





HANDLING BIG DATA

When Physics meets Genomics!



Rita Coimbra Café com Física, UC, 13 February 2019

IBiGen – Genomics Facility







Meet the Team

- •1 Principal Investigator
- •1 Senior Computer Engineer
- •1 Bioinformatician
- •2 Wet-lab Technicians
- •1 Biostatistics Technician
- •2 Post-doc Fellowships
- 2 PhD Students





When Physics Meets Biology



Professor E. Schrödinger 26 Kincora Road, Clontarf, Dublin, Ireland.

Dear Professor Schödinger,

"The notion that life might be perpetuated by means of an instruction book inscribed in a secret code appealed to me" (Watson)



The Physical Aspect of the Living Cell

BY ERWIN SCHRÖDINGER ENIOR PROFESSOR AT THE DUBLIN INSTITUTE FOR ADVANCED STUDIES

Watson and I were once discussing how we came to enter the field of molecular biology, and we discovered that we had both been influenced by your little book, "What is Life?".

Life is a four "letters" word



How DNA is packaged into a cell?



When it comes to genome, smaller is sometimes better

Species	H.sapiens	M. musculus	D. Melanogaster	C. elegans	S.cerevisae
Genome size	3.2 Gbp	2.8 Gbp	140 Mbp	100 Mbp	12 Mbp
Protein coding genes	21,000	20,000	14,000	20,000	6,600
Choromosomes	46 (2n)	40 (2n)	8 (2n)	12 (2n)	16
Common name					

"Junk DNA" - Not so useless after all



Figure 5-10 What Is Life? A Guide To Biology © 2010 W.H. Freeman and Company

- About 3% of DNA codes for proteins
- About 40-50% is repetitive
- What is the function of the remaining 50%?



The Big Question: what makes us unique?





The Central Dogma of Molecular Biology



The genetic code

	Second Nucleotide Position						
	U	С	Α	G			
ition N	UUU Phenylalanine	UCU Serine	UAU Tyrosine	UGU Cysteine			
	UUC Phenylalanine	UCC Serine	UAC Tyrosine	UGC Cysteine			
	UUA Leucine	UCA Serine	UAA STOP	UGA STOP			
	UUG Leucine	UCG Serine	UAG STOP	UGG Tryptophan			
de Pos	CUU Leucine	CCU Proline	CAU Histidine	CGU Arginine			
	CUC Leucine	CCC Proline	CAC Histidine	CGC Arginine			
J Gi	CUA Leucine	CCA Proline	CAA Glutamine	CGA Arginine			
clee	CUG Leucine	CCG Proline	CAG Glutamine	CGG Arginine			
First Nu	AUU Isoleucine	ACU Threonine	AAU Asparagine	AGU Serine			
	AUC Isoleucine	ACC Threonine	AAC Asparagine	AGC Serine			
	AUA Isoleucine	ACA Threonine	AAA Lysine	AGA Arginine			
	AUG Methionine	ACG Threonine	AAG Lysine	AGG Arginine			
G	GUU Valine	GCU Alanine	GAU Aspartate	GGU Glycine			
	GUC Valine	GCC Alanine	GAC Aspartate	GGC Glycine			
	GUA Valine	GCA Alanine	GAA Glutamate	GGA Glycine			
	GUG Valine	GCG Alanine	GAG Glutamate	GGG Glycine			

- With 4 bases (A,C,T,G) there are 64 combinations for a 3letter code
- We only need 20 code words, one for each aminoacid

First draft of the Human Genome was just the beginning



"It's a history book - a narrative of the journey of our species through time. It's a shop manual, with an incredibly detailed blueprint for building every human cell. And it's a transformative textbook of medicine, with insights that will give health care providers immense new powers to treat, prevent and cure disease." (Francis Collins, director of NHGRI, 2001, about)

How much does it cost to sequence the genome?



Cost: ~ \$3 billion Time: 13 years



Cost: \$1000 Time: 1 day



The Future



The Omics cascade



Measuring Gene Expression with Microarrays



Gene Expression profiling with RNA-seq







Table of counts for all genes in Genome

Transcriptome Analysis of the Minho Ageing Cohort

Immunosenescence and cognitive decline



Patient characterization (114):

- → Age (> 55 years), sex, school years;
- → Clinical data;
- Cognitive profile (assessed by neurocognitive tests and mood).





MEDPERSYST: Synaptic networks and Personalized Medicine Approaches to understand Neurobehavioural Diseases Acrross the Lifespan

Characterization of the Minho Ageing Cohort



	Ind #1	Ind #2	Ind #3
Gene 1	18.9297	18.3781	18.5562
Gene 2	18.8174	18.8174	18.8174
Gene 3	17.9671	18.6491	18.7507

Clustering - unsupervised learning





Estimation of immune cells content by in silico approach



Differential Gene Expression Analysis





Gene Ontology (GO)





Ontology BP

0

MF

https://xuegen564s18.weebly.com/gene-ontology.html

Pathway Analysis





Data Integration of the Minho Ageing Cohort



Increasing Resolution

N ~ 20,000 transcripts



Gene expression profilling is influenced by DNA variants?



Hongyu Zhao⁸

Acknowledgements









Biomedical Informatics and Technologies Group







UNIÃO EUROPEIA

Thank you for your attention !!!



Why Gene Expression

Gene expression profiling



Differential gene expression



Biological function: (GO enrichment and Pathway Analysis)



Find case/control discriminative signatures

Develop diagnostic tests. Identify new drug targets

Advantages of RNA-seq

RNA-seq

Microarrays

- High cost (but continually reducing); high data storage;
- High dinamic range (no saturation);
- High sensitivity, low technical variation high tecnical reproducibility;
- Not limited by a prior knowledge the organism's genome
- Identify novel transcripts, splice junctions,
 SNPs and non-coding RNAs.

- Relatively low-cost; ability to process a large # of samples;
- * Dinamic range limited by scanner;
- Low sensitivity and high background noise;
- Reliance upon existing knowledge about the genome sequence;
- **×** Limited to the probes on the chip.

Transcriptome Analysis of the Minho Ageing Cohort

Effector memory CD4⁺ T cells are associated with cognitive performance in a senior population

ABSTRACT

Cláudia Serre-Miranda, MSc* Susana Roque, PhD* Nadine Correia Santos, PhD Carles Portugal Nunga

Objective: Immunosenescence and cognitive decline are common markers of the aging process. Taking into consideration the heterogeneity observed in aging processes and the recently described link between lymphocytes and cognition, we herein explored the possibility of an association between alterations in lymphocytic populations and cognitive performance.



Biological question:

Is there an association between alterations in peripheral blood mononuclear cells (PBMCs) and cognitive performance in healthy aging Individuals ?

