Bioinformatics
Core Facility

UNIVERSITY OF GOTHENBURG

Identifying a disease causing mutation

Marcela Davila

2015-03-05

The Sahlgrenska Academy

Molecular Genetics & Genomic Medicine

Open Access

ORIGINAL ARTICLE

Whole exome sequencing reveals mutations in *NARS2* and *PARS2*, encoding the mitochondrial asparaginyl-tRNA synthetase and prolyl-tRNA synthetase, in patients with Alpers syndrome

The study

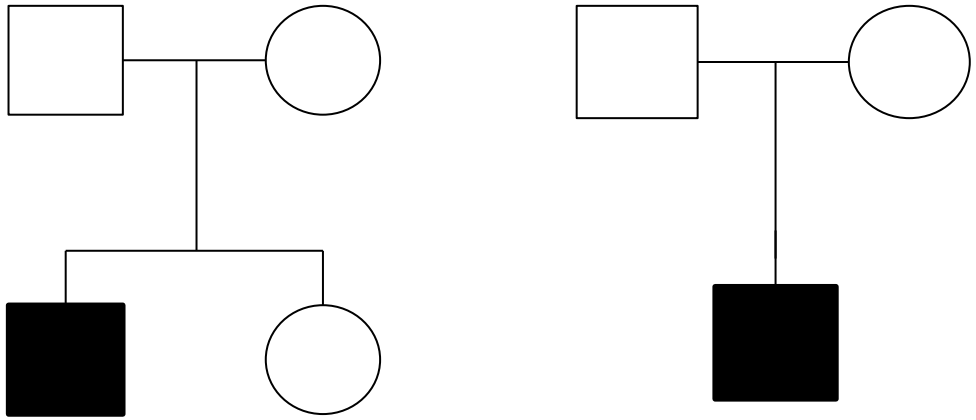
Alpers syndrome: progressive neurodegenerative dissorder

POLG1 – Alpers Huttenlocher

FARS2 – encoding enzyme to charge mt tRNA with Phe

19 patients: 6 had POLG1 mutations

For this study:



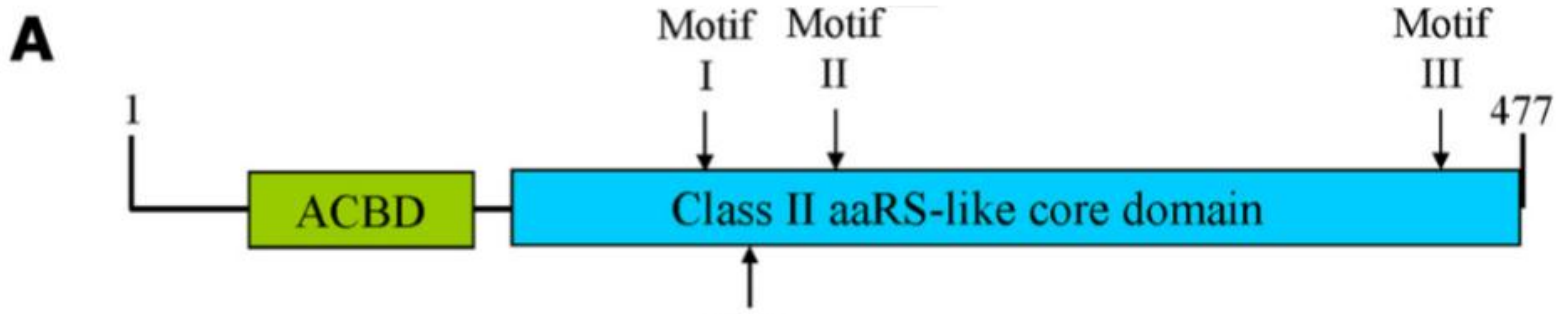
The study

Exome sequencing

	Patient I		Patient II	
	Variants	Genes	Variants	Genes
Total	124,631	15,978	129,098	16,015
Genes encoding mitochondrial protein	1698	671	1882	681
Allele frequency <3%	98	94	100	86
Predicted deleterious	32	27	18	18
Recessive pattern of inheritance	1	1	2	1

Mutations in *PARS2* (Pro) and *NARS2* (Asn)

NARS2

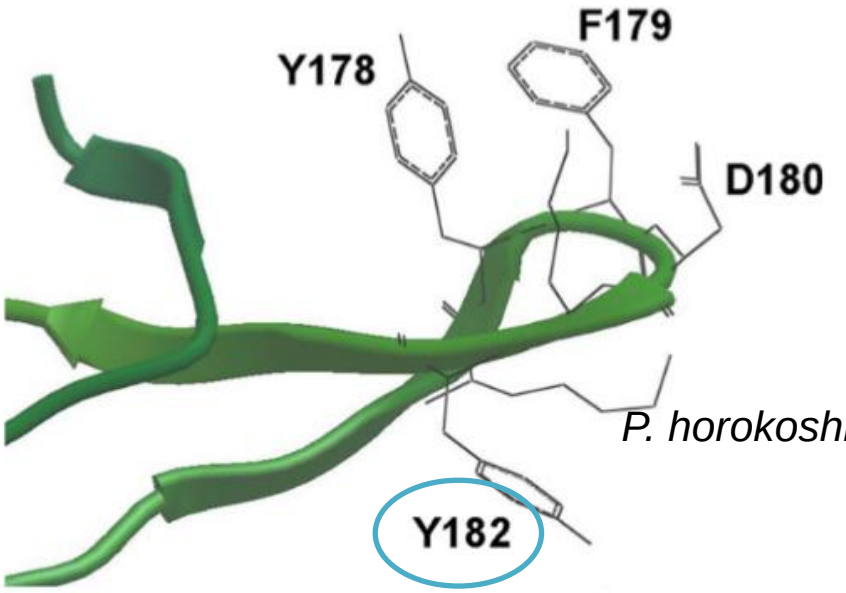


B

Mt h.sapiens	NFFNVPAFLT	218
Mt p.alecto	NFFNVPAFLT	218
Mt r.norvegicus	NFFDVPAFLT	218
Mt c.livia	HFFNVPAFLT	173
Mt d.rerio	HFFSVPAYLT	226
c.elegans	DYYGEPAYLT	293
s.cerevisiae	SYFGKPTYLT	214
p.horikoshii	KYFDKYAYLS	186

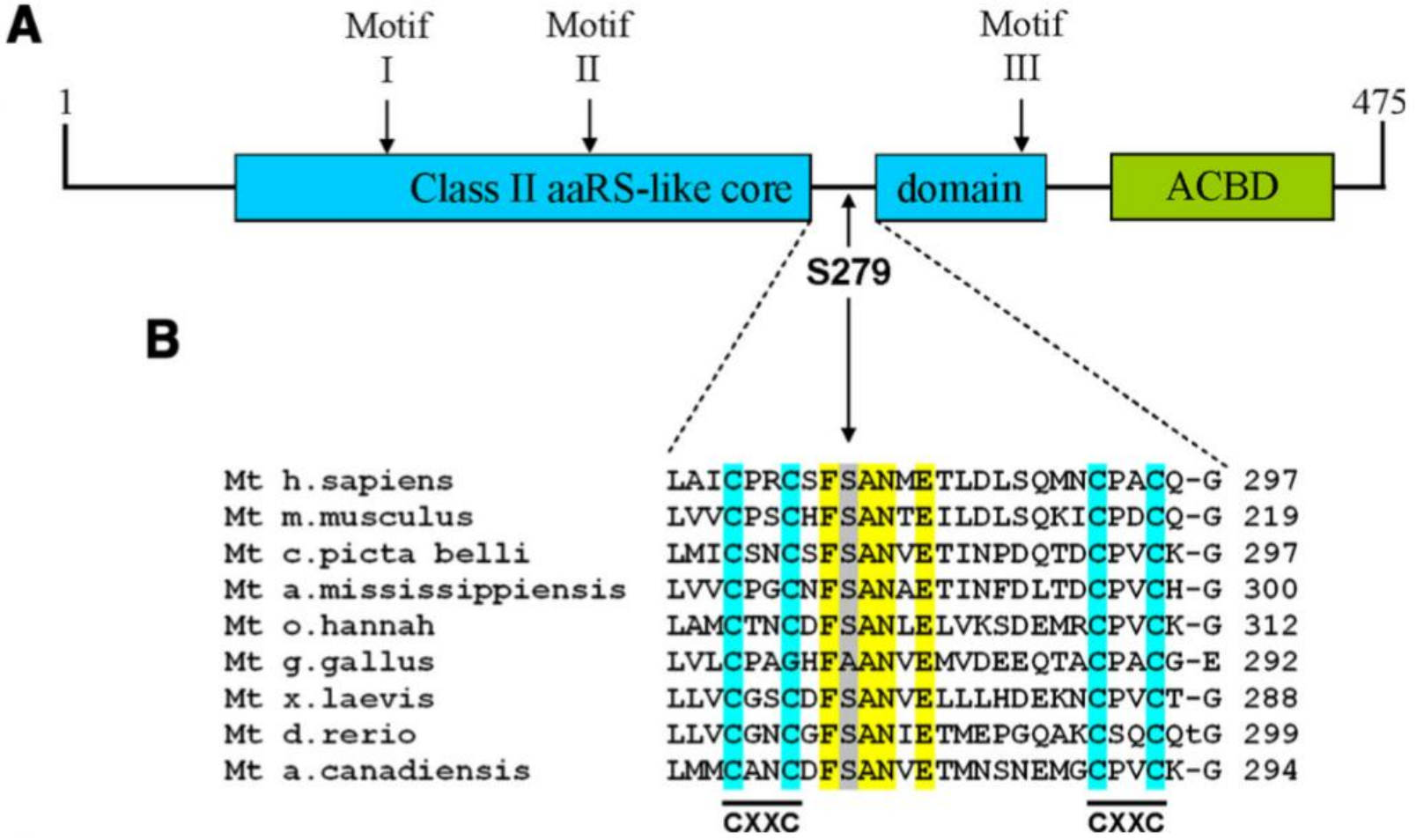
monomer-monomer interaction

C

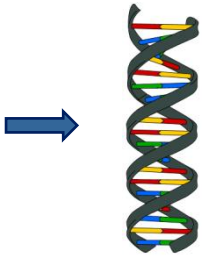


P. horokoshii

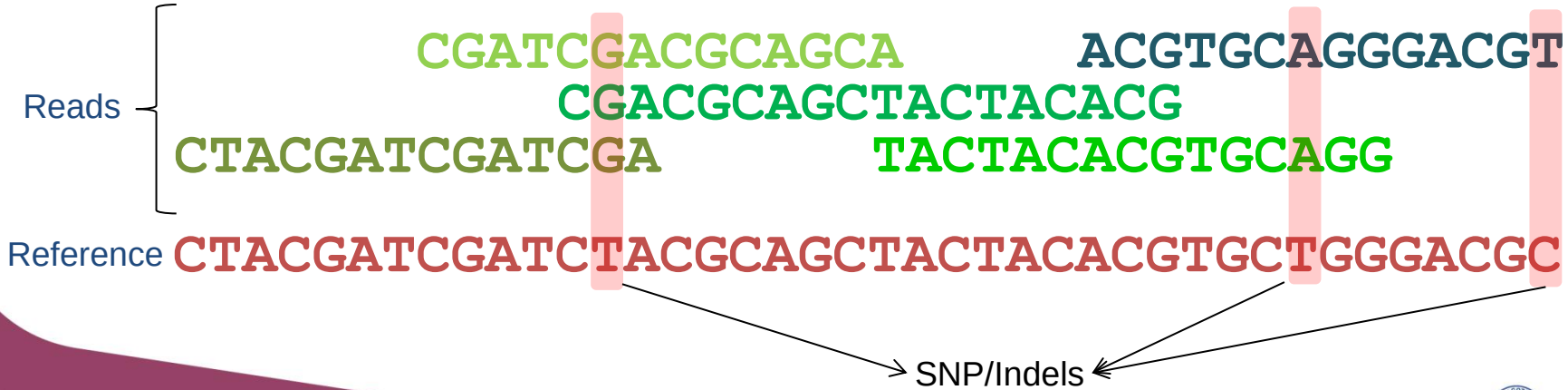
PARS2


A
Motif I
Motif II
Motif III
1
475
Class II aaRS-like core
domain
ACBD
S279
B
Mt h.sapiens LAICPRCSFSANMETLDLSQMNCPACQ-G 297
Mt m.musculus LVVCPSCHFSANTEILDLSQKICPDCQ-G 219
Mt c.picta belli LMICSNCSFSANVETINPDQTDCPVCK-G 297
Mt a.mississippiensis LVVCPGCNFSANAETINFDLTDCPVCH-G 300
Mt o.hannah LAMCTNCDFSANLELVKSDEMRCPVCK-G 312
Mt g.gallus LVLCPAGHFAANVEMVDEEQTACPACG-E 292
Mt x.laevis LLVCGSCDFSANVELLLHDEKNCPVCT-G 288
Mt d.rerio LLVCGNCGFSANIETMEPGQAKCSQCQtG 299
Mt a.canadiensis LMMCANCDFSANVETMNSNEMGCPVCK-G 294
CXXC
CXXC

Identification of SNVs

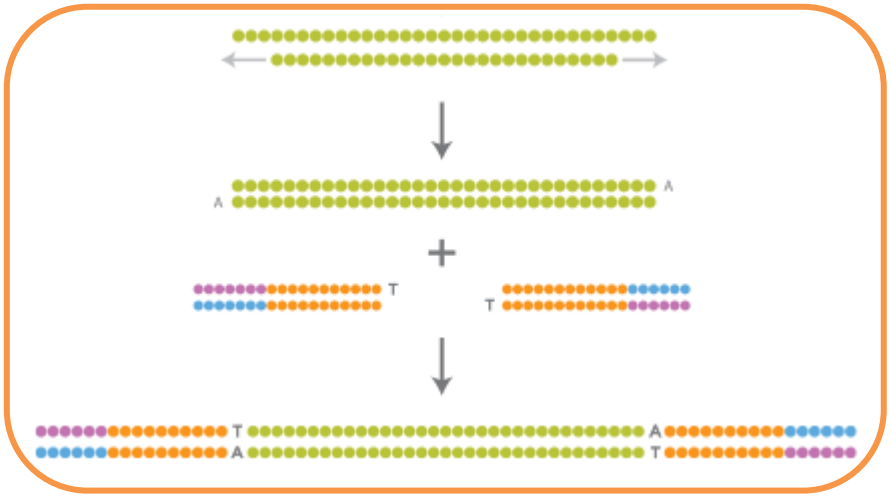


Sample preparation
sequencing

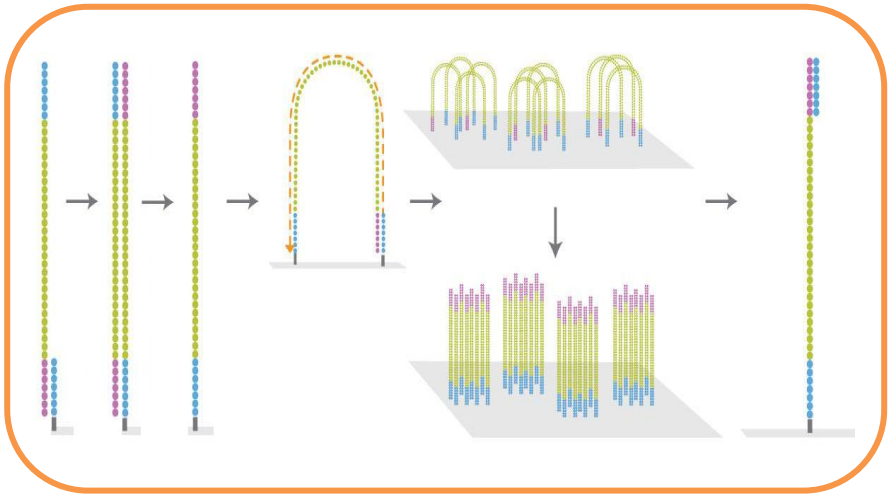


Sequencing: at the Lab

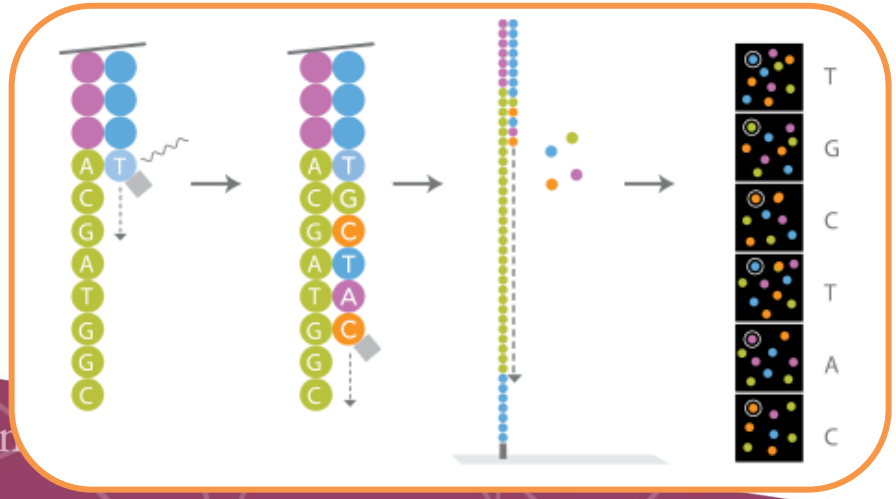
Library prep



Cluster generation



Sequencing, imaging and data generation

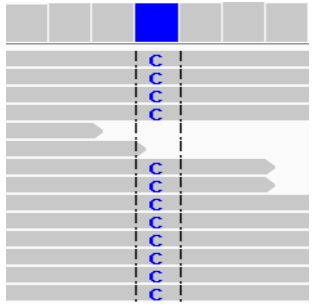


Sequencing: some applications

Quality Check
Quality Filter

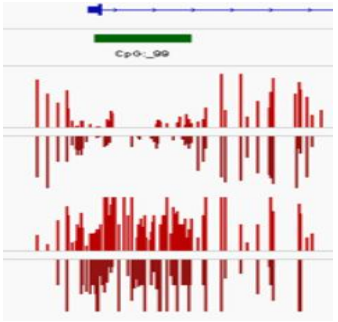
Mapping to ref genome
De novo assembly

Exome-seq



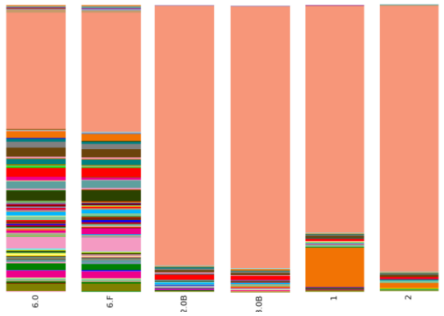
SNV detection

Methyl-seq



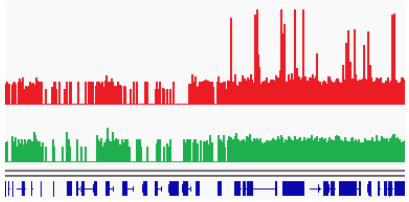
CpG patterns

Metagenomics



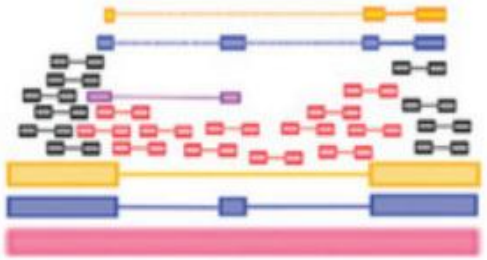
Taxa summary

ChIP-Seq



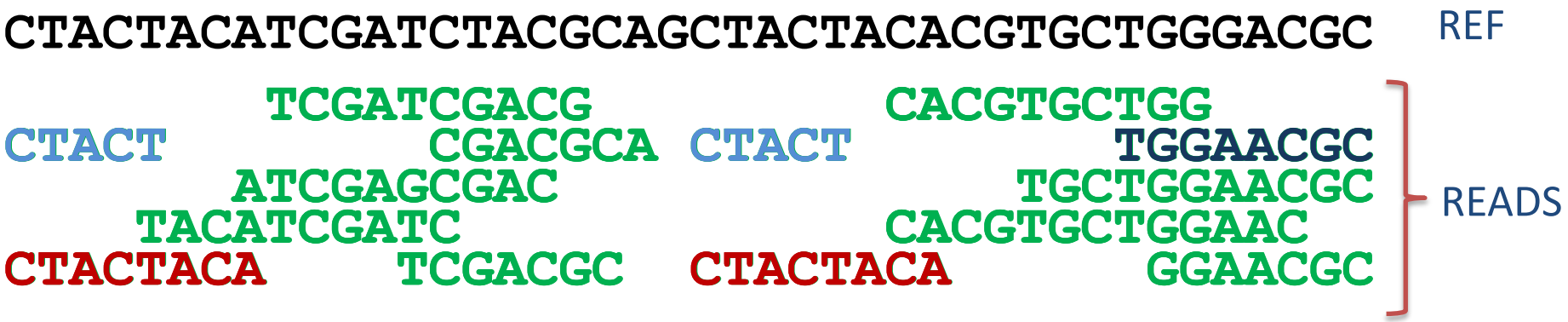
Peak detection

RNA-seq



Transcript estimation

Mapping to a reference genome



WHERE to place the reads?

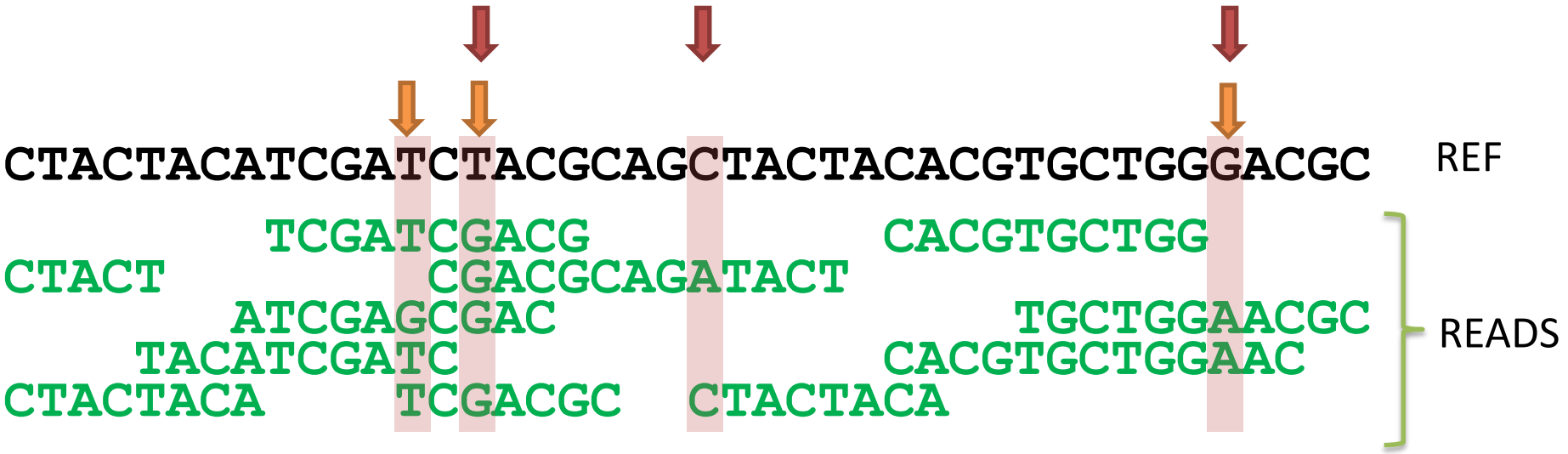
a) Unique reads
b) Everywhere possible
c) Choose one randomly
d) Use pair-end data

HOW to place the reads?

a) Ungapped
b) Gapped

Bfast, BioScope, **Bowtie**, **BWA,** CLC bio, CloudBurst, Eland/Eland2, GenomeMapper, GnuMap, Karma, **MAQ**, MOM, **Mosaik**, MrFAST/MrsFAST, NovoAlign, PASS, PerM, RazerS, RMAP, SSAHA2, Segemehl, …

Variant calling

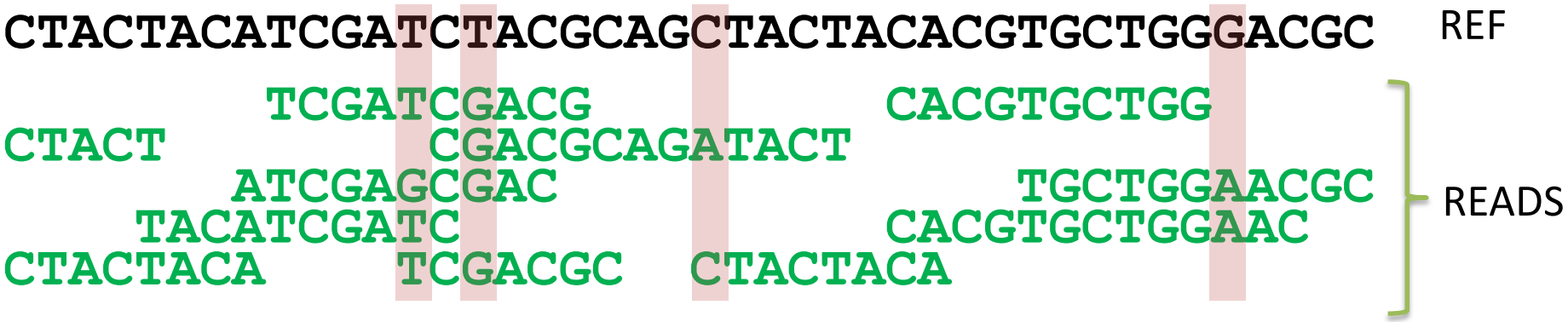


Is it a variant allele?

What is the most likely genotype?

SOAP2, **samtools**, **GATK**, Beagle, CRISP, Dindel, FreeBayes, SeqEM, VarScan, Mutect

Variant annotation



In which gene is it located?
Name, Description,
OMIM, Pathway, GO,
Expression profiles . . .

Where in the gene is it located?
Intron, exon, UTR,
intergenic region, splice site

Is there any AA change?

```
GAA -> GAG = E->E
GTT -> CTT = V->L
TGG -> TGA = W->X
TGA -> CGA = X->R
```

Is it a known SNP?

What impact does the AA change have?
Damaging, benign

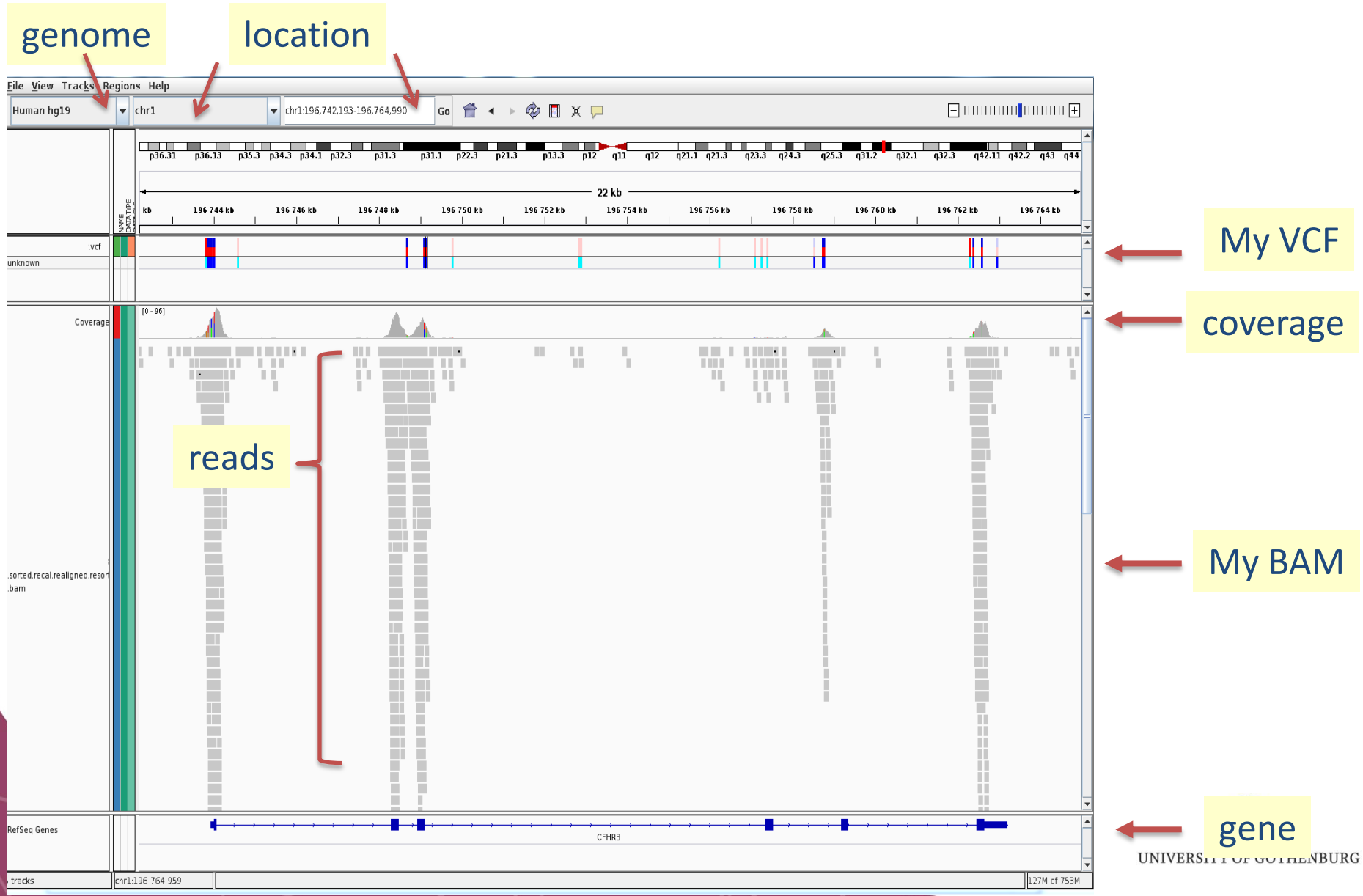
Annovar,
SIFT, PP2,
dbSNP,
GO,
KEGG,
OMIM
1000G

Variant list

CHR	POS	REF	OBS	ALLELE	GENE	DESCRIPTION	VARIANT_FUNCTION	EXONIC_FUNCTION
chr1	780785	T	A	homozygous	LOC643837	-	ncRNA_intronic	-
chr1	802496	C	T	heterozygous	FAM41C	-	downstream	-
chr1	887801	A	G	homozygous	NOC2L	Nucleolar complex protein 2 homolog	exonic	Synonymous
chr1	1265154	T	C	homozygous	GLTPD1	Glycolipid transfer protein domain-containing protein 1	downstream	-
chr1	151733327	T	C	heterozygous	MRPL9	39S ribosomal protein L9, mitochondrial	ncRNA_exonic	nonsynonymous
chr1	151733335	T	G	homozygous	MRPL9	39S ribosomal protein L9, mitochondrial	ncRNA_exonic	nonsynonymous
chr1	52306064	TCT	-	heterozygous	NRD1	Nardilysin	ncRNA_exonic	frameshift deletion
chr1	54605319	G	GC	homozygous	CDCP2	CUB domain-containing protein 2	exonic	frameshift substitution
chr3	189507518	C	CAGA	homozygous	TP63	Tumor protein 63	UTR5	-

AA_CHANGE_POS	AA_CHANGE	dbSNP	BUILD	SIFT	PP2	LRT	OMIM	CONSERVED
-		rs2977612	101					
-		rs10157494	119					conserved
NOC2L:uc001abz.3:exon10:c.T1182C:p.T394T	T => T	rs3828047	107					
-		rs307355	79					conserved
MRPL9:uc001eyv.2:exon6:c.A637G:p.I213V,	I => V	rs74228558	130	tolerated	bening	deleterius	611824	conserved
MRPL9:uc001eyv.2:exon6:c.A629C:p.E210A	K => Q	rs8480	52	damaging	damaging	neutral	611824	
NRD1:uc010ong.1:exon2:c.208_0del:p.70_0del,		rs145326984	134					
CDCP2:uc001cwv.1:exon4:c.1224_1224delinsGC,		rs66537746	130					
-		rs34201045	126					conserved

Data visualization: IGV



Variant Filtering

Sample → Seq →SNPs →

. . . → Filtering → candidate genes

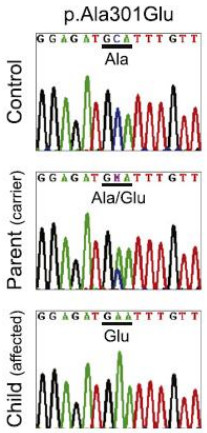
Control → Seq →SNPs →

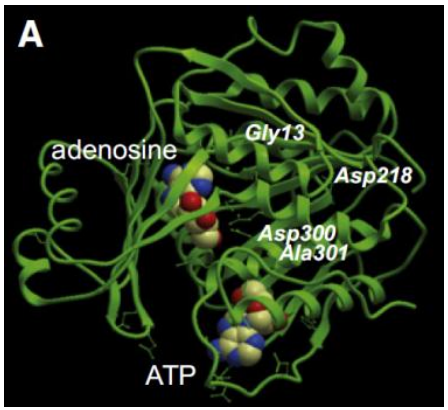
Making sense of the data

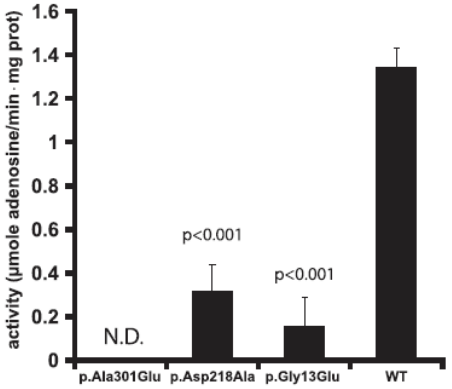
Exome sequencing cases → Coding variants → Disease model / Disease knowledge → **Candidate genes**

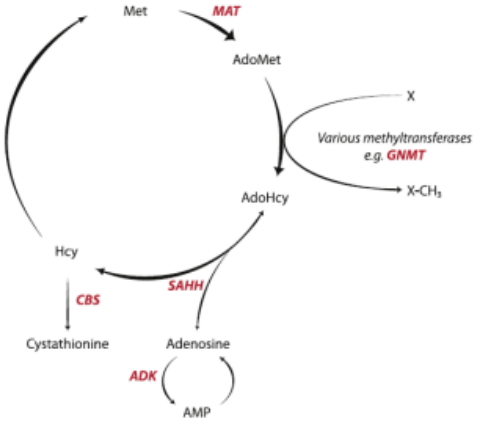
Familiy filters, Controls
Genetic variation DBs

Your real work begins…









Bioinformatics Core Facility

Contact information

Visiting address:
Medicinaregatan 3B, F1000

bioinformatics@gu.se

www.cf.gu.se/english/Bioinformatics/

Our main goals:

- Set up an interdisciplinary and collaborative environment
- Increase the understanding of statistical and bioinformatics and analysis
- Contribute to the development of a wide range of research projects

Bioinformatics Networks

Our activities

GOTBIN
GOTHENBURG BIOINFORMATICS NETWORK

Seminars

March 11th at 11:00
Carbohydrates in Bioinformatics
Miguel Rojas, Dept. of Biochemistry and Cell Biology
Place: F Andreasson, Medicinaregatan 11

April 16th at 13:00-14:30
Can I trust my network? Assessing network estimation uncertainty using local component resolution
José Sanchez, statistician at Bioinformatics Core Facility, GU
Place: Skagerak room, 3rd floor at Registercentrum, Medicinaregatan 18G

Confirmed speakers - 2015

Agatha Smialowska, BILS expert at Chalmers University of Technology

Katarina Truvé, BILS expert at Bioinformatics Core Facility, GU