

A Simple Text Mining Approach for Ranking Pairwise Associations in Biomedical Applications

Finn Kuusisto, PhD¹; John Steill, MS¹; Zhaobin Kuang, MS²;
James Thomson, VMD, PhD^{1,2}; David Page, PhD²; Ron Stewart, PhD¹

¹Morgridge Institute for Research, Madison, WI

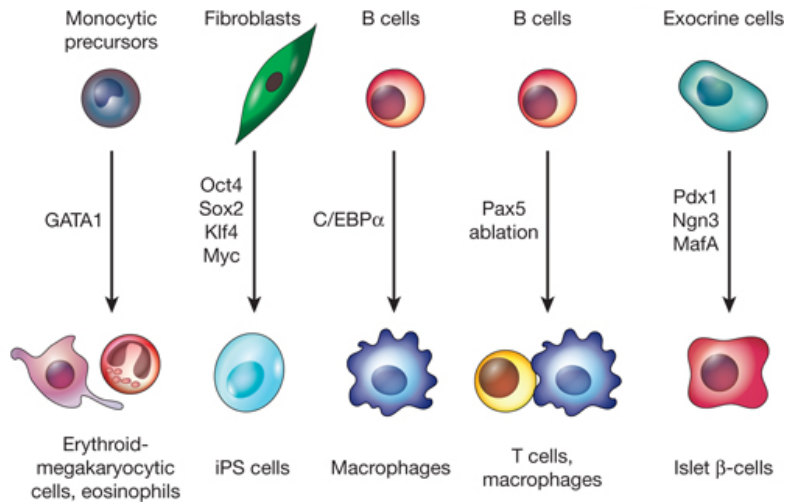
²University of Wisconsin, Madison, WI

Disclosure

Finn Kuusisto discloses that he has no relationships with relevant commercial interests.

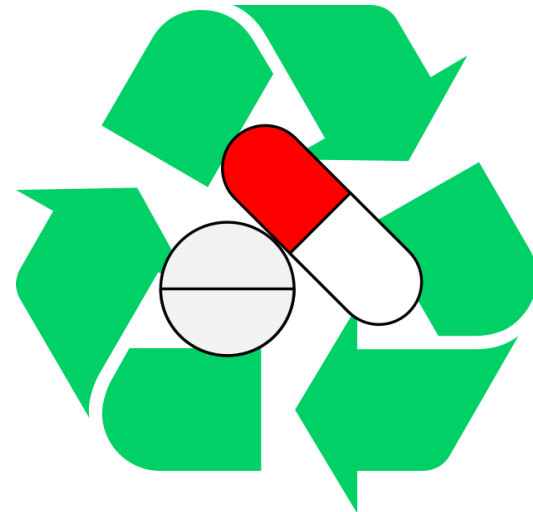
Arduous Scientific Discovery

Cell Reprogramming

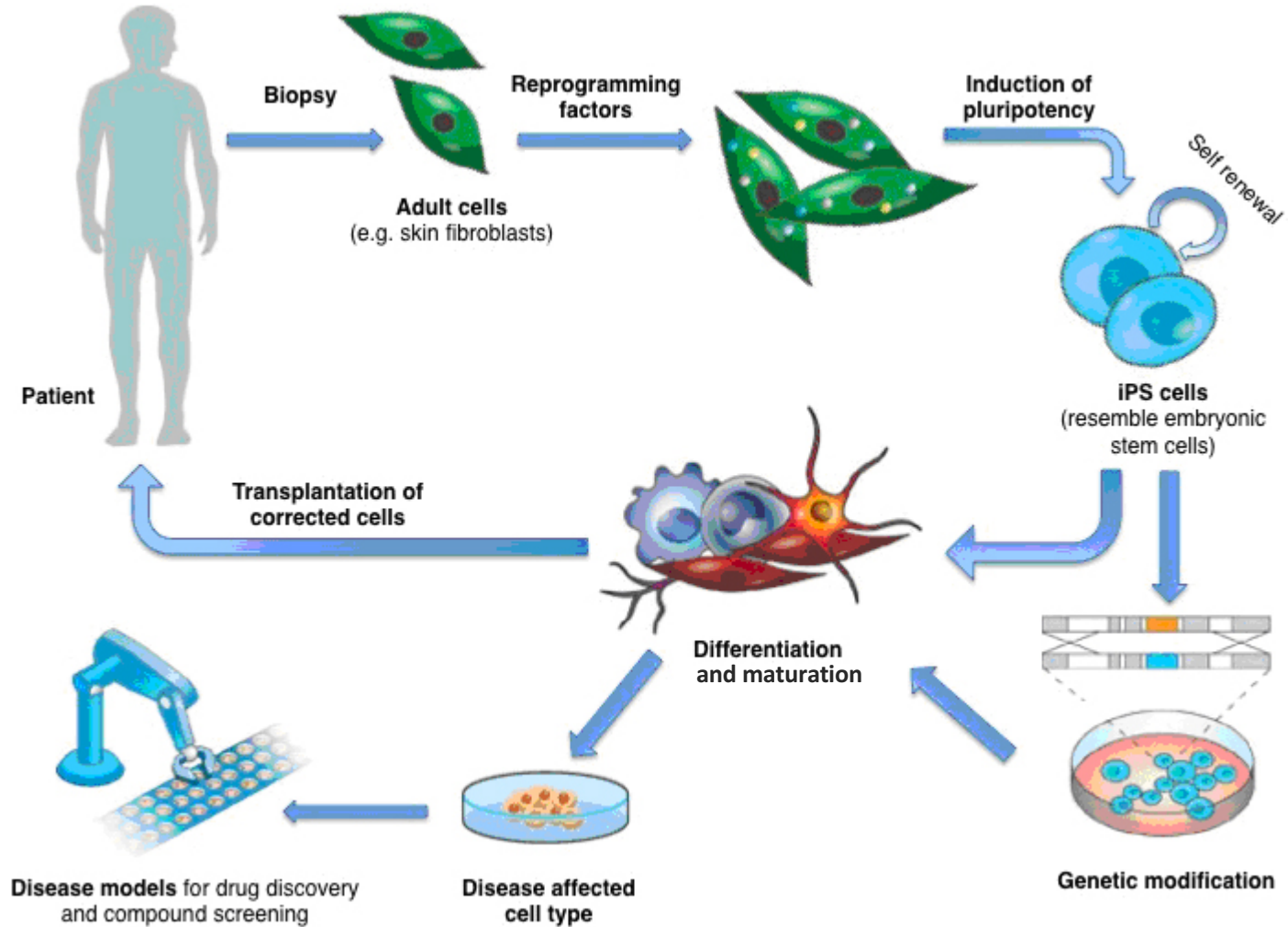


T. Graf and T. Enver. Forcing cells to change lineages. Nature (2009).

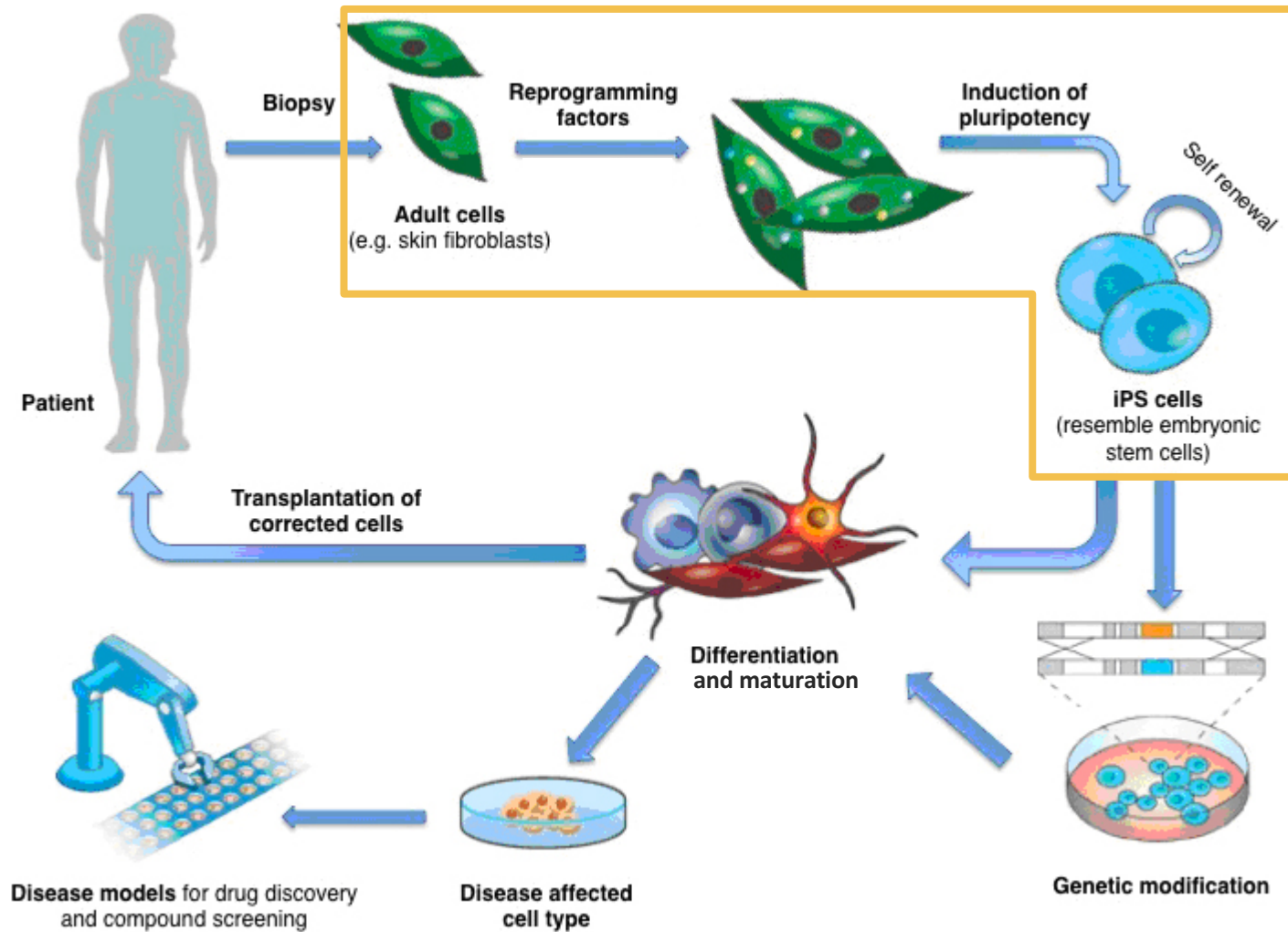
Drug Repurposing



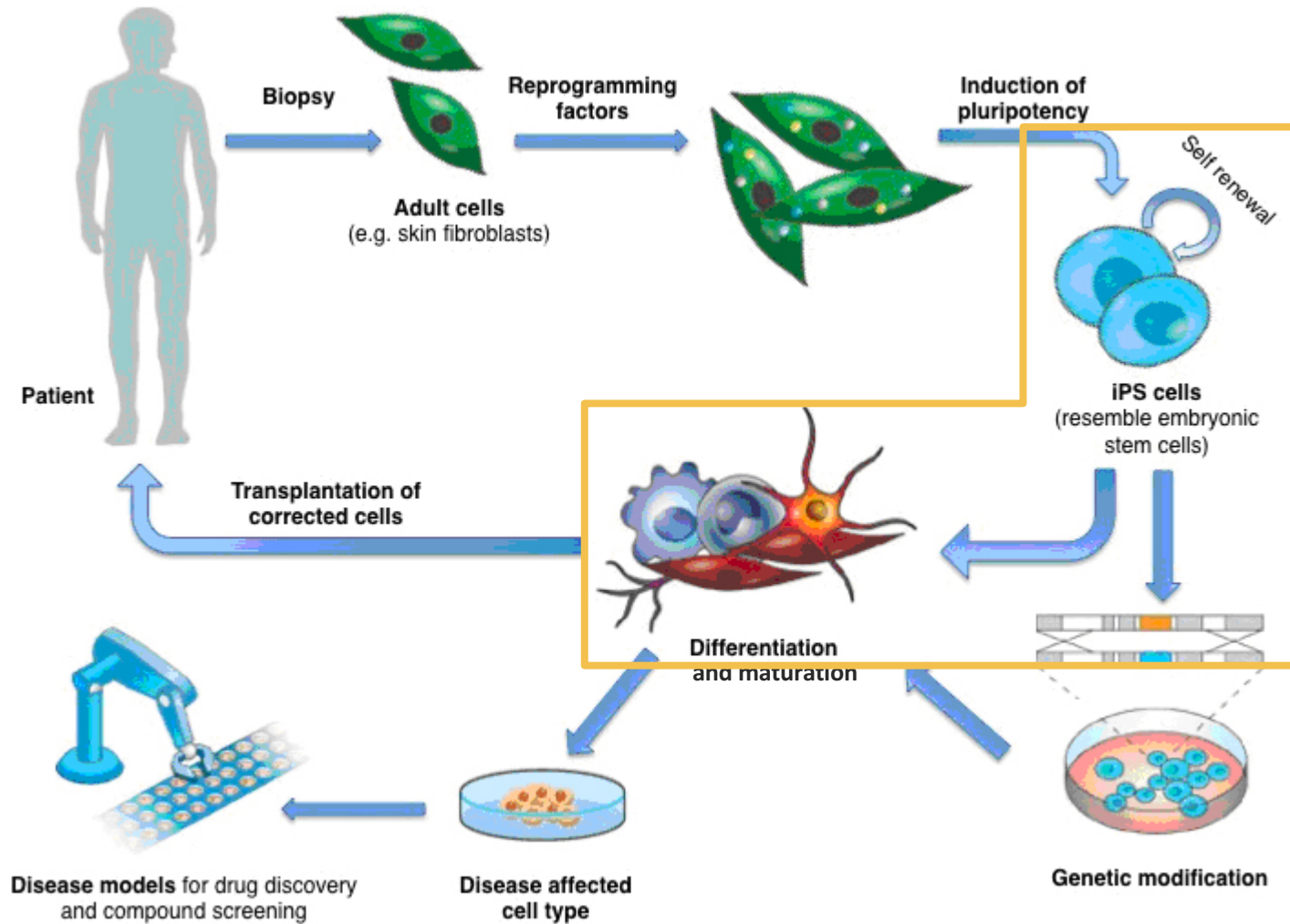
Cell Therapy and In Vitro Modeling



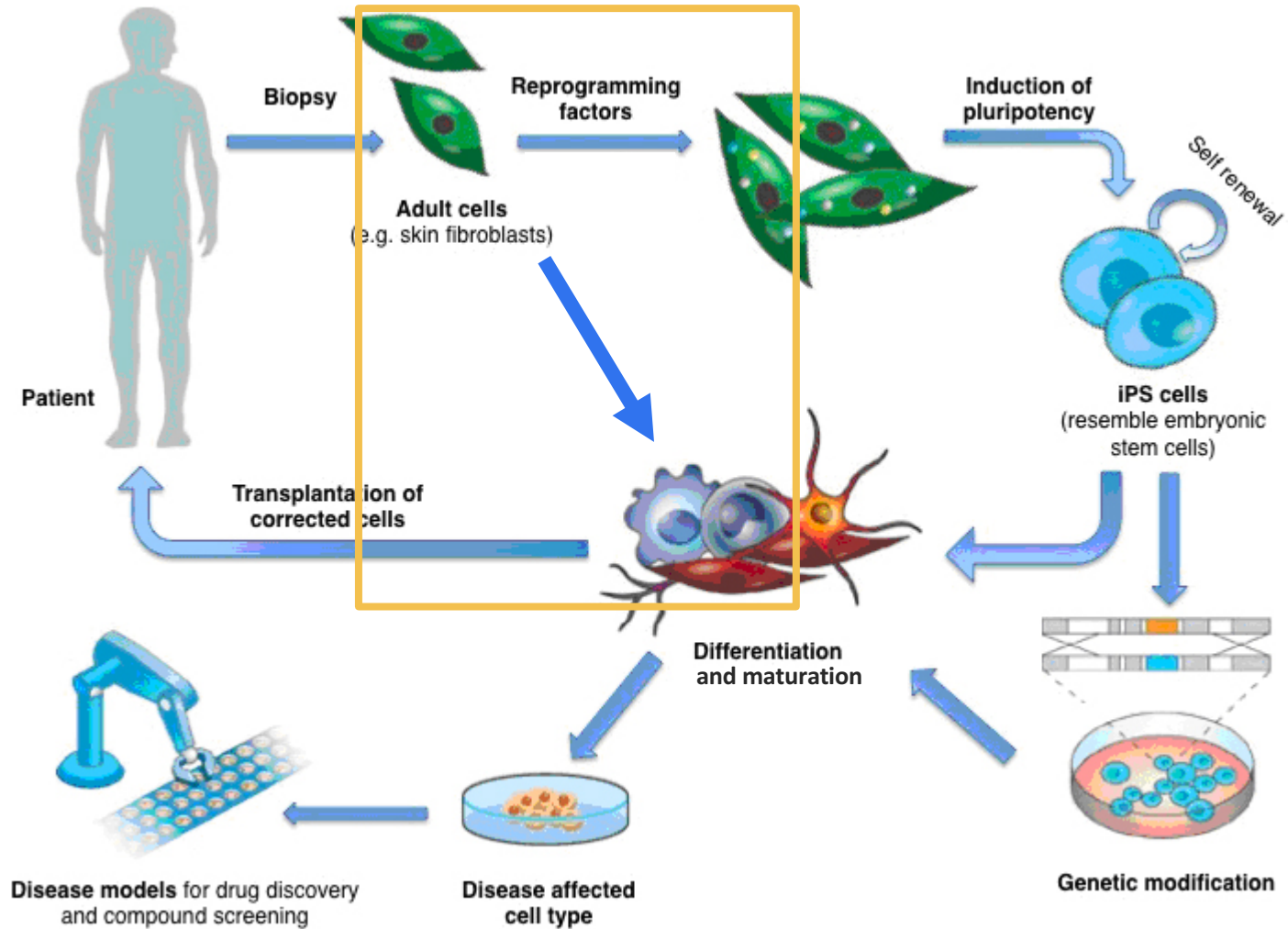
Cell Therapy and In Vitro Modeling



Cell Therapy and In Vitro Modeling



Cell Therapy and In Vitro Modeling



Current Process



embryonic stem cell transcription factors



Scholar

About 1,310,000 results (0.23 sec)

Articles

[\[HTML\] Core transcriptional regulatory circuitry in human embryonic stem cells](#)

LA Boyer, TI Lee, MF Cole, SE Johnstone, [SS Levine](#)... - *cell*, 2005 - Elsevier

Case law

... Genome-Wide Location Analysis in Human **Embryonic Stem Cells**. ... The results showed that one or more of the **stem cell transcription factors** occupied 1303 actively **transcribed** genes ... The importance of OCT4, SOX2, and NANOG for early development and ES **cell** identity led us ...

My library

Cited by 3560 [Related articles](#) [All 26 versions](#) [Web of Science: 2340](#) [Cite](#) [Save](#)

Any time

[\[HTML\] Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors](#)

[K Takahashi](#), [S Yamanaka](#) - *cell*, 2006 - Elsevier

Since 2017

Since 2016

Since 2013

Custom range...

... (C) Contribution of iPS-TTFgfp4-7 and iPS-TTFgfp4-3 **cells** to mouse **embryonic** development. ... **Embryos** were analyzed with a fluorescence microscope at E7.5 (upper panels, iPS-TTFgfp4-7) or E13.5 (lower panels, iPS-TTFgfp4-3 ... (D) The E13.5 chimeric **embryo** was sectioned ...

Cited by 16242 [Related articles](#) [All 72 versions](#) [Web of Science: 9590](#) [Cite](#) [Save](#)

Sort by relevance

[Embryonic stem cell lines derived from human blastocysts](#)

[JA Thomson](#), [J Itskovitz-Eldor](#), [SS Shapiro](#)... - ..., 1998 - [science.sciencemag.org](#)

Sort by date

... vitro, the ES **cells** differentiated when cultured in the absence of mouse **embryonic** fibroblast feeder ... largely restricted to the description of a limited number of sectioned **embryos** and to ... the egg cylinder all differ substantially from the corresponding structure of the human **embryo**. ...

Cited by 13810 [Related articles](#) [All 43 versions](#) [Web of Science: 7615](#) [Cite](#) [Save](#)

include patents

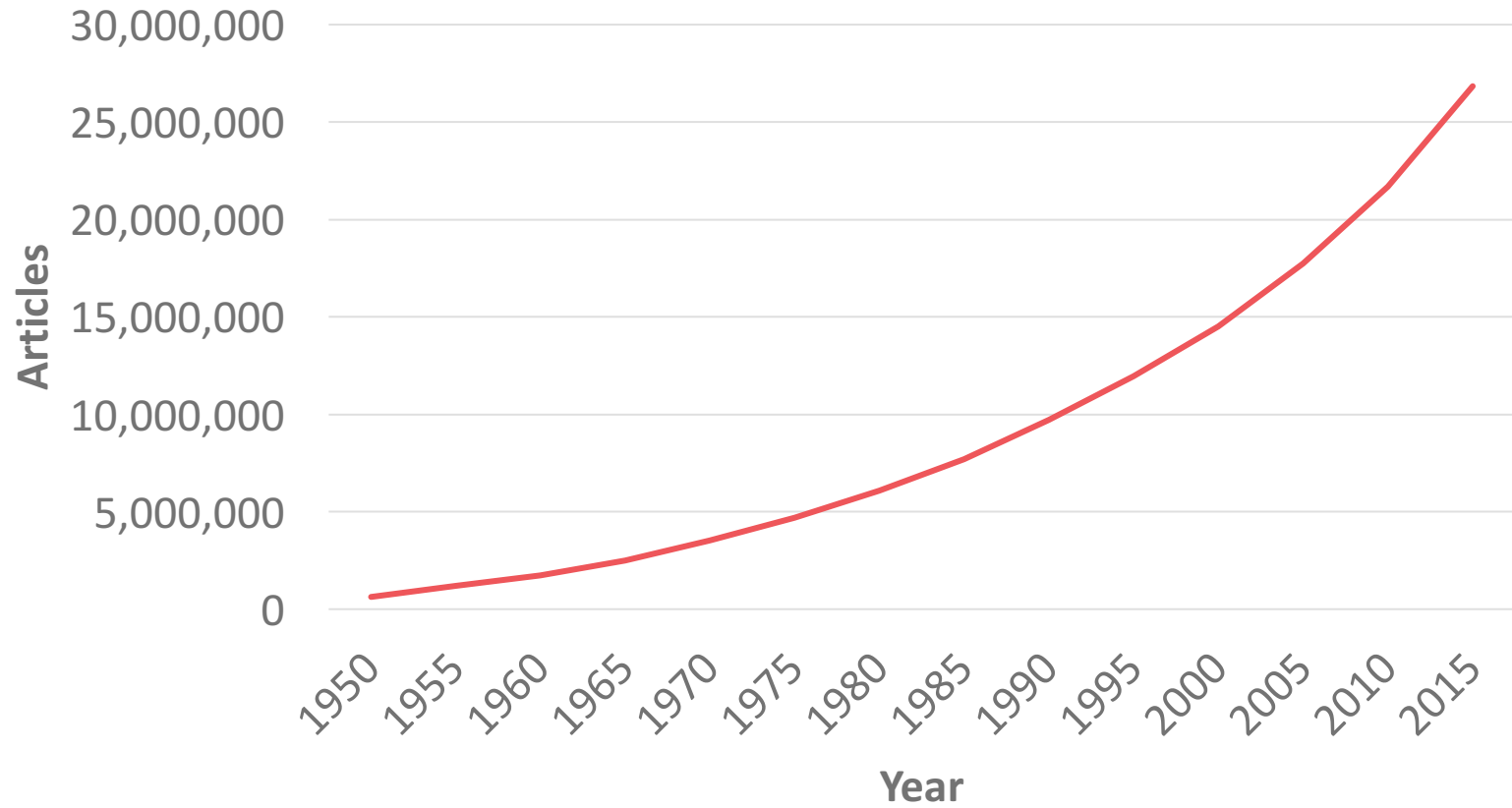
include citations

Create alert

[\[HTML\] Functional expression cloning of Nanog, a pluripotency sustaining factor in embrvonic stem cells](#)

Europe PubMed Central Over the Years

Number of EPMC Articles



Automate It

Have

A biomedical concept

e.g. *Embryonic stem cell*

Want

A ranking of targets

e.g. *Transcription factors*

Automate It

Have

A biomedical concept

e.g. *Embryonic stem cell*

Want

A ranking of targets

e.g. *Transcription factors*

NLP = HARD

Counting = EASY

KinderMiner

Target Terms

AATF	ABP1	...	ZXDC	ZYX
------	------	-----	------	-----

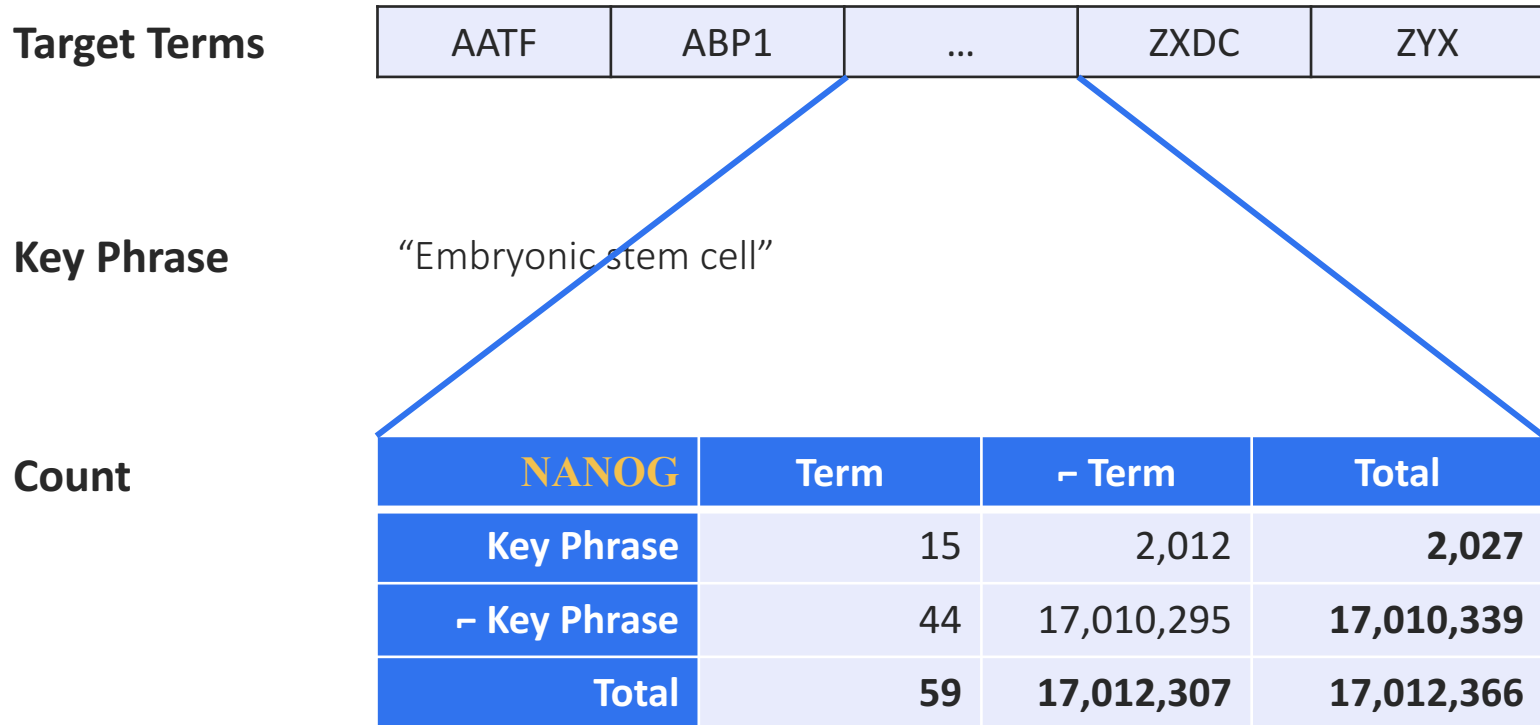
Key Phrase

“Embryonic stem cell”

KinderMiner

Target Terms	AATF	ABP1	...	ZXDC	ZYX
Key Phrase	"Embryonic stem cell"				
Count	AATF	Term	- Term	Total	
	Key Phrase	0	2,027	2,027	
	- Key Phrase	34	17,010,305	17,010,339	
	Total	34	17,012,332	17,012,366	

KinderMiner



KinderMiner

Target Terms	AATF	ABP1	...	ZXDC	ZYX
Key Phrase	"Embryonic stem cell"				
Count	ZYX	Term	- Term	Total	
	Key Phrase	0	2,027	2,027	
	- Key Phrase	93	17,010,246	17,010,339	
	Total	93	17,012,273	17,012,366	

KinderMiner

Output Rank

1. Compute article count contingency tables
2. Filter terms by one-sided Fisher Exact test
3. Sort terms by $\frac{\text{Key Phrase \& Term}}{\text{Term Total}}$

Example

NANOG + “Embryonic stem cell” + 2004

NANOG	Term	¬ Term	Total
Key Phrase	15	2,012	2,027
¬ Key Phrase	44	17,010,295	17,010,339
Total	59	17,012,307	17,012,366

One-sided FET p: 5.219e-46 Sort Ratio: $\frac{15}{59} = 0.254$

Experiments

Embryonic Stem Cell - 2004

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* (2006).

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* (2007).

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* (2007).

Cardiomyocyte - 2008

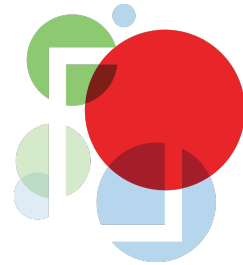
Ieda M, Fu JD, Delgado-Olguin P, Vedantham V, Hayashi Y, Bruneau BG, et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* (2010).

Addis RC, Ifkovits JL, Pinto F, Kellam LD, Estes P, Rentschler S, et al. Optimization of direct fibroblast reprogramming to cardiomyocytes using calcium activity as a functional measure of success. *J of Molecular and Cellular Cardiology* (2013).

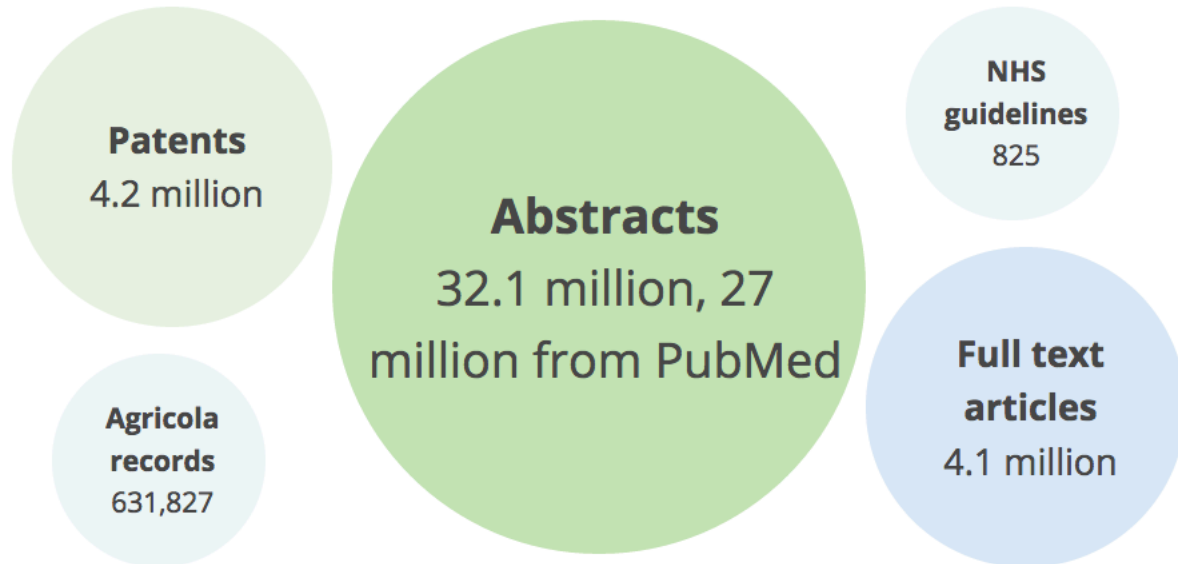
Hepatocyte - 2009

Huang P, He Z, Ji S, Sun H, Xiang D, Liu C, et al. Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* (2011).

Kogiso T, Nagahara H, Otsuka M, Shiratori K, Dowdy SF. Transdifferentiation of human fibroblasts into hepatocyte-like cells by defined transcriptional factors. *Hepatology International* (2013).



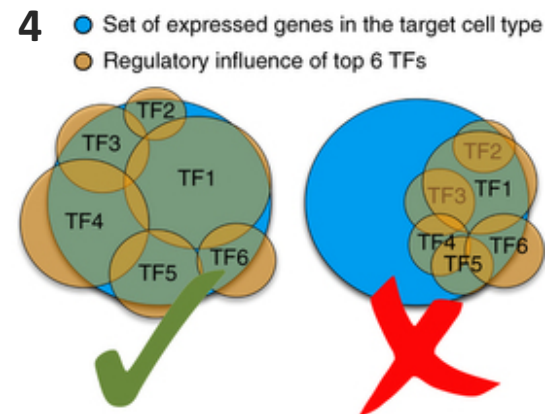
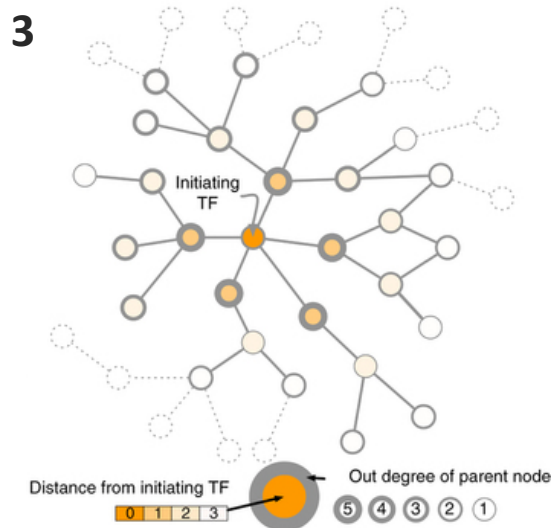
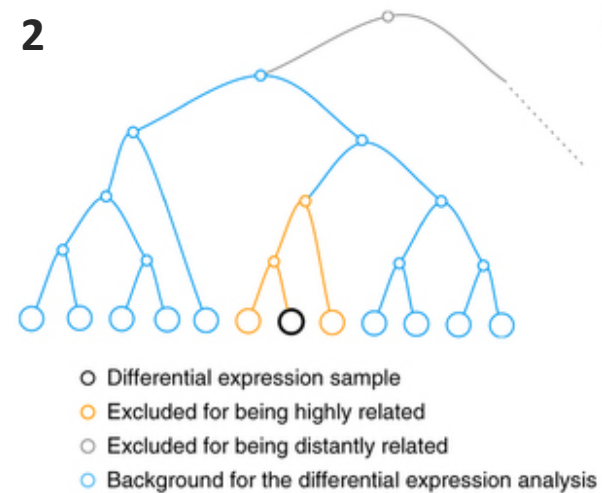
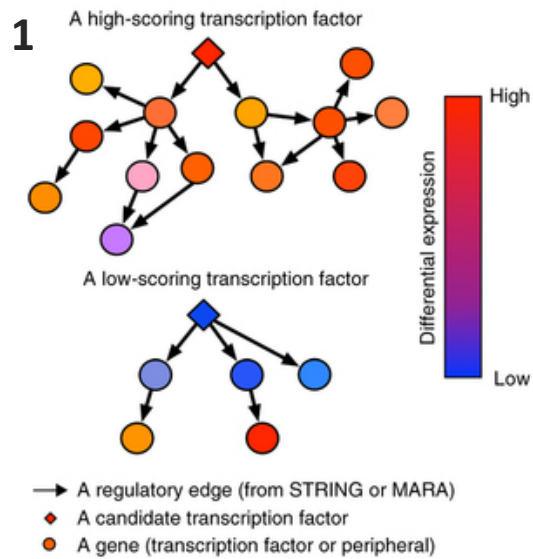
Europe PubMed Central



Reprogramming Results

"Embryonic Stem Cell" - 2004	"Cardiomyocyte" - 2008	"Hepatocyte" - 2009
NANOG	MESP1	HNF1A
UTF1	THRAP1	HNF1B
CBX4	TBX20	HNF4A
POU5F1	GATA4	ONECUT1
EZH1	NKX2-5	HNF4G
SOX1	TBX5	FOXA3
IRX4	GATA5	ONECUT3
FOXD3	MEF2C	FOXA1
MYF6	HAND2	FOXA2
HOXB4	CSRP3	TCF2
LMO2	IRX4	MLX
SOX2	HDAC9	NR0B2
EOMES	NFATC4	NR1I3
LMX1B	IRX5	NR1H4
LHX2	MKL2	HGBOX1
HOXD9	ISL1	NR1I2
HOXD11	GATA6	ONECUT2
OTX1	HAND1	TCF1
HAND1	HES2	CREB3L3
HOXB3	TBX18	CUTL2

Mogrify



O. Rackham et al., **A predictive computational framework for direct reprogramming between human cell types**. Nature Genetics (2016).

Reprogramming Results

"Embryonic Stem Cell" - 2004	"Cardiomyocyte" - 2008	"Hepatocyte" - 2009
★ NANOG	MESP1	HNF1A
★ UTF1	THRAP1	HNF1B
CBX4	★ TBX20	★ HNF4A
★ POU5F1	★ GATA4	★ ONECUT1
EZH1	★ NKX2-5	HNF4G
SOX1	★ TBX5	★ FOXA3
IRX4	GATA5	ONECUT3
★ FOXD3	★ MEF2C	★ FOXA1
MYF6	★ HAND2	★ FOXA2
HOXB4	CSRP3	TCF2
LMO2	IRX4	MLX
★ SOX2	HDAC9	★ NR0B2
EOMES	NFATC4	★ NR1I3
LMX1B	★ IRX5	★ NR1H4
LHX2	MKL2	HGBOX1
HOXD9	ISL1	NR1I2
HOXD11	★ GATA6	ONECUT2
OTX1	★ HAND1	TCF1
HAND1	HES2	★ CREB3L3
HOXB3	TBX18	CUTL2

Conclusions & Future Work

- Simple counting is surprisingly effective
- Could speed discovery without much effort

- Bayesian methods to adjust ranks when counts are low
- Compositions of multiple rankings
- Web application

Acknowledgements

- National Institutes of Health UH3TR000506-05
- National Institute of General Medical Sciences R01GM097618-05
- Marv and Mildred Conney

Thanks

Questions?