

# ***Genetical genomics applied to hematopoietic development using mouse models***

*Bioinformatics research into Illumina microarray data*

Presentation Bachelor Thesis

Danny Arends (s1276891)

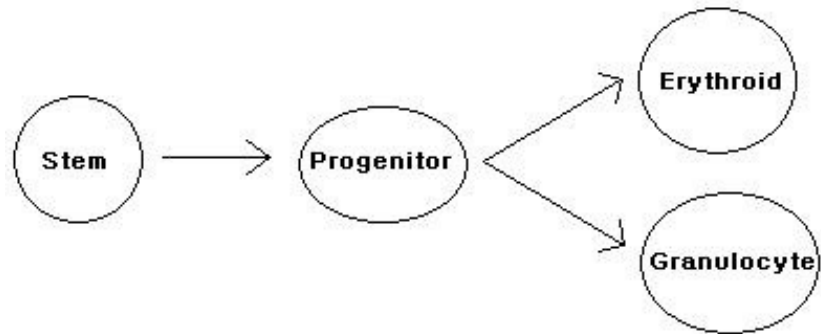
Supervision: Yang Li & Ritsert Jansen

# Presentation Overview

- Experiment
- Illumina
  - Technology overview
- Bioinformatics (Genetical Genomics)
- Results
  - Data quality
  - Differential expression
  - eQTL
- Discussion
- Future perspectives

# Experiment

- Stemcells
- Progenitory cells
  - Erythroid
  - Granulocyte
- Rna-expression measured in each cell type.
- Illumina sentrix beadarrays



# Experiment

- Question: Can we find gene(s) responsible for differentiation ?
- Inverse question: Can we find gene(s) responsible for proliferation of stem cells ?
- Hypothesis:  
Bioinformatics can help find regulatory elements and interactions governing differentiation/proliferation

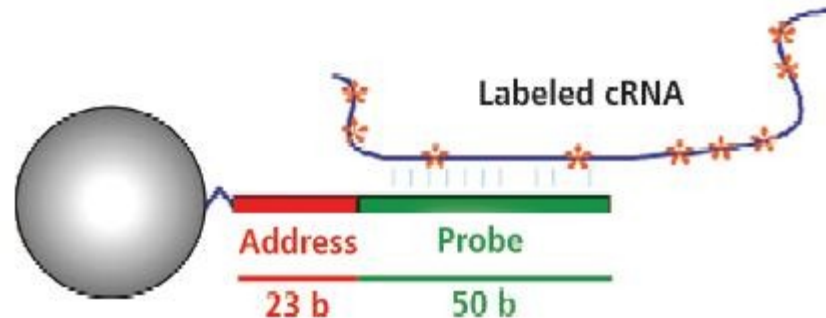
# Illumina

- Founded april 1998
- Innovative technology
- DNA / RNA arrays
- Sentrix Beadarray technology
- 9/1/2005, Mouse whole genome expression
- 9/1/2006, Rat whole genome expression

# Technology overview

## Manufacturing the arrays

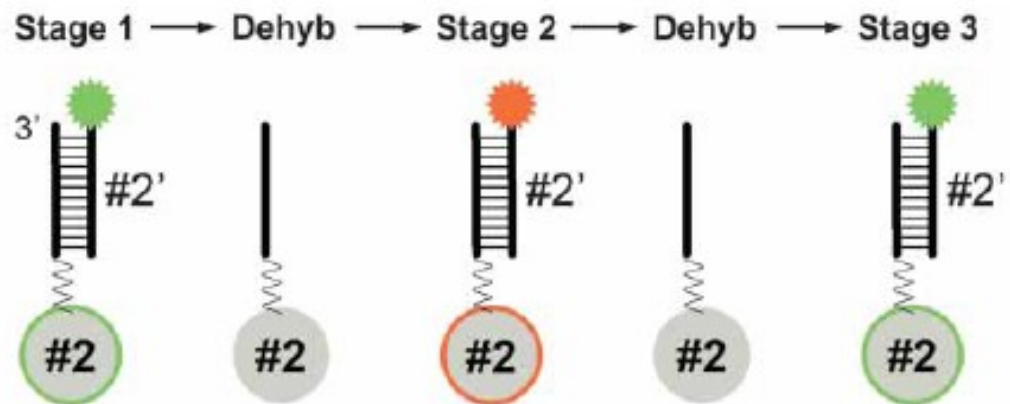
- Long oligonucleotides attached to glass beads
- 50 nucleotide probe
- 23 nucleotide address (bead ID)
- Pooled together in beadpools
- Random assembly in etched wells



# Technology overview

## Decoding the arrays

- Sequential hybridization with labeled decoders
- R-G-0-R-R-G-G-0
- 2 OFF states per code



# Technology overview

- Sentrix technology
  - Mouse (*mus musculus*)
  - 6 samples per sentrix microarray
  - 5 for this experiment (every 6<sup>th</sup> position failed)
  - Bead intensity measurements
- 62 samples on 13 arrays
  - 20 stem cells
  - 14 progenitor
  - 14 erythroid
  - 14 granulocytes

# Problems / Issues

- Random number of replicate beads()
- Decoding errors
  - Defect beads
  - Non detection
  - Wrong detection
- Unique arrays
- Batch effect / array effect
- Non functional arrays

# Bioinformatics

## (Genetical Genomics)

Data pre-processing

- Data exploration
- Quantile Normalization

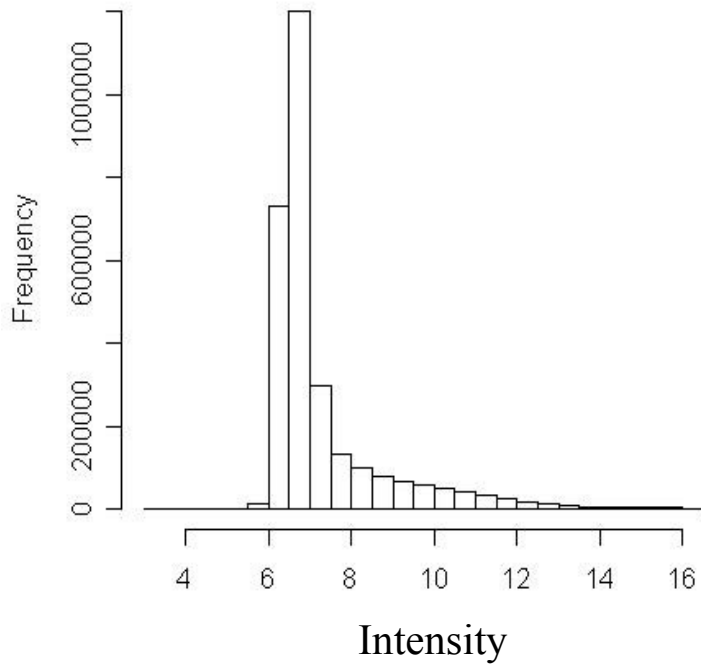
Cell type specific effect in expression (ANOVA analysis)

Genetical Genomics studies

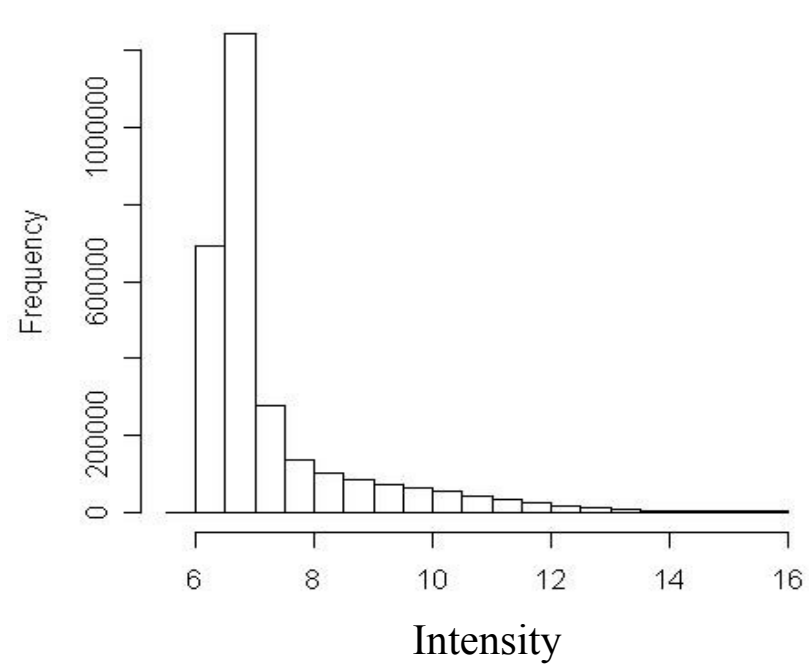
- Recombinant inbred lines
- Expression QTL mapping
  - Molecular markers
  - Expression

# Data exploration

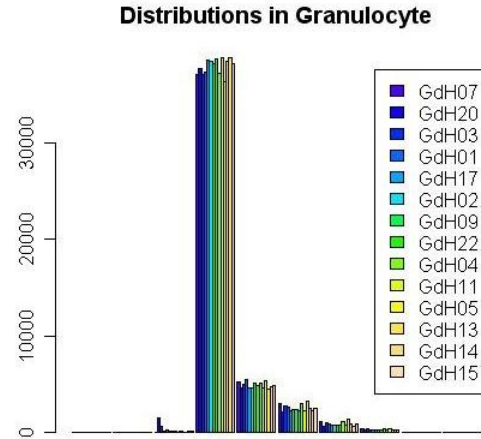
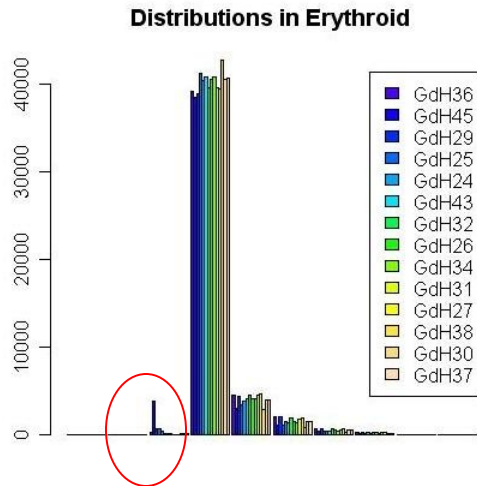
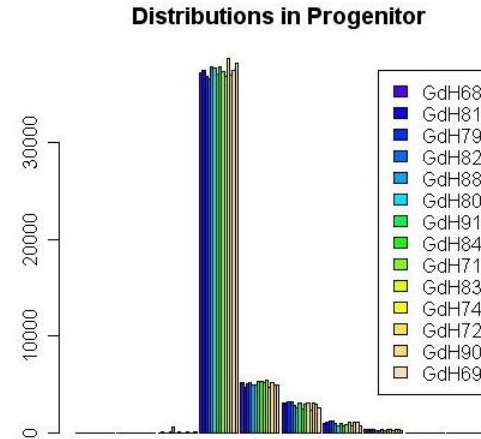
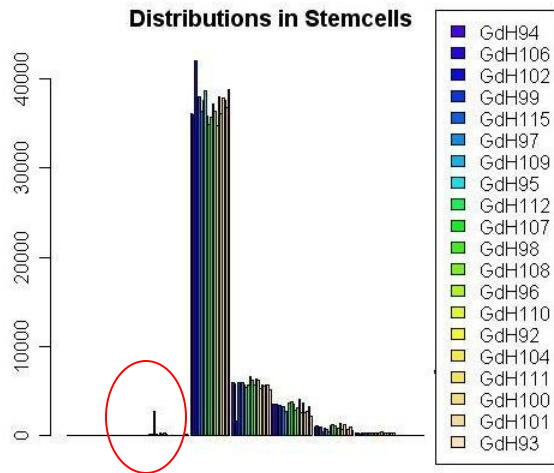
Before quantile normalization



After quantile normalization



# Data exploration

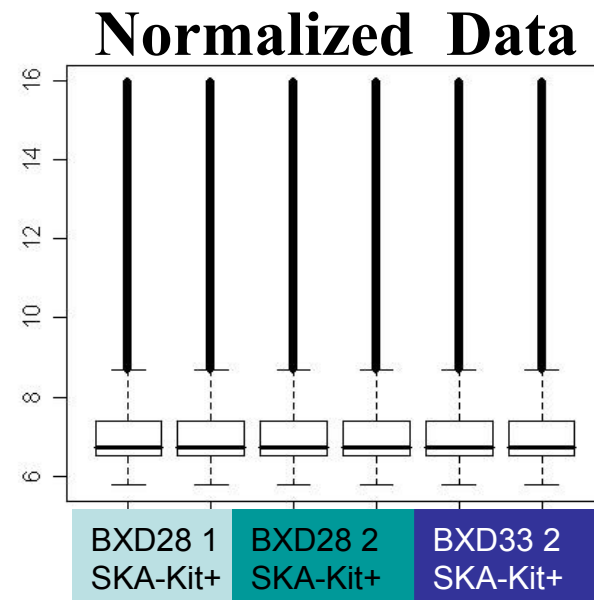
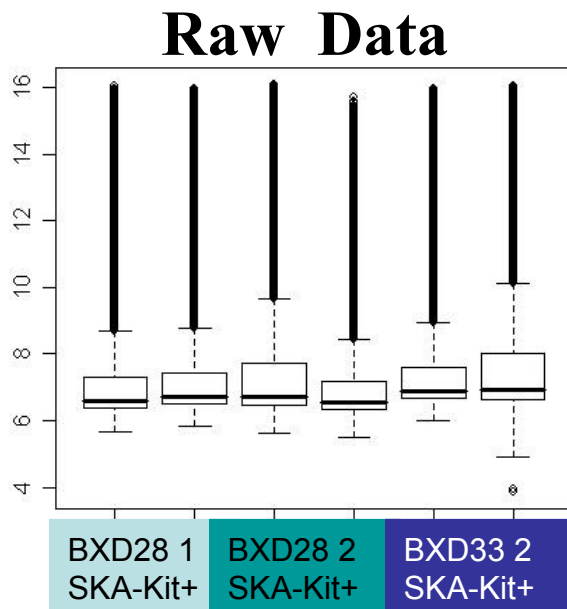


# Results histogram analysis

- No overhybridization seen
- Normalization doesn't seem to significantly change distribution
- Samples look equally distributed
- Stemcells and erythroid each seem to have 1 array with lower intensities

# Data exploration

- Biological replicates present
- Equal variation after normalization
- Boxplots of the biological replicates



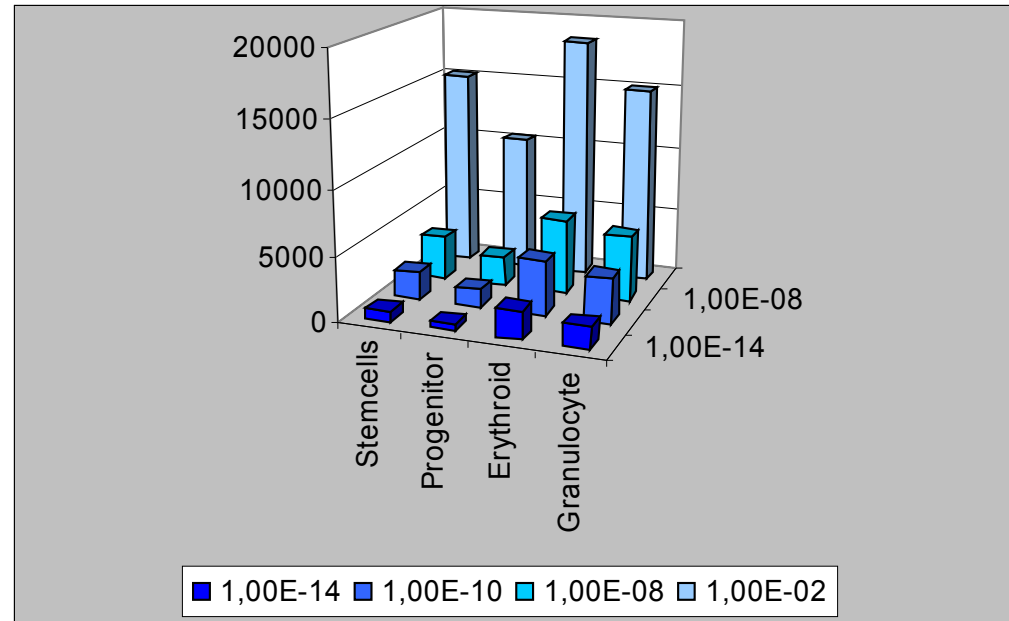
# Differential expression

- Differential expression using ANOVA
- ANOVA Model:  $y_i = \mu + C_i + \varepsilon$
- T-test for celltype specificity
- Test compared one cell type with the other 3 cell types

$$T = \frac{\bar{X} - \bar{Y}}{S \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

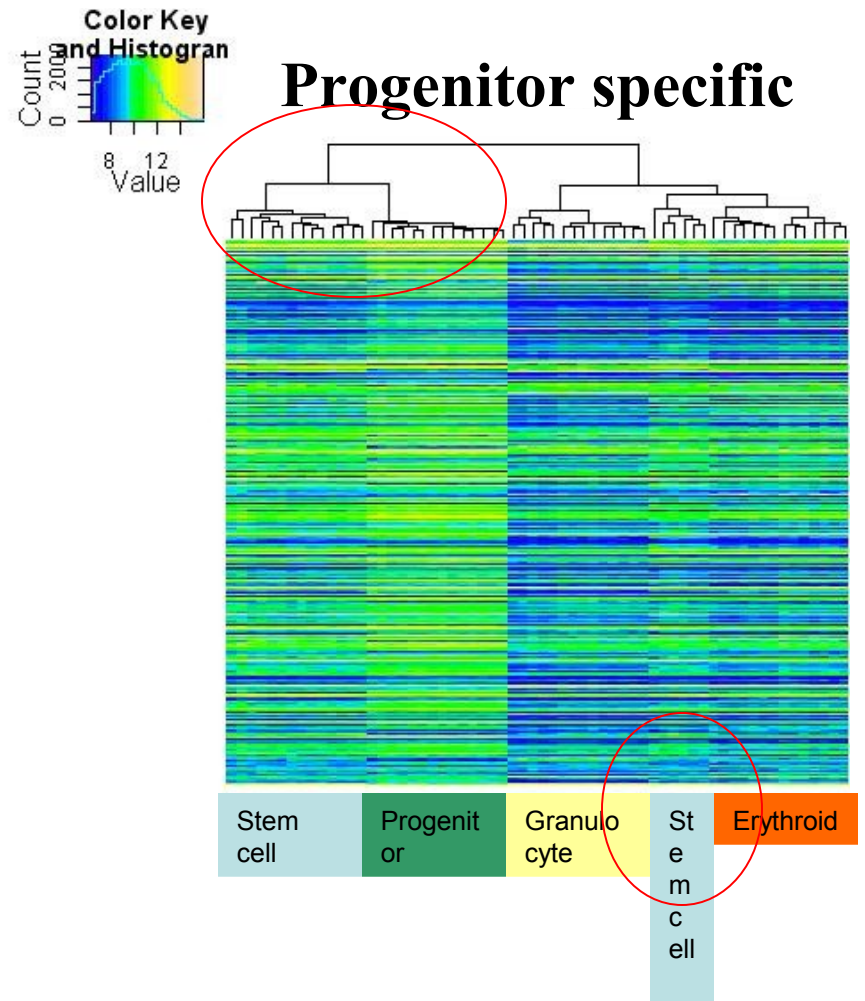
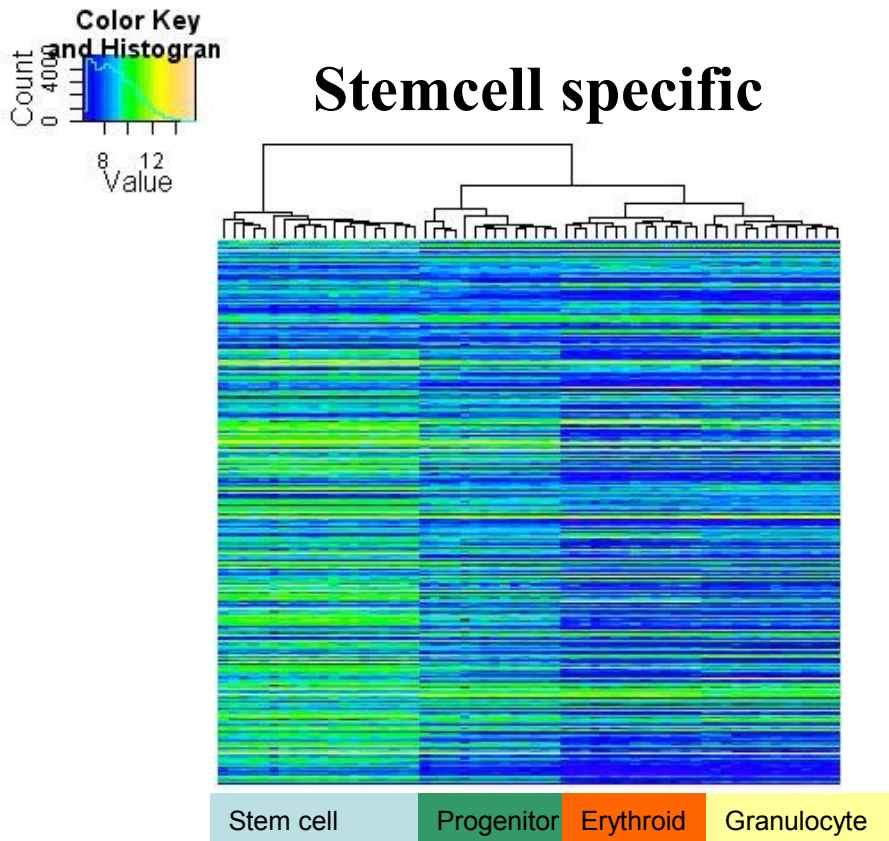
# Number of probes with cell specific effect

- Threshold vs. Hits

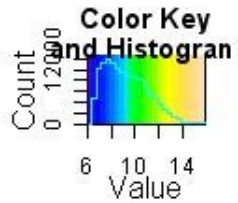


	1,0E-02	1,0E-08	1,0E-10	1,0E-14
Stemcells	15176	3340	2088	815
Progenitor	10416	2234	1383	501
Erythroid	18523	5769	4200	2128
Granulocyte	15028	5030	3460	1655

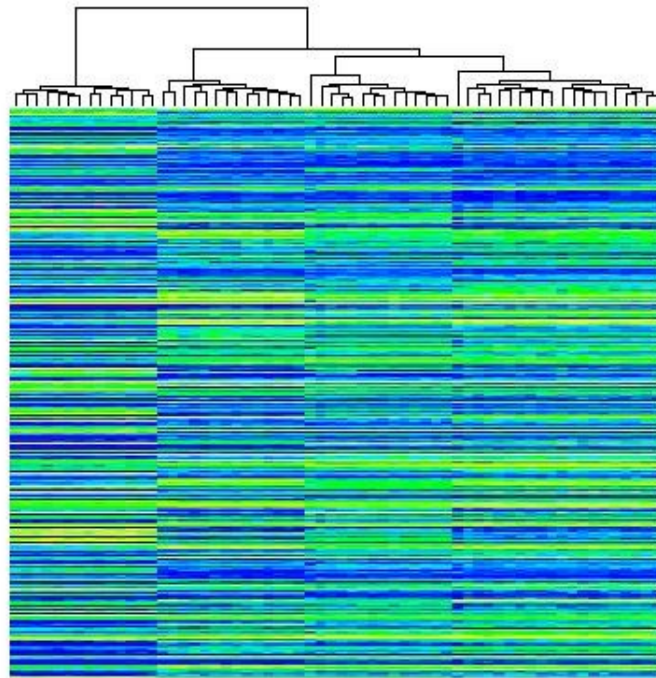
# Heatmap of gene expression for cell specific probes



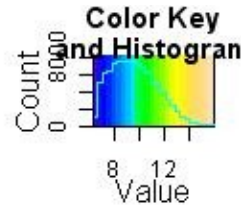
# Heatmap of gene expression for cell specific probes



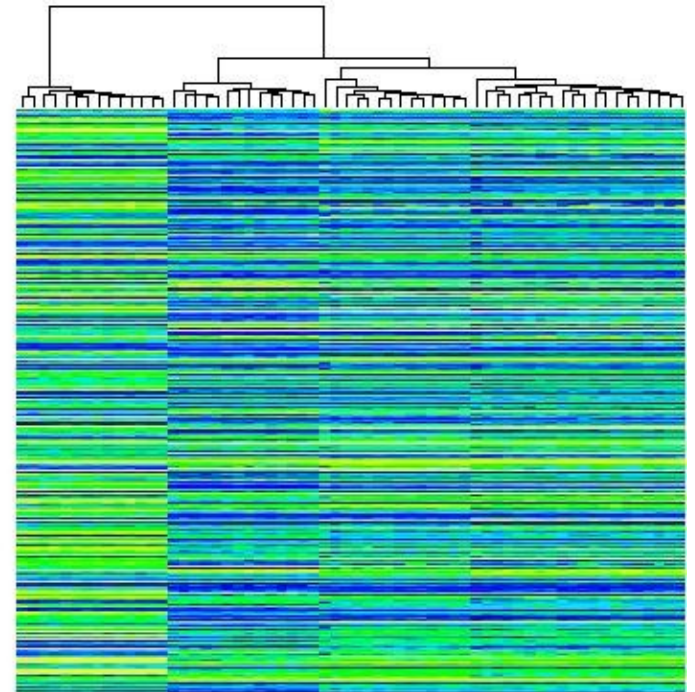
## Erythroid specific



Erythroid Granulocyte Progenitor Stem cell

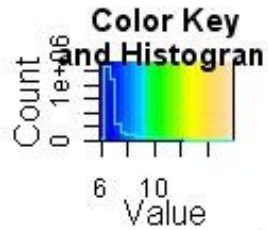


## Granulocyte specific

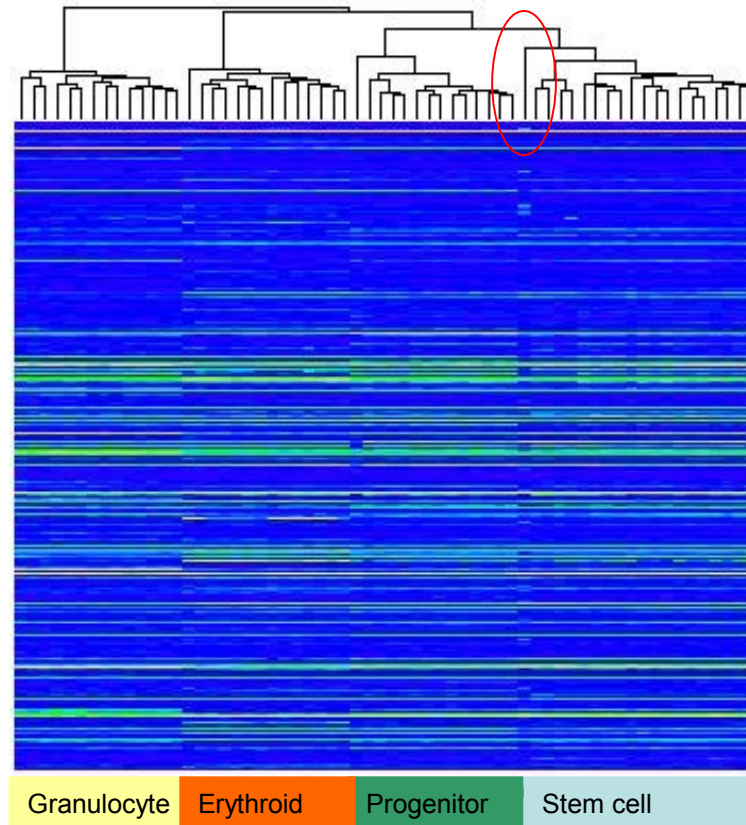


Granulocyte Erythroid Progenitor Stem cell

# Heatmap of gene expression for all probes

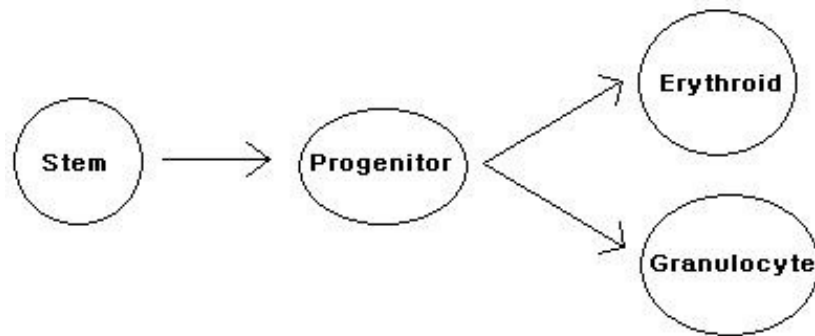


**All probes (no specificity)**



# Results heatmap analysis

- Differential expression exists
- Distance clustering shows outlier in stemcell
- Clustering into 4 distinct cell types
- Progenitor specific genes show overlap with type 1 (stemcells)
- Model seems to fit distance clustering for all probes



# goClustering

- Gene ontology groups
  - controlled vocabulary to describe **gene** and **gene** product attributes in any organism.
- Detect overrepresentation of goGroups
- Statistical process

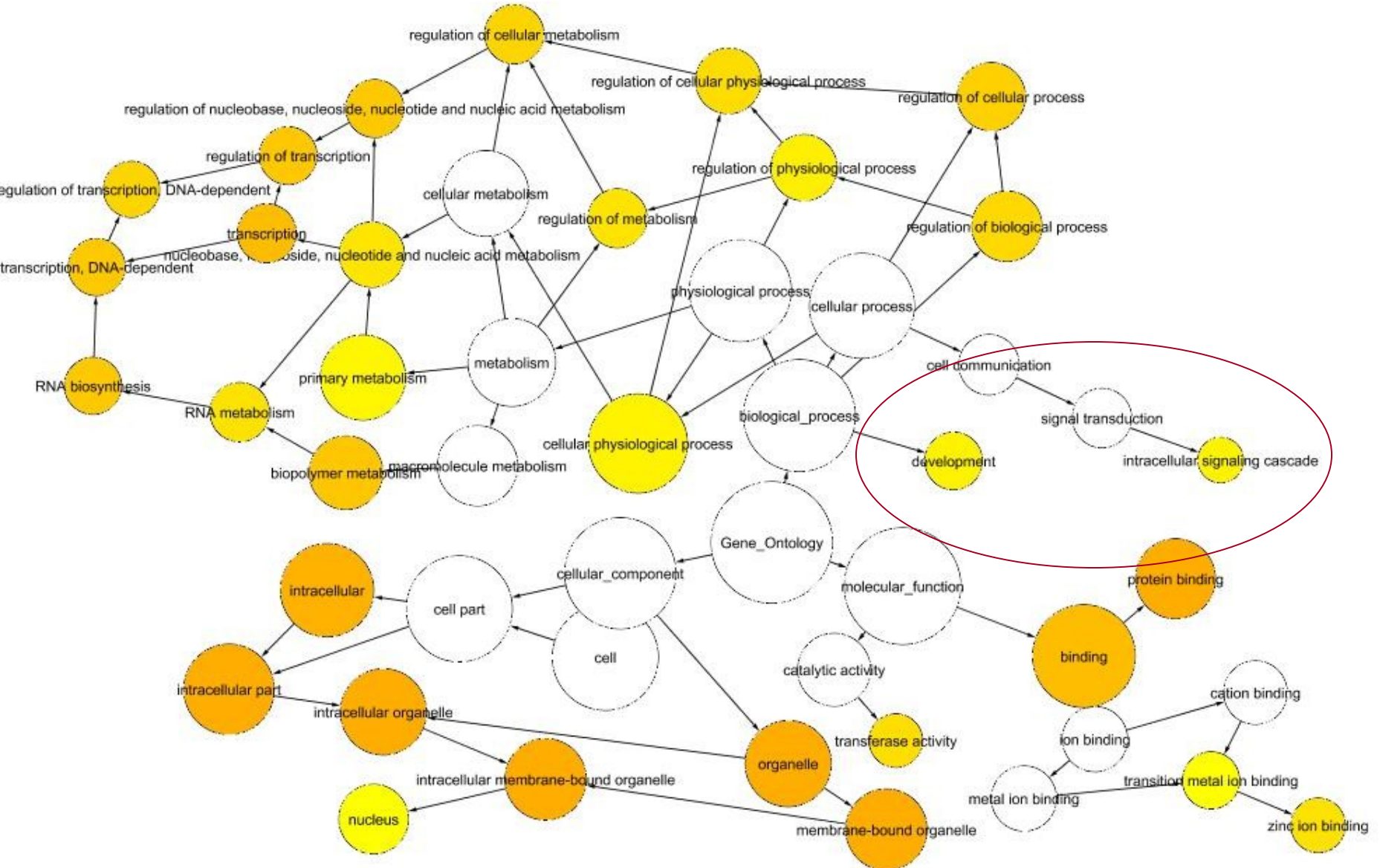
# Summary of ontology clustering

	Stem cells	Progenitor	Granulocyte	Erythroid
Biological process	Ubiquitin Cycle	cellular macromolecule metabolic process  <b>immune system process</b>	nucleobase, nucleoside, nucleotide  nucleic acid metabolic process  macromolecule metabolic process  <b>cell communication</b>	cellular carbohydrate metabolic process  methylation-dependent chromatin silencing  M phase  cell cycle  biogenic amine catabolic process  <b>autophagy</b>
Molecular Function	None Found	unfolded protein binding  monovalent inorganic cation transporter activity  phosphoric monoester hydrolase activity	nucleic acid binding  RNA binding  cytoskeletal protein binding  signal transducer activity  urea transporter activity  phospholipid-translocating ATPase activity  <b>microtubule motor activity</b>	cobalt ion transporter activity

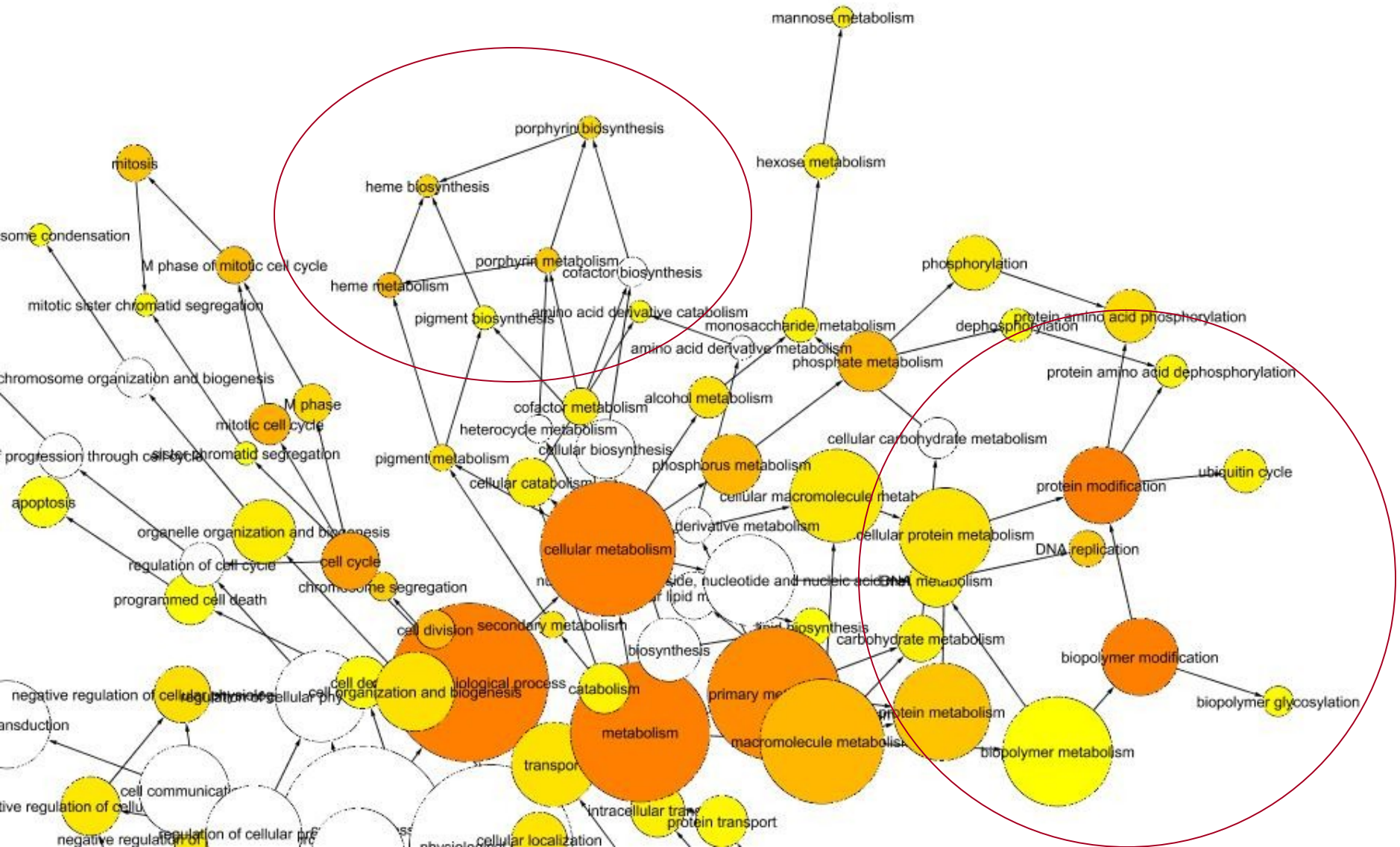
# Visualizing gene ontology

- Cytoscape for visualizing networks
- Using BINGO
- Database ontology search
- Hierarchical groups

# Gene ontology of stem cells



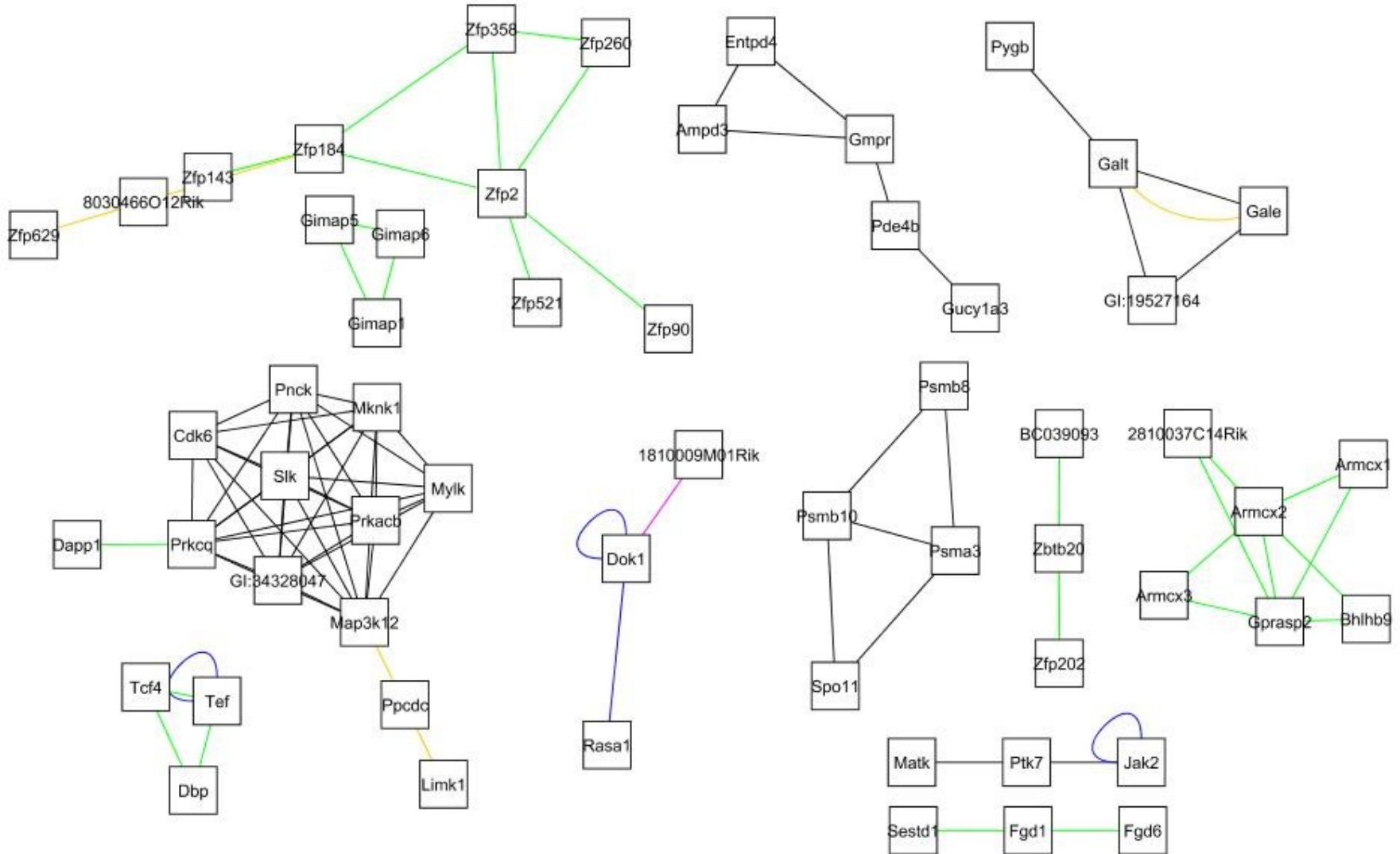
# Gene ontology of erythroid cells



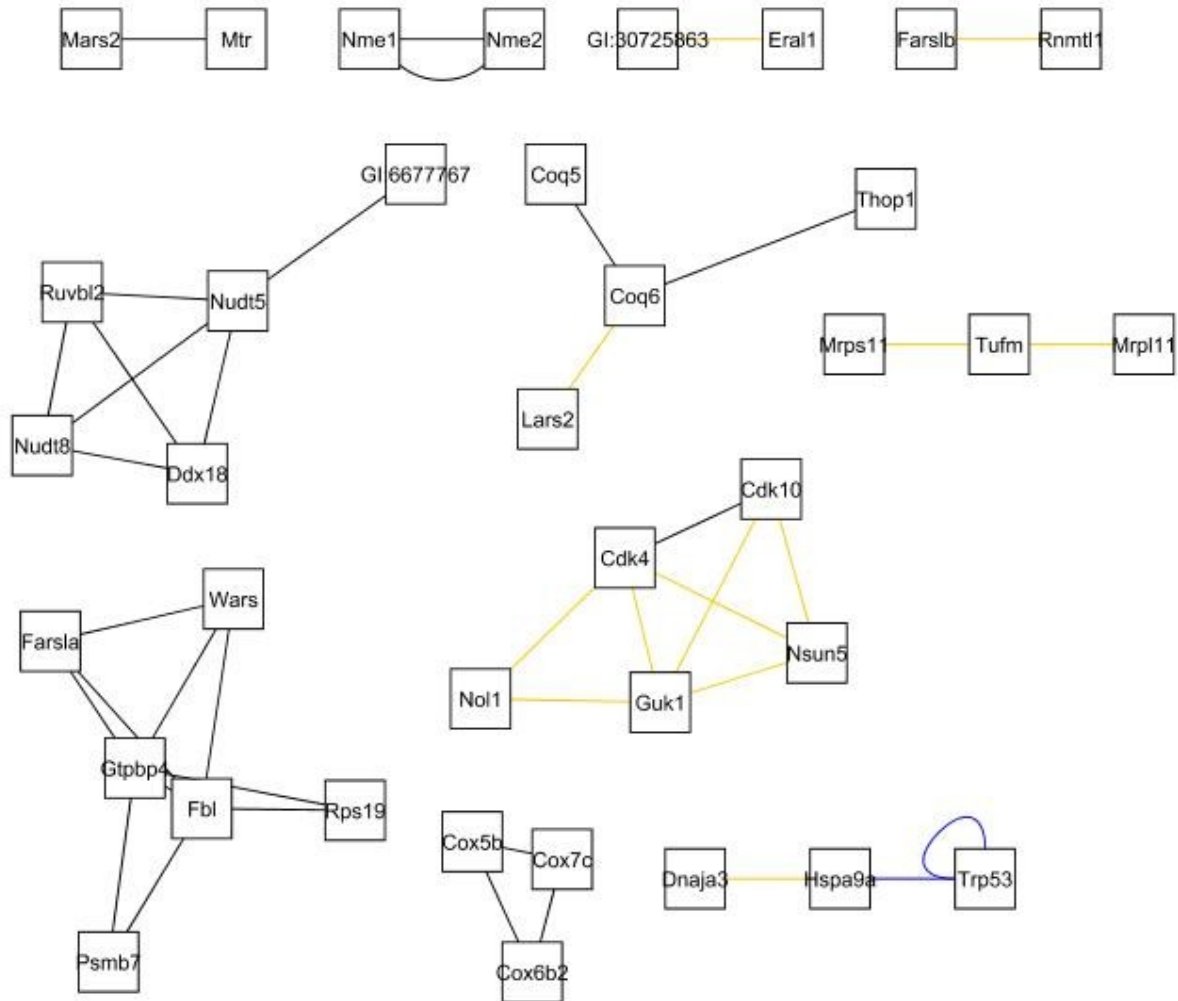
# Interaction networks

- Made by BioNetBuilder
- Public databases searched:
  - KEGG/ProLinks/BIND/DIP/BioGrid
- Stem: 110 interaction
- Prog: 48 interactions
- WBC: 469 interactions
- RBC: 552 interactions

# Stemcell interactions

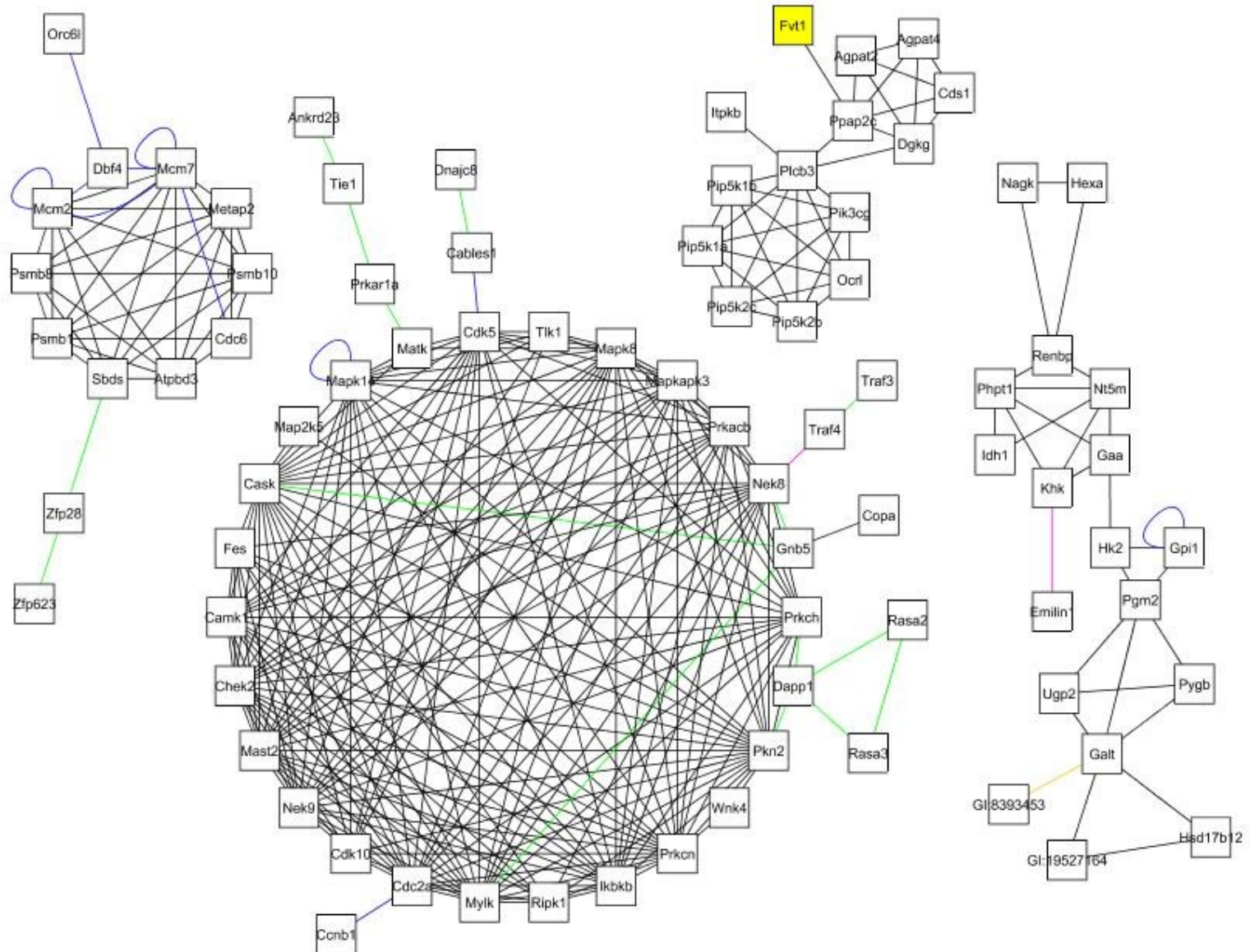


# Progenitor interactions





# RBC interactions

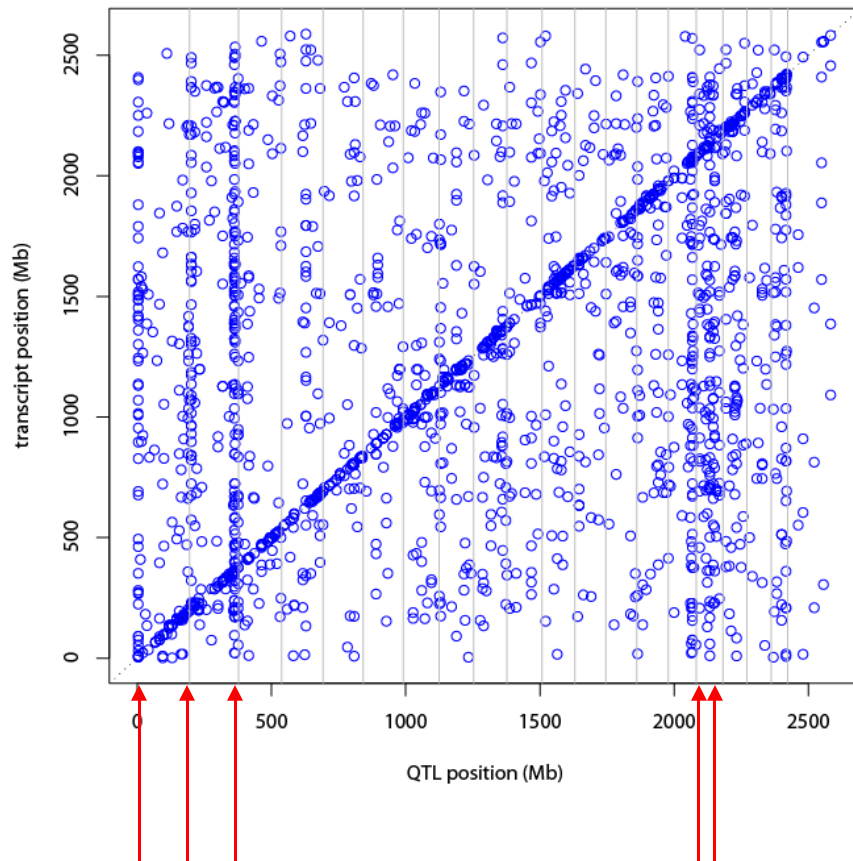


# eQTL mapping

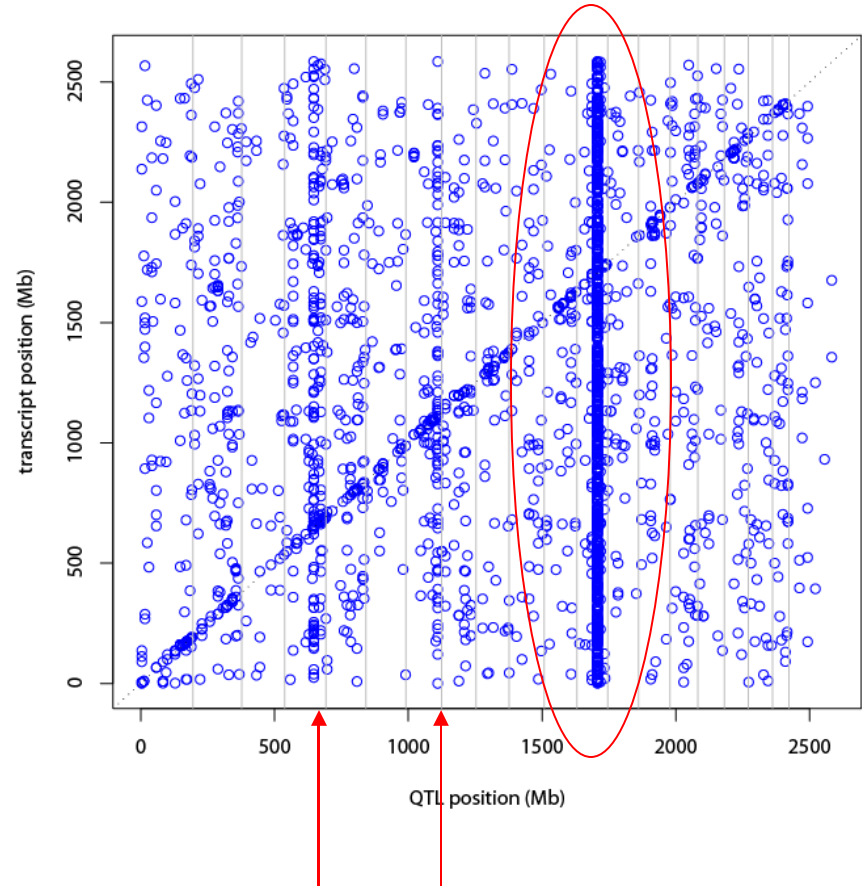
- ANOVA Model:  $y_i = \mu + Q_i + \varepsilon$
- Integration of genetic information and gene expression profiling
- *Cis* and *Trans* effects seen in all cell types
- Progenitor cell type has very strong transband

# Comparison of eQTL position with transcript position

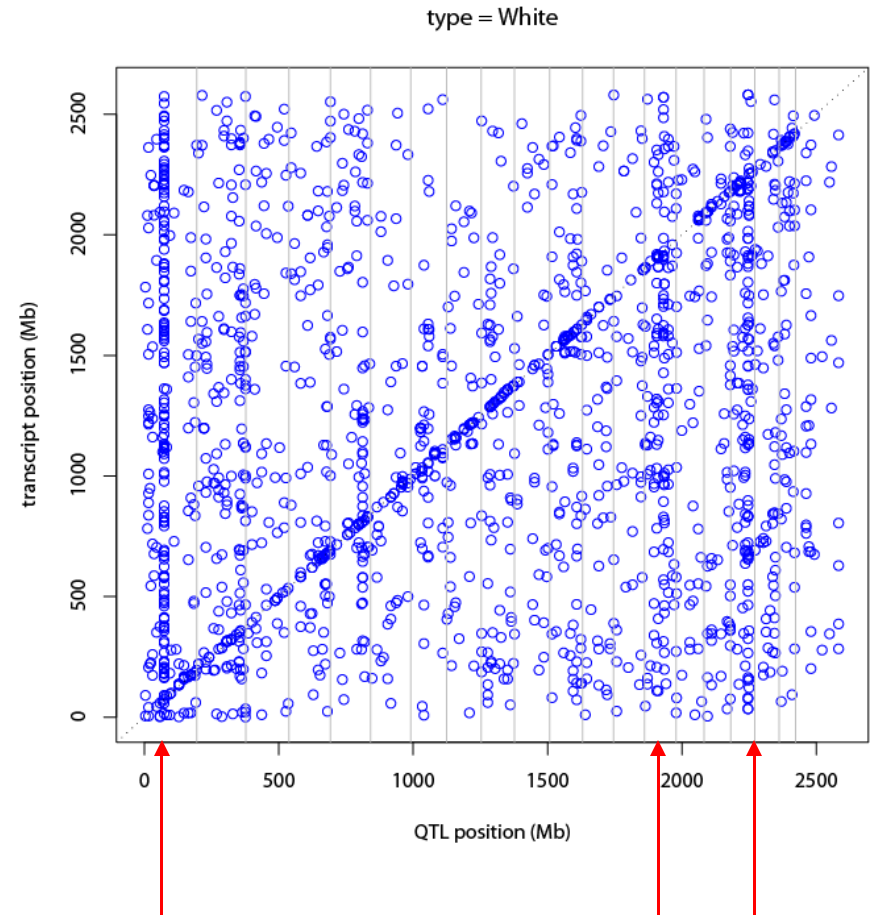
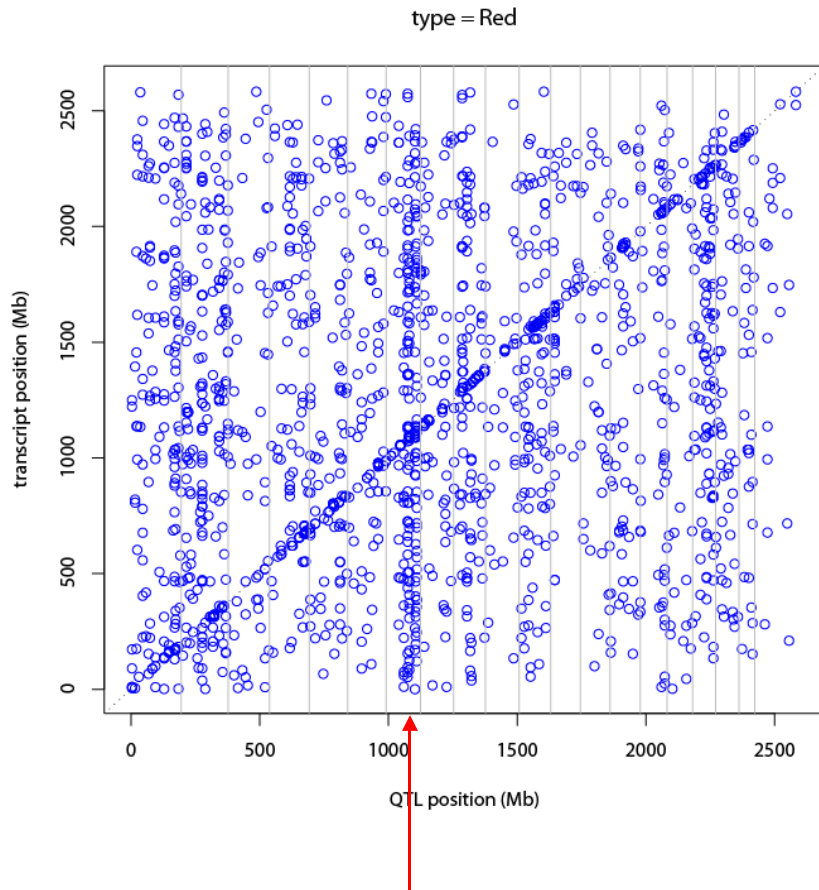
type = Stem



type = Pro



# Comparison of eQTL position with transcript position



# Conclusion / Discussion

- Illumina beadarray is a new technology
- Data obtained from illumina seem to be good
- Bioinformatics tools can be used to obtain knowledge about biological systems.
- Gene ontology can give researchers a clue in which pathway to look for regulation(?)
- Interaction networks
- eQTLs can give hints into regulatory mechanisms
- Changes in transbands seen between cell-types
  - Regulation ?

# Future perspectives

- Bead level analysis of Illumina data
- Comparison of Illumina with Affy
- Scaling up experiment
- Further eQTLs studies
- Combining eQTLs with gene ontology
- Combining eQTLs with known interactions
- More research into stemcell differentiation is needed(Biology).

# Questions ?

I'd like to thank the following people:

- From GBIC (Haren)
  - Ritsert Jansen
  - Yang Li
  - Brunno Tesson
  - Morris Swertz
  - Gonzalo vera Rodriguez
- From Stemcell biology (UMCG)
  - Leonid v Bystrykh
  - Gerald de Haan