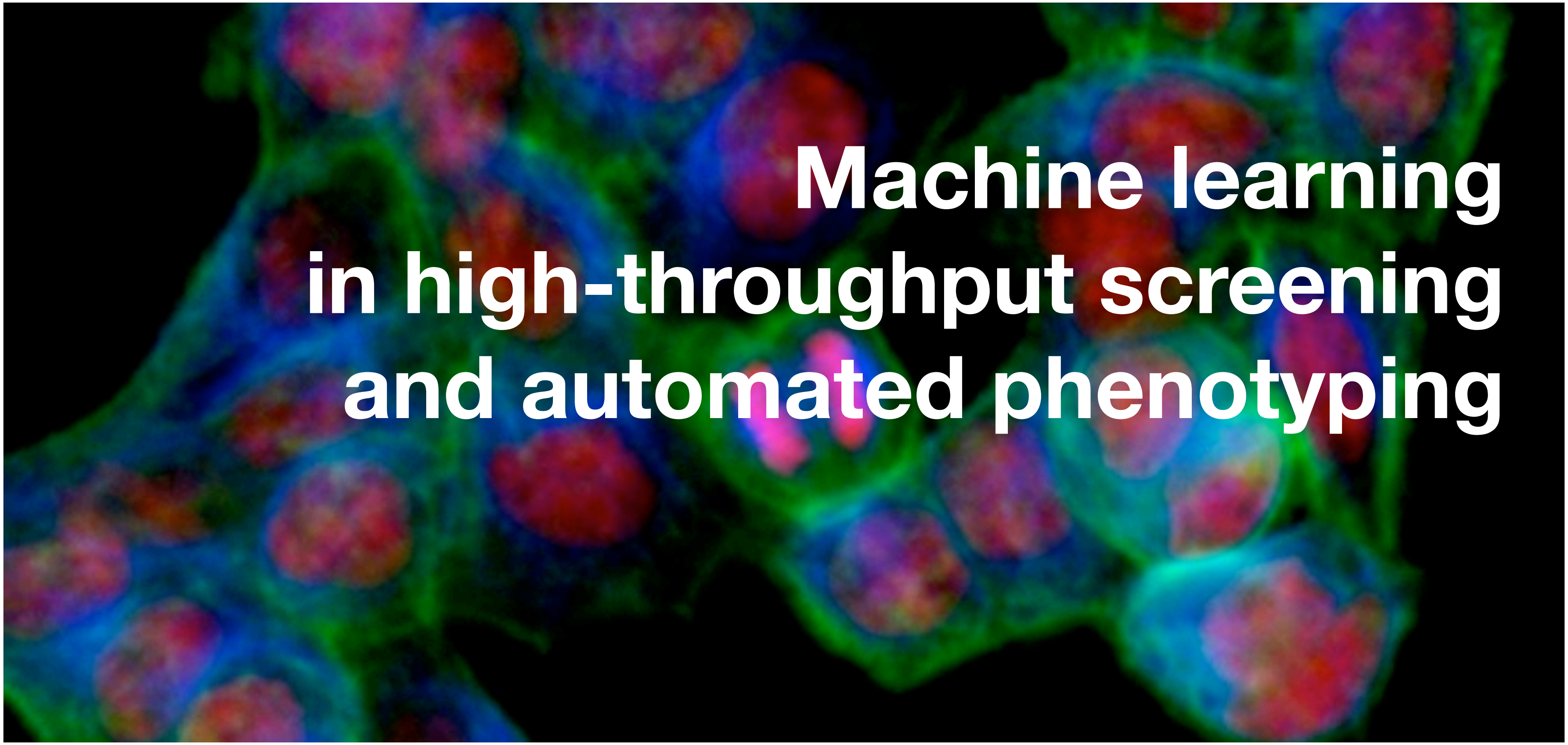


Machine learning in high-throughput screening and automated phenotyping

Wolfgang Huber

EMBL

Progress in science is driven by technology

Progress in science is driven by technology

Sequencing - DNA-Seq, RNA-Seq, ChiP-Seq, HiC

Microscopy & remote sensing- molecular interactions and life-cycles in single, live cells

Large scale perturbation libraries - RNAi, drugs

Progress in science is driven by technology

Sequencing - DNA-Seq, RNA-Seq, ChiP-Seq, HiC

Microscopy & remote sensing- molecular interactions and life-cycles in single, live cells

Large scale perturbation libraries - RNAi, drugs

We work on the methods in statistical computing, integrative bioinformatics and mathematical modelling to turn these data into biology.

Research areas

Gene expression

- Statistics - differential expression; alternative exon usage
- 3D structure of DNA (HiC & Co.)
- Single-cell transcriptomics and noise

Simon Anders, Aleksandra Pekoswka, Alejandro Reyes, Jan Swedlow; Tibor Pakozdi

collaborations with L. Steinmetz, P. Bertone, E. Furlong, T. Hiiragi

Cancer Genomics & Precision Oncology

- Somatic mutation detection (incl subclonal)
- Phylogeny inference

Julian Gehring, Paul Pyl

collaborations with C.v.Kalle/M.Schmid, H. Glimm (NCT); J. Korbel

Genetic Interactions, pharmacogenetics (reverse genetics)

- Large-scale combinatorial RNAi & automated microscopy phenotyping
- Cancer mutations & drugs

Joseph Barry, Bernd Fischer, Felix Klein, Malgorzata Oles

collaborations with M.Boutros (DKFZ), T.Zenz (NCT), M. Knop (Uni)

Basics of statistics

- Tools & infrastructure for software ‘publication’
- Teaching

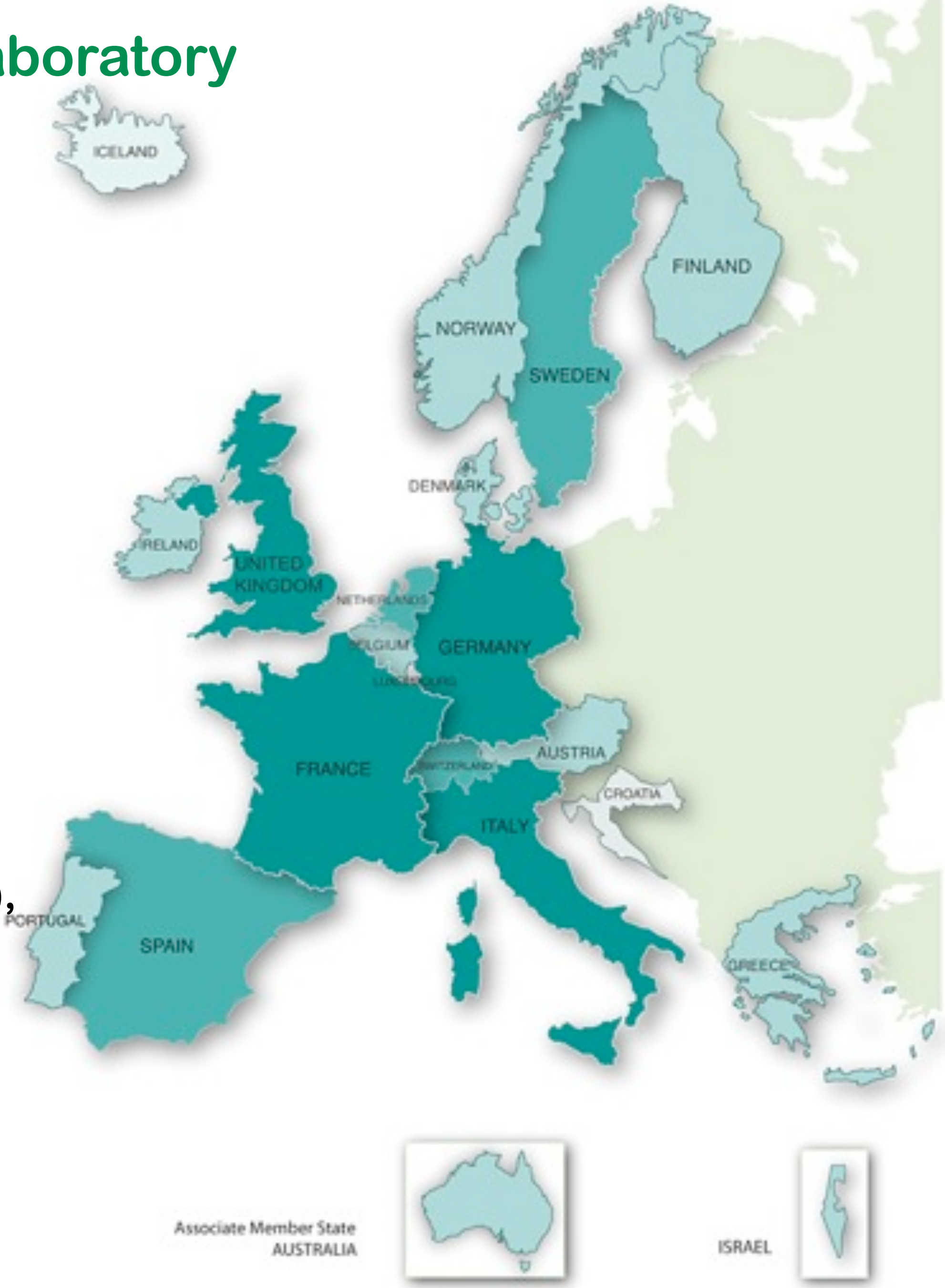
Bernd Klaus, Andrzej Oles

collaborations M.Morgan (FHCRC), R.Gentleman (Genentech)

European Molecular Biology Laboratory (EMBL)

European Intergovernmental Research Organisation

- 20 Member States
- Founded in 1974
- Sites in Heidelberg (D), Cambridge (GB), Roma (I), Grenoble (F), Hamburg (D)
- ca. 1400 staff (⊃1100 scientists) representing more than 60 nationalities



EMBL’s five missions

- Basic research
- Development of new technologies and instruments
- Technology transfer
- Services to the member states
- Advanced training

What can you do at EMBL?

Biology

Chemistry

Physics

Mathematics

Informatics

Engineering

www.embl.org/phdprogramme
www.embl.org/postdocs
www.embl.org/jobs

How do we know which genes do what?

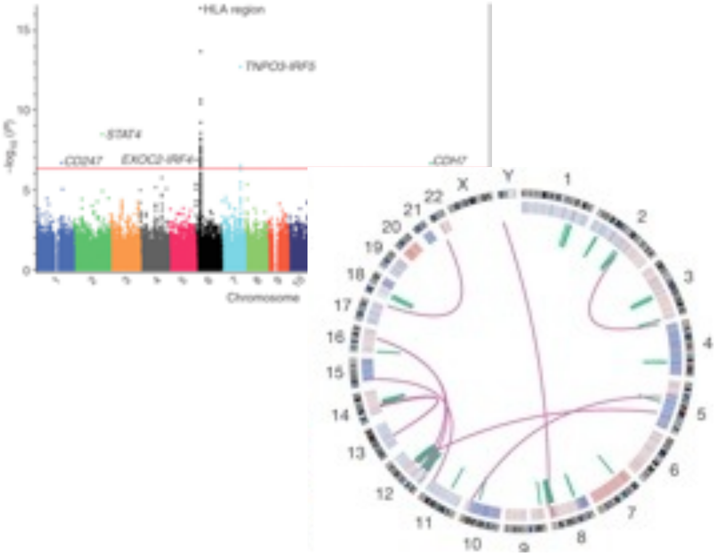
Forward genetics

from phenotypes to genes

→ genome-wide association studies

→ sporadic/rare mutations

→ cancer genome sequencing

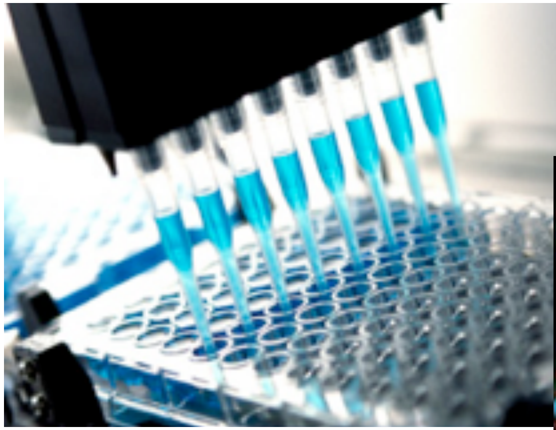


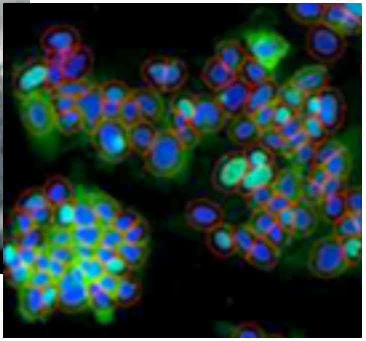
Reverse genetics

from genes to phenotypes

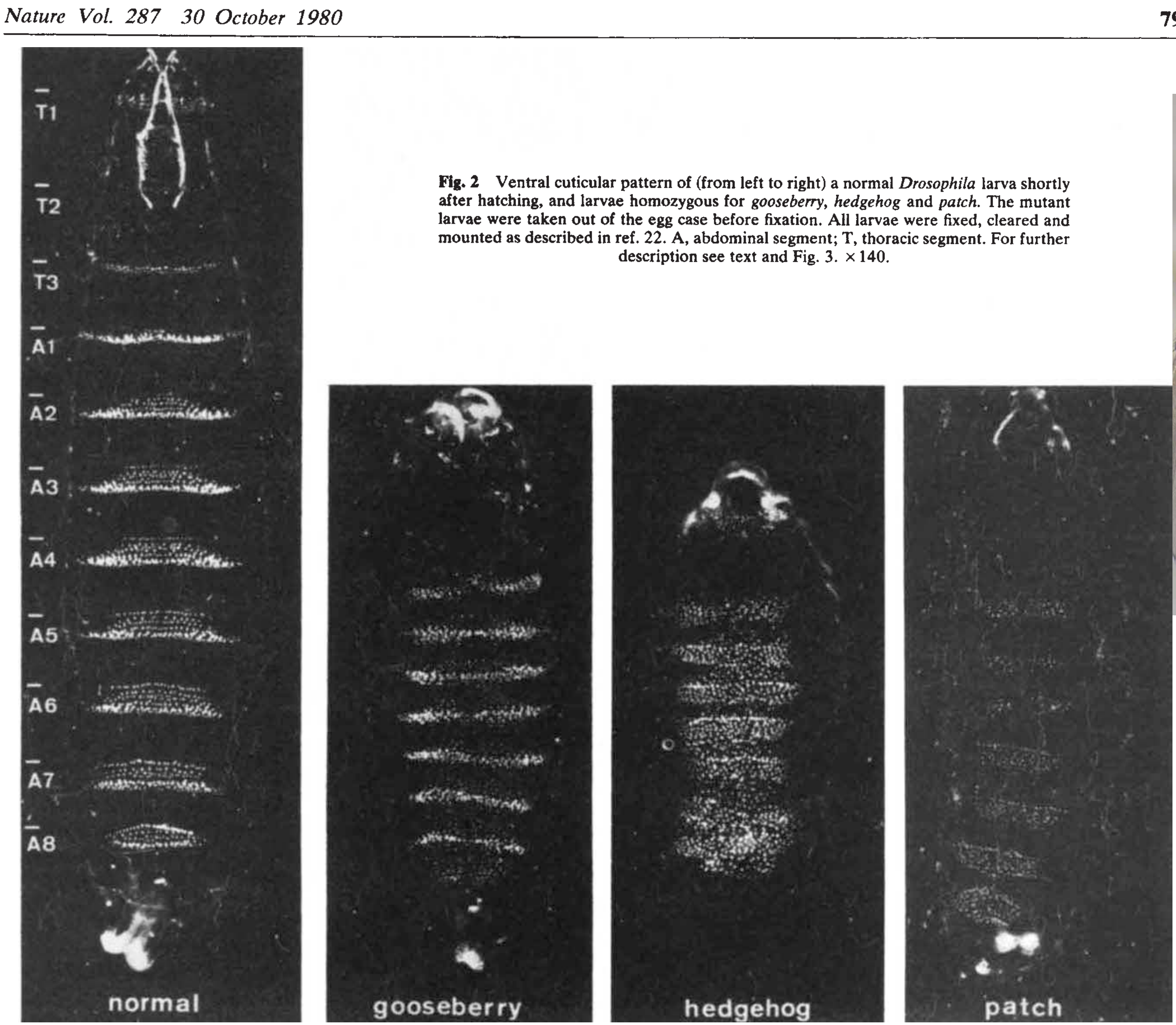
→ deletion libraries

→ high-throughput RNAi



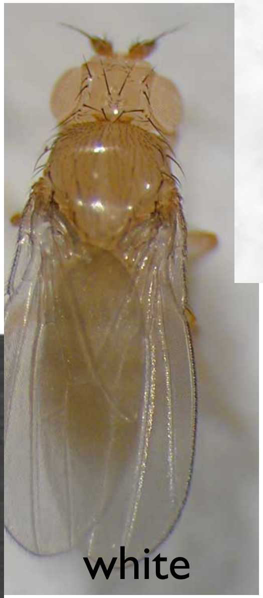


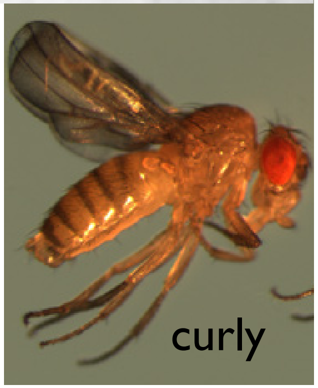
Forward genetics

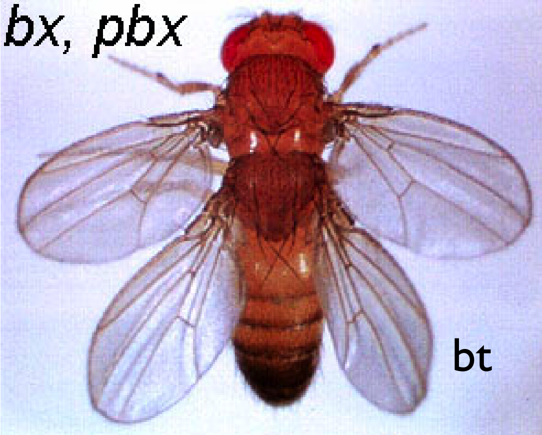


Nature Vol. 287 30 October 1980 797

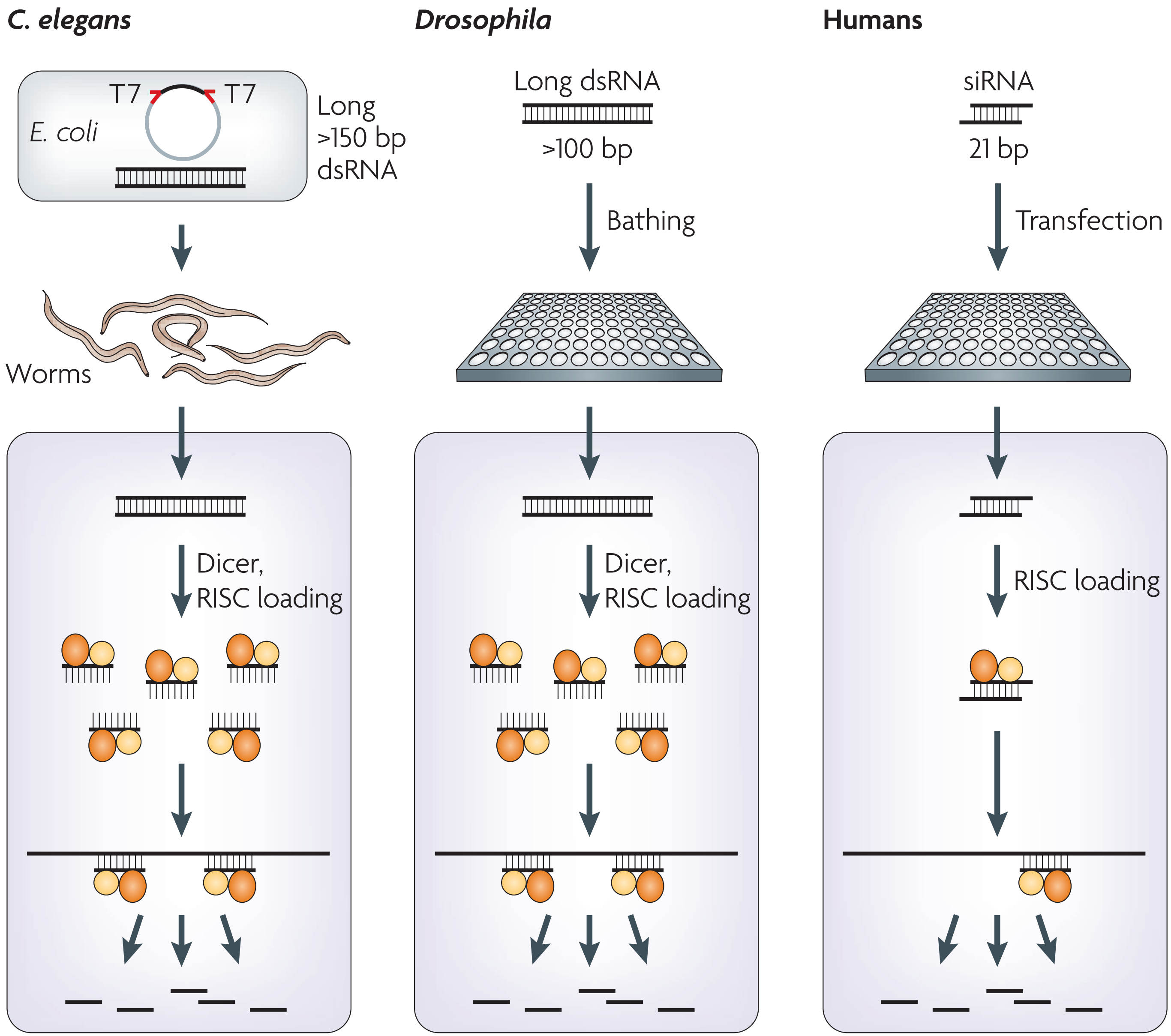
Fig. 2 Ventral cuticular pattern of (from left to right) a normal *Drosophila* larva shortly after hatching, and larvae homozygous for *gooseberry*, *hedgehog* and *patch*. The mutant larvae were taken out of the egg case before fixation. All larvae were fixed, cleared and mounted as described in ref. 22. A, abdominal segment; T, thoracic segment. For further description see text and Fig. 3. × 140.







Reverse genetics: RNA interference



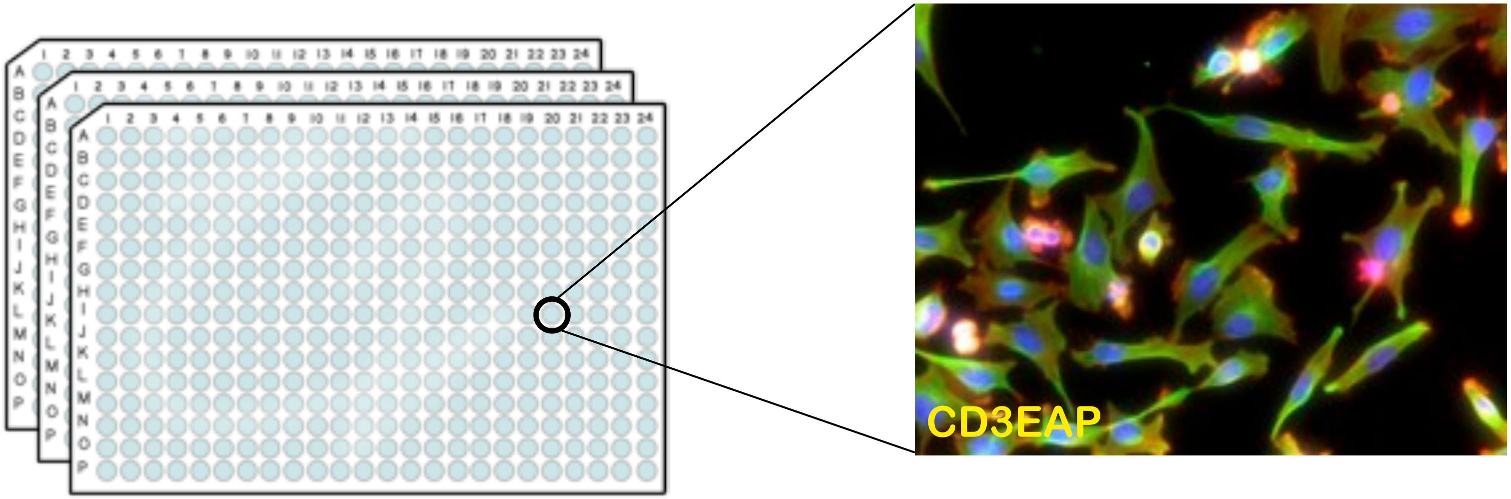
RNAi induced cell morphology phenotypes in human cells

with F. Fuchs, C. Budjan, Michael Boutros (DKFZ)

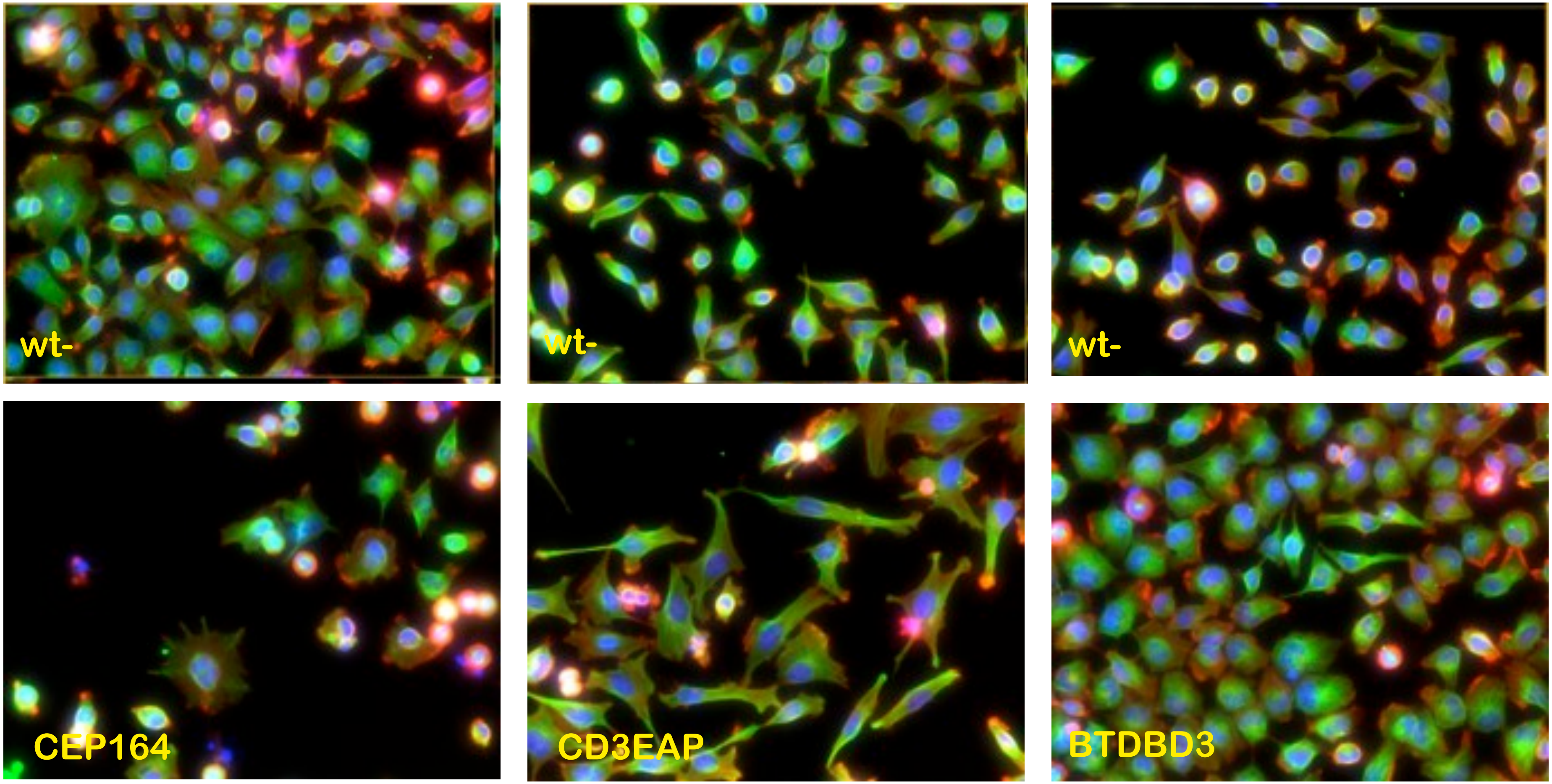
Genomewide RNAi library (Dharmacon, 22k siRNA-pools)

HeLa cells, incubated 48h, then fixed and stained

Microscopy readout: DNA (DAPI), tubulin (Alexa), actin (TRITC)

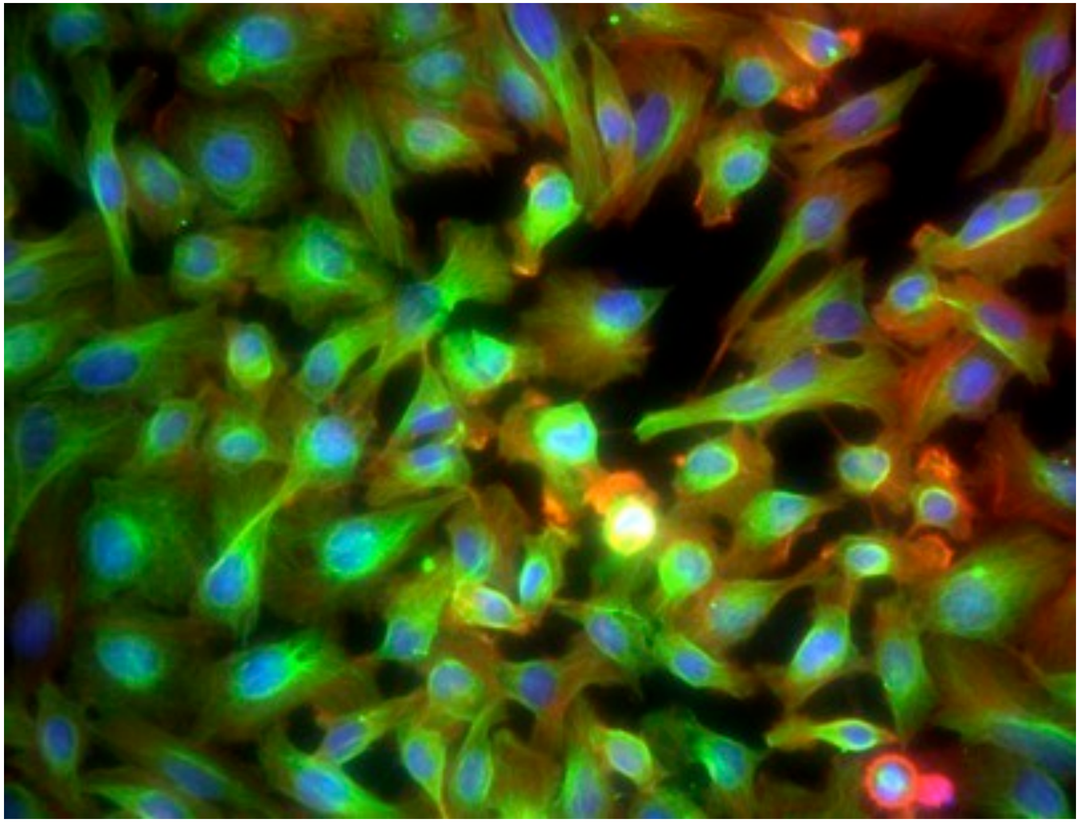


siRNA perturbation phenotypes are observed by automated microscopy



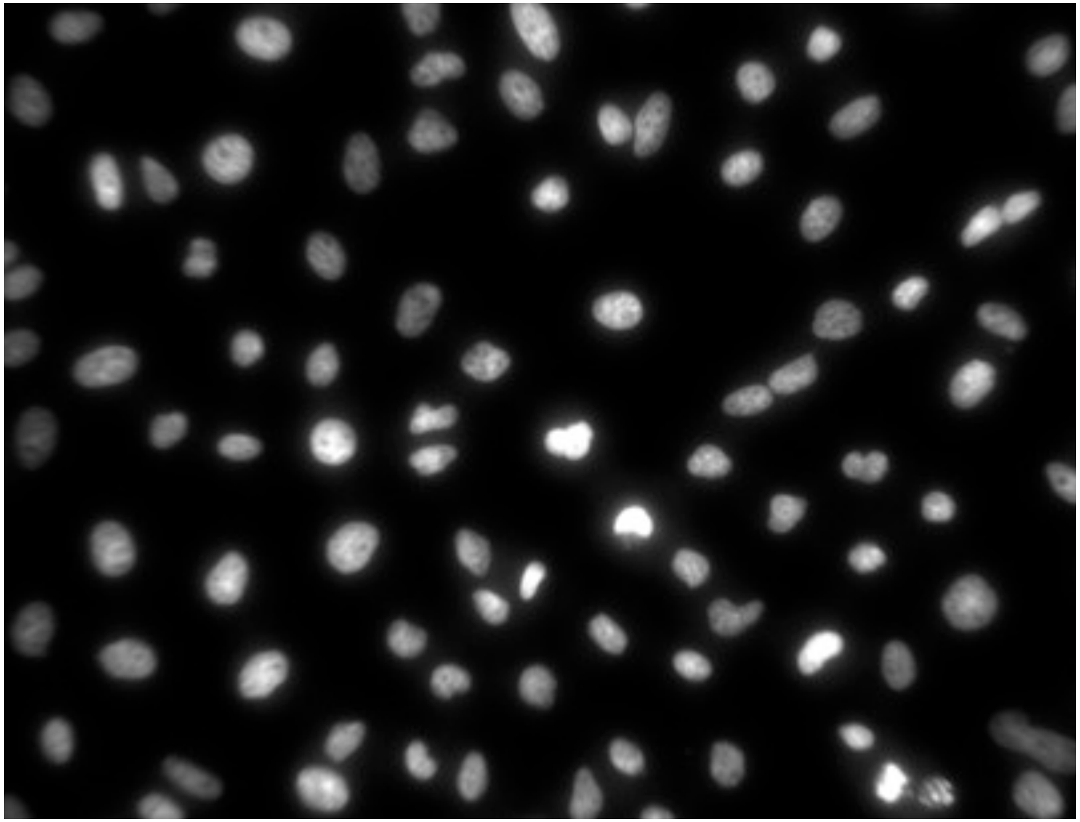
22839 wells, 4 images per well
each with DNA, tubulin, actin, 1344 x 1024 pixel at 12 bit

Cell segmentation



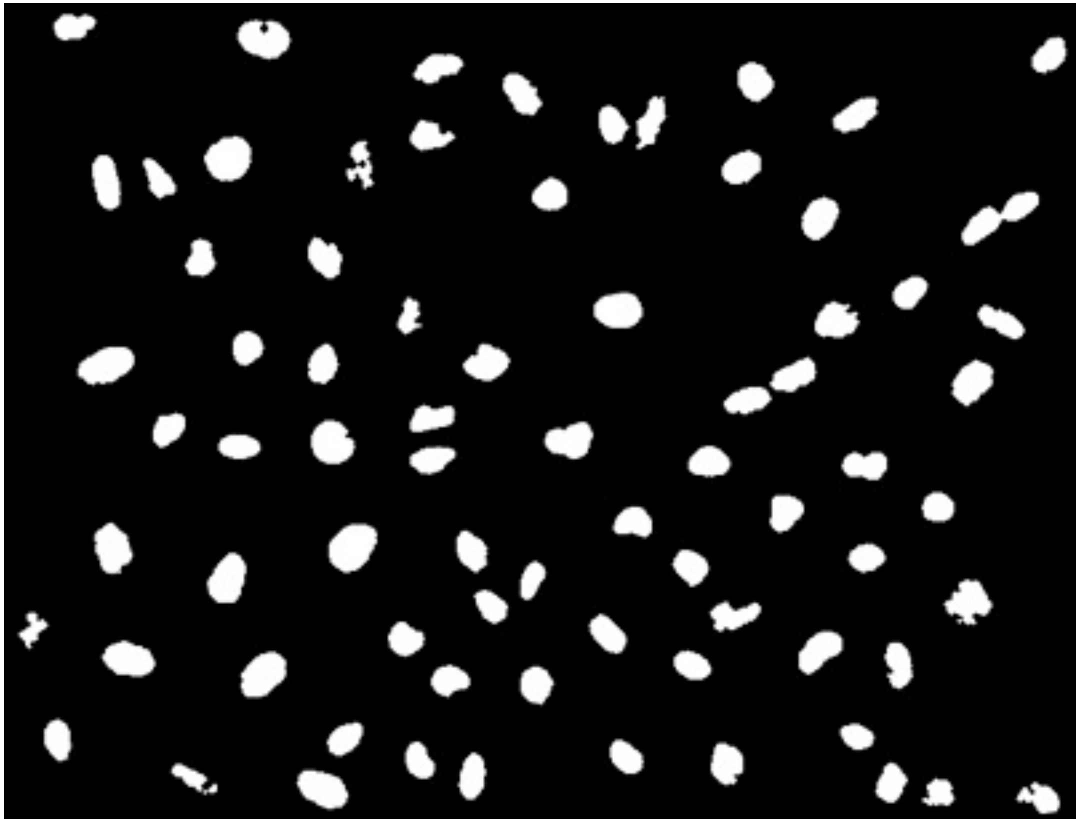
Cell segmentation

Adaptative thresholding + watershed



Cell segmentation

Adaptative thresholding + watershed

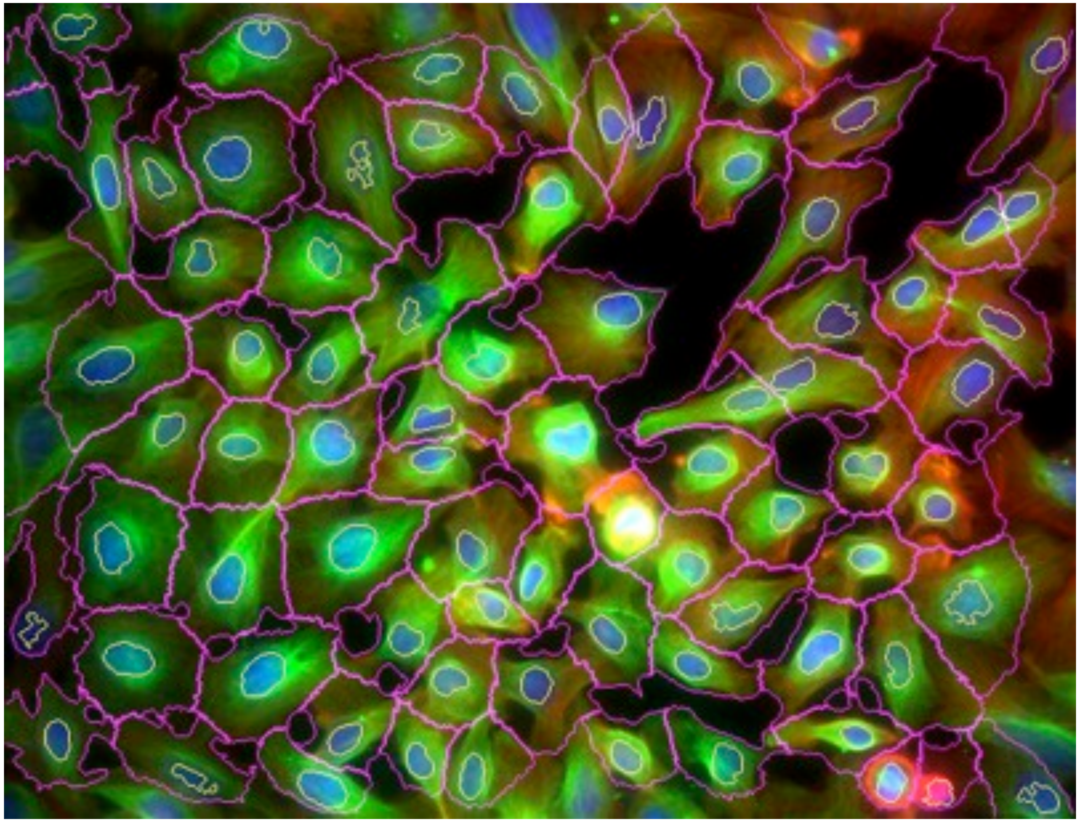


Cell segmentation

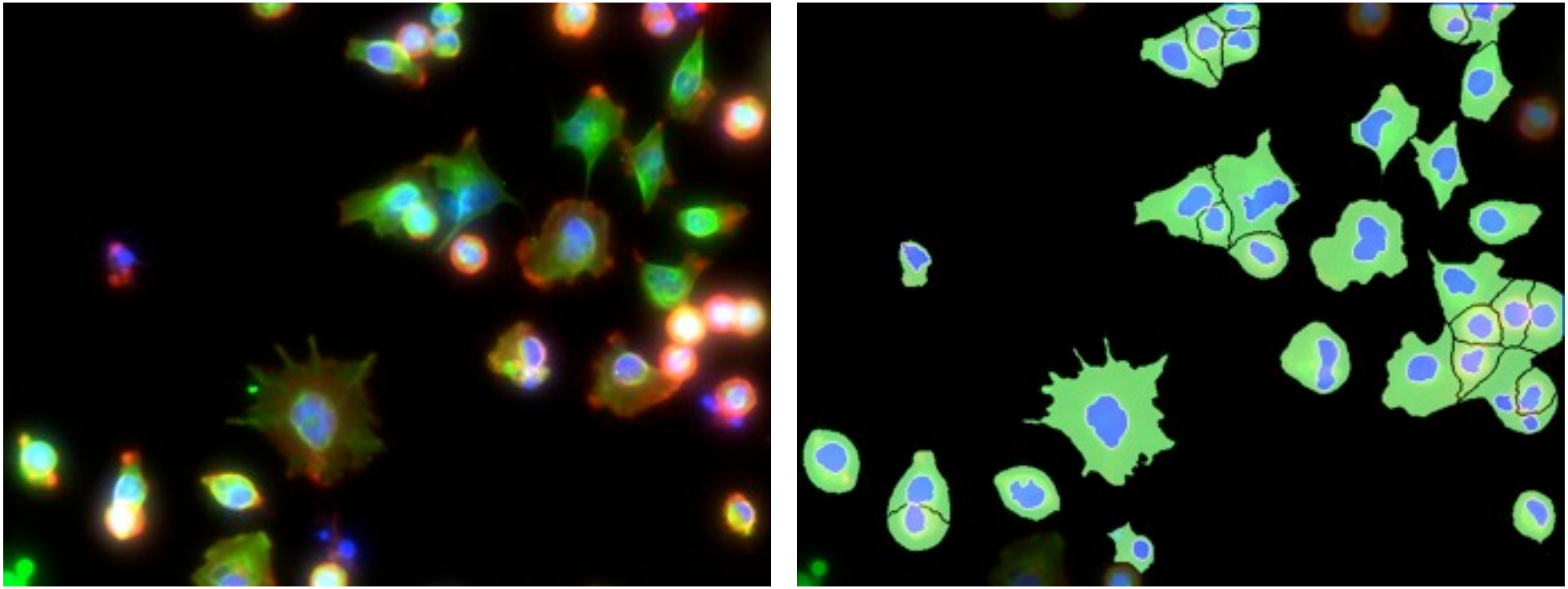
Adaptative thresholding + watershed

Voronoi segmentation using an image gradient based metric

R/Bioconductor package EBImage



Segmentation results



Fully automatic on all 88k images

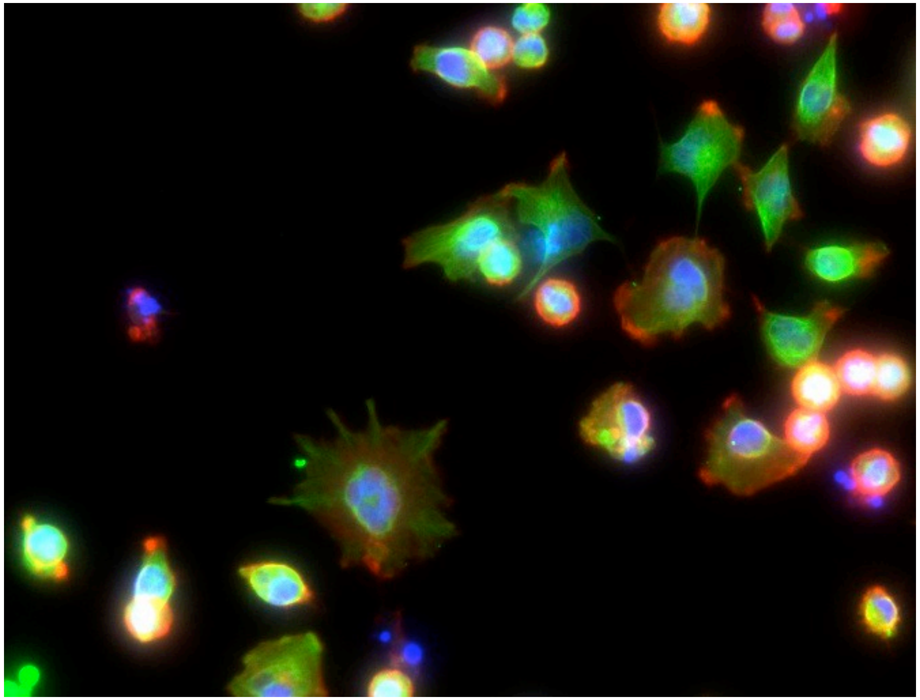
Detailed resolution of boundaries also for adjacent cells

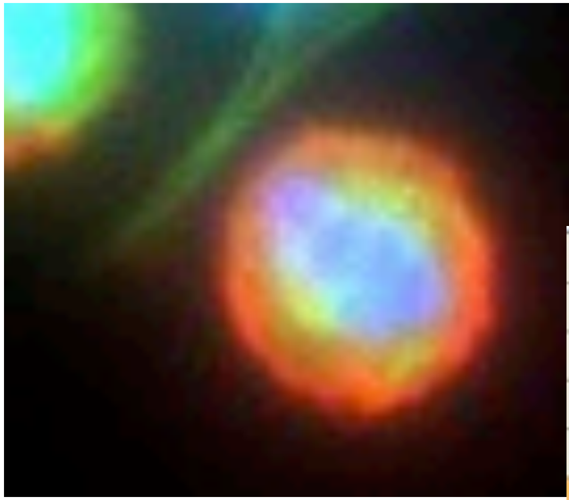
Would not deal with overlapping cells (multilayer, tissue)

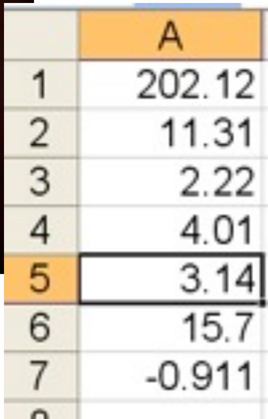
Extracting quantitative cell descriptors

translation and rotation invariant descriptors

- geometry (intensity, size, perimeter, eccentricity…)
- texture (Haralick, Zernike moments…) on each channel
- relative positions, joint distribution moments







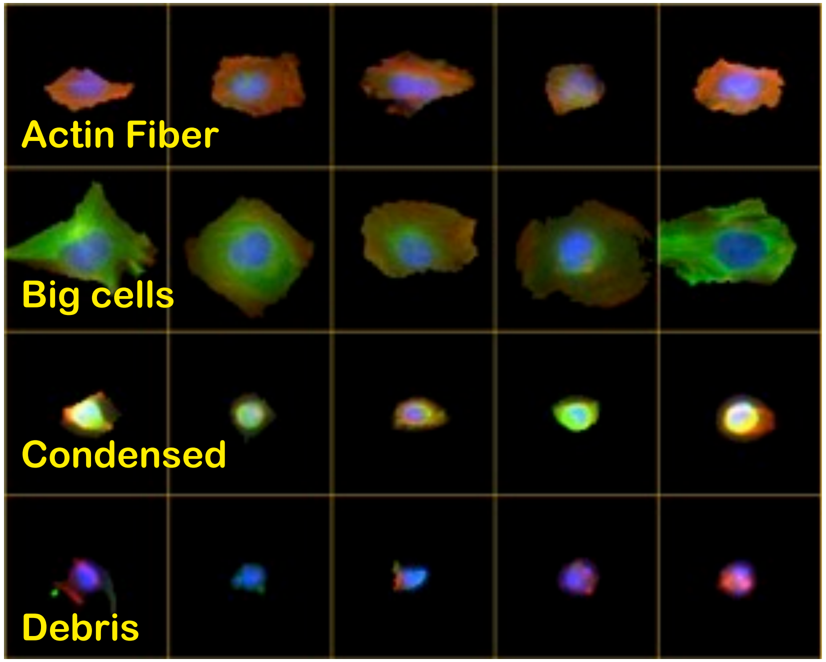
	A
1	202.12
2	11.31
3	2.22
4	4.01
5	3.14
6	15.7
7	-0.911

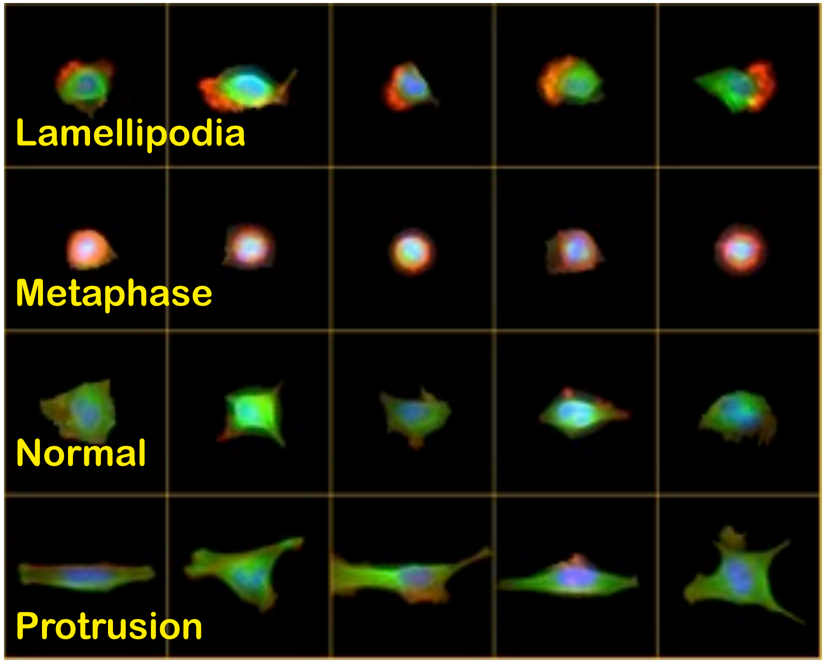
Cell classification

using the numeric descriptors

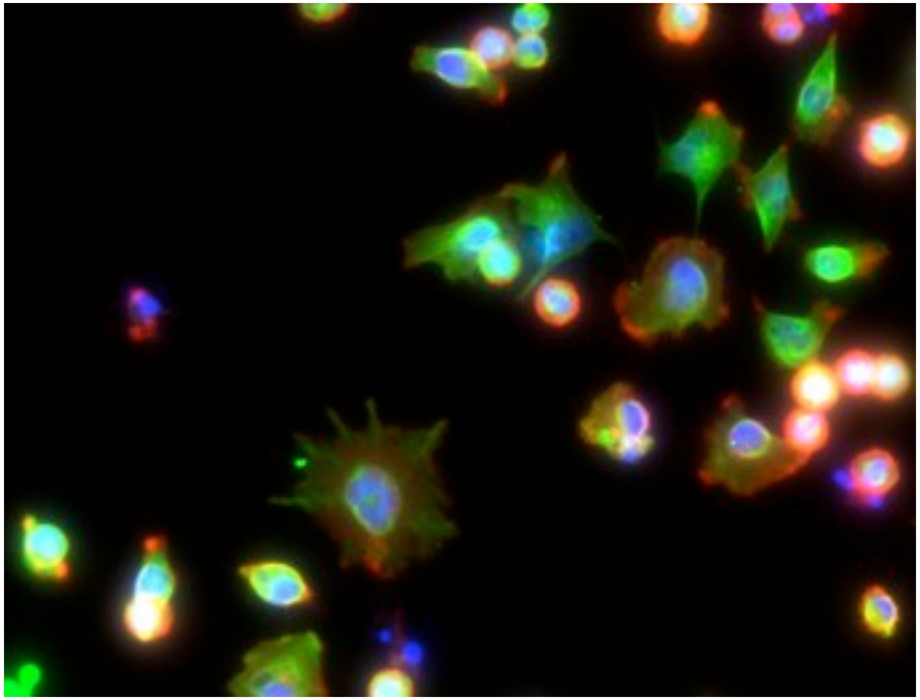
supervised learning, SVM

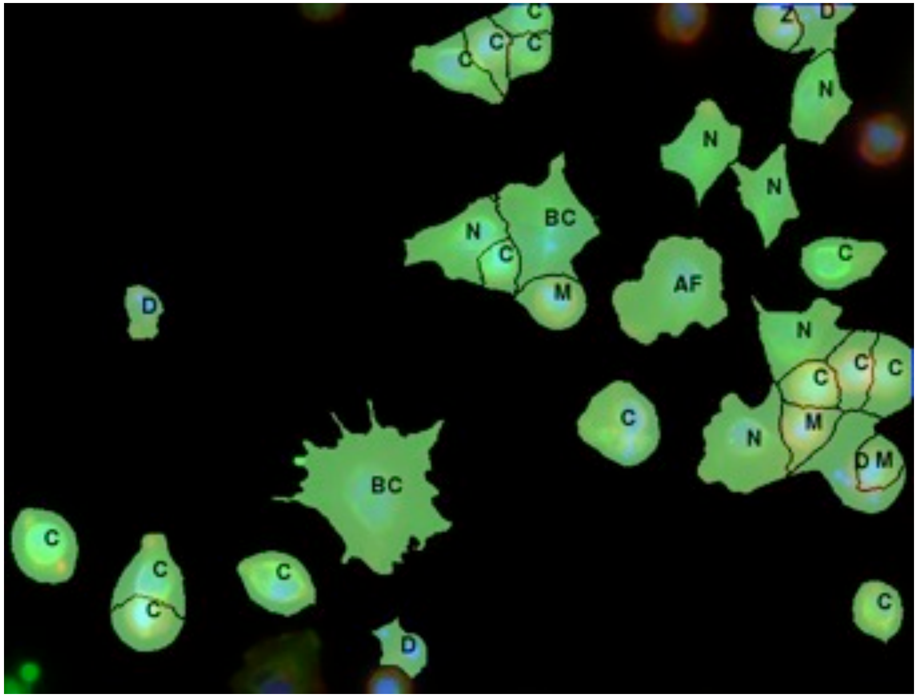
8 classes and a training set of ~3000 cells:



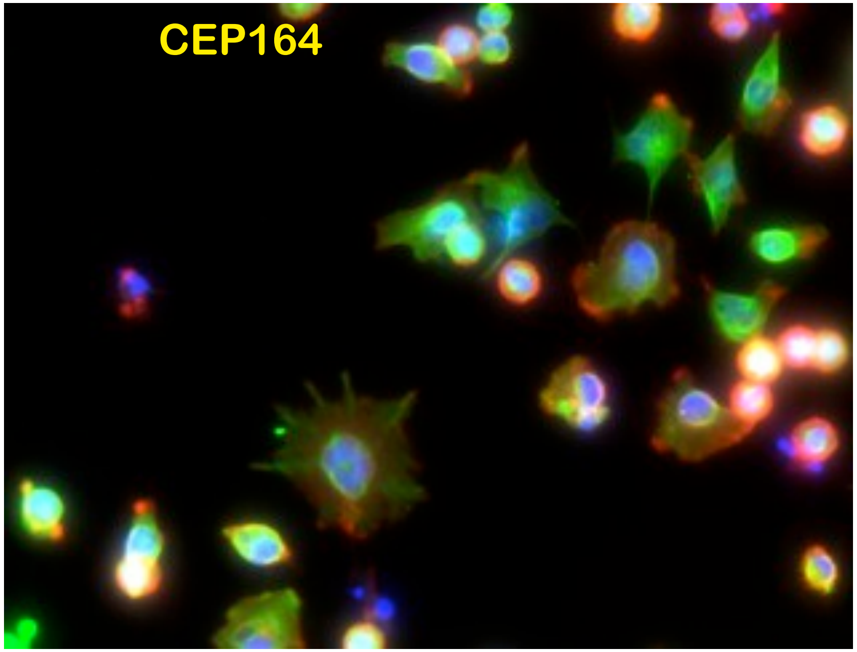


Cell classification





Each siRNA is characterized by its "phenotypic profile"

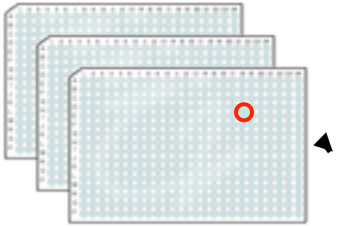


number of cells	128
average intensity	1054.8
average nuclear intensity	1225.6
average cell size	842.3
average nuclear size	278.7
average eccentricy	0.649
avg. nuclear / cell size	2.91
# AF (actin fibers)	2
# BC (big)	7
# M (mitotic)	15
# LA (lamellipodia)	0
# P (with protrusions)	17
# Z (telophase)	2

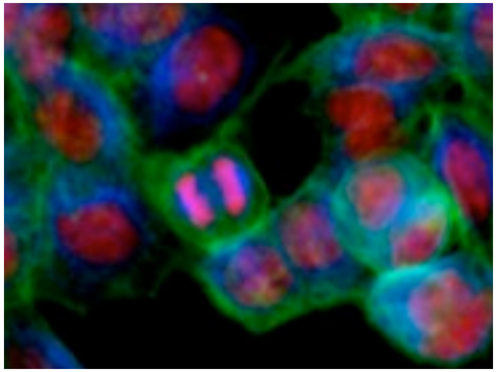
How do you measure distance and similarity in multidimensional phenotypic profile space?

Similarity depends on the choice and weighting of descriptors

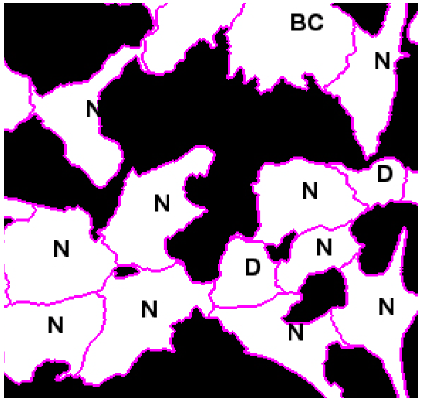
High-throughput RNAi and automated cellular phenotyping



RNAi or drug library



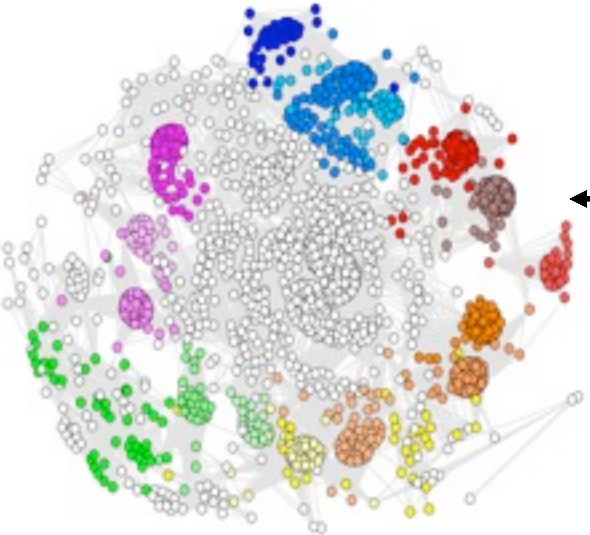
Segmentation



Feature extraction

```
         g.x        g.y    g.s g.p    g.pdm
 [1,] 123.1391   3.288660  194  67  9.241719
 [2,] 206.7460   9.442248  961 153 20.513190
 [3,] 502.9589   7.616438  219  60  8.286918
 [4,]  20.1919  22.358418 1568 157 22.219461
 [5,] 344.7959  45.501992 2259 233 35.158966
 [6,] 188.2611  50.451863 2711 249 28.732680
 [7,] 269.7996  46.404036 2131 180 26.419631
 [8,] 106.6127  58.364243 1348 143 21.662879
 [9,] 218.5582  77.299007 1913 215 25.724580
[10,]  19.1766  81.840147 1908 209 26.303760
[11,]   6.3558  62.017647  340  68 10.314127
[12,]  58.9873  86.034128 2139 214 27.463158
[13,] 245.1087  94.387405 1048 123 18.280901
[14,] 411.2741 109.198678 2572 225 28.660816
[15,] 167.8151 107.966014 1942 160 24.671533
[16,] 281.7084 121.609892 2871 209 31.577270
```

Quantitative cell and organelle features



multivariate phenotypic landscape

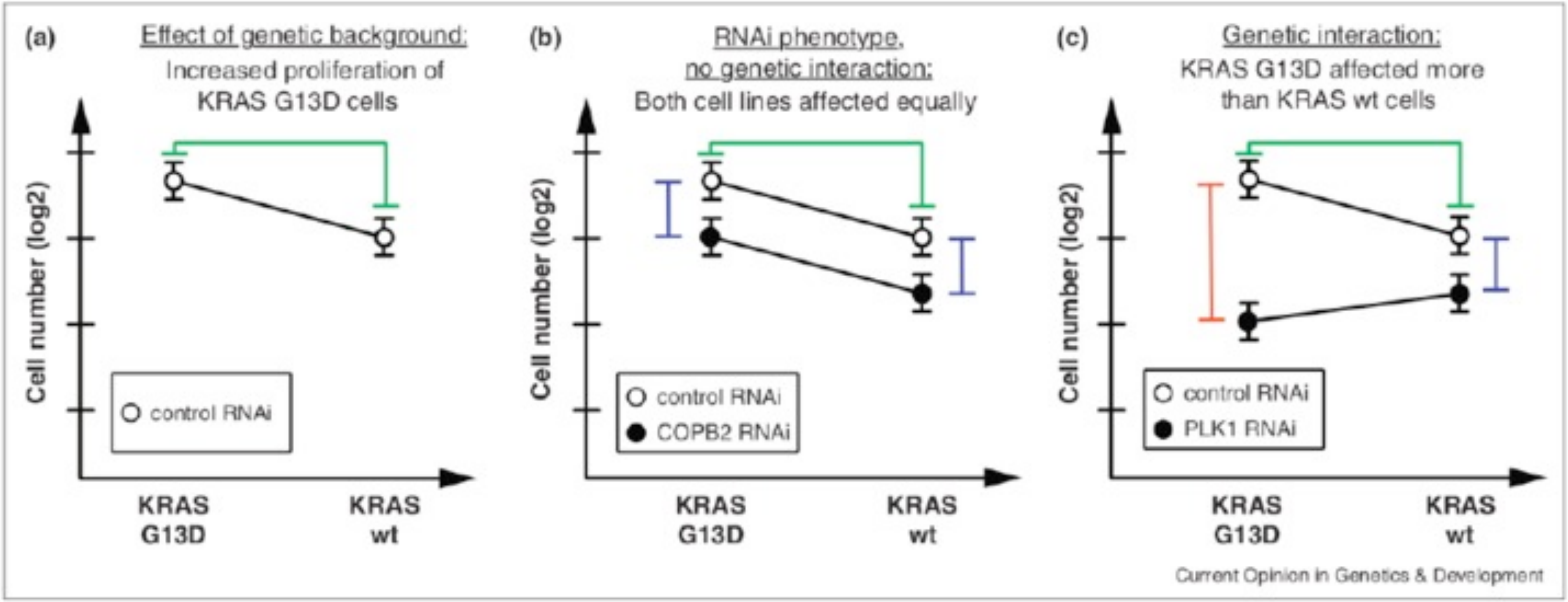
Michael Boutros

Gregoire Pau

Boutros, Bras, Huber, **Genome Biol.** 2006
Fuchs, Pau et al. **Mol. Sys. Biol.** 2010
Pau, Fuchs et al. **Bioinf.** 2010
Neumann et al. **Nature** 2010
Kuttenkeuler et al. **J. Innate Imm.** 2010

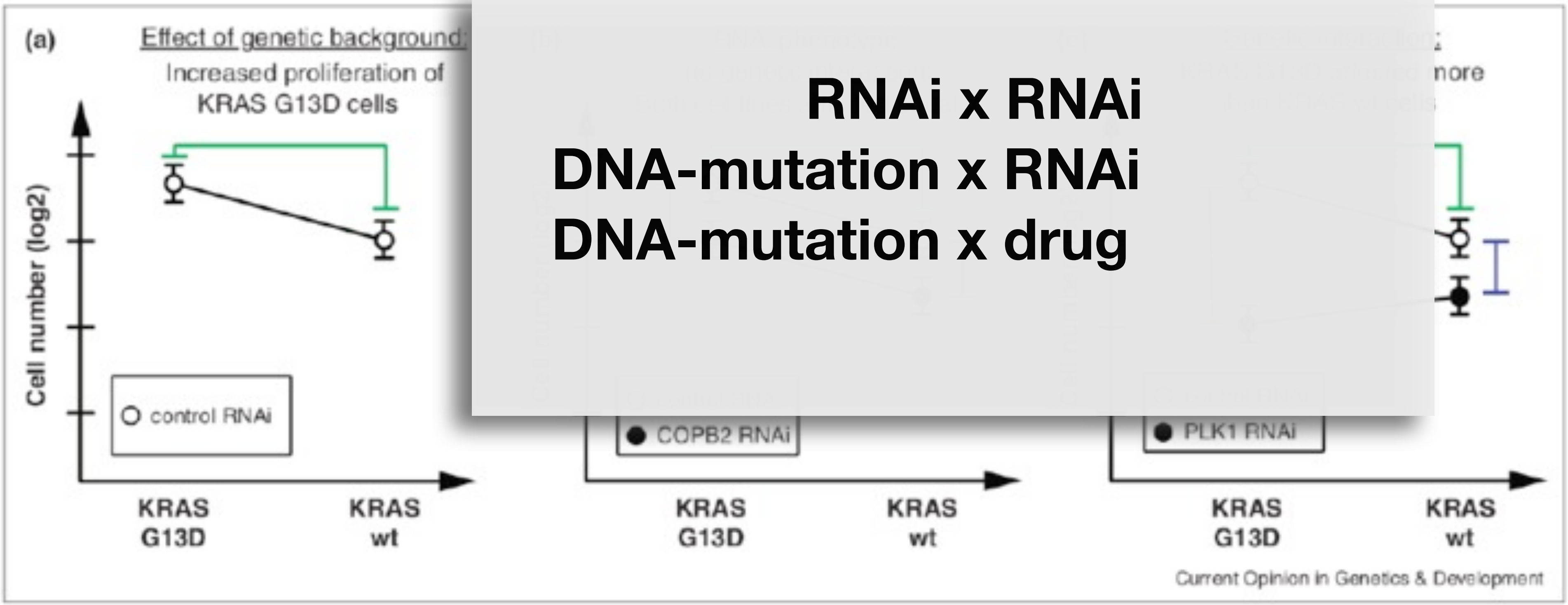
Axelsson et al. **BMC Bioinf.** 2011
Horn et al. **Nature Methods** 2011
Laufer et al. **Nature Methods** 2013

Genetic interactions



Luo, J. et al., Cell (2009).
Sandmann, T. & Boutros, M.
Current Opinion in Genetics & Development (2012)

Genetic interactions

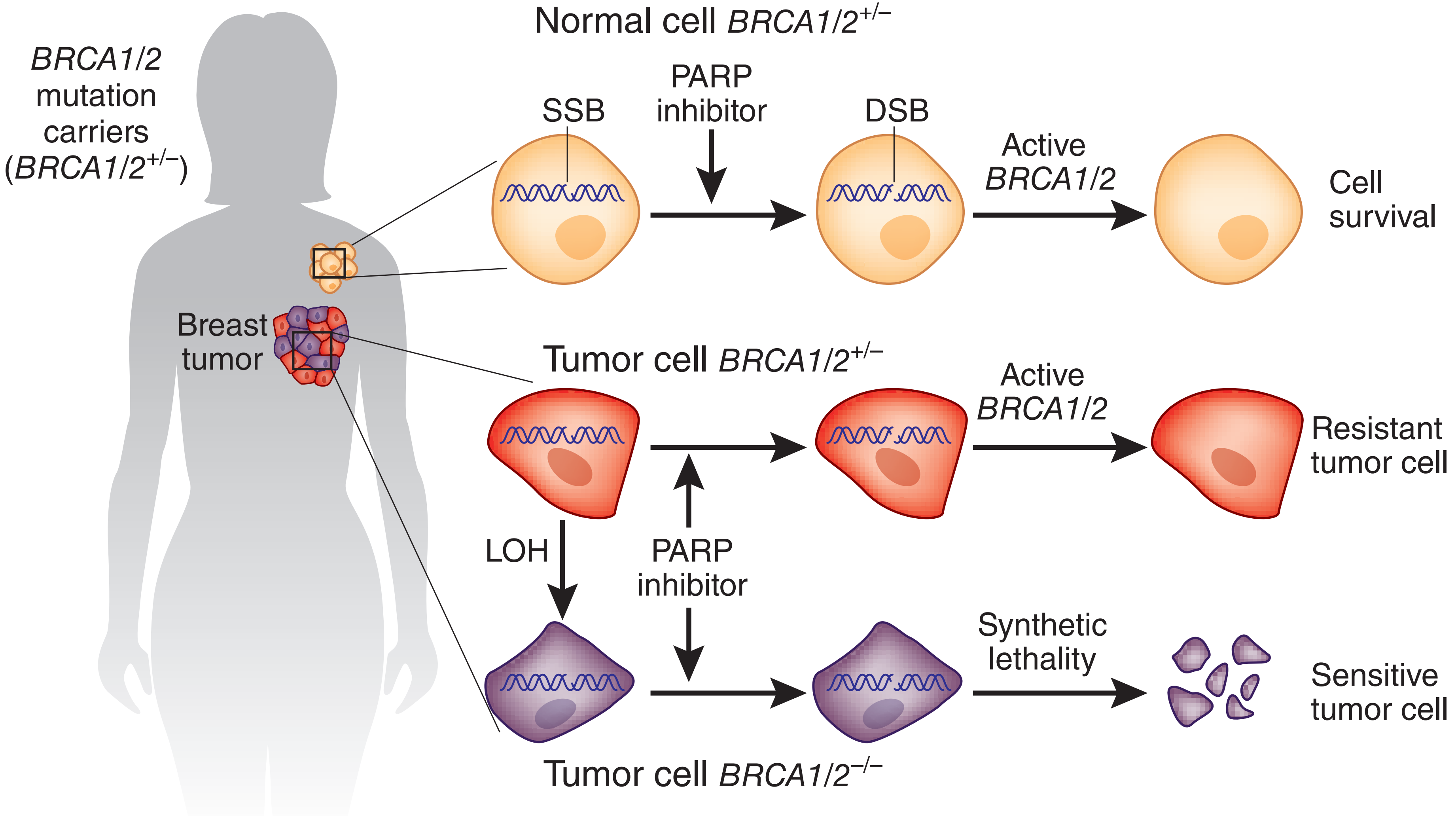


Current Opinion in Genetics & Development

RNAi x RNAi
DNA-mutation x RNAi
DNA-mutation x drug

Luo, J. et al., Cell (2009).
Sandmann, T. & Boutros, M.
Current Opinion in Genetics & Development (2012)

an example of synthetic lethality



atie Vicari

Genetic interactions capture nonlinearity of a system

$$y = f(x_1, \ldots, x_n).$$

phenotype **genotype**

Genetic interactions capture nonlinearity of a system

$$y = f(x_1, \ldots, x_n).$$

phenotype **genotype**

$$y - y^0 \quad = \quad \sum_{i=1}^{n} m_i (x_i - x_i^0)$$

Genetic interactions capture nonlinearity of a system

$$y = f(x_1, \ldots, x_n).$$

phenotype **genotype**

$$y - y^0 = \sum_{i=1}^{n} m_i (x_i - x_i^0) + \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} (x_i - x_i^0)(x_j - x_j^0) + \ldots,$$

buffering, sensitization, epistasis, ...

Epistasis as the primary factor in molecular evolution

Michael S. Breen[1], Carsten Kemena[1], Peter K. Vlasov[1], Cedric Notredame[1] & Fyodor A. Kondrashov[1,2]

25 OCTOBER 2012 | VOL 490 | NATURE | 535

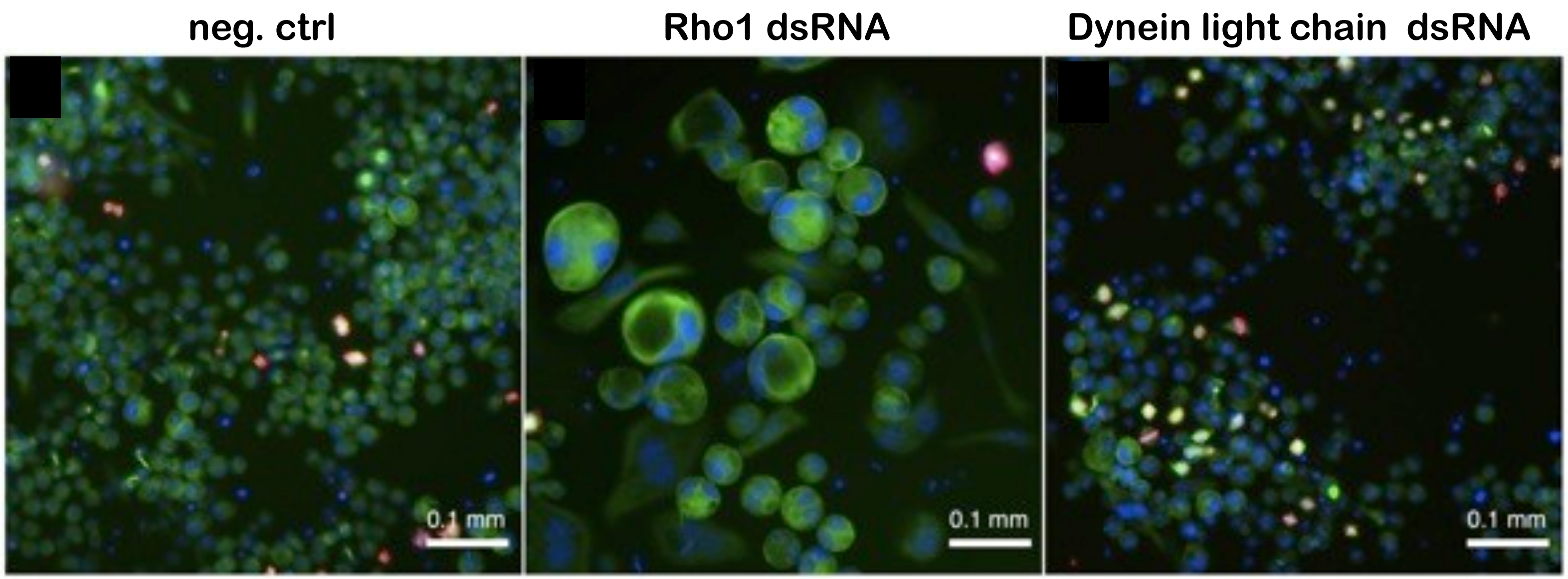
Comparing amino acid substitution rates over short and long evolutionary time indicates that fitness effect of almost all mutations depends is context-dependent:

Epistasis is pervasive.

Genetic interactions for multiple phenotypes

384-well plates, microscopy readout with 3 channels
(DAPI, phospho-His3, aTubulin)

Fly: 1367 x 72 genes (Nat. Methods 2011 & unpublished)
Human: 323 x 20 (Nat. Methods 2013)



Bernd Fischer

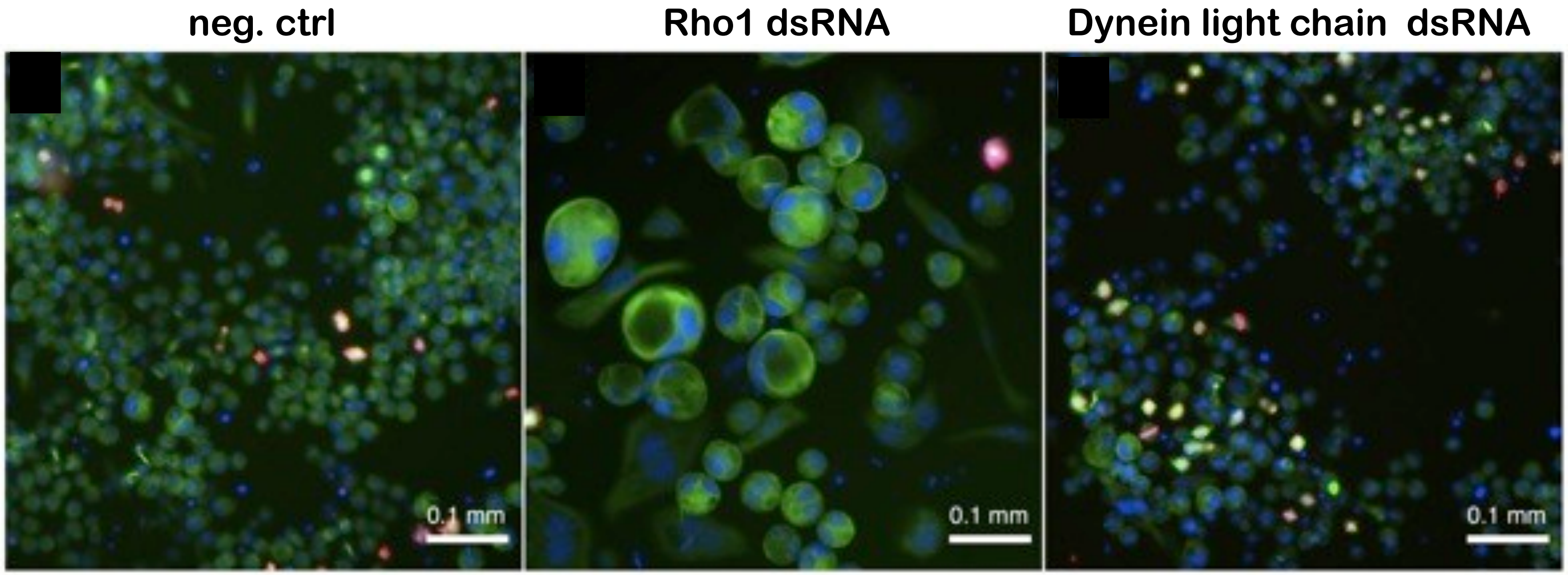
Horn*, Sandmann*, Fischer*, ..., Huber, Boutros. **Nature Methods** 2011

Genetic interactions for multiple phenotypes

384-well plates, microscopy readout with 3 channels (DAPI, phospho-His3, aTubulin)

Fly: 1367 x 72 genes (Nat. Methods 2011 & unpublished)

Human: 323 x 20 (Nat. Methods 2013)

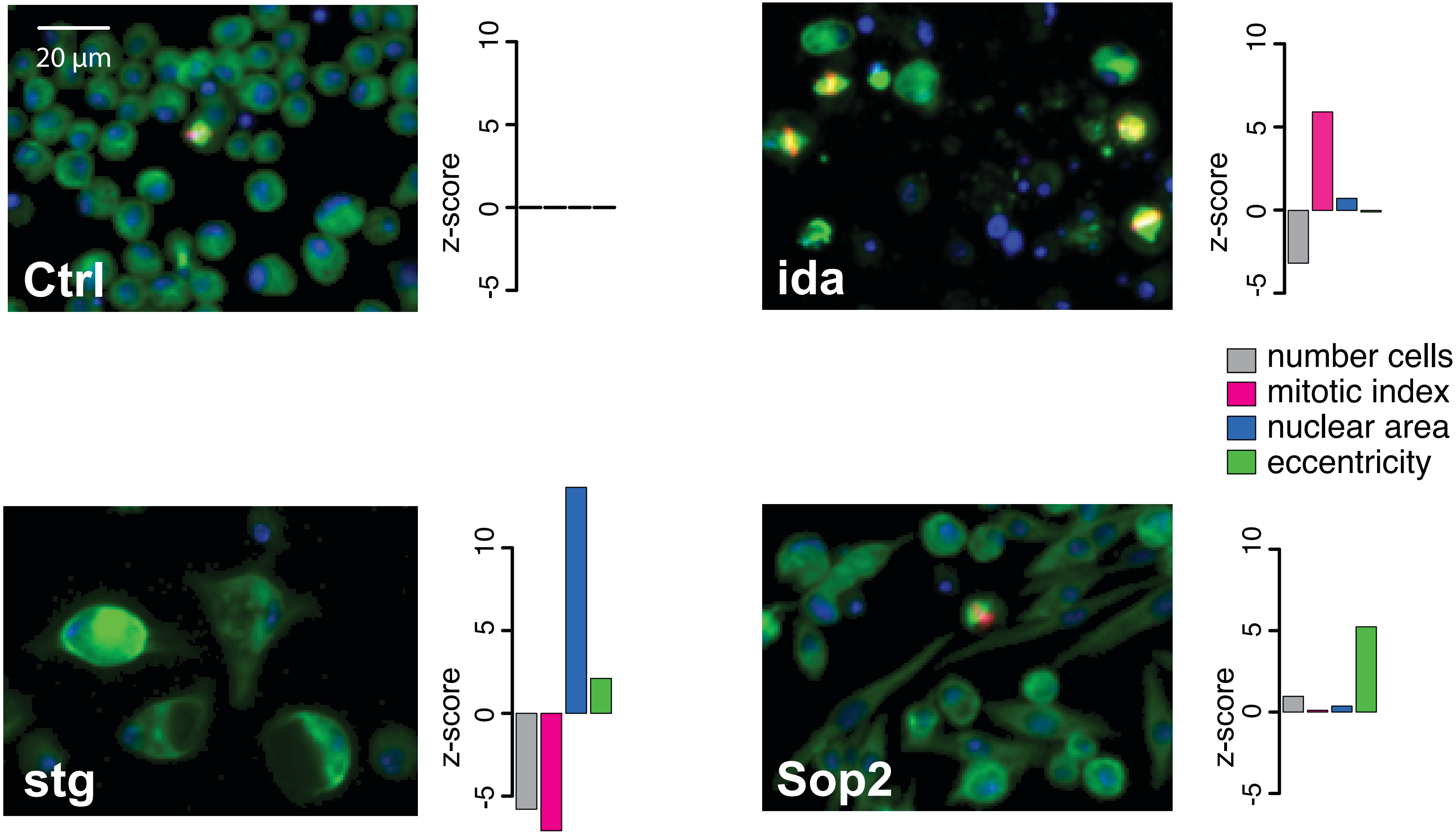


number of cells | area | mitotic index | shape | variances

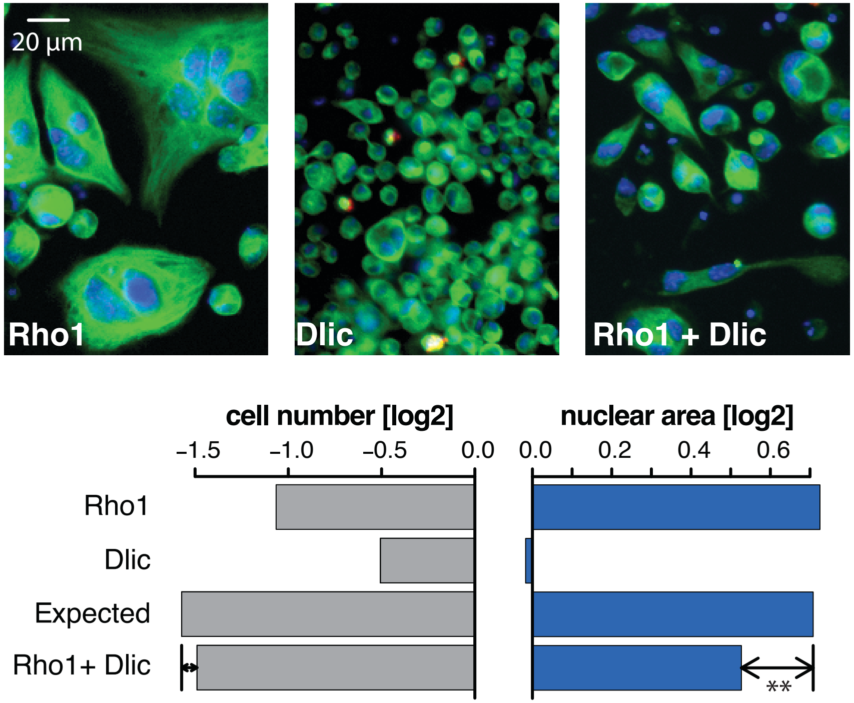
Horn*, Sandmann*, Fischer*, ..., Huber, Boutros. **Nature Methods** 2011

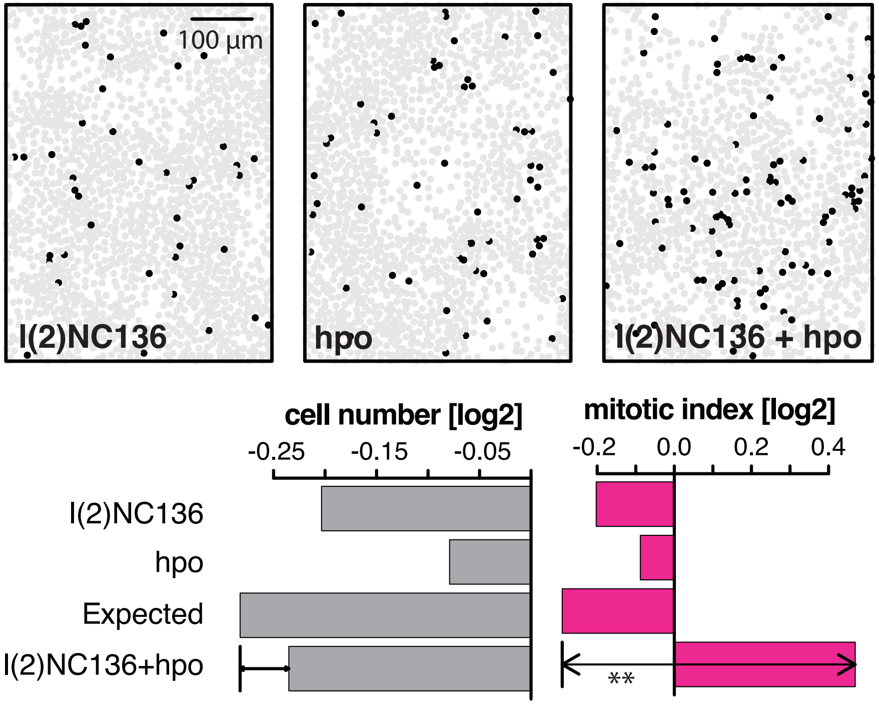
Bernd Fischer

Multiple phenotypes are observed

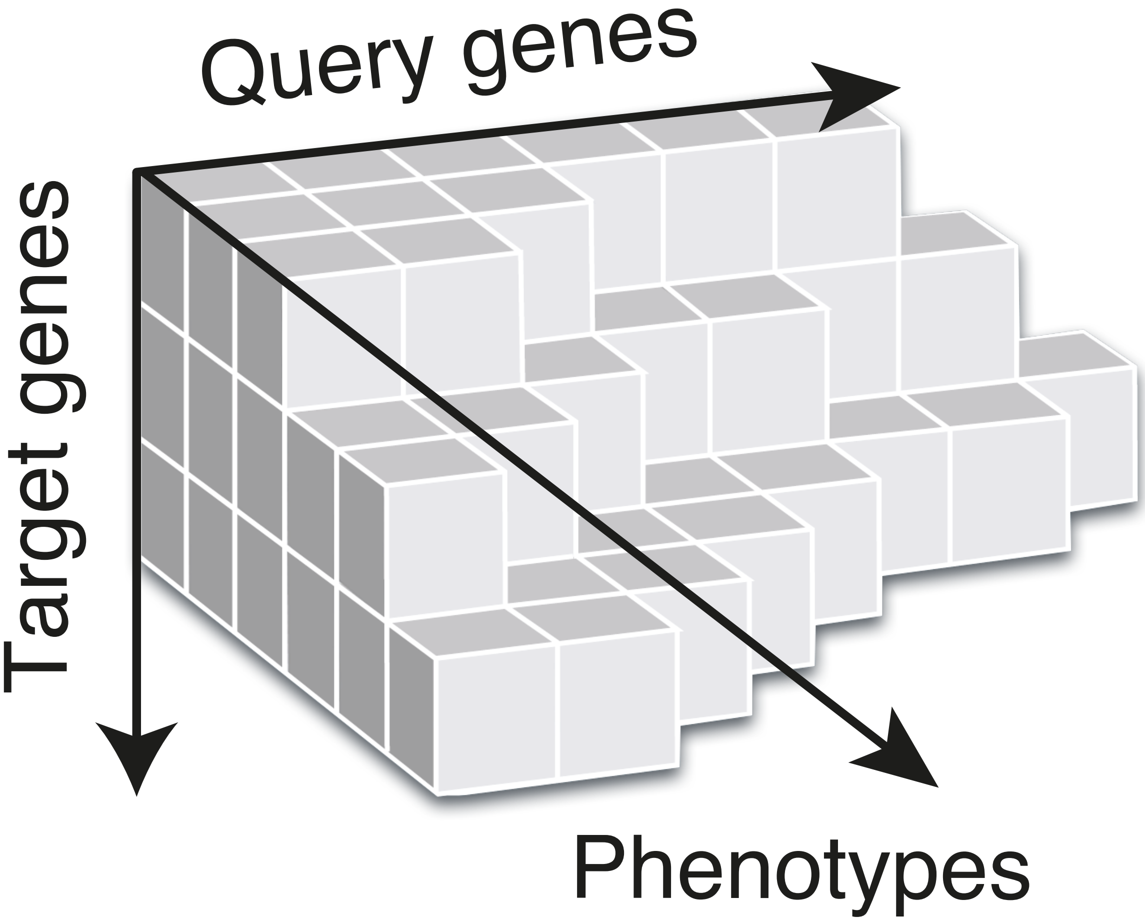


Distinct genetic interactions in multiple phenotypes





3D data cube



1293 target genes

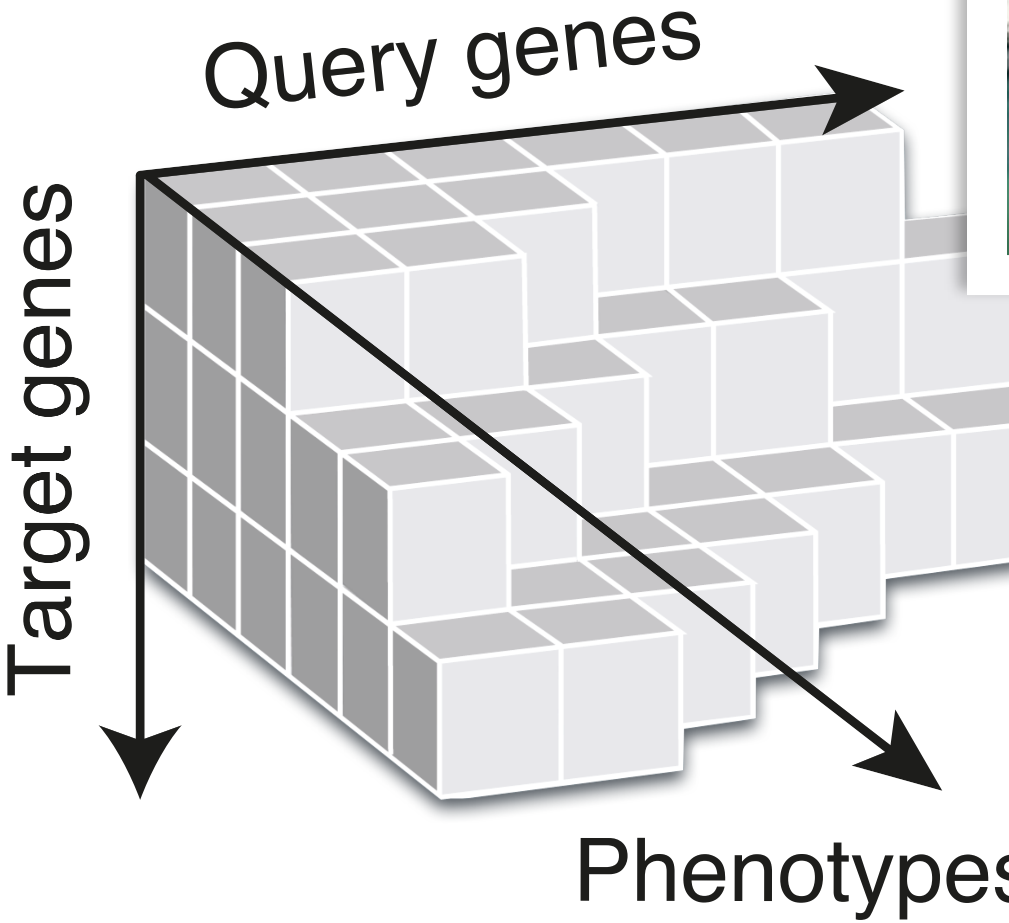
x 2 dsRNA

x 72 query genes

x 2 dsRNA

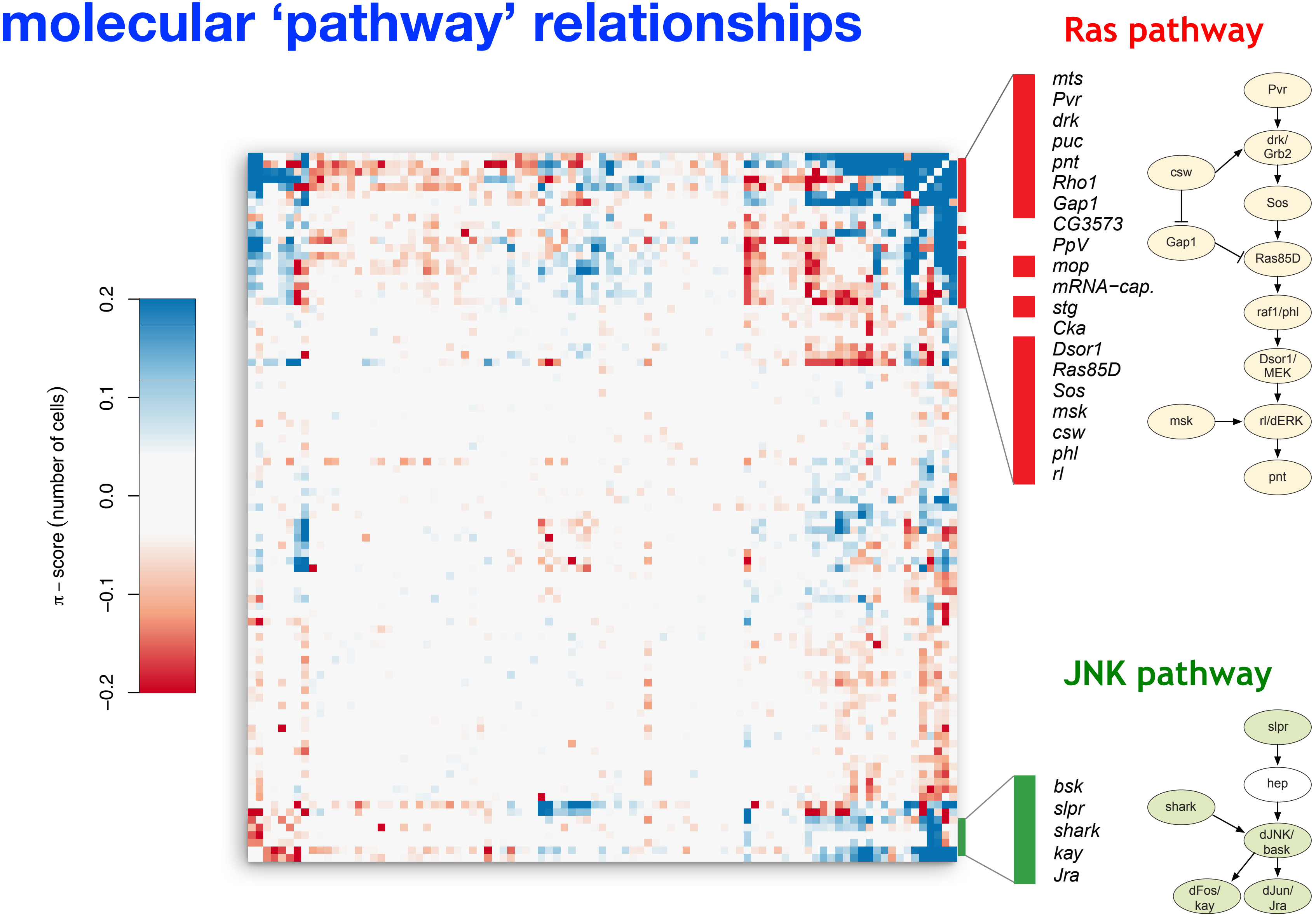
x 21 features

3D data cube



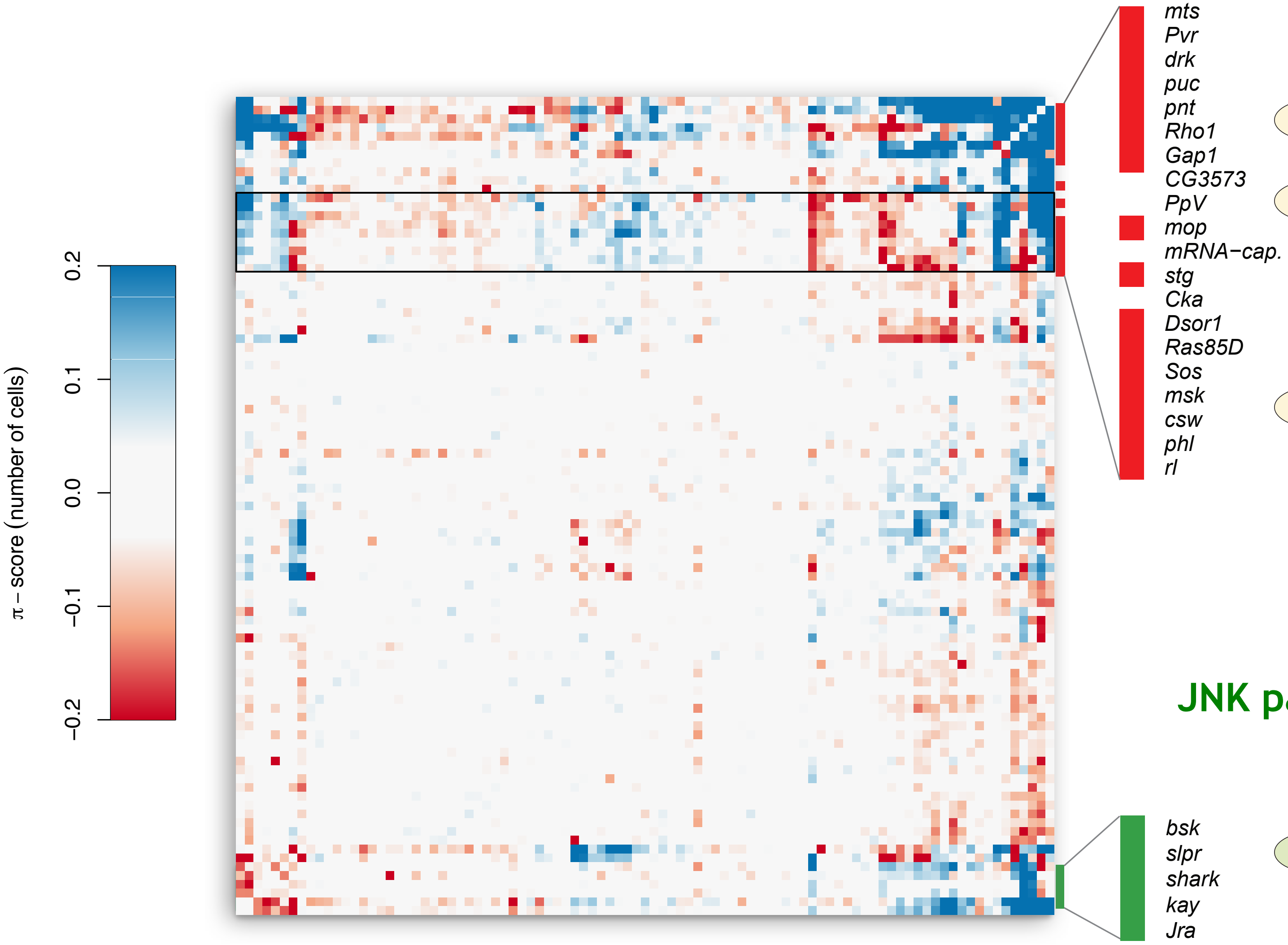
	1293	target genes
x	2	dsRNA
x	72	query genes
x	2	dsRNA
x	21	features

Similarity of interaction profiles reflects molecular ‘pathway’ relationships

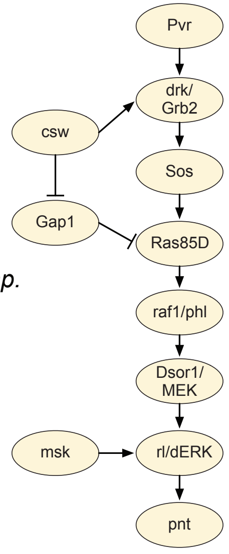


Horn*, Sandmann*, Fischer*, ..., Huber, Boutros. **Nature Methods** 2011

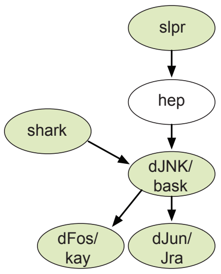
Similarity of interaction profiles reflects molecular ‘pathway’ relationships



Ras pathway

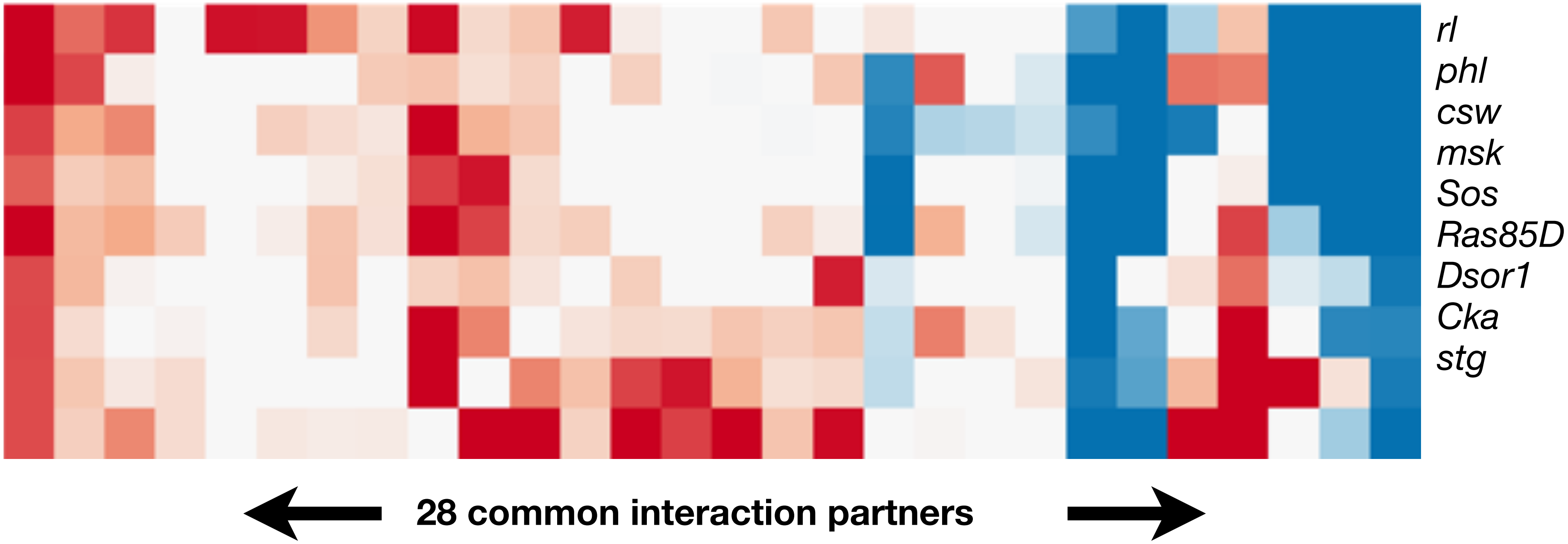


JNK pathway



Horn*, Sandmann*, Fischer*, ..., Huber, Boutros. **Nature Methods** 2011

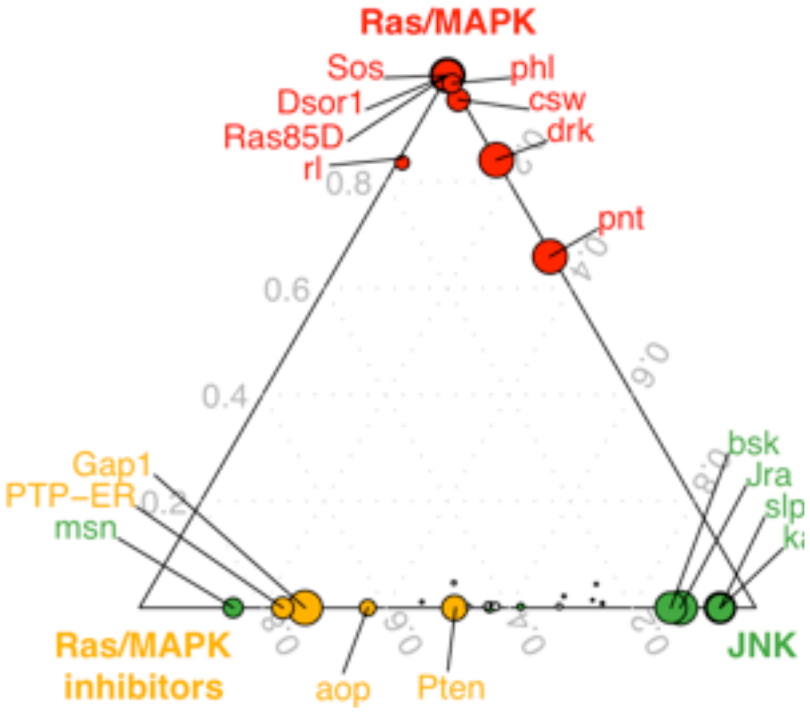
Similarity of interaction profiles reflects molecular ‘pathway’ relationships



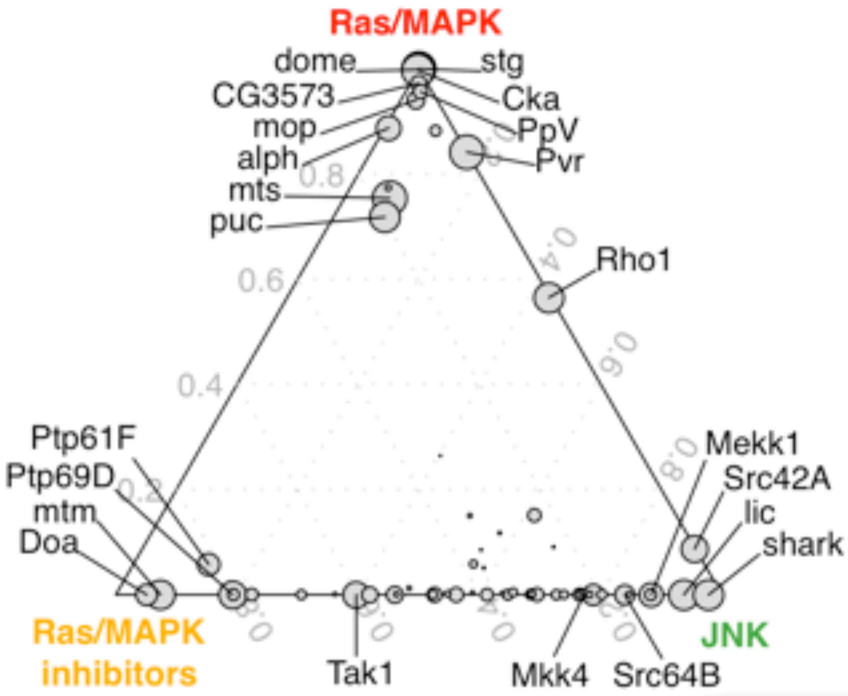
Horn*, Sandmann*, Fischer*, ..., Huber, Boutros. **Nature Methods** 2011

Interaction profiles predict functional roles

Classification of profiles by sparse linear discriminant analysis, 3 classes(posterior probabilities)



Performance on training set

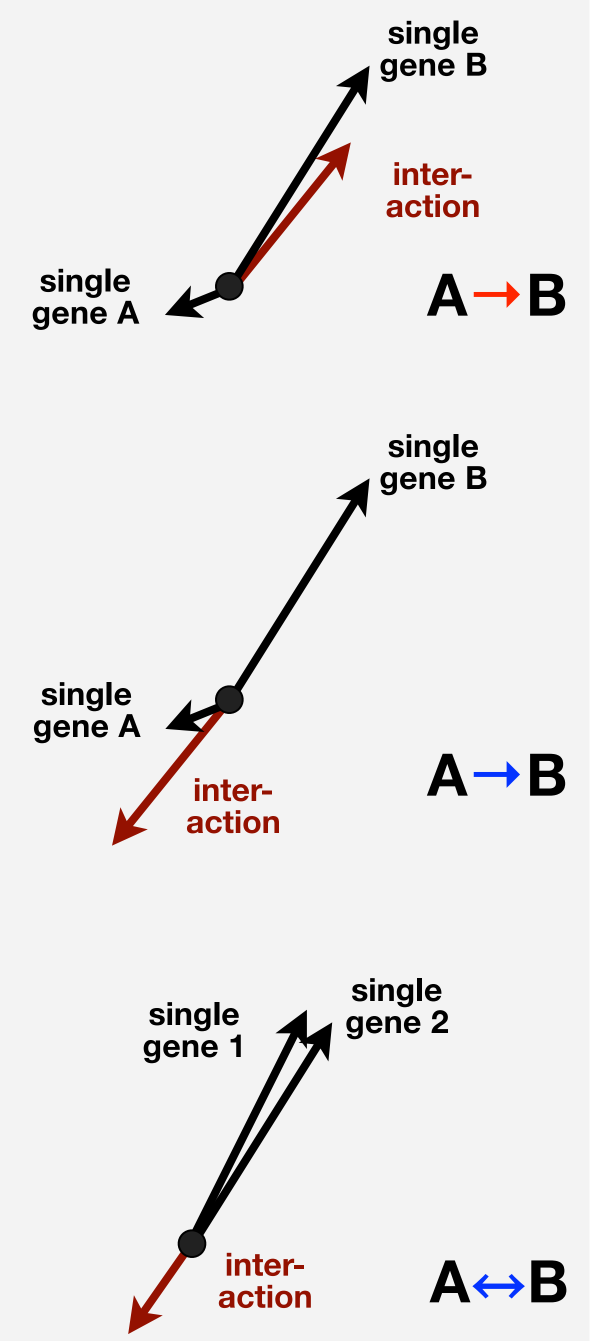


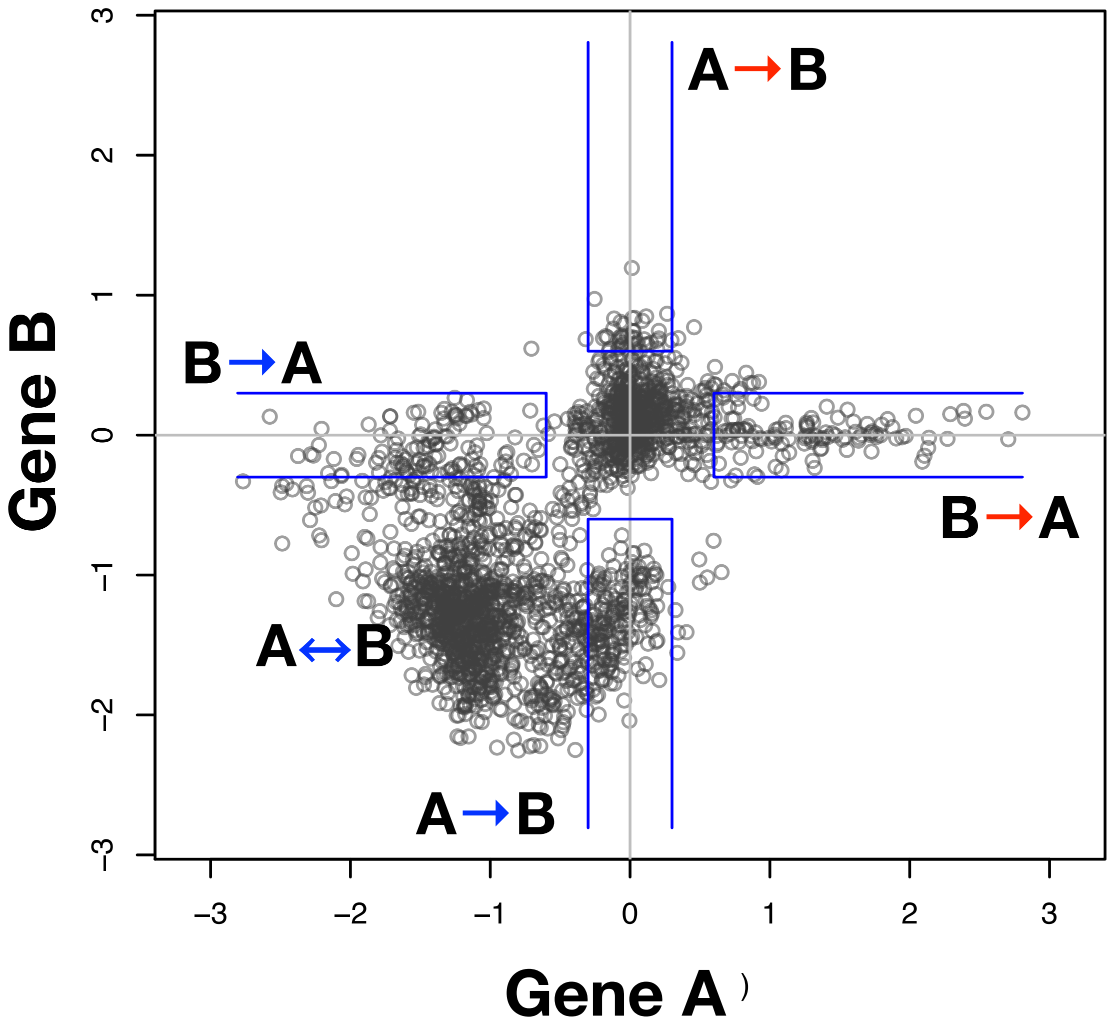
Performance on test set

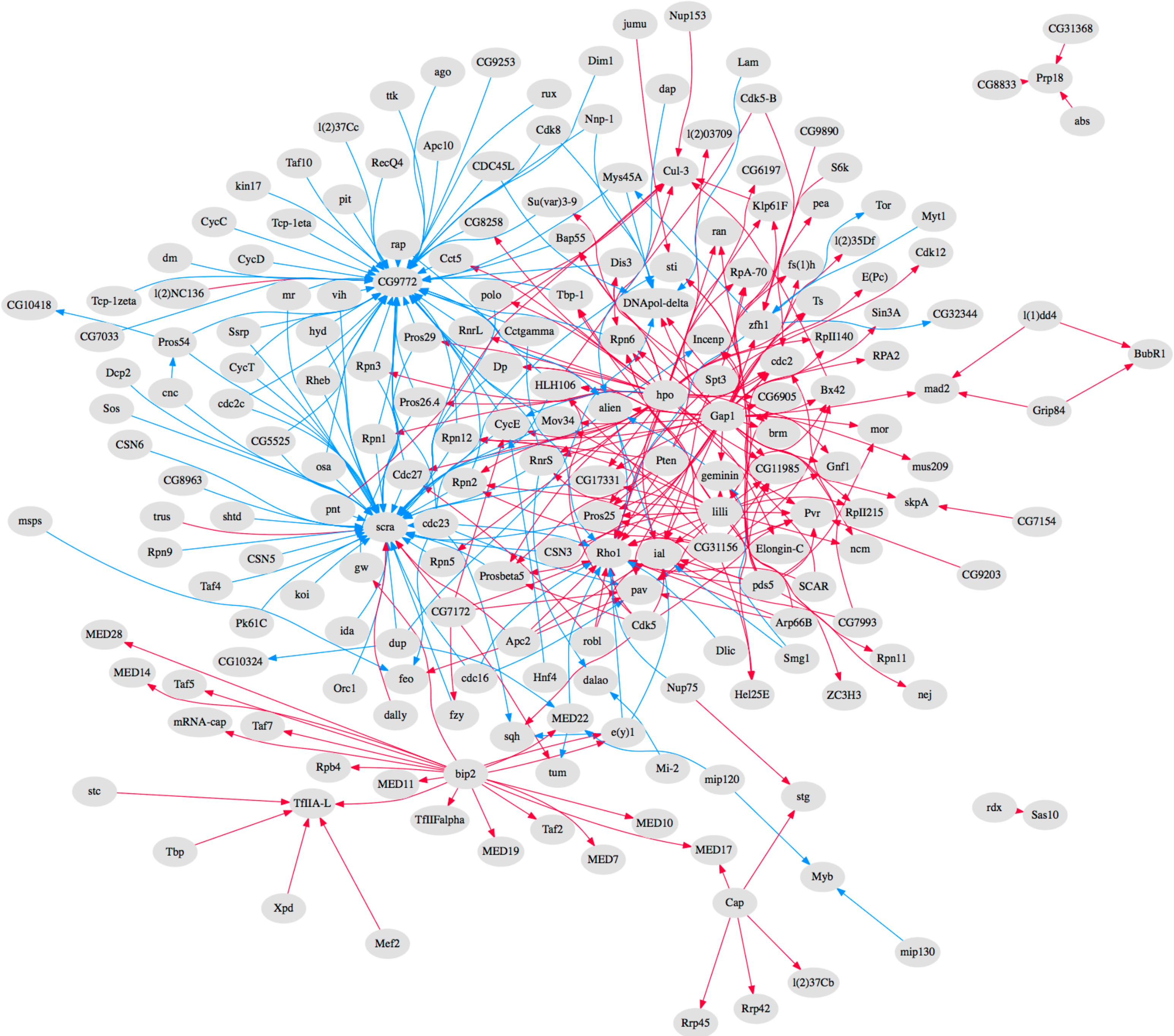
Horn*, Sandmann*, Fischer*, ..., Huber, Boutros. **Nature Methods** 2011

Bernd Fischer

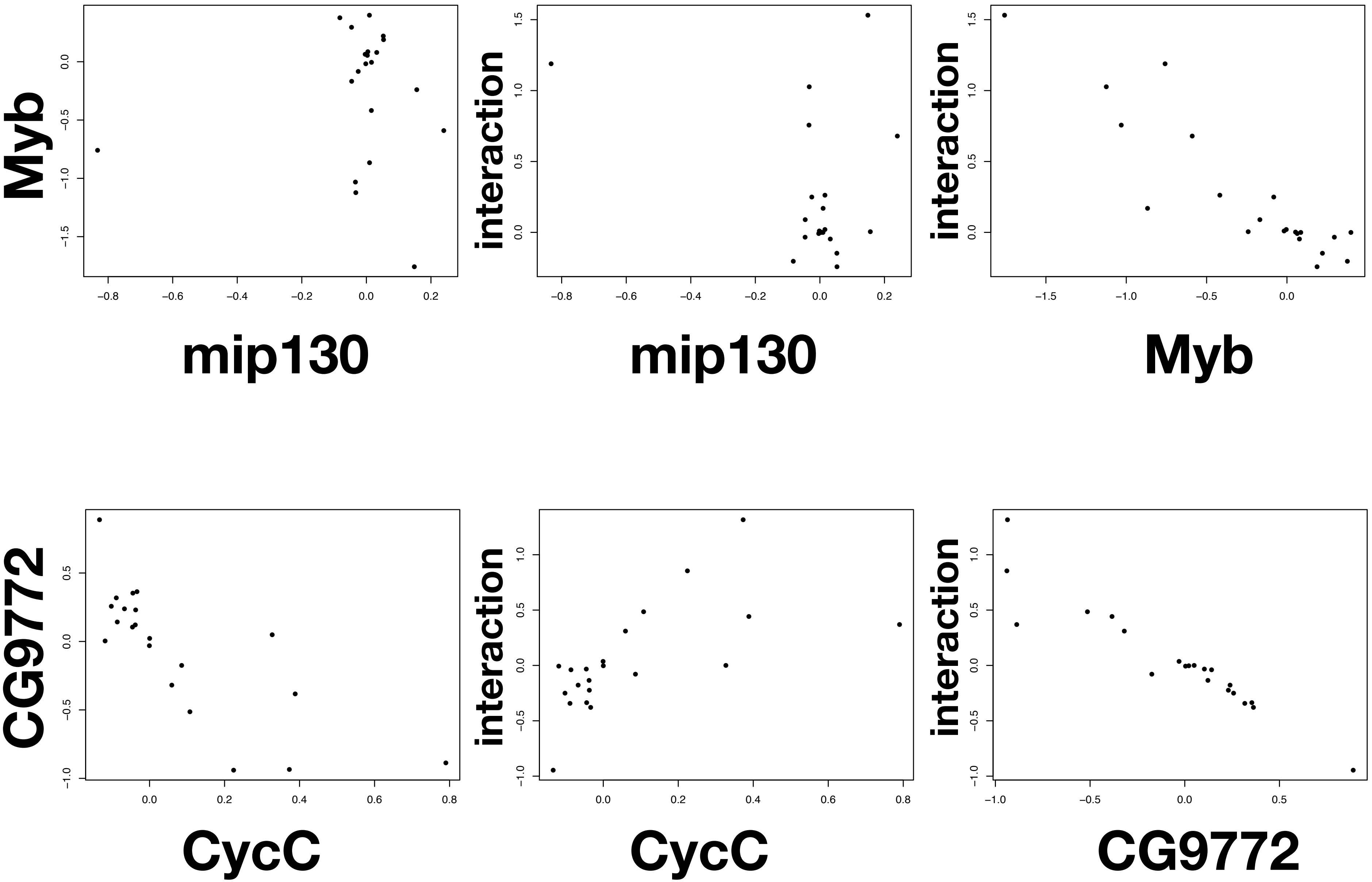
directed genetic interactions



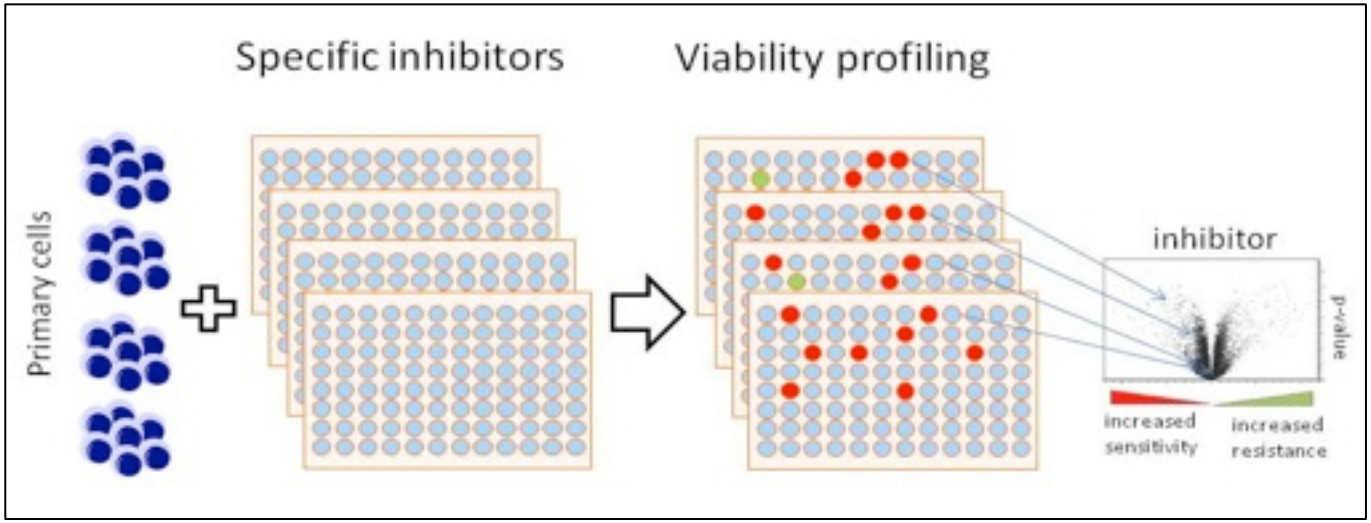


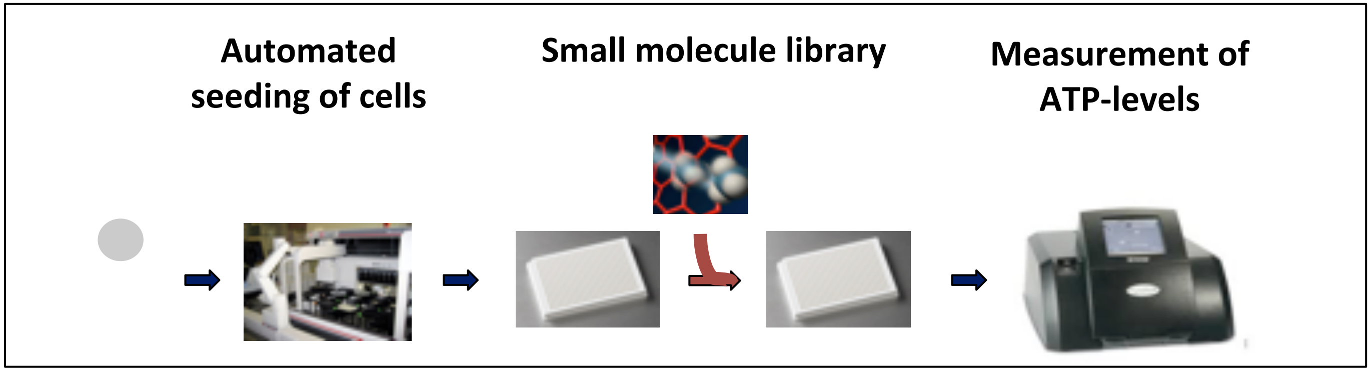

jumu
Nup153
CG31368
CG9253
Dim1
Lam
ago
rux
dap
Prp18
CG8833
ttk
Cdk5-B
abs
Nnp-1
l(2)37Cc
Cdk8
l(2)03709
CG9890
Apc10
Taf10
RecQ4
CDC45L
Mys45A
Cul-3
CG6197
S6k
kin17
pit
Su(var)3-9
Klp61F
pea
Tor
CycC
Tcp-1eta
CG8258
ran
Myt1
rap
Bap55
l(2)35Df
dm
CycD
Cct5
Dis3
sti
fs(1)h
Cdk12
CG9772
RpA-70
E(Pc)
Tcp-1zeta
l(2)NC136
mr
vih
polo
Tbp-1
DNApol-delta
Ts
CG10418
Sin3A
CG32344
l(1)dd4
CG7033
Ssrp
hyd
RnrL
Cctgamma
zfh1
Pros54
Pros29
Rpn6
Incenp
RpII140
BubR1
Dcp2
CycT
Rpn3
Dp
cdc2
RPA2
cnc
Rheb
HLH106
hpo
Spt3
Bx42
mad2
Sos
cdc2c
Pros26.4
alien
CG6905
Grip84
Gap1
CycE
Mov34
brm
mor
CSN6
CG5525
Rpn1
Rpn12
osa
RnrS
Pten
CG11985
Gnf1
mus209
CG8963
Cdc27
Rpn2
CG17331
geminin
skpA
pnt
lilli
Pvr
RpII215
CG7154
msps
trus
shtd
scra
cdc23
Pros25
ncm
Rpn9
CSN5
CSN3
Rho1
ial
CG31156
Elongin-C
Rpn5
gw
Prosbeta5
CG9203
Taf4
koi
pav
pds5
SCAR
Pk61C
CG7172
Cdk5
CG7993
Arp66B
MED28
ida
dup
Apc2
robl
Dlic
Smg1
Rpn11
CG10324
MED14
Taf5
Orc1
feo
cdc16
Hnf4
dalao
Nup75
Hel25E
ZC3H3
nej
dally
fzy
MED22
mRNA-cap
Taf7
sqh
e(y)1
Rpb4
bip2
tum
Mi-2
mip120
MED11
stc
TfIIA-L
stg
rdx
Sas10
TfIIFalpha
Taf2
MED10
MED17
Tbp
MED19
MED7
Myb
Cap
Xpd
mip130
l(2)37Cb
Mef2
Rrp42
Rrp45

Robust manifestation of epistasis through multivariate phenotypes



Genotype-dependence of drug sensitivity in lymphoma and leukemia





Automated seeding of cells

Small molecule library

Measurement of ATP-levels

Substance	Target
ABT-263 (Navitoclax)	Bcl-2
PCI-32765	Btk
CAL-101	PI3Kδ
SNS-032	CDK 2,7,9
Olaparib (AZD2281)	PARP
Fludarabine	purine analogue
Vorinostat	HDAC I, IIa, IIb, IV
Bortezomib (PS-341)	Proteasome
MS-275 (Entinostat)	HDAC I, III
Nutlin-3	MDM2
Enzastaurin	PKC
AZD6244 (Selumetinib)	MEK1/2
BIBW2992 (Afatinib)	EGFR/ERBB2
Deforolimus	mTOR
MK-1775	WEE1
GDC-0449	HH
AT13387	Hsp90
RO4929097	gamma-secretase
XAV-939	Wnt
AZD7762	CHK1/2
ON-01910	PLK
SP600125	JNK
LY2228820	p38 MAPK

etc.

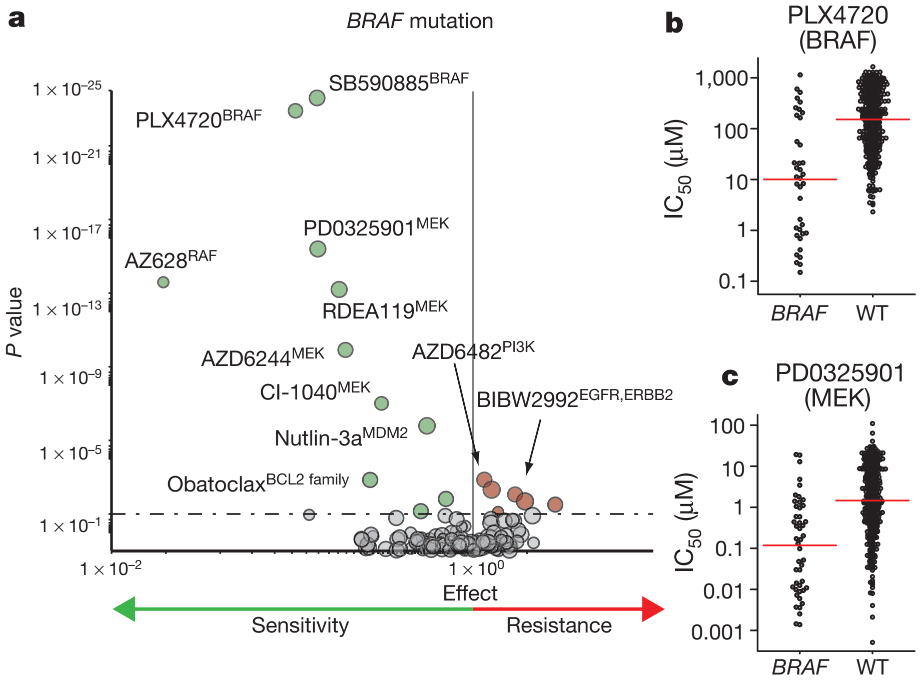
Thorsten Zenz, Leo Sellner, NCT

Drug screens in pan-cancer cell line panels

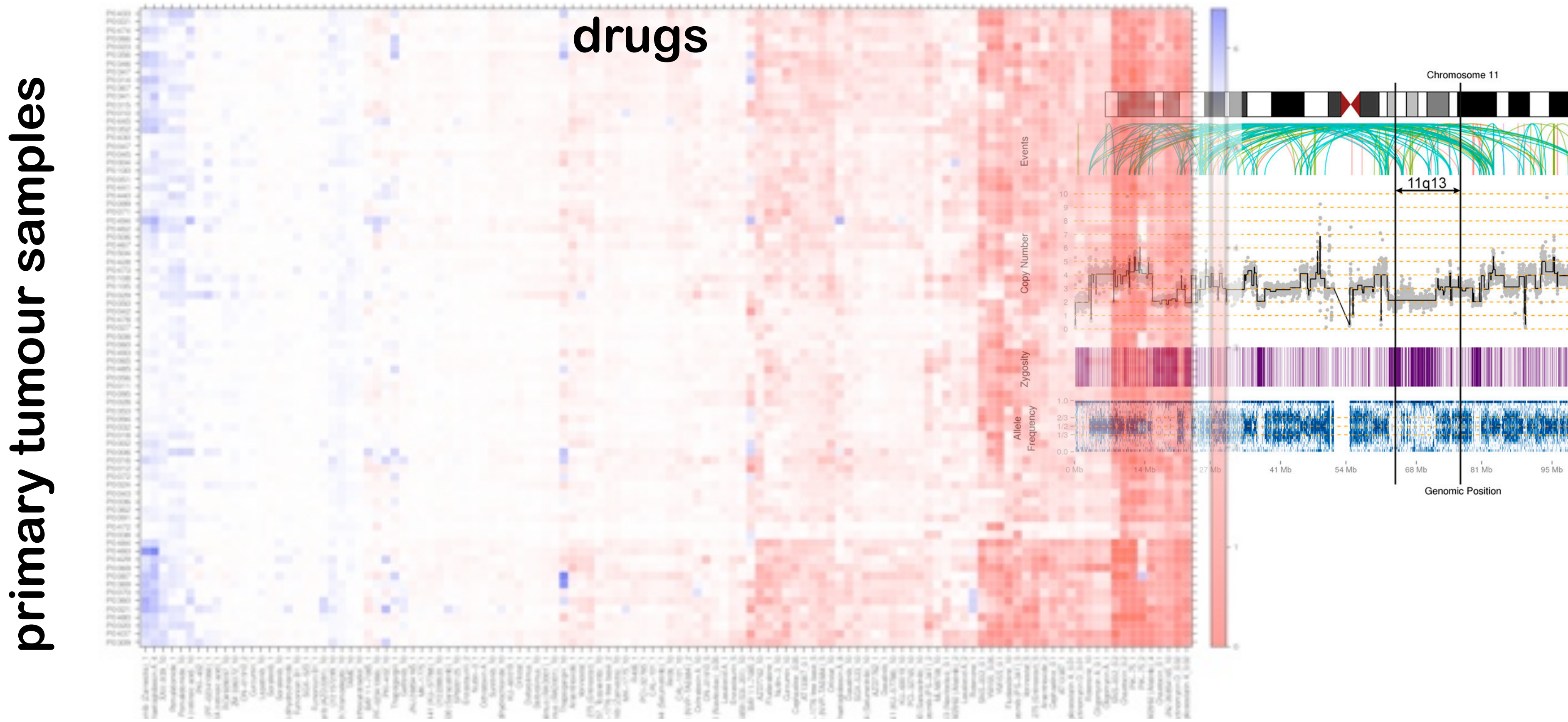
Garnett 2012: Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature.

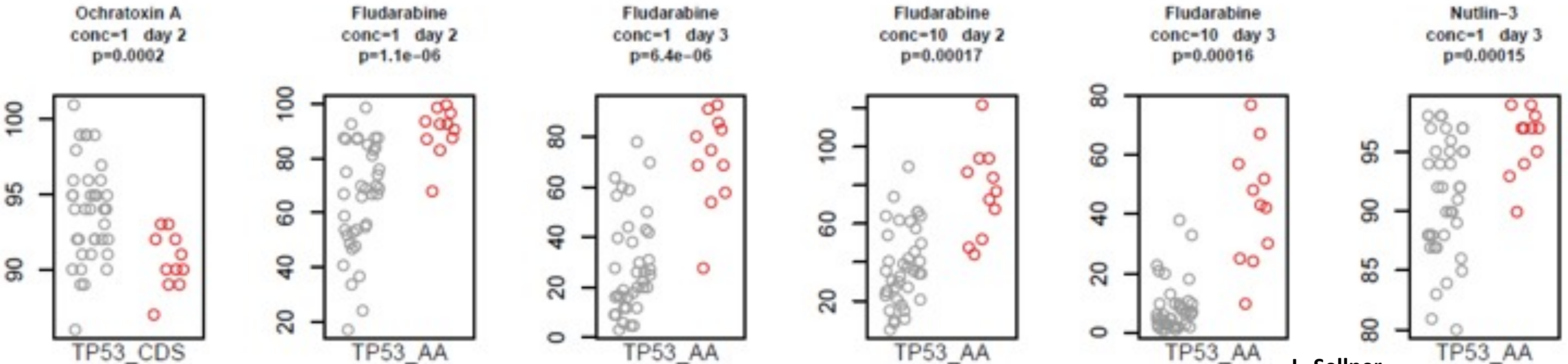
Barretina 2012: The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature.

Basu 2013: An Interactive Resource to Identify Cancer Genetic and Lineage Dependencies Targeted by Small Molecules. Cell.



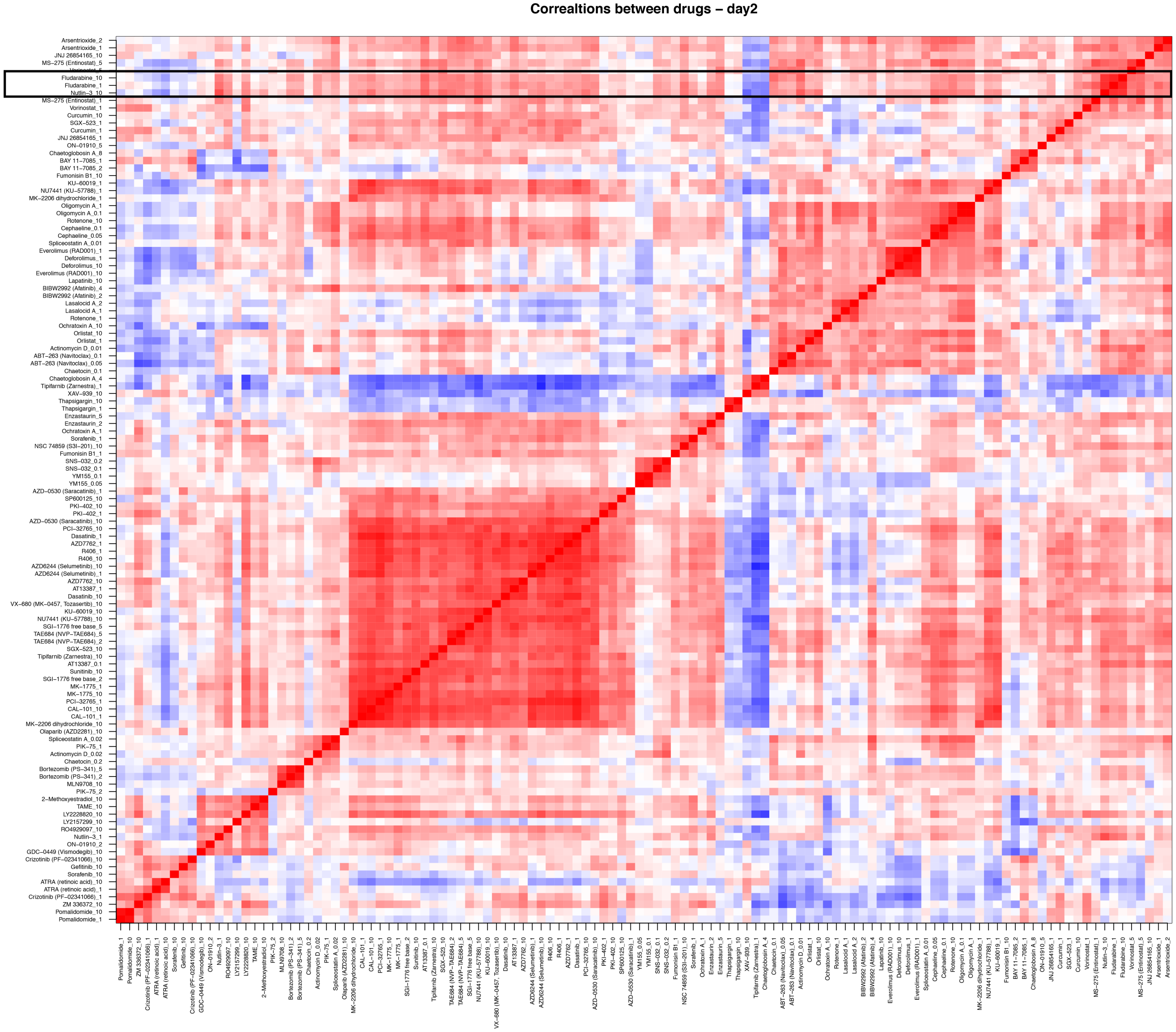
Association of (somatic) variants with drug response



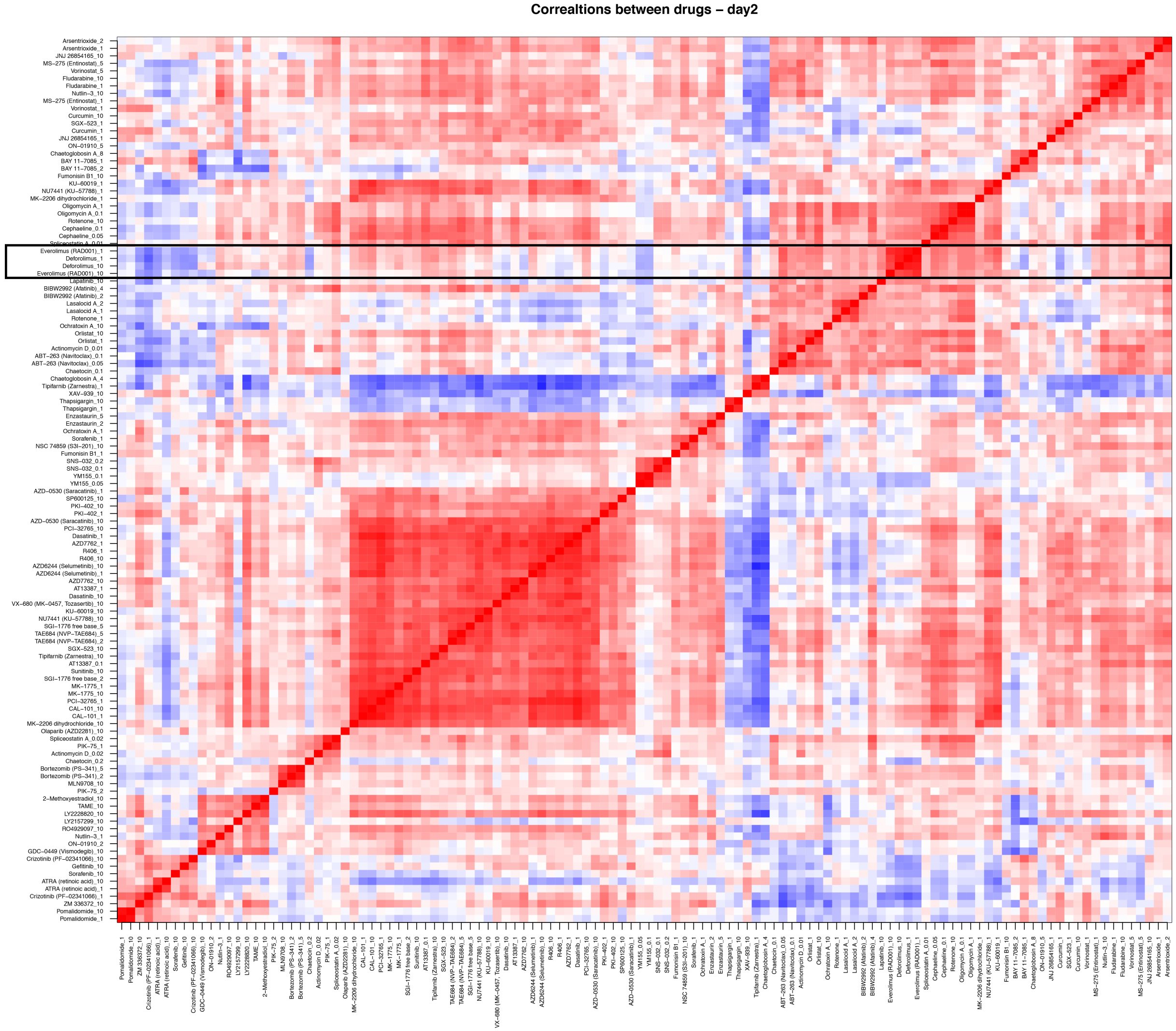


L. Sellner

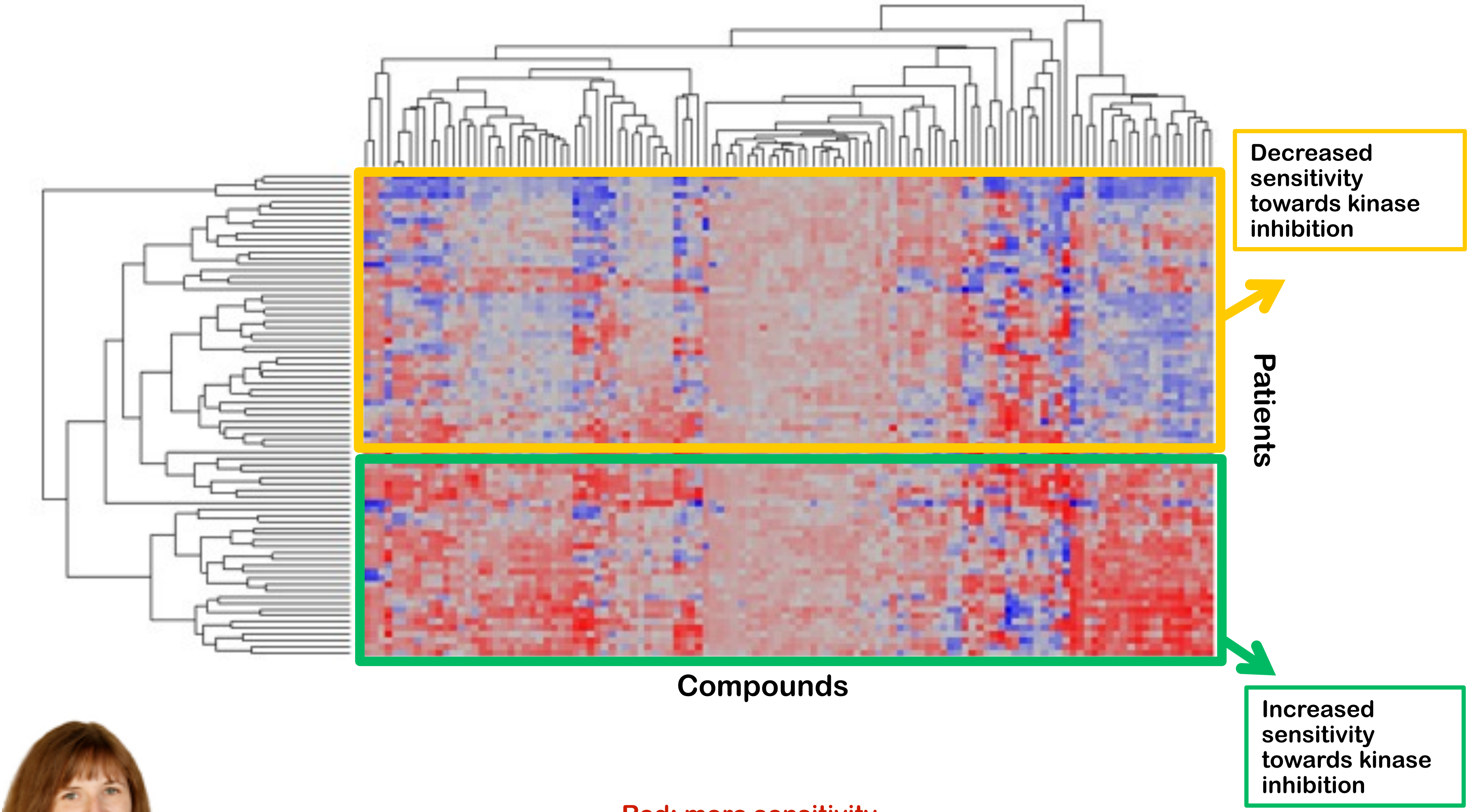
DAY 2 - Fludarabine, Nutlin-3


Correaltions between drugs – day2

DAY 2 - Everolimus (RAD001), Deferolimus



Clustering of patients and drugs according to drug response



Decreased sensitivity towards kinase inhibition

Patients

Compounds

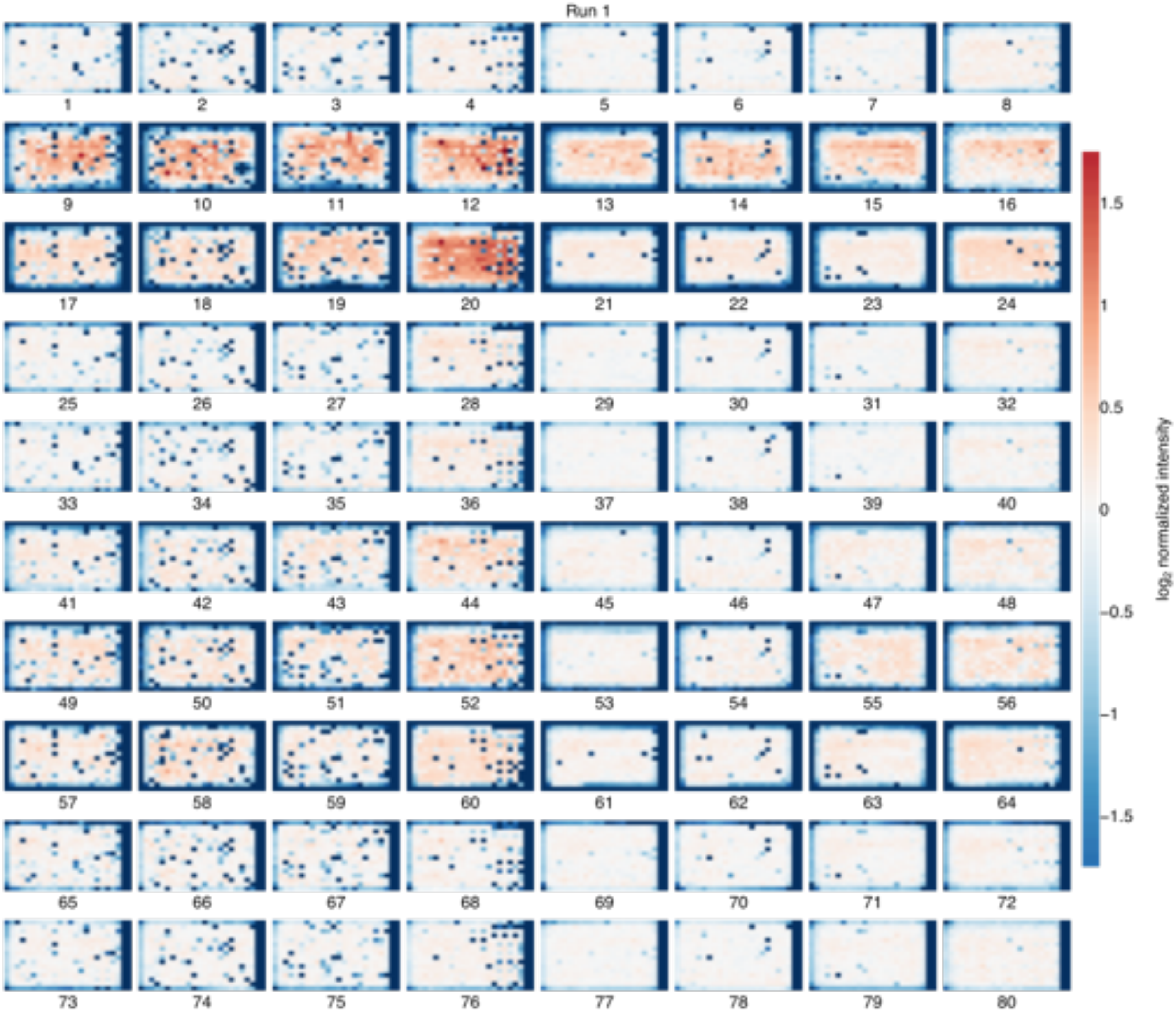
Increased sensitivity towards kinase inhibition

Red: more sensitivity

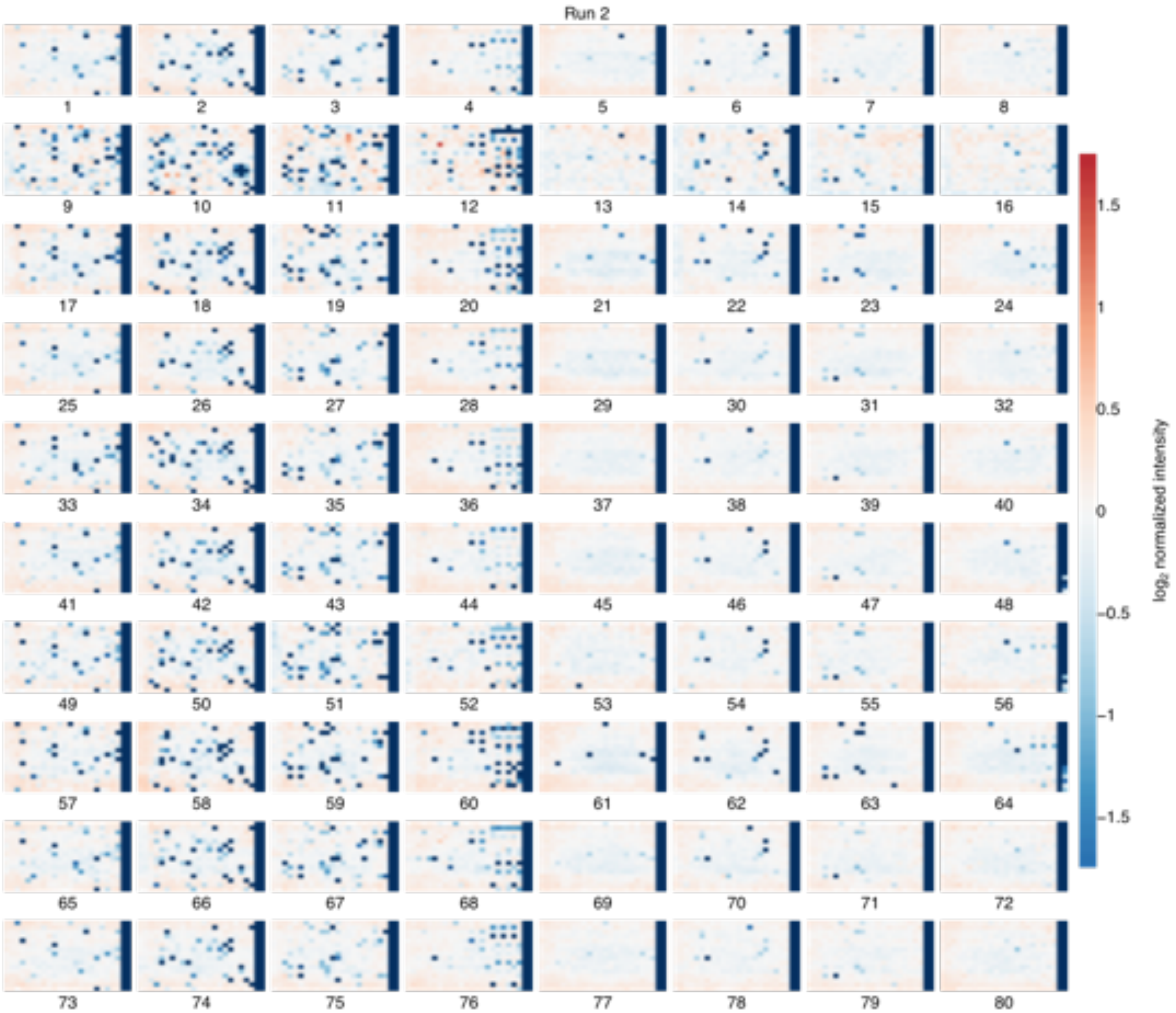
Blue: less sensitivity

M. Oles

CLL – EMBL screen, RUN I



CLL – EMBL screen, RUN II

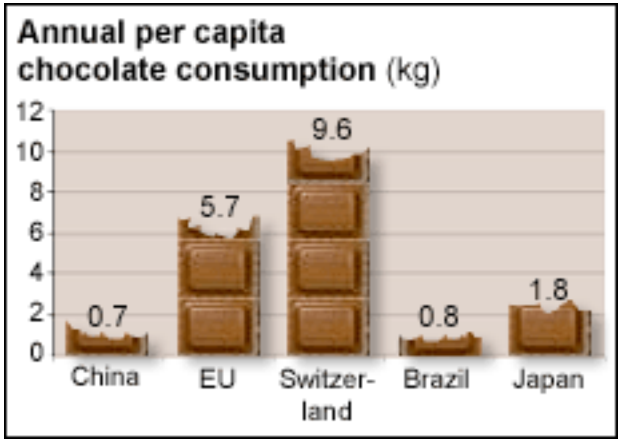


Summary

- Many opportunities for machine learners to make a real impact in biology, ‘precision’ medicine

The future is bright

- 3rd generation sequencing
- single cell everything
- super-resolution microscopy for proteomics
- HT TALEN, CRISPR
- 7 Billion humans to be genotyped, phenotyped (Google-glasses, watches), longitudinal omics
- “Big data”
- Multivariate statistical modelling has only just begun



The Bioconductor Project

Wolfgang Huber

EMBL

Bioconductor

OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Bioconductor

OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

International open source and open development software project for the analysis of genomic data

Objectives:

- Reduce barriers to entry into this interdisciplinary area
- Statistical methods for the analysis of genomic data
- Integrate meta-/other data in the analysis of experimental data
- Publication-quality graphics
- Facilitate reproducible research
- Training

Software: accessible, extensible, interoperable, transparent, well-documented

Approach: rapid development, code re-use, self-documenting datasets

The world’s largest bioinformatics project.

Collaborative software development

- open source
- open development
- interoperability
- code re-use

Code re-use

Writing good software is hard. Existing, well-used and maintained software contains fewer bugs.

Avoid re-implementation, rather produce interfaces

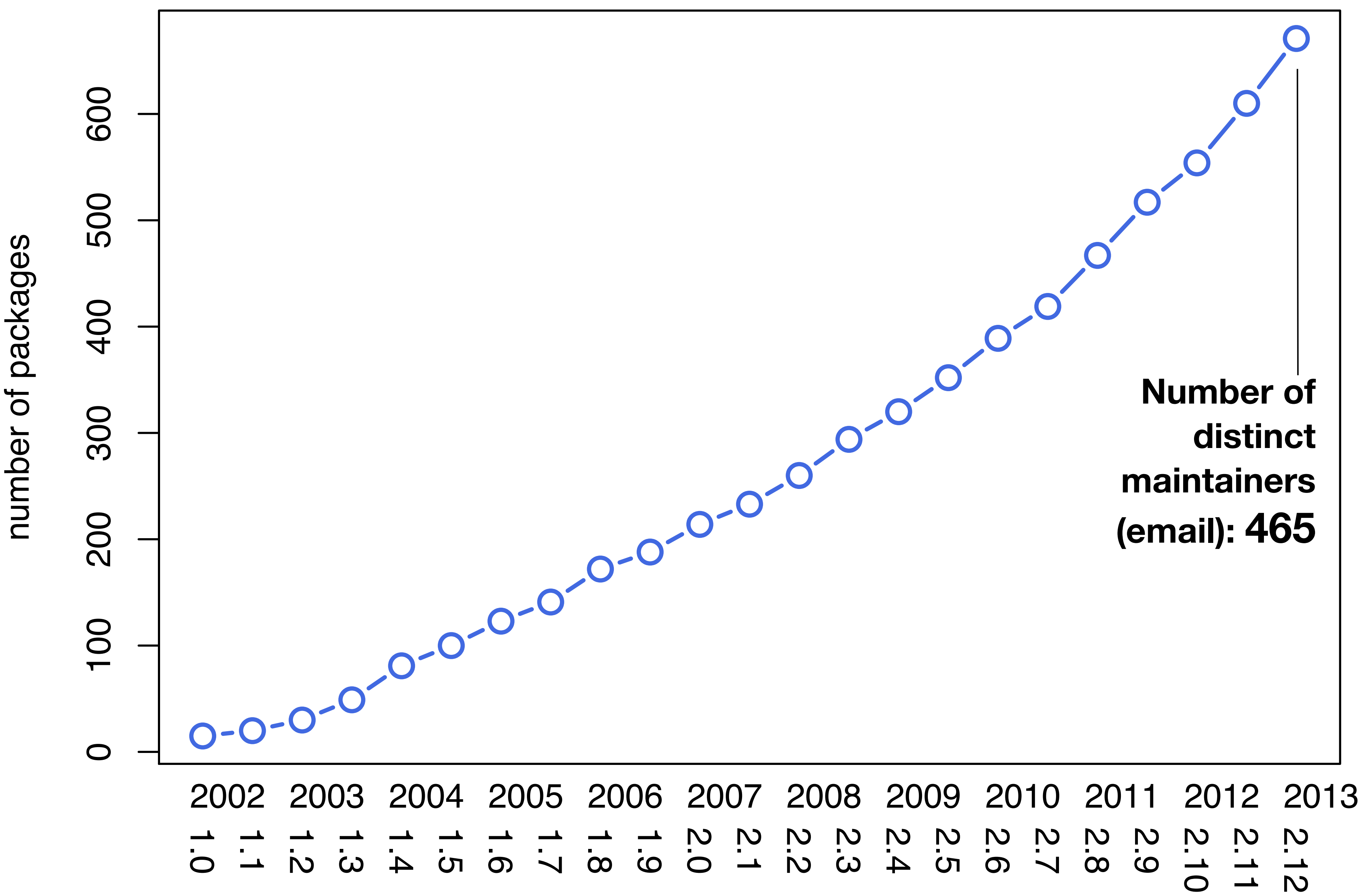
Developers can focus on new things

Software is dynamic and needs continuous maintenance and (re-)publication

Application domains changes (µarrays ... NGS ... 3GS)

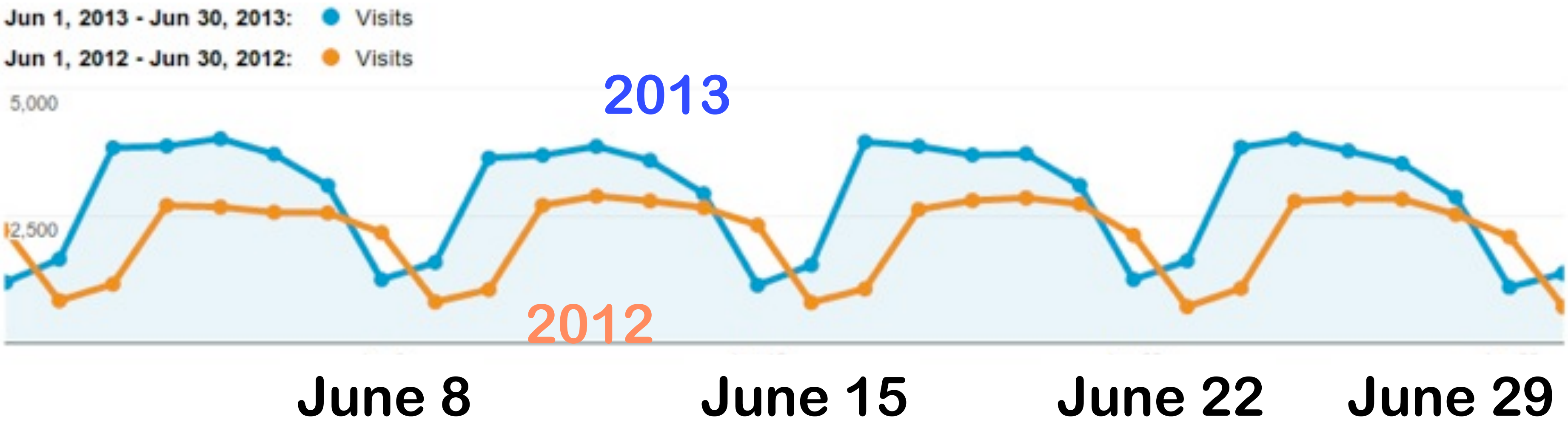
Software technologies change

Contributed Packages



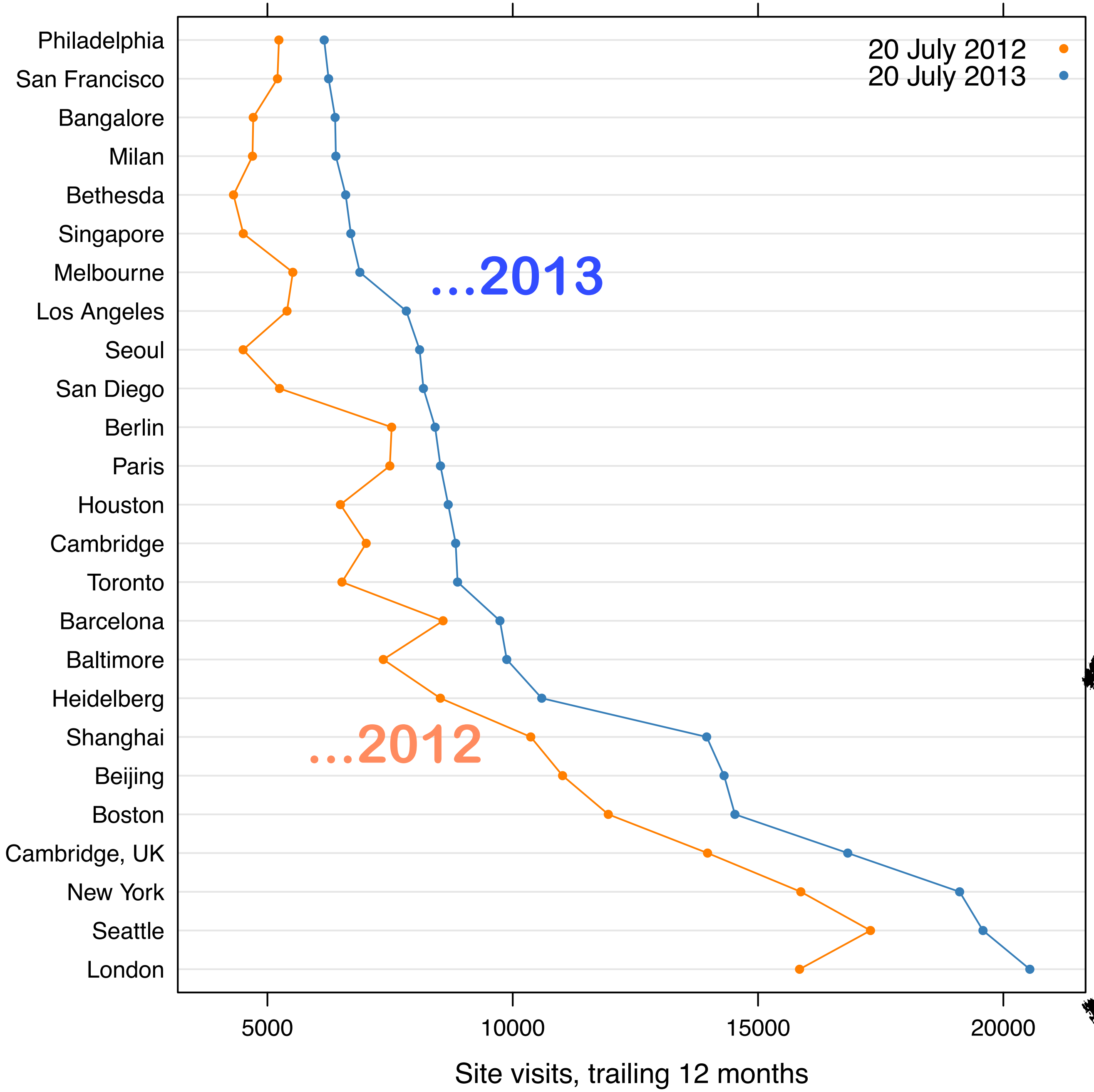
Number of distinct maintainers (email): 465

Site visits - per day



31% growth in the year
up to 6/2013

Site visits - by geographic location



Heidelberg
Shanghai
Beijing
Boston
Cambridge UK
New York
Seattle
London UK

A brief historical context

1970s John Chambers & colleagues develop the S language at Bell Labs - a language for computing with data & visualisation

FSF, GNU, Linux

1991 R. Ihaka and R. Gentleman, two professors at Uni Auckland, build an S interpreter on top of a Scheme interpreter (a Lisp dialect)

1990s: R project gathers a network of collaborators around the world, incl. package system, build server, rigorous ‘R CMD check’

1998 Coming out of the microarray technology (AML/ALL, cell cycle)

2001 Bioconductor project founded at Harvard, RG, VJC, soon R. Irizarry, S. Dudoit (Berkeley), W Huber (Heidelberg)

2002 Sweave, package vignettes

2004 “the book” (published early 2005)

2006... transformation to NGS

Language selection

R - high-level interpreted language, easy & quick prototyping

Packaging protocol

Statistical methods - tests, regression, ML

Visualisation

Parallel computing

Large user community (>> bioinformatics)

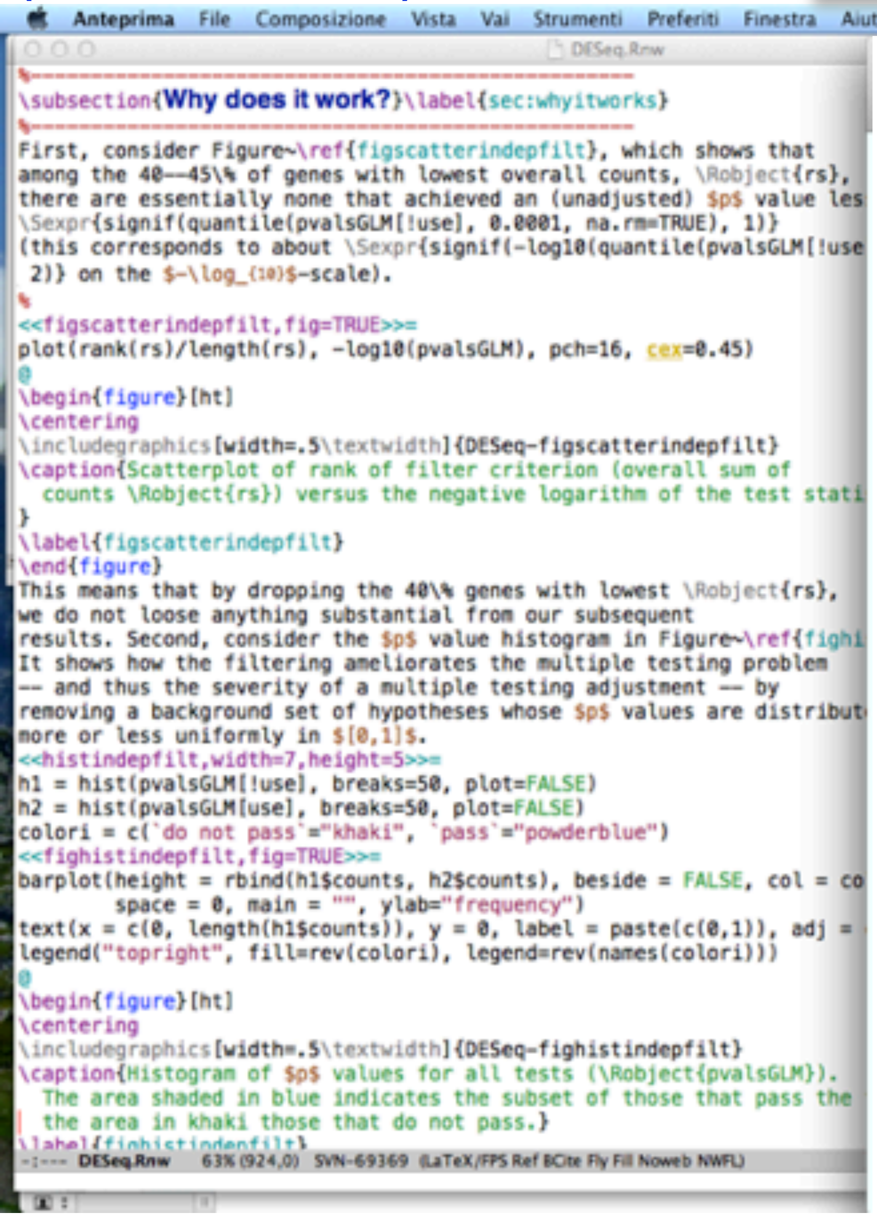
R: programming with data

(cf. Niklaus Wirth: algorithms and data structures = language)

Combined text and code markup (here: LaTeX & R)

Sweave

processed document (here: PDF)


Anteprima File Composizione Vista Vai Strumenti Preferiti Finestra Aiuto
Wed Oct 10 10:54 Wolfgang Huber
DESeq.Rnw

```
%----------------------------------------------------------
\subsection{Why does it work?}\label{sec:whyitworks}
%----------------------------------------------------------
First, consider Figure~\ref{figscatterindepfilt}, which shows that
among the 40--45\% of genes with lowest overall counts, \Robject{rs},
there are essentially none that achieved an (unadjusted) $p$ value les
\Sexpr{signif(quantile(pvalsGLM[!use], 0.0001, na.rm=TRUE), 1)}
(this corresponds to about \Sexpr{signif(-log10(quantile(pvalsGLM[!use
 2)} on the $-\log_{10}$-scale).
%
<<figscatterindepfilt,fig=TRUE>>=
plot(rank(rs)/length(rs), -log10(pvalsGLM), pch=16, cex=0.45)
@
\begin{figure}[ht]
\centering
\includegraphics[width=.5\textwidth]{DESeq-figscatterindepfilt}
\caption{Scatterplot of rank of filter criterion (overall sum of
  counts \Robject{rs}) versus the negative logarithm of the test stati
}
\label{figscatterindepfilt}
\end{figure}
This means that by dropping the 40\% genes with lowest \Robject{rs},
we do not loose anything substantial from our subsequent
results. Second, consider the $p$ value histogram in Figure~\ref{fighi
It shows how the filtering ameliorates the multiple testing problem
-- and thus the severity of a multiple testing adjustment -- by
removing a background set of hypotheses whose $p$ values are distribut
more or less uniformly in $[0,1]$.
<<histindepfilt,width=7,height=5>>=
h1 = hist(pvalsGLM[!use], breaks=50, plot=FALSE)
h2 = hist(pvalsGLM[use], breaks=50, plot=FALSE)
colori = c(`do not pass`="khaki", `pass`="powderblue")
<<fighistindepfilt,fig=TRUE>>=
barplot(height = rbind(h1$counts, h2$counts), beside = FALSE, col = co
        space = 0, main = "", ylab="frequency")
text(x = c(0, length(h1$counts)), y = 0, label = paste(c(0,1)), adj =
legend("topright", fill=rev(colori), legend=rev(names(colori)))
@
\begin{figure}[ht]
\centering
\includegraphics[width=.5\textwidth]{DESeq-fighistindepfilt}
\caption{Histogram of $p$ values for all tests (\Robject{pvalsGLM}).
  The area shaded in blue indicates the subset of those that pass the
  the area in khaki those that do not pass.}
\label{fighistindepfilt}
-:---  DESeq.Rnw    63% (924,0)   SVN-69369  (LaTeX/FPS Ref BCite Fly Fill Noweb NWFL)
```

```
ddsLocal <- estimateDispersions(dds, fitType="local")
plotDispEsts(ddsLocal)
```

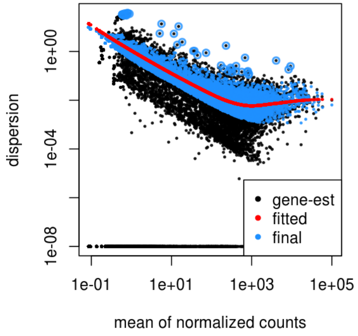


Figure 11: A dispersion estimate plot using a local regression fit is similar to that of Figure 10.

E.2 Mean dispersion

While RNA-Seq data tend to demonstrate a dispersion-mean dependence, this assumption is not appropriate for all assays. An alternative is to use the mean of all gene-wise dispersion estimates to benefit from information sharing across genes (Figure 12).

```
ddsMean <- estimateDispersions(dds, fitType="mean")
plotDispEsts(ddsMean)
```

E.3 Supply a custom dispersion fit

Any fitted values can be provided during dispersion estimation, using the lower-level functions described in the manual page for `estimateDispersionsGeneEst`. In the first line of the code below, the function `estimateDispersionsGeneEst` stores the gene-wise estimates in the metadata column `dispGeneEst`. In the last line, the function `estimateDispersionsMAP`, uses this column and the column `dispFit` to generate maximum *a posteriori* (MAP) estimates of dispersion. The modeling assumption is that the true dispersions are distributed according to a log-normal prior around the fitted values in the column `fitDisp`. The width of this prior is calculated from the data.

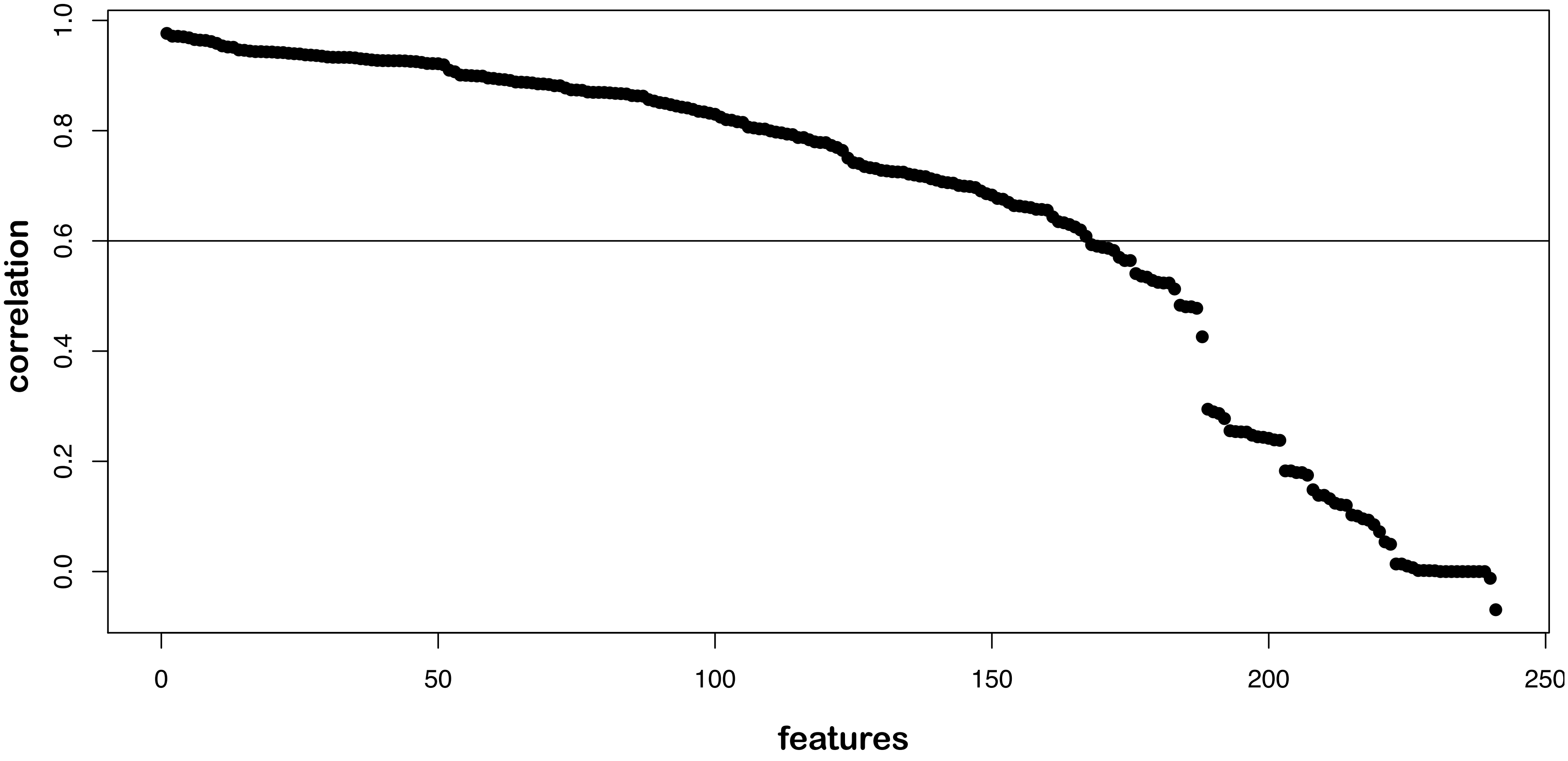
Good scientific software is like a good scientific publication

- Reproducible
- Peer-reviewed
- Easy to access by other researchers & society
- Builds on the work of others
- Others will build their work on top of it

Simon Anders
Joseph Barry
Bernd Fischer
Julian Gehring
Bernd Klaus
Felix Klein
Michael Love
Malgorzata Oles
Aleksandra Pekowska
Paul-Theodor Pyl
Alejandro Reyes
Jan Swedlow
Collaborators

Michael Boutros (DKFZ)
Thorsten Zenz (NCT)
Christof von Kalle (NCT)
Hanno Glimm (NCT)
Lars Steinmetz (EMBL/Stanford)
Robert Gentleman (Genentech)
Martin Morgan (FHCRC)
Jan Korbel (EMBL)

Quality control of features



**Quality criterium:
Correlation of interaction profiles between replicates
and number missing values
162 features passed QC**

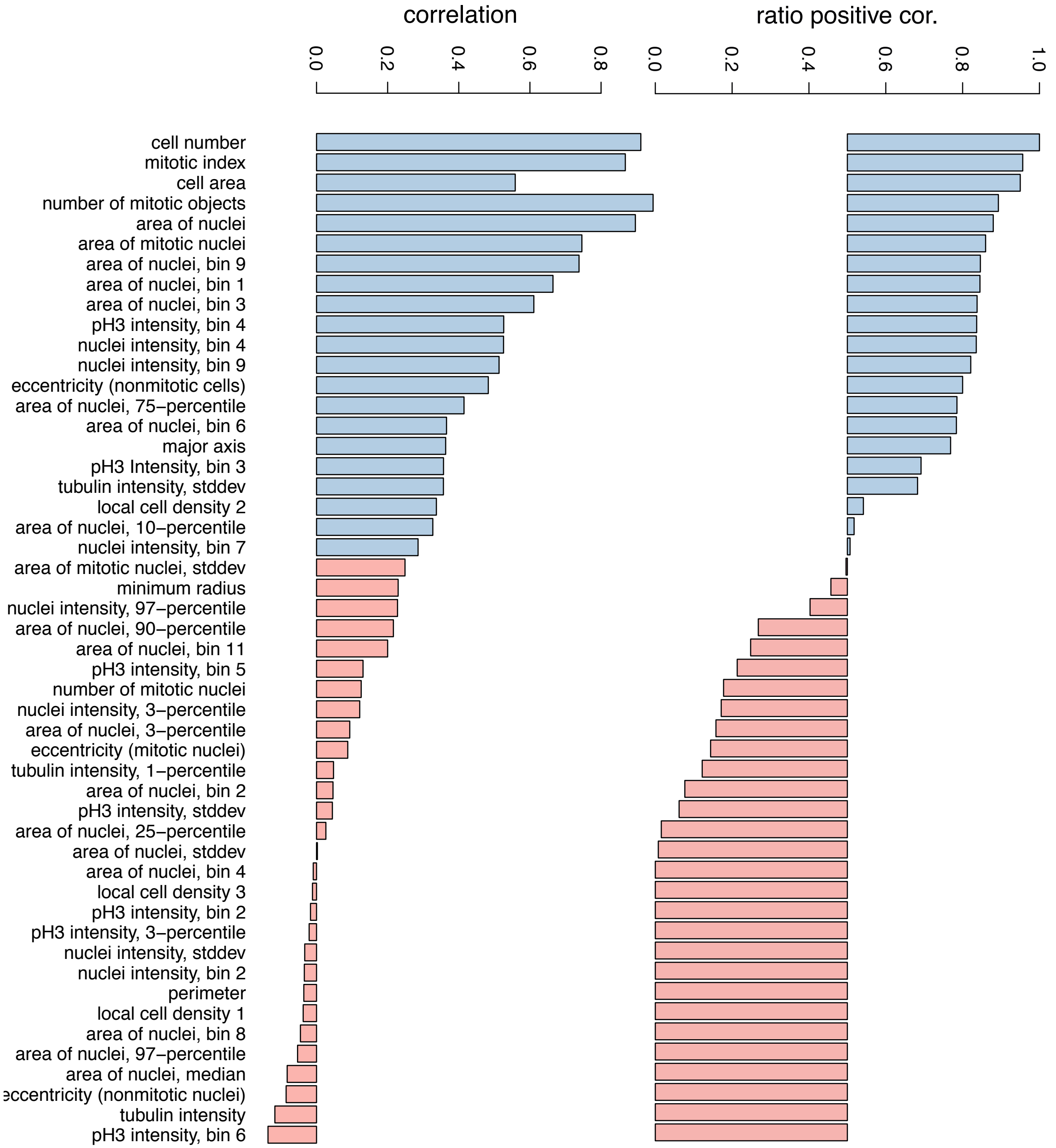
21 non-redundant features

Selection procedure:

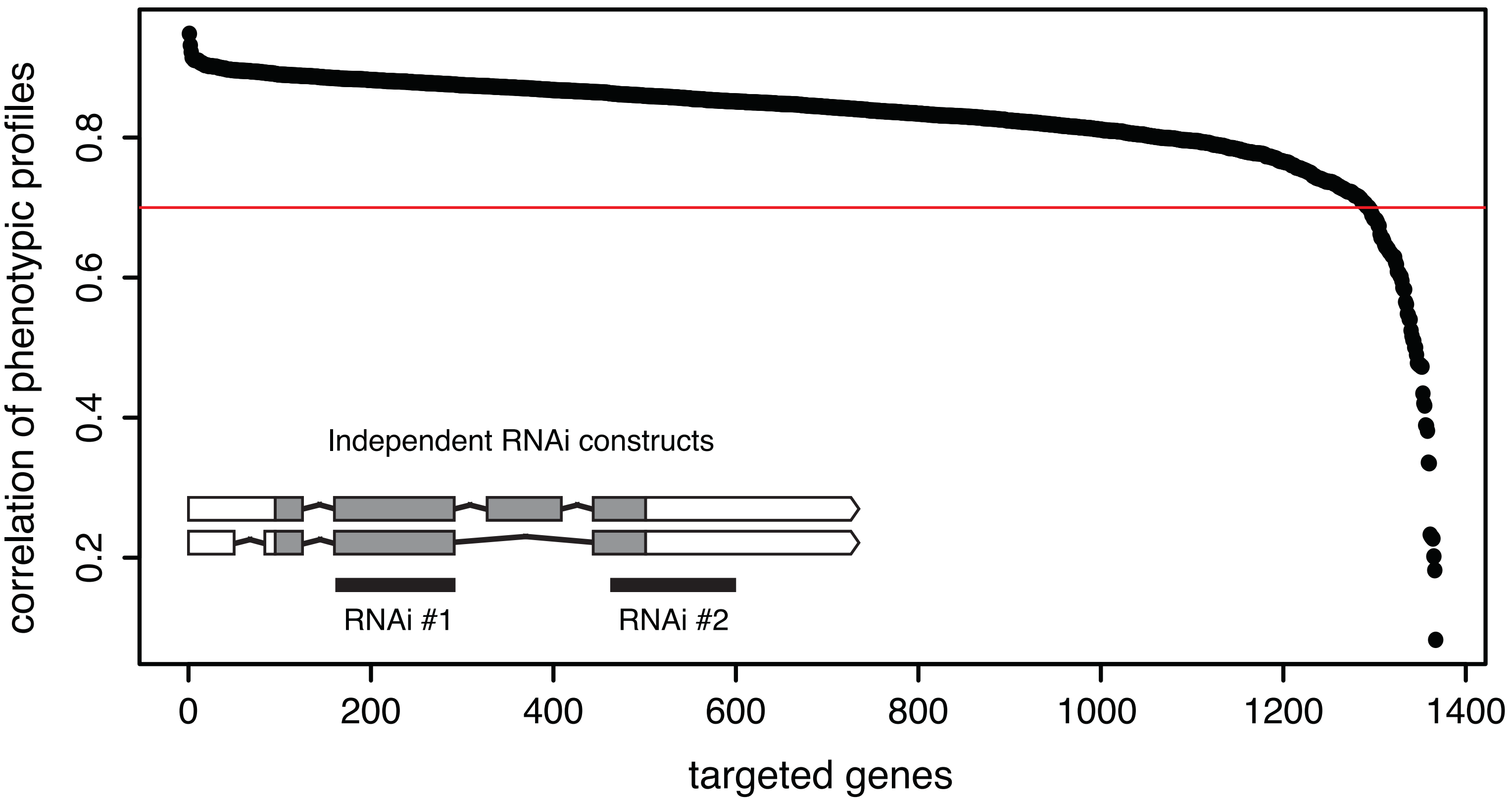
For each feature, determine component not yet spanned by previously selected features

Select the feature with highest S/N

Stop criterion



Quality control of dsRNA designs



possible off-target effects

2 independent dsRNA designs per gene

quality criterion:

cor. of multi-phenotype interaction profile between designs

1293 genes passed QC