Multi-resolution network modelling of T-cells for precision medicine of multiple sclerosis

Mika Gustafsson, Assoc Professor
Translational bioinformatics
Linköping University
Motivation

• Most medications start late, are ineffective for most patients and costly.
• Patients who appear to have the same disease could be sub-types of the same disease, with different mechanisms.
• Altered function or expression of thousands of genes in different cells changing as disease progresses.

Aim is to develop network based methods for identification of predictive markers of complex diseases, so treatment can be early and individualised, ideally before disease becomes symptomatic.
Multiple sclerosis (MS) and T-cells as models

• Complex disease with 25% concordance of monozygotic twins
• >100 SNPs exists and T-cell differentiation is the mostly enriched
• Relapse-remitting MS have recurrent non-symptomatic phases, so patients can be followed with and without symptoms
• T-cells can be activated into a ‘disease alike state’, and short-time dynamics analysed (Hellberg, Cell Reports 2016)
Many undiscovered disease genes have small individual effects

Integrative disease models needed, e.g. networks
Interactome is a hairball
Interlinked omics -> coarse-graining needed

Gustafsson et al, Genome Med 2014
GWAS co-localize in PPI disease modules

- GWAS is static and limited to genetics
- Gene expression profiling of a disease relevant cell type can increase sensitivity and includes epi-genetic effects
How do you validate modules containing hundreds of genes?

• Genomic concordance, i.e. overlap across omics
• GWAS was used for validating topological disease modules in DREAM 2016 by Marbach
• Clinical studies and functional animal studies of combinations of module gene products as biomarkers

Hellberg et al, Cell Reports, 2016
Bruhn et al, Science Translational Medicine, 2014
Gustafsson et al, Genome Med, 2014
nine ongoing studies on disease modules
Performed in collaboration with Z Lubovac, Skövde (modules)
Preliminary results

Different methods perform best on different data-sets

Multiple Sclerosis (Sawcer Nature Genetic 2011)

Rheumatoid arthritis CD4 (Okada 2014 European)
Modules from the same methods most similar
Preliminary results using four methods on 24 diseases:

Conensus module outperforms all individual methods

https://gitlab.com/Gustafsson-lab/MODifieR
Summary: Modules

• Modules concept is vaguely defined
• Strong method bias
• Assessment needed for development of new methods and integrating new data types
Gene regulatory networks (GRNs) provide directionality

• Modules are time-independent, undirected and does not provide mechanism
• GRNs give possibility to go upstream leading to early disease markers [1]
• **Data compression:** ≈7% of protein-coding genes are transcription factors (TFs), some TFs are master regulators
• TF targets can be predicted from DNA sequence and the functional targets can be inferred by time-series and perturbation experiments

Gustafsson et al., *A N Y A S*, 2009
Gustafsson et al., *CMGRN*, 2008
Gustafsson et al, *IEEE Comput biol/Bioinfo*, 2005
Hubs TFs in T-cell diff were GWAS

- TFs degree might be easier to infer than individual interactions
- mRNA is often used as a proxy of TF activity relied on average healthy GRNs, i.e. were not individualised

Gustafsson et al, Science Translational Medicine, 2015
GRNs for analyzing multiple sclerosis (MS)

- MS has >100 reported SNPs that are annotated as eQTLs, which are in T-cell genes, but surprisingly those target genes are not differentially expressed

- Instead those genes exhibited the highest variance in relapsing patients

- We hypothesize that personalized control for the upstream TF activity will resolve this

Collaboration with J Ernerudh, LiU (MS), S Baranzini, USA (MS) and R Lahesmaa, Turku (high-throughput biology)
42 TFs enriched for target MS-genes

• TFs were surprisingly not among target genes -> no genetic feedbacks
Individualised GRN for each MS patients

"MS TFs" $m \sim 700$

"MS SNPs" $n = 1806$

"MS genes" $n = 244$

Dynamic Response

Remission

Low Disease Activity

High

Relapse
Model details

- Targeted DNA sequencing of 1800 MS associated eQTLs in 48 patients/controls
  -> 48 different prior TF-target networks
  \[ S_{ij}^k = \begin{cases} 
    \gamma_{ij} & \text{if patient } k \text{ has SNP } (i,j) \\
    0, & \text{otherwise} 
  \end{cases} \]

- 96 gene microarrays from active and non-active T-cells,

\[
X_i = mRNA_i[ACTIVE] - mRNA_i[REST]
\]

\[
\min_A (X - S \cdot A)^2 \quad \gamma_{ij} = \text{sign}(\min_{S_{ij}}|X - SX|_2)
\]

\[
\sum |A_{jk} - X_{jk}| < \eta_j \quad \text{L1 Deviation between mRNA of TF and its activity}
\]

\[
\sum |S_{ij}| < \lambda_j \quad \text{Sparse network}
\]
Removing putative edges and allow for deviation between TF activity and mRNA expression lowers cross-validation error.
Some TFs are hubs in patients
Some TFs that predicted diff activated
Combining topology and diff activation
Summary: patients specific GRNs

- Longitudinal data opens the door of individual networks
- Preliminary data suggests: Hub upstream TFs can be used to cluster MS patients
- Validation experiments are ongoing, i.e. ATAC-seq and western-blotting
- Generally applicable principle using paired longitudinal data at two disease associated states
- Ongoing SSF project: Use in vitro differentiation time-scale of ~24h and gradual disease recovery during 9 months pregnancy to fit a nonlinear dynamic model
Realistic ODE modelling including feedbacks is needed for predicting effect of drugs

• Early hub TFs are potential bottlenecks whose dysregulation could be early markers for disease [1]

• We aim to identify their upstream early activity regulation, which could give earlier markers and predict drug responses

• Requires detailed ODE modelling of a core set of TFs, their upstream signaling paths and their thousands of target genes

• We therefore collect high-throughput proteomics, phospho-proteomics and transcript variants

Mechanistic modelling of the effect of TFs and their upstream regulators in health is possible using LASSIM

gitlab.com/Gustafsson-lab/lassim
Proteomics, transcriptomics and phospho-proteomics of six time-points of Th1 differentiation, combined with inhibition data and priori knowledge modelling to model MS treatment in vitro.

Collaboration with M Kim, Korea (mass-spec) and C Nestor, LiU (T-cell biology)
Preliminary results

- 7000 proteins and 240,000 transcript variants detected, corresponding to about 15,000 unique genes that are differentially expressed at some time-point of Th1 differentiation

- Awaiting perturbation and phsopho-proteomics
• Modules are powerful integration tools that increase functional relevance, but still not well-defined and do not give mechanism
• Gene regulatory networks may identify upstream early master regulators
• Nonlinear modelling of those factors for predicting drug effects
• Simultaneous integration of those concepts should ideally add synergistically
2015- my group includes medical doctor, computer scientists, theo. physicist, immunologists, and engineers

Hard work performed by
• Tejaswi Venkata Badam
• Dirk de Weerd
• Simon Wu
• Julia Åkerström
• Rasmus Magnusson
• Andreas Tjärnberg
• Olof Rundquist

Thank you!