Public Tools for Analysis & Systems Microscopy

Bernd Fischer
What makes us different?

- Genome-wide association studies (GWAS) have been enormously successful to identify many new causal players
- ... but have not produced models that can explain / predict complex phenotypes at the level of individuals
- Rare polymorphisms.
- Genetic interactions, epistasis: “system is more than the sum of its parts”.
Definition of Genetic Interactions

Phenotype (cell number)

Interactions

12.5% 25% 50% 75% 100%

-2 aggravating (negative)

0 interaction score ($\pi_{A,B}$)

+1 alleviating (positive)
Imaging

- Cells grow in incubator for some days
- Fixate and stain with DAPI, a -Tubulin, …
- Image with automated fluorescence microscope

Joint Work with Thomas Horn, Thomas Sandmann, Michael Boutros, DKFZ
**Image Processing**

- Segmentation of Nuclei (in DAPI and pH3 channel)

- Propagation of Segmentation from nucleus to cell body (region growing in α-Tubulin channel)

- Extraction of features (intensity, area, texture, …) per cell, summary per well (mean, sd, quantiles, local cell density).

<table>
<thead>
<tr>
<th>Feature</th>
<th>DAPI</th>
<th>pH3</th>
<th>Tubulin</th>
</tr>
</thead>
<tbody>
<tr>
<td>intensity</td>
<td>0.41</td>
<td>0.33</td>
<td>0.0645</td>
</tr>
<tr>
<td>area</td>
<td>52</td>
<td>51</td>
<td>104</td>
</tr>
<tr>
<td>roundness</td>
<td>0.91</td>
<td>0.91</td>
<td>0.72</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
R/Bioconductor Package EBImage

- Read/Write fluorescence images in R
- Segment Images of cell-based assays
  - by adaptive thresholding
    (e.g. for segmentation of well separated nuclei)
  - region growing on a secondary channel
  - (e.g. starting from the nucleus segment extent the region to the cell body)
- Feature extraction
  - cell positions
  - intensity features
  - Shape features (area, outline, ratio of largest to smallest extension, ..)
  - Haralick and Zernicke texture features (is the intensity uniformly distributed or is there a spatial distribution of the intensity values)
- Available as open source from http://www.bioconductor.org
Organization of a high throughput screen
R/Bioconductor package cellHTS2
cellHTS2 screen plots
cellHTS2 control plots
Automated Classification Mitotic Classes

inter
pro
prometa
meta
early ana
ana
telo
R/Bioconductor package imageHTS

- organization of high throughput screen from cellHTS2
- integration of EBImage for image segmentation and feature extraction
- produces webpages for human classification one can click on single cells and assign a class label
- automatic classification of cells for the whole image screen
- downstream statistical analysis
Design of Combinatorial knock-down Screen

- Subset of 93 Drosophila kinases and phosphatases
- each targeted by two independent dsRNA designs
- validation of knock-down by qPCR
- 96 plates (~37,000 wells)
- 4,600 distinct gene pairs

(Horn, Sandmann, Fischer et al., Nat. Methods, 2011)
Screen Plot of Read-out (Number of Cells)

- within screen replicates (cor=0.968)
- independent daRNA designs (cor=0.902)
- between screen replicates (cor=0.948)
Estimating Genetic Interactions

- For many phenotypes, the main effects (single gene) are multiplicative for non interacting genes i, j: \( d_{ij} = \sigma \cdot \mu_i \cdot \mu_j \)

- Additive on logarithmic scale

\[
\log d_{ijk} = w + m_i + m'_j + g_{ij} + \epsilon_{ijk}
\]

- Estimation of main effects (assume that interactions are rare)

\[
(\hat{m}, \hat{m}') = \arg\min_{m, m'} \sum_{i, j} \left\| \log d_{ij} - w - m_i - m_j \right\|_1
\]

- Detect Genetic Interactions: Compare \( \log d_{ijk} \) to \( w - \hat{m}_i - \hat{m}_j \) (t-test)
Clustering of Interaction Map

RasMAPK-pathway

JNK-pathway

10/10/2011
Classification

cross-validation

functional prediction of new genes

cross-validated posterior probabilities of the classifier are shown
R/Bioconductor package RNAinteract

- High Level Analysis for Genetic Interaction Screens
- Based on cellHTS2 screen representation
- Quality control plots (screen plots, …)
- Estimation of single effects
- Estimation of genetic interaction
- Clustering and Classification

- bioconductor data package RNAinteractMAPK
- all data used in (Horn, Sandmann, Fischer et al., Nat. Methods, 2011)
- Sweave file (paper vignette) with complete data processing, analysis, and production of all plots in the paper and its supplement
CellCognition (http://www.cellcognition.org)

Easy to install, usable without programming knowledge
Contains an automatic classification
Automatically detected mitotic events
Feature Representation

- Preprocessing and Data Analysis by Michael Held:
  - Segmentation, tracking, feature extraction, SVM classification

- 150 dim. feature vectors

- SVM classification

- Class probability vector

- Control, Mad2, Bub1
A Hidden Markov Model

Transition probability:
prob. of cell cycle phase at time $t$, given cell cycle phase at time $t-1$

Emission probability:
prob. of SVM predicted cell cycle phase at time $t$, given cell cycle phase at time $t$
Reduced Classification Error

SVM

HMM error correction

color corresponds to predicted mitotic phase

time
Summary

- **EBImage** (http://www.bioconductor.org)
  Oleg Sklyar, Gregoire Pau, Mike Smith, Wolfgang Huber

- **cellHTS2** (http://www.bioconductor.org)
  Ligia Bras, Wolfgang Huber, Michael Boutros, Gregoire Pau, Florian Hahne

- **imageHTS** (http://www.bioconductor.org)
  Gregoire Pau, Xian Zhang, Michael Boutros, Wolfgang Huber

- **RNAinteract** (http://www.bioconductor.org)
  Bernd Fischer, Thomas Horn, Thomas Sandmann, Michael Boutros, Wolfgang Huber

- **cellcognition** (http://www.cellcognition.org)
  Michael Held, Thomas Walter, Bernd Fischer, Jan Ellenberg, Daniel Gerlich

- http://www.mitocheck.org (EU project)
- http://www.systemsmicroscopy.org (EU project)