Bioinformatics
Core Facility

UNIVERSITY OF GOTHENBURG

Identifying a disease causing mutation

Marcela Davila

2014-02-19

The Sahlgrenska Academy

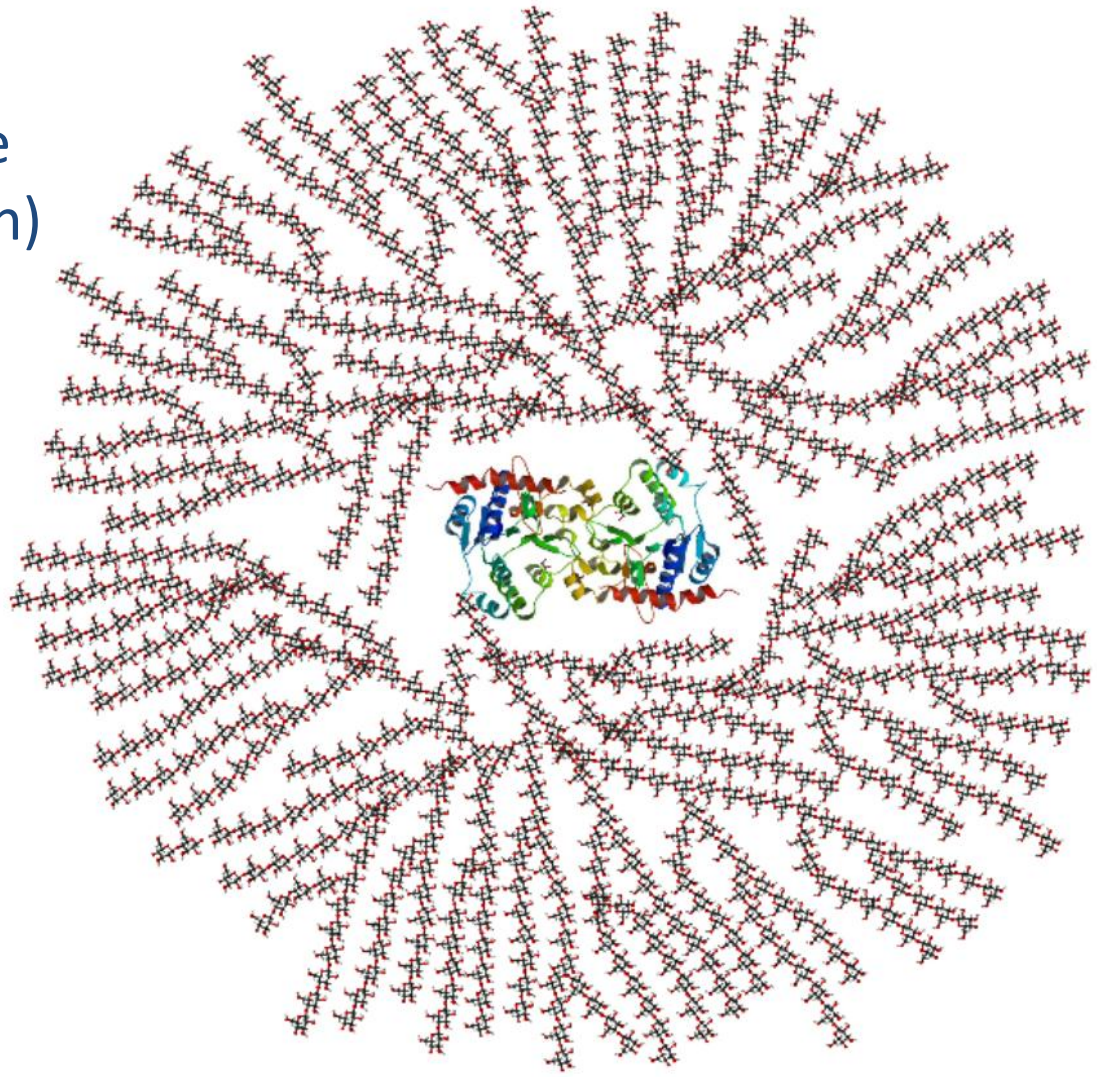
Polyglucosan Body Myopathy Caused by Defective Ubiquitin Ligase RBCK1

American *Neurological* Association and the Child *Neurology* Society

Glycogen

Polysaccharide of glucose
+ glycogenin (core protein)
Energy storage

Metabolism:
- Diabetes
- Hypoglycemia
- GSD
  (glycogen storage diseases)

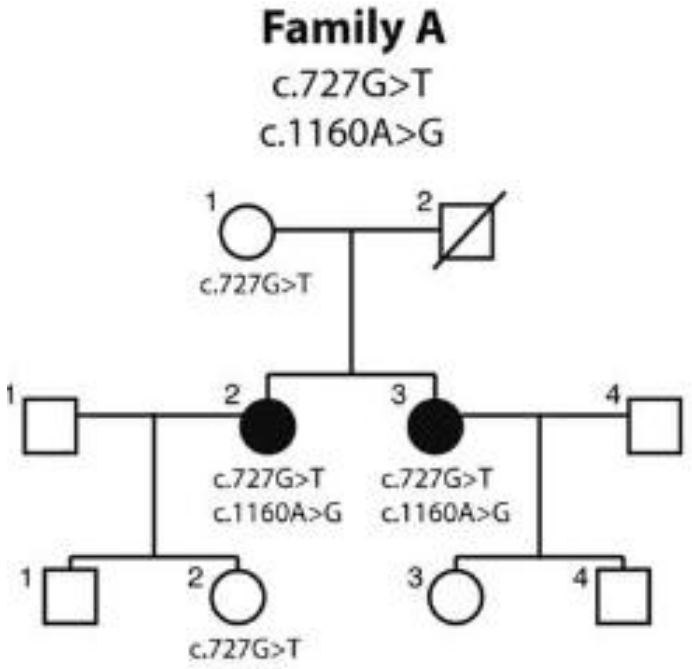


The study

Polyglucosan accumulation
No gene defects known

Exome sequencing

Mutations in RBCK1
E3 ubiquitin ligase

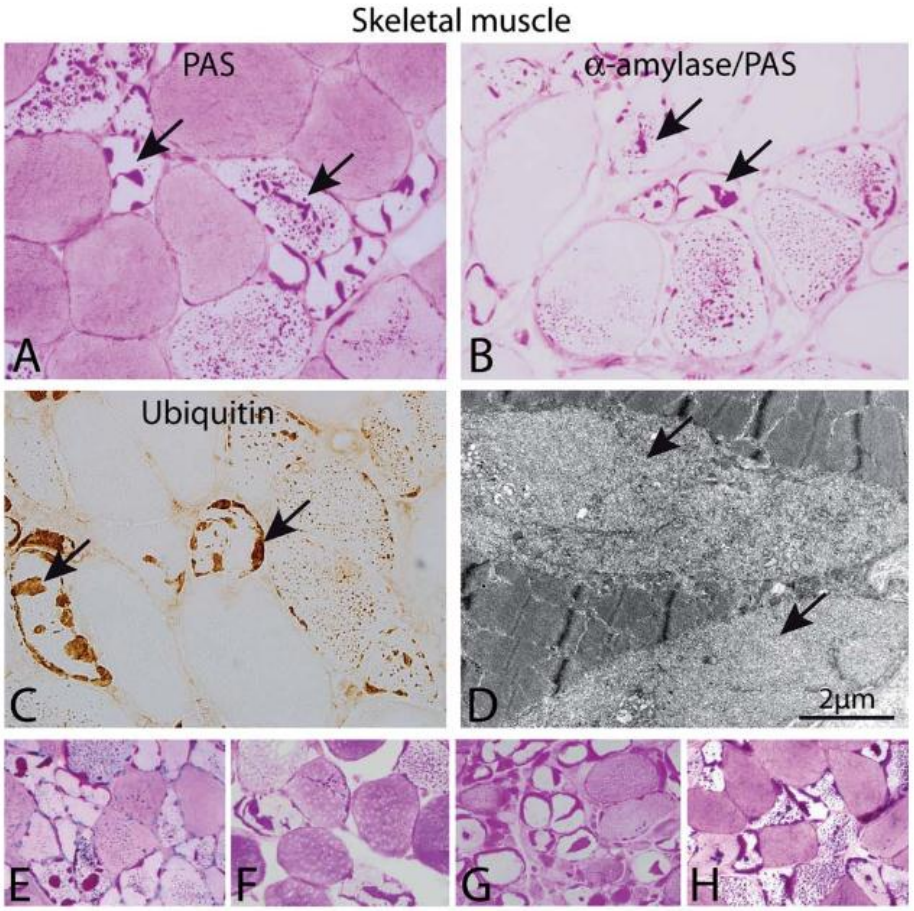


Screening in additional patients

Morphologic Analysis: muscle

Lack the normal intermyofibrillar glycogen but show accumulation of PAS-positive material

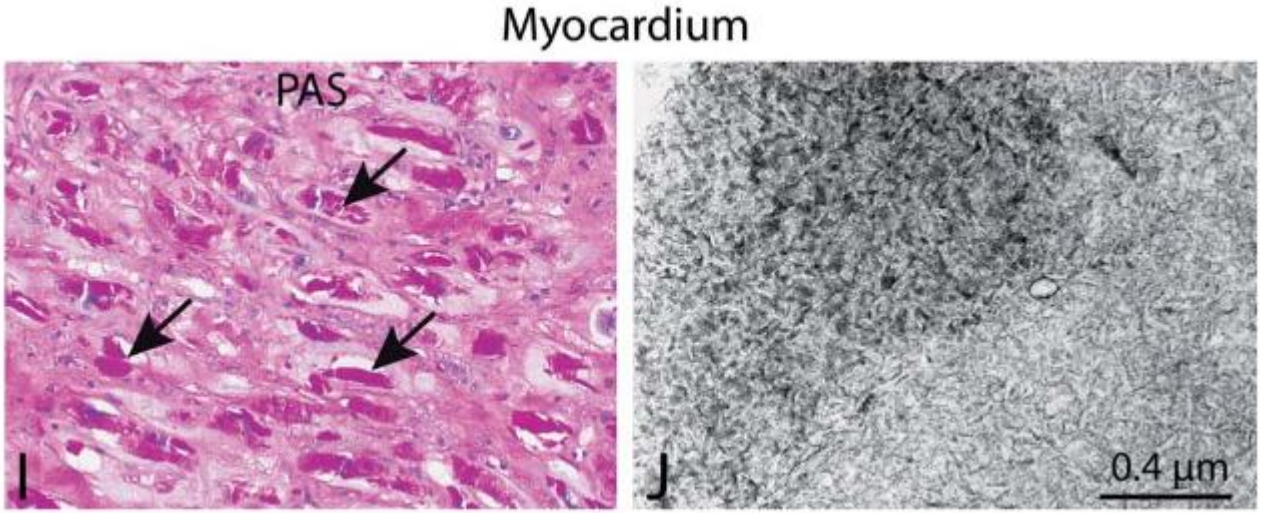
Storage material is not removed by amylase treatment



Storage material is not removed by amylase treatment

Storage material is filamentous

Morphologic Analysis: heart



Clinical findings

TABLE. Clinical Findings

Patient	AII:2	AII:3
Gender	F	F
Age, yr	50	47
Age at onset, yr	12	16
Initial symptoms	Leg weakness	Leg weakness
Walking ability	Wheelchair	Walks with aid
Facial muscle weakness/ptosis	No/no	No/no
Serum CK	×2	Normal
Scoliosis/ contractures	No	No
Pulmonary vital capacity, l [%]	1.8 [56]	2.6 [76]
Cardiomyopathy	DCM	No
Ejection fraction	35–40%	ND
Heart transplantation	No	No

[a]See also Schoser et al.[6]
[b]See also de la Blanchardière et al.[23]
CK = creatine kinase; DCM = dilated cardiomyopath

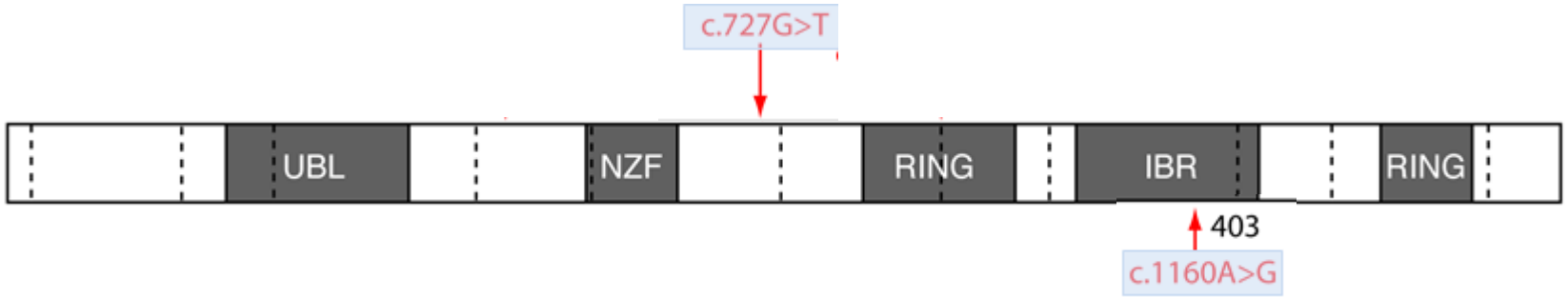
Molecular genetics

Exome sequencing (Family A)

missense mutation p.Asn387Ser – allele1

stop gain p.Glu243 – allele2

RBCK1



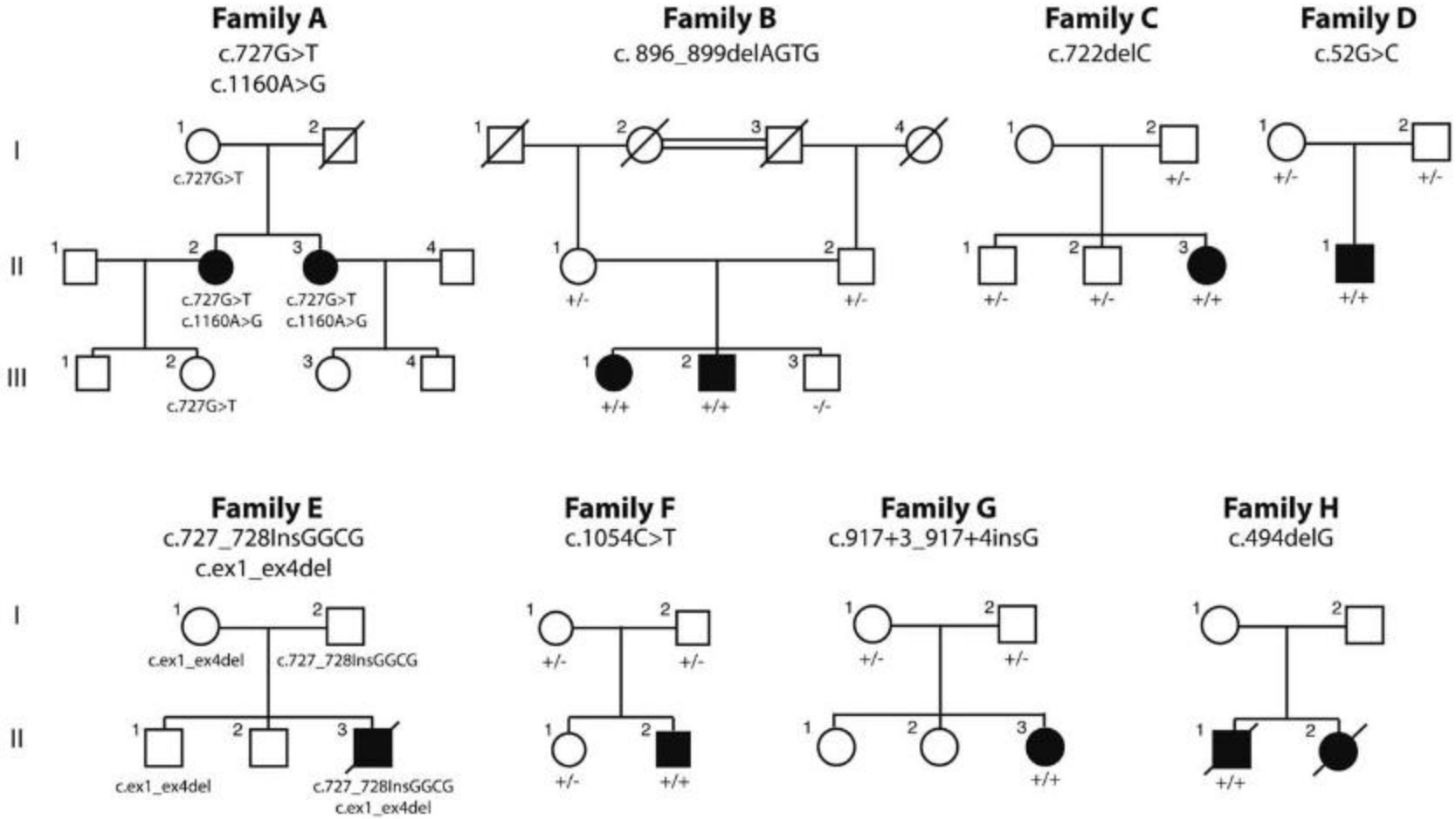
Additional cases

TABLE. Clinical Findings

Patient	AII:2	AII:3	BIII:1[a]	BIII:2[a]	CII:3	DII:1	EII:3	FII:2	GII:3	HII:1[b]
Gender	F	F	F	M	F	M	M	M	F	M
Age, yr	50	47	24	19	29	19	20 (expired)	26	32	15 (expired)
Age at onset, yr	12	16	6	5	Childhood	Childhood	4	12	17	9
Initial symptoms	Leg weakness	Leg weakness	Leg weakness	Difficulty running	Difficulty running	ND	Difficulty running	Leg weakness	Leg weakness	Difficulty running
Walking ability	Wheelchair	Walks with aid	Walks without aid	Walks without aid	ND	Walks without aid	Wheelchair	ND	Minimally restricted	Walked without aid
Facial muscle weakness/ptosis	No/no	No/no	No/mild	No/no	Yes/no	No/ND	No/mild	No/mild	Yes/no	No/mild
Serum CK	×2	Normal	×5	×6	Normal	ND	Normal	×1.5	Normal	Normal
Scoliosis/ contractures	No	No	No	No	No	ND	Yes/no	Yes/no	Yes/no	No/yes
Pulmonary vital capacity, l [%]	1.8 [56]	2.6 [76]	1.5 [52]	1.6 [66]	2.6 [71]	ND	2.0 [40]	2.8 [54]	3.64 [87.9]	ND
Cardiomyopathy	DCM	No	DCM	DCM	DCM	DCM	DCM	DCM	No	DCM
Ejection fraction	35–40%	ND	<20%	18%	25%	15%	23%	15%	ND	28%
Heart transplantation	No	No	Age 14 years	Age 13 years	No	No	No	Age 20 years	No	Age 15 years

[a]See also Schoser et al.[6]
[b]See also de la Blanchardière et al.[23]
CK = creatine kinase; DCM = dilated cardiomyopathy; F = female; M = male; ND = not determined.

Additional cases



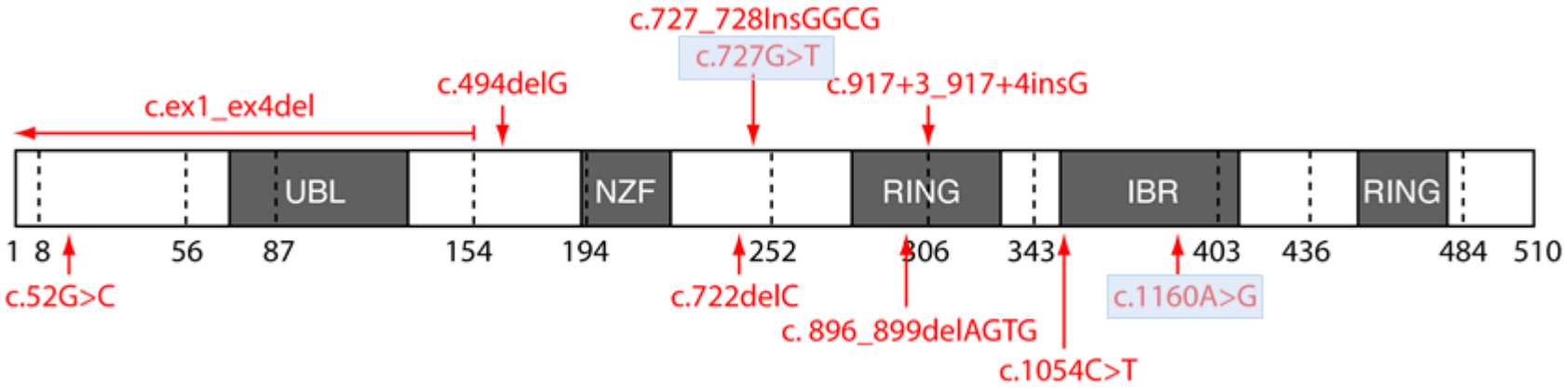
Molecular genetics

Exome sequencing (Family A)

missense mutation p.Asn387Ser – allele1

stop gain p.Glu243 – allele2

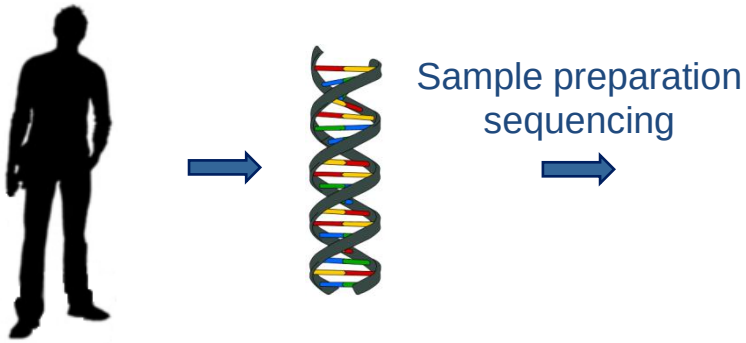
RBCK1



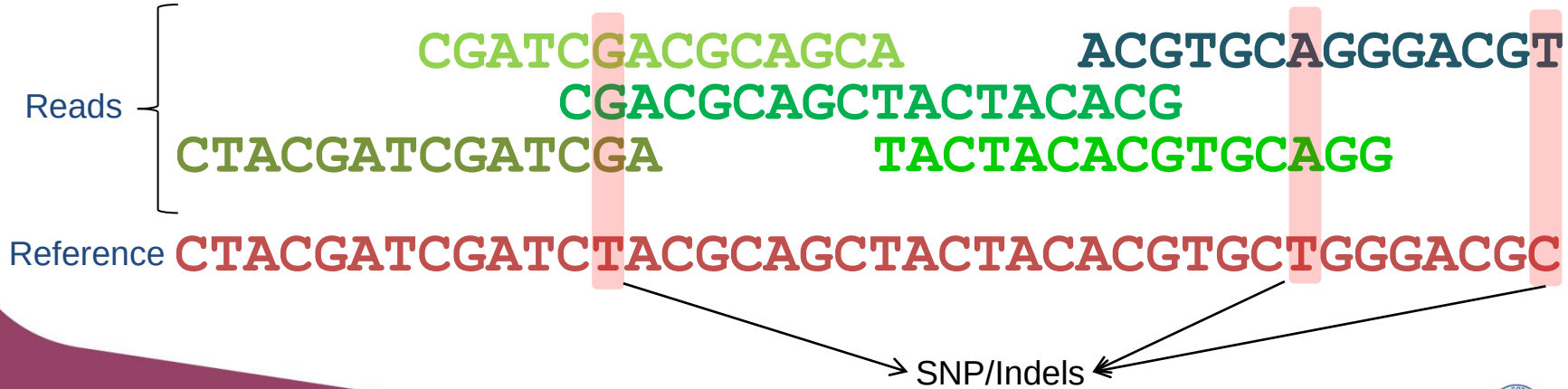
Others: stop gain, frameshifting deletions, or insertions

Homozygous or compound heterozygous

Identification of SNVs

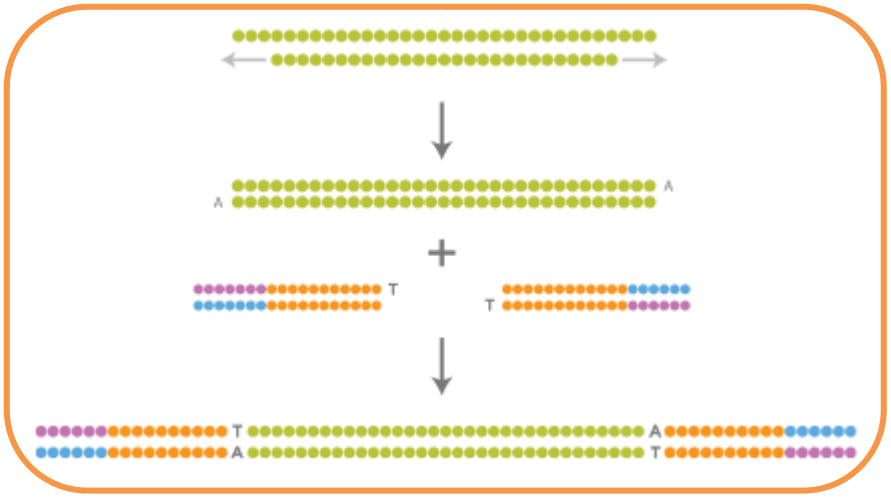


illumina · Roche · Oxford NANOPORE Technologies · Applied Biosystems · Helicos BioSciences Corporation · PACIFIC BIOSCIENCES

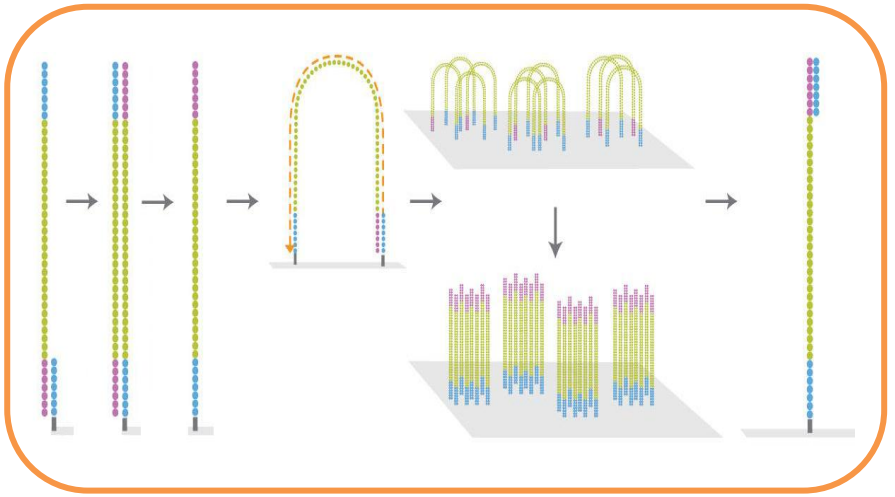


Sequencing: at the Lab

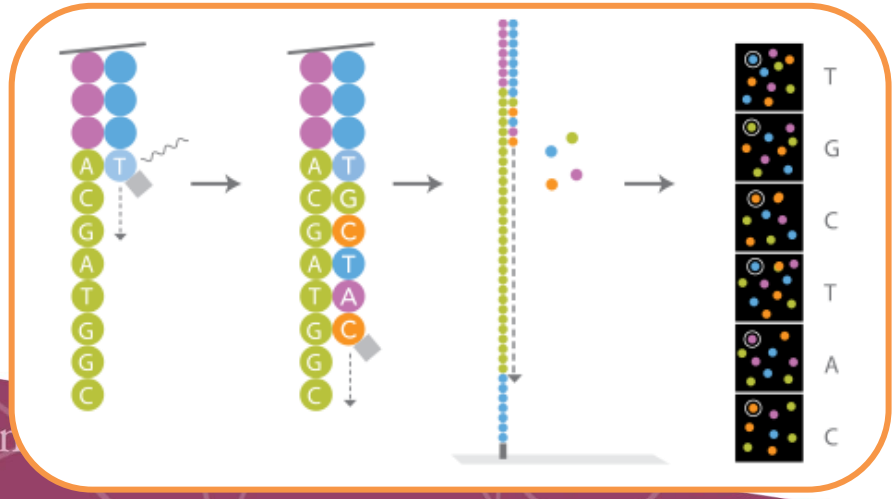
Library prep



Cluster generation



Sequencing, imaging and data generation

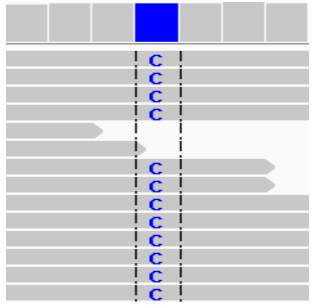


Sequencing: some applications

Quality Check
Quality Filter

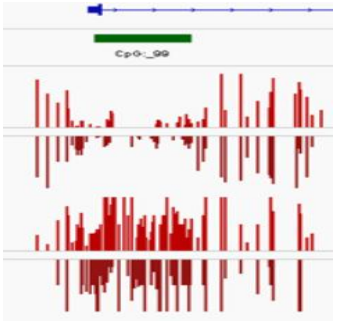
Mapping to ref genome
De novo assembly

Exome-seq



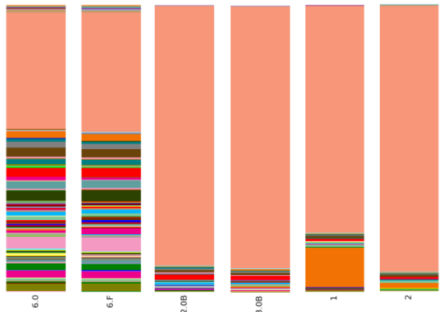
SNV detection

Methyl-seq



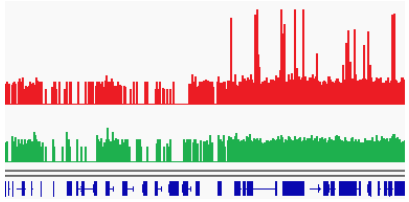
CpG patterns

Metagenomics



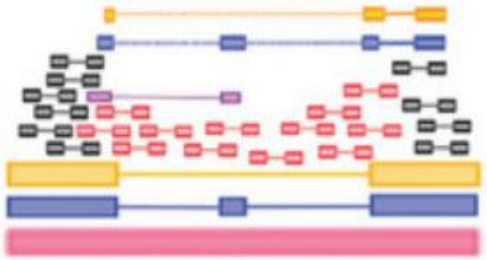
Taxa summary

ChIP-Seq



Peak detection

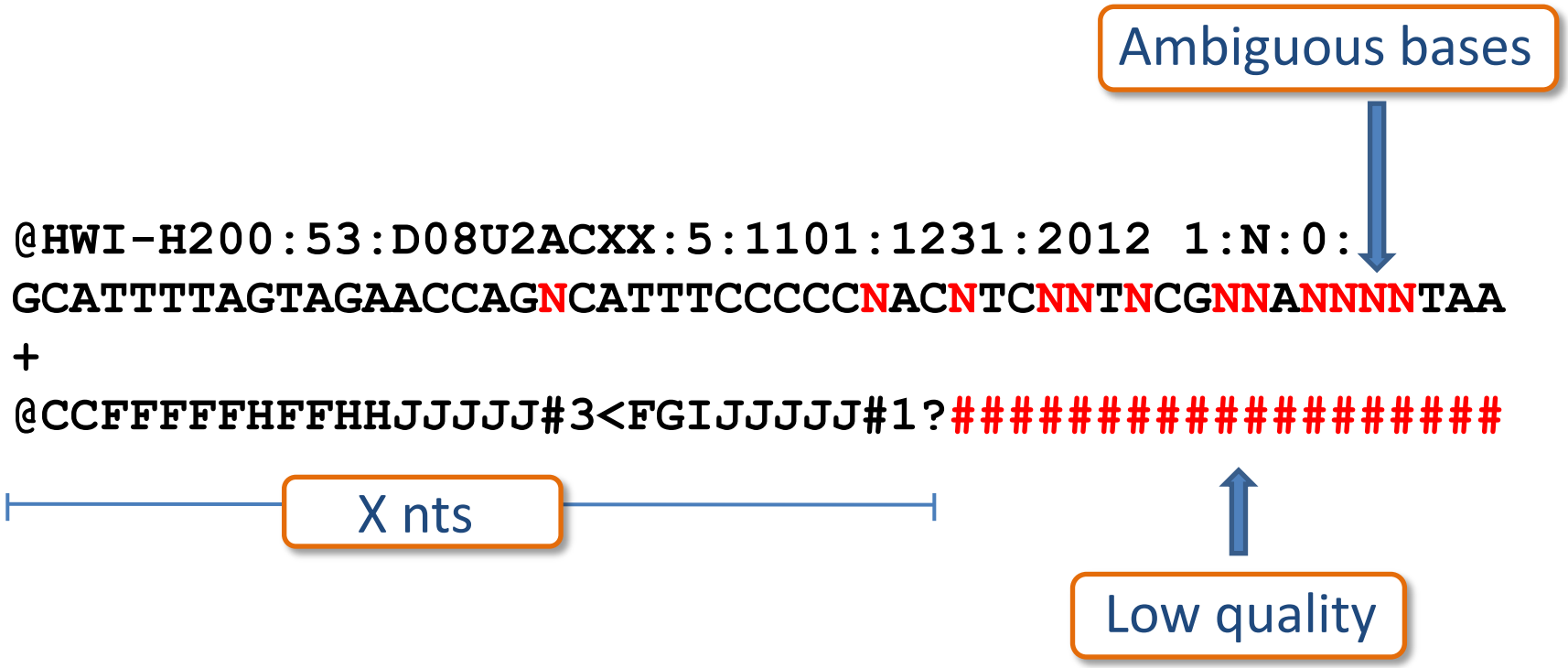
RNA-seq



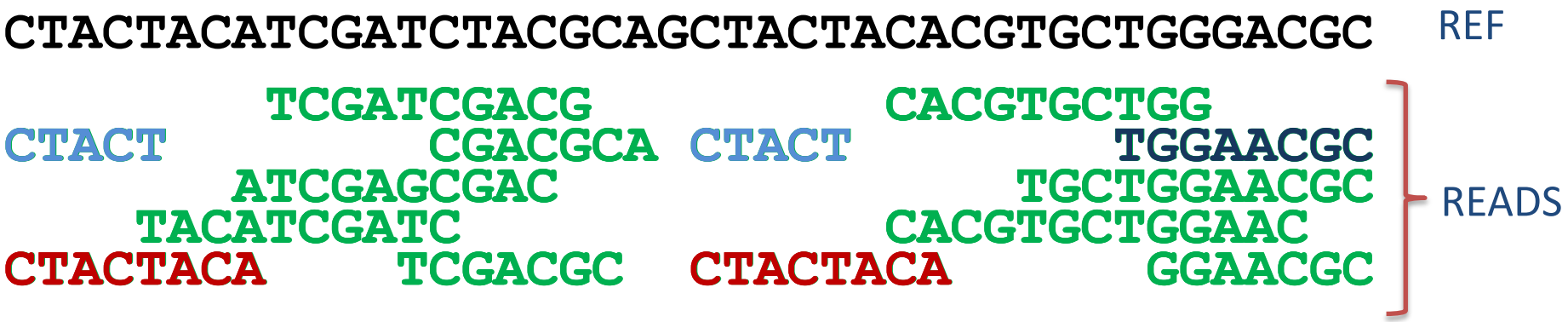
Transcript estimation

Quality filter

A collection of command line tools for Short-Reads FASTA/FASTQ files preprocessing. FastX, PRINSeq



Mapping to a reference genome



WHERE to place the reads?

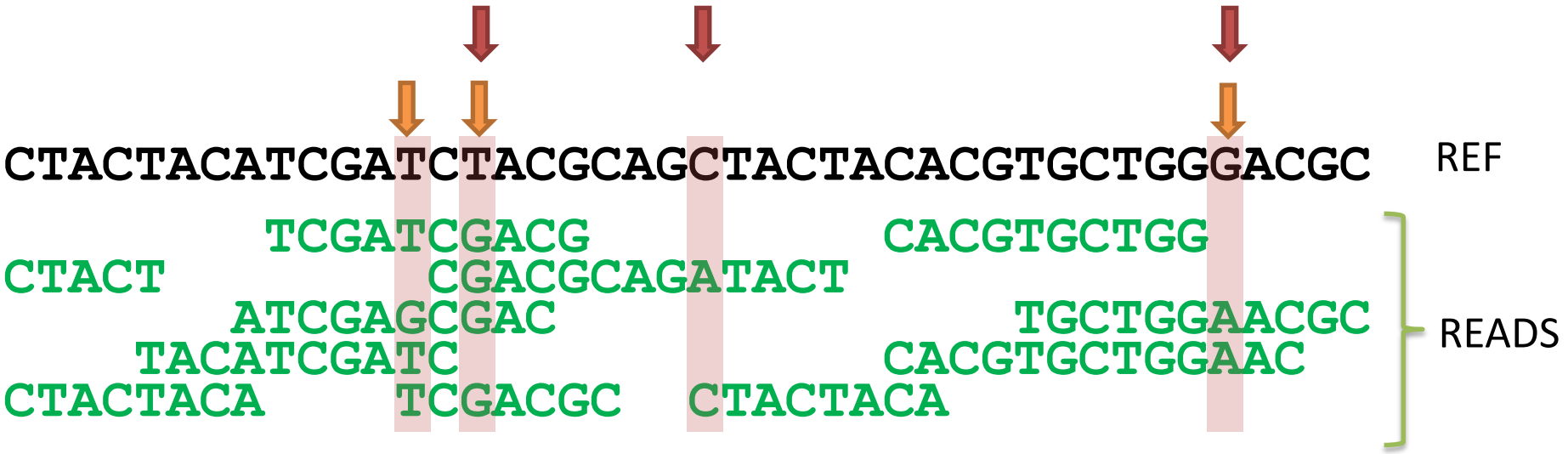
a) Unique reads
b) Everywhere possible
c) Choose one randomly
d) Use pair-end data

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

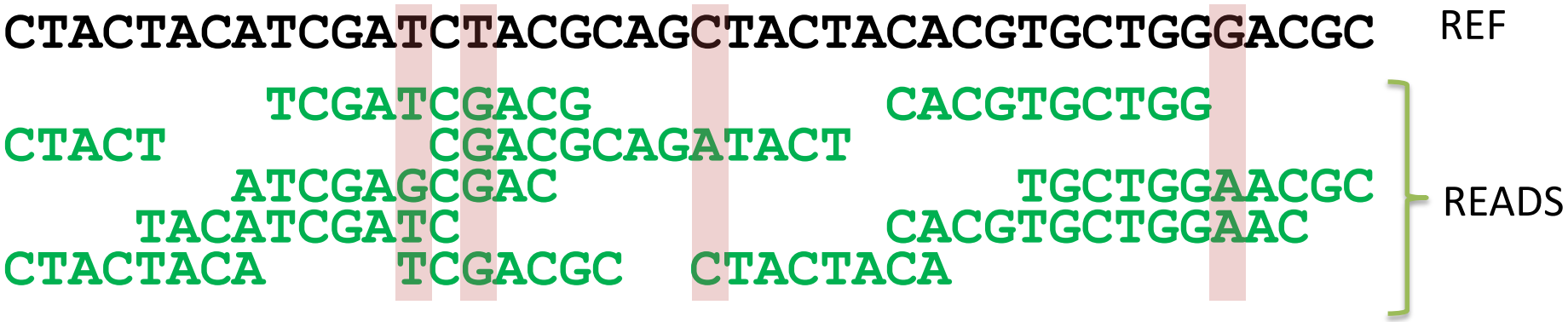


Is it a variant allele?

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

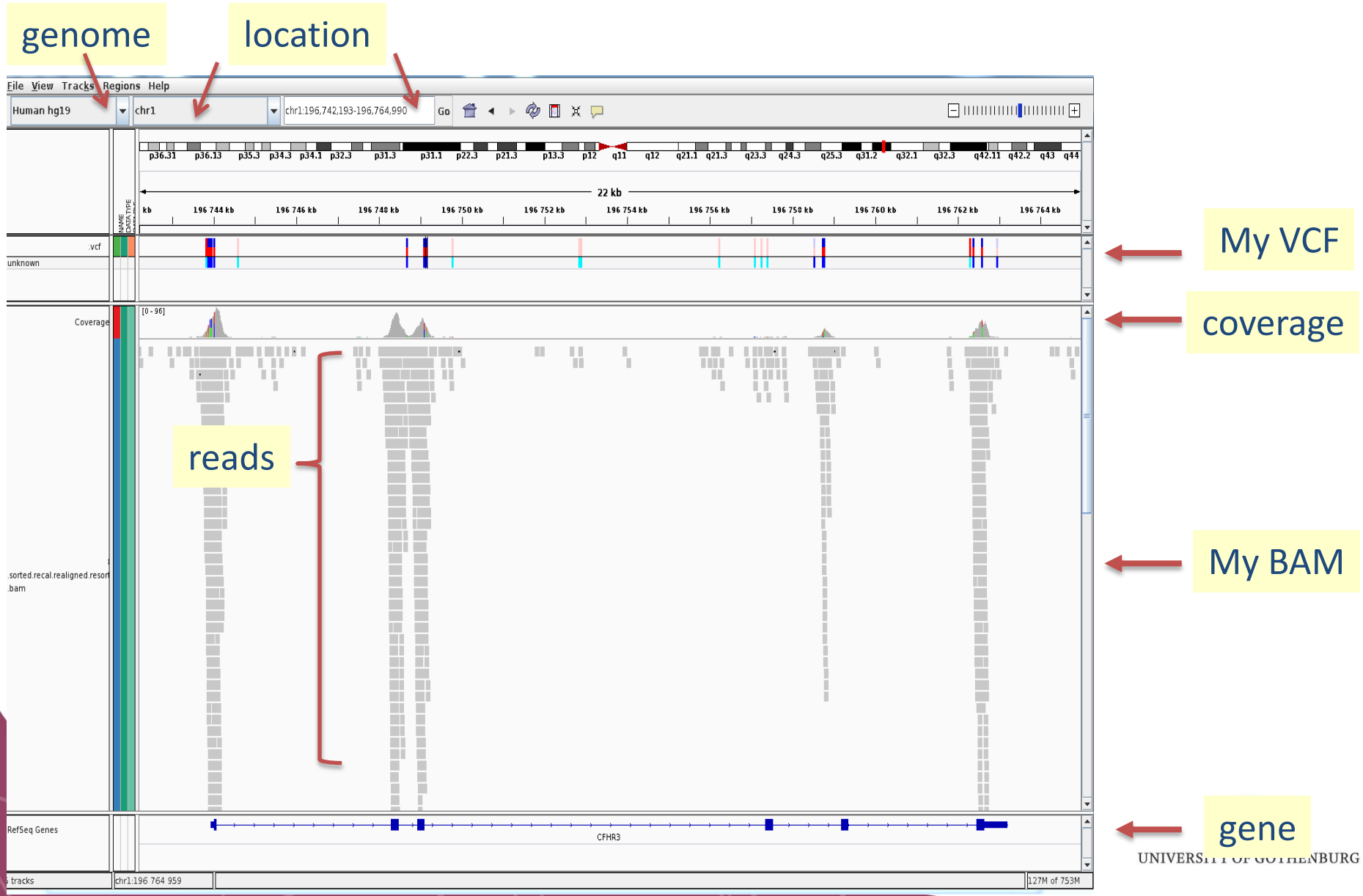
Annovar,
SIFT, PP2,
dbSNP,
GO,
KEGG,
OMIM
1000G

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

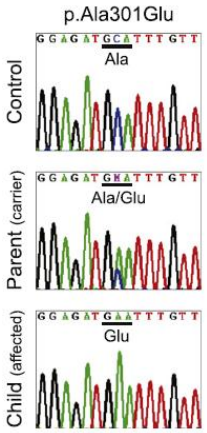
Control → Seq →SNPs →

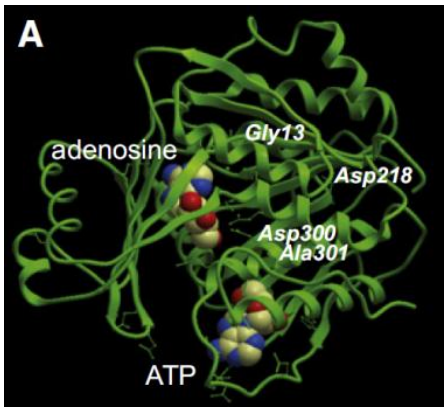
Making sense of the data

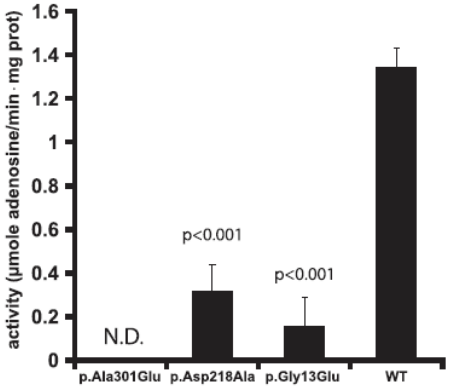
Exome sequencing cases → Coding variants → Disease model / Disease knowledge → **Candidate genes**

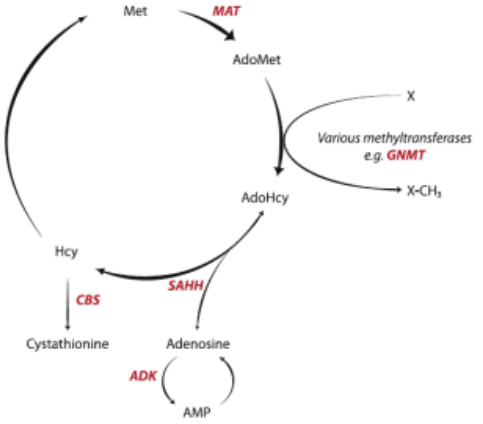
Familiy filters, Controls
Genetic variation DBs

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:

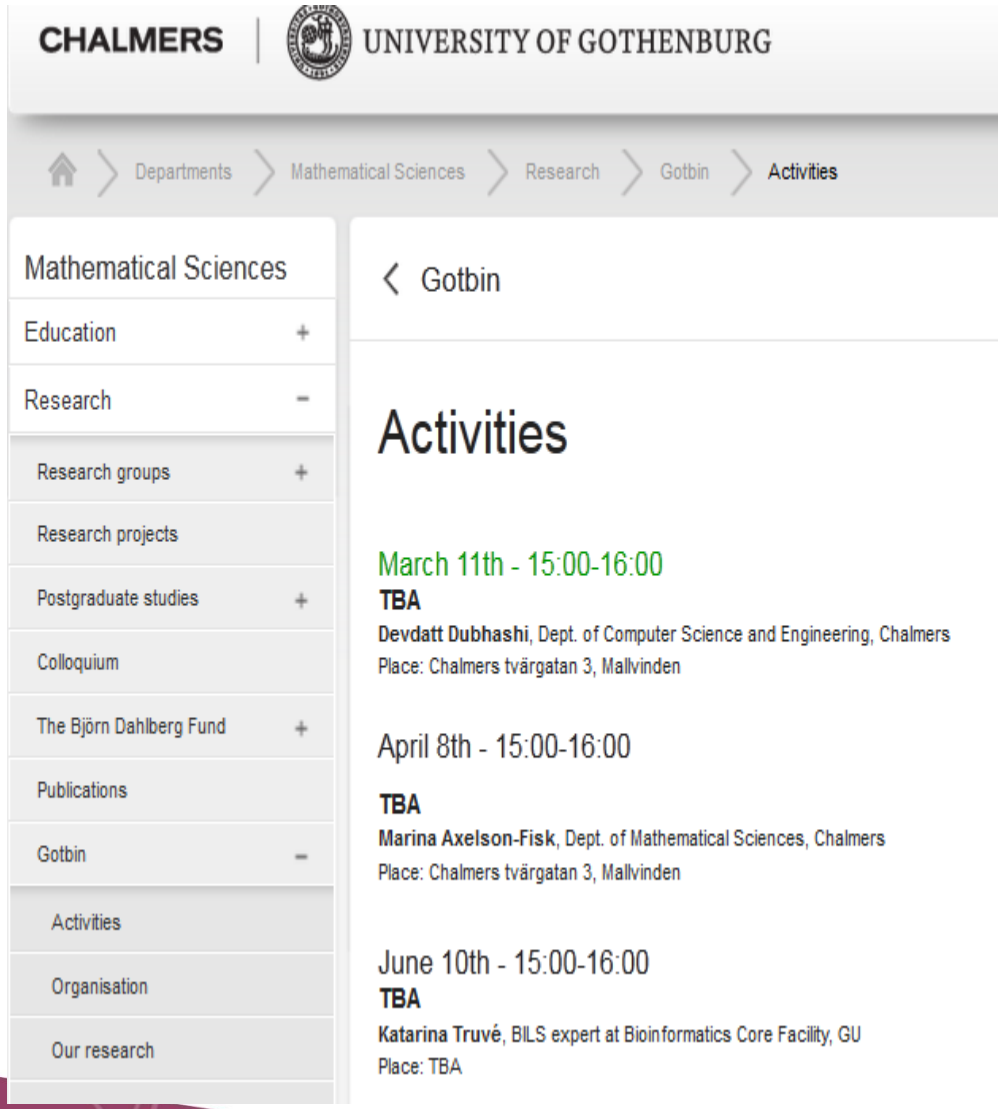
- Set up an interdisciplinary and collaborative environment
- Contribute to the development of a wide range of research projects
- Increase the understanding of statistical and bioinformatics and analysis

http://www.chalmers.se/gotbin

GOTBIN
GOTHENBURG BIOINFORMATICS NETWORK

To promote and forward bioinformatics in Gothenburg

GoBiG – Gothenburg Bioinformatics Group for PhD Students



BUNSA

Bioinformatics User Network at the Sahlgrenska Academy

forum that brings together researchers that deal with bioinformatics questions within biomedical research at the SA

- List different research interests and researchers
- Promote communication
- Promote knowledge transfer
- Provide commont repository of bioinformatics methods

Technical meetings

bioinformatics@gu.se