Cell-Based Microarrays: Current Status, Future Prospects

Tom Freeman
Scottish Centre of Genome Technology and Informatics (GTI),
University of Edinburgh Medical School
Reverse Transfection Methodology

- full-length cDNA clone expression library
- plasmid purification
- gelatin, transfection reagent
- array cDNA onto glass slide
- grow cells over array until confluent
- cells of interest (HEK293T)
- visualise sub-cellular localisation and/or transfection-dependent changes in cell phenotype
Applications of Reverse Transfection Gain of Function Assays

• high throughput screen for protein sub-cellular localisation

• protein interaction studies

• in vitro screen for genes that induce desired effect on cell phenotype

Our interest:
• potential for use in identifying candidate genes for use in gene therapy - search for pro-apoptotic genes for treatment of disease - collaboration with Prof. Andrew Miller (Genetic Therapies Centre, Imperial College, London)
Technology Set Up and Investigation into Gene Tagging Strategies

• cDNA clones were selected from the full-length MGC clone collection (IRAT) by searching for genes with interesting functional activity e.g. factors, receptors, kinases

• search refined using databases (EMBL, Locuslink) to pull out sequence and description of function

• 20 genes chosen due to differences in protein size and known function

• open reading frame (ORF) transferred into Gateway (Invitrogen) cloning system

• ORF tagged at both N- and C-terminal with GFP
16 of the 20 genes were successfully transferred into Gateway vectors

The genes could be divided into 3 classes:

1. C- and N-terminal sub-cellular localisations agreed (6/16)
2. C- and N-terminal sub-cellular localisations did not agree (5/16)
3. no fusion protein observed with N-terminal construct (5/16)
GROUP 1: N and C-terminal same localisation

- **ATF4**
  - C-terminal tag: Nucleus
  - N-terminal tag: Nucleus
  - C-terminal tag: Punctate cytoplasm/perinuclear cluster
  - N-terminal tag: Punctate cytoplasm/perinuclear cluster

- **CALM2**
  - C-terminal tag: Nucleus/Cytoplasm
  - N-terminal tag: Nucleus/Cytoplasm
  - C-terminal tag: Punctate nucleus/Cytoplasm
  - N-terminal tag: Punctate nucleus/Cytoplasm

- **CDK7**
  - C-terminal tag: Nucleus/Cytoplasm
  - N-terminal tag: Nucleus/Cytoplasm
  - C-terminal tag: Punctate nucleus/Cytoplasm
  - N-terminal tag: Punctate nucleus/Cytoplasm

- **CDK9**
  - C-terminal tag: Nucleus
  - N-terminal tag: Nucleus
  - C-terminal tag: Punctate cytoplasm/perinuclear cluster
  - N-terminal tag: Punctate cytoplasm/perinuclear cluster

- **IL17BR**
  - C-terminal tag: ER
  - N-terminal tag: ER
  - C-terminal tag: Nucleus (no nucleolous)
  - N-terminal tag: Nucleus (no nucleolous)

- **NFIB**
  - C-terminal tag: Nucleus
  - N-terminal tag: Nucleus
  - C-terminal tag: Punctate cytoplasm/perinuclear cluster
  - N-terminal tag: Punctate cytoplasm/perinuclear cluster

Bar = 5 µm
GROUP 2: N and C-terminal different localisations

PTPN11
- C-terminal tag: Punctate cytoplasm
- N-terminal tag: Nucleus/Cytoplasm

TNFR
- C-terminal tag: Tubules/vesicles
- N-terminal tag: Nucleus

SF10B
- C-terminal tag: Nuclear membrane/Cytoplasm
- N-terminal tag: Nucleus

TGIF
- C-terminal tag: Nucleus (no nucleolus)/punctate cytoplasm
- N-terminal tag: Nucleus/Cytoplasm

MARKL1
- C-terminal tag: Mitochondria
- N-terminal tag: Nuclear aggregates

SIP2-28
- C-terminal tag: Nuclear membrane/Cytoplasm
- N-terminal tag: Nucleus/Cytoplasm

Bar = 5 μm
GROUP 3: C-terminal localisations only

- CXADR
- NFIL3
- PPARG
- STK15
- CDKN1B

Bar = 5 μm
Conclusions

- technique works!
- in all cases sub-cellular localisation of C-terminal GFP construct correct as compared to that reported/predicted
- localisations same as those observed with well-based assays
- GFP provides good positive control for transfection and can be visualised in living cells
- other tags e.g. His-tag require post-processing of the arrays to allow visualisation of transfected protein which can damage the cell monolayer

Construction of High Density Reverse Transfection Array

Why?
• not much point otherwise
• limited reagents

Issues
• transfer of ORF’s from MGC clones into Gateway system time consuming and expensive for large numbers of genes
• transfer may introduce errors in ORF
• gene-tags can effect localisation (therefore function)

Strategy
• use MGC clones (IRAT) direct for construction of high density array
2,796 MGC IRAT clones used for plasmid prep.

High Density Gain of Function Screen

Functional assays
- phospho-tyrosine (Ab)
- apoptosis inducers (TUNEL)

Verification and functional characterisation

Reverse transfection
- grids outlined by GFP
- arrayed in quadruplet on each slide, each sub-grid being surrounded by a GFP-only expression vector
- plasmids arrayed onto slide
- array comprised of 10,080 features, 1,959 clones (post-array scan)

<table>
<thead>
<tr>
<th>GENE</th>
<th>12 hour (TUNEL)</th>
<th>12 hour (CASP3)</th>
<th>24 hour (TUNEL)</th>
<th>24 hour (CASP3)</th>
<th>36 hour (TUNEL)</th>
<th>36 hour (CASP3)</th>
<th>48 hour (TUNEL)</th>
<th>48 hour (CASP3)</th>
<th>60 hour (TUNEL)</th>
<th>60 hour (CASP3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSTB</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>30</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>MGC5439</td>
<td>0</td>
<td>20</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>C22orf23</td>
<td>30</td>
<td>10</td>
<td>30</td>
<td>10</td>
<td>60</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>AF1Q</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>CCBP2</td>
<td>N/A</td>
<td>10</td>
<td>1</td>
<td>20</td>
<td>1</td>
<td>20</td>
<td>1</td>
<td>30</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>LOC134285</td>
<td>N/A</td>
<td>20</td>
<td>5</td>
<td>20</td>
<td>5</td>
<td>20</td>
<td>5</td>
<td>25</td>
<td>30</td>
<td>65</td>
</tr>
<tr>
<td>EXOC7</td>
<td>N/A</td>
<td>20</td>
<td>1</td>
<td>30</td>
<td>1</td>
<td>40</td>
<td>1</td>
<td>40</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>STK3</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>50</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>ACO1</td>
<td>40</td>
<td>10</td>
<td>40</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>40</td>
<td>30</td>
<td>70</td>
<td>50</td>
</tr>
<tr>
<td>XBP1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>10</td>
<td>5</td>
<td>60</td>
<td>10</td>
<td>60</td>
<td>70</td>
</tr>
<tr>
<td>STS</td>
<td>N/A</td>
<td>20</td>
<td>N/A</td>
<td>20</td>
<td>N/A</td>
<td>30</td>
<td>N/A</td>
<td>50</td>
<td>N/A</td>
<td>60</td>
</tr>
<tr>
<td>Mock transfection</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

10/79 gene ‘hits’ verified as inducing apoptosis – 7 classified (annotated gene name), 3 unclassified

- 1 gene known to be associated with apoptotic pathway
- 1 over-expression gives rise to apoptosis through indirect route
- 2 considered interesting by our collaborators on apoptosis
- 6 no previous association to apoptosis
Cell-based Microarrays

- RNAi library
- cDNA library
- Compound library
- Microtitre plate
- Transfection reagent
- Microarray
- Array onto microscope slide
- Cells of interest
- Visualisation
- Phenotype assay
- Whole array imaging
- Cell culture

Graphical representation of the process involving cell-based microarrays, including the steps from library preparation to visualisation and imaging.
Cell-based ‘arrays’

- RNAi library
- cDNA library
- compound library
- microtitre plate
- array
- Store
- transfection reagent
- cell culture
- phenotype assay
- plate assay
- cell imaging
Final Comments on Cell-Based Arrays

- reverse transfection microarray technology is a challenge to set up and in our hands proved to be a temperamental technique – many variables
- requires significant resources to in terms of access to high quality reagent resources, printing and scanning technology
- many aspects simplified by use of plate-based assays
- need robust phenotype assays and good verification/functional characterisation pipeline
- Having said this, the technology can provide high throughput functional screen properly resourced
Found Genes but by What Mechanism Do They Induce Apoptosis?

<table>
<thead>
<tr>
<th>GENE</th>
<th>Percentage of cells undergoing apoptosis</th>
<th>12 hour</th>
<th>24 hour</th>
<th>36 hour</th>
<th>48 hour</th>
<th>60 hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSTB</td>
<td>TUNEL CASP3 TUNEL CASP3 TUNEL CASP3 TUNEL CASP3</td>
<td>5 0 5 30</td>
<td>10 10 10 10 10</td>
<td>10 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGC5439</td>
<td>0 20 10 20</td>
<td>10 20 10 40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C22orf23</td>
<td>30 10 30 10</td>
<td>60 10 40 40 40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF1Q</td>
<td>10 10 10 10</td>
<td>10 10 10 10 60 40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCBP2</td>
<td>N/A 10 1 20</td>
<td>1 20 1 30 10 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC134285</td>
<td>N/A 20 5 20</td>
<td>5 20 5 25 30 65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXOC7</td>
<td>N/A 20 1 30</td>
<td>1 40 60 50 60 60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STK3</td>
<td>0 20 0 20</td>
<td>10 20 10 50 40 60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACO1</td>
<td>40 10 40 10</td>
<td>40 20 40 30 70 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XBPI</td>
<td>1 2 1 2</td>
<td>10 5 60 10 60 70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STS</td>
<td>N/A 20 N/A 20</td>
<td>N/A 30 N/A 50 N/A 60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mock</td>
<td>2 2 2 2</td>
<td>2 2 3 2 4 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• 3 genes selected, STS and mock transfection samples then subjected to expression analysis on duplicate Affymetrix U133 2plus arrays at 12, 24 and 48 h (30 chips total)
Plots of Affymetrix Expression data of 3 Transfected Genes Across All Conditions
Up and Down Regulation of Gene Expression in Experimental Conditions vs. Time Matched Mock Transfection

1. gene P/A based on MAS5 and normalised by RMA (Bioconductor)
2. gene must be scored P in both replicate chips in at least one of 15 conditions
3. pairwise comparisons performed between each test condition and time-matched control (gene must be P in 1 of conditions)
4. 1-sided ANOVA performed with false discovery rate of 0.05 and error model estimates. List further filtered on fold change (>1.4)

<table>
<thead>
<tr>
<th></th>
<th>Hour</th>
<th>No. genes</th>
<th>% I+D</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACO1</td>
<td>12</td>
<td>53</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>90</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>63</td>
<td>22</td>
</tr>
<tr>
<td>STK3</td>
<td>12</td>
<td>73</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>76</td>
<td>21</td>
</tr>
<tr>
<td>XBP1</td>
<td>12</td>
<td>163</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>185</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>222</td>
<td>78</td>
</tr>
<tr>
<td>STS</td>
<td>12</td>
<td>320</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>366</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>282</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>149</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>40</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>82</td>
<td>46</td>
</tr>
</tbody>
</table>
Gene Tree of Differentials
Logic Mapping of Cellular Pathways

Use of standardised notion to depict biological entities and interactions, and standard layout rules

Drawn using Edinburgh Pathway Notation and Editor

Each interaction taken from literature with restriction that each must be cited in 2 papers from different laboratories

Contains:

• 61 proteins
• 17 genes
• 24 protein complexes

Map currently being expanded to cover other related signalling pathways

Current Apoptosis Pathway Map
Network analysis

- Network analysis has been used to model many types of data in order to better understand complex inter-relationships (edges) between entities (nodes).

- In biology network graphs have been used in the study of sequence similarity, protein structure, evolutionary relationships, protein interactions etc.
Lehner and Fraser

2,913 Nodes, 9,254 Edges
Graph from CytoScape
Network analysis and clustering of gene expression data

• Analysis of large expression datasets a common interest
• Data prone to noise (experimental and biological)
• Current clustering approaches are numerous, frequently based on pair-wise clustering approaches, usually slow and often ineffective at dealing with a large number of genes
• Relationship between clusters unclear
• Experimental design often non-optimal for pair-wise statistical approaches
Graph Paradigm for Gene Expression Data

- Genes (nodes) are connected to each other in a network based on their level of co-expression (edges)

- Co-expression measured using a correlation measure (e.g. Pearson, Spearman)

- Development of program BioLayoutExpress:
  - data import
  - Pearson calculations
  - 3D graph layout
  - link to Markov CLustering (MCL) algorithm
  - expression and annotation viewers
  - annotation statistical mining

Sanger, Team 101
Anton Enright
Stijn van Dongen
Russell Grocock
Markus Brosch
EBI
Leon Goldovsky
### Table 1: Expression Levels of Genes Across Tissues

<table>
<thead>
<tr>
<th>Gene</th>
<th>Tissue1</th>
<th>Tissue2</th>
<th>Tissue3</th>
<th>Tissue4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene1</td>
<td>100%</td>
<td>95%</td>
<td>50%</td>
<td>100%</td>
</tr>
<tr>
<td>Gene2</td>
<td>99%</td>
<td>55%</td>
<td>60%</td>
<td>100%</td>
</tr>
<tr>
<td>Gene3</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>100%</td>
</tr>
<tr>
<td>Gene4</td>
<td>40%</td>
<td>50%</td>
<td>55%</td>
<td>100%</td>
</tr>
<tr>
<td>Gene5</td>
<td>4%</td>
<td>2%</td>
<td>5%</td>
<td>4%</td>
</tr>
</tbody>
</table>

### Graphical Representation

- **Gene1**
- **Gene2**
- **Gene3**
- **Gene4**
- **Gene5**

### Pearson Calculations

- **Gene1**: 2.5 billion Pearson calculations
- **Gene2**: 50,000
- **Gene3**: 50,000
- **Gene4**: 50,000
- **Gene5**: 50,000

### Percentage Expression

<table>
<thead>
<tr>
<th>Gene</th>
<th>Expression 1</th>
<th>Expression 2</th>
<th>Expression 3</th>
<th>Expression 4</th>
<th>Expression 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene1</td>
<td>100%</td>
<td>99%</td>
<td>58%</td>
<td>38%</td>
<td>23%</td>
</tr>
<tr>
<td>Gene2</td>
<td>99%</td>
<td>100%</td>
<td>64%</td>
<td>46%</td>
<td>31%</td>
</tr>
<tr>
<td>Gene3</td>
<td>58%</td>
<td>64%</td>
<td>100%</td>
<td>97%</td>
<td>13%</td>
</tr>
<tr>
<td>Gene4</td>
<td>38%</td>
<td>46%</td>
<td>97%</td>
<td>100%</td>
<td>16%</td>
</tr>
<tr>
<td>Gene5</td>
<td>23%</td>
<td>31%</td>
<td>13%</td>
<td>16%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Network Analysis of hCMV Infection of Human Macrophages

Experimental design and workup of samples by Christian Sinzger, Tübingen, Germany

Questions addressed:

• What is the transcriptional response of human macrophages to infection by hCMV?

• What is the difference in response to a productive (VlhE) and non-productive (VlhF) infection?

• How specific is the response to viral infection?

Analysis used 90 Affymetrix U133A chips

Data QC and statistical analysis by Thorsten Forster
## Experimental design

### Time (PI)

<table>
<thead>
<tr>
<th>Time</th>
<th>Donor</th>
<th>E20041</th>
<th>E30025</th>
<th>E30079</th>
<th>E30035</th>
<th>E30066</th>
<th>14R_026</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 h</td>
<td>✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 h</td>
<td>✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 h</td>
<td>✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 h</td>
<td>✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 h</td>
<td>✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 h</td>
<td>✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 h</td>
<td>✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓ ✓✓✓✓✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Legend:**
- **Mock infection**
- **VlhE infection**
- **VlhF infection**
- **Asp infection**
- **Cbv infection**
- **Fil infection**
- **Sta infection**
Expression data (normalised and annotated)

Gene to gene Pearson correlation calculated for every probe set on the array

Pearson correlations >0.7 saved

Pearson correlation file >0.7 filtered based on user defined threshold (0.7-1.0)

Edges drawn between nodes (genes) based on correlations > than selected threshold

Singletons and graphs with <4 members removed

Network graphs laid out in tiled arrangement and clustered
Sinzger all data
0.80 Pearson Layout
7,079 nodes, 131,715 edges
Sinzger all data
0.80 Pearson Layout
Clustered 1.7 MCL Inflation
Cluster 2 – Down in VlhE (VlhF)
Cluster 4 – Up in Donor E66

Cluster 8 – Up in Donor E35
Cluster 7 – Up in Sta early
Cluster 25 – Up in Asp (8 h)
hCMV Infectome

Sinzger data 0.80 Pearson 1.7 MCL ‘Interesting clusters’

(1,2,3,7,10,13,22,23,25,27,29)
Graphical Display of Statistical Hits and Pathway Genes

- Statistical analysis of differential expression VlhE and VlhF vs. mock control
- interferon pathway and apoptosis genes
VIhE differentials on all genes
Up early
Up late
Down early
Down late
Sinzger data 0.80 Pearson ‘Interesting clusters’ Graph
VlhE Differentials
Sinzger data 0.80 Pearson ‘Interesting clusters’ Graph
Interferon Pathway Genes
Network to IFN Pathway

- Cluster 3
- Cluster 2
- Cluster 7
- Cluster 43
- Cluster 77
- No Class
A Systems Biology Approach

**Manipulate**
Knock-in RNAi compound GM

**Measure**
apoptosis markers microarray

**Model**
networks pathways

**Mine**
data analysis tools databases literature
Acknowledgements

Ella Palmer, RFCGR, UK

Andy Miller, Genetic Therapies Centre, Imperial College, UK

Christian Sinzger, University of Tubingen, DE

Anton Enright, Sanger Institute, Cambridge, UK

Peter Ghazal, GTI, Edinburgh, UK

Interested in learning more?
Wellcome Trust Advanced Course in
‘Functional Genomics and Systems Biology’
Sanger Institute, Cambridge UK, June 2007