Computationally Driven Learning of Disease Mechanisms to Predict Patient Outcome

Francesco Falciani
Centre for Computational Biology and Modelling
Institute of Integrative Biology
University of Liverpool
UK
The Challenge

- Increased use of high-throughput biological techniques
  - Generation of extraordinary amount of data
  - However, single omics analyses only provide information about one layer of a cellular system
    - Panels of candidate genes
  - Further, organ cross talk

*Fig. 3.2* Schematic diagram of interactions among various functional layers in a cellular system. Blank arrow, flow of biological information; dashed line, possible interaction between various biomolecular species
A computational perspective
Insulin Sensitivity

• The functional integrity of the muscle-skeletal system requires continuous physical exercise, which is therefore essential for healthy tissue homeostasis.

• Up to 25% of perfectly healthy subjects who complete a supervised exercise-training program demonstrate no measurable response.

• Up to 15% demonstrate even an adverse response to exercise.

The question is:

We currently do not understand the mechanisms underlying such heterogeneity and we cannot predict the complex spectrum of phenotypes from the molecular state of the muscle.
HERITAGE STUDY

Made available by Claude Bouchard
HERITAGE STUDY
Made available by Claude Bouchard

**Discovery cohort**

- 99 Caucasian families (n=500)
- Sedentary at baseline
- Disease-free
- Age range: 17-65 yrs
- 20-week fully supervised training (individually customized)
- +100 biochemical/physiological read-outs collected pre/post
- Pronounced inter-individual variability in training response for many key variables (e.g. VO2_max, HR, insulin sensitivity...)

**Validation cohort**

Matthijs Hesselink
Matthijs K.C. Hesselink, Ph.D.
Professor of Human Movement Sciences
Maastricht University
Step 1: Identification of genetic variants associated to ∆SI

GWAS SNP genotyping

Expression profiling

Identification of biological pathways enriched in ∆SI-associated SNPs and potentially controlling mRNA expression in cis

KEGG Pathways
**Step 1**: Identification of genetic variants associated to ΔSI

- GWAS SNP genotyping
- Expression profiling

**Step 2**: Identification and validation of transcription factors controlling baseline gene expression linked to ΔSI

1. *in silico* identification of TFs potentially controlling the expression of gene expression signatures correlated to ΔSI.
2. Selection of TFs controlled by pathways affected by ΔSI SNPs
3. Experimental validation of MEF2

- C2C12 myotubes ± MEF2A-siRNA

**Identification of biological pathways enriched in ΔSI-associated SNPs and potentially controlling mRNA expression in cis**

**KEGG Pathways**
Step 1: Identification of genetic variants associated to ΔSI
- GWAS SNP genotyping
- Expression profiling

Identification of biological pathways enriched in ΔSI-associated SNPs and potentially controlling mRNA expression in cis

Step 2: Identification and validation of transcription factors controlling baseline gene expression linked to ΔSI
- *in silico* identification of TFs potentially controlling the expression of gene expression signatures correlated to ΔSI.
- Selection of TFs controlled by pathways affected by ΔSI SNPs
- Experimental validation of MEF2 C2C12 myotubes ± MEF2A-siRNA

Step 3: Development of a MEF2 gene expression signature predictive of DSI
- Development of mRNA based predictor of ΔSI
- Validation of molecular predictor using independent clinical exercise training cohort

From Davidsen PK et al.
Training-induced change in SI plotted by mean family rank in the HERITAGE study

94 families with an average of 4.3 members. Each bar represents the range of training responses within a family (both parents and offspring). Red horizontal dotted line denotes family means.
From Davidsen PK et al.
KEGG pathways enriched for SNP variants whose genotype additively associate with \( \Delta \text{SI} \)

<table>
<thead>
<tr>
<th>KEGG pathway</th>
<th>Gene Count</th>
<th>SNP count *</th>
<th>FDR</th>
<th>Functional group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac muscle contraction</td>
<td>77</td>
<td>639</td>
<td>&lt;0.001</td>
<td>Calcium</td>
</tr>
<tr>
<td>Inositol phosphate metabolism</td>
<td>57</td>
<td>463</td>
<td>0.017</td>
<td>Calcium</td>
</tr>
<tr>
<td>Arachidonic acid metabolism</td>
<td>59</td>
<td>200</td>
<td>0.023</td>
<td>Calcium</td>
</tr>
<tr>
<td>Galactose metabolism</td>
<td>27</td>
<td>130</td>
<td>&lt;0.001</td>
<td>Carbohydrate</td>
</tr>
<tr>
<td>Starch and sucrose metabolism</td>
<td>55</td>
<td>140</td>
<td>0.002</td>
<td>Carbohydrate</td>
</tr>
<tr>
<td>Pentose phosphate pathway</td>
<td>27</td>
<td>132</td>
<td>0.017</td>
<td>Carbohydrate</td>
</tr>
<tr>
<td>Fructose and mannose metabolism</td>
<td>36</td>
<td>207</td>
<td>0.017</td>
<td>Carbohydrate</td>
</tr>
<tr>
<td>Focal adhesion</td>
<td>200</td>
<td>1575</td>
<td>0.023</td>
<td>Cell communication</td>
</tr>
<tr>
<td>Adherens junction</td>
<td>73</td>
<td>658</td>
<td>0.026</td>
<td>Cell communication</td>
</tr>
</tbody>
</table>

*Observed number of cis-SNPs associated with basal mRNA abundance

\(~90\% \text{ fall outside protein-coding regions}\)
Inference of TF activities

Gene expression correlated to DIS

GSEA approach

Involved in spliceosome complex formation - big gene set (only ~65% expressed in muscle)

1. Splicing factor-1
2. Myocyte enhancer factor (MEF)-2

Hypothesized that MEF2 could be a key driver of the pleiotropic transcriptional response linked to ΔSI
The calcium-dependent MEF2A TF drives the transcriptional signature associated to ∆SI

Can this MEF2A-dependent signature recapitulate the transcriptional signature linked to ∆SI in HERITAGE???
Predicting insulin sensitivity

Most predictive model:
(all possible multivariate linear regression)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Regression coefficient</th>
<th>Standard error</th>
<th>t-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>VO$_{2\max}$</td>
<td>-0.03</td>
<td>0.004</td>
<td>-0.68</td>
<td>0.50</td>
</tr>
<tr>
<td>Gender</td>
<td>-2.15</td>
<td>0.83</td>
<td>-2.59</td>
<td>0.013</td>
</tr>
<tr>
<td>CAMK2D</td>
<td>40.73</td>
<td>27.98</td>
<td>1.46</td>
<td>0.15</td>
</tr>
<tr>
<td>CAMK2G</td>
<td>74.78</td>
<td>32.04</td>
<td>2.33</td>
<td>0.02</td>
</tr>
<tr>
<td>HDAC4</td>
<td>-34.94</td>
<td>22.37</td>
<td>-1.56</td>
<td>0.13</td>
</tr>
<tr>
<td>CAMK2D: CAMK2G</td>
<td>-12.08</td>
<td>3.41</td>
<td>-3.54</td>
<td>0.001</td>
</tr>
<tr>
<td>CAMK2D: HDAC4</td>
<td>6.94</td>
<td>1.63</td>
<td>4.27</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CAMK2G: HDAC4</td>
<td>-0.40</td>
<td>2.90</td>
<td>-0.14</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Genes do not act in isolation, but interact with each other collectively

\[
\Delta SI = a\theta_1 + b\theta_2 + c\theta_3 + d\theta_1\theta_2 + e\theta_1\theta_3 + f\theta_2\theta_3 + sex + VO_2_{\max} + \varepsilon
\]
Calcium-related KEGG Pathways

Genes with an orange background contain DNA variants (within a 20-kb flanking window on either side) nominally associated to ∆SI. The basal mRNA abundance of genes with a red font are associated with DNA variants. Adapted from Davidsen PK et al.
This study thereby extends the notion that transcriptomic profiling captures critical information and process prognostic information about complex physiological processes.
Insulin Signalling

Published links:

1. Cell-cycle-regulated activation of Akt kinase by phosphorylation at its carboxyl terminus
2. E2F1 Regulates Cellular Growth by mTORC1 Signaling
3. Torin1-mediated TOR kinase inhibition reduces Wee1 levels and advances mitotic commitment in fission yeast and HeLa cells
4. PI3K regulation of the SKP-2/p27 axis through mTORC2
Live cell imaging of cell cycle timings in C2C12 myoblasts

FUCCI G     FUCCI O

The Multiphoton 2 (MP2) incubated confocal microscope at The Centre For Cell Imaging, Liverpool
The Experiment

The diagram illustrates the duration of the cell cycle with the timing of mitosis and the effects of insulin treatment. The experiment compares "No treatment" to "treatment between mitoses" and "chronic treatment" using FUCCI fluorescence for comparison. The cell cycle stages are G1, G1/S, S, G2, and M, with the duration of each stage highlighted for comparison.
Insulin and cell cycle

- Insulin 20nM
- Treated S-M
- Treated G1
- No treatment

Duration of cell cycle (h)

Time of birth relative to treatment (h)
Insulin and cell cycle

- No treatment: 100%
- Treated G1: 93%
- Treated S-M: 81%
- Chronic treatment: 97%

Significance levels:
- p=0.01
- p<0.0001
- p<0.001
- p<0.005

Range of % dividing cells.
The Mathematical Model

A.

- Insulin
- PI3K
- pS6K
- mTOR2
- AKT
- TSC1
- TSC2
- mTORC2
- Thr 308
- Ser 473

mTOR signalling

B.

Species level

Graph showing species level over time with different lines indicating different species:
- Cyclin E
- Cyclin B
- Wee1
- Cyclin A
- FUCCI O

C.

- Significant shorter
- No change

Graph showing species level over time with different areas indicating different treatments:
- Control
- Treatment between mitoses
- Chronic treatment
Space of parameters

No. of parameter sets

- Total: 11035
- Significant change: 843
- Fitting: 215
- Non-fitting: 430
- "Broken": 198

Legend:
- Blue: Treatment between mitoses
- Red: Chronic treatment
Space of parameters

B

1874 (66.5%)

42 (1.5%)

905 (32%)

Shorter (S)

Longer (L)

Relative time

Species level

C

Control

Insulin

Species level

Species level

Shorter (S)

Longer (L)

Broken (B)
Experimental validation
Experimental validation

A

Duration of cell cycle (h)

Time of birth relative to treatment (h)

mTOR inhibition

ONLY

No treatment

chronic

B

Duration of cell cycle (h)

Time of birth relative to treatment (h)

mTOR inhibition

ONLY

No treatment

chronic

p<0.0001

+ INS

Experimental validation
Experimental validation

- **A**: mTOR inhibition
  - No treatment
  - Chronic
  - nsc
  - p<0.0001

- **B**: mTOR inhibition + INSULIN
  - No treatment
  - Chronic
  - nsc
  - p<0.0001

- **C**: MEK inhibition
  - No treatment
  - Chronic
  - nsc

- **D**: MEK + mTOR inhibition
  - No treatment
  - Chronic
  - nsc

- **E**: MEK + mTOR inhibition + INSULIN
  - No treatment
  - Chronic
  - nsc
Experimental validation

A. mTOR inhibition ONLY

B. MEK inhibition ONLY

C. MEK + mTOR inhibition ONLY

D. MEK inhibition + INSULIN

E. MEK + mTOR inhibition + INSULIN

F. MEK inhibition + INSULIN
Primary satellite cells

A. PRIMARY SATELLITE CELLS

B. REL. CHANGE IN CELL CYCLE LENGTH

C. C2C12

D. REL. CHANGE IN CELL CYCLE LENGTH

ΔNT/NT  ΔBM/NT  ΔC/BM

Δnon/non  Δbetween/non  Δchronic/between
RAS
MEK
Insulin
Differentiation
IRS1
mTOR
AKT
Proliferation
Preparation for expansion
Cell Cycle
Expansion

Satellite cell on Muscle fiber

Progenitor activation

Differentiation
Fusion
New muscle fiber

A

B

Insulin

IRS1

RAS
MEK

Differentiation
p21
Cell Cycle

Proliferation

Expansion

Generation 1
(treatment during cycle)

Generation 2
(born into insulin)

Satellite cell on Muscle fiber

Mitosis

Preparation for expansion

C

Insulin

IRS1

LOW SETRUM

RAS
MEK

p21

Proliferation

Differentiation

Generation 2
(born into insulin)
Acknowledgements

University of Liverpool

Francesco Falciani
Kim Clarke*
Nil Turan*
Peter Davidson*
John Herbert*
John Ankers*
Michela Bonomo*

Collaborators

Stefano Biffo (DIBIT, Italy)
Stuart Egginton (Leeds, UK)
Dan Tennant (Birmingham, UK)
Rob Beynon (Liverpool, UK)
Malcom Jackson (Liverpool, UK)
Mark A. Sarzynski (Pennington, USA)
Claude Bouchard (Pennington, USA)
Matthijs K. C. Hesselink (Maastricht University Centre, The Netherland)
Patrick Schrauwen (Maastricht University Centre, The Netherland)