Journal Reading

Presenter: PGY2 張家甄
Supervisor: VS 張家銘醫師

International Journal of *Molecular Sciences*

Review

Innovations in Genomics and Big Data Analytics for Personalized Medicine and Health Care: A Review

Mubashir Hassan [1,2,*], Faryal Mehwish Awan [3], Anam Naz [1], Enrique J. deAndrés-Galiana [4], Oscar Alvarez [5], Ana Cernea [5], Lucas Fernández-Brillet [5], Juan Luis Fernández-Martínez [4] and Andrzej Kloczkowski [2,6,*]

Academic Editors: Ian A. Nicholls and Vladimir N. Uversky

Int. J. Mol. Sci. **2022**, *23*(9), 4645; https://doi.org/10.3390/ijms23094645

Received: 28 February 2022 / Revised: 6 April 2022 / Accepted: 18 April 2022 / Published: 22 April 2

Outline

1. Introduction
2. The Conceptualization of Big Data
3. Computational Approaches toward Personalized Medicine
4. Machine Learning Perspectives on Personalized Medicine
5. Modeling Genetic Data with Translational Purposes
6. Data Mining Tools/Algorithms and Their Applications for Personalized Medicine
7. Heterogeneity, a Huge Challenge in Big Data Analysis
8. Role of Big Data in Accelerating Digital Healthcare

Outline

9. Big Data Applications in Health Care
10. Electronic Health Records
11. Health Big Data as a Key Player for Informed Strategic Planning
12. Advanced Risk and Disease Management through Big Data
13. Developing New Therapies and Big Data
14. Impediments of Big Data in Health Care
15. Conclusions and Future Prospects

01

Introduction

>2000 yrs ago
Hippocrates

1956
Favism
Deficiency in the metabolic enzyme G6PD → selective toxicity

1985
Renato Dulbecco
sequence the human genome for advance cancer research

1988
Genentech Inc.
sequenced the entire human growth hormone locus

1990
Human Genome Project (HGP)

Introduction

1990

Human Genome Project (HGP)

1994

rHGH replacement therapy

Earliest registry of a companion molecular diagnostics (CMDx) test

1998

FDA approved Herceptin(trastuzumab) and HerceptTest

the first “official” CMDx

Introduction

- Cutting-edge biochemical advances
  - single-nucleotide polymorphisms (SNPs), genotyping, and biochips

- Variations such as SNPs, insertions and deletions, structural variants, and copy number variations in the human genome
  - Cancer, diabetes, and neurodegenerative and cardiovascular diseases

- The association between antimalarial drugs and G6PD deficiency
  - the first examples of personalized therapy

Introduction

- **Personalized or precision medicine**
  - An individual's genetic profile
    → prevention, diagnosis, and treatment of disease
  - Based on a **data-driven approach**: medical, genetic, behavioral, and environmental information
  - Scientific advancements:
    - high-throughput, high-resolution data-generating technologies
      → cost-effective analysis of big datasets
  - A need for new computational approaches
  - Aim: provide the right treatments to the right patients at the right time

Introduction

- Overview
  - Recent advancements
  - Update on important developments in the analysis of big data
  - Future strategies for personalized medicine

  - **Identify key conceptual and infrastructural challenges**

02

The Conceptualization of Big Data

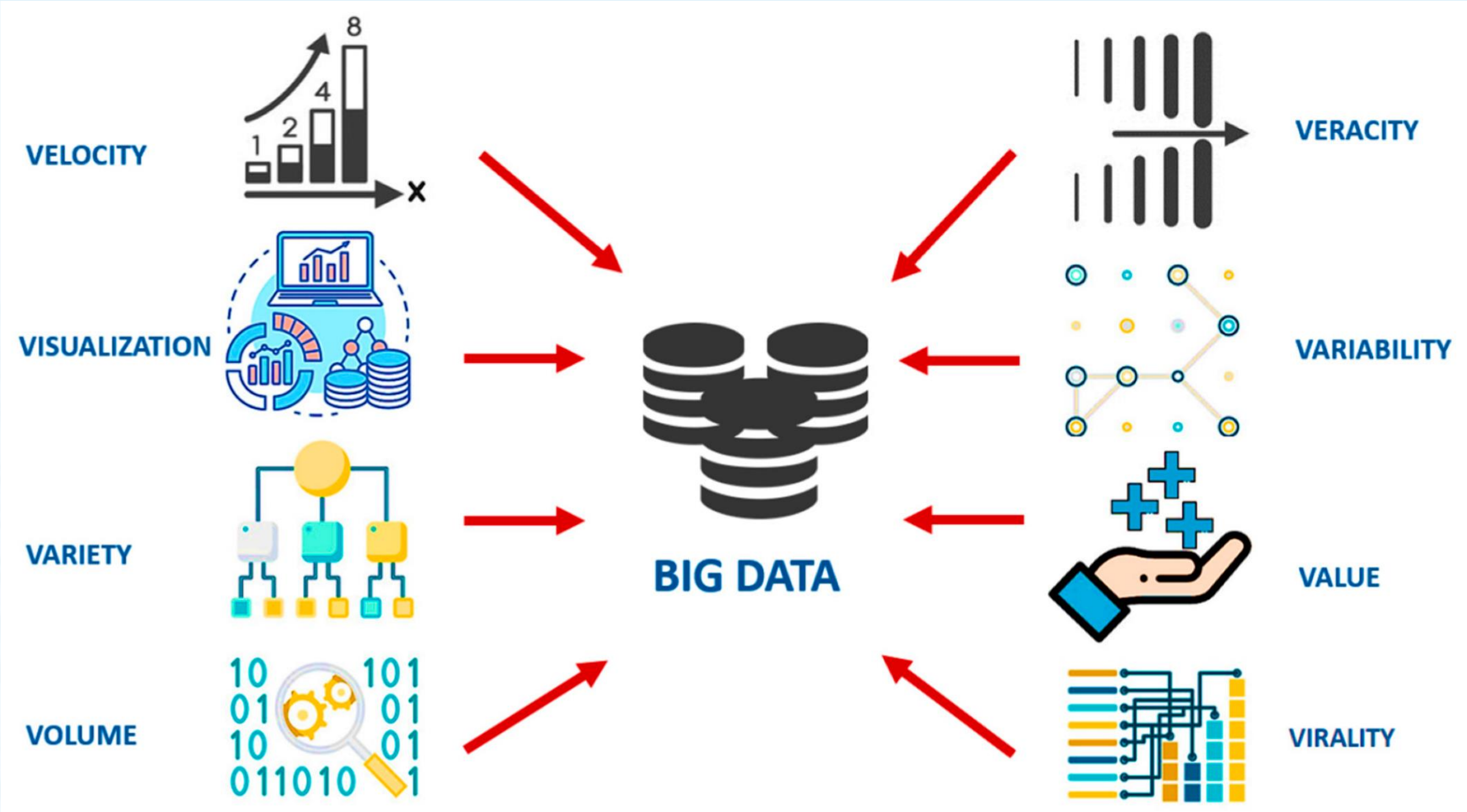


Figure 1. Representation of distinct dimensions of big data

The Conceptualization of Big Data

- Definition
  - High volume, high diversity biological, clinical, environmental, and lifestyle information
  - Collected from **single individuals to large cohorts**
  - At one or several time points

- Sources
  - medical records of patients, results of medical examinations, and hospital records

The Conceptualization of Big Data

- Advances in technology
  - Sequencing of DNA, RNA, and the characterization of proteins

- Urgently needed
  - high-end computing solutions, along with appropriate infrastructure

- For future
  - advanced machine learning algorithms and techniques
    - such as deep learning, and cognitive computing
    - multi-view big data analysis → explain an event or predict an outcome

The Conceptualization of Big Data

- **Difficulties and challenges**
  - lack of validation via prospective clinical trials, unsatisfactory performance of predictive models, and difficulties in interpreting complex models

  - The number of examples (patients) is usually very small in relation to the number of genes.

    → Needs robust sampling methods

  - Data formatting and the storing
    - Standard genomic data formats: **FASTQ**, **BAM/CRAM**, and **VCF files**
    - Still incomplete standardization

      → incompatibility between the inputs and outputs

      → **ExAC**, **GNOMAD**, and **the Beacon Network databases**

    - **** Downstream data formats:** non-uniform analysis → different outcomes

03

Computational Approaches toward Personalized Medicine

- Aim of personalized medicine
  - Right treatments to the right patients

- Rapid development in various high-throughput technologies
  - **molecular and cellular biology-related data**
  - different aspects of personalized medicine such as **diagnosis, prognosis, and pharmacogenomics**
  - refine the existing disease maps, development of a new predictive model

Computational Approaches toward Personalized Medicine

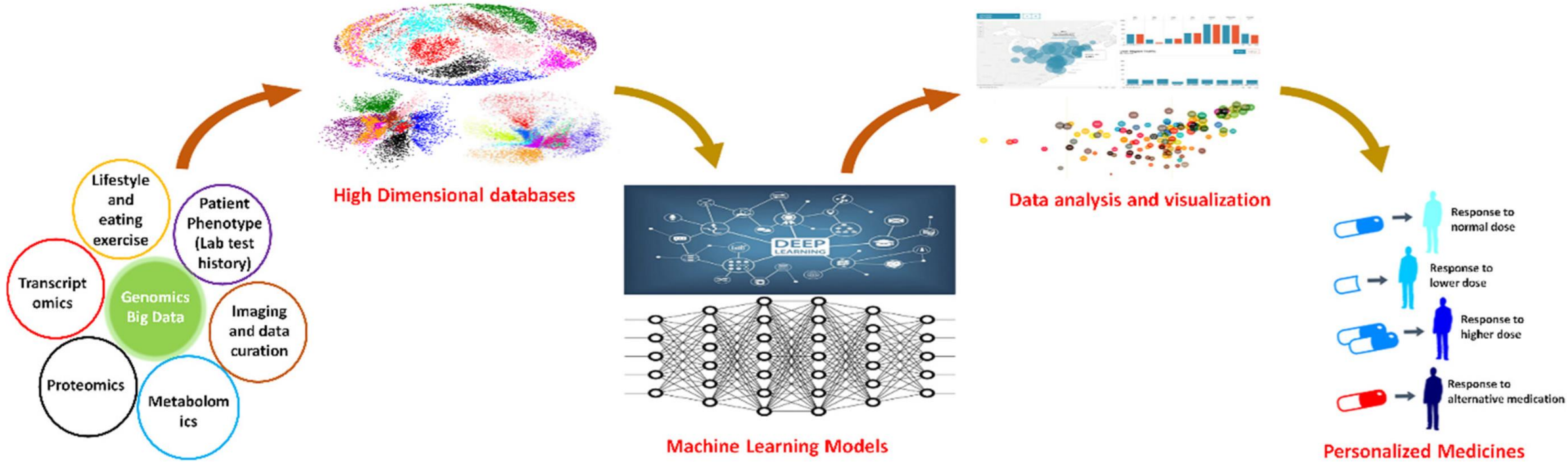


Figure 2. The overall computational approach for personalized medicine.

Computational Approaches toward Personalized Medicine

- Current computational models
  - Disease modeling, biomarker research, assessment of drug efficacy and safety

  - **Two types of models**
    - Mechanistic models
      - structural representation of the governing physiological processes

    - Data derived models
      - data-driven approaches (machine learning (ML), deep learning (DL)
      - Algorithms, AI

Computational Approaches toward Personalized Medicine

- Examples of different models

- Molecular Interaction Maps (MIMs)
  - the physical and causal interactions based on knowledge
  - different mechanistic pathways and regulatory modules

  - **Identify network static properties**

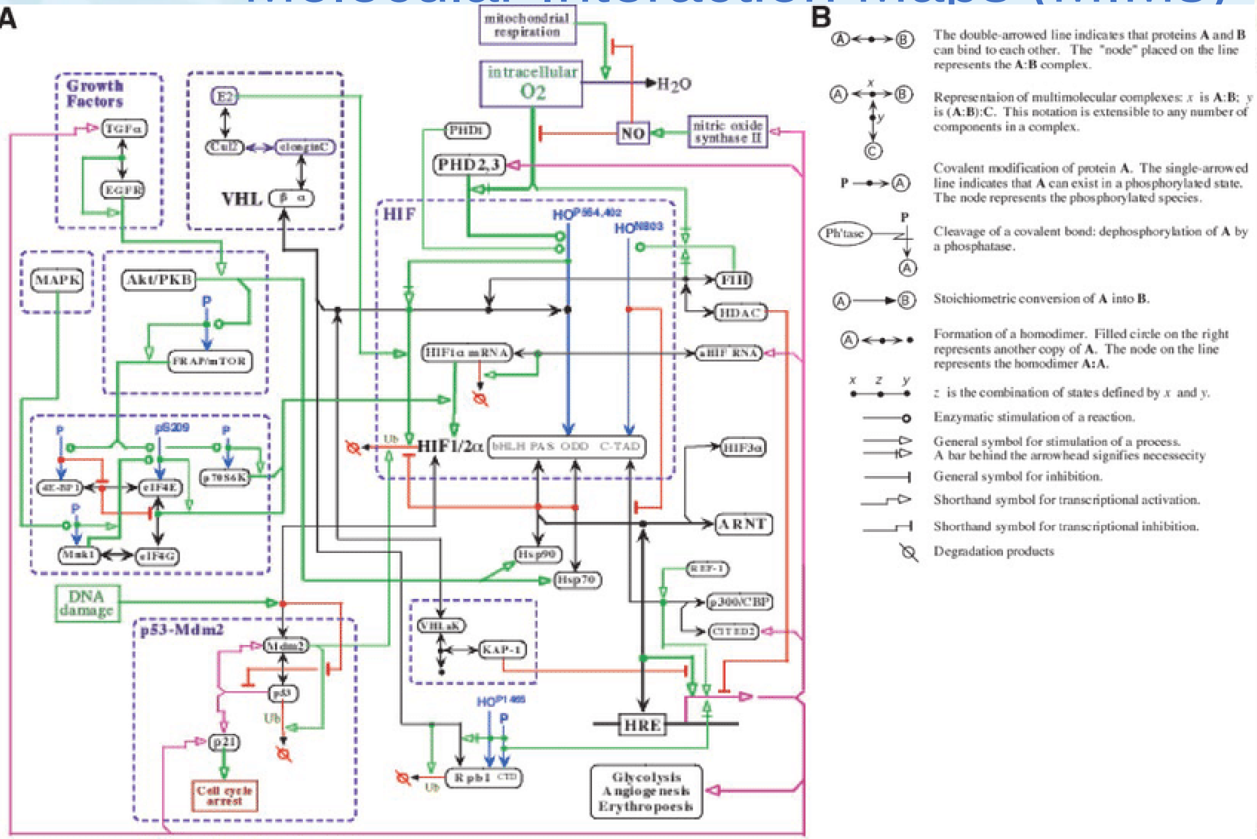
    (i) the identification of critical nodes; (ii) community detection;
    (iii) prediction of hidden links

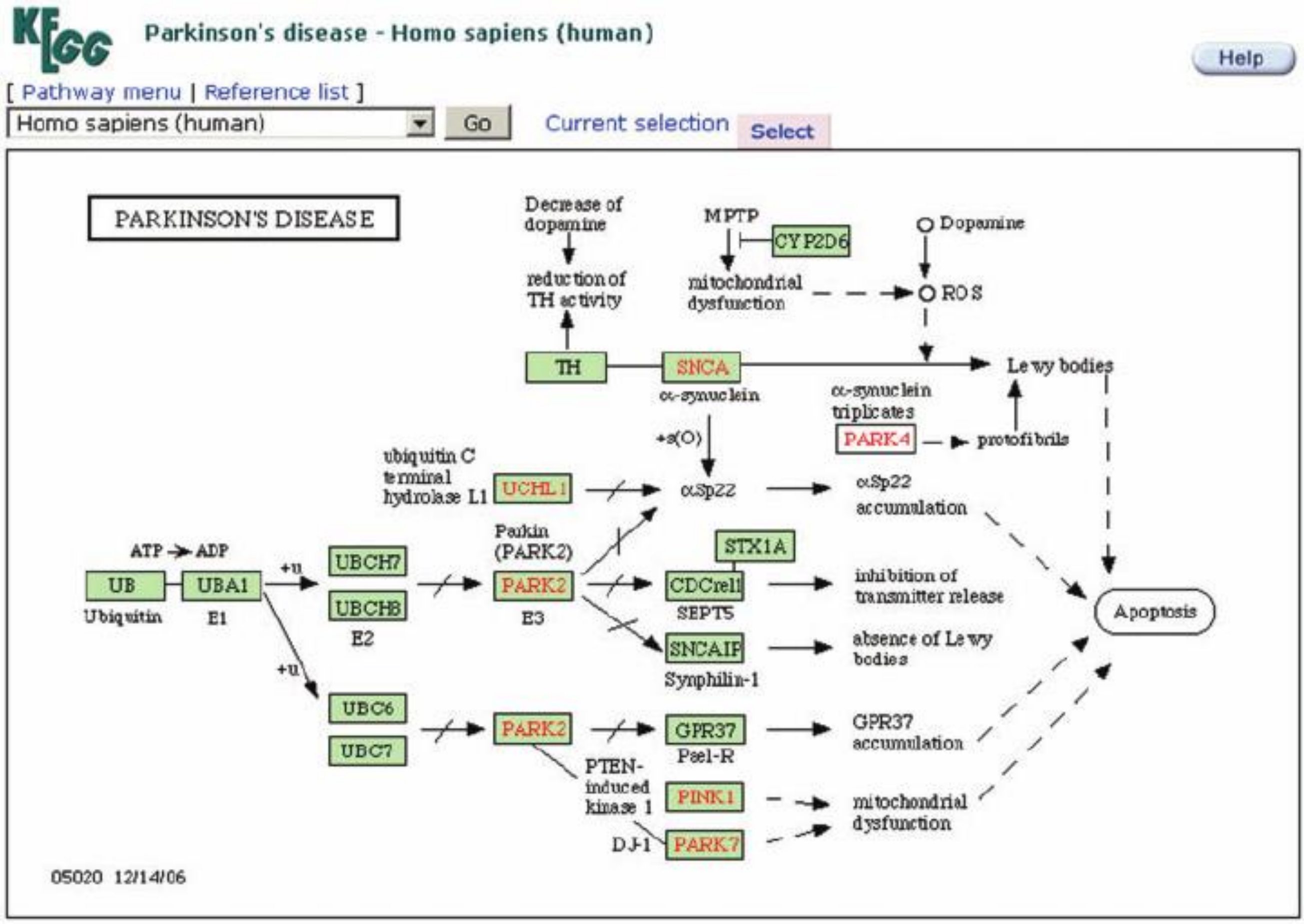
  - provide the **simplest mechanistic visualization** of data

Computational Approaches toward Personalized Medicine

- Examples of different models

- Molecular Interaction Maps (MIMs)





Computational Approaches toward Personalized Medicine

- Examples of different models

- Constraint-Based Models
  - **Genome-scale metabolic (GEM) models**

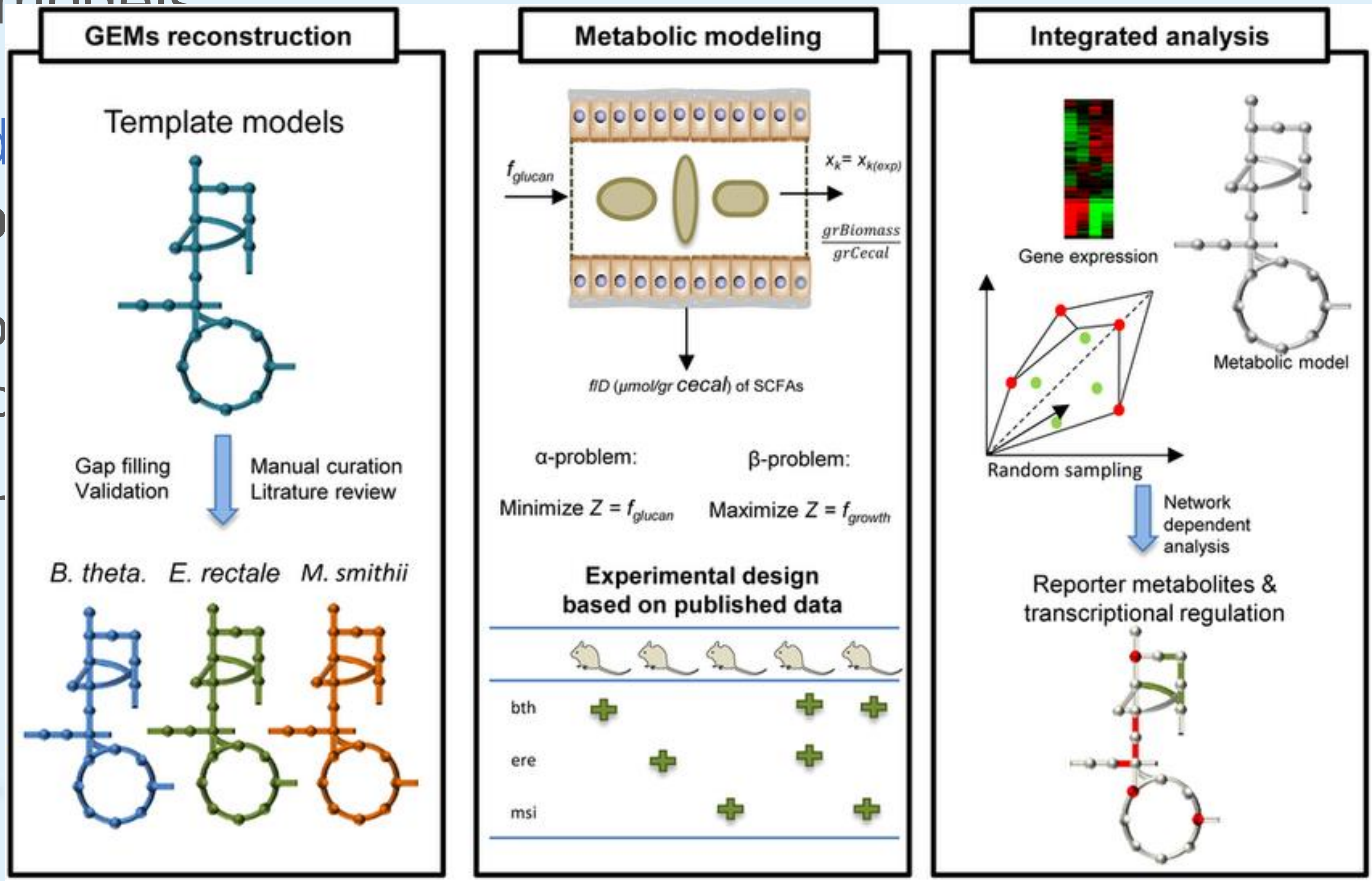
  - mathematical framework

    → the metabolic capacities of a cell, wide analysis of genetic perturbations

  - In multiple medical domains: cancer, obesity, Alzheimer’s disease

Computational Approaches toward Personalized Medicine

- Examples of different models
- Constraint-Based Mod
  - **Genome-scale metab**
  - mathematical framewo
    → the metabolic capac
  - In multiple medical dor



https://www.researchgate.net/profile/Adil-Mardinoglu/publication/256188965/figure/fig3/AS:279027651760151@1443536784980/Genome-scale-metabolic-models-GEMs-for-gut-microbiota-are-reconstructed-based-on-three.png

Computational Approaches toward Personalized Medicine

- Examples of different models

- Boolean Models (BMs)
  - **The simplest logic-based models**
    - two possible states: 1 (ON, activation) or 0 (OFF, inactivation)
    - Regulators (upstream) ↔ Targets (downstream): AND, OR, and NOT
    - Not require detailed kinetic data

    - Application to large biological systems: **cancer research**

Computational Approaches toward Personalized Medicine

- Examples of different models

- Quantitative Models (QMs)
  - Ordinary differential equation based
    - Quantitative behavior of a biochemical reaction with time
      → responds to different stimuli or perturbations
    - Explains the **biological-systems dynamics** in detail and applies to a single pathway
    - Application: Individual biomarker discovery, drug response, tailored treatment

Computational Approaches toward Personalized Medicine

- Examples of different models

- Pharmacokinetic Models
  - Surrogate for drug-induced responses
  - pharmacokinetic (PK) modeling or physiologically based PK (PBPK) modeling

04

Machine Learning Perspectives on Personalized Medicine

- Application: massive data collected through genome sequencing
- Aim: precisely define of treatment method
- Integration of the assorted patient data → biomarker discovery for various disease diagnoses
- Examples:
  - MammaPrint prognostic test: formalin-fixed-paraffin-embedded (FFPE) or fresh tissue for microarray analysis
  - Bejnordi et al. : detect metastases in LNs in stained tissue sections of breast cancer
  - Madani et al. : A machine learning echocardiography algorithm, diagnosis of cardiac disease

Machine Learning Perspectives on Personalized Medicine

- Machine learning and AI approaches
  - genetic, genomic epigenomic, transcriptomic, metabolomic data, medical images, biobanks data, electronic health records (EHR)
  - Two problems of interests
    - **Regression**: predict the value of continuous and real value quantities
      - the level of cholesterol in blood based on other biomarkers
    - **Classification problem**: predict the label of a set of individuals in a broad class
      - patients who have a survival time greater than the average from the rest

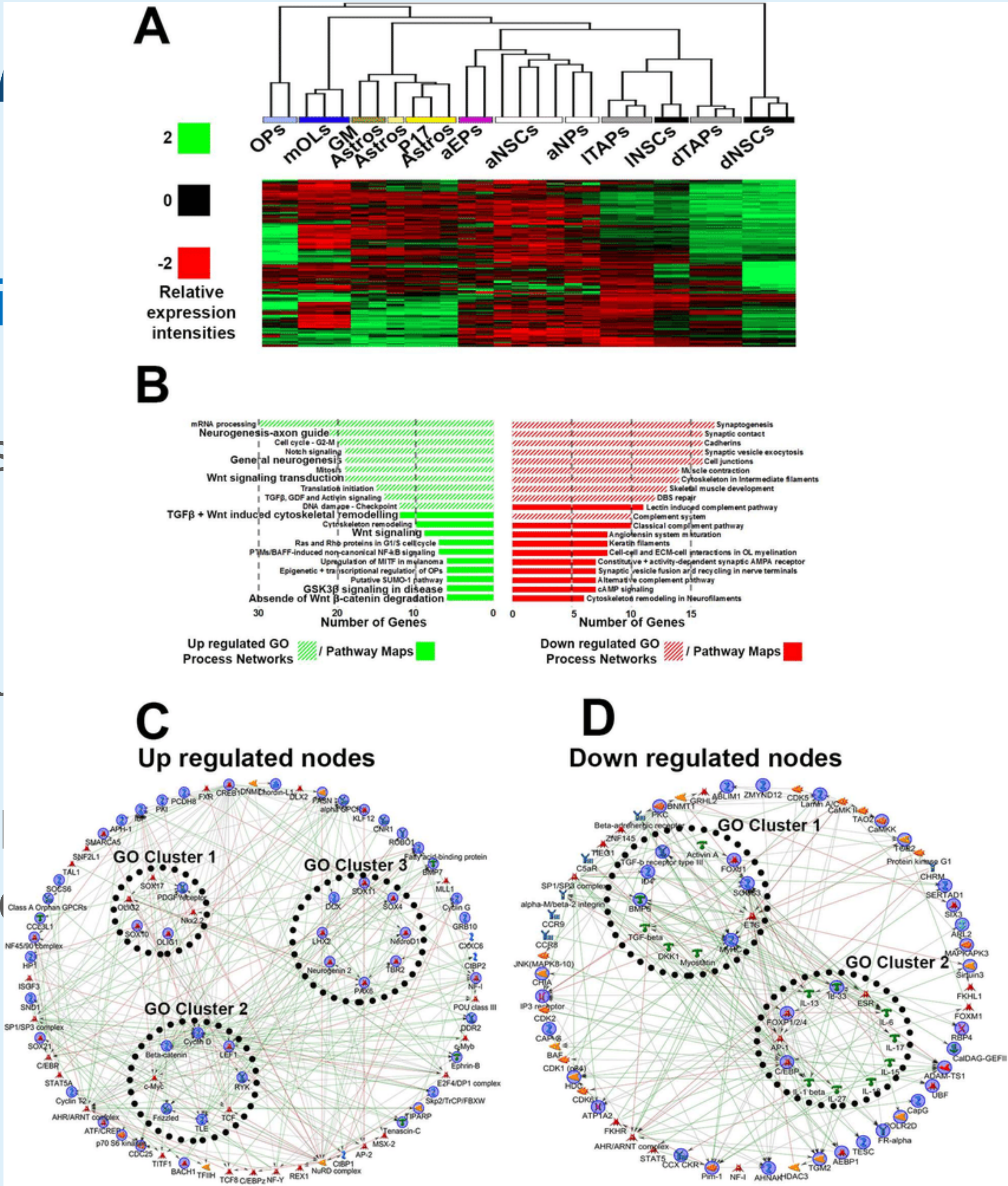
05

Modeling Genetic Data with Translational Purposes

- The genetic and epigenetic regulation of the altered pathways
  - Main topics in **pharmacogenomics**
  - mutation in the DNA impacts the transcriptome and the proteome downstream
  - **Transcriptomics**
    - how gene expressions, genetic pathways, regulatory networks are altered in each phenotype
    - Connectivity map (CMAP): Broad Institute → drug repositioning
    - new therapeutic uses in FDA approved compounds

Modeling Genetic Data w
Translational Purposes

- The genetic and epigenetic regulati
  - Main topics in **pharmacogenomics**
  - mutation in the DNA impacts the trans downstream
  - **Transcriptomics**
    - how gene expressions, genetic pa altered in each phenotype
    - Connectivity map (CMAP): Broad I
    - new therapeutic uses in FDA appr



https://www.researchgate.net/publication/315825259/figure/fig4/AS:480883833610243@1491663048643/Connectivity-map-CMAP-pathway-and-network-analysis-for-AR-A014418-target-genes-Gene.png

Modeling Genetic Data with Translational Purposes

- ***de novo* drug design**

- **Gene therapy**
  - Replacing a mutated gene that causes disease with a healthy copy of the gene.
  - Inactivating, or “knocking out,” a mutated gene that is functioning improperly.
  - Introducing a new gene into the body to help fight a disease.

- This promising treatment technique **remains risky**.

06

Data Mining Tools/Algorithms and Their Applications for Personalized Medicine

- The use of **multimodal data** helps in a deeper analysis of large datasets
- Algorithms → terminal node in the final predictions from big data
  - Lee: **Person-centered data** mining algorithm
    → Integrate both genetic information and baseline profiles
  - Ulyantsev: “MetaFast” → analyze metagenomes from novel environmental niches

  - Algorithms **is not to replace physicians**, but to provide them with tools that support their decisions.

Data Mining Tools/Algorithms and Their Applications for Personalized Medicine

- Pattern-Based Approaches
  - Discovery of sequential patterns
  - Pattern-based approaches (e.g., clustering and temporal pattern analysis)

  - **Temporal data**
    - depend on time series, with or without a sequence of events
    - the time-based quantitative measurements or sequence of temporal events related to particular clinical study

Data Mining Tools/Algorithms and Their Applications for Personalized Medicine

- Network Mining for Personalized Medicine and Health Care
  - The medical industry collects data
    - Most of which are **electronic health records (EHRs) collected by HIPAA** covered health care facilities
    - Guarded carefully by the Health Insurance Portability and Accountability Act (HIPAA) and are not available openly
    - not shared centrally to prevent the misuse of big data

Data Mining Tools/Algorithms and Their Applications for Personalized Medicine

- Big Data Management Problems in Precision Medicine and Health Care
  - Different types of barrier: philosophical, legal, and practical exist
  - Several issues: collecting and standardizing the heterogenous data, data curation; data de-identification and anonymization; legal consents
  - **G&T-seq**: simultaneously obtain both transcriptomic and genomic information from a single cell
  - Shortage of bioinformaticians

Data Mining Tools/Algorithms and Their Applications for Personalized Medicine

- Significance of Next Generation Informatics for Big Data in Precision Medicine Era
  - Transform biomedical data into useful drug development information
  - Apply the knowledge for decision support in clinical practice
  - Multiple applications
    - Proteomics, genomics, clinical prognosis, cancer treatment, aging, analysis of defective pathways, and drug repositioning

07

Heterogeneity, a Huge Challenge in Big Data Analysis

Figure 3. Big data challenges in recent times.

08

Role of Big Data in Accelerating Digital Healthcare

- Empower digital health care
  - Timely access of clinicians to the entire scope of a patient’s health information
  - Make more accurate predictions of where a patient’s health is trending
  - Provide more opportunities for proactive intervention
  - Be beneficial for the patients especially for those who can utilize telemedicine and remote patient monitoring

09

Big Data Applications in Health Care

- Two major divisions of health care big data
  - **Vital** and **social** data

10

Electronic Health Records

- Most significant application of big data in medicine and health care
    - Reporting demographics, medical history, allergies, and laboratory test results in digital forms

11

Health Big Data as a Key Player for Informed Strategic Planning

- Better understating of these data and better strategic plans can cure more patients in the most diverse areas

12

Advanced Risk and Disease Management through Big Data

- Tackling the hospitalization risk for particular patients with chronic diseases

→ prevent deterioration, provide accurate preventative care

→ reduce hospital admissions

13

Developing New Therapies and Big Data

- Identify the potential strengths and weaknesses in clinical trials or therapeutic processes
- Development of new ground breaking drugs and innovative, forward-thinking therapies

14

Impediments of Big Data in Health Care

- Medical data have been collected across different states, hospitals, and administrative departments using different protocols.
  - → standard regression-based methods

- Privacy: a hacker could identify an individual including financial and other confidential information
  - → most countries have created legislative principles: e.g. HIPAA in USA

15

Conclusions and Future Prospects

- The big data paradigm shift is significantly transforming health care and biomedical research.
  - Four dimensions of volume, velocity, variety, and veracity,
  - referring to scale, rate, forms, and content of generated data
- Genomics data
  - address personalized health care issues
  - help to propose new drugs for the treatment of gene related disorders
- Advanced machine learning approaches
  - artificial intelligence and deep learning → future toolbox

Thank You!