



Harnessing Big Data and Precision Medicine – Intelligent & Integrative Data Analysis for Health Impact



Presented by: **Zeeshan Ahmed, Ph.D.**

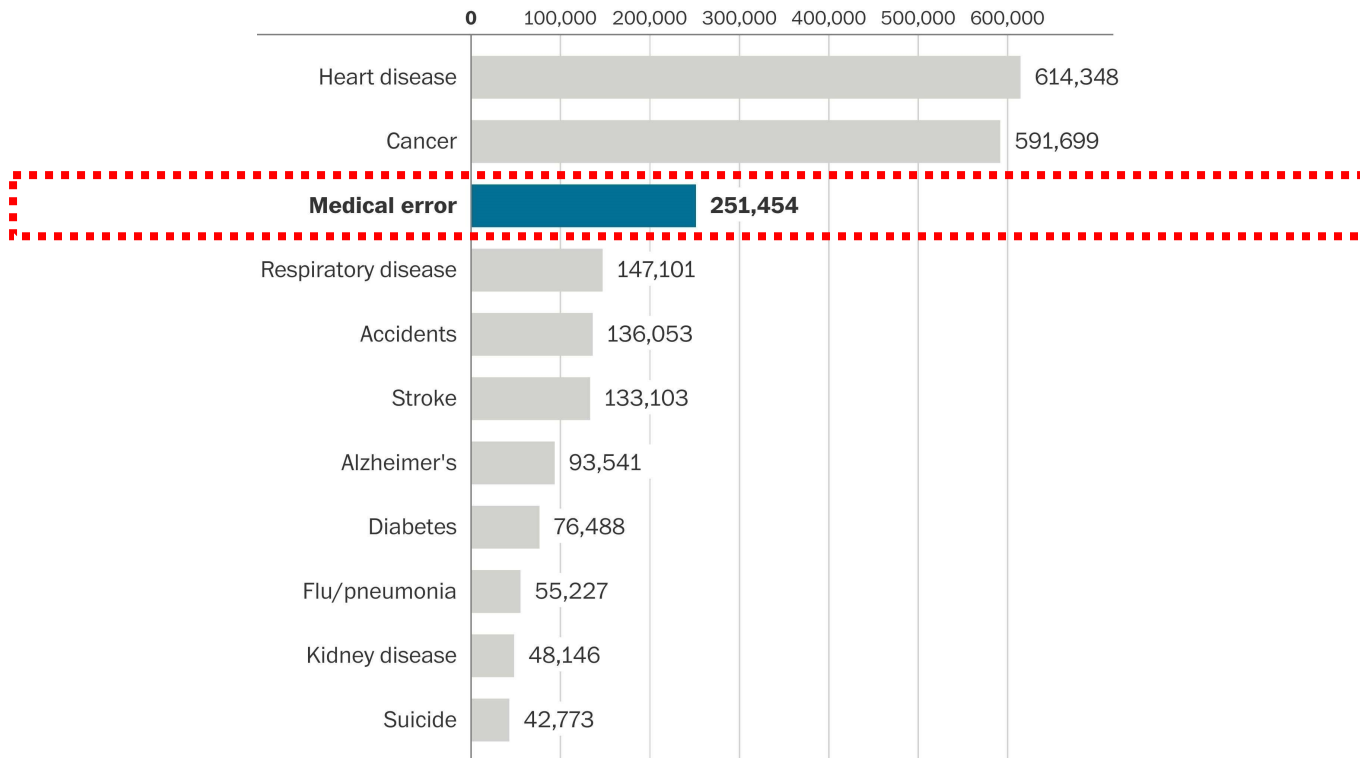


**OAK RIDGE
INSTITUTE
FOR SCIENCE
AND EDUCATION**

One of three leading causes of deaths

Death in the United States

Johns Hopkins University researchers estimate that medical error is now the third leading cause of death. Here's a ranking by yearly deaths.



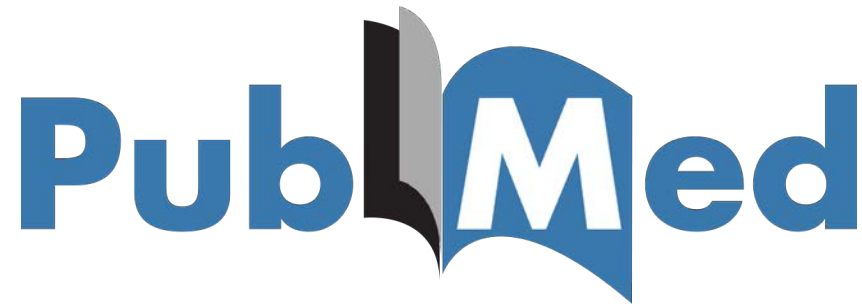
Source: National Center for Health Statistics, BMJ

THE WASHINGTON POST

Articles in PMC - NCBI

Medication error: Over 200,000

Delayed treatment: Over 600,000



URL: <https://www.ncbi.nlm.nih.gov/pmc/>



Precision Medicine World Conference #PMWCintl Nov 20
FDA approves a #PrecisionMedicine drug that targets a key genetic driver of the aggressive cancer b6jy20b0k1. Learn the latest in #PrecisionMedicine and #PrecisionMedicine Jan 20-23 at b6jy20b0k1 #PMWC19



Trending

The NEW ENGLAND JOURNAL OF MEDICINE
Perspective

A New Initiative on Precision Medicine

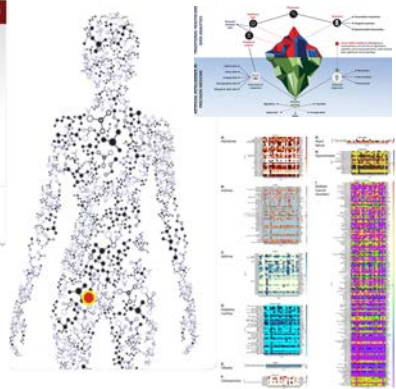
Francis S. Collins, M.D., Ph.D., and Harold Varmus, M.D.
N Engl J Med 2015; 372:793-795 | February 26, 2015 | DOI: 10.1056/NEJMp1500523

IN THIS SUPPLEMENT
• Editorial
• Article
• Sponsor page

Precision medicine integrates research disciplines and clinical practice to guide individualized patient care. The highlight examines the emerging framework that is bringing together researchers, clinical laboratories, clinicians and patients in a precision-medicine ecosystem: progress in pharmacogenomics and gene therapy towards personalized treatments, and the need to redesign clinical trials to match patients with the most suitable trial.

Credit: NIH Sponsor

Widening the Path of Precision Cancer Medicine b2zpm/kanjkt

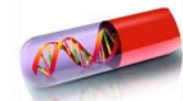


NCI CCR GU Malignancies @nciccr_gmb · Nov 25
William Figg1 and colleagues publish on precision medicine applications in prostate cancer. @PCF_Science



Precision medicine applications in prostate cancer
Aided by developments in diagnostics and therapeutics, healthcare is increasingly moving toward precision medicine, in which treatment is customized to each individual.

ncbi.nlm.nih.gov



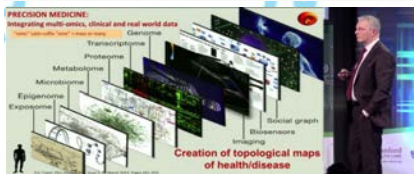
Rutgers Researchers Use Artificial Intelligence to Predict Cardiovascular Disease

Precision Med US @PrecisionMedUS · Nov 25
The latest The Precision Medicine Daily! paper by/PrecisionMedUS. Thanks to @FractureHealth @stratigist72 @ERStalk @HealthTech Foundation

Algorithms will out
dialconomy.com They anticipate the needs paper.

Precision Medicine

Understanding Health Better Will Take "All of Us"



POST Online Media @poandpo · 4h
Genomic study brings us closer to precision medicine for type 2 diabetes
#genomicstudy #ins-sickess-an... #health #people #medicine #patients #doctors #research #information #trending poandpo.com



Genomic study brings us closer to precision medicine for type 2 diabetes...
Most patients diagnosed with type 2 diabetes are treated with a "one-size-fits-all" protocol that is not tailored to each person's physiology and may

poandpo.com

National Human Genome Research Institute @genome.gov · Oct 10
Next-generation sequencing, a technique used to rapidly scan a person's genome, is within reach for medical applications. #NHGRI's Dr. Adams explains how it could impact the diagnosis and treatment of patients in @FrontMedicine #NGS #sequencing #genetics bit.ly/2Ea5q33

Next-Gen Sequence in Clinical Diagnosis

Dr. David Adams with NEJM's Stephen Morrissey

Precision Medicine World Conference Retweeted
John Miles @MilesJohn · Oct 15
The Impact of Precision #medicine. #PMWCintl @precisionmedcin @PrecisionMed @PrecisionMedicine @Thomas_Wilkins

It is already changing lives. We're seeing this very, very big impact of being able to save lives or extend lives. That's what I get excited about because, from an epidemiological perspective, to be able to watch whole populations of people live better, live longer and get cured of their illnesses through a targeted approach, that's super exciting to me.

Kristin Potbier

PMWC®

SILICON VALLEY JAN. 20-23

PRECISION MEDICINE WORLD CONFERENCE 2019

cancerclinicalresearch @cancerclinicalr · 1m
FDA approves 'precision medicine' drug for different cancers with same mutation
#PrecisionMedicine #Cancers #Mutation
washingtonpost.com/health/2018/11...
alliedacademies.org/journal-cancer...

Machine Learning

Predicting everything that can be predicted from the genetic code

Precision Medicine World Conference @PMWCintl · Nov 2
C.June @Penn #PMWC19 Honoree/Presenter:
Clinical implementation of #cancer #immunotherapy challenges: a tumor needs to be fully characterized by sequencing at the DNA/RNA levels, but only a minority of patients have their tumors fully characterized...."



Interview with Carl June, University of Pennsylvania...
Q: The track theme is on the topic "How do we accelerate and deliver on the promise of cancer immunotherapy?"
What are some key promises regarding immune-
pmwcintl.com



10 health IT happenings we're thankful for in 2018
From innovations and emerging technology to the promise of regulatory relief for hospitals, healthcare has plenty to be grateful for this week and year ...
healthcareitnews.com



The Future of Precision Medicine

No One-Size-Fits-All Artificial Intelligence Approach Works for Prevention, Diagnosis or Treatment Using Precision Medicine



“Doctors have always recognized that every patient is unique, and doctors have always tried to tailor their treatments as best they can to individuals. You can match a blood transfusion to a blood type — that was an important discovery. What if matching a cancer cure to our genetic code was just as easy, just as standard? What if figuring out the right dose of medicine was as simple as taking our temperature?”

President Obama, January 30, 2015

<https://obamawhitehouse.archives.gov/precision-medicine>

Precision Medicine Initiative



Institute for Health, Health Care Policy and Aging Research
Rutgers University-New Brunswick

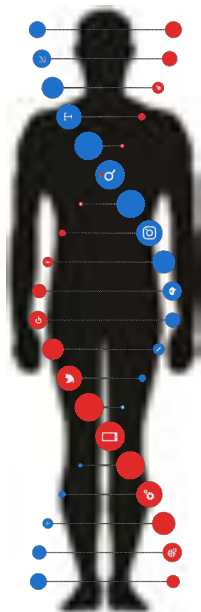


Robert Wood Johnson Medical School
Rutgers Biomedical and Health Sciences

Precision Medicine ?



Genomics



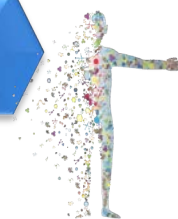
Clinics



Phenotype

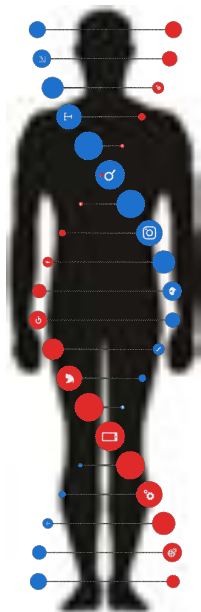


Metabolomics





Genomics



Clinics



Phenotype

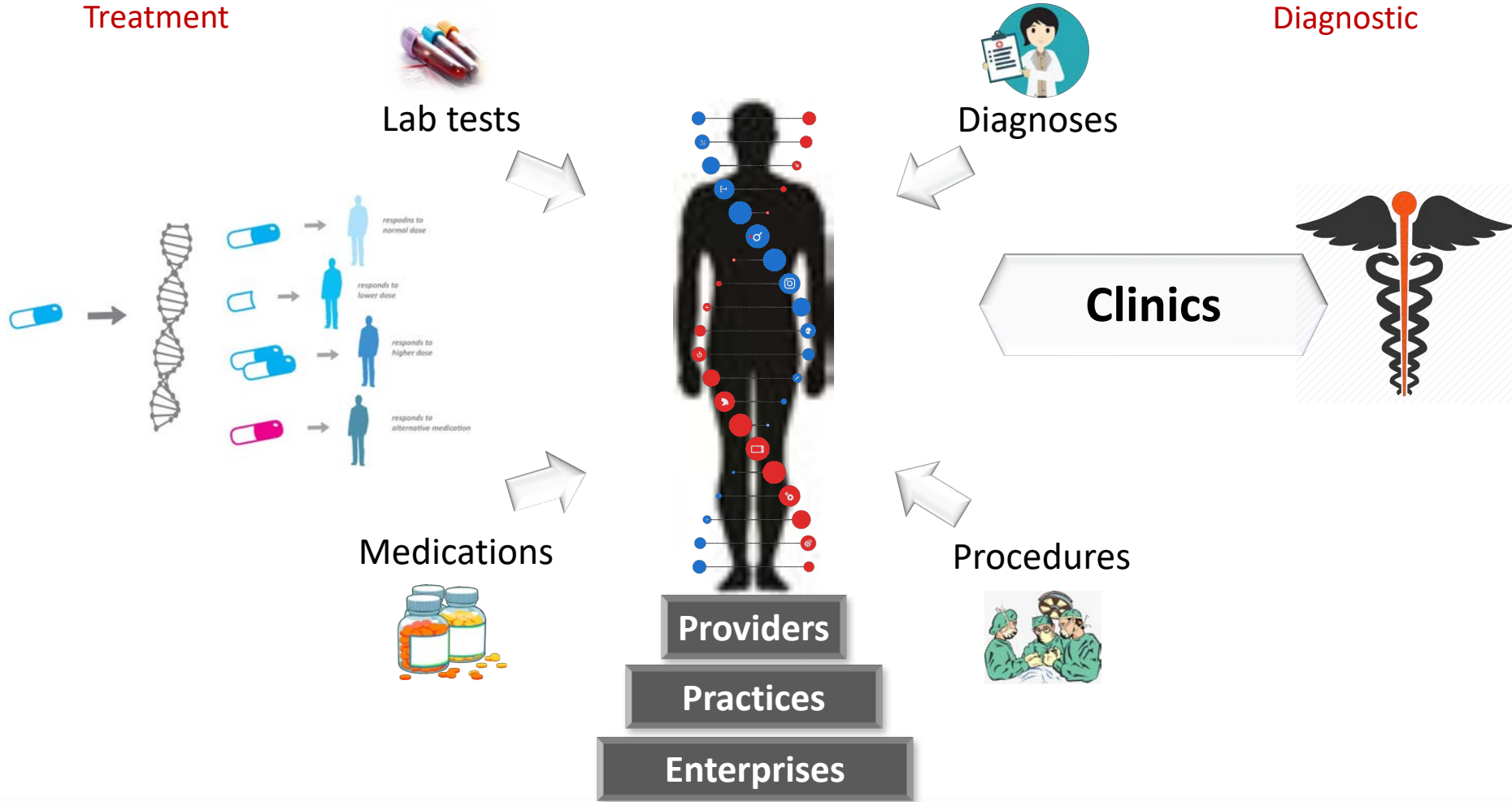
Metabolomics



Healthcare Data and Precision Medicine

Personalized Treatment

Predictive Diagnostic



Integrating Sequence Data and Precision Medicine

Personalized Treatment

Predictive Diagnostic

A portion of DNA molecule that serves as the basic unit of heredity.

Permanent alteration of the nucleotide sequence of the genome.

Genes

Mutations

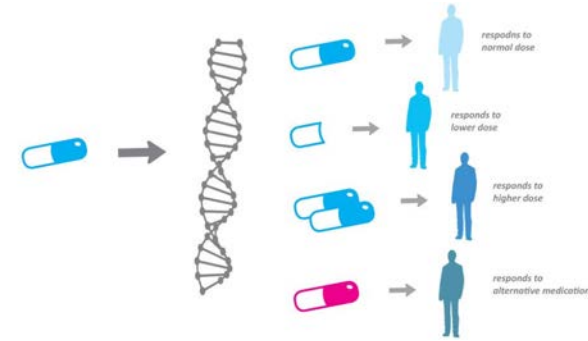
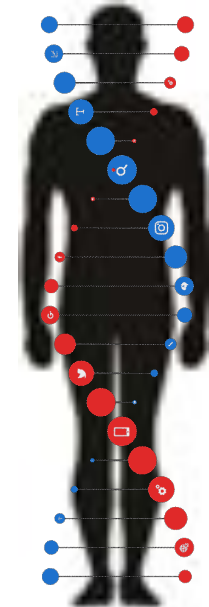
Genomics

SNPs

Transcripts

DNA sequence variation occurring in a single nucleotide – A (Adenine), C (Cytosine), G (Guanine), T (Thymine).

Protein sequence transcribed from DNA that forms different types of RNA and protein.



DNA

RNA

Protein

100 Years of evolving gene–disease complexities and scientific debutants

Saman Zeeshan*, Ruoyun Xiong*, Bruce T. Liang and Zeeshan Ahmed

Corresponding author. Zeeshan Ahmed, Department of Genetics and Genome Sciences, School of Medicine, University of Connecticut Health Center, 263 Farmington Ave, Farmington, CT 06032, USA. Tel.: +1-860-679-2643; Fax: +1-860-679-8345; E-mail: (zahmed@uchc.edu)

*These authors are equally contributing first authors.

Abstract

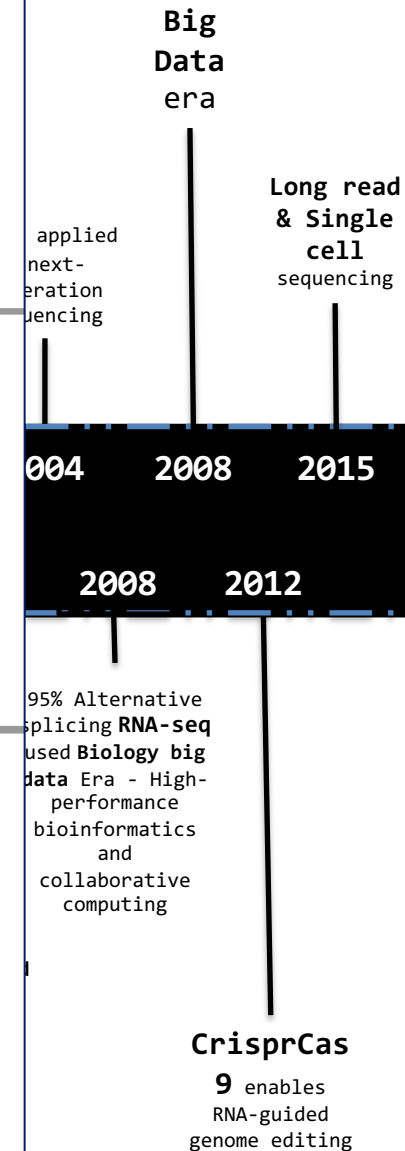
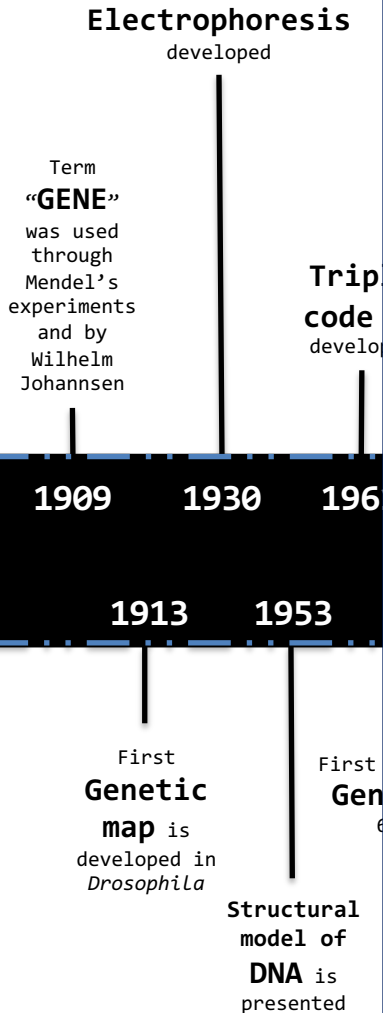
It's been over 100 years since the word 'gene' is around and progressively evolving in several scientific directions. Time-to-time technological advancements have heavily revolutionized the field of genomics, especially when it's about, e.g. triple code development, gene number proposition, genetic mapping, data banks, gene–disease maps, catalogs of human genes and genetic disorders, CRISPR/Cas9, big data and next generation sequencing, etc. In this manuscript, we present the progress of genomics from pea plant genetics to the human genome project and highlight the molecular, technical and computational developments. Studying genome and epigenome led to the fundamentals of development and progression of human diseases, which includes chromosomal, monogenic, multifactorial and mitochondrial diseases. World Health Organization has classified, standardized and maintained all human diseases, when many academic and commercial online systems are sharing information about genes and linking to associated diseases. To efficiently fathom the wealth of this biological data, there is a crucial need to generate appropriate gene annotation repositories and resources. Our focus has been how many gene–disease databases are available worldwide and which sources are authentic, timely updated and recommended for research and clinical purposes. In this manuscript, we have discussed and compared 43 such databases and bioinformatics applications, which enable users to connect, explore and, if possible, download gene–disease data.

Key words: gene; disease; databases; bioinformatics; precision medicine

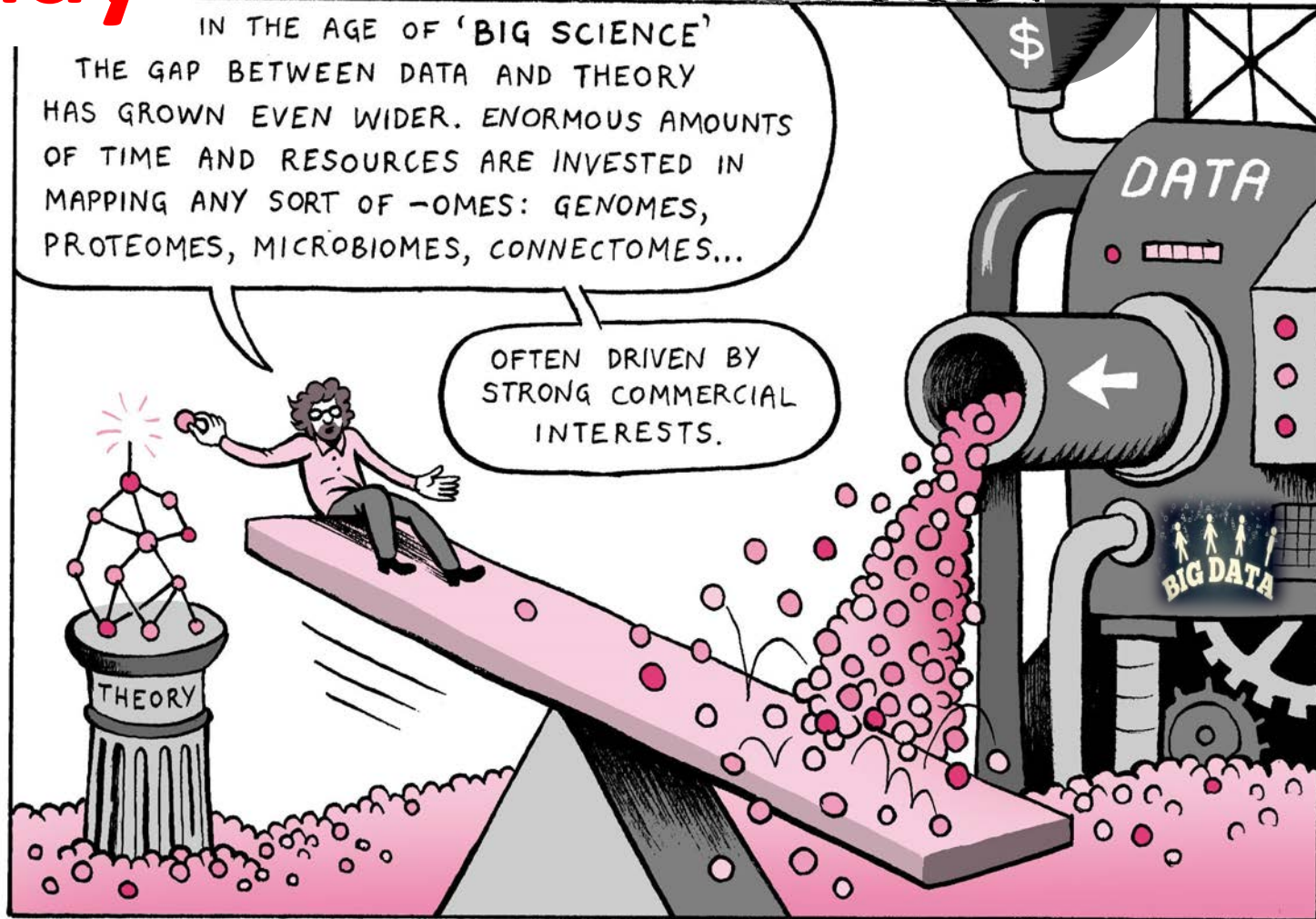
Introduction

Despite all of the scientific knowledge, much of medicine is still based on the treatment of symptoms and performing learned trials based on treatments, which works for most patients. Genetic research is assisting in producing tailored solutions to each individual, rather than what works for the average population, and understanding who is at risk for critical diseases like diabetes, high blood pressure or cancer. The variability in human genome sequence is a result of the biological code responsible for the development and functioning of a human being [1–6]. The complexity of human deoxyribonucleic acid (DNA) is a measure of the information contained within the DNA, and the maximal information possible in a solution of genomic DNA purified from a tissue or cell is equivalent to the total number of base pairs (bps)

present in the haploid genome [7–12]. The majority (~62%) of the human genome comprises of intergenic regions, the non-protein coding parts of the genome that lie between genes, used to be called 'junk DNA', but now genome research over the past few years has revealed functions associated with these regions, suggesting that every part of the genome may have some importance [13–23]. Intergenic DNA may also include gene regulatory sequences [24–31], such as promoters [32–37], enhancers [38–43] and silencers [44–46] that have yet to be characterized. Ribonucleic acid (RNA) [47–54] is the transcribed form of DNA and messenger RNA (mRNA) is the protein-coding form of RNA [55,56]. Non-coding RNAs, such as transfer RNA (tRNA) [57–59], micro RNA (miRNA) [60–62], ribosomal RNA (rRNA) [63–67] and long non-coding RNA (lncRNA) [68–71], play various roles in the cell, from protein translation to gene regulation.



Today



**Research
Investigators**

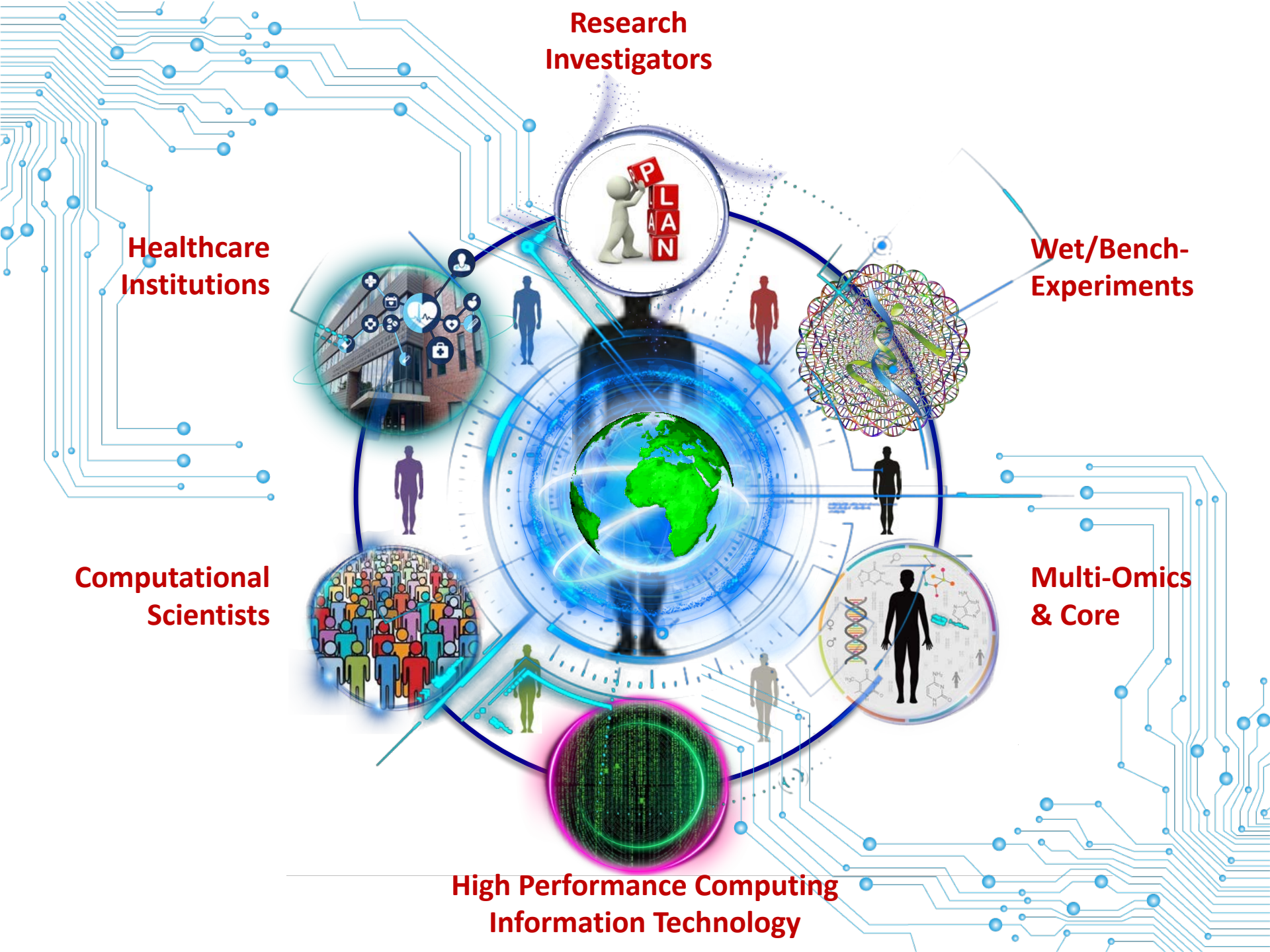
**Wet/Bench-
Experiments**

**Multi-Omics
& Core**

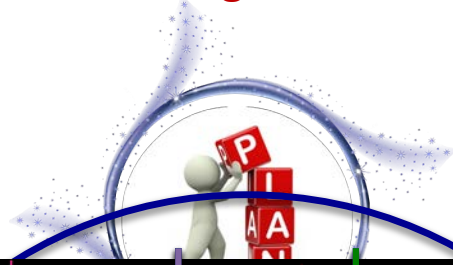
**Healthcare
Institutions**

**Computational
Scientists**

**High Performance Computing
Information Technology**



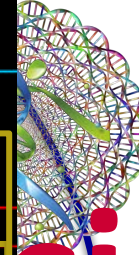
**Research
Investigators**



**Healthcare
Institutions**



**Wet/Bench-
Experiments**



Precision Medicine ?



**Computational
Scientists**



**Multi-Omics
& Core**



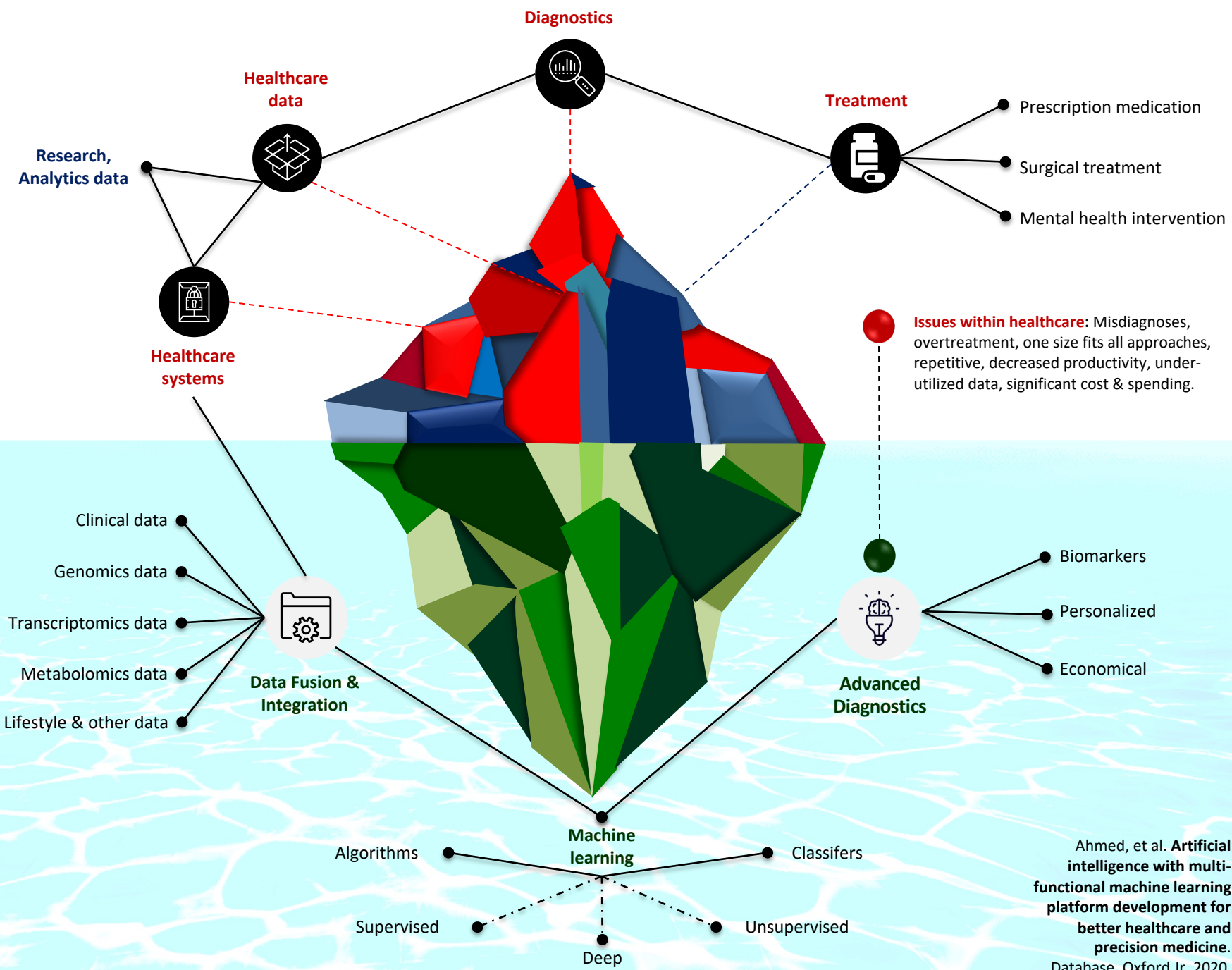
**High Performance Computing
Information Technology**





Precision Medicine ?

How such heterogenous data can be integrated, analyzed, and interpreted to support precision medicine and translational research in clinical settings and variable research environments?



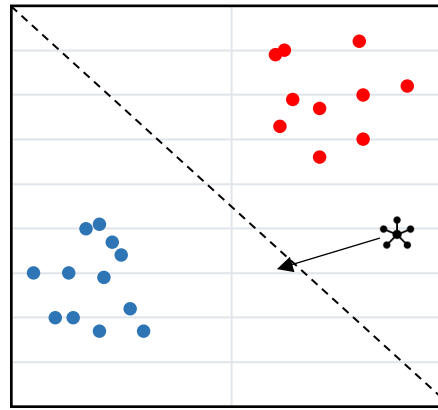
Ahmed, et al. **Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine.** Database. Oxford Jr. 2020.

Artificial Intelligence (AI) & Machine Learning (ML) Application

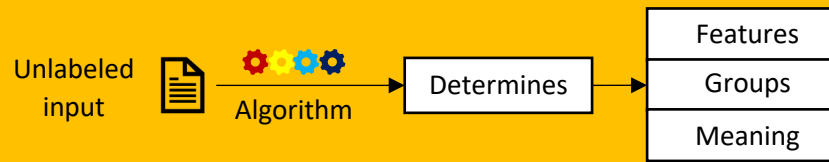
1. Data Classification.



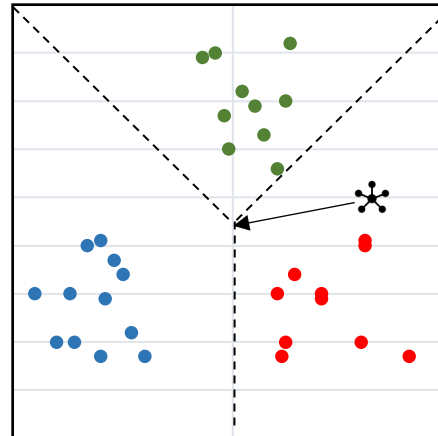
A training a model utilizing a set of labeled data to distinguish between positive and negative results e.g., determining if a biopsy sample is cancerous or not.



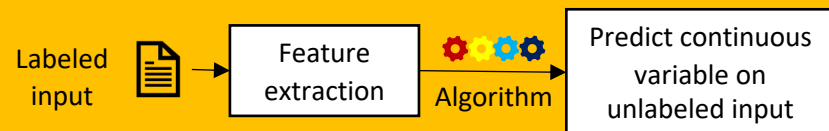
2. Data Cluster.



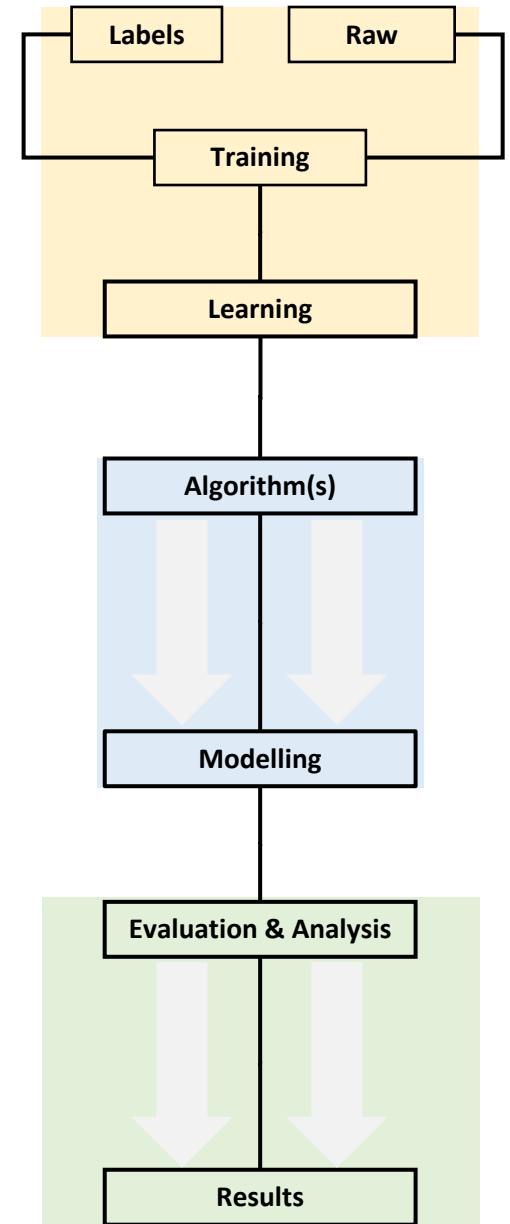
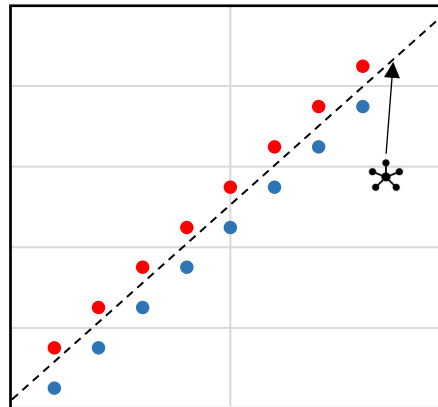
A model utilized to determine if any distinctive patterns are present without any determined outcome e.g., what is the prevalence of disease recurrence in a certain population due to pollution or chemical spill.



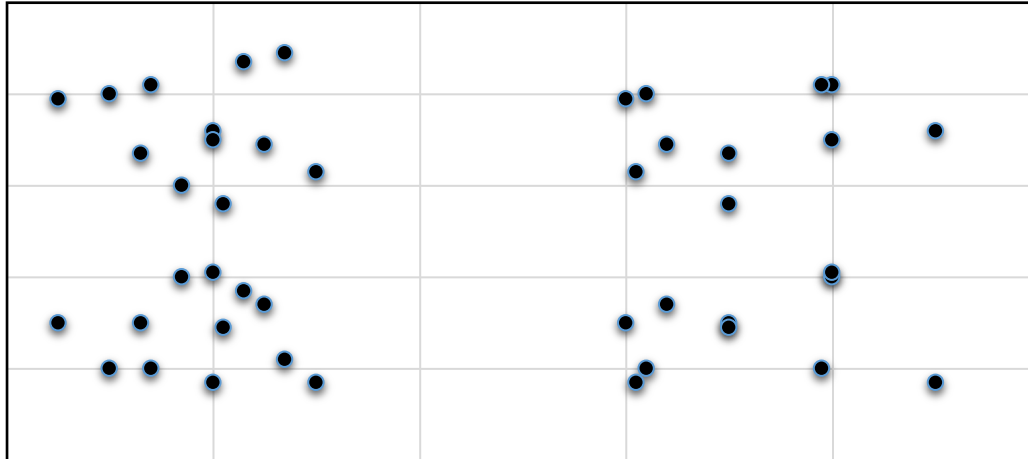
3. Data Regression.



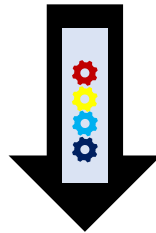
A predictive model used to examine and apply similar features obtained from a labeled data set to another data to make an accurate prediction e.g., how long before a patient is readmitted to the hospital following his/her discharge.



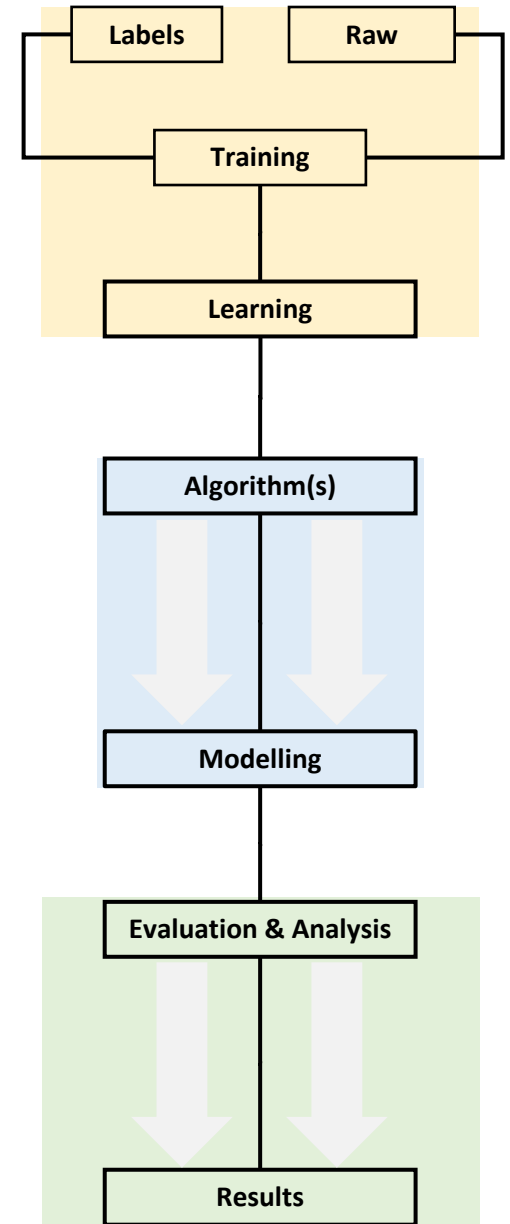
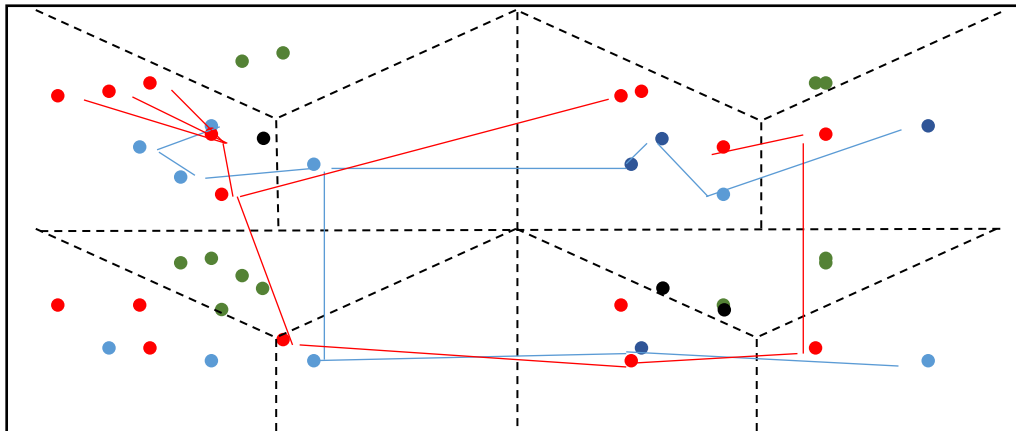
Artificial Intelligence (AI) & Machine Learning (ML) Application



Raw inputs reflecting non associated illness and symptoms expressed by one individual or distinct population.



Following the application of machine learning algorithms to multiple layers of data, we are able to generate meaningful connection between previously unrelated inputs



● Positive result

● Negative result

● Common relationship between dataset

✱ Rules determined by algorithms

Which AI/ML approach/algorithm is appropriate?

Artificial intelligence and machine learning approaches using gene expression and variant data for personalized medicine

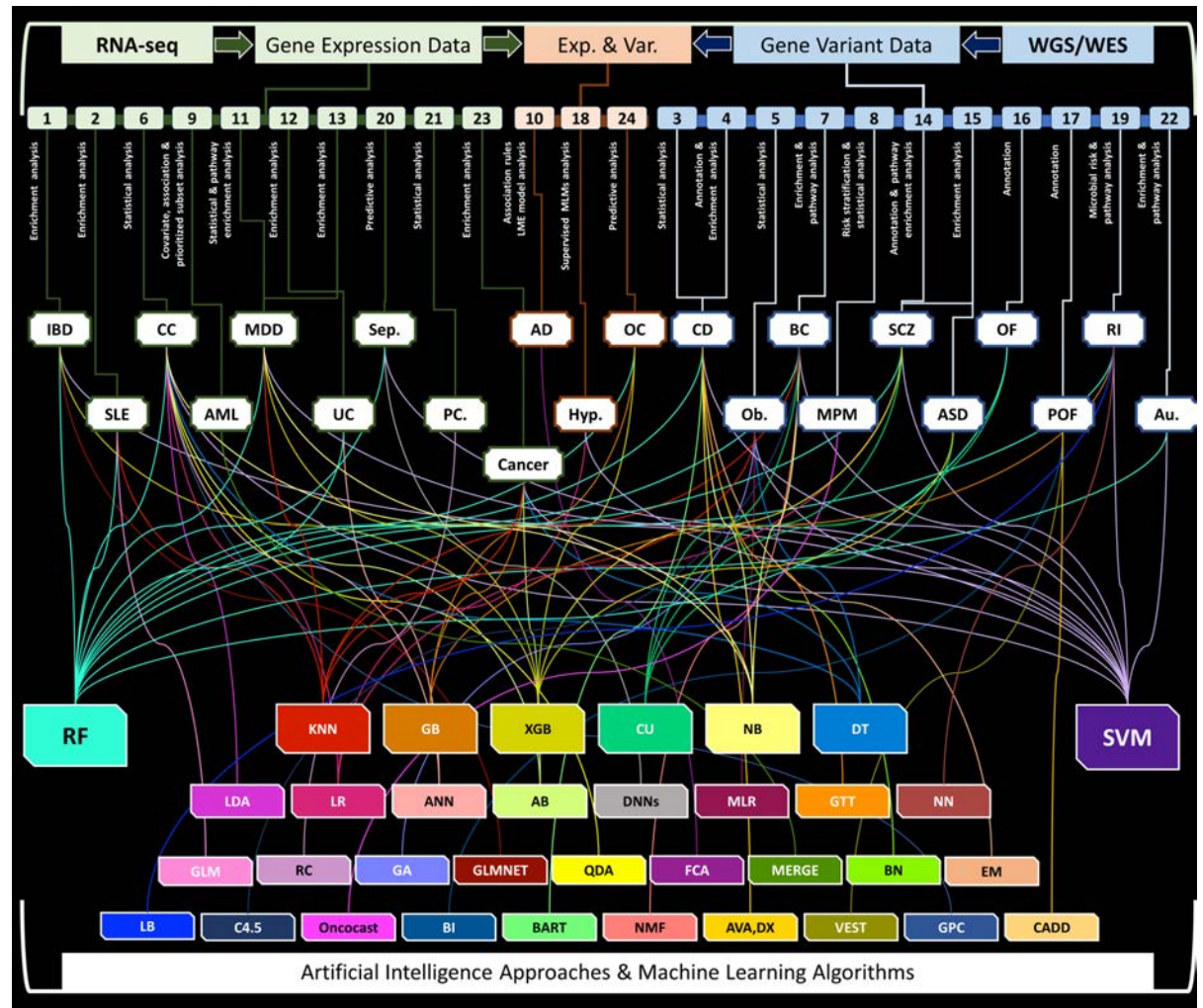
Sreyas Vadapalli¹, Habiba Abdelhalim¹, Saman Zveshan and Zeehan Ahmed¹

Abstract
Precision medicine uses genetic, environmental and lifestyle factors to more accurately diagnose and treat disease in specific groups of patients, and it is considered one of the most promising medical efforts of our time. The use of genetics is arguably the most data rich and complex components of precision medicine. The grand challenge today is the successful assimilation of genetics into precision medicine that translates across different ancestries, diverse diseases and other distinct populations, which will require clever use of artificial intelligence (AI) and machine learning (ML) methods. Our goal here was to review and compare scientific objectives, methodologies, datasets, data sources, ethics and gaps of AI/ML approaches used in genomics and precision medicine. We selected high quality literature published within the last 7 years that were indexed and available through PubMed Central. Our scope was narrowed to articles that reported application of AI/ML algorithms for statistical and predictive analyses using whole genome and/or whole exome sequencing for gene variants, and RNA-seq and microarrays for gene expression. We did not limit our search to specific diseases or data sources. Based on the scope of our review and comparative analysis criteria, we identified 32 different AI/ML approaches applied in variable genomics studies and report widely adapted AI/ML algorithms for predictive diagnosis across several diseases.

Keywords: artificial intelligence, machine learning, gene expression, gene variant, predictive analysis

Introduction
Genetic studies can reveal biomarkers that diagnose, determine risk and predict treatment outcomes for a wide variety of diseases [1]. Most genetic research investigates biological insights, disease mechanisms and disease risks by comparing healthy and diseased populations, which can overlook individual and subgroup variations [2]. DNA and RNA sequencing (RNA-seq) are the two most used methods in genetic research. Genetic variation, which encompasses DNA (gene) and RNA (gene expression) differences, is a fundamental element to understanding the genetics of diseases [3]. DNA sequencing can identify associations between genomic variants and diseases [4, 5], while RNA-seq can identify associations between RNA expression variations and diseases [6]. Combining multiple gene variants and/or gene expression differences into polygenic biomarkers can increase predictive power. Indeed, high and low polygenic scores from DNA can assess the probability of getting diseases [7]. While promising, the grand challenge here is analyzing the huge volume of known (and unknown) variants and using this information to diagnose, determine risk and predict treatment outcomes within diverse groups of humans [4]. This challenge is being met with precision medicine, which aims to translate this vast pool of genetic data to enhance disease outcomes accurately and safely [8]. However, the successful implementation of precision medicine remains difficult for heterogeneous ancestry groups and other distinct populations [9]. The convergence of genomics data and staggering developments in artificial intelligence (AI) and machine learning (ML)

12022 Arxiv, 12 on arXiv:1903.08181v1[cs.LG]. Preprint. DOI: 10.1101/2019.03.08.290239. This version posted March 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Which AI/ML approach/algorithm is appropriate?

OXFORD

Briefings in Bioinformatics, 2022, 1–25
<https://doi.org/10.1093/bib/bbab294>
 Review

Artificial intelligence and machine learning approaches using gene expression and variant data for personalized medicine

Sreyas Vaidapalli¹, Habiba Abdelhalim¹, Saman Zeeshan and Zeeshan Ahmed¹

Corresponding author: Zeeshan Ahmed, Department of Medicine, Robert Wood Johnson Medical School, Rutgers Biomedical and Health Sciences, 125 Paterson St, New Brunswick, NJ 08902-1226, USA. Tel: +1 848 940 7866; Fax: +1 732 932 1253; E-mail: z.ahmed@rwjms.rutgers.edu
¹Sreyas Vaidapalli and Habiba Abdelhalim contributed equally.

Abstract

Precision medicine uses genetic, environmental and lifestyle factors to more accurately diagnose and treat disease in specific groups of patients, and it is considered one of the most promising medical efforts of our time. The use of genetics is arguably the most data rich and complex components of precision medicine. The grand challenge today is the successful assimilation of genetics into precision medicine that translates across different ancestries, diverse diseases and other distinct populations, which will require clever use of artificial intelligence (AI) and machine learning (ML) methods. Our goal here was to review and compare scientific objectives, methodologies, datasets, data sources, ethics and gaps of AI/ML approaches used in genomics and precision medicine. We selected high quality literature published within the last 5 years that were indexed and available through PubMed Central. Our scope was narrowed to articles that reported application of AI/ML algorithms for statistical and predictive analyses using whole genome and/or whole exome sequencing for gene variants, and RNA-seq and microarrays for gene expression. We did not limit our search to specific diseases or data sources. Based on the scope of our review and comparative analysis criteria, we identified 32 different AI/ML approaches applied in variable genomics studies and report widely adapted AI/ML algorithms for predictive diagnosis across several diseases.

Keywords: artificial intelligence, machine learning, gene expression, gene variant, predictive analysis

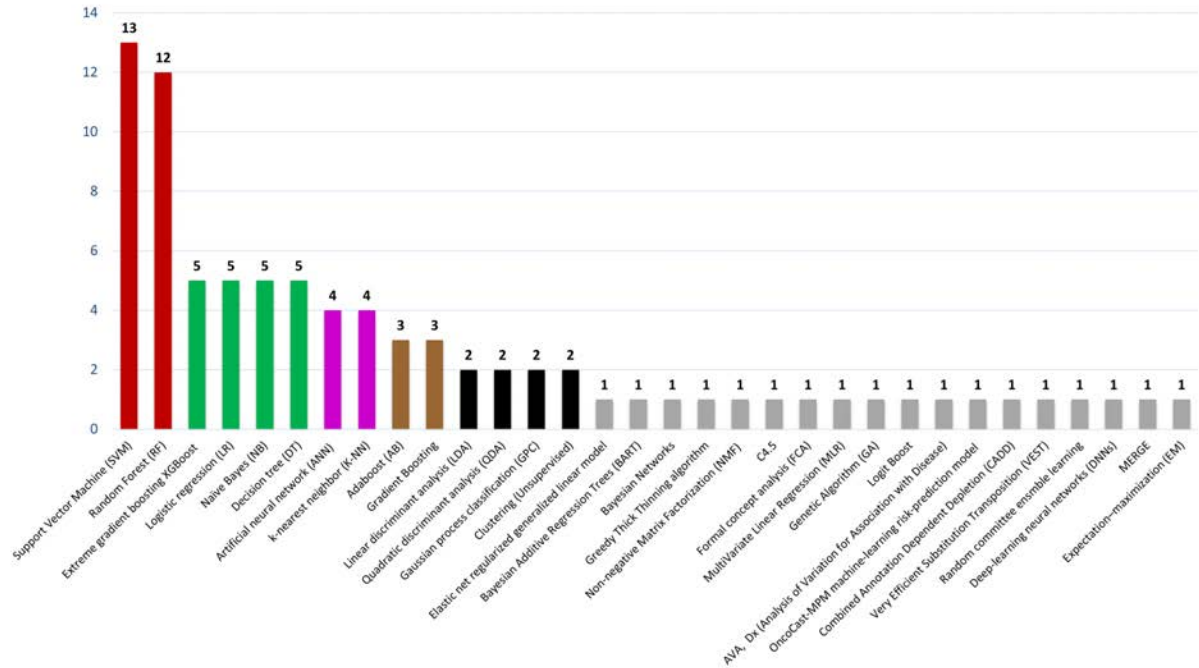
Introduction

Genetic studies can reveal biomarkers that diagnose, determine risk and predict treatment outcomes for a wide variety of diseases [1]. Most genetic research investigates biological insights, disease mechanisms and disease risks by comparing healthy and diseased populations, which can overlook individual and subgroup variations [2]. DNA and RNA sequencing (RNA-seq) are the two most used methods in genetic research. Genetic variation, which encompasses DNA (gene) and RNA (gene expression) differences, is a fundamental element to understanding the genetics of diseases [3]. DNA sequencing can identify associations between genetic variants and diseases [4, 5], while RNA-seq can identify associations between RNA expression variations and diseases [6]. Combining multiple gene variants and/or

gene expression differences into polygenic biomarkers can increase predictive power. Indeed, high and low polygenic scores from DNA can assess the probability of getting diseases [7]. While promising, the grand challenge here is analyzing the huge volume of known (and unknown) variants and using this information to diagnose, determine risk and predict treatment outcomes within diverse groups of humans [4]. This challenge is being met with precision medicine, which aims to translate this vast pool of genetic data to enhance disease outcomes accurately and safely [8]. However, the successful implementation of precision medicine remains difficult for heterogeneous ancestry groups and other distinct populations [9]. The convergence of genomics data and staggering developments in artificial intelligence (AI) and machine learning (ML)

Sreyas Vaidapalli is a research assistant at the Ahmed Lab, Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers University-New Brunswick, NJ.
 Habiba Abdelhalim is a research assistant at the Ahmed Lab, Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers University-New Brunswick, NJ.
 Saman Zeeshan is a bioinformatics research scientist and postdoctoral research associate at the Rutgers Cancer Institute of New Jersey, Rutgers University-New Brunswick, NJ.
 Zeeshan Ahmed is an assistant professor of Medicine - Therapeutic Track and Core Member at the Rutgers Institute for Health, Health Care Policy and Aging Research, and Department of Medicine - Division of General Internal Medicine, Rutgers Robert Wood Johnson Medical School, Rutgers Biomedical and Health Sciences, Rutgers University-New Brunswick. Zeeshan Ahmed is an adjunct assistant professor at the Department of Genetics and Genome Sciences, UC Davis School of Medicine, UC Davis Health, CA and a full academic member of the Rutgers Microbiology and Molecular Genetics, Center for Cancer Health Equity, Rutgers Cancer Institute of New Jersey, Rutgers Biomedical and Health Sciences, Rutgers University-New Brunswick.
 Received: March 8, 2022. Revised: April 2, 2022. Accepted: April 26, 2022.
 © The Author(s) 2022. Published by Oxford University Press. All rights reserved. For permissions, please email: journals.permissions@oup.com

Downloaded from <https://academic.oup.com/bib/advance-article-abstract/doi/10.1093/bib/bbab294/6511061> by guest on 12 May 2022



Precision Medicine Project @ Ahmed Lab

1. Technology Development

2. Multi/Disease Research

I. High Performance & Secure Computing Frameworks

for secure clinical and multi-omics/genomics data acquisition, processing, modelling, integration, sharing, and management.

II. Biomedical Informatics Applications

for electronic healthcare records (EHR) extraction, transfer, loading (ETL), and analysis. Patient recruitment, consenting, sample collection and management.

III. Bioinformatics Tool, Scripts, and Pipelines

for standard multi-omics/genomics (e.g., expression, variant, enrichment, pathway, metabolic flux etc.) data analysis and deep phenotyping.

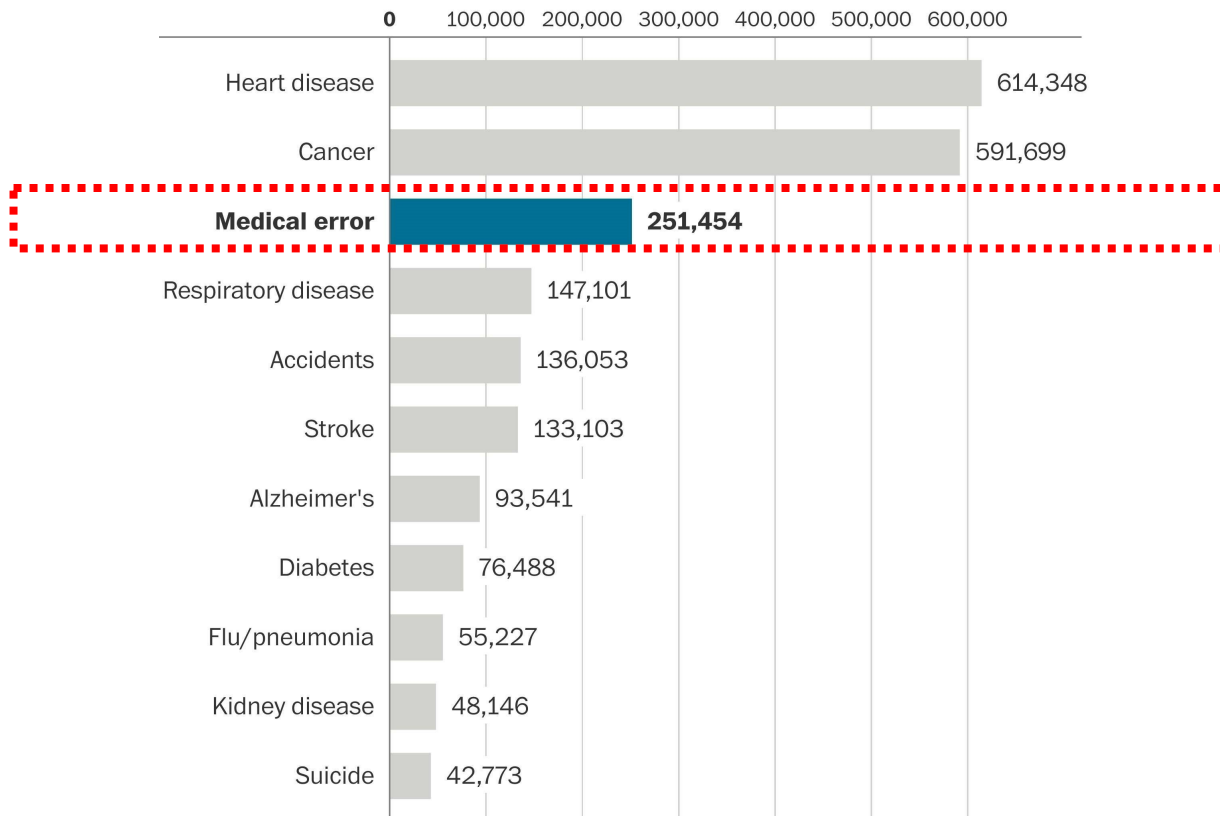
IV. Artificial Intelligence (AI) and Machine Learning (ML) Approaches

for the identification of patterns revealing predictive and modifiable risk factors to support earlier diagnosis of targeted disorders and sequela.

Multi/Disease Research @ Ahmed Lab

Death in the United States

Johns Hopkins University researchers estimate that medical error is now the third leading cause of death. Here's a ranking by yearly deaths.



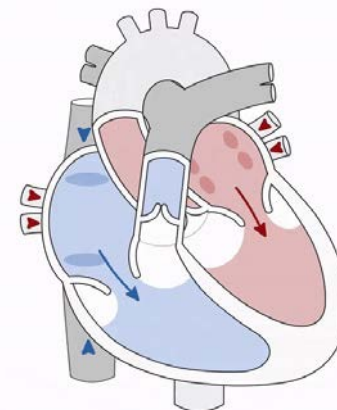
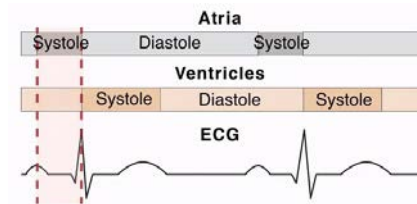
Source: National Center for Health Statistics, BMJ

THE WASHINGTON POST



Cardiovascular disease (CVD)

- **Heart Failure (HF)** and **Atrial Fibrillation (AF)** are among the most common manifestations of CVD and contribute to about 45% of all CVD deaths. *(Dickinson et al., 2014)*
 - **AF is an arrhythmic disorder** in the atrium of the heart, which can cause irregular heart rhythms. *(Staerk et al., 2017)*
 - **HF is a chronic disorder**, which weakens heart muscle and affects the regular function of the heart impairing its ability to pump enough oxygen-rich blood. *(Kalogirou et al., 2020)*
- Due to the complex nature, progression, inherent genetic makeup, and heterogeneity in CVDs, personalized treatments are critical for CVD patients.
- To improve the deciphering of CVD mechanisms, it will be necessary to systematically investigate known and identify novel genes that are responsible for the CVD development.
- Studying genetic insight with the application of Artificial Intelligence (AI), Machine Learning (ML), and state-of-the-art bioinformatics approaches can accelerate the processes of discovering disease causing variants and decode genetics of complex phenotypes to predict, prevent, and treat CVD.



Atrial Systole

Bioinformatics & CVD

Patel et al. *Human Genetics* (2023) 17:47
<https://doi.org/10.1186/s40246-023-00498-0>

Human Genetics

REVIEW

Open Access

Genomic approaches to identify and investigate genes associated with atrial fibrillation and heart failure susceptibility

Kush Ketan Patel^{1†}, Cynthia Venkatesan^{1†}, Habiba Abdelhalim³, Saman Zeeshan², Yuichiro Arima³, Suvi Linna-Kuosmanen^{4,5,6} and Zeeshan Ahmed^{2*}

Abstract

Atrial fibrillation (AF) and heart failure (HF) contribute to about 45% of all cardiovascular disease (CVD) deaths in the USA and around the globe. Due to the complex nature, progression, inherent genetic makeup, and heterogeneity of CVDs, personalized treatments are believed to be critical. To improve the deciphering of CVD mechanisms, we need to deeply investigate well-known and identify novel genes that are responsible for CVD development. With the advancements in sequencing technologies, genomic data have been generated at an unprecedented pace to foster translational research. Correct application of bioinformatics using genomic data holds the potential to reveal the genetic underpinnings of various health conditions. It can help in the identification of causal variants for AF, HF, and other CVDs by moving beyond the one-gene one-disease model through the integration of common and rare variant association, the expressed genome, and characterization of comorbidities and phenotypic traits derived from the clinical information. In this study, we examined and discussed variable genomic approaches investigating genes associated with AF, HF, and other CVDs. We collected, reviewed, and compared high-quality scientific literature published between 2009 and 2022 and accessible through PubMed/NCBI. While selecting relevant literature, we mainly focused on identifying genomic approaches involving the integration of genomic data; analysis of common and rare genetic variants; metadata and phenotypic details; and multi-ethnic studies including individuals from ethnic minorities, and European, Asian, and American ancestries. We found 190 genes associated with AF and 26 genes linked to HF. Seven genes had implications in both AF and HF, which are *SYNPO2L*, *TTN*, *MTSS1*, *SCNSA*, *PITX2*, *KLHL3*, and *AGAP5*. We listed our conclusion, which include detailed information about genes and SNPs associated with AF and HF.

Keywords Genes, Genetic loci, Atrial fibrillation, Heart failure, Cardiovascular diseases, Genomics, Multi-OMICs

[†]Kush Ketan Patel and Cynthia Venkatesan are equally contributing first authors

*Correspondence:

Zeeshan.Ahmed@rutgers.edu

¹Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers University, 112 Paterson St, New Brunswick, NJ, USA

²Rutgers Cancer Institute of New Jersey, Rutgers University, 195 Little Albany St, New Brunswick, NJ, USA

³Developmental Cardiology Laboratory, International Research Center for Medical Sciences, Kumamoto University, 2-2-1 Honjo, Kumamoto City, Kumamoto, Japan

⁴A. Viikari Institute for Molecular Sciences, University of Eastern Finland, 70211 Kuopio, Finland

⁵Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁶Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

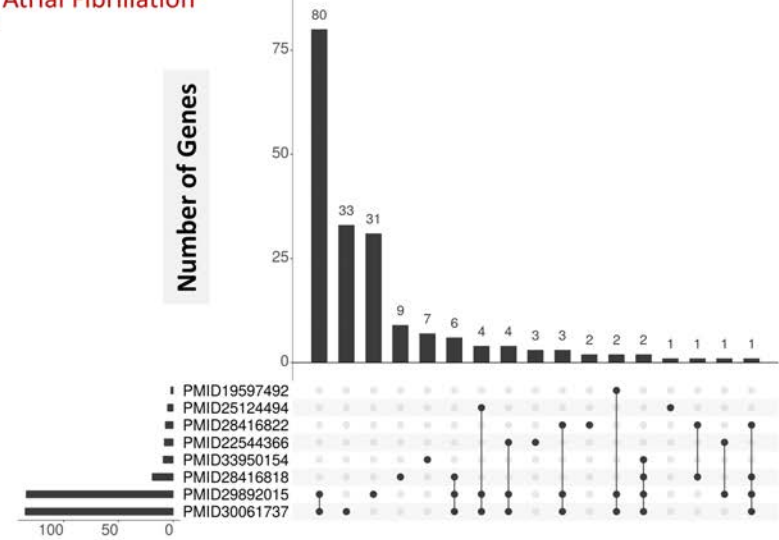
⁷Department of Genetics and Genome Sciences, UConn Health, 400 Farmington Ave, Farmington, CT, USA

⁸Department of Medicine, Robert Wood Johnson Medical School, Rutgers Biomedical and Health Sciences, 125 Paterson St, New Brunswick, NJ, USA



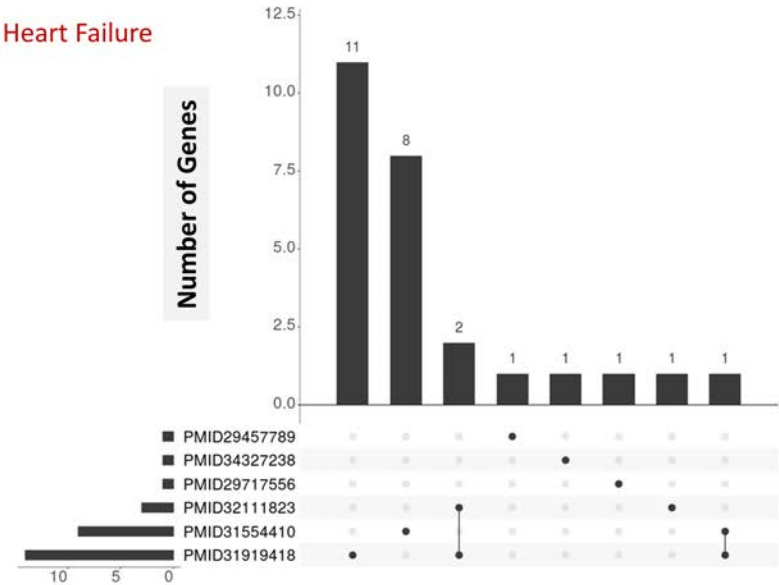
© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Number of Genes



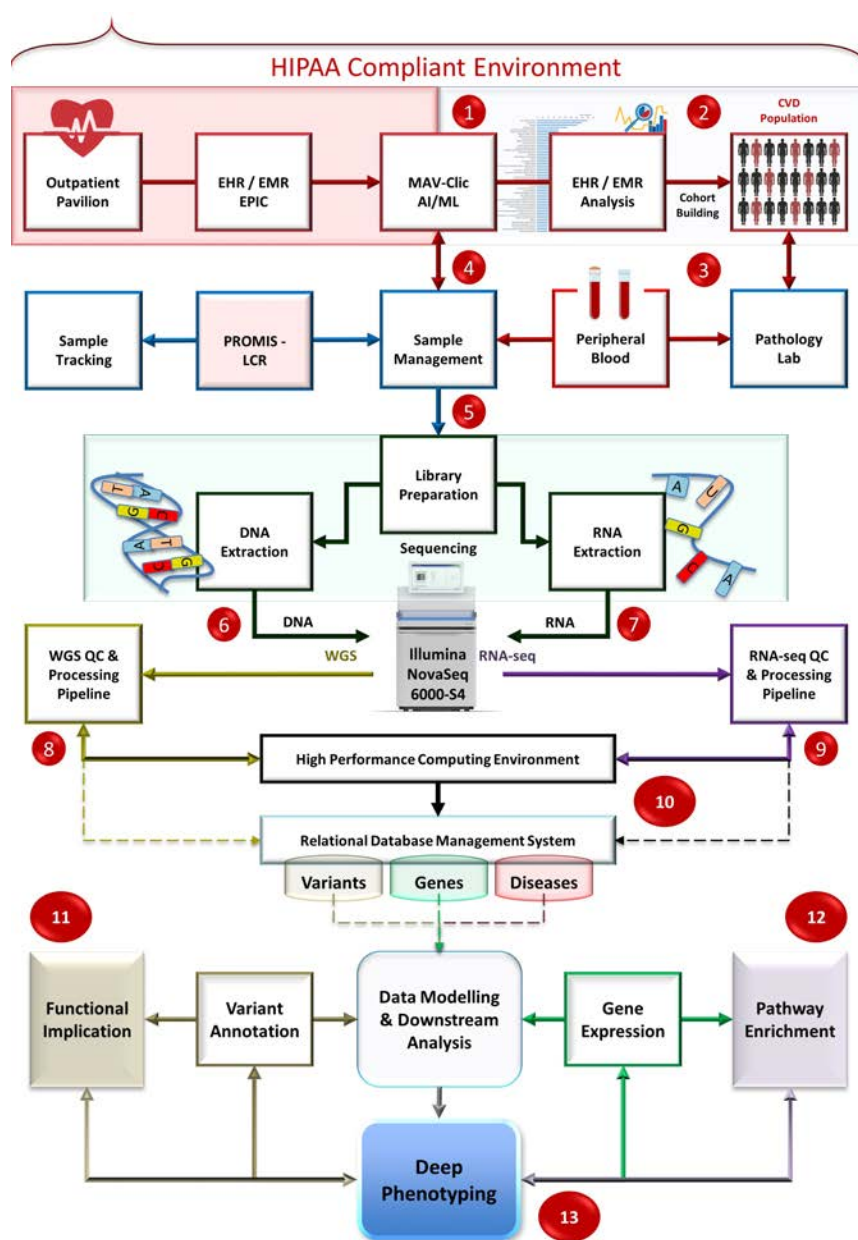
B. Heart Failure

Number of Genes



CVD Study Design

1. EHR extraction from EPIC
2. CVD cohort building
3. Consenting patients
4. Sample collection & management
5. Sequence data generation
6. Gene expression analysis (RNA-seq)
7. Gene-disease annotation
8. Variant analysis & validation (WGS)
9. AI/ML ready data generation (CIGT)
10. Predict disease with high accuracy



Cohort Building & MAV-clic

The CVD cohort include **40 male and 21 female individuals (n=61)**, aged between 45 to 92, with self-described race (42 Whites, 7 Blacks or African Americans, 1 Asian, and 11 of unknown race). In addition, the PI has built a control set, which included healthy individuals (n=10); 5 males and 5 females; out of which 9 were White race and 1 unknown race; aged between 28 to 78.

Database Notes

MAV-clic: management, analysis, and visualization of clinical data

Zeeshan Ahmed,¹ Minjung Kim² and Bruce T. Liang³

¹Department of Genetics and Genome Sciences, Institute for Systems Genomics, School of Medicine, University of Connecticut Health Center, Farmington, Connecticut, USA, ²The Pat and Jim Calhoun Cardiology Center, School of Medicine, University of Connecticut Health Center, Farmington, Connecticut, USA and ³Way Neag Distinguished Professor of Cardiovascular Biology and Medicine, Director Pat and Jim Calhoun Cardiology Center, Dean UConn School of Medicine, University of Connecticut Health Center, Farmington, Connecticut, USA.

Corresponding Author: Zeeshan Ahmed, PhD, Assistant Professor & Assistant Director Bioinformatics Precision Medicine Program, Department of Genetics and Genome Sciences, School of Medicine, University of Connecticut Health Center, 400 Farmington Ave, Farmington, CT 06032, USA (zrahmed@uconn.edu)

Received 30 January 2018; Revised 18 July 2018; Editorial Decision 3 November 2018; Accepted 22 November 2018

ABSTRACT

Objective: Develop a multifunctional analytics platform for efficient management and analysis of healthcare data.

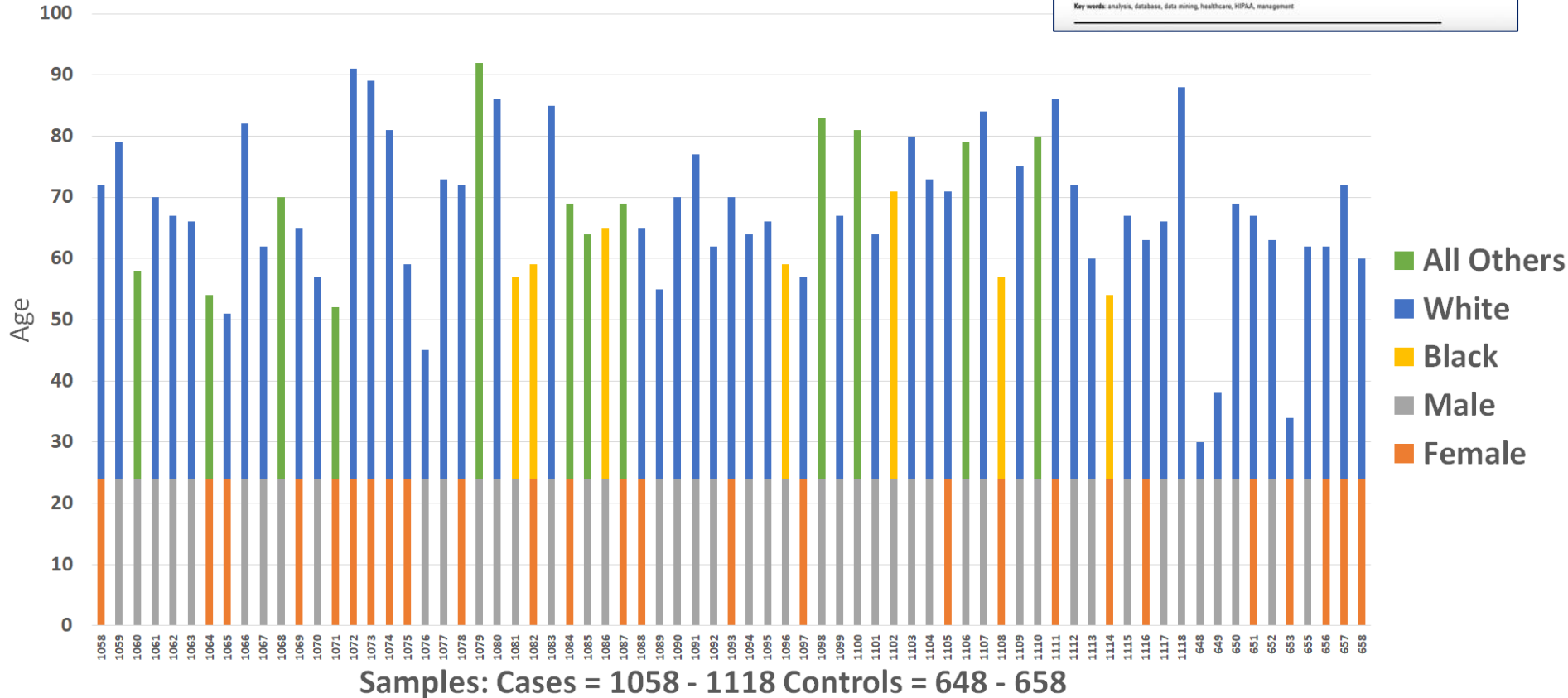
Materials and Methods: Management, Analysis, and Visualization of Clinical Data (MAV-clic) is a Health Insurance Portability and Accountability Act of 1996 (HIPAA)-compliant framework based on the Butterfly Model. MAV-clic extracts, cleanses, and encrypts data then restructures and aggregates data in a deidentified format. A graphical user interface allows query, analysis, and visualization of clinical data.

Results: MAV-clic manages healthcare data for over 800 000 subjects at UConn Health. Three analytic capabilities of MAV-clic include: creating cohorts based on specific criteria; performing measurement analysis of subjects with a specific diagnosis and medication; and calculating measure outcomes of subjects over time.

Discussion: MAV-clic supports clinicians and healthcare analysts by efficiently stratifying subjects to understand specific scenarios and optimize decision making.

Conclusion: MAV-clic is founded on the scientific premise that to improve the quality and transition of health-care, integrative platforms are necessary to analyze heterogeneous clinical, epidemiological, metabolomics, proteomics, and genomics data for precision medicine.

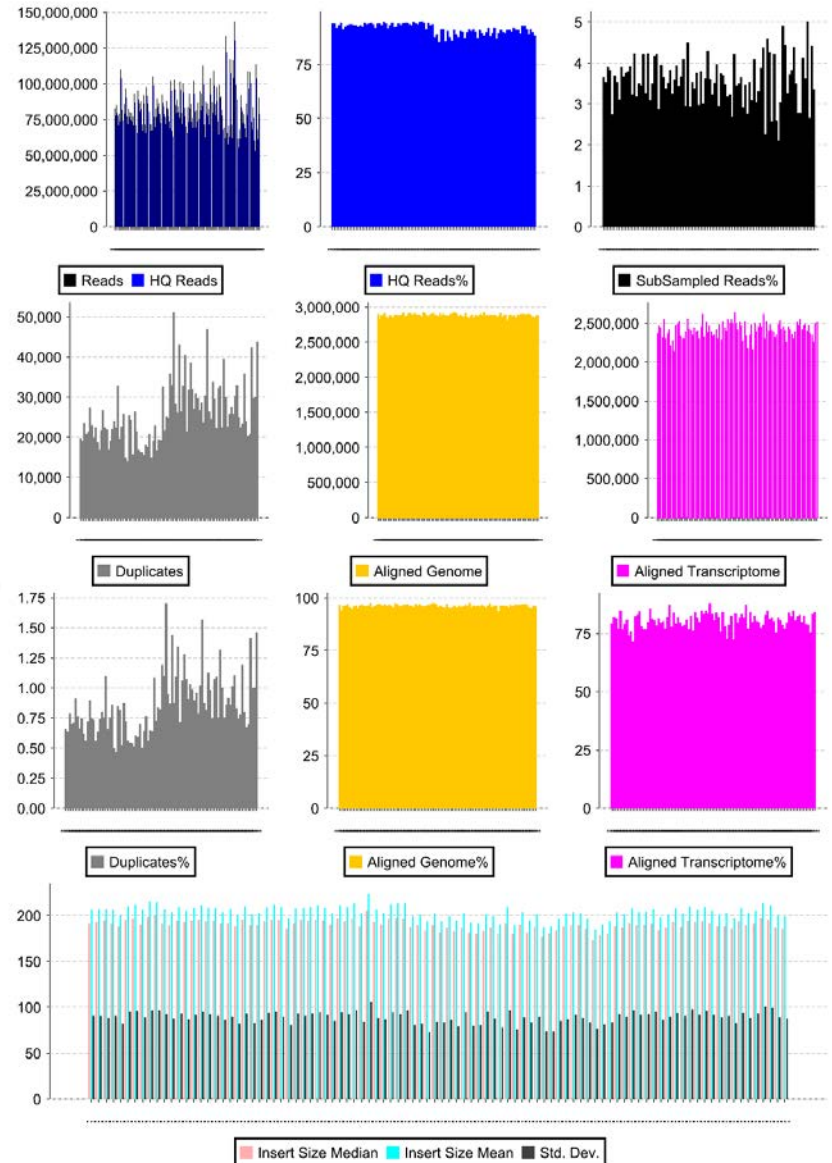
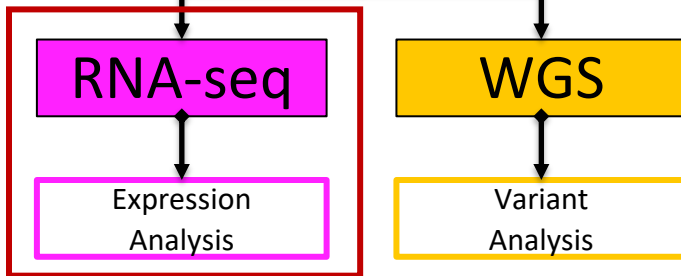
Key words: analysis, database, data mining, healthcare, HIPAA, management



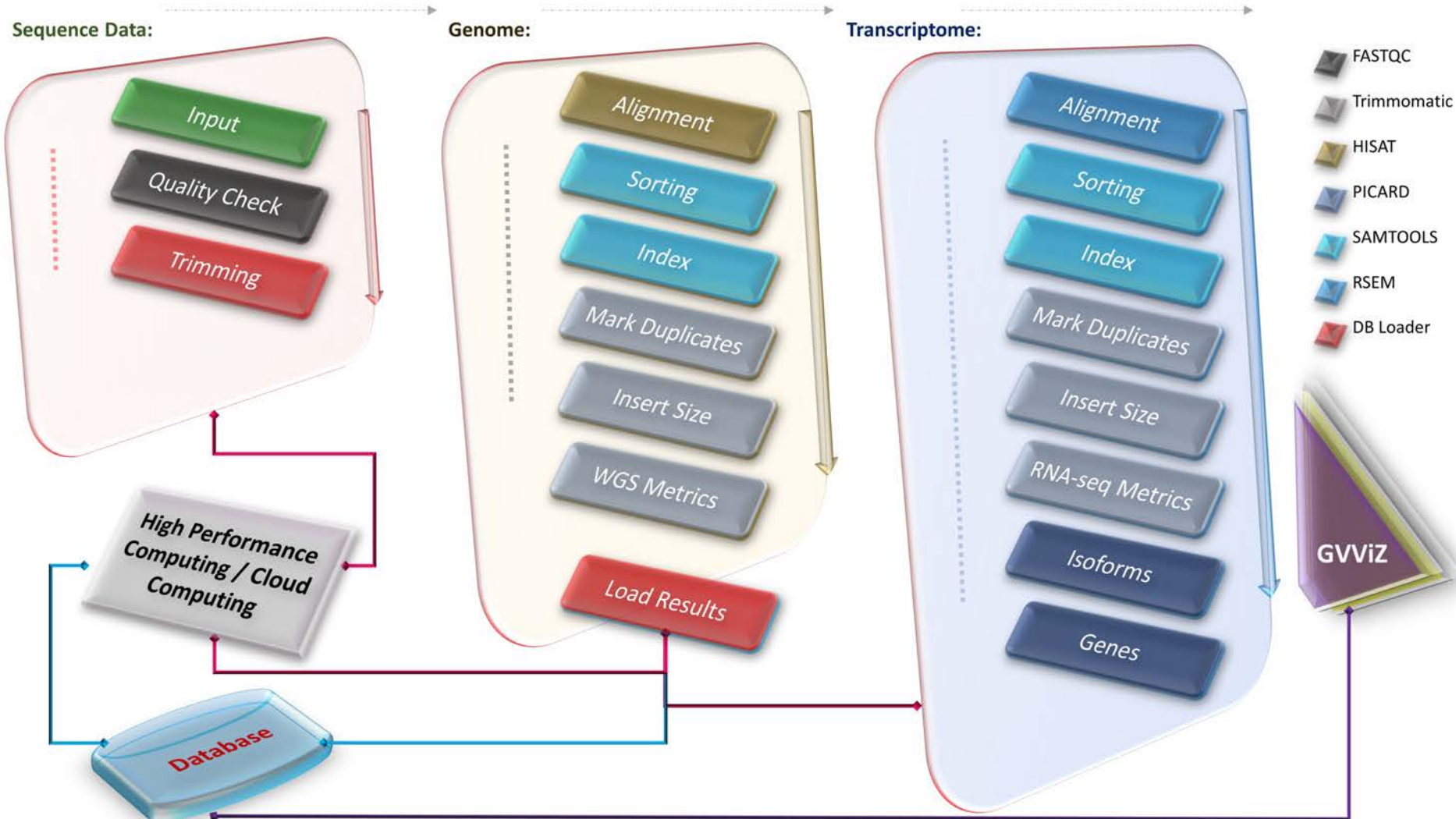
Sequence Data Generation

QC Report using PROMS-MED

ILLUMINA NOVASeq 6000-S4



GVViZ – RNA-seq Data Analysis



GViZ – Demo and Download Information



URL: <https://www.youtube.com/watch?v=xORroYpk8Nw>

URL: <https://github.com/drzeeshanahmed/GViZ-Public>

Ahmed et al. Human Genomics (2021) 15:37
<https://doi.org/10.1186/s40246-021-00336-1>

Human Genomics

PRIMARY RESEARCH

Open Access

Advancing clinical genomics and precision medicine with GViZ: FAIR bioinformatics platform for variable gene-disease annotation, visualization, and expression analysis

Zeeshan Ahmed^{1,2*}, Eduard Gibert Renart¹, Saman Zeeshan³ and XinQi Dong^{1,2}

Abstract

Background: Genetic disposition is considered critical for identifying subjects at high risk for disease development. Investigating disease-causing and high and low expressed genes can support finding the root causes of uncertainties in patient care. However, independent and timely high-throughput next-generation sequencing data analysis is still a challenge for non-computational biologists and geneticists.

Results: In this manuscript, we present a findable, accessible, interactive, and reusable (FAIR) bioinformatics platform, i.e., GViZ (visualizing genes with disease-causing variants). GViZ is a user-friendly, cross-platform, and database application for RNA-seq-driven variable and complex gene-disease data annotation and expression analysis with a dynamic heat map visualization. GViZ has the potential to find patterns across millions of features and extract actionable information, which can support the early detection of complex disorders and the development of new therapies for personalized patient care. The execution of GViZ is based on a set of simple instructions that users without a computational background can follow to design and perform customized data analysis. It can assimilate patients' transcriptomics data with the public, proprietary, and our in-house developed gene-disease databases to query, easily explore, and access information on gene annotation and classified disease phenotypes with greater visibility and customization. To test its performance and understand the clinical and scientific impact of GViZ, we present GViZ analysis for different chronic diseases and conditions, including Alzheimer's disease, arthritis, asthma, diabetes mellitus, heart failure, hypertension, obesity, osteoporosis, and multiple cancer disorders. The results are visualized using GViZ and can be exported as image (PNG/TIFF) and text (CSV) files that include gene names, Ensembl (ENSG) IDs, quantified abundances, expressed transcript lengths, and annotated oncology and non-oncology diseases.

* Correspondence: zahmed@iuh.rutgers.edu

¹Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers University, 112 Paterson Street, New Brunswick, NJ, USA

²Department of Medicine, Robert Wood Johnson Medical School, Rutgers Biomedical and Health Sciences, 125 Paterson Street, New Brunswick, NJ, USA

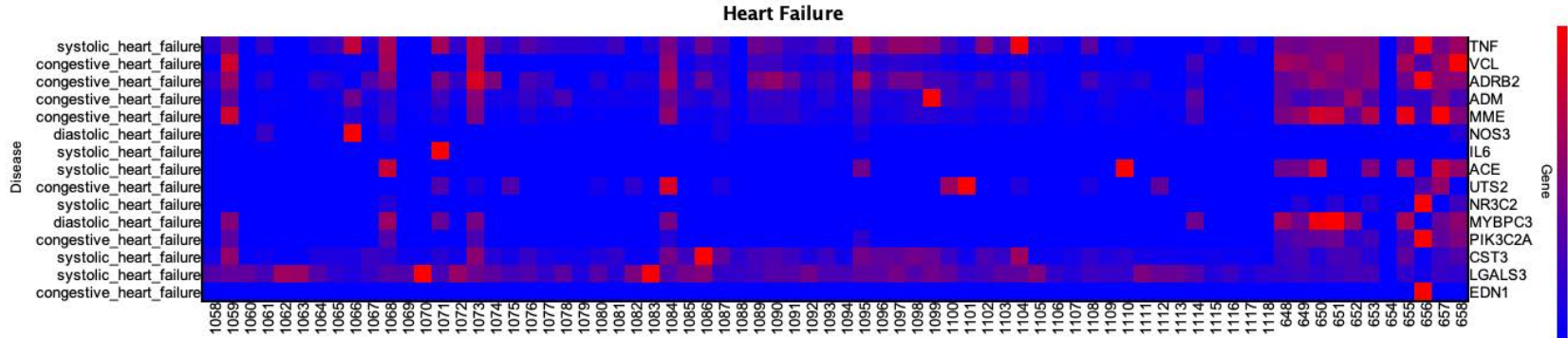
Full list of author information is available at the end of the article



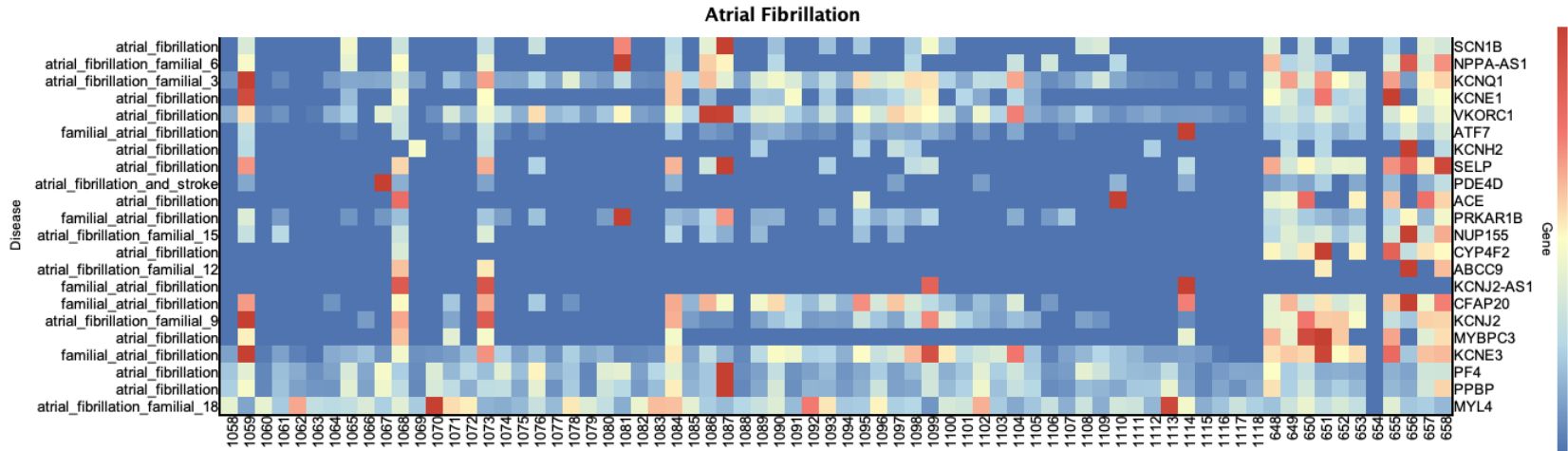
© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

GViZ & Gene-disease annotation

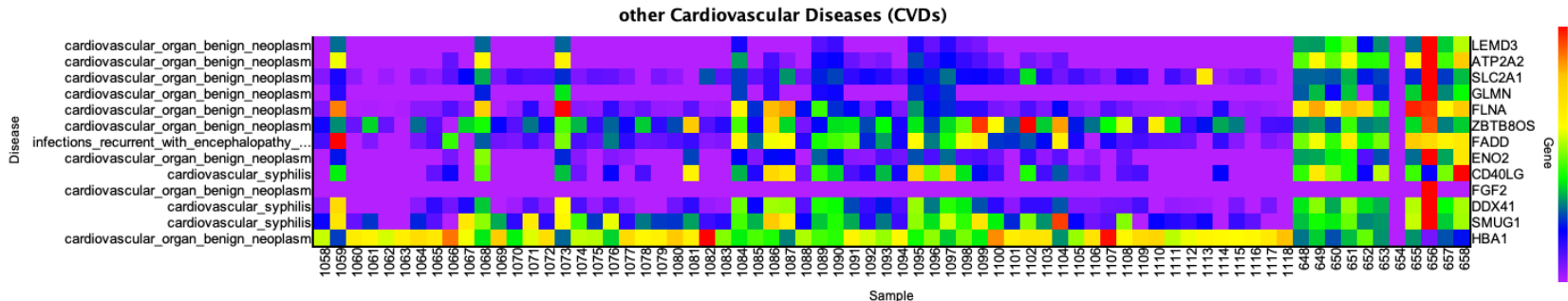
A.



B.



C.

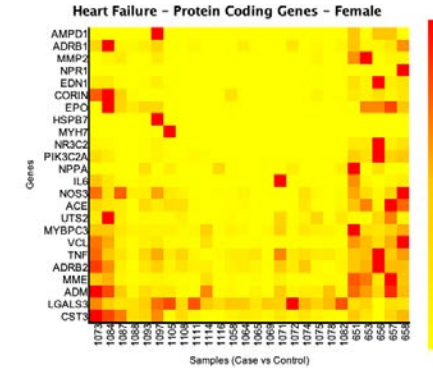
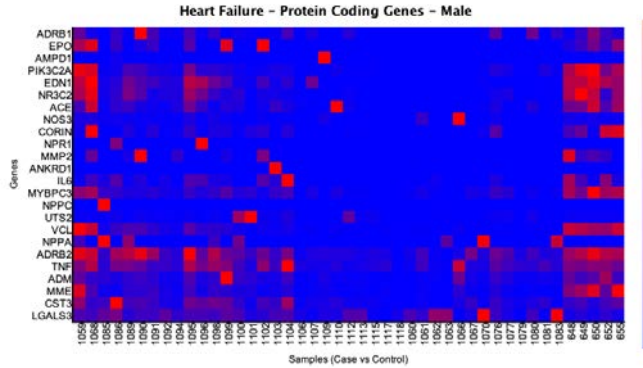


Gender-based gene expression analysis

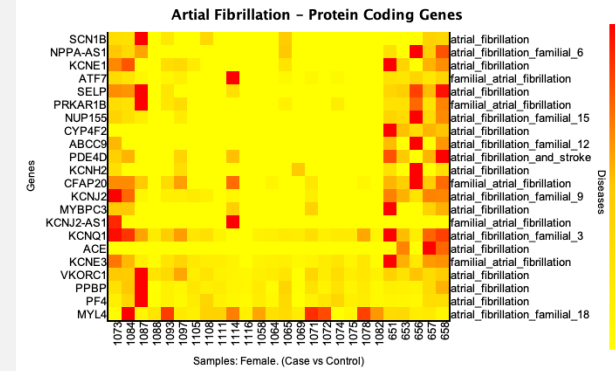
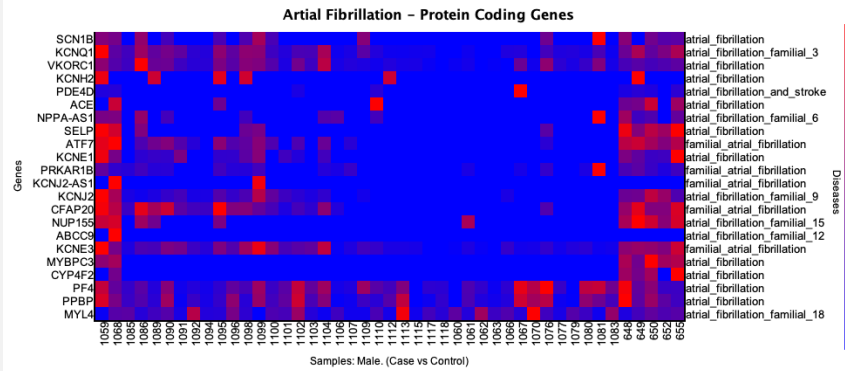
MALE

FEMALE

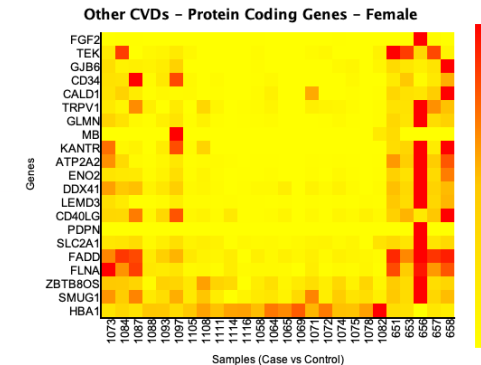
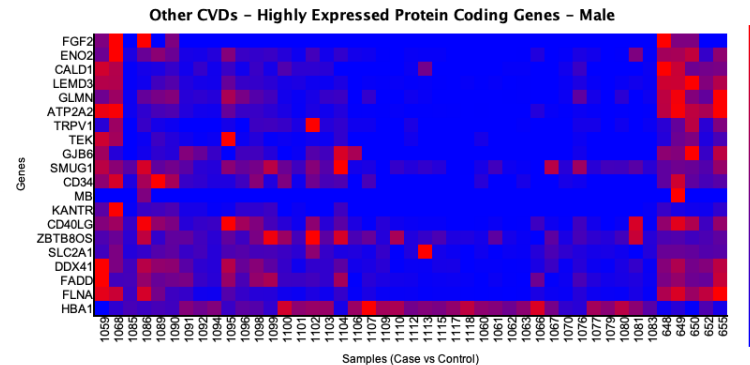
A. HF



B. AF

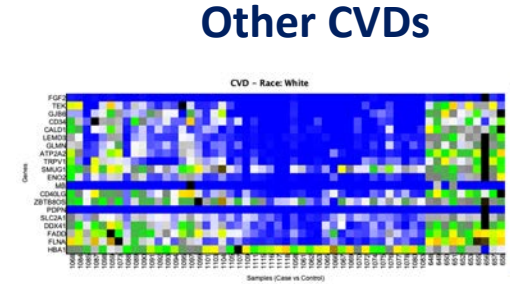
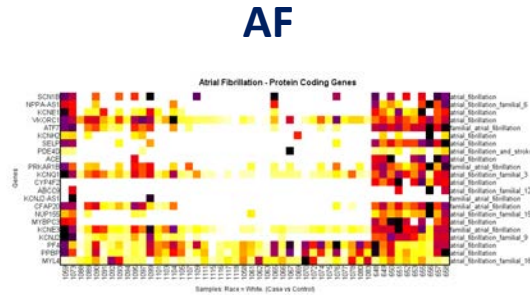
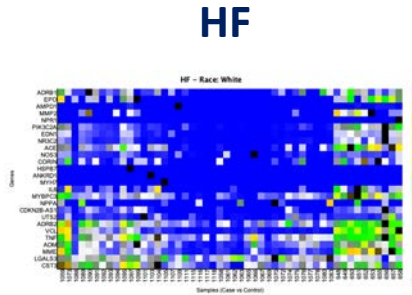


C. other CVDs

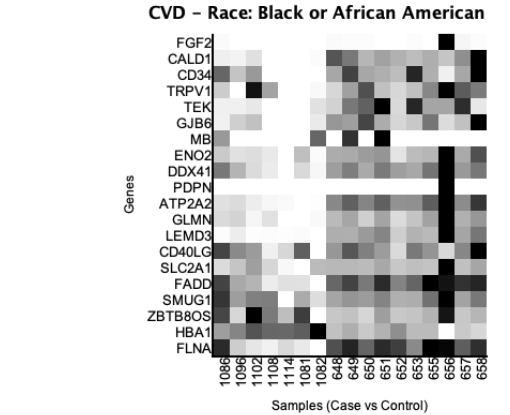
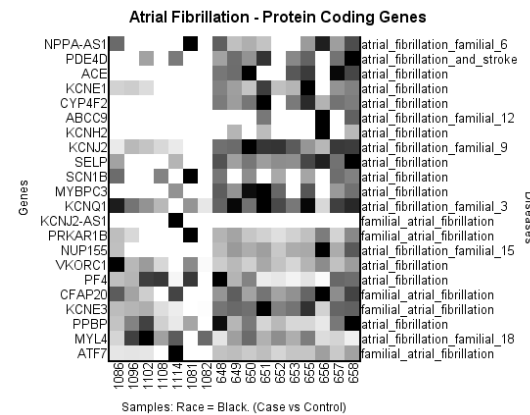
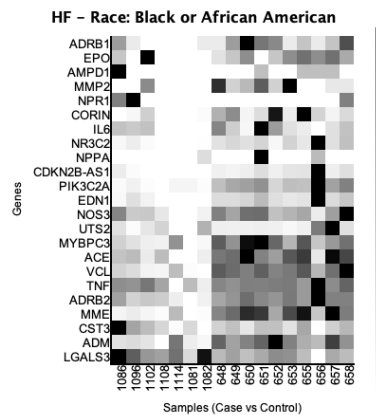


Race-based gene expression analysis

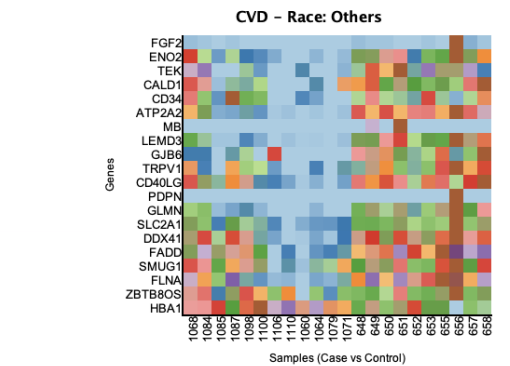
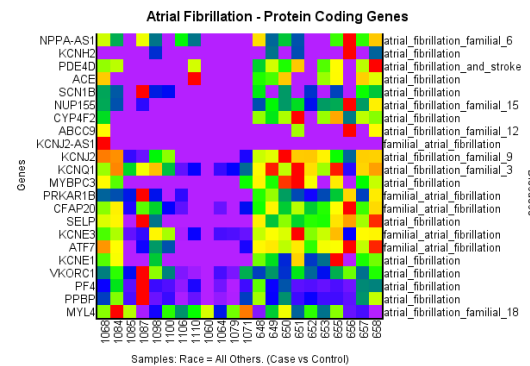
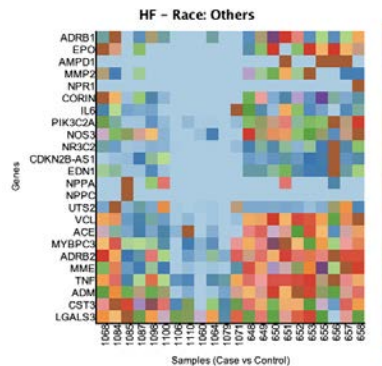
A. White



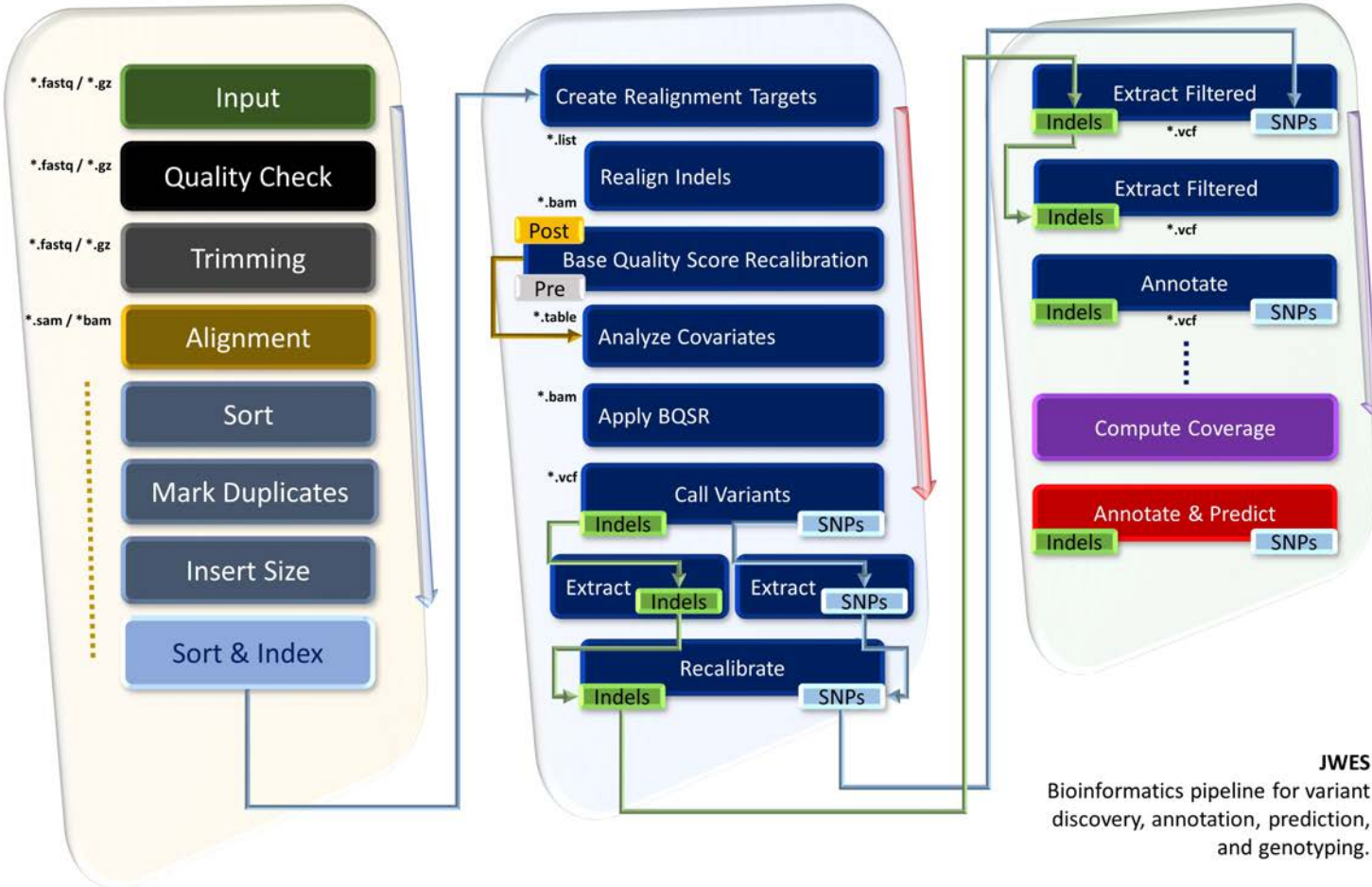
B. Black



C. All others



JWES: Variant Analysis



Download URL: <https://github.com/drzeeshanahmed/JWES-DB>

Publications at CVD Gene Expression, Annotation, and Variant Analysis

Ahmed et al. *Human Genomics* (2021) 15:67
<https://doi.org/10.1186/s13024-021-00367-8>

Human Genomics

PRIMARY RESEARCH

Open Access

RNA-seq driven expression and enrichment analysis to investigate CVD genes with associated phenotypes among high-risk heart failure patients

Zeeshan Ahmed^{1,2,4,5*}, Saman Zeeshan³ and Bruce T. Liang⁵

Abstract

Background: Heart failure (HF) is one of the most common complications of cardiovascular diseases (CVDs) and among the leading causes of death in the US. Many other CVDs can lead to increased mortality as well. Investigating the genetic epidemiology and susceptibility to CVDs is a central focus of cardiology and biomedical life sciences. Several studies have explored expression of key CVD genes specially in HF, yet new targets and biomarkers for early diagnosis are still missing to support personalized treatment. Lack of gender-specific cardiac biomarker thresholds in men and women may be the reason for CVD underdiagnosis in women, and potentially increased morbidity and mortality as a result, or conversely, an overdiagnosis in men. In this context, it is important to analyze the expression and enrichment of genes with associated phenotypes and disease-causing variants among high-risk CVD populations.

Methods: We performed RNA sequencing focusing on key CVD genes with a great number of genetic associations to HF. Peripheral blood samples were collected from a broad age range of adult male and female CVD patients. These patients were clinically diagnosed with CVDs and CMS/HCC, HF, as well as including cardiomyopathy, hypertension, obesity, diabetes, asthma, high cholesterol, anemia, chronic kidney, joint pain, dizziness and giddiness, osteopenia of multiple sites, chest pain, osteoarthritis, and other diseases.

Results: We report RNA-seq driven case-control study to analyze patterns of expression in genes and differentiating the pathways, which differ between healthy and diseased patients. Our in-depth gene expression and enrichment analysis of RNA-seq data from patients with mostly HF and other CVDs on differentially expressed genes and CVD annotated genes revealed 4,885 differentially expressed genes (DEGs) and regulation of 41 genes known for HF and 23 genes related to other CVDs, with 15 DEGs as significantly expressed including four genes already known (FLNA, CST3, LGA53, and HBA1) for HF and CVDs with the enrichment of many pathways. Furthermore, gender and ethnic group specific analysis showed shared and unique genes between the genders, and among different races. Broadening the scope of the results in clinical settings, we have linked the CVD genes with ICD codes.

Conclusions: Many pathways were found to be enriched, and gender-specific analysis showed shared and unique genes between the genders. Additional testing of these genes may lead to the development of new clinical tools to improve diagnosis and prognosis of CVD patients.

*Correspondence: zahmed@rti.rutgers.edu

¹Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers University, 112 Paterson Street, New Brunswick, NJ 08901, USA
Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Received: 11 May 2021 / Revised: 23 June 2021 / Accepted: 27 June 2021
DOI: 10.1186/s13024-021-00367-8

LETTER TO THE EDITOR

RNA-seq-driven expression analysis to investigate cardiovascular disease genes with associated phenotypes among atrial fibrillation patients

To the Editor

Atrial fibrillation (AF) is defined as the high-frequency excitation of the atrium, resulting in both dyssynchronous atrial contraction and the irregularity of ventricular excitation.¹ According to its condition, AF disease is divided into two sub-types paroxysmal and persistent. In contrast to persistent AF, paroxysmal AF is diagnosed in the first phase of the disease, which later progresses to persistent AF.¹ Furthermore, AF includes risk factors such as obesity, diabetes, smoking and a sedentary lifestyle and is prevalent in the older males of European ancestry. Previous studies have shown that both heart failure (HF) and cardiovascular diseases (CVD) contribute to an increased risk of AF.² In this study, we investigated genes responsible for AF with sub-disease groups through transcriptomic analysis (Additional file 1: High-resolution figures). It was conducted as a continuation of our thorough CVD research focusing on HF³ performed on 61 CVD patients (Sample IDs: 1058-1108) and 10 patients without CVD (Control IDs: 648-658) (Additional file 2: population details). When grouped by gender and race, there were 40 males and 21 females, 42 Whites, 7 Blacks (Blacks or African Americans), 1 Asian, 1 Decline to Answer, 2 others, and 8 NA (Table 1 and Figure 1A). Peripheral blood samples were used for RNA extraction, and sequencing was performed using Illumina NovaSeq 6000-S4 to assess the RNA quality.⁴ An efficient data management system (PROMIS-LCR) with data extraction, transfer and loader system (ETL), created by the authors,⁵ was used for patient recruitment and consent tracking as well as dealing with the multi-omics data, respectively.⁶ We also created a publicly available gene-disease database, PAS-Gen, which includes over 59,000 protein-coding and non-coding genes, and over 90,000 classified gene-disease associations, to ease the gene-disease visualization for researchers, medical practitioners and pharmacists.

First, the transcriptomic data analysis involved the development of an RNA-seq processing pipeline that contained four operating parts: (I) data pre-processing, (II) data quality checking, (III) data storage and management and (IV) data visualization (Additional file 1: High-resolution figures).⁷ The analysis of transcripts per million (TPM) was performed to normalize the RNA-seq data by using the visualizing genes with disease-causing variants environment with the findable, accessible, intelligent and reproducible approach (Additional file 2: AF analysis - gene expression data). It reveals all genes annotated with their associated clinical AF phenotype using gene-disease association.^{8,9} This expression analysis was expanded to visualize the classification of protein- and non-coding genes in detail as gender- and race-based. First, we looked across the AF-annotated genes to identify protein- and non-coding genes together and found 71 genes related to AF and relative diseases (Additional file 2: Complete Gene List). Next, we observed expression in protein-coding genes and found 22 genes associated with direct and relative AF diseases, which are denominated as AF phenotypes (SCN1B, NPPA-AS1, KCNQ1, KCNE1, VKORC1, AIFP2, KCNH2, SLC6, PPI4B, ACE, PRKAR1B, NUPR1, CYP4F2, ARCC9, KCNH2-AS1, CPAPB, KCNJ2, MYBP3, KCNE3, PPI4, PPP1R, MYL4) (Figure 1B and Table 2). After the initial analysis, differential gene expression analysis was implemented to further investigate AF genes. Of the protein-coding genes, seven AF-associated genes (MYL4, PPP1R, PPI4, KCNE3, VKORC1, KCNQ1 and CYP4F2) showed differentially regulated expression (Figure 1C). A previous study has reported some of these genes (GLIS4, KCNAs, KCNE2, KCNE1, KCNQ1, KCNH2, NPPA and SCN5A) as novel genes for familial AF in the absence of mutations, whereas mutations in MYL4 have been strongly associated with AF disease in humans.⁶

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Author(s). *Clinical and Translational Medicine* published by John Wiley & Sons Australia, Ltd on behalf of Shanghai Institute of Clinical Bioinformatics.

Clin. Transl. Med. 2021;11:e76.
<https://doi.org/10.1002/ctm2.1076>

www.clinicaltrialsjournal.com/journal/ctm2 | 1 of 8

Received: 5 April 2021 / Revised: 11 May 2021 / Accepted: 11 May 2021
DOI: 10.1002/ctm2.1076

RESEARCH ARTICLE

Investigating genes associated with cardiovascular disease among heart failure patients for translational research and precision medicine

Zeeshan Ahmed^{1,2,4}, Saman Zeeshan³, Nicholas Persaud¹, William Degroot¹, Habiba Abdelhalim¹, Bruce T. Liang^{5*}

¹Institute for Health, Health Care Policy and Aging Research, Rutgers University, New Brunswick, New Jersey, USA

²Department of Medicine, Robert Wood Johnson Medical School, Rutgers Biomedical and Health Sciences, New Brunswick, New Jersey, USA

³Rutgers Cancer Institute of New Jersey, Rutgers University, New Brunswick, New Jersey, USA

⁴Department of Genetics and Genome Sciences, UConn Health, Farmington, Connecticut, USA

⁵Pat and Jim Calsson Cardiology Center, UConn Health, Farmington, Connecticut, USA

⁶UConn School of Medicine, University of Connecticut, Farmington, Connecticut, USA

Correspondence: Zeeshan Ahmed, Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers University, 112 Paterson Street, New Brunswick, 08901, NJ, USA. Email: zahmed@rti.rutgers.edu

Funding information: School of Medicine, UConn Health, CT

Abstract

Background: Cardiovascular disease (CVD) is a leading cause of premature mortality in the United States and the world. CVD comprises several complex and mostly heritable conditions, which range from myocardial infarction to congenital heart disease. The risk factors contributing to the development of CVD and response to therapy in an individual patient are highly variable. Here, we report our findings from an integrative analysis of gene expression, disease-causing gene variants and associated phenotypes among CVD populations, with a focus on high-risk heart failure (HF) patients.

Methods: We built a cohort using electronic health records of consented patients with available samples and then performed high-throughput whole genome and RNA sequencing of key genes responsible for HF and other CVD pathologies. Our in-depth gene expression analysis revealed differentially expressed genes associated with HF and other CVDs. We performed a variant analysis of whole genome sequence data of CVD patients and identified genes with altered gene expression with functional and non-functional mutations in these genes.

Results: Our results highlight the importance of investigating the mechanisms of CVD progression through multi-omics datasets. Next, we performed splice mutation and variant distribution analysis of genes associated with HF and other CVD. We implemented Jensen-Shannon divergence (JSD)-based method and identified HBA1, FADD, ADRB2, NPPB, ADRB1, ADR and NPPC genes with the greatest variance based on their JSD scores. Our study provided evidence that applying integrative data analysis approach involving genomics and transcriptomics data will not only help understand the pathophysiology of CVD diseases but also reduce heterogeneity in disease subtypes.

KEYWORDS

cardiovascular disease, expression, gene, genome, heart failure, RNA-seq, variant

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Author(s). *Clinical and Translational Medicine* published by John Wiley & Sons Australia, Ltd on behalf of Shanghai Institute of Clinical Bioinformatics.

Clin. Transl. Med. 2021;11:e76.
<https://doi.org/10.1002/ctm2.1076>

www.clinicaltrialsjournal.com/journal/ctm2 | 1 of 8



OAK RIDGE
INSTITUTE
FOR SCIENCE
AND EDUCATION

Implementing Artificial Intelligence (AI) & Machine Learning (ML)

1. **Generate AI/ML ready data** with the integration of existing bioinformatics tools and development of new pipelines for the multi-omics and clinical data management, integration, annotation, and sharing.
2. **Predict CVD with high accuracy** with knowledge-driven approach based on known genetic evidence establishing association by implementing best fitting AI/ML algorithms for deep phenotyping and predictive analytics.

Software Impacts 16 (2023) 100493

Contents lists available at ScienceDirect

Software Impacts

Journal homepage: www.journals.elsevier.com/software-impacts

Original software publication

Hyggia: AI/ML pipeline integrating healthcare and genomics data to investigate genes associated with targeted disorders and predict disease

William DeGroot^{a,b}, Vignesh Venkat^{a,b}, Widnie Pierre-Louis^a, Habiba Abdelhalim^a, Zeeshan Ahmed^{a,b,c}

^a Division for Health, Health Care Policy and Aging Research, Rutgers University, 112 Patterson Street, New Brunswick, NJ, USA
^b Department of Medicine, Robert Wood Johnson Medical School, Rutgers Biomedical and Health Sciences, 125 Patterson Street, New Brunswick, NJ, USA
^c Department of Statistics, Rutgers University, 112 Patterson Street, New Brunswick, NJ, USA

ARTICLE INFO

ABSTRACT

Keywords:
Hyggia
Artificial Intelligence
Machine Learning
Genomics
Bioinformatics
Predictive analysis

Due to the advancements in sequencing technologies, genomics data is developing at an unmatched pace and levels to foster translational research. Over ten million genomics datasets have been produced and publicly shared in the year 2022. Genome-wide association studies (GWAS) have remarkably assisted in understanding the genetic basis of human disease by uncovering millions of loci associated with various complex phenotypes. However, GWAS are unable to predict disease and detect all the heritability explained by single nucleotide polymorphisms (SNPs) and can only target specific variants. The righted use of the artificial intelligence (AI) and machine learning (ML) techniques can accelerate our ability to leverage and extend the information contained within the original data, and model patient-specific genomics data against publicly available association repositories for understanding how coding and non-coding genomic variations are connected to disease mechanisms. The grand challenge here is assimilation of genetics into precision medicine that translates across different ancestries, diverse diseases, and other distinct populations with the implementation of effective AI/ML methods. We present first AI/ML ready pipeline i.e., Hyggia, integrating genomics and clinical data to investigate genes associated with the targeted disorders and predict disease with high accuracy. Hyggia can utilize broad dataset sizes with heterogeneous levels of granularity and offer a supervised approach to analyze integrated gene expression and multivariate clinical data. It includes the Random Forest based model for regression analysis and predict without hyperparameter tuning. We trained and tested our model across variable disorders and using diverse datasets. Hyggia is an open-source and simple to use pipeline, which does not strong require computational background to execute.

Code metadata

Current Code Version	Hyggia v1.0.2
Permanent Link to Repository	https://github.com/SoftwareImpacts/IMPACT-2023-36
Repository Capabilities	https://codemeta.org/impacts/2023-36
Legal license	GNU General Public License (GPL)
Code Versioning System	Git
Software Code Language	Python 3.10.9
Compilation Requirements, Dependencies	python, pandas, numpy, matplotlib, seaborn
Support email for questions	zahmed@du.rutgers.edu

1. Introduction

Precision and genomics medicine is driven by the paradigm shift of empowering clinicians to predict the most appropriate course of action

The code (and data) in this article has been certified as Reproducible by Code Ocean: (<https://codeocean.com>). More information on the Reproducibility Badge Initiative is available at <https://www.elsevier.com/locate/ympev>.

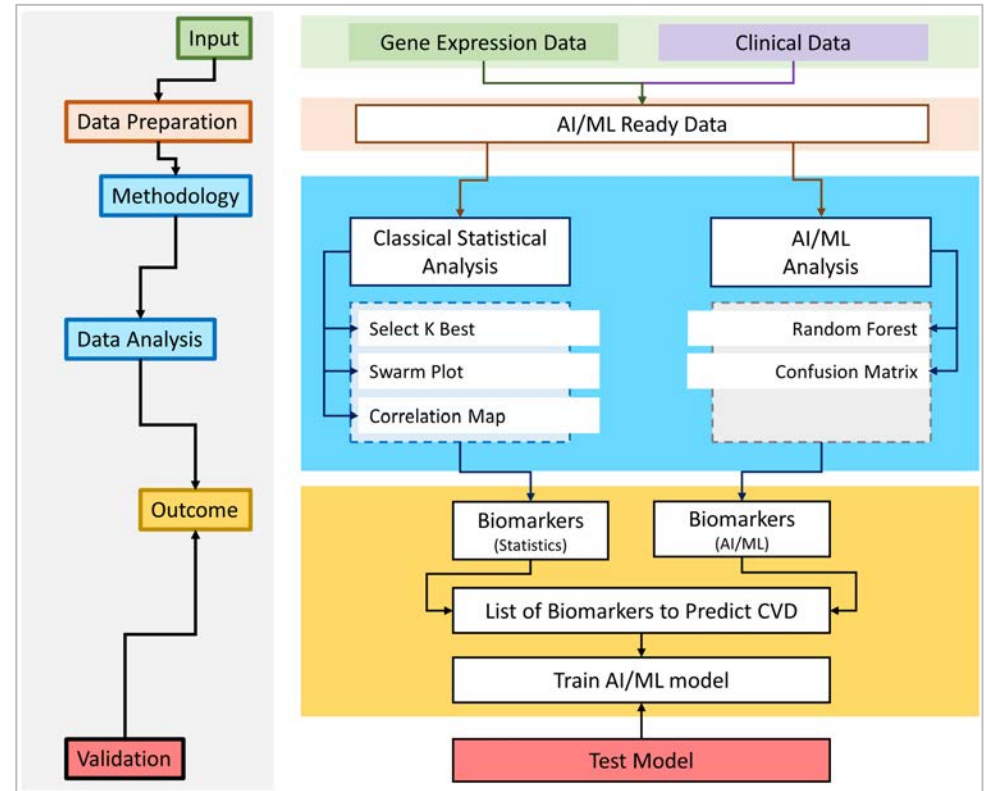
* Correspondence to: Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers University, 112 Patterson Street, New Brunswick, 08901, NJ, USA.
E-mail address: zahmed@du.rutgers.edu (Z. Ahmed).
† Equally contributing first authors.

<https://doi.org/10.1016/j.simp.2023.100493>
Received 20 January 2022; Received in revised form 4 March 2022; Accepted 6 March 2022

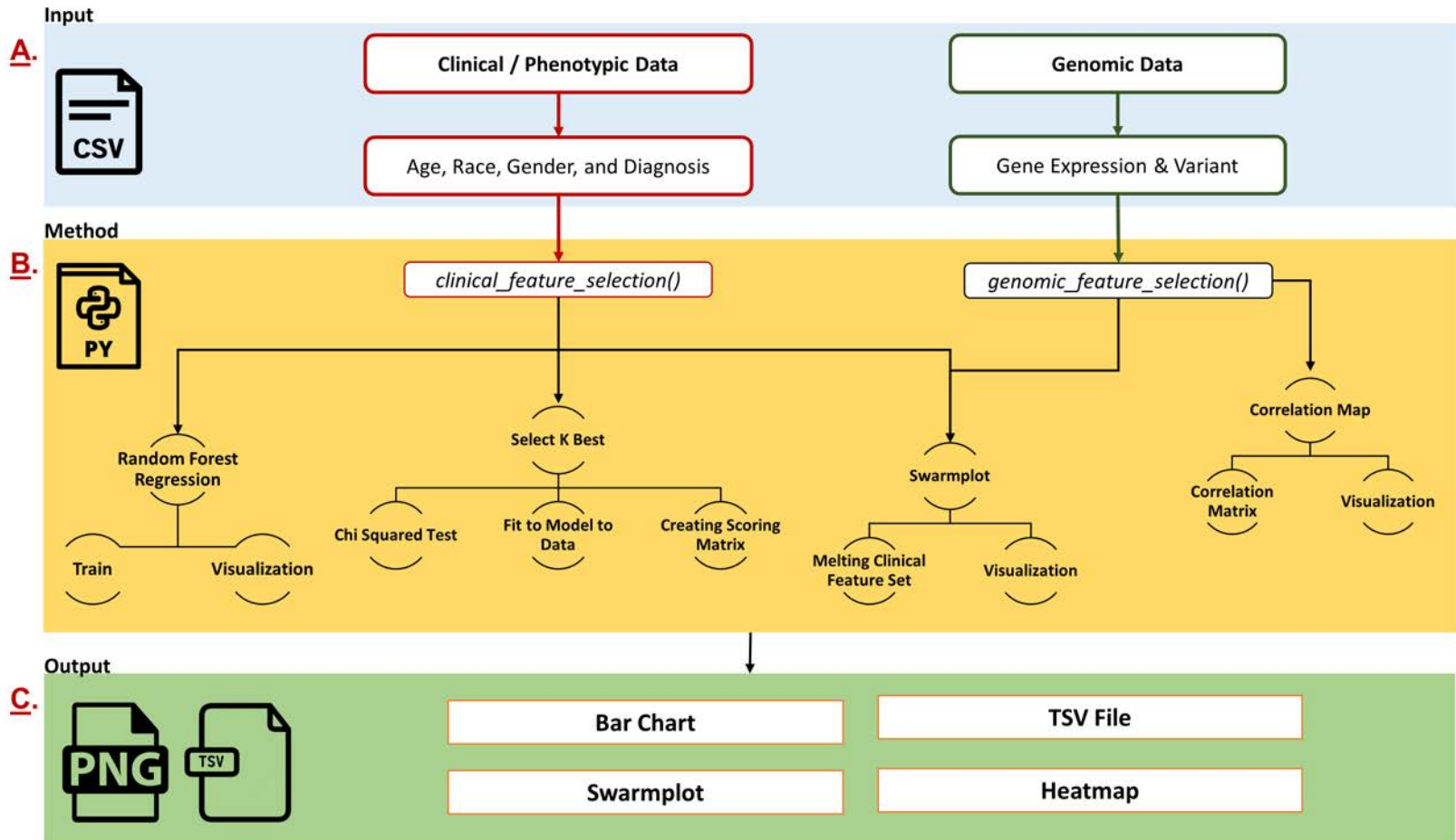
2665-9638/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Implementing Artificial Intelligence (AI) & Machine Learning (ML)

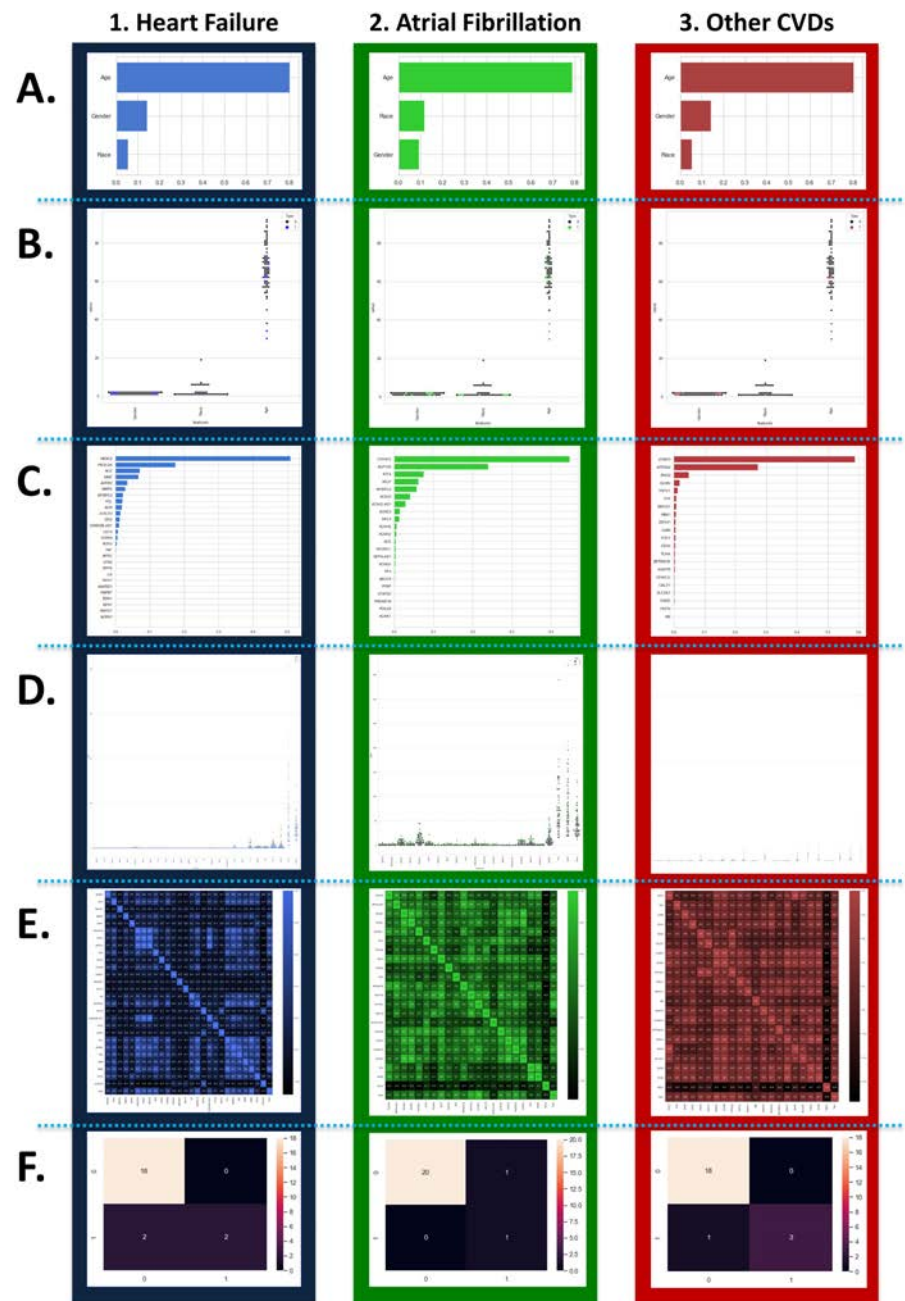
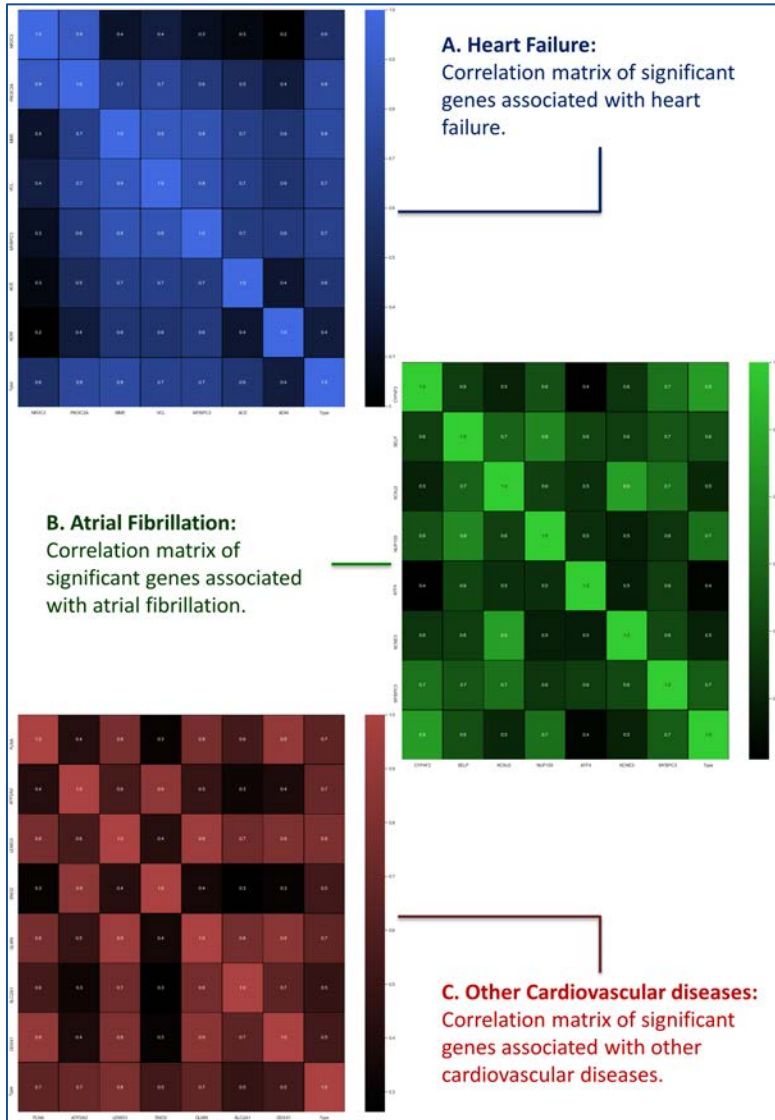
1. **Generate AI/ML ready data** with the integration of existing bioinformatics tools and development of new pipelines for the multi-omics and clinical data management, integration, annotation, and sharing.
2. **Predict CVD with high accuracy** with knowledge-driven approach based on known genetic evidence establishing association by implementing best fitting AI/ML algorithms for deep phenotyping and predictive analytics.



Hygieia Design



AI/ML Analysis



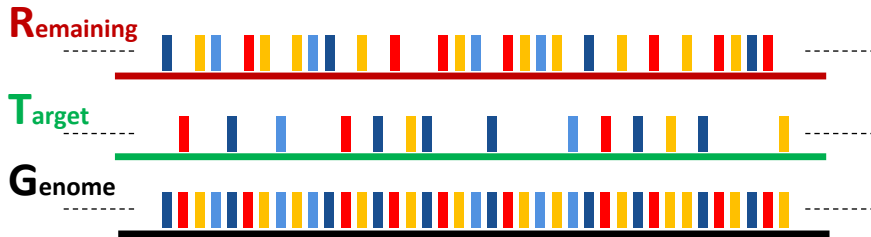
(A) Population distribution based on clinical features; (B) Correlation matrix; (C) Gene ranking; (D) Feature Swarm Plot; (E) Correlation matrix of genes; (F) Confusion matrix of genes.

AI/ML Analysis

- We used our **open-source AI/ML ready pipeline i.e., Hygieia**, which is based on the Random Forest (RF) for regression analysis and predicting disease without requiring hyperparameter tuning.
- We **trained our model on different cross-sections** of the three different matrices based on HF, AF, and other CVDs.
- We uncovered an interesting correlation between age, gender, race, and diagnosis. During our analysis, it was observed that **age and gender appeared to have a high correlation in HF and other CVDs while age, and race were highly correlated in AF.**
- We observed the most significant genes associated with HF, AF, and other CVDs based on the RF feature importance global variable. **A score was assigned to each gene, which represents the feature importance for the model in stratifying CVD patients.**
- Visible data clusters were observed for the genes highly correlated, downregulated and with altered expression in CVD patients compared to healthy individuals. Our model was able to correctly classify individuals as CVD patients and predict CVD with **95% accuracy.**
- We observed and reported overlapping in **significant results produced in gene expression, variant, phenotypic, and predictive analyses**, which include genes associated with HF, AF, and other CVDs.

AI/ML Analysis

Identify new predictive biomarkers using data-driven approach with a novel nexus of AI/ML algorithms to reveal biological patterns explaining the causal relationship based on genomic, transcriptomic, and phenotypic data of patients with CVD.



Investigating genes associated with heart failure, atrial fibrillation, and other cardiovascular diseases, and predicting disease using machine learning techniques for translational research and precision medicine

Vignesh Venkat^{a,1}, Habiba Abdelhalim^{a,1}, William DeGroat^a, Saman Zeeshan^b, Zeeshan Ahmed^{a,c,*}

^a Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers University, 112 Paterson St, New Brunswick, NJ, USA

^b Rutgers Cancer Institute of New Jersey, Rutgers University, 195 Little Albany St, New Brunswick, NJ, USA

^c Department of Medicine, Robert Wood Johnson Medical School, Rutgers Biomedical and Health Sciences, 125 Paterson St, New Brunswick, NJ, USA

ARTICLE INFO

Keywords:
Artificial intelligence
Atrial fibrillation
Cardiovascular diseases
Gene expression
Heart failure
Machine learning
Predictive analysis

ABSTRACT

Cardiovascular disease (CVD) is the leading cause of mortality and loss of disability adjusted life years (DALYs) globally. CVDs like Heart Failure (HF) and Atrial Fibrillation (AF) are associated with physical effects on the heart muscles. As a result of the complex nature, progression, inherent genetic makeup, and heterogeneity of CVDs, personalized treatments are believed to be critical. Rightful application of artificial intelligence (AI) and machine learning (ML) approaches can lead to new insights into CVDs for providing better personalized treatments with predictive analysis and deep phenotyping. In this study we focused on implementing AI/ML techniques on RNA-seq driven gene-expression data to investigate genes associated with HF, AF, and other CVDs, and predict disease with high accuracy. The study involved generating RNA-seq data derived from the serum of consented CVD patients. Next, we processed the sequenced data using our RNA-seq pipeline and applied GAVIZ for gene-disease data annotation and expression analysis. To achieve our research objectives, we developed a new Findable, Accessible, Intelligent, and Reproducible (FAIR) approach that includes a five-level biostatistical evaluation, primarily based on the Random Forest (RF) algorithm. During our AI/ML analysis, we have fitted, trained, and implemented our model to classify and distinguish high-risk CVD patients based on their age, gender, and race. With the successful execution of our model, we predicted the association of highly significant HF, AF, and other CVDs genes with demographic variables.

1. Introduction

Cardiovascular disease (CVD) is the leading causes of mortality and loss of disability adjusted life years (DALYs) globally [1–3]. The World Health Organization (WHO) states that over 75% of premature CVDs are preventable with a better understanding of risk factors and gene-disease associations [2]. CVDs like Heart Failure (HF) and Atrial Fibrillation

(AF) are associated with physical impacts on the heart muscles [1, 3]. HF occurs due to weak heart muscles that impact the efficiency of pumping blood to the body's cells [1]. While AF occurs due to the high-frequency excitation of the atrium, resulting in both dysynchronous atrial contraction and the irregularity of ventricular excitation [3, 4]. Genomics studies done using genome-wide association studies (GWAS) have aided in disease prediction [5, 6], discovery of genetic loci and alleles

Abbreviations: Artificial intelligence, (AI); Atrial fibrillation, (AF); Cardiovascular diseases, (CVDs); Computerized tomography, (CT); Differentially expressed genes, (DEGs); Electronic health records, (EHR); Extract, transfer, and load, (ETL); Fragments per kilobase million, (FPKM); Genome-wide association studies, (GWAS); Heart failure, (HF); Institutional review board, (IRB); Machine learning, (ML); Mean expressed transcript lengths, (METL); Next generation sequencing, (NGS); Normalized enrichment score, (NES); Random forest, (RF); Reads per kilobase of transcript per million mapped reads, (RPKM); RNA-sequencing, (RNA-seq); Sci-kit learn, (Sklearn); Support vector machine, (SVM); Transcripts per million, (TPM); Visualizing genes with disease-causing variants, (GVVIZ); World Health Organization, (WHO); Whole genome sequencing, (WGS); Whole exome sequencing, (WES).

* Corresponding author at: Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers University, 112 Paterson Street, New Brunswick 08901, NJ, USA.

E-mail address: zahmed@rh.rutgers.edu (Z. Ahmed).

¹ Equally contributing first authors.

<https://doi.org/10.1016/j.ygeno.2023.110584>

Received 19 November 2022; Received in revised form 6 February 2023; Accepted 11 February 2023

Available online 20 February 2023

0888-7543/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Traditional Bioinformatics and AI/ML

- **Gene Expression Analysis (GEA) and Genome-wide association studies (GWAS)** have been useful in providing deeper understanding of the genetic basis of human diseases by uncovering millions of loci associated with various complex phenotypes.
- **These are unable to predict disease and detect all the heritability** explained by SNPs, as they can only target specific variants of disease.
- These limitations are not exclusive to GEA and GWAS, as **no method to date can identify all the genetic components of complex diseases.**
- In addition, a persistent challenge in multi-genomic data analysis lies in the handling, **integration, and standardization of large volumes of sequencing data.**
- **The proficient and synergistic implementation of AI/ML techniques holds the promise of fostering an augmented comprehension of CVD at the systemic level, unveiling the intricacies of genomic regulatory networks.**

Global impact of precision medicine

The appropriate utilization of artificial intelligence (AI) and machine learning (ML) methodologies can yield novel understandings of complex traits, enabling improved personalized treatments through predictive analysis and deep phenotyping.

Human Genomics

OPINION ARTICLE

Practicing precision medicine with intelligently integrative clinical and multi-omics data analysis

Zeehan Ahmed^{1,2,3,4*}

Abstract
Precision medicine aims to empower clinicians to provide the most appropriate course of action for each patient with complex chronic (e.g., cancer, diabetes, cardiovascular), and COVID-19 (with a progressive etiological of the clinical, molecular, and genetic factors) to play a diagnosis, more effective and personalized medical treatment are anticipated for many diseases. Understanding patient's metabolism and genetic make-up inconspicuous with these data will significantly lead to determining pathogenesis, diagnosis, prognosis, and predictive biomarkers and path alterations providing optimal and personalized care for chronic and targeted chronic and acute diseases. In clinical strategy, we seek to timely model clinical and multi-omics data to find suitable patterns across millions of features to identify underlying hidden patterns, modifiable risk factors, and actionable interventions that support early detection and prevention of complex diseases, and development of new therapies for better patient care. It is important to include quantitative phenotypic measurements, evaluate variants in unique genes and interpret using ACMG guidelines, and frequency of pathogenic and likely pathogenic variants without disease indications, and observe additional disease carrier with a phenotypic manifestation in individuals. First, ensuring security to electronic notes, we need to build and train machine learning algorithms models for recognizing precise multi-omics heterogeneous data to identify high-risk care variants and make medical advice predictions. The goal today is to facilitate implementation of precision medicine to improve the traditional symptom-driven practice of medicine, and allow wider interventions using predictive diagnosis and tailored better personalized medicines, the strongly evidenced, automated implementation of cutting-edge technologies, utilizing machine learning (ML) and artificial intelligence (AI) approaches for the multi-omics data integration, multi-scale economic development of knowledgegraph of disease prediction for decision support, and best strategies for dealing with adverse clinical trials.

Keywords Precision medicine, Omics, Genomics, Metabolomics, Integrative analysis, Artificial intelligence, Machine learning

Background
Since the beginning of scientific discoveries, it has been central to understand the cause of disease and determine (1). Path is one of the key triggers for patients to seek diagnosis and treatment. However, when dealing with some of the life-threatening diseases, patients may not feel pain. To identify and help patients with known disease and symptoms, and those heading toward late stages of novel diseases (e.g., COVID-19, Alzheimer's, diabetes, heart disease), acute (e.g., flu, stroke, heart attack), and complex (e.g., cancer) diseases, it is essential to provide timely personalized treatment [2–3]. One ending understanding of the complex nature has led us to

Conclusion
Precision medicine is driven by the paradigm shift of empowering clinicians to predict the most appropriate course of action for patients with complex diseases and to improve routine medical and public health practice. Understanding patients' multi-omics make-up in conjunction with the clinical data will lead to determining predispositions, diagnostic, prognostic, and predictive biomarkers and to optimal path providing personalized care for diverse and targeted chronic, acute, and infectious diseases. Precision medicine promotes integrating collective and individualized clinical data with patient-specific multi-omics data to develop therapeutic strategies and knowledge bases for predictive and personalized medicine in diverse populations. Artificial intelligence approaches and machine learning algorithms will add additional capabilities to precision medicine that will leverage and extend the information contained within the original data and facilitate making patient-specific multi-omics data against publicly available annotation data for better understanding disease mechanisms. This chapter

Footnote
*Correspondence: z.ahmed@rutgers.edu

Page 1 of 13
Progress in Molecular Biology and Translational Science, Volume 190, Issue 1, 2022, Pages 101–120
© Springer 2022. Springer Inc. All rights reserved.

Progress in Molecular Biology and Translational Science

Precision medicine with multi-omics strategies, deep phenotyping, and predictive analysis

Zeehan Ahmed^{1,2,3,4*}

Abstract
Precision medicine is driven by the paradigm shift of empowering clinicians to predict the most appropriate course of action for patients with complex diseases and to improve routine medical and public health practice. Understanding patients' multi-omics make-up in conjunction with the clinical data will lead to determining predispositions, diagnostic, prognostic, and predictive biomarkers and to optimal path providing personalized care for diverse and targeted chronic, acute, and infectious diseases. Precision medicine promotes integrating collective and individualized clinical data with patient-specific multi-omics data to develop therapeutic strategies and knowledge bases for predictive and personalized medicine in diverse populations. Artificial intelligence approaches and machine learning algorithms will add additional capabilities to precision medicine that will leverage and extend the information contained within the original data and facilitate making patient-specific multi-omics data against publicly available annotation data for better understanding disease mechanisms. This chapter

Footnote
*Correspondence: z.ahmed@rutgers.edu

Page 1 of 13
Progress in Molecular Biology and Translational Science, Volume 190, Issue 1, 2022, Pages 101–120
© Springer 2022. Springer Inc. All rights reserved.

ARTICLE IN PRESS

Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine

Zeehan Ahmed^{1,2,3,4*}, Khalid Mohamed⁵, Saman Zeehan⁶ and XinQi Dong^{7,8}

Abstract
Precision medicine is one of the recent and powerful developments in medical care, which has the potential to improve the traditional symptom-driven practice of medicine, allowing earlier interventions using advanced diagnostics and tailoring better and economically personalized treatments. Identifying the best pathway to personalized and population medicine involves the ability to analyze comprehensive patient information together with broader aspects to monitor and distinguish between sick and relatively healthy people, which will lead to a better understanding of biological indicators that can signal shifts in health. While the complexities of disease at the individual level have made it difficult to utilize healthcare information in clinical decision-making, some of the existing constraints have been greatly minimized by technological advancements. To implement effective precision medicine with enhanced ability to positively impact patient outcomes and provide real-time decision support, it is important to harness the power of electronic health records by integrating disparate data sources and discovering patient-specific patterns of disease progression. Useful analytic tools, technologies, databases, and approaches are required to augment networking and interoperability of clinical, laboratory and public health systems, as well as addressing ethical and social issues related to the privacy and protection of healthcare data with effective balance. Developing multifunctional machine learning platforms for clinical data extraction, integration, management and analysis can support clinicians by efficiently streamlining

Footnote
*Correspondence: z.ahmed@rutgers.edu

Page 1 of 13
Database, Volume 19, Issue 1, 2022, Pages 101–120
© Springer 2022. Springer Inc. All rights reserved.

frontiers | Frontiers in Genetics

Artificial Intelligence, Healthcare, Clinical Genomics, and Pharmacogenomics Approaches in Precision Medicine

Habiba Alkhalafah¹, Anusha Borker², Madeline Loo³, Rivy Jain⁴, Arthun Nae⁵, Anusha Prasad⁶, Anshu Patel⁷, Sagnika Venkatesh⁸, Cynthia Venkatesh⁹, Pragna Wadhwa¹⁰, Matthew Olatunji¹¹, Abhinav Puri¹², Vikram Puri¹³, John Kulkarni¹⁴, Marc Akpan¹⁵, Joseph Kulkarni¹⁶, David Mena¹⁷, Mayank Patel¹⁸, Nishi Patel¹⁹, Theres Pappas²⁰, Zeyu Fu²¹, Prabhu Saravali²², Rajesh Venkatesh²³, Sreyasi Bhat²⁴, Sanya Verma²⁵ and Zeehan Ahmed^{26*}

Abstract
Precision medicine has greatly aided in monitoring health outcomes using earlier diagnosis and better prognosis for chronic diseases. It makes use of clinical data associated with the patient as well as their multi-omics/genomic data to reach a conclusion regarding how a physician should proceed with a specific treatment. Compared to the symptom-driven approach in medicine, precision medicine considers the critical fact that all patients do not react to the same treatment or medication in the same way. When considering the intersection of traditionally distinct areas of medicine, that is, artificial intelligence, healthcare, clinical genomics, and pharmacogenomics—what has them together is their impact on the development of precision medicine as a field and how they each contribute to patient-specific, rather than symptom-specific patient outcomes. This study discusses the impact and integration of these different fields in the scope of precision medicine and how they can be used in preventing and predicting acute or chronic diseases. Additionally, this study also discusses the advantages as well as the current challenges associated with artificial intelligence, healthcare, clinical genomics, and pharmacogenomics.

Keywords artificial intelligence, healthcare, clinical genomics, pharmacogenomics, precision medicine

INTRODUCTION
Precision medicine is the utilization of healthcare tools to create specialized treatments that consist of optimal actions for the patient, based on the data available (Khan et al., 2011; Prabhu, 2011; Gammeter et al., 2016; Gauding and Phillips, 2016; Ahmed et al., 2016a, Ahmed, 2020; Eshwari, 2020; Fakhri et al., 2020; Ahmed et al., 2012a). An ethical, genetic, and metabolic data become easier to obtain and interpret in relation to complex chronic diseases such as cancer. Disease treatment will become more effective (El-Gohary et al., 2017; Prabhu, 2017; Gauding and Phillips, 2016; Cozzani and Luchini, 2016; Bhat et al., 2016; Ahmed et al., 2016a, Ahmed, 2018; Fakhri et al., 2018; Ahmed et al., 2017a). In the current state of healthcare, healthcare professionals tend to divide their attention

Footnote
*Correspondence: z.ahmed@rutgers.edu

Page 1 of 13
Frontiers in Genetics, Volume 12, Article 781878, 2021
© Copyright © 2021 Ahmed et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). See the frontmatter for more details.



Lab Members (2020 – 23)



Active

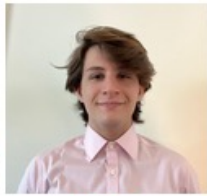


Zeshan Ahmed, PhD
PI/Lead Lab Director

Assistant Professor of Medicine & Core Director, Institute for Health, Health Care Policy and Aging Research, Department of Medicine, Rutgers Robert Wood Johnson Medical School, Rutgers Biomedical and Health Sciences



Habiba Abdelhain, BS
Senior Research Assistant/Scientist
Project Manager



William DeGroot
Research Assistant/Scientist



Dinesh Menche
Scientific Software Engineer



Kush Patel
Research Assistant/Scientist



Anirudh Pappu
Research Assistant/Scientist



Ragharanandh Wable
Research Assistant/Scientist



Sophia Dumas
Research Assistant/Scientist



Atharv Jayprakash
Research Assistant/Scientist



Oluwafemi Omediran
Research Assistant/Scientist



Oluchi Nwanoko
Research Assistant/Scientist



Atharva Hananaradhan Bhanani
Research Student/Scientist

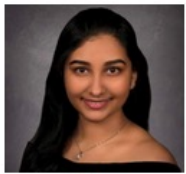
Alumni



Achuth Suresh Nair
Research Assistant/Scientist

Project: Participated in Multiple Bioinformatics, Genomics, and Precision Medicine Projects.

Alumni



Ishant Mishra
Research Assistant/Scientist

Project: Functional and non-functional studies on analysis of COVID-19 genes for precision medicine.

Alumni



Cynthia Venkatesan
Research Student/Scientist

Project: Multi-omics approaches to identify and investigate genes associated with heart failure.

Alumni



Mussoori Loth
PhD Student

Project: Implementing Clinical Decision in Precision Medicine for COVID-19.

Alumni



Rishi Jain
Research Student

Project: Learning role of AI, Healthcare, and Clinical Decision Approaches in Precision Medicine.

Alumni



Kush K. Patel, MD Candidate
Research Student

Project: Building Director White Association Studies for Precision Medicine.

Alumni



Vinayak Vimal
Research Assistant/Scientist

Project: SVM using gene expression data for predicting cardiovascular disease with high accuracy.

Alumni



Sneha Vasudevan, BS/MS Tech.
Research Student

Project: AI & ML approaches using gene expression and varied data for precision medicine.

Alumni



Anshu Baran, BS
Research Student

Project: Investigate COVID-19 genes with associated phenotypes among all of Florida's patients.

Alumni



Saad Ishaq
Research Student

Project: Learning SVM and Decision tree approaches for gene expression and precision medicine.

Alumni



Shreyas Solla
Research Student

Project: Learning SVM and Decision tree approaches for gene expression and precision medicine.

Alumni



Khushi Mohamed, MS
Research Student

Project: AI Clinical Intelligence in Healthcare Data Analysis and Precision Medicine.

Alumni



Shreyas Ashok
Research Student

Project: Functional and non-functional analysis of COVID-19 genes among high-risk Indian patients.

Alumni



Nicholas Perandi, MD Candidate - M2
Research Student

Project: Investigating functional mutations in human genes among high-risk COVID-19 patients.

Alumni



Rishad Perandi, MD Candidate - M2
Research Student

Project: Investigating methylation for identifying potential candidates for better COVID-19 treatment (BCT).

Alumni



Jihong Dang
MS PhD Candidate MD

Project: Investigating genetic and diagnostic clinical state of gene expression with an epidemiologic datasets.

Alumni



Ruzoyi Xiong
MS PhD Candidate MD

Project: Analyzing gene cluster systems and their clinical applications for Precision Medicine.

Alumni



Justin Thomas Phanula, MS
Research Student

Project: MDR of Pacific Islander Microbiome Evolution Data for identifying Genetic Epidemiological Correlations.

Alumni



Waleed Pierre-Louis
Research Assistant/Scientist

Project: Intelligent and integrative genomics data analysis with machine learning algorithms for precision medicine.

Alumni



Rachel Mae Hunter
Research Assistant/Scientist

Project: Role of COVID-19 and integrative risk scoring using big data for precision medicine.

Alumni



Edward Gilbert Newark, PhD
Research Student

Project: Scientific workflow development for RNA and DNA-seq data processing and visualization.

Alumni



Samuel Asareh, BS
Research Student

Project: Build mobile systems development for monitoring and triaging patients to bring virtual care channels.

Alumni



Hussain Patel, BS
PhD Research Associate

Project: Exploring Genomic and Academic Medication Systems To Facilitate Patient Transformation.

Alumni



Deepshika Mishra, PhD
PhD Research Associate

Project: COVID-19 data processing, management, and gene-based discovery, prediction, and genotyping.

Alumni

Collaborations

UCONN HEALTH



Bruce T. Liang, M.D., F.A.C.C.

Dean, UConn School of Medicine, UConn Health
Director, Pat and Jim Calhoun Cardiology Center

Partho P. Sengupta, M.D.

Chief of the Division of Cardiovascular Disease & Hypertension.
Chief of Cardiology, Robert Wood Johnson Medical School.

Daniel Fine, D.M.D

Chair, Department of Oral Biology
Rutgers School of Dental Medicine

Michal Schnaider Beeri, Ph.D.

Director, Herbert and Jacqueline Krieger Klein Alzheimer's Research Center

Danielle Dick, Ph.D.

Director, Rutgers Addiction Research Center (RARC)

Naveena Yanamala, Ph.D.

Director of Data Science and Machine Learning Research, RWJMS

Olga F. Jarrín Montaner, Ph.D.

Director, Community Health and Aging Outcomes Lab

Tammy Chung, Ph.D.

Director, Center for Population Behavioral Health, Rutgers IFH

Suvi Linna-Kuosmanen, PhD.

Massachusetts Institute of Technology (MIT).

Yuichiro Arima, M.D., PhD.

Kumamoto University, Japan.

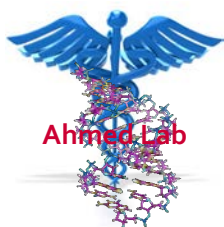
RUTGERS

Institute for Health, Health Care
Policy and Aging Research



Rutgers Robert Wood Johnson Medical School
Rutgers Biomedical and Health Sciences

Grants/Funding





RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY



Zeeshan Ahmed, Ph.D.

Email: < zahmed@ifh.rutgers.edu >

Ahmed Lab: < <https://promis.rutgers.edu/> >

Precision Medicine Project: < <https://sites.rutgers.edu/precision-medicine/> >

