R & the Study of Rare Diseases: Using Government Databases & Molecular Datasets to Set Research Priorities

Simina Boca, PhD, MHS

@siminaboca

smb310@georgetown.edu

www.github.edu/SiminaB

Associate Professor

Innovation Center for Biomedical Informatics, Georgetown University

The Government and Public Sector R Conference

Washington, DC

December 4, 2020

Rare diseases

- A rare disease is often defined as a disorder that affects fewer than 200,000 individuals in the United States.
- Over 70% of rare diseases are genetic.
  - Of those, 70% start in childhood and often lead to a substantially reduced life expectancy.
  - Eg: cystic fibrosis, sickle cell anemia, muscular dystrophies, Huntington's disease, some inherited cancer syndromes
- Together, rare diseases affect 25-30 million individuals in the United States.
- What are some good data resources and research approaches for combating rare diseases?

Zebra ribbon: https://commons.wikimedia.org/wiki/File:Zebra_ribbon.svg

(David Richfield)

“When you hear hoofbeats behind you, don't expect to see a zebra”

Theodore Woodward (1914-2005)
Researcher at University of Maryland

Challenges with clinical trials in rare diseases

- It can be very challenging to perform clinical trials in this space:
  - Small number of individuals affected by any one disease
  - Often progressive nature: How do you choose which outcomes to consider?
  - Ethical and implementational challenges with performing research on children.
- The FDA allows for specific flexibilities when evaluating new drugs for the treatment of rare diseases:
  - Eg by allowing the use of biomarkers as surrogate endpoints in some instances.

Zebra ribbon: https://commons.wikimedia.org/wiki/File:Zebra_ribbon.svg

(David Richfield)

Duchenne Muscular Dystrophy (DMD)

- Devastating X-linked single-gene disorder.
  - Mutations are in the *DMD* gene.
  - Different mutations may lead to different phenotypes.
  - Certain therapies target only individuals with specific changes.
- Affects about 1 in 5,000 newborn males.
- Leads to muscle loss and eventual loss of ambulation and death in late teens to early 20s or 30s, generally due to cardiac or respiratory problems.
- Steroid treatment can prolong ambulation but chronic use can cause many problems.

Understanding the clinical trial landscape for DMD

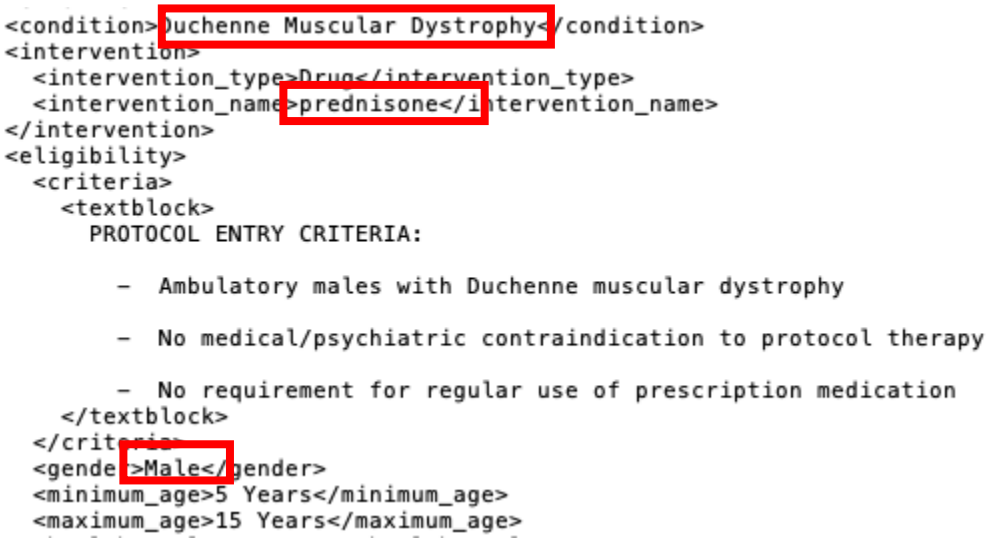
- A few therapies are approved or conditionally approved in US for a subset of individuals, with others in clinical trials.
- Many different trials for DMD patients:
  - Multiple trials are competing over a small number of patients.
  - Other patients may not be served at all (older patients, rarer mutations)

-> Need to assess what trials are performed in DMD space.

- Eventual goals:
  - Develop a product to allow researchers, clinicians, and patients to stay up-to-date with ongoing drug development in the DMD.
  - Prioritize research focusing on individuals who do not currently have many available clinical trial options.

Understanding the clinical trial landscape for DMD

- Summer student Bin Hai (2019) downloaded xml files for DMD-related clinical trials from https://clinicaltrials.gov/, the government database of registered clinical trials.
- Used XML package to extract information from xml files:

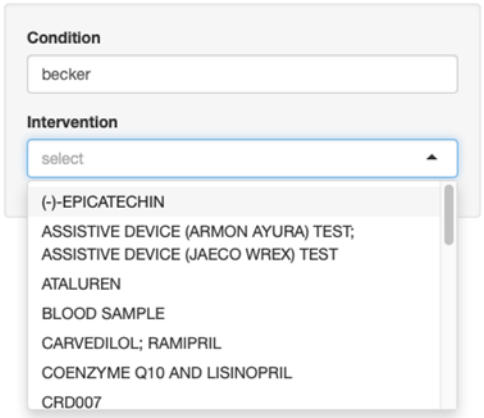


```
<condition>Duchenne Muscular Dystrophy</condition>
<intervention>
  <intervention_type>Drug</intervention_type>
  <intervention_name>prednisone</intervention_name>
</intervention>
<eligibility>
  <criteria>
    <textblock>
      PROTOCOL ENTRY CRITERIA:

        -  Ambulatory males with Duchenne muscular dystrophy

        -  No medical/psychiatric contraindication to protocol therapy

        -  No requirement for regular use of prescription medication
    </textblock>
  </criteria>
  <gender>Male</gender>
  <minimum_age>5 Years</minimum_age>
  <maximum_age>15 Years</maximum_age>
```

- Performed data cleaning, including for redundant drug names

Understanding the clinical trial landscape for DMD

- Used R/shiny to design and created a web interface to explore DMD-related clinical trials https://bh658.shinyapps.io/curation_dmd_clinicaltrials/



Total trials: 36

Gender	count
All	18
Female	1
Male	17

Allocation	count
Non-Randomized	4
Randomized	10
NA	22

Study Rank	NCT ID	Title	Allocation	Condition	Intervention Type	Intervention
213	NCT00005574	Gentamicin Treatment of Muscular Dystrophy	NA	BECKER MUSCULAR DYSTROPHY; DUCHENNE MUSCULAR DYSTROPHY	Drug	GENTAMICIN
254	NCT00104078	Study Evaluating MYO-029 in Adult Muscular Dystrophy	Randomized	BECKER MUSCULAR DYSTROPHY; FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY; LIMB-GIRDLE MUSCULAR DYSTROPHY	Drug	STAMULUMAB
231	NCT00390104	Molecular Analysis of Patients With Neuromuscular Disease	NA	BECKER MUSCULAR DYSTROPHY; DUCHENNE MUSCULAR DYSTROPHY; LIMB-GIRDLE MUSCULAR DYSTROPHY; NEUROMUSCULAR; DISORDER, HEREDITARY	NA	NA
134	NCT00592553	Phase 2b Study of PTC124 in	Randomized	BECKER MUSCULAR DYSTROPHY; DUCHENNE	Drug	ATALUREN

Searching for biomarkers of disease progression and drug response

- Given the difficulty with choosing appropriate outcomes, there has been a focus on finding potential biomarkers of disease progression and drug response.
- We considered metabolomics – the study of metabolites (small biomarkers):
  - Can be detected in various biofluids and tissues.
  - Have various biological functions.
  - Can be influenced by genetics and environmental factors.

Boca SM et al. "Discovery of metabolic biomarkers for Duchenne Muscular Dystrophy within a natural history study."
PLOS ONE, 2016. https://doi.org/10.1371/journal.pone.0153461

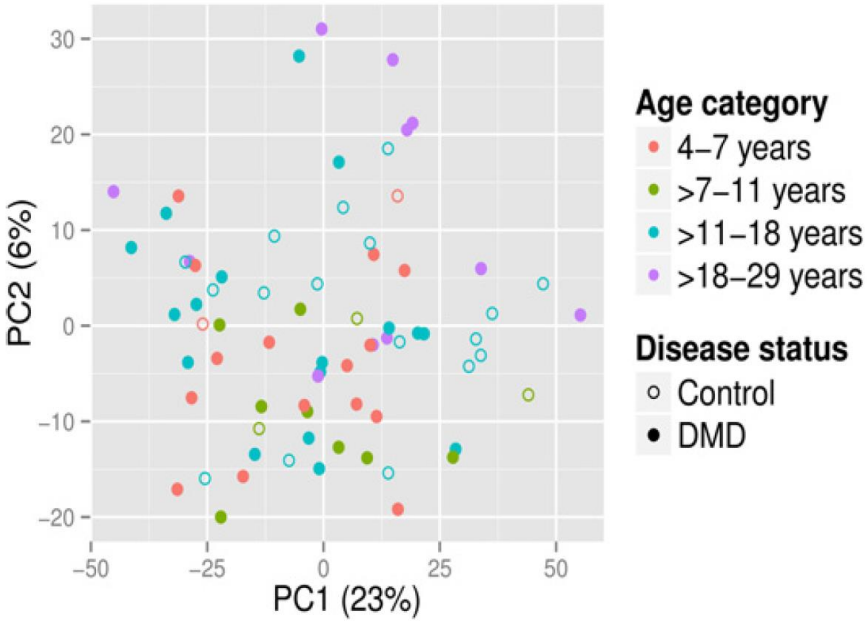
Thangarajh M et al. "Discovery of potential urine-accessible metabolite biomarkers associated with muscle disease and corticosteroid response in the mdx mouse model for Duchenne." *PLOS ONE*, 2019. https://doi.org/10.1371/journal.pone.0219507

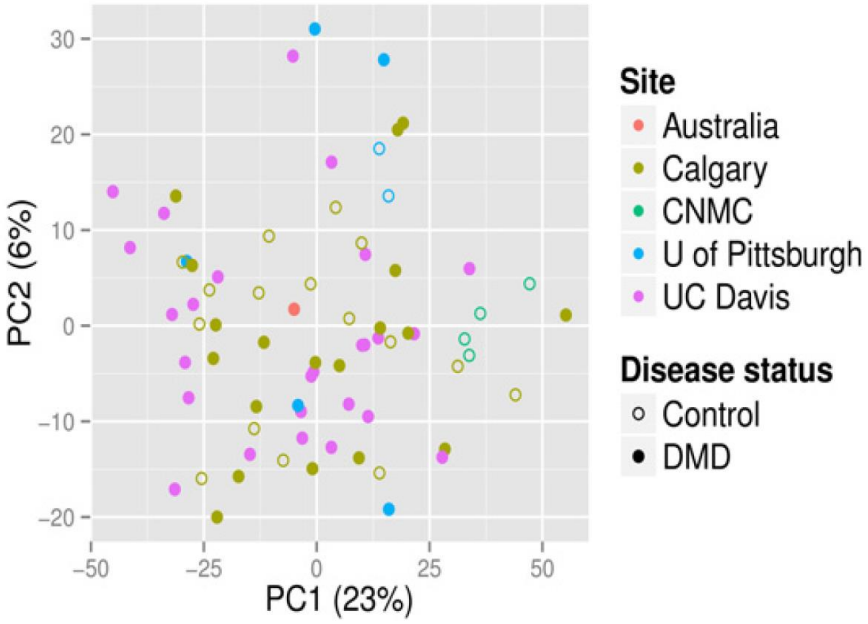
Overview of study

- Use untargeted metabolomics from serum samples of 51 DMD patients and 22 healthy controls.
- A total of 2,203 metabolites detected.

Study site	DMD patients	Healthy controls
Alberta Children's Hospital (Calgary)	19	16
University of California, Davis (UC Davis)	26	0
University of Pittsburgh / Children's Hospital of Pittsburgh of UPMC (U of Pittsburgh)	5	2
Children's National Medical Center (CNMC)	0	4
Children's Hospital at Westmead (Australia)	1	0
Median age in years (minimum, maximum)	11.4 (4, 28.7)	13.7 (6, 17.8)
Total by age category		
4–7 years	15	2
> 7–11 years	8	3
> 11–18 years	17	17
> 18–29 years	11	0
Total	51	22

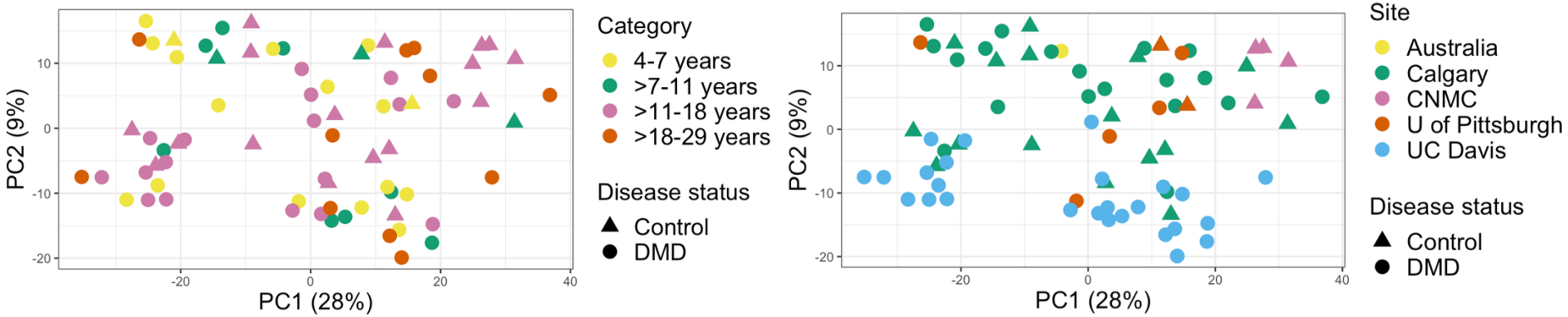
Principal components analysis





Principal components analysis: What I've learned since!

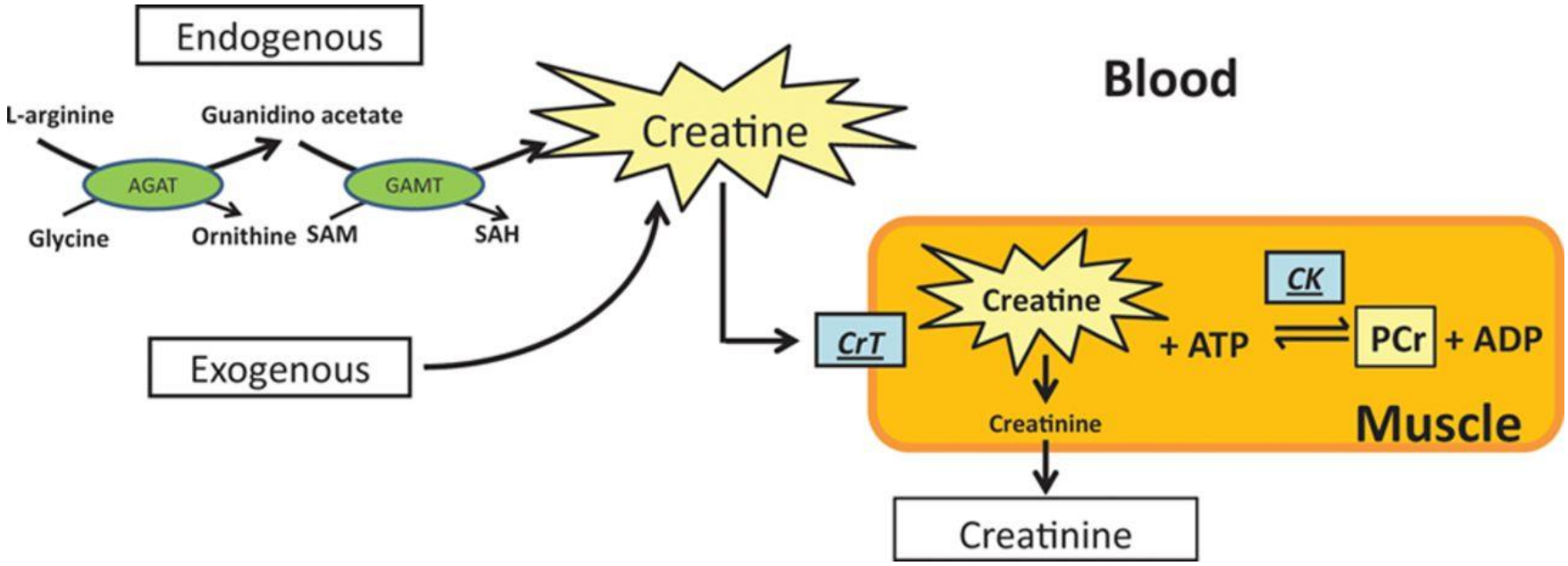
- Use colorblind-friendly palette!
- Don't scale and use same distances on both axes!
  - For example, see Nguyen and Holmes. "Ten quick tips for effective dimensionality reduction." *PLOS Computational Biology*, 2019. https://doi.org/10.1371/journal.pcbi.1006907



Metabolic biomarkers in DMD

- Fourteen metabolites were found significantly altered (FDR of 1%) between DMD patients and healthy controls.
  - Models adjusted for age and site, considered age x disease interaction
- Metabolites validated using MS/MS:
  - Higher in cases:
    - Creatine, L-arginine (likely validated)
  - Higher in controls:
    - Creatinine, 5a-DHT, testosterone sulfate

Creatine pathway and DMD

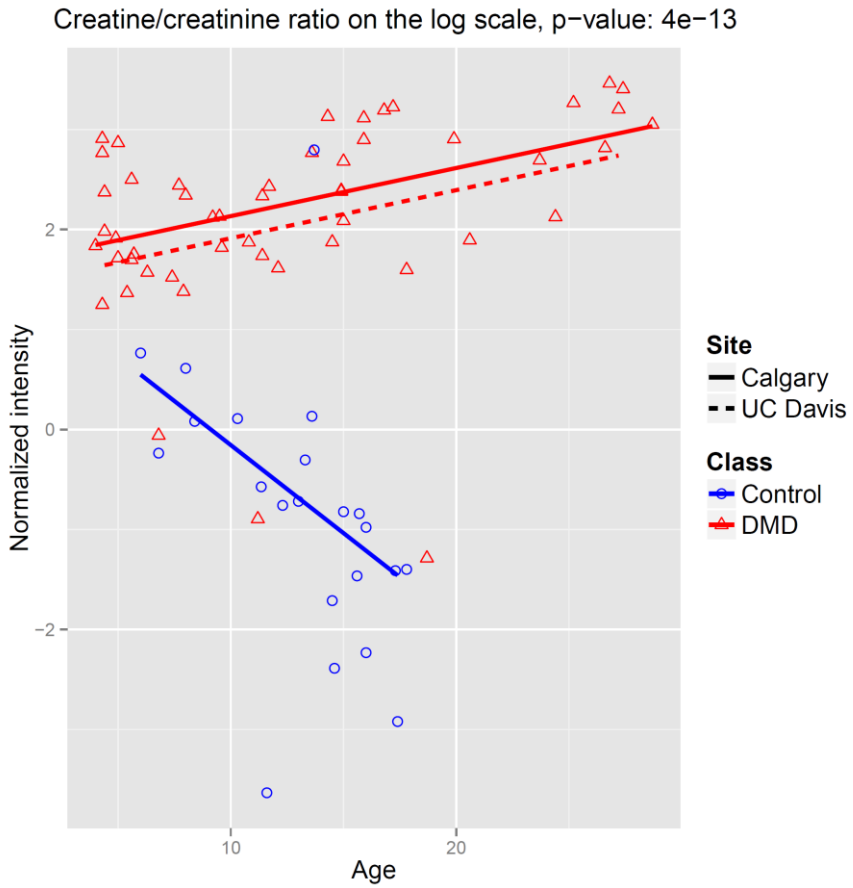


CK = creatine kinase

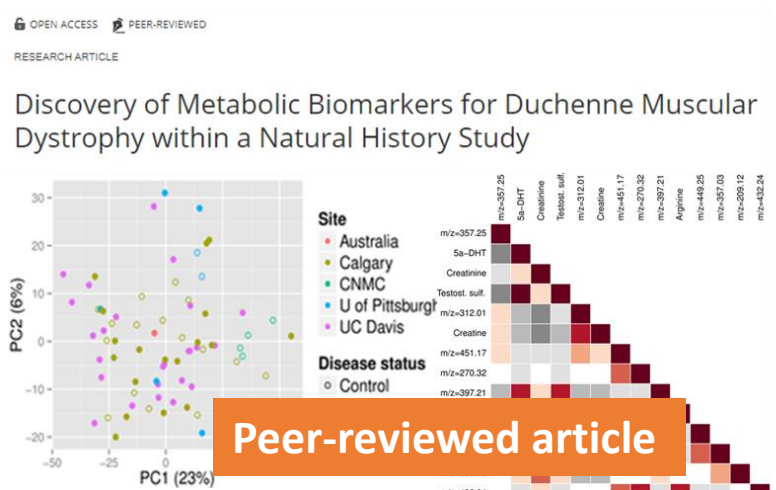
From: https://www.ahajournals.org/doi/full/10.1161/CIRCRESAHA.113.300974

Metabolic biomarkers in DMD

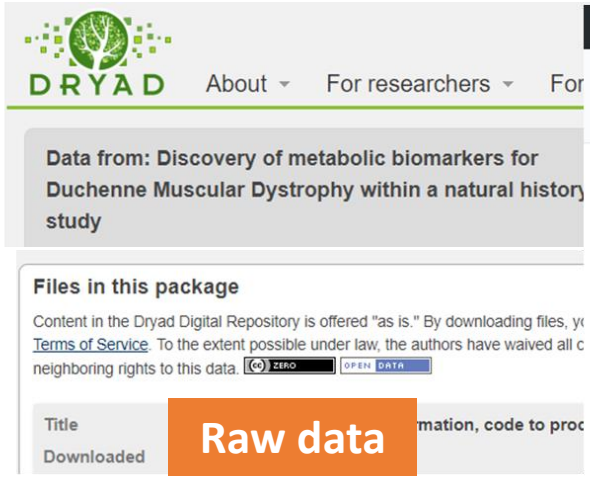
- Also considered the ratio between creatine and creatinine as a possible biomarker:



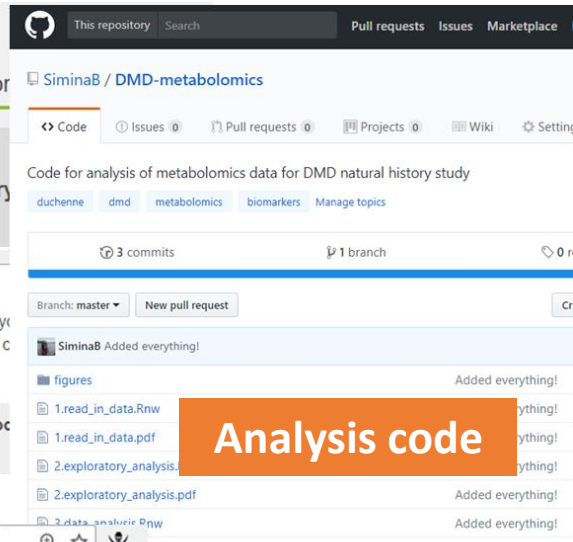
Study Ecosystem: Beyond the paper



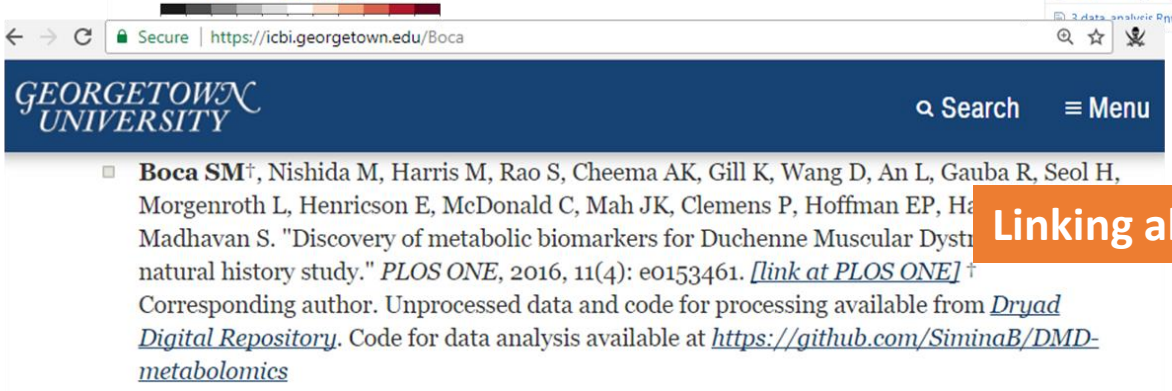
Peer-reviewed article



Raw data



Analysis code



Linking all components

Reproducibility: Powered by

Conclusions

- The 2016 study was the first comprehensive metabolomic study for DMD.
  - Represents one of the first steps towards finding metabolic surrogate biomarkers of disease progression.
  - Adds to the list of possible non-invasive blood circulating biomarkers.
- The 2020 study provided additional biomarkers, which may be related to either the mechanisms of muscle injury in DMD or prednisolone treatment.
- Future goals include:
  - Embedding these types of studies into clinical trials!
  - Connecting this type of study with genetic data and focusing it on the areas of greatest need.

Acknowledgments

Maki Nishida
Michael Harris
Shruti Rao
Subha Madhavan
Bin Hai
Bohyung Yoon
Xintong Liu

Amrita Cheena
Kirandeep Gill
Habtom Ressom
Zhenzhi Li
Rency Varghese

Mathula Thangarajh
Aiping Zhang
Haeri Seol
Lauren Morgenroth
Erik Henricson
Craig McDonald
Jean Mah
Paula Clemens
Eric Hoffman
Yetrib Hathout
Kanneboyina Nagaraju

- P30CA051008 (Proteomics and Metabolomics Core at Lombardi), R01AR062380 (PI: McDonald), P50AR060836 (PI: Clemens), American Academy of Neurology/American Brain Foundation Clinical Research Training Fellowship (PI: Thangarajh), MDA353094 (PI: Hathout)
- Cooperative International Neuromuscular Research Group (CINRG)
- Participants and families

GEORGETOWN UNIVERSITY

Georgetown University's

MS in Health Informatics and Data Science

https://healthinformatics.georgetown.edu/

R-Ladies

Washington DC

https://twitter.com/RLadiesGlobal

https://twitter.com/RLadiesDC