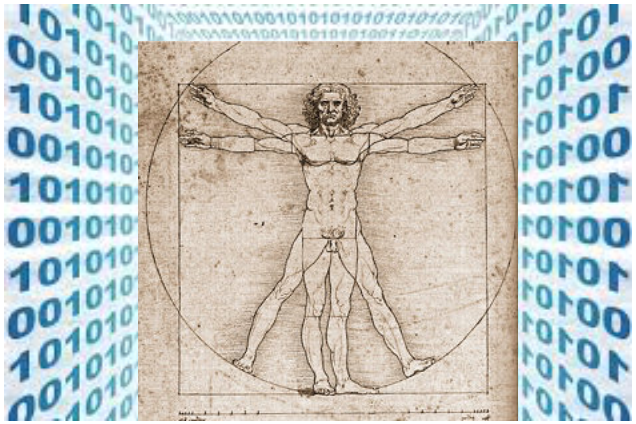


Introduction to biomedical data

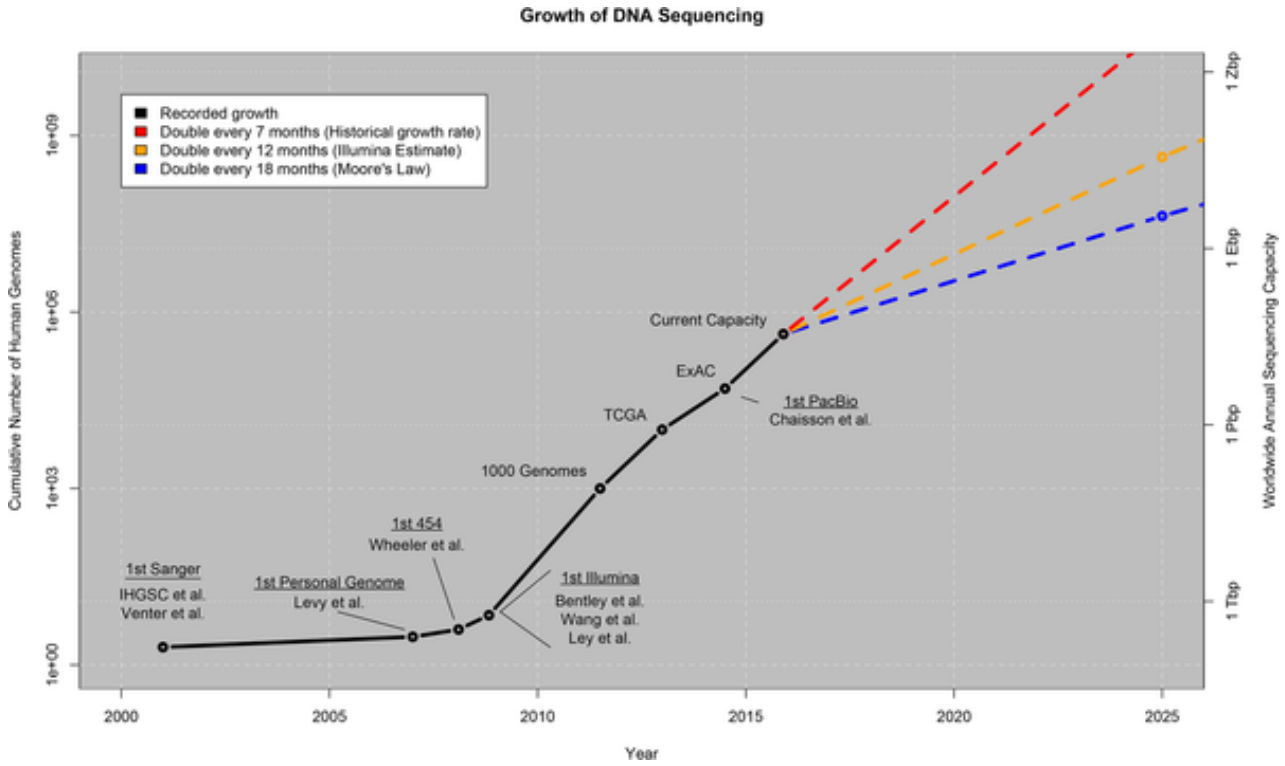
MERJA HEINÄNIEMI

ASSOCIATE PROFESSOR IN BIOINFORMATICS

INSTITUTE OF BIOMEDICINE, SCHOOL OF MEDICINE

UNIVERSITY OF EASTERN FINLAND

New challenges that bring us together

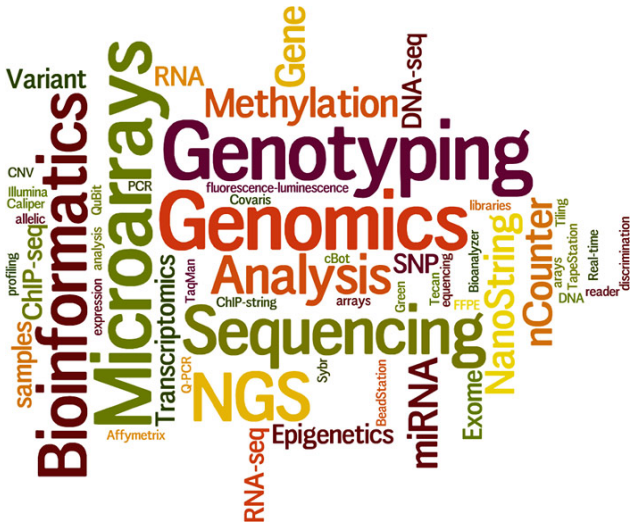


Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, et al. (2015) Big Data: Astronomical or Genomical?. PLoS Biol 13(7): e1002195. doi:10.1371/journal.pbio.1002195

PLOS | BIOLOGY

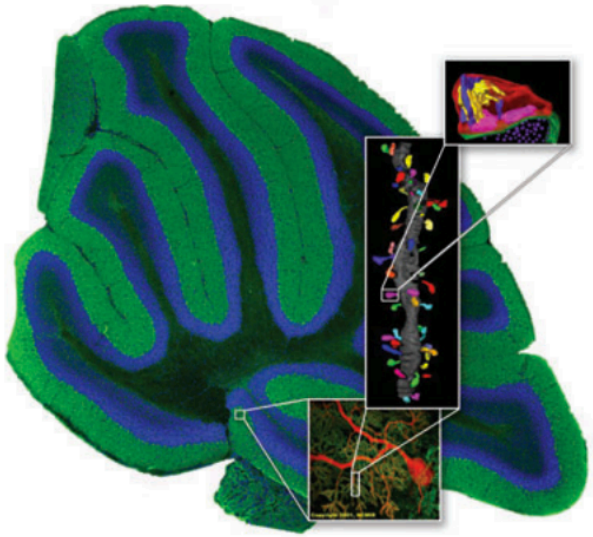
Big data in biomedicine

Genomics from genotyping and gene expression to deep- sequencing data



Big data in biomedicine

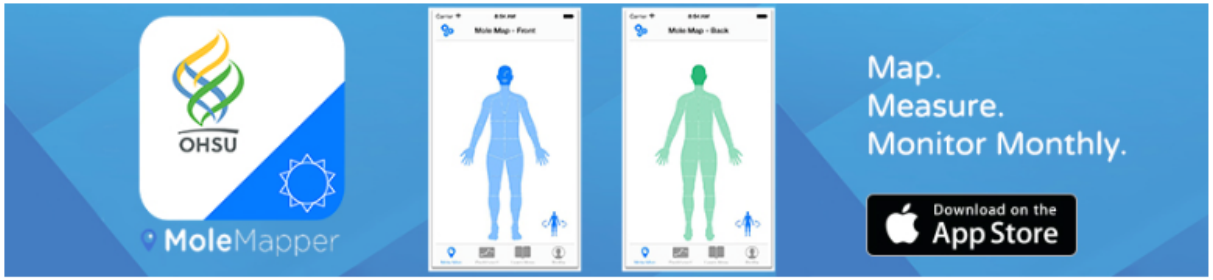
Imaging digital pathology, video recordings, integrating across imaging modalities



Big data in biomedicine

Health electronic medical records, pharmacy prescription information, insurance records

Personal recordings



Mole Mapper uses the phone's camera to track moles based on measurement, location and monitoring over time - **melanoma**

The Parkinson's app takes advantage of the phone's microphone, gyroscope and accelerometer to collect and track health and symptoms

Other patient-driven communities include https://www.patientslikeme.com

Sage BIONETWORKS

Collaborating with the UEF summer school to arrange the challenge for 2nd week

Big data in biomedicine

- Genomics data matrices: ***long & thin***

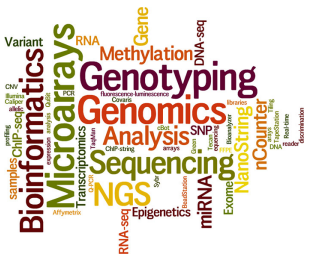
  10^4 to 10^6 variables ; 10 to 10^2 patients

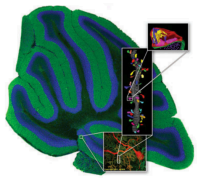
- Health record data matrices: ***short & broad***

  10 to 10^2 variables ; 10^4 to 10^6 patients

- Biomedical images:

  high-resolution digital images and movies





Future challenges

Modeling these together

Bioinfo track lectures: molecular data

Wednesday
Introduction to biomedical data - Merja Heinäniemi
Cell type deconvolution problem, Petri Pölönen
Lunch
Unsupervised dimensionality reduction, Juha Mehtonen
Multiview dimensionality reduction, Robert Ciszek + Break
Deep neural network applications, overview, Merja Heinäniemi
Deep neural network configurations for DNA motif analysis, Juha Mehtonen

Learning objectives

Domain knowledge	Basic molecular biology
Data types	RNA-sequencing and other omics data
ML applications	Unsupervised: dimensionality reduction Supervised: deep neural networks
In preparation of 2nd week	*Cell type deconvolution problem*

Bioinfo track lectures: imaging data

http://www.jussitohka.net

Learning objectives

Domain knowledge	Imaging modalities
Data types	MRI, tissue sections
ML applications	Mainly supervised methods

+ Integrating molecular and imaging data in multiscale models

http://www.cs.tut.fi/~ruusuvuo/

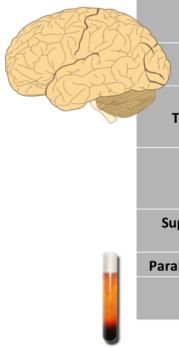
Bioinfo track independent assignment

You will analyze and model data generated by AMPAD-AD, *the Accelerating Medicines Partnership-Alzheimer's Disease*

Scientific hypothesis: neuro-degenerative diseases such as Alzheimer's disease affect the different brain cell types, resulting e.g. in neuronal loss in certain brain regions

Question: Which cell types are up-, down-regulated or unchanged in AD?

Human Tissue



HUMAN TISSUE	Diagnosis	Assay
Prefrontal Cortex	• Alzheimer's Disease • Mild Cognitive Impairment • Parkinson's Disease • Amyothrophic Lateral Sclerosis • Corticobasal Degeneration • Frontotemporal Dementia • Dementia with Lewy Bodies	• RNAseq • Gene Expression array • miRNA array • ChIPseq • DNA Methylation array • Proteomics • Confocal Imaging • SNP genotypes • Proteomics • Whole Exome Seq
Visual Cortex	• Alzheimer's Disease	• Gene Expression Array • SNP genotypes
Temporal Cortex	• Alzheimer's Disease • Progressive Supranuclear Palsy • Parkinson's Disease	• RNAseq • SNP genotypes
Cerebellum	• Alzheimer's Disease • Progressive Supranuclear Palsy • Parkinson's Disease	• RNAseq
Superiour Temporal Gyrus	• Alzheimer's Disease	• RNAseq • Whole Exome Seq
Parahippocampal Gyrus	• Alzheimer's Disease	• RNAseq
Serum	• Alzheimer's Disease • Mild Cognitive Impairment	• Metabolomics

Data: RNA-sequencing from postmortem tissue samples collected across 7 different brain regions
+ Associated metadata (clinical and technical)

Bioinfo track independent assignment

We can distinguish different cell types from measuring RNA levels or from tissue images

The problem: tissue samples represent the average RNA level profile across all cells (and cell types) – these are called bulk measurements

Bioinfo track independent assignment (1)

Molecular data *New type of data is now available: single cell RNA-sequencing (scRNAseq)*

The aim is to predict the proportion of different cell types from bulk RNASeq using clever approach utilizing data from single cell RNASeq -> cell type deconvolution problem

Challenge: ***Can you come up with a clever way to utilize scRNAseq to deconvolve the signal in the Alzheimer brain tissue samples?***

Bioinfo track independent assignment (2)

Imaging data *Brain imaging data is available: Allen Brain Atlas in situ hybridization*

The aim is to perform automated image annotation to find out which cell types / tissue sub-structures express the genes that change their level in Alzheimer's disease patients (RNA-seq profiles)

Challenge: ***Can you come up with a clever way to utilize Allen Brain Atlas data to link the gene expression changes in the Alzheimer brain to specific brain regions?***

Bioinfo track independent assignments

Choose one of the proposed tasks (molecular data / imaging data)

Evaluation will be based on:

- clever choice of methodology
- clever use of data

Demonstrating preliminary success will not be a requirement to pass the course but used to set up a leader board. You can expect that your approach should be able to detect neuronal loss in Alzheimer's disease in the cortex.