High Content siRNA Screening Cytoskeletal Reorganization
María Montoya PhD. Cellomics Unit CNIC
Human Genome Project

- 30,000 genes (Many Questions)
  - Gene function?
  - Role in disease
  - Gene regulation
- Systems Approach

Cellular Knowledge Generation in the Post-Genomic Era

- Genomics Building Blocks
- Proteomics Structure
- Cellomics Function

- Cellomics is a neologism meaning the “study of cells” or the “knowledge of cell phenotype and function”
LOSS OF FUNCTION GENETIC SCREENING: LOF (siRNA)

Loss of function genetic screening: Identify gene function through inactivation of a gene or its corresponding mRNA

siRNA collections: screen a small set of genes targeting a specific gene family or components of a molecular pathway; kinase, GTPase, cell-cycle, cytoskeleton...

Genome Wide siRNA libraries; Large scale screens

RNAi: Sequence specific process for silencing gene expression at the post-transcriptional level.

Approaches to RNAi mediated gene knock down in mammalian cells

Mechanism of experimental RNAi
CELLULAR ASSAY (FreedomEVO TECAN)

- siRNA plated on 384 plates
- Dispense cells (Hela)
- Transfect Rab8Q67-GFP DNA
- Fix and Stain: Cell Mask/Dapi/Phalloidin Vinculin

IMAGE ACQUISITION (OPERA PE)

- Ch1: Hoechst
- Ch2: GFP
- Ch3: Actin
- Ch4: Vinculin

IMAGE ANALYSIS (DEFINIENS)

1. Automated object detection
   - Cells
   - Nuclei
   - Border/Inner
   - Spots = FA
   - Polymerized Actin

2. Feature extraction
   - Actin
   - Spots (vinculin)
   - Morphometric

DATA PROCESSING (KNIME)

- Machine learning based phenotype classification algorithms: IB3, NaiveBayes, Decision Trees (j48), SVM
- Quality Control
- Feature selection
- Phenotype classification
  - Class prediction of GFP+ cells

Ctrl

Hit

Rab8
Cellular Assay: Workflow for reverse transfection of synthetic siRNAs on multiwell plate format

1. Synthetic siRNA source plates
2. Transfer siRNA + transfection reagent + substrates to assay plates
3. Dry assay plates (Speed-vac)
4. Plate cells on top of dried assay plates
5. Incubate 24-96h To allow transfection
6. Plate processing: Fix/Perm/Stain
7. Image analysis (Definiens)
8. Data Mining (KNIME)
9. High content Imaging
siRNA plated on 384 plates

Dispense cells (Hela)

Transfect Rab8Q67-GFP DNA

Fix and Stain:
- Cell Mask
- Dapi
- Phalloidin
- Vinculin

1. Automated object detection
   - Cells
     - Nuclei
     - Border/Inner
     - Spots=FA
     - Polymerized Actin

2. Feature extraction

Ch1: Hoechst
Ch2: GFP
Ch3: Actin
Ch4: Vinculin
Initial Background Segmentation
- Normalized Darkness
Nuclei Segmentation

- Normalized Darkness
- MinMax filter. Multiresol Segmentation

Image Analysis (Définien):

- Upper Level
- Cell/Backgrnd
- New Level
- Nuclei

Nuclei Segmentation Final
Cell Segmentation

- Pix Based Obj Resizing (grow nucl seeds)
Cell Segmentation

- Pix Based Obj Resizing (grow nucl seeds)
Cell Segmentation

- Pix Based Obj Resizing (grow nucl seeds)
Cell Segmentation

- Pix Based Obj Resizing (grow nucl seeds)
Cell Segmentation

- Pix Based Obj Resizing (grow nucl seeds)
Cell Segmentation

* Pix Based Obj Resizing (grow nucl seeds)
Cell Segmentation

- Pix Based Obj Resizing (grow nucl seeds)
Cell Segmentation

- Pix Based Obj Resizing (grow nucl seeds)
Cell Segmentation

- Pix Based Obj Resizing (grow nucl seeds)

Refine GFP+ Cell borders
Cell Segmentation

- Pix Based Obj Resizing (grow nucl seeds)
Find highly contrasted spots
Segmentation Cell Border/Inner

- Distance Map

- Upper Level
- UpNew Level
- New Level
- Lower Level

Cell/Backgrnd

Spots

Nuclei

Border/Inner

Sementation (Border/Inner)
Segmentation High Intensity Actin

Upper Level

UpNew Level

New Level

Lower Level

Level 0

Cell/Backgrnd

Spots

Nuclei

Border/Inner

High Intensity Polymerized Actin
Segmentation Fibrillar Actin

- **Upper Level**: Cell/Backgrnd
- **UpNew Level**: Spots
- **New Level**: Nuclei
- **Lower Level**: Border/Inner
- **Level 0**: HighIntActin
- **Level sub0**: FibrillarActin

**Fibrillar polymerized actin**
Cell classification and final QC

- GFP+
- Cell
- Poor spreading cell

Analysis

\{ \text{GFP- cells} \}

\{ \text{GFP+ cells} \}

Round Cells

Cells at Image Border
1. Automated object detection
   - Cells
     - Nuclei
     - Border/Inner
       - Spots = FA
       - Polymerized Actin

2. Feature extraction
   - Actin
   - Spots (vinculin)
   - Morphometric
DEVELOPMENT OF IMAGE ANALYSIS TOOL FOR AUTOMATIC CLASSIFICATION OF CYTOSKELETAL REARRANGEMENT PHENOTYPES: IMAGE QUANTITATIVE ANALYSIS

- Actin (51)
- Feature Extraction (79)
- Spot (21...)
- Morphom (7)

- Area (Border/Inner)
- Elliptic Fit
- Spreading evaluation
- Sphericity
DEVELOPMENT OF IMAGE ANALYSIS TOOL FOR AUTOMATIC CLASSIFICATION OF CYTOSKELETAL REARRANGEMENT PHENOTYPES: IMAGE QUANTITATIVE ANALYSIS

- **Actin** (51)
- **Feature Extraction** (79)
- **Spot** (21...)
- **Morphom** (7)

**Segmentation**
FA = Spots

- Distances FA To cell Border (6)
- Areas Localization Border/Inner (15)

**Vinculin texture analysis**
Shape texture based on subobjects

- Area
- Density
- Assimetry
- Direction
Definiens image analysis

2. Feature extraction cell based

- Actin (51)
- Spot (21...)
- Morphom (7)

Feature Extraction (79)

Segmentation polymerized Actin

- Area High Intensity 70/85Q
- Polymerized fibrillar structures
- Relative Areas Border/Inner (14)
- Relative Areas Border/Inner (10)

Actin texture analysis

- S.Dev.
- Diff. I (75Q-25Q) (6)

Edge extraction algorithms

- Mean
- S. Dev. (14)

- Mean
- S. Dev. (7)
cells

Transfect siRNA

24h

Transfect Rab8Q67-GFP DNA

24h

Fix and Stain:
Cell Mask/Dapi/Phalloidin Vinculin

Segmented images

Actin Intensity
normalization

Definiens

Analyze
79 Features

Definiens

.csv file/well

.arff file/image

R/Excel

Extract
biologically meaningful information

Phenotype classification

R

WEKA

test classifiers

train classifiers

Ch1:cell/nuclei
Ch2:nuclei
Ch3:actin
Ch4: FA
Ch5:GFP

Ch1:cell/nuclei
Ch2:nuclei
Ch3:actin
Ch4: FA
Ch5:GFP

LOST IN TRANSLATION

Segmented images

Actin Intensity normalization

Definiens

Analyze 79 Features

Definiens

.csv file/well

.arff file/image

R/Excel

Extract biologically meaningful information

Phenotype classification

R

WEKA

test classifiers

train classifiers

Ch1:cell/nuclei
Ch2:nuclei
Ch3:actin
Ch4: FA
Ch5:GFP

LOST IN TRANSLATION

Segmented images

Actin Intensity normalization

Definiens

Analyze 79 Features

Definiens

.csv file/well

.arff file/image

R/Excel

Extract biologically meaningful information

Phenotype classification

R

WEKA

test classifiers

train classifiers

Ch1:cell/nuclei
Ch2:nuclei
Ch3:actin
Ch4: FA
Ch5:GFP
Machine learning based phenotype classification algorithms: IB3, NaiveBayes, Decision Trees (j48), SVM

Class prediction of GFP+ cells

Quality Control

Feature selection

Phenotype classification

Ctrl

Rab8

Hit
Several rules were implemented as a fuzzy score based on biological knowledge of the subject. Some of these rules are:

- **Spreading Evaluation**
  - Number of pixels within a given range
  - Relative nuclear area
  - Shape
  - ...

- **Confluency**
  - Number of Neighbour cells
  - Relative Border to Background
  - Number of pixels

- **Focusness**
  - Staining for Actin
  - Staining for Vinculin

Each cell has a score 0-1 which is a classification value based on for each of the features evaluated...
DATA QUALITY CONTROL (KNIME) BASED ON DEFINIENS FUZZY SCORES
LOST IN TRANSLATION

Machine learning based phenotype classification algorithms: IB3, NaiveBayes, Decision Trees (j48), SVM

Class prediction of GFP+ cells

Quality Control

Feature selection

Phenotype classification

Control (Ctrl)

Rab8

Hit
Supervised learning

• **Training set**: Paired nº of GFP+ and GFP- cells from a set of images from the same experiment
  - GFP+ cell (300+)
  - GFP- cell (300+)

• **Feature vector**:
  - Cell attributes (Features)
  - Sets of Features
    - ALL (79)
    - Actin (51)
    - Spots (21)
    - Morphometric (7)

• **Classification algorithm**
  - Decision tree (J48)
  - Naive Bayes
  - K Nearest Neighbors (IB3/IB7)
  - Neural Networks NN
  - Support vector machines

• **Output data: confusion matrix**
  - Sensitivity (How well GFP+ cells are classified)
  - Specificity (How well control cells are classified)
  - Accuracy (proportion of total number of predictions correct, GFP + and -)
DATA PROCESSING (KNIME): SUPERVISED CLASSIFICATION FOR FEATURE SELECTION

K medoids and Hierarchical clustering

Distance Matrix generation

Data gathering

Feature selection filtering

Training and classification

Ambiguous elements elimination

Model analysis by hierarchical clustering (under construction)
### IB7

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>SPOT</th>
<th>MORF</th>
<th>ACTIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>80.51%</td>
<td>84.41%</td>
<td>61.04%</td>
<td>81.81%</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>75.55%</td>
<td>100.00%</td>
<td>62.50%</td>
<td>84.21%</td>
</tr>
<tr>
<td>Specificity</td>
<td>82.35%</td>
<td>75.51%</td>
<td>59.45%</td>
<td>79.48%</td>
</tr>
</tbody>
</table>
DEVELOPMENT OF IMAGE ANALYSIS TOOL FOR AUTOMATIC CLASSIFICATION OF CYTOSKELETAL REARRANGEMENT PHENOTYPES: IMAGE QUANTITATIVE ANALYSIS

- Feature Extraction (79)
- Spot (21…)
- Morphom (7)
- Actin texture analysis
- Segmentation polymerized Actin

- Area High Intensity 70/85Q
- Polymerized fibrillar structures
- Relative Areas Border/Inner (14)
- Relative Areas Border/Inner (10)
- S.Dev. (14)
- S. Dev. (7)
- Mean (14)
- Mean (7)

Edge extraction algorithms:
Elimination of redundant features based in evaluation of classification performance

- Actin segmentation based in areas higher than 70Q vs 85Q intensity
- Texture of actin based in Lcanny vs Arithmetics
- Spot localization at Border/Inner
Feature elimination (backward) output

**CE**: Classification Error. Based in the mean error that a feature introduces to a series of classification routines (Naive Bayes/J48 Tree/Neural networks/...).
Feature filtering

- **All (79)**
- **NR: NonRedundant (53):** All except redundant features.
- **CE (57):** Classification Error. Based in the mean error that a feature introduces to a series of classification routines (Naive Bayes/J48 Tree/Neural networks/)
  - CE: Features with classification error < 0
- **CEFL (57):** Feature Logic. Re arrange CE list by removing sets of features of same root even with low CE and viceversa
- **CEFLPh (48):** Phenotype. Rearrange CEFL list removing those with output is not in accodance with expected phenotype
- **CEFLPhNR (31):** Phenotype. Rearrange CEFLPh list removing redundant features
- **Bal (24):** Balanced: remove actin features o balance with the number of features of FA.
Summary

1. Automated object detection
   - Cells
     - Nuclei
     - Border/Inner
     - Spots = FA
     - Polymerized Actin

2. Feature extraction
   - Actin
   - Spots (vinculin)
   - Morphometric

Feature filtering

- KNIME as a data exploration platform for screening development
- KNIME as a tool for experimentalists to take a look at your (HCS) data
- You don't need to be “the analyst” for taking advantage of KNIME capabilities

Machine learning based phenotype classification algorithms: IB3, NaiveBayes, Decision Trees (j48), SVM

Quality Control
Feature selection
Phenotype classification
Class prediction of GFP+ cells

Rab8

Ctrl
Hit