

Bioinformática 101

Felipe Simao

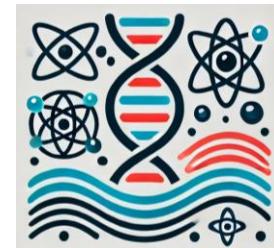
06.11.2024

Biología

Bioquímica



Biofísica



Química

Físico-química



Física

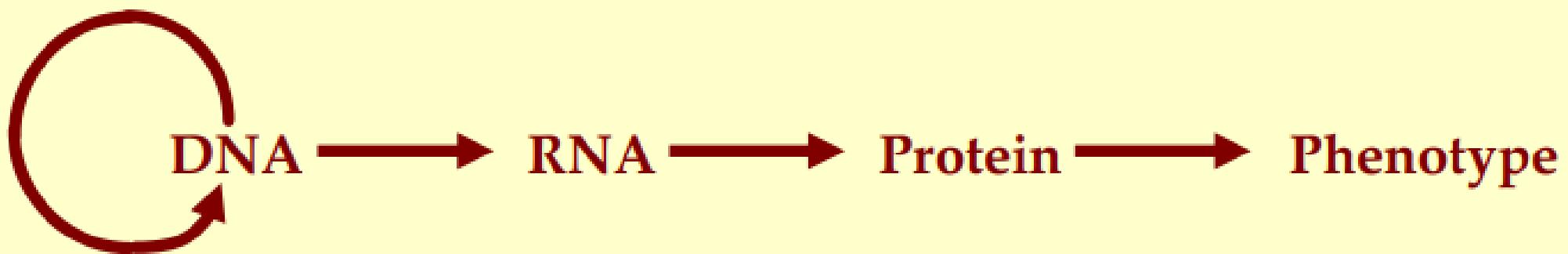


A bioinformática é um campo multidisciplinar que utiliza métodos de informática, matemática, estatística e engenharia nas áreas de estudo da biologia; especialmente quando lidando com grandes quantidades de dados.

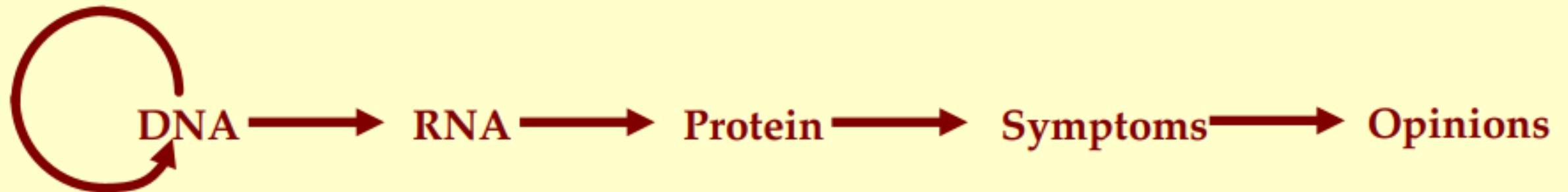
- "Eu não acho que toda a computação biológica seja bioinformática, por exemplo, a modelagem matemática não é bioinformática, mesmo quando conectada a problemas relacionados à biologia. Na minha opinião, a bioinformática tem a ver com a gestão e o uso subsequente da informação biológica, particularmente a informação genética."

- Richard Durbin

Dogma central da Biologia Molecular



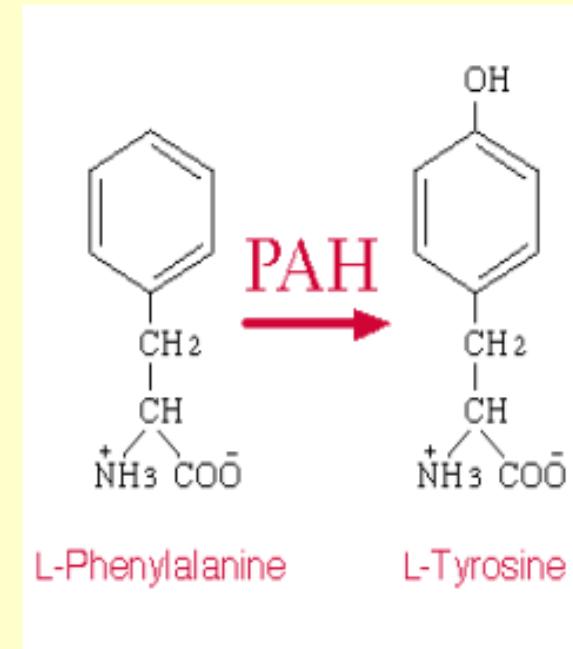
Dogma central da Medicina



Dogma central da Bioinformática

Genetic Information → Molecular Structure → Biochemical Function → Phenotype (Symptoms)

MVHLTPPEEKT
AVNALWGKVN
VDAVGGEALG
RLLVVYPWTQ
RFFESFGDLS
SPDAVMGNPK
VKAHGKKVLG
AFSDGLAHLD
NLKGTFSQLS
ELHCDKLHVD
PENFRLLGKV
LVCVLARNFG
KEFTPQMQAA
YQKVVAGVAN
ALAHKYH



Dogma central da Bioinformática

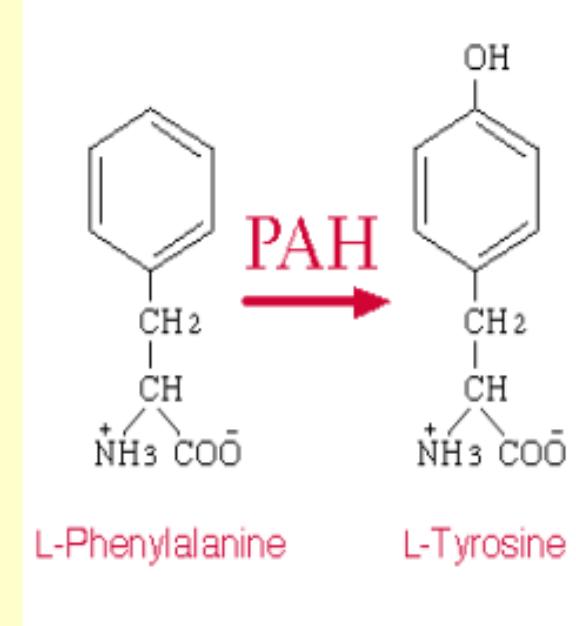
Genetic Information

MVHLTPPEEKT
AVNALWGKVN
VDAVGGEALG
RLLVVYPWTQ
RFFESFGDLS
SPDAVMGNPK
VKAHGKKVLG
AFSDGLAHLD
NLKGTFSQLS
ELHCDKLHVD
PENFRLLGNV
LVCVLARNFG
KEFTPQMQAA
YQKVVAGVAN
ALAHKYH

Molecular Structure



Biochemical Function



Phenotype (Symptoms)

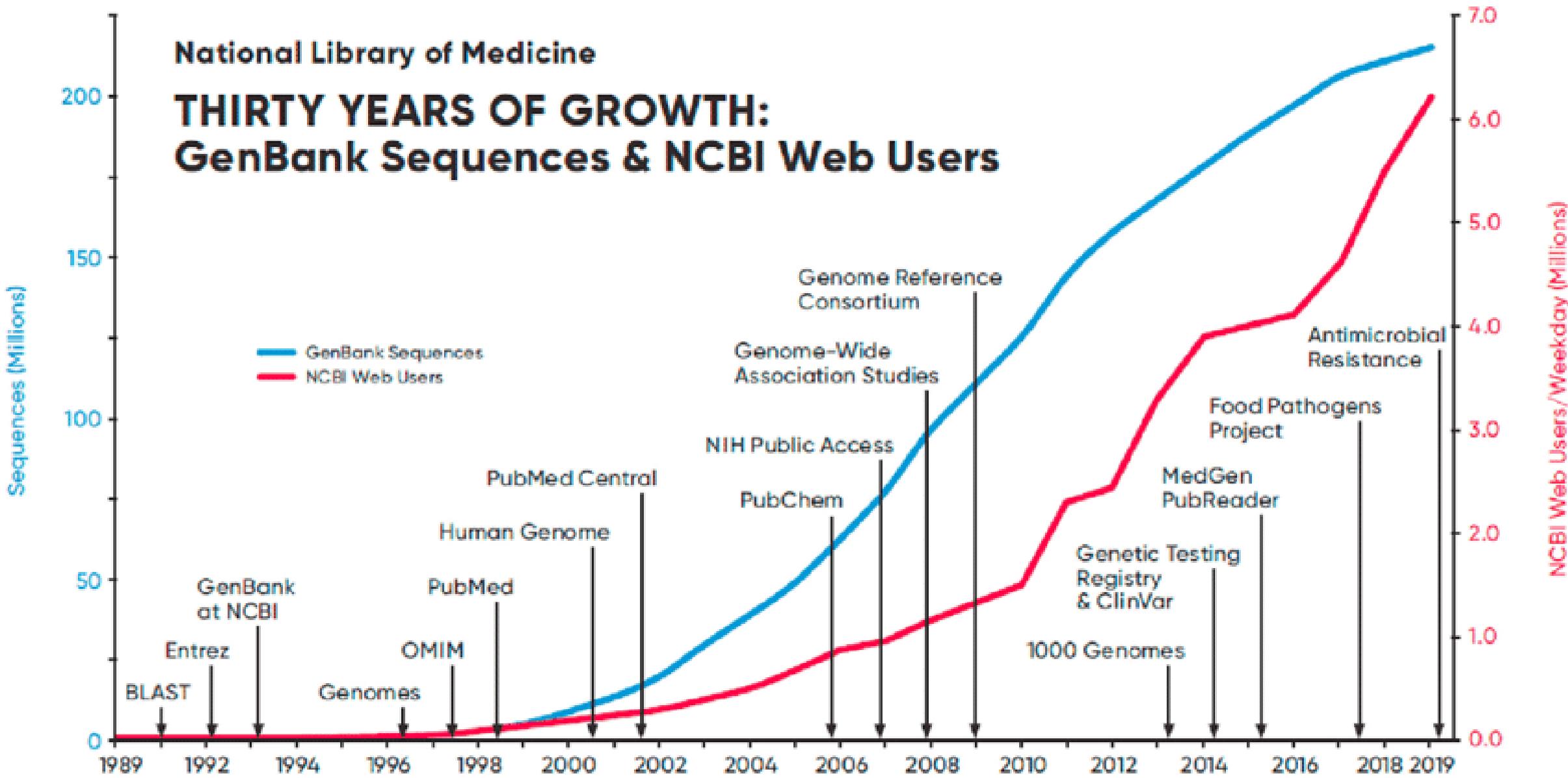


Porque a bioinformática é importante?

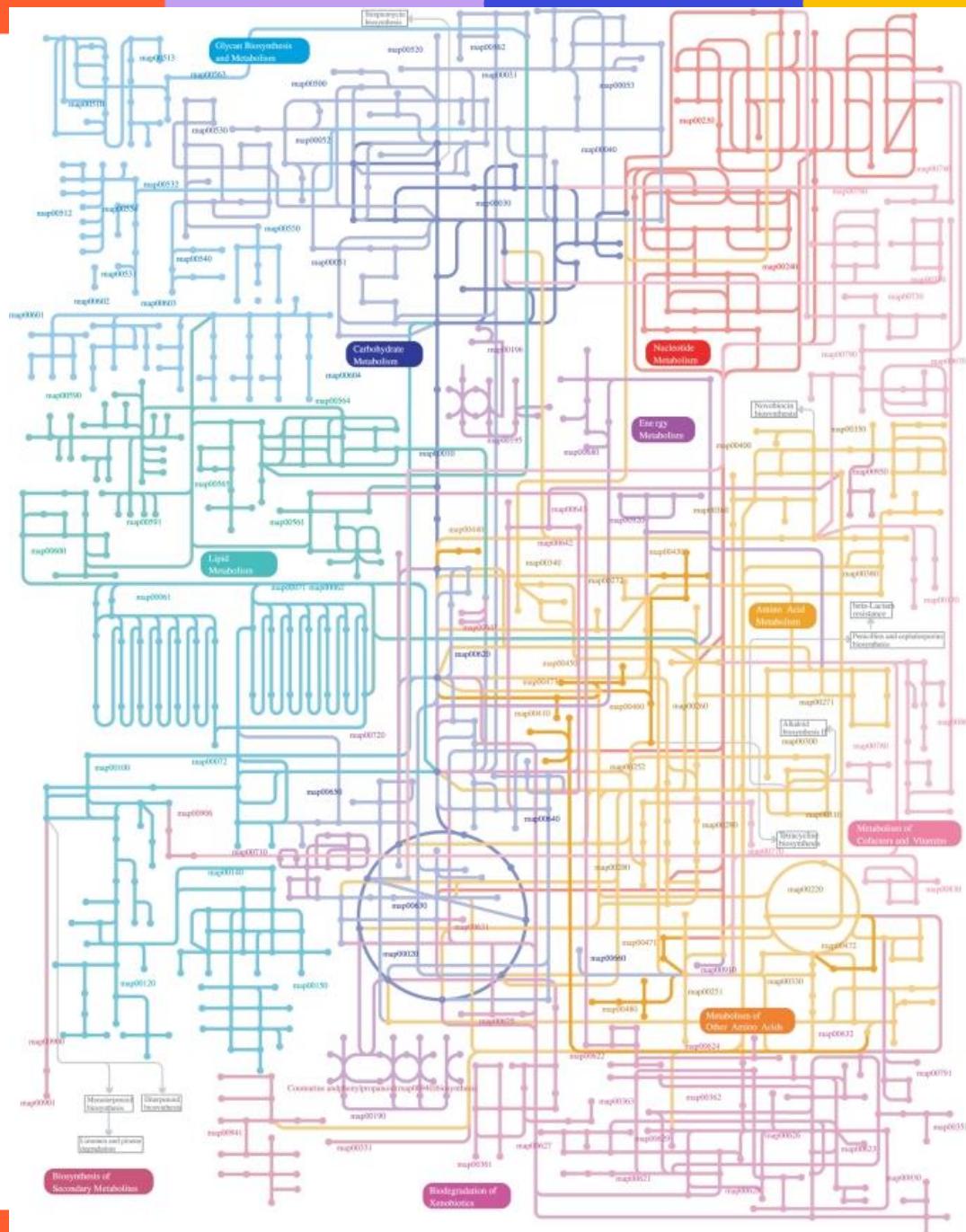
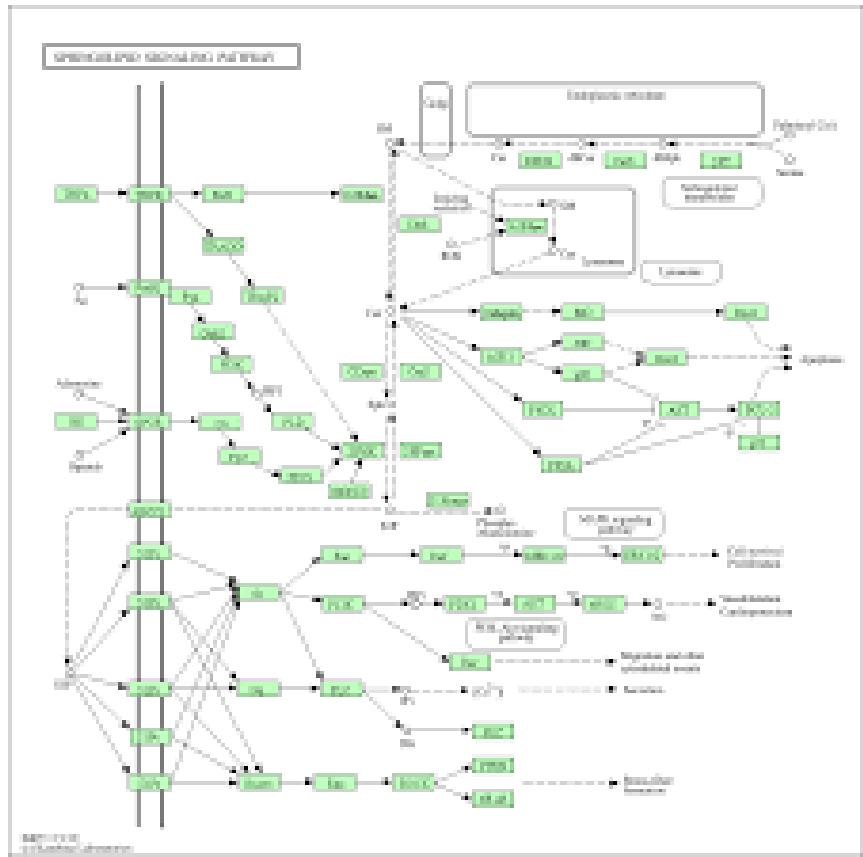
- Novos métodos produzem quantidades astronômicas de dados
- Métodos avançados de análise de dados em larga escala são necessários para derivar insights biológicos
- Formas de produção de dados (ex. NGS) geram dados ruidosos com muitos valores faltados
- Mudança de paradigma de utilizar ciência de dados na biologia

National Library of Medicine

THIRTY YEARS OF GROWTH: GenBank Sequences & NCBI Web Users



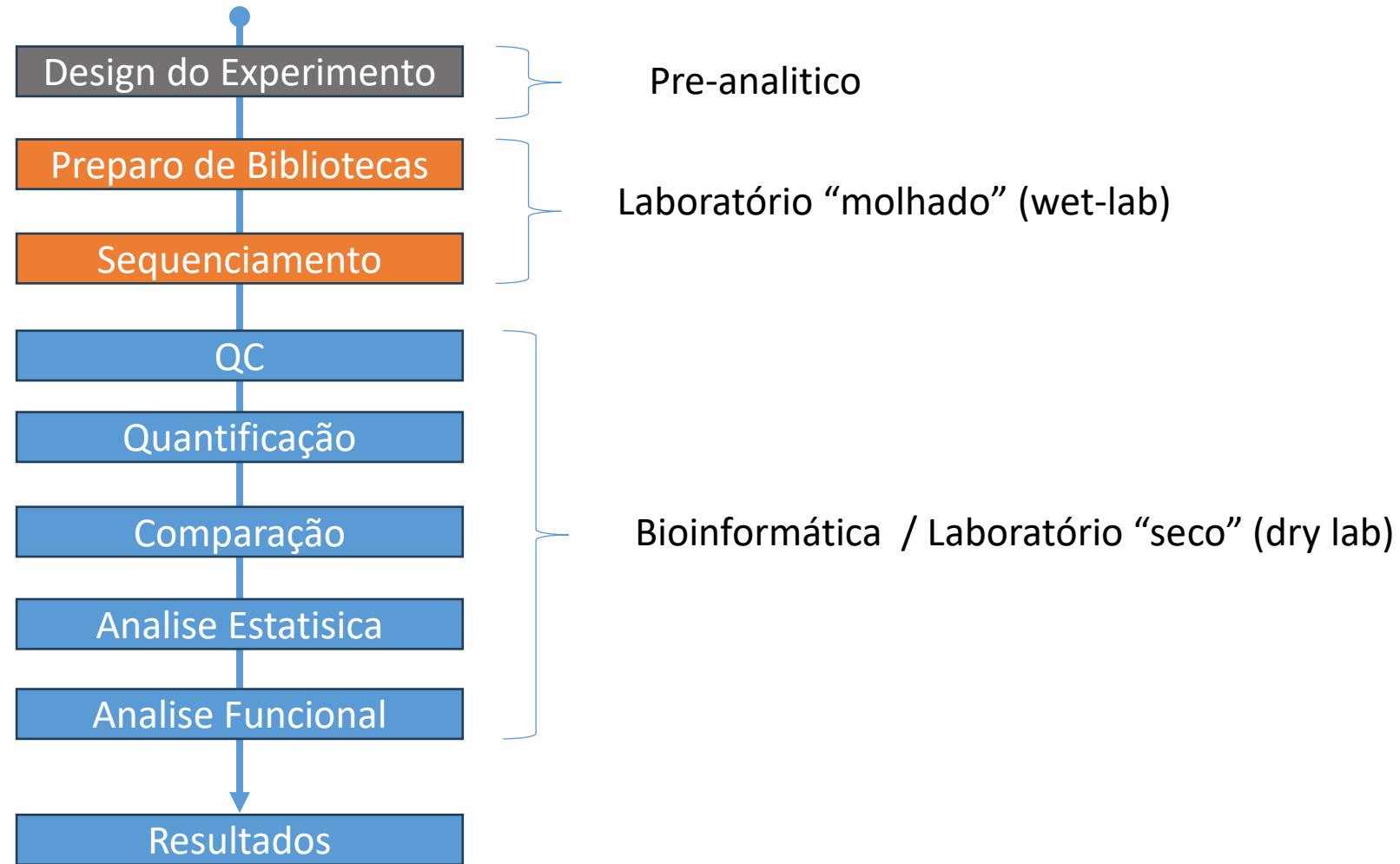
Mapa de vias metabólicas no KEGG



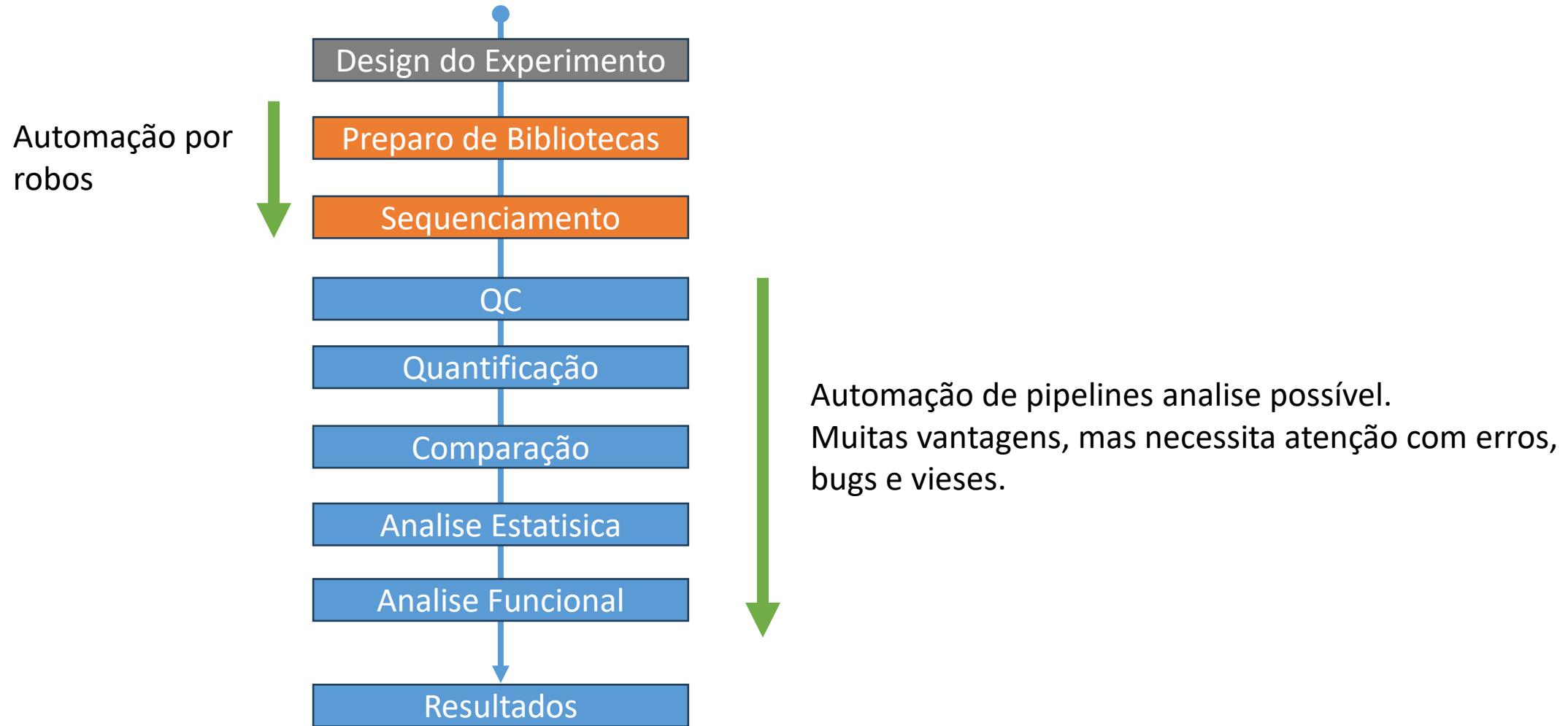
Métodos comuns

- Montagem de sequencias (ex. projeto genoma humano)
- Alinhamento de sequencias
- Chamada de variantes
- Previsão de risco, resistência, sensibilidade...
- Previsão de função / fenótipo
- Genômica comparativa
- Analises filogenéticas
- Analises estruturais

Bioinformática em um projeto NGS

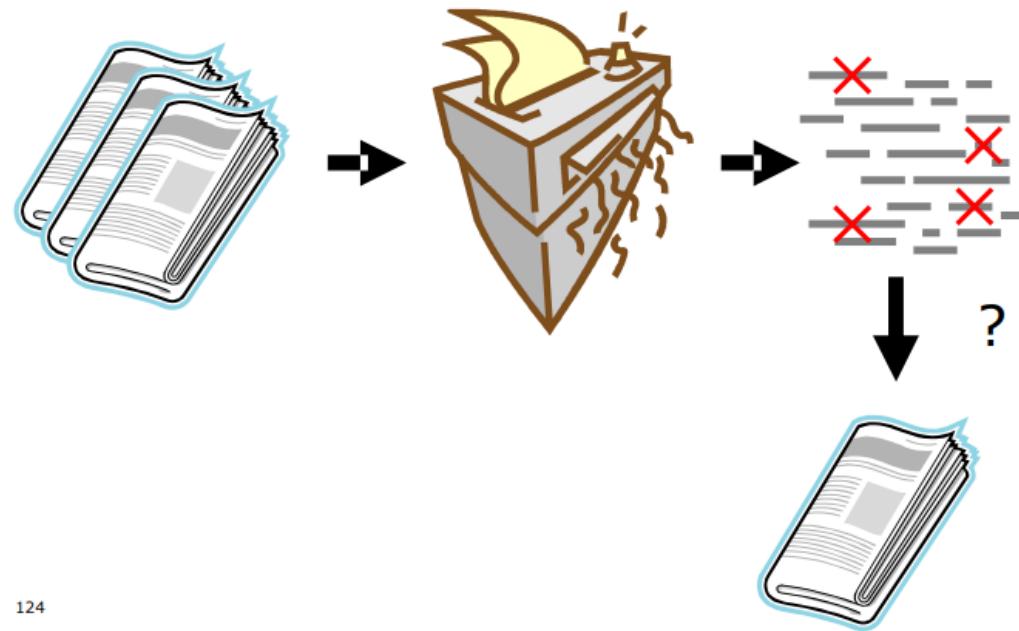


Bioinformática em um projeto NGS

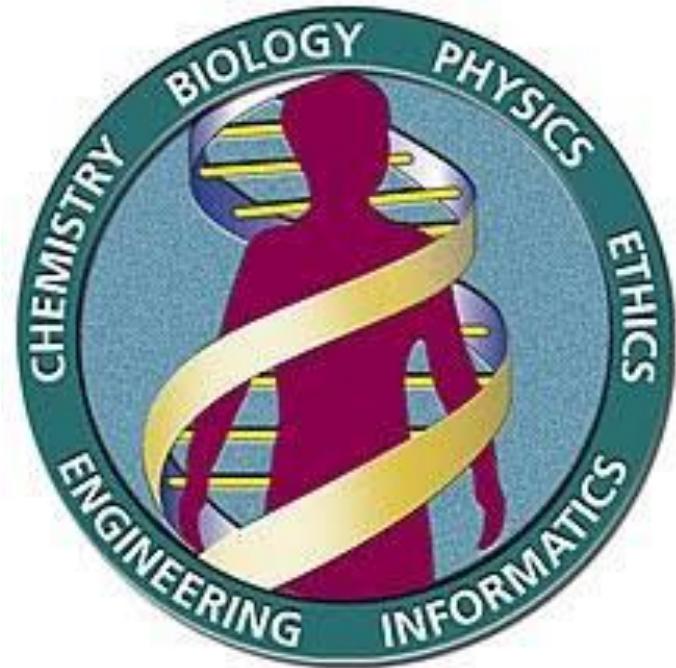


Projeto Genoma Humano

- Montagem de sequencias

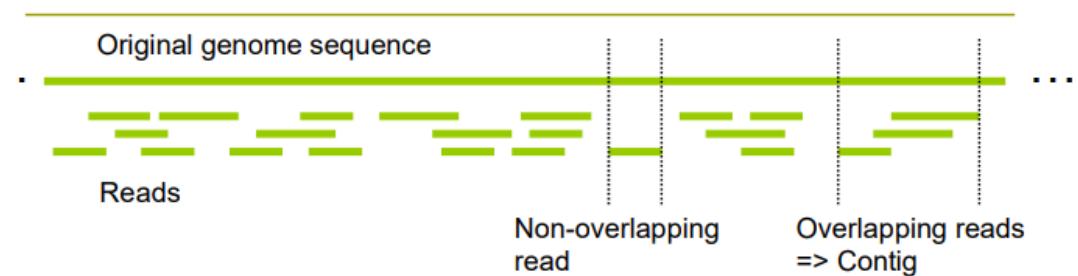


124



Projeto Genoma Humano

- Problema NP-completo
- Leituras podem ser originarias de ambas as fitas do DNA
- Taxa de erro em leituras não trivial
- Regiões repetitivas ou com homologia dificultam estratégias
- Baixa cobertura em certas regiões



- Ordering of the reads is initially unknown
- Overlaps resolved by aligning the reads
- In a 3×10^9 bp genome with 500 bp reads and 5x coverage, there are $\sim 10^7$ reads and $\sim 10^7(10^7-1)/2 = \sim 5 \times 10^{13}$ pairwise sequence comparisons

Overlap

AA ACT TT CC GGG G C G C G C C
A A G C T T C G C G A G G T A T T A T
C G A G G T A T T A T C G A A T A G G

G A C G C T A G T G C G T G T A T T A

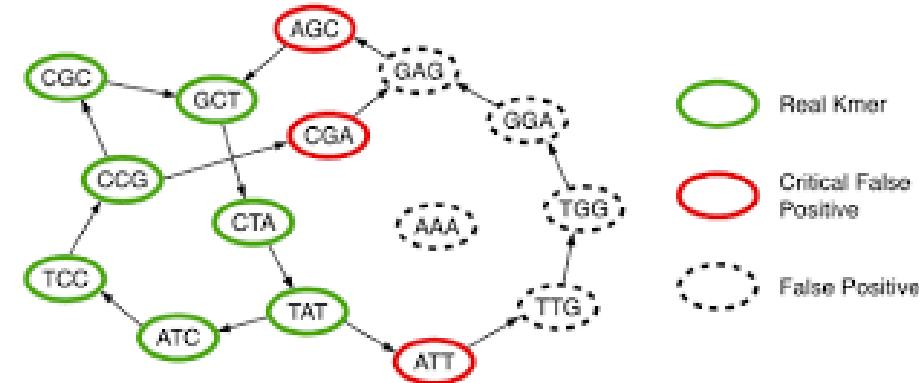
C G T A A G C T T C G C G A G G T A T

Layout

G A C G C T A G T G C G T G T A T T A
C G A G G T A T T A T C G A A T A G G
A A G C T T C G C G A G G T A T T A T
C G T A A G C T T C G C G A G G T A T

Consensus

C G T A A G C T T C G C G A G G T A T T A T C G A A T A G G



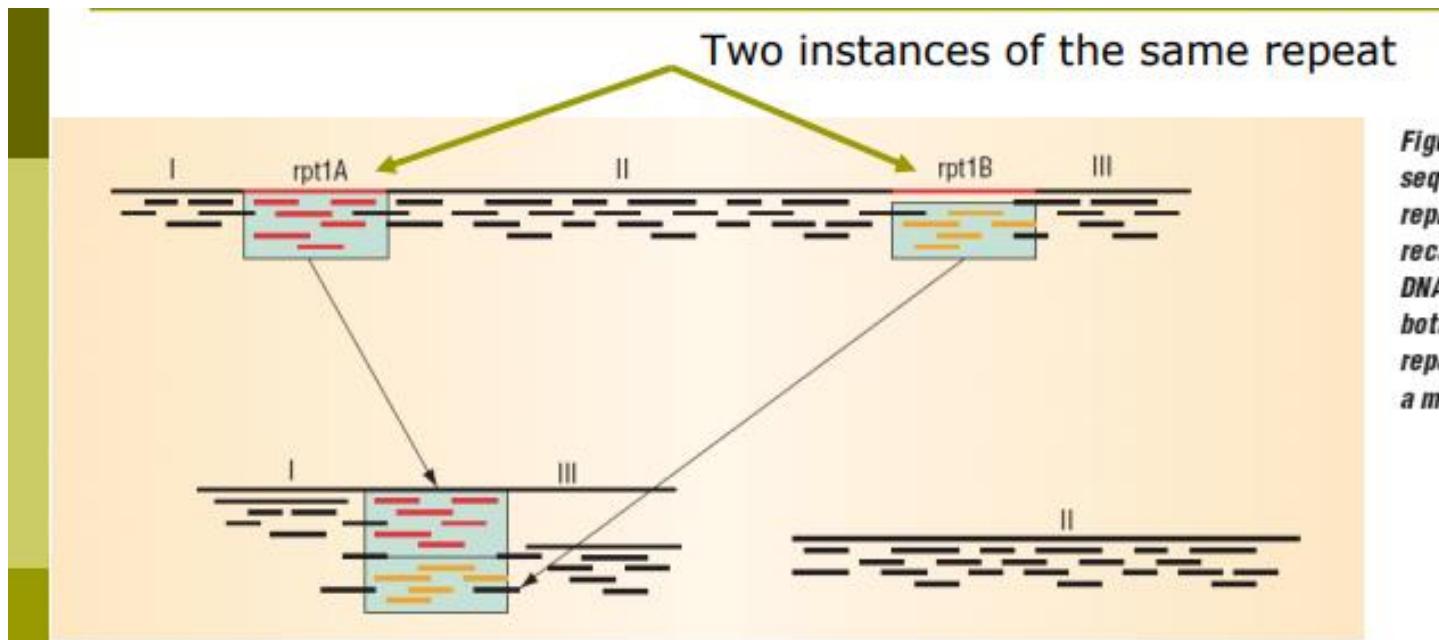


Figure 2. Repeat sequence. The top represents the correct layout of three DNA sequences. The bottom shows a repeat collapsed in a misassembly.

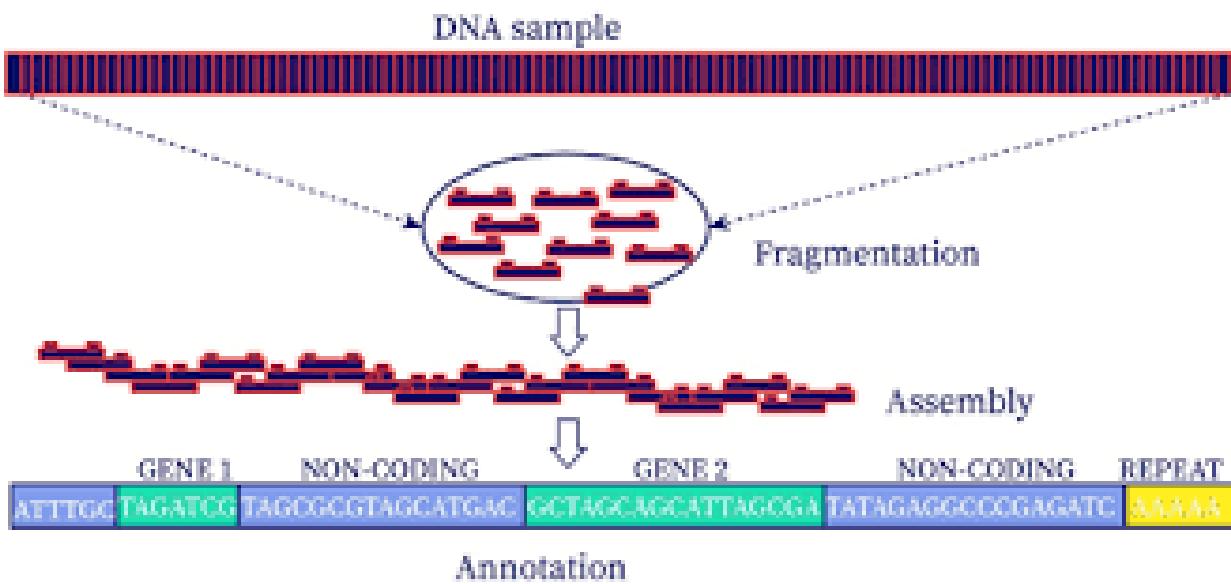


Hepatitis delta virus, complete genome

1 atgagccaag ttccgaacaa ggattcgccg ggaggataga tcagcgcgg agaggggtga
61 gtcggtaaag agcattggaa cgtcgagat acaactccc agaaggaaaa aagagaaagc
121 aagaagcgg a tgaatttccc cataacgcca gtgaaactct aggaagggg aagagggaaag
181 gtggaagaga aggaggcggg cctcccgatc cgaggggccc ggcggccaag tttggaggac
241 actccggccc gaagggttga gagtaccca gagggaggaa gccacacgga gtagaacaga
301 gaaatcacct ccagaggacc cttcagcga acagagagcg catcgcgaga gggagtagac
361 catagcgata ggaggggatg ctaggagttg gggagaccc aagcgaggag gaaagcaaag
421 agagcagcgg ggctagcagg tgggtgttcc gccccccgag agggacgag tgaggcttat
481 cccggggAAC tcgacttatac gtccccacat agcagactcc cgacccctt ttcaaagtga
541 ccgagggggg tgactttgaa cattggggac cagtggagcc atggatgct cctcccgatt
601 ccccccAAC tccttcccc caagggtcgc ccaggaatgg cgggacccca ctctgcaggg
661 tccgcgttcc atcctttttt acctgatggc cggcatggc ccagccctt cgctggcgcc
721 ggctgggcaa cattccgagg ggaccgtccc ctcggtaatg gcgaatggg cccacaaatc
781 tctctagctt cccagagaga agcgagagaa aagtggctct cccttagcca tccgagtgga
841 cgtgcgttcc cttcggatg cccaggtcgg accgcgagga ggtggagatg ccatgcccac
901 ccgaagagga aagaaggacg cgagacgcaa acctgcgagt gaaacccgc ttatttact
961 ggggtcgaca actctgggaa gaggaggag ggtcggttgg gaagagtata tcctatgggaa
1021 atccctgtct tcctttatg tccagtcctt ccccggtccg agtaaagggg gactccgggaa
1081 ctcccttgcattt gctggggacg aagccggccc cggcgctcc ctcgttcca ctttcgaggg
1141 gttcacacc cccaaacctgc gggccggcta ttcttcttcc cttctctcg tcttcctcg
1201 tcaacccctt aagtccctt tcctccctt tgcgtggat tttccccc cccgatagct
1261 gtttctttt gttctcgagg gccttccttc gtcgggtatc ctgcctctcc ttgtcggtga
1321 atccctccctt ggaaggcctc ttccctaggatc cggagtctac ttccatctgg tccgttccgg
1381 ccctttcgc cgggggagcc ccctctccat ctttatctt cttccgaga attcccttga
1441 tttttcccaag ccaggatgt tcatcctcaa gtttcttgcattt accttccggaa
1501 ggtctcttc gatccctt aacttcttc ttccgctcac ccactgctcg agaacctt
1561 ctctcccccc gcggttttc ttcccttcgg gcccggctcat ctgcacttag aggcgcacgg
1621 cctcagtaact cttaactttt tctgtaaaga ggagactgct ggcctgtcg cccaaagtcc
1681 ag

Projeto Genoma Humano

- Anotação do genoma



Identificação de Variantes

- Detecção de diferenças em relação ao genoma referência.
- Principais usos clínicos atualmente no diagnóstico de doenças genéticas e no tratamento de câncer.
- Exemplo de testes clínicos: Exoma Completo (WES), painéis multigenicos, microarrays, exame pre-natal não invasivo (NIPT/NIFTY)

Exoma Completo e painéis

- Exoma: regiões codificadoras do genoma humano (apenas 1,5% do DNA total)
- Em média, cada pessoa apresenta 40 mil variantes em seu DNA
- ~85% das doenças genéticas são causadas por alterações genéticas no Exoma
- Sequenciamento do exoma completo: identifica variantes que podem ser patológicas e estar relacionadas ao quadro clínico do paciente

Exames Possíveis

- Exoma completo
- Pré-Natal
- Trio
- Painel Personalizado
- Gene único personalizado

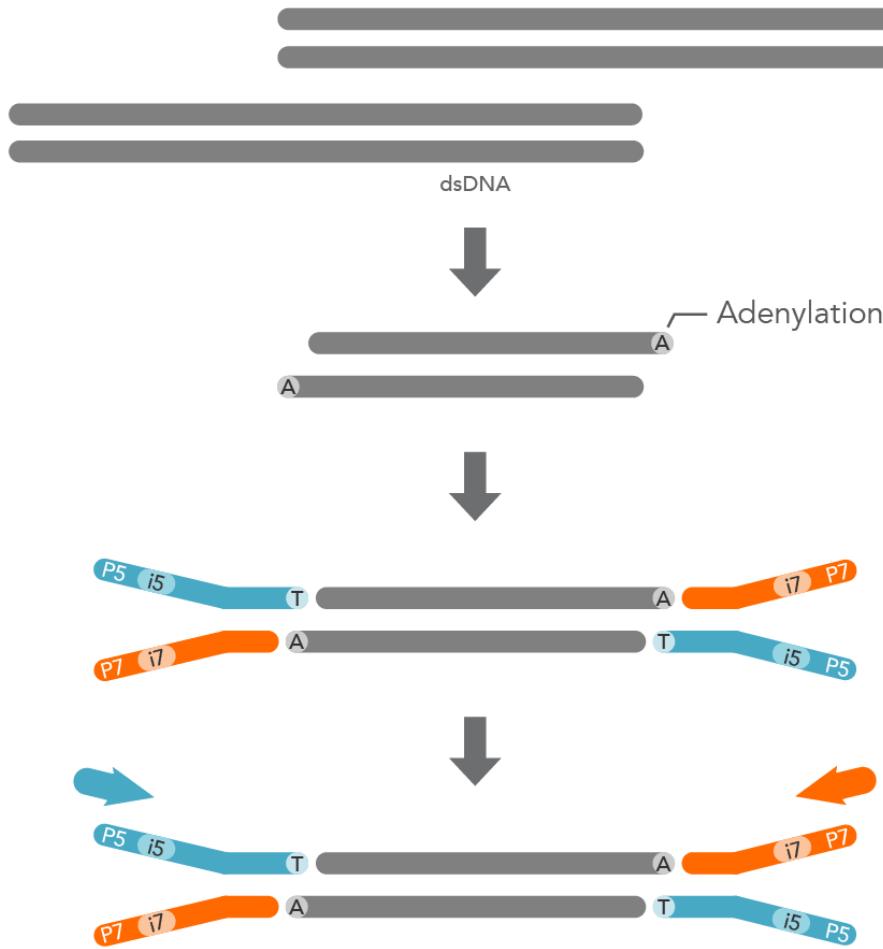
Detecção

- SNVs, Indels
- CNVs, Inserções Alu

	MicroArray (CGH)	Painel de Genes ou Único Gene	Exoma Completo	Genoma Completo
Análise de todos os genes conhecidos	✗	✗	✓	✓
Análise de grandes deleções e duplicações	✓	✗	✗	✓
Análise dos genes clinicamente mais relevantes	✗	✓	✓	✓
Análise de Achados incidentais	✗	✗	✓	✓

O exame de sequenciamento de exoma completo não deve ser considerado apenas como um único exame definitivo para análise de doenças, e pode ser usado em conjunto e não limitado apenas com outros testes como microarranjos (CGH), painéis de genes ou único gene.

Sequenciamento: Preparo de Biblioteca



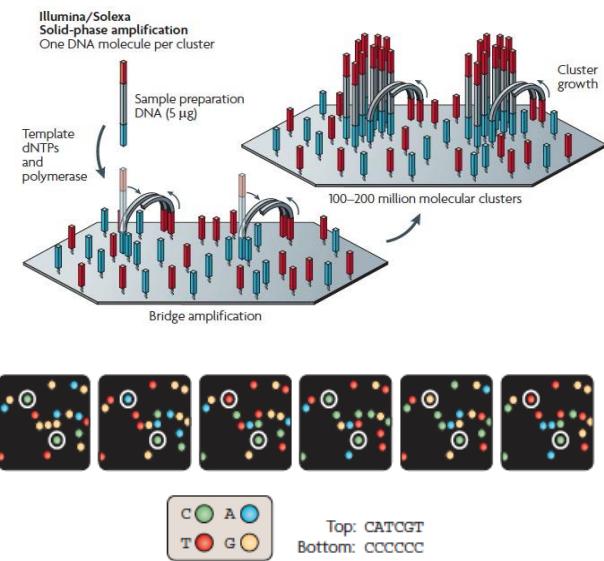
- Fragmentação do DNA
- Reparação (end repair) e 3' adenilação
- Ligação de adaptadores
 - Primers de sequenciamento
 - Primer de index
 - Index (código de barras ind.)
- (Amplificação por PCR)
- Seleção por tamanho e Purificação
- Controle de Qualidade
- Normalização (concentração)



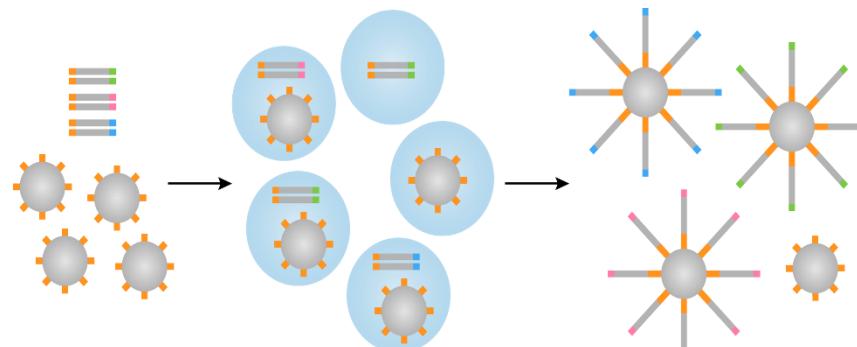
Tecnologias de Sequenciamento



Bridge PCR (Cluster)



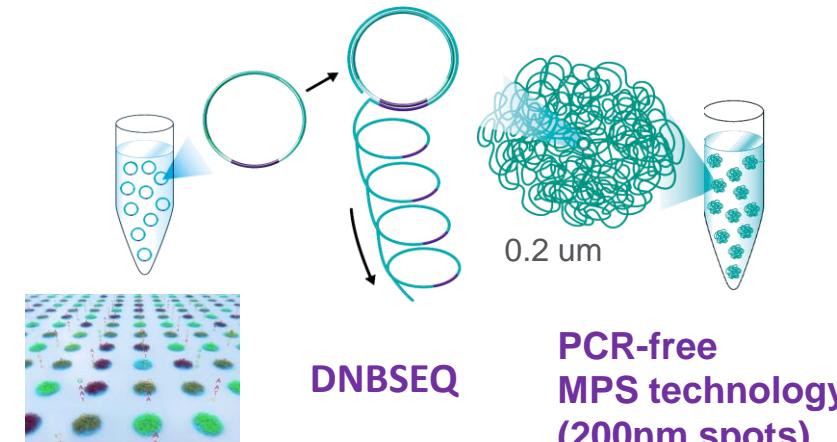
Emulsion PCR (Bead)



R. Drmanac, R. Crkvenjakov, [Prospects for a Miniaturized, Simplified and Frugal Human Genome Project](#), Scientia Yugoslavica 161 (1-2), 97-107, 1990.

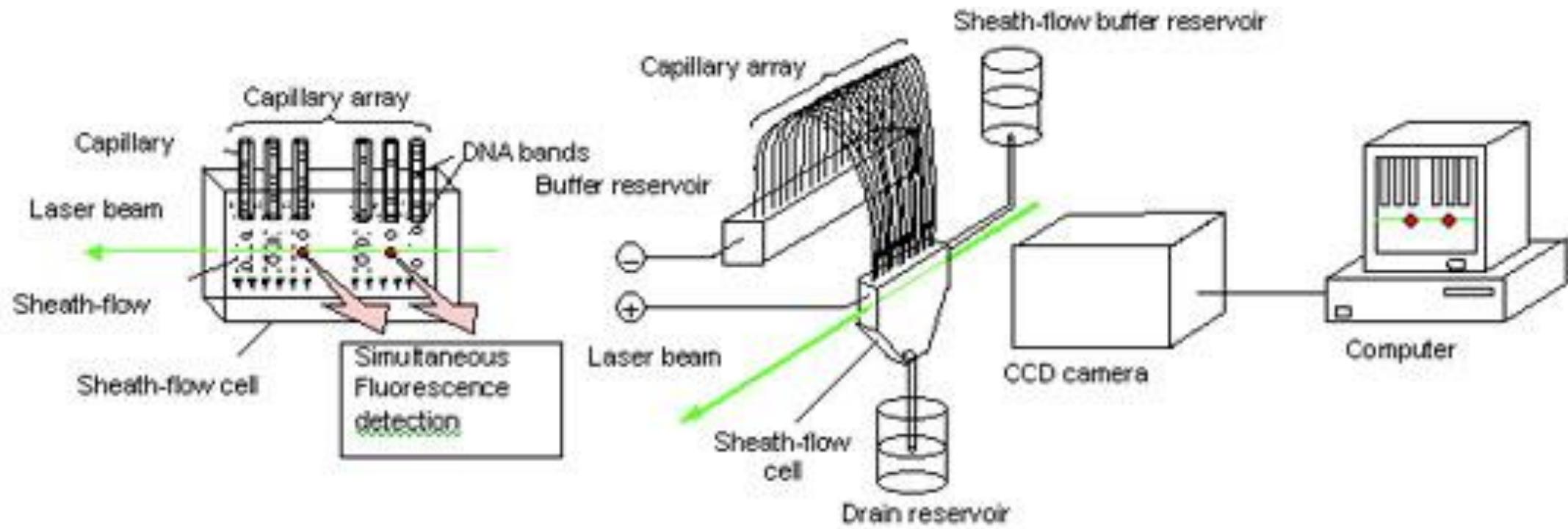


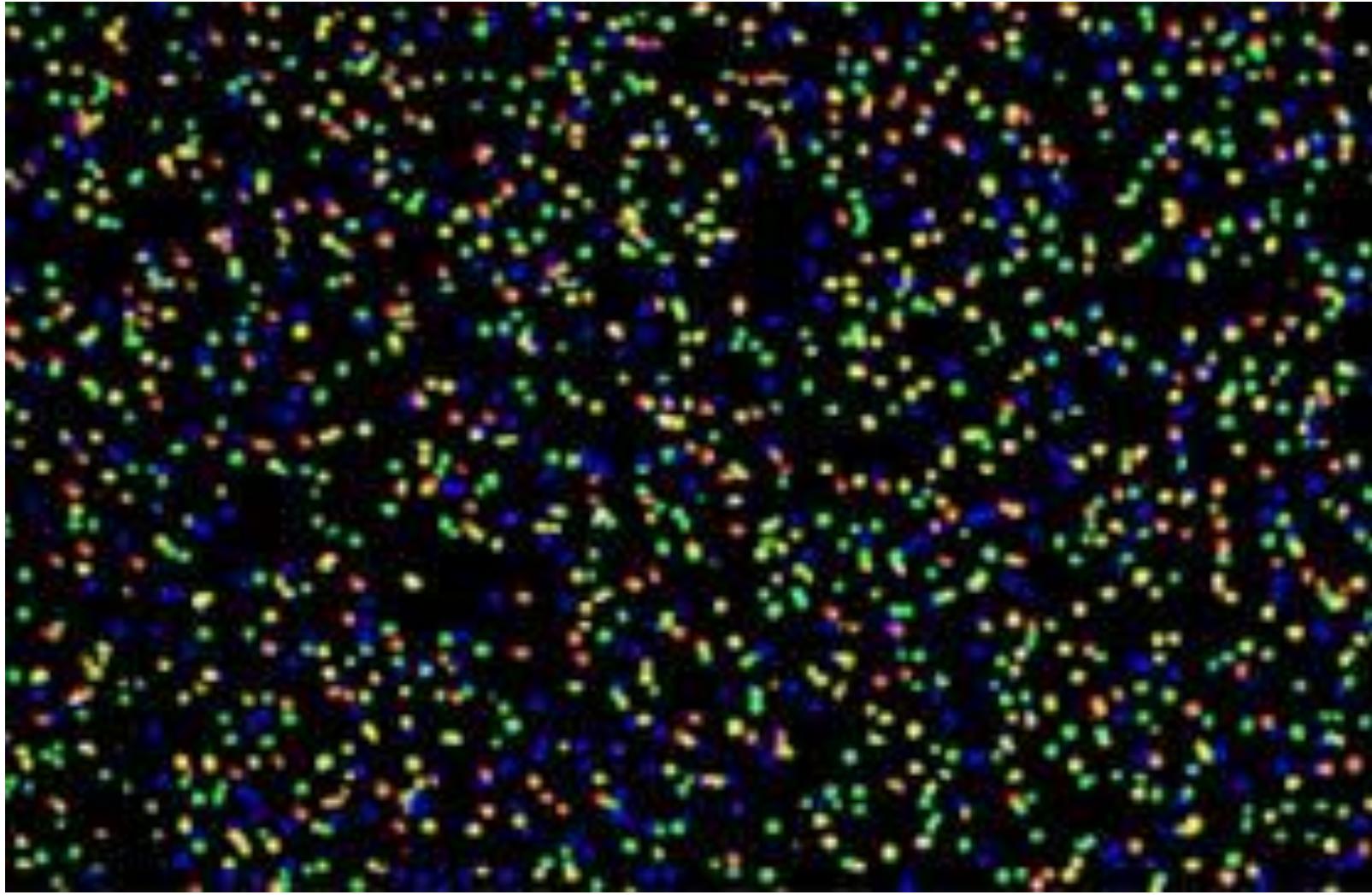
**Rolling Circle replication (RCR)
(DNA Nanoball - "DNB")**



**PCR-free
MPS technology
(200nm spots)**

R. Drmanac, et al., [Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays](#), Science, 5 November 2009 (10.1126/science.1181498)





Leituras de sequenciamento (formato: FASTQ)

Header

Sequence

Quality

```
@HWI-ST227:389:C4WA2ACXX:7:1204:2272:59979
GGAGGAAGGTCTCGCTCCTCTTCATATAAGGGAAATGGCTGAAT
+
FFFFHHHHHJIJJJJJJJIJJJIGIGIGGIJJJIJJJJJJIII
@HWI-ST227:389:C4WA2ACXX:7:1205:15214:42893
GAGGATCCCAGGGAGGAAGGTCTCGCTCCTCTTCATCTAAGGGA
+
12BAFB?A:3<AE1@<FF;1*@EG*)?0?DBD>9BF9B*?#####
@HWI-ST227:389:C4WA2ACXX:8:2208:2467:44624
AAAGAGGGAGAGAGGACCATCCTCCCTGGGATCCTCAGAAGTCTACT
+
BDDA:DB?2AA@FC>F?EEGC<FED>GFD;?GBB?<?F99*/9?9?
```

Alinhamento

- Quais posições em duas sequencias são equivalentes
- Uso na genômica clínica:

comparação entre a sequência genética do paciente
difere em relação ao genoma humano de referencia

Indel: insertion or
deletion of a base
with respect to the
ancestor sequence

acgttctag
-actcttag

Mismatch: substitution
(point mutation) of
a single base

Alinhamento

(a)

\$ a c a a c g
a a c g \$ a c
a c a a c g \$
a c a a c g \$ → a c g \$ a c a → g c \$ a a a c
c a a c g \$ a
c g \$ a c a a
g \$ a c a a c

(c)

a a c a a c a a c
\$ a c a a c g \$ a c a a c g \$ a c a a c g
a a c g \$ a c a a c g \$ a c a a c g \$ a c
a c a a c g \$ a c a a c g \$ a c a a c g \$
a c g \$ a c a a c g \$ a c a a c g \$ a c a
c a a c g \$ a c a a c g \$ a c a a c g \$ a
c g \$ a c a a c g \$ a c a a c g \$ a c a a
g \$ a c a a c g \$ a c a a c g \$ a c a a c

(b)

g c g a c g a a c g c a a c g a c a a c g
\$ a c a a c g \$ a c a a c g \$ a c a a c g \$ a c a a c g \$ a c a a c g \$ a c a a c g
a a c g \$ a c a a c g \$ a c
a c a a c g \$ a c a a c g \$
a c g \$ a c a a c g \$ a c a
c a a c g \$ a c a a c g \$ a c a a c g \$ a c a a c g \$ a c a a c g \$ a c a a c g \$ a
c g \$ a c a a c g \$ a c a a c g \$ a c a a c g \$ a c a a c g \$ a c a a c g \$ a c a a
g \$ a c a a c g \$ a c a a c

Burrows-Wheeler Transform

Alinhamento

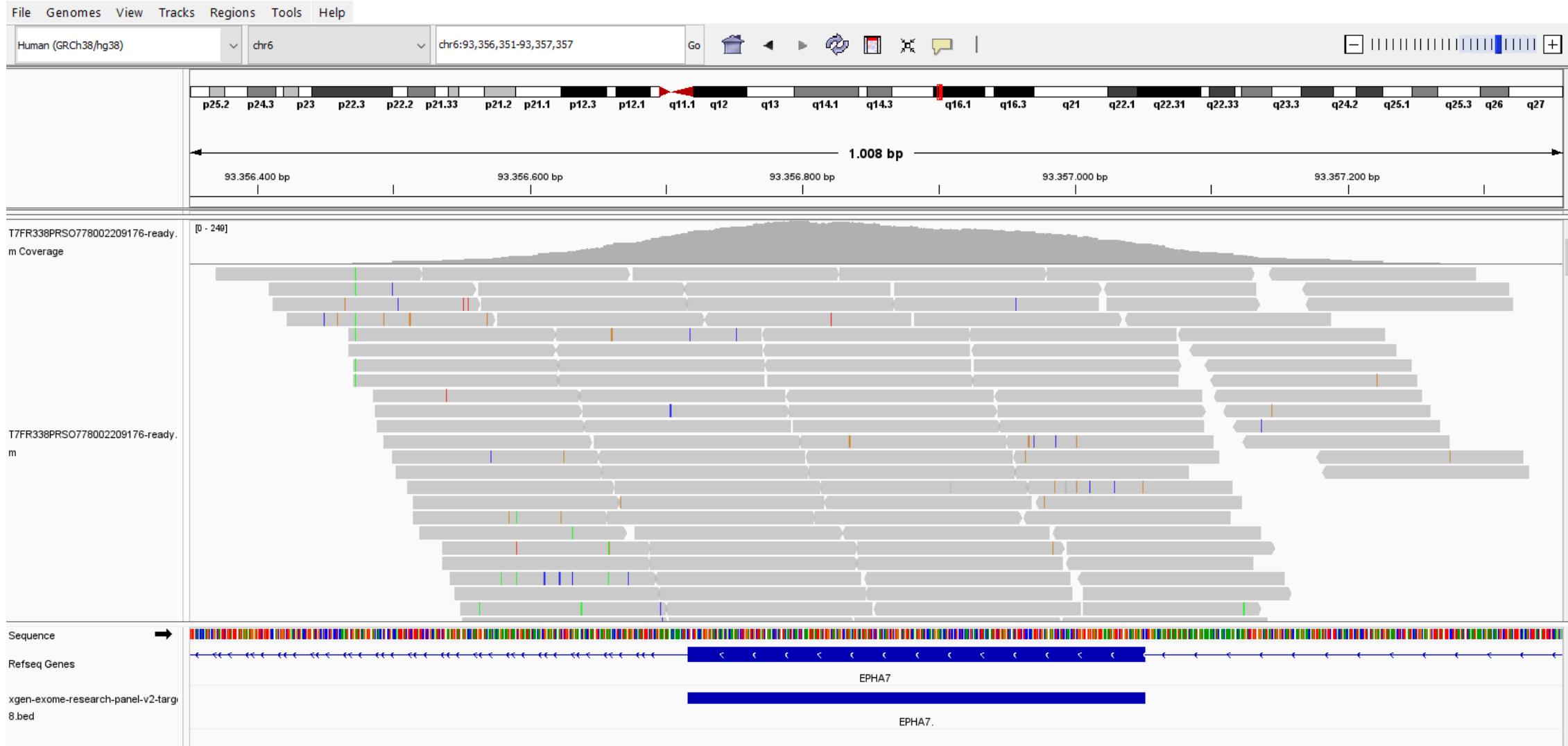
Transformation				
1. Input	2. All rotations	3. Sort into lexical order	4. Take the last column	5. Output
^BANANA\$	^BANANA\$ \$^BANANA A\$^BANAN NA\$^BANA ANA\$^BAN NANA\$^BA ANANA\$^B BANANA\$^	ANANA\$^B ANA\$^BAN A\$^BANAN BANANA\$^ NANA\$^BA NA\$^BANA ^BANANA\$ \$^BANANA	ANANA\$^B ANA\$^BAN A\$^BANAN BANANA\$^ NANA\$^BA NA\$^BANA ^BANANA\$ \$^BANANA	BNN^AA\$A

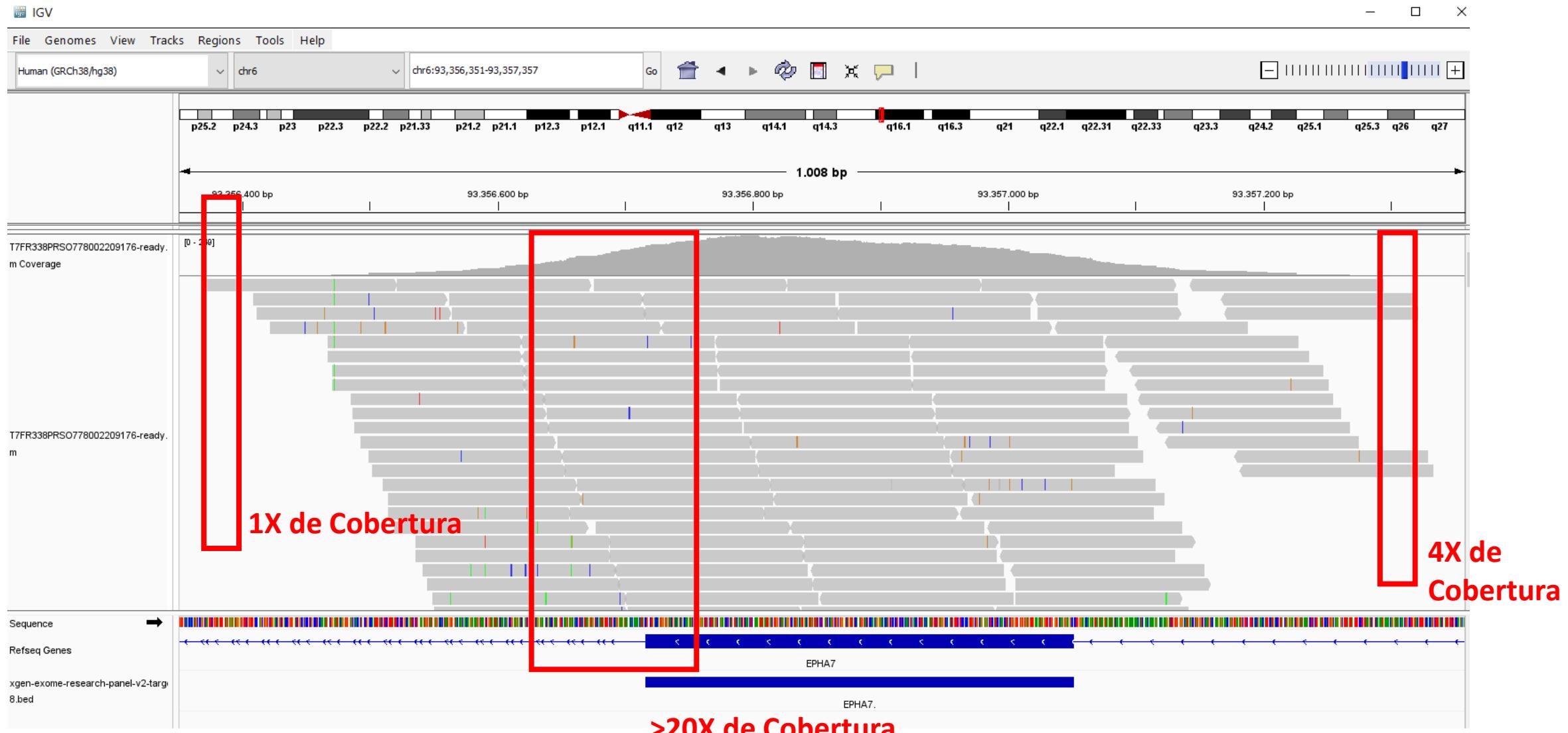
Burrows-Wheeler Transform



IGV

- □ ×





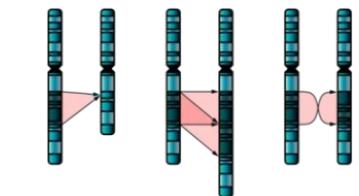
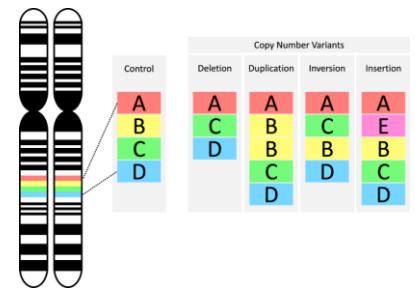
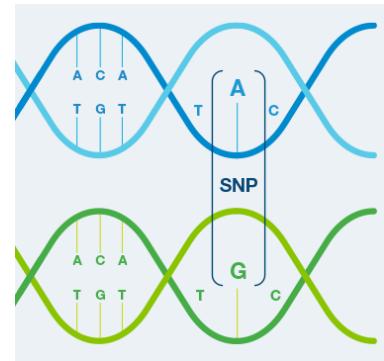
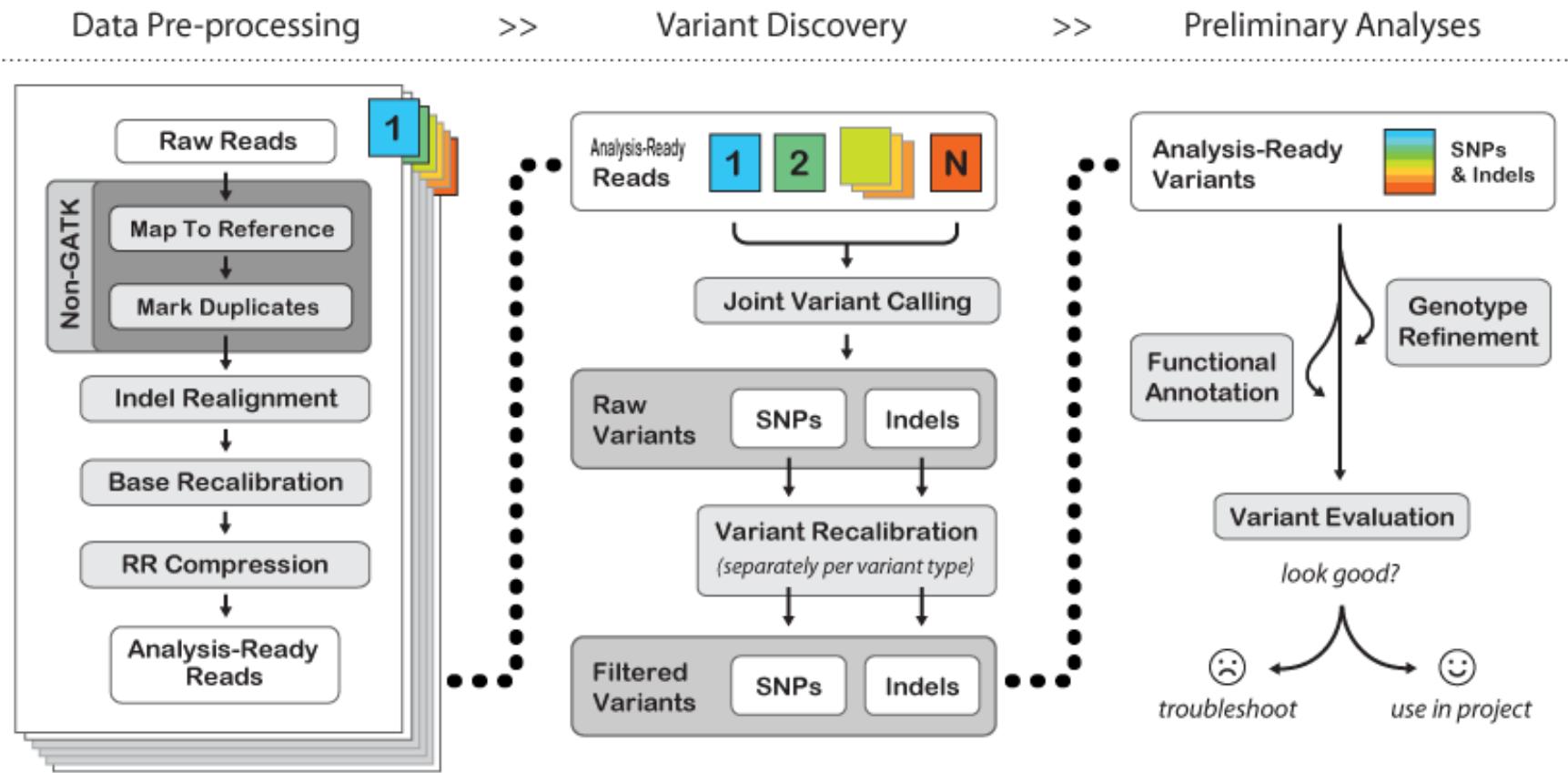
Alinhamento em formato SAM/BAM

>sp|P51490|OPSB_BOVIN G Blue-sensitive opsin (BOP) (Blue cone photoreceptor pigment)
Length=349

Score = 297 bits (761), Expect = 4e-80, Method: Composition-based stats.
Identities = 151/327 (46%), Positives = 213/327 (65%), Gaps = 1/327 (0%)

Query	35	STRGPPEGPNYHIAPRWVYHLSVWMIFVVIASVFTNGLVLAATMFKKLRHPLNWILVN	94
		S GP++GP YH+AP W +HL +V+M FV N VL AT++++KLR PLN+ILVN	
Sbjct	17	SLVGPWDGPQYH LAPVWAFHL QAVFMGFVFFVGTPLNATVLVATLRYRKL RQPLNYILVN	76
Query	95	LAVADLAETVIASTISVVNVQVYGYFVLGHPMCVLEGYTVSLCGITGLWSLAIISWERWMV	154
		+++ + + I + YGYFV G +C LE + G+ WSLA +++ER+++	
Sbjct	77	VSLGGPIYCI PSV FIVPITS CYGYFV PGRHVCALEAFLGCTAGLVTGWSLAFLA PERYII	136
Query	155	VCKPPGNVRFDAKLAIVGIAFSWIWA AVWTAPP IFGWSRYWPHGLKTSCGP DVFGSSYP	214
		+CKPGN RF +K A++ + +W + PP FGWSR+ P GL+ SCGP D ++ +	
Sbjct	137	ICKPPGNPRFSSKHALMV VVATWTIGIGVSIPPF PGWSRFVPEGLQCSCGP DWYT VGTKY	196
Query	215	GVQSYMIVLMVTCCITPLSII VLCYLQVWL AIRAVAKQQKESESTQKA EKVTRM VVV MV	274
		+ Y L + C I PLS+I Y Q+ A+RAVA QQ+ES STQKAE+EV+ M VV MV	
Sbjct	197	YSEYYTWFLPIFCYIVPLSLICFSYSQ LL GALRAVAAQQQESASTQKA EREVSHM VVV MV	256
Query	275	LAFCPCWG PYAFFACFAA ANPGYPFHPLMAALPAFFAKSATIYNPV IYVFMNRQFRNCIL	334
		+FC C+ PYA A + N + +PAFF+KSA +YNP+IY FMN+QFR CI+	
Sbjct	257	GSFCLCYTPYA ALAMYIVNNRNHGVDLRLVTIPAFFSKSACVYNPIIYCFMN KQFRACIM	316
Query	335	QLF-GKKVDDGSELSSASKTEVSSVSS	360
		++ GK + D SELSS+ KTEVS+VSS	
Sbjct	317	EMVC GKPMTDE SELSSSQKTEVSTVSS	343

Chamada de variantes



VCF

Format Specification

Metadata

Column Headings

1st Row of Data

maize_sub5k.vcf - Notepad

File Edit Format View Help

```
##fileformat=VCFv4.1
##FORMAT<ID=GT,Number=1>Type=String>Description="Genotype"
##INFO<ID=DP,Number=1>Type=Integer>Description="Total Depth"
##INFO<ID=NZ,Number=1>Type=Integer>Description="Number of taxa with data"
##INFO<ID=AD,Number=.,Type=Integer>Description="Total allelic depths in order listed"
##INFO<ID=AN,Number=.,Type=Integer>Description="Total number of alleles in order listed"
##INFO<ID=AQ,Number=.,Type=Integer>Description="Average phred base quality for alleles in order listed"
##INFO<ID=GN,Number=.,Type=Integer>Description="Number of taxa with genotypes AA,AB,BB or AA,AB,AC,BB,BC,CC"
##INFO<ID=HT,Number=1>Type=Integer>Description="Number of heterozygotes"
##INFO<ID=EF,Number=1>Type=Float>Description="EF=het_frequency/(presence_frequency * minor_allele_frequency)"
##INFO<ID=PV,Number=.,Type=Float>Description="p-value from segregation test between AB or AB, AC, BC if 2 al"
##INFO<ID=MAF,Number=1>Type=Float>Description="Minor allele frequency"
##INFO<ID=MAF0,Number=1>Type=Float>Description="Minor allele frequency from unimputed HapMap 3.2.1 on 1210 t"
##INFO<ID=IBD1,Number=0>Type=Flag>Description="only one allele present in IBD contrasts; based on 916 taxa o"
##INFO<ID=LLD,Number=0>Type=Flag>Description="Site in local LD with GBS map (from 916 taxa of HapMap 3.1.1)"
##INFO<ID=NI5,Number=0>Type=Flag>Description="Site with 5bp of a putative indel from 916 taxa of HapMap 3.1."
##INFO<ID=INHMP311,Number=0>Type=Flag>Description="Site peresent in HapMap3.1.1"
##INFO<ID=ImpHomoAccuracy,Number=1>Type=Float>Description="Fraction of homozygotes imputed back into homozyg"
##INFO<ID=ImpMinorAccuracy,Number=1>Type=Float>Description="Fraction of minor allele homozygotes imputed bac"
##INFO<ID=DUP,Number=0>Type=Flag>Description="Site with heterozygotes frequency > 3%"
##ALT<ID=DEL>Description="Deletion"
##ALT<ID=INS>Description="Insertion"
##Comment="Genotypes called on HapMap3.2.1 sites, some INFO parameters - see descriptions - taken directly fr
##HapMapVersion="3.2.1"
```

CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	282set_A188	282set_A214NGoodman-B
87	282set_CML311	282set_CML314	282set_CML321	282set_CML322	282set_CML323	282set_CML328	282se			
1W	282set_MS1334	282set_MS153	282set_Mo17	282set_Mo18W	282set_Mo24W	282set_Mo44	282se			
	82set_SC55	282set_SD40	282set_SD44	282set_Sg1533	282set_Sg18	282set_T232	282set_T234			
10	267893	.	C	T	.	.	DP=1796;NZ=256;AD=1792,4;AN=511,1;AQ=32,30;GN=255,1,0			
0	0/0:3,0 0/0:10,0		0/0:4,0	0/0:14,0			0/0:5,0 0/0:5,0 0/0:3,0 0/0:6,0 0/0:4,0 0/0:9,0 0/0:5			
,	0/0:5,0 0/0:2,0 0/0:14,0			0/0:10,0			0/0:3,0 0/0:7,0 0/0:8,0 0/0:5,0 0/0:9,0 0/0:15,0			
10	299244	.	C	A	.	.	DP=1192;NZ=247;AD=1189,3;AN=493,1;AQ=33,34;GN=246,1,0			
,	0/0:5,0 0/0:7,0 0/0:4,0 0/0:3,0 0/0:5,0 0/0:6,0 0/0:4,0 0/0:3,0 0/0:2,0 0/0:4,0 0/0:2,0 0/0:8,0 .									
0:6,0	0/0:12,0		0/0:5,0 0/0:1,0 0/0:1,0 0/0:5,0 0/0:3,0 0/0:4,0 0/0:4,0 0/0:6,0 0/0:1,0 0/0:7,0 0/0:2							

Ln 1, Col 1 100% Windows (CRLF) UTF-8



All Cases > [REDACTED] PP_1

857701 [REDACTED] XTPP_1

Hyb Panel v2 | Batch: Jun 26, 2024 | Assignee:

@dasa.com.br

Status: Resolved

Follow Case

Analysis Results

6 Clinically relevant variants

Quality Control

All Samples Passed QC

Workbench

- Additional Information
UPD: No UPD was detected

Variants

- Variants
6 clinically relevant variants detected

ROH | Add variant

Coverage Report

VUS to Report (1)

★ Congenital heart defects, multiple types, 8, with or without heterotaxy (Autosomal Dominant) was found to have a Medium connection to... ★ +1 more

Report Preview

Gene	Allele	Frequency	Internal	Community	Confidence	Prediction	Inheritance	Details
SMAD2	Heterozygote (M) p.R330del c.987_989del	N/A	N/A undefined Hom	N/A	High AB: 53.38%	Benign	AD 8 Conditions	VUS NR ... WB R

Chr18:47848482-CCTT-C | NM_005901.6 | Non Frameshift | Exon 8

Waiting for assessment

Causal Variants

Secondary Findings

Not Relevant

CASE DATA



Patient Inform

[REDACTED]
Patient information missing

Add / Edit Infor

Phenotypes

Click the tag to disable phenotype
Cardiovascular

Abnormal ventric

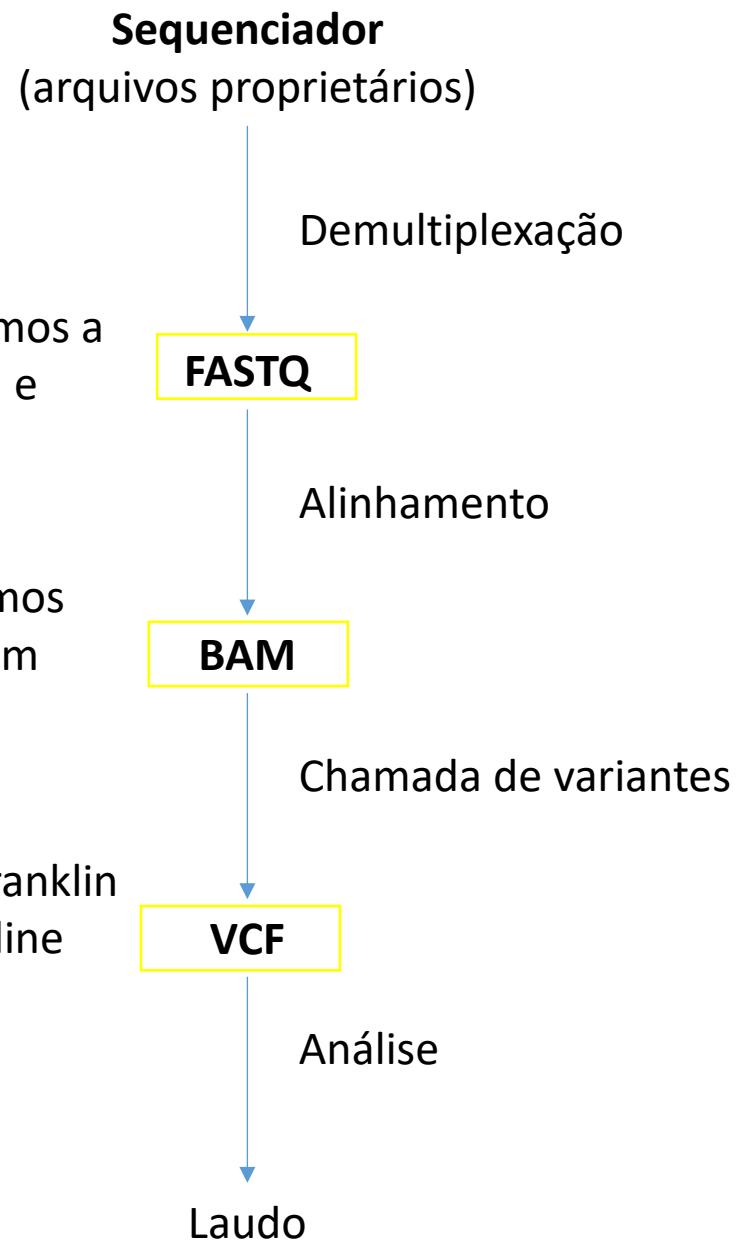
Pulmonary valve a

Dados Brutos

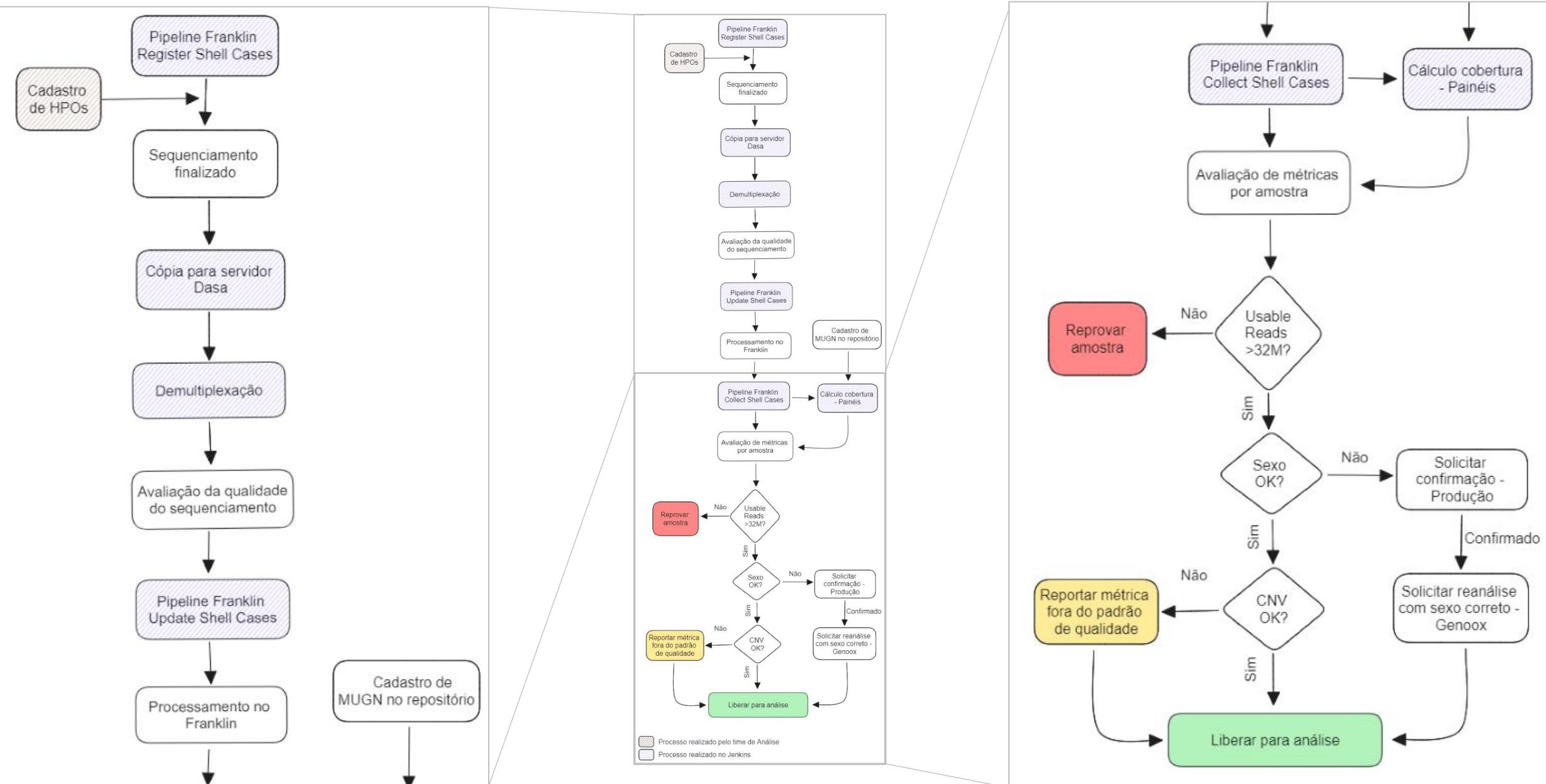
Se estiverem interessados somente no **FASTQ**, faremos a demultiplexação usando o software da própria MGI e transferência para o cliente

Se estiverem interessados somente no **BAM**, podemos considerar realizar o alinhamento internamente (sem utilizar Franklin)

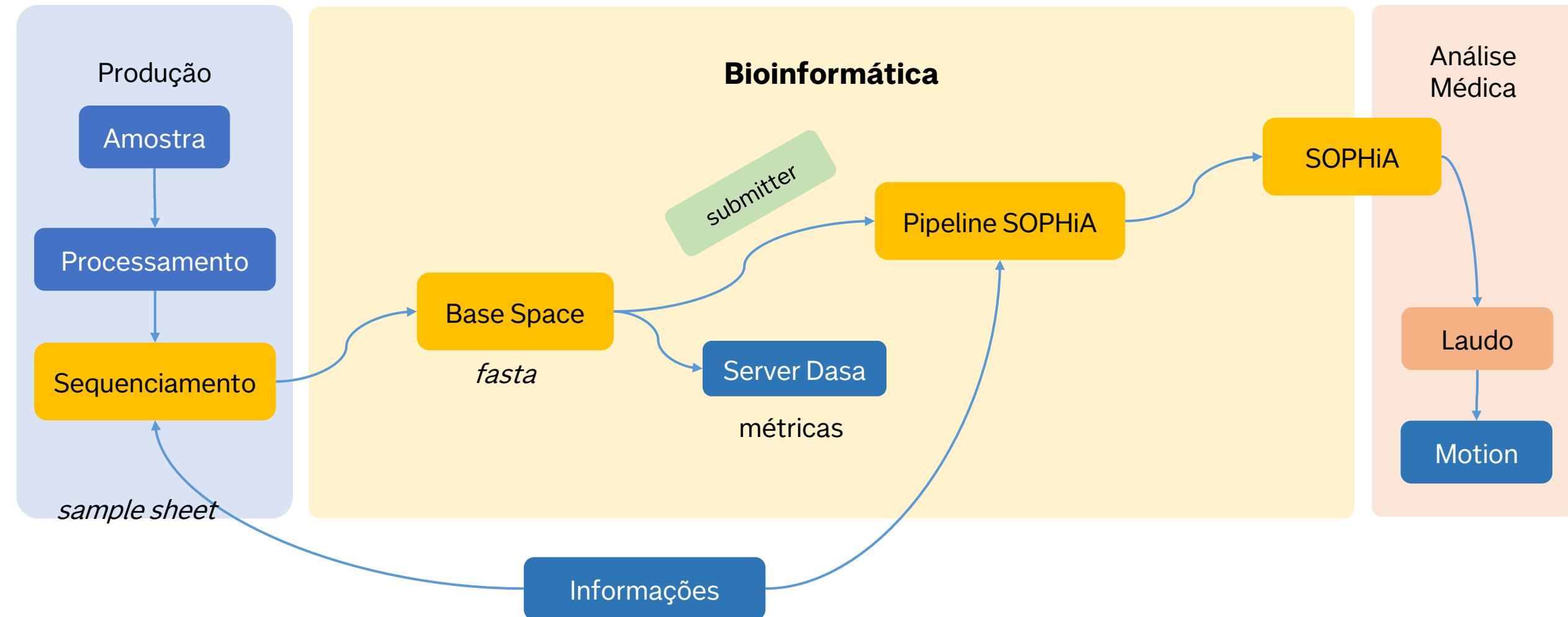
Se estiverem interessados no **VCF**, utilizaremos o Franklin para Genoma e Exoma. Para Exoma Tumoral o pipeline precisa ser desenvolvido.



Fluxograma Rotina Exoma IDT



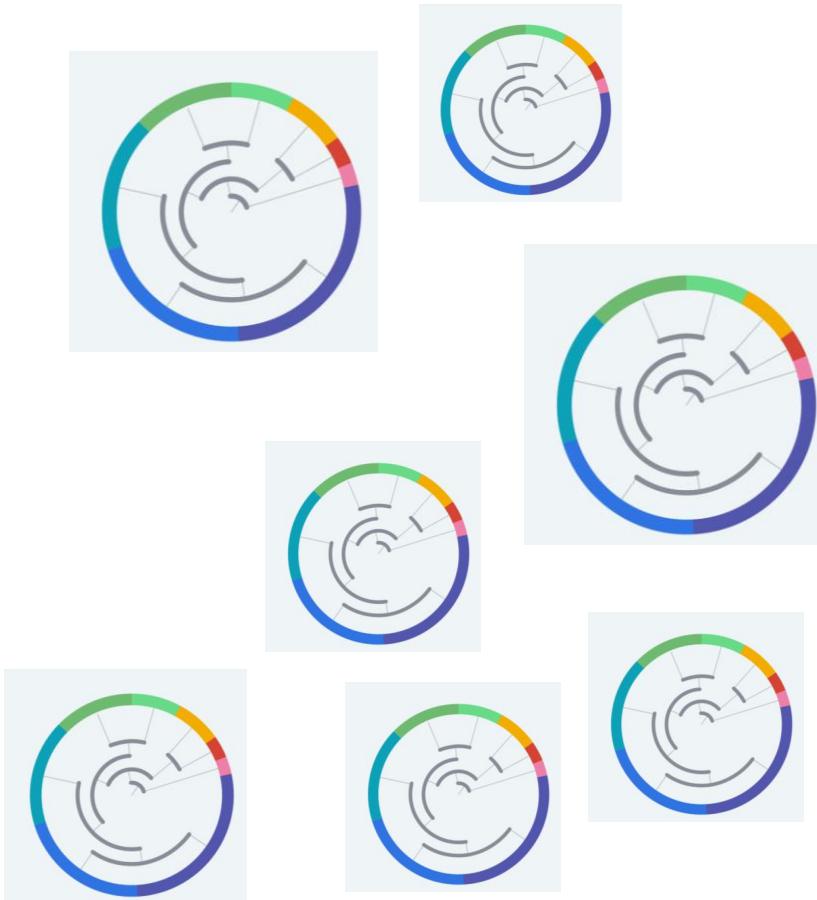
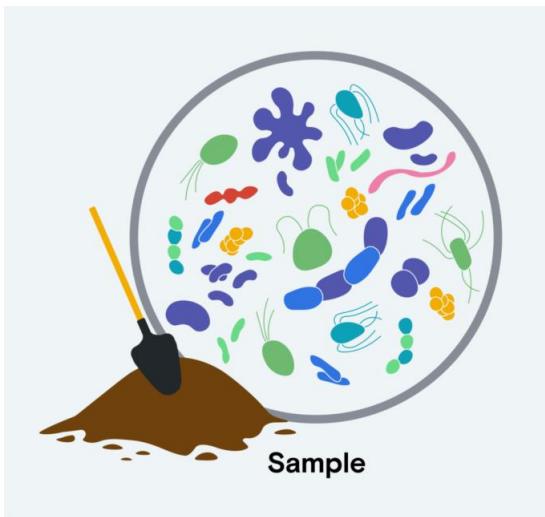
SOPHiA



Exemplos de outros tipos projetos na bioinformática

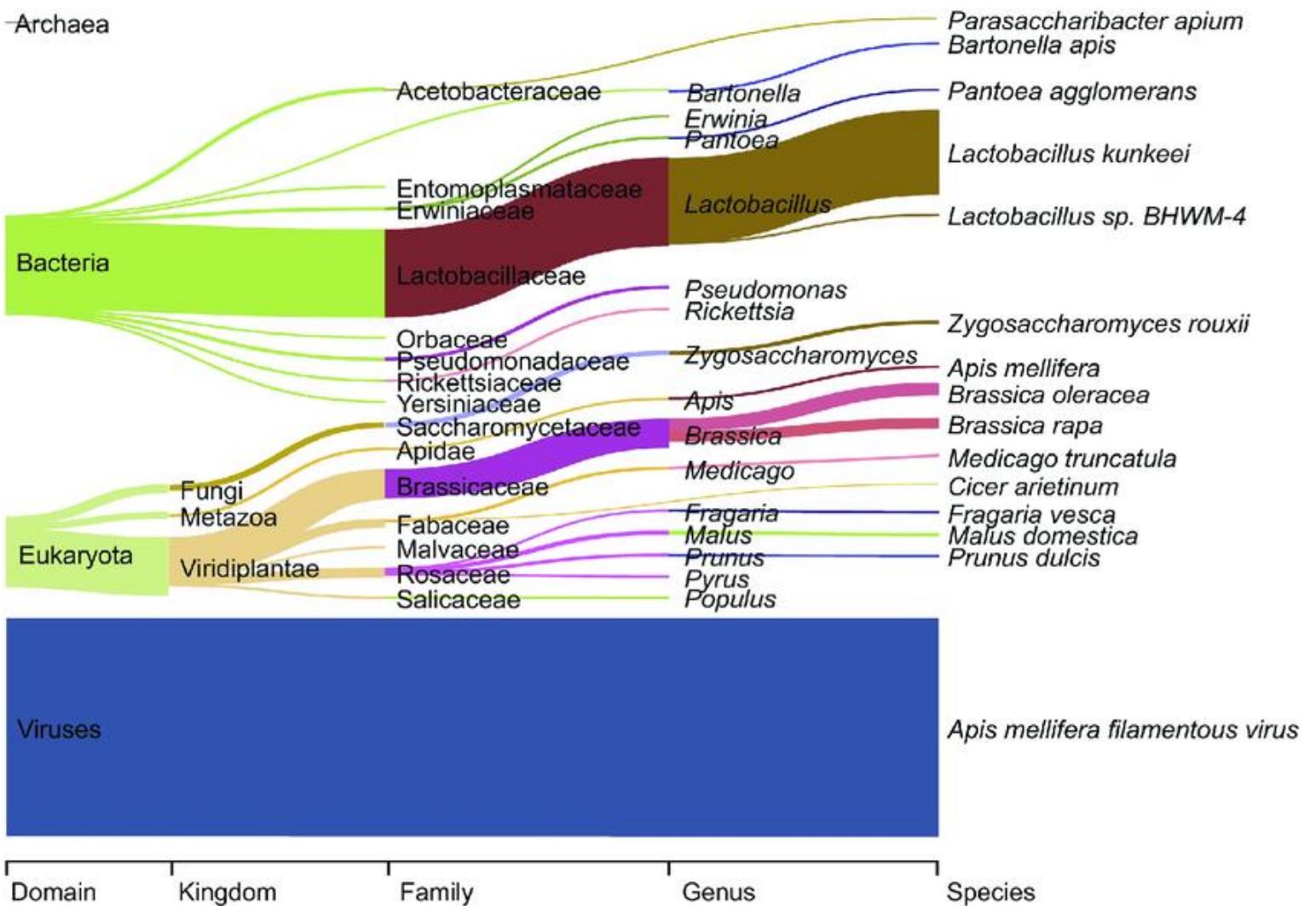
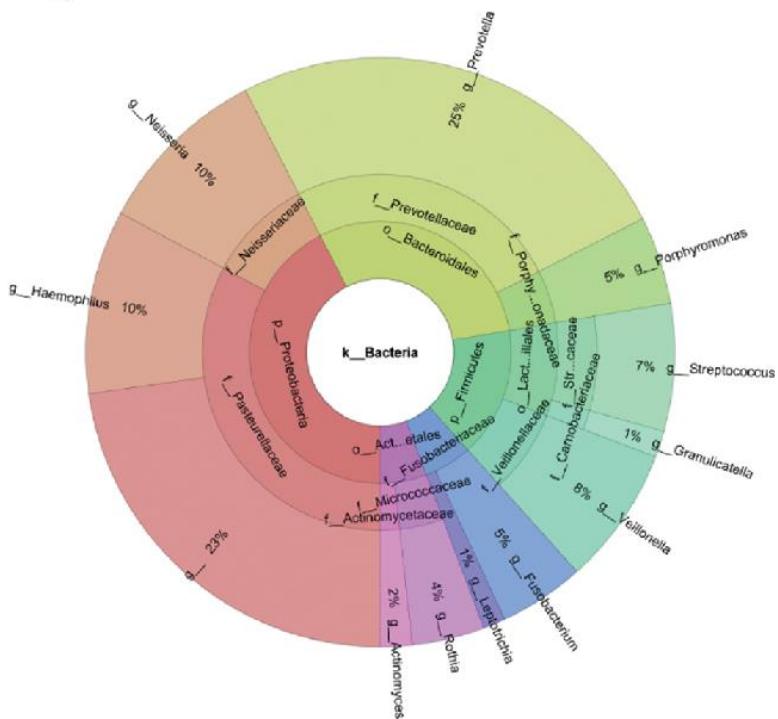
- Estudos de associação genômica ampla (GWAS)
- Risco poligênico (PRS)
- Genômica comparativa
- Farmacogenomica
- Multi-Omicos (ex. proteômica, epigenômica...)
- Metagenômica
- Analises estruturais
- Desenvolvimento de ferramentas e algoritmos
- Bancos de dados
- Biologia de sistemas
- Drug discovery

Metagenômica

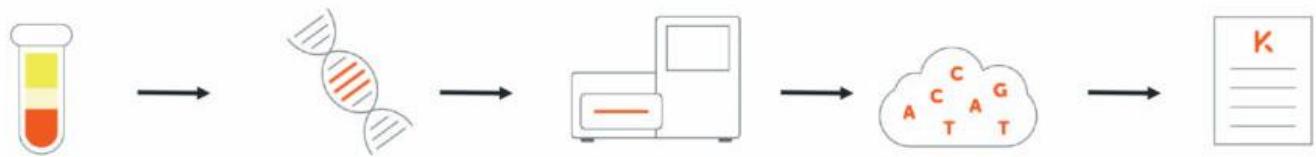


Metagenômica

B



Metagenômica



- Não necessita cultura ou isolamento de colônias
- Possível realizar utilizando cfDNA isolado do plasma
 - Mesma amostra pode ser utilizada para biopsias líquidas em contexto oncológico.
- Possibilidade de resultados conclusivos em menos de 48h
- Resultados já incluem perfil de virulência e resistência a anti-microbianos
- Principais Desvantagens: 1) Custo; 2) não cobertura por planos de saúde.

Farmacogenética e Farmacogenômica

- A farmacogenética e a farmacogenômica são campos inter-relacionados que estudam a interação entre medicamentos e o material genético.
- Embora frequentemente usados de forma intercambiável, eles têm distinções importantes:
 - Farmacogenética: Personalizar a terapia medicamentosa com base no perfil genético do paciente para maximizar a eficácia e minimizar os efeitos colaterais.
 - Farmacogenômica: Desenvolver medicamentos mais eficazes e seguros ao considerar a variabilidade genética entre diferentes populações.

PharmOne – medicamentos avaliados

Psiquiatria e neurologia

- alpazolam
- amitriptilina
- anfetamina
- aripiprazol
- atomoxetina
- brexipipraxol
- buspirona
- citalopram
- clobazam
- clomipramina
- clonazepam
- clozapina
- desipramina
- dexmetilfenidato
- dextroanfetamina
- dextrometorfano
- quinidina
- diazepam
- donepezila
- doxepina
- duloxetina
- escitalopram
- eszopiclona
- fenitoína
- fluoxetina
- flupentixol
- fluvoxamina
- galantamina
- guanfacina
- haloperidol
- iloperidona
- imipramina
- lisdexanfetamina
- metilfenidato
- mirtazapina
- moclobemida
- nortriptilina
- olanzapina
- paroxetina
- perfenazina
- pimozida
- protriptilina
- quetiapina
- risperidona
- sertralina
- tetrabenazina
- tioridazina
- trazodona
- venlafaxina
- vortioxetina
- zuclopentixol

Cardiologia

- acenocumarol
- atorvastatina
- carvedilol
- clopidogrel
- flecainida
- metoprolol
- nebivolol
- prasugrel
- propafenona
- propanolol
- simvastatina
- ticagrelor
- timolol
- varfarina

Endocrinologia

- anticoncepcionais com estrógeno
- eligustate
- glibenclamida
- gliclazida
- glimepirida
- saxagliptina
- tolbutamida

Dor

- buprenorfina
- carisoprodol
- celecoxibe
- codeína
- diclofenaco
- fentanila
- flurbiprofeno
- hidrocodona
- meloxicam
- oxicodona
- tramadol

Demais terapias

- cevimelina
- ciclosporina
- eliglustate
- fesoterodina
- gefitinibe
- mirabegrona
- ondansetrona
- sirolimo
- tacrolimo
- tamoxifeno
- tansulosina
- tolterodina
- tropisetrona

Gastroenterologia

- dexlansoprazol
- esomeprazol
- lansoprazol
- omeprazol
- pantoprazol
- rabeprazol

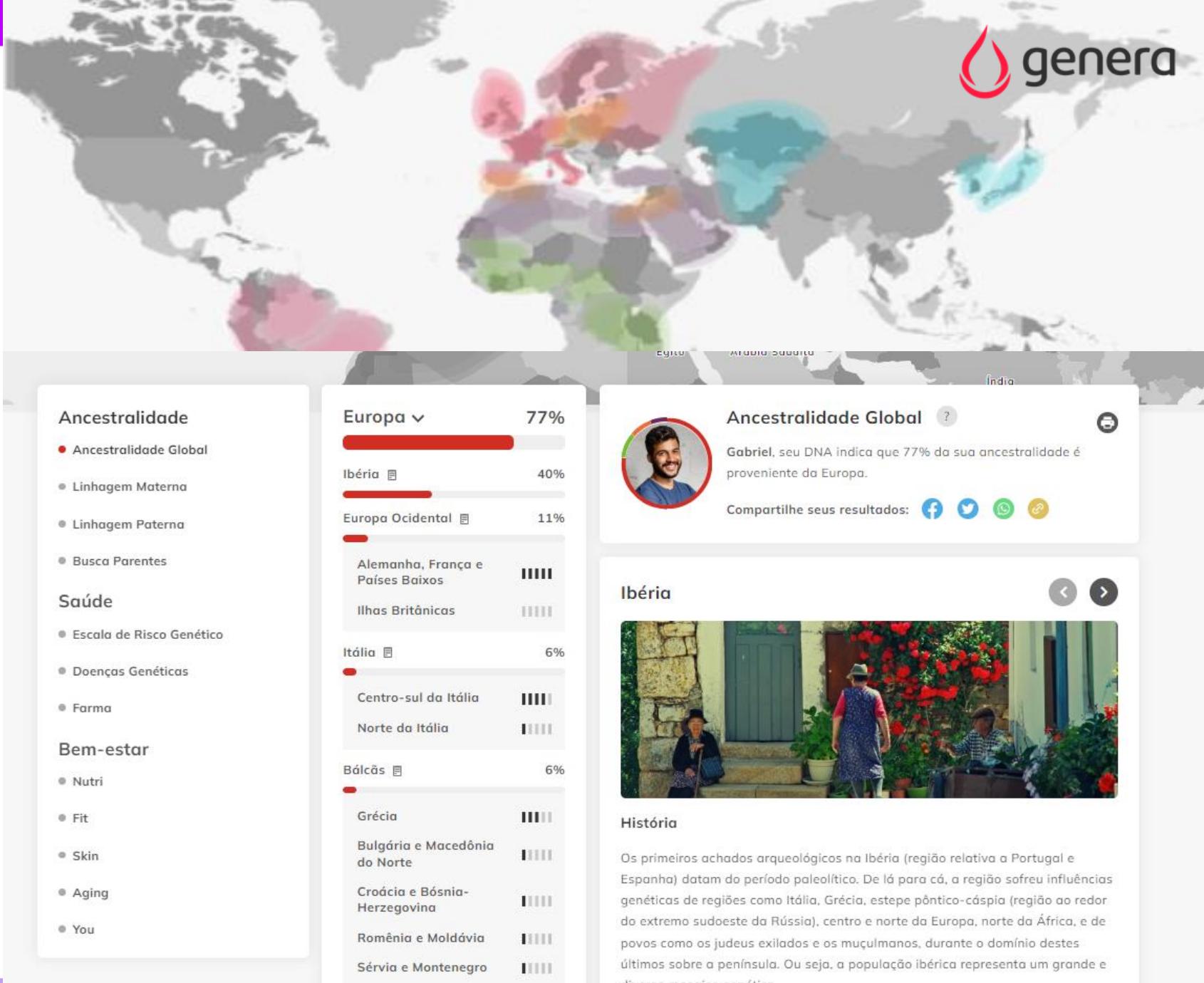
Antifúngicos

- cetoconazol
- voriconazol

Genera

Primeiro laboratório de genética a oferecer uma plataforma de testes de Ancestralidade e de Saúde personalizada.

<https://descubra.genera.com.br/ancestralidade>



Ancestralidade

- Ancestralidade Global
- Linhagem Materna
- Linhagem Paterna
- Busca Parentes

Saúde

- Escala de Risco Genético
- Doenças Genéticas
- Farma

Bem-estar

- Nutri
- Fit
- Skin
- Aging
- You



Resultados

Cuidados relevantes

Atorvastatina (Lípitor®)

Predisposição para menor taxa de resposta ao tratamento

Hidroclorotiazida (Clorana®)

Predisposição para menor taxa de resposta ao tratamento

Nicotina

Sem predisposição para abandonar o hábito de fumar com reposição de nicotina

Opioides

Predisposição para menor taxa de resposta ao tratamento

Salbutamol (Aerolin®) e Salmeterol (Serevent®)

Predisposição para menor taxa de resposta ao tratamento

Sildenafil (Viagra®)

Predisposição para menor taxa de resposta ao tratamento

Atorvastatina (Lípitor®)

A atorvastatina faz parte de uma classe de medicamentos chamada de estatinas. É frequentemente utilizada para diminuir os níveis de colesterol LDL no sangue, conhecido também como "colesterol ruim", e para a prevenção da aterosclerose, que é o acúmulo de placas de gorduras nos vasos sanguíneos chamados de artérias. Há, no mercado, diversos medicamentos que têm a atorvastatina como princípio ativo, tais como Atorless®, Citalor®, Kolevas®, Lipigran®, Lipistat®, Lípitor® e medicamentos genéricos com a atorvastatina cálcica.

Entendendo seu resultado

O gene chamado *APOE*, localizado no cromossomo 19, desempenha um importante papel no transporte e na eliminação de colesterol no sangue. As pessoas que possuem o alelo T no marcador rs7412 nesse gene apresentam predisposição para maior taxa de resposta ao tratamento com atorvastatina.

SNP

rs7412

Resultado conforme genótipos

c,c

c,t

t,t

Nível de evidência	Moderado
Gene	<i>APOE</i>
Seu genótipo	c,c
População de estudo: Asiática, europeia e norte-americana	

Seu resultado
Predisposição para menor taxa de resposta ao tratamento

Ancestralidade

- Ancestralidade Global
- Linhagem Materna
- Linhagem Paterna
- Busca Parentes

Saúde

● Escala de Risco Genético

- Doenças Genéticas
- Farma

Bem-estar

- Nutri
- Fit
- Skin

● Câncer colorretal

Risco de desenvolver a doença até os 95 anos



Risco aumentado*

5,40%

Entre 4,71% e 5,89%

*Em relação ao risco médio da população brasileira

● Colite ulcerativa

Risco de desenvolver a doença até os 95 anos



Risco aumentado*

0,62%

Entre 0,48% e 0,81%

*Em relação ao risco médio da população brasileira

● Diabetes tipo 2

Risco de desenvolver a doença até os 95 anos



Risco aumentado*

38,90%

Entre 30,79% e 46,90%

*Em relação ao risco médio da população brasileira

● Doença celíaca

Risco de desenvolver a doença até os 95 anos



Risco aumentado*

0,87%

Entre 0,44% e 0,97%

● Doença de Crohn

Risco de desenvolver a doença até os 95 anos



Risco aumentado*

0,39%

Entre 0,30% e 0,50%

● Doença intestinal inflamatória

Risco de desenvolver a doença até os 95 anos



Risco aumentado*

0,69%

Entre 0,53% e 0,89%



NIPT (*Non-Invasive Prenatal Test*)

- Trissomias:
 - Patau's syndrome (Chr 13)
 - Edwards' syndrome (Chr 18)
 - Down's syndrome (Chr 21)
- Triploidia (trissomia de todos Chrs)
- Aneuploidia do Cromossomos sexuais:
 - XO / Monosity X (Turner's)
 - XXY (Klinefelter Syndrome)
 - XYY (Jacob's Syndrome)
 - XXX (Triple X)
- 22q11.2 microdeleção



Diversos outros exames envolvem processos de bioinformática...

PGT-A
HRD
CARIÓTIPO
EXOMA
CGH ARRAY
NIPT AMPLIADO
IMUNOFENOTIPAGEM
NIPT BÁSICO
CÂNCER HEREDITÁRIO - NGS
MAOV
HRR
BRCA1E2
PHARMAONE

GUARDANT
ENDOPREDICT
SIGNATERA
METHYLBRAIN
HSL500
FOUNDATION
RETINOPATIA
TARGET1
PAINEL NGS PARA CARDIOMIOPATIAS
MICROBIOMA VAGINAL
PCGT
BABYGENES
METHYLSARC
NIPT MULTIGENES



Exemplos de outros tipos projetos na bioinformática

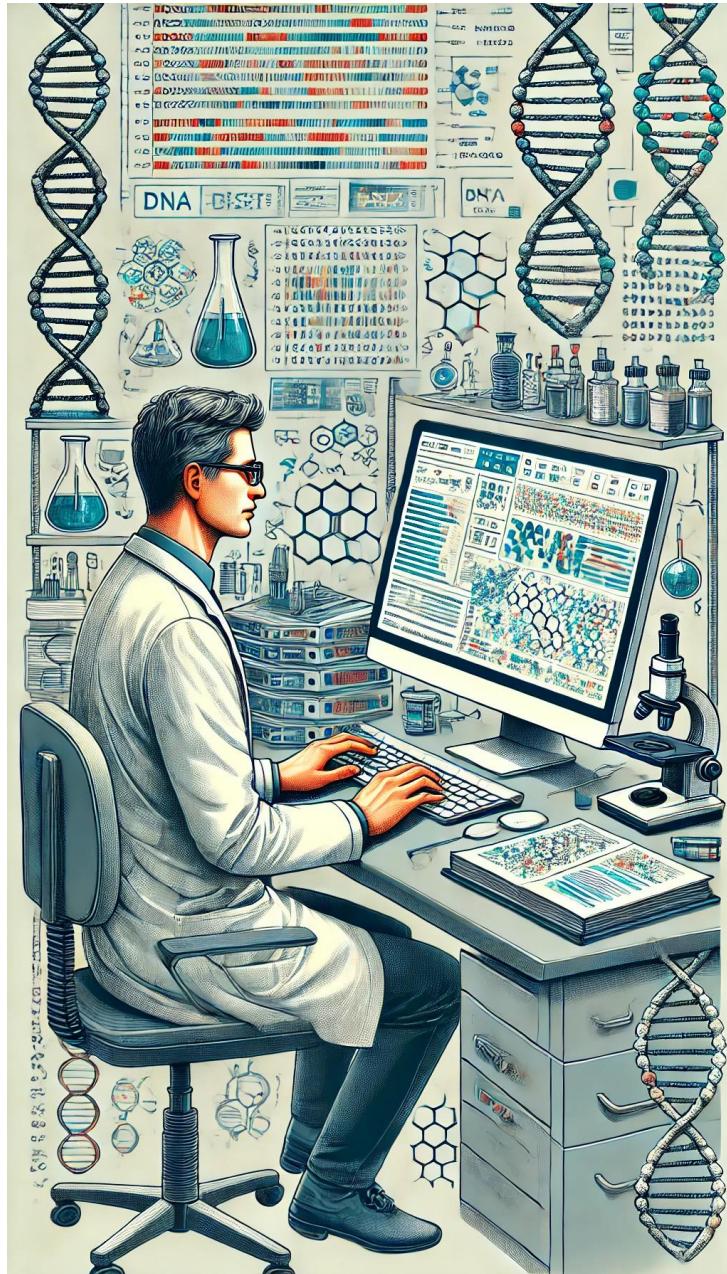
- Estudos de associação genômica ampla (GWAS)
- Risco poligênico (PRS)
- Genômica comparativa
- Multi-Omicos (ex. proteômica, epigenômica...)
- Analises estruturais
- Desenvolvimento de ferramentas e algoritmos
- Bancos de dados
- Biologia de sistemas
- Drug discovery

A bioinformática é diversa!

- Campo interdisciplinar
- Enorme variedade de projetos
 - Requerem diversas combinações de skills diferentes
- Muitos caminhos diferentes podem formar um “bioinformata”
 - Farmacêuticos, biólogos, engenheiros cientistas de computação/TI, físicos, matemáticos...

Equipe de bioinformática Dasa

- 5x Biólogos (5x)
- 3x Farmacêuticos
- 3x Cientistas da Computação/TI
- 3x Biomédicos
- 2x Engenheiros de Biotecnologia / Bioprocessos
- 2x Informática Biomédica
- 1x Engenheiro Químico
- ...



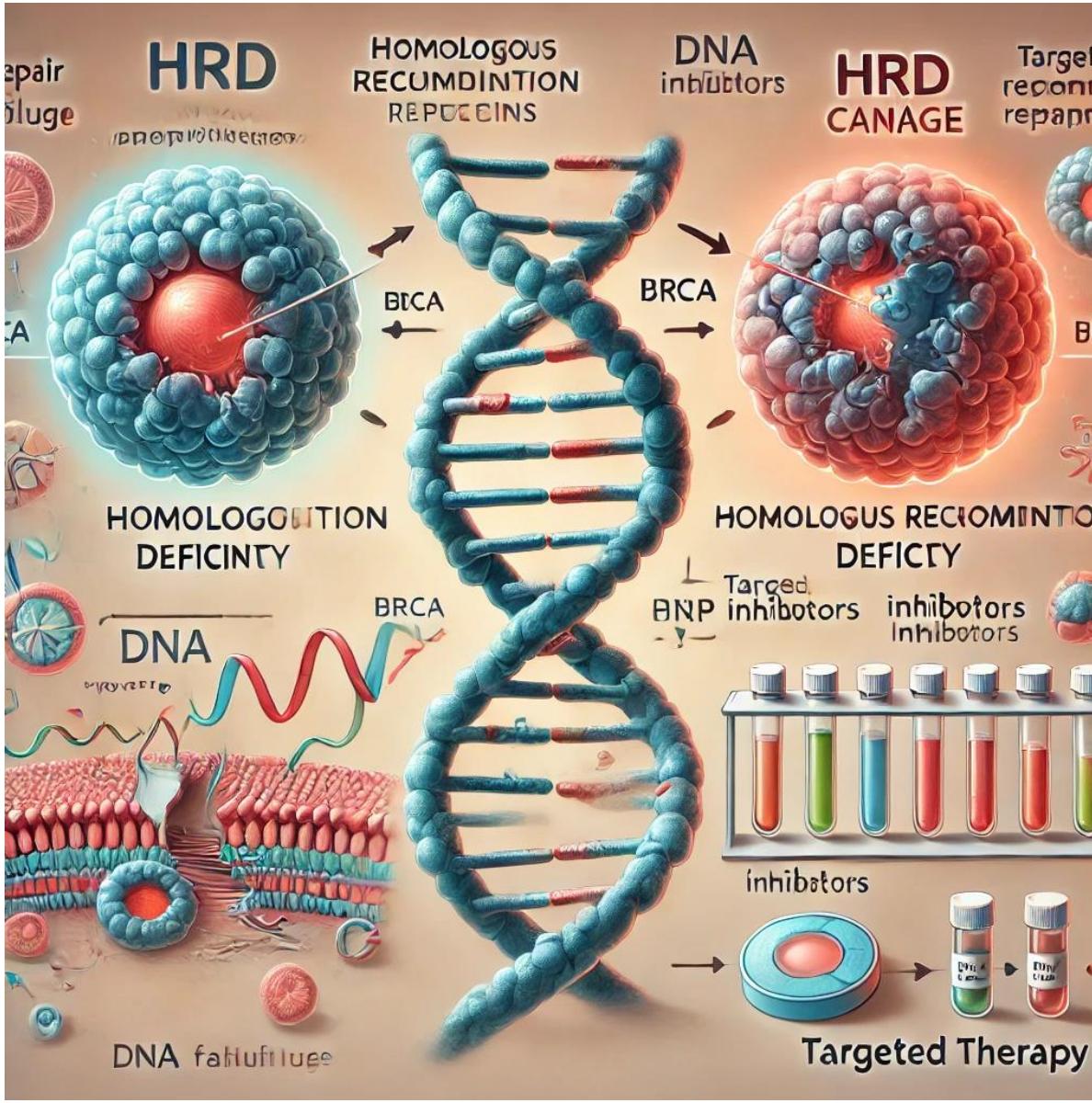
Skill set típico

- Conhecimentos em biologia molecular, genética, bioquímica...
- Linguagens de programação: Python, Bash
 - Pacotes para computação científica e estatística: R, Matlab
- Pipelines: snakemake, WDL, nextflow
- Estruturas de dados, bancos de dados, AI
- (Bio)Estatística, HPC e métodos para análise de dados em grande escala
- Soft skills: comunicação, escrita científica, resiliência!
- Não existe fórmula ou um único caminho para a bioinformática!



HRD

by ChatGPT

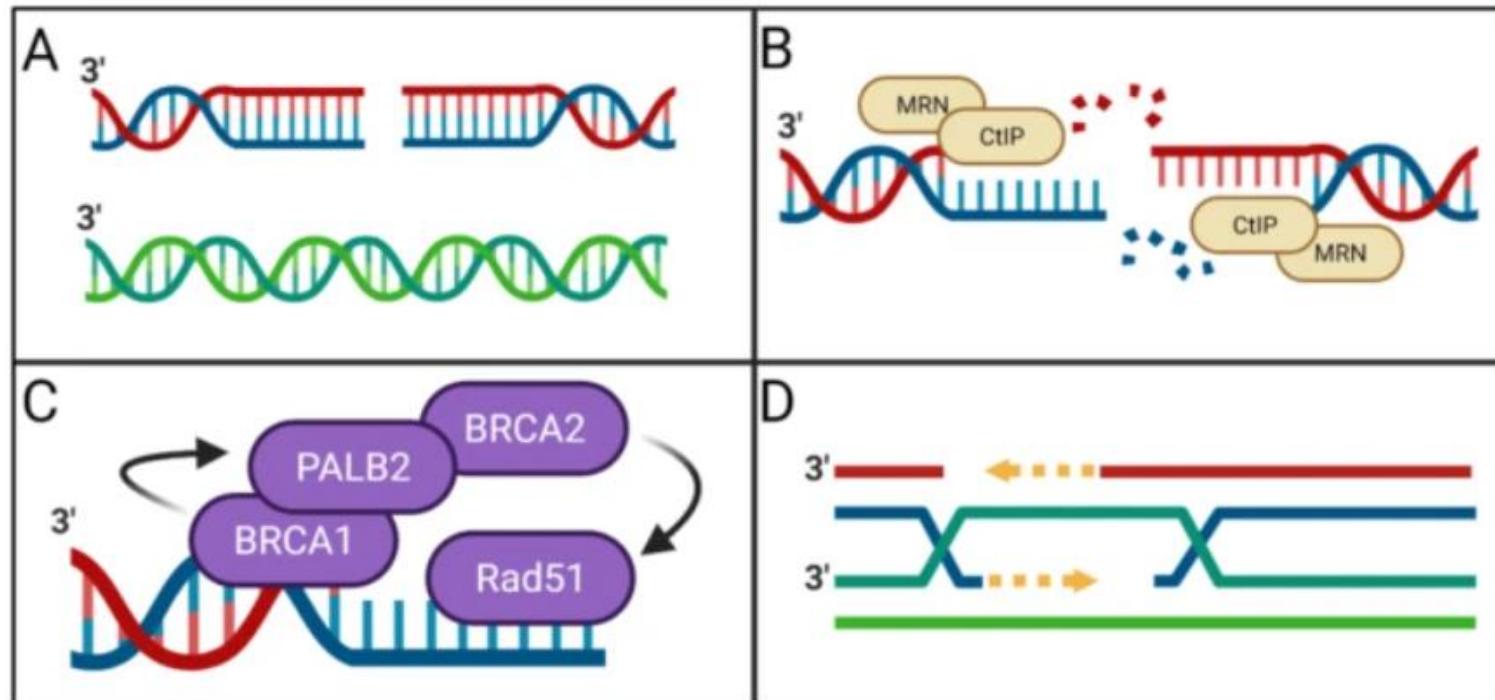


HRD

by ChatGPT

Homologous Recombination Deficiency (HRD)

- Inabilidade da célula em reparar quebras no DNA de fita dupla (dsDNA) utilizando genes da vida de Reparo por Recombinação Homóloga (HRR)



Creeden, 2021

Homologous Recombination Deficiency (HRD)

- SOPHiA *Homologous Recombination (Deficiency) Solution*
- **HRS** 16 HRR genes:
 - ATM, BARD1, **BRCA1, BRCA2**, BRIP1, CDK12, CHEK1, CHEK2, FANCL, PALB2, PPP2R2A, RAD51B, RAD51C, RAD51D, RAD54L, TP53.
- **HRD** 28 genes (14 genes do HRS + 14 genes):
 - **AKT1, ATM, BARD1, BRCA1, BRCA2, BRIP1, CCNE1, CDK12, CHEK1, CHEK2, ESR1, FANCA, FANCD2, FANCL, FGFR1, FGFR2, FGFR3, MRE11, NBN, PALB2, PIK3CA, PPP2R2A, PTEN, RAD51B, RAD51C, RAD51D, RAD54L, TP53**
 - **IpWGS** (genoma completo baixa profundidade, sem captura) para score de Instabilidade Genômica (**GI score**)

Homologous Recombination Deficiency (HRD)

- HRD é uma condição em que a via de reparo de DNA através da recombinação homóloga está comprometida, levando à instabilidade genômica
- Definição do status HRD tem implicações diretas no prognostico e nas drogas de escolha para tratamento!
 - HRD+ → Suscetibilidade a tratamentos com inibidores de PARP, certas drogas de platina
- Teste recomendado para tumores de ovário e mama
 - Uso adicional em casos específicos de câncer de próstata e pâncreas.

Como funciona o teste de HRD

- Testes realizados:
 - Analise de 28 biomarcadores (genes)
 - Sequenciamento com enriquecimento por captura hibrida
 - Analise de sequenciamento de genoma completo de baixa cobertura (IpWGS)
- Métricas avaliadas:
 - Status HRD
 - Derivado do GI Score (*Genome Instability Score*)
 - Mutações nos biomarcadores
 - Outros indicadores de instabilidade genômica
 - large-scale state transitions (LST), loss of heterozygosity (LOH) e telomeric allelic imbalance (TAI).

Homologous Recombination Deficiency (HRD)

Análise de mutações (SNVs e indels)
nos genes do painel de HRR

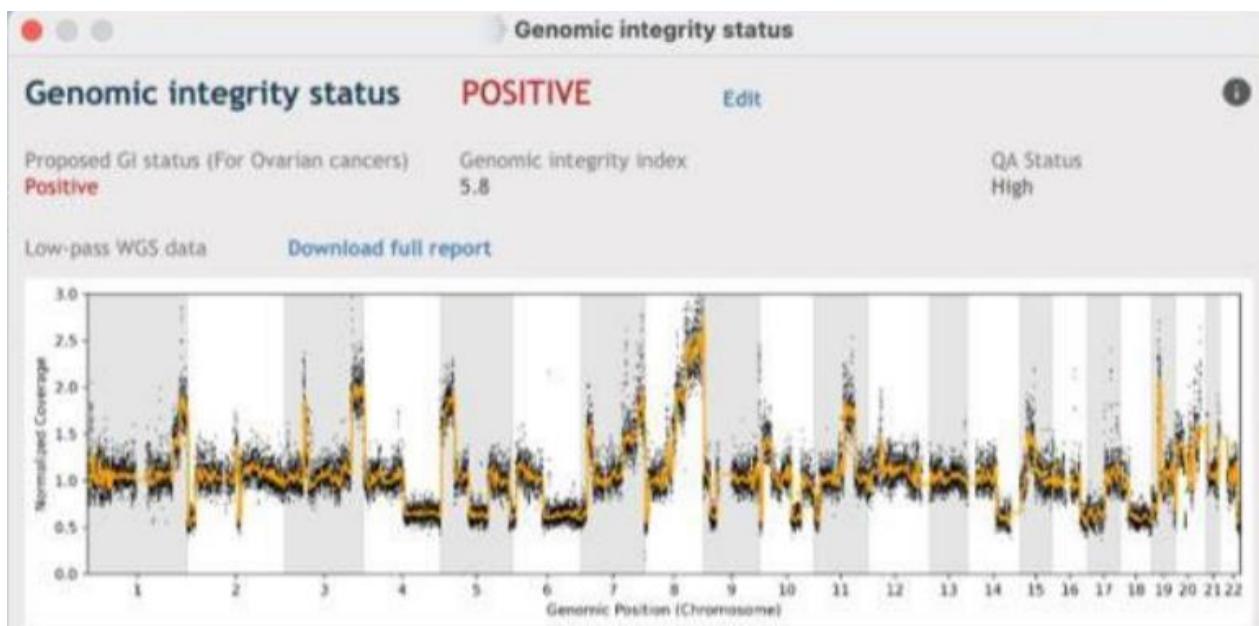
Análise de cicatrizes do genoma
(CNVs de larga escala) por lpWGS

Reported Variants

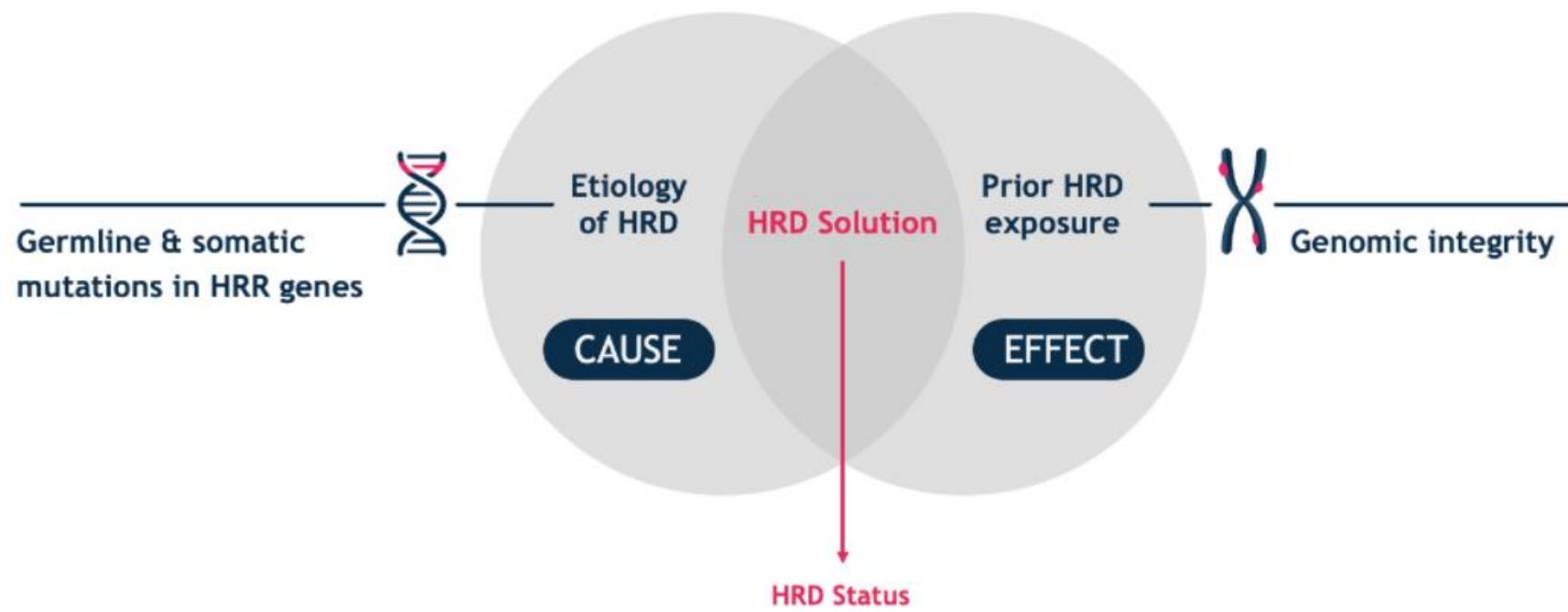
BRCA VARIANTS

In Report	Type	Gene(s)	Chromosome	Consequence	Details	BRCA pathogenicity	BRCA score
BRCA	INDEL	BRCA1	chr17	frameshift	c.1112del	Pathogenic	Positive
	SNP	BRCA1	chr17	missense	c.1638G>A	unclear	Undetermined
	SNP	BRCA1	chr17	missense	c.1112C>T		Undetermined
	SNP	BRCA2	chr13	missense	c.5713C>T		Undetermined
	SNP	BRCA2	chr13	splice_donor_->+4	c.7805+4C>T	unclear	Complex case
	SNP	BRCA2	chr13	missense	c.8182G>A	Benign	Negative
	SNP	BRCA2	chr13	missense	c.77C>T	unclear	Undetermined
	SNP	BRCA2	chr13	missense	c.5670G>A		Undetermined
	SNP	BRCA2	chr13	missense	c.5501G>A		Undetermined
	SNP	BRCA2	chr13	missense	c.3733G>A		Undetermined
	SNP	BRCA2	chr13	missense	c.2875G>A	unclear	Undetermined
	SNP	BRCA2	chr13	missense	c.8314G>A	unclear	Undetermined
	SNP	BRCA2	chr13	missense	c.8611G>A	unclear	Undetermined
	SNP	BRCA2	chr13	intronic	c.7806-14T>C	Benign	Negative
	INDEL	BRCA1	chr17	intronic	c.548-58del	Benign	Negative

Close



Status de HRD: positivo ou negativo

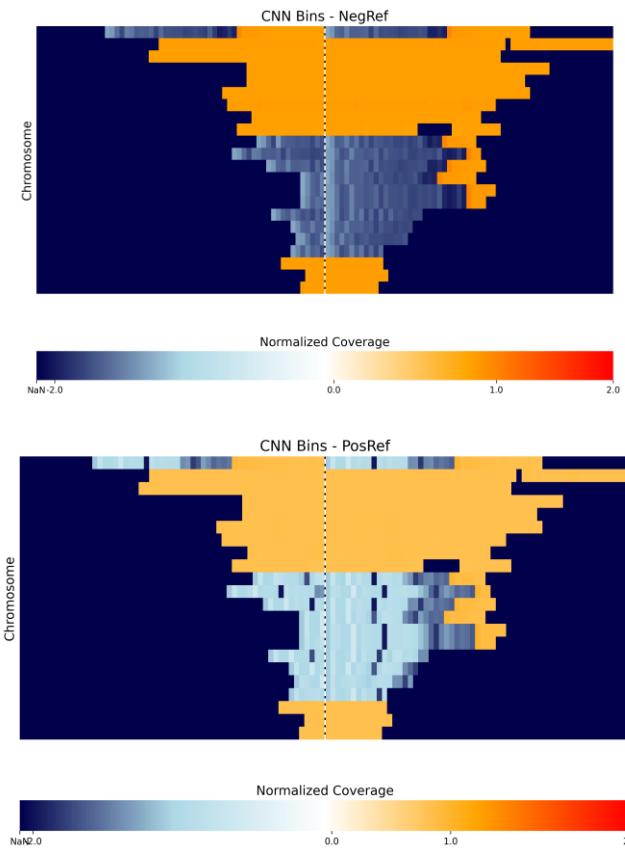


Pré-processamento dos dados para a Sprint

- Downsampling de dados de sequenciamento completo
 - Redução de cobertura total (>30X) para “low-pass” (3X)
- Alinhamento contra genoma de referencia (hg38)
- Calculo de cobertura em bins de 3 Mbp (mega base-pairs)
- Normalização de métricas
- Heatmaps

Dataset para o Sprint

- 554 amostras de genoma completo de baixa cobertura (*lpWGS**)
 - 301 Positivas
 - 253 Negativas
- Heatmap de cobertura genômica
- Dados de cobertura em formato TSV
- Objetivo: cálculo do status HRD/GI Score através de uma CNN implementada em Python



**Low-pass Whole Genome Sequencing*

Dasa
genómica

genera

