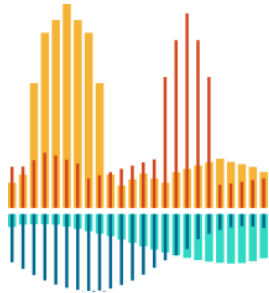


GENOMIC RESEARCH ENABLED BY ELECTRONIC HEALTH RECORDS



CLEVELAND
INSTITUTE FOR
COMPUTATIONAL
BIOLOGY



November 14, 2019

Dana C. Crawford, PhD

Associate Professor

Population and Quantitative Health Sciences

Genetics and Genome Sciences

PRECISION MEDICINE

THE CONCEPT



RELAX
AND BE
PATIENT

Right Patient

Right Drug

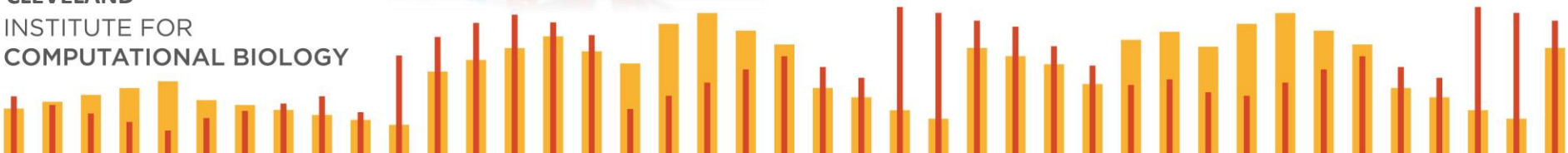


Right Dose

First Time

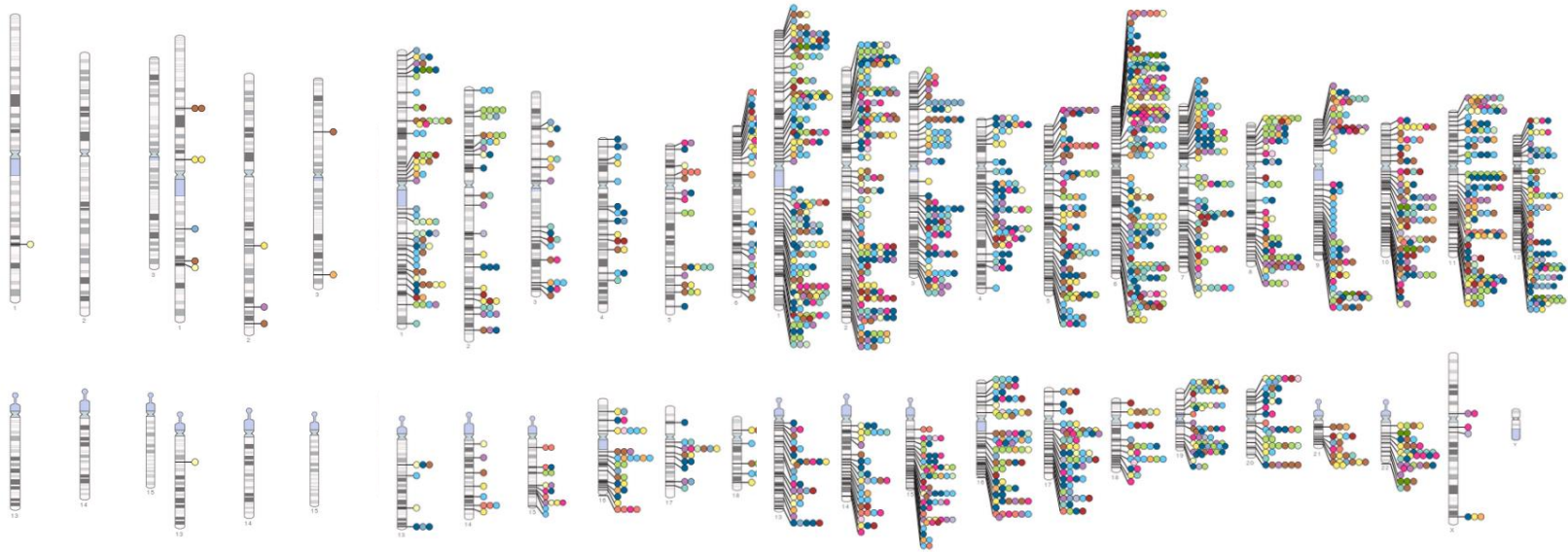


CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY

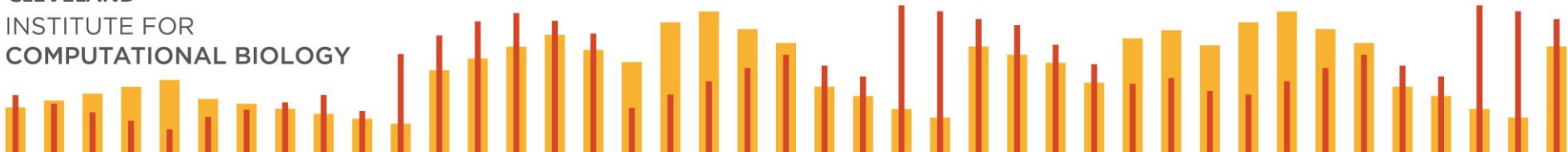


PRECISION MEDICINE RESEARCH

MUCH DATA COLLECTED



CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY



<https://www.ebi.ac.uk/gwas/>

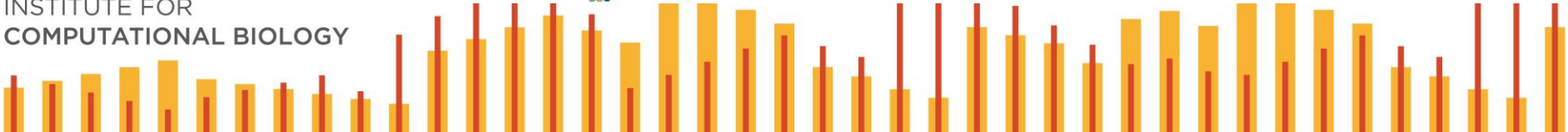
GWAS as of
10/14:

7,796 publications

159,202 SNP-
Trait Associations



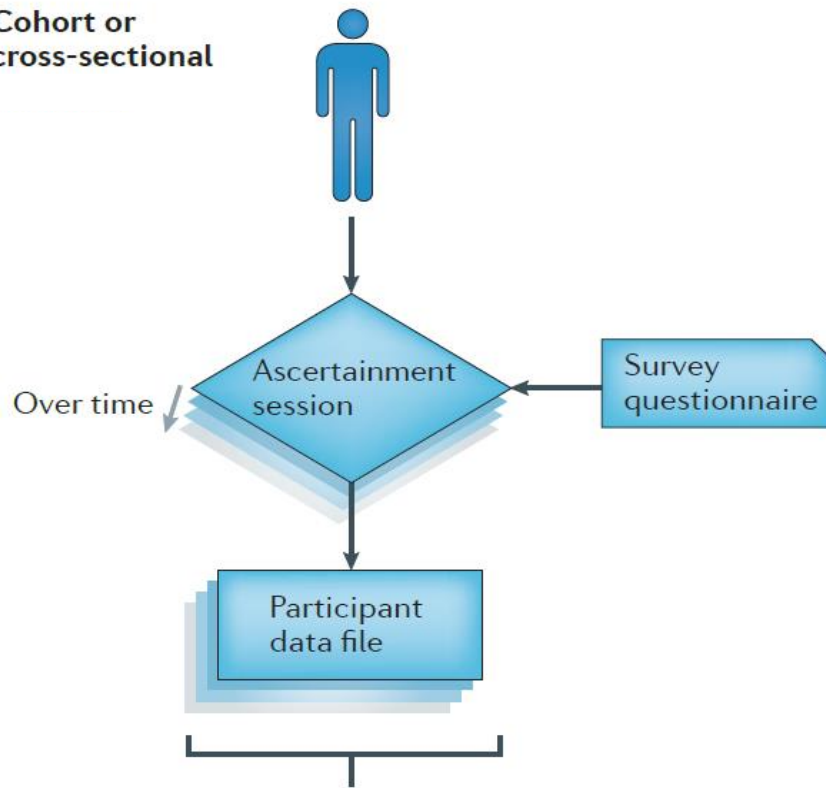
CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY



PRECISION MEDICINE RESEARCH

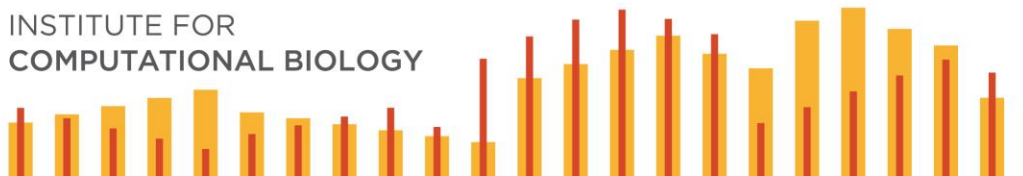
MUCH DATA COLLECTED

Cohort or
cross-sectional



Bush, Oetjens, Crawford (2016) *Nat Rev Genet* 17(3):129-45

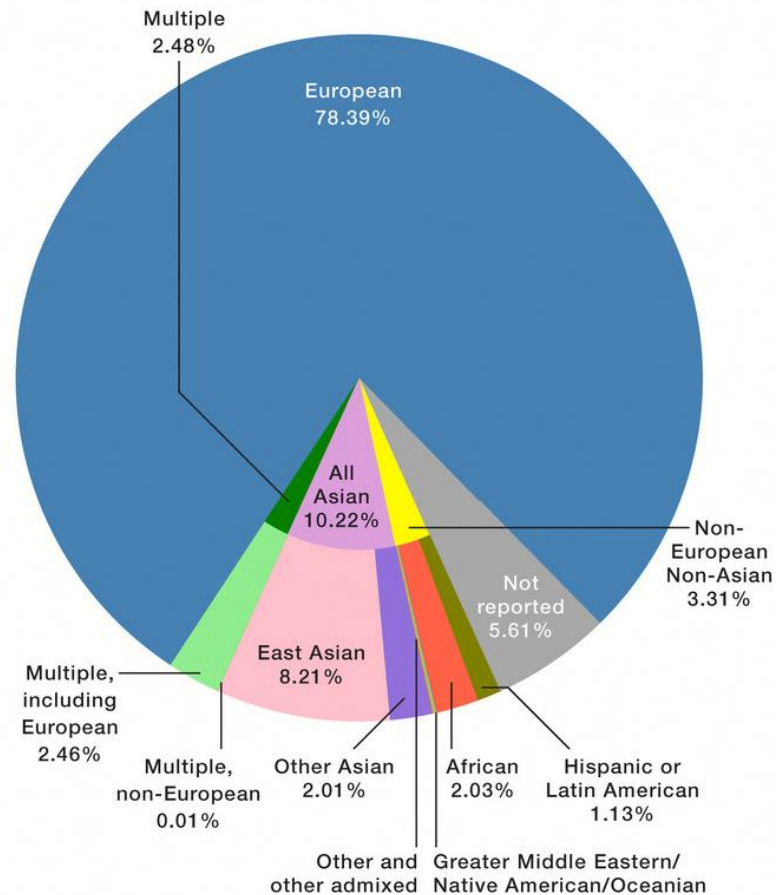
CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY



Participant epidemiology-based measures

PRECISION MEDICINE RESEARCH MORE DATA NEEDED

Ancestry category distribution
of individuals in GWAS catalog



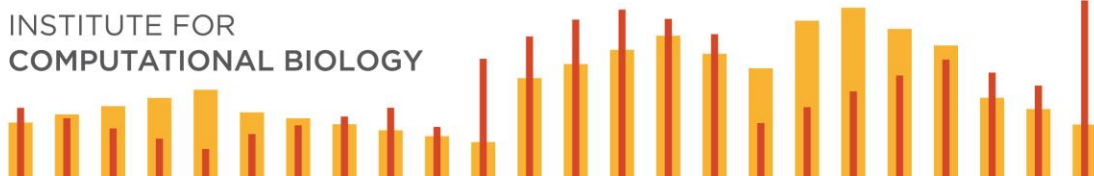
Leading Edge
Commentary

The Missing Diversity in Human Genetic Studies

Sirugo, Williams, Tishkoff (2019) *Cell* 177(1):26-31

Cell

CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY





1948–2014: Original cohort (Gen 1): physical examinations, blood samples and lifestyle-related data, collected approximately every other year

1971–present: Offspring cohort (Gen 2): physical examinations, blood samples and lifestyle-related data, collected approximately every 4–7 years

2002–present: Third-Generation cohort (Gen 3): physical examinations, blood samples and lifestyle-related data, collected approximately every 4–7 years

1948

1971

1979

1982

1998

1999

2000

2002

2006

2009

2014

2015

2019

Echocardiography

Exercise testing

Familial clustering studies

Cardiac CT

Linkage studies

Vascular tonometry

Brain MRI

Chest and abdominal CT

Cardiac MRI

Genome-wide association studies

DNA methylation and gene expression studies

Cardiopulmonary exercise testing

Risk factor identification and multivariable risk assessment

Circulating biomarkers, tonometry, echocardiography, coronary and whole-body CT and cardiac MRI

Familial risk, DNA and multi-omics measures

Andersson et al (in press) *Nat Rev Cardiol*

Participant epidemiology-based measures

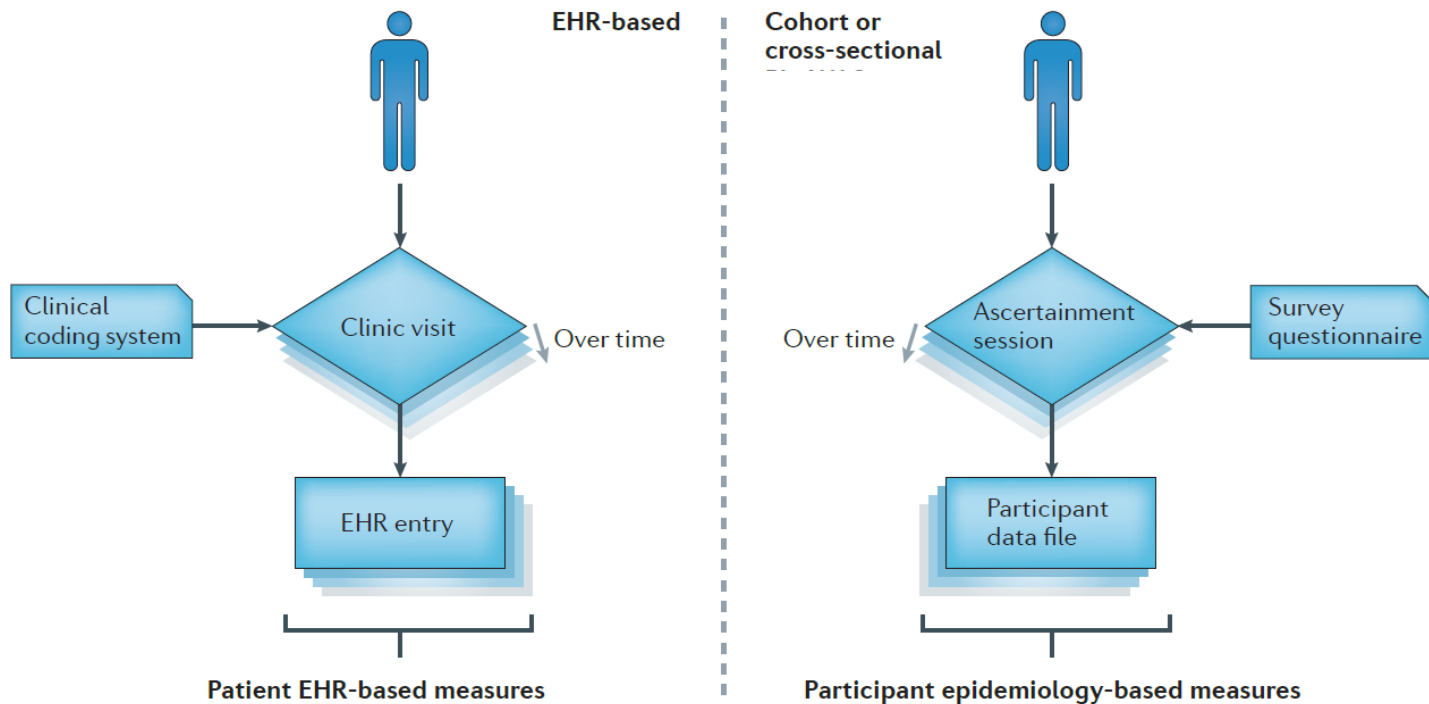
CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY



ELECTRONIC HEALTH RECORDS

ACCELERATING PRECISION MEDICINE RESEARCH

Bush, Oetjens, Crawford (2016) *Nat Rev Genet* 17(3):129-45

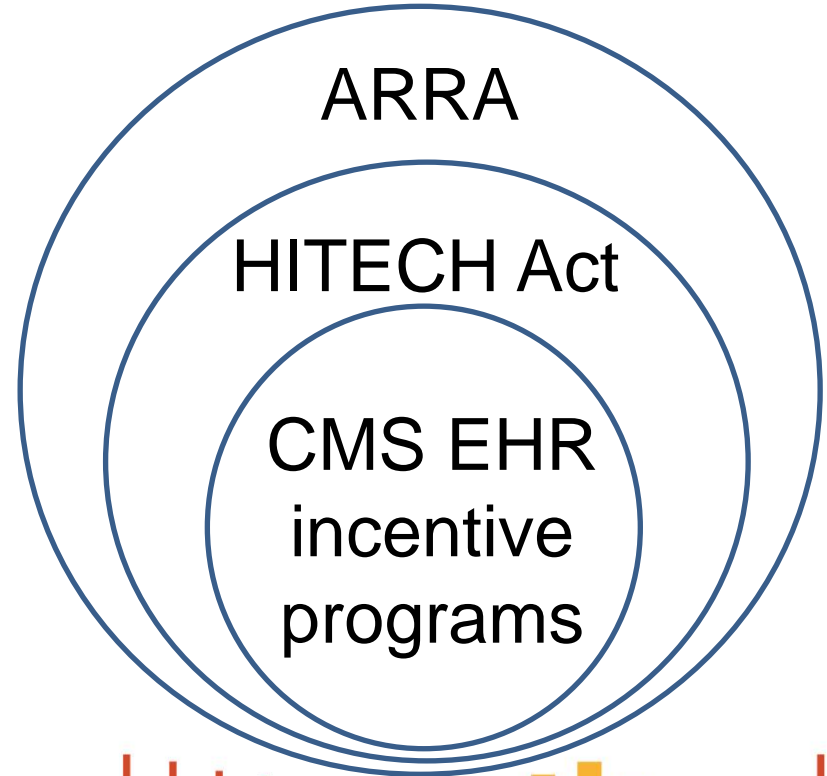


THE RAPID RISE OF EHRs

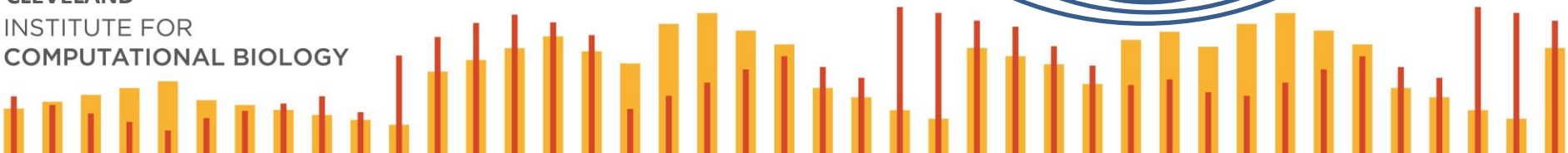


ARRA

American Recovery and
Reinvestment Act of 2009



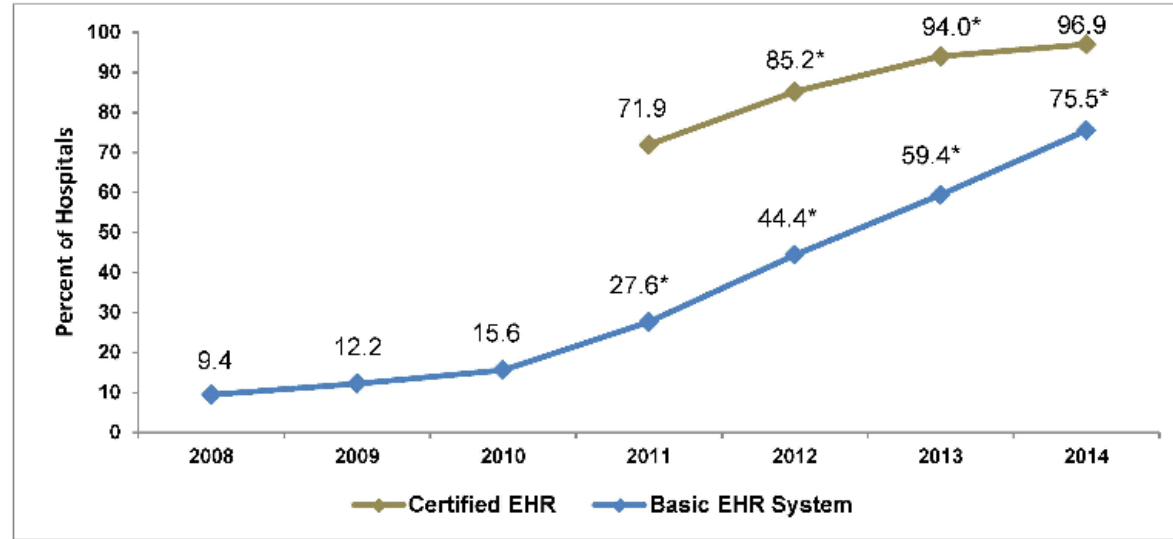
CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY



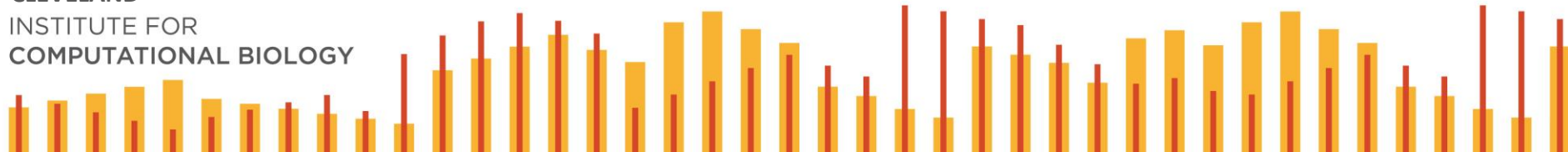
THE RAPID RISE OF EHRs



96% of reporting US hospitals have at least a basic EHR



<https://www.healthit.gov/sites/default/files/data-brief/2014HospitalAdoptionDataBrief.pdf>



ELECTRONIC HEALTH RECORDS

ACCELERATING PRECISION MEDICINE RESEARCH

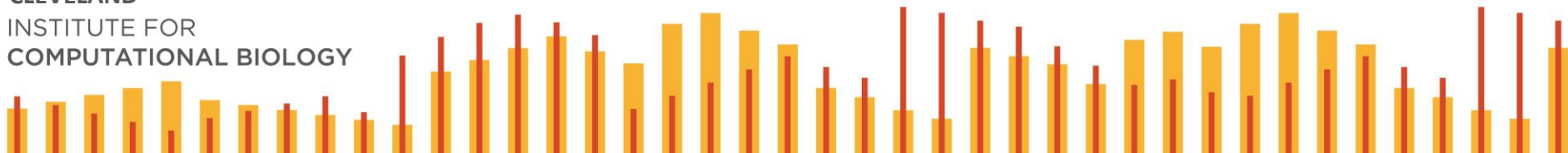
Patient EHR-based measures

250.0 T2DM	Yes
411.1 coronary syndrome	Yes
414.01 coronary artery disease	No
278.01 obesity	Yes
Alanine aminotransferase	15.6 units per l
Blood albumin	3.7 g per dl
Aspartate aminotransferase	22 units per l
Bicarbonate (HCO ₃)	24 mEq per l
Carbon dioxide (CO ₂)	27 mEq per l
Blood cholesterol	240 mg per dl
Blood creatinine	1.2 mg per dl

Participant epidemiology-based measures

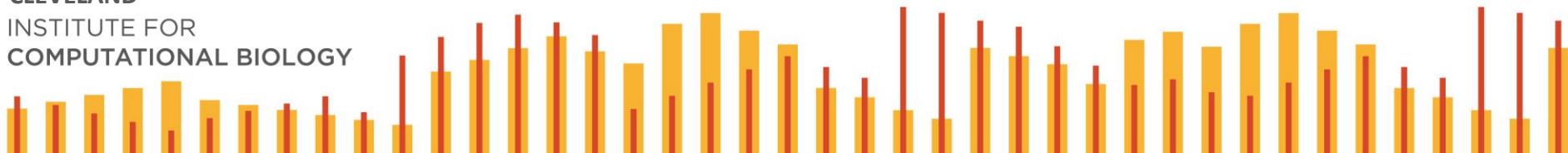
Ever had diabetes?	Yes
Cancer ever diagnosed?	Yes
Ever smoked?	No
Allergic to gluten?	No
Allergic to peanuts?	Yes
Current weight	240 lb
Current height	5'8"
Green vegetables per week	2–4 servings
Red meat per week	6–8 servings
Blood cholesterol	275 mg per dl
Exercise time per week	30 min

Bush, Oetjens, Crawford (2016) *Nat Rev Genet* 17(3):129-45



COMPUTABLE PHENOTYPING

- Demographics Structured and unstructured text
- Vitals Structured
- Medical History Structured data and unstructured text
- Medical encounter Structured data and unstructured text
- Orders and prescriptions Structured data
- Laboratory tests Structured data



COMPUTABLE PHENOTYPING

STRUCTURED DATA

Bush, Oetjens, Crawford (2016) *Nat Rev Genet* 17(3):129-145

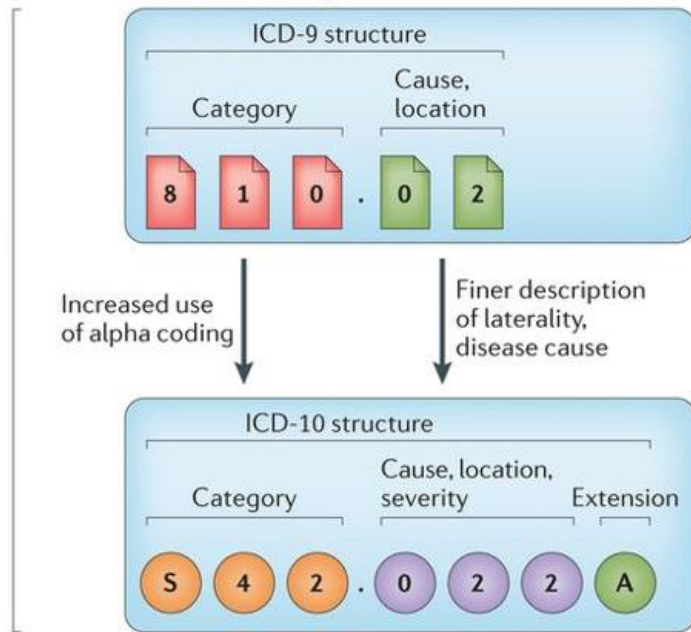
Billing Codes

ICD-9-CM

ICD-10-CM

5× more diagnosis codes
in ICD-10-CM
(14,025 versus 69,823)

Gross anatomy of ICD-9 and ICD-10 codes



CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY



RULES-BASED ALGORITHMS

Billing codes
Procedure codes
Problems lists

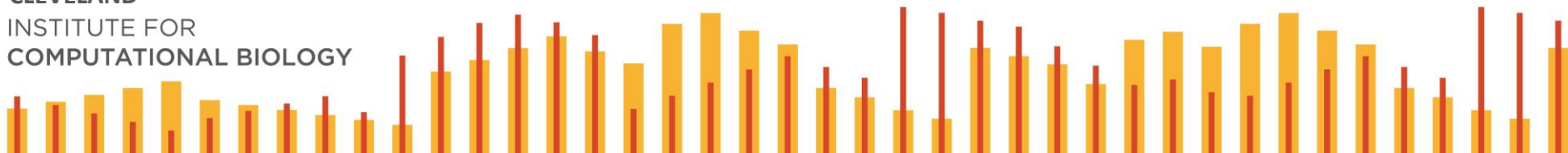
Disease
Diagnosis

EXCLUDE

Laboratory
Values

Medications

Pendergrass and Crawford (2019)
Curr Proc Hum Genet 100:e80



MILLION VETERAN PROGRAM (MVP)

- Funded by Department of Veterans Affairs Office of Research and Development

- Observational cohort of US Veterans (2011)

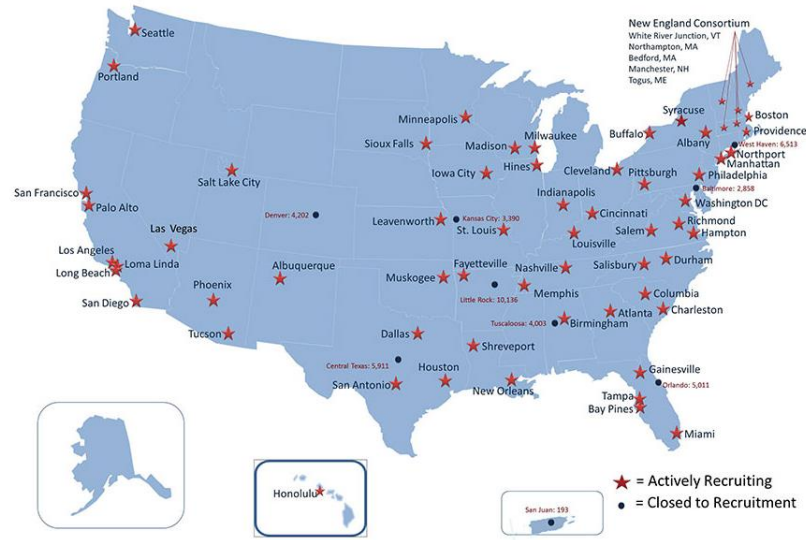
Biospecimens

Genome-wide data

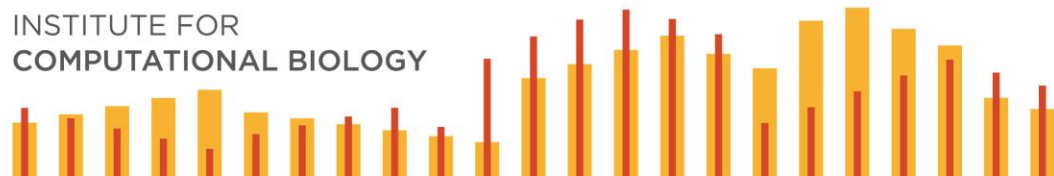
Surveys

Electronic health records

<https://www.research.va.gov/MVP/default.cfm>



CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY



RULES-BASED ALGORITHMS IN MVP

Halladay et al (2019) *AMIA Joint Summits Transl Sci Proc* 2019:153-162

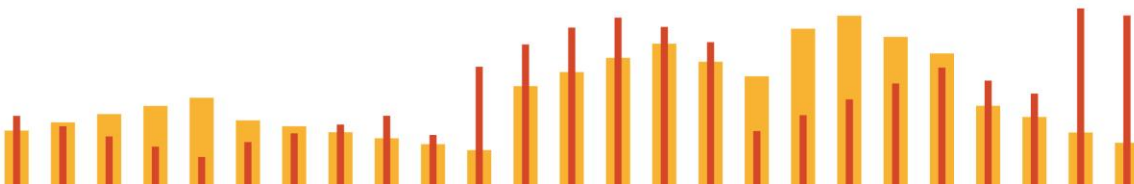


	Not Hispanic (with genotypes)	Hispanic (with genotypes)	Unknown (with genotypes)
African American	94,891 (65,983)	1,279 (861)	695 (457)
Native American/ Alaska Native	4,562 (3,101)	962 (647)	29 (18)
Asian	4,884 (3,132)	251 (167)	38 (21)
Pacific Islander	1,778 (1,126)	474 (307)	54 (27)
White	350,142 (247,301)	23,227 (16,171)	1,814 (1,171)
Other and Unknown	7,413 (5,106)	6,741 (4,615)	4,793 (2,742)
TOTAL	463,670 (325,749)	32,934 (21,949)	7,423 (4,436)

AGE-RELATED MACULAR DEGENERATION (AMD)

- Leading cause of adult-onset blindness
- Characterized by
 - Deterioration of the macula
 - Loss of central vision
- Two forms of AMD
 - Atrophic (dry)
 - Neovascular or exudative (wet)

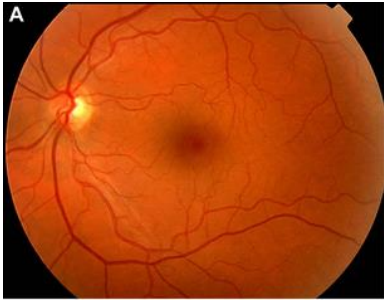
From mugumogu (on YouTube)



CLINICAL PHENOTYPE OF AMD

Fundus images showing

- Deposition of drusen within the retina
- Abnormal blood vessel growth (wet AMD)



Normal

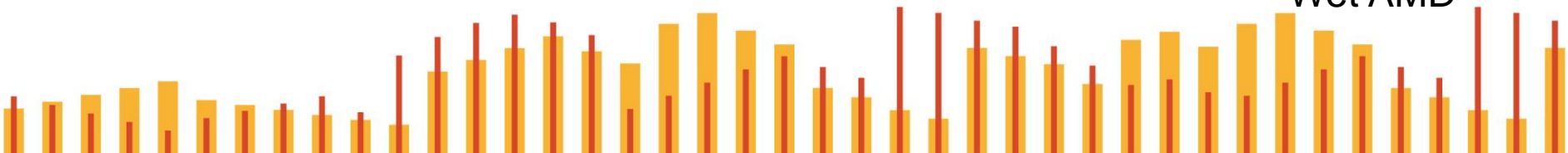


Dry AMD



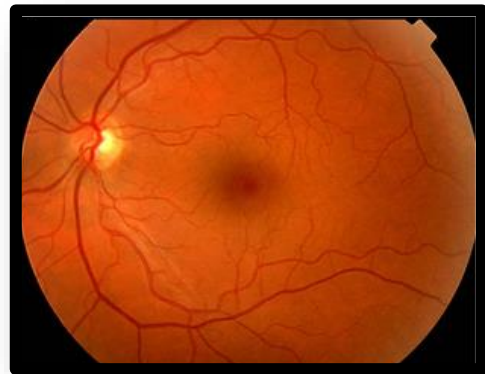
Wet AMD

(2005) *PLoS Medicine* 3(1):e38

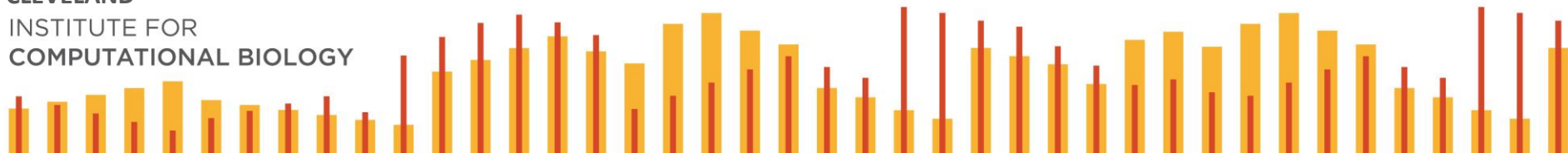


RULES-BASED ALGORITHMS AMD IN MVP

- MVP EHR hosted by Veterans Informatics and Computing Infrastructure (VINCI)
- No ophthalmology imaging data
 - Fundus photography
 - Fundus autofluorescence imaging
 - Optical coherence tomography
- Must rely EHR structured data
 - ICD-9-CM/10-CM codes
 - CPT codes



(2005) *PLoS Medicine* 3(1):e38



AMD case definition	AMD control definition
≥65 years of age	≥65 years of age
AND	AND
At least one mention within the last two years of CPT code 92004 OR CPT code 92014	At least one mention within the last two years of CPT code 92004 OR CPT code 92014
AND	AND
At least two mentions or only at the most recent visit to the Eye Clinic of ICD-9-CM codes 362.51 or 362.52 OR ICD-10-CM codes H35.31 or H35.32	Absence of ICD-9-CM codes 362.51 and 362.52 OR ICD-10-CM codes H35.31 and H35.32
AND	
Absence of ICD-9-CM code 362.55 OR ICD-10-CM code H35.389	Halladay et al (2019) <i>AMIA Joint Summits Transl Sci Proc</i> 2019:153-162

AMD ALGORITHM OPTIMIZATION

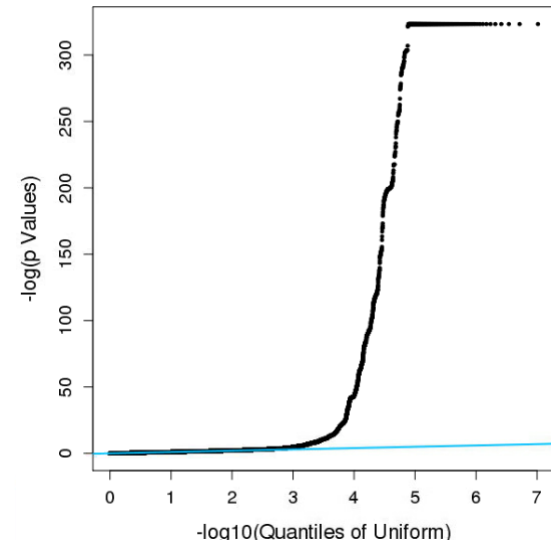
Algorithm (2 codes/case)	Case / Control age	CFH rs10801555 Odds ratio	CFH rs1080155 p- value	ARMS2 rs10490924 Odds ratio	ARMS2 rs10490924 p-value
1	65/65	1.775	1.81e-290	1.680	5.57e-201
2	More exclusions 65/65	1.775	6.90e-289	1.681	8.91e-201
1	50/60	1.746	5.85e-374	1.659	5.47e-259
2	50/60	1.747	2.42e-372	1.663	1.49e-259
3	50/65	1.689	2.73e-428	1.613	5.91e-293
3	50/60	1.667	6.40e-423	1.607	2.59e-298
4	More inclusions/ no eye exam 50/65	1.661	2.25e-431	1.600	2.26e-303
4	50/60	1.647	4.56e-424	1.588	2.36e-299

RULES-BASED ALGORITHMS AND IN MVP

Unpublished

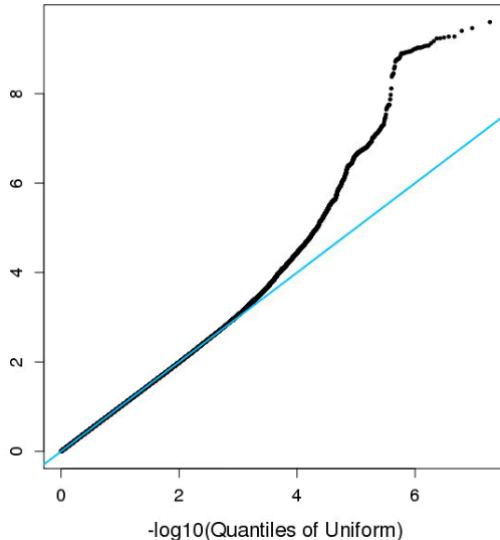
European-Americans

$n_{\text{cases}} \sim 18\text{K}$



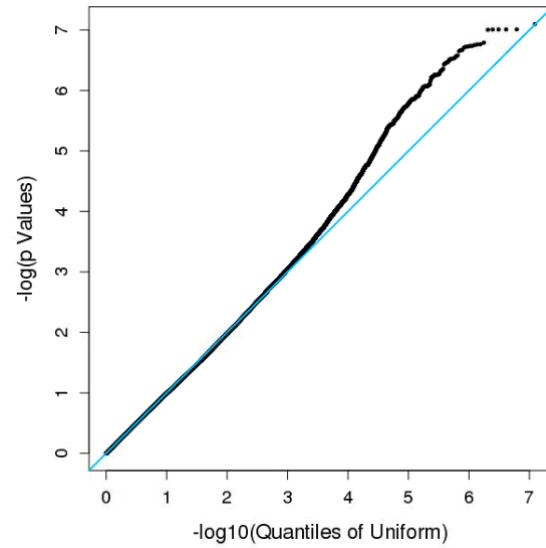
African Americans

$n_{\text{cases}} \sim 1.4\text{K}$



Hispanics

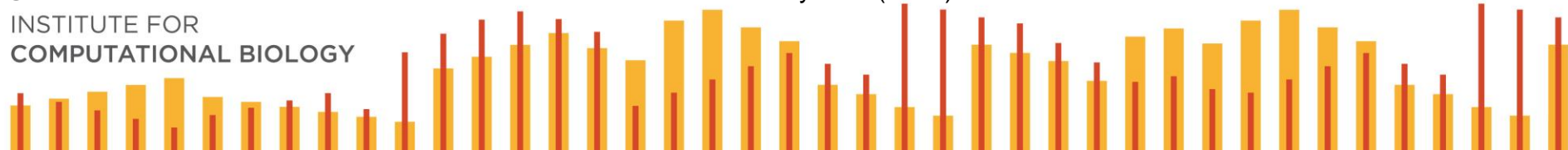
$n_{\text{cases}} \sim 600$



SUMMARY AMD AND MPV



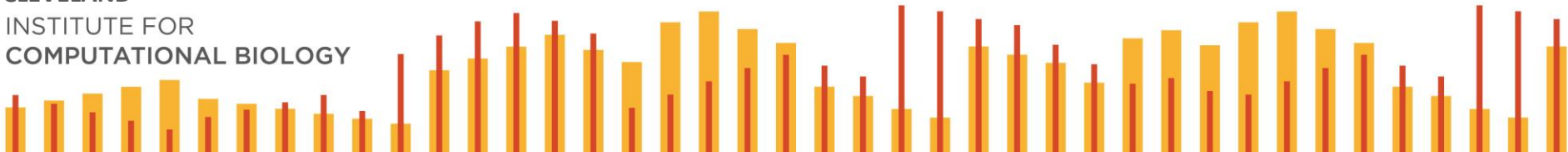
- EHR codes alone can define AMD
With high confidence
- Genetics can help optimize algorithms
Optimization based on smallest p-value (discovery)



SUMMARY ALGORITHMS AND GENETICS

- Requires EHRs linked to biobanks with DNA
- Requires phenotypes with known genetics
- Requires manual review

All will improve with more data and innovation!



ACKNOWLEDGEMENTS*

*This presentation does not represent the views of the Department of Veterans Affairs or the United States Government.

Eric Konicki

Matthew D. Anger

Chris Halladay

MVP Core Teams

Neal Peachey

Jack Sullivan

Paul Greenberg

MVP participants

Tamer Hadi

Steven Fliesler

Wen-Chih Wu

Local sites

Rob Igo, Jr.

Sudha Iyengar



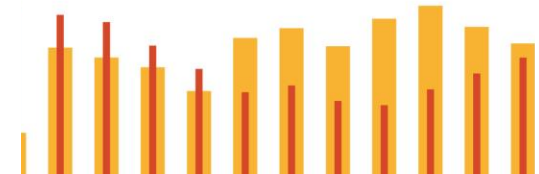
101 BX003364

Research to Prevent Blindness



Providence VA
Medical Center

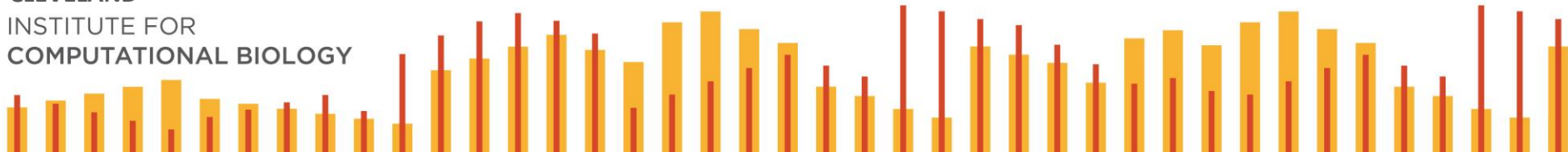
CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY



PRECISION MEDICINE RESEARCH IS MULTIDISCIPLINARY

Biostatistics
Computer Genomics
Medicine Omics
Human Biomedical
Clinical Data Big
Bioethics Science
Bioinformatics
Informatics
Genetics

CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY



WHAT DEGREES AND TRAINING DO YOU NEED?



Genetics and Molecular Biology
Emory University
2000

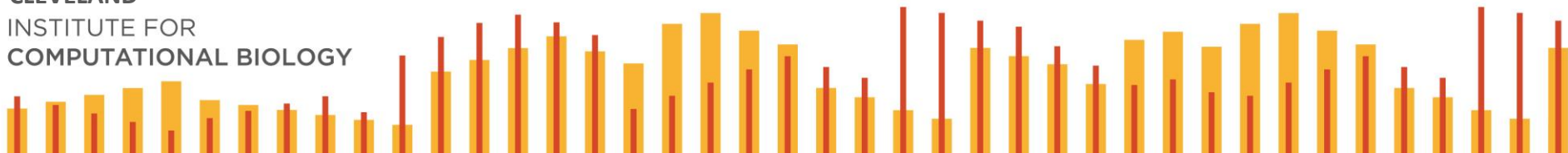


EIS 2000-2002



Genome Sciences
University of Washington
2002-2006

CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY



WHAT DEGREES AND TRAINING DO YOU NEED?

PHDs IN HUMAN GENETICS

Logan



Research
Scientist
(Academia)

Janina



Industry
Scientist

Matt



Staff
Scientist
(Industry)

Jennifer



Technical
Writer

Nicole



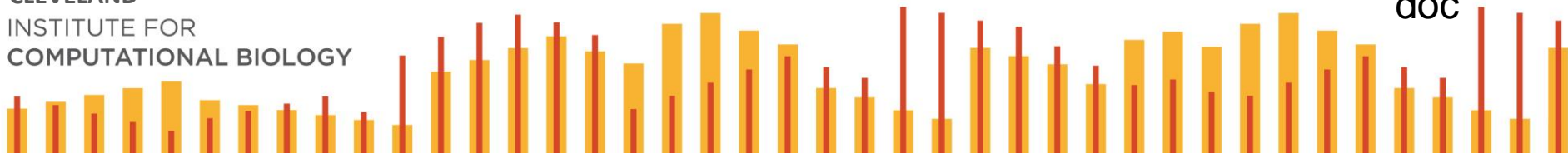
Staff
Scientist
(Industry)

Brittany



NIH Post-
doc

CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY



WHAT DEGREES AND TRAINING DO YOU NEED?

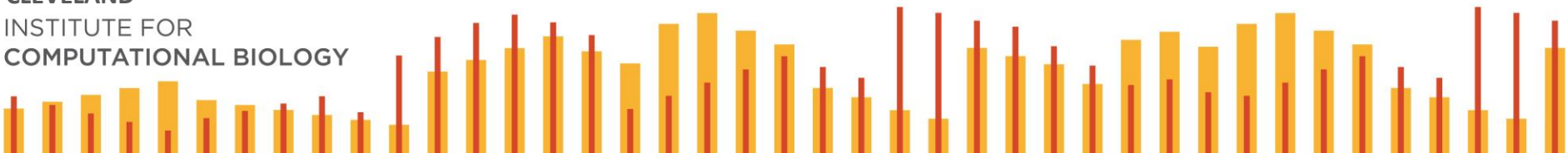
BIOETHICS AND HUMAN GENETICS



Aaron Goldenberg, PhD
Associate Professor
Case Western Reserve University



Jessica Cooke Bailey, PhD
Assistant Professor
Case Western Reserve University



WHAT DEGREES AND TRAINING DO YOU NEED?

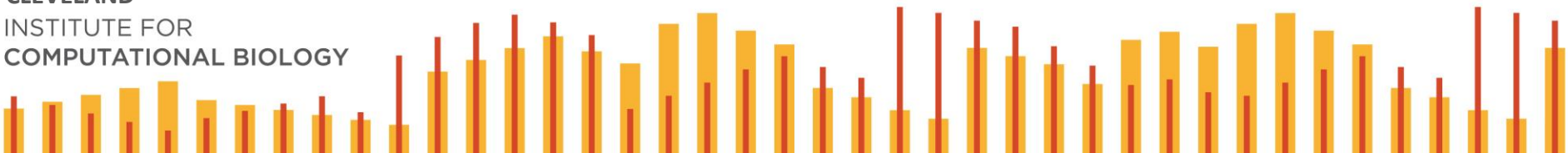
BIOINFORMATICS, COMPUTATIONAL BIOLOGY, COMPUTER SCIENCE



Sarah Pendergrass, PhD, MS
Staff Scientist
Genentech



Will Bush, PhD, MS
Associate Professor
Case Western Reserve University



WHAT DEGREES AND TRAINING DO YOU NEED?

DOCTORATE OF MEDICINE



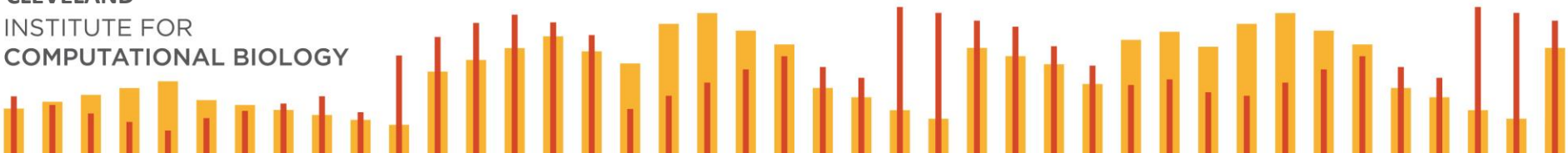
John Sedor, MD



John O'Toole, MD

Nephrologists
Cleveland Clinic

CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY



WHAT DEGREES AND TRAINING DO YOU NEED?

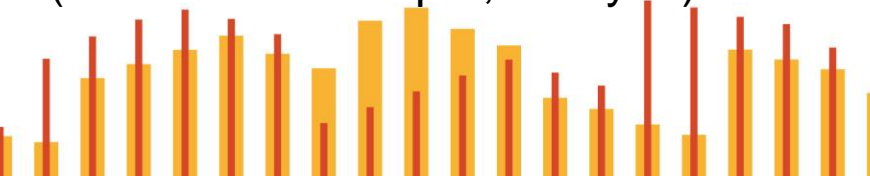
BIOSTATISTICS, DATA MANAGEMENT/ARCHITECTS

Kristin Brown-Gentry, MS
(Health Outcomes Scientist, Magellan Health)

Robert Goodloe, MS
(Consultant Statistician, Eli Lilly)

Bob McClellan, BS
(Senior Application Developer,
Vanderbilt University Medical Center)

Jonathan Boston, BS
(Software Developer, Cicayda)



QUESTIONS?



Pilot streams December 10, 2019

Season 1 (46 Chromosomes and a Mule)
streams February 4, 2020

Janina Jeff, PhD

CLEVELAND
INSTITUTE FOR
COMPUTATIONAL BIOLOGY

