

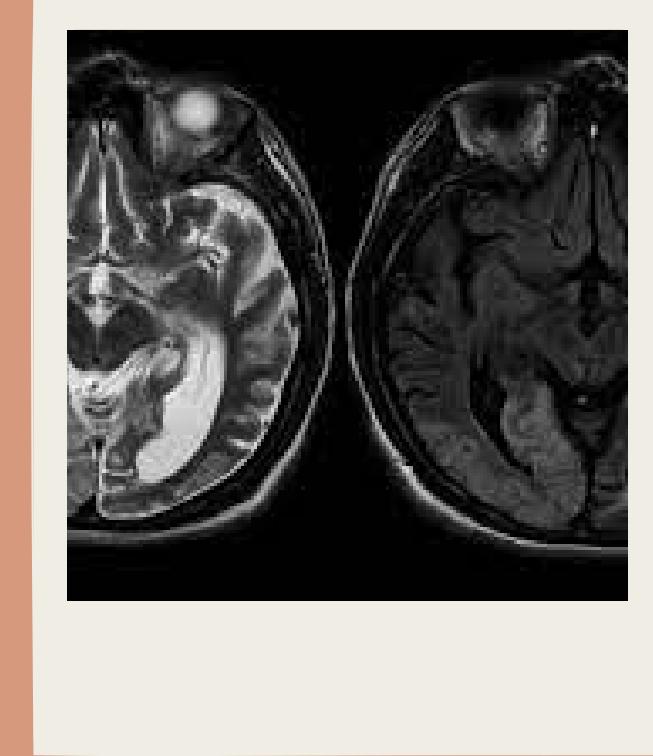
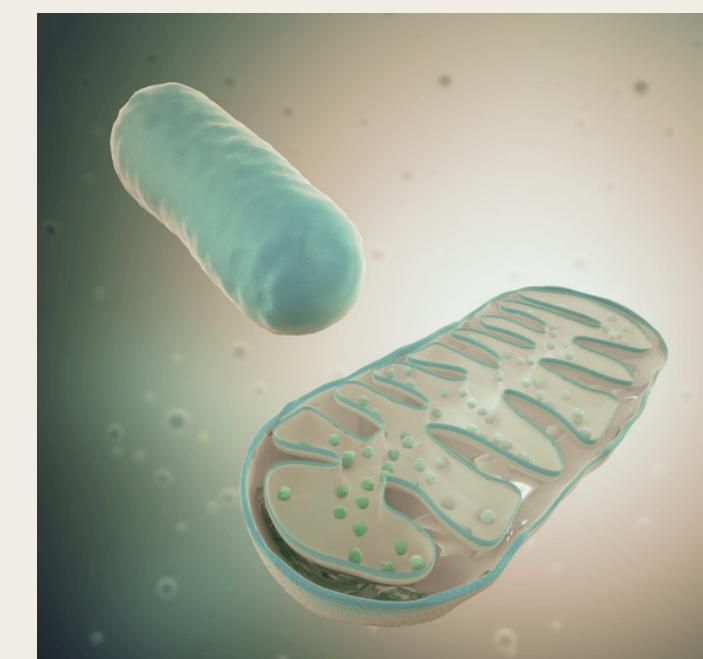
# Variabilidad en el genoma mitocondrial humano en personas sanas

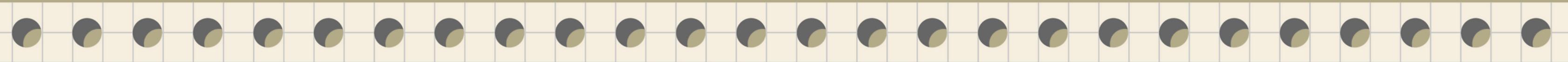


Mayela Fosado

# Mutaciones puntuales Deleciones Inserciones pequeñas

Enfermedades neuromusculares,  
encefalopatías: neuropatía óptica  
hereditaria de Leber, síndrome MELAS,  
síndrome de Leigh y diabetes con sordera



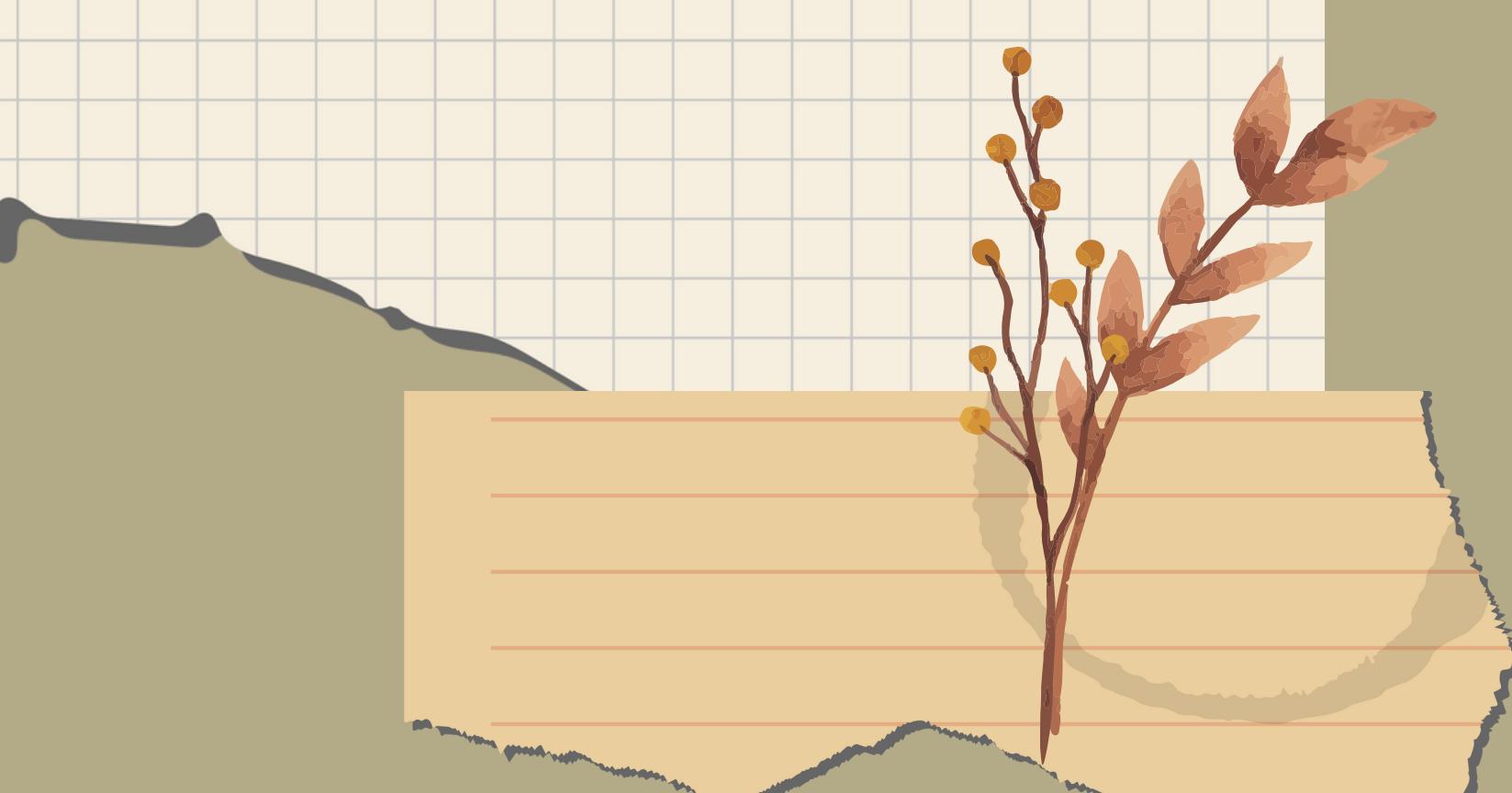
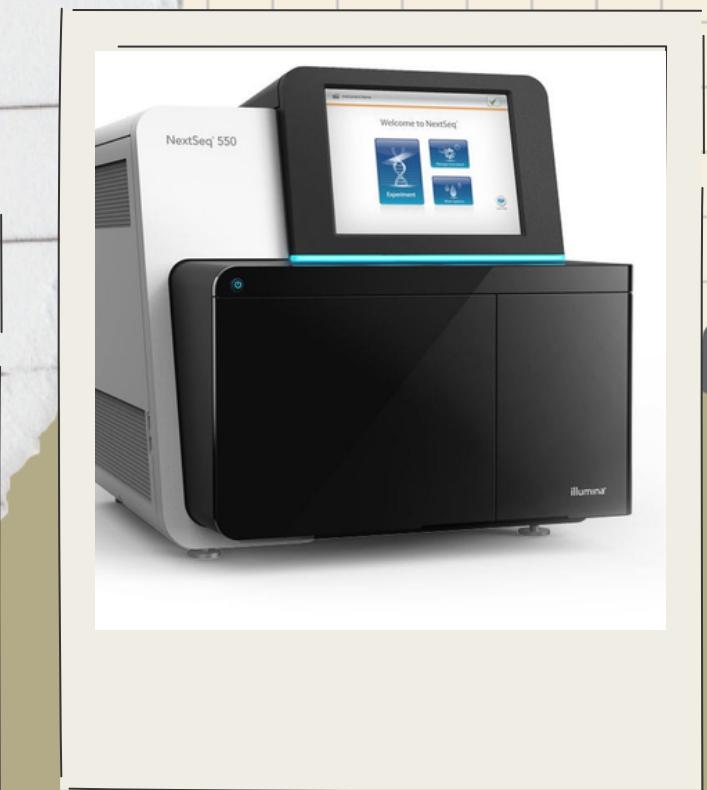
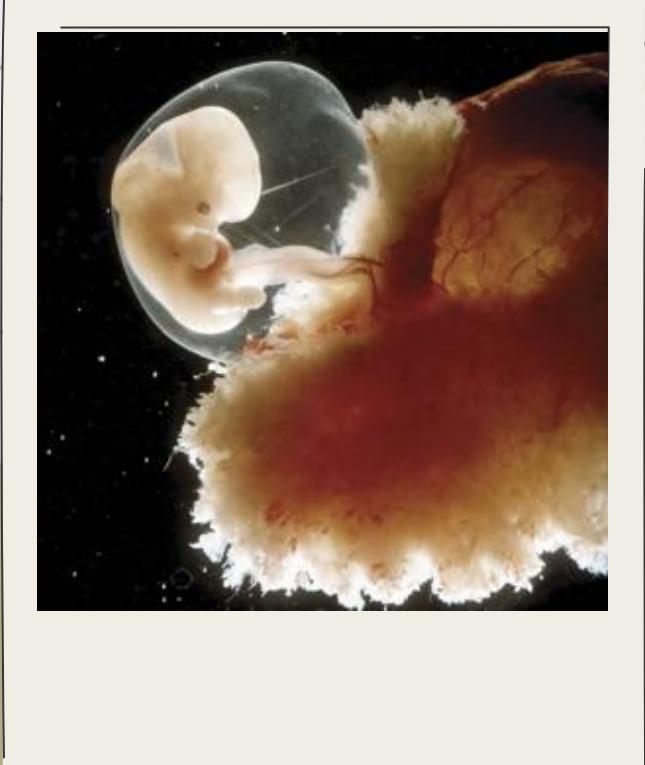
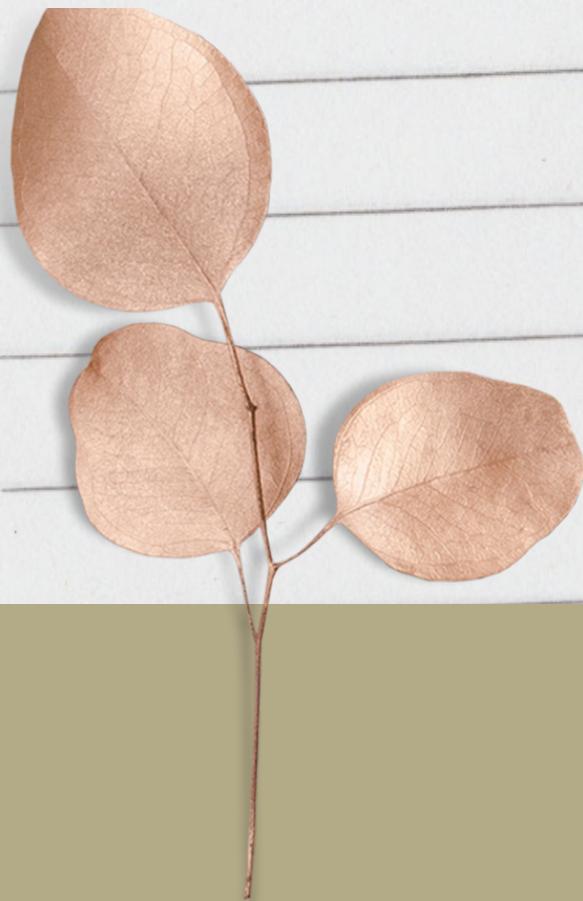


# Obtención de muestras

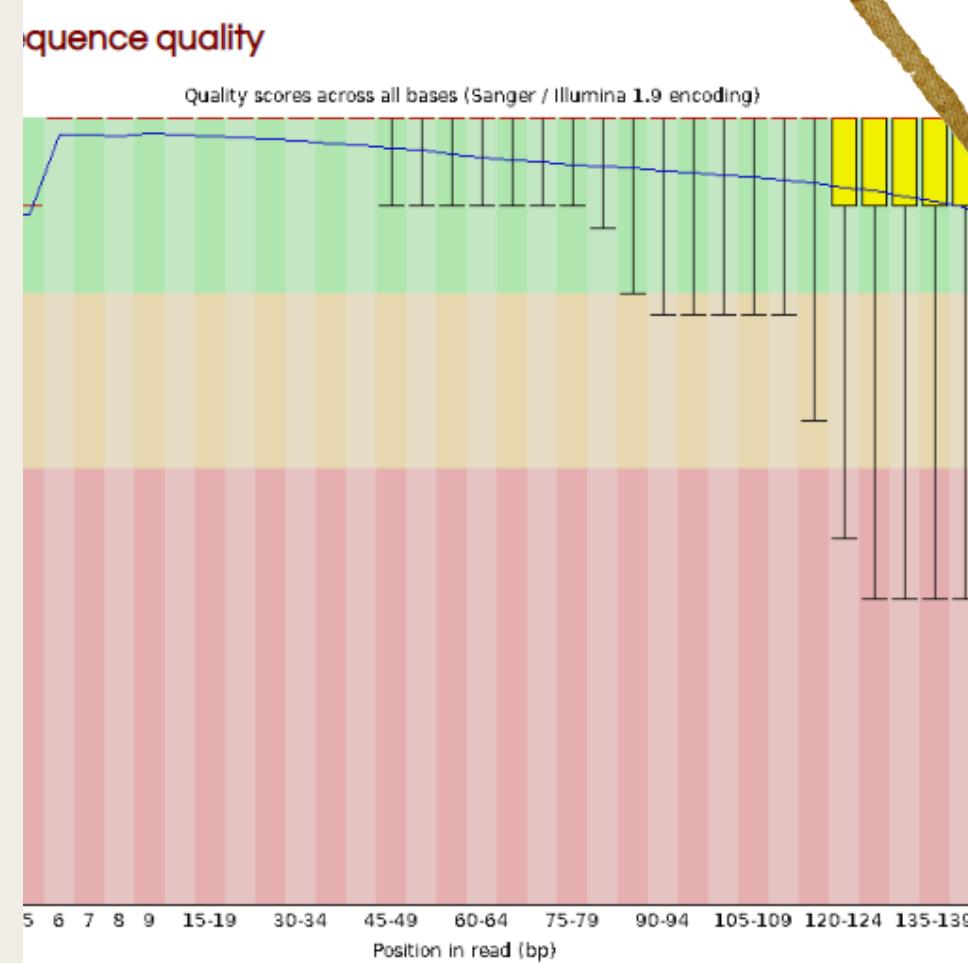
Muestra de placenta

Clínica Cer, Corregidora, Qro.

Illumina NextSeq 550, Duke University



# FastQC



Revisar la calidad  
de las secuencias

# FastQC

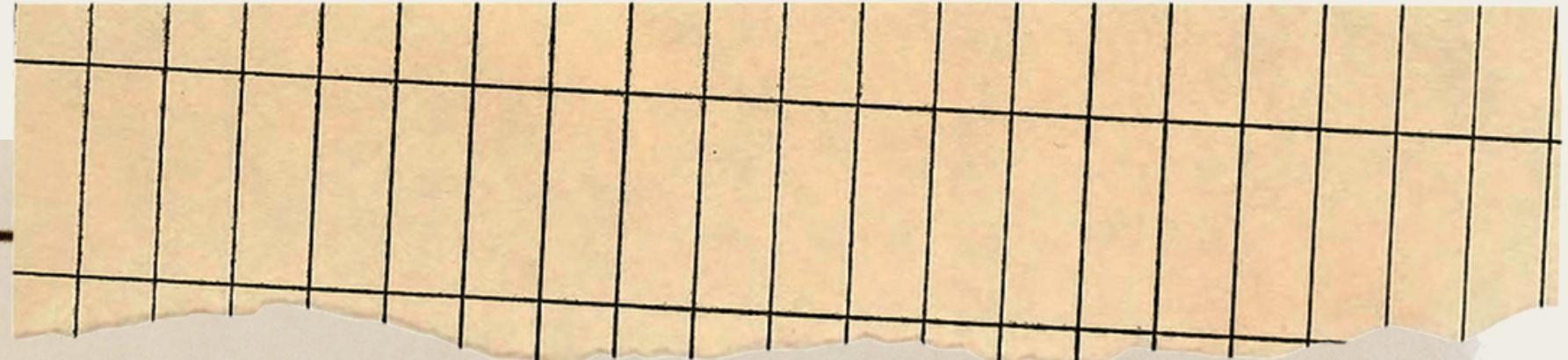
```
#$ -S /bin/bash
#
# Your job name
#$ -N fastQC
#
# Use current working directory
#$ -cwd
#
# Join stdout and stderr
#$ -j y
#
# Send an email after the job has finished
#$ -m e
#$ -M maye.fosado@gmail.com
#
# If modules are needed, source modules environment (Do not delete the next line):
. /etc/profile.d/modules.sh
#
# Add any modules you might require:
module load fastqc/0.11.3
#
#pe (Parallel environment)request a parallel environment (MPI, OpenMP). Set your number of requested slots here.
#A continuación, número de hilos pedidos a máquina
#$ -pe mpich 1
#
# Write your commands in the next line
fastqc /mnt/Timina/alfredvar/alfredvar/seqdata/SENASICA_7/BC50_S37_R1_001.fastq
```

# Trimmomatic

Eliminar secuencias de

Baja calidad

Recortar extremos de baja  
calidad



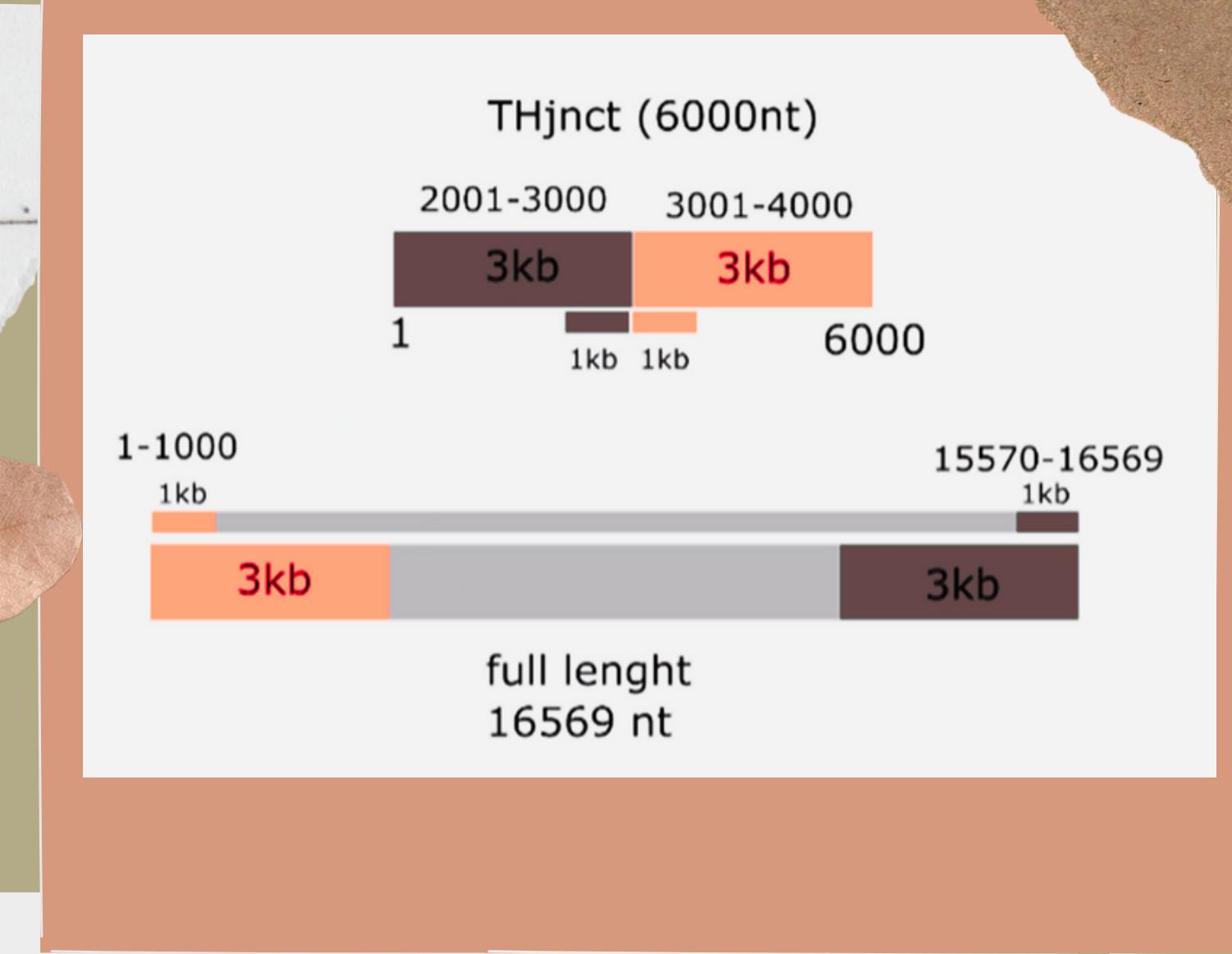
# Trimmomatic

```
t$ -S /bin/bash
t
t Your job name
t$ -N trimm.B
t
t Use current working directory
t$ -cwd
t
t Join stdout and stderr
t$ -j y
t
t
t If modules are needed, source modules environment (Do not delete the next line):
. /etc/profile.d/modules.sh
t
t Add any modules you might require:
module load trimmomatic/0.33
t
tpe (Parallel environment)request a parallel environment (MPI, OpenMP). Set your number of requested slots here.
tA continuación, número de hilos pedidos a máquina
t$ -pe mpich 1
t
t Write your commands in the next line

trimmomatic PE -trimlog trimlog.BC50 BC50_S37_R1_001.fastq BC50_S37_R2_001.fastq BC50_R1_trimP.fastq BC50_R1_trimU.fastq BC50_R2_trimP.fastq BC50_R2_trimU.fastq HEADCROP:9
trimmomatic PE -trimlog trimlog.BC51 BC51_S38_R1_001.fastq BC51_S38_R2_001.fastq BC51_R1_trimP.fastq BC51_R1_trimU.fastq BC51_R2_trimP.fastq BC51_R2_trimU.fastq HEADCROP:9
trimmomatic PE -trimlog trimlog.BC52 BC52_S39_R1_001.fastq BC52_S39_R2_001.fastq BC52_R1_trimP.fastq BC52_R1_trimU.fastq BC52_R2_trimP.fastq BC52_R2_trimU.fastq HEADCROP:9
trimmomatic PE -trimlog trimlog.BC54 BC54_S41_R1_001.fastq BC54_S41_R2_001.fastq BC54_R1_trimP.fastq BC54_R1_trimU.fastq BC54_R2_trimP.fastq BC54_R2_trimU.fastq HEADCROP:9
trimmomatic PE -trimlog trimlog.BC55 BC55_S42_R1_001.fastq BC55_S42_R2_001.fastq BC55_R1_trimP.fastq BC55_R1_trimU.fastq BC55_R2_trimP.fastq BC55_R2_trimU.fastq HEADCROP:9
trimmomatic PE -trimlog trimlog.BC56 BC56_S43_R1_001.fastq BC56_S43_R2_001.fastq BC56_R1_trimP.fastq BC56_R1_trimU.fastq BC56_R2_trimP.fastq BC56_R2_trimU.fastq HEADCROP:9
trimmomatic PE -trimlog trimlog.BC57 BC57_S44_R1_001.fastq BC57_S44_R2_001.fastq BC57_R1_trimP.fastq BC57_R1_trimU.fastq BC57_R2_trimP.fastq BC57_R2_trimU.fastq HEADCROP:9
trimmomatic PE -trimlog trimlog.BC58 BC58_S45_R1_001.fastq BC58_S45_R2_001.fastq BC58_R1_trimP.fastq BC58_R1_trimU.fastq BC58_R2_trimP.fastq BC58_R2_trimU.fastq HEADCROP:9
```



# Segemehl



Alineamiento de las secuencias al genoma  
de referencia

```
module load segemehl/0.3.4

segemehl.x -x chrM.hg38A.idx -d chrM.hg38A.fa
segemehl.x -x chrM.hg38.THjnct.idx -d chrM.hg38.THjnct.fa
```

# Segemehl

```
#!/bin/bash
#
# Your job name
#$ -N BC50j
#
# Use current working directory
#$ -cwd
#
# Join stdout and stderr
#$ -j y
#
# If modules are needed, source modules environment (Do not delete the next line):
. /etc/profile.d/modules.sh
#
# Add any modules you might require:
module load segemehl/0.3.4
#
#pe (Parallel environment)request a parallel environment (MPI, OpenMP). Set your number of requested slots here.
#A continuación, número de hilos pedidos a máquina
#$ -pe mpich 8
#
# Write your commands in the next line
segemehl.x -S -b -t 8 -g RGmt -d chrM.hg38.THjncf.fa -i chrM.hg38.THjncf.idx -q /mnt/Timina/alfredvar/alfredvar/seqdata/SENASICA_7/BC50_R1_trimP.fastq -p /mnt/Timina/alfredvar/alfredvar/seqdata/SENASICA_7/BC50_R2_trimP.fastq -o BC50j.bam
```

# Manejo de datos & Análisis de datos

## Samtools

```
#$ -N samtools.B
#
# Use current working directory
#$ -cwd
#
# Join stdout and stderr
#$ -j y
#
#
# If modules are needed, source modules environment (Do not delete the next line):
. /etc/profile.d/modules.sh
#
# Add any modules you might require:
module load samtools/1.3.1
#
#pe (Parallel environment)request a parallel environment (MPI, OpenMP). Set your number of requested slots here.
#A continuación, número de hilos pedidos a máquina
#$ -pe mpich 1
#
# Write your commands in the next line
samtools sort BC50A.bam -o BC50A.bam.sort
samtools sort BC50.j.bam -o BC50.j.bam.sort
samtools sort BC51A.bam -o BC51A.bam.sort
samtools sort BC51.j.bam -o BC51.j.bam.sort
samtools sort BC52A.bam -o BC52A.bam.sort
samtools sort BC52.j.bam -o BC52.j.bam.sort
samtools sort BC54A.bam -o BC54A.bam.sort
samtools sort BC54.j.bam -o BC54.j.bam.sort
samtools sort BC55A.bam -o BC55A.bam.sort
samtools sort BC55.j.bam -o BC55.j.bam.sort
samtools sort BC56A.bam -o BC56A.bam.sort
samtools sort BC56.j.bam -o BC56.j.bam.sort
samtools sort BC57A.bam -o BC57A.bam.sort
samtools sort BC57.j.bam -o BC57.j.bam.sort
samtools sort BC58A.bam -o BC58A.bam.sort
samtools sort BC58.j.bam -o BC58.j.bam.sort
```

# Mutect2

```
module load gatk/4.2.3.0

gatk CreateSequenceDictionary -R chrM.hg38A.fa
gatk CreateSequenceDictionary -R chrM.hg38.THjncf.fa
```

## Identificación de variantes



# Mutect2

```
#$ -S /bin/bash
#
# Your job name
#$ -N BC50A.mt2
#
# Use current working directory
#$ -cwd
#
# Join stdout and stderr
#$ -j y
#
# If modules are needed, source modules environment (Do not delete the next line):
. /etc/profile.d/modules.sh
#
# Add any modules you might require:
module load gatk/4.1.6.0
#
#pe (Parallel environment)request a parallel environment (MPI, OpenMP). Set your
number of requested slots here.
#A continuación, número de hilos pedidos a máquina
#$ -pe mpich 4
#
# Write your commands in the next line
gatk Mutect2 \
-R chrM.hg38A.fa \
-I BC50A.bam.sort \
-O BC50A.vcf.gz \
-L chrM \
--mitochondria-mode \
--active-probability-threshold 0.000001 \
--base-quality-score-threshold 6 \
--callable-depth 5 \
--downsampling-stride 1 \
--fir2-max-depth 1000000 \
--force-active \
--max-mnp-distance 0 \
--max-population-af 1 \
--max-reads-per-alignment-start 0 \
--min-base-quality-score 30 \
--dont-use-soft-clipped-bases \
--phred-scaled-global-read-mismapping-rate 30
```

# Filtrar datos obtenidos

```
#!/bin/bash
# this filters mutect2 output of mapping to TH junction, selects single nucleotide events, inverts abundant alleles if
necesary, eliminates strand bias (0.66), selects >=10/D, selects 1kb from each tail and head correcting nucleotide co
ordinates, calculates allele frequency

for i in $(ls -1 BC5?j.vcf.gz | awk -F '\t' '{print $1}')
do
zcat $i.vcf.gz |\
grep ^chrM |\
awk '{print $2"\t"$4"\t"$5"\t"$10}' |\
awk -F ':' '{print $1"\t"$2"\t"$4"\t"$7}' |\
grep ' 0/1 '|\
awk -F ',' '{print $1"\t"$2"\t"$3"\t"$4"\t"$5}' |\
awk '{if ($5 < $6) print
$1"\t"$3"\t"$2"\t"$4"\t"$6"\t"$5"\t"$7"\t"$10"\t"$11"\t"$8"\t"$9; else print
$0}' |\
awk '($1>1000) && ($1 < 15570) && ($11 ? $10/$11<=1.5 : 0) && ($10 ? $11/$10<=1.5 :
0) && ($8/$9<=1.5) && ($9/$8<=1.5) && ($10>=5 && $11 >=5 {print $0"\t"$6/$7})'
$1.fltrd.1
done
```

```
#!/bin/bash
# this filters mutect2 output of full A lenght mapping, selects single nucleotide events, inver
necesary, eliminates strand bias (0.66), selects >=10/D, eliminates 1kb from each tail and head
quency

for i in $(ls -1 BC5?A.vcf.gz | awk -F '\t' '{print $1}')
do
zcat $i.vcf.gz |\
grep ^chrM |\
awk '{print $2"\t"$4"\t"$5"\t"$10}' |\
awk -F ':' '{print $1"\t"$2"\t"$4"\t"$7}' |\
grep ' 0/1 '|\
awk -F ',' '{print $1"\t"$2"\t"$3"\t"$4"\t"$5}' |\
awk '{if ($5 < $6) print
$1"\t"$3"\t"$2"\t"$4"\t"$6"\t"$5"\t"$7"\t"$10"\t"$11"\t"$8"\t"$9; else print
$0}' |\
awk '($1>1000) && ($1 < 15570) && ($11 ? $10/$11<=1.5 : 0) && ($10 ? $11/$10<=1.5 :
0) && ($8/$9<=1.5) && ($9/$8<=1.5) && ($10>=5 && $11 >=5 {print $0"\t"$6/$7})'
$1.fltrd.1
done
```

# Juntar variantes

```
grep "" BC5*fltrd.l | awk '$12>0.0025 {print $0}' | sed 's/.fltrd.l:/ /g' | sort -k2n,2 > z_Finalists.human
```

BC54j	146	T	C	0/1	27647	372	28819	12626	15821	163	289	0.0132767
BC55j	146	Ci	Ti	0/1	5011	15	5026	2348	2663	9	6	0.00298448
BC54j	153	A	G	0/1	26307	386	26693	12186	14121	178	288	0.0144607
BC54j	235	A	G	0/1	12153	326	12479	5721	6432	160	166	0.0261239
BC55j	235	Gi	Ai	0/1	4373	11	4384	2312	2861	6	5	0.00250912
BC54j	247	Gi	GAi	0/1	9294	277	9571	5313	3981	155	122	0.0289416
BC54j	663	A	G	0/1	20312	646	20958	11989	8323	359	287	0.0308236
BC54A	1736	A	G	0/1	33927	1162	35089	13973	19954	469	693	0.0331158
BC58A	3196	G	A	0/1	36962	146	37108	16691	28271	78	76	0.00393446
BC58A	3669	A	G	0/1	26435	147	26582	13054	13381	77	78	0.00553006
BC58A	3849	G	A	0/1	43400	218	43610	23081	28319	112	98	0.00481541
BC54A	4248	T	C	0/1	30257	1846	31303	16730	13527	579	467	0.0334153
BC55A	4248	Ci	Ti	0/1	13592	41	13633	7657	5935	23	18	0.00300741
BC55A	4769	Gi	Ai	0/1	3202	11	3213	1470	1732	6	5	0.00342359
BC54A	4824	A	G	0/1	25154	446	25600	14257	18897	210	236	0.0174219
BC56A	6504	G	A	0/1	43812	529	44341	23522	28290	261	268	0.0119303
BC58A	6821	Ti	Ci	0/1	23429	15869	39298	12159	11270	8264	7605	0.403812
BC54A	7003	T	C	0/1	31493	1483	32976	13104	18389	659	824	0.0449721
BC57A	7138	T	C	0/1	45995	237	46232	26408	19587	131	186	0.00512632
BC54A	7196	Ai	Ci	0/1	24090	973	25063	12540	11550	485	488	0.0388222
BC54A	7724	A	T	0/1	30394	1165	31559	15170	15224	559	606	0.036915
BC55A	7724	Ti	Ai	0/1	14696	46	14742	7317	7379	27	19	0.00312034
BC54A	7859	G	A	0/1	27383	944	28327	14081	13302	468	476	0.0333251
BC54A	8027	G	A	0/1	26190	963	27153	15018	11172	525	438	0.0354657
BC58A	8034	T	C	0/1	24592	1132	25724	14223	10369	659	473	0.0440056
BC54A	8701	Gi	Ai	0/1	31794	1227	33021	15302	16492	594	633	0.0371582
BC54A	8794	C	T	0/1	31898	611	32509	13756	18142	258	353	0.0187948
BC55A	9845	T	C	0/1	14864	141	15005	7761	7103	77	64	0.00939687
BC52A	12385	C	T	0/1	36021	116	36137	18148	17873	58	58	0.00321001
BC52A	12417	C	CA	0/1	36705	94	36799	208605	16100	55	39	0.00255442
BC57A	12417	C	CA	0/1	31423	93	31516	17590	13833	49	44	0.00295088
BC54A	12717	Ti	Ci	0/1	27977	1216	29193	13590	14387	583	633	0.0416538
BC54A	13263	Gi	Ai	0/1	27666	1115	28781	12920	14746	554	561	0.0387408
BC54A	14470	T	C	0/1	17960	602	18562	8441	9519	258	344	0.0324319
BC54A	15043	Ai	Gi	0/1	27688	1085	28773	12906	14782	508	577	0.037709
BC54A	15301	Ai	Gi	0/1	21947	917	22864	11029	18918	484	433	0.0401067
BC54A	15487	Ti	Ai	0/1	26844	1071	27915	11152	15692	498	573	0.0383665
BC54j	16293	A	G	0/1	5063	51	5114	2690	2373	28	23	0.00997262
BC54j	16519	Ci	Ti	0/1	24826	1008	25834	11088	13738	434	574	0.0390183

```
awk '{print "samtools faidx chrM.hg38.fa chrM:$2-10\""$2+10" > x\"$1\".$2"}'  
z_Finalists.human > z_Finalists.human.extract.list
```

```
samtools faidx chrM.hg38.fa chrM:136-156 > xBC54j.146  
samtools faidx chrM.hg38.fa chrM:136-156 > xBC55j.146  
samtools faidx chrM.hg38.fa chrM:143-163 > xBC54j.153  
samtools faidx chrM.hg38.fa chrM:225-245 > xBC54j.235  
samtools faidx chrM.hg38.fa chrM:225-245 > xBC55j.235  
samtools faidx chrM.hg38.fa chrM:237-257 > xBC54j.247  
samtools faidx chrM.hg38.fa chrM:653-673 > xBC54j.663  
samtools faidx chrM.hg38.fa chrM:1726-1746 > xBC54A.1736  
samtools faidx chrM.hg38.fa chrM:3186-3206 > xBC58A.3196  
samtools faidx chrM.hg38.fa chrM:3659-3679 > xBC50A.3669  
samtools faidx chrM.hg38.fa chrM:3839-3859 > xBC58A.3849  
samtools faidx chrM.hg38.fa chrM:4238-4258 > xBC54A.4248  
samtools faidx chrM.hg38.fa chrM:4238-4258 > xBC55A.4248  
samtools faidx chrM.hg38.fa chrM:4759-4779 > xBC55A.4769  
samtools faidx chrM.hg38.fa chrM:4814-4834 > xBC54A.4824  
samtools faidx chrM.hg38.fa chrM:6494-6514 > xBC56A.6504  
samtools faidx chrM.hg38.fa chrM:6811-6831 > xBC58A.6821  
samtools faidx chrM.hg38.fa chrM:6993-7013 > xBC54A.7003  
samtools faidx chrM.hg38.fa chrM:7128-7148 > xBC57A.7138  
samtools faidx chrM.hg38.fa chrM:7186-7206 > xBC54A.7196  
samtools faidx chrM.hg38.fa chrM:7714-7734 > xBC54A.7724
```

# Extraer los datos con la coordenada de la variante y la secuencia extraída

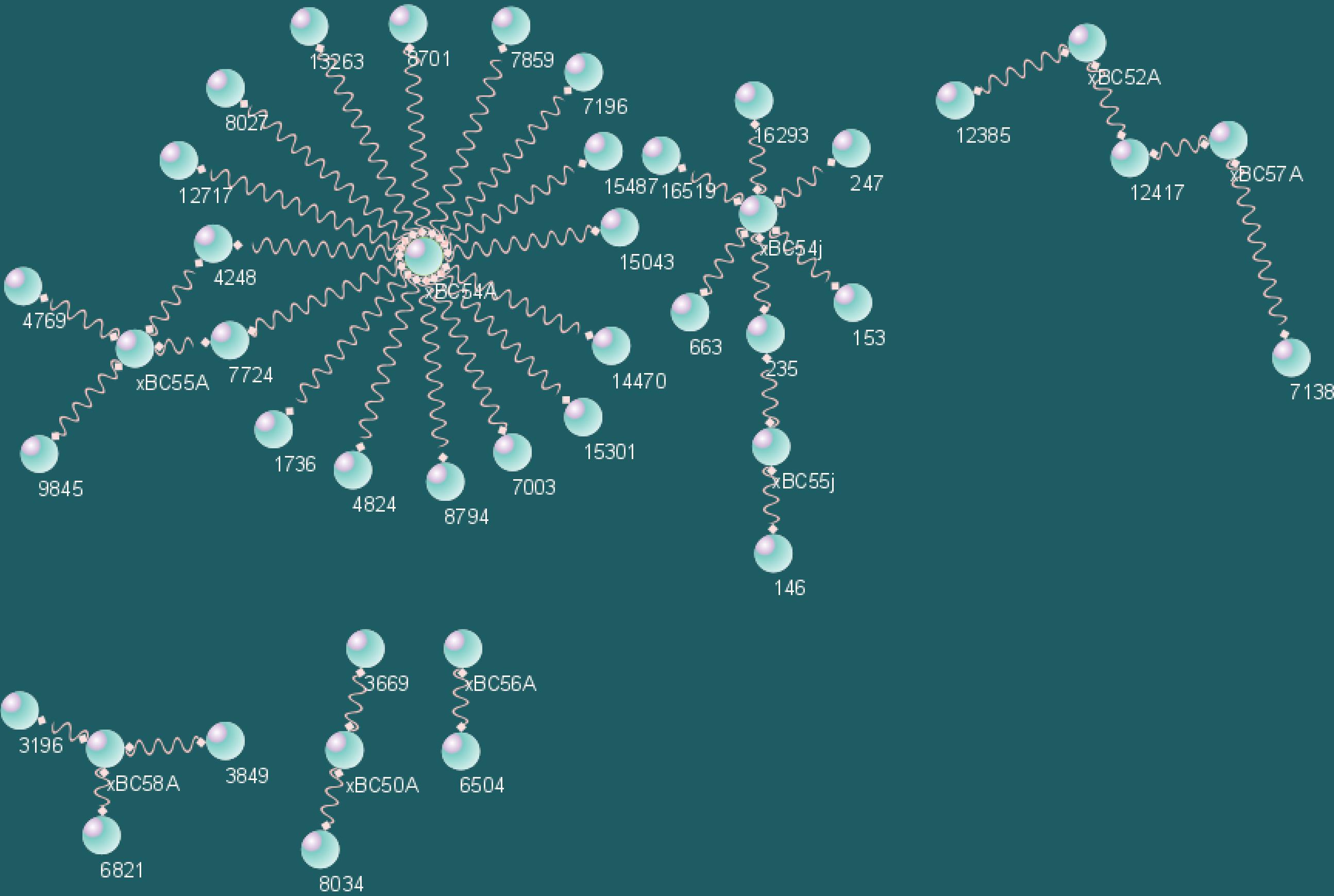
```
grep [ATGC] xBC5* | awk -F"[. :]" '{print $1"\t"$2"\t"$3}' | sort -k2n,2 >  
z_Finalists.human.flanks
```

```
xBC54j 146 GATTCCCTGCCCATCCTTATT  
xBC55j 146 GATTCCCTGCCCATCCTTATT  
xBC54j 153 GCCTCATCCTTATTATTTATCG  
xBC54j 235 GTAGGACATAATAATAACAAT  
xBC55j 235 GTAGGACATAATAATAACAAT  
xBC54j 247 AATAACAATTGAATGTCTGCA  
xBC54j 663 GTTGGTCCTAGCCTTCTAT  
xBC54A 1736 CATTACCCAAATAAAGTATA  
xBC58A 3196 TCTCAACTTAGTATTATACCC  
xBC58A 3669 GATCAGGGTGAGCATCAAAC  
xBC58A 3849 CATGACCCCTGGCCATAATAT  
xBC54A 4248 TCTCCAGCATTCCCCCTCAAA  
xBC55A 4248 TCTCCAGCATTCCCCCTCAAA  
xBC55A 4769 TAATCATAATAGCTATAGCAA  
xBC54A 4824 CCCAGAGGTTACCCAAGGCAC  
xBC56A 6584 ATCTCTCCCAGTCCTAGCTGC  
xBC58A 6821 CATATTCACCTCCGCTACCA  
xBC54A 7003 GACATCGTACTACACGACACG  
xBC57A 7138 TAGCCAAAAATCCATTCACT  
xBC54A 7196 AACACTTCTCGGCTATCCG  
xBC54A 7724 CCTAACACTCACAAACAAA  
xBC55A 7724 CCTAACACTCACAAACAAA  
xBC54A 7859 CGAGGTCAACGATCCCTCCCT  
xBC54A 8027 CCCGATTGAAGCCCCCATTG  
xBC58A 8034 GAAGCCCCCATTGTATAATA  
xBC54A 8701 ACAAAATGATAACCATAACAA  
xBC54A 8794 ACTCCTGCCCACTCATTTAC  
xBC55A 9845 TTGGCTCAACTTCCCTCACTA  
xBC52A 12385 TTCCCTAATTCCCCCATCCT  
xBC52A 12417 TTAACCCCTAACAAAAAAA  
xBC57A 12417 TTAACCCCTAACAAAAAAA  
xBC54A 12717 TCCTAATTACCATACTAATCT  
xBC54A 13263 CTTCAAGTCAACTAGGACTCA  
xBC54A 14470 CTGAGTATATCCAAAGACAA  
xBC54A 15043 TACACATGGGGCGAGGCCTAT  
xBC54A 15301 ACTTCATTTGCCCTTCATTA  
xBC54A 15487 TATTCTCACCAAGACCTCCTAG  
xBC54j 16293 AAACCTACCCACCCCTAACAG  
xBC54j 16519 TACTTCAGGGTCATAAAGCCT
```

```
awk '(print $1"\t"$2"\t"substr($3,1,10)"\t"substr($3,11,1)"\t"substr($3,12,10))'  
z_Finalists.human.flanks > z_Finalists.human.flanks.split
```

```
xBC54j 146 GATTCCCTGCC T CATCCTATT  
xBC55j 146 GATTCCCTGCC T CATCCTATT  
xBC54j 153 GCCTCATCCT A TTATTTATCG  
xBC54j 235 GTAGGACATA A TAATAACAAT  
xBC55j 235 GTAGGACATA A TAATAACAAT  
xBC54j 247 AATAACAATT G AATGTCTGCA  
xBC54j 663 GTTGGTCCT A GCCTTCTAT  
xBC54A 1736 CATTACCC A ATAAGTATA  
xBC58A 3196 TCTCAACTT G TATTATACCC  
xBC58A 3669 GATCAGGGTG A GCATCAAAC  
xBC58A 3849 CATGACCCCTT G GCCATAATAT  
xBC54A 4248 TCTCCAGCAT T CCCCCCTCAA  
xBC55A 4248 TCTCCAGCAT T CCCCCCTCAA  
xBC55A 4769 TAATCATAAT A GCTATAGCAA  
xBC54A 4824 CCCAGAGGTT A CCCAAGGCAC  
xBC56A 6584 ATCTCTCCC A TCCTAGCTGC  
xBC58A 6821 CATATTCAC C TCCGCTACCA  
xBC54A 7003 GACATCGTAC T ACACGACACG  
xBC57A 7138 TAGCCAAAAA T CCATTTCACT  
xBC54A 7196 AACACTTCTT C GGCCTATCCG  
xBC54A 7724 CCTAACACTC A CAACAAA  
xBC55A 7724 CCTAACACTC A CAACAAA  
xBC54A 7859 CGAGGTCAAC G ATCCCTCCCT  
xBC54A 8027 CCCGATTGAA G CCCCCATTG  
xBC58A 8034 GAAGCCCCCA T TCGTATAATA  
xBC54A 8701 ACAAAATGATA A CCATACACAA  
xBC54A 8794 ACTCCCTGCCT C ACTCATTAC  
xBC55A 9845 TTGGCTCAAC T TTCCCTACTA  
xBC52A 12385 TTCCCTAATT C CCCCCATCCT  
xBC52A 12417 TTAACCCCTAA C AAAAAAAA  
xBC57A 12417 TTAACCCCTAA C AAAAAAAA  
xBC54A 12717 TCCTAATTAC C ATACTAATCT  
xBC54A 13263 CTTCAAGTCA A CTAGGACTCA  
xBC54A 14470 CTGAGTATAT C CAAAGACAA  
xBC54A 15043 TACACATGGG G CGAGGCCTAT  
xBC54A 15301 ACTTCATTT G CCCTTCATTA  
xBC54A 15487 TATTCTCACCA G GACCTCCTAG  
xBC54j 16293 AAACCTACCC A CCCTTAACAG  
xBC54j 16519 TACTTCAGGG T CATAAAGCCT
```

# Red



# Variabilidad en el genoma mitocondrial humano en personas sanas

