The Whole Genome Sequencing (WGS)

Introduction-Définitions

Le séquençage du génome entier est une technologie relativement nouvelle qui nous permet de "lire" le code génétique entier d'une personne ou d'un organisme.

Quel est son fonctionnement ?

Quel sens faut –il donner à la vaste quantité d'informations que le séquençage du génome entier fournit ?

Et comment cette nouvelle technologie de séquençage sera-t-elle utilisée au profit des patients?

Le processus de séquençage du génome entier (WGS)

- Le WGS est une procédure de laboratoire qui détermine l'ordre des bases dans le génome d’un organisme dans un seul processus.
- Le WGS fournit une empreinte ADN très précise qui peut aider de relier les cas entre eux, ce qui permet de détecter et de résoudre plus rapidement une épidémie.

Intérêts du WGS

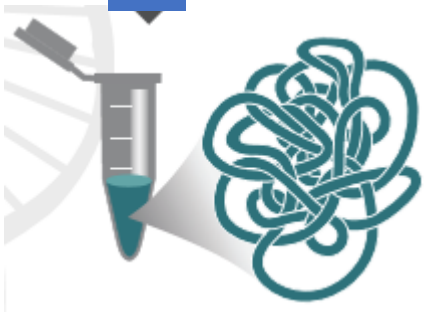
- **Grâce à un algorithme de recherche de gènes, nous pouvons découvrir une partie importante des gènes**
- **Comprendre la structure d'un génome**
- **Comprendre l'évolution du génome- Recherche de gènes associés à maladies**

Séquençage de l’exon (Alternative)

- Le séquençage du génome entier est une approche de force brute pour résoudre les problèmes lorsqu'il existe une base génétique au cœur d'une
- Cependant, le séquençage de l'exome entier est une alternative moins coûteuse au séquençage du génome entier
- Dans le séquençage de l'exome, seules les régions codantes et productrices d'exon de l'ADN sont séquencées

Extraction ADN

Traitement avec des produits chimique de bactéries des cellules d'une plaque de gélose pour libérer l'ADN qui est ensuite purifié.



Découpage ADN

ADN est découpé en petits de taille connue
Soit par des enzymes soit par rupture

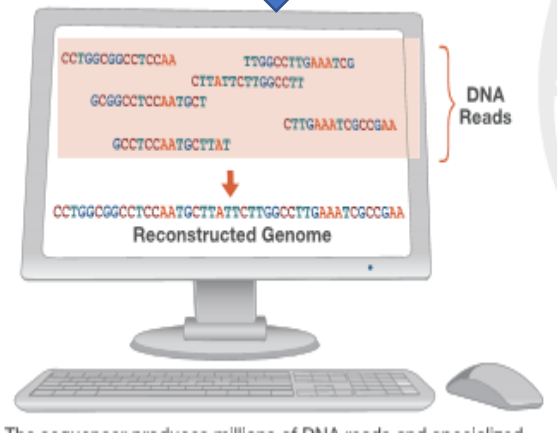
Préparation de librairie d'ADN

copies de chaque fragment d'ADN par PCR . L'ensemble des fragments générés dans un appareil PCR est appelé "bibliothèque ou librairie d'ADN".

Séquençage des librairies d'ADN

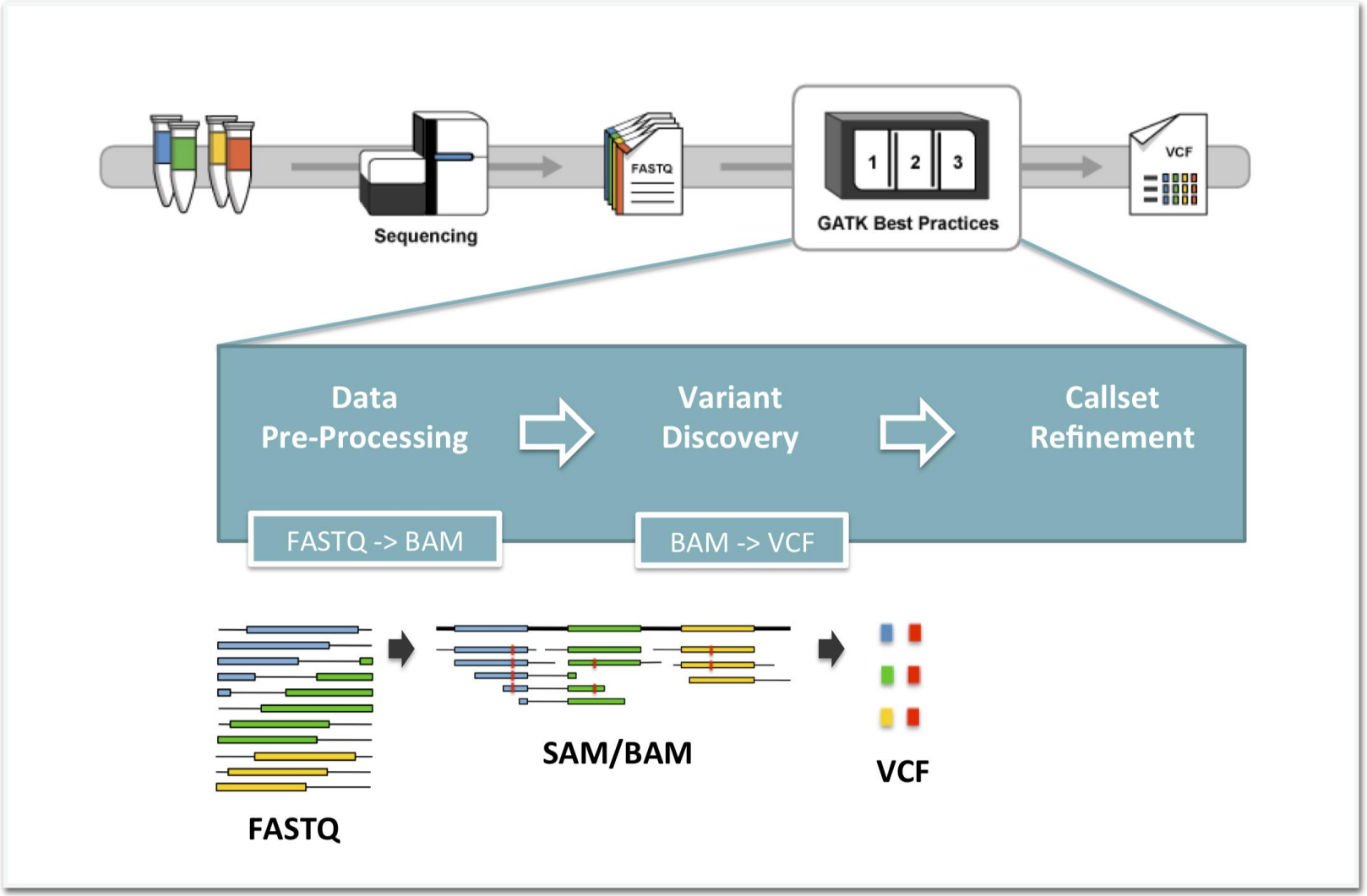
Séquençage des librairies d'ADN. La combinaison nucléotide est déterminée et chaque est appelé Reads

Analyse des séquences d'ADN

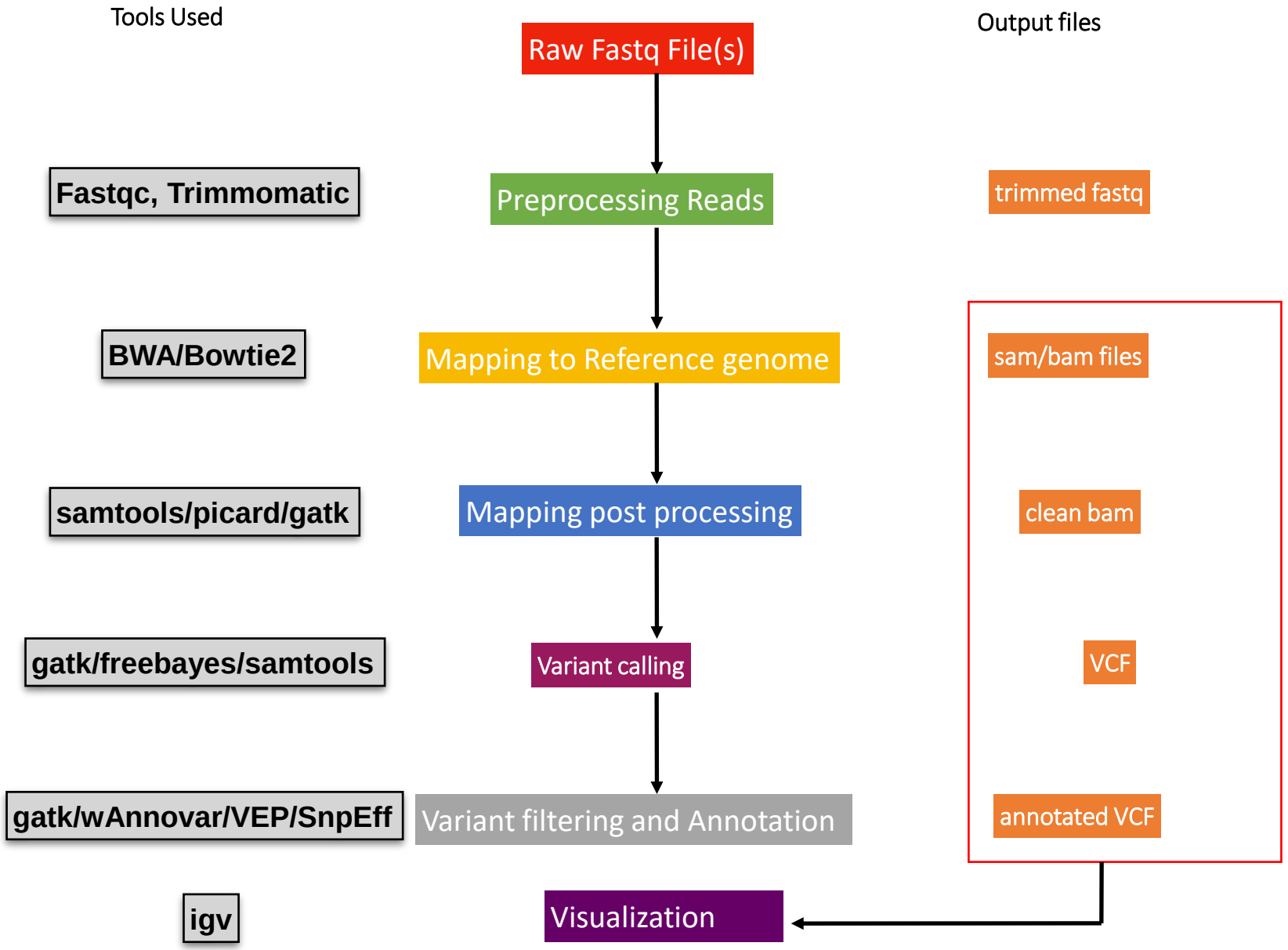


Des millions d'ADN lus et spécialisés. Des programmes informatiques sont utilisés pour les assembler dans le bon ordre, comme les pièces d'un puzzle. Une fois terminée, la séquence du génome contenant des millions de nucléotides (en un ou quelques gros morceaux) est prête pour une analyse plus approfondie.

Workflow: aperçu

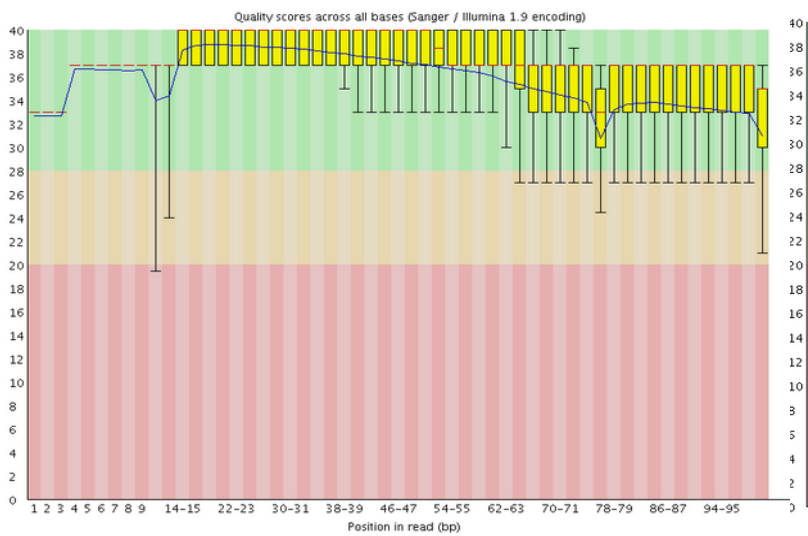


Flux Basique

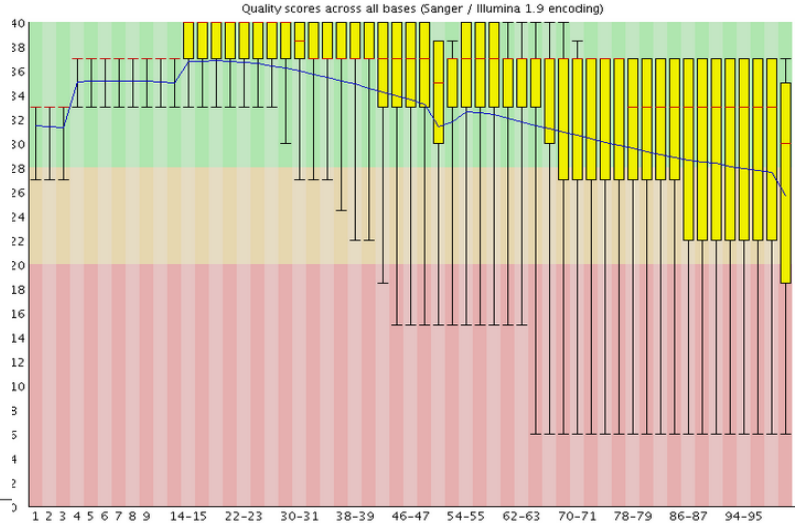


Fastq Qualité: Evaluation en utilisant FastQC

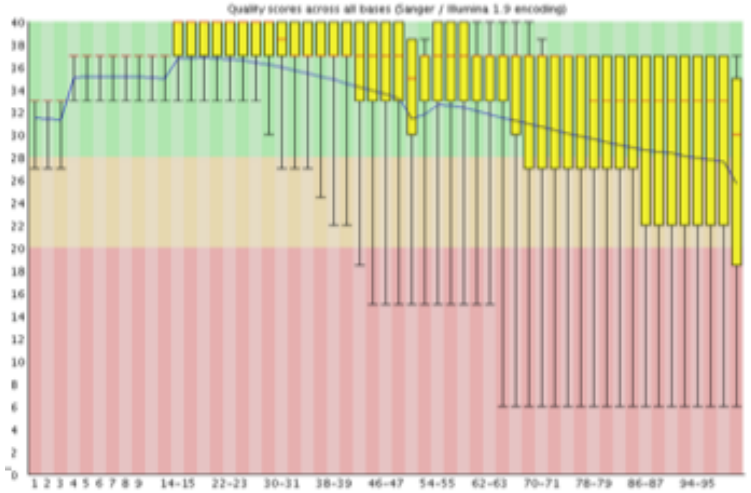
Good quality



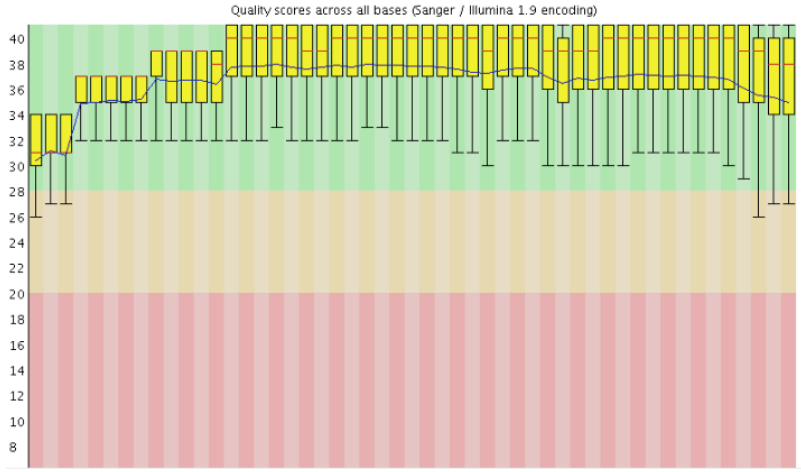
Not so good, requires trimming



Before trimming



After trimming



Fichiers générés

fastq: Text file with unaligned sequence information

SAM: Sequence Alignment Map with aligned sequence information

BAM: Binary SAM

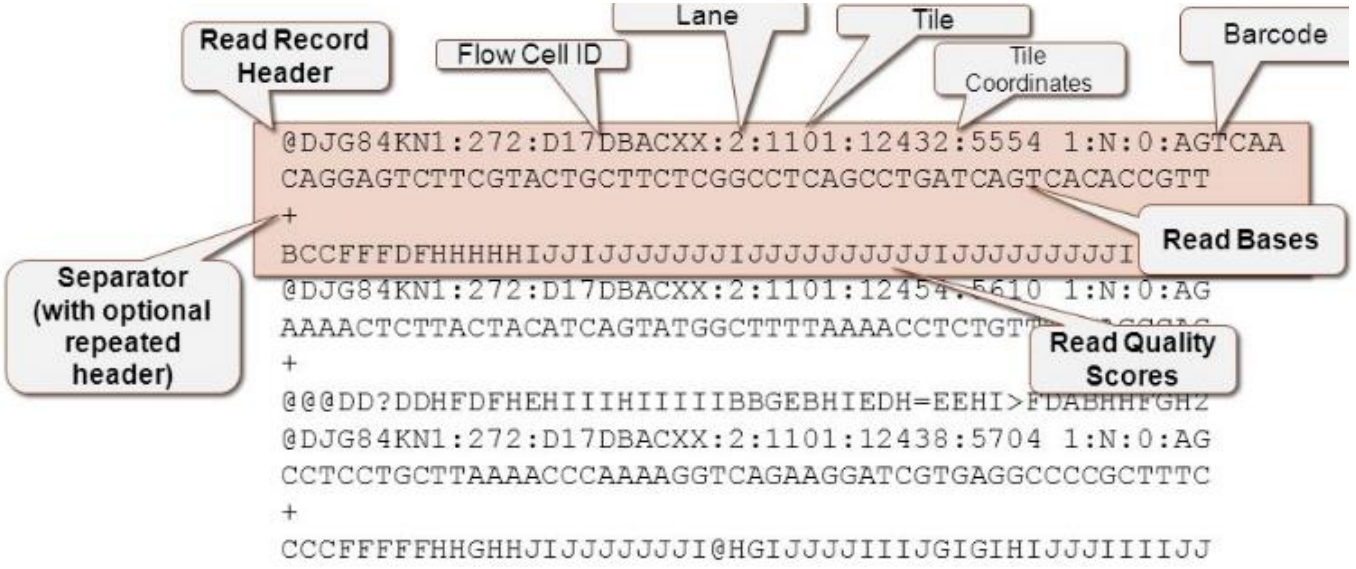
VCF: Variant Call Format file storing sequence variation information

BCF: Binary VCF

fastq et SAM/BAM formats

Fastq format

```
@HWI-ST1341:109:C8MUAACXX:8:1101:5151:1996 1:N:0:ACTGAT
NTTGCCCAGGCTGGAGTGCAGTGGCTATTCACAGGCGCGATCCCACTACTGATCAGCACGGGAGTTTTGACCTGCTCCGTTTCCGACCTGGGCCGGTTCAC
+
#0<BFFFFFFFFFFFFFFFIFBBFBFIIIIIFIIIFF7FFFFFFIFFIIIIFFFFFFFIFFBF707<BBBB<<BBBBBFFBBBBFBFB<BBBFFFFF07BB<
@HWI-ST1341:109:C8MUAACXX:8:1101:7410:1973 1:N:0:ACTGAT
NTCACTACCTCCCCGGGTCGGGAGTGGGTAATTTGCGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTTTCTCAGGCTCCCTCTCCGGAATAGATCGGAA
+
#0<FFFFFFFFFFIIIIFFFII<FFFFIFFFFFIIFIIIFFFFFFFBFBBBBBBBBBBBBBBBBBBB7BBBBBFBBBBBBBBBBB<BBBBBBBFBBBBB
@HWI-ST1341:109:C8MUAACXX:8:1101:8071:1973 1:N:0:ACTGAT
NCTATGTGCATACGGCCTGATGGCTCGAAACCATCATAACAGATTAATTTCCTTTTCAATAATAATTTAACTCTCAATTCATCCGGTTGAATACATTCCGA
+
#0<FFFFFFFFFFIIIIIIIIIFIIIIIIIIIIIIIIIIIIIIIIIIIIIIIFFIIIFFFIIIIIIFFFFFFFFFFBFFFBFFFBFFBBFFFFFFFFFFB7
@HWI-ST1341:109:C8MUAACXX:8:1101:8657:1999 1:N:0:ACTGAT
NGACGGGCCGCTCGGCGGACGGACGGACGGAATCGCGCCGGCCCGACCGCCCGCCCGACGCTCCGTCGGGAGACGGGCCCGGCGAGGGGGAGAGGCGACGG
+
#0<FFFFFFFFFFIIIIIIIIFFFBFBBFFFFFFFBBBBBFFFBFFB07BFF<BBFF7BBFFFFB<B7<77<<BBB77<BBFFF7<BBFF077BBBF<BFB
```



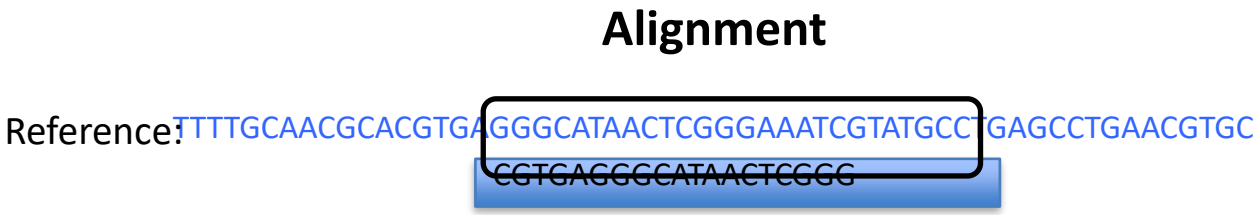
SAM/BAM format

Each row describes a single alignment of a raw read against the reference genome.
Each alignment has 11 mandatory fields, followed by any number of optional fields.

DNA Data

Reference: TTTTGCAACGCACGTGAGGGCATAACTCGGGAAATCGTATGCCTGAGCCTGAACGTGC

DNA Data


Alignment
Reference:
TTTTGCAACGCACGTGAGGGCATAACTCGGGAAATCGTATGCCTGAGCCTGAACGTGC
CGTGAGGGCATAACTCGGG

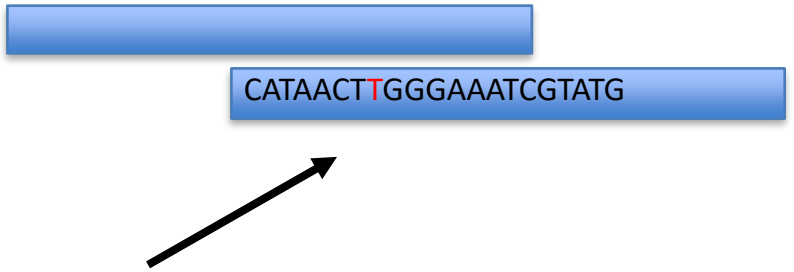
DNA Data

Reference: TTTTGCAACGCACGTGAGGGCATAACTCGGGAAATCGTATGCCTGAGCCTGAACGTGC

Perfect match to the reference

DNA Data

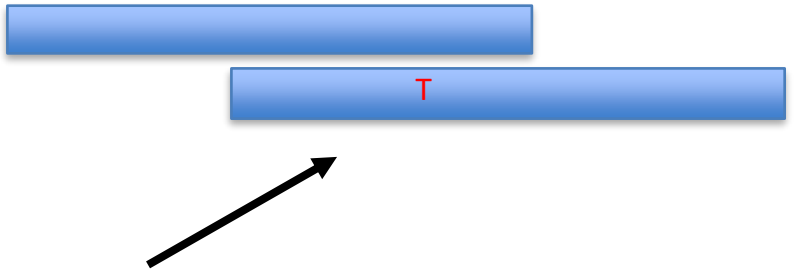
Reference: TTTTGCAACGCACGTGAGGGCATAACTCGGGAAATCGTATGCCTGAGCCTGAACGTGC



One mismatch to the reference

DNA Data

Reference: TTTTGCAACGCACGTGAGGGCATAACTCGGGAAATCGTATGCCTGAGCCTGAACGTGC

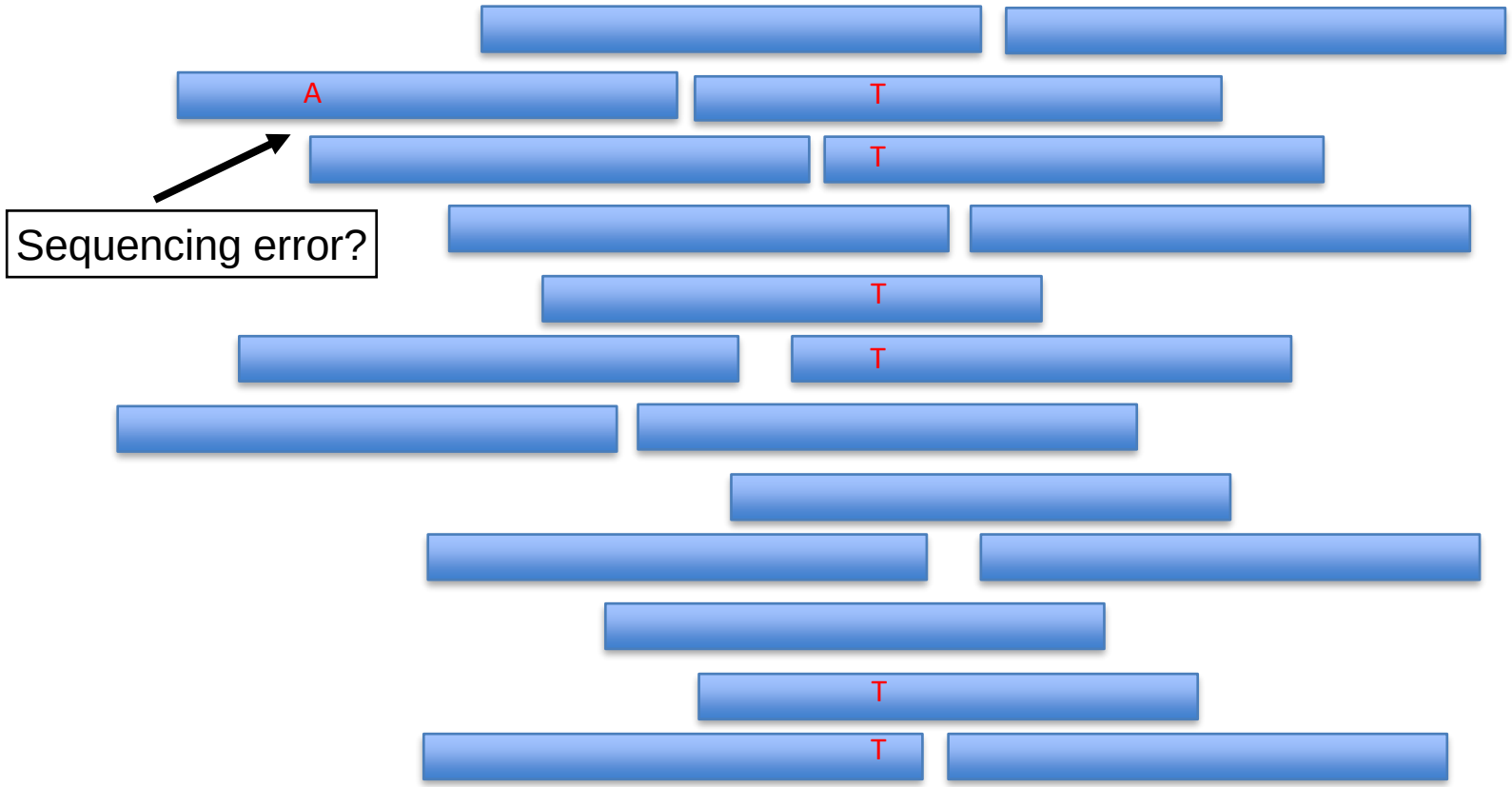


One mismatch to the reference

DNA Data

Variant Call: C/T = heterozygote

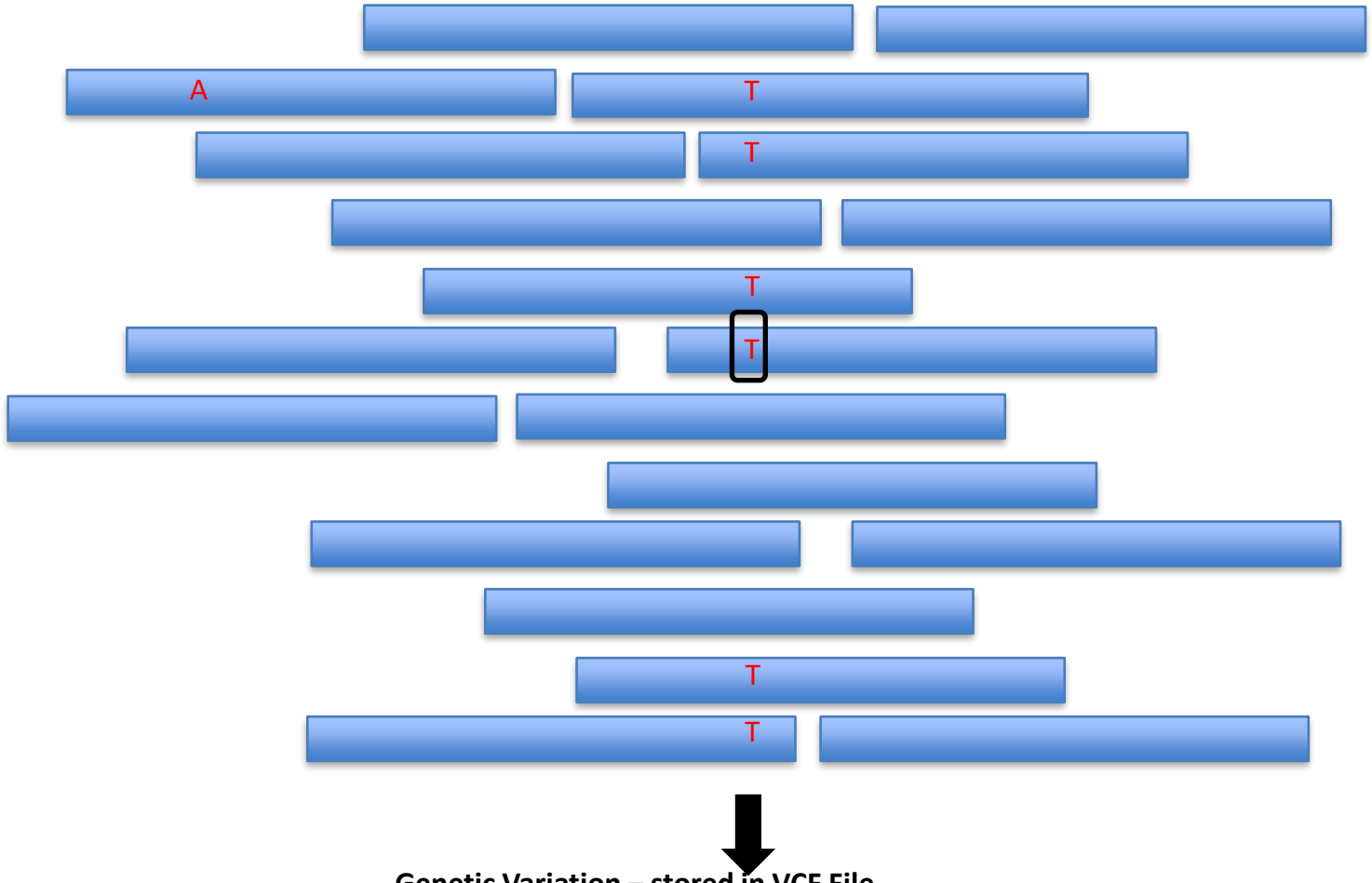
Reference: TTTTGCAACGCACGTGAGGGCATAACTCGGGAAATCGTATGCCTGAGCCTGAACGTGC



DNA Data

Variant Call: C/T = heterozygote

Reference: TTTTGCAACGCACGTGAGGGCATAACTCGGGAAATCGTATGCCTGAGCCTGAACGTGC



Genetic Variation – stored in VCF File

VCF File Format

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID        REF    ALT     QUAL FILTER INFO                              FORMAT      NA00001        NA00002        NA00003
20     14370   rs6054257 G      A       29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:.,.
20     17330   .         T      A       3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3   0/0:41:3
20     1110696 rs6040355 A      G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2   2/2:35:4
20     1230237 .         T      .       47   PASS   NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20     1234567 microsat1 GTC    G,GTCT  50   PASS   NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4       0/2:17:2       1/1:40:3
```

Field	Number	Type	Description
AD	R	Integer	Read depth for each allele
ADF	R	Integer	Read depth for each allele on the forward strand
ADR	R	Integer	Read depth for each allele on the reverse strand
DP	1	Integer	Read depth
EC	A	Integer	Expected alternate allele counts
FT	1	String	Filter indicating if this genotype was “called”
GL	G	Float	Genotype likelihoods
GP	G	Float	Genotype posterior probabilities
GQ	1	Integer	Conditional genotype quality
GT	1	String	Genotype
HQ	2	Integer	Haplotype quality
MQ	1	Integer	RMS mapping quality
PL	G	Integer	Phred-scaled genotype likelihoods rounded to the closest integer
PQ	1	Integer	Phasing quality
PS	1	Integer	Phase set

Other genotype keys

VCF File Format

1.1 An example

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID        REF    ALT     QUAL FILTER INFO                              FORMAT      NA00001        NA00002
20     14370   rs6054257 G      A       29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51
20     17330   .         T      A       3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3
20     1110696 rs6040355 A      G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2
20     1230237 .         T      .       47   PASS   NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51
20     1234567 microsat1 GTC    G,GTCT  50   PASS   NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4       0/2:17:2
```

Genetic Variation

http://samtools.github.io/hts-specs/VCFv4.3.pdf

VCF File Format

Main Fields:

- CHROM : Chromosome dans le génome de référence
- POS : Position dans le chromosome (ou contig)
- ID : Identificateur - le plus souvent le numéro rs de dbSNP
- REF : Base(s) de référence - peut être multiple pour l'insertion/suppression
- ALT : Base(s) alternative(s) - peut inclure plusieurs bases alternatives
- QUAL : Qualité - probabilité basée sur le phred que la ou les bases alternatives soient fausses
- FILTRE : Filtre - PASS ou raison de l'échec selon les critères de filtrage
- INFO : Informations complémentaires - détails normalement dans les métadonnées
- FORMAT : Informations détaillant les types de données affichées
- DONNÉES SNP : normalement le génotype plus d'autres types de données

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	NA00001	NA00002	NA00003
20	14370	rs6054257	G	A	29	PASS	NS=3;DP=14;AF=0.5;DB;H2	GT:GQ:DP:HQ	0|0:48:1:51,51	1|0:48:8:51,51	1/1:43:5:.,.
20	17330	.	T	A	3	q10	NS=3;DP=11;AF=0.017	GT:GQ:DP:HQ	0|0:49:3:58,50	0|1:3:5:65,3	0/0:41:3
20	1110696	rs6040355	A	G,T	67	PASS	NS=2;DP=10;AF=0.333,0.667;AA=T;DB	GT:GQ:DP:HQ	1|2:21:6:23,27	2|1:2:0:18,2	2/2:35:4
20	1230237	.	T	.	47	PASS	NS=3;DP=13;AA=T	GT:GQ:DP:HQ	0|0:54:7:56,60	0|0:48:4:51,51	0/0:61:2
20	1234567	microsat1	GTC	G,GTCT	50	PASS	NS=3;DP=9;AA=G	GT:GQ:DP	0/1:35:4	0/2:17:2	1/1:40:3

VCF File Format

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID        REF  ALT     QUAL FILTER INFO                              FORMAT      NA00001        NA00002        NA00003
20     14370   rs6054257 G    A       29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:.,.
20     17330   .         T    A       3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3   0/0:41:3
20     1110696 rs6040355 A    G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2   2/2:35:4
20     1230237 .         T    .       47   PASS   NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20     1234567 microsat1 GTC  G,GTCT  50   PASS   NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4       0/2:17:2       1/1:40:3
```

0->Ref 1->1st Alt 2->2nd Alt N->Nth Alt

G/G Hom-REF

A/G Het

A/A Hom-Alt

Chr 20
Pos 14370

rs6054257

Reference G
Alternative A

Variant
Quality 29
(0.126%)

Pass

Variant Info:
3 samples
14 reads total
Allele Freq = 50%
Variant In dbSNP
Variant in Hapmap2

Expect Info on:
Genotype
Quality
Read Depth
Haplotype Quality

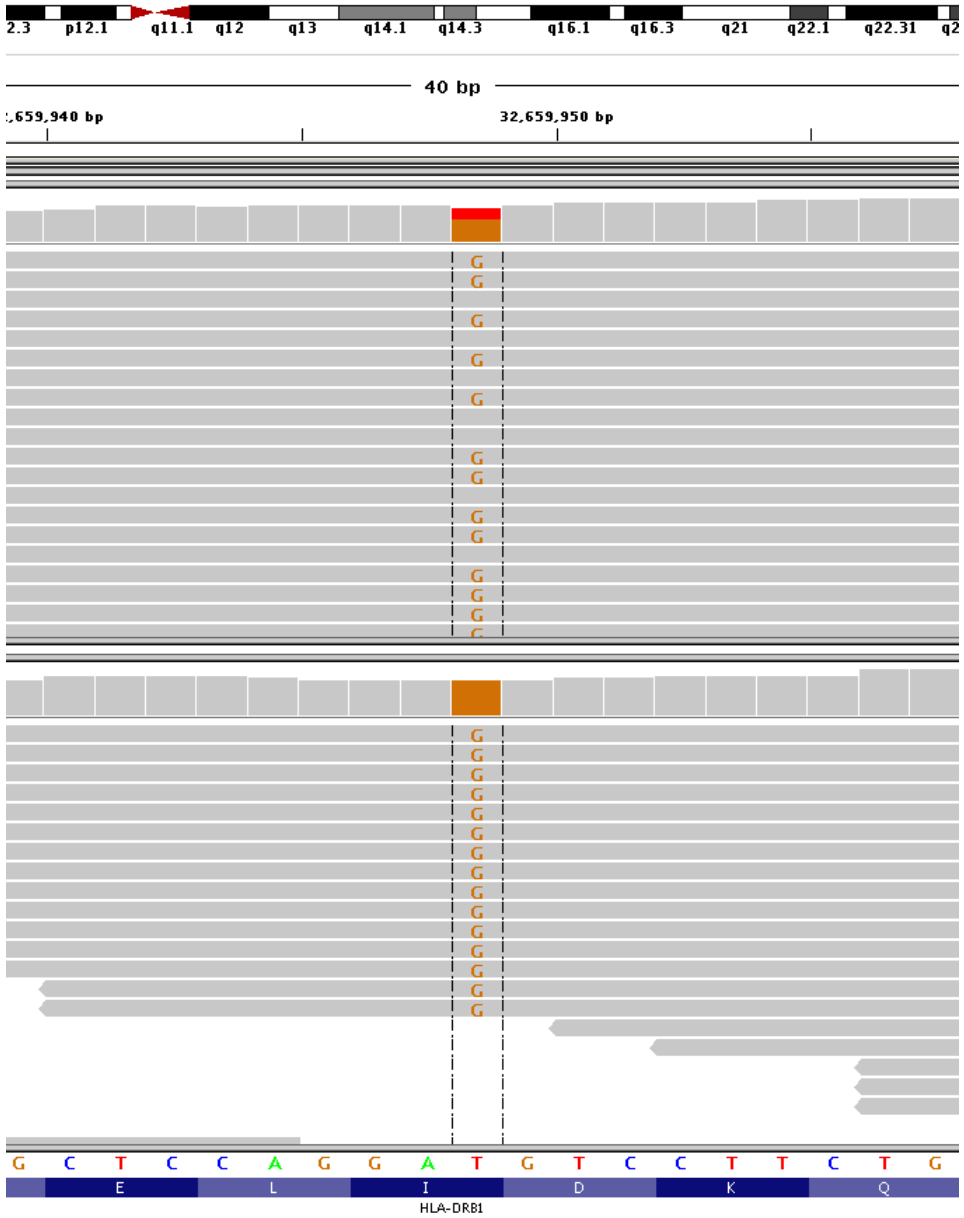
Info:
Genotype = Heterozygote
Quality = 48 (0.0125%)
Read Depth = 8
Haplotype Quality = 51&51

Visualisation des données de l'AND sur IGV

Individual 1
Genotype G/T

Individual 2
Genotype G/G

Reference



VCF File Format

- Format de fichier standard dans le domaine de la génétique / génomique
- Fichier de sortie généré par presque tous les logiciels d'appel de variantes
- Format de fichier utilisé dans les outils en aval pour l'analyse des données

1.1 An example

File and Genome Information

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID        REF    ALT     QUAL FILTER INFO                              FORMAT      NA00001        NA00002
20     14370   rs6054257 G      A       29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51
20     17330   .         T      A       3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3
20     1110696 rs6040355 A      G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2
20     1230237 .         T      .       47   PASS   NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51
20     1234567 microsat1 GTC    G,GTCT  50   PASS   NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4       0/2:17:2
```

VCF File Format

1.1 An example

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID        REF    ALT     QUAL FILTER INFO                              FORMAT      NA00001        NA00002
20     14370   rs6054257 G      A       29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51
20     17330   .         T      A       3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3
20     1110696 rs6040355 A      G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2
20     1230237 .         T      .       47   PASS   NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51
20     1234567 microsat1 GTC    G,GTCT  50   PASS   NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4       0/2:17:2
```

Field Descriptions

VCF File Format

1.1 An example

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID        REF    ALT     QUAL FILTER INFO                              FORMAT      NA00001        NA00002
20     14370   rs6054257 G      A       29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51
20     17330   .         T      A       3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3
20     1110696 rs6040355 A      G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2
20     1230237 .         T      .       47   PASS   NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51
20     1234567 microsat1 GTC    G,GTCT  50   PASS   NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4       0/2:17:2
```

Genetic Variation

Annotation du Variant

Explorer Fichier VCF généré à partir du whole exome sequence d'un patient

Quelles mutations pourraient être importante pour comprendre la maladie ?

Réaliser l'annotation du Variant

Annotation du Variant

Quel type d'information pourrait nous intéresser ?

Annotation fonctionnelle : où se situe la variante et quel type de changement entraîne-t-elle : faux-sens, non-sens, intronique, intergénique.

Information sur la fréquence : quelle est la fréquence de la mutation dans la population générale ? 1000 données sur les génomes, données EXac.

Conservation évolutive : La position du site muté est-elle importante ? par exemple, le score du GERP (Un score utilisé pour calculer la conservation de chaque nucléotide dans un alignement multi-espèces)

Information sur les gènes : s'il s'agit d'une variante exonique, dans quel gène se trouve-t-elle ?

Impact prévu : savons-nous comment la mutation pourrait influencer la structure de la protéine ? Polyphène, SIFT.

Informations sur l'association : la mutation influence-t-elle l'expression d'un gène ou a-t-elle été précédemment associée à une maladie ?

Il existe de nombreux outils permettant d'annoter les fichiers VCF avec des informations fonctionnelles

Découverte de Mutations

Pipeline de base :

- **Recherchez les variantes dans les régions de codage.**
- **Les mutations absurdes (arrêt des pertes/arrêt des gains) ont plus de chances d'être fonctionnellement importantes.**
- **La mutation est-elle censée être dommageable ?**
- **La variante se produit-elle à une position qui est conservée pour l'évolution ?**
- **La variante est-elle à faible fréquence dans les études de population ?**
- **Le gène a-t-il un sens ?**

Next steps: **Validation and functional tests…** **Back to the lab!**

Outils pour VCF Annotation

wANNOVAR — http://wannovar.wglab.org

Launch VeP — https://asia.ensembl.org/Tools/VEP

SnpEff
Genomic variant annotations and functional effect prediction toolbox. — http://snpeff.sourceforge.net

SnpSift
Filter and manipulate annotated files — http://snpeff.sourceforge.net/SnpSift.html

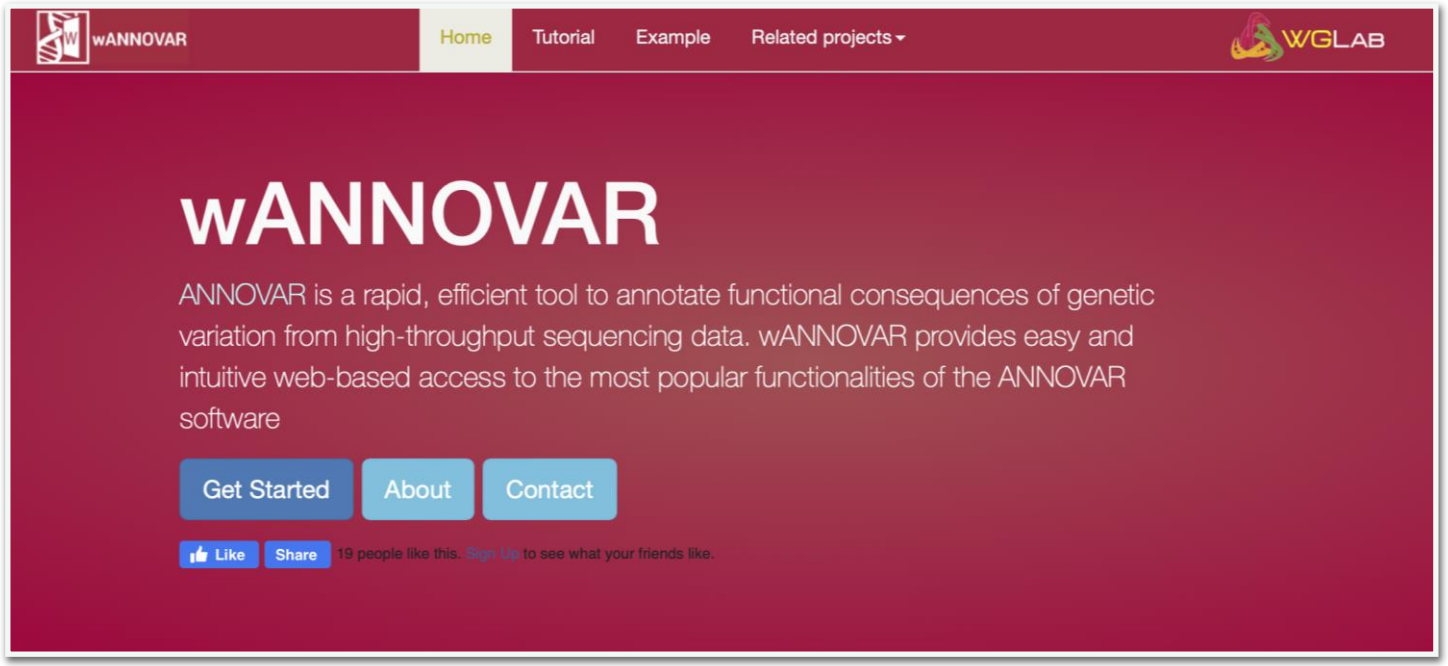
SeattleSeq Annotation 151 — https://snp.gs.washington.edu/SeattleSeqAnnotation151/

Ingenuity Variant Analysis — QIAGEN — https://www.qiagenbioinformatics.com/products/ingenuity-variant-analysis/

moon by Diploid — http://www.diploid.com/moon

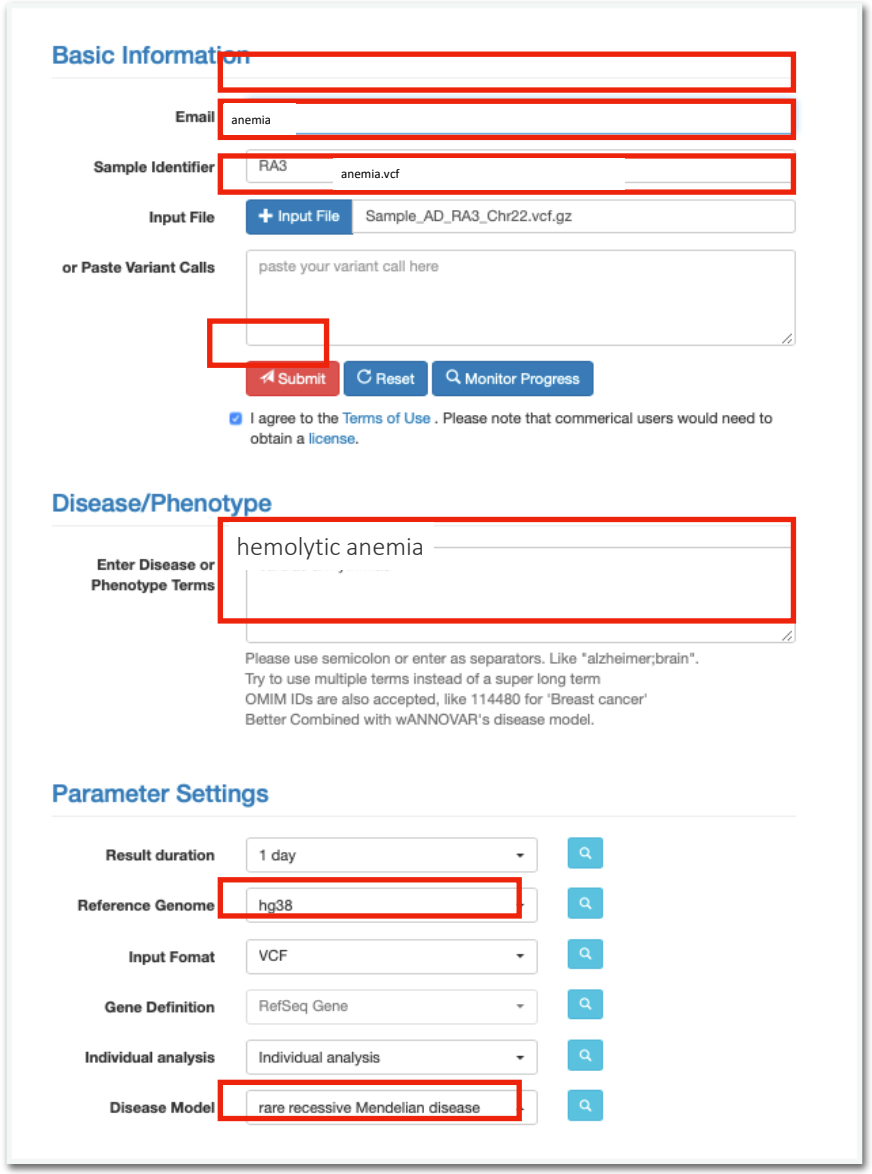
Franklin by Genoox — https://franklin.genoox.com/home

VCF Annotation: wANNOVAR



http://wannovar.wglab.org

VCF Annotation: wANNOVAR



Your email address

Job name

Upload your VCF/Input

Submit

Phenotype

Reference genome

Disease Model

VCF Annotation: wANNOVAR

Sample identifier = anemia
File_name=anemia.vcf
File_format=vcf4
Reference_genome=hg19
Disease_model=rare recessive disease
Processed variants=30726
phenotype is hemolytic anemia

exome summary results	view	CSV file	TXT file
genome summary results	view	CSV file	TXT file

ANNOVAR filtering results:

(click to view details about this pipeline)
Initially, 11361 variants were fed into the annotation pipeline and 0 variants were detected as invalid input.

Step1:254 variants	Identify missense, nonsense and splicing variants	download
Step2:25 variants	Remove variants in the 1000 Genomes Project(ALL) with MAF>0.01	download
Step3:22 variants	Remove variants in gnomAD exome database with MAF>0.01	download
Step4:4 genes	Compile a list of candidate genes based on diseaes model	download variants_info

download all filtering results

Phenotype/disease Prioritization Result:

Exonic variant list from the wANNOVAR output after filtration, the last filter step with variants left will be used. (Total: 4)	Variant List
Gene list from the wANNOVAR output, input into Phenolyzer (Total: 4)	Input Gene List
The prioritized genes from Phenolyzer (Total: 3)	Result Gene List
The network visualization	Show

VCF Annotation: wANNOVAR

SNP/Gene Informations

Chr	Start	End	Ref	Alt	Func	Gene	GeneDetail	ExonicFunc	AAChange
22	16783675	16783675	G	T	exonic	XKR3		nonsynonymous SNV	XKR3:NM_001318251:exon4:c.C1324A:p.H442N,XKR3:NM_175878:exon4:c.C1324A:p.H442N
22	16784014	16784014	A	G	exonic	XKR3		synonymous SNV	XKR3:NM_001318251:exon4:c.T985C:p.L329L,XKR3:NM_175878:exon4:c.T985C:p.L329L
22	16784234	16784234	A	C	exonic	XKR3		nonsynonymous SNV	XKR3:NM_001318251:exon4:c.T765G:p.F255L,XKR3:NM_175878:exon4:c.T765G:p.F255L
22	16784304	16784304	G	A	exonic	XKR3		nonsynonymous SNV	XKR3:NM_001318251:exon4:c.C695T:p.P232L,XKR3:NM_175878:exon4:c.C695T:p.P232L
22	16966101	16966101	C	T	exonic	GAB4		synonymous SNV	GAB4:NM_001037814:exon6:c.G1287A:p.K429K
22	16970162	16970162	G	C	exonic	GAB4		nonsynonymous SNV	GAB4:NM_001037814:exon4:c.C718G:p.P240A
22	17108356	17108356	G	A	exonic	IL17RA		synonymous SNV	IL17RA:NM_001289905:exon12:c.G1035A:p.K345K,IL17RA:NM_014339:exon13:c.G1137A:p.K379K
22	17109290	17109290	G	A	exonic	IL17RA		nonsynonymous SNV	IL17RA:NM_001289905:exon12:c.G1969A:p.A657T,IL17RA:NM_014339:exon13:c.G2071A:p.A691T
22	17120087	17120087	G	A	exonic	CECR6		synonymous SNV	CECR6:NM_031890:exon1:c.C1041T:p.F347F
22	17120858	17120858	C	T	exonic	CECR6		synonymous SNV	CECR6:NM_031890:exon1:c.G270A:p.V90V
22	17159155	17159155	G	A	exonic	HDHD5		nonsynonymous SNV	HDHD5:NM_033070:exon1:c.C97T:p.R33C
22	17209519	17209519	G	A	exonic	ADA2		synonymous SNV	ADA2:NM_001282225:exon2:c.C159T:p.N53N,ADA2:NM_001282226:exon2:c.C159T:p.N53N,ADA2:NM_(
22	17726847	17726847	A	G	exonic	BCL2L13		synonymous SNV	BCL2L13:NM_001270727:exon5:c.A699G:p.S233S,BCL2L13:NM_001270730:exon5:c.A285G:p.S95S,BC
22	19039100	19039100	A	G	exonic	DGCR2		nonsynonymous SNV	DGCR2:NM_001173533:exon9:c.T1295C:p.V432A,DGCR2:NM_001173534:exon9:c.T1286C:p.V429A,DGC
22	19122176	19122176	G	A	exonic	DGCR2		synonymous SNV	DGCR2:NM_001173533:exon1:c.C31T:p.L11L,DGCR2:NM_001173534:exon1:c.C31T:p.L11L,DGCR2:NM_
22	19132032	19132032	C	T	exonic	TSSK2		synonymous SNV	TSSK2:NM_053006:exon1:c.C633T:p.C211C
22	19132173	19132173	C	T	exonic	TSSK2		synonymous SNV	TSSK2:NM_053006:exon1:c.C774T:p.S258S
22	19132238	19132238	C	T	exonic	TSSK2		nonsynonymous SNV	TSSK2:NM_053006:exon1:c.C839T:p.T280M
22	19132425	19132425	G	A	exonic	TSSK2		synonymous SNV	TSSK2:NM_053006:exon1:c.G1026A:p.R342R
22	19134359	19134359	G	A	exonic	DGCR14		nonsynonymous SNV	DGCR14:NM_022719:exon10:c.C1268T:p.A423V
22	19135152	19135152	C	T	exonic	DGCR14		synonymous SNV	DGCR14:NM_022719:exon9:c.G1059A:p.E353E
22	19196584	19196584	T	C	exonic	CLTCL1		nonsynonymous SNV	CLTCL1:NM_001835:exon25:c.A3946G:p.M1316V,CLTCL1:NM_007098:exon25:c.A3946G:p.M1316V
22	19208992	19208992	G	A	exonic	CLTCL1		synonymous SNV	CLTCL1:NM_001835:exon21:c.C3372T:p.A1124A,CLTCL1:NM_007098:exon21:c.C3372T:p.A1124A
22	19210386	19210386	T	C	exonic	CLTCL1		synonymous SNV	CLTCL1:NM_001835:exon20:c.A3189G:p.A1063A,CLTCL1:NM_007098:exon20:c.A3189G:p.A1063A
22	19444271	19444271	G	A	exonic	C22orf39		synonymous SNV	C22orf39:NM_173793:exon3:c.C423T:p.D141D
22	19524402	19524402	G	A	exonic	CLDN5		stopgain	CLDN5:NM_001130861:exon1:c.C109T:p.Q37X,CLDN5:NM_003277:exon2:c.C109T:p.Q37X
22	19763250	19763250	T	C	exonic	TBX1		synonymous SNV	TBX1:NM_005992:exon4:c.T420C:p.F140F,TBX1:NM_080646:exon4:c.T420C:p.F140F,TBX1:NM_08064
22	19765926	19765926	A	G	exonic	TBX1		synonymous SNV	TBX1:NM_005992:exon8:c.A933G:p.A311A,TBX1:NM_080646:exon8:c.A933G:p.A311A,TBX1:NM_08064
22	19766568	19766568	A	C	exonic	TBX1		nonsynonymous SNV	TBX1:NM_080647:exon9:c.A1189C:p.N397H
22	19779259	19779259	C	T	exonic	TBX1		nonsynonymous SNV	TBX1:NM_080646:exon9:c.C1049T:p.T350M
22	19802018	19802018	A	C	exonic	GNB1L		nonsynonymous SNV	GNB1L:NM_053004:exon7:c.T715G:p.W239G
22	19880248	19880248	C	T	exonic	TXNRD2		unknown	UNKNOWN
22	19880695	19880695	A	G	exonic	TXNRD2		unknown	UNKNOWN
22	19895461	19895461	T	G	exonic	TXNRD2		nonsynonymous SNV	TXNRD2:NM_001282512:exon11:c.A895C:p.S299R
22	19918988	19918988	G	A	exonic	TXNRD2		synonymous SNV	TXNRD2:NM_001282512:exon4:c.C246T:p.L82L
22	19919595	19919595	G	A	exonic	TXNRD2		synonymous SNV	TXNRD2:NM_001282512:exon3:c.C177T:p.A59A
22	19962712	19962712	C	T	exonic	COMT		synonymous SNV	COMT:NM_007310:exon1:c.C36T:p.H12H,COMT:NM_000754:exon3:c.C186T:p.H62H,COMT:NM_0011351(

VCF Annotation: wANNOVAR

Allele Frequencies

Cosmic

1000G ALL	1000G AFR	1000G AMR	1000G EAS	1000G EUR	1000G SAS	ExAC Freq	ExAC AFR	ExAC AMR	ExAC EAS	ExAC FIN	ExAC NFE	ExAC OTH	ExAC SAS	ESP6500si ALL	ESP6500si AA	ESP6500si EA	CG46	NCI60	dbSNP	COSMIC ID	COSMIC DIS
0.65	0.79	0.56	0.37	0.59	0.87	.	.	.	.	.	.	.	.	.	.	.	0.86	.	rs5748622	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0.35	.	rs9605145	.	.
0.69	0.79	0.63	0.5	0.63	0.88	0.0037	0.0094	0.0019	0.0046	0	0.0045	0.0086	0	.	.	.	0.87	0.033	rs5748623	COSM4591934	1
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0.42	.	rs9605146	.	.
0.83	0.98	0.84	0.84	0.74	0.68	0.7719	0.9492	0.8960	0.8381	0.7521	0.7368	0.7711	0.6951	0.82	0.94	0.75	0.84	0.78	rs4819925	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.12	0.25	0.089	.	0.12	0.1	0.1153	0.2365	0.0607	0.0003	0.1092	0.1229	0.1180	0.1092	0.16	0.23	0.12	0.18	0.18	rs879576	.	.
0.11	0.24	0.078	.	0.11	0.1	0.3023	0.4136	0.3	0.0043	0.3493	0.3580	0.2821	0.2349	0.13	0.18	0.1	0.13	0.041	rs41323645	COSM3749756	2
0.48	0.49	0.51	0.49	0.38	0.52	0.4307	0.4758	0.5099	0.4994	0.3898	0.3880	0.4218	0.4973	0.41	0.47	0.37	0.22	0.47	rs5994165	COSM3759017	1
0.33	0.58	0.24	0.13	0.34	0.28	0.5732	0.6341	0.5	0.3235	.	0.6482	0.5	0.5569	0.34	0.46	0.28	.	.	rs9606620	COSM4135112	1
0.21	0.32	0.18	0.035	0.28	0.19	0.4810	0.5077	0.4044	0.0898	0.4545	0.5210	0.5	0.4497	.	.	.	0.087	.	rs7287672	COSM3759018	1
0.31	0.27	0.34	0.074	0.49	0.4	0.4341	0.3020	0.2811	0.0669	0.4646	0.5280	0.4625	0.4237	0.45	0.31	0.52	0.36	0.5	rs362129	COSM4135116	1
0.67	0.58	0.65	0.93	0.54	0.66	0.5914	0.6022	0.7001	0.9319	0.5100	0.5228	0.6035	0.6385	0.55	0.6	0.53	0.59	0.49	rs4488761	.	.
0.43	0.57	0.41	0.34	0.36	0.43	0.3892	0.5304	0.3274	0.3463	0.2879	0.3818	0.3337	0.4407	0.42	0.51	0.38	0.43	0.39	rs2072123	COSM4135140	2
0.0024	.	0.0014	.	0.003	0.0082	0.0073	0.0004	0.0043	0	0	0.0047	0.0052	0.0129	0.0008	.	0.0012	.	.	rs150441373	.	.
0.16	0.12	0.13	0.12	0.17	0.27	0.1742	0.1157	0.1190	0.1339	0.1402	0.1768	0.1375	0.2759	0.16	0.11	0.18	0.15	0.21	rs45604134	COSM5425332	1
0.16	0.094	0.13	0.12	0.17	0.27	0.1720	0.0891	0.1179	0.1335	0.1400	0.1771	0.1367	0.2762	0.15	0.086	0.18	0.14	0.22	rs1052756	COSM5425333	1
0.28	0.36	0.22	0.21	0.19	0.35	0.2283	0.3298	0.2470	0.2107	0.2495	0.1864	0.1920	0.3234	0.23	0.32	0.19	0.26	0.23	rs1052763	.	.
0.22	0.29	0.15	0.14	0.18	0.27	0.1934	0.2661	0.1281	0.1529	0.1642	0.1809	0.1507	0.2813	0.2	0.25	0.18	0.22	0.23	rs1052773	.	.
0.1	0.17	0.086	0.11	0.055	0.071	0.0737	0.1740	0.0675	0.1042	0.0312	0.0626	0.0543	0.0597	0.095	0.17	0.057	0.11	0.074	rs712965	COSM4135146	1
0.17	0.12	0.13	0.12	0.19	0.28	0.1819	0.1192	0.0888	0.1327	0.1536	0.1920	0.1402	0.2862	0.17	0.12	0.2	0.15	0.21	rs2240111	COSM5425334	1
0.4	0.23	0.46	0.31	0.51	0.59	0.4897	0.2780	0.4803	0.3168	0.5217	0.5113	0.5126	0.6059	0.42	0.26	0.5	0.38	.	rs1061325	COSM4416574;COSM4416573	1
0.15	0.11	0.12	0.4	0.038	0.11	0.1216	0.1210	0.2101	0.4518	0.0762	0.0632	0.1105	0.1168	0.063	0.1	0.045	0.13	.	rs1060376	COSM4002092;COSM4002093	1
0.85	0.97	0.78	0.94	0.63	0.88	0.7356	0.9278	0.8406	0.9399	0.6749	0.6391	0.7133	0.8567	0.73	0.92	0.64	0.88	.	rs698423	.	.
0.0084	.	.	0.001	0.004	0.038	0.0095	0.0015	0.0024	0.0004	0.0009	0.0081	0.0115	0.0352	0.007	0.0014	0.0099	.	.	rs61748355	COSM4667547;COSM4667548	1
0.5	0.21	0.59	0.53	0.56	0.73	0.5912	0.3219	0.6234	0.6117	0.75	0.6353	0.5294	0.7577	.	.	.	0.35	.	rs885985	COSM3749757	1
0.23	0.023	0.28	0.46	0.22	0.24	0.2352	0.0442	0.3334	0.4699	0.1582	0.2276	0.2029	0.2273	0.16	0.044	0.23	0.25	0.28	rs41298814	COSM3759030;COSM3759029;COSM3759031	1
0.23	0.022	0.28	0.47	0.22	0.24	0.2276	0.0307	0.2558	0.4	0.175	0.2584	0.2097	0.2249	0.061	0.011	0.088	0.033	.	rs41298840	COSM4135151;COSM4135150;COSM4135149	1
0.23	0.022	0.28	0.47	0.22	0.25	0.3115	0.0842	0.4815	0.5494	0.4375	0.3785	0.2544	0.2628	0.11	0.03	0.15	.	.	rs72646967	COSM5428525	1
0.21	0.16	0.26	0.12	0.17	0.38	0.2111	0.1471	0.2397	0.1154	0.2028	0.1920	0.2115	0.3621	0.17	0.15	0.18	0.12	0.17	rs4819522	COSM1414906	1
0.21	0.035	0.24	0.48	0.12	0.22	0.2697	0.0763	0.4661	0.5818	0.1387	0.2226	0.2406	0.2633	0.088	0.039	0.11	0.2	0.057	rs2073770	COSM3759035	5
0.27	0.64	0.17	0.033	0.16	0.19	0.1853	0.5661	0.1439	0.0323	0.0962	0.1620	0.1852	0.1949	0.29	0.55	0.16	0.46	0.082	rs1139795	.	.
0.72	0.95	0.62	0.62	0.72	0.56	0.7012	0.9150	0.5511	0.6030	0.6915	0.7400	0.6950	0.5808	0.8	0.92	0.74	0.77	0.64	rs1139793	COSM5022107	1
0.25	0.55	0.16	0.052	0.17	0.2	0.1879	0.5139	0.1389	0.0655	0.1259	0.1665	0.1778	0.2041	0.27	0.48	0.17	0.33	0.066	rs5992495	COSM4985211;COSM4985210	12
0.17	0.37	0.12	0.013	0.15	0.11	0.1463	0.3763	0.0842	0.0126	0.1250	0.1510	0.1578	0.1154	0.21	0.34	0.15	0.054	0.049	rs11541479	.	.
0.6	0.39	0.63	0.88	0.51	0.65	0.5649	0.4302	0.6263	0.8092	0.5712	0.5426	0.6339	0.6157	0.45	0.4	0.48	0.48	0.34	rs5748470	.	.
0.37	0.29	0.38	0.27	0.5	0.44	0.4680	0.3332	0.4029	0.2730	0.5546	0.5206	0.4775	0.4536	0.46	0.34	0.52	0.35	0.43	rs4633	COSM149252	1
0.3	0.17	0.3	0.34	0.4	0.31	0.3392	0.2021	0.1938	0.3246	0.3061	0.3969	0.3699	0.3152	0.33	0.2	0.39	0.3	0.45	rs4818	COSM3759038	1
0.37	0.28	0.38	0.28	0.5	0.44	0.4727	0.3197	0.4075	0.2849	0.5723	0.5283	0.4827	0.4481	0.45	0.31	0.52	0.33	0.41	rs4680	COSM4997949	1
0.62	0.43	0.62	0.53	0.85	0.72	0.7512	0.4919	0.5634	0.5126	0.7949	0.8580	0.7882	0.7201	0.73	0.5	0.86	0.66	0.88	rs165815	.	.
0.36	0.39	0.35	0.23	0.34	0.48	0.4773	0.4792	0.4853	0.3223	0.5434	0.4628	0.4338	0.5426	0.34	0.35	0.34	0.4	0.15	rs2073748	COSM1751722	1
0.72	0.8	0.65	0.61	0.77	0.74	0.7375	0.7966	0.6050	0.5970	0.7350	0.7775	0.7127	0.7243	0.81	0.83	0.8	0.7	0.58	rs2073747	COSM3759040	1
0.42	0.6	0.37	0.23	0.35	0.49	0.3941	0.5702	0.3791	0.2295	0.4070	0.3706	0.3505	0.4846	0.41	0.53	0.34	0.4	0.3	rs2240717	COSM3759041	1
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.19	0.11	0.19	0.084	0.3	0.31	0.3405	0.1719	0.2569	0.0829	0.2857	0.3609	0.3333	0.3895	.	.	.	0.087	.	rs175181	.	.
0.24	0.13	0.22	0.18	0.43	0.27	0.3602	0.1671	0.2135	0.2023	0.4820	0.4432	0.4076	0.2709	0.35	0.16	0.44	0.17	0.34	rs7235	COSM149253	1
0.11	0.19	0.11	0.002	0.16	0.056	0.1236	0.1844	0.0800	0.0003	0.1331	0.1532	0.1211	0.0595	0.16	0.17	0.15	0.098	0.098	rs2228236	COSM3759047	1
0.53	0.62	0.5	0.33	0.65	0.54	0.5814	0.6200	0.4253	0.3594	0.6813	0.6310	0.6230	0.5700	0.65	0.66	0.65	0.28	0.2	rs874101	COSM5458095	1
0.24	0.34	0.24	0.039	0.41	0.15	0.3829	0.3801	0.2275	0.1100	0.5147	0.4423	0.4237	0.2629	0.29	0.26	0.3	0.022	0.041	rs874100	COSM4135166	1
0.52	0.61	0.55	0.56	0.47	0.38	0.4558	0.5871	0.5914	0.5310	0.2840	0.4344	0.4394	0.3946	0.5	0.58	0.45	0.51	0.51	rs4675	COSM149258;COSM3759056	1
0.49	0.6	0.51	0.55	0.43	0.33	0.4233	0.5750	0.5613	0.5144	0.2586	0.3978	0.4183	0.3529	0.46	0.57	0.41	0.47	0.4	rs165854	COSM149259	1

VCF Annotation: wANNOVAR

ClinVar information

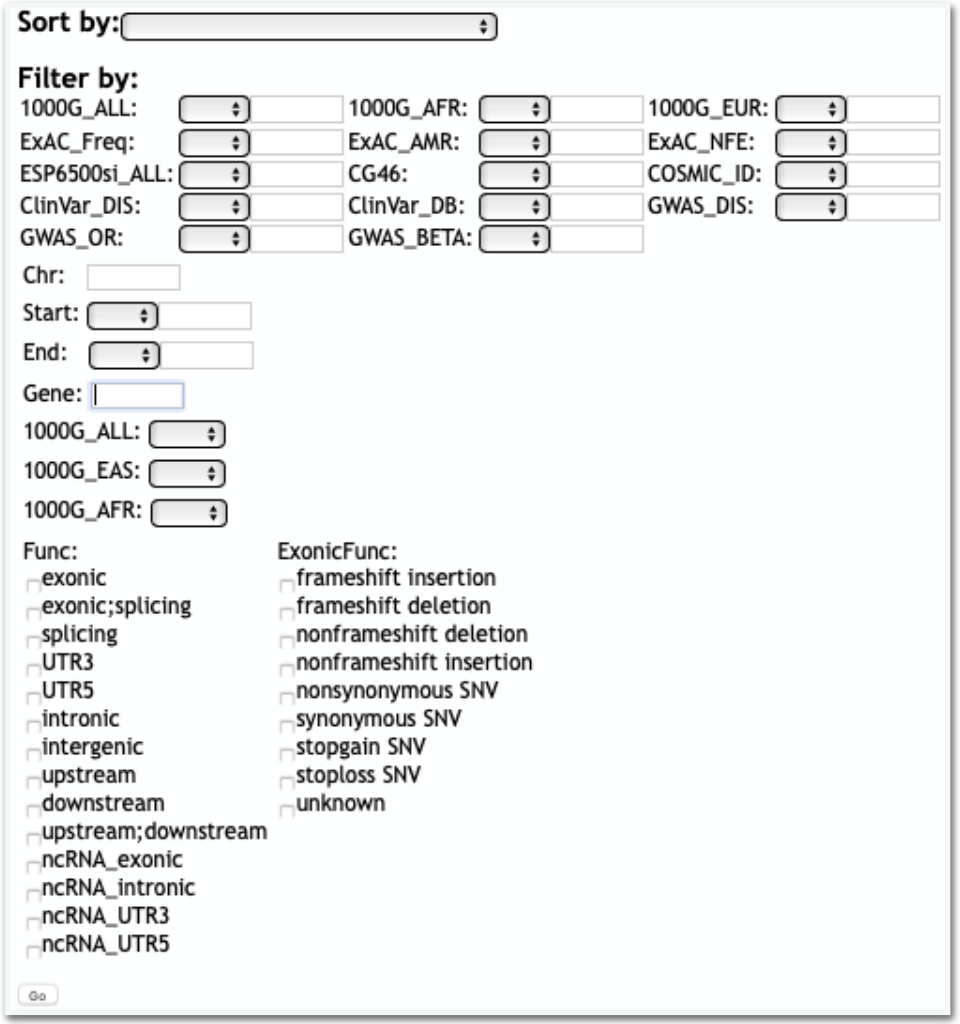
ClinVar DIS	ClinVar ID	ClinVar DB	ClinVar DBID
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
Familial_Candidiasis\x2c_Recessive	RCV000399063.1	MedGen	CN239217
Familial_Candidiasis\x2c_Recessive	RCV000385365.1	MedGen	CN239217
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
not_specified	RCV000454823.1	MedGen	CN169374
.	.	.	.
.	.	.	.
not_specified	RCV000456090.1	MedGen	CN169374

Annotation & Predictions

SIFT score	SIFT converted rankscore	SIFT pred	Polyphen2 HDIV score	Polyphen2 HDIV rankscore	Polyphen2 HDIV pred	Polyphen2 HVAR score	Polyphen2 HVAR rankscore	Polyphen2 HVAR pred	LRT score	LRT converted rankscore	LRT pred	MutationTaster score	MutationTaster converted rankscore	MutationTaster pred	MutationAssessor score	MutationAssessor score rankscore	MutationAssessor pred	FATHMM score	FATHMM converted rankscore	FATHMM pred	PROVEAN score	PROVEAN converted rankscore	PROVEAN pred	VEST3 score	VEST3 rankscore
0.329	0.132	T	0.898	0.475	P	0.094	0.286	B	0.224	0.035	U	1	0.090	P	0	0.065	N	0.17	0.605	T	-0.33	0.125	N	0.059	0.043
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1.0	0.010	T	0.0	0.026	B	0.0	0.013	B	0.001	0.396	N	1.000	0.207	P	-1.5	0.005	N	0.51	0.555	T	3.1	0.001	N	0.097	0.098
0.393	0.107	T	0.331	0.315	B	0.022	0.184	B	0.052	0.229	N	1	0.090	P	1.415	0.357	L	0.17	0.605	T	-5.49	0.857	D	0.34	0.401
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.934	0.022	T	0.0	0.026	B	0.0	0.013	B	0.204	0.165	N	1	0.090	N	0.715	0.187	N	1.7	0.269	T	-0.07	0.080	N	0.045	0.025
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.139	0.257	T	0.996	0.899	D	0.715	0.644	P	0.039	0.242	N	1	0.090	P	1.995	0.543	M	3.27	0.066	T	-0.79	0.219	N	0.075	0.065
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.202	0.784	T	0.925	0.494	P	0.257	0.376	B	0.620	0.057	N	1	0.090	P	0	0.065	N	1.81	0.282	T	-1.06	0.420	N	0.137	0.162
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.005	0.654	D	1.0	0.899	D	0.986	0.754	D	0.000	0.629	D	1.000	0.513	D	2.075	0.572	M	-1.61	0.823	D	-5.78	0.880	D	0.687	0.693
0.312	0.139	T	0.237	0.291	B	0.015	0.162	B	0.006	0.325	N	1	0.182	P	-1.28	0.007	N	-1.54	0.817	D	-1.16	0.297	N	0.085	0.088
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.002	0.721	D	0.976	0.764	D	0.482	0.627	P	0.439	0.126	N	0.158	0.351	P	2.095	0.583	M	0.7	0.515	T	-1.73	0.410	N	0.139	0.208
0.0	0.912	D	0.001	0.067	B	0.004	0.093	B	.	.	.	1	0.810	D	.	.	.	1.86	0.243	T	0.46	0.031	N	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1.0	0.010	T	0.0	0.026	B	0.0	0.013	B	0.039	0.242	N	1	0.090	N	-1.555	0.005	N	1.98	0.219	T	1.68	0.007	N	0.029	0.010
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.001	0.784	D	0.772	0.420	P	0.105	0.295	B	.	.	.	0.941	0.373	D	2.005	0.547	M	-0.64	0.721	T	-4.78	0.805	D	0.124	0.140
0.577	0.060	T	0.0	0.026	B	0.0	0.013	B	.	.	.	1	0.090	P	-0.345	0.033	N	0.04	0.621	T	0.11	0.057	N	0.016	0.003
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.229	0.183	T	0.001	0.067	B	0.001	0.040	B	.	.	.	1	0.090	P	1.7	0.440	L	-0.13	0.648	T	-0.81	0.223	N	0.016	0.277
0.571	0.062	T	0.001	0.067	B	0.001	0.040	B	.	.	.	1	0.090	P	-0.955	0.012	N	0.14	0.609	T	-0.22	0.105	N	0.007	0.000
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1.0	0.010	T	0.0	0.026	B	0.001	0.040	B	0.425	0.128	N	1	0.197	P	-0.255	0.038	N	1.0	0.414	T	-0.2	0.102	N	0.028	0.012
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.224	0.187	T	0.002	0.090	B	0.002	0.063	B	0.024	0.264	N	0.000	0.588	P	-0.345	0.033	N	-0.49	0.704	T	-0.84	0.229	N	0.061	0.045
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.266	0.162	T	0.002	0.090	B	0.004	0.093	B	0.005	0.333	N	0.955	0.379	D	0.885	0.217	L	1.87	0.241	T	-1.25	0.316	N	0.015	0.002
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
0.178	0.221	T	0.695	0.398	P	0.098	0.289	B	0.001	0.420	D	0.961	0.382	D	1.91	0.513	L	0.96	0.429	T	0.11	0.057	N	0.06	0.044
0.022	0.486	D	0.0	0.026	B	0.0	0.013	B	0.207	0.165	N	1	0.090	P	0.695	0.181	N	-2.8	0.911	D	-0.31	0.121	N	0.103	0.111
0.046	0.405	D	0.176	0.561	B	0.314	0.807	B	0.000	0.843	D	0.000	0.588	P	2.465	0.718	M	2.55	0.138	T	-2.15	0.486	N	0.624	0.652

VCF Annotation: wANNOVAR

Scroll Down the Page



VCF Annotation: wANNOVAR

Sample identifier = RA3
File_name=Sample_AD_RA3_Chr22.vcf.gz
File_format=vcf4
Reference_genome=hg38
Disease_model=rare recessive disease
Processed variants=11361
phenotype is cardiac arrhythmias

Basic Information

exome summary results	view	CSV file	TXT file
genome summary results	view	CSV file	TXT file

ANNOVAR filtering results:

(click to view details about this pipeline)
Initially, 11361 variants were fed into the annotation pipeline and 0 variants were detected as invalid input.

Step1:254 variants	Identify missense, nonsense and splicing variants	download
Step2:25 variants	Remove variants in the 1000 Genomes Project(ALL) with MAF>0.01	download
Step3:22 variants	Remove variants in gnomAD exome database with MAF>0.01	download
Step4:4 genes	Compile a list of candidate genes based on diseaes model	download variants_info

download all filtering results

Phenotype/disease Prioritization Result:

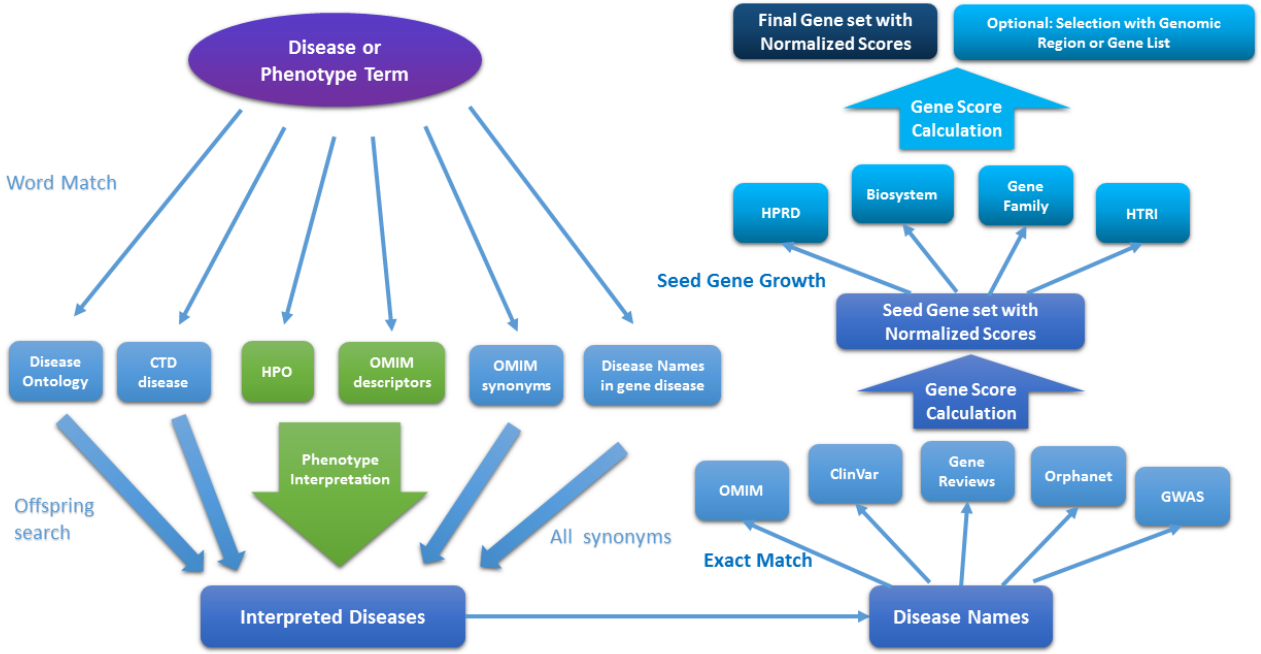
Exonic variant list from the wANNOVAR output after filtration, the last filter step with variants left will be used. (Total: 4)	Variant List
Gene list from the wANNOVAR output, input into Phenolyzer (Total: 4)	Input Gene List
The prioritized genes from Phenolyzer (Total: 3)	Result Gene List
The network visualization	Show

VCF Annotation: wANNOVAR + Phenolyzer

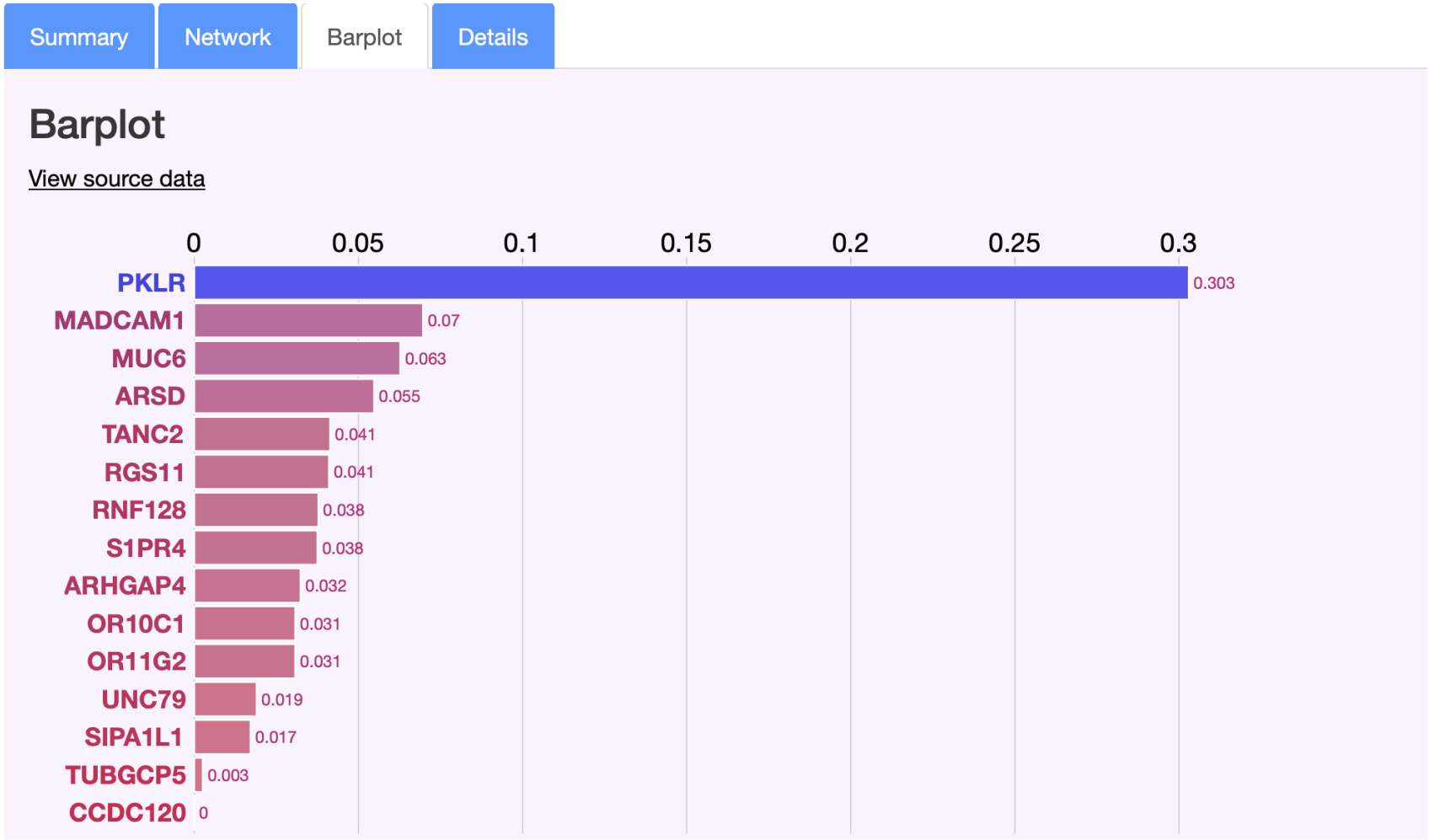
Phenolyzer

Phenotype Based Gene Analyzer, un outil axé sur la découverte de gènes à partir de termes de maladies/phénotypes spécifiques défini par l'utilisateur

Website: http://phenolyzer.wglab.org



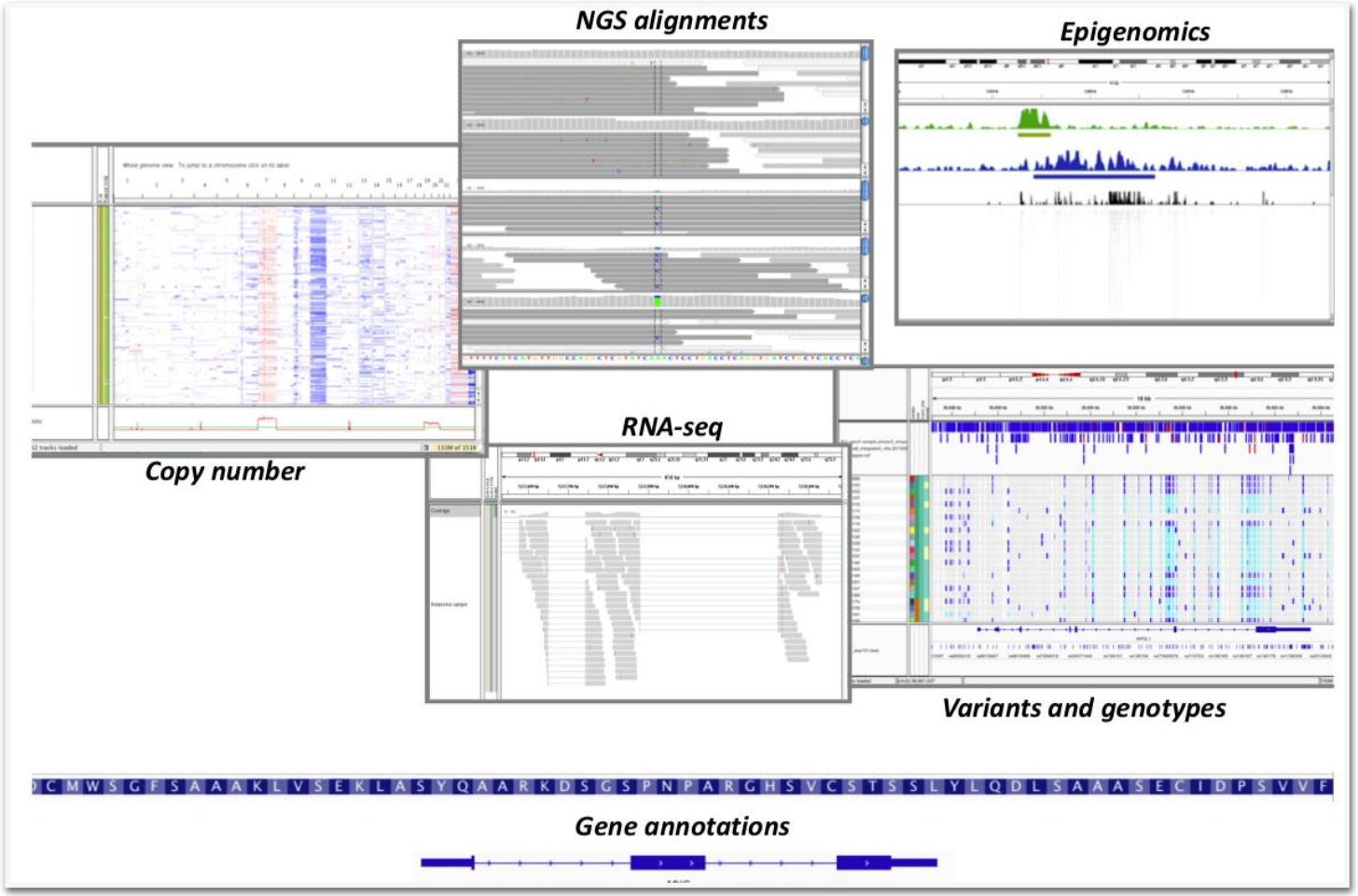
VCF Annotation: wANNOVAR + Phenolyzer



Phenolyzer

Integrative Genomics Viewer (IGV)

L'Integrative Genomics Viewer (IGV) est un outil de visualisation très performant pour l'exploration interactive de grands ensembles de données génomiques intégrées de différents types



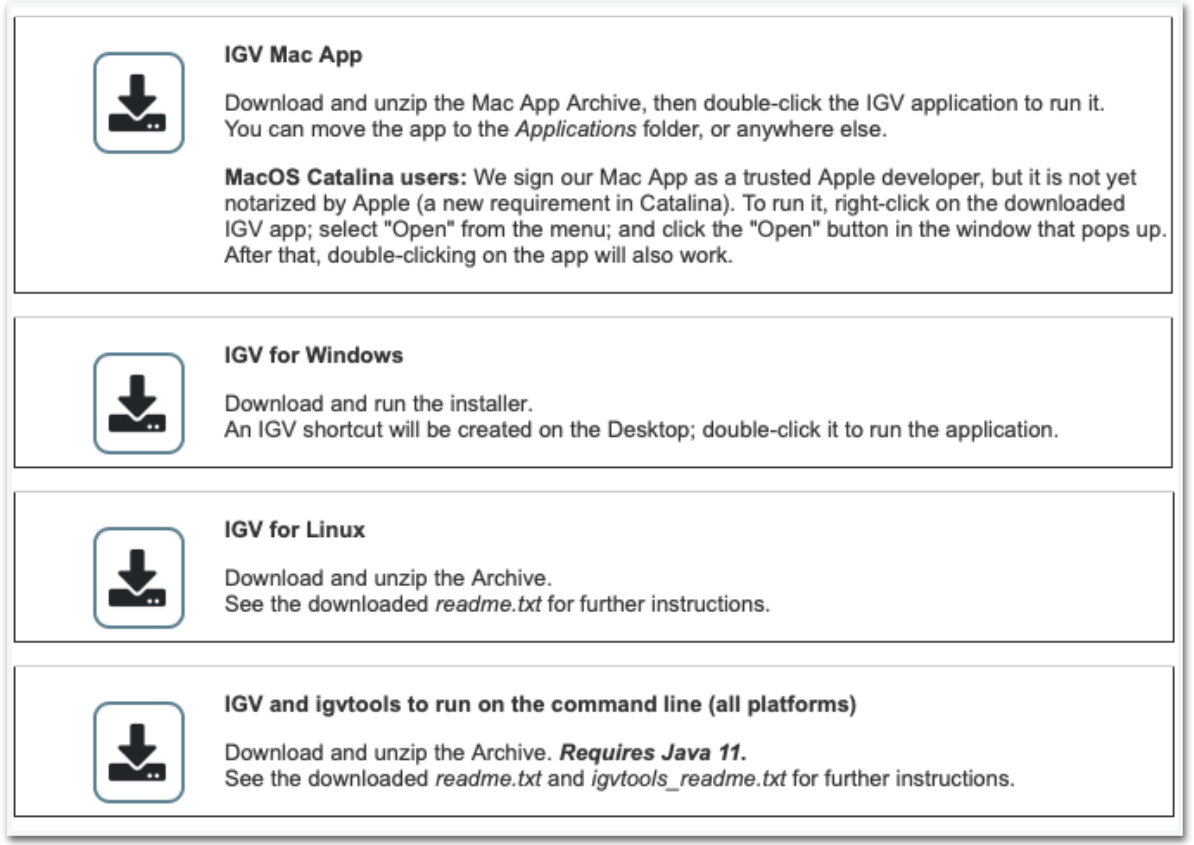
Integrative Genomics Viewer (IGV)

File Formats

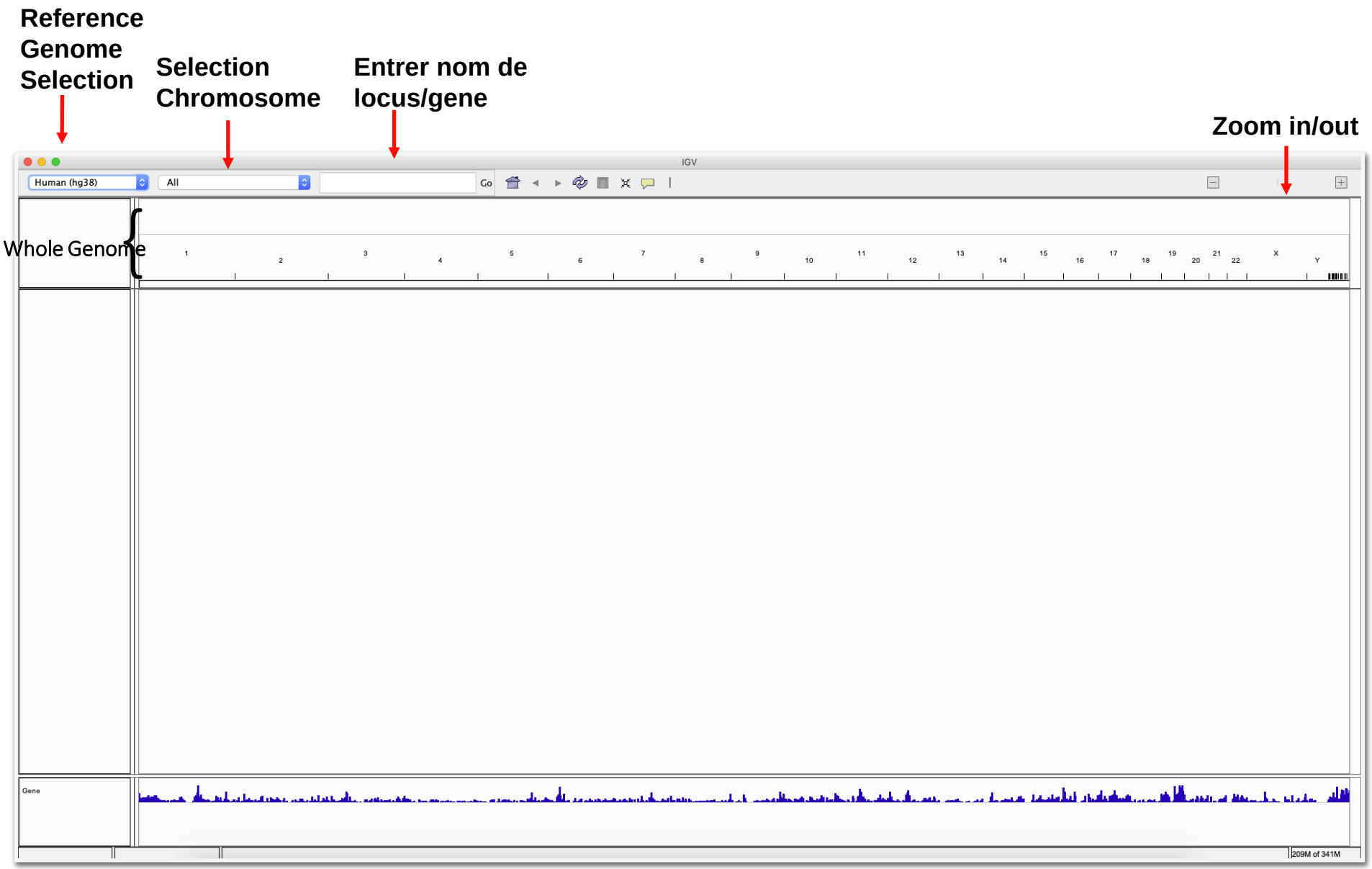
- BAM
- BED
- BEDPE
- BedGraph
- bigBed
- bigWig
- Birdsuite Files
- broadPeak
- CBS
- Chemical Reactivity Probing Profiles
- chrom.sizes
- CN
- Custom File Formats
- Cytoband
- FASTA
- GCT
- CRAM
- genePred
- GFF/GTF
- GISTIC
- Goby
- GWAS
- IGV
- LOH
- MAF (Multiple Alignment Format)
- MAF (Mutation Annotation Format)
- Merged BAM File
- MUT
- narrowPeak
- PSL
- RES
- RNA Secondary Structure Formats
- SAM
- Sample Info (Attributes) file
- SEG
- SNP
- TAB
- TDF
- Track Line
- Type Line
- VCF
- WIG

Download & install

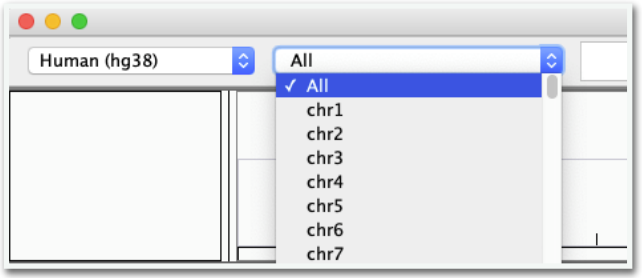
http://software.broadinstitute.org/software/igv/download



Integrative Genomics Viewer (IGV)



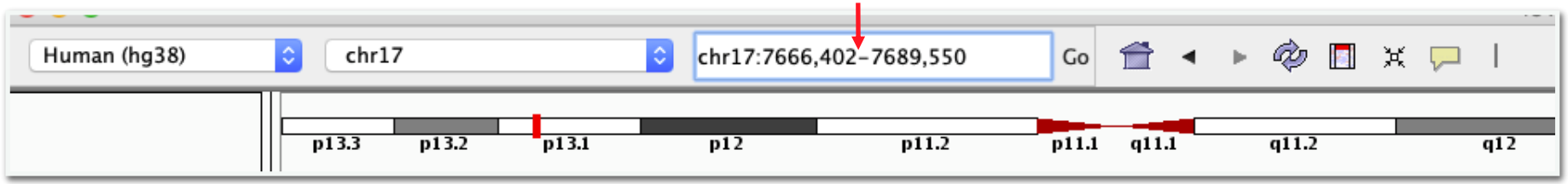
IGV Basic panels: SelectionChromosome



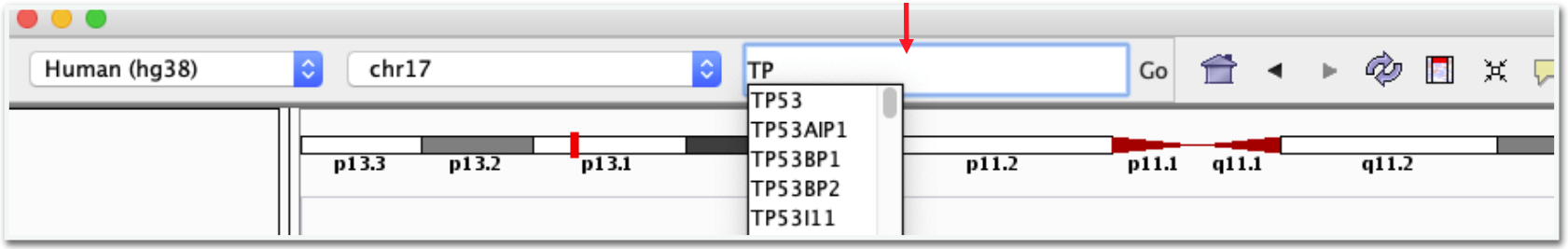
Affiche tous les chromosomes du génome de référence sélectionné

Panels Basic IGV : nom Locus/gène

Chromosome:Start-end



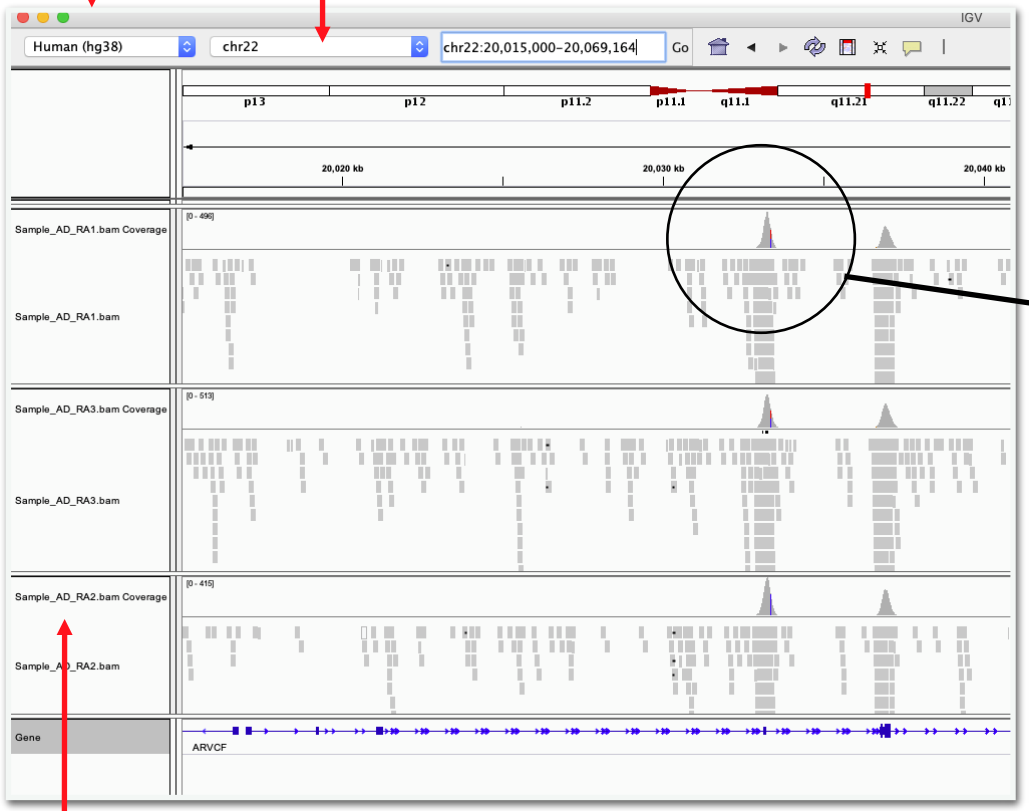
Entrer nom du Gène



IGV: Vue du Gène d'intérêt

Reference Genome Selection

Chromosome Selection



BAM FILES

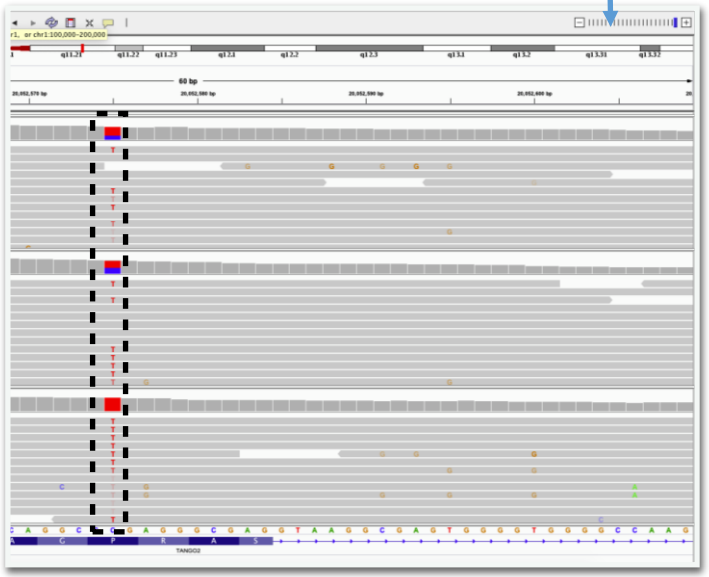
Coverage

Aligned reads mapped to the location

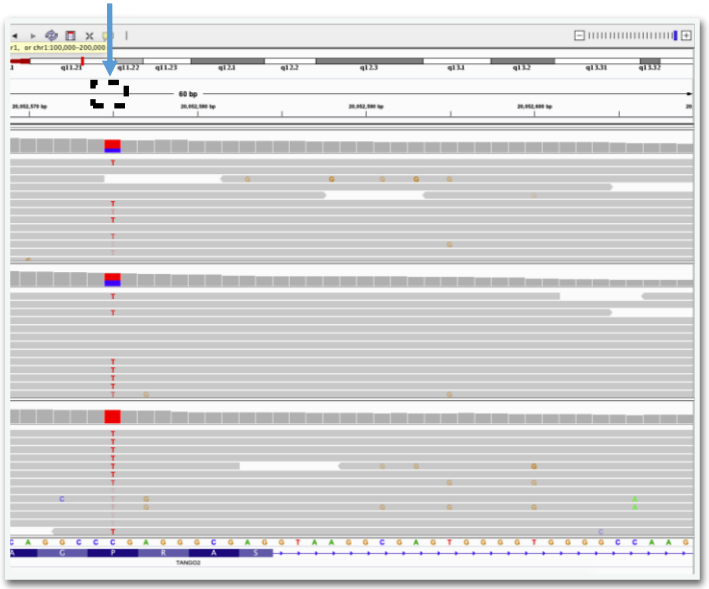
Variation

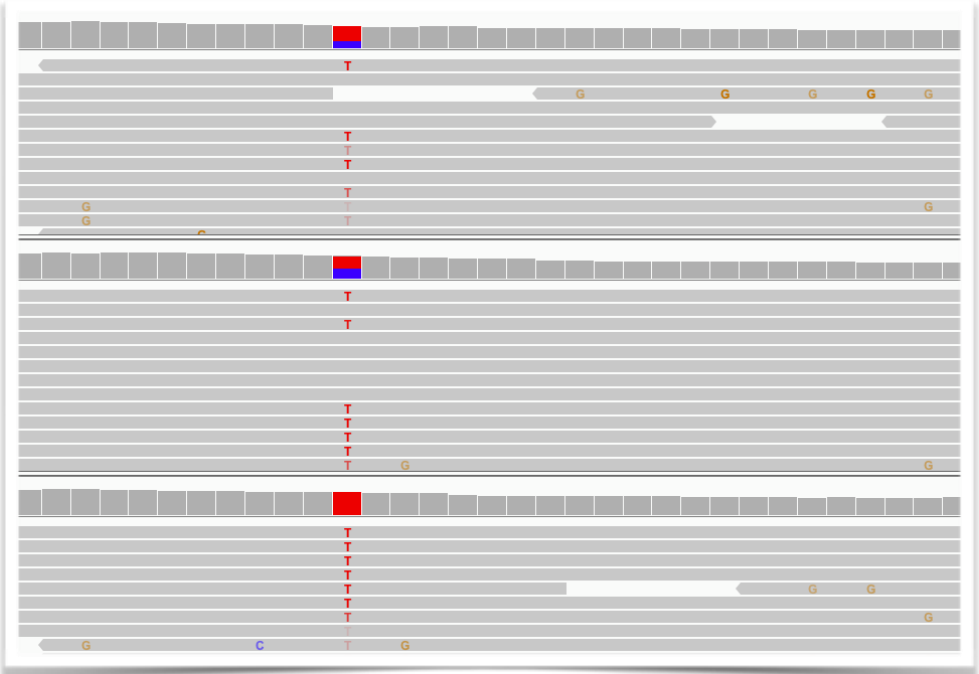
IGV: Zoom

Zoom in/out

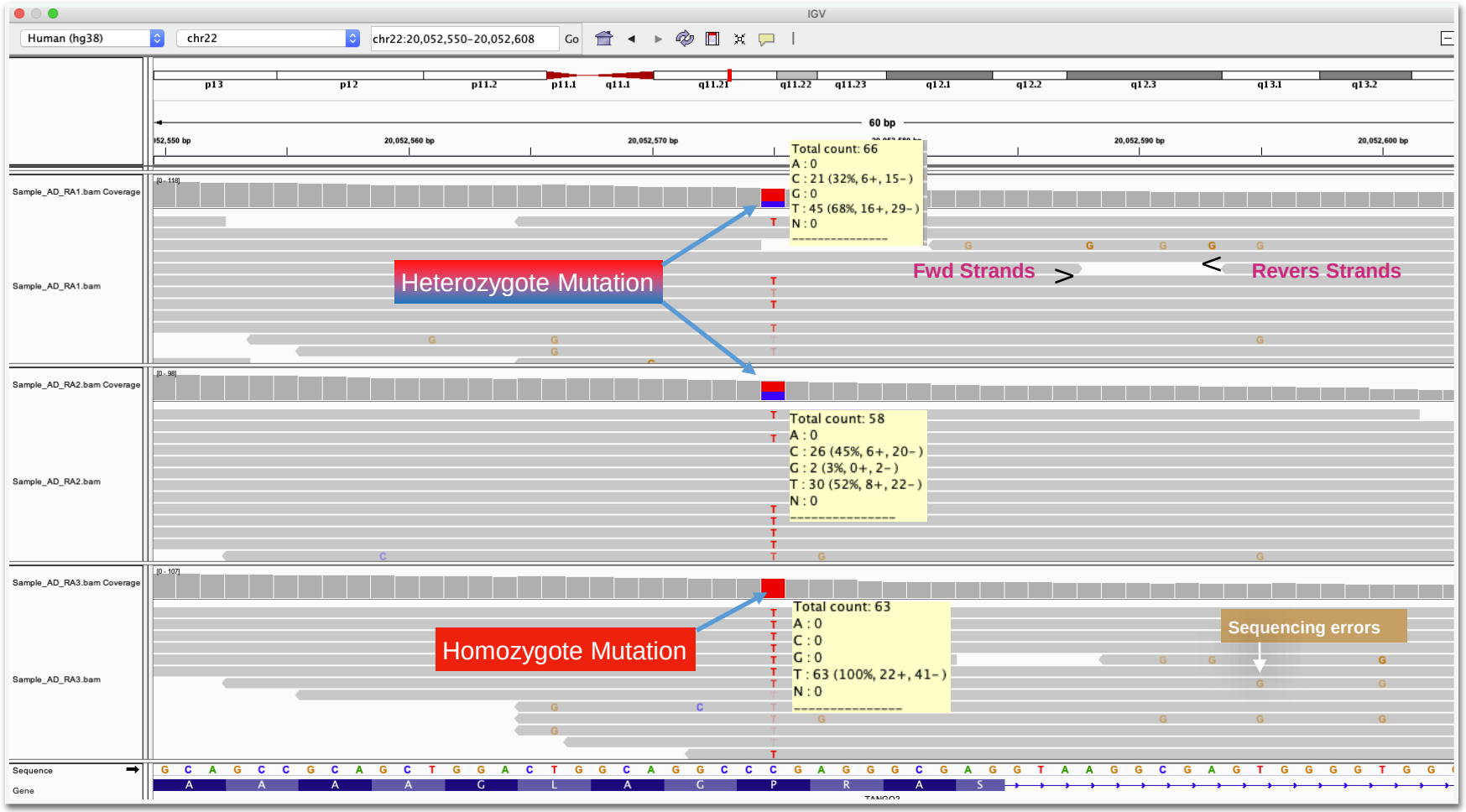


Click and Drag



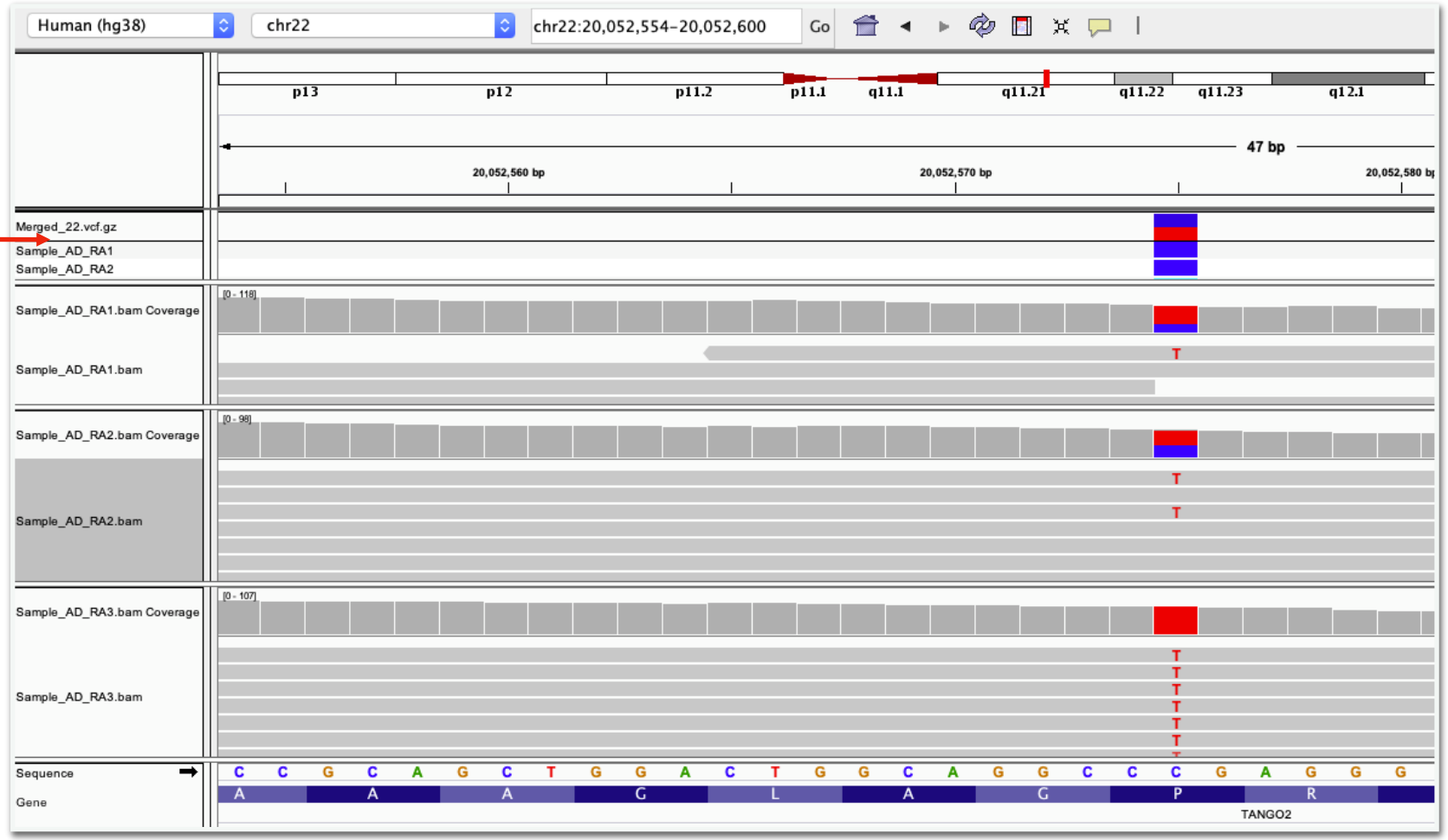


IGV: Zoom


IGV
Human (hg38)
chr22
chr22:20,052,550-20,052,608
Go
60 bp
Sample_AD_RA1.bam Coverage
Sample_AD_RA1.bam
Total count: 66
A : 0
C : 21 (32%, 6+, 15-)
G : 0
T : 45 (68%, 16+, 29-)
N : 0
Heterozygote Mutation
Fwd Strands
Revers Strands
Sample_AD_RA2.bam Coverage
Sample_AD_RA2.bam
Total count: 58
A : 0
C : 26 (45%, 6+, 20-)
G : 2 (3%, 0+, 2-)
T : 30 (52%, 8+, 22-)
N : 0
Sample_AD_RA3.bam Coverage
Sample_AD_RA3.bam
Total count: 63
A : 0
C : 0
G : 0
T : 63 (100%, 22+, 41-)
N : 0
Homozygote Mutation
Sequencing errors
Sequence
Gene

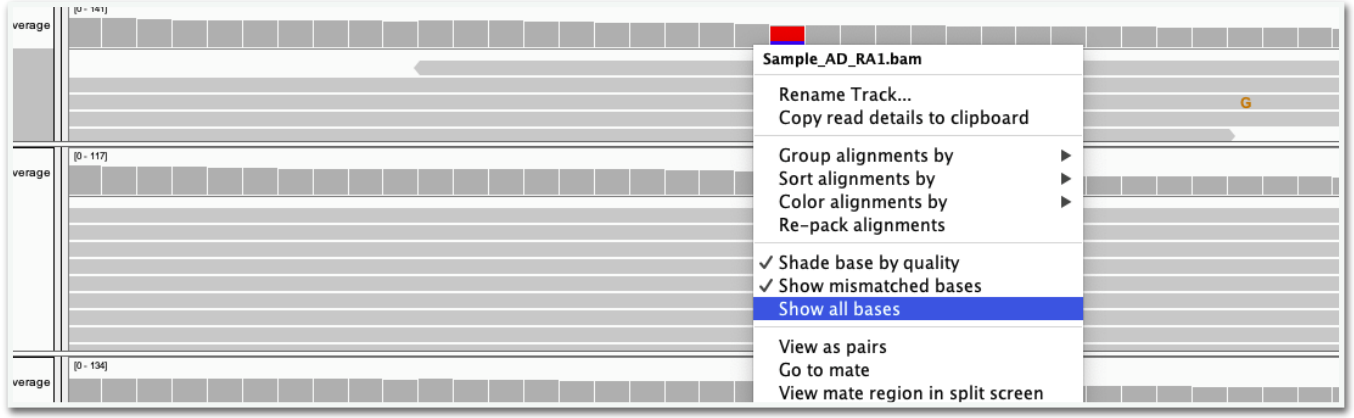
IGV: Charger le VCF

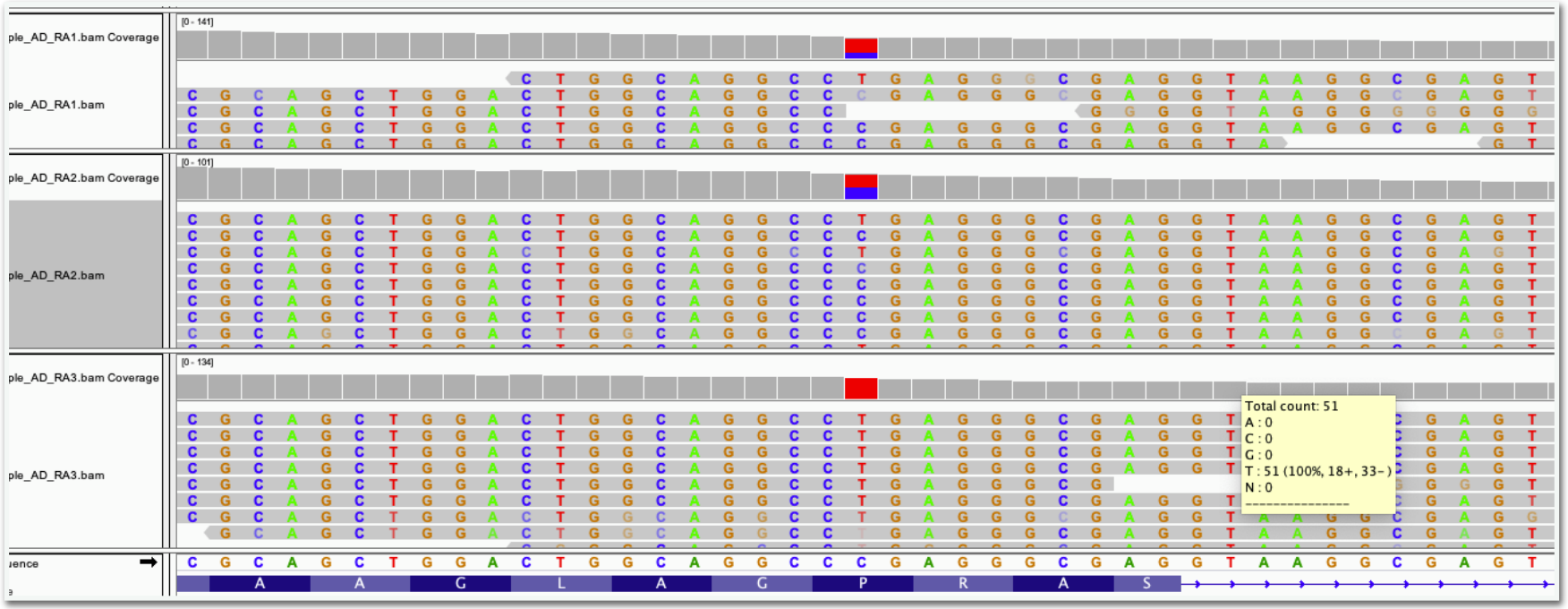
VCF file is loaded as new tracks



IGV: View All bases

Step1:





Merci