Biomedical Data Science:
Mining and Modeling

Biosensors and Wearables

Jason Liu

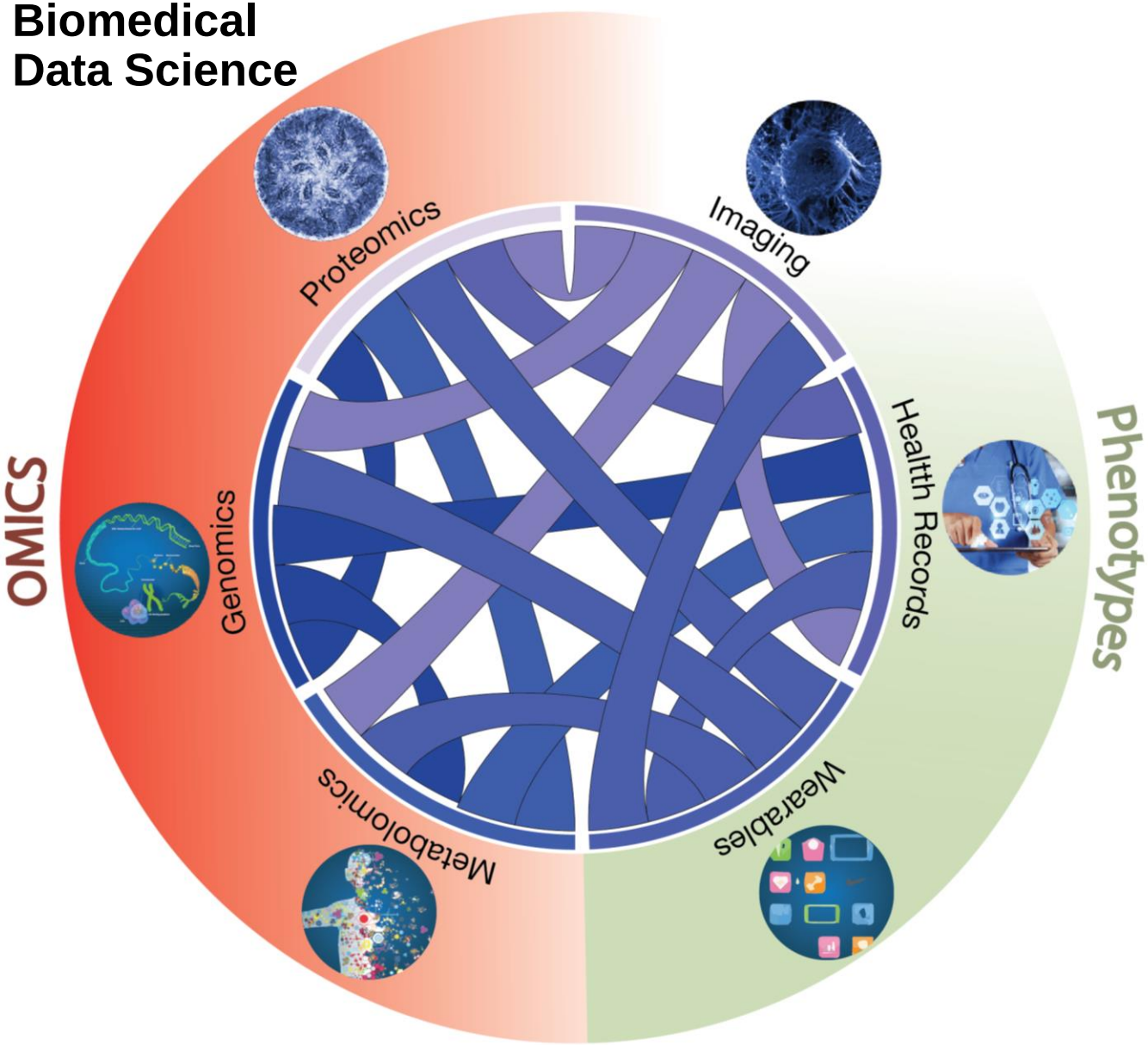
Lecture for CBB 752

Biomedical Data Science: Mining and Modeling

Spring 2023

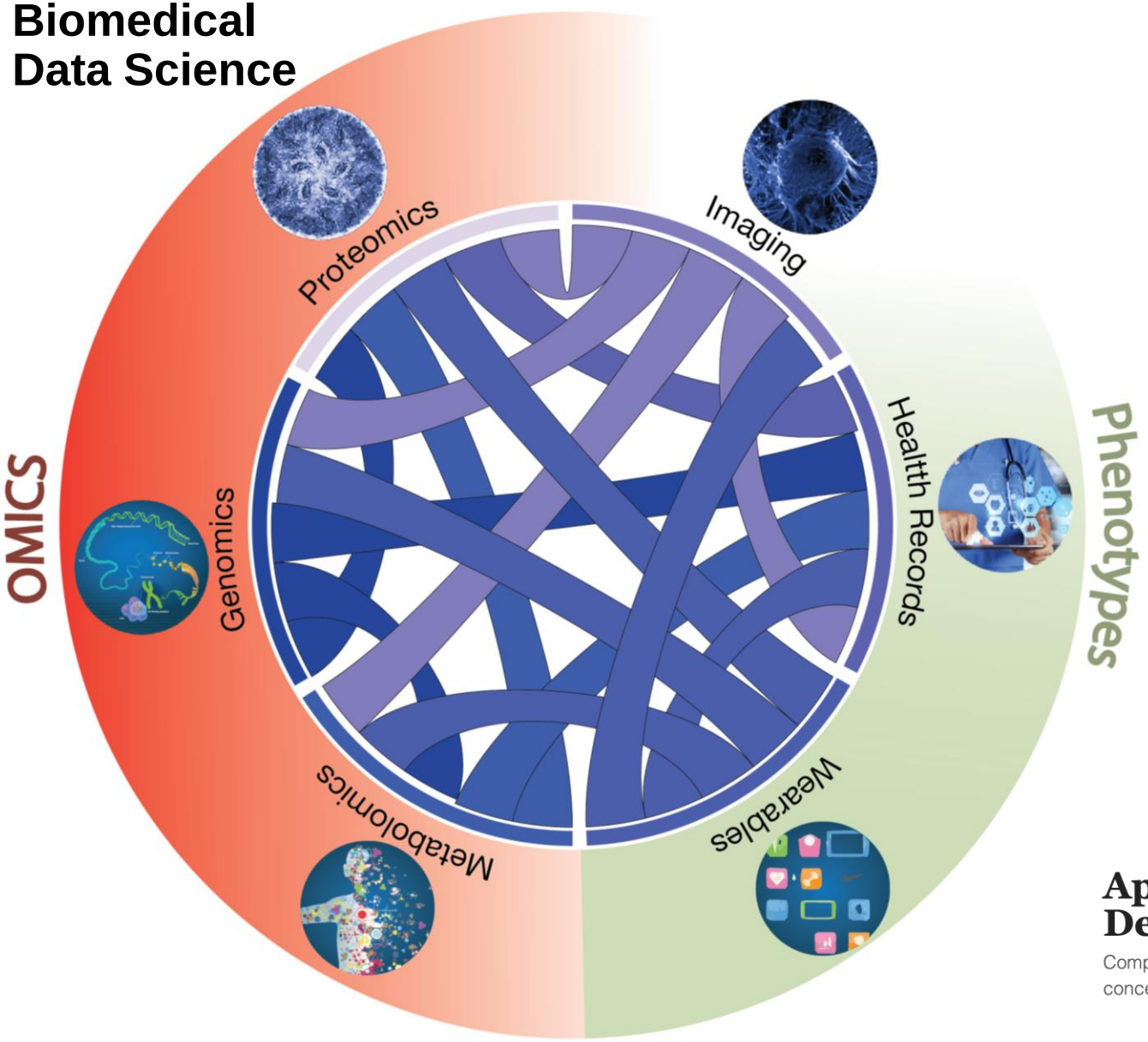
Yale University

Data Drivers for Biomedical Data Science



[Navarro et al. Genome Biol. (’19)]

Data Drivers for Biomedical Data Science



[Navarro et al. Genome Biol. ('19)]

The New York Times — July 7, 2021

Fitbits Detect Lasting Changes After Covid-19

Some people recovering from a coronavirus infection had an elevated heart rate for months, according to a new study.

Thomas R. Insel

PERSPECTIVES

Digital phenotyping: a global tool for psychiatry

2018

VIEWPOINT

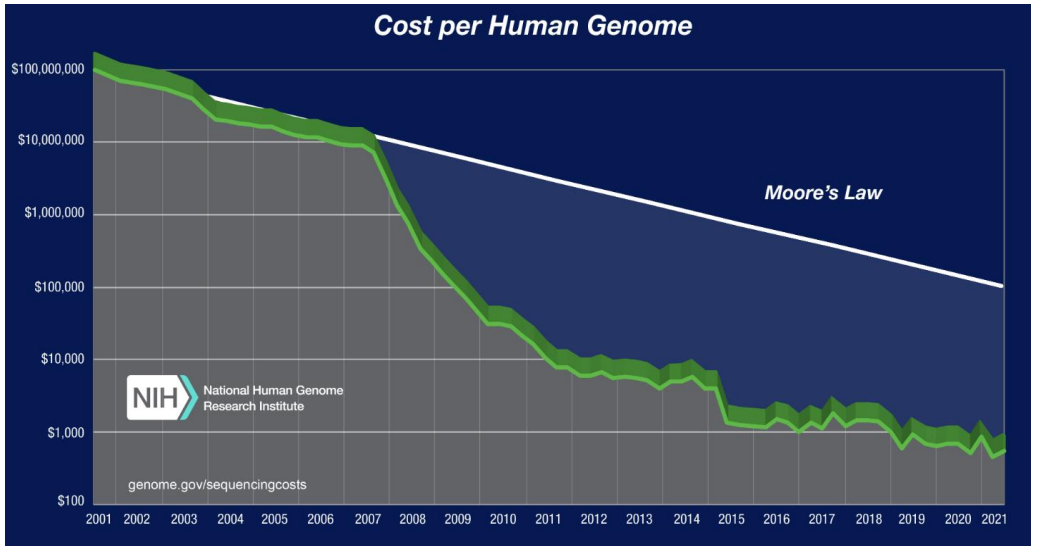
Digital Phenotyping
Technology for a New Science of Behavior

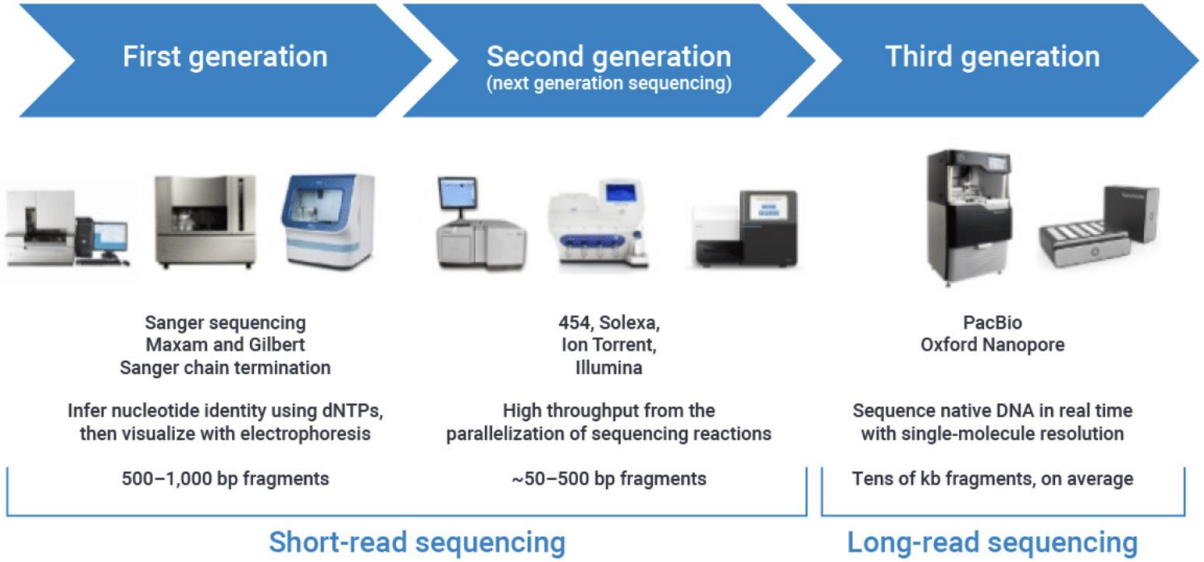
THE WALL STREET JOURNAL. — Sep 21, 2021

English Edition ▾ | Print Edition | Video | Audio | Latest Headlines | More ▾

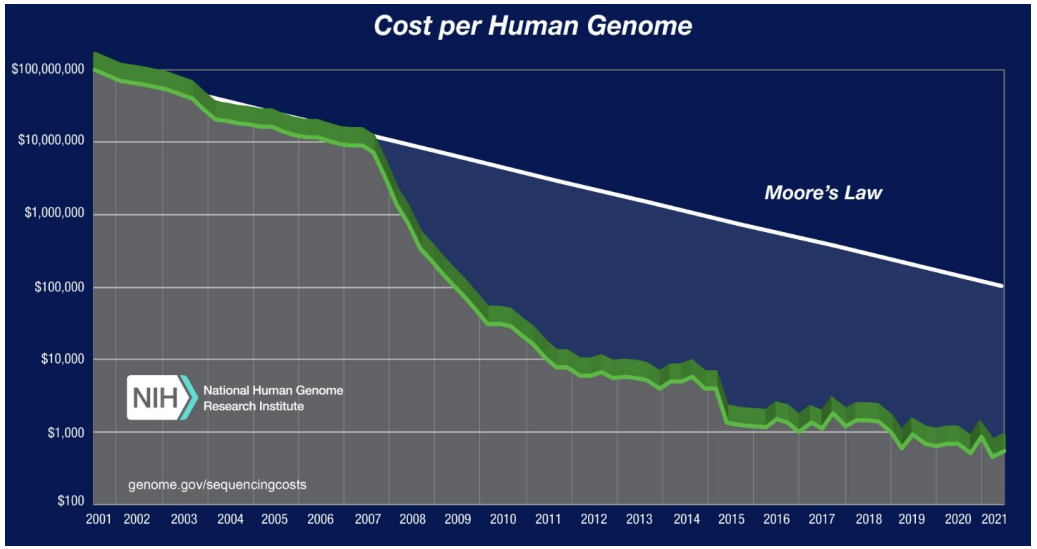
Apple Is Working on iPhone Features to Help Detect Depression, Cognitive Decline

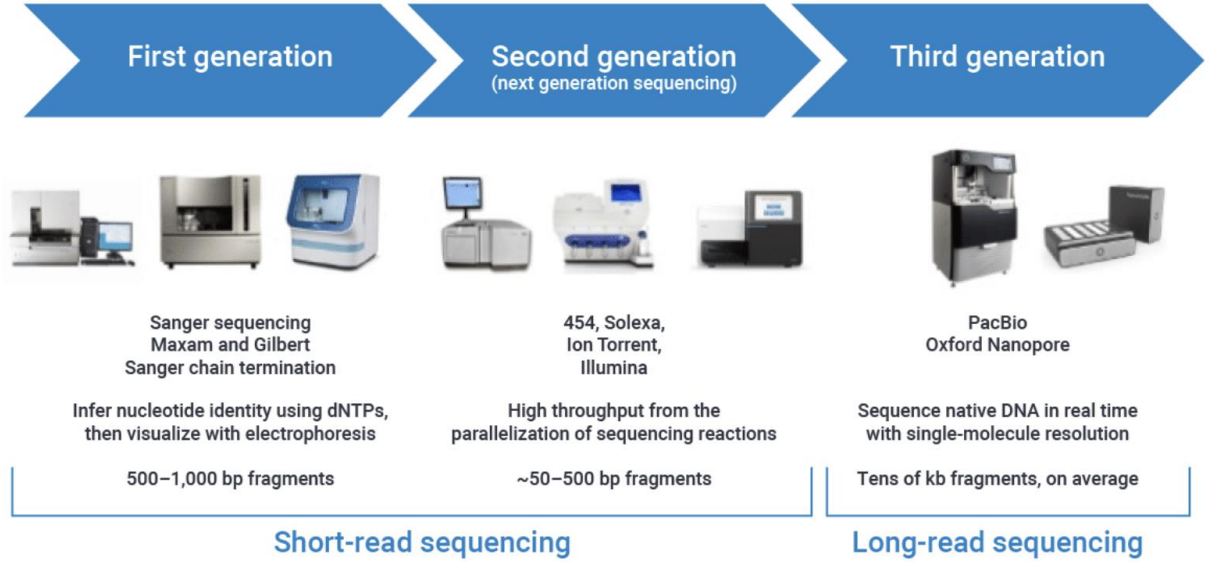
Company is working with UCLA, Biogen to see if sensitive data like facial expressions, typing metrics could signal mental-health concerns

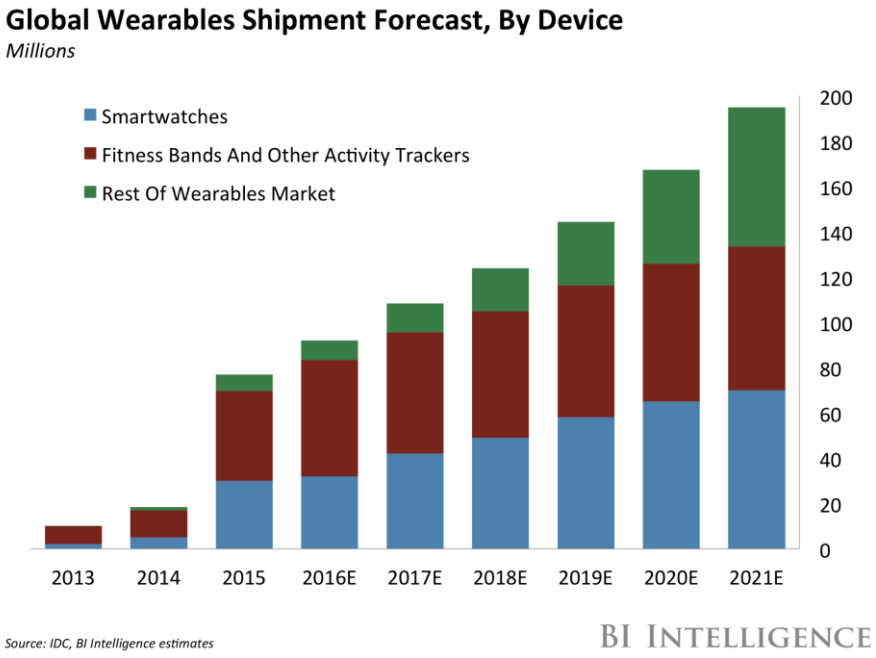




[Adapted from NIH and PacBio websites]







1982

Manpo-Meter (Pedometer)

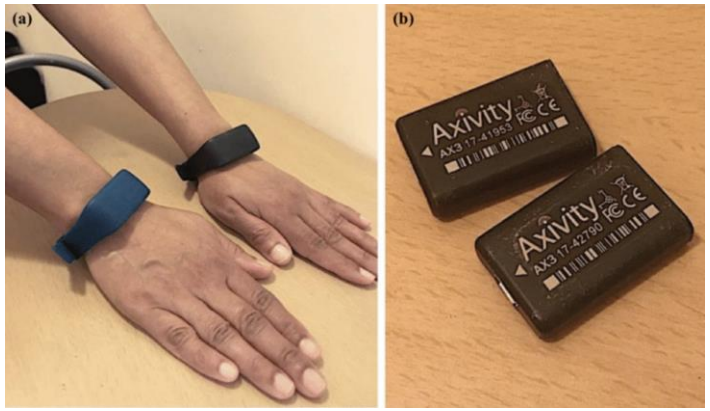
Polar Sport Tester PE 2000

Garmin

Apple

FitBit

biobank uk
Enabling scientific discoveries that improve human health



Present

[Adapted from NIH, PacBio, BI Intelligence, UKBB websites]

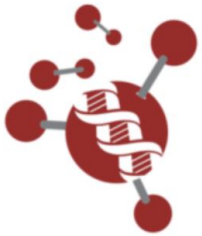
biobank uk

Enabling scientific discoveries that improve human health

NIH National Institute of Mental Health

Adolescent Brain Cognitive Development®

Teen Brains. Today's Science. Brighter Future.



iPOP

INTEGRATIVE PERSONAL OMICS PROFILING

mindstrong

Measurement in mental health.

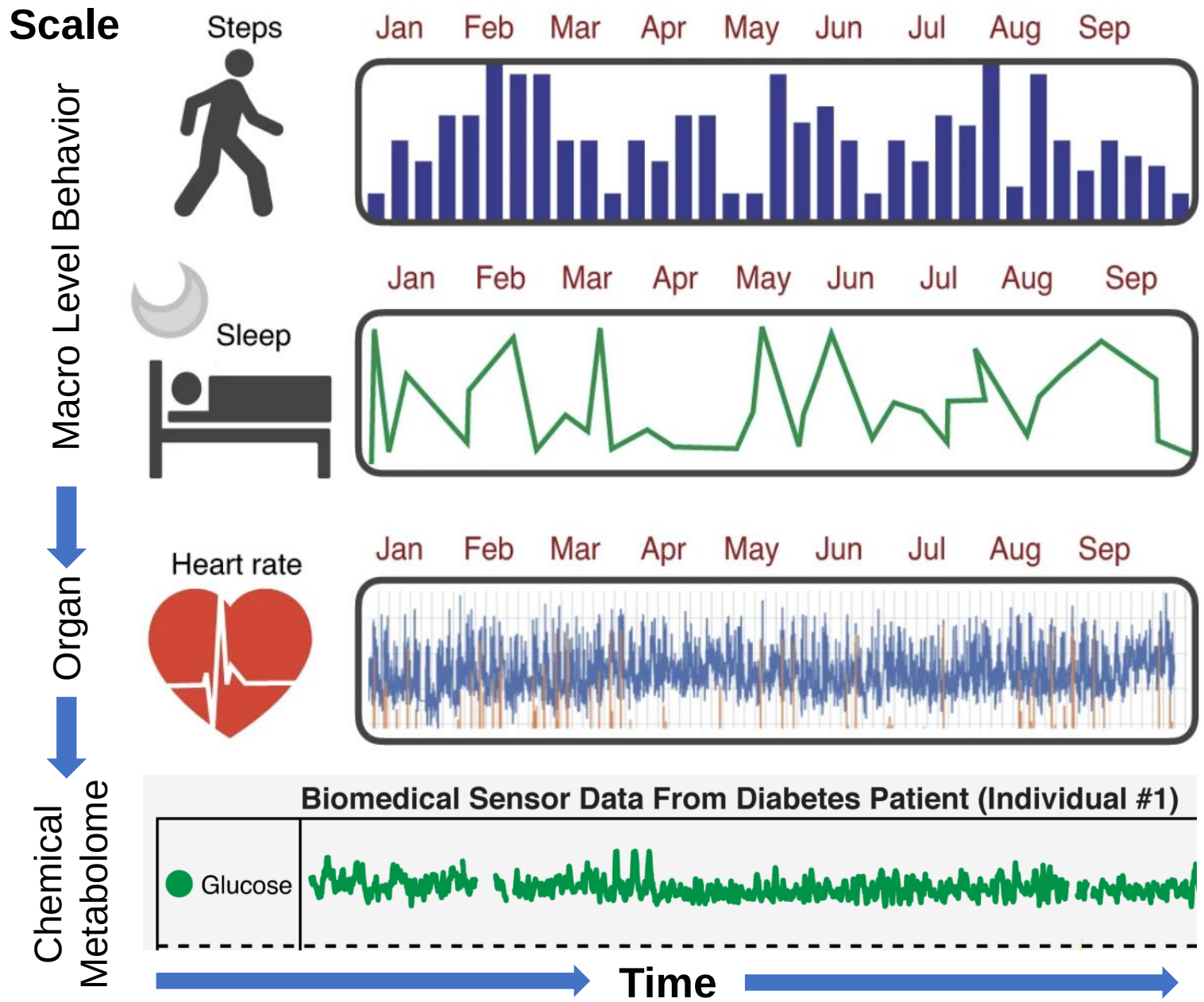
biobank uk
Enabling scientific discoveries that improve human health

NIH National Institute of Mental Health

Adolescent Brain Cognitive Development®
Teen Brains. Today's Science. Brighter Future.

IPOP
INTEGRATIVE PERSONAL OMICS PROFILING

mindstrong
Measurement in mental health.

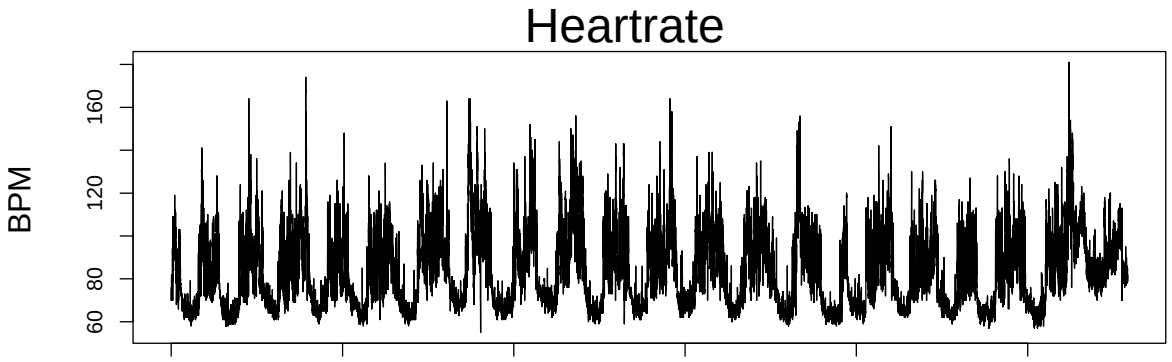


[Adapted from Mishra et al. Nat. BME ('20); Liu et al. PLOS CB ('21)]

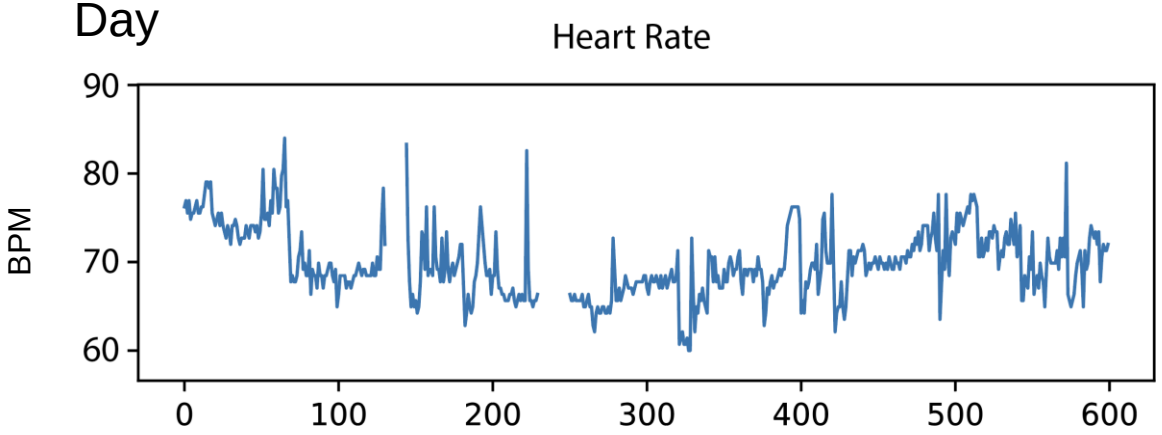
- Data and Signal Processing
- Feature Engineering
  - Can we extract meaningful information?
- Modeling
  - Can we address real world clinical and biological questions?
  - Can we refine phenotype-to-genotype linkages?

Biosensor and Wearable Time Series Strategies for Signal Processing

Weeks



Day

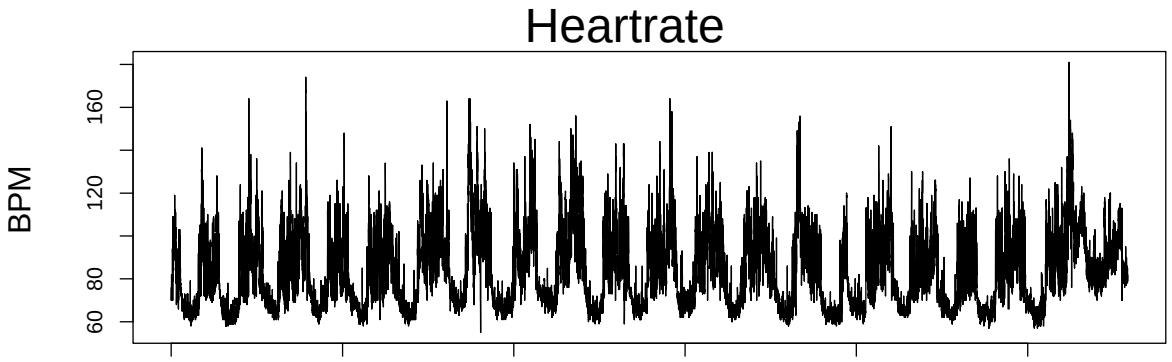


As we zoom in, challenges begin to arise

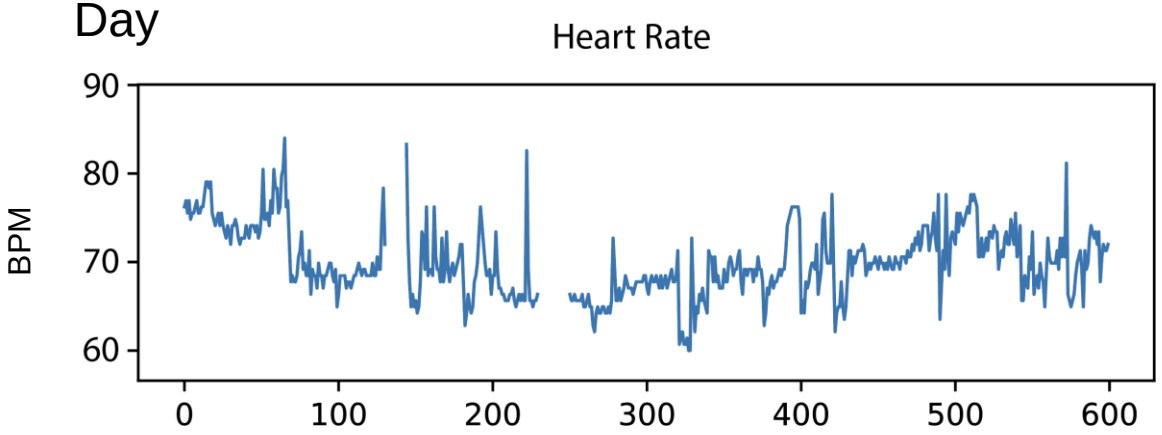
[NIMH ABCD Wearable Time Series Example]

Biosensor and Wearable Time Series Strategies for Signal Processing

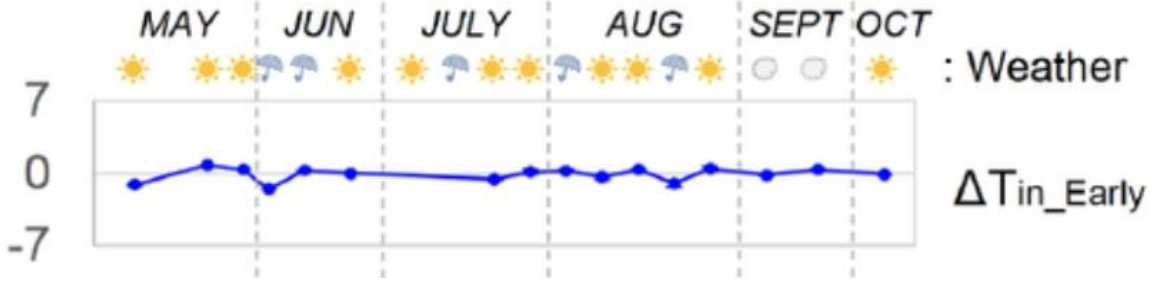
Weeks



Day



Further complicated due to other data streams

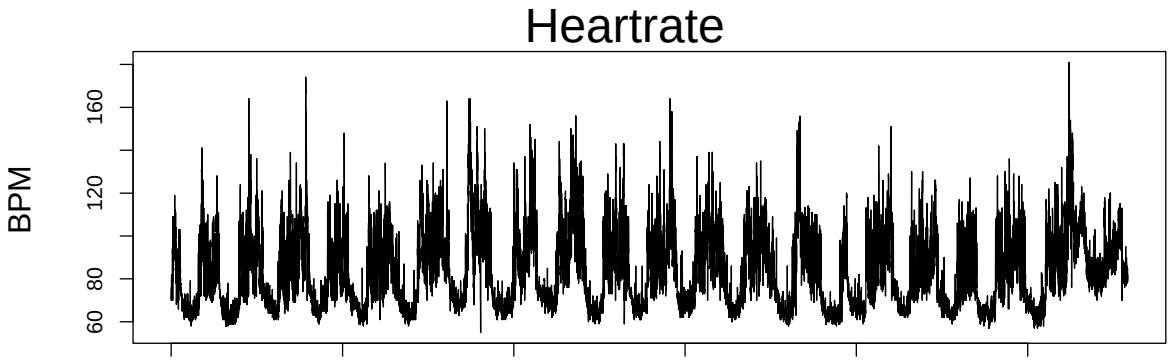


Covariates

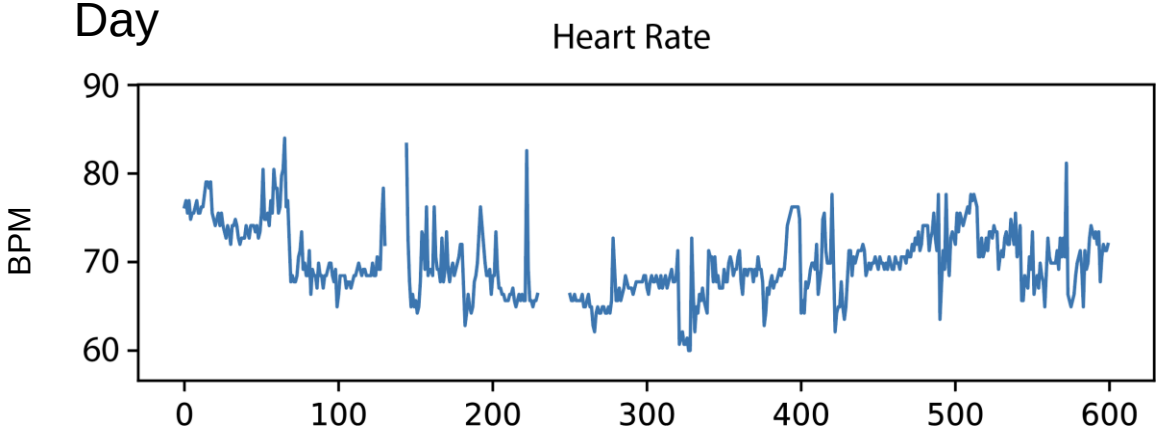
[NIMH ABCD Wearable Time Series Example]

Biosensor and Wearable Time Series Strategies for Signal Processing

Weeks

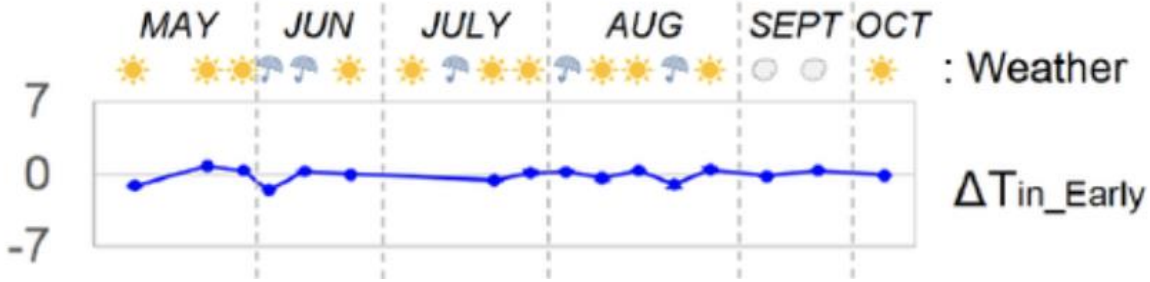


Day



Challenges

1. Missingness
2. Alignment/Sampling
3. De-noising

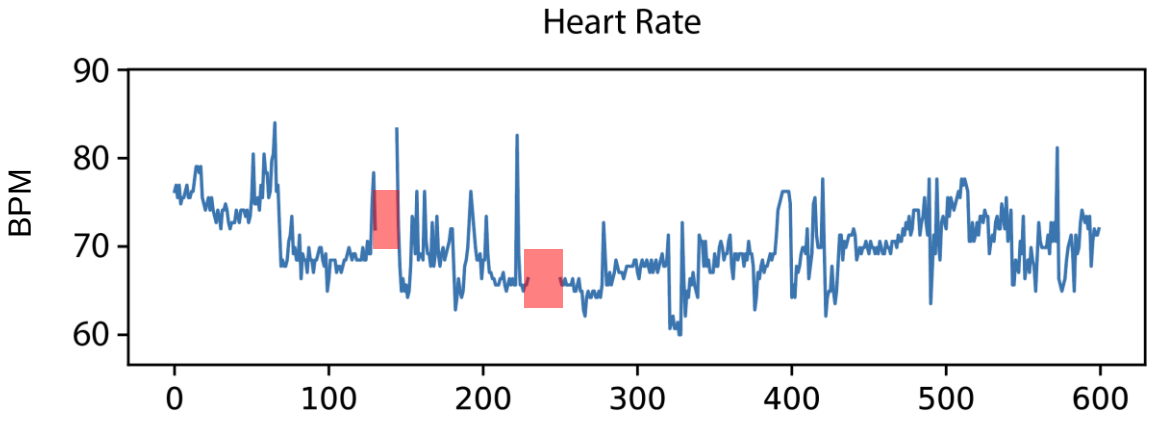


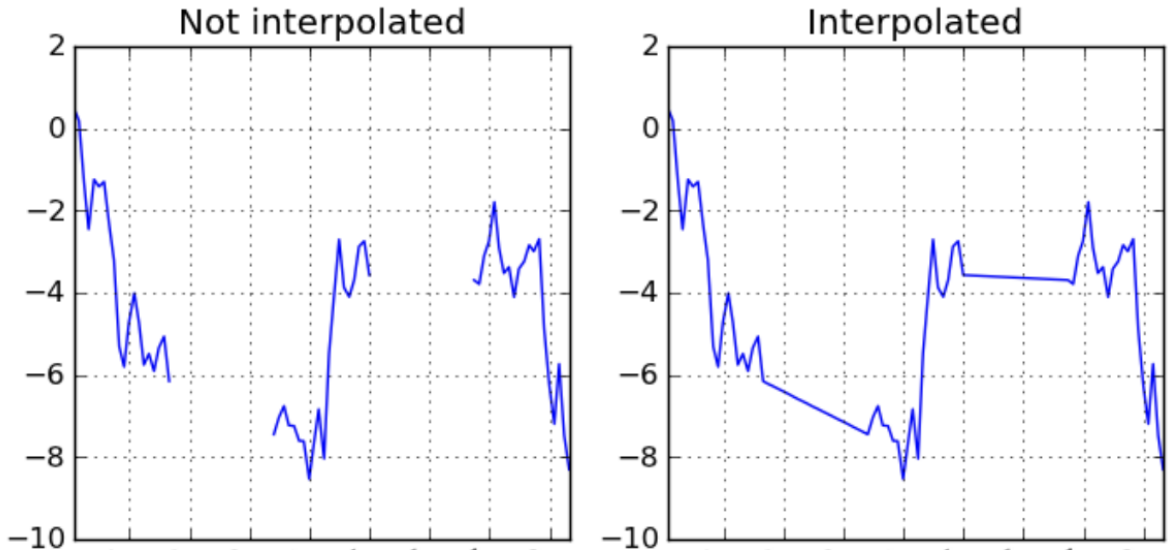
Covariates

[NIMH ABCD Wearable Time Series Example]

Processing the Data: Missingness

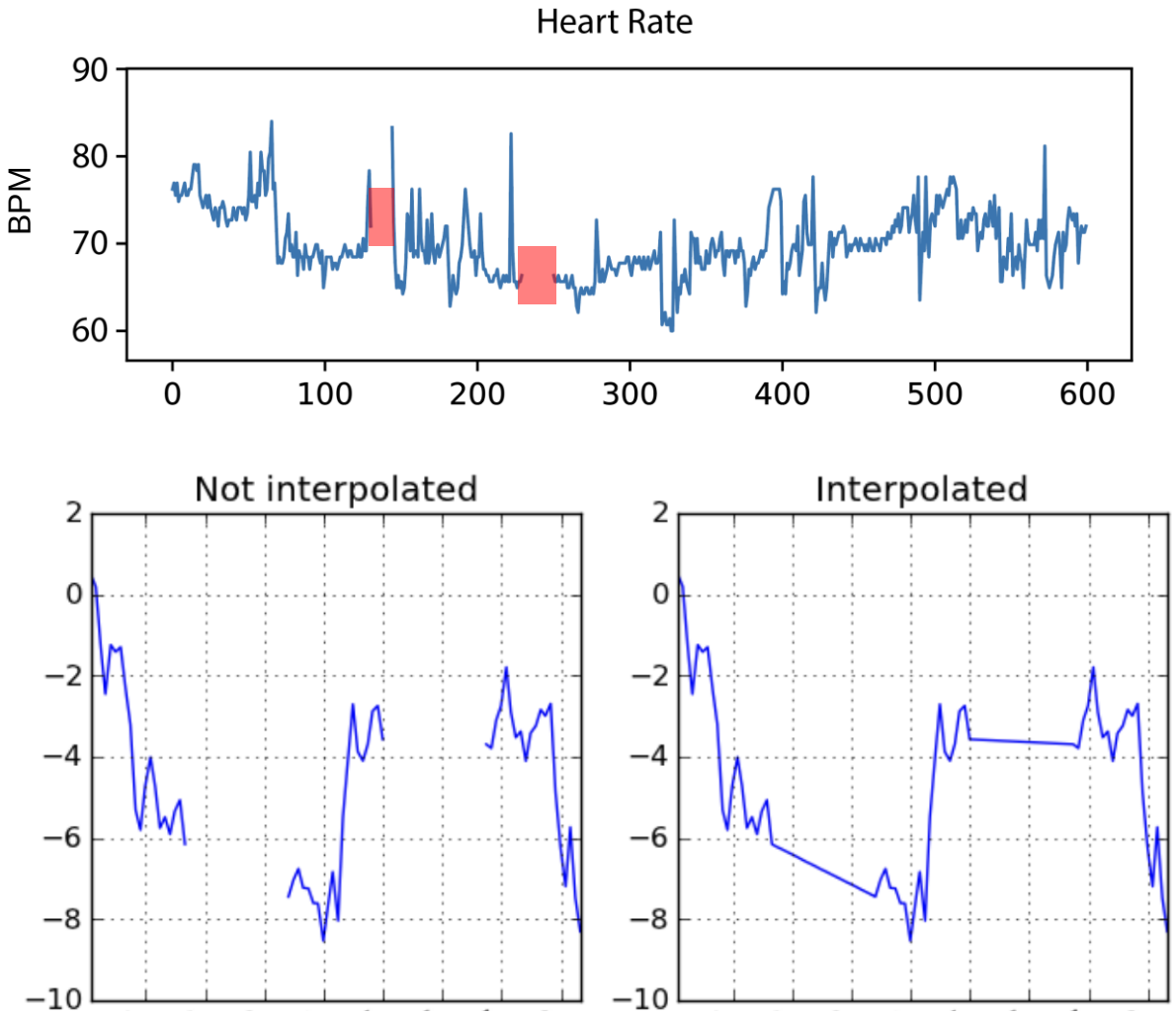
Example 1



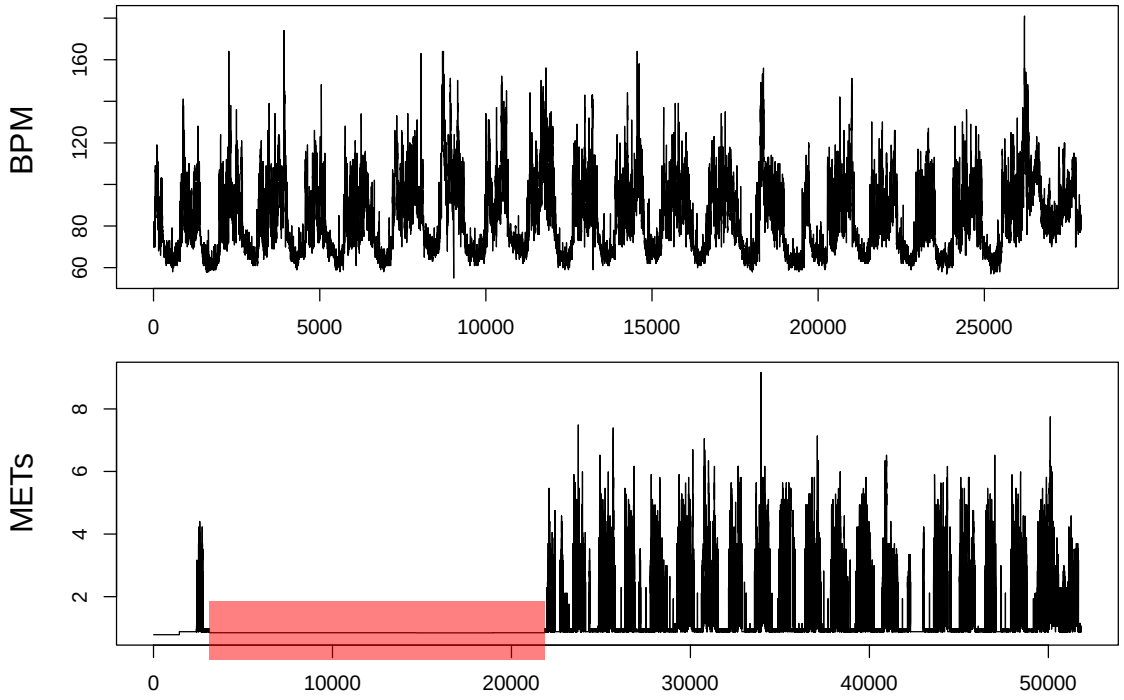


Processing the Data: Missingness

Example 1

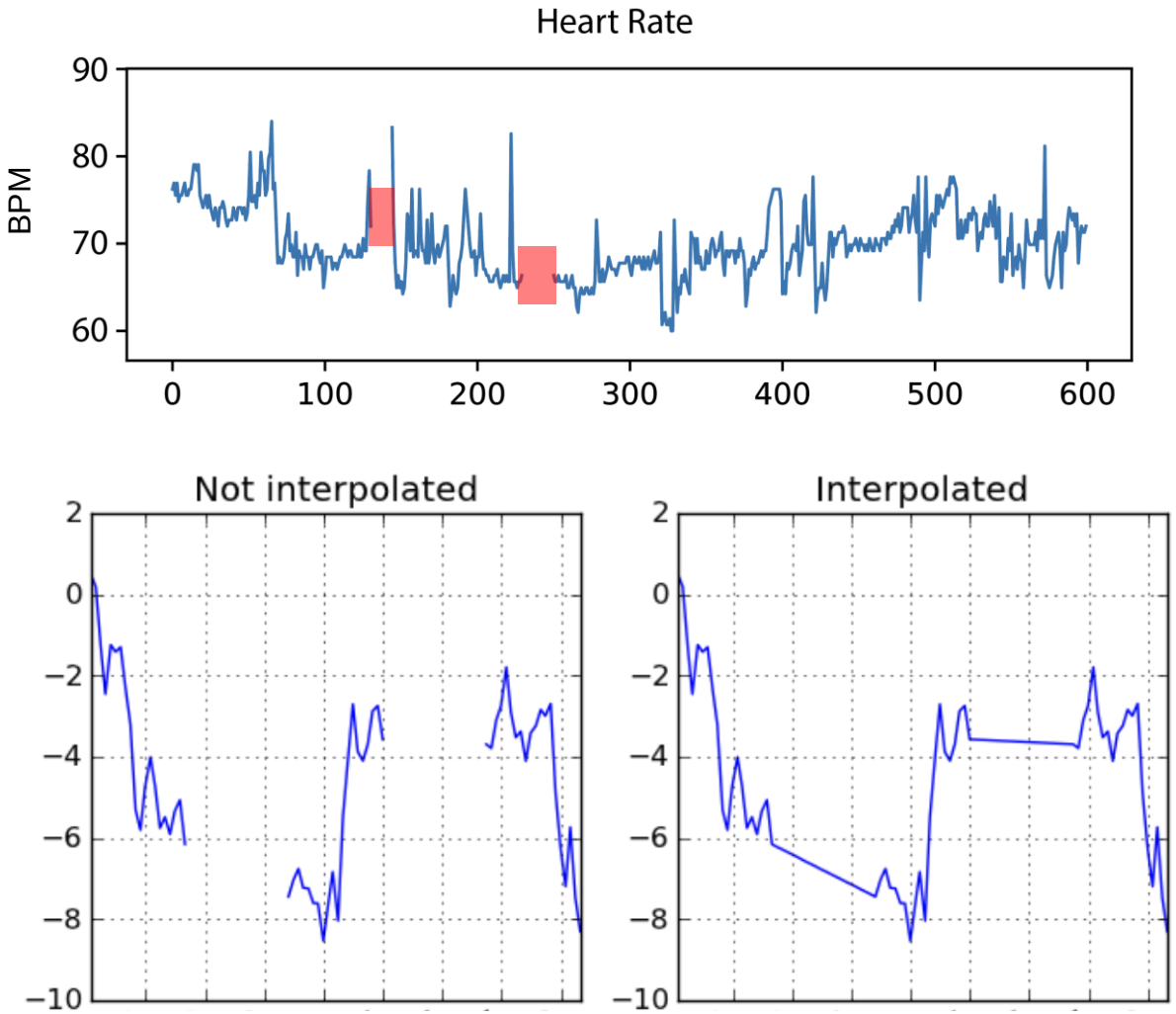


Example 2

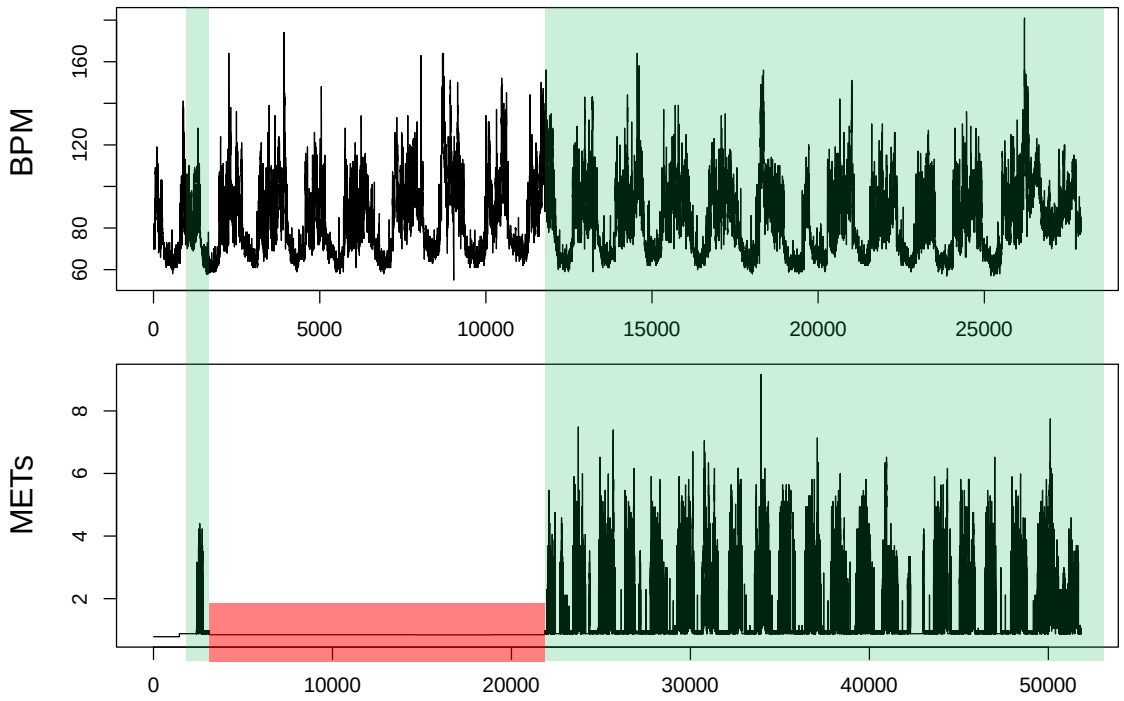


Processing the Data: Missingness

Example 1

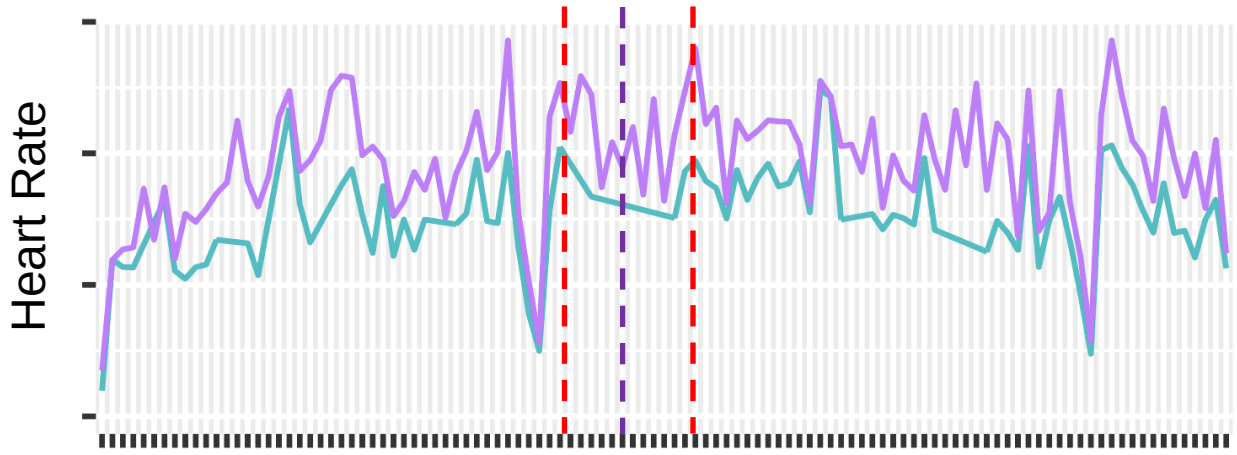


Example 2



Processing the Data: Sampling/Alignment

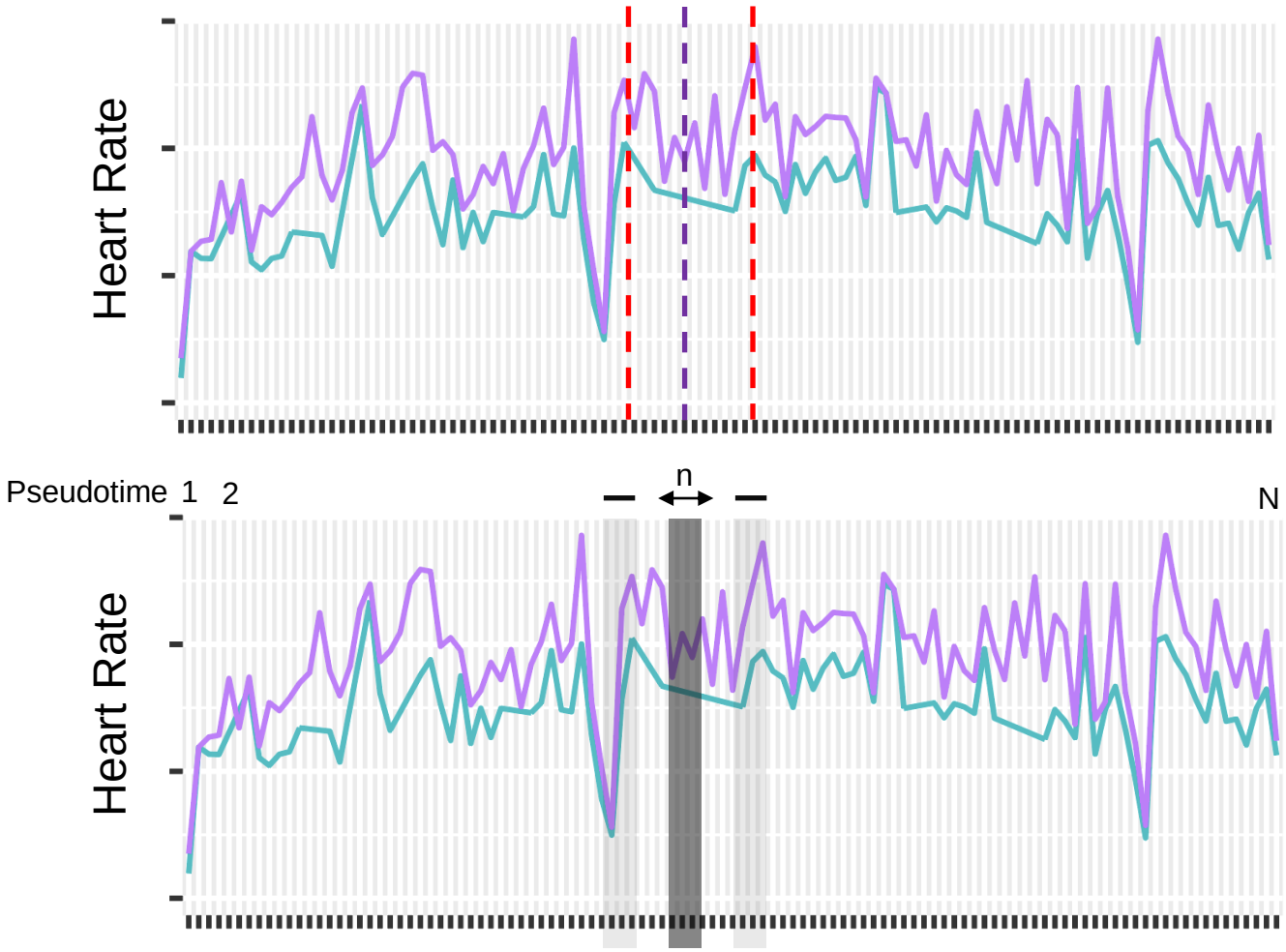
Example 1



[Adapted and modified from Snyder Lab MOVES Data]

Processing the Data: Sampling/Alignment

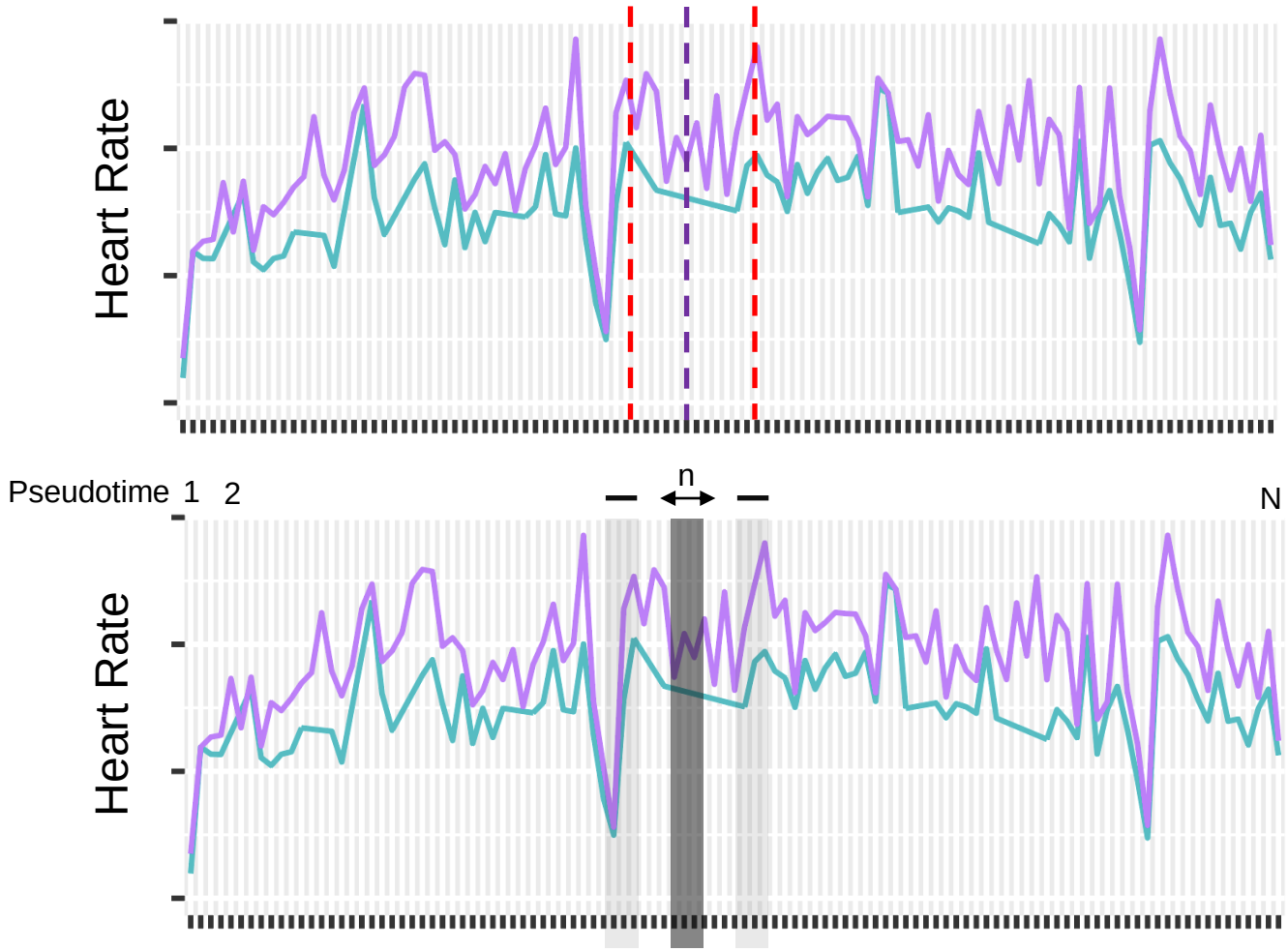
Example 1



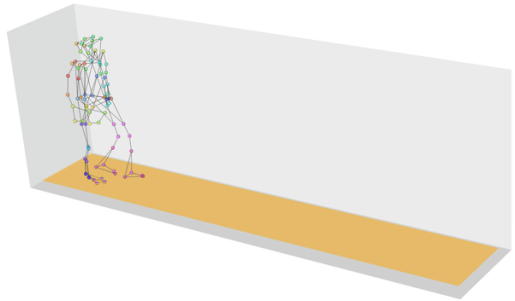
[Adapted and modified from Snyder Lab MOVES Data]

Processing the Data: Sampling/Alignment

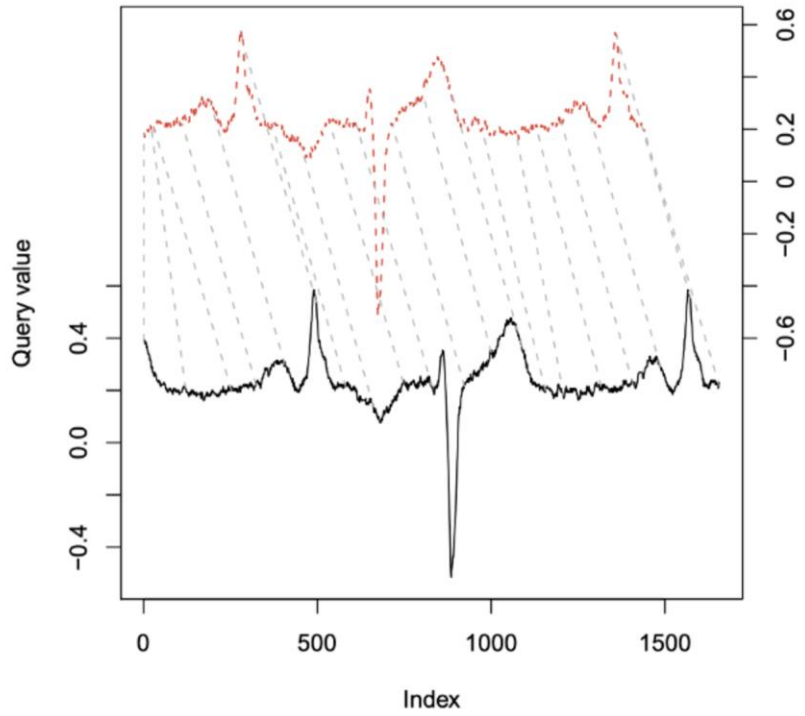
Example 1



Example 2



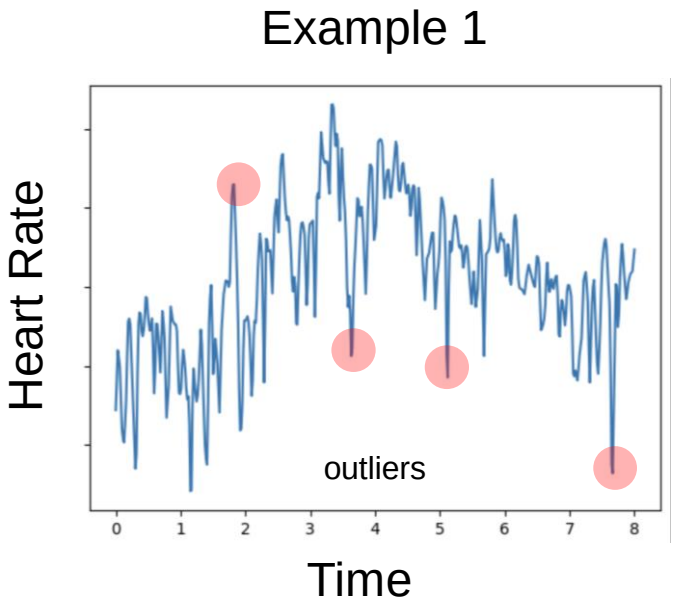
Dynamic Time Warping (DTW)



[Adapted and modified from Snyder Lab MOVES Data]

[Top: Olsen et al. J. Royal Stat Soc. (‘18); Bottom: Giorgino, Jstatsoft (‘19)]

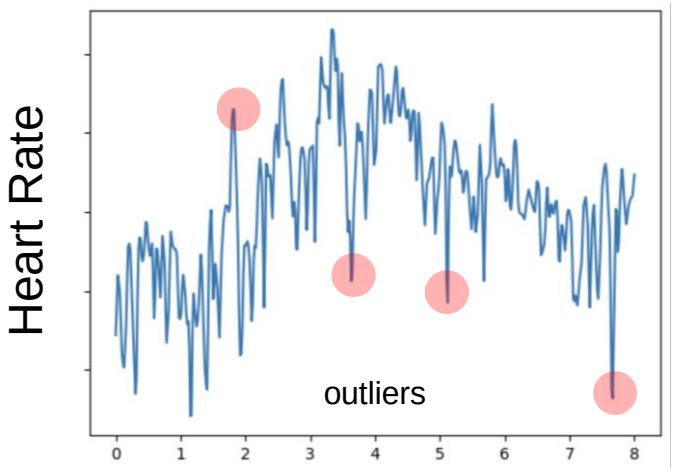
Processing the Data: Denoising

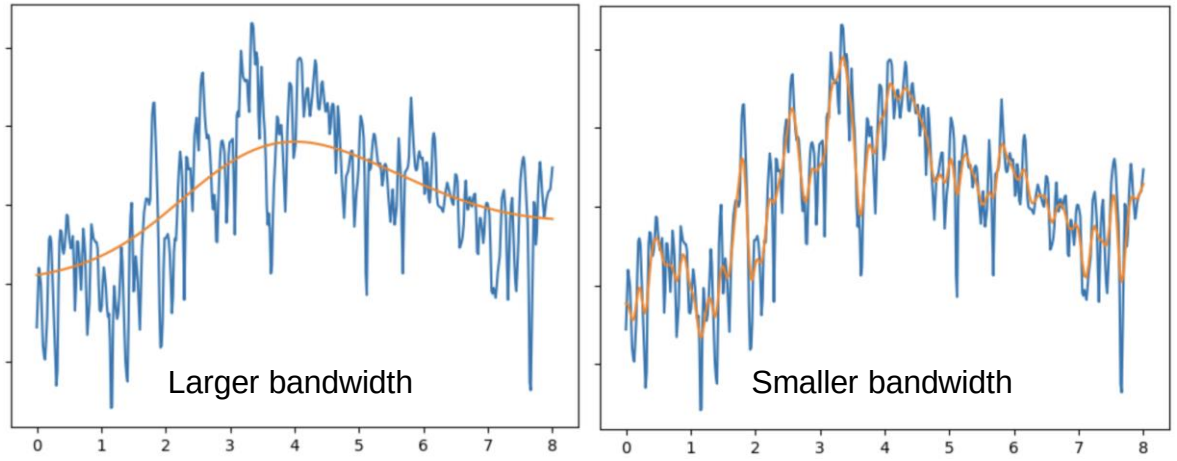


[Adapted and modified from SciKit-FDA]

Processing the Data: Denoising

Example 1



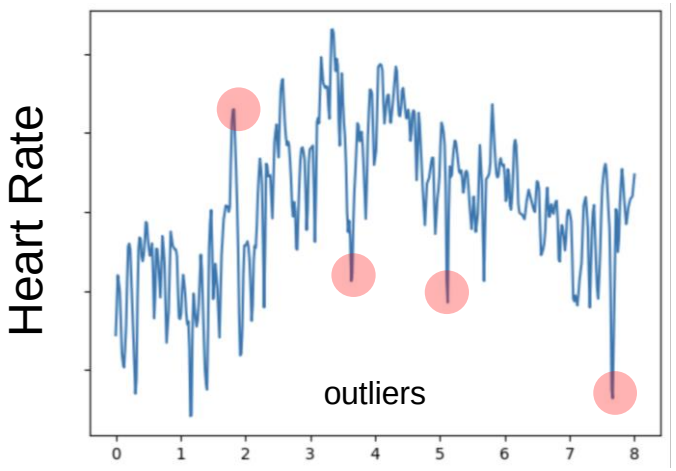


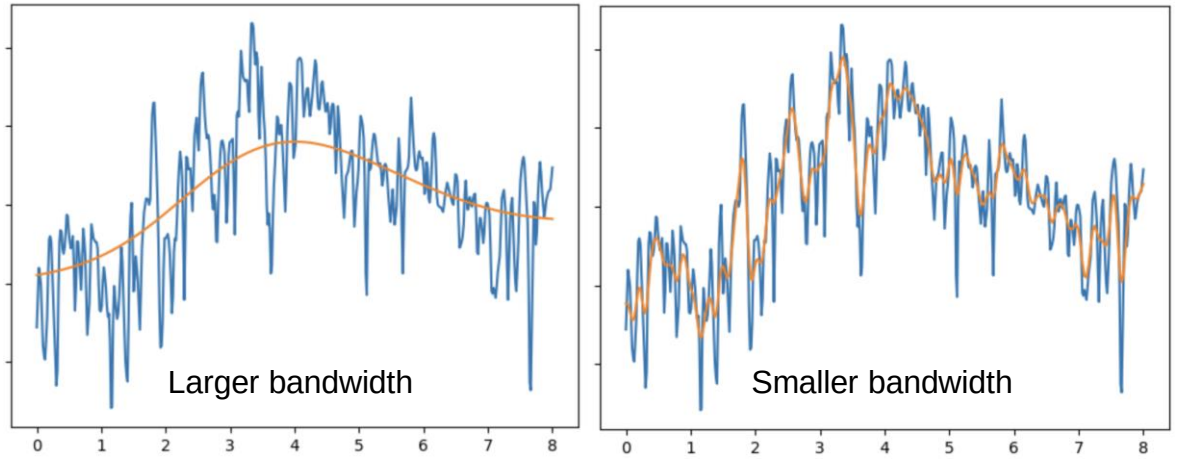
Kernel Smoother

[Adapted and modified from SciKit-FDA]

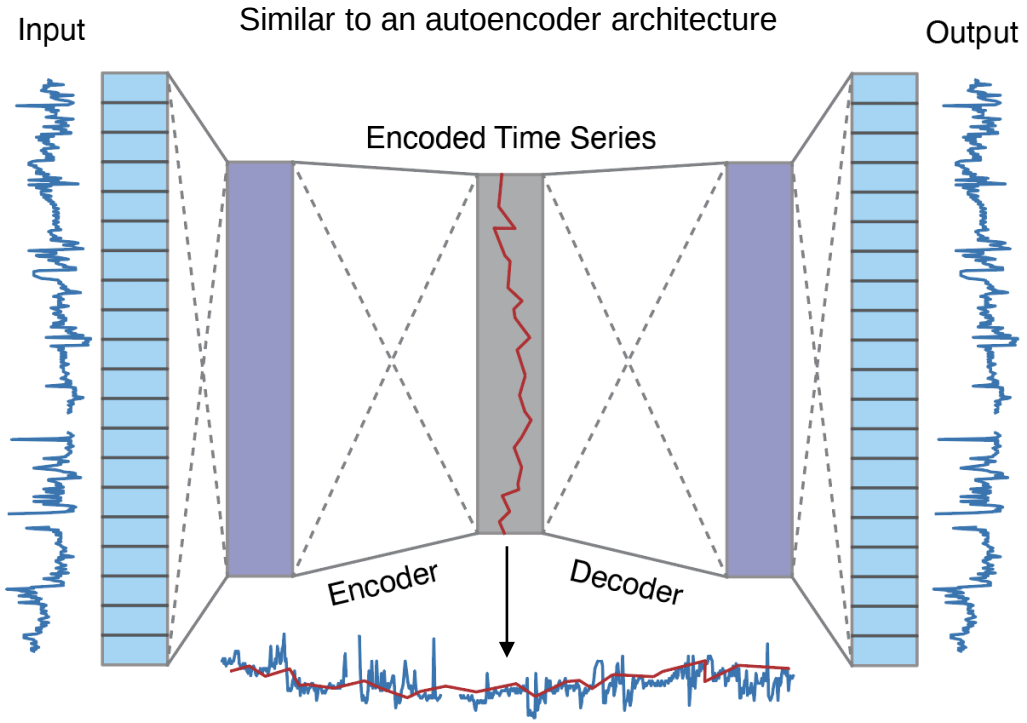
Processing the Data: Denoising

Example 1





Example 2

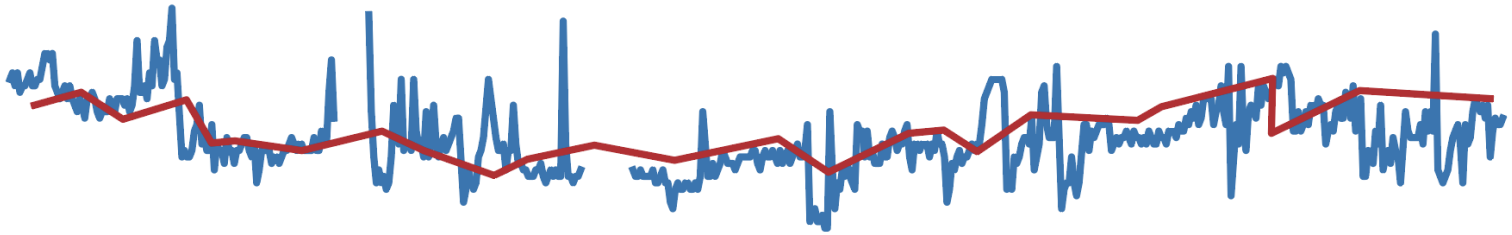


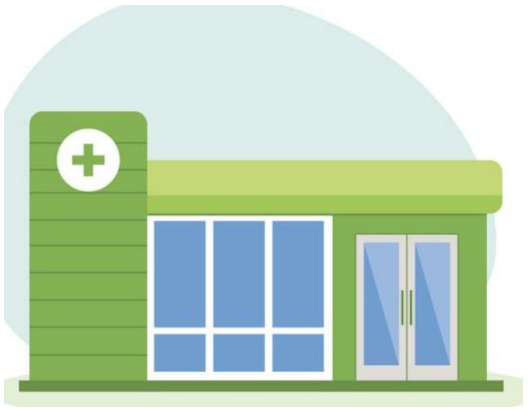
Encoded Time Series Retains Signal and Removes Outliers

[Adapted and modified from SciKit-FDA]

Wearables and Biosensors

Feature Engineering and Modeling





Wearable Sensors in Biomedical and Clinical Research

ARTICLES
https://doi.org/10.1038/s41551-020-00640-6

nature biomedical engineering

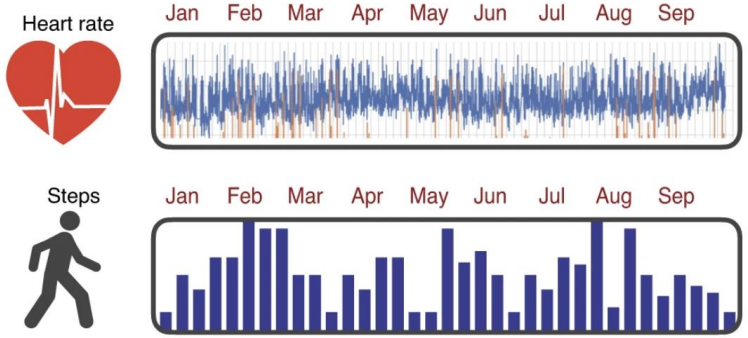
Check for updates

Pre-symptomatic detection of COVID-19 from smartwatch data

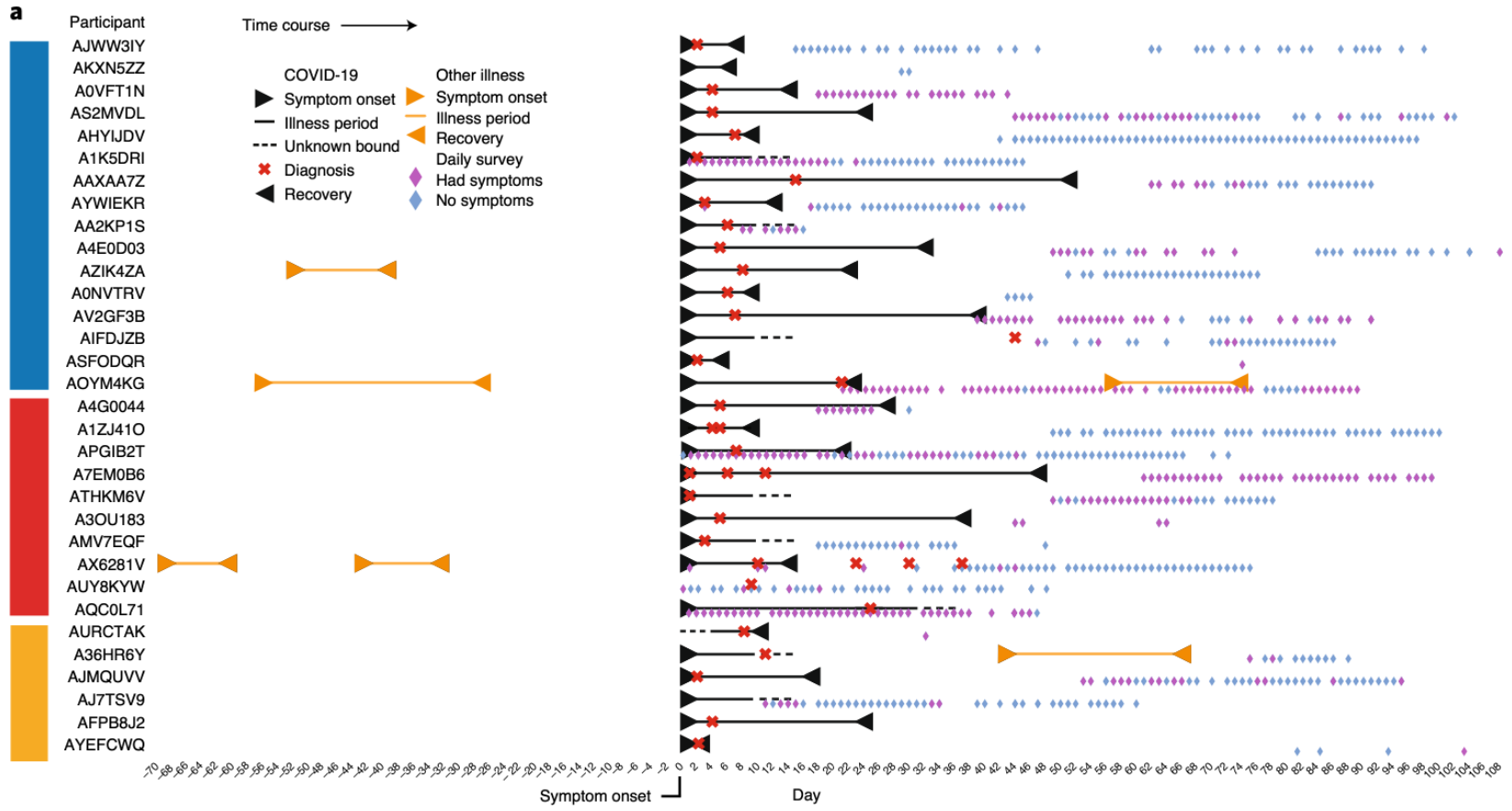
Tejaswini Mishra[1,3], Meng Wang[1,3], Ahmed A. Metwally[1,3], Gireesh K. Bogu[1,3], Andrew W. Brooks[1,3], Amir Bahmani[1,3], Arash Alavi[1,3], Alessandra Celli[1], Emily Higgs[1], Orit Dagan-Rosenfeld[1], Bethany Fay[1], Susan Kirkpatrick[1], Ryan Kellogg[1], Michelle Gibson[1], Tao Wang[1], Erika M. Hunting[1], Petra Mamic[1], Ariel B. Ganz[1], Benjamin Rolnik[1], Xiao Li[2]✉ and Michael P. Snyder[1]✉

[Mishra et al. Nat. BME ('20)]

Data Collected



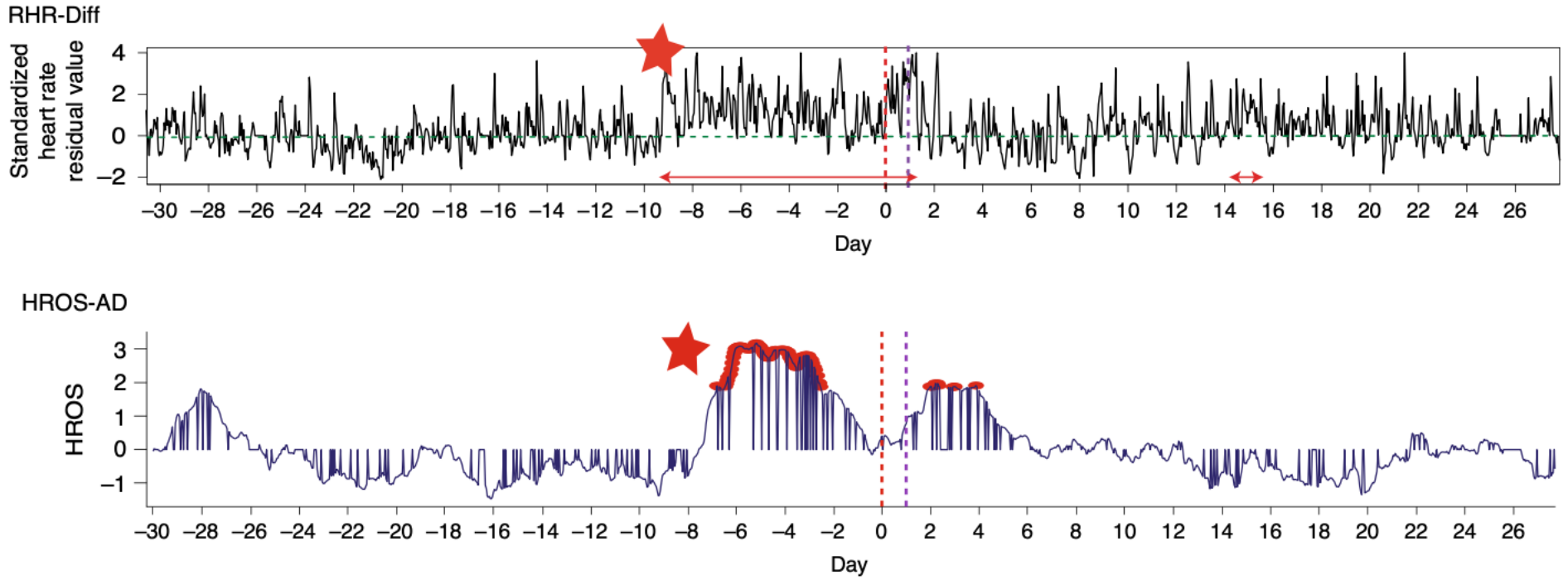
Clinical Metadata



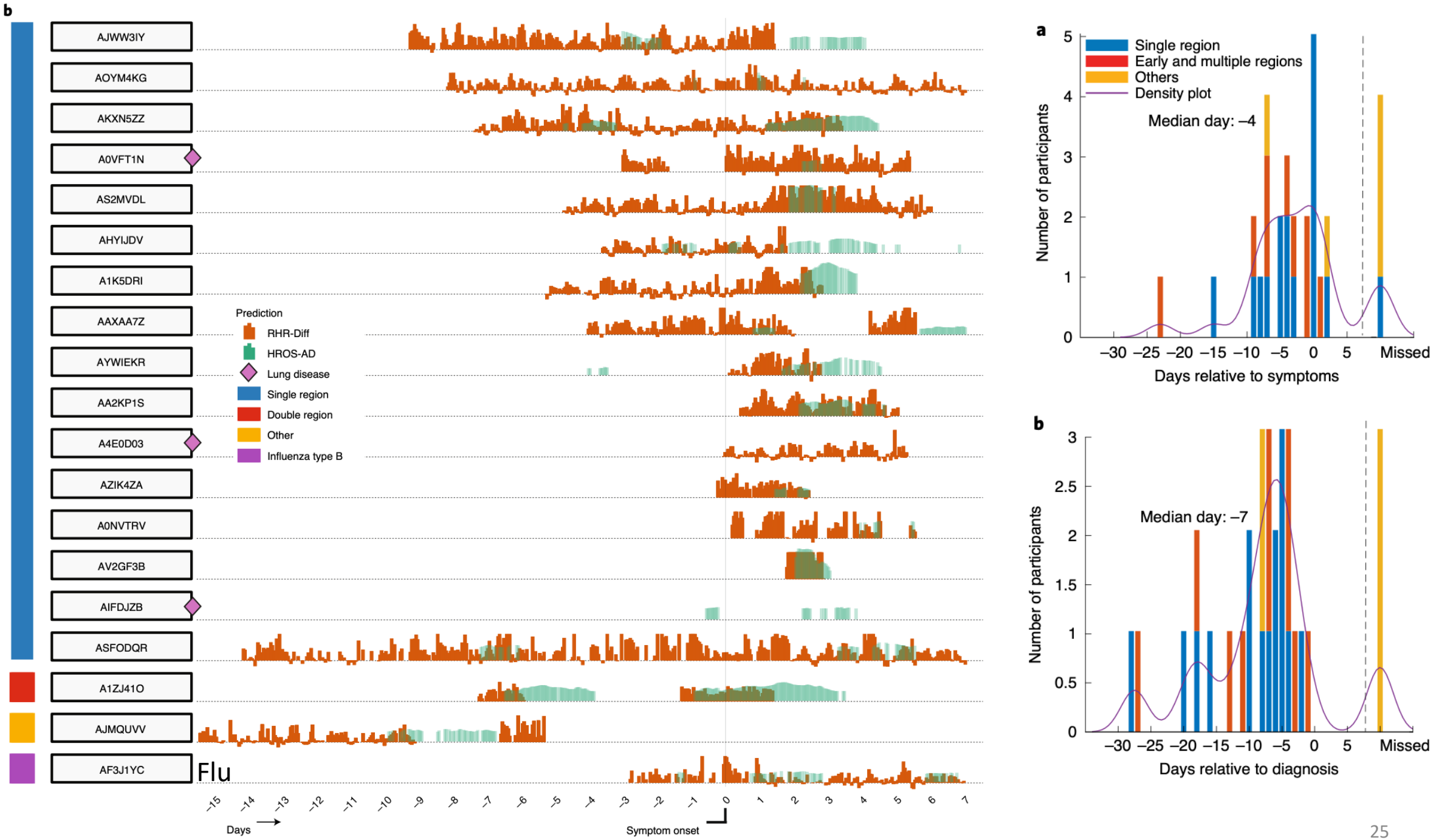
[Mishra et al. Nat. BME ('20)]

Interpretable Feature Engineering

- Resting Heart Rate (RHR)
- Heart rate over steps ratio (HROS)



[Mishra et al. Nat. BME ('20)]



[Mishra et al. Nat. BME ('20)]

Wearable Sensors in Biomedical and Clinical Research: Evaluating Personalized Interventions

PLOS COMPUTATIONAL BIOLOGY

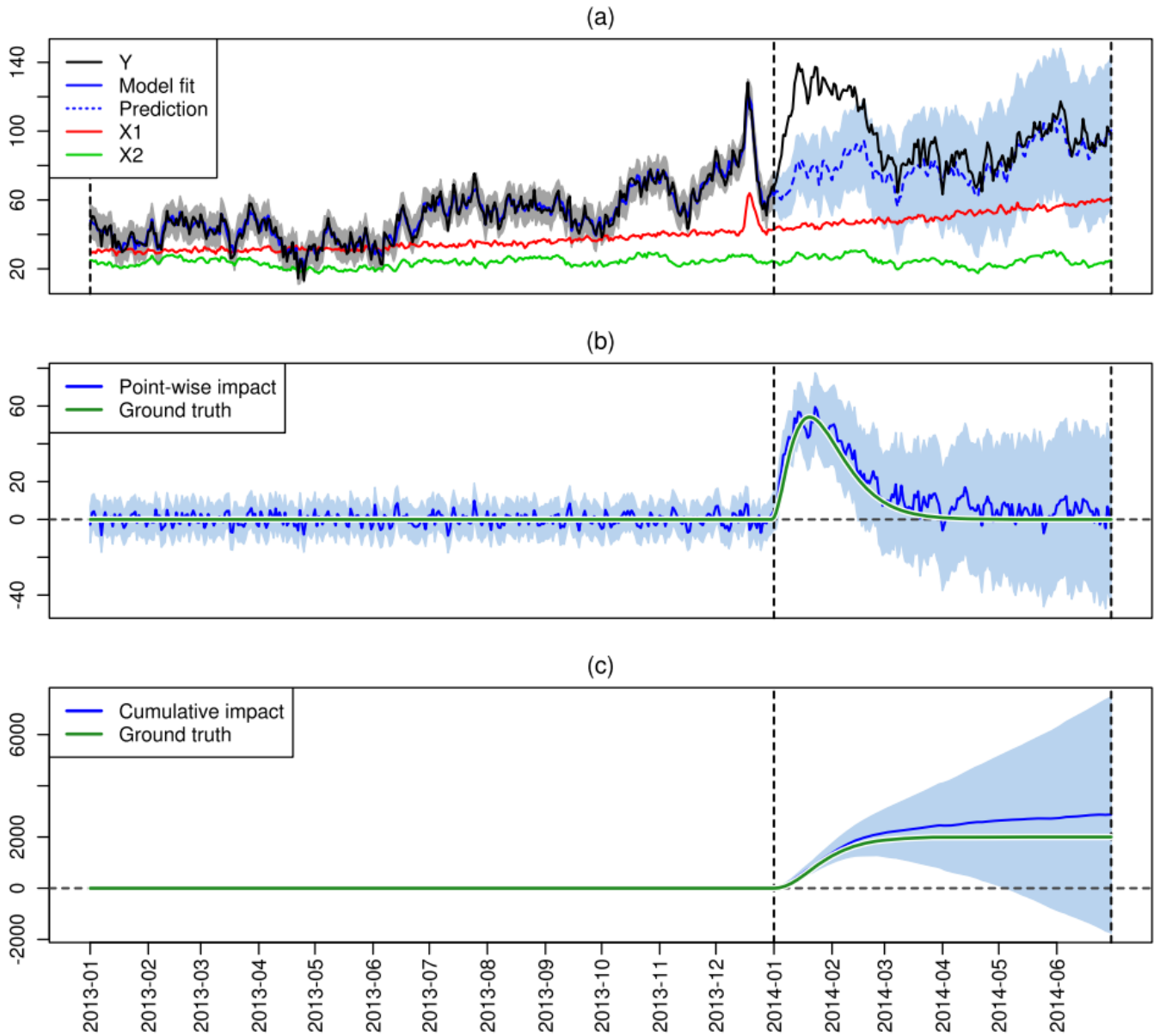
RESEARCH ARTICLE

Bayesian structural time series for biomedical sensor data: A flexible modeling framework for evaluating interventions

Jason Liu[1,2‡], Daniel J. Spakowicz[3,4‡], Garrett I. Ash[5,6], Rebecca Hoyd[3], Rohan Ahluwalia[1,2], Andrew Zhang[1,2], Shaoke Lou[1,2], Donghoon Lee[7,8], Jing Zhang[9], Carolyn Presley[3], Ann Greene[10], Matthew Stults-Kolehmainen[11,12], Laura M. Nally[10], Julien S. Baker[13,14], Lisa M. Fucito[15,16,17], Stuart A. Weinzimer[10,18], Andrew V. Papachristos[19], Mark Gerstein[1,2,20,21]*

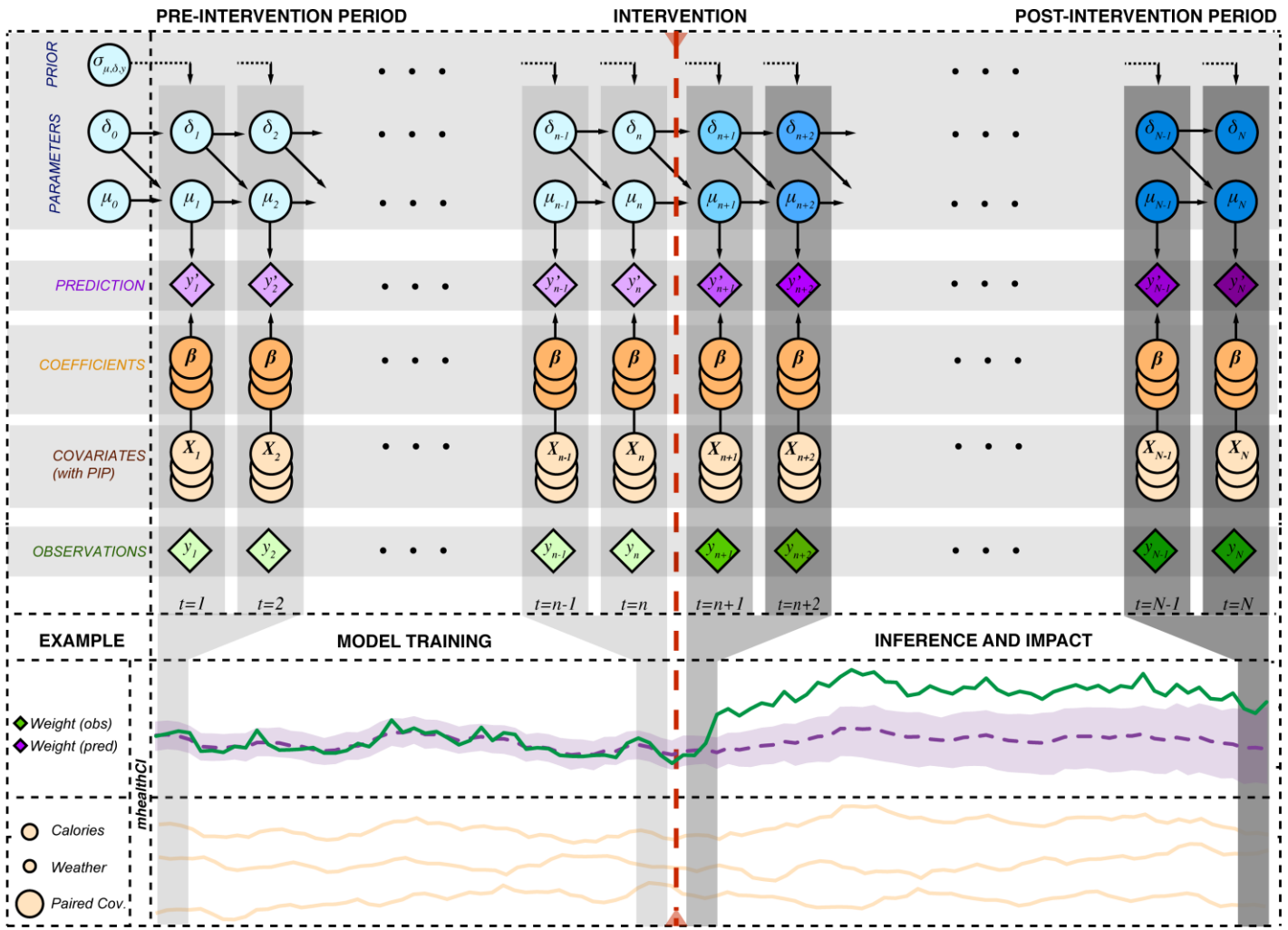
[Liu*, Spakowicz* et al. PLOS Computational Biology ('21)]

Bayesian Structural Time Series and Causal Impact
MODELING



[Broderson et al. Annals of Applied Stat. ('15)]

Using a Bayesian Structural Time Series Framework for Modeling Biosensor Data to Evaluate Interventions



y_t
y'_t } Impact

X_t

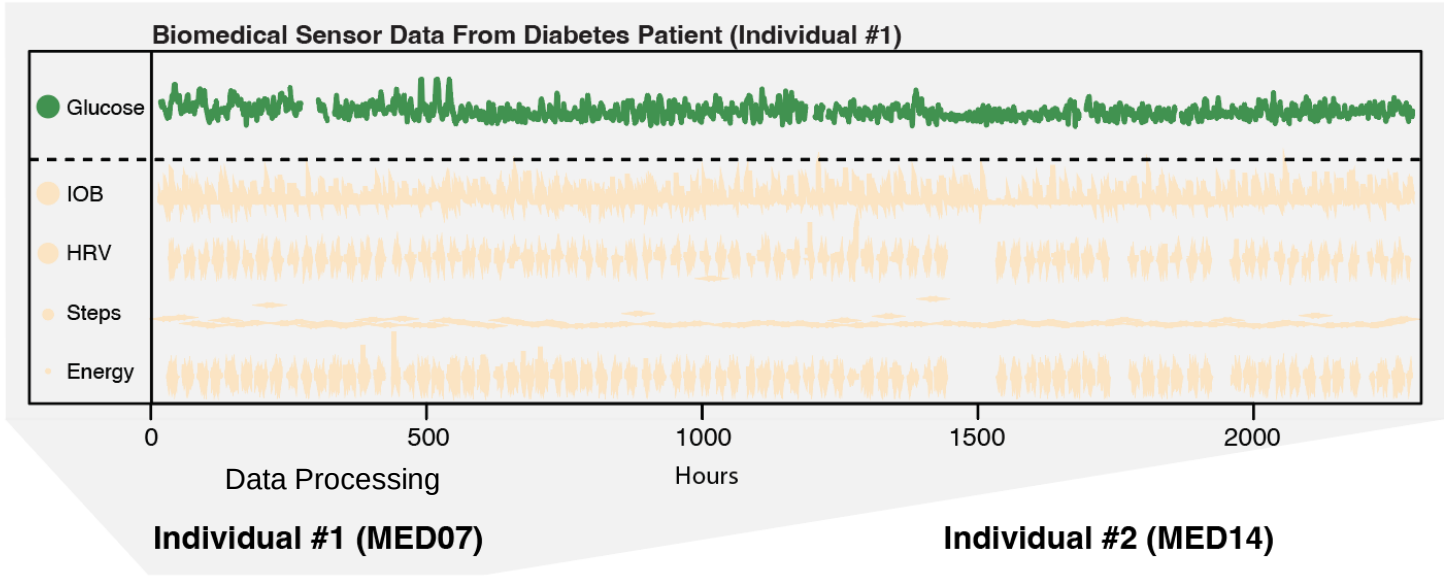
$$y'_t = \mu_t + X_t\beta + e_t, e_t \sim N(0, \sigma_e^2)$$
$$\mu_{t+1} = \mu_t + \delta_t, \delta_t \sim N(0, \sigma_\delta^2)$$

- y_t : weight
- X_t : Covariates (calories, weather, etc.)
- e_t : error term
- μ_t : local level (unobserved trend)
- δ_t : slope

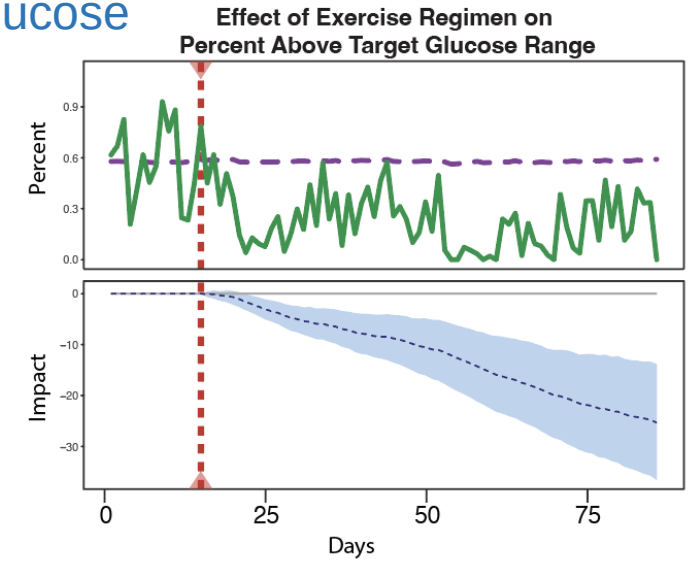
$$P(y'_{n+1:N} | y_{1:n})$$

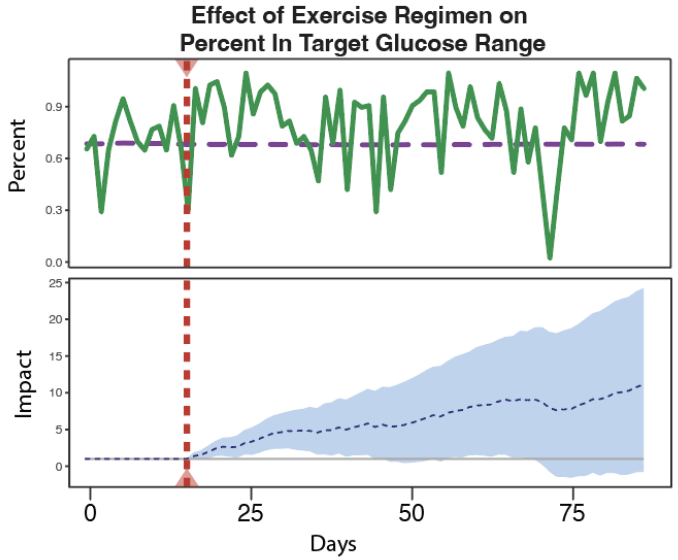
[Liu*, Spakowicz* et al. PLOS Computational Biology ('21)]

Evaluating The Efficacy of Exercise Regimens in Diabetes Patients



Green: Glucose
Dotted red: Exercise
Blue confidence:
Impact of exercise on glucose



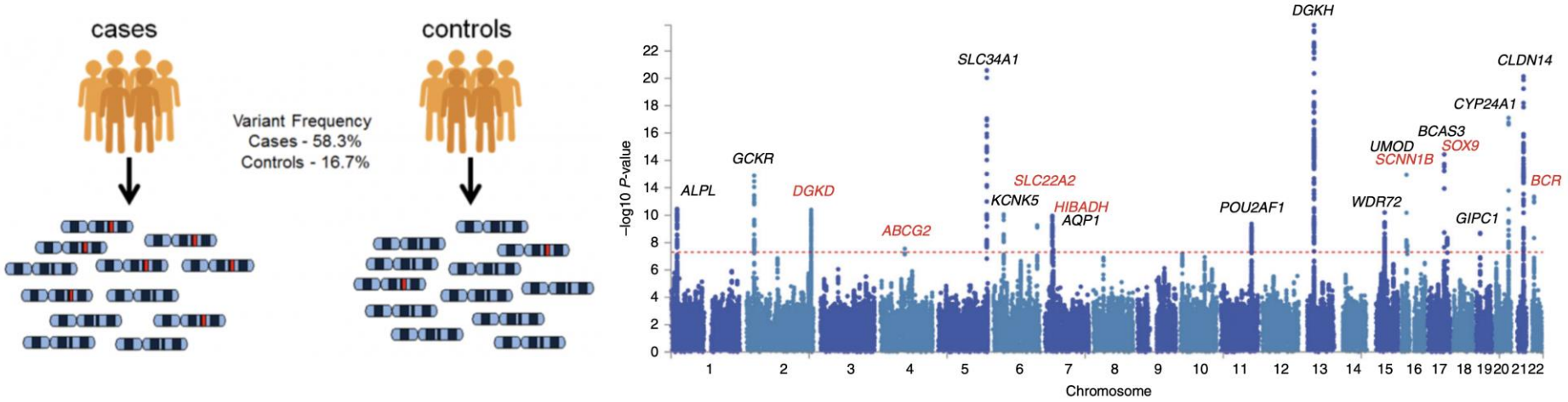


Validated Blood sugar levels: HbA1c 8.6% to 6.9%

HbA1c 5.3% to 5.2%

[Liu*, Spakowicz* et al. PLOS Computational Biology ('21)]

Exploring phenotype-to-genotype linkages



[Adapted from EMBL-EBI and Wikipedia: GWAS]

Exploring phenotype-to-genotype linkages