BIOINFORMATICS

dr. Nadia Pisanti

Dipartimento di Informatica

NEW GENERATION SEQUENCING (NGS)

Thanks to new cheaper technologies (Illumina, SOLiD, Roche, ION Torrent), thousands of *desktop sequencers* sit in hospitals and biology labs.

Ten years after the first one, sequencing a whole human genome takes
- from $2.7 billions to ~$1000
- from 13 years to few days

Sequencers are improving at a faster rate than computers are, and spew out much more data than geneticists can analyse.

This opens the way to new challenges in medicine and biology, and as many in computer science for collecting, storing, and extracting useful biological information from such datasets...

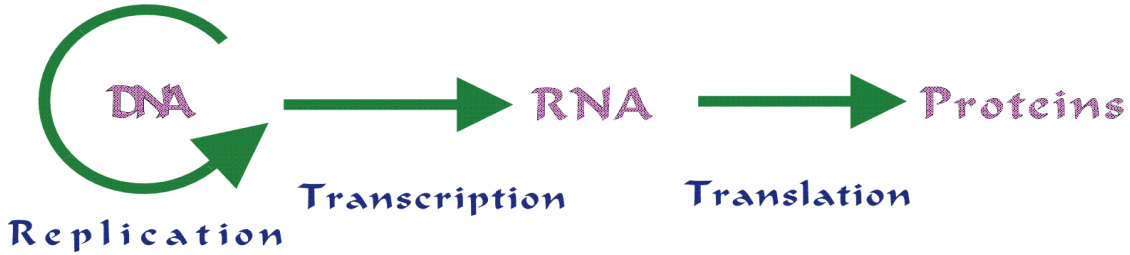
NEW *in silico* CHALLENGES FOR BIOLOGY

Metagenomics is the (comparative) study of microbial genetic material sampled their natural environment (ground, sea, animals gut) without prior culturing. NGS as a new tool to view the microbial world having the potential to revolutionise understanding of the entire living world.

Analysis and comparison of metagenomic data require fast and accurate new computational and statistical approaches.

Transcriptome sequencing (RNA-Seq) allows understanding regulatory mechanisms in gene expression for protein synthesis.

Detecting different ways and frequency of RNA fragments arrangement requires sophisticated text algorithms on very specific data structures.



NEW *in silico* MEDICAL APPLICATIONS

NGS has allowed to identify hundreds of genetic variants that cause (or increase the risk for) diseases.

Their discovery and association to the disease requires the analysis of lots of data with sophisticated methods (indexing, text algorithms, efficient data structures) and technologies.

Personalised medicine: genome guided targeted treatment.

The association of genomic variants to the responsiveness to drugs requires ICT tools for a fast detection of such variants in patients to make (cancer) pharmacogenomics a current practice.

Risk assessment in healthy individuals. At population level this can allow tasks prediction for health assistance, as well as prevention strategies.

Huge "regional" ICT projects exist for this.

OUR CURRENT RESEARCH 1/2

MOBILOMICS: comparative analysis of the mobile elements in genomes and how they drive evolution (dr.Pisanti, prof.Grossi, dr.Marangoni, dr.Menconi).

RNA-Seq & GRAPH ALGORITHMS: design and application of an optimal algorithm for the detection on RNA data of differential expression of genes on specific data structures used to store NGS data (dr.Pisanti, prof.Grossi, Univ. Florence, INRIA Lyon).

[as a side effect, we have conceived an optimal algorithm for the detection of cycles in undirected graphs improving the state-of-the-art that dates back to 1974]

PRE-CLINICAL DATA ANALYSIS: creation of a knowledge base for prediction and biosimulation in pre-clinical data for cerebral ischemia (dr.Pisanti, dr.Zini, dr.Trasciatti, dr.Bonaretti).

CANCER TREATMENT INVESTIGATION: comparative analysis of genomic and proteomic data to study differential gene expression in cancer cells with and without treatment (dr.Pisanti, dr.Zini, dr.Trasciatti, dr.Bonaretti).

*HAPLOTYPE PHASING: a*n algorithm for investigating how genomes alleles are arranged when inherited from mother and father (dr.Pisanti & CWI Amsterdam).

OUR CURRENT RESEARCH 2/2

INFERENCE OF REPETITIONS IN GENOMES: extraction and analysis of regularities in genomic sequences (dr.Pisanti, prof.Grossi, INRIA Lyon).

VIRTUAL CELL: first model (VICE) of "virtual cell" with minimal genome (prof.Degano, dr.Marangoni, dr.Chiarugi).

MODELS OF CANCER CELLS METABOLISM: developed of computational models of metabolic pathways in cancer cells for making *in silico* predictions of therapies efficacy targeting cancer cells (prof.Barbuti, prof.Maggiolo, dr.Milazzo, prof.Minutolo).

MODEL FOR CALYX OF HELD: first discrete stochastic model of Calyx of Held, a large synapse in the auditory brainstem structure (prof.Degano, dr.Bracciali).

MODEL-BASED CANCER THERAPIES: development of patient-specific computational models of the pharmacodynamics of some cancer drugs for estimating the most suitable dosage of treatments based on patient genomic information (prof.Barbuti, pro.Maggiolo, dr.Milazzo, dr.Pardini, dr.Sameen, prof.Danesi, prof.Del Re).

OUR MAIN COLLABORATORS

INRIA France

CWI Amsterdam

Geneticists of AOUP Pisa and Dept. of Biology,

Dept. of Chemistry,

Dept. of Clinical and Experimental Medicine.

Galileo Research (formerly ABIOGEN Pharma)

Microsoft Research center & Trento University Research center: COSBI.

NOVARTIS Research Center in Siena

University of Edinburgh

University of Stirling