Ottawa Hospital
Research Institute
Institut de recherche
de l'Hôpital d'Ottawa

uOttawa
L'Université canadienne
Canada's university

Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis

Theodore J. Perkins
Oct 1, 2021

Connecting Network Structure to its Dynamics: Fantasy or Reality?
(BIRS Meeting 21w5005)

most of this talk comes from:

CellPress

Molecular Cell

Resource

Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis

Mark A. Gillespie,[1,7] Carmen G. Palii,[2,3,7] Daniel Sanchez-Taltavull,[2,3,4,7] Paul Shannon,[1] William J.R. Longabaugh,[1] Damien J. Downes,[5] Karthi Sivaraman,[2] Herbert M. Espinoza,[1] Jim R. Hughes,[5] Nathan D. Price,[1] Theodore J. Perkins,[2,3,8,*] Jeffrey A. Ranish,[1,6,8,*] and Marjorie Brand[2,3,8,9,*]

[1]Institute for Systems Biology, Seattle, WA 98109, USA

[2]Sprott Center for Stem Cell Research, Ottawa Hospital Research Institute, Ottawa, ON K1H8L6, Canada

[3]Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON K1H8L6, Canada

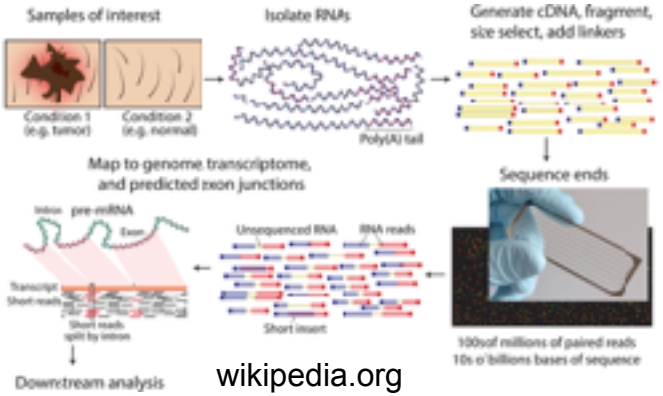
[4]Visceral Surgery and Medicine, Inselspital, Bern University Hospital, Department for BioMedical Research, University of Bern, Murtenstrasse 35, 3006 Bern, Switzerland

[5]MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DS, UK

[6]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

Motivation: Usually, gene expression is not measured in "absolute" units

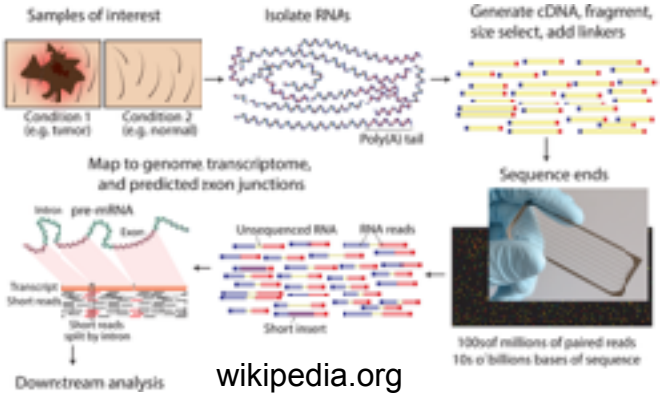
[sc]RNA-seq (rpm,fpkm,tpm)



wikipedia.org

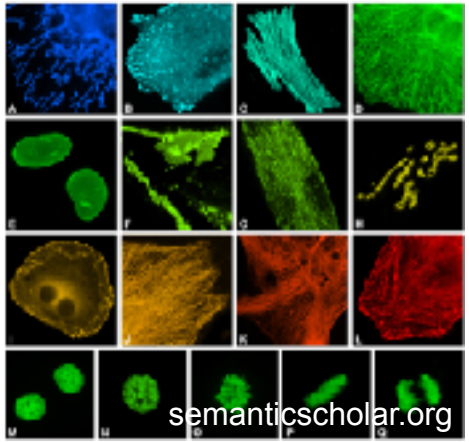
Motivation: Usually, gene expression is not measured in "absolute" units

[sc]RNA-seq (rpm,fpkm,tpm)



wikipedia.org

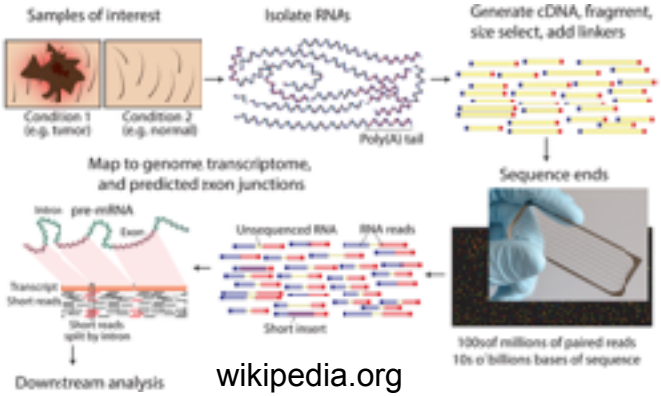
protein fluorescence (afu)



semanticscholar.org

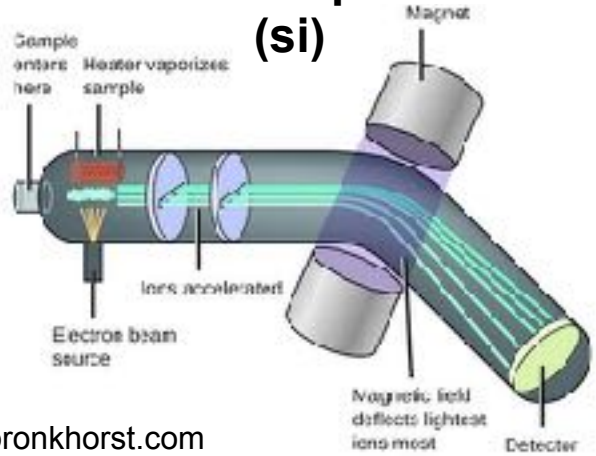
Motivation: Usually, gene expression is not measured in "absolute" units

[sc]RNA-seq (rpm,fpkm,tpm)



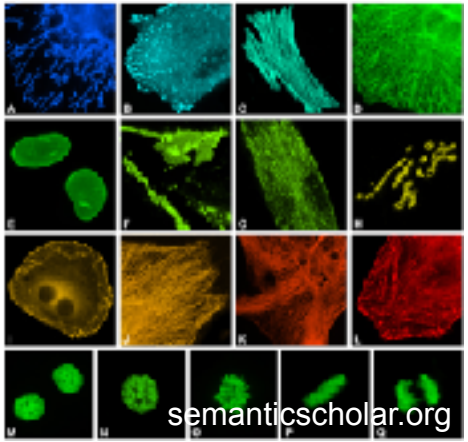
wikipedia.org

mass spec (si)



bronkhorst.com

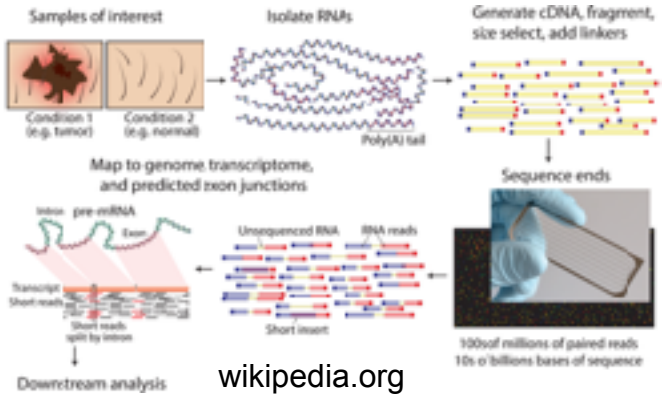
protein fluorescence (afu)



semanticscholar.org

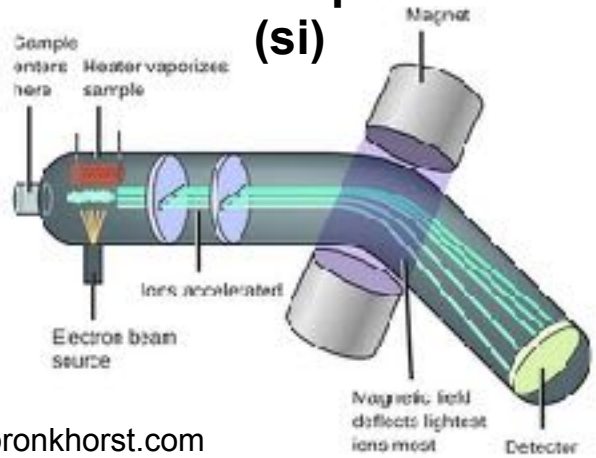
Motivation: Usually, gene expression is not measured in "absolute" units

[sc]RNA-seq (rpm,fpkm,tpm)



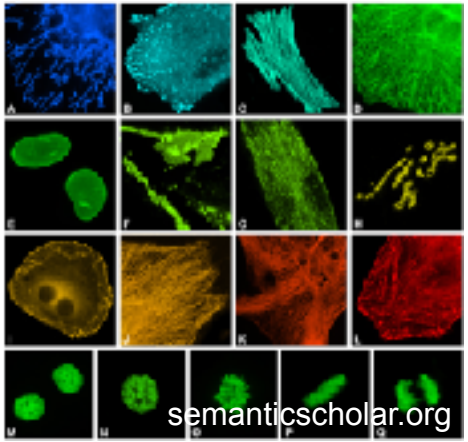
wikipedia.org

mass spec (si)



bronkhorst.com

protein fluorescence (afu)



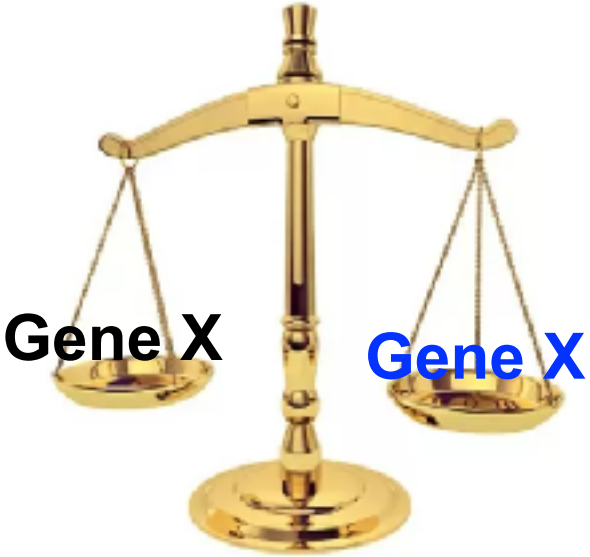
semanticscholar.org

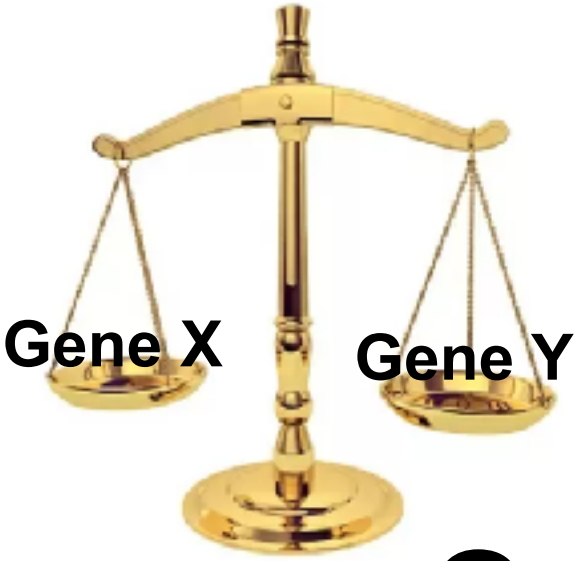
=> although: spike-ins, calibration, stochastic noise, normalization ...

Consequences

Absolute expression ?

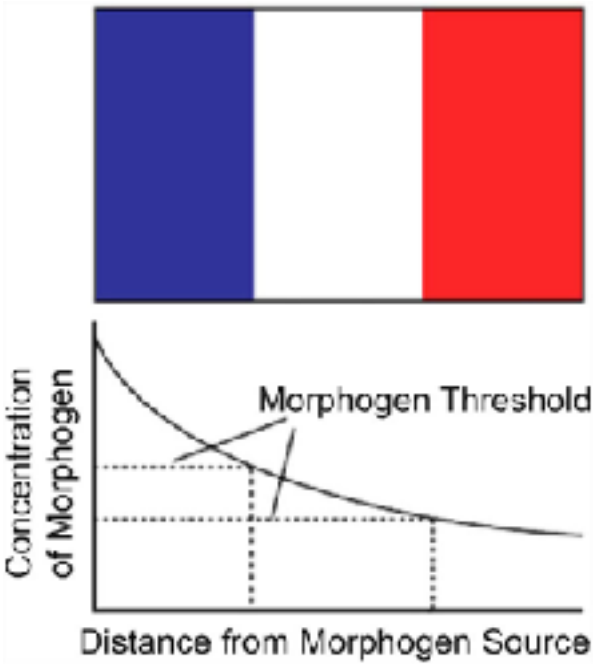
Differential expression ?



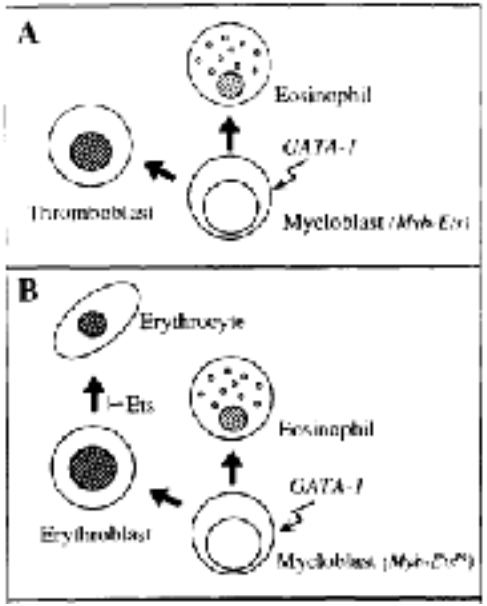


Stoichiometry ?

Importance of transcription factors dosage for cell fate decisions



Wolpert/Gardiner



○ No GATA1

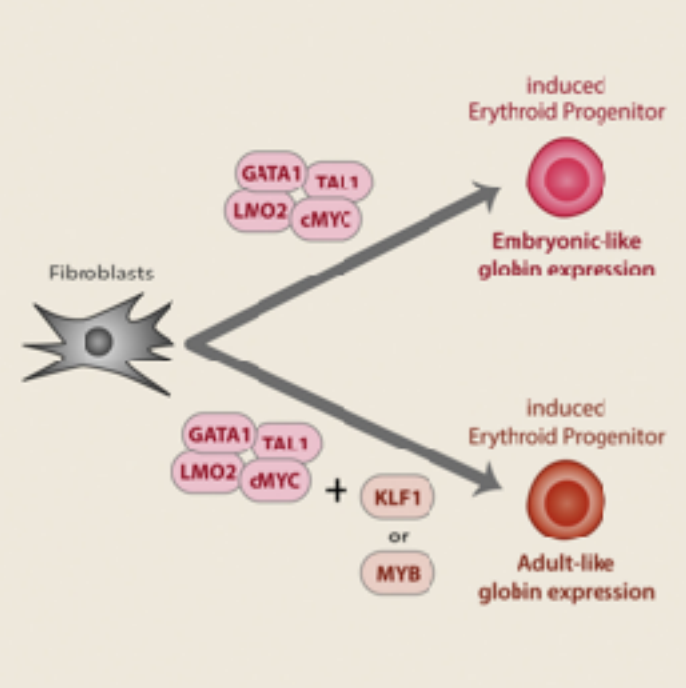
▨ Intermediate levels of GATA1

● High levels of GATA1

Kulessa, Frampton & Graf G&D (9): 1250-1262, 1995

→ **Different amounts of GATA1 protein promote alternate cell fates**

Transcription factors stoichiometries in cell fate decisions



2:1:1:1 GTLM ratio → exclusively red colonies

1:1:1:2 GTLM ratio → almost no red colonies

→ **The stoichiometry of TFs is key for reprogramming efficiency**

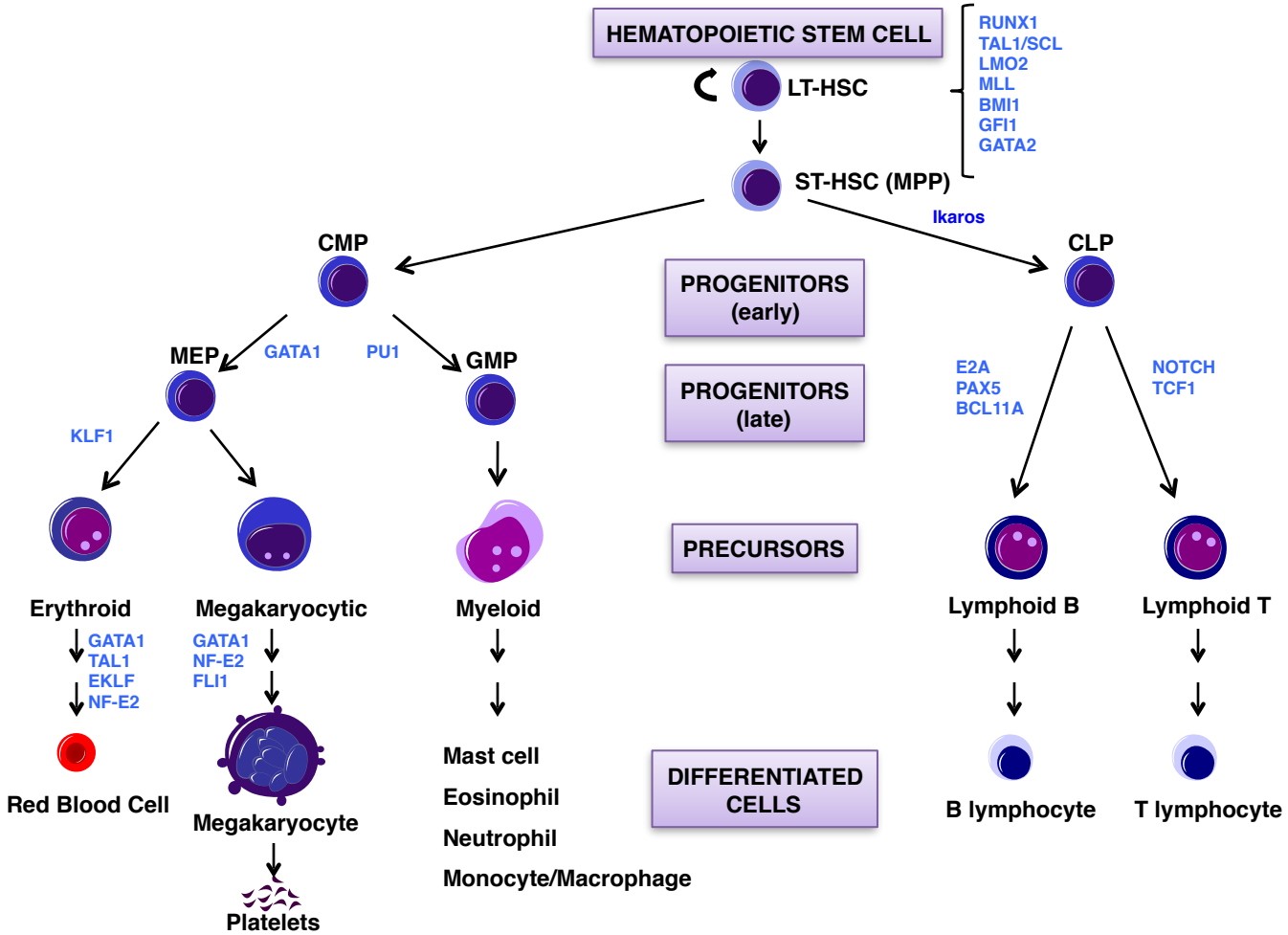
Capella-Garcia et al. Cell Reports 15(11): 2550-2562, 2016

Outline

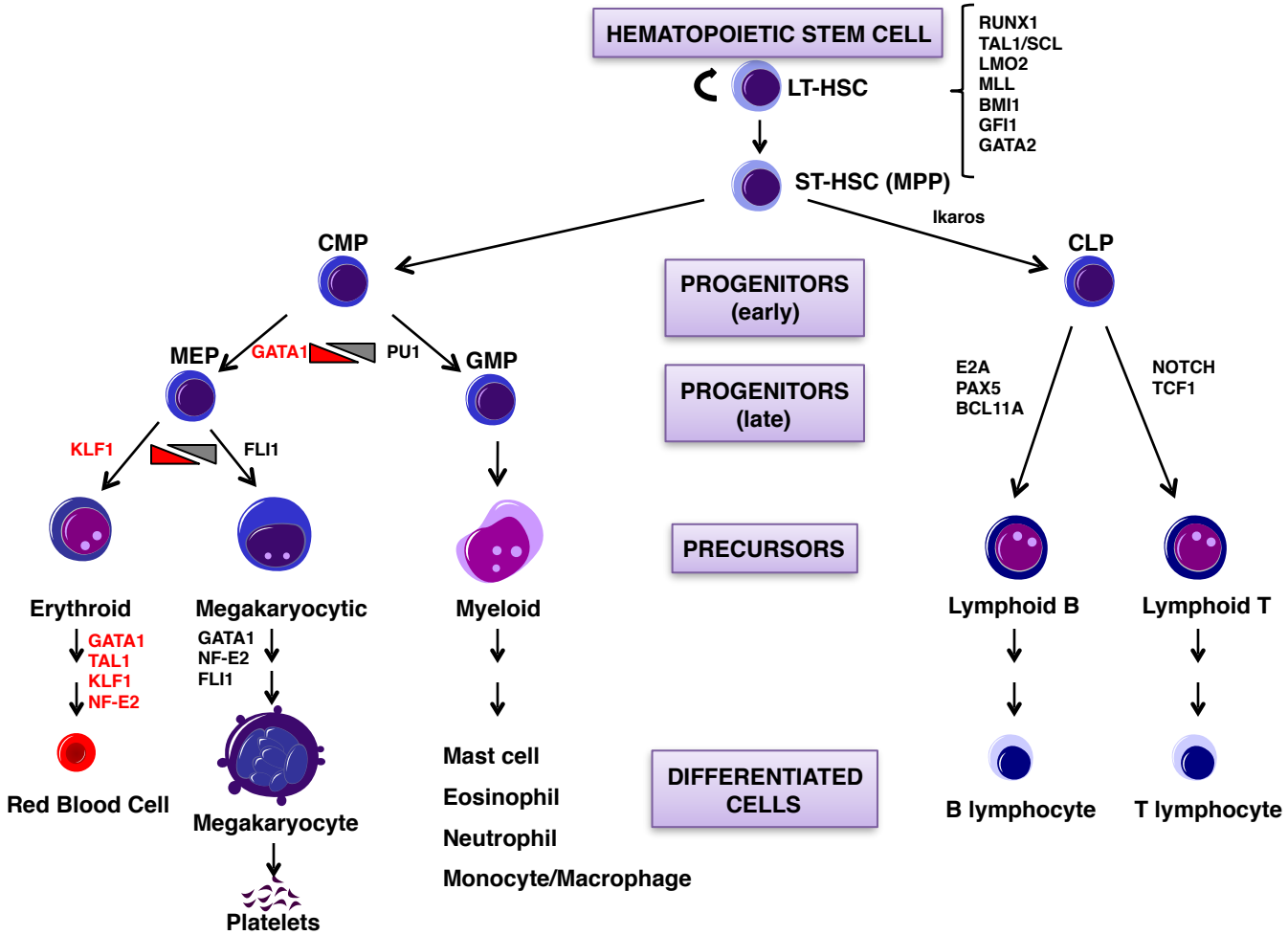
- Blood cell hierarchy
- Experiment design
- High-throughput absolute quantitative proteomics approach
- Immediate observations
- Gene regulatory network modeling

Blood Cell Hierarchy

Blood cell hierarchy

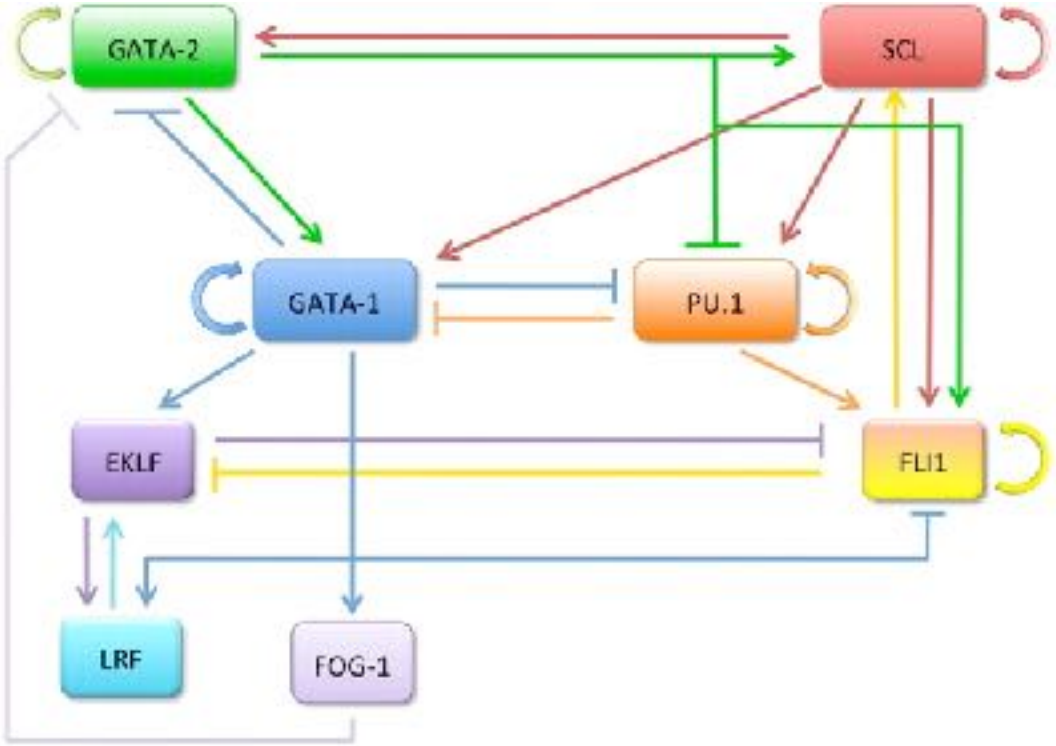


Changes in the relative levels of transcription factors drive erythropoiesis



Network model of cell fate choice in MEP based on lineage-specifying TFs cross-antagonism

Erythroid Fate



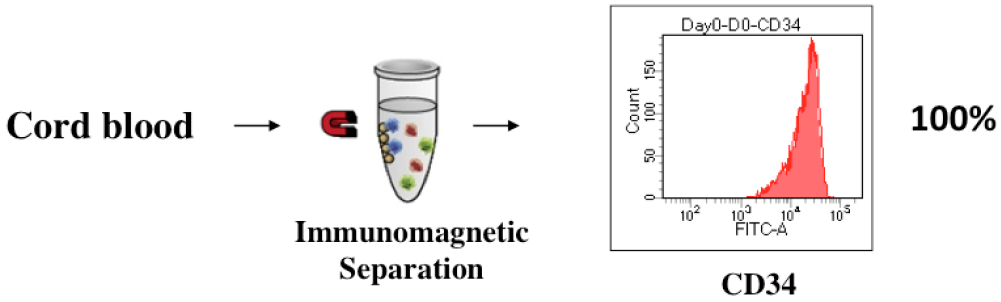
Megakaryocytic Fate

L.C. Doré, and J. D. Crispino, Blood 118:231-239, 2011

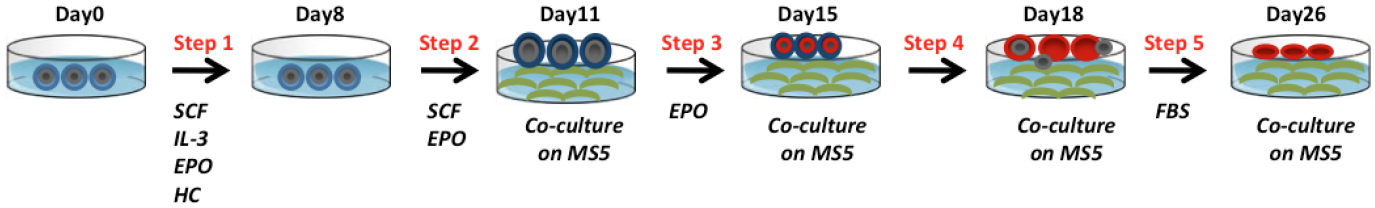
Experimental design

ex vivo human erythropoiesis

STEP 1 ***Positive Isolation of CD34+ human hematopoietic stem and progenitor cells (HSPCs)***



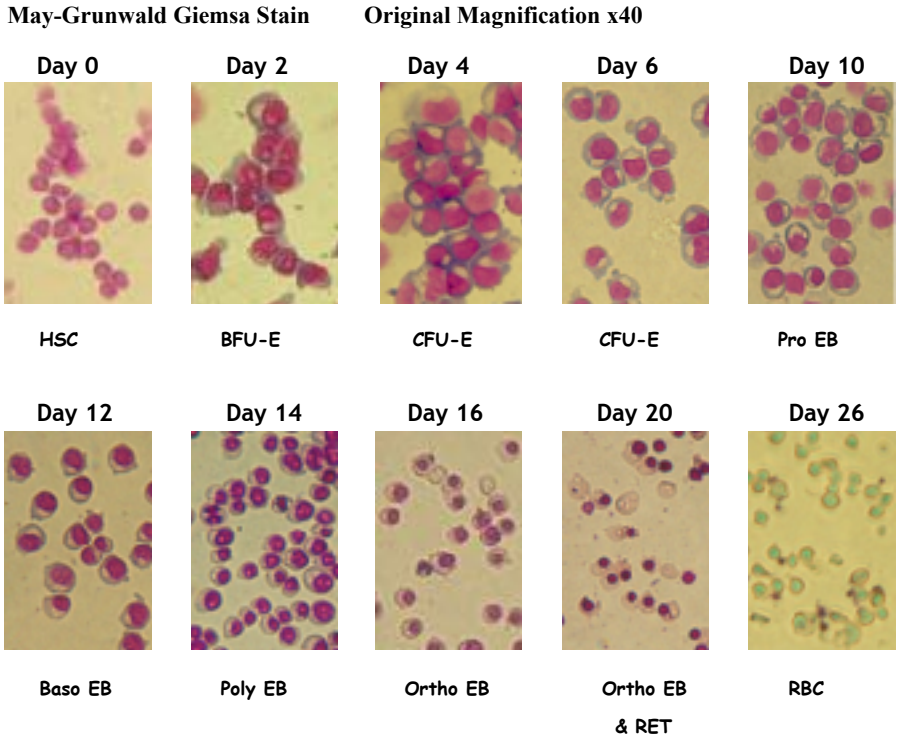
STEP 2 ***Serum-free Liquid Culture and Erythroid Differentiation***

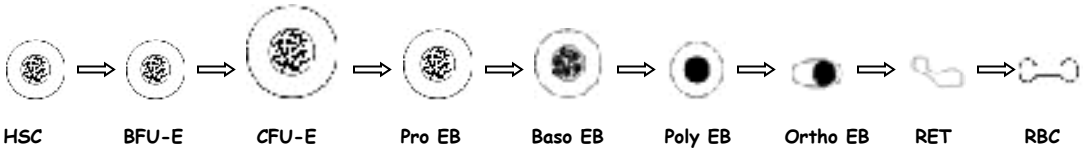


Giarratana et al. Nature Biotechnology, 23 (1), 2005

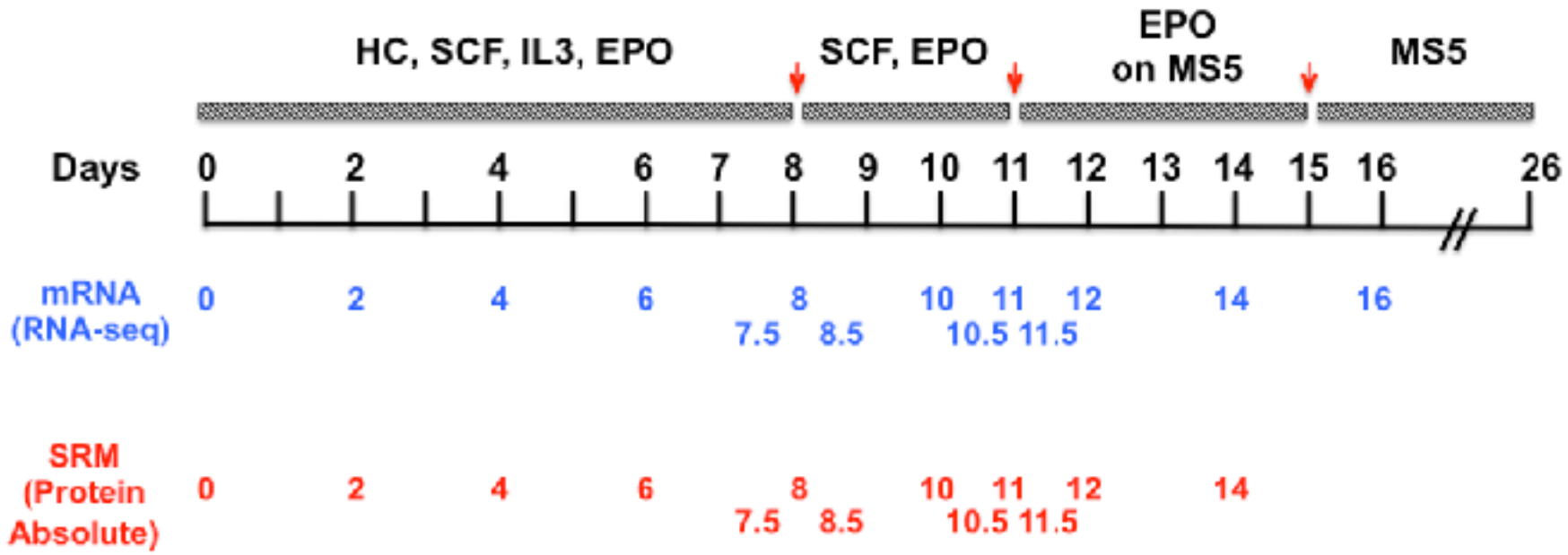
Palii et al. J Vis Exp. 53, 2011

Cell morphology during human *ex vivo* erythropoiesis





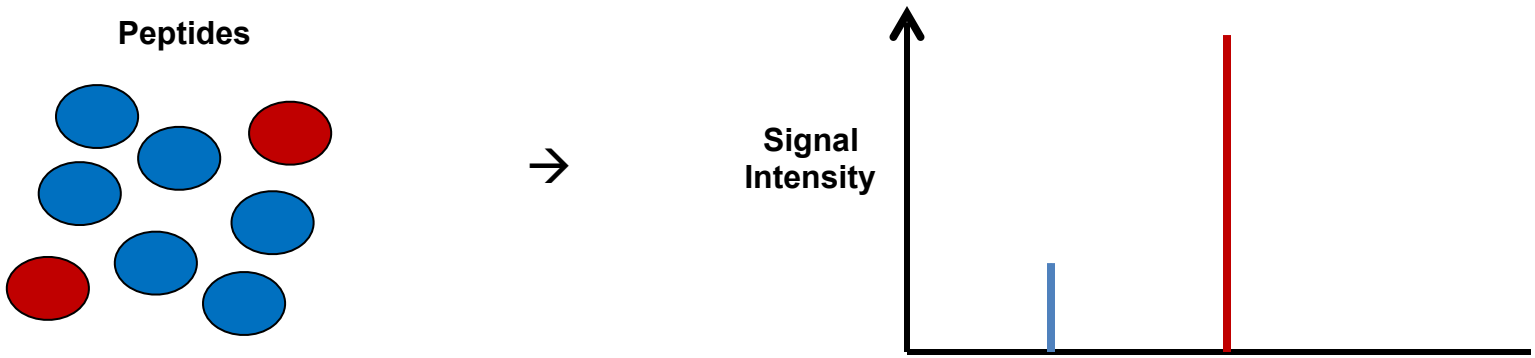
Data acquisition



+ iTraq, ATAC-seq, CyTOF, ...

quantitative proteomics approach

Mass Spectrometry is not inherently quantitative

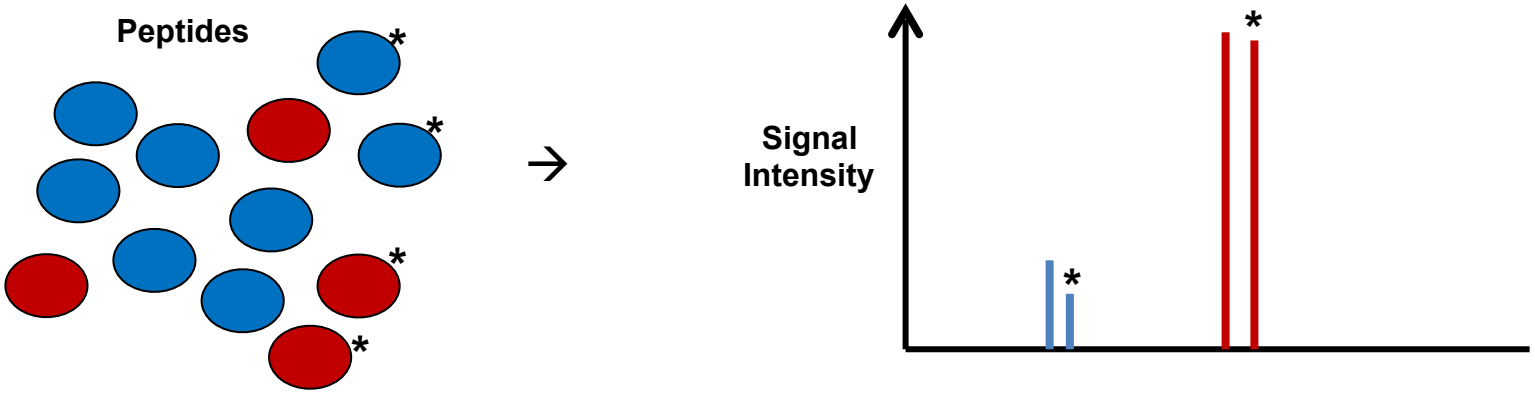


<u>For quantification</u>:

-> the <u>"proteomic ruler" method</u> to estimate copy number of proteins per cells (Wisniewski, J.R., Hein, M.Y., Cox, J., and Mann, M. (2014). Mol Cell Proteomics *13*, 3497-3506.

-> uses MS signals from histones as an internal standard to estimate total protein amounts and total MS signal to estimate the abundance of individual proteins

Selected Reaction Monitoring (SRM) coupled with spiking of isotopically-labeled AQUA peptides provides absolute quantification of proteins

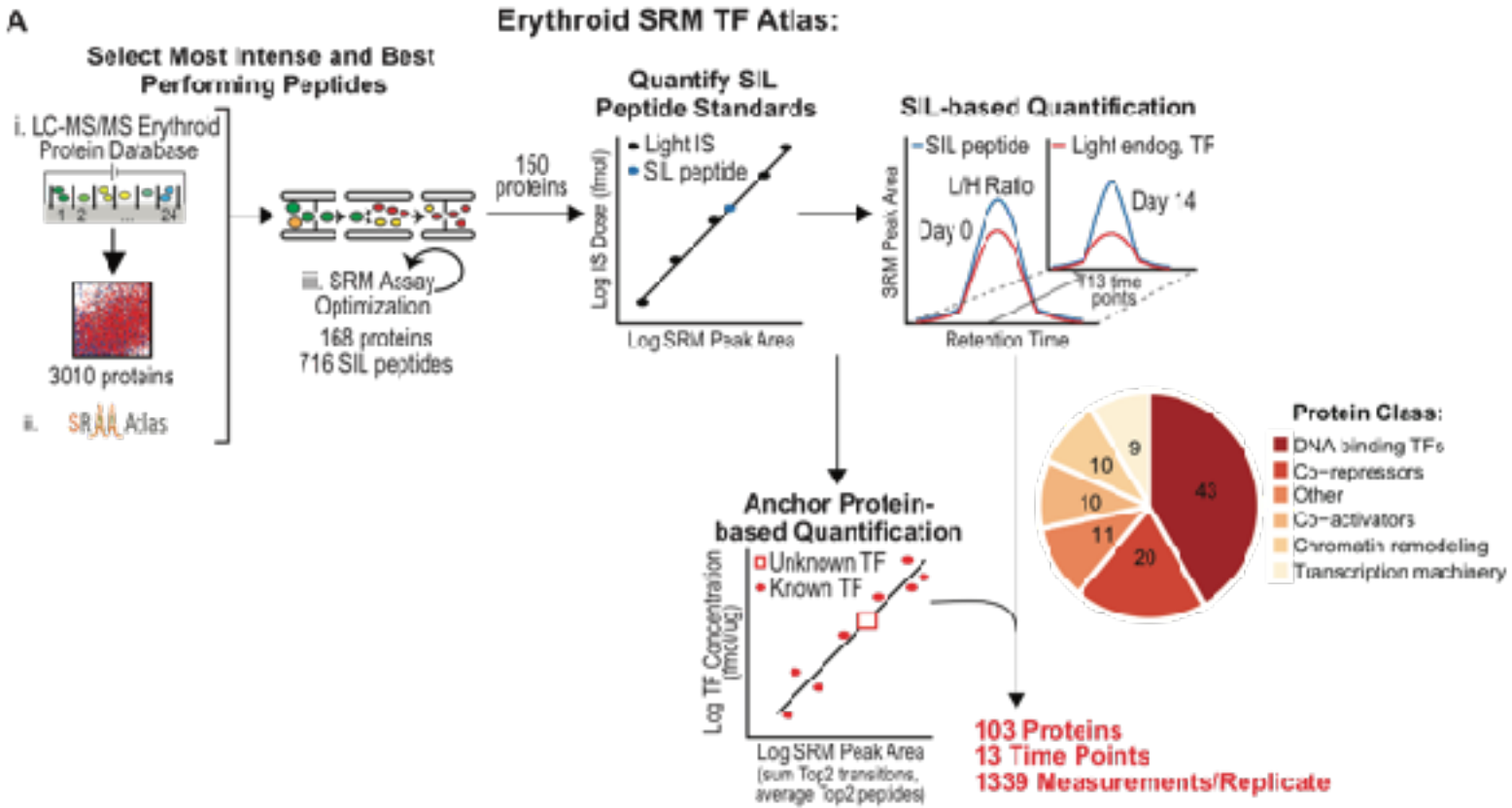


<u>For absolute quantification</u>:

-> spike-in of known amount of isotopically labelled peptides to be used as internal controls

-> each peptide is quantified using an isotopically labelled version of itself (SIL peptide)

Absolute quantification of erythroid TFs using SRM

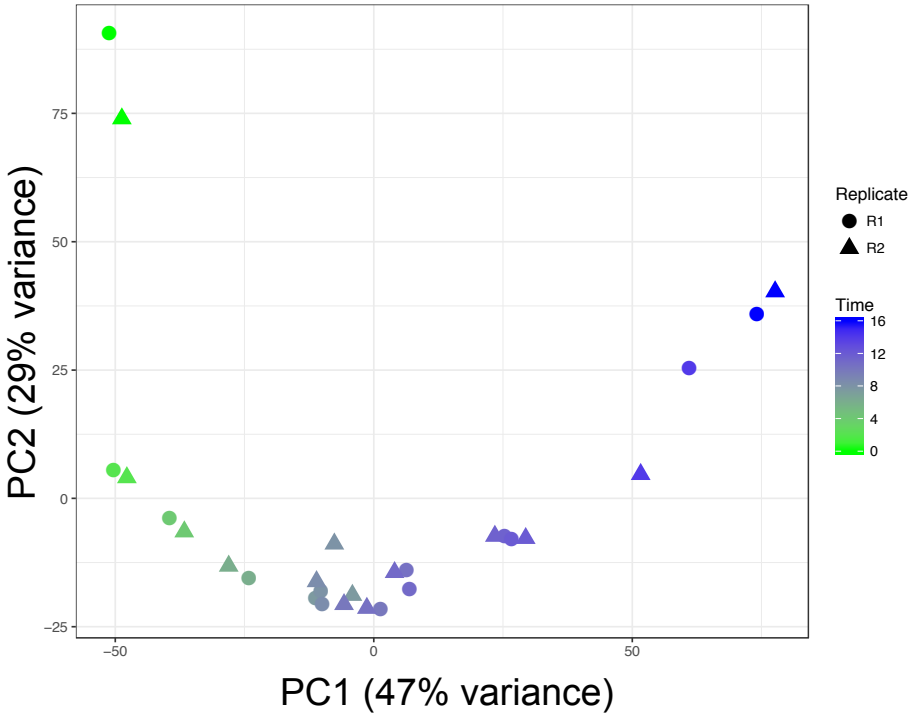


Gillespie et al. (2020) STAR Protocols 1(3): 100216

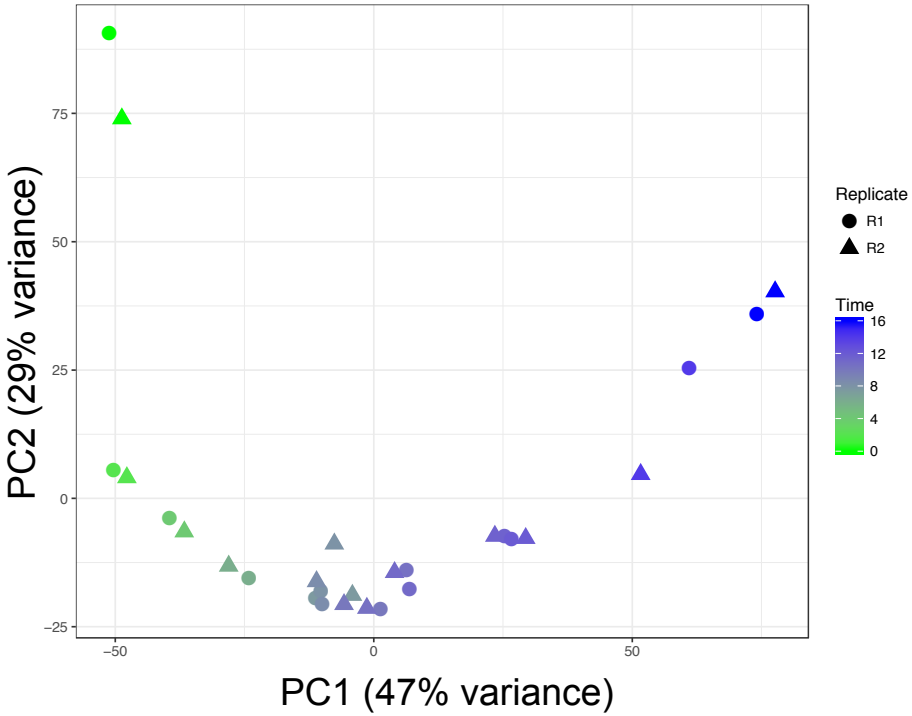
data + immediate observations

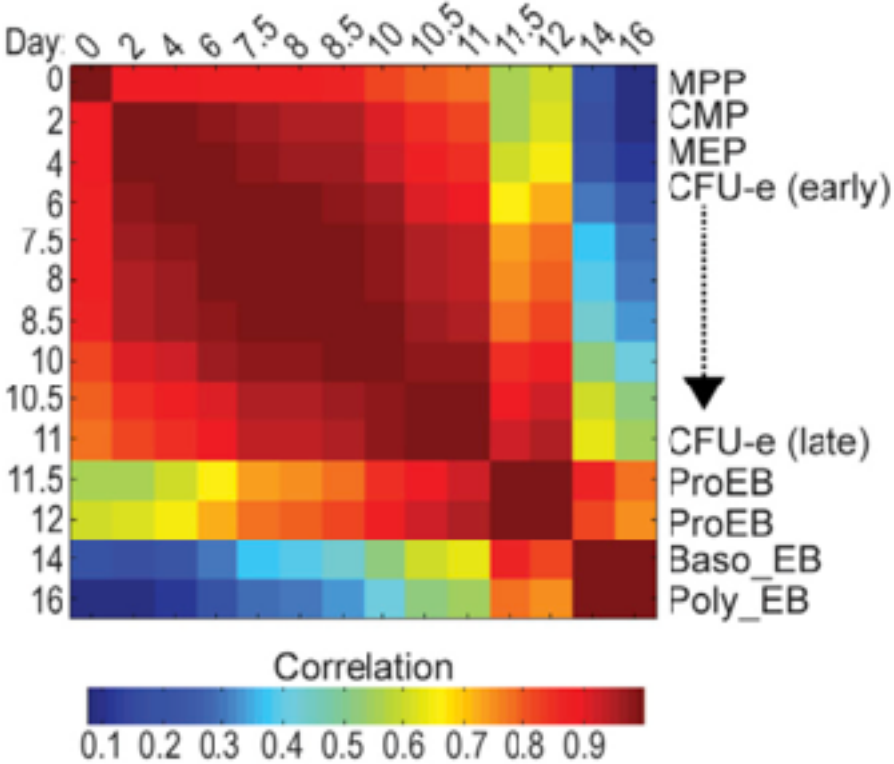
Interact with data at: https://hoodlab.shinyapps.io/tf-srm-rna/

mRNA visualization

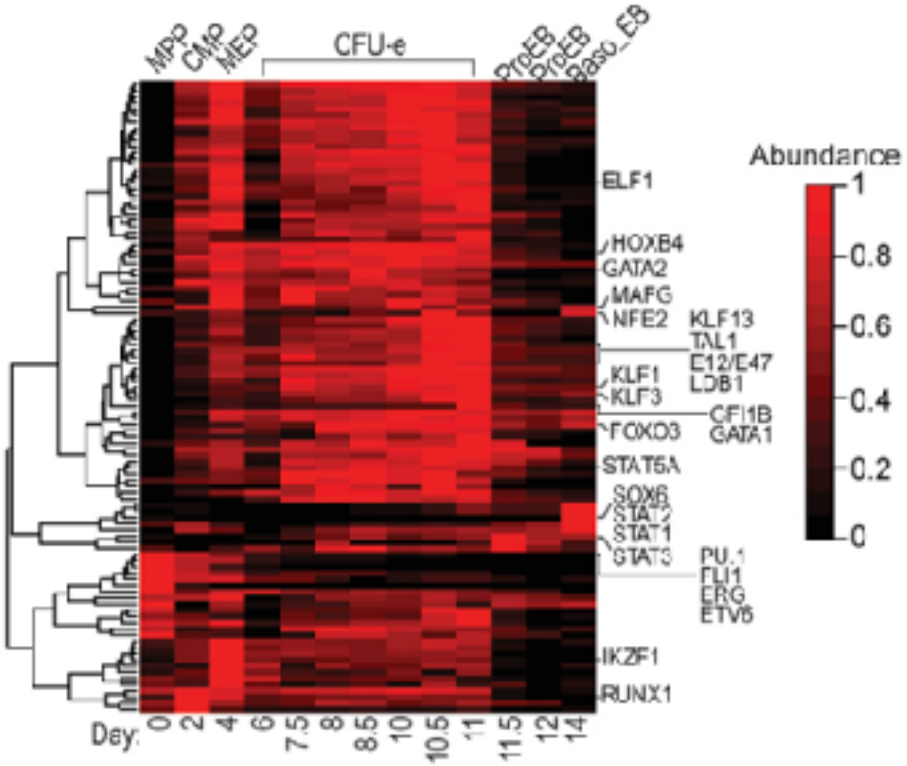


mRNA visualization

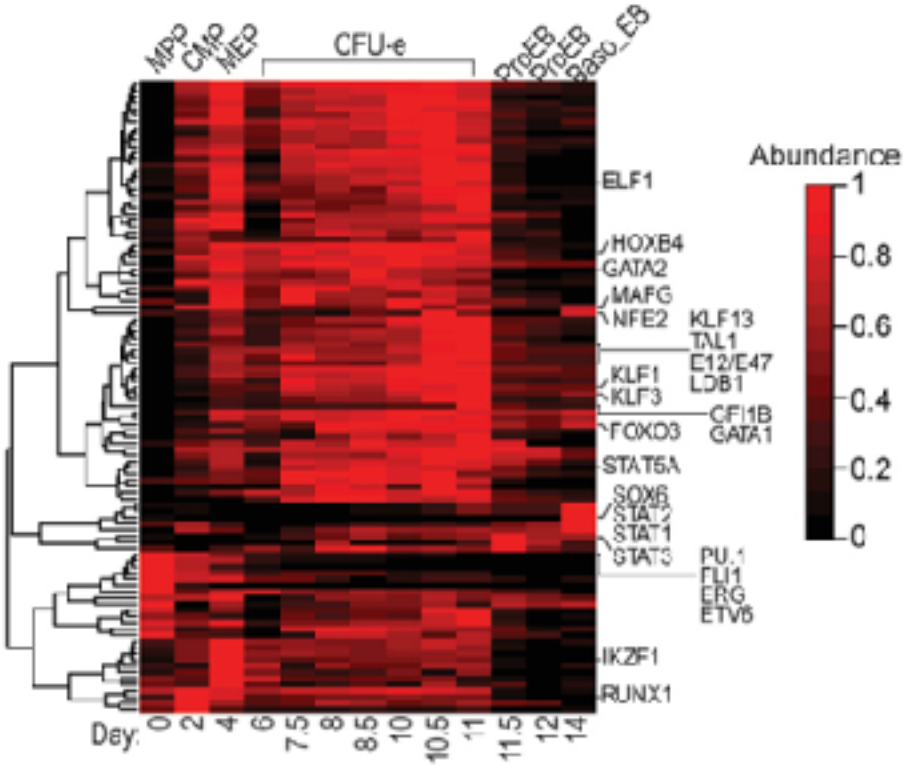


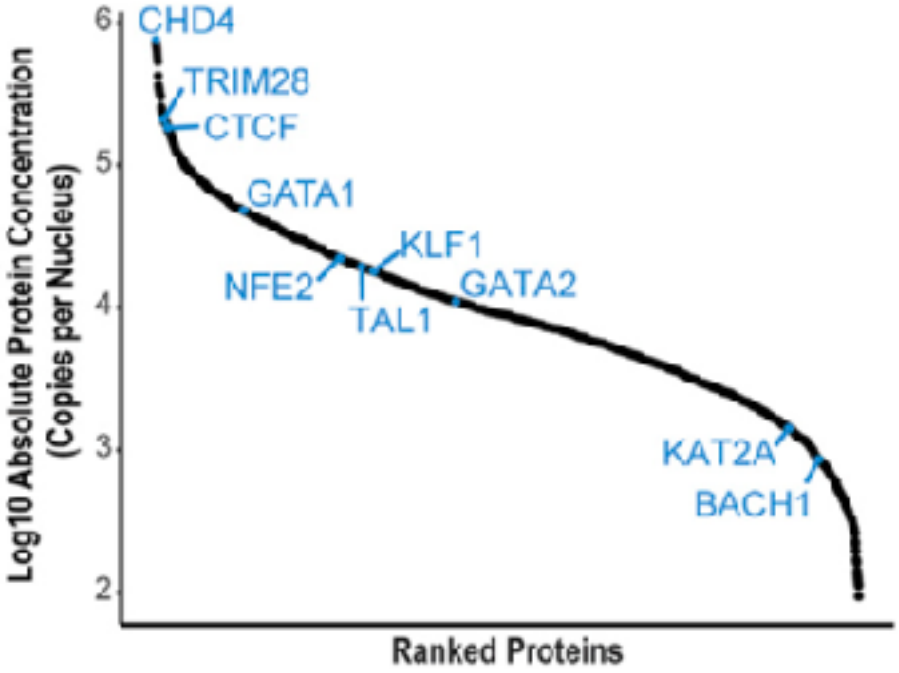


protein data

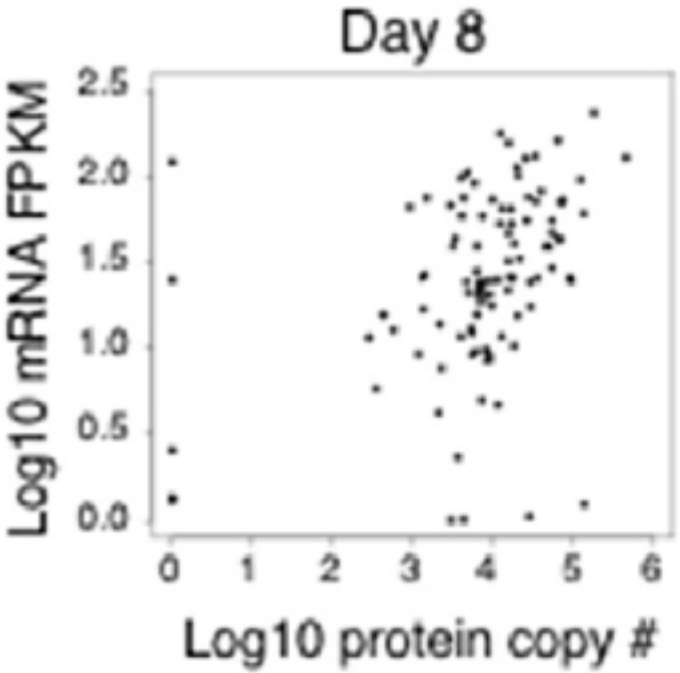


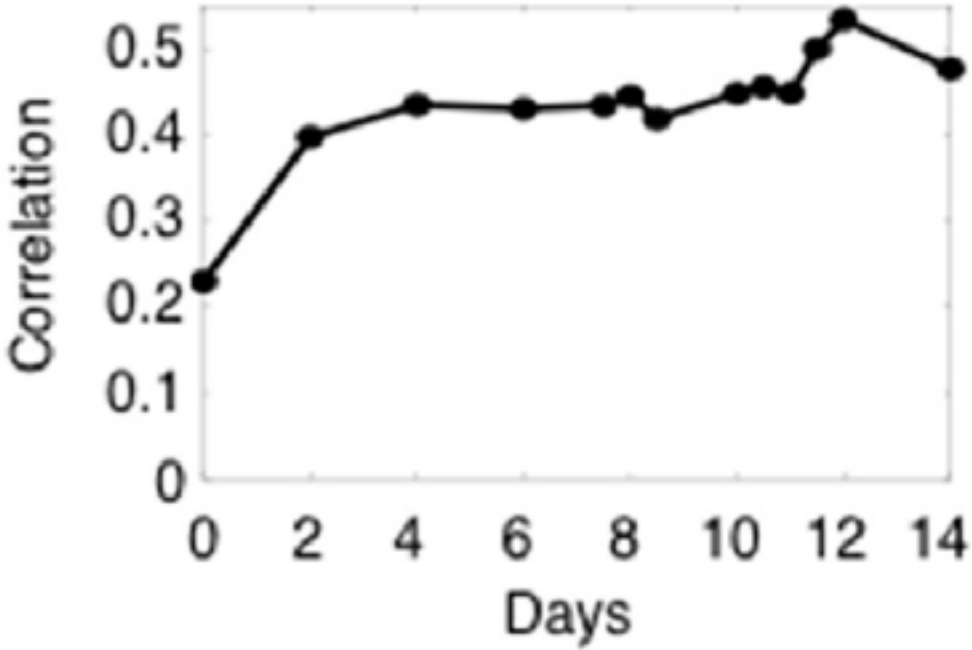
protein data



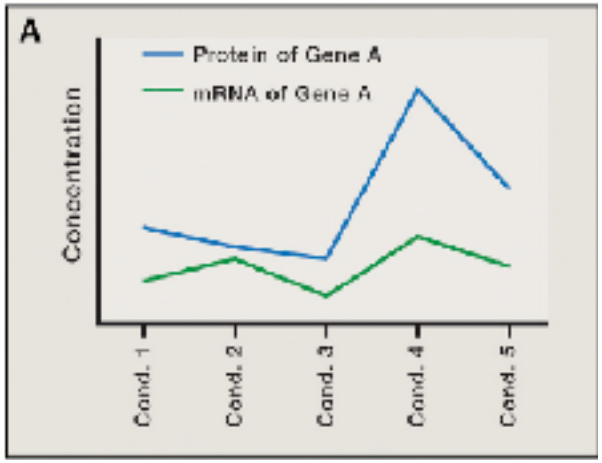


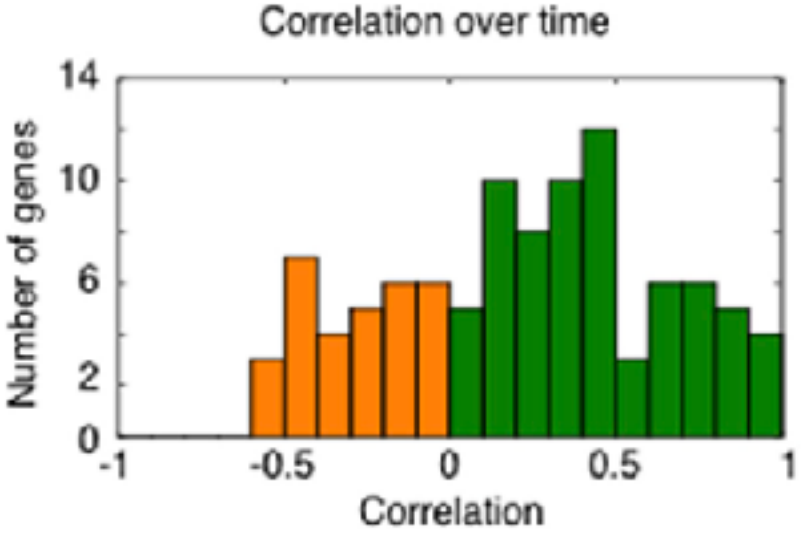
mRNA-protein correlation across genes at each timepoint

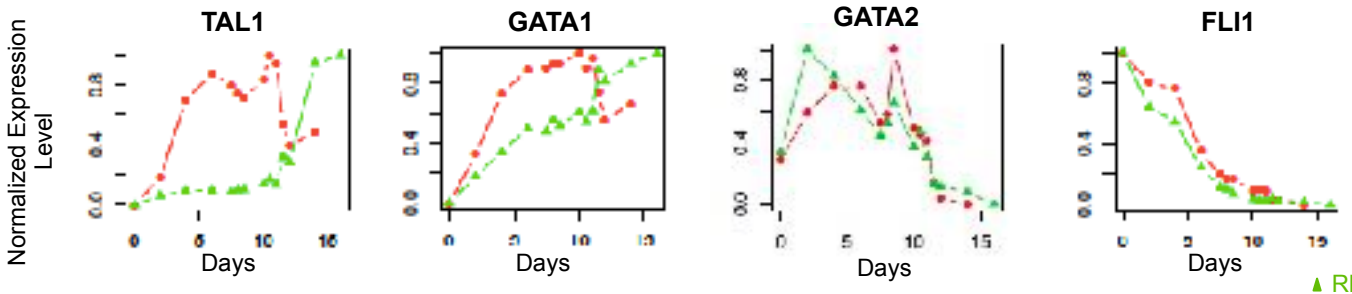




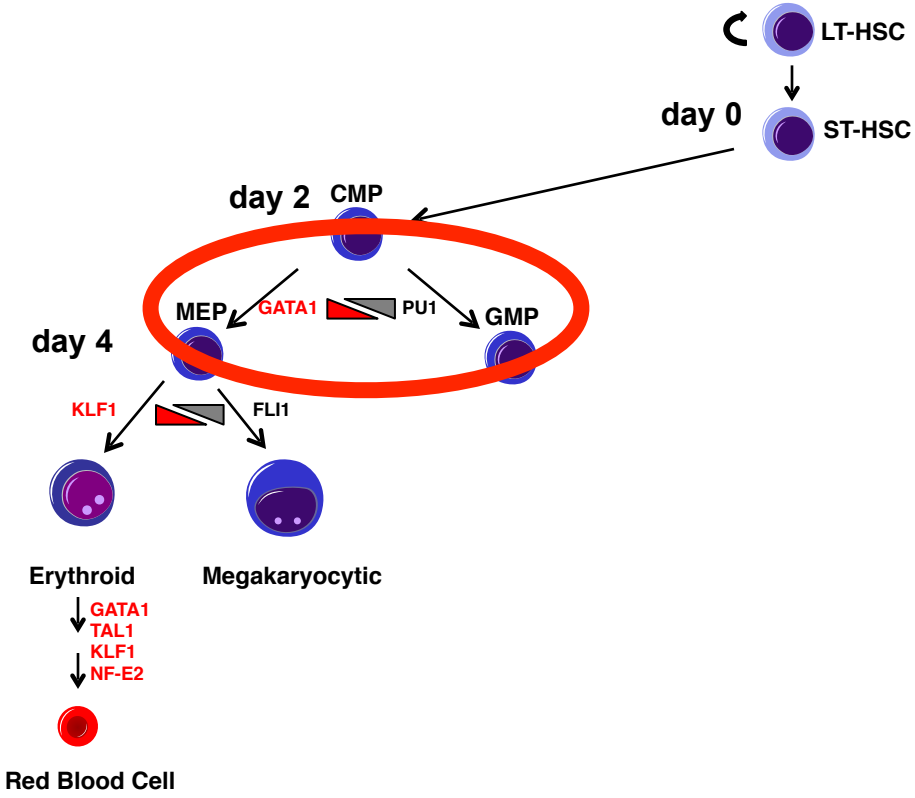
correlation between mRNA and protein across time

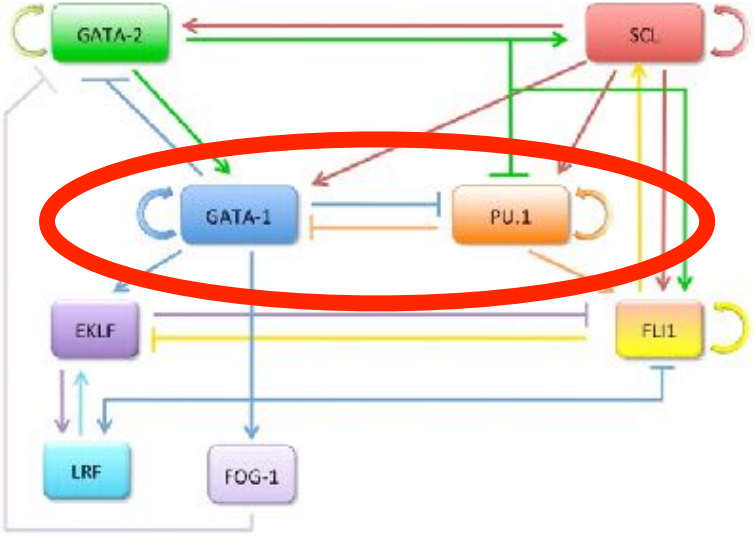




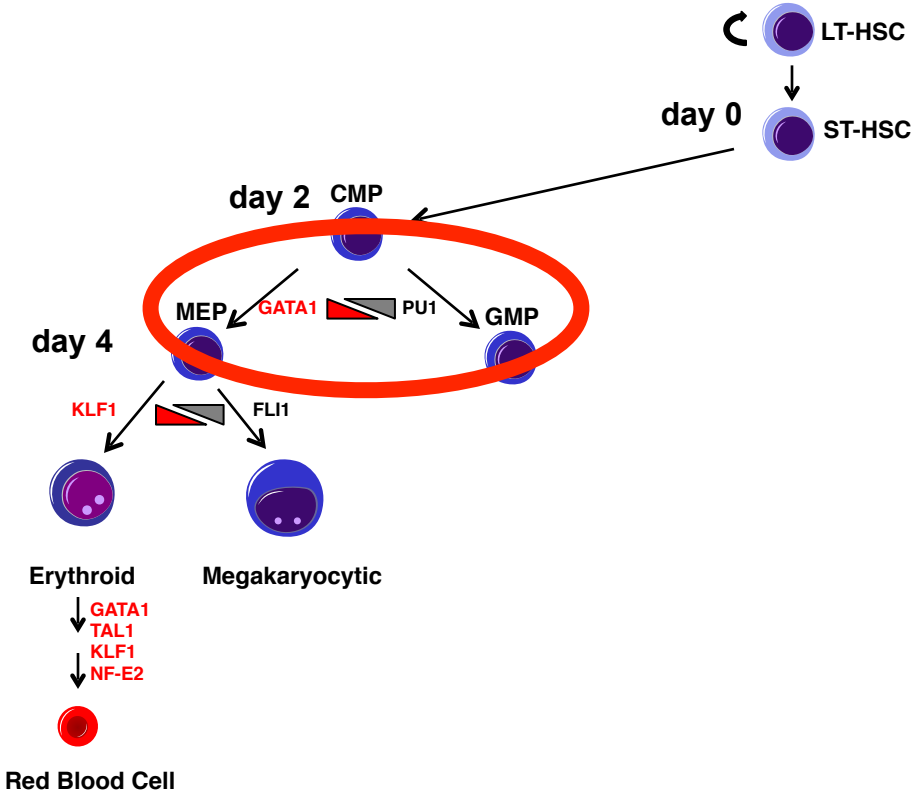


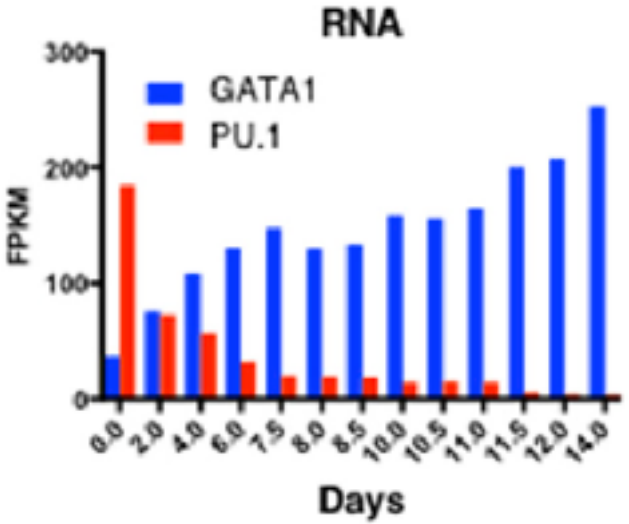
"unfair" competition in regulatory switches



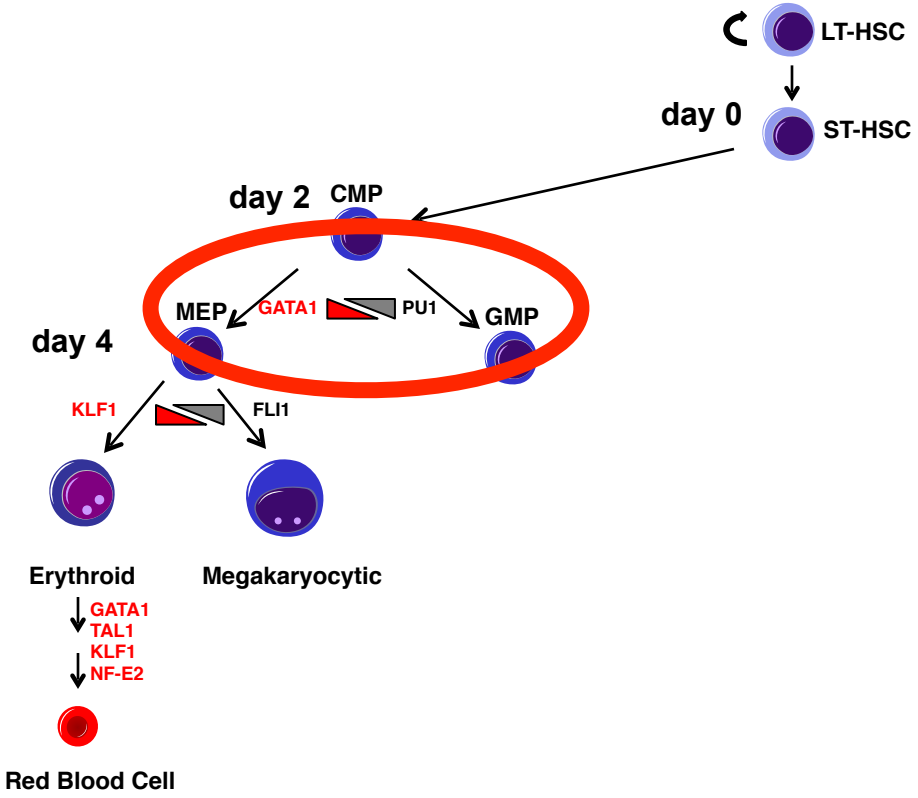


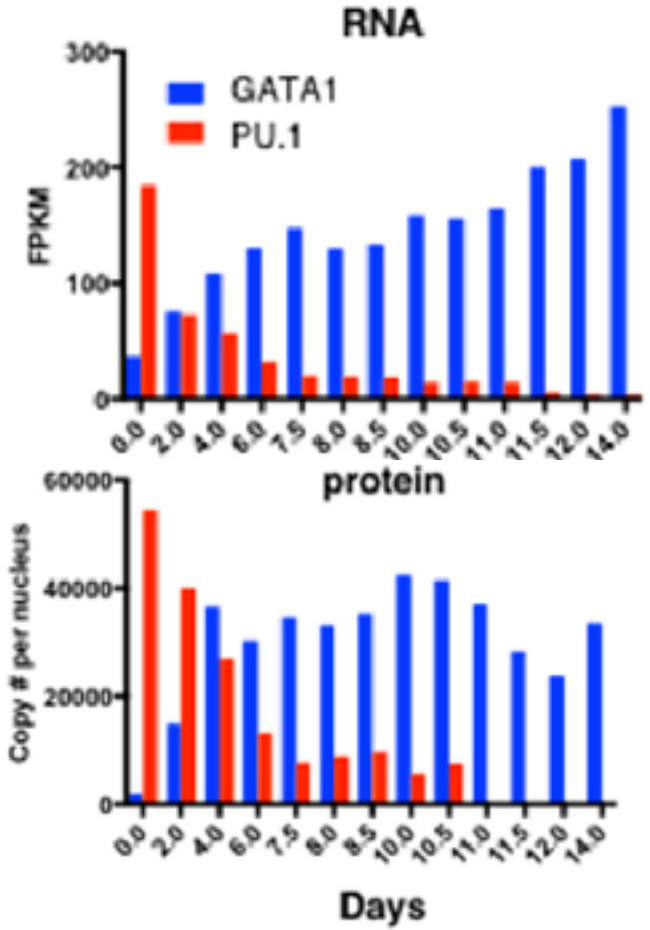
"unfair" competition in regulatory switches



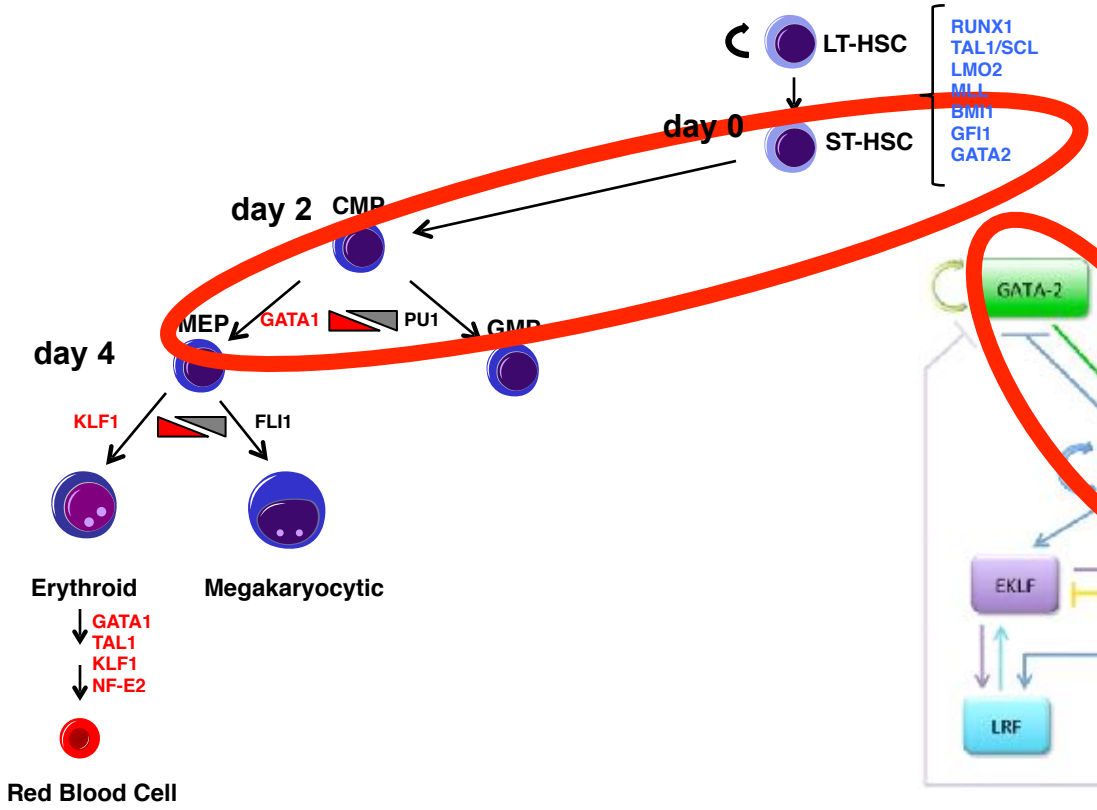


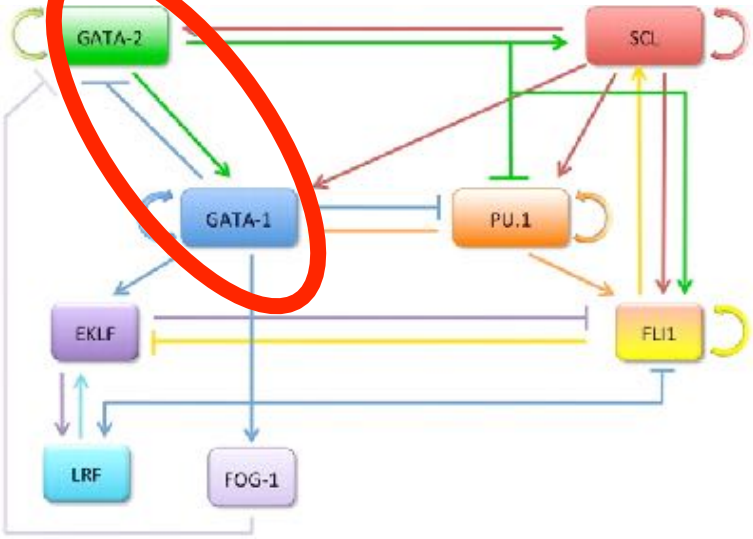
"unfair" competition in regulatory switches



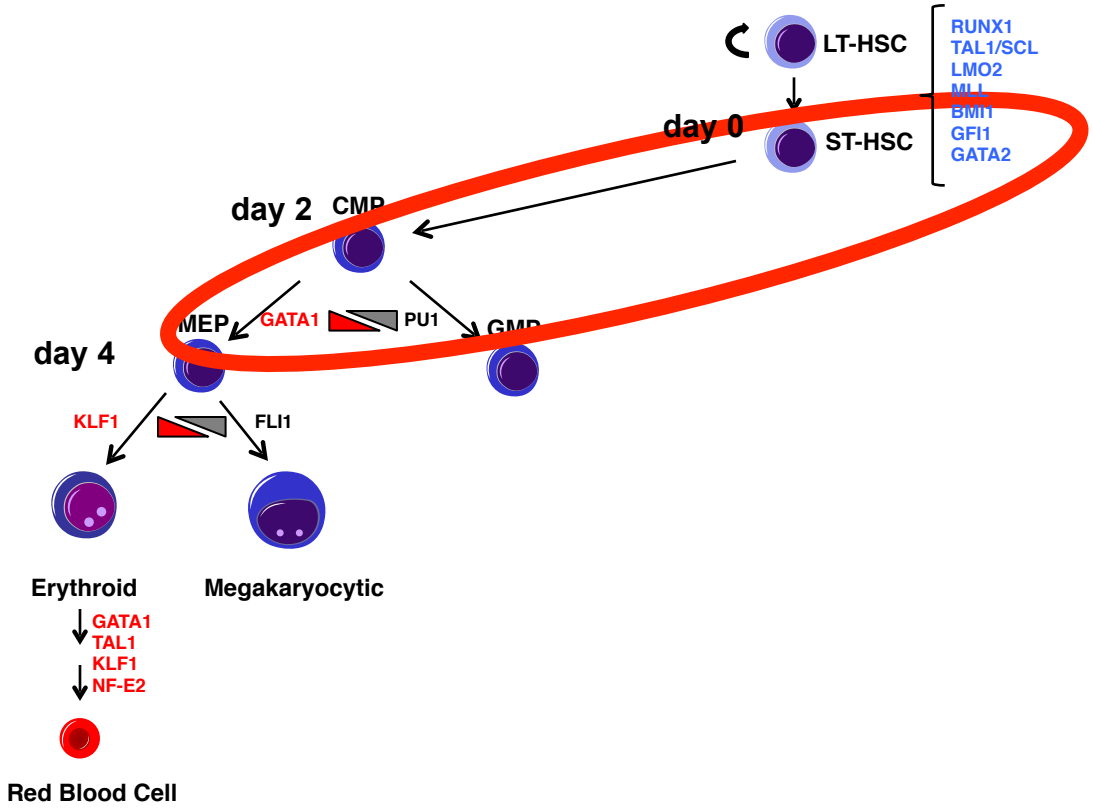


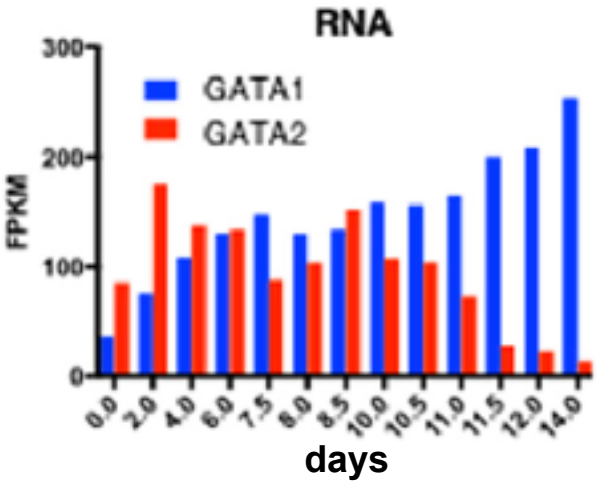
"unfair" competition in regulatory switches



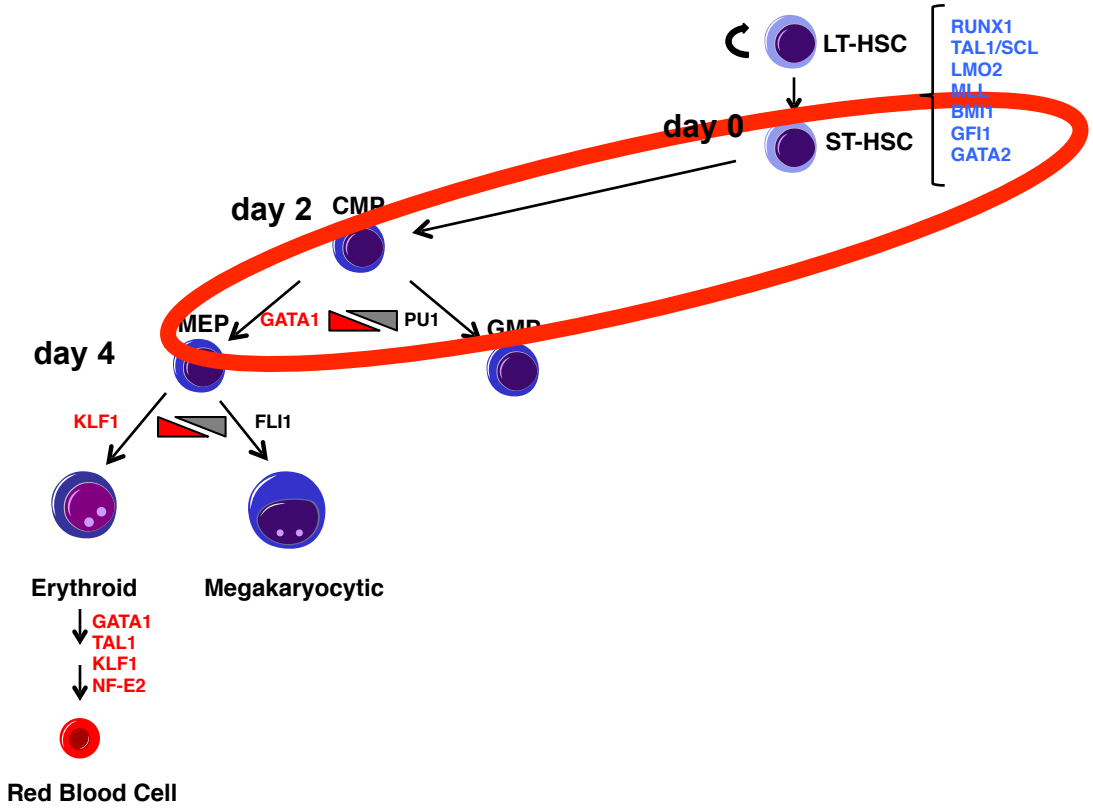


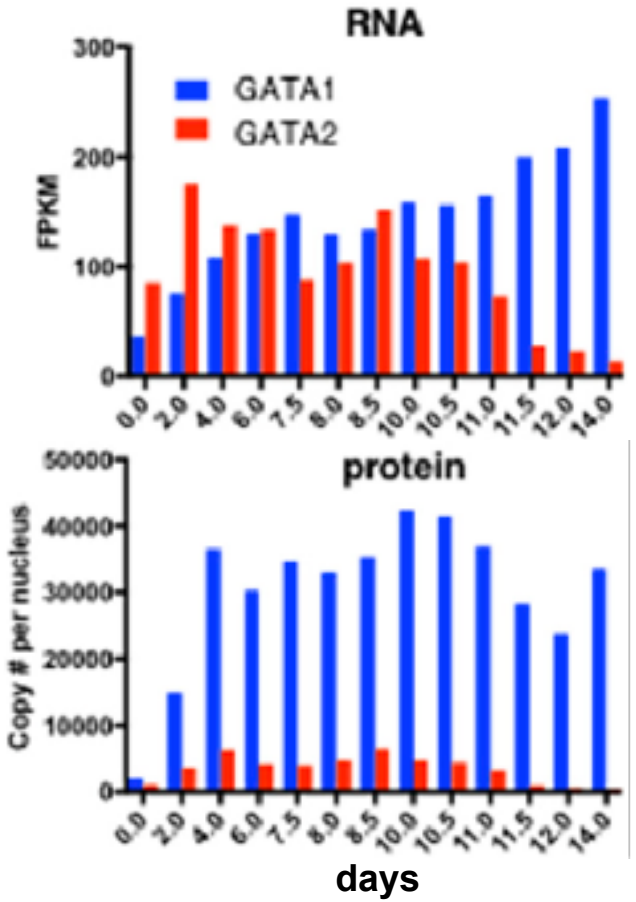
"unfair" competition in regulatory switches



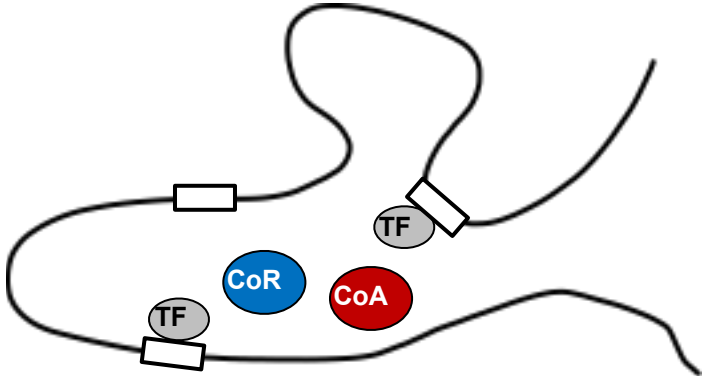


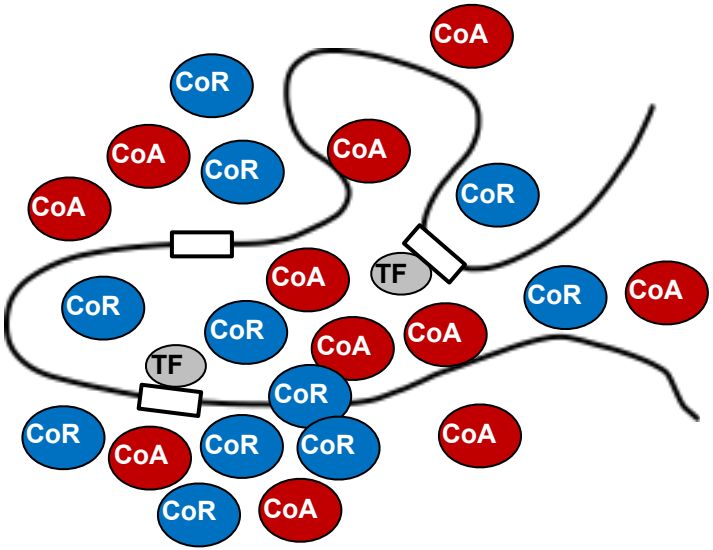
"unfair" competition in regulatory switches



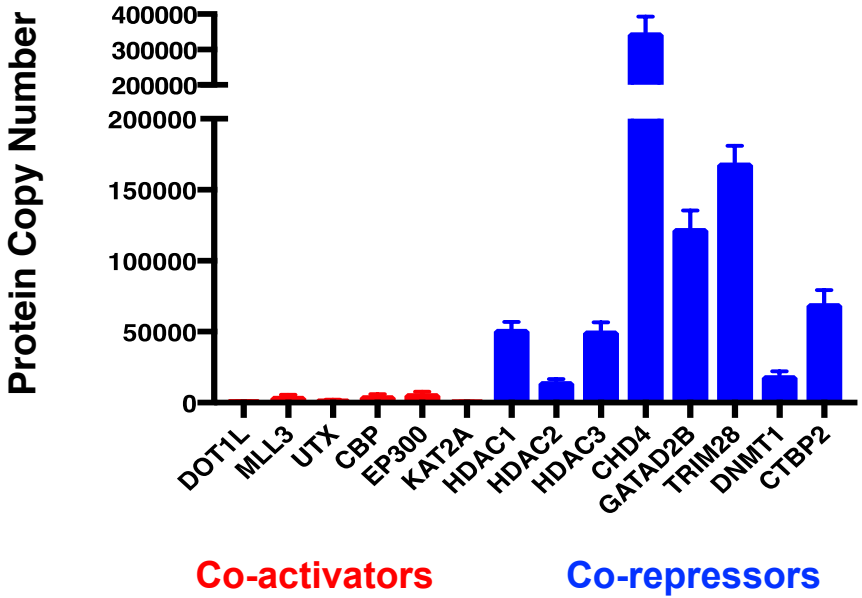


The stoichiometry between TFs and cofactors is unknown

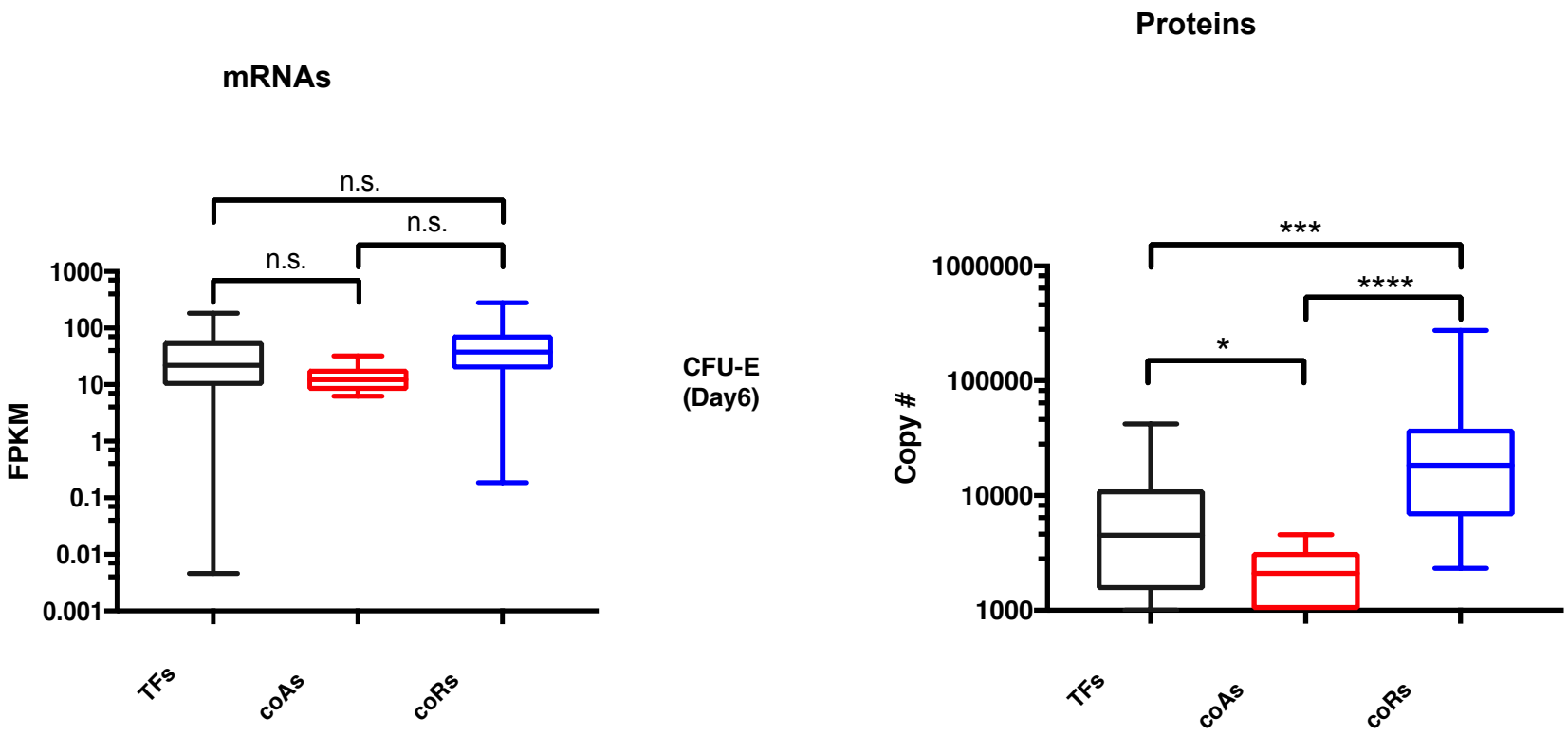




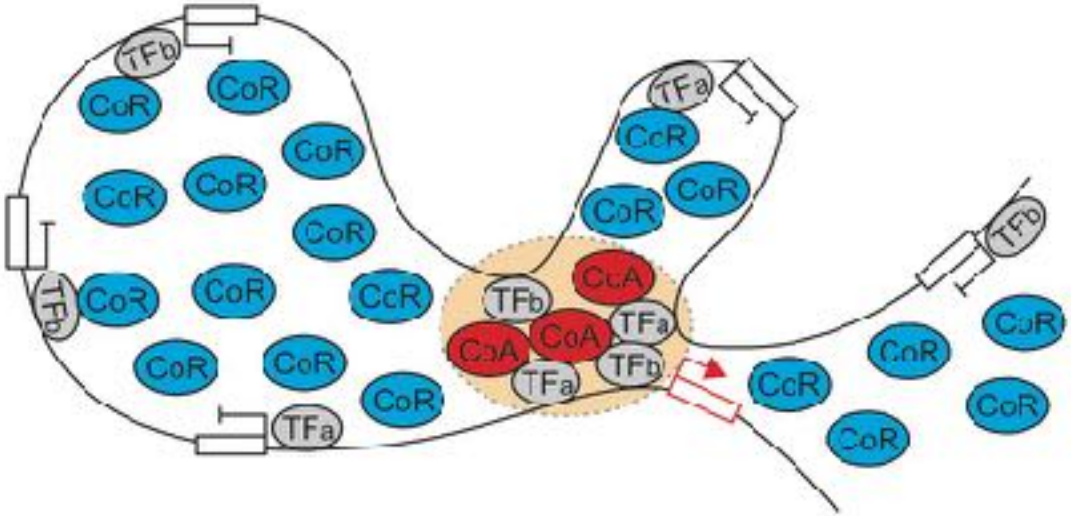
Co-activators are limiting compared to co-repressors



Co-activators are limiting compared to co-repressors



Model



Gillespie, Palii, Sanchez-Taltavull et al. (2020) Mol. Cell 78: 960-974

→ Restricting the abundance of co-activators in a highly repressive nuclear environment may be an important mechanism for concerted gene regulation

→ Important for the cell fate decision process by ensuring that only a limited number of genes can be expressed thereby preventing high level co-expression of genes from different lineages in multipotent progenitors

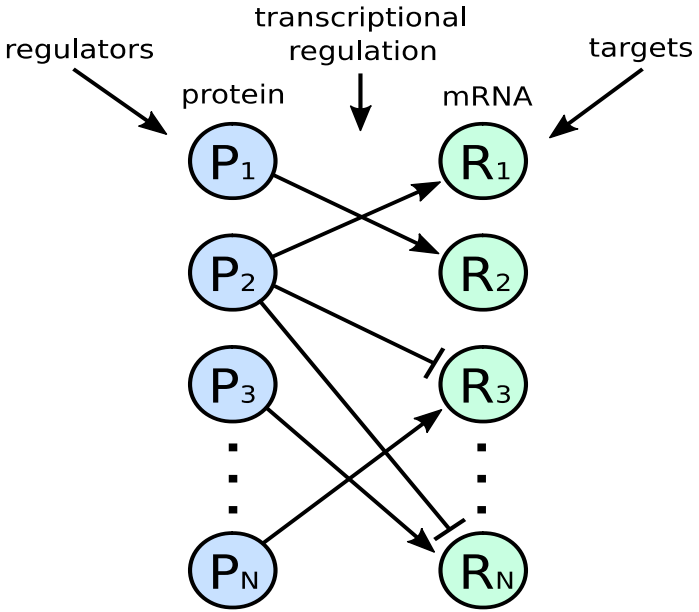
How do the numbers of active enhancers and co-activator molecules compare?

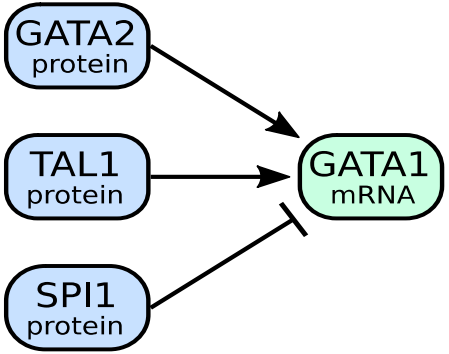
Day	Active Enhancers	CBP	P300	MLL3	MLL4
8	9,092	6,234	4,825	5,454	8,027
10	5,143	7,478	7,084	6,505	7,683
12	6,000	1,953	7,717	1,091	1,103

→ the formation of active enhancers in the nucleus may depend on the availability of co-activators molecules

→ **Major discrepancies between mRNA and protein abundances for master regulators of erythropoiesis suggest that gene regulatory networks should not be limited to mRNA but should integrate proteins**

Building a dynamic network model of erythroid commitment that incorporates quantitative changes in TFs protein levels over time

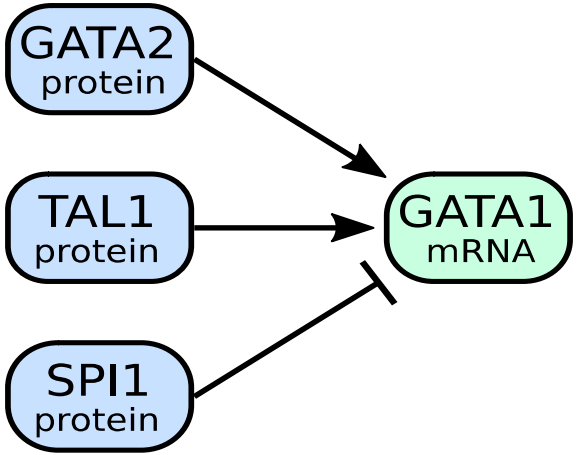




Model focuses on explaining transcriptional regulation: $dR_i/dt = f(P,\theta_i)$

Each gene separate. *Not* a closed loop model!

Building a dynamic network model of erythroid commitment that incorporates quantitative changes in TFs protein levels over time



$$\frac{dR_{GATA1}}{dt} = \frac{K_{GATA2}P_{GATA2} + K_{TAL1}P_{TAL1}}{1 + K_{SPI1}P_{SPI1}} - \lambda R_{GATA1}$$

R_{GATA1} = mRNA abundance of GATA1

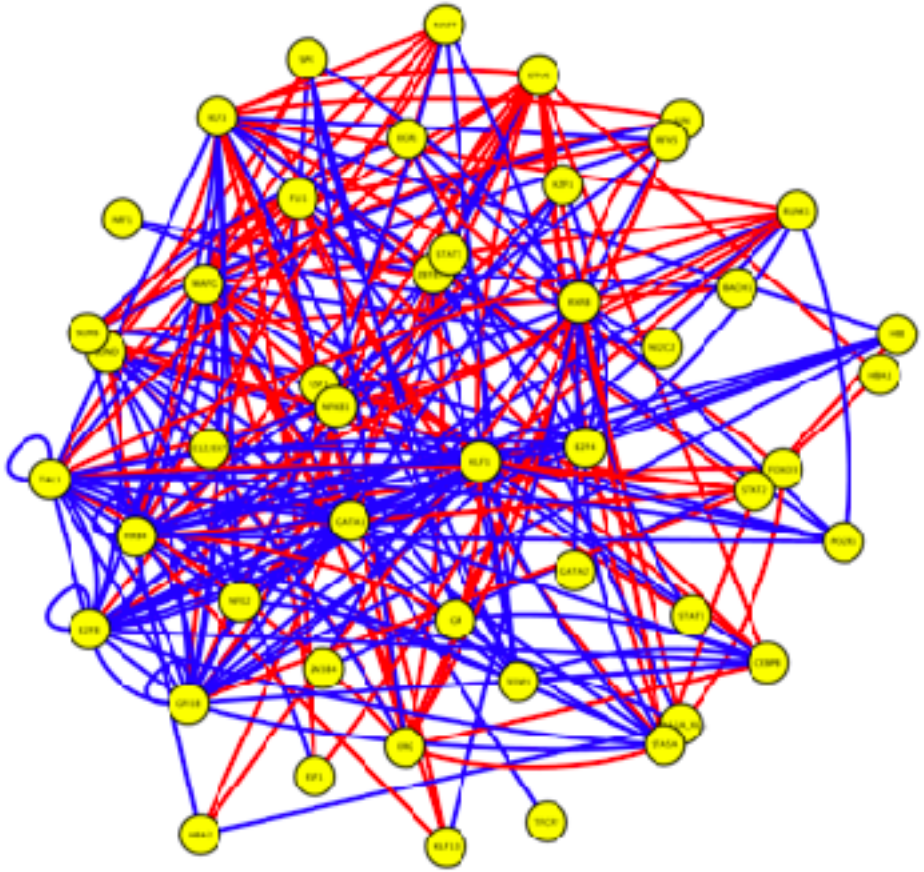
P_{GATA2}, P_{TAL1}, P_{SPI1} = protein abundance of GATA2, TAL1, SPI1

K_{GATA2}, K_{TAL1}, K_{SPI1} = regulatory parameters

λ = mRNA decay parameter

Model parameters optimized so simulated R matches observed R

network model? correlations (as usual) abound!

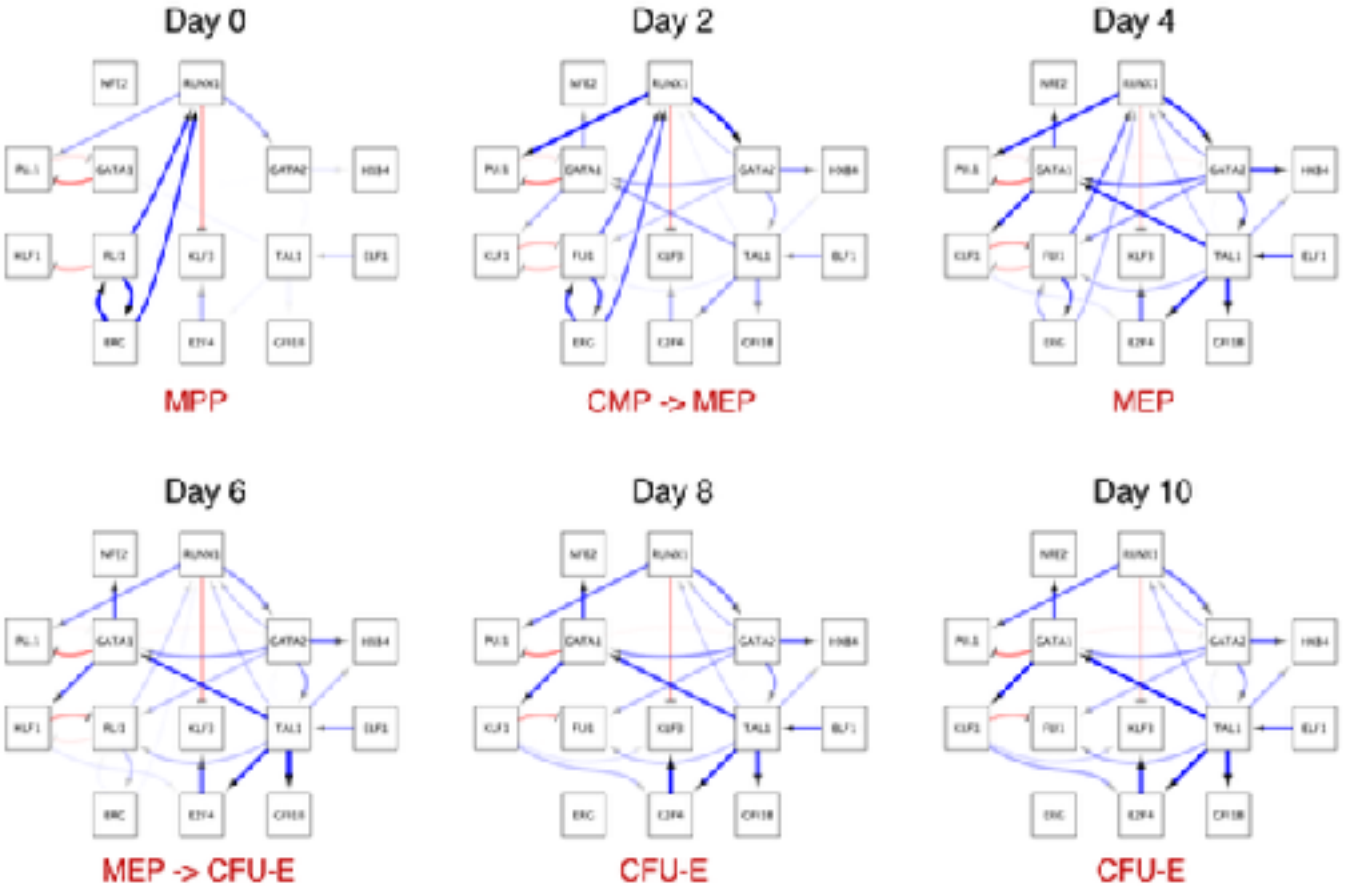


network model? prior knowledge + a few correlations

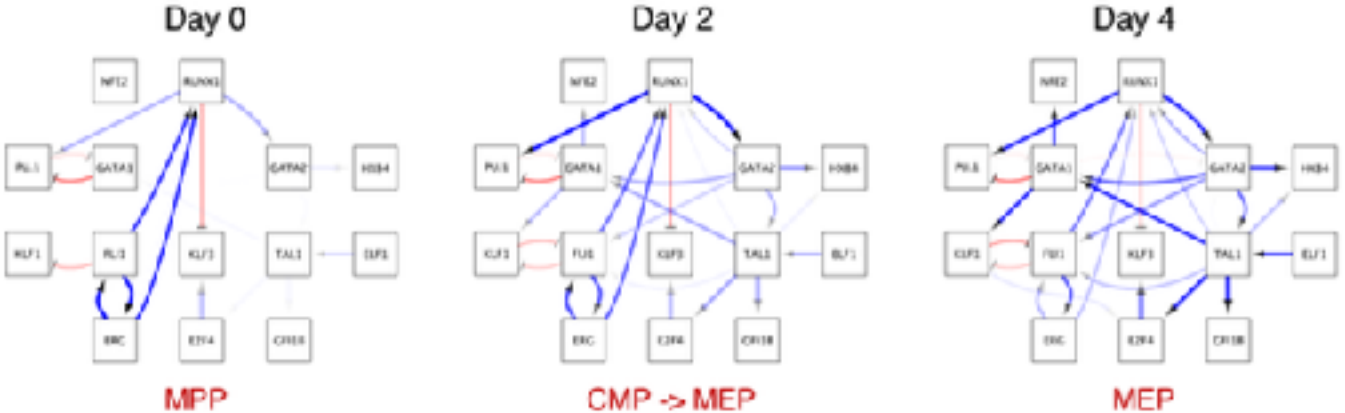
- We focused on "core" erythropoiesis TFs where regulatory links were known: ELF1, ERG, FLI1, GATA1, GATA2, GFI1B, KLF1, NFE2, RUNX1, TAL1, SPI1

- Added E2F4, HXB4, KLF3 with links to top 3 positive correlated and top 1 negative correlated genes (correlated means regulator protein to regulatee mRNA)

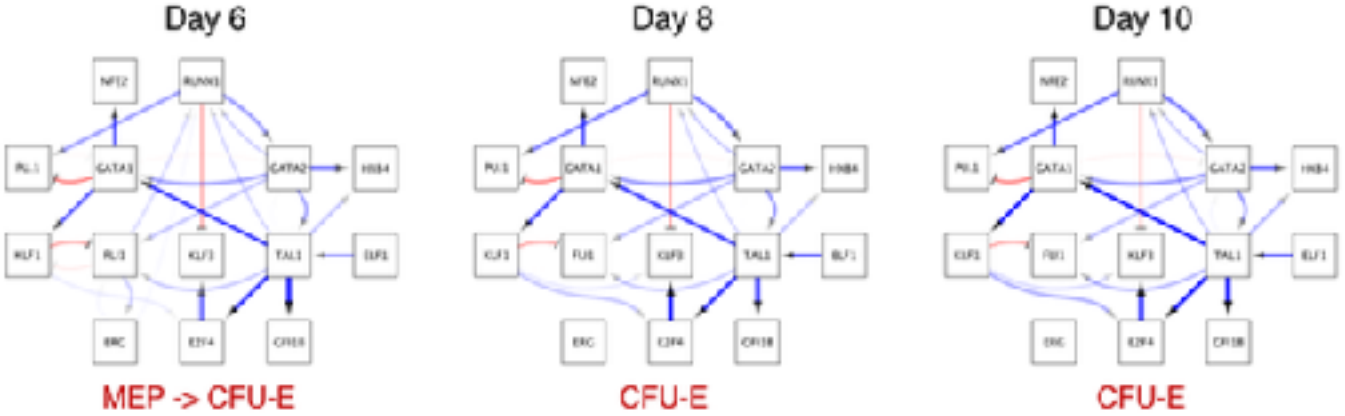
network model, with regulatory influence as function of time



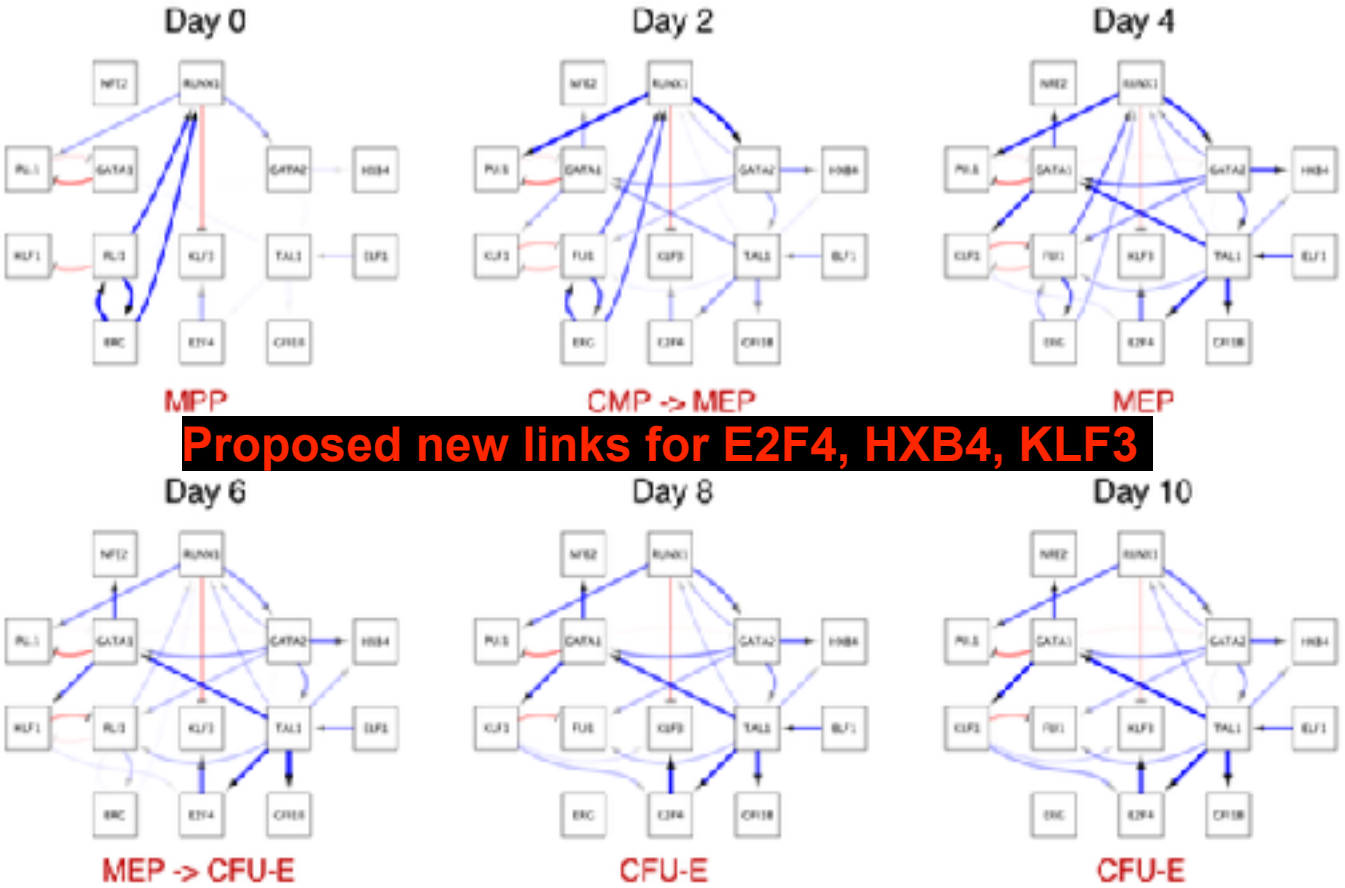
network model, with regulatory influence as function of time



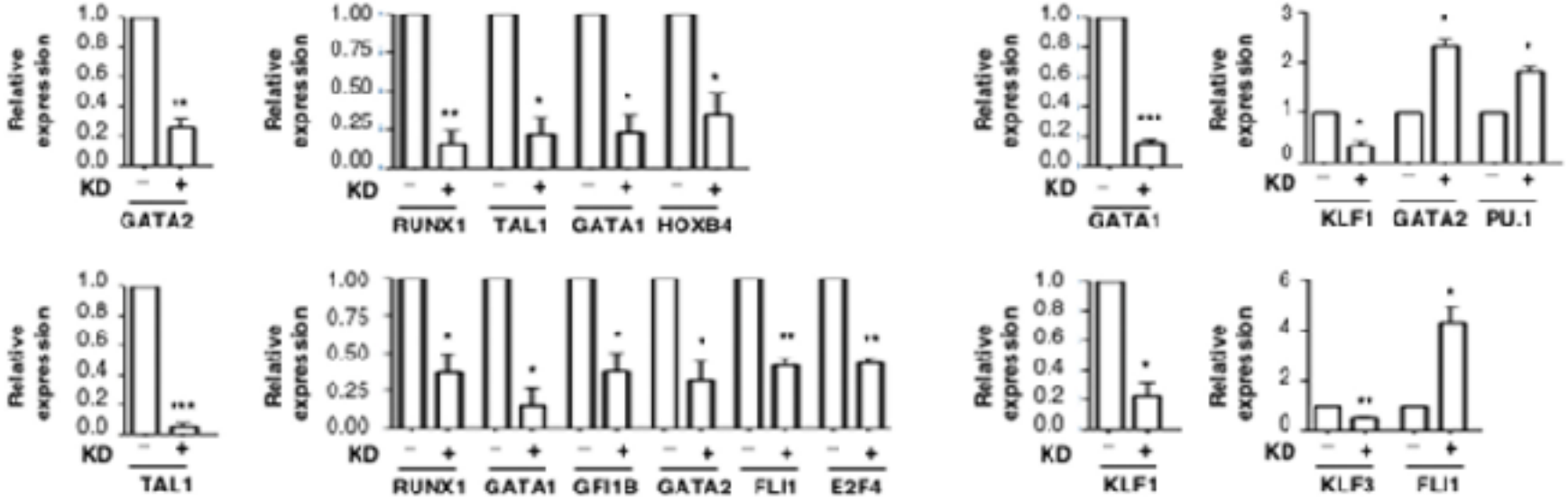
Timing of PU1 |--| GATA1and KLF1 |--| FLI1 approximately correct



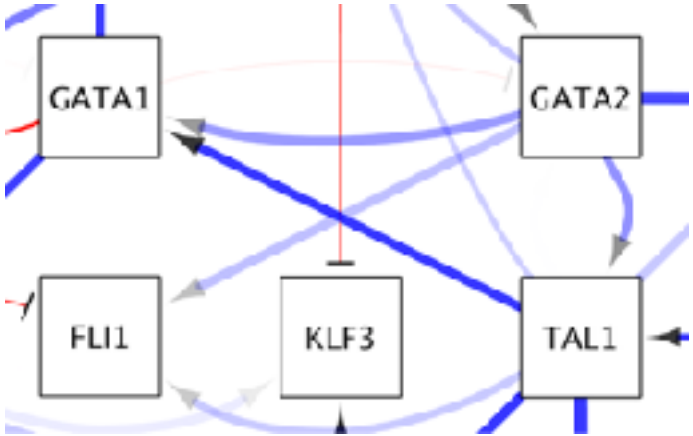
network model, with regulatory influence as function of time



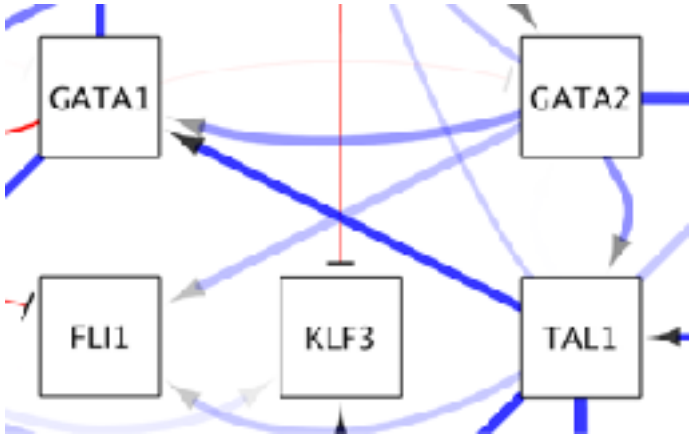
testing by knockdowns

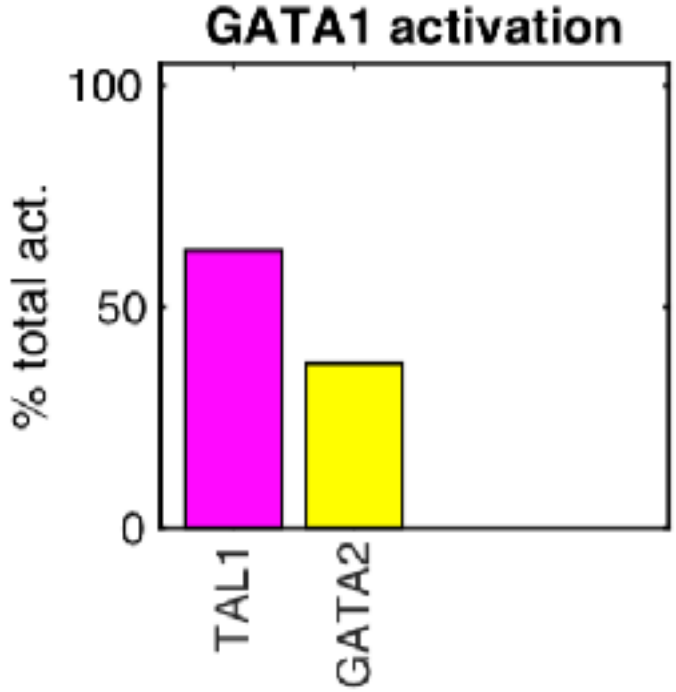


decomposing regulatory contributions



decomposing regulatory contributions

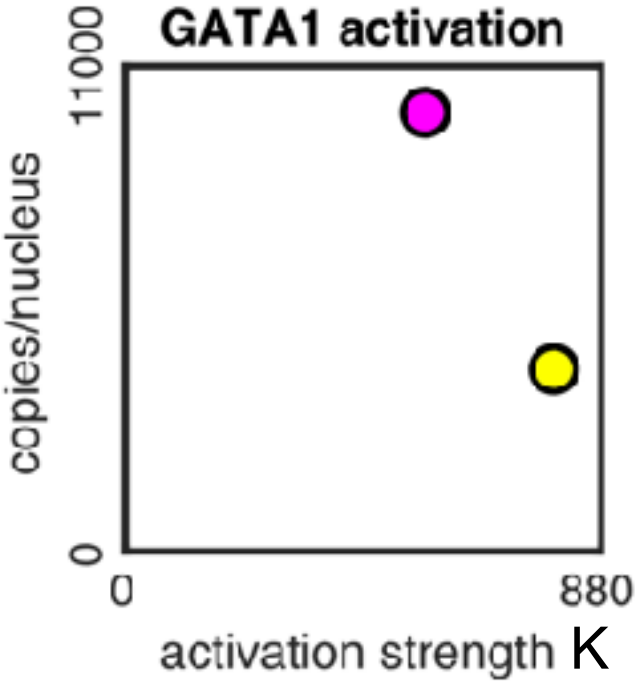


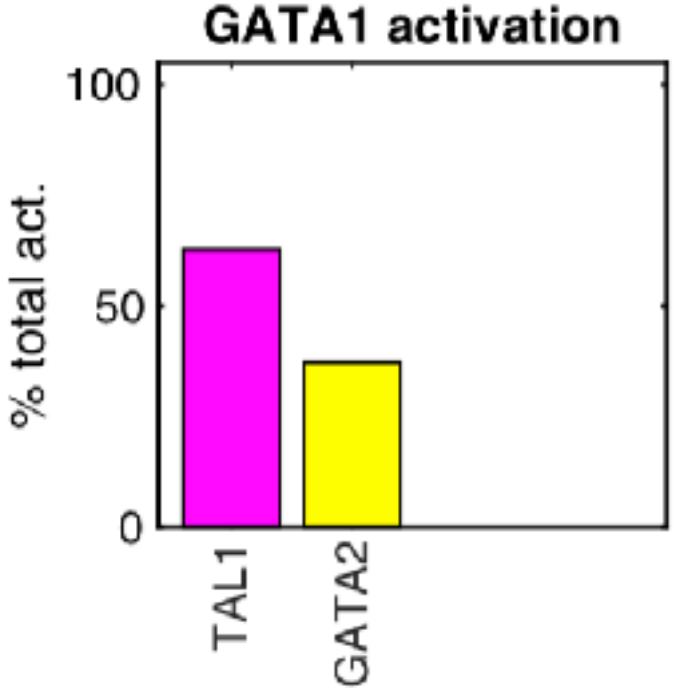


total act X → Y =

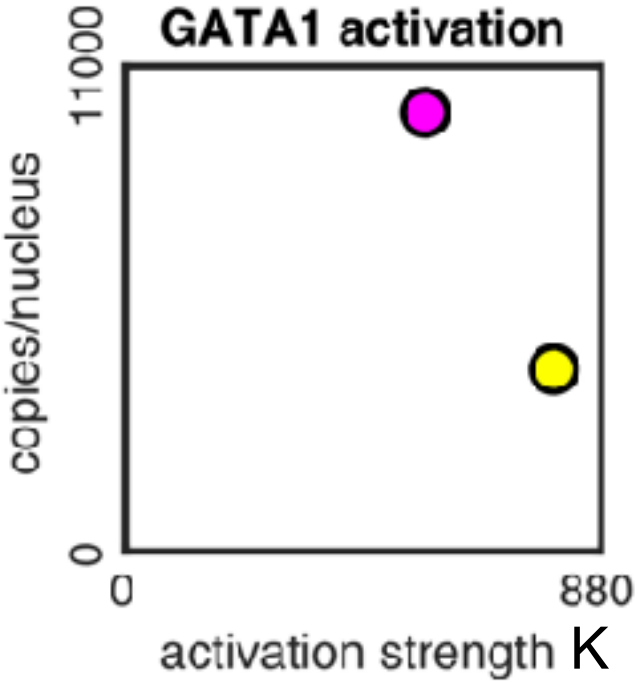
$$\int_{t-0}^{T} K_{X,Y} P_X$$

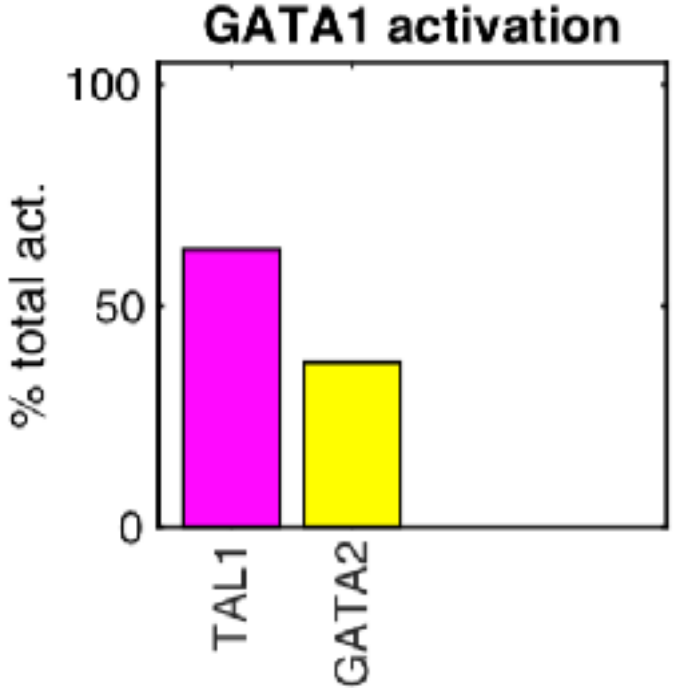
decomposing regulatory contributions





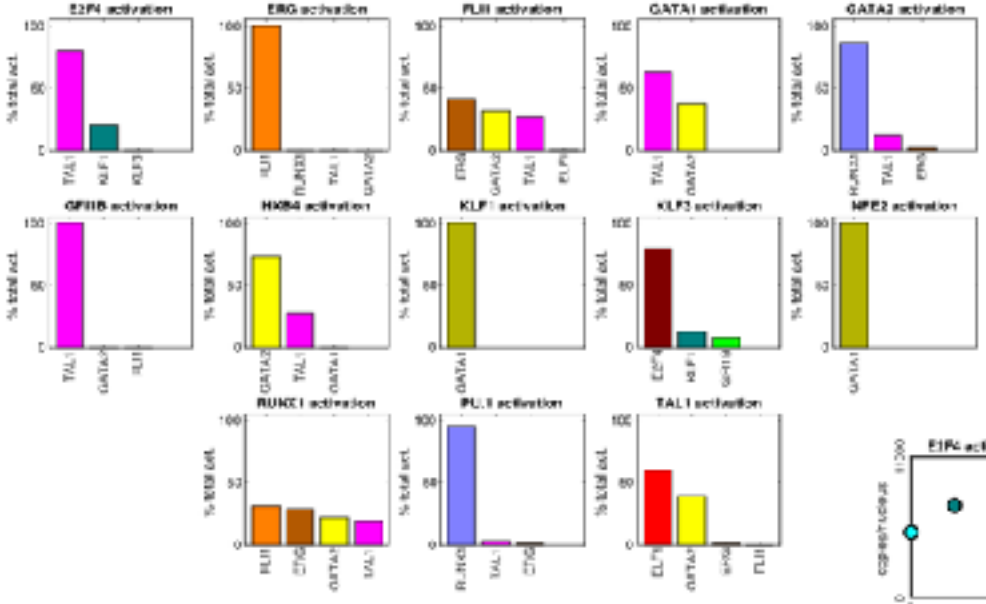
decomposing regulatory contributions

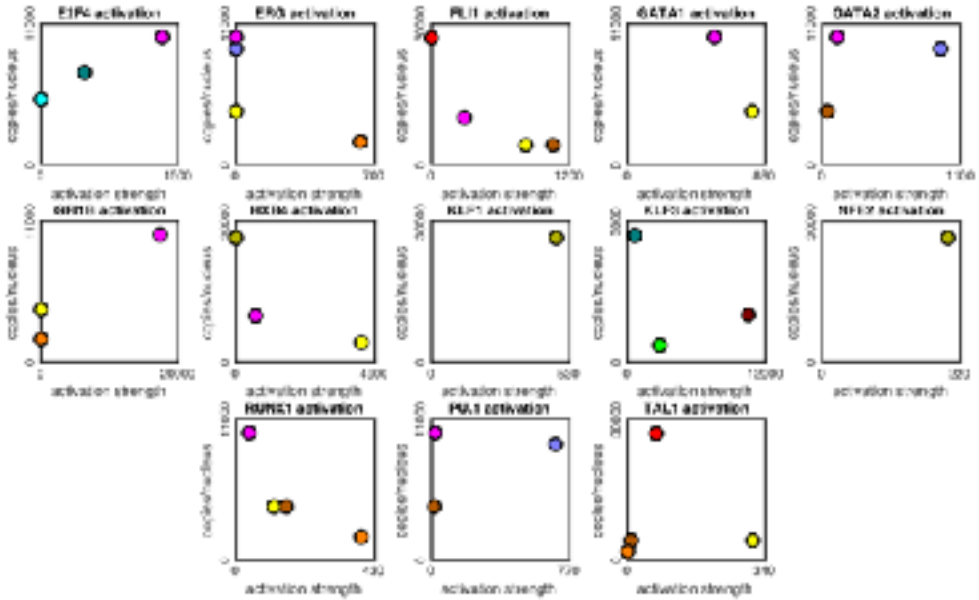




Tal1 influence achieved by more molecules at weaker per-molecule influence!

decomposing regulatory contributions





Conclusions

- Most gene expression measurements are not absolute, but we can make them absolute.
  - TF abundances differ by orders of magnitude
  - Co-activators and co-repressors even more so.
- mRNA and protein levels not always correlated across genes and across time
- We can focus on transcriptional regulation by modeling RNA as function of protein
  - Can "decompose" regulation into abundance and regulatory strength per molecule
  - Computationally efficient, because each gene modeled separately

Acknowledgements

Coauthors

Mark A Gillespie
Carmen G Palli
Damiel Sanchez-Taltavull
Paul Shannon
William J R Longabaugh
Damien J Downes
Karthi Sivaraman
Herbert M Espinoza
Jim R Hughes
Nathan D Price
Theodore J Perkins
Jeffrey A Rainish
Marjorie Brand

Perkins lab

Aseel Awdeh
R Matt Tanner
Justin Chitpin
Aarthie Senathirajah
Soroush Fard
Xun Xun Shi
Renad Al-Ghazawi

Funding

NIH National Institutes of Health

GenomeCanada

NSERC CRSNG

CIHR IRSC

Abbreviation cheat sheet

HSC = hematopoietic stem cell
LT-HSC = long-term HSC
ST-HSC = short-term HSC = MPP = multi-potent progenitor
CMP = common myeloid progenitor
MEP = megakaryocyte/erythrocyte progenitor
BFU-E = burst-forming unit-erythroid
CFU-E = colony-forming unit-erythroid
Pro EB = proerythroblast
Base EB = basophilic erythroblast
Poly EB = polychromatophilic
Ortho EB = orthochromatic erythroblast
Ret = reticulocytes
RBC = red blood cell