

Week 1

Tuesday, 7 September 2021 22:04

WHY GENOMICS

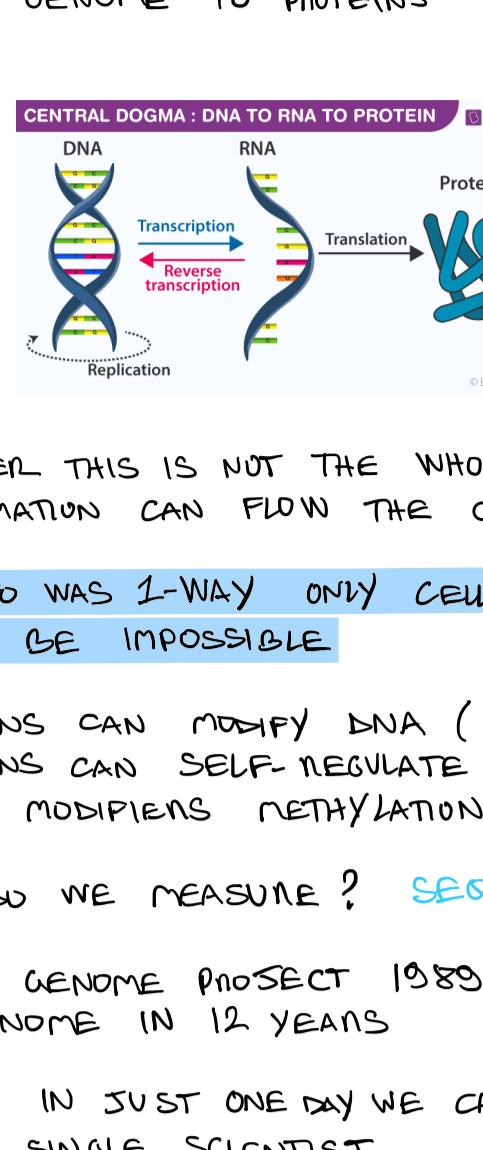
STUDYING THE GENOMES INSIDE US

GENOMES GIVE US OUR DEVELOPMENT & BIOLOGY

WE LOOK NEARLY DIFFERENT, HOWEVER ONLY 99,9% IDENTICAL

SMALL AMOUNT OF CHANGES IN GENOME CREATES DIVERSITY

STUDYING WHAT IS DRIVING OUR DIFFERENCES



- DIVERSITY
- DEVELOPMENT
- CELL SPECIALIZATION
- CANCER

ALL CELLS HAVE THE SAME DNA (GENOME)
SAME PROGRAM, DIFFERENT EXECUTION

CANCER IS ESSENTIALLY A GENETIC DISEASE
CANCER IS DEFINED BY STATIONARY CELLS

COMMON PHENOTYPE: DIVING WITHOUT CONTROL

MUTATION CHANGE IN THE GENOME

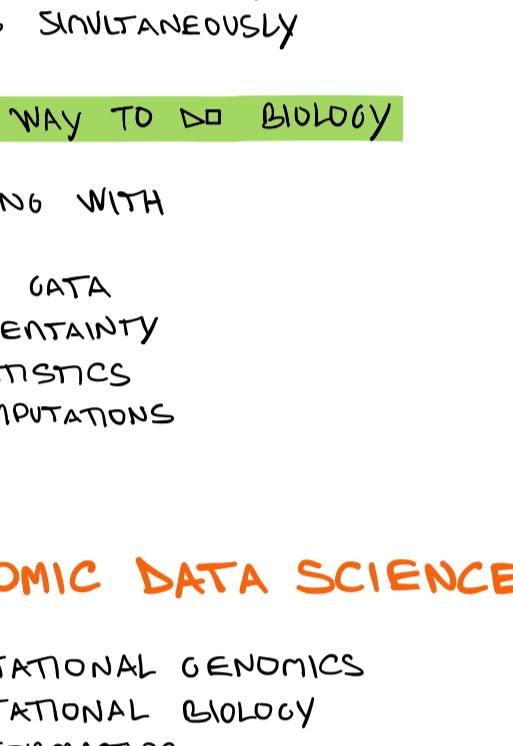
- DAMAGED DNA
- FAILED CELL REPLICATION

1-3 ERRORS X CELL DIVISION

HOW DOES A MUTATION TURN THE CELL CANCEROUS?
MUTATION AFFECTS GENE THAT CONTROLS CELL DIVISION

CENTRAL DOGMA (1 WAY INFORMATION FLOW)

FROM GENOME TO PROTEINS



HOWEVER THIS IS NOT THE WHOLE PICTURE
INFORMATION CAN FLOW THE OTHER WAY

IF INFO WAS 1-WAY ONLY CELL SPECIALIZATION
WOULD BE IMPOSSIBLE

PROTEINS CAN MODIFY DNA (TURNING GENES ON & OFF)

PROTEINS CAN SELF-REGULATE

OTHER MODIFICATIONS: METHYLATION (EPIGENETICS)

HOW DO WE MEASURE? SEQUENCING

HUMAN GENOME PROJECT 1989 - 2001

1 GENOME IN 12 YEARS

TODAY IN JUST ONE DAY WE CAN SEQUENCE HUNDREDS OF GENOMES
BY A SINGLE SCIENTIST

CANCER GENOMICS

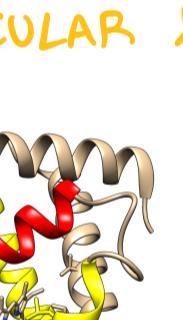
SEQUENCING THE ENTIRE GENOME OF A TUMOR

GENOMICS = BIG DATA

SEQUENCING IS GETTING CHEAPER \$

TODAY: 1K GENOME

NEW TECH FROM ILLUMINA

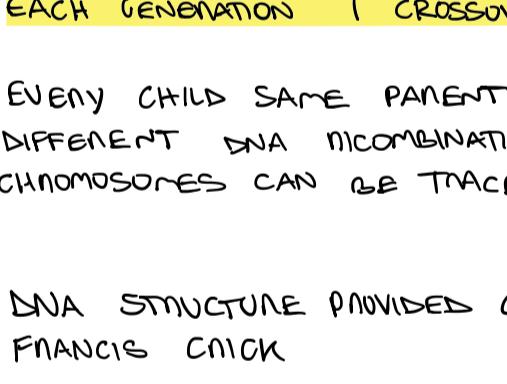


MORE EXPERIMENTS ARE NOW FEASIBLE

NCBI NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

SRA SEQUENCE READ ARCHIVE

WHAT IS GENOMICS



STRUCTURE (SEQUENCE)

3BP IN HUMAN GENOME

23 PAIRS OF CHROMOSOMES

22 AUTO + 1 SEX

TELOMERS, CENTROMERS

FUNCTION

WHAT DOES DNA DO?

WHAT HAPPENS DURING DEVELOPMENT

EVOLUTION

HOW GENOMES CHANGE OVER LONG TIME FRAMES

METAGENOMICS, PHYLOGENETICS

COMPARATIVE GENOMICS

ALL SPECIES SHARE SOME SEQUENCE IN COMMON

TO PERFORM SIMILAR FUNCTION (METABOLISM)

DNA REPLICATION, CELL DIVISION

MAPPING

WHERE ARE GENES LOCATED IN THE SEQUENCE?

- HEREDITABLE UNIT
- PART OF GENOMES THAT ENCODES A PROTEIN OR HAS REGULATORY FUNCTIONS

LITTLE BITS OF SEQUENCE THAT TURN INTO FUNCTIONAL ELEMENTS

GENETICS ONE GENE AT A TIME

GENOMICS ALL GENES AT ONCE

TECHNOLOGY IS THE REAL DRIVER

HIGH THROUGHPUT

MEASURING ACTIVITY OF HUNDREDS OF GENES SIMULTANEOUSLY

NEW WAY TO DO BIOLOGY

DEALING WITH

- BIG DATA
- UNCERTAINTY
- STATISTICS
- COMPUTATIONS

GENOMIC DATA SCIENCE

COMPUTATIONAL GENOMICS

COMPUTATIONAL BIOLOGY

BIOINFORMATICS

STATISTICAL GENOMICS



- SUBJECT (HUMAN)
- SAMPLE (SKIN CELL)
- PREPARE SAMPLE FOR ANALYSIS
- SEQUENCING
- ANALYSIS OF SHORT FRAGMENTS (READS)
- ALIGNING TO REFERENCE GENOME
- ANALYSIS OF GENOME
- DATA IN DATABASE

IN THE FUTURE WE WILL HAVE MANY REFERENCE GENOMES

2 COPIES OF CHROMOSOMES (MUM & DAD)

HOW DO THOSE COPY DIFFER?

EXPERIMENTAL DESIGN

QUESTION TO ANSWER

HOW MUCH DATA?

HOW MANY SUBJECTS?

WHAT KIND OF DATA DO WE GET?

CHIP-SEQ 3 NEW

METHYL-SEQ 2 NEW

STUDYING OTHER THINGS OF THE GENOME

EXPERIMENTS ARE EXPENSIVE & TIME CONSUMING

YOU MUST BE SURE THAT THE DATA HELPS YOU TO ANSWER YOUR QUESTION

ALIGNMENT IS ASSEMBLY

HOW MUCH YOUR GENE WAS PRESENT

PBQ PROCESSING & NORMALIZATION

SEQUENCING TECH MAKES MISTAKES

COLLECTING THE DATA INTRODUCES BIASES

BIASES CAN BE SYSTEMATIC

COMPUTATIONAL & STATISTICAL METHODS

REMOVING SYSTEMATIC BIASES & BIASES

STATISTICS & MACHINE LEARNING

DEVELOPING SOFTWARE

RNA-SEQ AND

EXPERIMENTAL PROTOCOL (PARADIGM)

CAPTURED GENES TURNED ON IN A SET OF CELLS

STANDARDIZATION

POPULATION GENOMICS

WHY DID HE DEVELOP CANCER? (INDIVIDUAL)

WHAT MAKES A PARTICULAR GROUP OF INDIVIDUALS MORE SUSCEPTIBLE TO A PARTICULAR TYPE OF DISEASE?

WHY SOME PEOPLE ARE RESISTANT TO X?

WHY SOME PEOPLE EXHIBIT ACQUIRED TRAITS?

INTEGRATIVE BIOLOGY

SYSTEM BIOLOGY

MORE MEASURES INTEGRATED TOGETHER

TO MAKE BIOLOGICAL DECISIONS

GENOME

PROTEOME

METABOLOME

TRANSCRIPTOME

MOLECULAR & CELL BIOLOGY

EUKARYA, BACTERIA, ARCHAEA

PROKARYOT

MITOCHONDRIA HAS ITS OWN DNA

INDEPENDENT PROKARYOT ASSIMILATED BY ANCESTOR OF ALL EUKARYOTIC CELL

23 PAIRS OF CHROMOSOMES (MUM & DAD)

HOW DO THOSE COPY DIFFER?

EXPERIMENTAL DESIGN

QUESTION TO ANSWER

HOW MUCH DATA?

HOW MANY SUBJECTS?

WHAT KIND OF DATA DO WE GET?

CHIP-SEQ 3 NEW

METHYL-SEQ 2 NEW

STUDYING OTHER THINGS OF THE GENOME

EXPERIMENTS ARE EXPENSIVE & TIME CONSUMING

YOU MUST BE SURE THAT THE DATA HELPS YOU TO ANSWER YOUR QUESTION

ALIGNMENT IS ASSEMBLY

HOW MUCH YOUR GENE WAS PRESENT

PBQ PROCESSING & NORMALIZATION

SEQUENCING TECH MAKES MISTAKES

COLLECTING THE DATA INTRODUCES BIASES

BIASES CAN BE SYSTEMATIC

COMPUTATIONAL & STATISTICAL METHODS

REMOVING SYSTEMATIC BIASES & BIASES

STATISTICS & MACHINE LEARNING

DEVELOPING SOFTWARE

RNA-SEQ

EXPERIMENTAL PROTOCOL (PARADIGM)

CAPTURED GENES TURNED ON IN A SET OF CELLS

STANDARDIZATION

POPULATION GENOMICS

WHY DID HE DEVELOP CANCER? (INDIVIDUAL)

WHAT MAKES A PARTICULAR GROUP OF INDIVIDUALS MORE SUSCEPTIBLE TO A PARTICULAR TYPE OF DISEASE?

WHY SOME PEOPLE ARE RESISTANT TO X?

WHY SOME PEOPLE EXHIBIT ACQUIRED TRAITS?

1000 TIGER

FIRST COMPLETE BACTERIAL GENOME SEQUENCED

A. INFLUENZAE 1.8 BILLION (10^9) bp

S. CRAIG VENTER, H. SMITH

WHOLE SHOTGUN GENOME SEQUENCING

NO MAPS & LIBRARIES

TAKE WHOLE GENOME

FINISH IT INTO TINY PIECES

RANDOMLY SEQUENCED PIECES

BY OVERSAMPLING YOU CAN PUT YOUR READS INTO AN ASSEMBLER (SOFTWARE)

SEQUENCING EVERY PART OF THE GENOME MULTIPLE TIMES OVER

EVOLUTION FOR THE FIELD OF MICROBIAL GENOMICS

1000 TIGER

NHI VS APPLIED BIOSYSTEMS VS CELENA GENOMICS

(S. C. VENTER)

WHICH GENOME WAS SEQUENCED?

DIFFERENCE BETWEEN HUMANS

1 EVERY 100 BASE PAIRS bp

HUMAN STANDARD DNA

MOLECULE OF 12 VOLUNTEERS

NON-EUROPEAN ORIGIN

NOT FROM A SINGLE PERSON!

HOW MANY GENES DO WE HAVE?

BETTER DISEASE TREATMENTS