Session 4 - Genomics: Core, Allied & Applied Fields

Genomics, Human Genome Project (HGP), WGS, WES, Genomic and Personalised Medicine, Genome India Project (GIP), Proteomic and Transcriptomic Analysis, Human Microbiome Project, 1000-Genome Project, Telomere-to-Telomere (T2T) Project

Genomics

Genomics Overview

- Genomics is the study of an organism's entire DNA set, including all genes, their functions, and interactions with each other and the environment.
- Objectives include understanding and curing diseases, predicting individual disease risks, and studying drug interactions pre-clinical trials.
- **Applications of Genomics:** Personalized medicine, forensic science, genealogical assessment, and agricultural improvements.

Modern Tools Used in Genomics

1. **Next Generation Sequencers**
    - Facilitate rapid, efficient DNA sequencing for whole-genome analysis.
2. **RNA-Seq**
    - Sequences RNA molecules to study gene expression and regulation.
3. **Single Cell Sequencing Kits**
    - Enable genome study of individual cells, providing insights into cellular diversity and disease mechanisms.
4. **Bisulfite Sequencing / Methyl-Seq**
    - Used for studying DNA methylation, crucial for gene regulation and disease association.

Genomics in Practical Applications

Application Area	Description	Example
Forensic Science	Genomic analysis for individual identification using DNA samples.	The first use in 2001 helped the FBI solve an anthrax mailing case.
Agriculture	Utilizing genomics for plant breeding and improving crop varieties.	Assessing genotype-phenotype relationships for complex traits in crops.
Crop Biofortification	Enhancing nutritional and phytochemical profiles of crops.	Developing crops with higher nutrient levels for improved health benefits.
Personalized Medicine	Creating tailored treatment plans based on genetic profiles.	Targeted therapies for cancer patients based on genetic mutations.
Conservation Biology	Using genomics for species conservation and biodiversity studies.	Identifying genetic diversity in endangered species for conservation strategies.
Infectious Disease Control	Studying pathogen genomics to understand and control outbreaks.	Genomic analysis of COVID-19 for tracking virus mutations and spread.
Pharmacogenomics	Understanding how genes affect an individual's response to drugs.	Tailoring antidepressant treatments based on genetic testing for drug metabolism.
Prenatal Screening	Detecting genetic disorders in fetuses through genomic tests.	Non-invasive prenatal testing (NIPT) for Down syndrome and other chromosomal abnormalities.
Evolutionary Biology	Tracing the evolutionary history and relationships of species.	Genomic studies revealing the evolutionary history of the human species.
Biotechnology	Using genomics in genetic engineering and synthetic biology.	Developing genetically modified organisms (GMOs) for agriculture or medicine.

Human Genome Project (HGP)

- The **Human Genome Project (HGP)** was an international scientific endeavor aimed at understanding human DNA.
- Launched in October 1990 and completed in April 2003, it is considered one of the most significant scientific projects in history.

Objectives of HGP

1. Determine sequences of all base pairs in the human genome.
2. Identify all genes in the human genome.
3. Store this information in databases.
4. Develop efficient genome analysis technology.
5. Enhance tools for data analysis.
6. Transfer related technologies to the private sector.

7. Address ethical, legal, and social issues (ELSI) related to the project.

Collaboration and Sequencing

- Coordinated by the U.S. Department of Energy (DOE) and the National Institutes of Health, with contributions from international entities.
- A collaborative international effort involving researchers worldwide.
- Produced a genome sequence covering over 90% of the human genome, using available technologies at that time.

Findings of the HGP

- The human genome contains 3.2 billion nucleotide bases (A, C, T, and G) and approximately 25,000 genes.
- The average gene consists of 3,000 bases, with dystrophin being the largest known human gene at 2.4 million base pairs.
- Chromosome 1 has the most genes, while the Y chromosome has the fewest.
- Functions of over 50% of discovered genes remain unknown.
- Human genome sequence is almost exactly the same (99.9%) in all people.
    - However, millions of locations identified where single-base DNA differences occur in humans, crucial for understanding common diseases.
- Only about 2% of the genome encodes instructions for protein synthesis.
- Repeat sequences, constituting at least 50% of the human genome, are thought to have no direct functions but play a role in chromosome structure and dynamics.
- Genes are concentrated in random areas with vast expanses of noncoding DNA between them.
- Over 40% of predicted human proteins share similarities with fruit-fly or worm proteins.

Whole Genome Sequencing

Definition and Purpose

- **Whole Genome Sequencing (WGS)** is a genetic test that determines the entire DNA sequence of an organism, encompassing both coding and non-coding regions.
- WGS provides insights into an individual's ancestry, health, traits, and drug response.

Tools Utilized in WGS

- Sequencing machines like **Sanger sequencing**, **Illumina sequencing**, and **nanopore sequencing**.
- **Bioinformatics software** for aligning DNA fragments, assembling genomes, and identifying variants.

- Databases and algorithms for genome annotation, interpretation, and comparison.

Advantages

- Offers a comprehensive view of the genome, detecting rare or novel variants.
- Aids in diagnosing diseases, especially those influenced by multiple genes or complex interactions.
- Enables **personalized medicine** by customizing treatments based on genetic profiles.

Challenges

- Can be expensive and time-intensive.
    - The cost of WGS has decreased from $10,000 in 2008 to under $1,000, excluding data storage, analysis, and interpretation costs.
    - Duration varies from weeks to months, with some companies offering services within three weeks.
- Generates substantial data volumes, posing storage, analysis, and interpretation challenges.
- Raises ethical, legal, and social concerns like privacy and consent issues.

Whole Exome Sequencing (WES)

- Sequences all protein-coding gene regions (exome).
- Identifies genetic variants associated with diseases, traits, and drug responses.
- A cost-effective alternative to WGS, focusing on 1-2% of the genome with most known disease-related variants.
- Limited by its inability to detect non-coding variants and the requirement for target enrichment methods.

Genomic and Personalised Medicine

Genomic Medicine

- Utilizes genome information to guide healthcare, helping in disease diagnosis, risk prediction, treatment choice, and complication prevention.

Personalised Medicine

- A broader medical approach tailored to individual characteristics, including genes, environment, lifestyle, and biomarkers.
- Genomic medicine, focusing on genes and genomes, is a subset of personalised medicine.

Clinical Applications of Genomic Medicine

Clinical Applications	Description
Cancer Treatment	Genomic testing identifies mutations or biomarkers for targeted cancer therapies, like EGFR gene mutations.
Rare Disease Diagnosis	Genomic sequencing aids in diagnosing rare genetic disorders, exemplified by **whole exome sequencing (WES)** identifying mutations in syndromes like Cornelia de Lange, Kabuki, or Noonan.
Pharmacogenomics	- Predicts individual drug responses, adjusting dosages to avoid adverse reactions, such as in patients with TPMT gene variants.

Benefits of Genomic Medicine in Cancer Treatment

1. **Personalized Treatment Plans**: Tailoring treatment plans based on genetic variations related to cancer.
2. **Early Detection and Prevention**: Identifying genetic predispositions to cancer for early intervention.
3. **Improved Patient Outcomes**: Personalizing treatment plans for more effective healthcare.
4. **Reduced Healthcare Costs**: Using genomic tools for targeted therapies, reducing unnecessary treatments.
5. **Increased Patient Satisfaction**: Enhancing patient involvement in their healthcare decisions.
6. **Advancements in Cancer Research**: Contributing to a deeper understanding of cancer's genetic basis and new therapy development.

Genome India Project (GIP)

- The Genome India Project (GIP) is led by the Indian Institute of Science's Centre for Brain Research, involving over 20 institutions across India.
- Initiated in 2020, the project focuses on understanding genetic variations and disease-causing mutations in the Indian population.
- The Department of Biotechnology, Ministry of Science and Technology, targets sequencing 10,000 Indian genomes by the end of 2023.

Objectives of GIP

- Developing a comprehensive catalog of genetic variations in the Indian population, including SNPs and structural variations.
- Creating a reference haplotype structure for Indians to aid future genome-wide association studies.
- Designing genome-wide arrays for research and diagnostics at affordable costs.

- Establishing a biobank for DNA and plasma for future research use.

Impact on Various Fields

- The GIP is expected to significantly influence biotechnology, agriculture, and healthcare sectors.
- By cataloging Indian genomes, the project will aid in developing personalized medicine and predicting health and disease outcomes.
- Reflects India's advancement in gene therapies and precision medicine, leading towards next-generation medicine.

Implications for Healthcare

1. **Precision Healthcare**
    - Developing personalized medicine based on genomic information to anticipate and modulate health and disease outcomes.
    - Mapping disease propensities to genetic variations for targeted interventions.
2. **Predictive Diagnostics**
    - Enhancing understanding of diseases prevalent in the Indian population to support the development of predictive diagnostic markers and precision medicine.
3. **Targeted Preventive Care**
    - Identifying individuals at higher risk for certain diseases, allowing early intervention and targeted preventive care.
4. **Boost to Biotechnology Sector**
    - Expanding India's biotechnology sector, fostering growth in valuable companies and research institutions.
5. **Designing Genome-Wide Association Chips**
    - Building a catalog of genetic variations to facilitate the design of genome-wide association chips and support large-scale research.

Proteomic and Transcriptomic Analysis

Proteomic Analysis

- Focuses on the study of proteins, including their functions and interactions within cells, tissues, or organisms.
- Provides insights into molecular mechanisms of biological processes and aids in identifying therapeutic targets for drug development.

Transcriptomic Analysis

- Involves studying mRNA transcripts, the intermediary molecules between DNA and proteins.
- Helps identify gene expression levels and regulation, offering insights into molecular mechanisms of biological processes.

Applications in Healthcare

1. **Disease Biomarker Identification**
    - Detecting biomarkers for diseases like Alzheimer's disease through differential gene and protein expression analysis in patients versus healthy individuals.
2. **Drug Resistance Mechanisms**
    - Investigating molecular mechanisms of drug resistance, such as resistance in *Pseudomonas aeruginosa*.
3. **Environmental Adaptation Studies**
    - Examining adaptive responses to environmental challenges

Importance of Integrated Analysis

- Integrated proteomic and transcriptomic analyses provide a comprehensive understanding of molecular mechanisms in biological processes.
- Analysis of transcriptomic or proteomic data alone may not accurately characterize molecular processes.
- Integrated analyses are crucial for identifying potential therapeutic targets and developing new drugs or treatments.

Human Microbiome Project

The Human Microbiome Project (HMP) was a research initiative launched in 2007 to understand the role of microorganisms in the human body.

It has two key phases: **mapping the microbiome** and **understanding microbial function**.

The project has shown that changes in the gut microbiota are associated with conditions like obesity, diabetes, and Crohn's disease.

It has also revealed the vast diversity of microbial species in the human body and how individual variations in microbiome composition can impact health and disease management.

1000-Genome Project

1. **Launched in 2008**, the **1000 Genomes Project** was an international effort to map human genetic variation.
2. The **1000 Genomes Project** was named as such because, at its inception, the goal was to sequence the genomes of approximately 1,000 individuals.

3. Its primary aim was to **sequence the genomes** of a large, diverse population to understand genetic variations.
4. The project finally carried out **sequencing over 2,500 individuals** from 26 different populations worldwide.
5. A key achievement was the **identification of more than 88 million genetic variants**, including SNPs, insertions, and deletions.
6. It created a **comprehensive catalogue of genetic variants**, greatly expanding the known variants.
7. The project enhanced understanding of **human genetic diversity** across various populations.
8. It provided significant insights into **human evolutionary history** and migration patterns.
9. The project **drove advancements in genomic research**, especially in sequencing technology and data analysis.
10. The data serves as a crucial reference for **genetic studies in biomedical research**, aiding in disease and drug response research.
11. The 1000 Genomes Project's contributions are pivotal for the future of **genetics and personalized medicine**.

Telomere-to-Telomere (T2T) Project

Objectives and Focus

- The **Telomere-to-Telomere (T2T) Project** is an open, community-based initiative aimed at creating the first complete assembly of a human genome.
- It addresses the remaining **8% of the human genome**, which includes complex regions like segmental duplications, ribosomal rRNA gene arrays, and satellite arrays.
- The project's ultimate goal is to enhance technology for producing high-quality, complete assemblies of diploid human genomes.

Achievements

- In March 2022, the T2T Consortium published the **first complete, gapless sequence of a human genome**.
- This achievement unveiled the final DNA sequences, making up a total of **3.055 billion base pairs**.

Methodologies and Tools

- The T2T Project has developed tools like **Iris** and **Jasmine** for refined genomic analysis.
- It utilized **high-coverage, ultra-long-read nanopore sequencing** to improve the telomere-to-telomere sequence.

Implications and Future Prospects

- The completion of the human genome sequence has profound implications for understanding human genetic variation and the genetic basis of complex diseases.
- It opens new avenues for research in **personalized medicine** and disease prevention.