MaRA – an Automatic Microarray R-based Analysis Pipeline
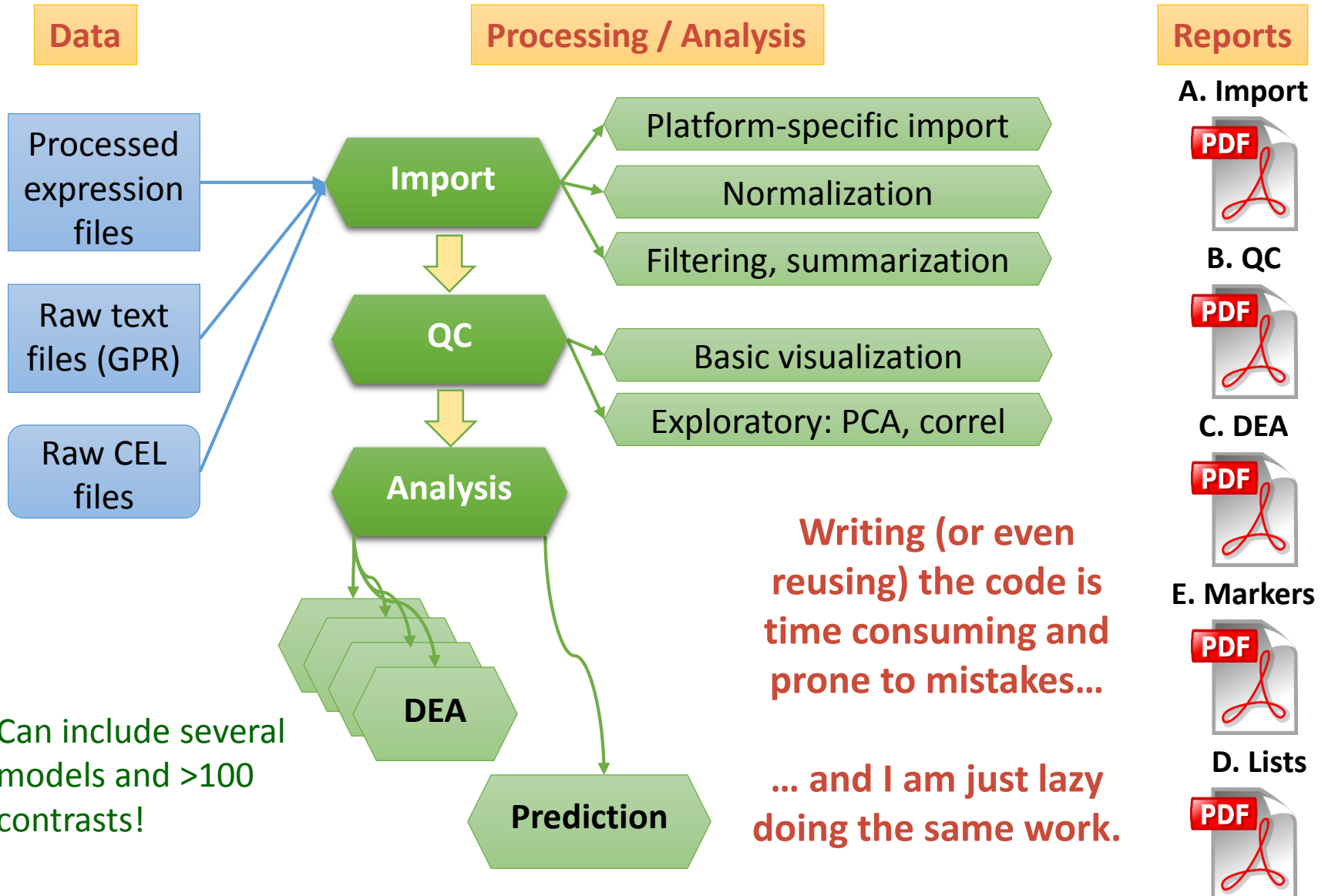
Petr Nazarov

LabMeeting, the 17th of January 2017

MaRA = a "dream" in Belarusian

- **Pipeline** in microarray analysis

- **Features** of MaRA
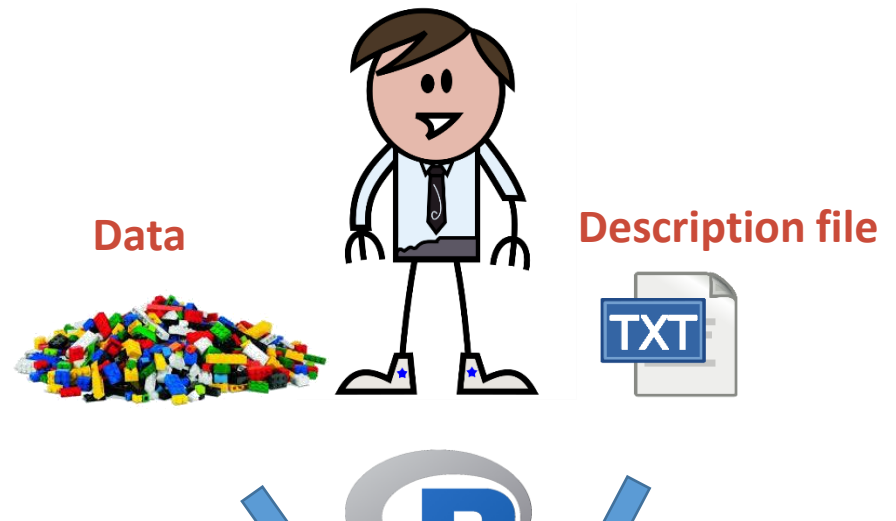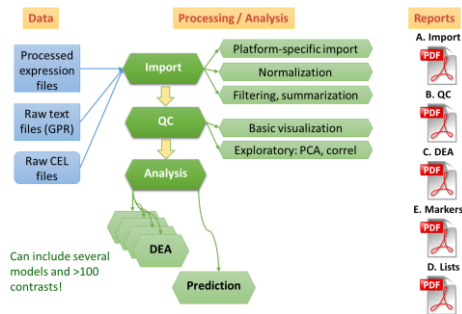
- **Example** – HepMirSTAT project (prof. Iris Behrman)

MaRA (**M**icro**a**rray **R**-based **A**nalysis) is an advanced flexible pipeline for automated analysis of microarray data and reporting the results.

Features
- R-based pipeline
- Scalable & flexible analysis
- Automatic processing and reporting:

Data

Description file

Reports & results

Drawbacks
- No GUI (is in development)

Development
- R / Bioconductor
- Qt C++ (GUI in development)

Images originate and belong to LEGO® ☺

Example of a Description File: Standard INI Configuration

```ini
[Project]
Title      = Project Pit Ullmann
Names      = P.Ullmann, E.Letellier, S.Haan
Platform = Affymetrix HuGene 2.0 ST
Description = Effect of treatments on 2 cell lines
```

```ini
[Analysis]

Model = Cells + Treatment + Cells*Treatment

CL18_clust.vs.ctrl    = CL18,Cluster  - CL18,Ctrl
CL620_clust.vs.ctrl  = CL620,Cluster - CL620,Ctrl
Ctrl_18.vs.620       = CL18,Ctrl - CL620,Ctrl
Cluster_18.vs.620 = CL18,Cluster  - CL620,Cluster
diffCL_Cluster.vs.Ctrl = (CL18,Cluster - CL18,Ctrl) - (CL620,Cluster - CL620,Ctrl)
```

```ini
ExpressionThreshold = 5
```

```ini
[Colors]
red      = CL18,Ctrl
orange  = CL18,Cluster
blue    = CL620,Ctrl
cyan    = CL620,Cluster
```

Example of a Description File: Standard INI Configuration

SDE genes found in contrasts can be intersected/united/excluded: &, |, !, ()

```
[Lists]
FDR = 0.01
FC = 1
list_Colon_cancer.v.healthy = Colon_HT29.vs.NCM460 & Colon_HCT116.vs.NCM460
list_Skin_cancer.v.healthy  = Skin_MelJuso.vs.NHEM & Skin_A375.vs.NHEM
list_Liver_cancer.v.healthy  = Liver_Hep3B.vs.PH5CH8 & Liver_Huh7.vs.PH5CH8
list_B1_cancer.v.healthy = Colon_HT29.vs.NCM460 & Colon_HCT116.vs.NCM460 & S
```

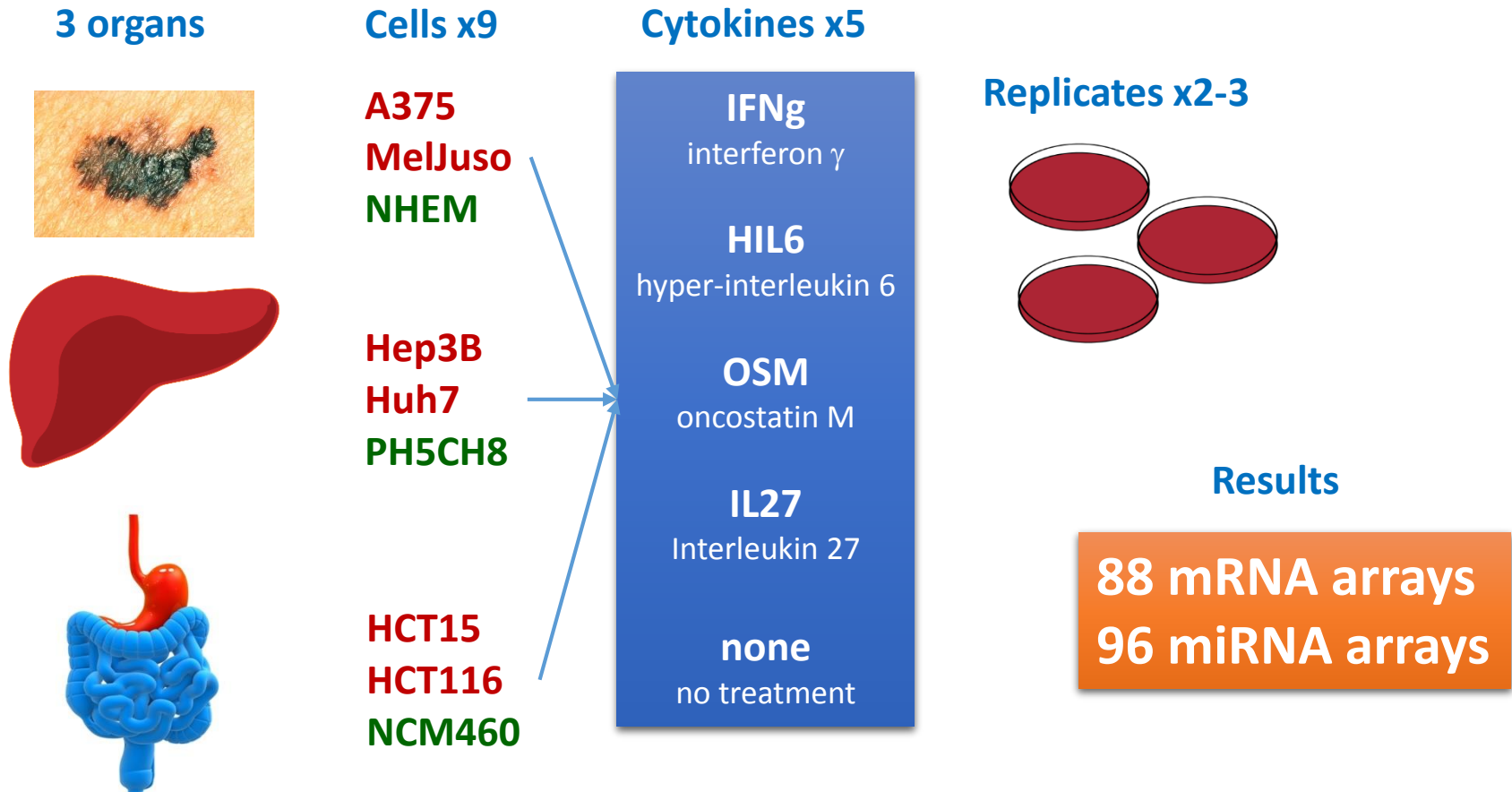## Additional analysis – looking for markers in groups of samples

```
[Markers]
AUC = 0.99
mark_Colon  = *,*,*,colon  - (*,*,*,skin + *,*,*,liver)
mark_Skin   = *,*,*,skin   - (*,*,*,colon + *,*,*,liver)
mark_Liver  = *,*,*,liver  - (*,*,*,skin + *,*,*,colon)
mark_Liver.vs.Skin  = *,*,*,liver  - *,*,*,skin
mark_Liver.vs.Colon  = *,*,*,liver  - *,*,*,colon
mark_Skin.vs.Colon  = *,*,*,skin   - *,*,*,colon
mark_Cancer = *,*,cancer,* - *,*,normal,*
```

Example: HepMirSTAT

Project of University: cytokines in CL at mRNA & miR level

Investigation of signaling after cytokine stimulation in cell lines originated from 3 organs – skin, liver, colon.

PI - prof. **Iris BEHRMANN**, UniLu

3 organs

Cells x9

A375
MelJuso
NHEM

Hep3B
Huh7
PH5CH8

HCT15
HCT116
NCM460

Cytokines x5

IFNg
interferon γ

HIL6
hyper-interleukin 6

OSM
oncostatin M

IL27
Interleukin 27

none
no treatment

Replicates x2-3

Results

88 mRNA arrays
96 miRNA arrays

x2 as miRNA and mRNA datasets should be investigated!

What is an effect of **each** cytokine in **each** cell line?

Can we see tissue-specific response to cytokines?

Do we have marker genes for cancer / healthy cells ?

Do we have specific genes for **each** cytokine?

Any common tendency between cytokines signaling?

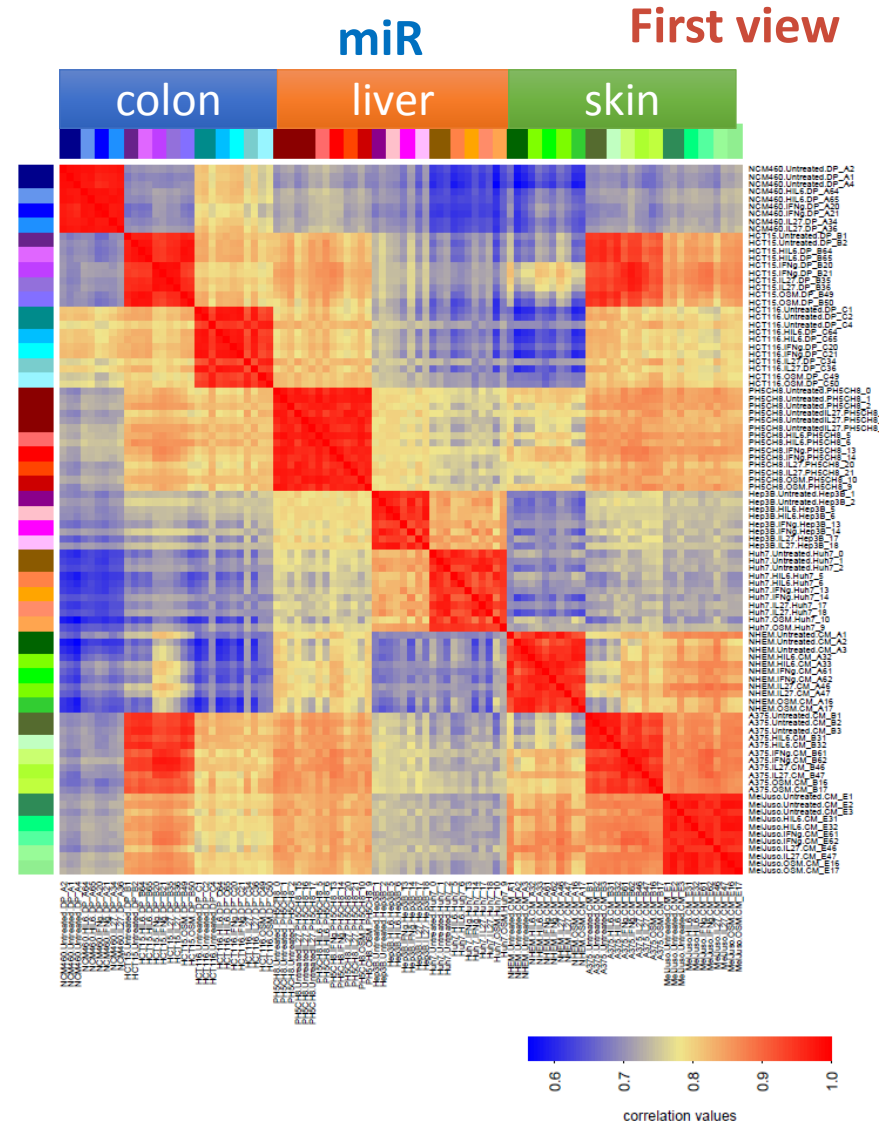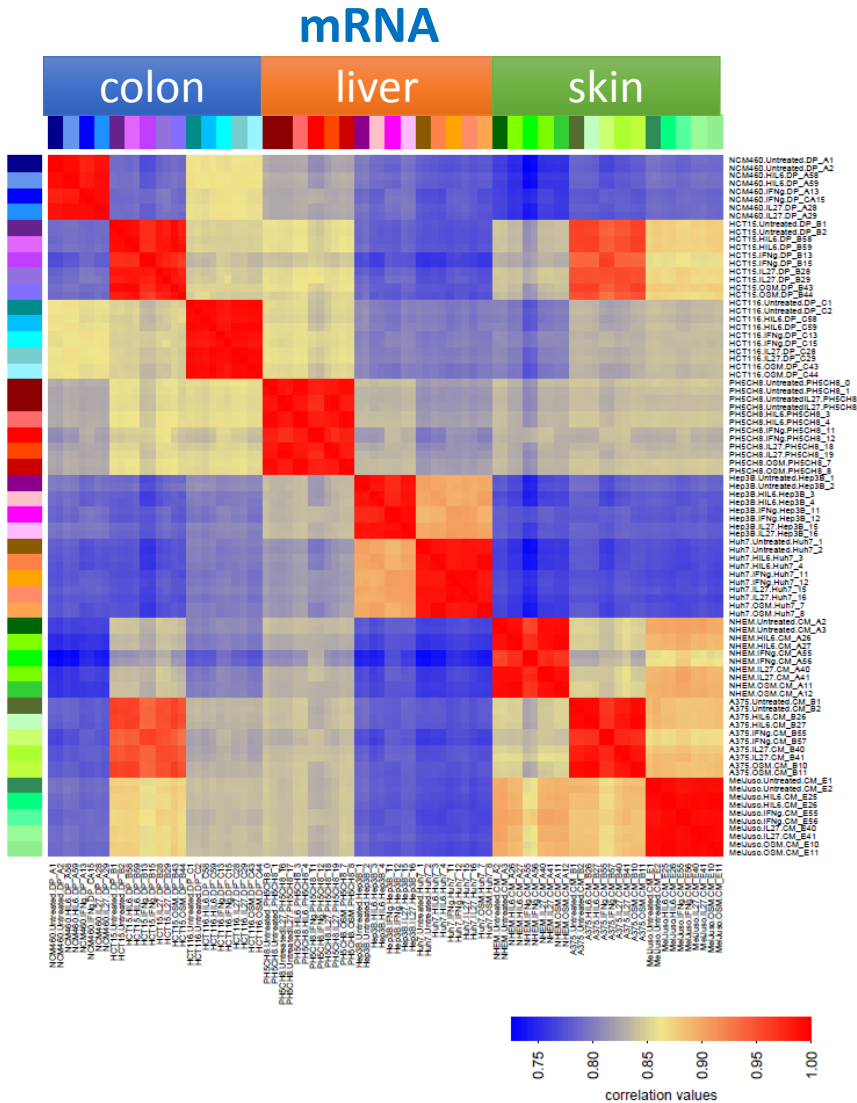Can we see cancer-specific response to cytokines?

~ 100 samples

92 comparisons

TXT

14 lists of markers

miRNA vs mRNA vs TargetScan

31 intersection lists

5 reports ☺

8

mRNA   **miR**   **First view**

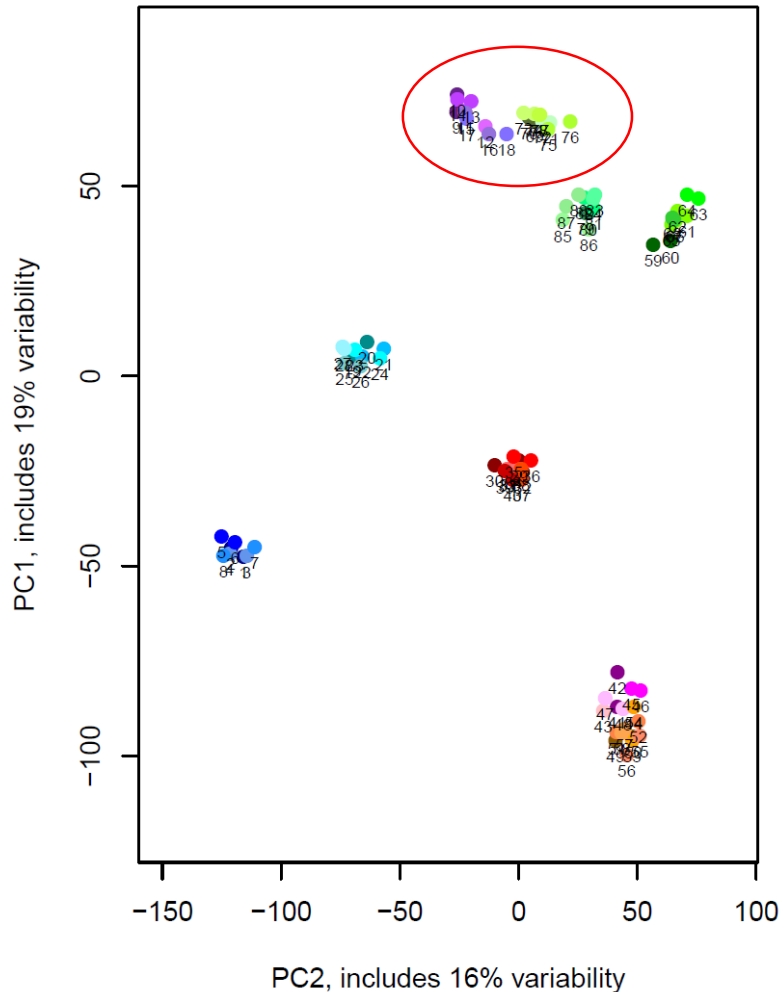Noise in miR data is higher. Cytokine effect is much smaller then cell type effect.

No organ grouping?   HCT15 is correlated to A375?

Why is HCT15 similar to A375?

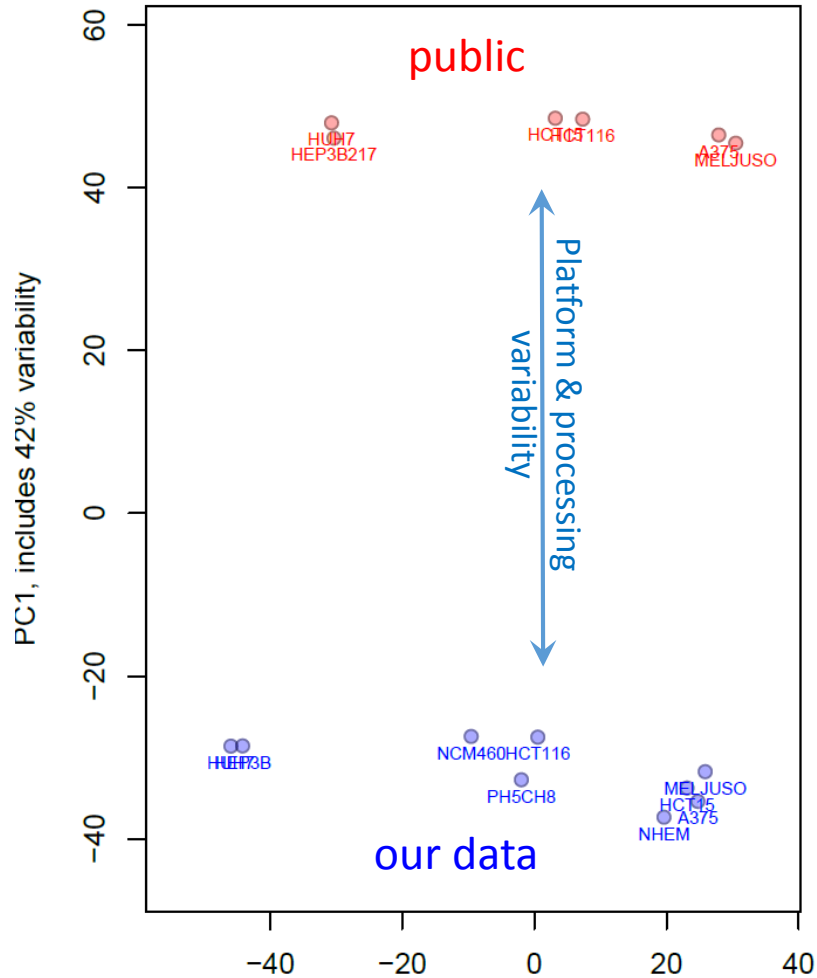**Principle component analysis (PCA) (35% variability)**

Legend:
- HCT15,HIL6
- HCT15,IFNg
- HCT15,IL27
- HCT15,OSM
- HCT15,Untreated
- HCT116,HIL6
- HCT116,IFNg
- HCT116,IL27
- HCT116,OSM
- HCT116,Untreated
- NCM460,HIL6
- NCM460,IFNg
- NCM460,IL27
- NCM460,Untreated
- Hep3B,HIL6
- Hep3B,IFNg
- Hep3B,IL27
- Hep3B,Untreated
- Huh7,HIL6
- Huh7,IFNg
- Huh7,IL27
- Huh7,OSM
- Huh7,Untreated
- PH5CH8,HIL6
- PH5CH8,IFNg
- PH5CH8,IL27
- PH5CH8,OSM
- PH5CH8,Untreated
- PH5CH8,UntreatedIL27
- A375,HIL6
- A375,IFNg
- A375,IL27
- A375,OSM
- A375,Untreated
- MelJuso,HIL6
- MelJuso,IFNg
- MelJuso,IL27
- MelJuso,OSM
- MelJuso,Untreated
- NHEM,HIL6
- NHEM,IFNg
- NHEM,IL27
- NHEM,OSM
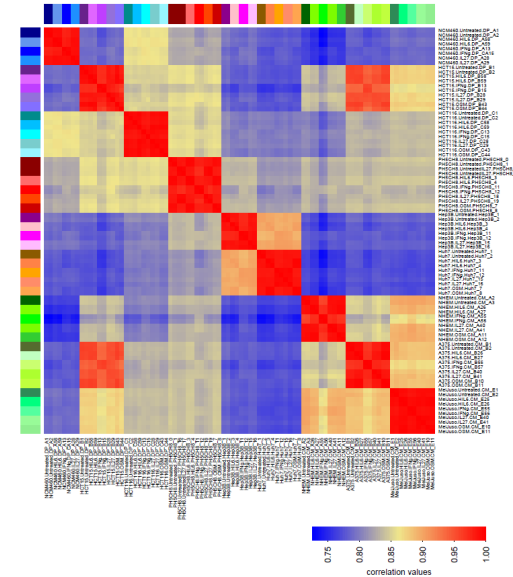
Let's have a deeper look with **Cancer Cell Line Encyclopedia (CCLE)** data

10

PCA 1–2
(60% variability)

public

Platform & processing variability

our data

PCA 2–3
(29% variability)

Our HCT15

Skin cluster

Colon cluster

HCT116
HCT15

Liver cluster

Public HCT15
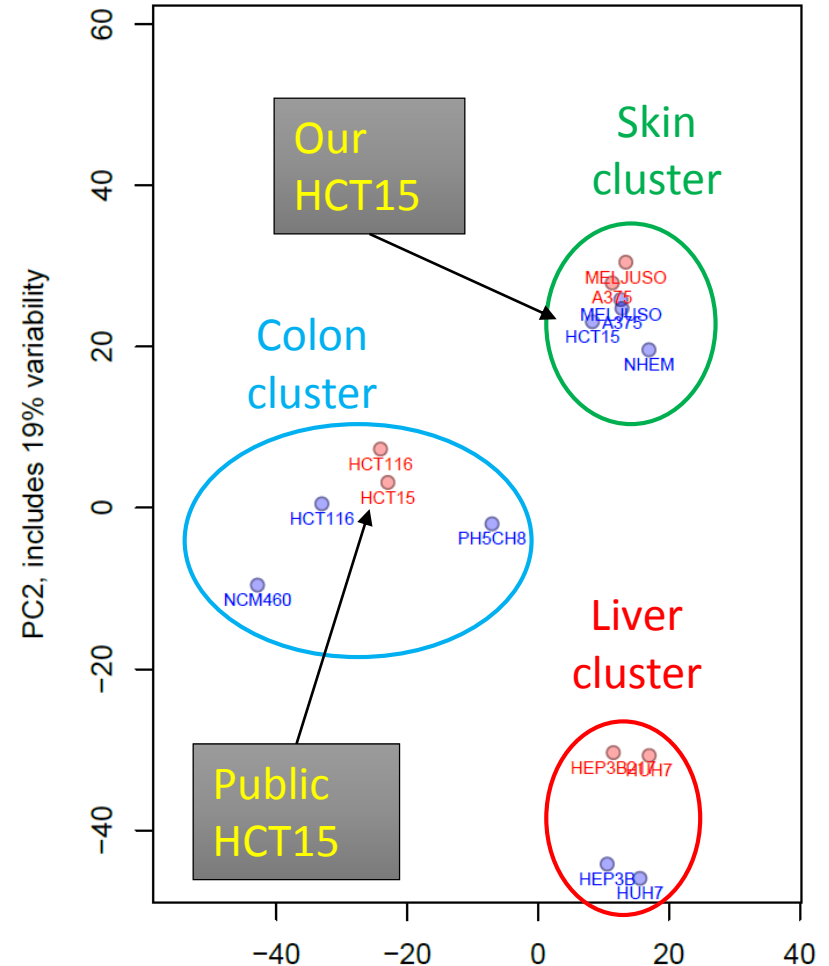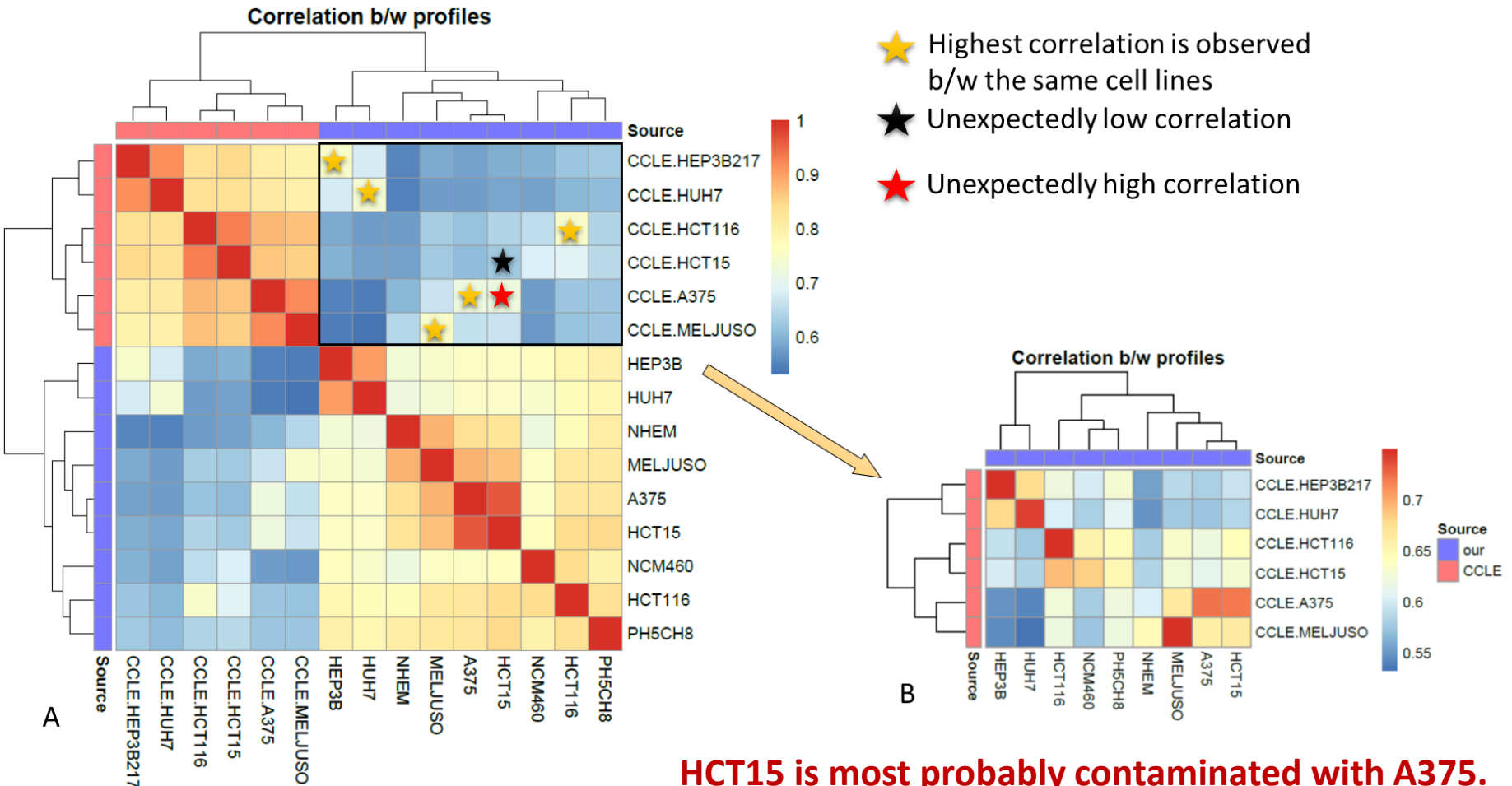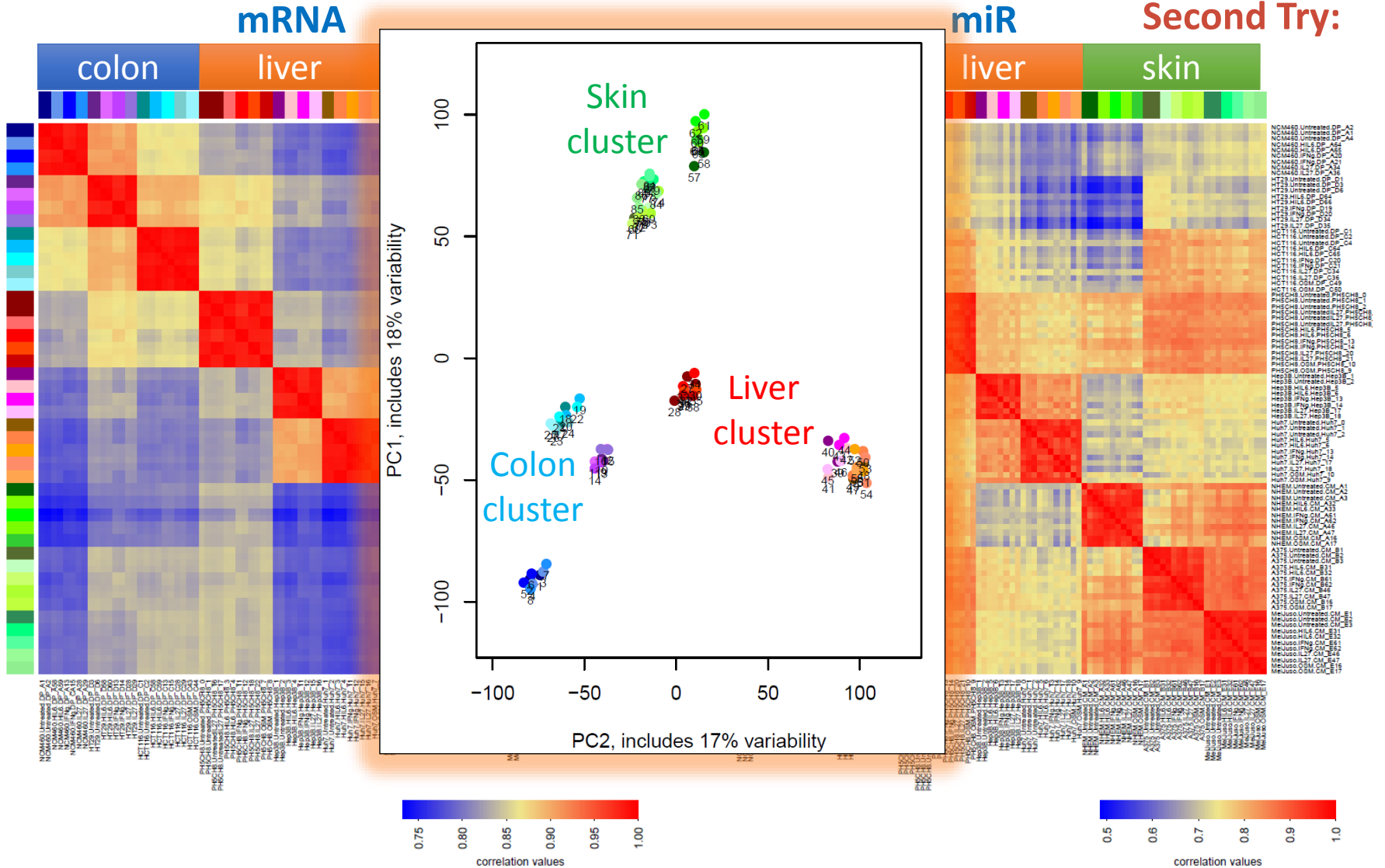
PC 1 captures b/w platform differences

PC 2-3 capture tissue differences

HCT15 is most probably contaminated by A375

★ Highest correlation is observed b/w the same cell lines
★ Unexpectedly low correlation
★ Unexpectedly high correlation

HCT15 is most probably contaminated with A375. We replaced it with HT29

HepMirSTAT

Second Try:

It is much better now!

LUXEMBOURG INSTITUTE OF HEALTH — RESEARCH DEDICATED TO LIFE

A375_IFNg = A375,IFNg − A375,Untreated

	|logFC|>0	|logFC|>0.5	|logFC|>1	|logFC|>2
FDR<0.05	8415	1826	324	52
FDR<0.01	6640	1813	324	52
FDR<0.001	5027	1768	323	52

Q–Q Plot

Volcano–plot.
324 with |logFC|>1 and FDR<0.01
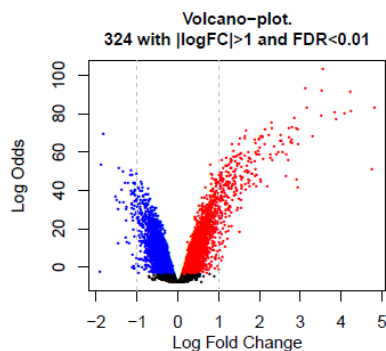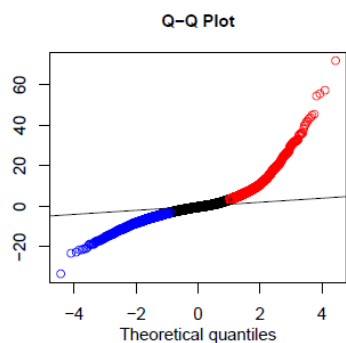
Top 25 probes. Look for the complete list in DE=A375_IFNg.txt

| logFC	AveExpr	adj.P.Val	Gene.Symbol	gene_assignment
4.8199	5.8536	0	GBP4	NM_052941 // GBP4 // guanylate binding protein 4 // 1p22.2 // 115361 //
4.2209	5.1924	0	CD274	NM_014143 // CD274 // CD274 molecule // 9p24 // 29126 /// ENST00000
3.547	5.809	0	IRF1	NM_002198 // IRF1 // interferon regulatory factor 1 // 5q31.1 // 3659 /// E
4.2401	6.9842	0	GBP1	NM_002053 // GBP1 // guanylate binding protein 1, interferon–inducible
4.0816	5.2839	0	IDO1	NM_002164 // IDO1 // indoleamine 2,3–dioxygenase 1 // 8p12–p11 // 36
3.524	4.8014	0	GBP5	NM_001134486 // GBP5 // guanylate binding protein 5 // 1p22.2 // 11536
3.8389	5.7626	0	GBP1P1	NR_003133 // GBP1P1 // guanylate binding protein 1, interferon–inducibl
3.8732	5.7346	0	XAF1	NM_017523 // XAF1 // XIAP associated factor 1 // 17p13.1 // 54739 /// N
3.1251	9.0768	0	STAT1	NM_007315 // STAT1 // signal transducer and activator of transcription 1
3.5092	6.5973	0	GBP3	NM_018284 // GBP3 // guanylate binding protein 3 // 1p22.2 // 2635 /// E
3.1604	4.6776	0	LGALS17A	NR_034156 // LGALS17A // Charcot–Leyden crystal protein pseudogen
4.7524	5.4888	0	CXCL10	NM_001565 // CXCL10 // chemokine (C–X–C motif) ligand 10 // 4q21 //
3.3013	5.6671	0	SAMD9L	NM_152703 // SAMD9L // sterile alpha motif domain containing 9–like //
2.8489	5.7948	0	CIITA	NM_000246 // CIITA // class II, major histocompatibility complex, transac
2.9338	7.1773	0	TAP1	NM_000593 // TAP1 // transporter 1, ATP–binding cassette, sub–family E
2.8952	6.464	0	ACSL5	NM_016234 // ACSL5 // acyl–CoA synthetase long–chain family membe
2.7947	4.6455	0	PDCD1LG2	NM_025239 // PDCD1LG2 // programmed cell death 1 ligand 2 // 9p24.2
2.7202	6.6156	0	A2M	NM_000014 // A2M // alpha–2–macroglobulin // 12p13.31 // 2 /// ENST0
2.9483	6.9096	0	GBP2	NM_004120 // GBP2 // guanylate binding protein 2, interferon–inducible
2.7911	6.6802	0	BST2	NM_004335 // BST2 // bone marrow stromal cell antigen 2 // 19p13.1 // (
2.718	5.7971	0	UBD	NM_006398 // UBD // ubiquitin D // 6p21.3 // 10537 /// ENST000003770
2.7975	5.7914	0	OAS2	NM_001032731 // OAS2 // 2'–5'–oligoadenylate synthetase 2, 69/71kDa
2.7613	6.4274	0	AREG	NM_001657 // AREG // amphiregulin // 4q13.3 // 374 /// ENST00000264
2.5133	5.9309	0	EPSTI1	NM_001002264 // EPSTI1 // epithelial stromal interaction 1 (breast) // 13
2.2894	5.7031	0	MAOB	NM_000898 // MAOB // monoamine oxidase B // Xp11.23 // 4129 /// ENS

A375_IFNg = A375,IFNg − A375,Untreated

standartized expression values

Lists generated
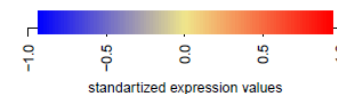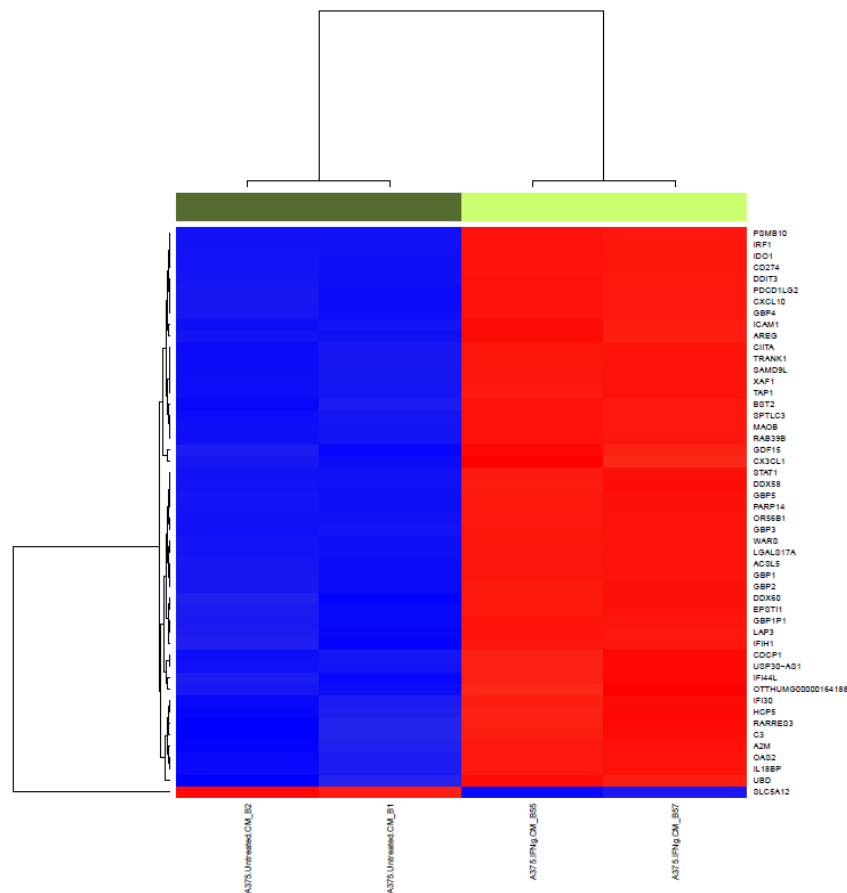
List: list_C2_IFNg

Formula: A & B & C & D & E & F & G & H & I

	List	Signif	Up	Down
A	A375_IFNg	324	263	61
B	HCT116_IFNg	100	98	2
C	HT29_IFNg	236	209	27
D	Hep3B_IFNg	221	191	30
E	Huh7_IFNg	142	138	4
F	MelJuso_IFNg	67	67	0
G	NCM460_IFNg	195	169	26
H	NHEM_IFNg	521	268	253
I	PH5CH8_IFNg	190	180	10
=>	list_C2_IFNg	19	19	0

Majority of the genes indeed comes from IFNg pathways.

15

DEA overview
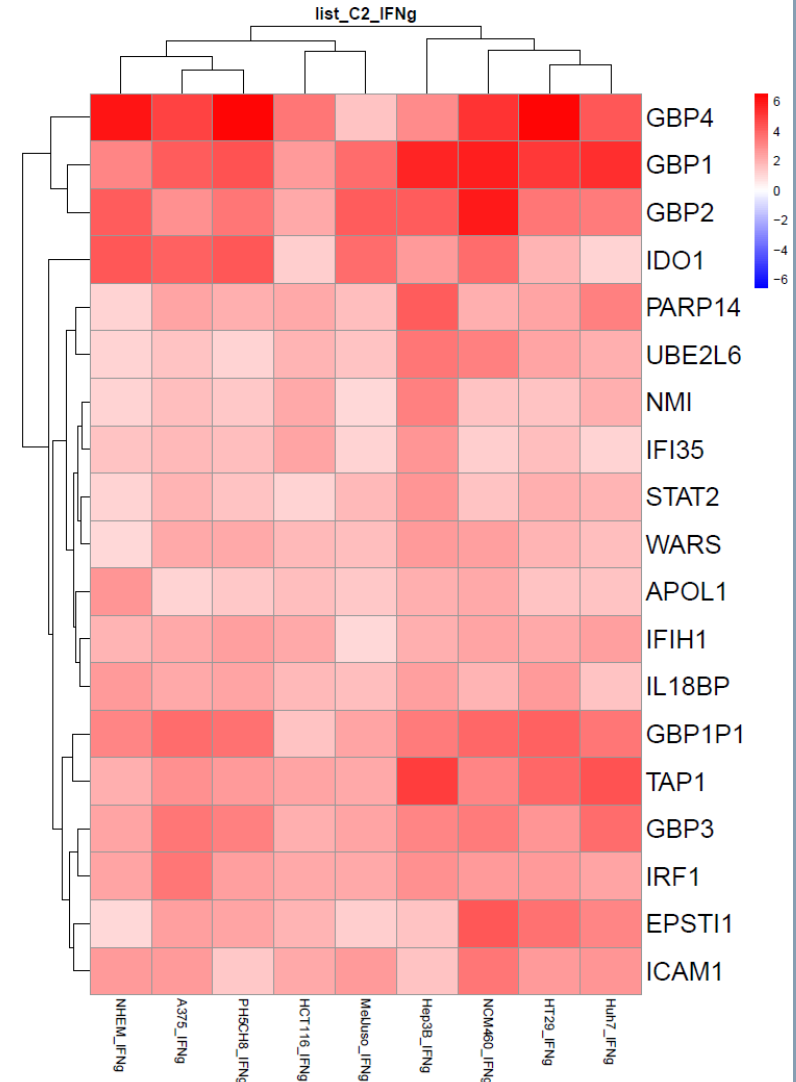
mRNA

Cell line	OSM	HIL6	IL27	IFNg
MelJuso	2	3	9	67
HCT116	0	0	25	100
PH5CH8	35	3	34	190
HT29		43	25	236
NCM460		34	79	195
Hep3B		94	102	221
A375	8	4	124	324
Huh7	134	88	106	142
NHEM	29	34	54	521

miR

Cell line	OSM	HIL6	IL27	IFNg
MelJuso	0	1	2	13
HCT116	6	0	0	0
PH5CH8	1	0	1	6
HT29		3	0	18
NCM460		3	1	30
Hep3B		0	3	3
A375	12	1	56	44
Huh7	13	18	0	4
NHEM	10	6	3	24

FDR = 1e-2, |logFC|>1

Common: only with IFNg stimulation – 19 mRNAs

And the rest of the interpreting is done at Uni side – we are waiting for an update meeting with them. In December they were presenting 2 posters.

LUXEMBOURG
INSTITUTE
OF **HEALTH**
RESEARCH DEDICATED TO LIFE