Genomic data based drug discovery/repurposing

Di Wu
Oct 2015
Three aspects, three stories

• Identifying ‘cell of origin’ of breast-cancer subtypes from normal mammary cell types. Help find relevant Drug targets.
• Drug sensitivity across lung cancer subtypes.
• Integrating GWAS risk SNPs with public drug databases for drug repurposing.
Breast cancer is a heterogeneous disease

Are there different cells of origin for the different cancer subtypes?

Normal cell types in human mammary glands

Most breast cancers initiate in ductal tissue

Cross-section of duct

Stem cell \[\rightarrow\] Luminal progenitor \[\rightarrow\] Luminal
Microarray datasets

- Breast tumor subtypes
- Human normal mammary cell types

Data from collaboration with Jane Visvader and Geoff Lindeman’s lab in WEHI
4 cell types in normal human mammary gland ducts

CD49f^{-}EpCAM^{+}

CD49f^{+}EpCAM^{+}

Homogeneous

CD49f^{hi}EpCAM^{-}

Luminal Progenitor

Heterogeneous

CD49f^{-}EpCAM^{-}

Mature Luminal

ML

LP

Stroma

Stromal (mainly fibroblast)

Stroma

In vivo

Mammary Stem Cell (MaSC)-enriched

Xenotransplantation

MS

Elgene Lim
Illumina human beadchips

stem cell
luminal progenitor
mature luminal
stromal cell

Illumina human beadchips
Normal cell types have distinct expression profiles

Multidimensional Scaling (MDS) plot
Breast cancer subtype data

Expression profiles of human breast tumors from C Perou’s group.  
2-color Agilent Human Microarray

<table>
<thead>
<tr>
<th>Cancer subtype</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basal-like</td>
<td>33</td>
</tr>
<tr>
<td>Claudin-low</td>
<td>5</td>
</tr>
<tr>
<td>HER2+/ER-</td>
<td>14</td>
</tr>
<tr>
<td>Luminal A</td>
<td>23</td>
</tr>
<tr>
<td>Luminal B</td>
<td>14</td>
</tr>
<tr>
<td>Normal breast-like</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>94</strong></td>
</tr>
</tbody>
</table>

Herschkowitz et al, Genome Biology (2007)
Do the tumors show traces of transcriptional signature from normal mammary cell types?

Tumor subtypes

Normal mammary cell types

Which tumor subtype is most similar to which cell type?
Challenging to create a link between both datasets and find the traces statistically

Heatmap was made based on the normalized expression value of 100 genes with largest variability across normalized combined data.
Signature gene sets for normal cell types
Signature genes for normal cell types

- **Stroma**
  - 846 up
  - 397 down

Significantly up (or down) in one cell type versus other types

Discriminatory strength of each signature gene is average log-fold change, versus the other cell types

Fold change > 2, FDR p < 0.05
Considering one positive LP signature gene expression in tumors

CHI3L1

basal-like

normal-like

tumor subtypes

Expression

Basal, Normal, Claudin-low, ERBB2, luminal B, luminal A.
Considering another positive LP signature gene expression in tumors

TFCB2L1

tumor subtypes

Expression

- Basal
- Normal
- Claudin-low
- ERBB2
- luminal B
- luminal A
Combining two LP positive signature genes

<table>
<thead>
<tr>
<th>logFC</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.18</td>
<td>CHI3L1</td>
</tr>
<tr>
<td>2.5</td>
<td>TFCB2L1</td>
</tr>
</tbody>
</table>

Combination of CHI3L1 and TFCB2L1 genes shows a significant logFC increase.

Box plot showing the signature score distribution across different tumor subtypes (Basal, Normal, Claudin-low, ERBB2, Luminal B, Luminal A).
Combining all LP signature genes

Signature score in tumors =
  average log-expression of **positive** LP signature genes
- average log-expression of **negative** LP signature genes
(weighted by size of log-fold-change between LP vs other cell types)

Highest expression of luminal progenitor gene signature in basal-like breast cancer → cell of origin for basal tumor

higher scores
more like luminal progenitor

$t$ test
* $p < 0.05$, ** $p < 0.01$, *** $p = 0.001$
Score is the average log-expression of tumour samples weighted by log-fold-change between cell types

For the gene $g$ in gene set $G_i$ for cell type $i$

$$S_{i,j} = \frac{\sum_g x_{gij} y_{g,j}}{\sum_g x_{g,j}}$$

- $X_{g,i}$ is average log fold change between cell type $i$ to other cell types
- $y_{g,j}$ is log expression in tumour sample $j$

Higher scores = high similarity between that cell type and tumor samples
Similarities between normal mammary cell types and tumor subtypes

- Lum Prog signature
- MaSC enriched signature
- Mature Lum signature
- Stroma signature

** t test, * p < 0.05, ** p < 0.01, *** p = 0.001
Concordance

Tumor subtypes

Normal-like
Claudin-low

Basal-like

Luminal A, B

Normal mammary cell types

MaSC enriched

Luminal Progenitor

Mature Luminal
Gene ranking gives expression barcodes

Rank genes by differential expression (between 2 tumour types)

Gene ranking gives expression barcodes

<table>
<thead>
<tr>
<th>Gene 1</th>
<th>Gene 2</th>
<th>Gene 3</th>
<th>Gene 4</th>
<th>Gene 5</th>
<th>Gene 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene 7</td>
<td>Gene 8</td>
<td>Gene 9</td>
<td>Gene 10</td>
<td>Gene 11</td>
<td>Gene 12</td>
</tr>
<tr>
<td>Gene 13</td>
<td>Gene 14</td>
<td>Gene 15</td>
<td>Gene 16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tumor subtype A

Positive signature genes of a cell type

Tumor subtype B

Genome-wide barcode plot

Negative signature genes of a cell type
LP signature is higher in basal tumors

<table>
<thead>
<tr>
<th></th>
<th>Tumor</th>
<th>Wilk P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td></td>
<td>0.025</td>
</tr>
<tr>
<td>Like</td>
<td></td>
<td>&lt; 10^-6</td>
</tr>
<tr>
<td>Claudin</td>
<td></td>
<td>&lt; 10^-6</td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td>&lt; 10^-4</td>
</tr>
<tr>
<td>ERBB2</td>
<td></td>
<td>&lt; 10^-6</td>
</tr>
<tr>
<td>Basal</td>
<td></td>
<td>&lt; 10^-6</td>
</tr>
<tr>
<td>like</td>
<td></td>
<td>&lt; 10^-6</td>
</tr>
<tr>
<td>Lum B</td>
<td></td>
<td>&lt; 10^-6</td>
</tr>
<tr>
<td>Lum A</td>
<td></td>
<td>&lt; 10^-6</td>
</tr>
</tbody>
</table>
Permuting genes vs permuting arrays

Permuting array labels
Tests significance of gene set in isolation: self-contained test

Permuting gene labels
Tests ranking of gene set relative to other genes: competitive test

Goeman & Bühlmann 2007
Correlation Adjusted MEan RAnk gene set test (*camera*)

Variance of mean-rank ($\bar{r}$) of G genes

$$\bar{r} = \sum_{g=1}^{G} r_g$$

is determined only by the average correlation

$$\text{var} \bar{r} = \sigma^2 \left\{ 1 + (G - 1) \bar{\rho} \right\}$$

where

$$\bar{\rho} = \text{corr}(r_i, r_j) = \frac{\sum_{i \neq j} \rho_{ij}}{G(G - 1)}$$

Nucleic Acid Research, 2012 May

*Camera: a competitive gene set test accounting for inter-gene correlation*

Di Wu^{1,3,*} and Gordon K. Smyth^{1,4,*}
LP signature is higher in basal tumors

Correcting for inter-gene correlation
Permuting genes vs permuting arrays

Permuting array labels
Tests significance of gene set in isolation: self-contained test

Permuting gene labels
Tests ranking of gene set relative to other genes: competitive test

Goeman & Bühlmann 2007
Roast, rotation gene set test

• Used with linear models (allows complex designs)
• Replaces array permutation with random rotation of residuals (avoids granularity of P-values, handles small sample sizes)
• Assumes multivariate normality, but proves to be highly robust to deviations from normality

ROAST: rotation gene set tests for complex microarray experiments
Di Wu1,2, Elgene Lim1, François Vaillant1, Marie-Liesse Asselin-Labat1, Jane E. Visvader1,2 and Gordon K. Smyth1,2,*

1The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052 and
2The University of Melbourne, Victoria 3010, Australia

Bioinformatics, July 2010
Two steps: projection and rotation

- **Project** data onto the null-hypothesis residual space (eliminates coefficients not interested in)
- Random **rotation** of the remaining orthogonal space (coefficient-of-interest + residuals)
- Monte-Carlo p-value
Roast – select genes

Genes in set
Roast – projection

Project data onto space orthogonal to nuisance parameters in linear model

\[ X \]

- design matrix
- expression matrix
Roast – projection

Project data onto space orthogonal to nuisance parameters in linear model
Roast – projection

Discard nuisance parameters

- Design matrix
- Nuisance parameters
- Test parameter
- Independent residuals
Roast – rotation

Rotate effects to simulate null hypothesis

Rotation matrix

\[ R^T R I = \]

Parallel rotation preserves correlation structure
Roast

• Allows gene-wise weights and choice of test statistics

• Can combine positive and negative signature genes in one test, using gene-wise directional weights
### LP signature is higher in basal tumors

<table>
<thead>
<tr>
<th>Tumor</th>
<th>Wilk</th>
<th>camera</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal-like</td>
<td>0.025</td>
<td>0.27</td>
</tr>
<tr>
<td>Claudin Low</td>
<td>&lt; 10^{-6}</td>
<td>0.02</td>
</tr>
<tr>
<td>ERBB2</td>
<td>&lt; 10^{-6}</td>
<td>&lt; 10^{-4}</td>
</tr>
<tr>
<td>Lum B</td>
<td>&lt; 10^{-6}</td>
<td>&lt; 10^{-4}</td>
</tr>
<tr>
<td>Lum A</td>
<td>&lt; 10^{-6}</td>
<td>&lt; 10^{-3}</td>
</tr>
</tbody>
</table>

### P-values

Allowing for inter-gene correlation.
Summary

Luminal progenitor might be the cell of origin for basal-like breast tumors.
Drug target in luminal progenitor signature for basal-like cancer

<table>
<thead>
<tr>
<th></th>
<th>48</th>
<th>95</th>
<th>95</th>
<th>64</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>58</td>
<td>15</td>
<td>75</td>
<td>74</td>
</tr>
<tr>
<td>C</td>
<td>68</td>
<td>95</td>
<td>35</td>
<td>84</td>
</tr>
<tr>
<td>B</td>
<td>49</td>
<td>05</td>
<td>45</td>
<td>44</td>
</tr>
<tr>
<td>P</td>
<td>39</td>
<td>05</td>
<td>65</td>
<td>44</td>
</tr>
<tr>
<td>B</td>
<td>49</td>
<td>84</td>
<td>85</td>
<td>84</td>
</tr>
<tr>
<td>E</td>
<td>28</td>
<td>24</td>
<td>95</td>
<td>84</td>
</tr>
<tr>
<td>E</td>
<td>68</td>
<td>84</td>
<td>85</td>
<td>84</td>
</tr>
<tr>
<td>E</td>
<td>46</td>
<td>74</td>
<td>64</td>
<td>84</td>
</tr>
<tr>
<td>E</td>
<td>66</td>
<td>54</td>
<td>24</td>
<td>84</td>
</tr>
</tbody>
</table>

c-KIT: proto-oncogene, mutations associated with leukaemia and gastrointestinal stromal tumour. Inhibitors in clinical use.
Our other stories of relating gene expression datasets

Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways

Elgene Lim1, Di Wu1, Bhupinder Pal1, Toula Bouras1, Marie-Liesse Asselin-Labat1, Francois Vaillant1, Hideo Yagita3, Geoffrey J Lindeman1,4, Gordon K Smyth1,2 and Jane E Visvader1,2

Control of mammary stem cell function by steroid hormone signalling

Marie-Liesse Asselin-Labat1, Francois Vaillant1, Julie M. Sheridan1, Bhupinder Pal1, Di Wu2, Evan R. Simpson3, Hisataka Yasuda4, Gordon K. Smyth1,2, T. John Martin5, Geoffrey J. Lindeman1,6,7 & Jane E. Visvader1,2
Three aspects, three stories

• Identifying ‘cell of origin’ of breast-cancer subtypes from normal mammary cell types. Help find relevant Drug targets

• Drug sensitivity across lung cancer subtypes

• Integrating GWAS risk SNPs with public drug databases for drug repurposing
Gene-expression data integration to squamous cell lung cancer subtypes reveals drug sensitivity

D Wu*1,2, Y Pang3, M D Wilkerson4, D Wang2, P S Hammerman5,6 and J S Liu*,1

1Department of Statistics, Harvard University, Cambridge, MA, USA; 2Centre for Cancer Research, Monash Institute of Medical Research, Monash University, Clayton, Victoria, Australia; 3Department of Biochemistry, University of Washington, Seattle, WA, USA; 4Department of Genetics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 5The Eli and Edythe L. Broad Institute of Massachusetts, Institute of Technology and Harvard University, Cambridge, MA, USA and 6Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA
Associate SqCC subtypes with airway cell culture time
Table 2. Classification results of the 28 SqCC cell lines to SqCC subtypes

<table>
<thead>
<tr>
<th>Cell line</th>
<th>First subtype</th>
<th>Second subtype</th>
<th>First subtype</th>
<th>Permutation $P$</th>
<th>Second subtype</th>
<th>Permutation $P$</th>
<th>Predicted</th>
<th>With drug data</th>
</tr>
</thead>
</table>

Gene expression

Classification of microarrays to nearest centroids

Alan R. Dabney
Department of Biostatistics, University of Washington, Seattle 98195, USA
## Table 2. Classification results of the 28 SqCC cell lines to SqCC subtypes

<table>
<thead>
<tr>
<th>Cell line</th>
<th>First subtype</th>
<th>Second subtype</th>
<th>First subtype</th>
<th>Permuatiaon P</th>
<th>Second subtype</th>
<th>Permutation P</th>
<th>Predicted</th>
<th>With drug data</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCI-H226</td>
<td>Secretory</td>
<td>Primitive</td>
<td>Secretory</td>
<td>0.16</td>
<td>Basal</td>
<td>0.457</td>
<td>Secretory</td>
<td>y</td>
</tr>
<tr>
<td>EBC-1</td>
<td>Secretory</td>
<td>Basal</td>
<td>Basal</td>
<td>0.426</td>
<td>Primitive</td>
<td>0.277</td>
<td>Basal</td>
<td>y</td>
</tr>
<tr>
<td>VMRC-LCP</td>
<td>Basal</td>
<td>Classical</td>
<td>Basal</td>
<td>0.033</td>
<td>Classical</td>
<td>0.32</td>
<td>Basal</td>
<td>y</td>
</tr>
<tr>
<td>HCC-1588</td>
<td>Basal</td>
<td>Primitive</td>
<td>Basal</td>
<td>0.082</td>
<td>Secretory</td>
<td>0.595</td>
<td>Basal</td>
<td>y</td>
</tr>
<tr>
<td>RERF-LC-AI</td>
<td>Primitive</td>
<td>Secretory</td>
<td>Primitive</td>
<td>0.07</td>
<td>Secretory</td>
<td>0.311</td>
<td>Primitive</td>
<td>y</td>
</tr>
<tr>
<td>SK-MES-1</td>
<td>Primitive</td>
<td>Secretory</td>
<td>Basal</td>
<td>0.293</td>
<td>Secretory</td>
<td>0.318</td>
<td>Secretory</td>
<td>y</td>
</tr>
<tr>
<td>NCI-H520</td>
<td>Classical</td>
<td>Primitive</td>
<td>Primitive</td>
<td>0.076</td>
<td>Classical</td>
<td>0.678</td>
<td>Primitive</td>
<td>y</td>
</tr>
<tr>
<td>HCC-15</td>
<td>Classical</td>
<td>Secretory</td>
<td>Classical</td>
<td>0.061</td>
<td>Basal</td>
<td>0.528</td>
<td>Classical</td>
<td>y</td>
</tr>
<tr>
<td>SW 900</td>
<td>Primitive</td>
<td>Basal</td>
<td>Secretory</td>
<td>0.109</td>
<td>Basal</td>
<td>0.25</td>
<td>Secretory</td>
<td>y</td>
</tr>
<tr>
<td>LUDLU-1</td>
<td>Basal</td>
<td>Classical</td>
<td>Classical</td>
<td>0.024</td>
<td>Basal</td>
<td>0.077</td>
<td>Classical</td>
<td>y</td>
</tr>
<tr>
<td>HARA</td>
<td>Secretory</td>
<td>Classical</td>
<td>Basal</td>
<td>0.308</td>
<td>Secretory</td>
<td>0.557</td>
<td>Secretory</td>
<td>y</td>
</tr>
<tr>
<td>KNS-62</td>
<td>Basal</td>
<td>Classical</td>
<td>Basal</td>
<td>0.118</td>
<td>Classical</td>
<td>0.551</td>
<td>Basal</td>
<td>y</td>
</tr>
<tr>
<td>EPLC-272H</td>
<td>Basal</td>
<td>Primitive</td>
<td>Basal</td>
<td>0.114</td>
<td>Secretory</td>
<td>0.339</td>
<td>Basal</td>
<td>y</td>
</tr>
<tr>
<td>HCC-95</td>
<td>Classical</td>
<td>Secretory</td>
<td>Classical</td>
<td>0.009</td>
<td>Basal</td>
<td>0.667</td>
<td>Classical</td>
<td>y</td>
</tr>
<tr>
<td>LC-1/sq-SF</td>
<td>Classical</td>
<td>NA</td>
<td>Classical</td>
<td>0.013</td>
<td>Basal</td>
<td>0.688</td>
<td>Classical</td>
<td>y</td>
</tr>
<tr>
<td>SW 1573</td>
<td>Secretory</td>
<td>Primitive</td>
<td>Primitive</td>
<td>0.108</td>
<td>Secretory</td>
<td>0.277</td>
<td>Primitive</td>
<td>y</td>
</tr>
<tr>
<td>LC-1F</td>
<td>Classical</td>
<td>Basal</td>
<td>Classical</td>
<td>0.012</td>
<td>Primitive</td>
<td>0.539</td>
<td>Classical</td>
<td>y</td>
</tr>
<tr>
<td>LOU-NH91</td>
<td>Secretory</td>
<td>Primitive</td>
<td>Secretory</td>
<td>0.19</td>
<td>Primitive</td>
<td>0.231</td>
<td>Secretory</td>
<td>y</td>
</tr>
<tr>
<td>LK-2</td>
<td>Classical</td>
<td>Primitive</td>
<td>Primitive</td>
<td>0.074</td>
<td>Classical</td>
<td>0.585</td>
<td>Primitive</td>
<td>y</td>
</tr>
<tr>
<td>RERF-LC-Sq-1</td>
<td>Primitive</td>
<td>Classical</td>
<td>Secretory</td>
<td>0.293</td>
<td>Primitive</td>
<td>0.209</td>
<td>Primitive</td>
<td>y</td>
</tr>
<tr>
<td>HLF-a</td>
<td>Primitive</td>
<td>Classical</td>
<td>Secretory</td>
<td>0.063</td>
<td>Primitive</td>
<td>0.233</td>
<td>Secretory</td>
<td>y</td>
</tr>
<tr>
<td>Calu-1</td>
<td>Primitive</td>
<td>Secretory</td>
<td>Secretory</td>
<td>0.08</td>
<td>Basal</td>
<td>0.583</td>
<td>Secretory</td>
<td>y</td>
</tr>
<tr>
<td>NCI-H2170</td>
<td>Basal</td>
<td>Classical</td>
<td>Classical</td>
<td>0.114</td>
<td>Primitive</td>
<td>0.085</td>
<td>Classical</td>
<td>y</td>
</tr>
<tr>
<td>Sq-1</td>
<td>Primitive</td>
<td>NA</td>
<td>Secretary</td>
<td>0.129</td>
<td>Basal</td>
<td>0.166</td>
<td>Secretory</td>
<td>y</td>
</tr>
<tr>
<td>HCC-2814</td>
<td>Classical</td>
<td>Basal</td>
<td>Classical</td>
<td>0.048</td>
<td>Basal</td>
<td>0.369</td>
<td>Classical</td>
<td>y</td>
</tr>
<tr>
<td>HCC-1897</td>
<td>Secretory</td>
<td>Primitive</td>
<td>Secretory</td>
<td>0.07</td>
<td>Basal</td>
<td>0.39</td>
<td>Secretory</td>
<td>y</td>
</tr>
<tr>
<td>NCI-H1385</td>
<td>Classical</td>
<td>Primitive</td>
<td>Primitive</td>
<td>0.069</td>
<td>Classical</td>
<td>0.352</td>
<td>Primitive</td>
<td>y</td>
</tr>
<tr>
<td>NCI-H1869</td>
<td>Basal</td>
<td>NA</td>
<td>Secretary</td>
<td>0.244</td>
<td>Basal</td>
<td>0.303</td>
<td>Basal</td>
<td>y</td>
</tr>
</tbody>
</table>

Abbreviations: SpaNC = classification method to nearest centroids; SqCC = squamous cell lung cancer.
Less drug sensitivity in secretory SqCC cell line due to low proliferation
Three aspects, three stories

- Identifying ‘cell of origin’ of breast-cancer subtypes from normal mammary cell types. Help find relevant Drug targets
- Drug sensitivity across lung cancer subtypes
- Integrating GWAS risk SNPs with public drug databases for drug repurposing
Drug Repurposing

Approved small molecules treating other diseases

• Two nice properties of drug repurpose
  – Cost effectiveness
  – Less safety issues
Two well-known examples

Thalidomide
- 1950-1960: Morning sickness
- (1994) 2006: multiple myeloma

Viagra (sildenafil)
- Re-purpose
  - Pulmonary hypertension
  - 1998: erectile dysfunction
Why genetics-based drug repurposing?

- Genetics help better understand disease mechanism
- Previous drug repurposing is mostly not genetics based
- More identified risk SNPs based on GWAS in past 6-7 years

(According to NIH GWAS catalog database, www.genome.gov/26525384)

Use of genome-wide association studies for drug repositioning
Sanseau et al, Nature Biotechnology 2012
Our goals

• Genetics based drug repurposing

• How genetics guides drug discovery
Simple strategies to use genetics guiding drug repurposing

GWAS: genome-wide association studies
1. Is a drug target also a risk gene?
   → direct overlap method

2. Is a PPI neighbor of drug target also a risk gene?
   → 1 way PPI method

3. Are the neighbors significantly overlapped?
   → 2 way PPI method

Risk gene in one disease
PPI neighbors

> 1,000 drugs
> 500 target genes
> 2,500 drug-gene pairs
Results
1/7 are indicated for RA
direct overlap

92 SNPs → 55 significant genes

<table>
<thead>
<tr>
<th>generic Name</th>
<th>drug ID</th>
<th>tarG</th>
<th>action</th>
<th>SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denileukin diftitox</td>
<td>DB00004</td>
<td>IL2RA</td>
<td>binder</td>
<td>rs706778</td>
</tr>
<tr>
<td>Intravenous Immunoglobulin</td>
<td>DB00028</td>
<td>C5</td>
<td>binder</td>
<td>rs1953126, rs3761847, rs881375</td>
</tr>
<tr>
<td>Basiliximab</td>
<td>DB00074</td>
<td>IL2RA</td>
<td>antibody</td>
<td>rs706778</td>
</tr>
<tr>
<td>Muromonab</td>
<td>DB00075</td>
<td>CD247</td>
<td></td>
<td>rs840016</td>
</tr>
<tr>
<td>Daclizumab</td>
<td>DB00111</td>
<td>IL2RA</td>
<td>antibody</td>
<td>rs706778</td>
</tr>
<tr>
<td>Eculizumab</td>
<td>DB01257</td>
<td>C5</td>
<td>antibody</td>
<td>rs1953126, rs3761847, rs881375</td>
</tr>
<tr>
<td>Ipilimumab</td>
<td>DAP001561</td>
<td>CTLA4</td>
<td>antibody</td>
<td>rs231735, rs3087243</td>
</tr>
</tbody>
</table>

Target genes: IL2RA, C5, CTLA4

Rheumatoid Arthritis (RA)

Direction?
4/10 of 1 way PPI expansion are for RA (top 10, 8/159 total)

<table>
<thead>
<tr>
<th>Generic Name</th>
<th>drug ID</th>
<th>target</th>
<th>action</th>
<th>nPPI</th>
<th>opG</th>
<th>hyp.p</th>
<th>snpid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abatacept</td>
<td>DB01281</td>
<td>CD80</td>
<td>antagonist</td>
<td>5</td>
<td>CTLA4</td>
<td>0.08</td>
<td>rs231735, rs3087243</td>
</tr>
<tr>
<td>Antithymocyte globulin</td>
<td>DB00098</td>
<td>CD86</td>
<td>Inhibitor</td>
<td>6</td>
<td>CTLA4</td>
<td>0.09</td>
<td>rs231735, rs3087243</td>
</tr>
<tr>
<td>Abatacept</td>
<td>DB01281</td>
<td>CD86</td>
<td>antagonizer</td>
<td>6</td>
<td>CTLA4</td>
<td>0.09</td>
<td>rs231735, rs3087243</td>
</tr>
<tr>
<td>Sulfacetamide</td>
<td>DAP001191</td>
<td>MPI</td>
<td>Inhibitor</td>
<td>6</td>
<td>PTPN2</td>
<td>0.09</td>
<td>rs2847297, rs1893217</td>
</tr>
<tr>
<td>Sulfoxone</td>
<td>DAP001192</td>
<td>MPI</td>
<td>Inhibitor</td>
<td>6</td>
<td>PTPN2</td>
<td>0.09</td>
<td>rs2847297, rs1893217</td>
</tr>
<tr>
<td>Sulfanilamide</td>
<td>DAP001193</td>
<td>MPI</td>
<td>Inhibitor</td>
<td>6</td>
<td>PTPN2</td>
<td>0.09</td>
<td>rs2847297, rs1893217</td>
</tr>
<tr>
<td>Rituximab</td>
<td>DB00073</td>
<td>MS4A1</td>
<td>antibody</td>
<td>8</td>
<td>CD40</td>
<td>0.12</td>
<td>rs4810485</td>
</tr>
<tr>
<td>Ibritumomab</td>
<td>DB00078</td>
<td>MS4A1</td>
<td>antibody</td>
<td>8</td>
<td>CD40</td>
<td>0.12</td>
<td>rs4810485</td>
</tr>
<tr>
<td>Tositumomab</td>
<td>DB00081</td>
<td>MS4A1</td>
<td>antibody</td>
<td>8</td>
<td>CD40</td>
<td>0.12</td>
<td>rs4810485</td>
</tr>
<tr>
<td>Intravenous Immunoglobulin</td>
<td>DB00028</td>
<td>C5</td>
<td>binder</td>
<td>13</td>
<td>C5</td>
<td>0.18</td>
<td>rs881375</td>
</tr>
</tbody>
</table>
LETTER

Genetics of rheumatoid arthritis contributes to biology and drug discovery

A list of authors and their affiliations appears at the end of the paper

Yukinori Okada¹-³*, Di Wu⁴-⁵, Gosia Trynka¹-³, Towfique Raj²,³,⁶, Chikashi Terao⁷,⁸, Katsunori Ikari⁹, Yuta Kochi¹⁰, Koichiro Ohmura⁸, Akari Suzuki¹⁰, Shinji Yoshida⁹, Robert R. Graham¹¹, ...

Ryo Yamada⁶², Michiaki Kubo⁶³, Jun Liu⁴, Sang-Cheol Bae⁴⁵, Jane Worthington⁵⁷,⁵⁸, Leonid Padyukov⁶⁴, Lars Klareskog⁶⁴, Peter K. Gregersen⁶⁵, Soumya Raychaudhuri¹-³,⁶⁶, Barbara E. Stranger⁶⁷,⁶⁸, Philip L. De Jager²,³,⁶, Lude Franke²⁵, Peter M. Visscher¹⁹,²⁰, Matthew A. Brown¹⁹, Hisashi Yamanaka⁹, Tsuneyo Mimori⁸, Atsushi Takahashi⁶⁹, Huji Xu¹⁸, Timothy W. Behrens¹¹, Katherine A. Siminovitch²¹-²³, Shigeki Momohara⁹, Fumihiko Matsuda⁷,⁷⁰,⁷¹, Kazuhiko Yamamoto¹⁰,⁷², and Robert M. Plenge¹-³,*.
ROC curves determine the best rank

For RA

ROC: receiver operating characteristic
Drug and disease information not enough \(\rightarrow\) False negatives

- **FN**: disease’s drugs not in repurpose table
- Drug and disease are **not connected by genetics**
  - Type 1 Diabetes
  - Anti-TNFα for RA
- Drug and diseases are connected by genetics, but **not through**
  - identified risk genes
  - PPI neighbors
Reasons to call false positives

• FP: drugs not known for the disease, but in repurpose table; can be rep drugs

• Threshold for risk genes maybe not high enough. Non-relevant genes were falsely considered as risk genes

• The drug action directions is not clear

• Drugs with the same target gene in the same action direction may be indicated for different diseases
Result summary

• We repurposed approved drugs for 286 disease-relevant phenotypes
• ROC and significant tests for possible combination of methods
• True disease drugs are enriched in repurposing table
• About 40% of risk SNPs are used for repurposing
Three aspects, three stories

• Identifying cell of origin of breast-cancer subtypes form normal mammary cell types. Help find relevant Drug targets

• Drug sensitivity across several lung cancer cell types

• Integrating GWAS risk SNPs with public drug databases for drug repurposing
Acknowledgments

Harvard Statistics

Jun S Liu
Ke Deng
Ming Hu
Yushu Pang

Harvard Biostatistics

Tianxi Cai

BWH, Harvard Medical School

Robert Plenge (MERCK)
Yukinori Okada
Dorothee Diogo
Bobby Reddy
Gang Li
Soumya Raychaudhuri
Towfique Saj
Gyan Srivastava

Broad Institute

Paul Clemons
Vlado Dancik

Beijing Jiaotong University

Xuezhong zhou

Bioinformatics Division

WEHI

Gordon Smyth
Terry Speed

Victoria Breast Cancer Research Consortium

WEHI

Jane Visvader
Geoff Lindeman
François Vaillant
Marie-Liesse Asselin-Labat

Dana-Farber Cancer Institute

Elgene Lim

Dana Farber Cancer Institute

Broad Institute

Matthew Meyerson
Peter Hammerman
Heidi Greulich

University of North Carolina

Chapel Hill
Matthew D. Wilkerson
Neil Hayes

Texas A&M University

Alan Dabney

Monash University

Bryan Williams
Die Wang