

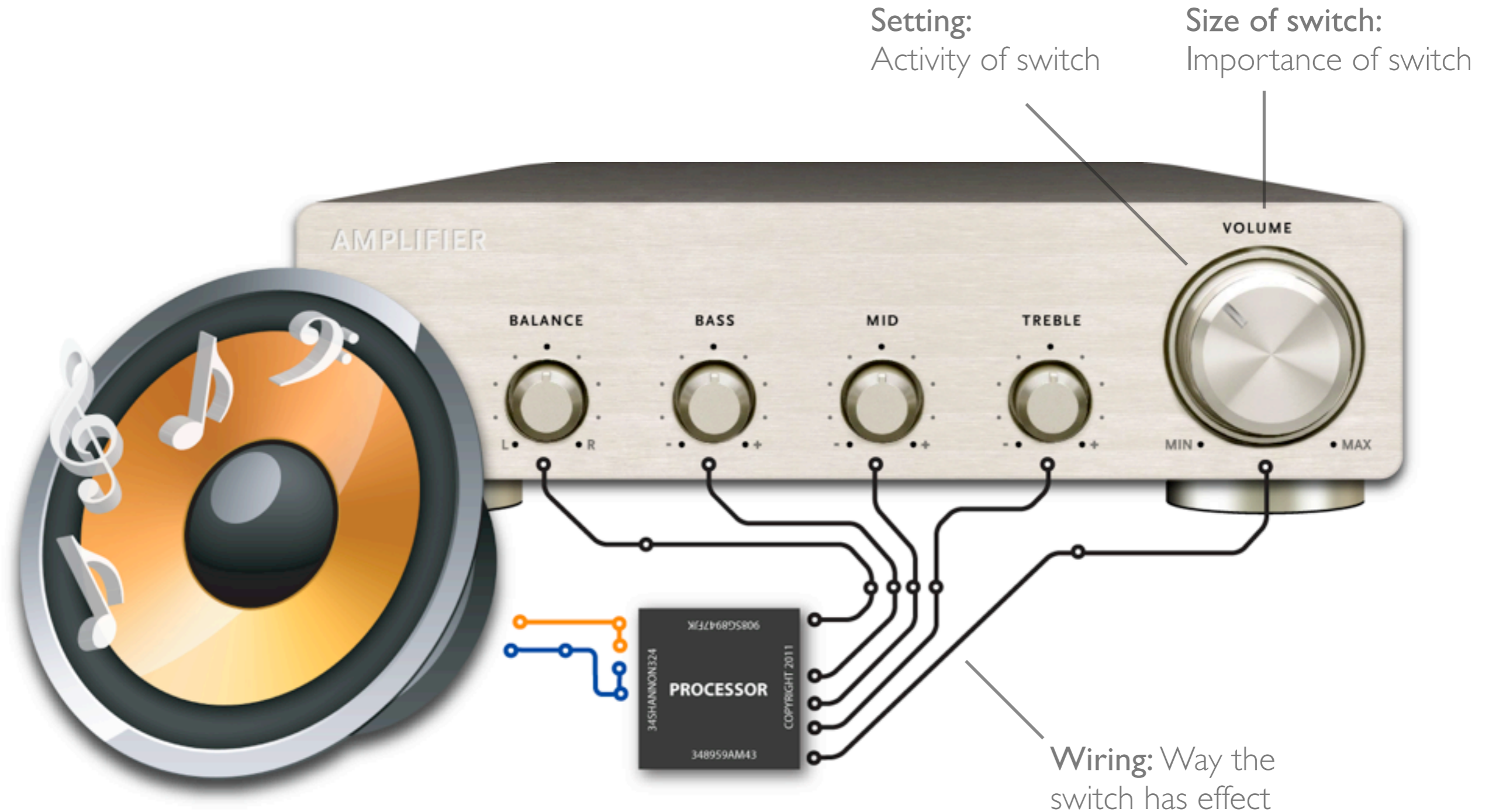


umcg

Department of genetics

Lude Franke > Integrating different omics data

Amplifier can change many aspects of music

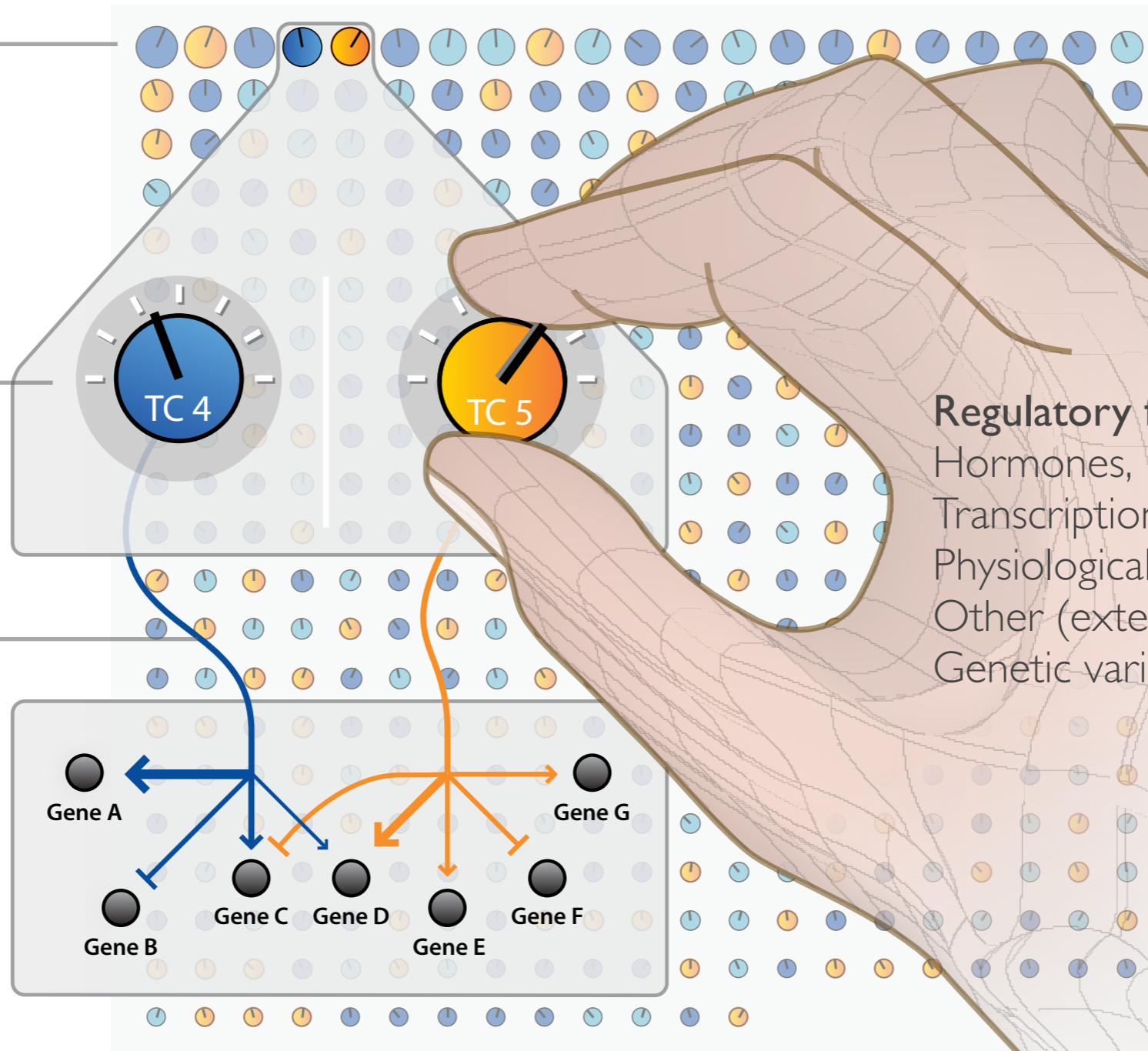


A control panel that determines gene expression?

Size of switch:
Importance

Setting: State of
a certain sample

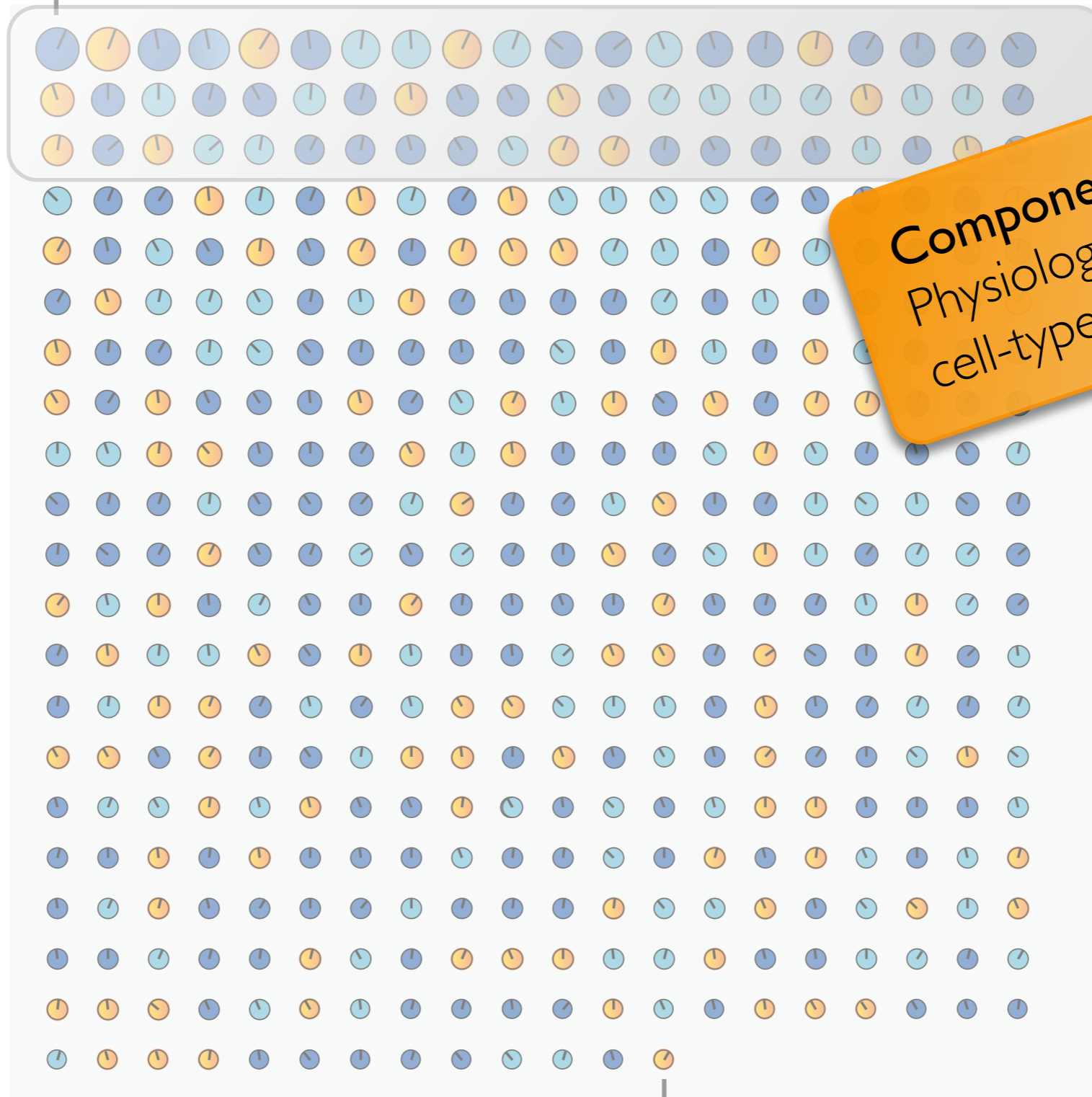
Wiring: Effect on
individual genes



Regulatory factors:
Hormones,
Transcription factors,
Physiological factors,
Other (external) stimuli
Genetic variation

800 'transcriptional components': Component 1 - 50

Component 1

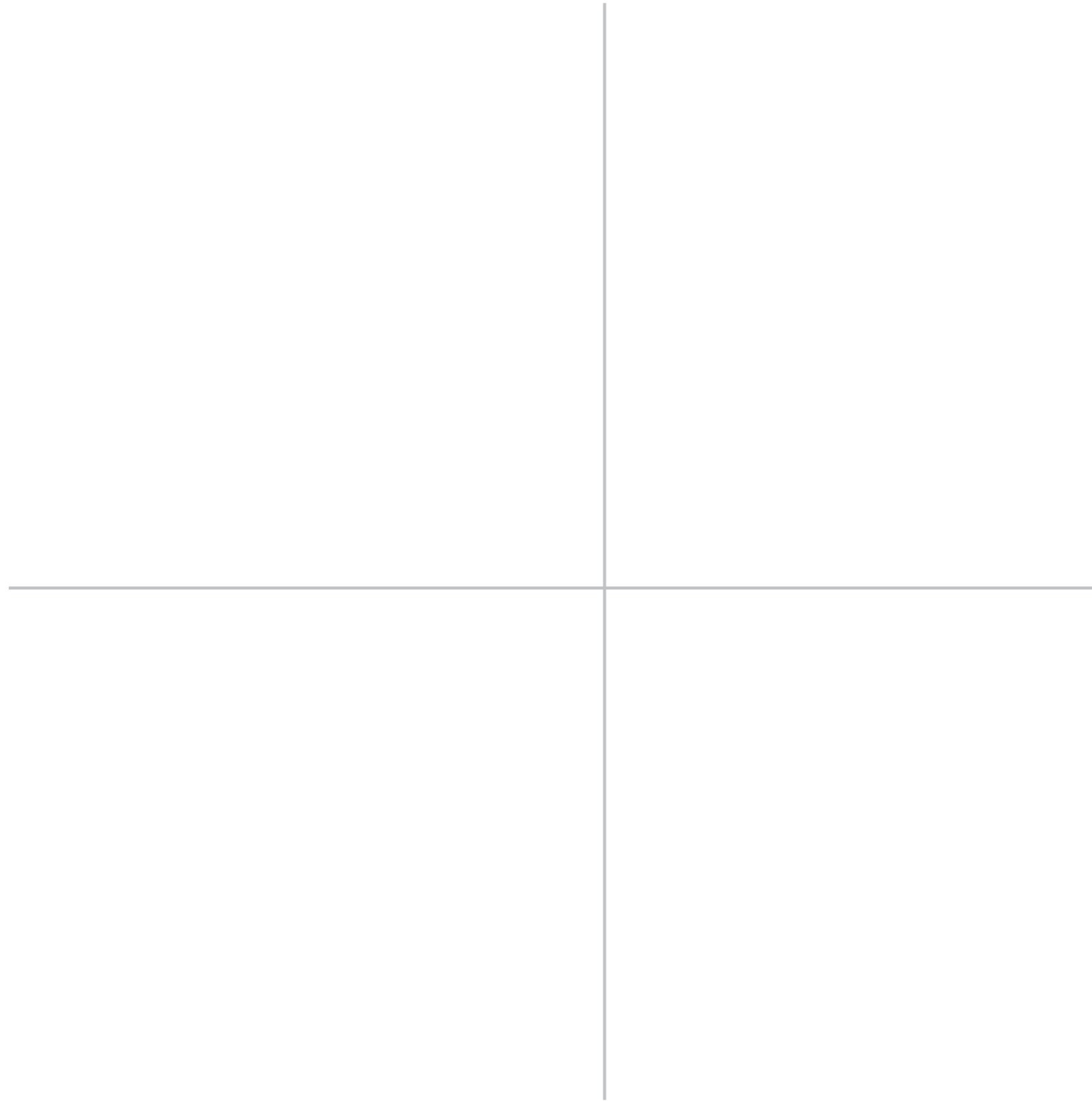


Components 1 - 50:
Physiology, metabolism,
cell-type differences

Component 800

Component 1 and 2

Transcriptional Component 2



Transcriptional Component 1

Component 1 and 2

Transcriptional Component 2

Transcriptional Component 1



Component 1 and 2

Transcriptional Component 2

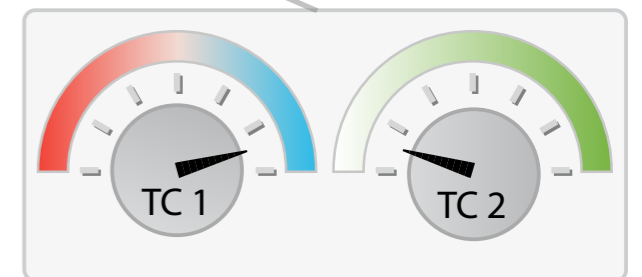
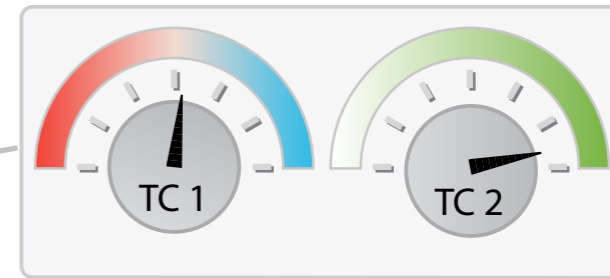
Transcriptional Component 1



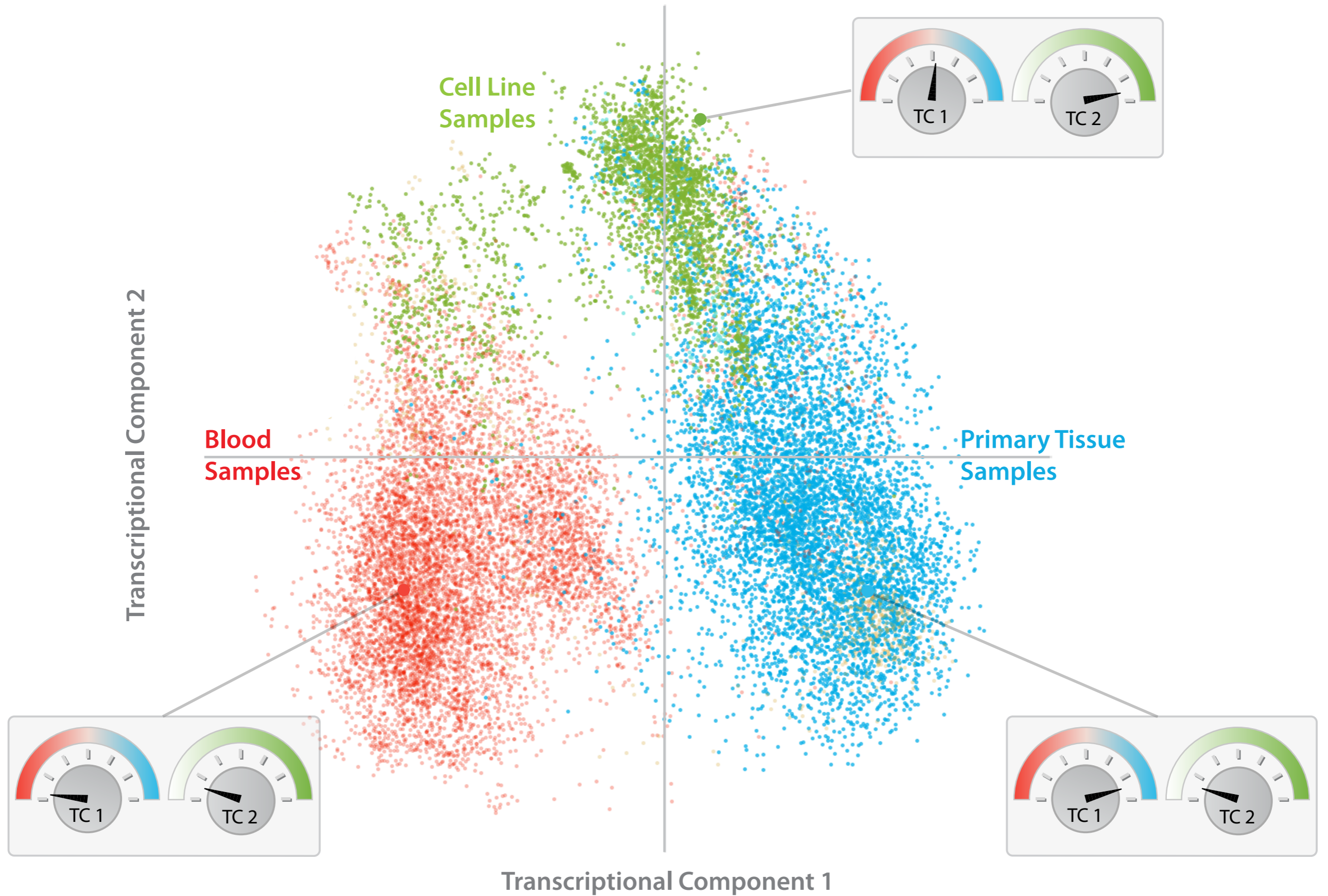
Component 1 and 2

Transcriptional Component 2

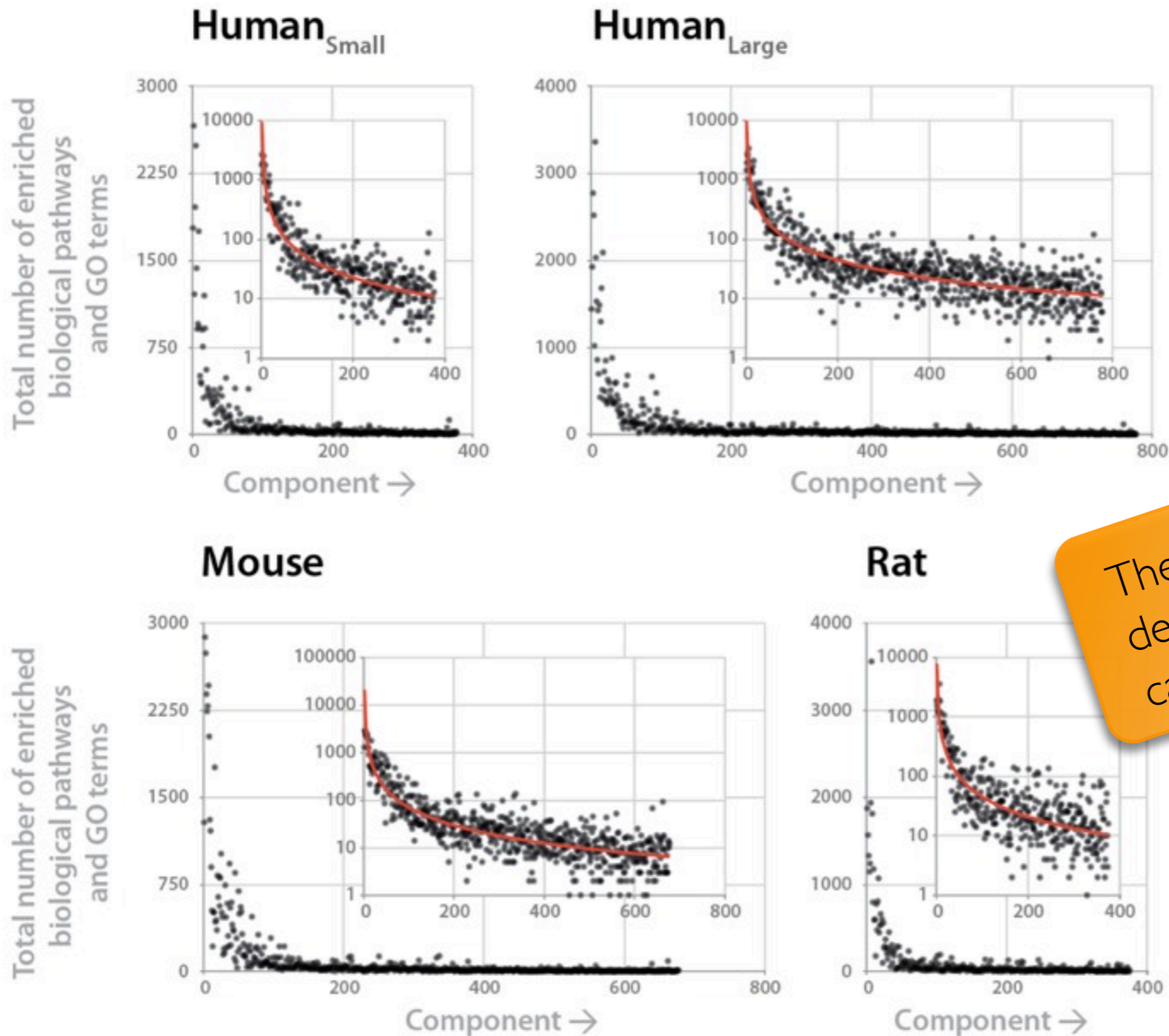
Transcriptional Component 1



Component 1 and 2

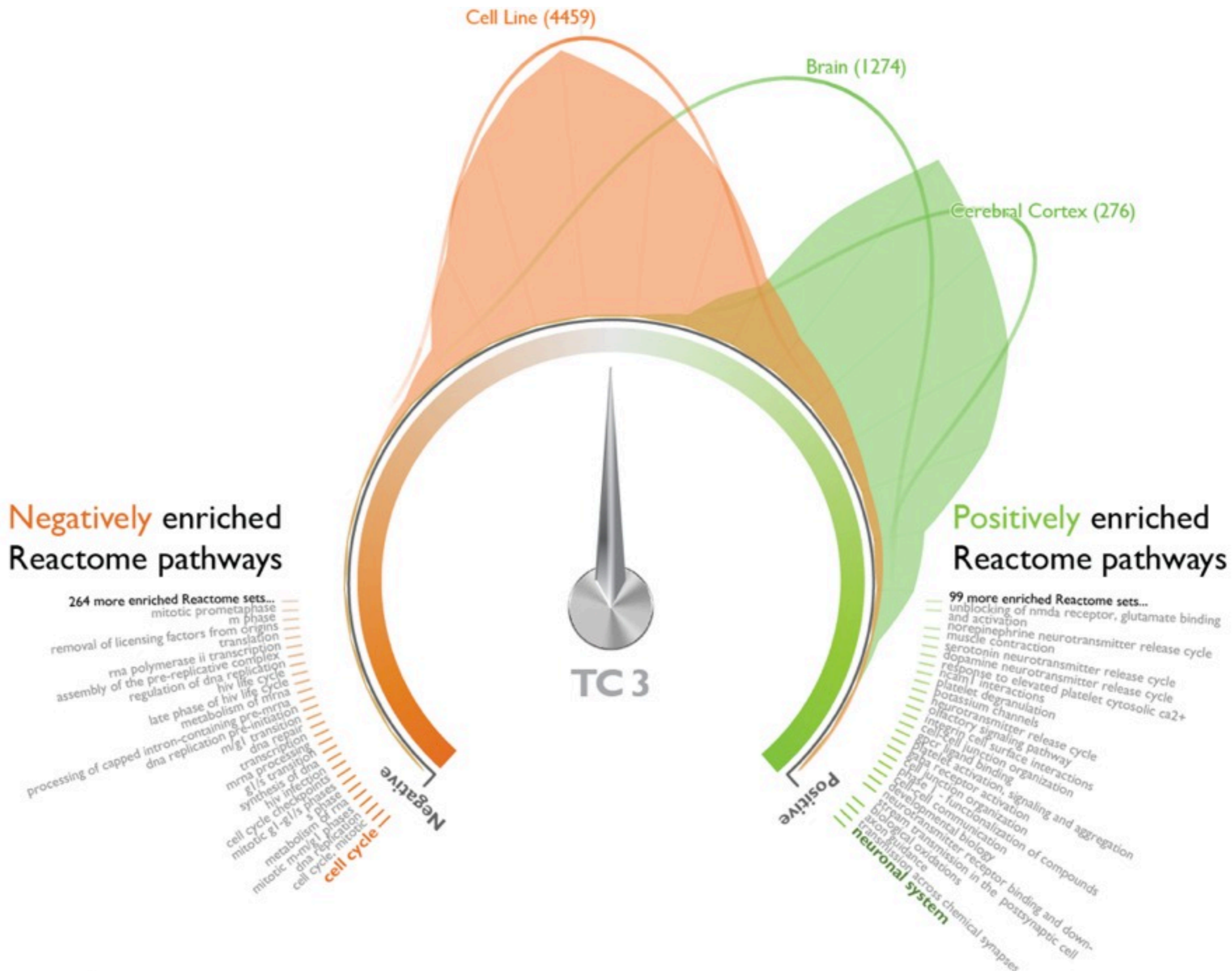


Three different species



These components describe clear biology, can we use this?

Transcriptional component 3

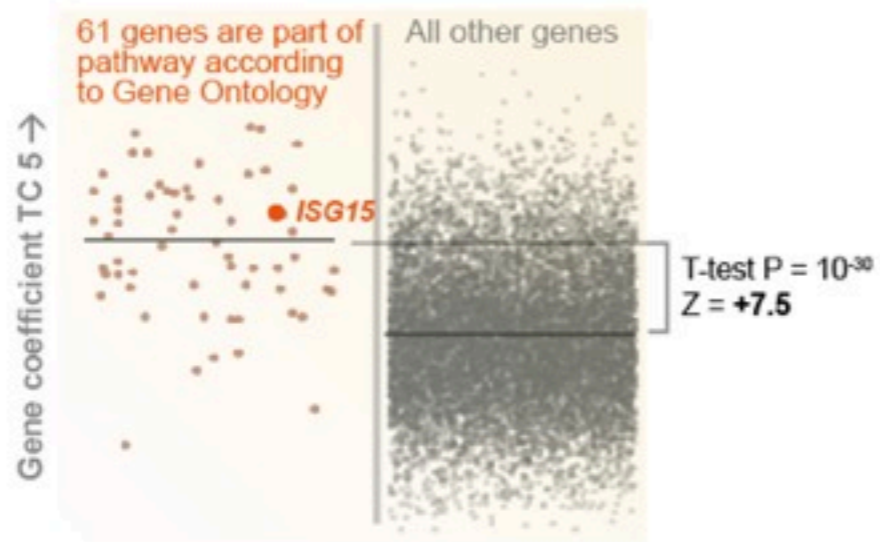


Predict function of genes

1

Build profile for a pathway (GO: *Type I interferon-mediated signaling pathway*)
Assess each of the 2,200 transcriptional components (TCs), perform T-Test per TC:

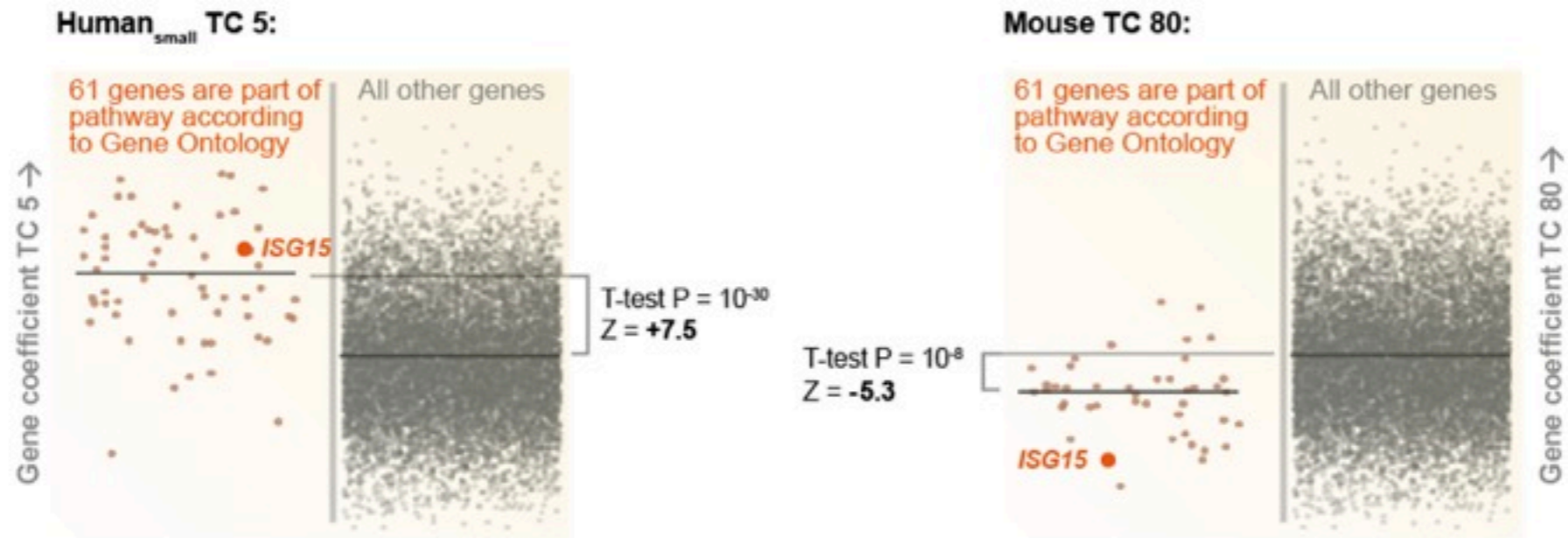
Human_{small} TC 5:



Predict function of genes

1

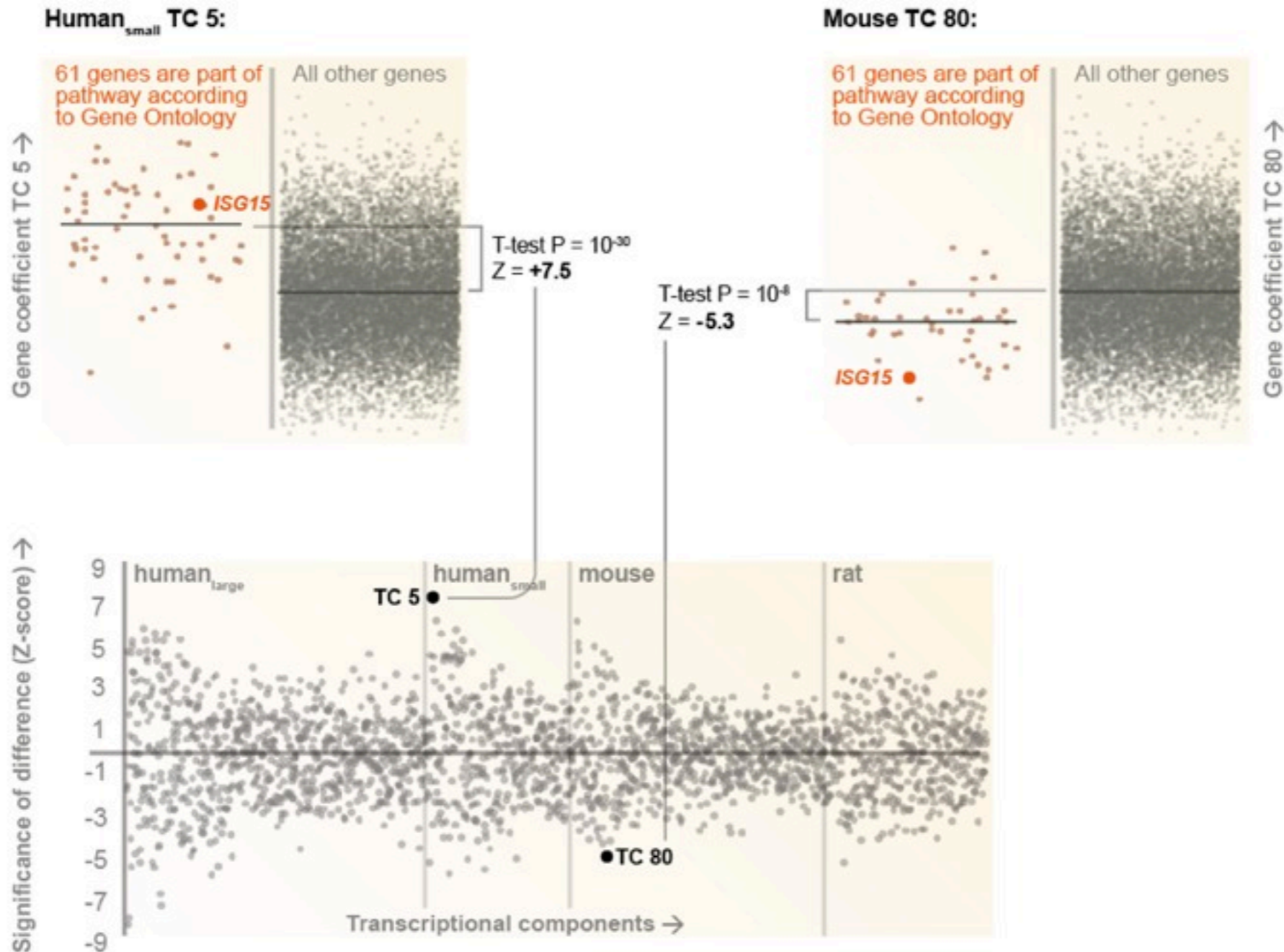
Build profile for a pathway (GO: *Type I interferon-mediated signaling pathway*)
Assess each of the 2,200 transcriptional components (TCs), perform T-Test per TC:



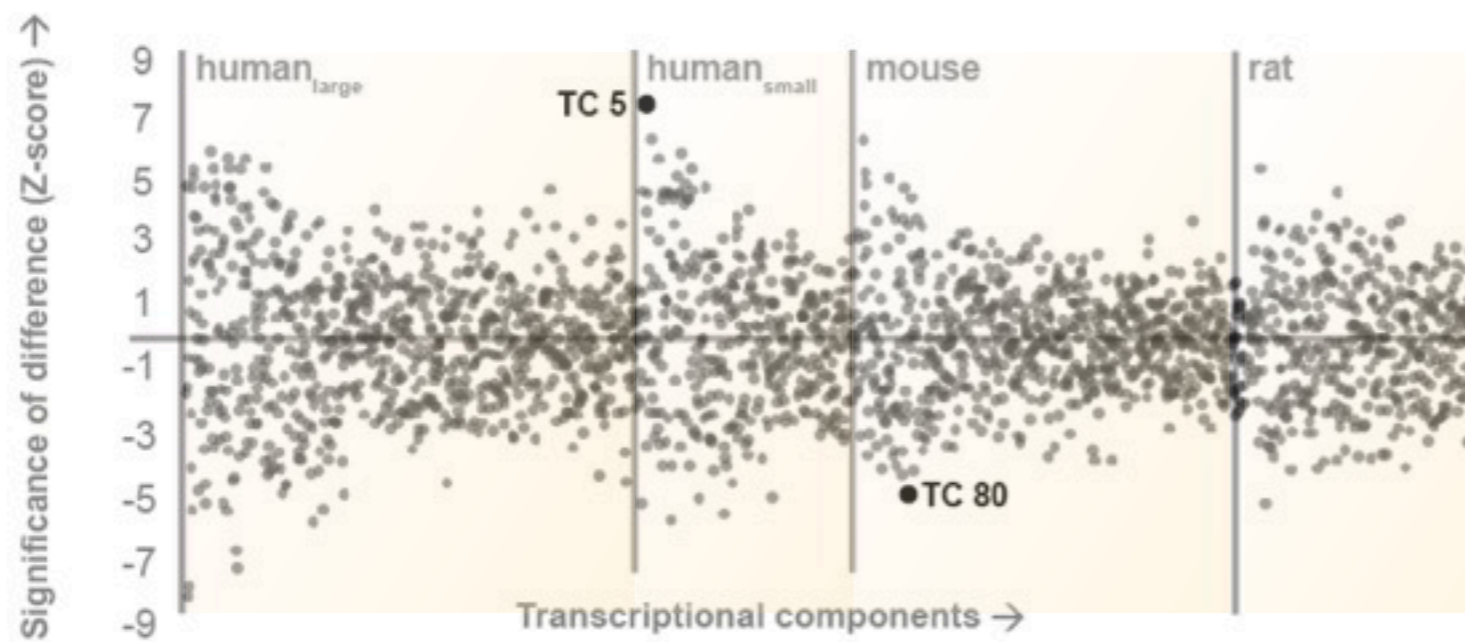
Predict function of genes

1

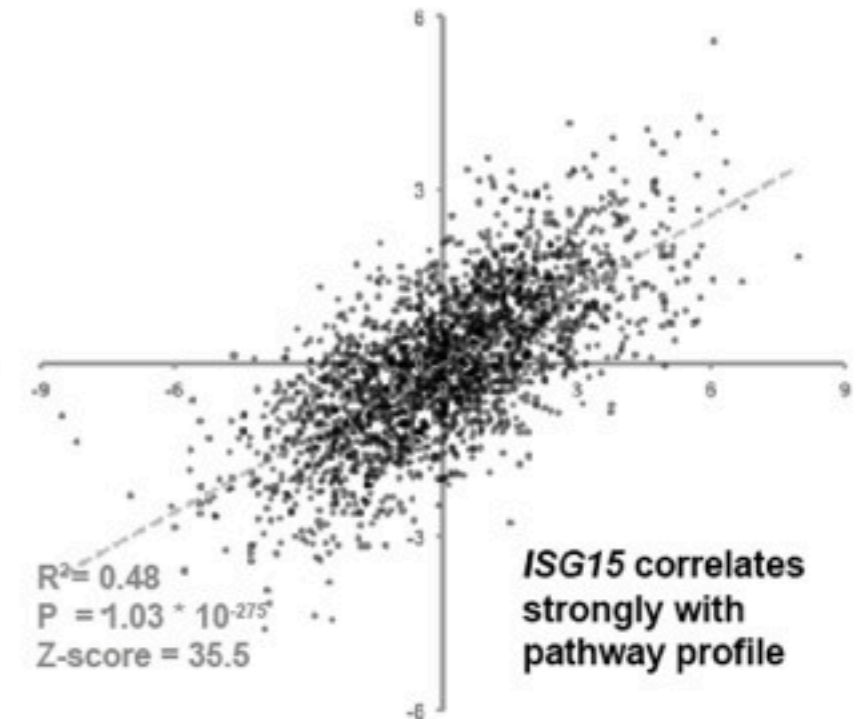
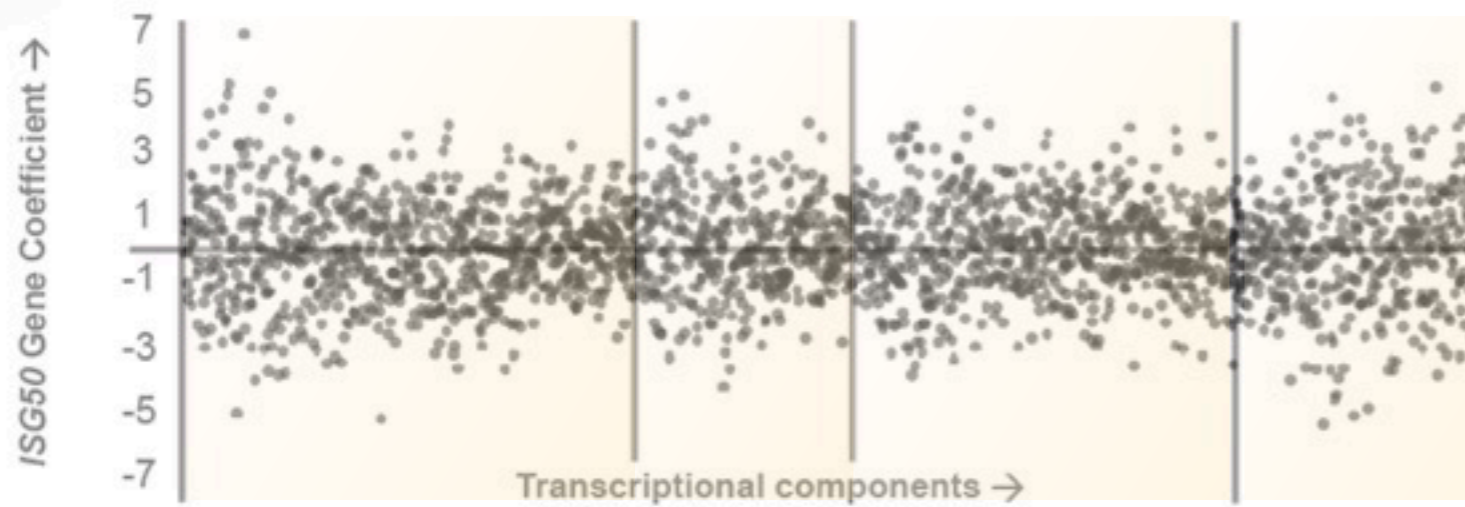
Build profile for a pathway (GO: *Type I interferon-mediated signaling pathway*)
Assess each of the 2,200 transcriptional components (TCs), perform T-Test per TC:



Predict function of genes

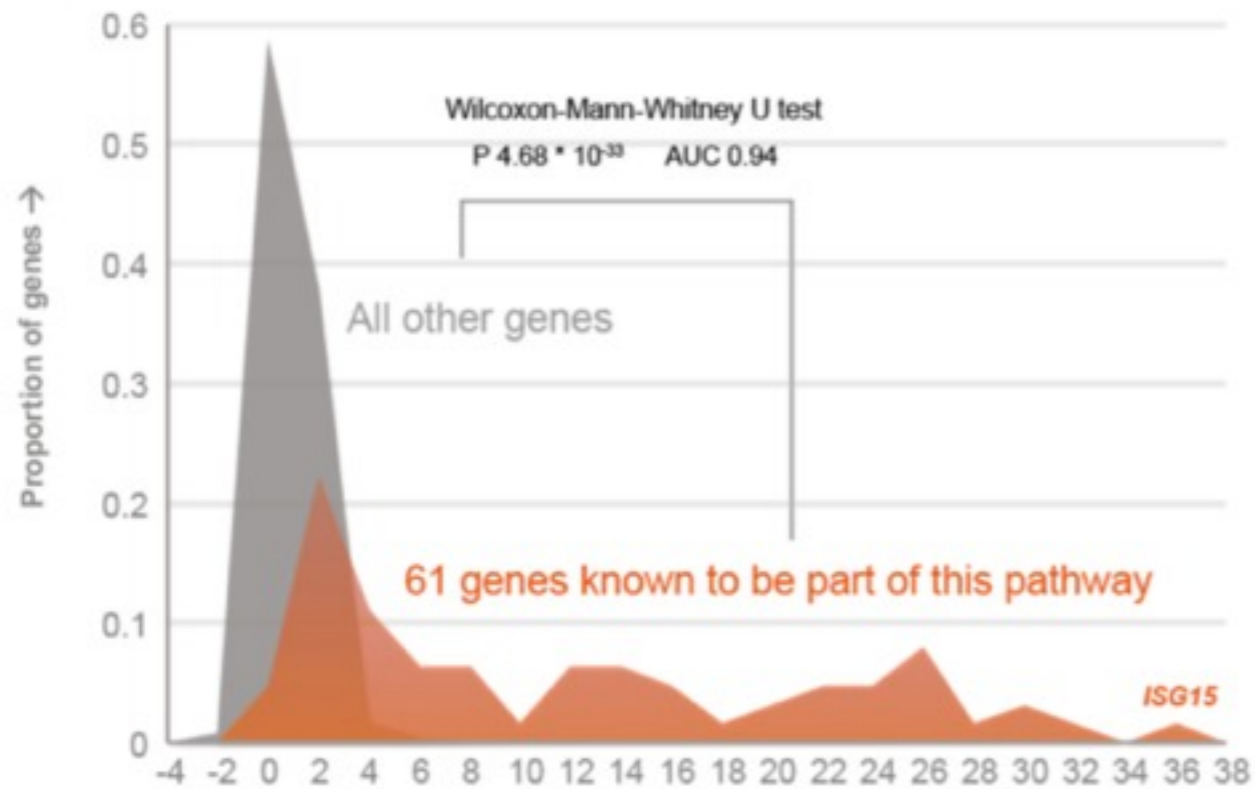


② Correlate individual genes with this profile. Example for *ISG15*:



Predict function of genes

③ Assess performance



Assess predictions yourself: Look at genenetwork.nl/genenetwork

Example: TP53

The screenshot shows a web browser window titled "Gene Network" with the URL <http://129.125.165.109:8080/GeneNetwork/?gene=tp53>. The page displays the "Gene Network" interface for the gene TP53 (Tumor protein p53). It includes a search bar, a "Get" button, and a "method about" link. The main content area is divided into tabs for "Predicted function", "Tissues", and "Network". Under "Predicted function", there are sub-tabs for "GO biological process", "GO cellular component", "GO molecular function", "KEGG", "BioCarta", "Reactome", "TFBS", and "MicroRNA". The "GO biological process" tab is active, showing a table of predicted functions for TP53.

Term	P-value	Direction	Annotated
signal transduction by p53 class mediator resulting in induction of apoptosis	1.25E-17	+	+
DNA damage response, signal transduction by p53 class mediator resulting in induction...	5.52E-13	+	+
response to UV	1.68E-11	+	+
induction of apoptosis by intracellular signals	3.09E-10	+	+
DNA damage response, signal transduction resulting in induction of apoptosis	5.22E-9	+	+
positive regulation of axonogenesis	6.79E-8	+	-
nuclear mRNA splicing, via spliceosome	9.37E-8	+	-
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	9.37E-8	+	-
RNA splicing, via transesterification reactions	1.88E-7	+	-
nuclear-transcribed mRNA poly(A) tail shortening	2.62E-7	+	-
mRNA 3'-end processing	4.67E-7	+	-
mRNA catabolic process	5.21E-7	+	-
induction of apoptosis	8.97E-7	+	+
induction of programmed cell death	9.12E-7	+	+
positive regulation of protein deacetylation	1.19E-6	+	+
DNA biosynthetic process	1.21E-6	+	-
nuclear-transcribed mRNA catabolic process	1.37E-6	+	-
serine family amino acid biosynthetic process	1.77E-6	+	-
RNA catabolic process	1.85E-6	+	-
RNA 3'-end processing	2.30E-6	+	-

Download all predictions for TP53

© 2006-2012 UMCG Department of Genetics

Example: TP53

Gene Network

TP53 Tumor protein p53

Predicted function | Tissues | Network

GO biological process | GO cellular component

Term
signal transduction by p53 class mediator result
DNA damage response, signal transduction by response to UV
induction of apoptosis by intracellular signals
DNA damage response, signal transduction response to DNA damage
positive regulation of axonogenesis
nuclear mRNA splicing, via spliceosome
RNA splicing, via transesterification reactions
RNA splicing, via transesterification reactions
nuclear-transcribed mRNA poly(A) tail shortening
mRNA 3'-end processing
mRNA catabolic process
induction of apoptosis
induction of programmed cell death
positive regulation of protein deacetylation
DNA biosynthetic process
nuclear-transcribed mRNA catabolic process
serine family amino acid biosynthetic process
RNA catabolic process
RNA 3'-end processing

Download all predictions for TP53

Gene Network

TP53 Tumor protein p53

Predicted function | Tissues | Network

Tissue	# samples	AUC	P-value
Retinal Pigment Epithelium	12	0.91	8×10^{-7}
Neural Stem Cells	11	0.88	2×10^{-5}
Umbilical Veins	113	0.86	8×10^{-41}
Astrocytes	12	0.84	4×10^{-5}
Endothelial Cells	196	0.84	4×10^{-60}
Veins	133	0.83	5×10^{-40}
Induced Pluripotent Stem Cells	35	0.82	3×10^{-11}
Cell Line, Transformed	102	0.82	3×10^{-29}
Trophoblasts	11	0.82	2×10^{-4}
HEK293 Cells	100	0.82	4×10^{-28}
Pluripotent Stem Cells	47	0.78	2×10^{-11}
Blood Vessels	171	0.77	4×10^{-34}
Embryoid Bodies	11	0.77	2×10^{-3}
HT29 Cells	17	0.74	8×10^{-4}
Oocytes	15	0.73	2×10^{-3}
Colon, Sigmoid	27	0.72	8×10^{-5}
Blastocyst	14	0.71	6×10^{-3}
Myocytes, Smooth Muscle	141	0.71	4×10^{-18}
Muscle Cells	146	0.70	5×10^{-17}
Foreskin	69	0.70	1×10^{-8}

Download all tissue data for TP53

© 2006-2012 UMCG Department of Genetics

Example: TP53

Gene Network

TP53 Tumor protein p53

Predicted function Tissues Network

GO biological process GO cellular component

Term
signal transduction by p53 class mediator resulting in induction of apoptosis
DNA damage response, signal transduction by p53 class mediator resulting in induction of apoptosis
response to UV
induction of apoptosis by intracellular signals
DNA damage response, signal transduction resulting in induction of apoptosis
positive regulation of axonogenesis
nuclear mRNA splicing, via spliceosome
RNA splicing, via transesterification reactions
RNA splicing, via transesterification reactions
nuclear-transcribed mRNA poly(A) tail shortening
mRNA 3'-end processing
mRNA catabolic process
induction of apoptosis
induction of programmed cell death
positive regulation of protein deacetylation
DNA biosynthetic process
nuclear-transcribed mRNA catabolic process
serine family amino acid biosynthetic process
RNA catabolic process
RNA 3'-end processing

Download all predictions for TP53

Gene Network

TP53 Tumor protein p53

Predicted function Tissues Network

Tissue	# samples	AUC
Retinal Pigment Epithelium	12	0.91
Neural Stem Cells	11	0.88
Umbilical Veins	113	0.86
Astrocytes	12	0.84
Endothelial Cells	196	0.84
Veins	133	0.83
Induced Pluripotent Stem Cells	35	0.82
Cell Line, Transformed	102	0.82
Trophoblasts	11	0.82
HEK293 Cells	100	0.82
Pluripotent Stem Cells	47	0.78
Blood Vessels	171	0.77
Embryoid Bodies	11	0.77
HT29 Cells	17	0.74
Oocytes	15	0.73
Colon, Sigmoid	27	0.72
Blastocyst	14	0.71
Myocytes, Smooth Muscle	141	0.71
Muscle Cells	146	0.70
Foreskin	69	0.70

Download all tissue data for TP53

Gene Network

TP53 Tumor protein p53

Predicted function Tissues Network

Color genes based on GO biological process 93 genes shown

- DNA damage response, signal transduction by p53 class mediator resulting in induction of apoptosis
- signal transduction by p53 class mediator resulting in induction of apoptosis
- response to UV
- induction of apoptosis by intracellular signals
- DNA damage response, signal transduction resulting in induction of apoptosis
- weak positive coexpression
- strong positive coexpression

Download image (PDF)

VEL blood group, gene unknown

- GWAS on red blood cell traits (Van der Harst *et al*, Nature 2012)

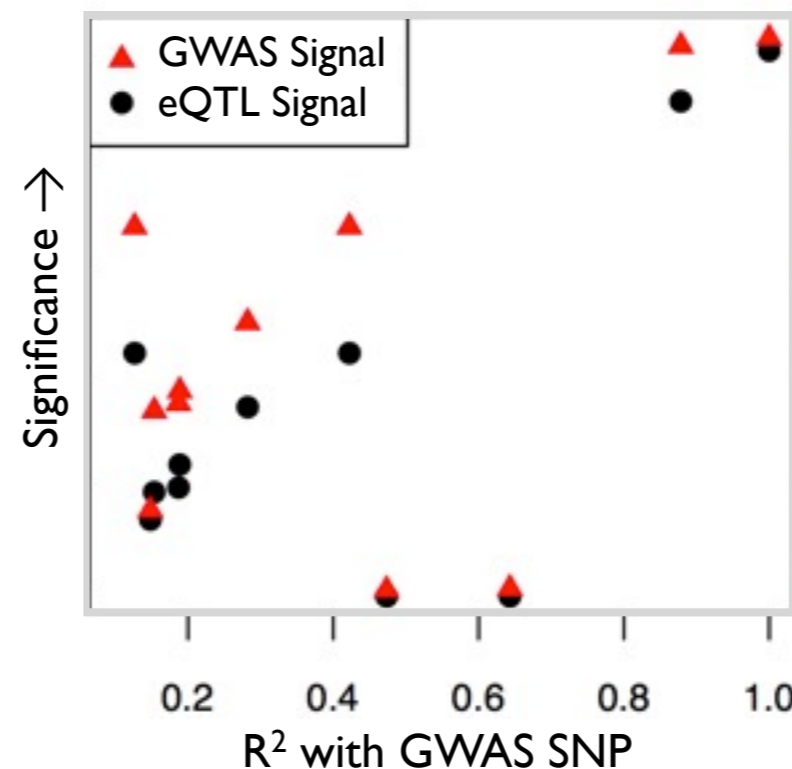
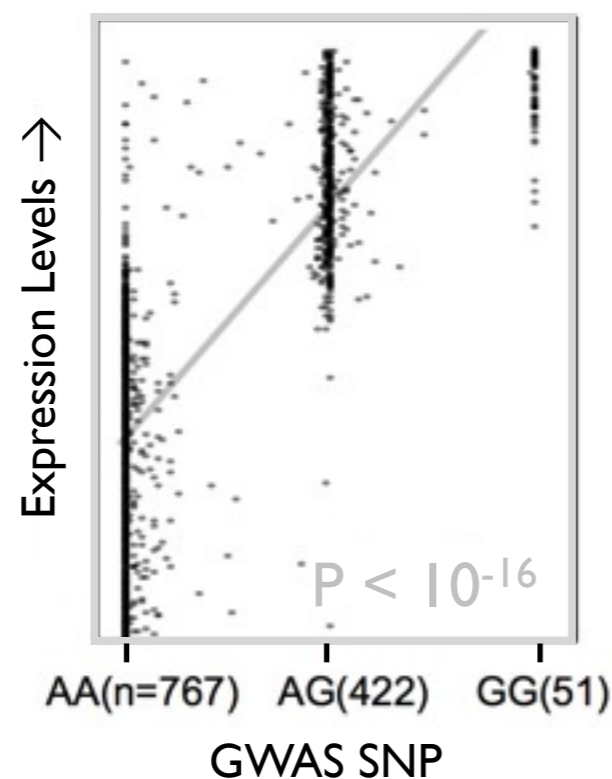
ARTICLE

nature

doi:10.1038/nature11677

Seventy-five genetic loci influencing the human red blood cell

- In one locus the SNP strongly affected gene expression of SMIM1, a gene without known function:



VEL blood group, *SMIM1* gene

- Most significantly predicted GeneNetwork function: 'Hemoglobin metabolic process' ($P = 10^{-16}$)
- Homozygous *SMIM1* 17bp deletion found in 63 out of 69 individual (Mutation frequency $< 1\%$ in entire population)
- Zebrafish knock-down: reduced number of red blood cells

Control fish

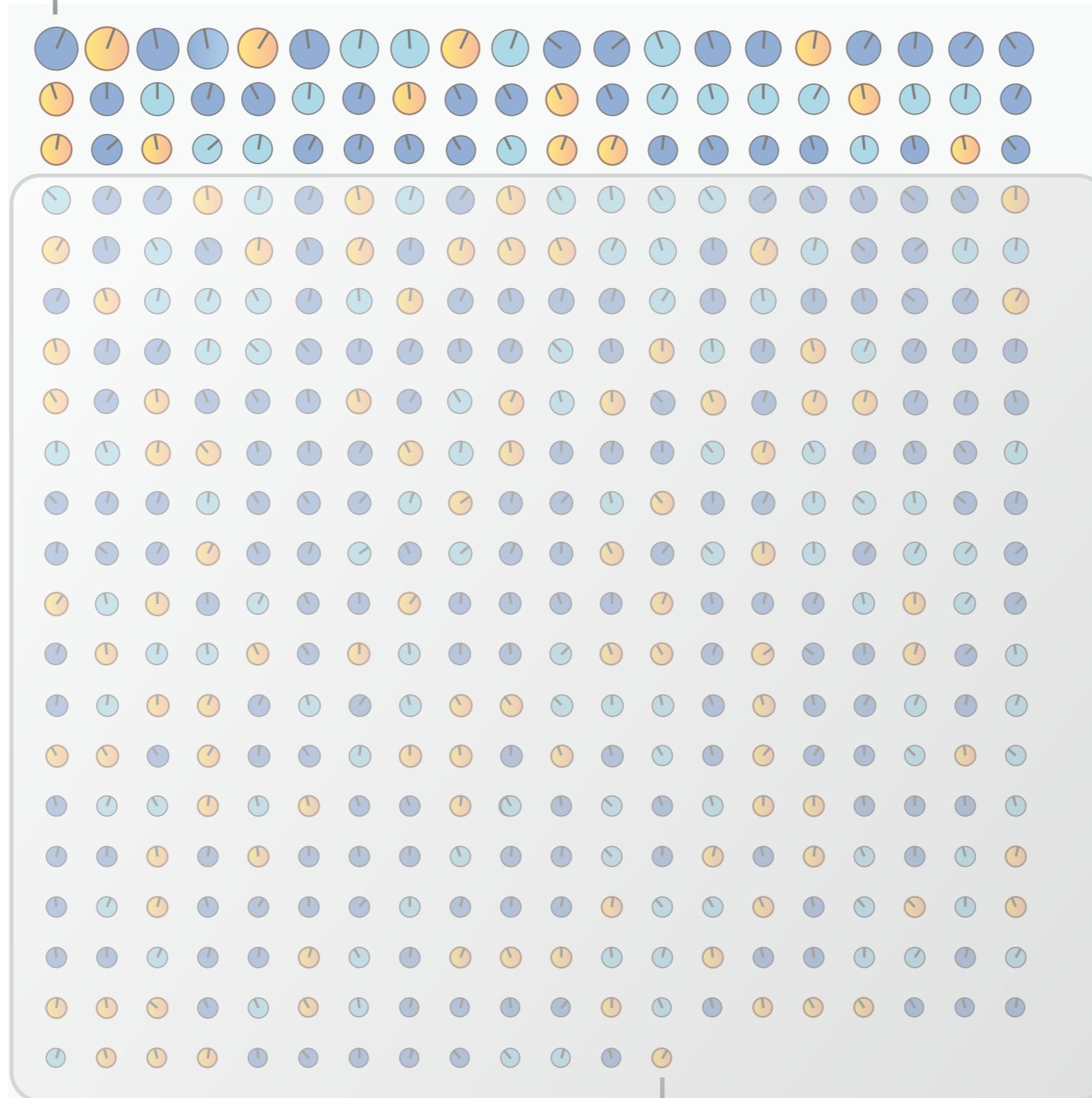


SMIM1 knock-down



Components 51 - 800

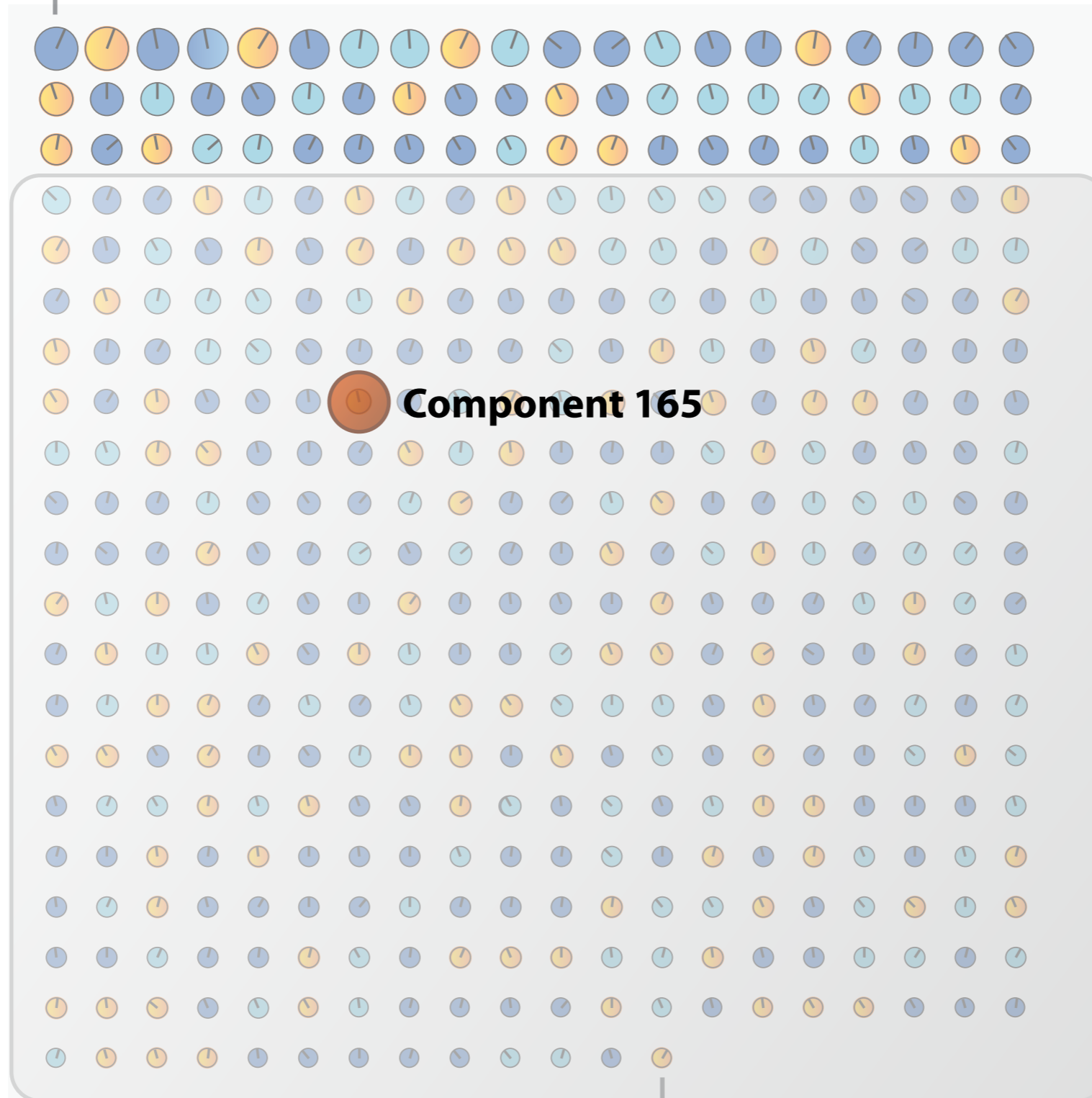
Component 1



Component 800

Components 51 - 800

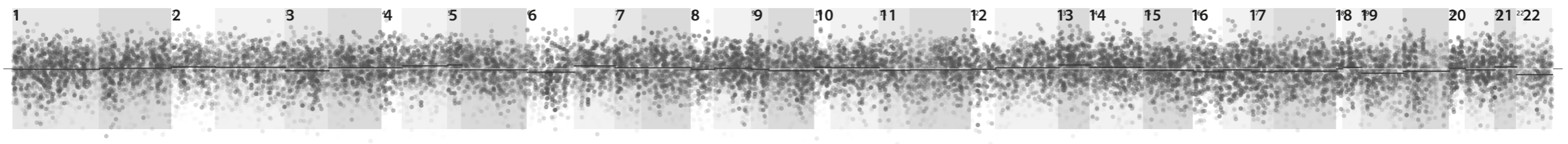
Component 1



Component 800

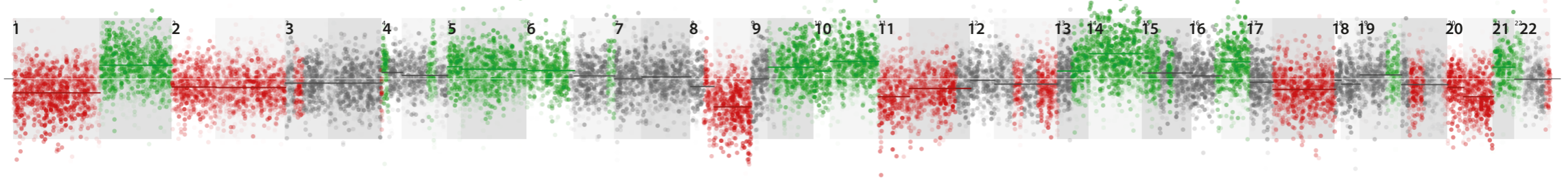
Some component show weird behaviour

TC 1: No cytogenetic effect, zero autocorrelation

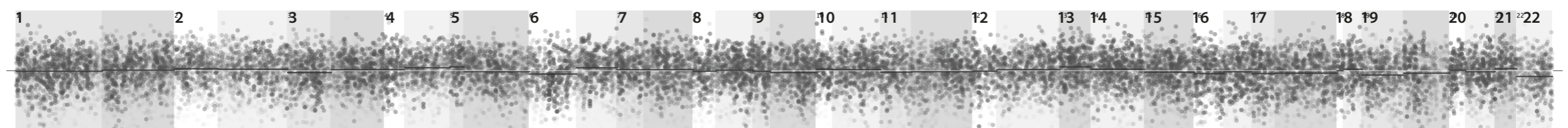


Some component show weird behaviour

TC 165: Strong cytogenetic effects, high autocorrelation

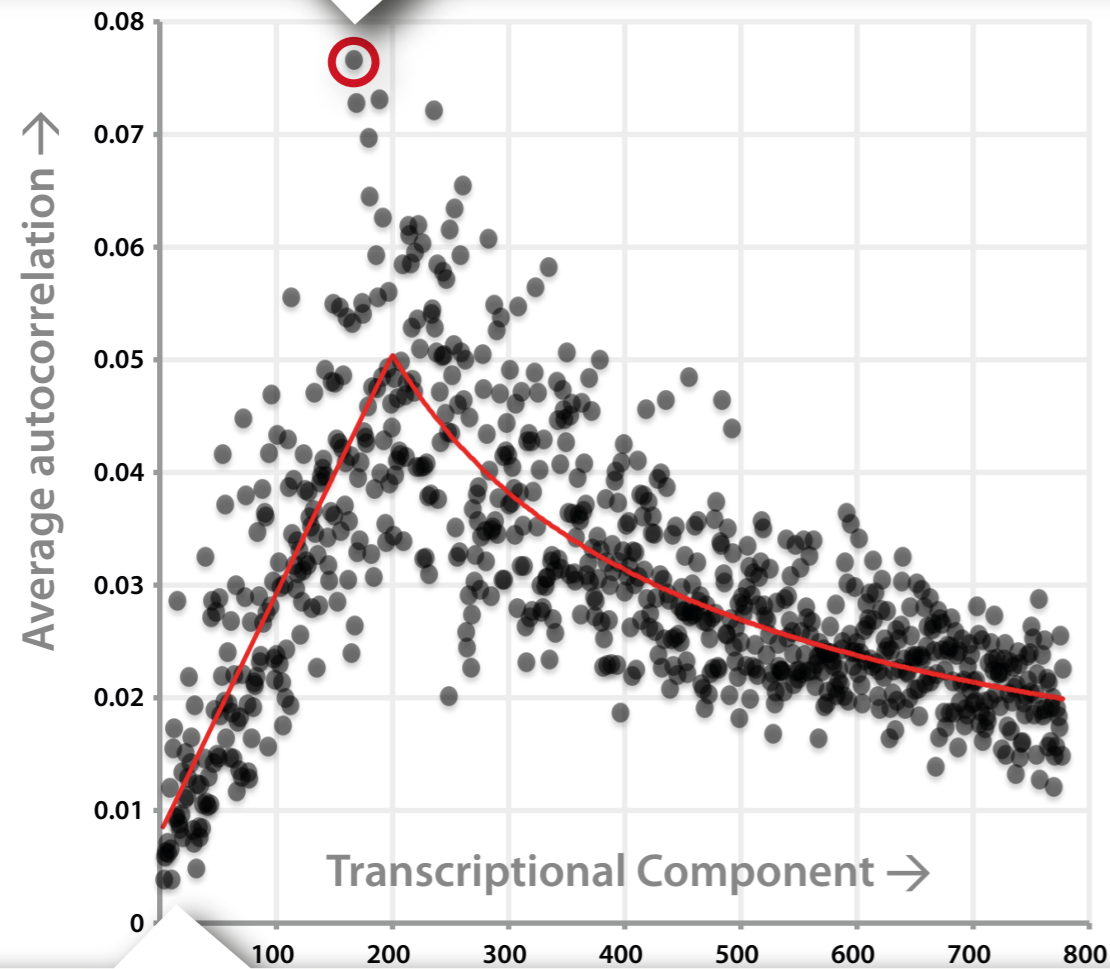
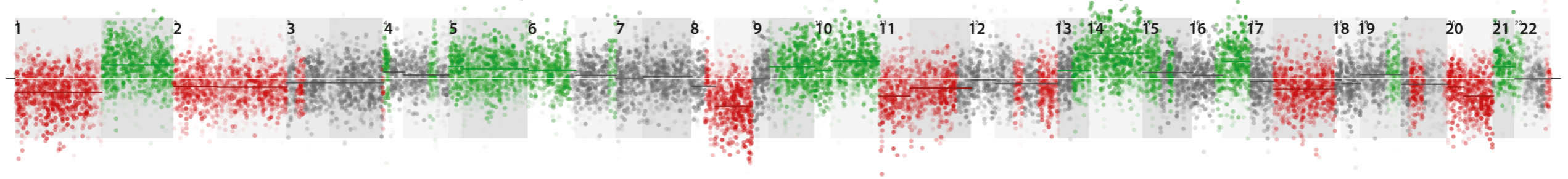


TC 1: No cytogenetic effect, zero autocorrelation

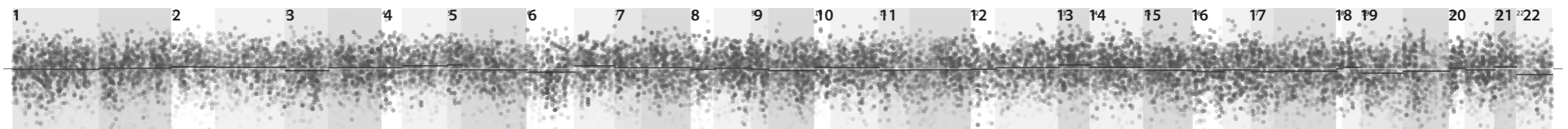


Some component show weird behaviour

TC 165: Strong cytogenetic effects, high autocorrelation



TC 1: No cytogenetic effect, zero autocorrelation



Identify non-cancer, physiological TCs

Take 50% of 37,427 samples with least genomic instability

PCA on 54,000 x 54,000
probeset correlation matrix
(dedicated multi-threaded
optimized C++ code, ensuring
numerical stability)

718 highly robust Principal Components (Cronbach's Alpha > 0.7)

Treat these as orthogonal
covariates, correct the expres-
sion data of all 37,427 samples,
take residual expression data

'Cytogenetic' RNA expression, only use 50%
of 37,427 with most genomic instability

Detection cytogenetic aberration in expression data

Chromosome

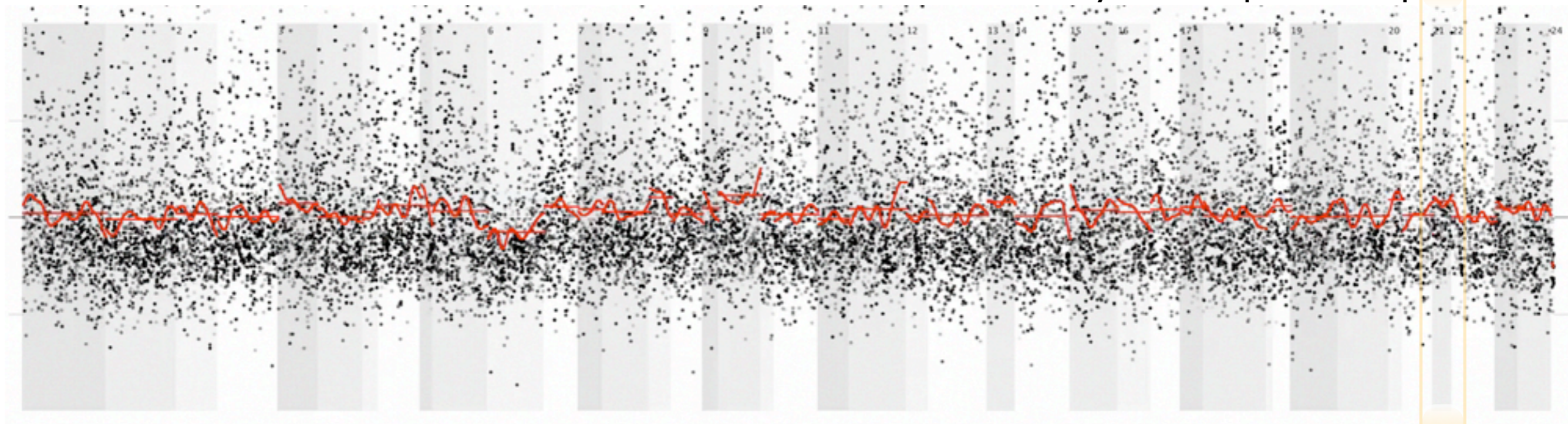
Down Syndrome patient: dup 21



Detection cytogenetic aberration in expression data

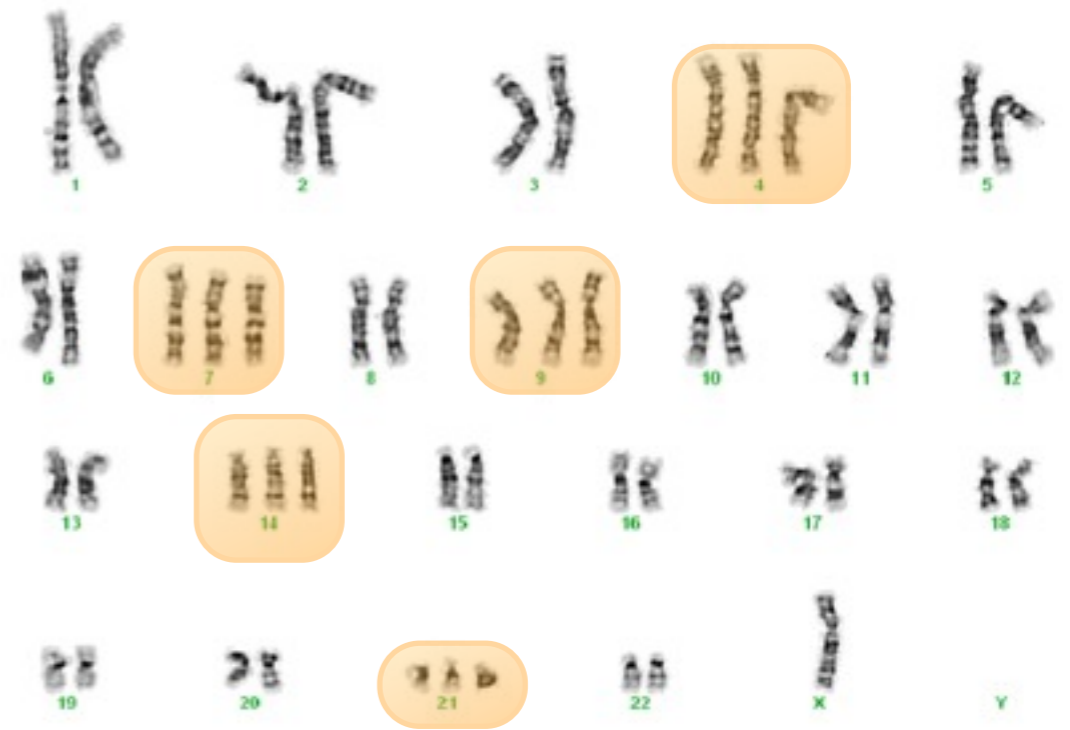
Chromosome

Down Syndrome patient: dup 21



Identifying five chromosome duplications

Karyogram
HapMap LCL



Chromosome

4

7

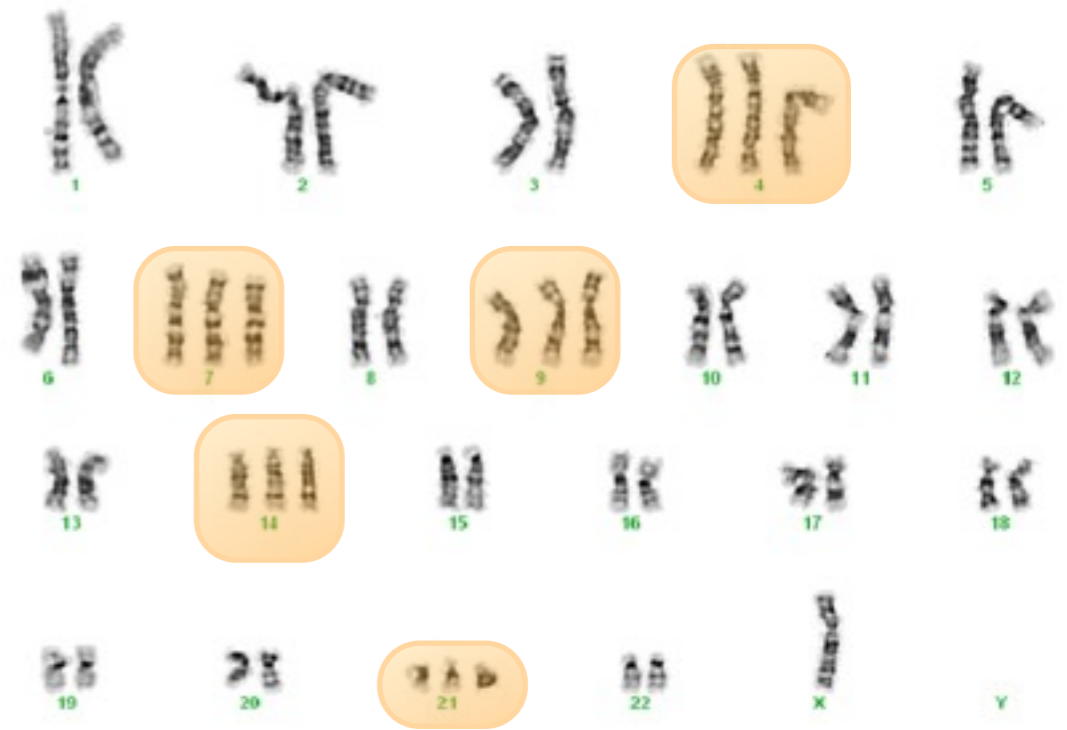
9

14

21

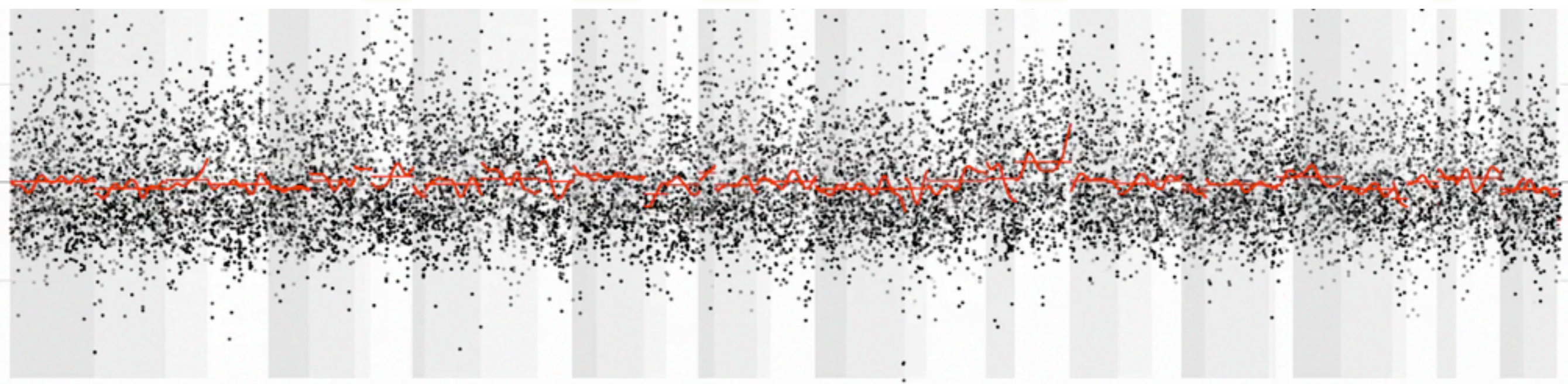
Identifying five chromosome duplications

Karyogram
HapMap LCL



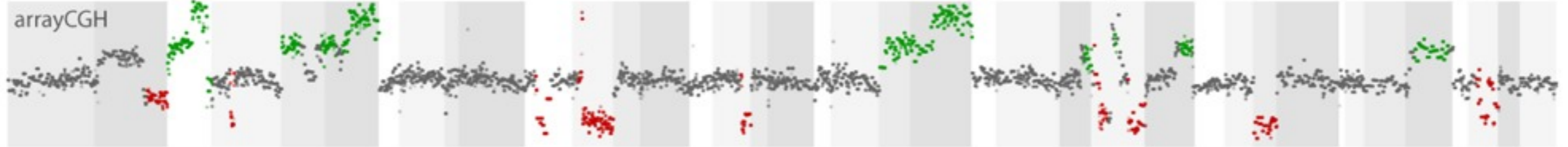
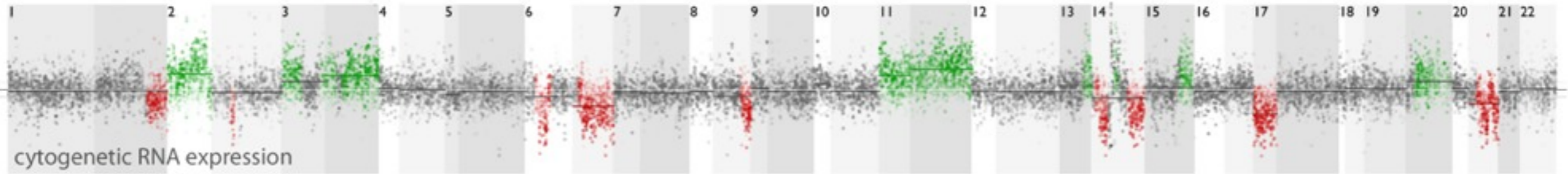
Chromosome

4 7 9 14 21

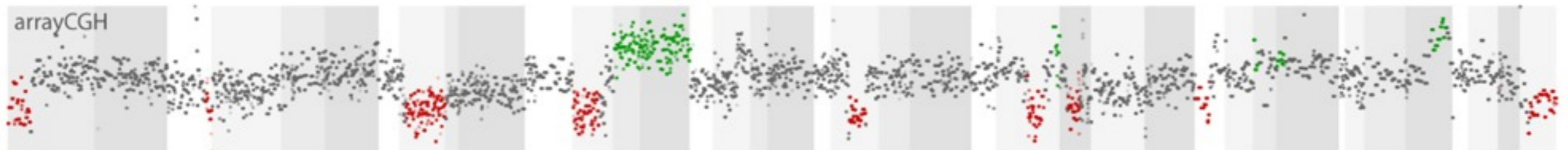
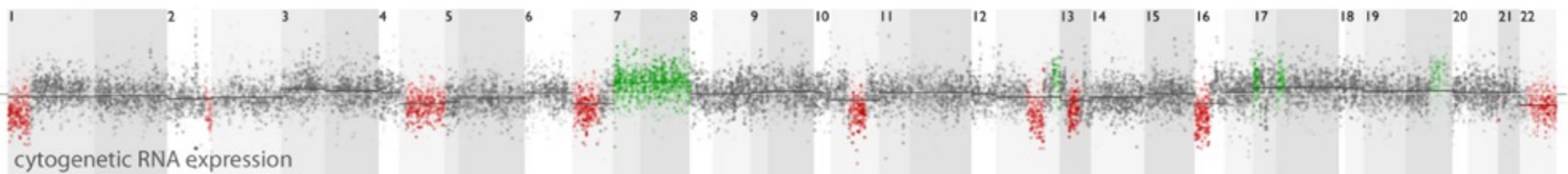


Comparison of arrayCGH and cytogenetic RNA profiles

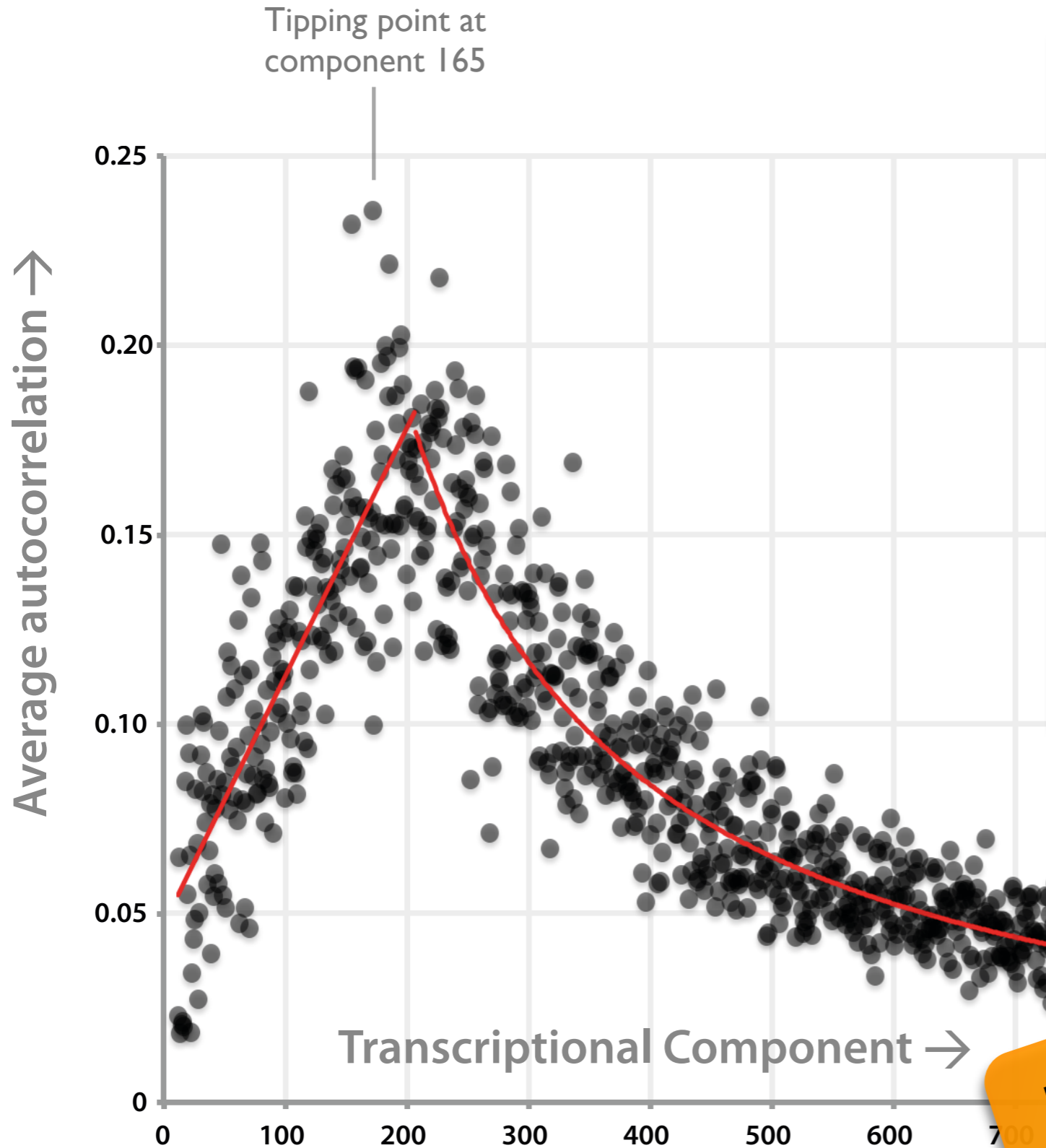
GSM274996



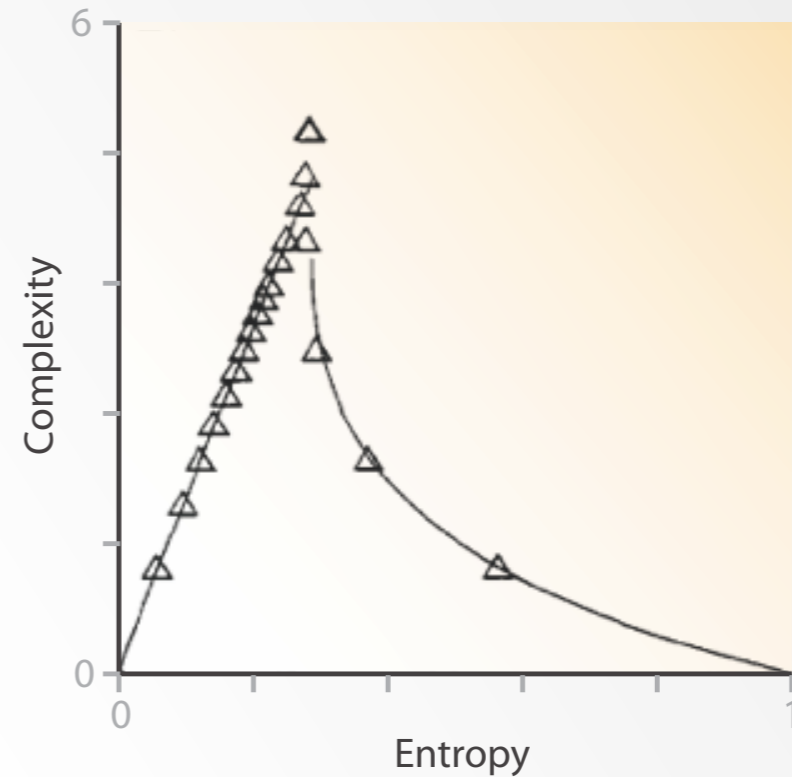
GSM275008



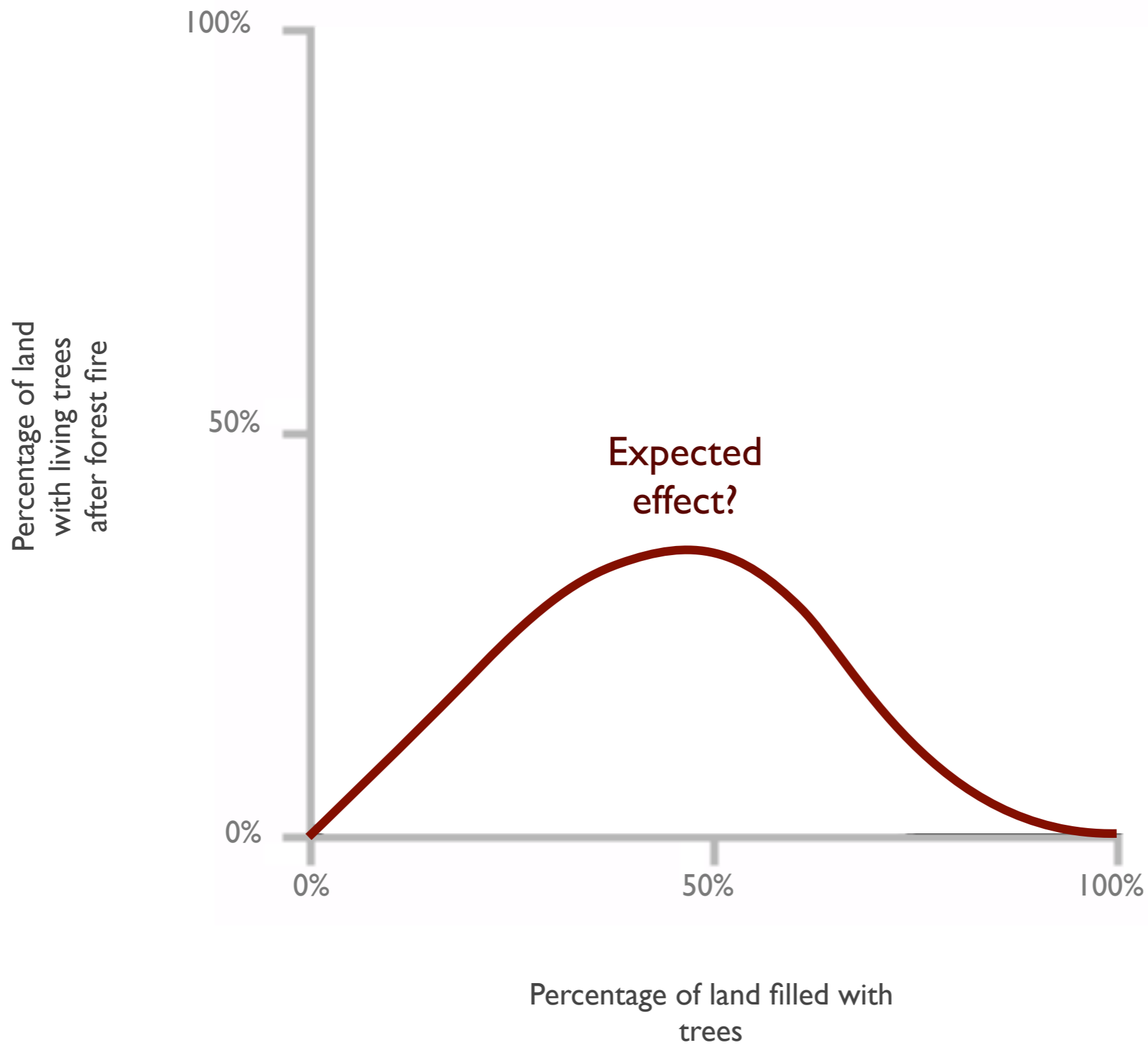
Amount of cytogenetic aberrations

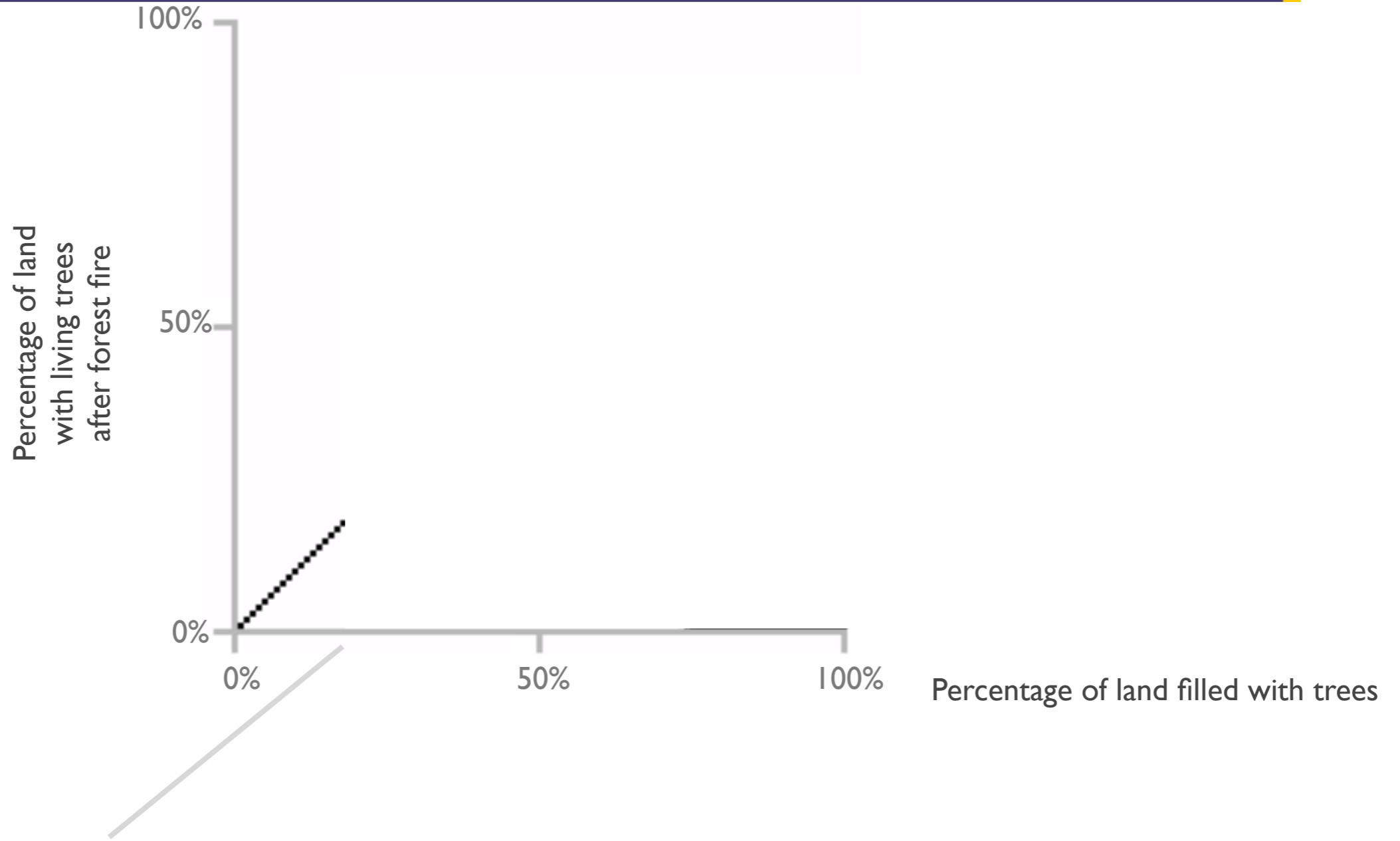


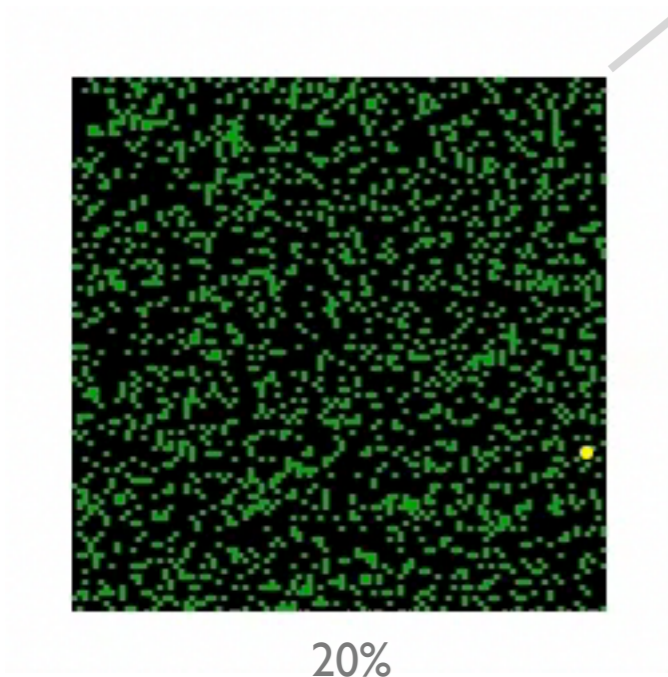
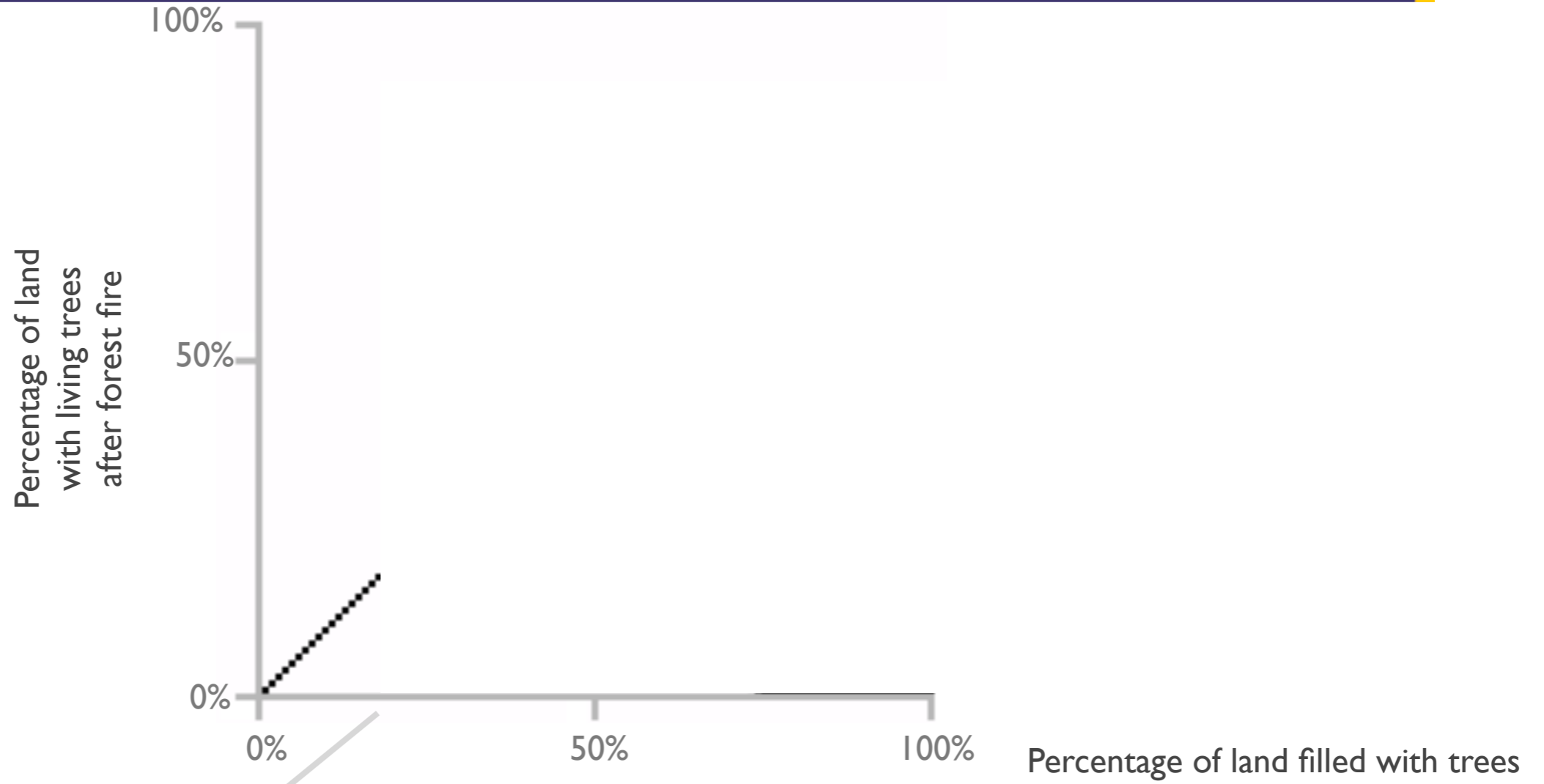
Transition to chaos in the logistic map
Crutchfield et al, 1990

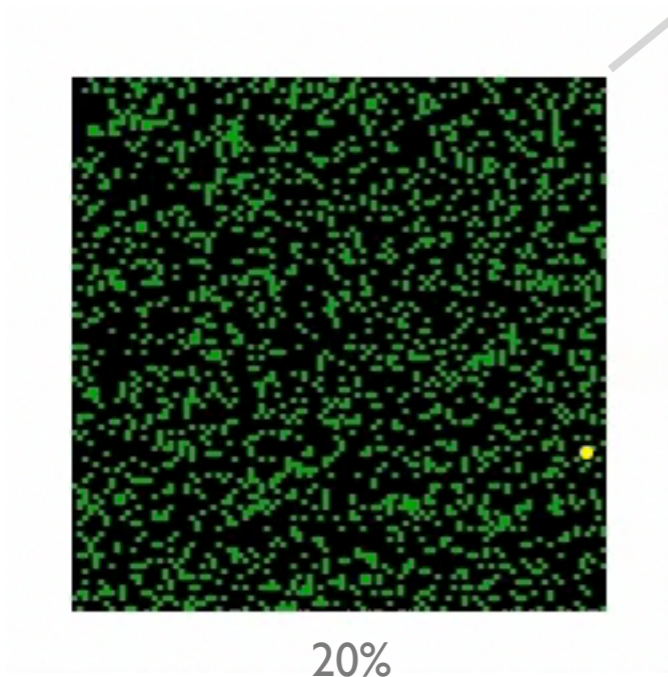
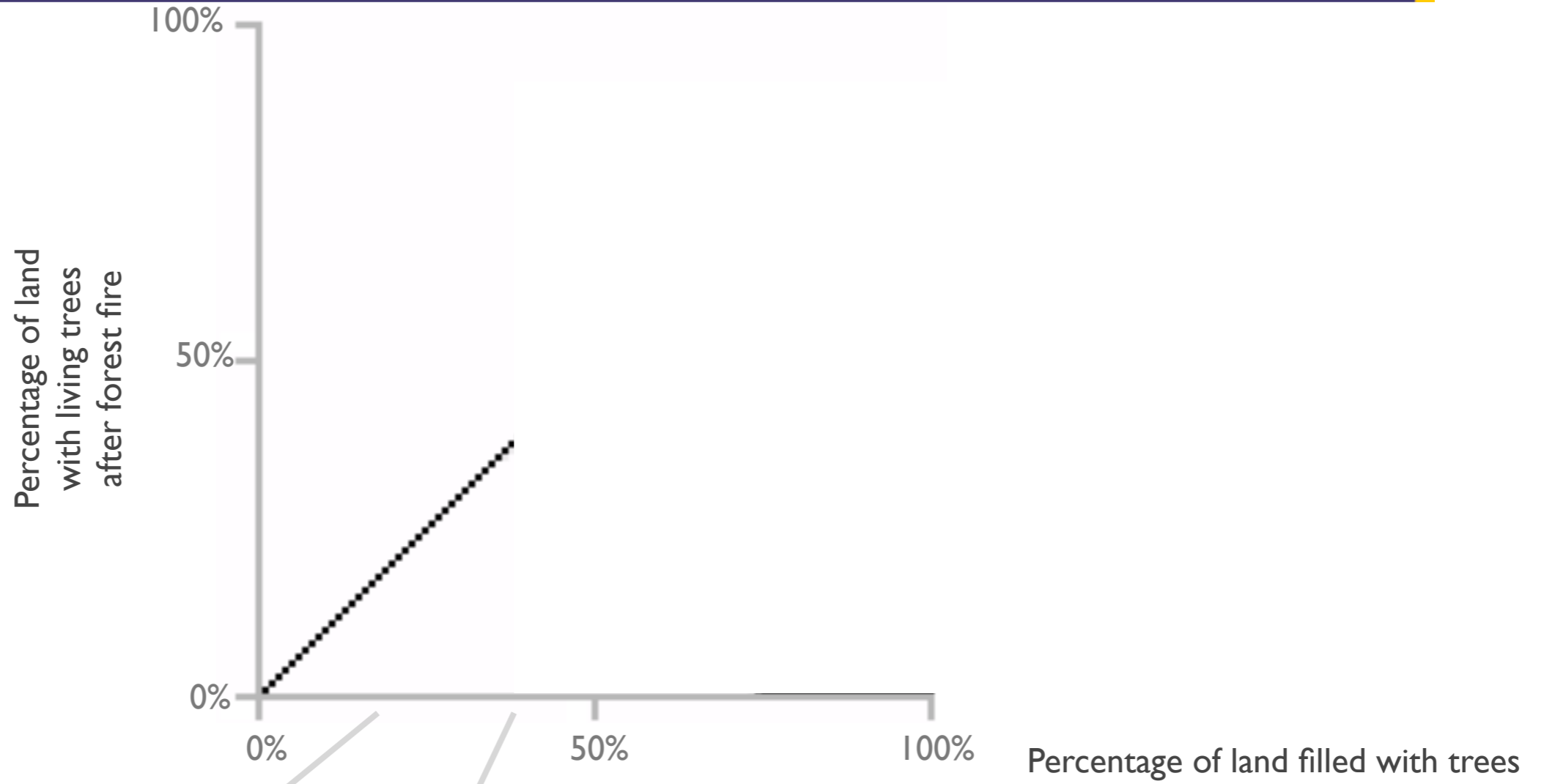


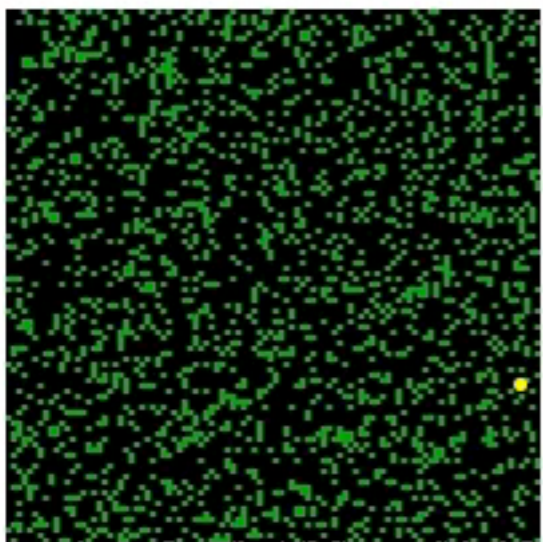
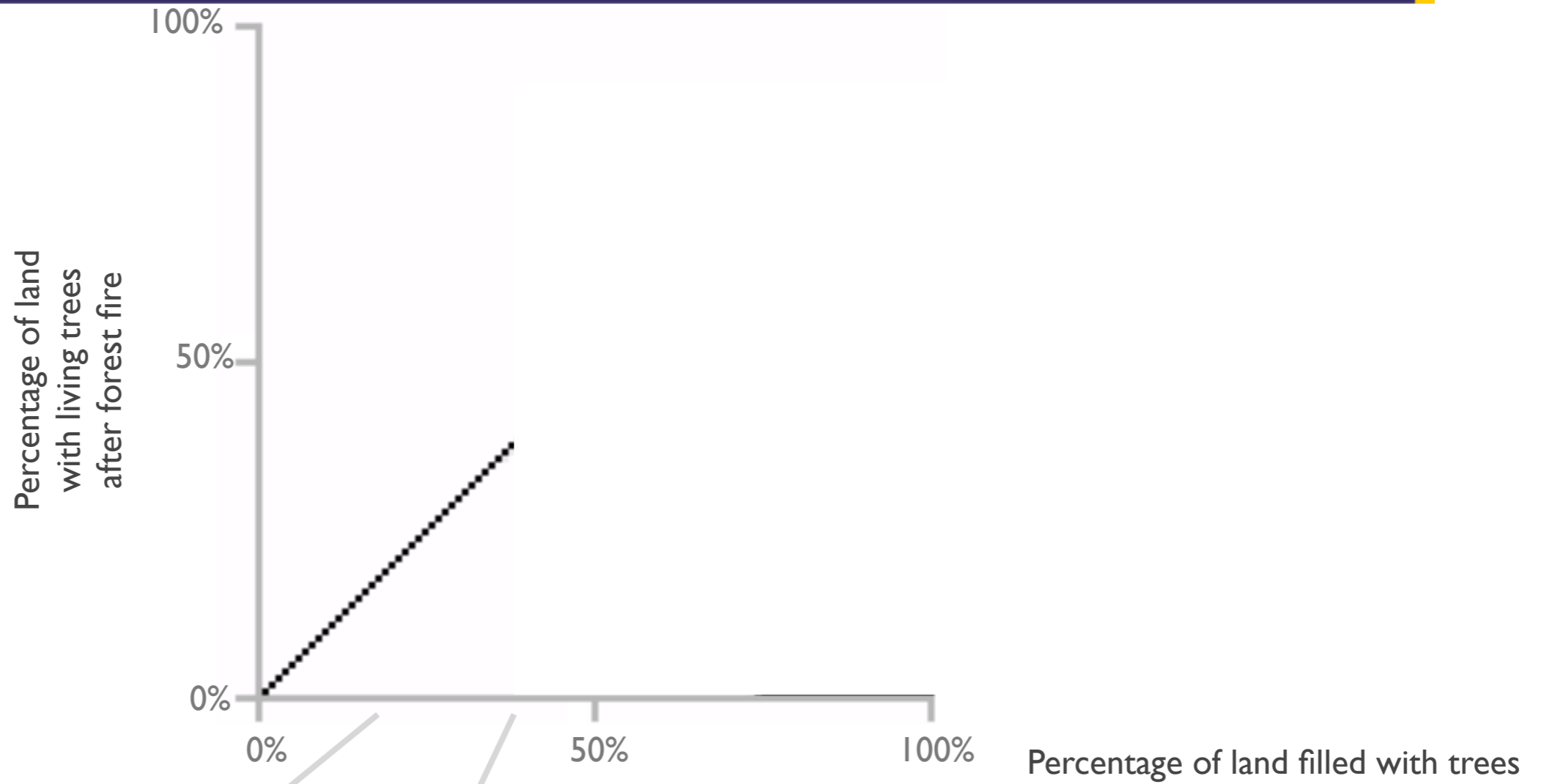
Distribution identical to simulations in complexity theory



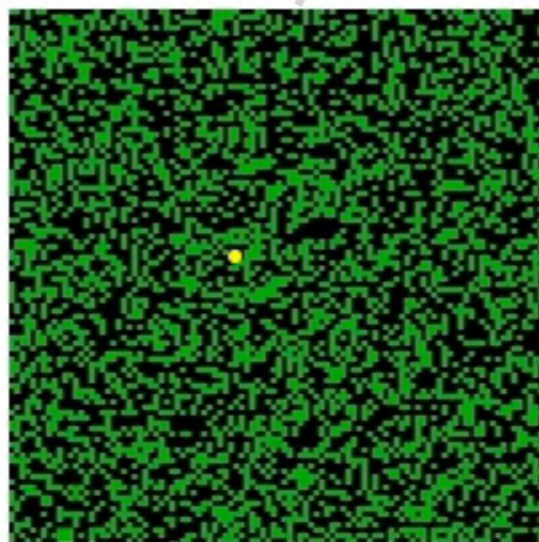




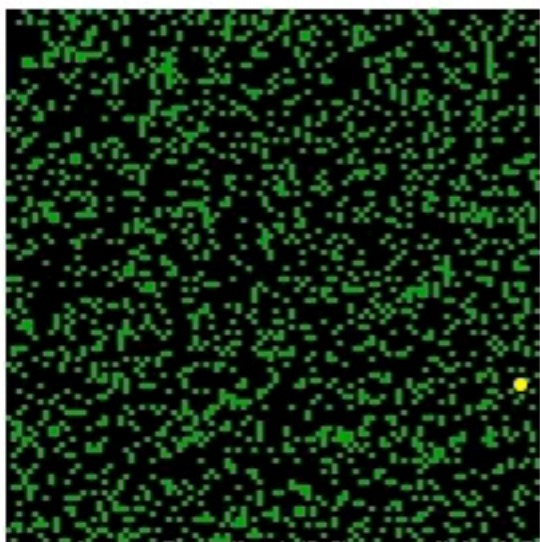
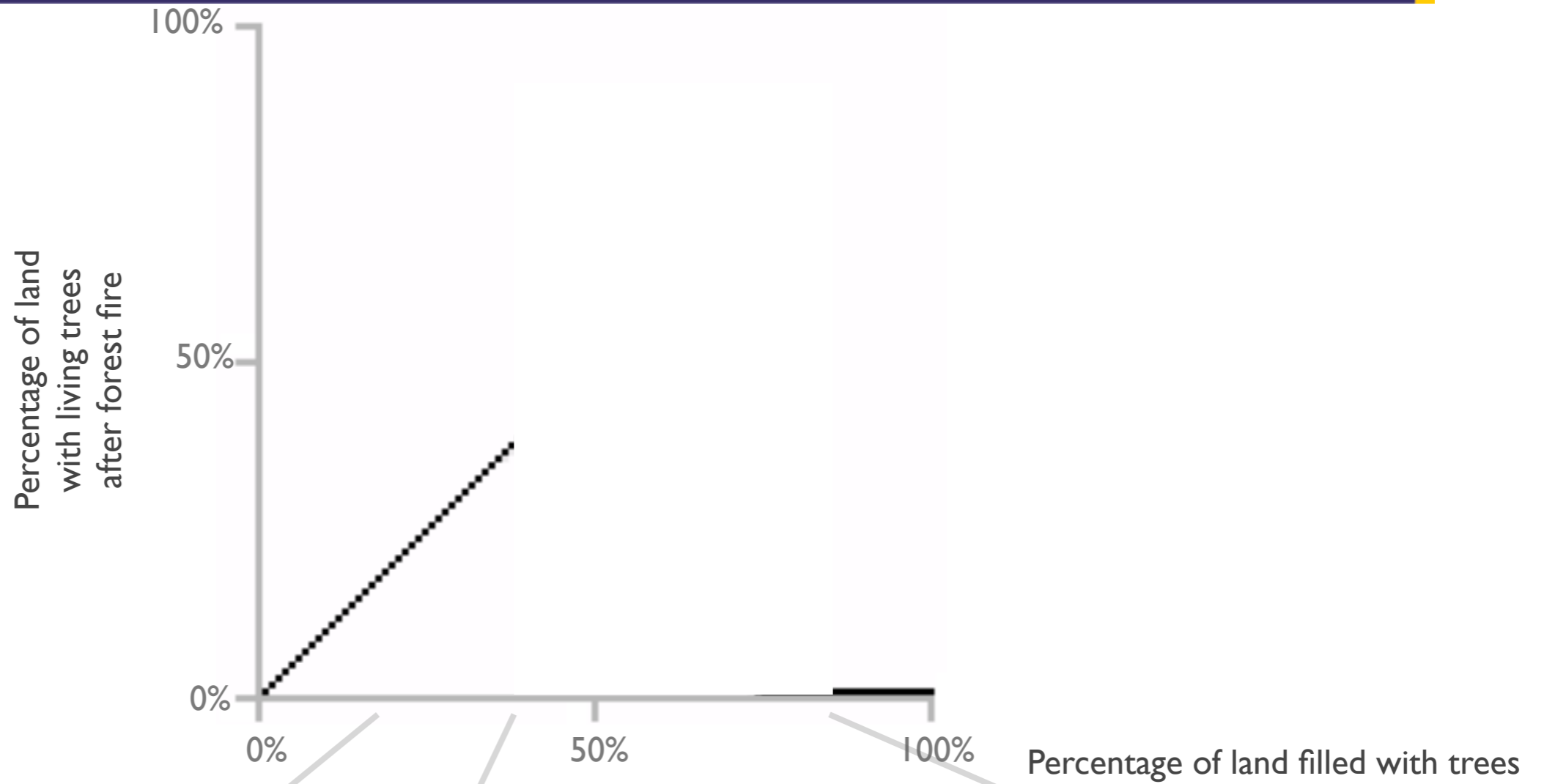




20%



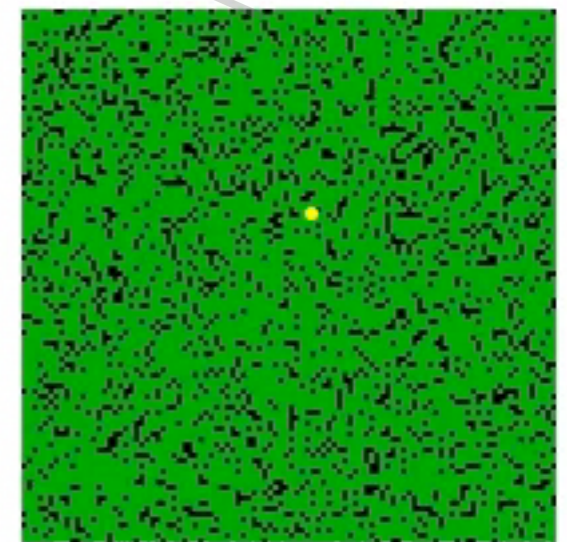
40%



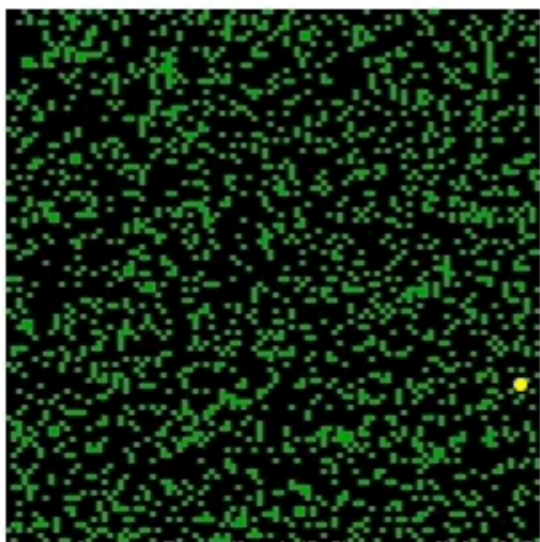
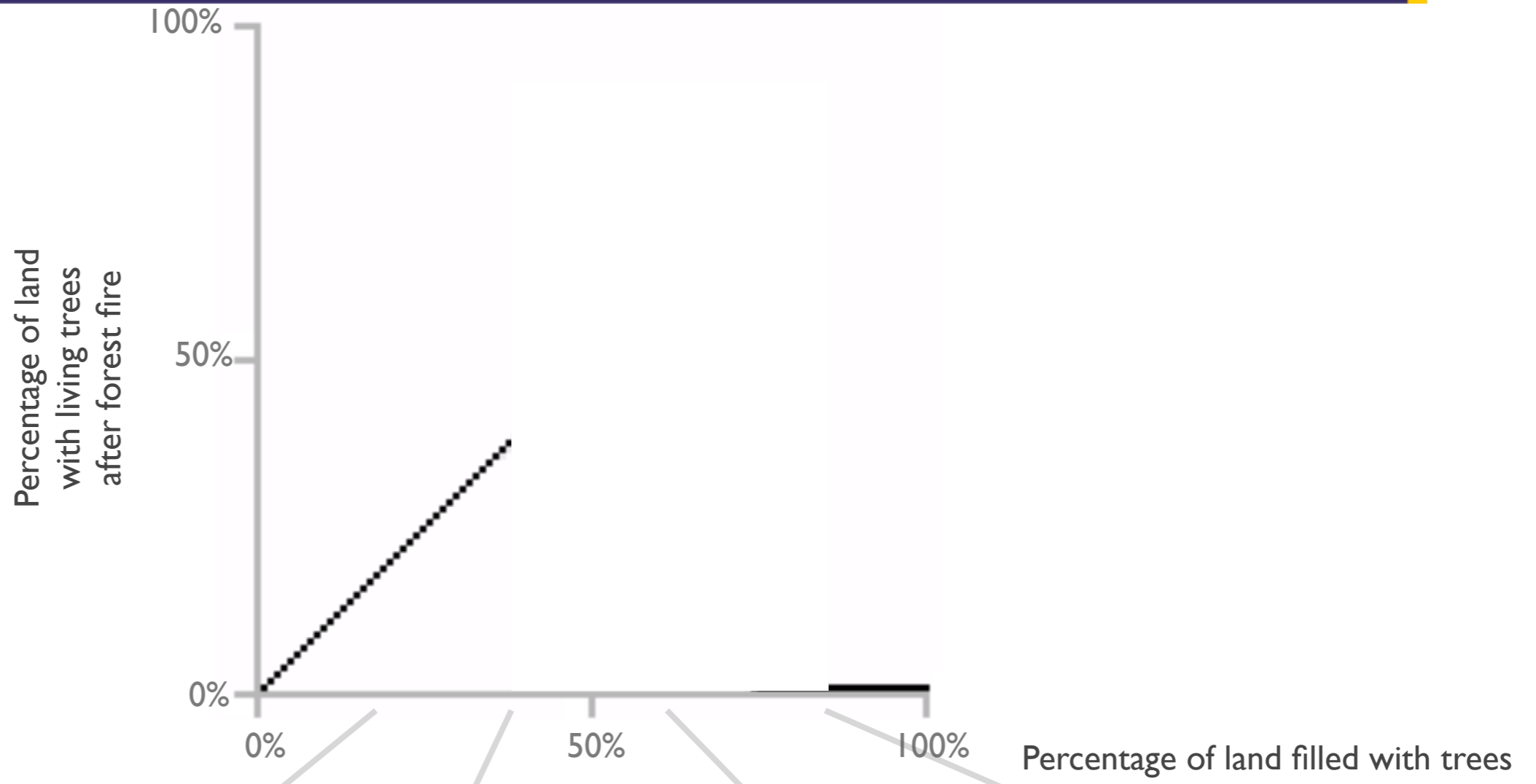
20%



40%



80%



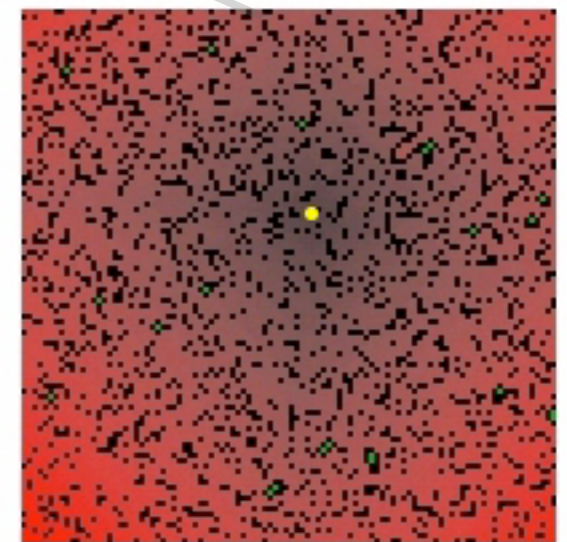
20%



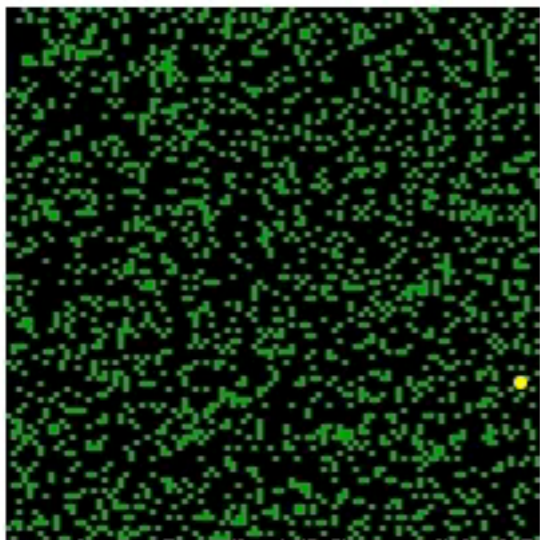
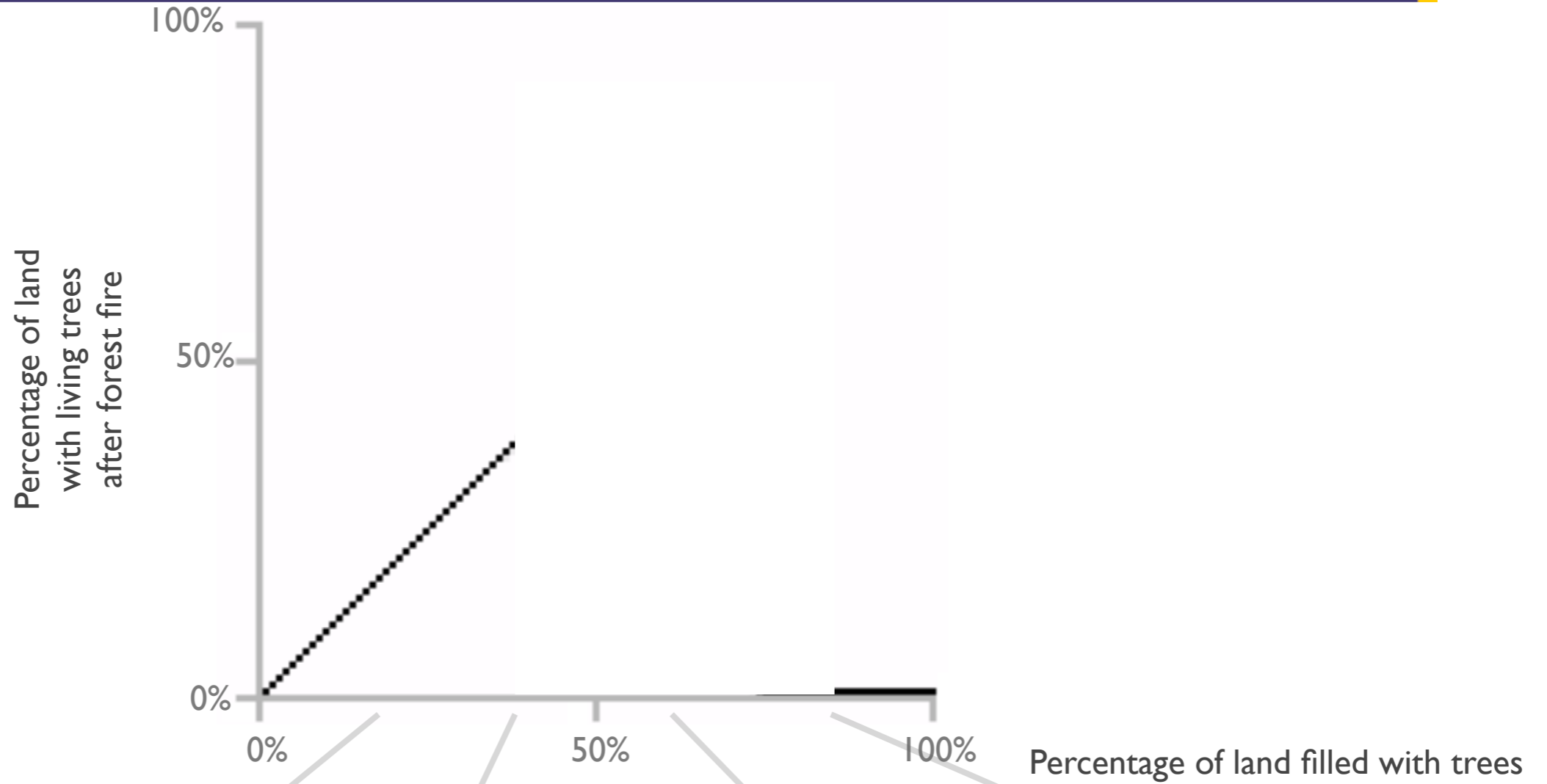
40%



60%



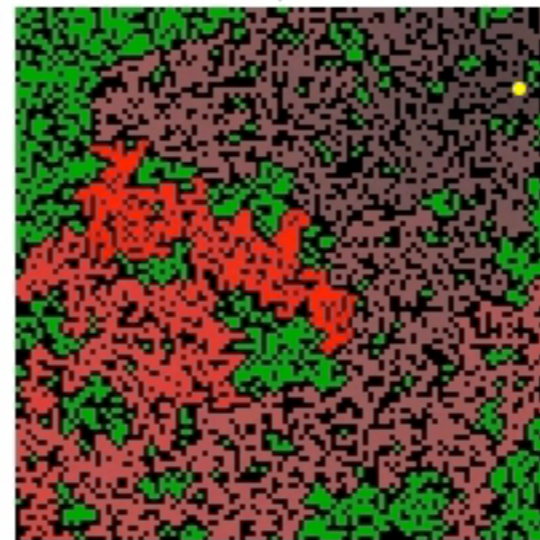
80%



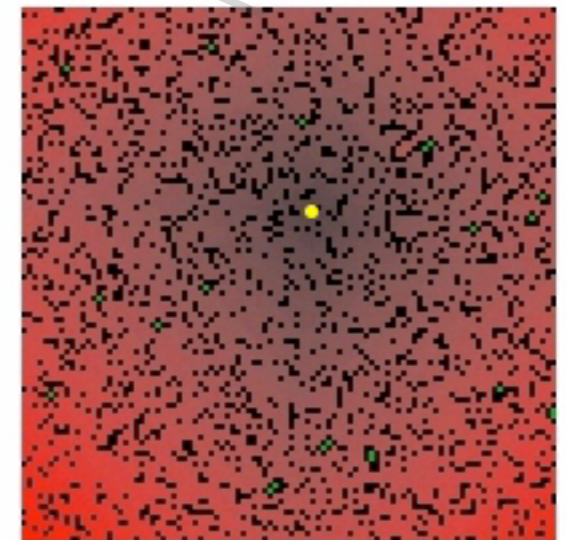
20%



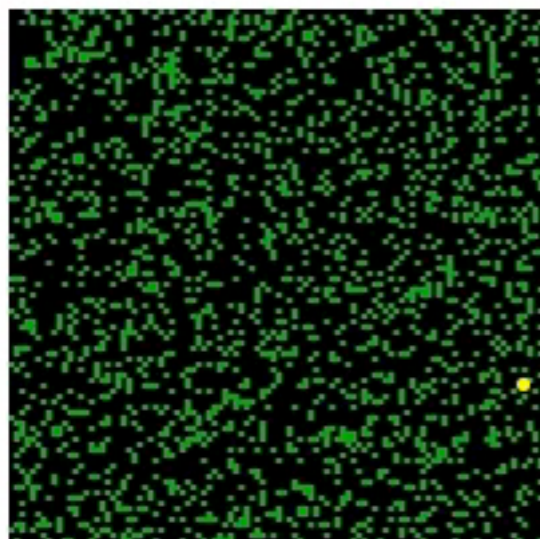
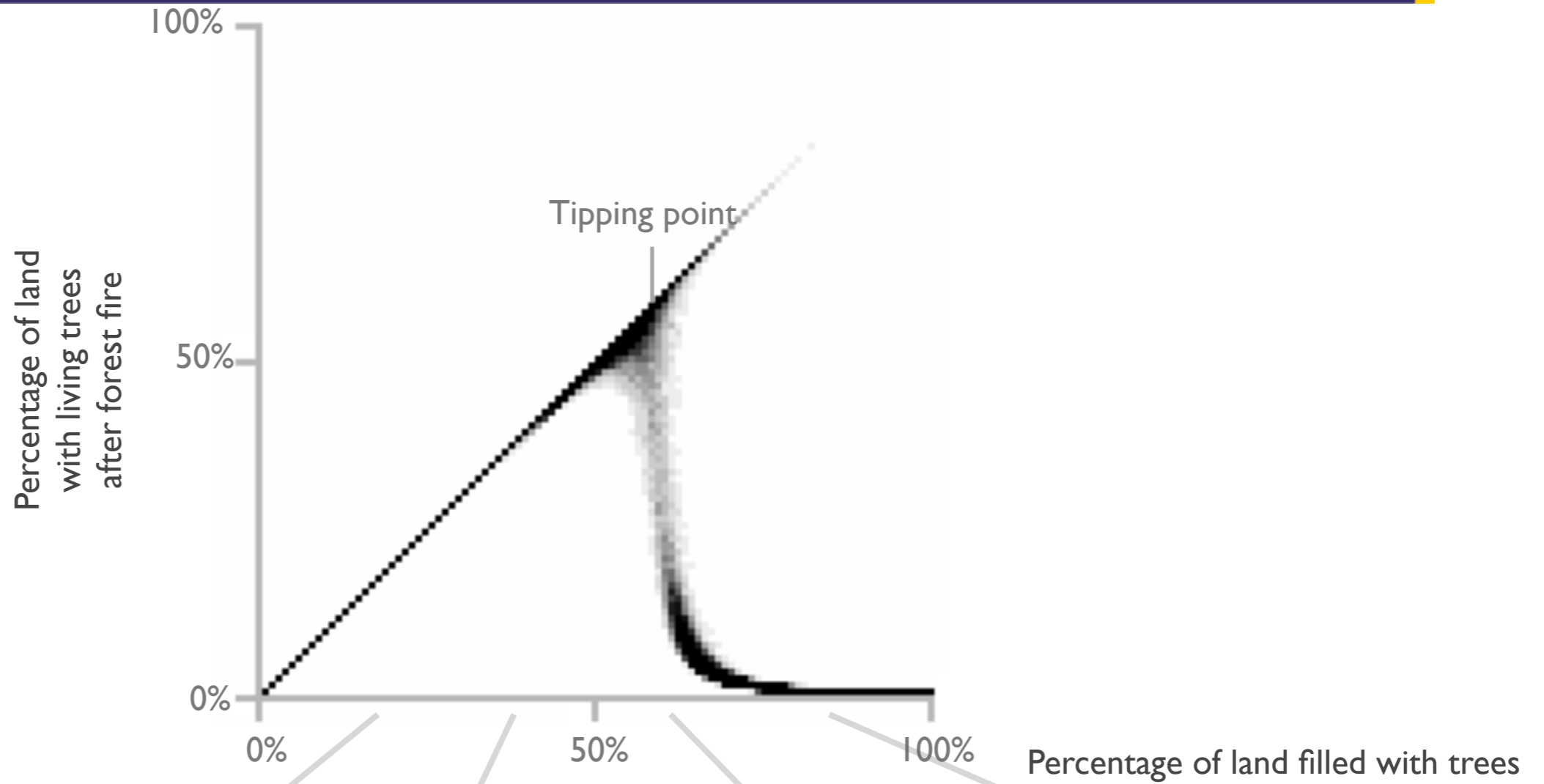
40%



60%



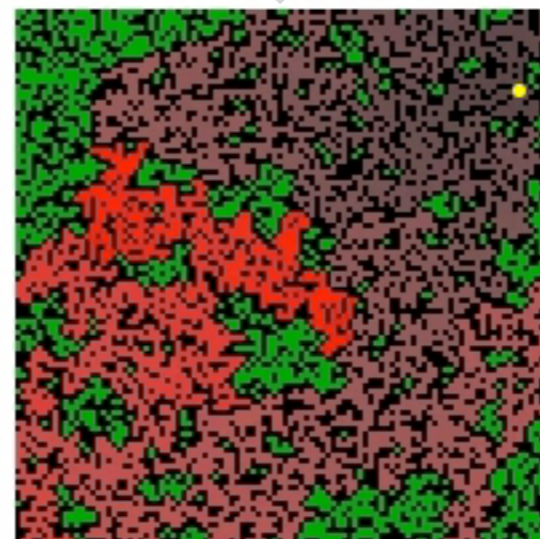
80%



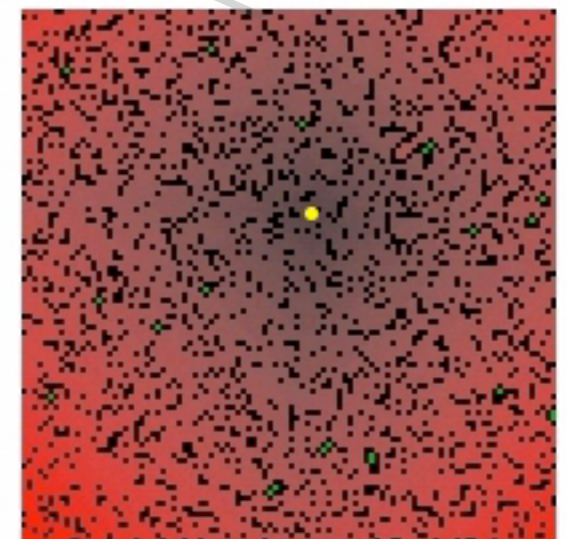
20%



40%



60%

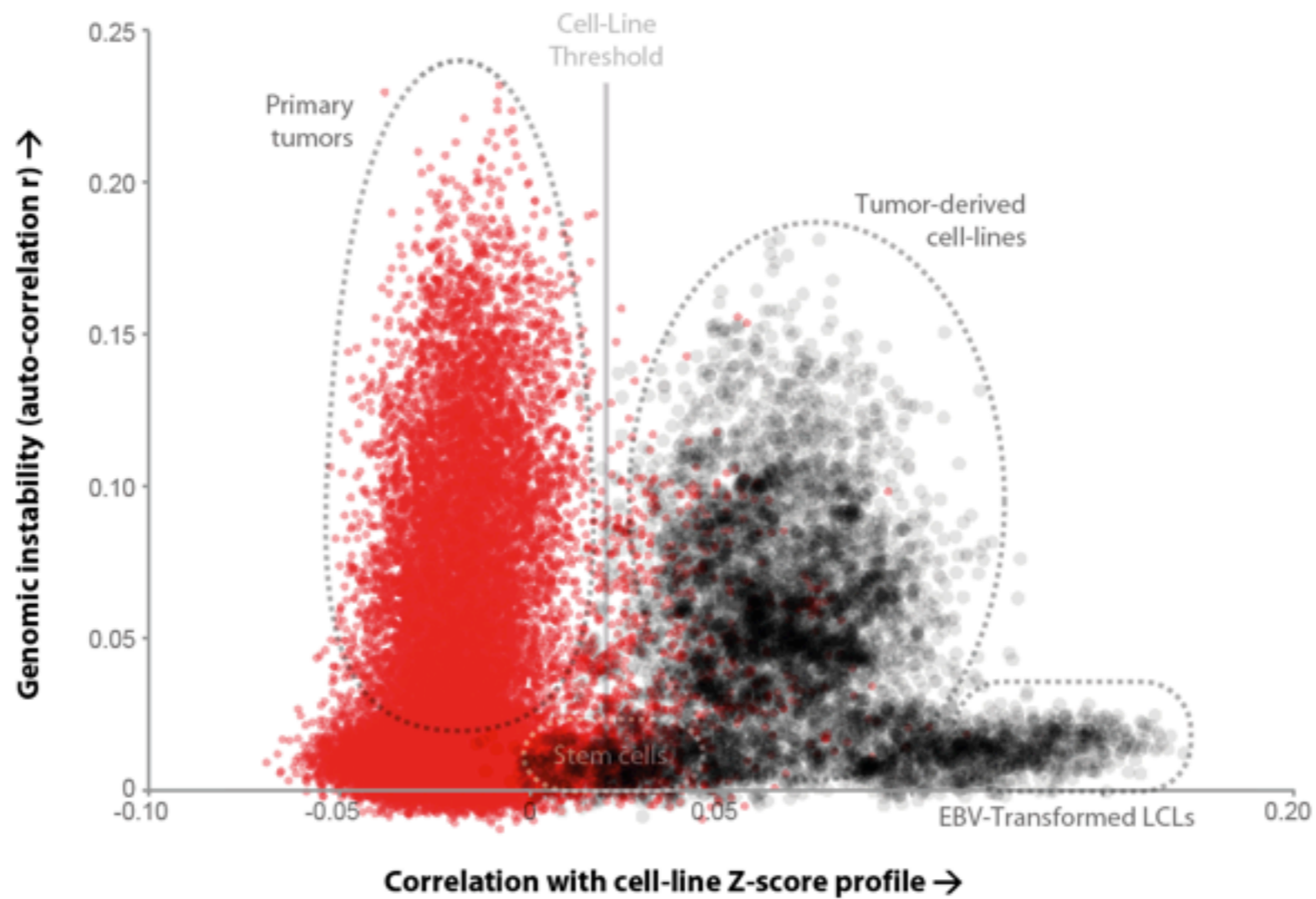
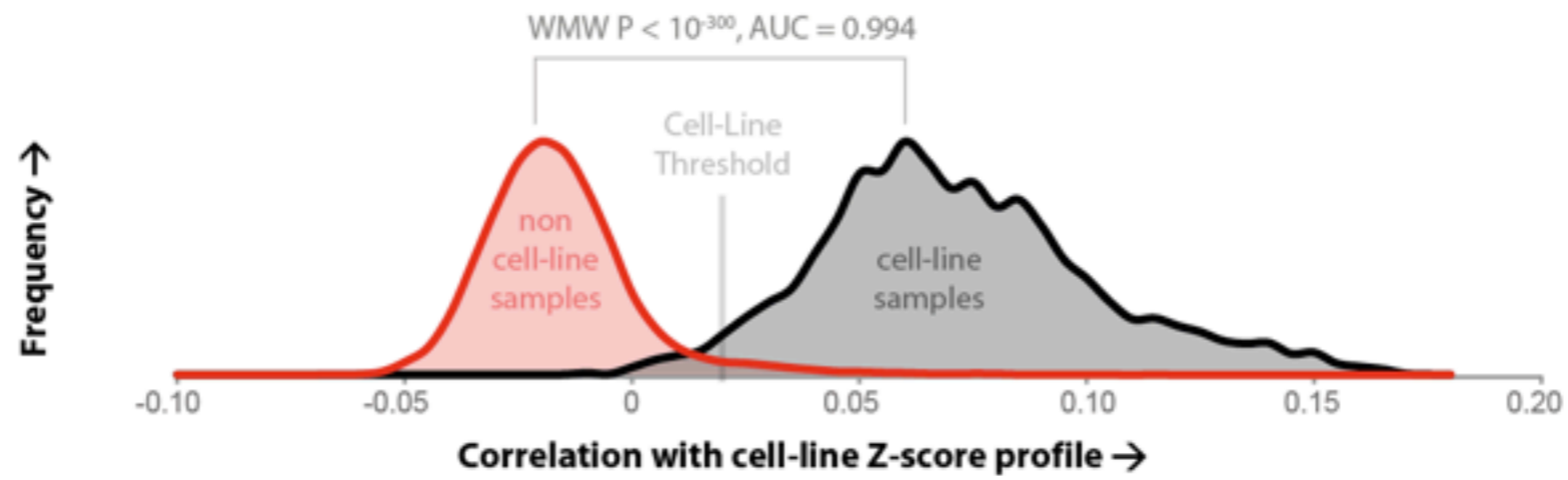


80%

Find out which are cell line samples

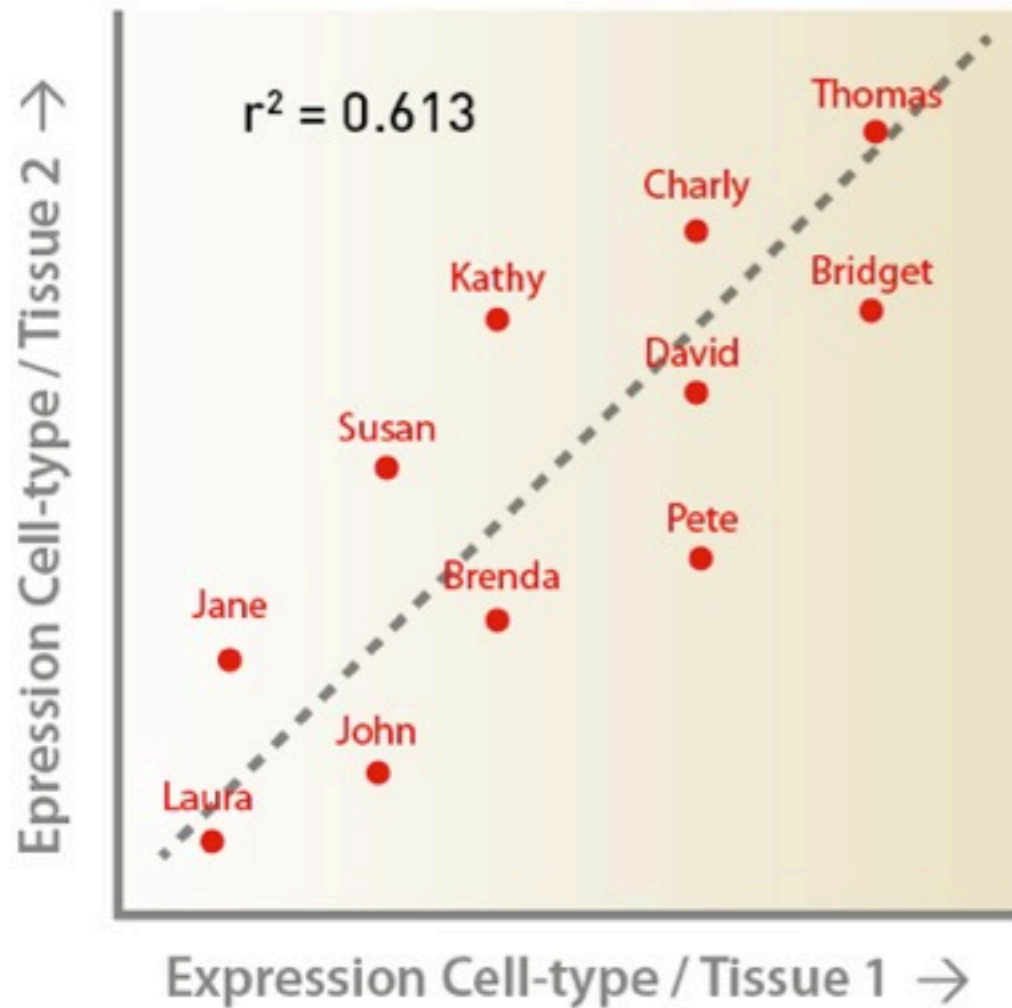
- Text mining analysis on Affymetrix U133 Plus 2.0 platform (54,000 probesets): 7,319 cell line samples
- Find additional samples that have been missed by text mining but which look similar to cell lines. Can we do this?
- Strategy: Perform for every component a T-Test, and denote T value per component. We have 777 components, and thus have a vector of 777 T values. Subsequently we correlate all 37,427 samples with this profile.

Predict cell line samples

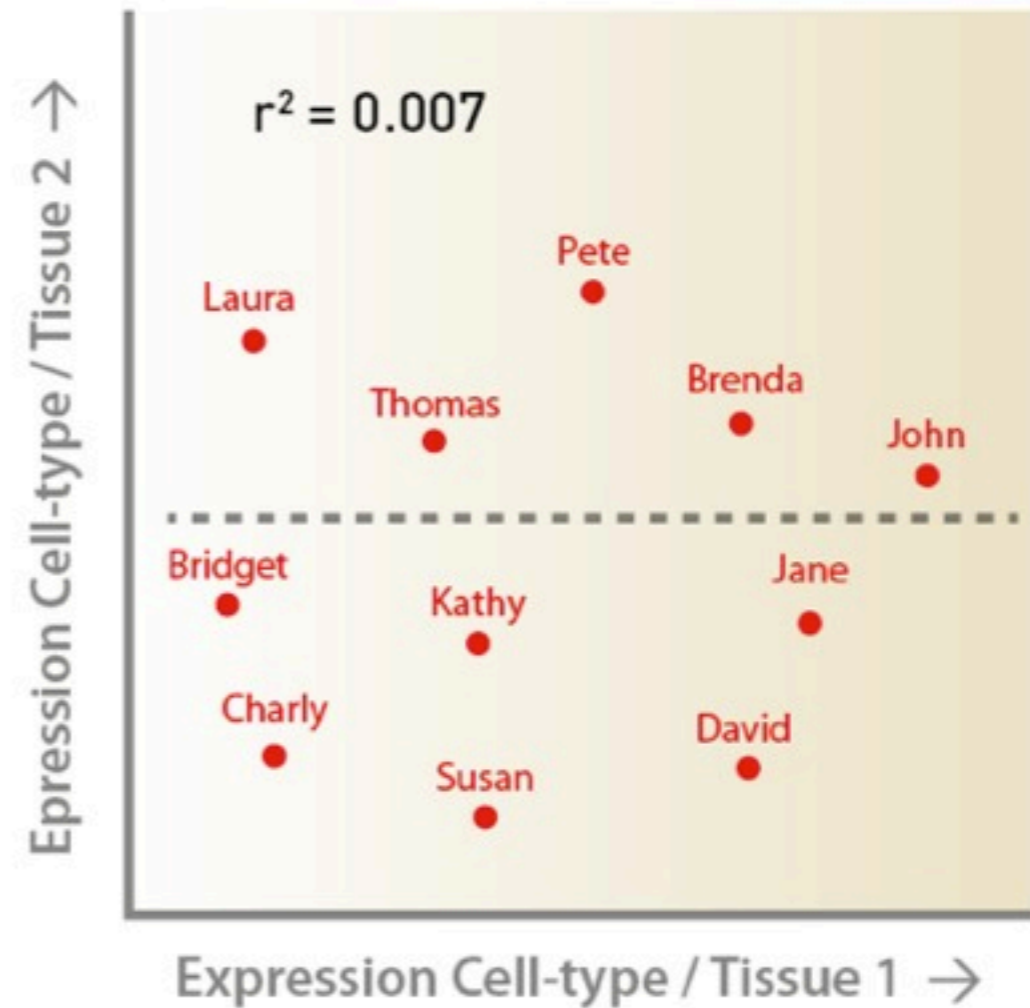


Identify related individuals

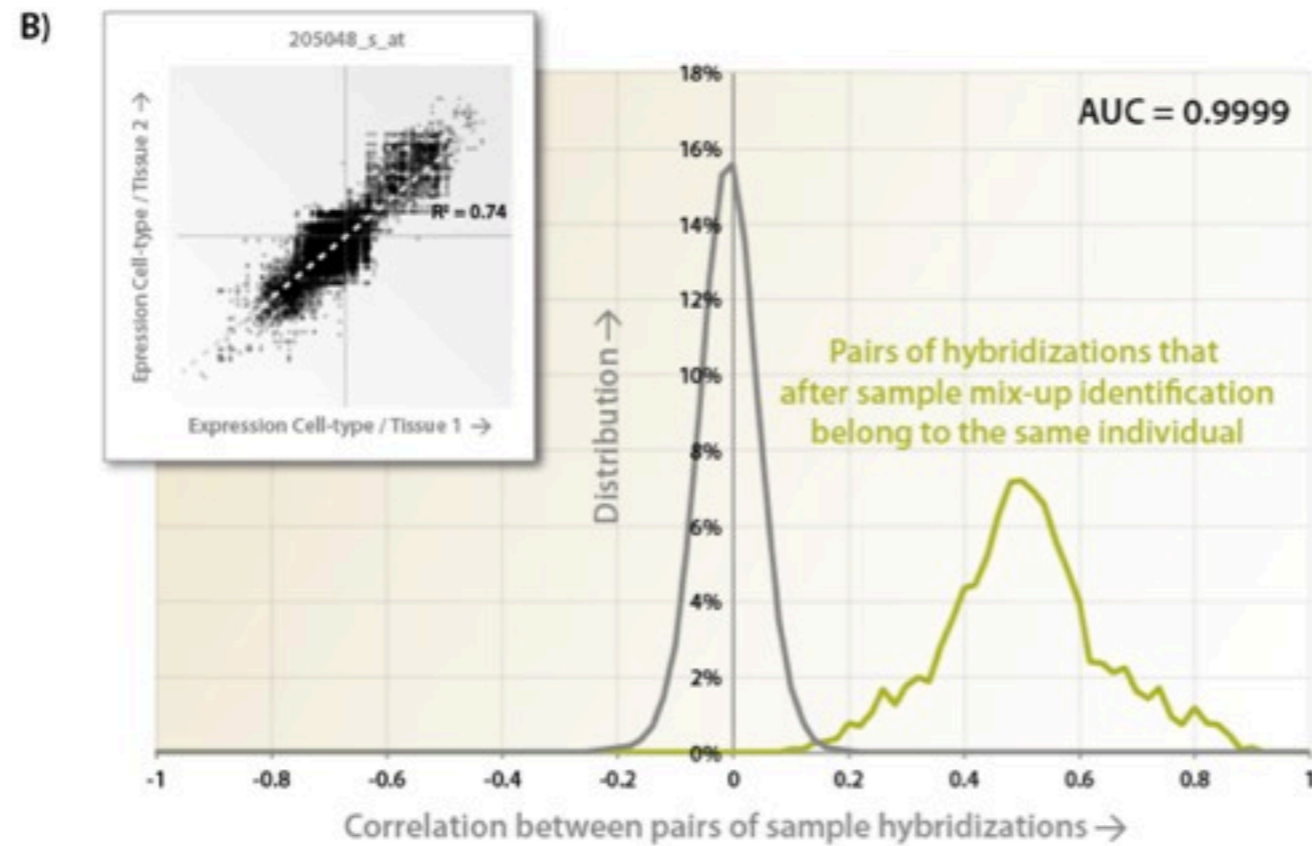
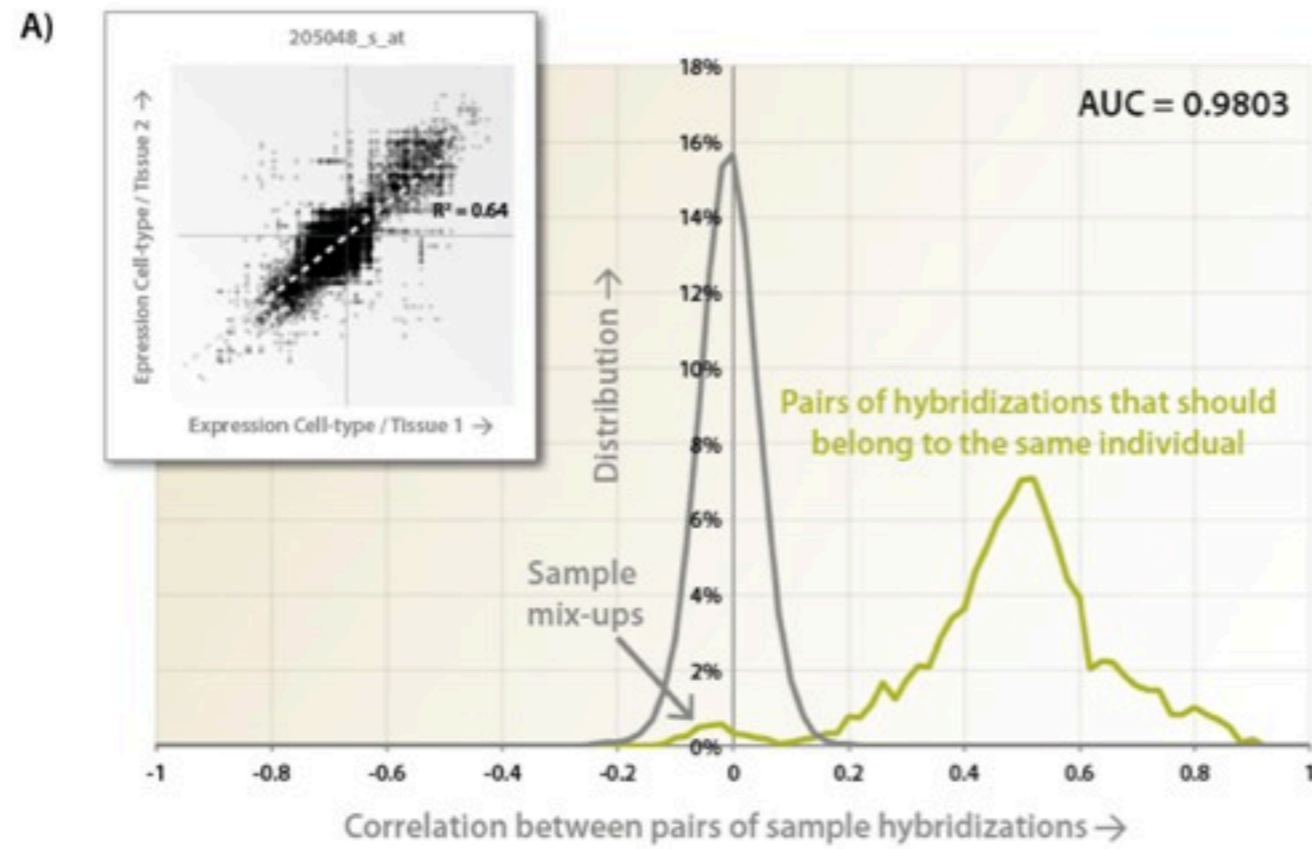
Probeset, informative for relating genetically identical samples:



Probeset, **not** informative for relating genetically identical samples:

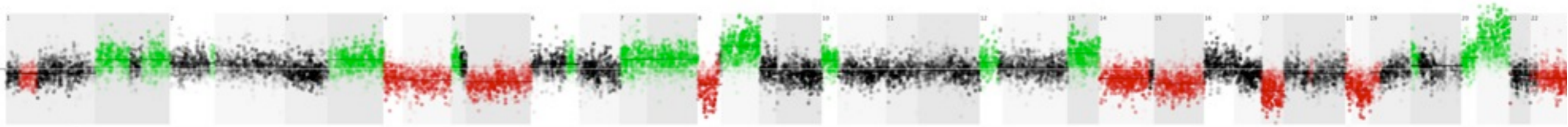


Identify related individuals

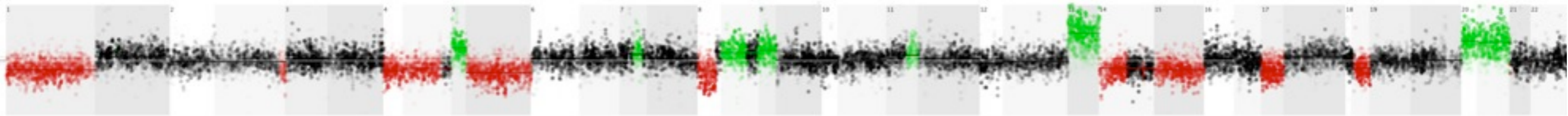


Combinations of dels/dups in many tumors

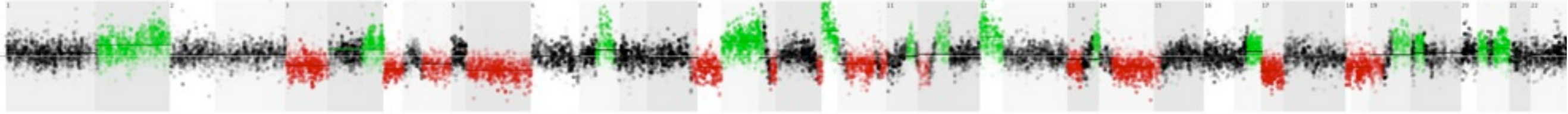
Profile most often identified in 16,172 cancer samples:



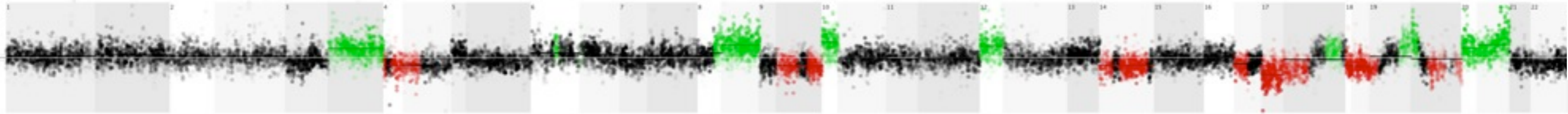
Colorectal tumor (GSM358505):



Breast tumor (GSM411312):

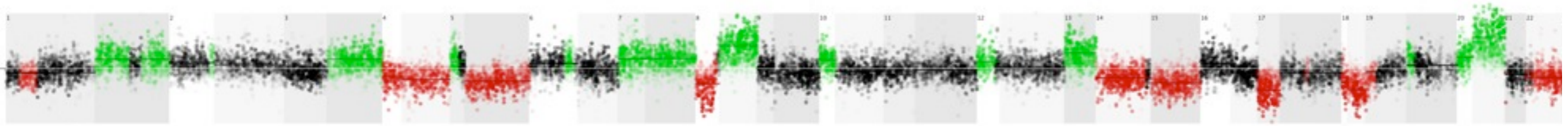


Ovarian tumor (GSM249825):

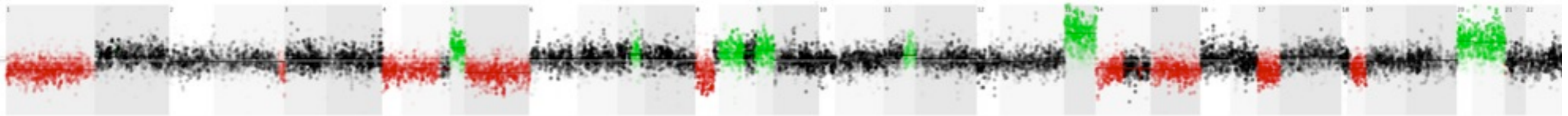


Combinations of dels/dups in many tumors

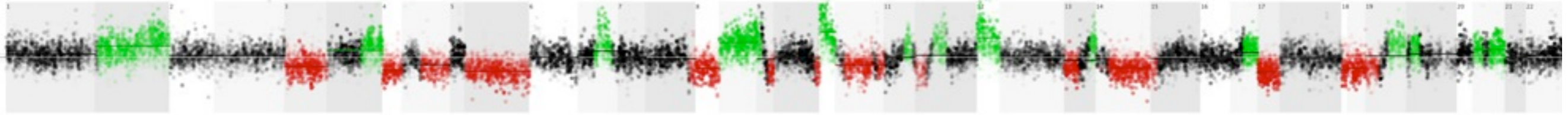
Profile most often identified in 16,172 cancer samples:



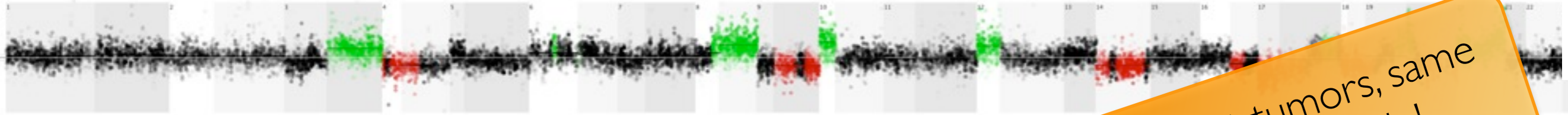
Colorectal tumor (GSM358505):



Breast tumor (GSM411312):



Ovarian tumor (GSM249825):



Different tumors, same cytogenetic profile!

Trans-eQTL mapping in 16,172 samples

Cytogenetic signature:

1 2 3 4



Enriched Pathway



Proliferative signaling



Resisting cell death



Genome maintenance



invasion & metastasis



Angiogenesis



Inflammation



Growth suppressors



Immortality



Immune defense

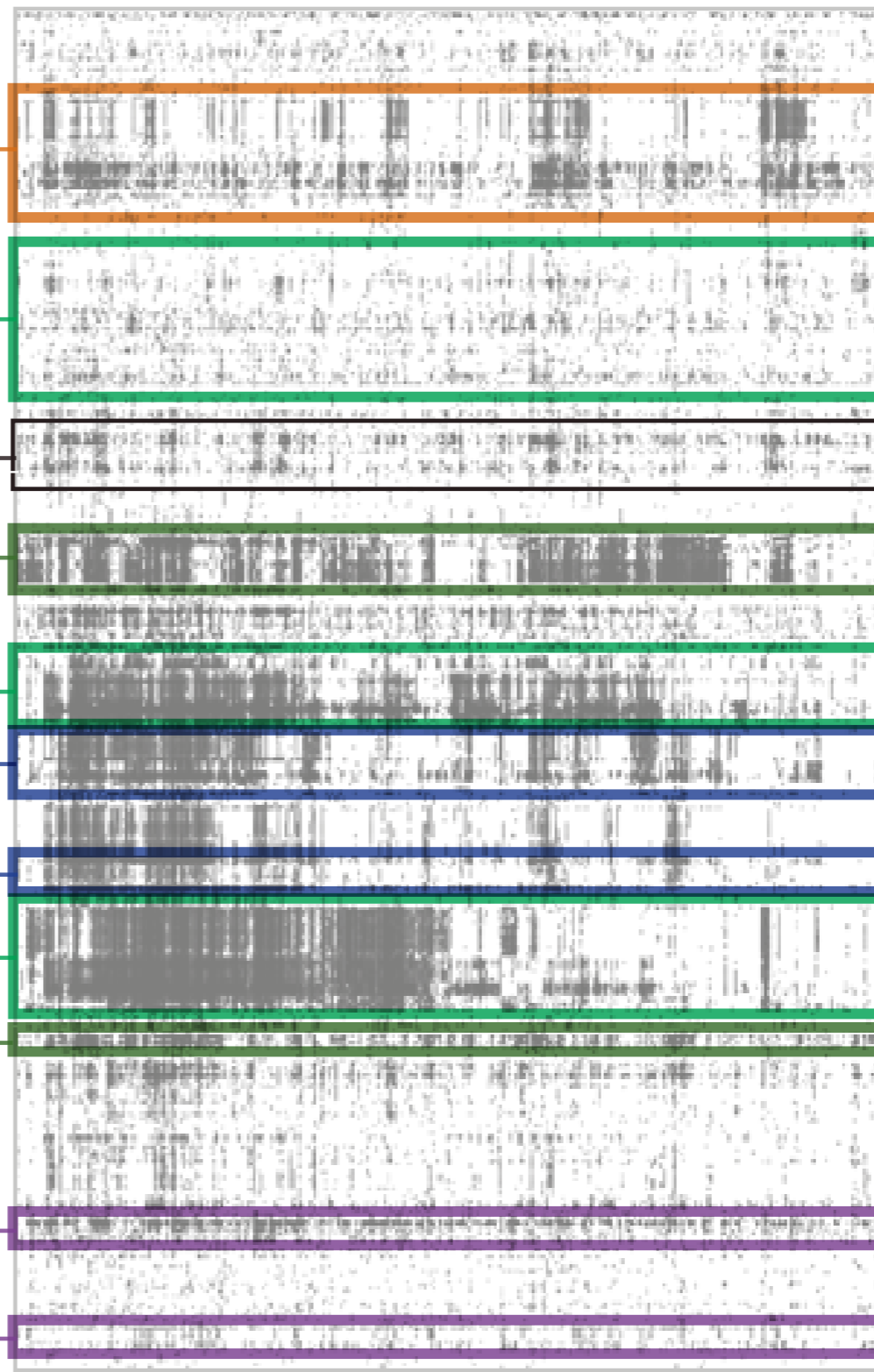


RNA processing/translation



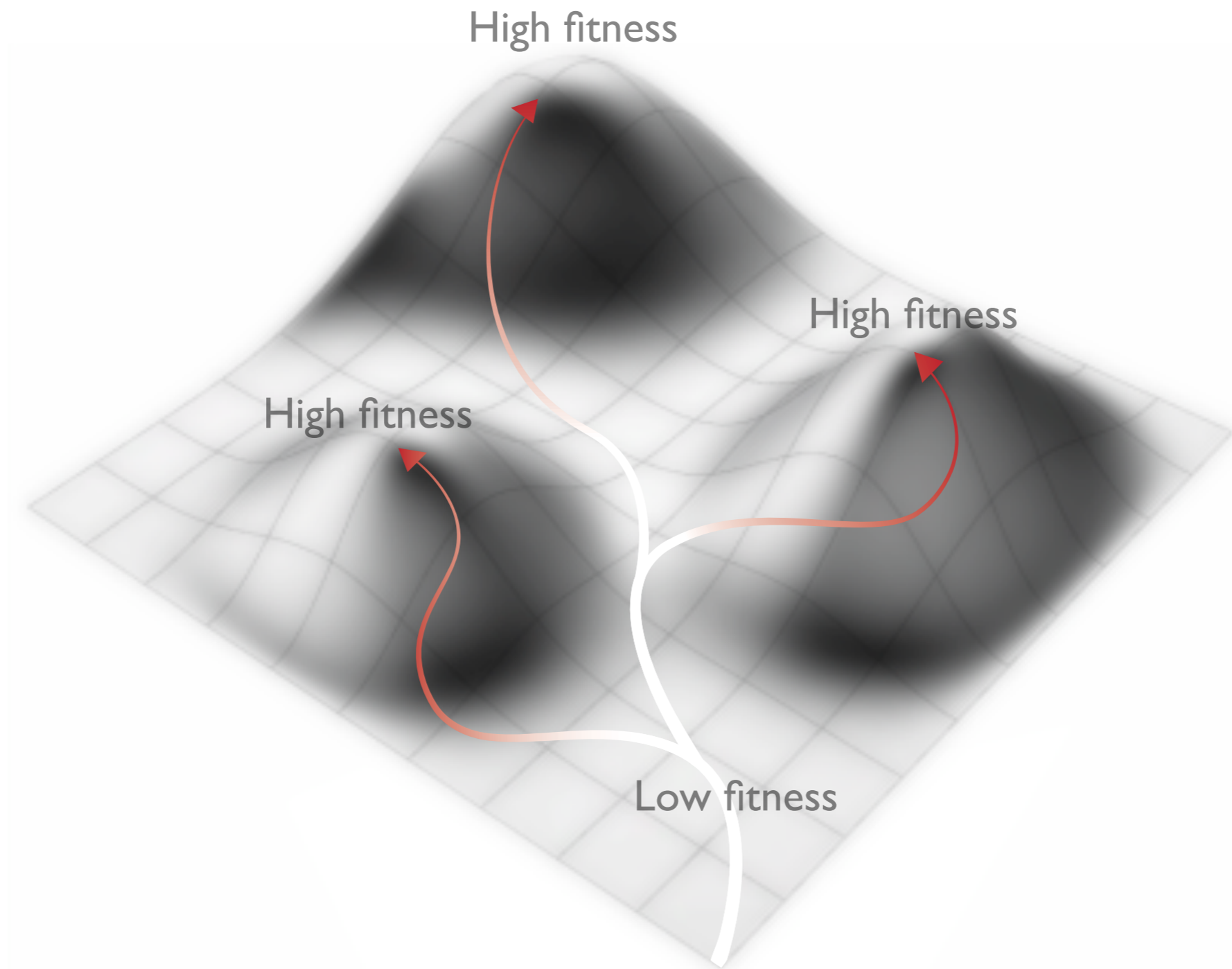
Metabolism

All 500 cytogenetic signatures



Reactome pathways

Strong dependencies exist within cancers



Cancer development

- Reanalysis of publicly available data can reveal new insight into transcriptional regulation and genomic instability in cancer
- Many avenues for bioinformaticians and statistical geneticists!