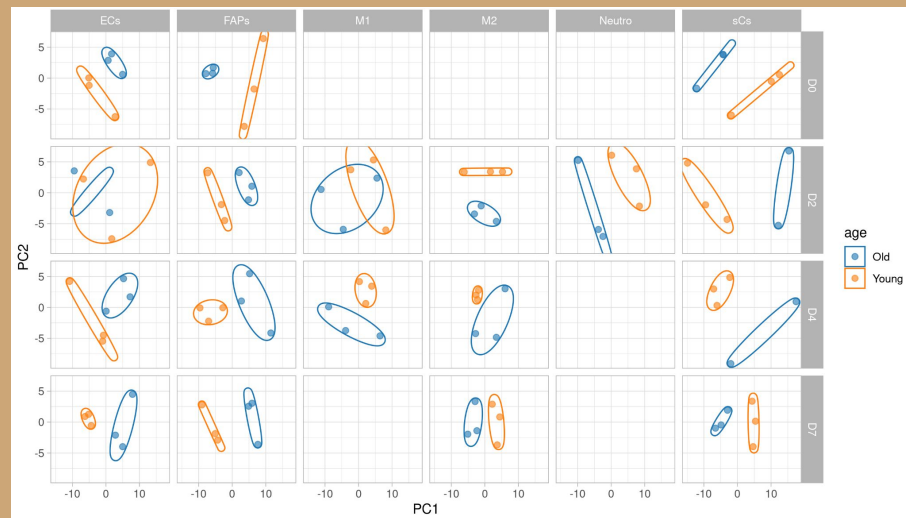
Expression slightly grouping by age

Major Steps

Experimental design

Pretreatment data

Preliminary exploration data

upstream analysis

Static analysis

downstream analysis

Dynamic analysis

- Differential Expressed Genes
- Gene Set Enrichment Analysis
- Gene Tissues Specificity (tau)
- Ligands Receptors Analysis
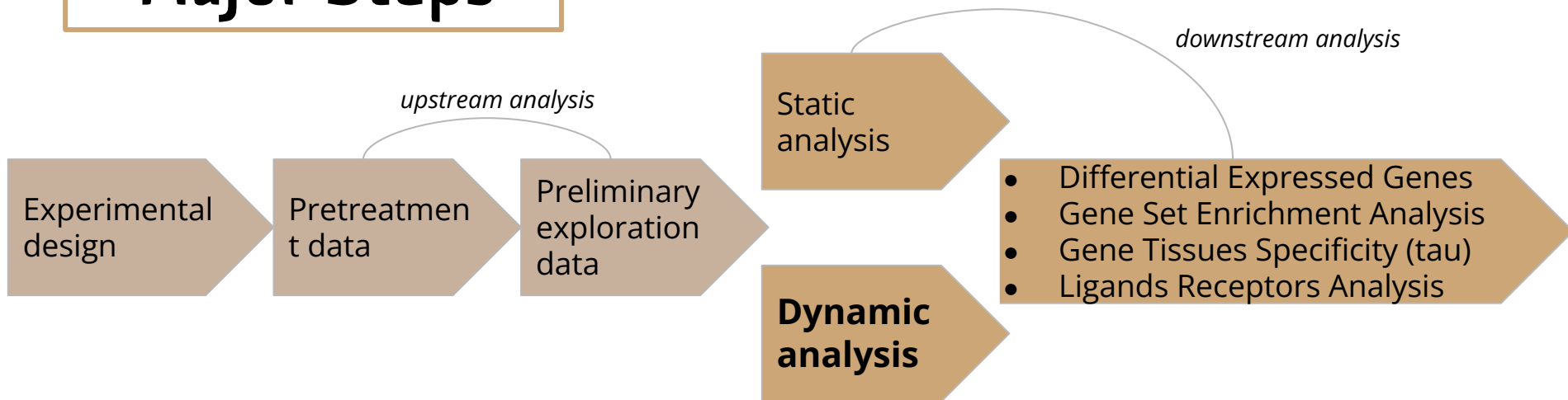
Differential expression between 2 time points:

Cell sequenced	D0	D2	D4	D7
Ecs				
FAPs				
Mac-M1				
Mac-M2				
Neutrophils				
MuSc				

Major Steps

Experimental design → Pretreatment data → Preliminary exploration data

upstream analysis

Static analysis

downstream analysis

Dynamic analysis

- Differential Expressed Genes
- Gene Set Enrichment Analysis
- Gene Tissues Specificity (tau)
- Ligands Receptors Analysis
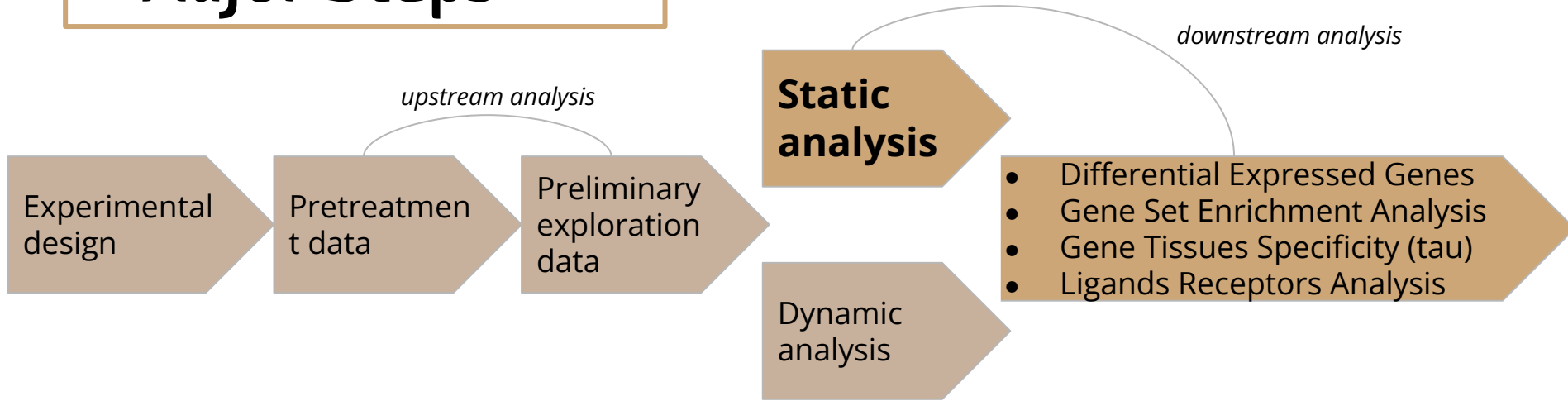
Axis of analysis

1. Young vs Old
   1.1. Cell type:
   - Endothelial Cell - ECs
   - Neutrophils
   - Muscle Stem Cell - MuSCs
   - Macrophage 1 - Mac_M1
   - Macrophage 2 - Mac_M2
   - FibroAdipogenic Progenitors - FAPs
   1.2. Days post injury : D0, D2, D4 et D7

Pretreatment raw data

fastq: ask william

•nf-core: Localisation pipeline William
- Mapping
  -STAR
  -genome ref GRCm38
  (mm10)
- Counting
  -FeatureCount
  (raw counts)
  -StringTie
  (TPM,RPKM et FPKM)
- Quality Control
  -report : MultiQC

total_results_nfcore/:
•Stringtie/FPKM/genes/*
•FeatureCounts/
•merged_gene_counts.ot
→ Matrix counts/TPM/FPKM

•scripts/get_expre_to_csv_rdf.R:
→ Merge 4 batchs rawcount and merges les files StringTie
•scripts/get_biotypes_bulk.R :
→ association avec type transcrit + homogénéisation gène
•scripts/get_protcod_TPKMolIiCounts.R:
→ Filter1 : Protein coding only (Gene Annotation : Ensembl & ncbireference)
•scripts/libraries_kept_andPCA.R
→ Filter2 : 22 Technical sample removing
→ Filter3 : Gene express only (1673 removed)
→ Preliminary aexploration with PCA

data/
•prefiltered_counts.rds
•prefiltered_TPM.rds
•metadata.rds=deesignMatrice
•plotsPrelim=PCA

Specificity Cell Analysis - Tau Index

data/
•metadata.rds
•prefiltered_TPM.rds
•COV/matrix.csv

DEG Analysis

data/
•metadata.rds
•prefiltered_counts.rds
•COV/matrix.csv

•scripts/exam_Inter_cond_sta.R :
→ Keep genes with on less 3 samples with raw count >= 5
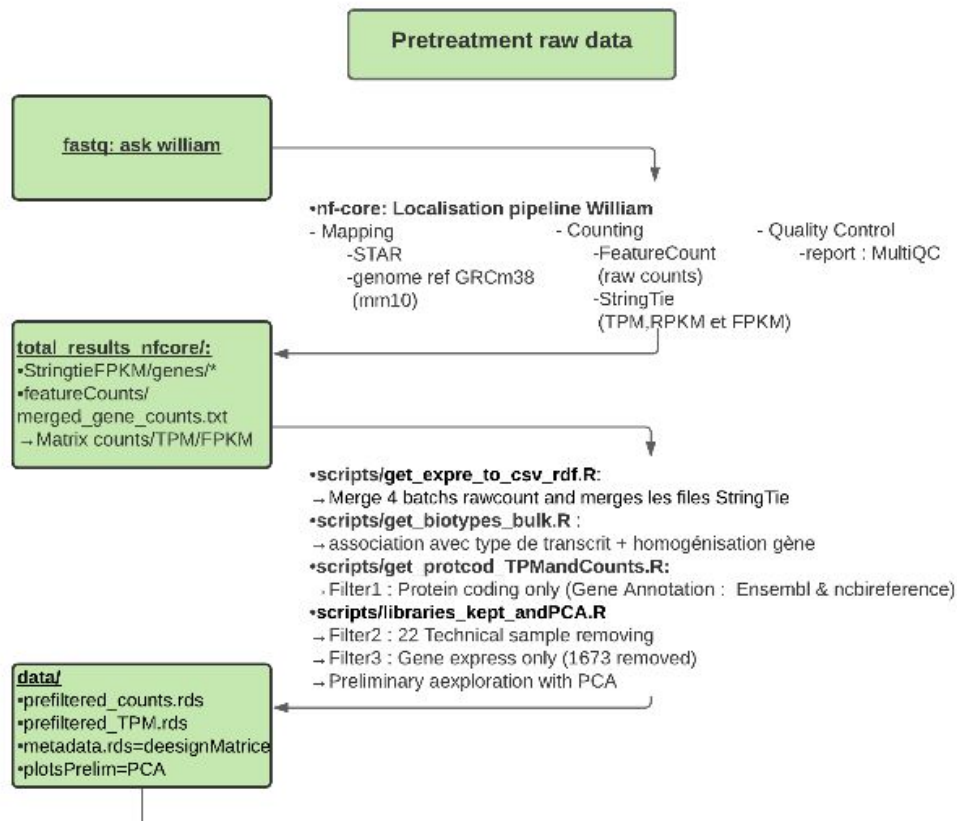→ DEG with DESeq2 on Young-Old for each 18 conditions
  TypeCell.DayPostInjury
  →softFilters padj<0.5 & |fc|> 0.1
  →finalFilters padj<0.05 & |fc| >=1.2

**exam_INTER_conditions/
static/:**
•rds/shot_rds_full.rds
•csv/shot_dataframe_softfilter.csv
•csv/shot_dataframe_finalfilter.csv
•volcano_base/MeanInfo.png

GSEA Analysis

**exam_INTER_conditions/
static/:**
•rds/shot_rds_full.rds
•R[msgdb(species = 'Mus
musculus", category = 'C2',
subcategory: 'CP:Reactome')

•scripts/fgsea_dayandcellhype.R :
→ fGEA on each days and cellType, keep top (min padj) 500
  genes with a regulation Up and 500 Down
  - filters = top 15 min padj of pathway UP et Down
•scripts/fgseaTotablesToplots.R
→ Trie sur padj, NES, number of gene in pathway

**exam_INTER_conditions/
static/GSEA:**
•rds/fgsea_bd_bct_full.rds
•rds/fgsea_bd_bct_filtered.rds
•rds/fgsea_matrices4heatmaps.rds
•csv/Total_GSE_pathways.xlsx
•plots & heatmaps

NATMI Part - Receptor Ligand interactions

data/
•metadata.rds
•prefiltered_TPM.rds
•COV/matrix.csv

•scripts/PrepareData/mp_prep4Natmi.R
→ remove genes with rowSum(coverage =0)
→ created table annot associated sample name with cellType

inDataNatmi/
•annot_AgeDay.txt
•TPM_AgeDay.txt

•scripts/natmi_cmd.sh
→ Run Natmi to detect LR
•scripts/natmi_diffdges.sh
→ Extract LR up/down/appeared/dispeared

natmi_edges.sh
→ Chordiagr + heatmap + network
•natmi_diffViz.sh
→ Chordiagr + fexatmap + network

natmiOut/AgeDay
•One file.xlsx by LR
•File with all EdgesLR
•All graph recap

natmiOut/DiffDay
•One file all_edges_mean
•One file appeared and all
from_x_to_y
• Same for disappeared up
down

•networks_explore/L_LR_networks:
→ Save dico, networks filter and non filter
→ dico(ages)[Days] = Edges.csv + ID uniqu Ligand & recepteur
→ network:Node1 = Id uniq Ligand from + symbol + cellType +
  specificity (rule/vlevel of average expression)
  Node2 = Id uniq Receptor to
  Edge = nodefrom , nodeto, coffrm, Edge average
  expression derived specificity

Filtre on Edge average
expression derived
specificity >= -1:
the product of the ligand
and receptor derived
specificity of average
expression values

scripts/Merge_Result_LR_pop_up_down.py
→ Table recap in natmiOut : on 1 sheet by
Day_typeCell concat info
appeared/disappeared/up/down

natmiOut/
•diff_LR_pop_up_down_regulation_
total.xlsx

networks_explore/graphobis/:
•dicto_dx_unfi.p
•dicto_dx_filtered.p
•dicto_lr.p : /name & /filered
•age_Days_igraph_filUnfi.ml

AgeinMice_shinyApp/scripts/aggreg_matrice.R:
→ Count/cumulate/caculated number/Weight/Ratio (Weight/count)
interaction from typeCell to type Cell

AgeinMice_shinyApp/Data
•aggreg_matrices.rds

Shiny App Muscle and Age

AgeinMice_shinyApp:
→ copy of exam_INTER_conditions
/static/csv/shot_dataframe_finalfilter
+ symbolID in
DE_copy/shot_dataframe.csv

AgeinMice_shinyApp:
→ copy of exam_INTER_conditions/
static/rds/fgsea_matrices4heatmaps
in Data

AgeinMice_shinyApp:
•Data/aggreg_matrices.rds
•copy of networks_explore/graphobjs/
age_Days_igraph_unfi.ml in
graphobj_copy

Explore network LR

Selected DAY

L-R pairs

DEG

Pathways

→_Load_: ChordDiagram + Table of count/weight/Ratio
Ligand from typeCellA to Receptor typeCellB
→_Table Young/Old_: Print interactive table lr.frame Young
and Old - filtre possible on column by name + order.
→_Cross_: do crossing -> print gene_cellType
Exclusive/Young/Old and shared

→_DEGs_: Print interactive table shot_dataframe
→_Intersect DII with loaded LR table_: Print
gene in DEG and LR table, Up regulated in
Young/Old DEG but no in table LR

Selected one or list of gene_cellType in Young or Old condition & number neighbour

Network

Animated:
→Nodes label = symbol
→Node value = specificty*10
→Node color = TypeCell
→Edges value = weight *10
→Edges color = Ligand TypeCell

Pretreatment raw data

fastq: ask william

•**nf-core: Localisation pipeline William**
- Mapping   - Counting   - Quality Control
    -STAR   -FeatureCount   -report : MultiQC
    -genome ref GRCm38   (raw counts)
    (mm10)   -StringTie
                                               (TPM,RPKM et FPKM)

total_results_nfcore/:
•StringtieFPKM/genes/*
•featureCounts/
merged_gene_counts.txt
→Matrix counts/TPM/FPKM

•scripts/**get_expre_to_csv_rdf.R**:
→ Merge 4 batchs rawcount and merges les files StringTie
•scripts/**get_biotypes_bulk.R** :
→ association avec type de transcrit + homogénisation gène
•scripts/**get_protcod_TPMandCounts.R:**
  ›Filter1 : Protein coding only (Gene Annotation : Ensembl & ncbireference)
•**scripts/libraries_kept_andPCA.R**
→ Filter2 : 22 Technical sample removing
→ Filter3 : Gene express only (1673 removed)
→ Preliminary aexploration with PCA

data/
•prefiltered_counts.rds
•prefiltered_TPM.rds
•metadata.rds=deesignMatrice
•plotsPrelim=PCA

Specificity Cell Analysis - Tau Index

data/
- metadata.rds
- prefiltered_TPM.rds
- COVmatrix.csv

DEG Analysis

data/
- metadata.rds
- prefiltered_counts.rds
- COVmatrix.csv

• **scripts/exam_Inter_cond_sta.R** :
  → Keep genes with on less 3 samples with raw count >= 5
  → DEG with DESeq2 on Young-Old for each 18 conditions
       TypeCell.DayPostInjury
       → softFilters padj<0.5 & |lfc|> 0.1
       →finalFilters padj<0.05 & |lfc| >=1.2

exam_INTER_conditions/
static/:
- rds/shot_rds_full.rds
- csv/shot_dataframe_solftfilter.csv
- csv/shot_dataframe_finalfilter.csv
- volcano_baseMeanInfo.png

GSEA Analysis

exam_INTER_conditions/
static/:
- rds/shot_rds_full.rds
- |R|msigdbr(species = "Mus musculus", category = 'C2', subcategory: 'CP:Reactome')

• **scripts/fgsea_dayandcelltype.R** :
  → fGEA on each days and cellType, keep top (min padj) 500
       genes with a regulation Up and 500 Down
       → filters = top 15 min padj of pathway UP et Down
• **scripts/fgseaTotablesToplots.R**
       → Trie sur padj, NES, number of gene in pathway
       → ....

exam_INTER_conditions/
static/GSEA:
- rds/fgsea_bd_bct_full.rds
- rds/fgsea_bd_bct_filtered.rds
- rds/fgsea_matrices4heatmaps.rds
- csv/Total_GSE_pathways.xlsx
- plots & heatmaps

NATMI Part - Receptor Ligand interactions

data/
- metadata.rds
- prefiltered_TPM.rds
- COVmatrix.csv

• **scripts/PrepareData/**tmp_prep4Natmi.R :
  → remove genes with rowSum(coverage =0)
  → created table annot_ associated sample name with cellType

inDataNatmi/
- annot_AgeDay.txt
- TPM_AgeDay.txt

• **scripts/**natmi_cmd.sh
  → Run Natmi to detect LR
• **scripts/**natmi_diffEdges.sh
  → Extract LR up/down/appeared/dispeared

natmiOut/AgeDay
- One file.xlsx by LR
- File with all EdgesLR
- All graph recap

• **networks_explore**/1_LR_networks:
  → Save dico, networks filter and non filter
  → dico[ages][Days] = Edges.csv + ID unique Ligand & recepteur
  → network:Node1 = Id uniq Ligand from + symbol + cellType +
         specificity (=derived of average expression)
       Node2 = Id uniq Receptor to
       Edge = nodefrom , nodeto, colfrm, Edge average
         expression derived specificity

Filtre on Edge average
expression derived
specificity >= -1:
the product of the ligand
and receptor derived
specificity of average
expression values.

networks_explore/graphobjs/:
- dictio_dx_unfi.p
- dictio_dx_filtered.p
- dictio_lr.p : .frame & .filtered
- age_Days_igraph_filt/unfi.ml

AgeinMice_shinyApp/scripts/aggreg_matrice.R:
→ count/cumulate/caculated number/Weight/Ratio (Weight/count)
interaction from typeCell to type Cell

AgeinMice_shinyApp/Data
- aggreg_matrices.rds

• natmi_edges.sh
  → Chordiagr + heatmap + network
• natmi_diffViz.sh
  → Chordiagr + heatmap + network

natmiOut/DiffDay
- One file.all_edges_mean
- One file appeared and all from_x_to_y
- Same for disappeared up down

scripts/Merge_Result_LR_pop_up_down.py
  → Table recap in natmiOut : on 1 sheet by
  Day_typeCell concat info
  appeared/disappeared/up/down

natmiOut/ .
- Diff_LR_pop_up_down_regulation_total.xlsx

Shiny App Muscle and Age

AgeinMice_shinyApp:
→ copy of exam_INTER_conditions
/static/csv/shot_dataframe_finalfilter
+ symbolID in
DE_copy/shot_dataframe.csv

AgeinMice_shinyApp:
‑ copy of exam_INTER_conditions/
static/rds/fgsea_matrices4heatmaps
in Data

AgeinMice_shinyApp:
•Data/aggreg_matrices.rds
•copy of networks_explore/graphobjs/
age_Days_igraph_unfi.ml in
graphobj_copy

Explore network LR

Selected DAY

L-R pairs

DEG

Pathways

→**Load**: ChordDiagrame + Table of count/weight/Ratio
Ligand from typeCellA to Receptor typeCellB
→**Table Young/Old**: Print interactive table lr.frame Young
and Old - filtre possible on column by name + order
→**Cross**: do crossing -> print gene_cellType
ExclusiveYoung/Old and shared

‑**DEGs**: Print interactive table shot_dataframe
→**Intersect DE with loaded LR table**: Print
gene in DEG and LR table, Up regulated in
Young/Old DEG but no in table LR

Selected one or list of gene_cellType in Young or Old condition & number neighbour

Network

Animated:
→ Nodes label = symbol
→ Node value = specifity*10
→ Node color = TypeCell
‑ Edges value = weight *10
‑ Edges color = Ligand TypeCell

Gene specificity to a type cell

Calculate Tau Specificity

Index de spécificité Tau (τ) calculate for each gene at a condition chosen age +Day postInjury :

$$\tau_{gene} = \frac{\sum_{i=1}^{n(nbTypeCell)} \left(1 - \frac{meanTPM_{igene}}{\max\limits_{1<i<n}(meanTPM_{igene})}\right)}{n - 1}$$

Table resultat:

GeneID	GeneSymbol	τ	Classification	whichMAX	nbMAX	maxlog10TPM	AgeDay

Gene classification:
- Tau >= 0.8 ~ specific
- 0.5 <=Tau < 0.8 ~ intermediate,
- Tau < 0.5 ~ housekeeping

Keep the cell type the more expressed

If equality, keep the 2 types

Gene specificity to a type cell
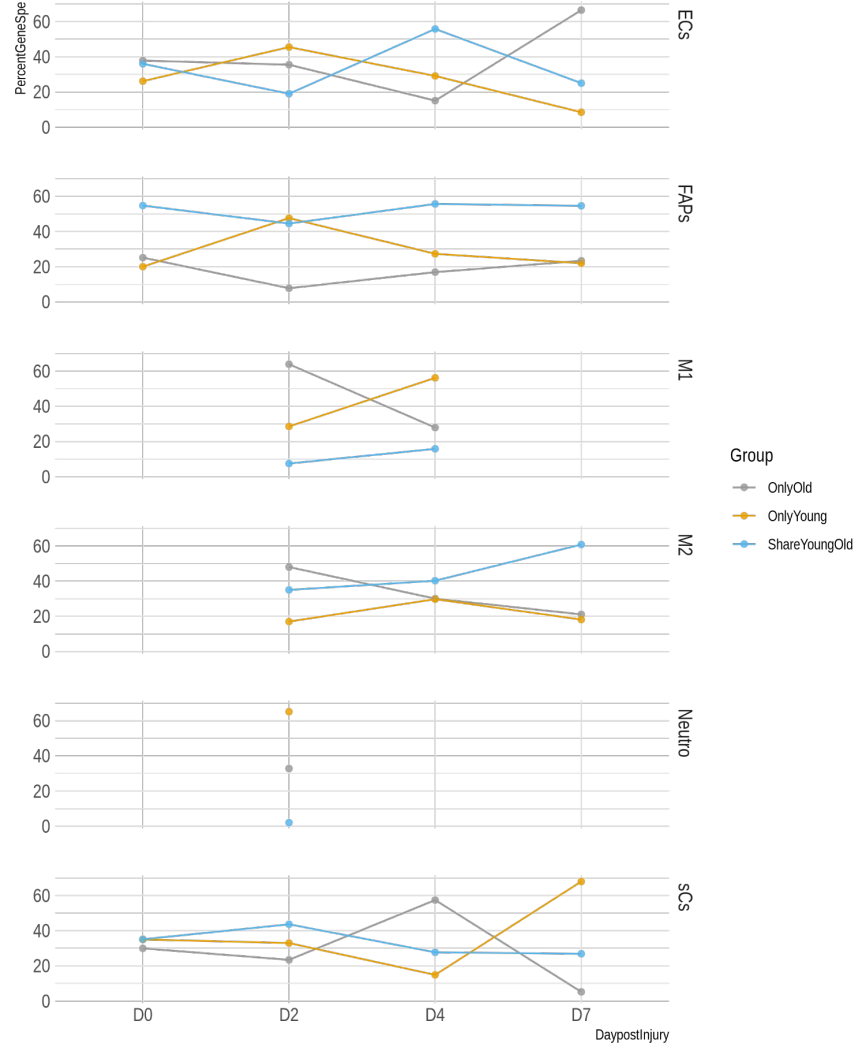
Without DEG analyse , the gene specificity dynamique change in young and old condition

Gene specificity to a type cell
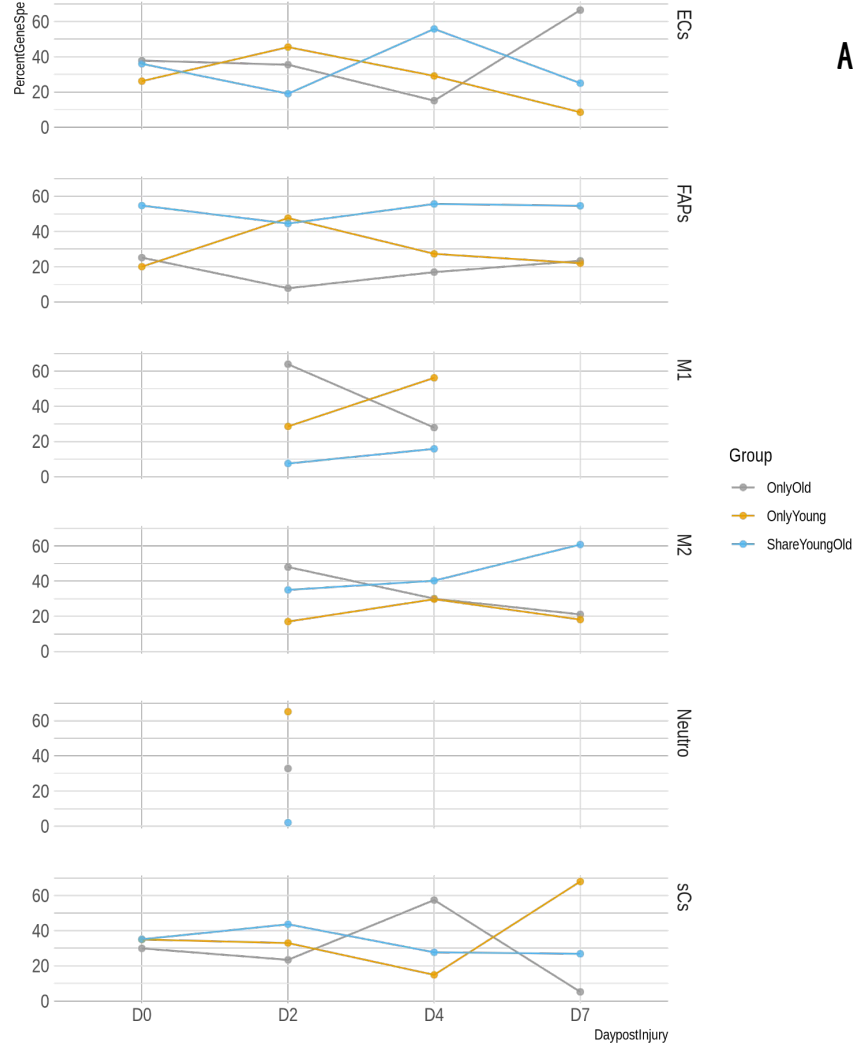
Are the specific genes in young condition the same as in old condition?

Without DEG analyse , the gene specificity dynamique change in young and old condition

Gene specificity to a type cell

Mac-M1

D2

OLD — YOUNG

291 (64%) | 34 (7.47%) | 130 (28.6%)

D4

447 (56.2%) | 126 (15.8%) | 222 (27.9%)

YOUNG — OLD

MuSCs

D4

OLD — YOUNG

896 (57.4%) | 432 (27.7%) | 232 (14.9%)

D7

1284 (68%) | 507 (26.8%) | 98 (5.19%)

YOUNG — OLD

Group
- OnlyOld
- OnlyYoung
- ShareYoungOld
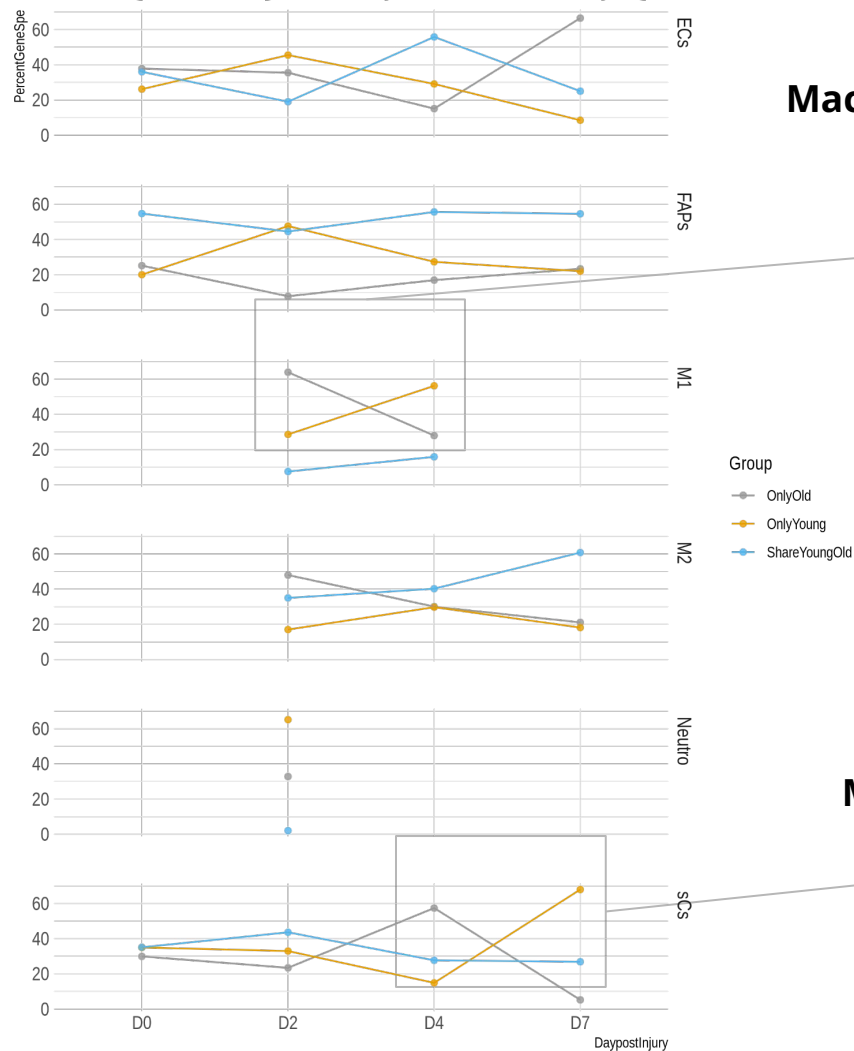
➔ What is maybe interesting ?
- Genes where the specificity changes between Y-O
- Genes where the
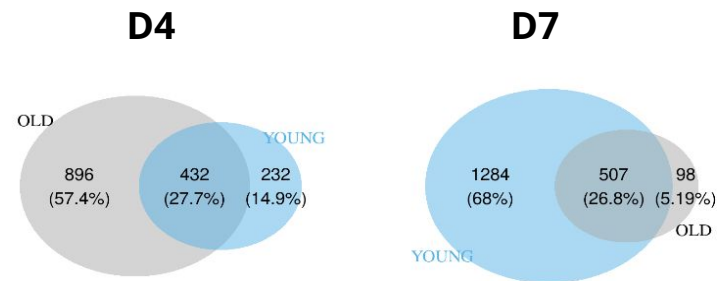
I. Differential Expressed Genes (DEG)

- On gene with on less 3 samples with raw count >= 5 (2 460 gene remove)

- DEG Old-Young for each 18 conditions TypeCell.DayPostInjury

- 25% genes are padj<0.05 on less one condition

I. Differential Expressed Genes (DEG)
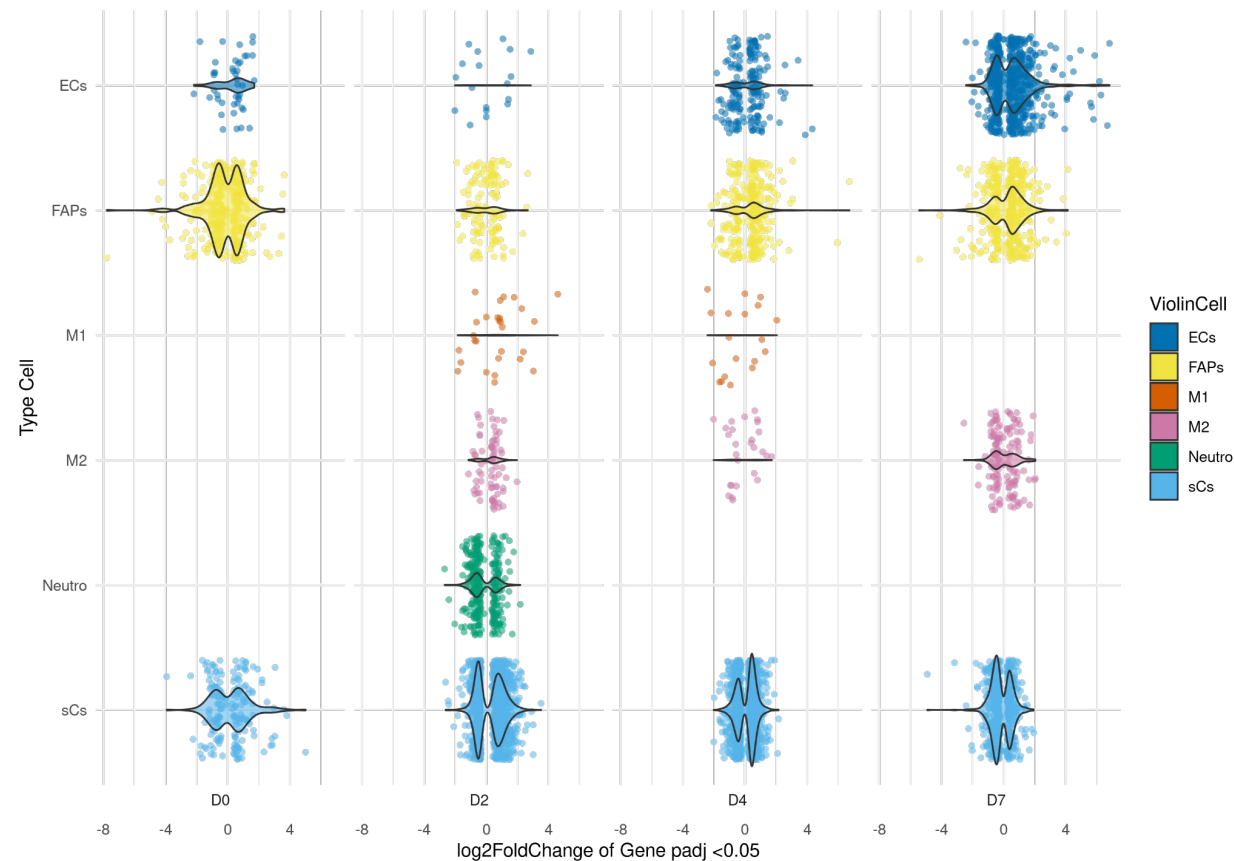
	ECs	FAPs	**sCs**	M1	M2	Neutro
D0	1.6	8.7	8.1	NA	NA	NA
D2	1.0	4.0	32.3	0.5	1.6	5.3
D4	4.7	6.6	**23.4**	0.3	0.6	NA
D7	10.4	7.0	10.1	NA	2.4	NA

Number Genes DE Young-Old on a TypeCell only on one Day
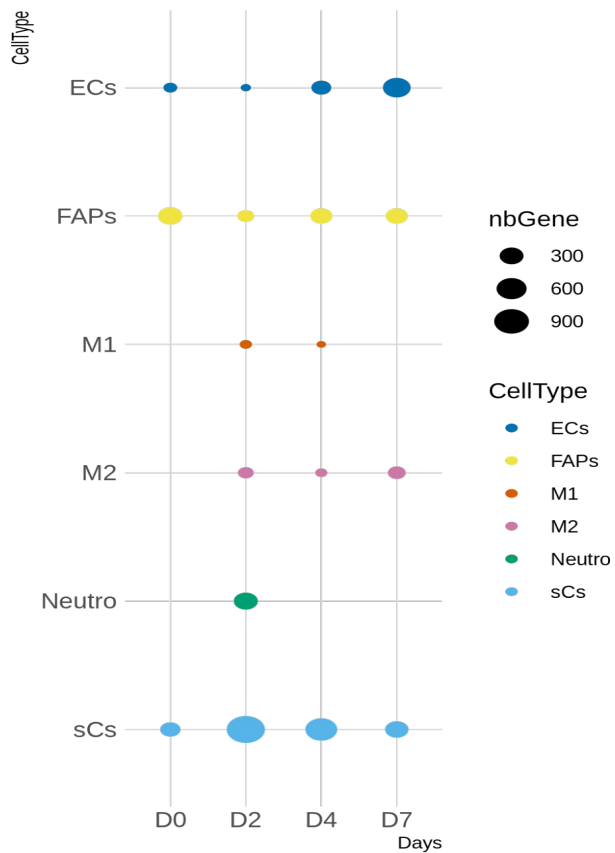
Number Genes DE Young-Old on a TypeCell only on more of one Day
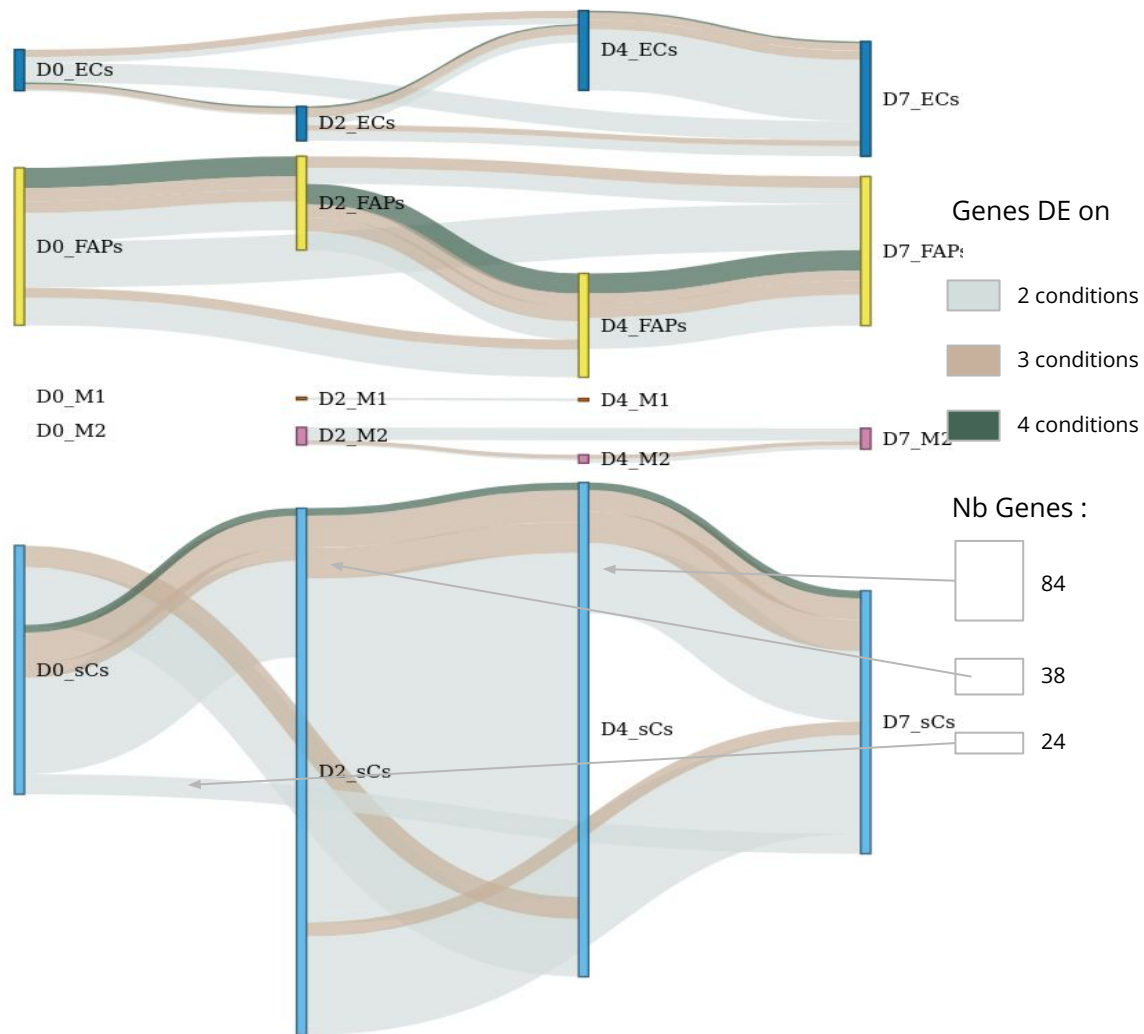
I. Differential Expressed Genes (DEG)
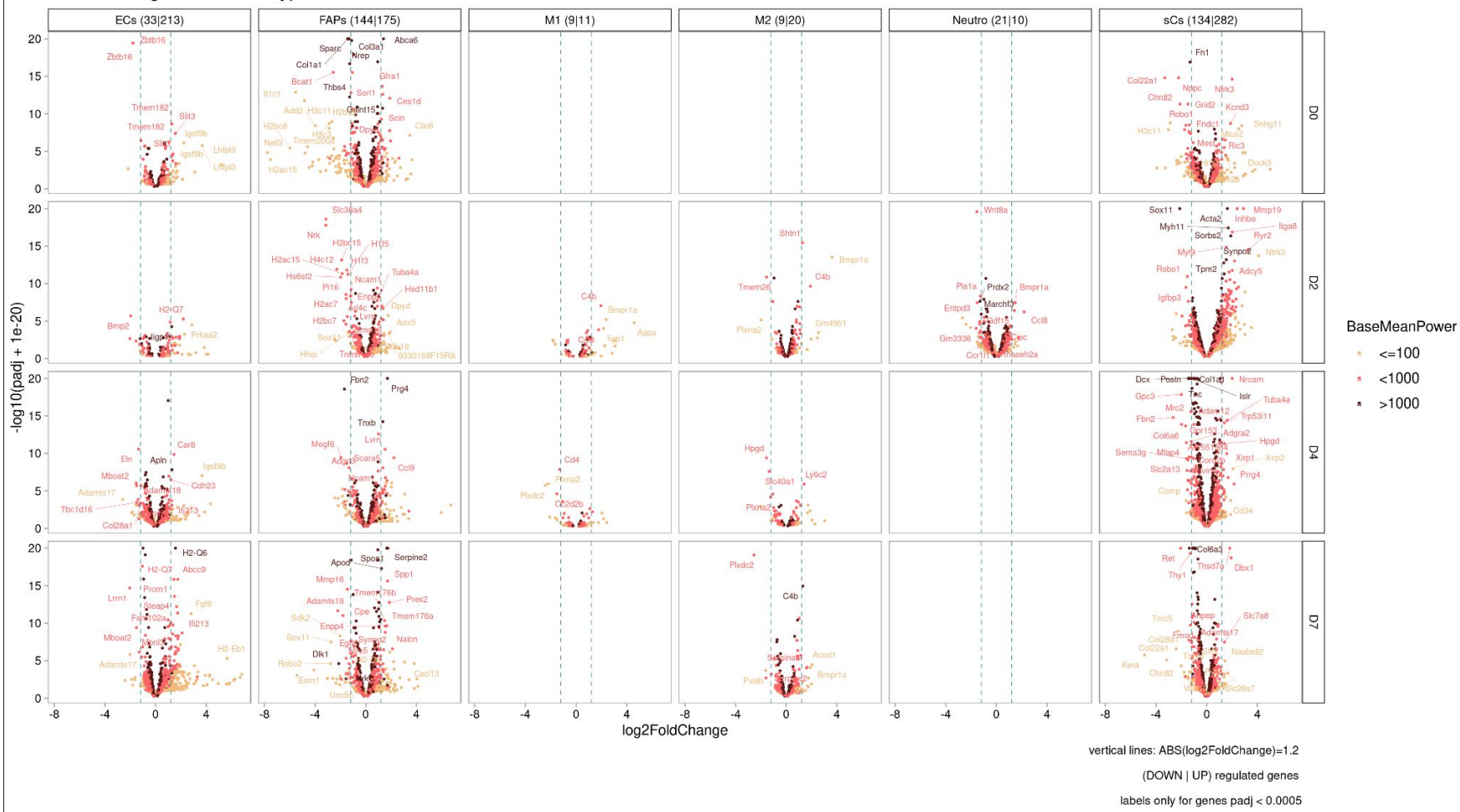
II. Genes Set Enrichment Analysis (GSEA)

➔ Algorithme : fGEA use in R

➔ Database:

◆ Genes Set from Molecular Signatures Database (MSigDB) species :

◆ species = Mus musculus

◆ category C2 = online pathway databases and the biomedical literature

◆ subcategory = Canonical pathways ,Reactome database

1. Before

● On each days and cellType

● On top (min padj) 500 genes with a regulation Up and 500 Down

● Sort by log2FoldChange

2. Now

● On each days and cellType

● On "all" genes

● Sort by signed of log2FoldChange * - log10(padj)

● Approx 3 000 genes with padj = NA

II. Genes Set Enrichment Analysis (GSEA)

GSEA Old vs Young (REACTOME)
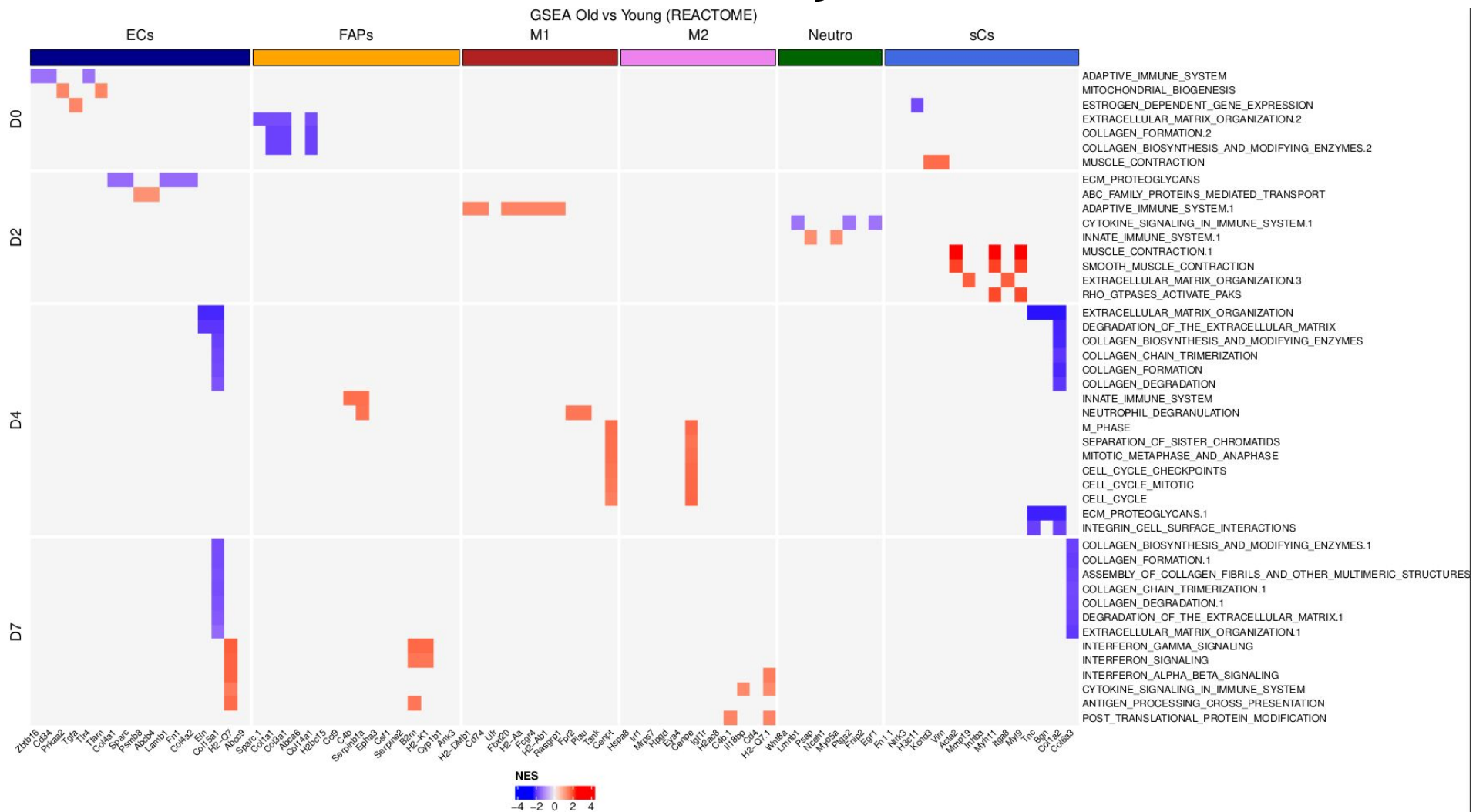
II. Genes Set Enrichment Analysis (GSEA)

file:///home/bioinfo/BulkAnalysis_plusNetwork/reportHTML/fgseaReport.html

NATMI Network Analysis Toolkit for Multicellular Interaction

Base on expression table (TPM) and database Ligand Receptor *ConnectomeDB2020 :*
- Calc specificity of Ligand and Receptor
- Edge weight = specificity Ligand * specificity Receptor

Représentation des réseaux: des graphes orientés pondérés

μ=14 « composantes connexes »
(taille <= 12)
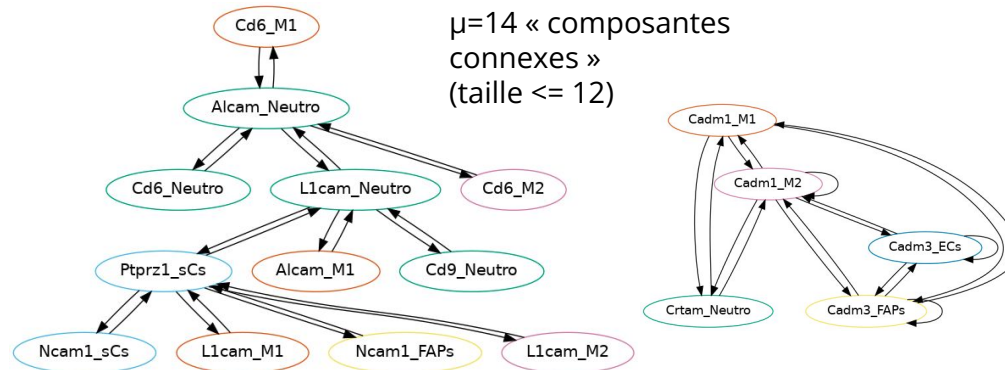
Bilan Static Analysis

Each gene has the following information for each DaysPostInjury and TypeCell:

- Expression intensity for Old and Young:
    - Raw count
    - TPM (mean replicat TPM) (Normalization by length gene and depth)
- DEG Young vs Old:
    - False Discovery Rate (FDR) with -> padj
    - Intensity and sense of DEG with -> log2FoldChange
- Pathways which it is imply :
    - False Discovery Rate (FDR) with -> padj
    - ((Intensity and sense of regulation of pathway with -> log2FoldChange))
- Sensibility at a type of cell
    - Tau
- Implication on a interaction Receptor Ligands

Objectives

- Created a central table with majors informations and the keys access to all specifics info locate in small table
  - For each genes DEG on less one

◆ Number of conditions DEG (padj<0.05)   ◆ Days   ◆ Type Cell   ◆ Code_illustrating trends ex:FAPs_1011

◆ Number of Gene Set Enrichment (padj<0.05)   ◆ Days   ◆ Type Cell   ◆ Code_illustrating trends ex:FAPs_Up0DownUp

◆ Number of Interaction Receptor Ligand   ◆ Days   ◆ Type Cell   ◆ NumberOnlyYoung/Old/Share ◆ Code trends O&Y

◆ Number Specific & intermediate   ◆ Days   ◆ Type Cell   ◆ NumberOnlyYoung/Old/Share ◆ Code trends O&Y

- Illustrating trends and try to cross trends

- Exploration of dataset with shiny app
  - Central table with filters -> allow to select a gene
  - Visualisation of his network for Old and Young conditions
  - Find a way to remove the 2 networks to highlight the differences

Objectives

Shiny Application
to explore datas