



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

Prof. Dr. Ralf Zimmer  
LFE Practical Informatics and Bioinformatics  
Department Institut für Informatik  
Ludwig-Maximilians-Universität München

**OS 13.12.2019**

**OS 25.11.2019**

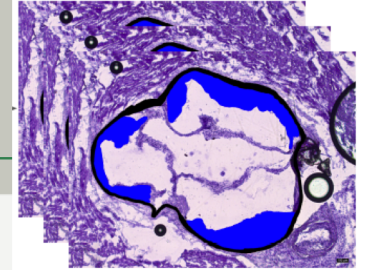
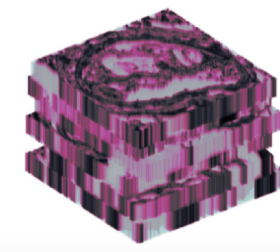
**OS 18.11.2019**





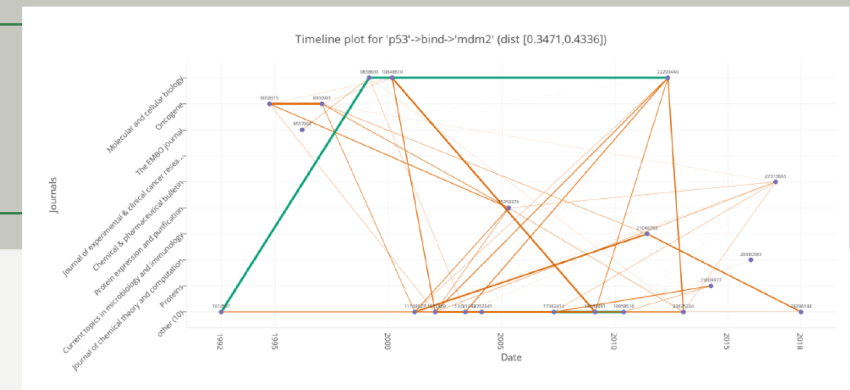
- **Fuzzy discretization, values & logic for HT data analysis**
- **Confidence Intervals**
  - in particular for GSEA/GNEA
- **Robust GSEA**
- **Splicing**
  - tool: MS data set => splicing candidates with estimation graphics
  - Splicing landscape in Cancer (CPTAC)
  - splicing for protein production (miRNA blocking of BP)
- **(MS)-EmpireS**
- **3D Reconstruction**
  - 3D index of heterogeneous HT and annotation data
- **Context based text mining**
  - miRExplore
  - time-line, recommendation systems, IF
- **ConceptMaker**
  - SVD, MOFA, umap
- **Systems Modeling of Expression (Yeast)**
- **Internet of Facts (IoF)**
- **Internet of Claims (IoC)**
  
- **DNA computing**
- **Brainbook**
  
- **Drug side effects and interactions (App)**
- **NGS/HT core center**
  
- **Cloud**
- **KI**
- **Climate Change**





- **3D Reconstruction**
  - 3D index of heterogeneous HT and annotation data
- **SFB1123 (Prof. Söhnlein, Celia Borja, Yue Wang)**
- **Kooperation in Z2 (Imhof, MSI)**
  - MS => Constantin ?
  - NGS data
  - single cell data
- **Bachelorarbeit Margaryta Olenchuk**
- **Vision:**
  - Integration of heterogeneous data
  - Spatio-temporal model
  - „Superresolution“: Inference/imputation

- **Context-based text mining**
  - miRexplore
  - Time-line
- **SFB1123: AtheMir**
- **Lehre: NEAP Master**
- **Resource for „contextual miRNA regulations“**
  - simple benchmark for context search
  - ... and recommendation/document similarity
  - as complete and accurate as possible
- **Time-line as „Killer-app“**
  - should be right !
  - new measure of scientific impact
  - find facts in contexts !





## Increased expression of miR-330-3p: a novel independent indicator of poor prognosis in human breast cancer

H. Wang, S.-H. Chen, P. Kong, L.-Y. Zhang, L.-L. Zhang, N.-Q. Zhang, H. Gu

Hospital of Clinical Medicine, Binzhou Medical University, Binzhou, Shandong, China. gg4646mm369@126.com

### ONCOLOGY

**OBJECTIVE:** Previous study has reported that miR-330-3p was highly expressed in breast cancer (BC) patients. However, the effect of miR-330-3p in BC progression remains largely unclear. The purpose of this study was to investigate the clinical significance of miR-330-3p expression in BC.

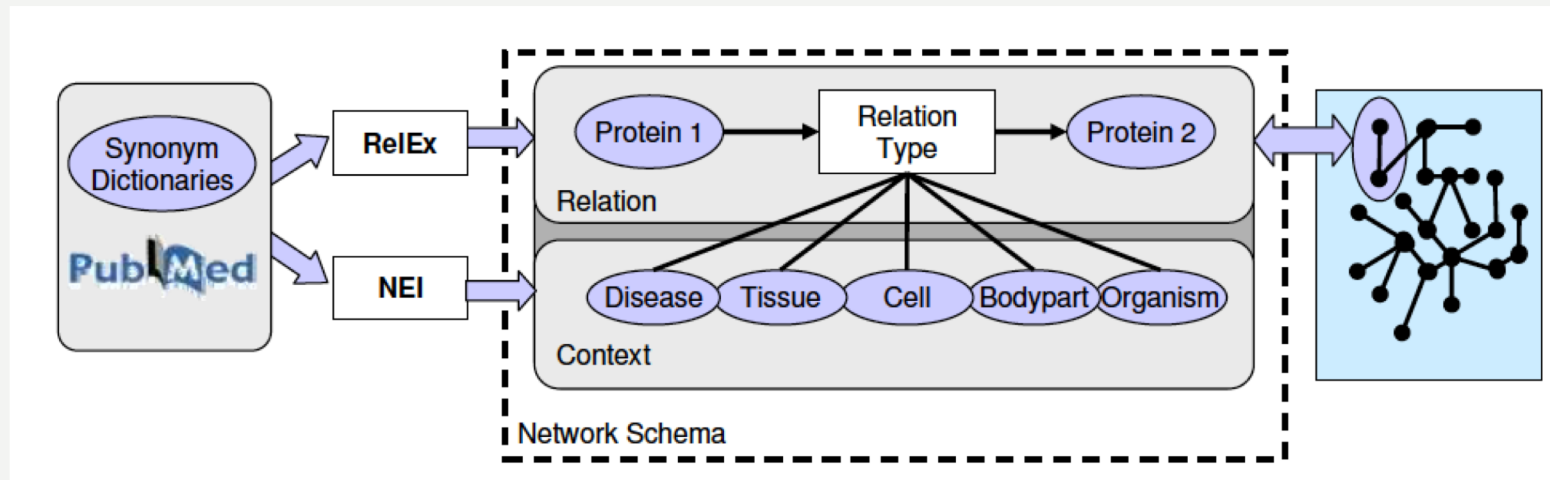
**PATIENTS AND METHODS:** The expression of miR-330-3p was detected by quantitative Real-time PCR in BC tissues and matched normal breast tissues. The association of miR-330-3p expression with clinicopathological factors of BC patients was also analyzed by  $\chi^2$ -test. Prognosis value of patients with BC was assessed by Kaplan-Meier method and Cox proportional hazards model, respectively.

**RESULTS:** Quantitative real-time PCR analysis showed that the expression level of miR-330-3p was significantly higher in BC specimens than that in corresponding noncancerous tissues ( $p < 0.01$ ). The levels of miR-330-3p were positively correlated with the status of TNM stage ( $p = 0.011$ ) and lymph node metastasis ( $p = 0.006$ ). Kaplan-Meier analysis revealed that 5-year overall survival of BC patients with high miR-330-3p expression was shorter compared to those patients with low miR-330-3p expression ( $p < 0.0001$ ). Moreover, univariate and multivariate regression analysis demonstrated that miR-330-3p was an independent prognostic factor in BC.

**CONCLUSIONS:** Our data suggest that miR-330-3p upregulation maybe concurrently associated with prognosis in patients with BC, suggesting that miR-330-3p may be a potential prognostic biomarker and therapeutic target for patients with BC.

- **False Positive in miRExplore:**
- **„The levels of miR-330-3p were positively correlated with the status of TNM stage ( $p = 0.011$ ) and lymph node metastasis ( $p = 0.006$ ).“**
- **But Context is very clear !**

- **Diss Katrin Fundel (2007)**



Increased expression of miR-330-3p “biomarker-of” poor prognosis in human breast cancer.

The context information from the sentence and the abstract which supports this relation is as follows:

**Evidence:** expression, increased expression, rtPCR

**Objects:** miR-330-3p

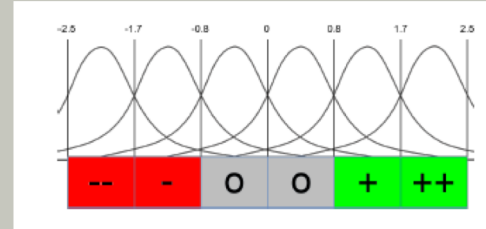
**Disease:** (human) breast cancer, TNM stage

**Species:** human

**Tissue:** breast cancer tissue, normal breast tissues

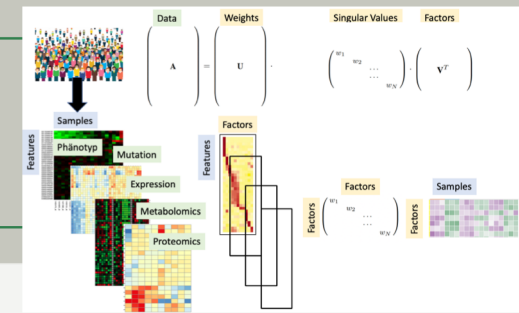
**Experimental technique:** prognosis, Kaplan-Meier method, Cox proportional hazards model, regression, x2-test

**Year:** Eur Rev Med Pharmacol Sci. (IF= 2.7)



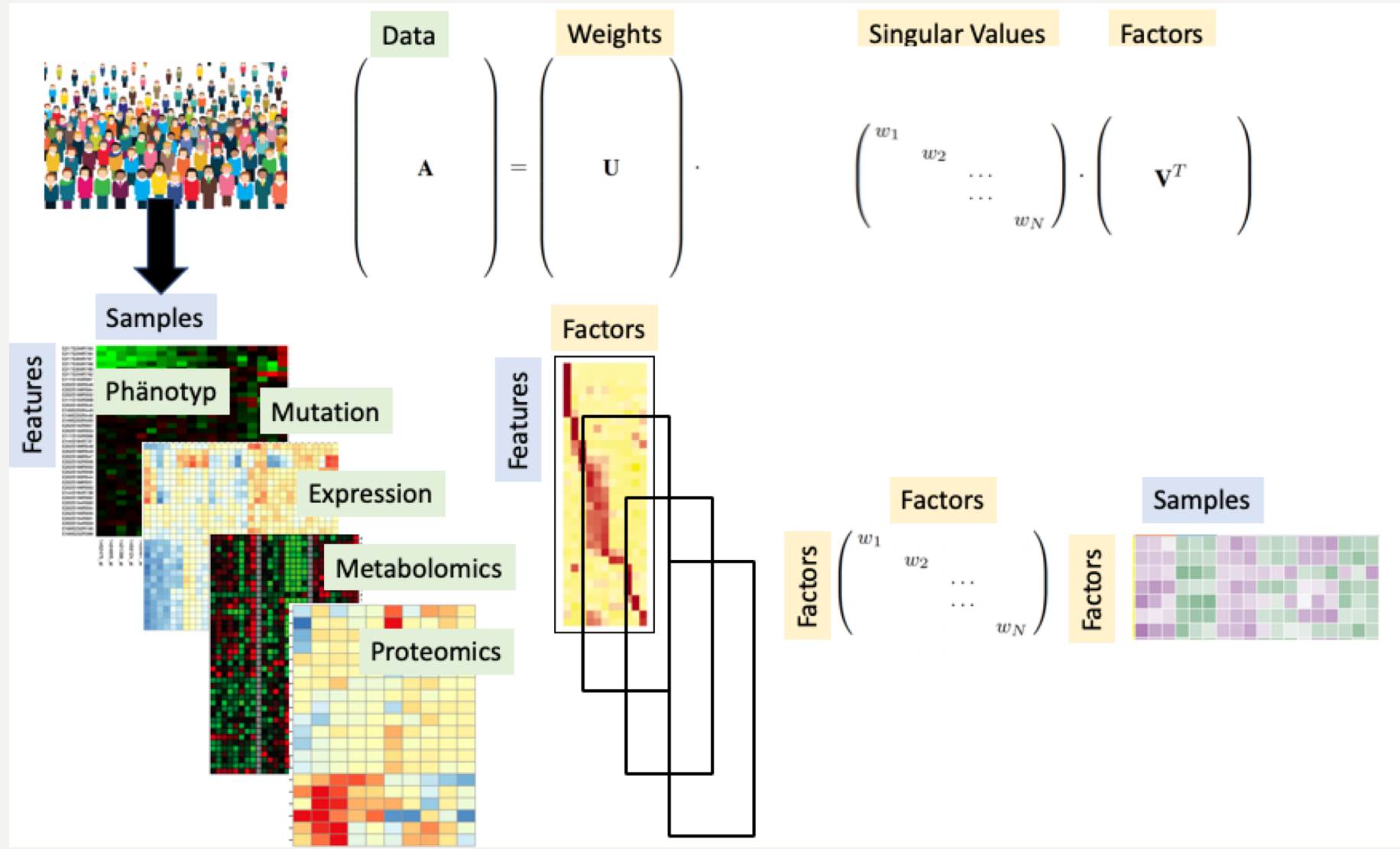
- **Fuzzy discretization, values & logic for HT data analysis (Lukas Diss)**
  - HT data description
  - Differential analysis
  - GSE analysis
- **SFB1123, Heatshock, NGS core**
- **Lehre: NEAP Master**
- **(MS)-Empire**
  
- **Combine Fuzzy Logic with context-dependent background distributions and local fold changes**
- **Combine Fuzzy Logic with confidence intervals and estimation graphics**
- **Extend to GSE, GPE (Evi Diss)**

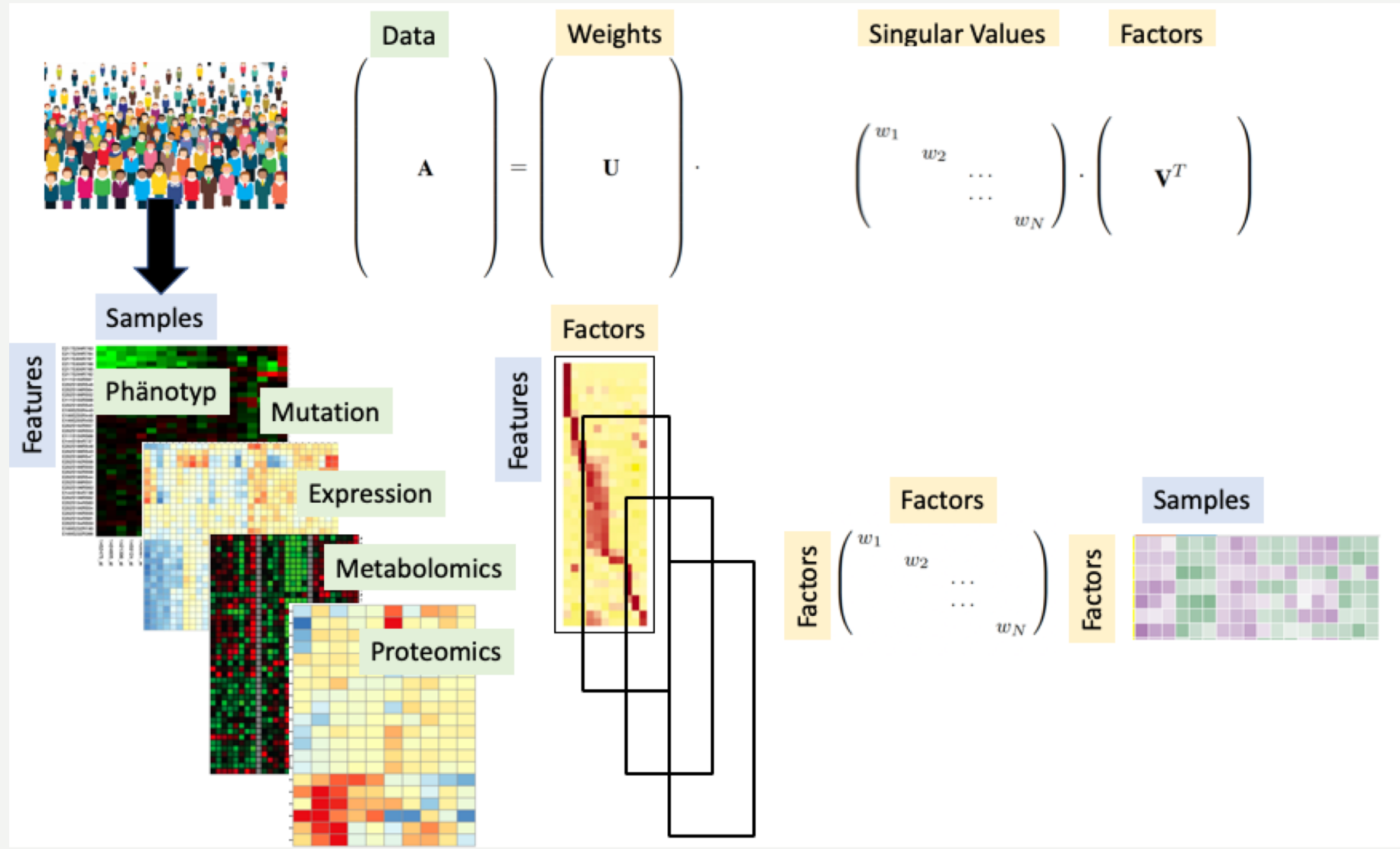




- **HT data analysis and heterogeneous data integration**
- **SFB1123 Athernetwork, Atheroindex**
- **Cancer Inadscape (TCGA)**
- **Encode: reproducibility ChipSeq, Atacseq etc**
- **Yeast Heatshock: Integration of data**
- **Can „classical“ integration methods work („latent“ methods)**
  - t-SNE, umap
  - MOFA
  - ConceptMaker
- **What about Bayesian Networks? Network Inference ?**


















## Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets

Ricard Argelaguet<sup>1,†</sup> , Britta Velten<sup>2,†</sup> , Damien Arno<sup>1</sup> , Sascha Dietrich<sup>3</sup> , Thorsten Zenz<sup>3,4,5</sup> ,  
John C Marionni<sup>1,6,7</sup> , Florian Buettner<sup>1,8,\*</sup> , Wolfgang Huber<sup>2,\*\*</sup>  & Oliver Stegle<sup>1,2,\*\*\*</sup> 

Multi-omics studies promise the improved characterization of biological processes across molecular layers. However, methods for the unsupervised integration of the resulting heterogeneous data sets are lacking. We present Multi-Omics Factor Analysis (MOFA), a computational method for discovering the principal sources of variation in multi-omics data sets. MOFA infers a set of (hidden) factors that capture biological and technical sources of variability. It disentangles axes of heterogeneity that are shared across multiple modalities and those specific to individual data modalities. The learnt factors enable a variety of downstream analyses, including identification of sample subgroups, data imputation and the detection of outlier samples. We applied MOFA to a cohort of 200 patient samples of chronic lymphocytic leukaemia, profiled for somatic mutations, RNA expression, DNA methylation and *ex vivo* drug responses. MOFA identified major dimensions of disease heterogeneity, including immunoglobulin heavy-chain variable region status, trisomy of chromosome 12 and previously underappreciated drivers, such as response to oxidative stress. In a second application, we used MOFA to analyse single-cell multi-omics data, identifying coordinated transcriptional and epigenetic changes along cell differentiation.



*Text Mining***Expert knowledge without the expert: integrated analysis of gene expression and literature to derive active functional contexts**

Robert Küffner\*, Katrin Fundel and Ralf Zimmer

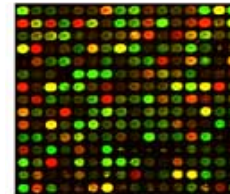
Department of Informatics, Ludwig-Maximilians-Universität München,  
Amalienstrasse 17 80333 München, Germany

# Understanding Complex Phenotypes and Diseases: e.g. Osteoarthritis



From Data to

~Disease Hypothesis

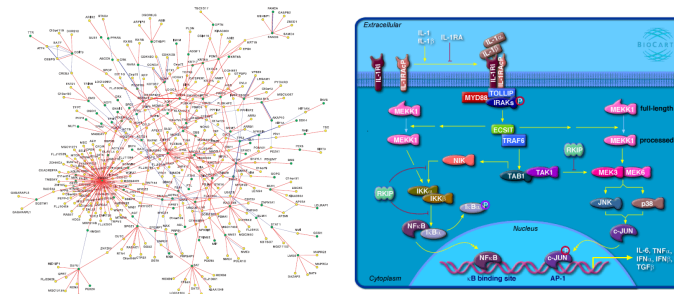


ConceptMaker

Concepts

Coherent Contexts

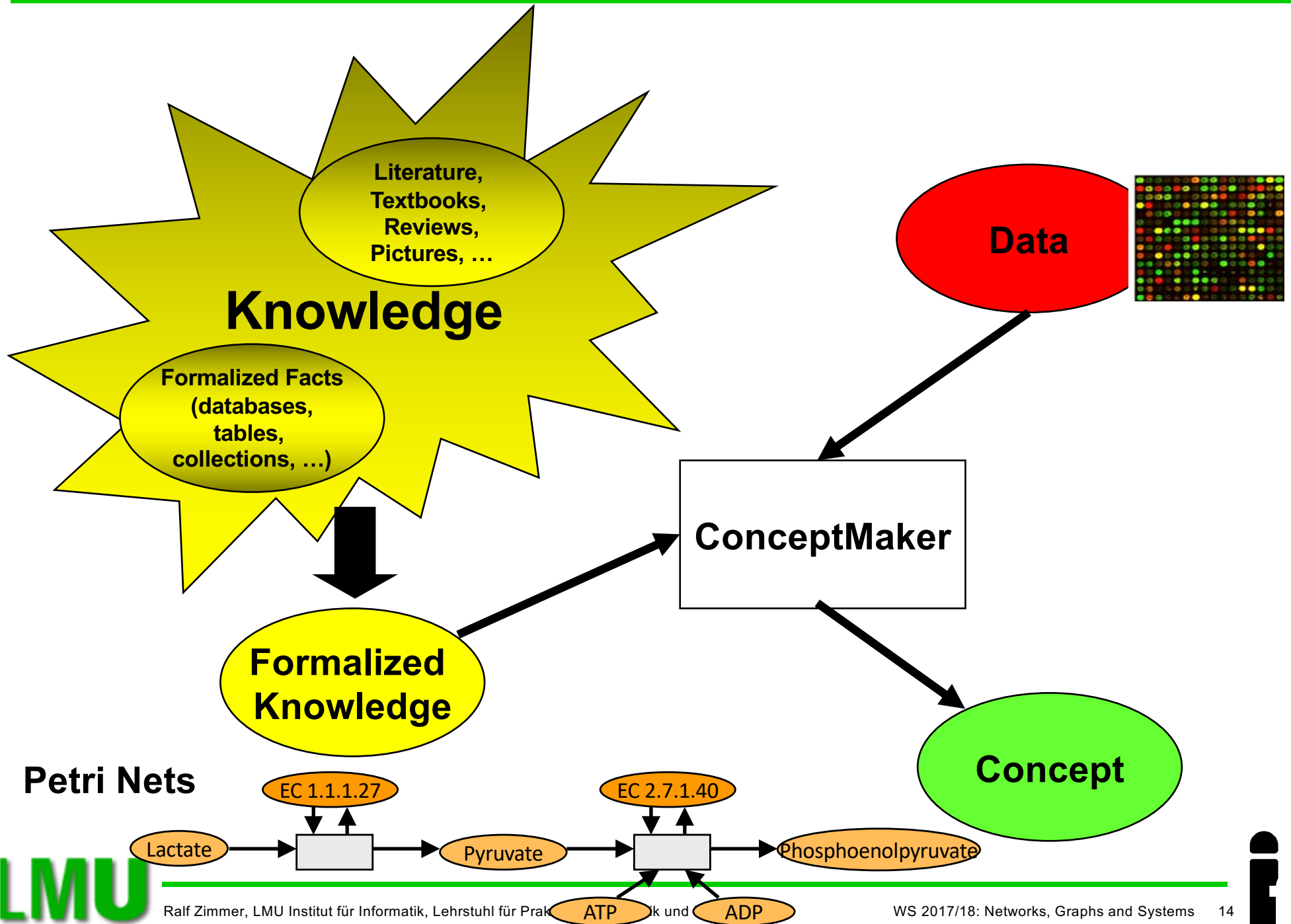
- Regulation
- Networks
- Literature
- Documents



Küffner, Fundel et al.  
2005-2006

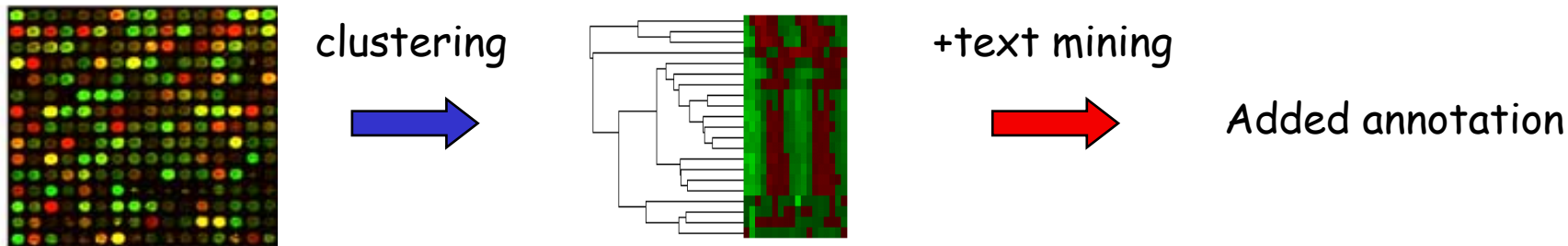
PubMed

# ConceptMaker: Concept



# Augmenting expression analysis with text mining

Approach 1, e.g. Masys et al, 2001:



Approach 2, e.g. Chaussabel&Cher, 2002 (Variant: GO-ORA):



ConceptMaker approach:



```

54 sc=11.0 sim=0.17 #= 3 neuroendocrin peptid
55 sc=10.8 sim=0.40 #= 4 growth hormone
56 sc=10.8 sim=0.48 #= 3 tubulin
57 sc=10.3 sim=0.36 #= 3 gadd45
58 sc=10.0 sim=0.32 #= 3 antiprolif
59 sc= 5.2 sim=0.41 #= 3 fibronectin iii

```

**Tree of clusters**

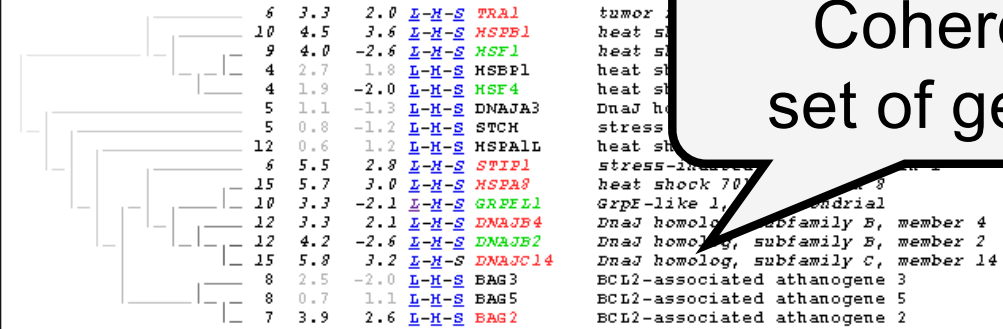
```

7 3 secretas complex
45 3 melanoma, experimenta
50 3 antizym
57 3 gadd45
56 3 tubulin
17 14 metallothionein
55 4 growth hormone
15 16 iqfbb
21 5 serum
47 4 antigens, cd98
28 8 antigens, cd20
49 5 groucho
40 5 high mobility group p
58 3 antiprolif
38 5 oxygenator
48 6 proap
37 12 ret
33 12 aloc
18 24 cyto
27 13 circ
5 59 wnt
23 17 rece
54 3 neu
35 7 plac
41 9 mac
46 4 rece
31 4 ccaat-enhanc
43 9 interleukin
8 4 urinary p
2 3 sl00 pr
53 8 heat-shock proteins 7
42 3 intergen
52 10 cofilin
19 18 wasp
6 6 exm protein
39 7 complement
16 15 cl inhibitor
12 6 hla-dr antigen
34 11 nadph oxidase
3 22 sialyltransferase
59 3 fibronectin iii
51 5 roundabout
1 51 lencin repeat
26 21 angiopoietin
22 10 receptors, fibroblast
10 46 activin
44 7 immediate-early prote
24 17 tetraspanin
11 33 integrin
30 6 tenascin
25 22 hyaluronan
13 25 sulfotransferase
29 11 protein-lysine 6-oxid
32 21 thrombospondin
4 29 matrix metalloprotein
14 15 muscular dystrophie
20 14 microfibril
36 6 extracellular matrix
9 21 laminin
0 30 collagen

```

[Cluster links](#)

**Gene tree**

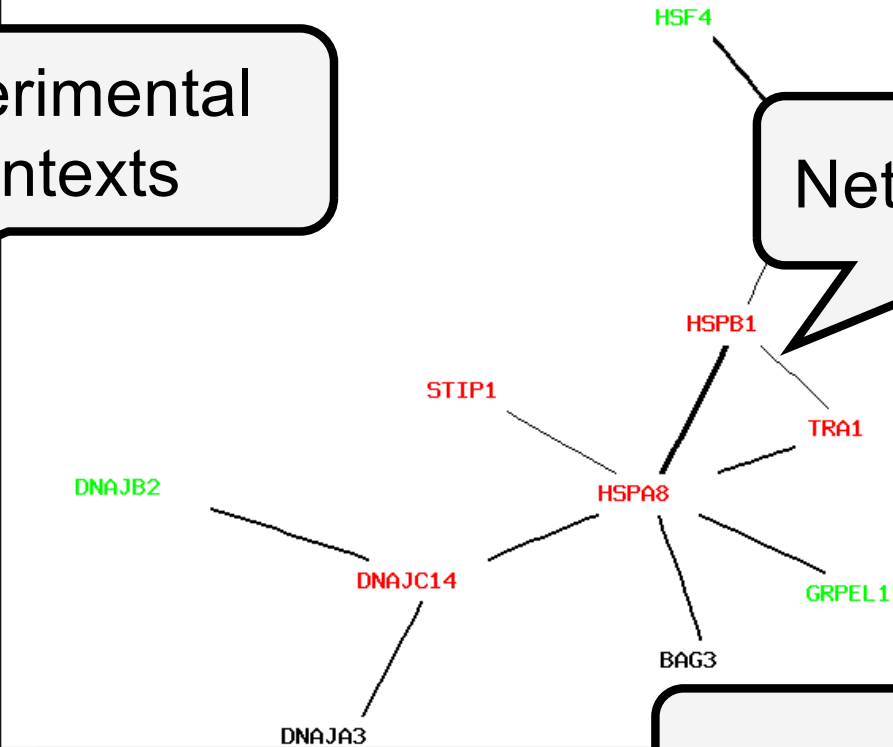


Coherent set of genes

**Gene links**

Experimental contexts

Network context



Annotations

1: HSPA8 heat shock 70kDa protein 8 [ Homo sapiens ]

GeneID: 3312

Summary

Official Symbol HSPA8

Official Full Name heat shock 70kDa protein 8

provided by HGNC

Summary  
Genomic regions, transcripts...  
Genomic context  
Bibliography



# Understanding Complex Phenotypes and Diseases: Data Analysis

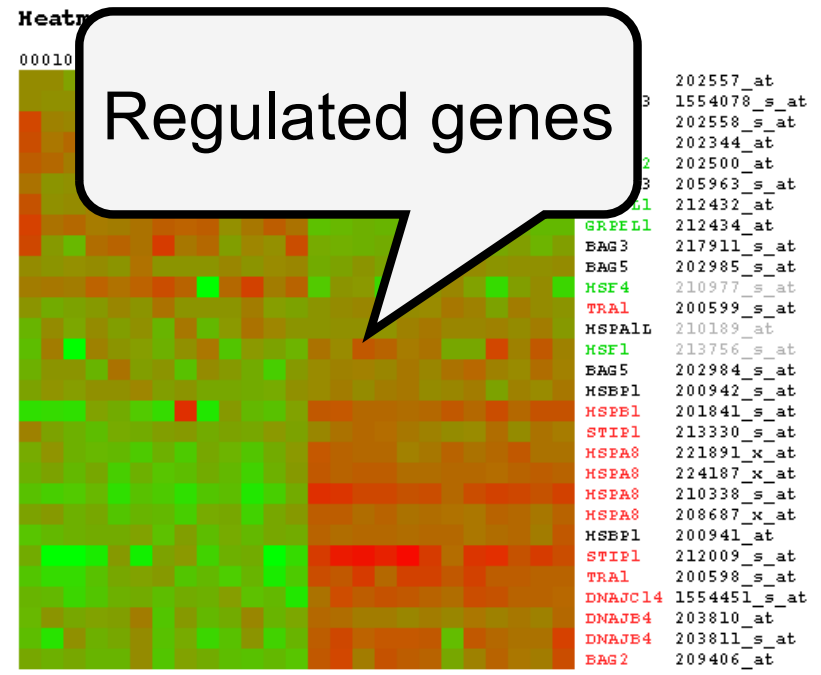
**Coherent set of genes**

**Experimental contexts**

**Network context**

**Annotations**

Gene tree showing clusters of genes related to HSP88. Network context diagram showing interactions between HSP88 and other genes like HSF4, HSF1, HSPB1, TRAF1, STIP1, DNAJB2, DNAJC14, DNAJB3, BAG3, and GRPSEL1. Annotations include: Official Symbol: HSP88, Official Full Name: heat shock 70kDa protein 8, GeneID: 3312, Entrez Gene Home, Summary, and Table of Contents.



**Literature contexts**

Search results for HSP88 in literature contexts, including:
 

- Stress-inducible, murine protein mSTH1. Characterization of the protein structure and binding domains for heat shock proteins and in vitro phosphorylation by different kinases.
- Isolation and characterization of a cDNA encoding a Xenopus 70-kDa heat shock cognate protein, Hsc70.1.
- Localization of the gene encoding the human heat shock cognate protein, HSP73, to chromosome 11.
- The human heat-shock protein family. Expression of a novel heat-inducible HSP70 (HSP70B) and isolation of its cDNA and genomic DNA.
- Structure and expression of the three MHC-linked HSP70 genes.
- Molecular chaperones as HSF1-specific transcriptional repressors.
- Modulation of Drosophila heat shock transcription factor activity by the molecular chaperone DROU1.
- Early transcriptional activation of the hsp70.1 gene by osmotic stress in one-cell embryos of the mouse.

1.7 1.6 L-H-S CD44 CD44 antigen interactions [help]

Clusters: [hyaluronan cell-matrix adhesion \(GO:0007160\)](#) [collagen binding \(GO:0005518\)](#)

#mutual	23	25	11	48	16	28	8	2	20	10	11
					253	1072	38	87	856	163	217
					15.8	15.7	13.7	11.9	10.8	10.5	10.4
					7.6 5.8	2.7 2.1	1.1 -1.4	6.4 6.1	1.3 1.6	2.6 1.8	3.1 2.1
					L-H-§	L-H-§	L-H-§	L-H-§	L-H-§	L-H-§	L-H-§
					VIL2 villin 2	SELL selectin L	PPG1 proteoglycan 1, secretory granule	TNFAIP6 tumor necrosis factor alpha-induced protein 6	ITGA4 integrin, alpha 4	MSN meslin	HSP62 heparan sulfate proteoglycan 2
							hyaluronan	hyaluronan	integrin	erm protein	laminin basement membrane (GO:0005604)

**Interactions & text contexts**



	1	1	4560	18026	30100
	1	10	303		
	1	11	190		
	1	111	1		
	1	118	2		
	1	12	165		
	1	121	1		

639	20	104	1	68
640	20	21	1	
641	20	25	1	
642	20	30	1	
643	20	31	2	
644	20	32	1	

- **Alternative Splicing**
  - **Detection (reliable, protein level, ... & transcript level)**
  - **Landscape (is splicing important?)**
  - **Tool**
- **SFB1123 MassSpec!**
- **Diss Constantin (MS-Empire, splicing project),  
Diss Markus (EmpireS)**
- **Use CPTAC proteomics data:**
  - **tool: MS (TMT?) data set => all reliable splicing candidates (with estimation graphics)**
  - **Apply to CPTAC and identify some relevant cases, check for known and unknown**
  - **Apply for landscape: => IoF**
  - **Atheroindex (SFB1123) as a (simple) application**





- **IoF & IoC**
  - **Internet of Facts => time-lines with evidences, facts = relations in contexts**
  - **Internet of Claims => facts which could be True, i.e. in related contexts, in generalized contexts, but are not (yet) established/validated**
  - **Resource**
- **SFB1123: AtheroNet, AtheroIndex**
- **NEAP recommendation systems**
- **NEAP relations, NEAP data search**
- **Context-based TextMining**
- **Robust data analysis: Transcriptomics (input), regulations (mechanism), mass spec (effect??)**

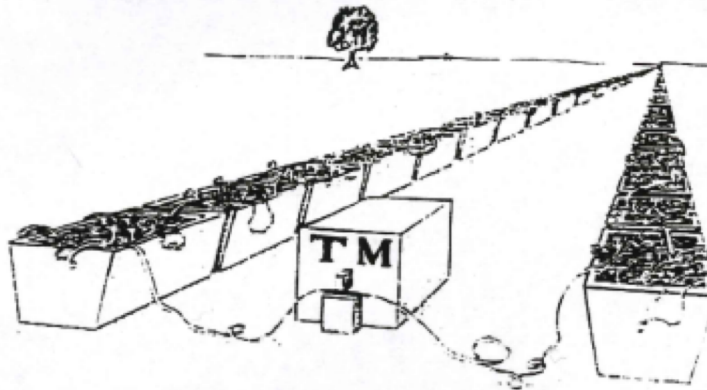




- **DNA Computing**
- **Lambda calculus:  $E = x \mid @ E E \mid \text{lambda } x. E$**
- **Semantic / meaning function = computation = string reduction**
- **$m (@ \text{lambda } x. E A) = E [x \leftarrow A]$**



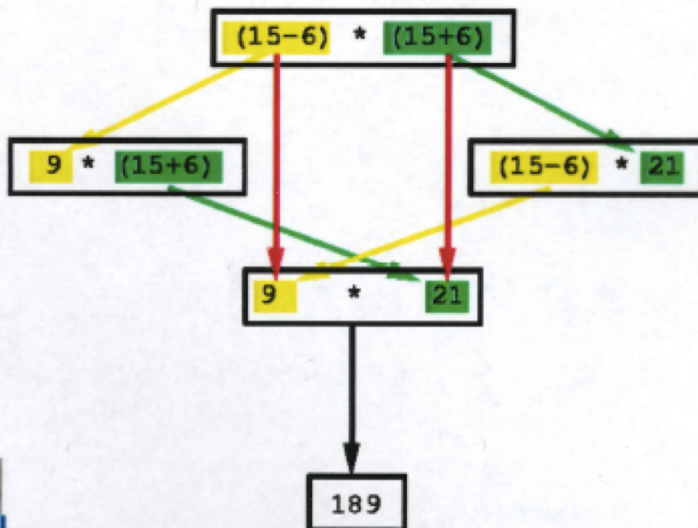
## Zustandstransformation



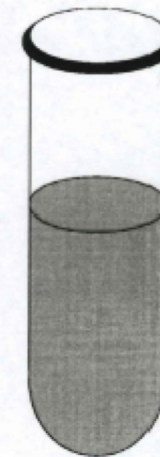
## Computer



## Ausdruckstransformation

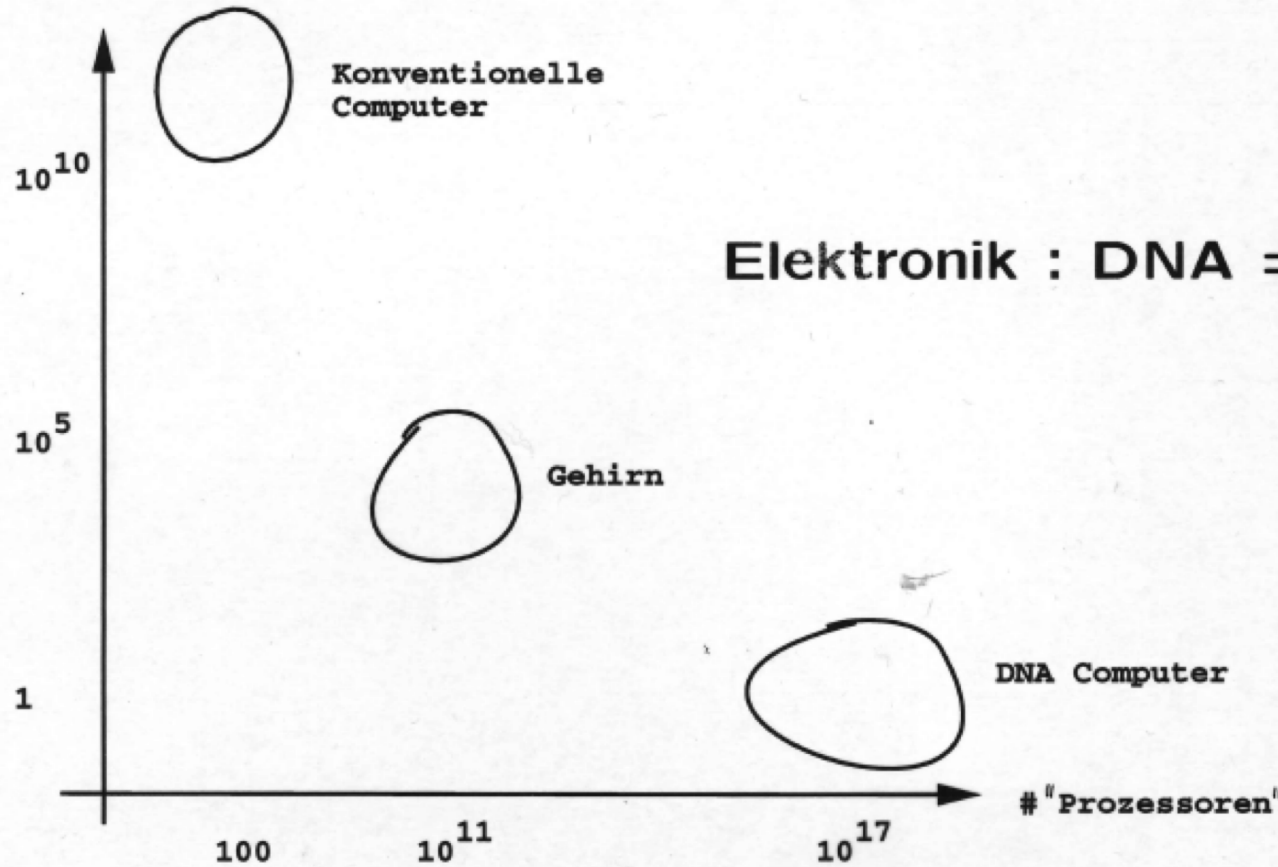


## DNA Computer





# Operationen/Stunde



Elektronik : DNA =  $10^9$  Energie  
 $10^{12}$  Speicher  
 $10^{15}$  Prozessoren

DNA Lösung 6g/l:  $6 \cdot 10^{19}$  Moleküle à 200 bp  $\approx 3 \cdot 10^9$  TeraByte/l  
 PCR Zyklus 1min:  $6 \cdot 10^{19}$  Ops/min  $\approx 1$  Mio TeraOps/sec





- DNA Computing

Article

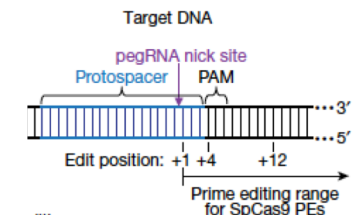
# Search-and-replace genome editing without double-strand breaks or donor DNA

<https://doi.org/10.1038/s41586-019-1711-4>

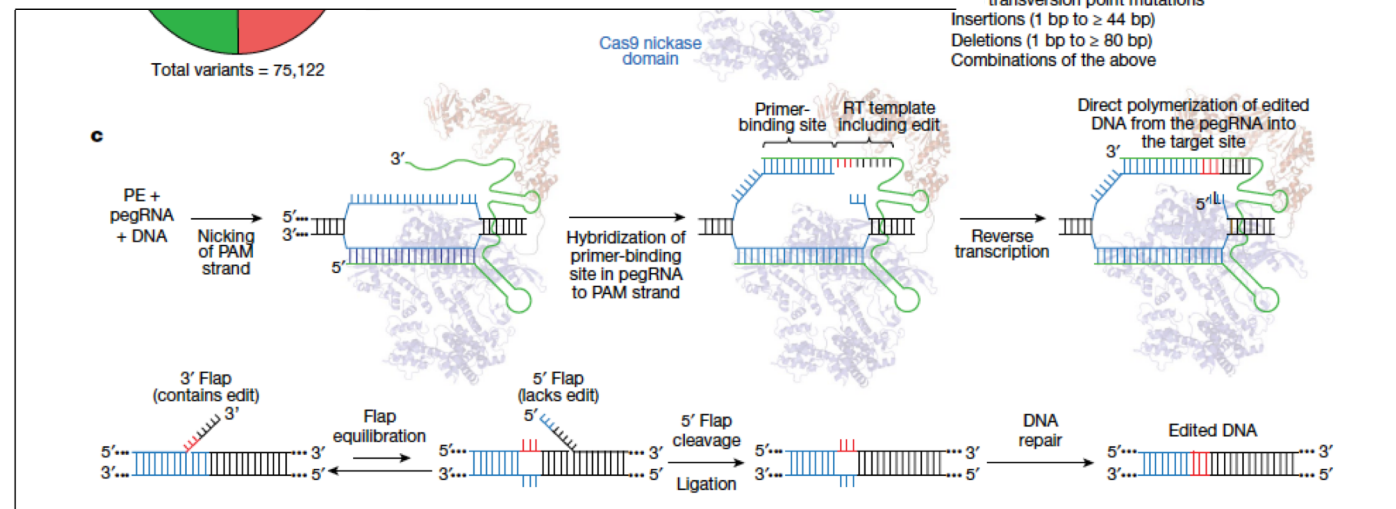
Received: 26 August 2019

Accepted: 10 October 2019

Andrew V. Anzalone<sup>1,2,3</sup>, Peyton B. Randolph<sup>1,2,3</sup>, Jessie R. Davis<sup>1,2,3</sup>, Alexander A. Sousa<sup>1,2,3</sup>, Luke W. Koblan<sup>1,2,3</sup>, Jonathan M. Levy<sup>1,2,3</sup>, Peter J. Chen<sup>1,2,3</sup>, Christopher Wilson<sup>1,2,3</sup>, Gregory A. Newby<sup>1,2,3</sup>, Aditya Raguram<sup>1,2,3</sup> & David R. Liu<sup>1,2,3\*</sup>



editing scope:  
 transition point mutations  
 transversion point mutations  
 Insertions (1 bp to ≥ 44 bp)  
 Deletions (1 bp to ≥ 80 bp)  
 Combinations of the above





• DNA Computing

Article | Published: 13 November 2019

**High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast**

Alex N. Nguyen Ba, Ivana Cvijović, José I. Rojas Echenique, Katherine R. Artur Rego-Costa, Xianan Liu, Sasha F. Levy & Michael M. Desai

Nature 575, 494–499(2019) | Cite this article

8552 Accesses | 213 Altmetric | Metrics

nature

Letter | Published: 16 August 2017

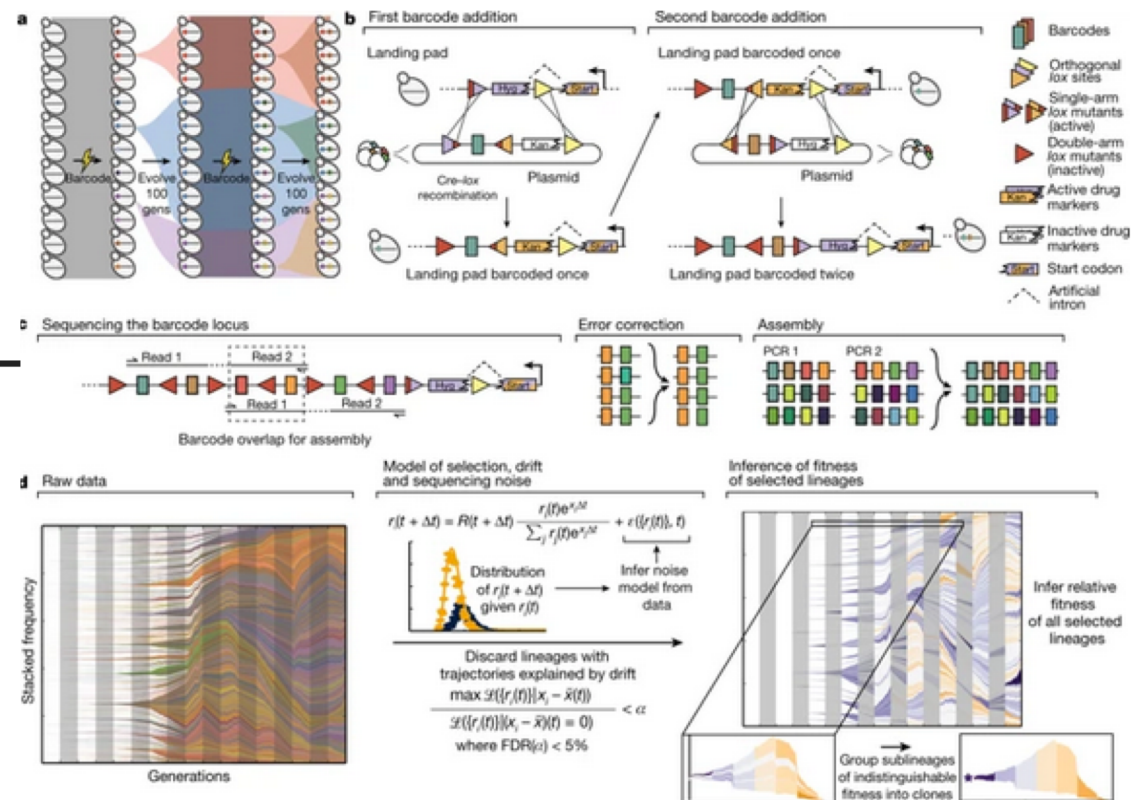
**Polylox barcoding reveals haematopoietic stem cell fates realized in vivo**

Weike Pei, Thorsten B. Feyerabend, Jens Rössler, Xi Wang, Daniel Postrach, Katrin Busch, Immanuel Rode, Kay Klapproth, Nikolaus Dietlein, Claudia Quedenau, Wei Chen, Sascha Sauer, Stephan Wolf, Thomas Höfer & Hans-Reimer Rodewald

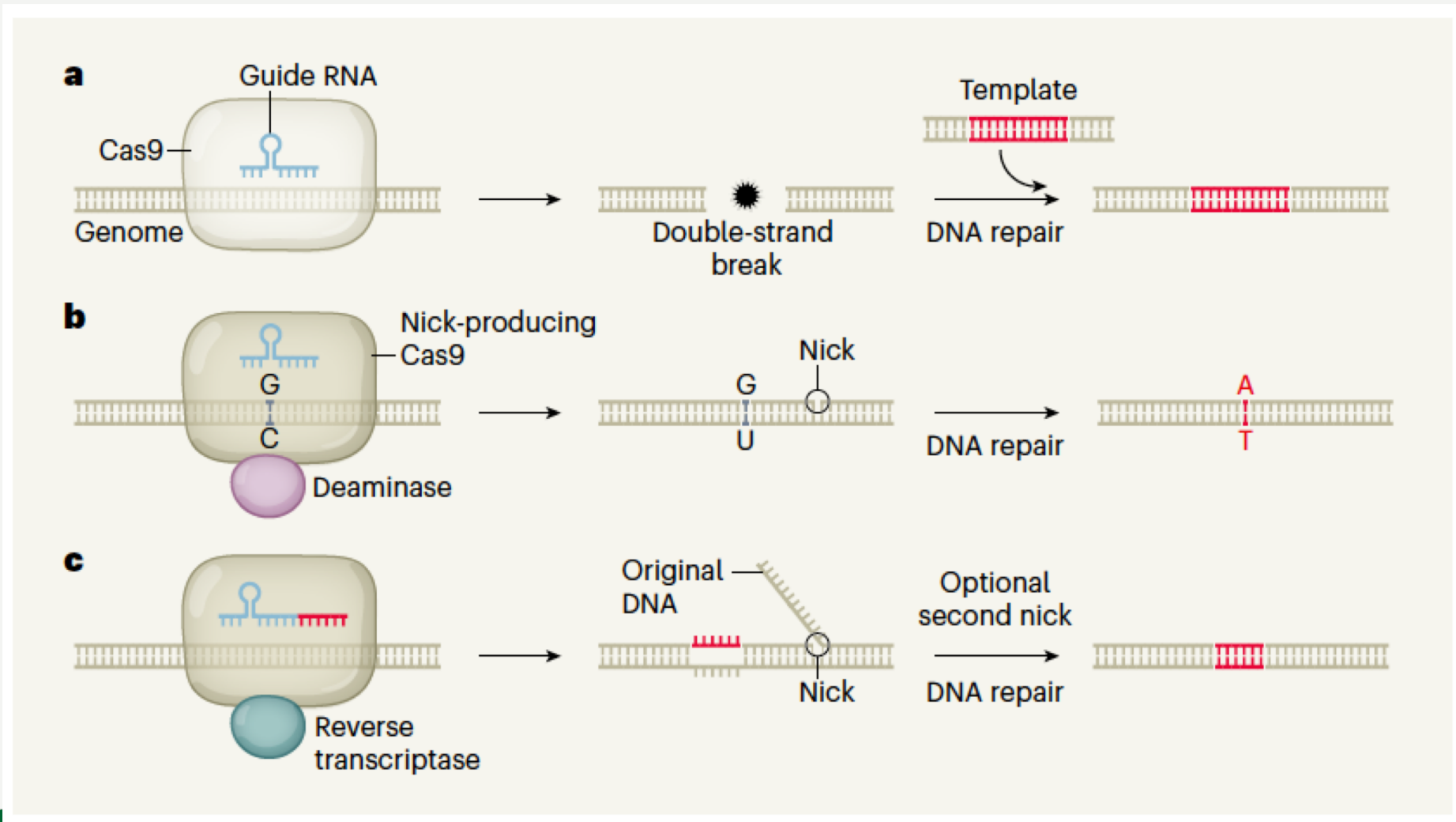
Nature 548, 456–460(2017) | Cite this article

3900 Accesses | 90 Citations | 123 Altmetric | Metrics

**Fig. 1: Renewable barcoding system and lineage dynamics.**









- **Climate neutral bioinformatics institute**
- **(First) Climate neutral university department**



