## Global Overview

In the following lines there are a brief description of what has been done so far.

**RNAseq data**

1. For each pair cell-condition, we kept only those genes that has row count greater than 100 counts in at least the 40% of the timePoints.
2. Normalization with upper quantile from EDAseq

**TimeClip**

Polynomial function 2° degree to catch time variations.

Once we get all the best path for each pathway we kept only those paths greater than 2 times the mean (preliminary).

Clique by clique approach.

Networks of the results.

**Membrane protein database (cell surface protein)**

Download the database from “A mass spectrometric-Derived cell surface protein atlas” and identification of the expressed and time dependent protein surface.

(**preliminary**)

**Secreted Protein**

Compile a list of secreted proteins using go cellular component and literature

**WIP**

**Ligand – receptor interactions**

1. Download and compile the human ligand to receptor database from “A draft network of ligand–receptor-mediated multicellular signalling in human”
2. Build the orthologous database
3. Developed a method to connect this piece of information:
   1. Gene by gene comparison for significant genes
   2. timeClip derived signals
   3. expression driven signals
4. analysis with timeClip
   1. cell-ibrid approach: once the ligand receptor pairs have been identified we isolate the pathway (if any) and analyze the upstream portion (the ligand dise) with one cell type expression profile and the downstream portion (receptor side) with another cell type (**WIP**)

**Analyzing specific pathways.**

ECM remodeling

Cell proliferation

**WIP**

### RNAseq data.

**New dataset with some replicated time points.**

**Expression data normalization.**

We inspect the row counts of the different time points to select the most similar runs. We averaged the similar distributions.

From the raw-count matrix, we filtered out all those genes (rows) that has less than **100 reads** in 60% of the time-point. This is needed to ensure a strong signals (reliable) across all time points.

The obtained matrix was normalized using upper quantile normalization (EDASeq R Package).

**Focus on RNAseq runs.**

For time point with replicates, as mentioned above we kept and averaged those timePoints replicates that has similar distribution. In the following figure the boxplots. All the green boxes were retained and when needed averaged.

We kept all the timePoints that have no replicates even if their distribution is not similar to the other runs.



Replicate ‘2days a’ has been excluded.



Overall good quality.



Overall good quality.



The ‘4days’ run distribution is globally quite different.



Overall good!



Overall good!

After upper quantile normalization, we perform a cluster analysis (Euclidean distance, clustering method average) on the normalize reads count.



This cluster analysis on FAP wt confirmed what we expected: 0 days (or undamaged) is quite apart from other days. There is only a big temporal flow: days 3 and 7 seems to be quite similar.



This clusters highlight some potential problems: days 1 and 10 are separated from the other structure that seems quite good.



This dendrogram is quite difficult to explain: 7 days timePoint is completely different dorm the others days. Moreover, there are two big branches that are quite separated and in one the time flow is hard to follow.



Also this dendrogram is quite difficult to explain with 7days and 0 days that are separated from the rest of the dendrogram. In this case the time flow is more accurate.



In this cell time days 2 is quite apart from the others.­­­



This dendrogramm shows that the distributions reflect the time flow.

### Single gene analysis

**FAP**

FAP\_wt dataset has 10 timepoints (undamaged, 1, 2, 3, 4, 5, 6, 7, 10, 14 days post damage). In the dataset we observed the expression of the main FAP markers Pdgfra and Ly6a (Sca-1).

Pdgfra has an expression peak at day 1 that last to day 2 and than its expression return to value similar to undamaged muscle.

Ly6a peak at days 1, 3 and 7 but the increment magnitude is more modest compare to Pdgfra.

These data could indicate that there is an expansion in FAP population.

In FAP\_CCR2\_ko we observed almost the same profile for both these markers.

Elevated presence of TGF-B (produced by pro-regenerative macrophages) favors FAP survival and switch to matrix-producing cells by blocking TNF induced apoptosis.

We then analyze Casp3 expression. In FAP\_wt, we saw a slight increase of Casp3 expression that starting at day 1 last to day 5.

From day 6 there is a decrease of it’s expression. The situation is similar in FAP\_CCR2\_ko.

**EC**

**….**

**Ligand receptor analysis.**

We downloaded the human receptor database from “A draft network of ligand–receptor-mediated multicellular signaling in human”. Then we “translate” the human genes into mouse genes.

This approach has a major drawback that the annotation is poor (orthologes with relation one2many and many2many) so every strong candidate need to be literature checked before proceeding with further analysis.

We put into contact the different cell types using ligand receptors pairs as bridges. This create a map between the ligands of specific cell type and the receptor of his own specific cell type or other cell types.

This analysis can be enriched in many different ways and combinations:

* using timeClip results to filter;
* using single gene timeClip approach to filter;
* using expression based (correlation or expression level).

**WIP** analysis with timeClip

* cell-ibrid approach: once the ligand receptor pairs have been identified we isolate the pathway (if any) and analyze the upstream portion (the ligand side) with one cell type expression profile and the downstream portion (receptor side) with another cell type.

**Results.**

For each condition we compute the discretization of the expression. We scan the entire expression distribution. Genes with at least the 20% of the time points >80 quantile of expression distribution were marked as “high”; on the other end, if more than the 80% of the timePoints were under the 30quantile of the expression distribution we marked the gene as “low” expressed. The remaining genes were marked as “medium”.

To build the intercellular network of ligand receptor, we kept only those with at least a medium in one condition (wt or ko).

On the based of simple expression (low, medium, high), we identify a network of 4567 possible relationships amongst the 3 cell type. Furthermore, for every pairs (in all conditions) we compute Spearman correlation and to every gene we attach the pvalue of the single gene analysis.

Filtering for those couples with either the ligand or the receptor that shows time dependencies, we obtained 1158 relationships.

Summing up: these relations represent a relation from one ligand in a cell (ligand expression of class “medium” in the given cell) that points to its receptor (receptor expression of class “medium”) in the same or another cell. Since all possible pairs has been contemplated a ligand receptor pairs can be repeated in different cells. Either the ligand or the the receptor must have a temporal behavior. A further selection can be made filtering using spearman correlation.

Please see files:

ec-ligand2all-cell-types.xlsx

fap-ligand2all-cell-types.xlsx

mc-ligand2all-cell-types.xlsx

These files have all the same format. The columns are the following:

1 Ligand symbol

2 receptor symbol

3 cell – ligand -> the cell type where the ligand is expressed (medium class)

4 cell – receptor -> the cell type where the receptor is expressed (medium class)

5 singleGene Pvalue ligand -> the ligand pvalue of timeClip single gene in the given cell type

6 Single gene pvalue receptor -> the receptor pvalue of timeClip single gene in the given cell type

7 Spearman correlation ligand receptor: in the respective cell types

8 relation id: the id of the edge

9 is Unique across cell types: Unique when the relation is exclusive of the cell type in the row; otherwise multi.

10 intercellular sharing: the other cell-cell pairs where the relation is present.

**Old Analyses**

**TimeClip Analysis**

We used timeCLip Step 2 approach on all keg and reactome pathways but given the nature of the expression data we chose to apply also a clique by clique analysis and then compare the cliques across the 6 condition (FAP\_wt, EC\_wt, FAP\_CCR2\_ko, EC\_CCR2\_ko, MC\_wt, MC\_ko\_CCR2).

**FAP Pathway Analysis.**

We applied only step2 of clipper analysis. We filter out the results with the relevance score below 2x median of the relevance distribution. This way we kept only the stronger signals. Then we build the pathway/network of the results: the paths are the node and we put an edge between two nodes if the paths share at least one edge (in the genes pathways).

We found that the paths Neurotropin / RAS signaling have high number of connection: 12.

This suggest us that those are the pathways that convey the signal.

**EC Pathway Analysis**

….

**WARNING: From now on there are some of my free annotation to help me keep track of how the work is going. Please Do not rely full on this annotation that could be incomplete.**

**Interesting Genes**

**Mainly FAP - side**

Inpp1: anti hypertrophy

Tfdp1: transcription factor inhibited by TGFB

Plcg1: cell aggregation

Ppp3cb: Vascularization

Slc2a1: Heart failure Hypertrophy

Acat: Acetil Coenzima Liver fibrosis is promoted by

Dvl2: Myoblast differentiation related

Agtr1a: receptor for Angiotensin II

*Surface Markers*

Angpt1 (F): stimulates HIF to recover ischemic injury, enhance proliferation of skeletal myoblast. Potentiate Orai1 (only EC) and calmodulin (Calm1) massively expressed in EC (Ca2+ signaling).

Agtr1a (F): angiotensin II receptor

Cacna1c (F and moderately EC-wt): calcium channel proliferation

Epha3 (F) maintenance of slow fiber identity

Ephb2 (F) cell contact / repulsion and involved in platelet control

Ephb3 (F) axonal plasticity and regrowth

Ncam1 (F expecially KO) cell motility and proliferation

Pdgfra (F) smooth muscle and vascularization

Ppap2a (Plpp1, common to all) lipid phosphate phosphatase

Ppap2b (Plpp3, common to all) phospholipid phosphatase (vascular related)

Slc30a1 (common to all): development zinc transporter

Slc39a14, Scl39a6 (common to all):

Slc8a1 Vascularization related (FAP, EC wt)

Tnfrsf10b (common to all, EC very expressed)

**FAP KO**

Dcn: Decorin Collgaen formation and TGFB dependent -> modulate IGF-IR

Acvr1: activin A receptor

**Mainly EC - side**

Cdc42: stimulated by VEGF through MAPK11-14 promote migration

Amotl2 receptors?

Kit/ Kitl promote arteriogenesis in myocardial infraction

Nos3: NO signling for VEGF

*Surface Markers*

Epha2 (E): Eph receptor a -> Eph/ephrin molecules – hub for signaling in endocytosis

Bdkrb2 (E): Hypothesis of antifibrosing-effects

Met (E) <- Hepatocyte Growth Factor (Hgf if expressed by FAP)

Il2rg (E)

Atp2a3 (E) (Serca 3) Sarco/Endoplasmic reticulum

**EC and FAP related**

Gnas: Adipose tissues metabolism / Progenitors controlling

Edn1: Endothelin

Ednra: Endothelin receptor

Orai1: calcium influx in epithelial cells

Mllt4: Cell to cell junction

Ywhab: prevent growth and cell adhesion

Rac1 – Vav1: Endocytosis

Pld1: Endocytosis

*Surface Markers*

Adcy6: adenilato Cyclase 6

Erbb2: triggers cardiac myocites regeneration by cardiomyocites dedifferentiation and proliferation

Fas: TNF receptor superfamiliy death cell receptor

Fgfr1: Fibroblast growth factor receptor (cell motility - cancer)

Igf1r: insulin like growth factor

Itga7: Muscle related

Itgav: vascular morphogenesis

Itgb1 (massively expressed): Angiogenesis

Itgb3 (moderately): impair endothelial regeneration

Itgb5 (moderately): vascular cell permeability

Pdgfrb: involved in fibrosis

Ptdss2: phosphatidil serine

Tgfb1:

Tnc: massively expressed in FAP (expecially FAP KO) knock put mice has inhibition of fibrous adhesion formation

Vtn (wt 5-6 giorni max; ko seems flat): vitronectin connected with integrin sig pathway, apoptotic cells and hedgehog and angiogenesis

**WT Secific**

Adcy9 (WT): adenilato cyclase 6

Calcrl: (downregulated early in WT): microvascular envolved and skeletal muscle related

Epha4: (main EC): cohoperate with fgf; regulates cell renewal and differentiation.

Gabbr1: (– follow temporal pattern) homeostasis in hippocampal networks

S1pr1 (massively expressed in EC): G coupled receptor involved in vascular and heart development.

S1pr3 (mainly FAP): angiogenesis; it’s knowout attenuate inflammation and fibrosis.

Vcan (massively in FAP) : wound repair and fibroblast migration.

**KO specific**

Cant1

Hsp90ab1

Hsp90b1

Hspg2

Ifnar1

Ifnar2

Il17rd

Il1r1

Il1rap

Itga1

Lama2

Lama4

Lamb2

Lamc1

Lifr

Lum

Myl12b

Myo9a

Nrp1

Plxna4

Polr1a

Stim1

Thbs1

Timp1

Vcam1

**EC to FAP “kiss”**

Gap Junction ???

Are they exchanging Ca2+ or cAMP?

**Second messangers in Endothelial Cells.**

AcetylCholine

NO

Arginine