Multi-omics analysis powered by massive data integration

RE(ACT) congress
08-02-2018

Kristina Hettne, PhD
BioSemantics Group, Human Genetics
LEIDEN UNIVERSITY MEDICAL CENTER
Duchenne muscular dystrophy (DMD)

- X-linked
  - Mutations in dystrophin - in DMD patients dystrophin is virtually absent; Becker muscular dystrophy patients have 10% to 40% of the normal amount
  - Muscle atrophy, inflammation, fibrosis, fat deposits
- Incidence 1:5000, costs: $80.000 - $120.000 p.p/p.y

https://kin450-neurophysiology.wikispaces.com/Duchenne+Muscular+Dystrophy
Muscle biopsies are often requested in interventional clinical trials to show proof of concept that the drug is working e.g. dystrophin restoration after gene therapy, but muscle biopsies are invasive for the patient.
Can bio-signature detection in blood offer an alternative?

• Identification of blood biomarkers able to monitor disease progression and response to therapy would enable:
  • Better marketing authorization of medicinal products for DMD and muscular dystrophies in general
  • The analysis of blood should support drug developers to show the drug is working (or not working) without taking muscle biopsies
Gain insights into the effect of dystrophin deficiency on a molecular level

identification of novel biomarkers for disease severity and progression
Identify groups of genes that remain active and drive the pathology
Dystrophin-deficient *mdx* mice

- Bred at the LUMC
- DMD mouse model
  - Nonsense mutation *dystrophin*
- Capable of producing functional analog *Utrophin*
  - Less severe symptoms

© Olga Veth
Experimental design

- *mdx* mice:
  - Samples at 6, 12, 18, 24 and 30 weeks of age
  - Sacrifice at week 30
  - 4 groups, 5 - 6 mouse per group
  - C57BL or hybrid
Multi-omics data integration, networks and knowledge discovery

Problems:
- Different sizes of data sets
- Multiple testing burden

Solution:

Network-based integration

Knowledge graphs

Rachel Cavill et al. Brief Bioinform, Dec 2015: bib.bbv090
Approach *mdx* study

1. Longitudinal transcriptomics
2. Longitudinal Metabolomics
3. Longitudinal Lipidomics

   - Calculate area under the curve (AUC) measurements

   - AUC Transcriptomics
   - AUC Metabolomics
   - AUC Lipidomics

   - Calculate WGCNA modules

   - Transcriptomics modules
   - Metabolomics modules
   - Lipidomics modules

   - Correlate WGCNA modules

   - Multi-omics modules

AUC = area under the curve
WGCNA = weighted gene co-expression analysis

\[ \text{data} \quad \text{process} \]
1. Calculate co-expression values

2. Cluster co-expression values

3. Relate modules (lists of genes, metabolites or lipids) to genotype, phenotype, or clinical parameters

Langfelder and Horvath, 2008
Approach

Longitudinal transcriptomics → Calculate area under the curve (AUC) measurements → AUC Transcriptomics

Longitudinal Metabolomics → Calculate area under the curve (AUC) measurements → AUC Metabolomics

Longitudinal Lipidomics → Calculate area under the curve (AUC) measurements → AUC Lipidomics

Calculate WGCNA modules → Transcriptomics modules → Correlate WGCNA modules → Multi-omics modules

Calculate WGCNA modules → Metabolomics modules → Correlate WGCNA modules → Multi-omics modules

Calculate WGCNA modules → Lipidomics modules → Correlate WGCNA modules → Multi-omics modules

AUC = area under the curve
WGCNA = weighted gene co-expression analysis

=data  =process
Metabolomics modules
Lipidomics modules
• Does a module have any relation to the disease?
• Calculate module eigengenes
  • The first principal component of the standardized expression profiles
• Module eigengenes not always normally distributed
  • Calculate significance by comparing the group means of the module eigengene values using the parametric Kruskal-Wallis test
  • Significance ($p < 0.05$) indicates a general difference in expression of the module’s molecules between one or more group pairs (e.g. wt vs mdx)
All metabolomics modules significant, all but one lipidomics module, no transcriptomics modules

Example: metabolomics eigengene plots

\[ mdx^{+/+} = \text{mdx/utron}^{+/+} \]
\[ mdx^{+-} = \text{mdx/utron}^{-/-} \]
Approach

Longitudinal transcriptomics → Calculate area under the curve (AUC) measurements → AUC Transcriptomics

Longitudinal Metabolomics → AUC Metabolomics

Longitudinal Lipidomics → AUC Lipidomics

Calculate WGCNA modules

Transcriptomics modules → Correlate WGCNA modules

Metabolomics modules

Lipidomics modules

Multi-omics modules

AUC=area under the curve
WGCNA=weighted gene co-expression analysis

=process

=process
Cross-correlation between module eigengenes reveals multi-omic correlations

Module eigengene: the first principal component of the standardized expression profiles

Encircled numbers: significant
Red line: positive correlation
Blue line: negative correlation

© Mohammed Charrouf
Cross-correlation filtered for three-way omics

Red line: positive correlation
Blue line: negative correlation

© Mohammed Charrout
Cross-correlation filtered for three-way omics

Turquoise module
- Highly connected
- Significant correlations to other omics
  - 0.55 ($p=0.01$)
  - 0.48 ($p=0.04$)
  - 0.47 ($p=0.04$)
  - 0.48 ($p=0.04$)

Red line: positive correlation
Blue line: negative correlation

\[
\begin{align*}
mdx+/+ & = mdx/utrn+/+ \\
mdx+/- & = mdx/utrn+/-
\end{align*}
\]
Overlap with longitudinal proteomics in DMD patients

*Proteins changed over time in serum from DMD patients from Spitali et al. 2018. Under review.*
Physiologic functional annotation with Euretos


Phenotype <-> Pathways <-> Proteomics <-> Metabolomics <-> Genetics
Euretos physiologic function annotation reveals annotation overlap

cytokine signaling process
glucose homeostasis
biological adaptation to stress
cytoprotection
negative regulation of circadian rhythm
inhibition of circadian rhythm
muscle contraction
activation of phosphoinositide 3-kinase cascade
endocrine physiology
cardiac muscle hypertrophy in response to stress

Longitudinal genes mdx turquoise

neutrophil degranulation
heterophil degranulation
fibrinolysis
immunity
energy metabolism
heat-shock response
uptake
physiologic function

Longitudinal proteins DMD
Functional annotation overlap does not have to mean gene overlap
Euretos physiologic function annotation reveals muscle-related processes

cytokine signaling process
glucose homeostasis
biological adaptation to stress
cytoprotection
negative regulation of circadian rhythm
inhibition of circadian rhythm
**muscle contraction**
glucose tolerance
activation of phosphoinositol 3-kinase cascade
endocrine physiology
**cardiac muscle hypertrophy in response to stress**
Digging deeper with gene-disease analysis: SCN4A from muscle contraction process

Located for keywords: SCN4A (homo sapiens) AND muscular dystrophy, duchenne

Featured Result: Review if SCN4A (homo sapiens) plays a role in muscular dystrophy, duchenne

Nav1.4 deregulation in dystrophic skeletal muscle leads to Na+ overload and enhanced cell death.

2008 08-01 Duchenne muscular dystrophy (DMD) is a hereditary degenerative disease manifested by the absence of dystrophin, a structural, cytoskeletal protein, leading to muscle degeneration a ...

PMID: 18625851 Source: Authors: Hirn, Carole (C); Shapovalov, George (G); Petermann, Olivier (O)

Role of SCN4A (homo sapiens) in muscular dystrophy, duchenne

SCN4A (homo sapiens) has the following RNA and protein expression patterns in healthy tissues. The top 3 tissues associated with muscular dystrophy, duchenne are highlighted. See the information panel above for more details.

RNA Expression (FPKM)

GTEx

- skeletal muscle structure
- visceral fat
- breast
- subcutaneous fat
- thyroid gland
- testis
- pituitary gland

Human Protein Atlas

- skeletal muscle structure
- adipose tissue
- thyroid gland
- testis

Illumina Body map

- skeletal muscle structure
- adipose tissue
- breast
- thyroid gland
- colon structure (body str.)

Fantom 5
• Disease-related biomolecular modules were found in mouse blood metabolomics and lipidomics data
• Cross-correlation between module eigengenes revealed multi-omic correlations
• Significantly correlated turquoise transcriptomics module was related to cytokine production, glucose homeostasis and muscle contraction

• In conclusion:
  • Multi-omics network analysis reveals a muscular dystrophy-related signature in mouse blood
Data management

- When published: data in public repositories, scripts on GitHub
- Now:
  - Scripts on LUMC GitLab
  - Metadata and result data files on FAIR Data Point
  - WGCNA web tool by Mohammed Charrout creates report with parameters used on GitHub
    - https://github.com/mochar/wgcna
Next steps

- Drug repurposing
  - In progress, integrative pipeline to match targets with drugs in place
- Uptake in the RD-Connect platform
  - Multi-omics analysis pipeline under construction
  - Will incorporate lessons learned from the past years multi-omics analyses:
    - **Huntington's Disease** (transcriptomics, metabolomics, lipidomics; mouse [1] and human [2])
    - **Duchenne muscular dystrophy** (proteomics, transcriptomics, metabolomics, lipidomics; mouse [3] and human [4])
    - **SCA3** (transcriptomics, metabolomics, lipidomics; mouse [5])
    - **Beta-thalassemia** (transcriptomics, proteomics, human [6])

Thank you

BioSemantics/Bioinformatics (LUMC)
Mohammed Charrout, Eleni Mina, Marco Roos, Mark Thompson, Kees Burger, Rajaram Kaliyaperumal, Barend Mons, Annika Jacobson
Peter A.C. 't Hoen (since Feb 2018 Radboud UMC)

DMD exon skip group (LUMC)
Pietro Spitali, Annemieke Aartsma-Rus

Medical Statistics (LUMC)
Roula Tsonaka

Profilomic/MedDay
Alexandre Seyer