



GOBIERNO
DE ESPAÑA

MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES



> BU-ISCIII

Análisis Secundario I: Variant Calling

Sara Monzón

BU-ISCIII

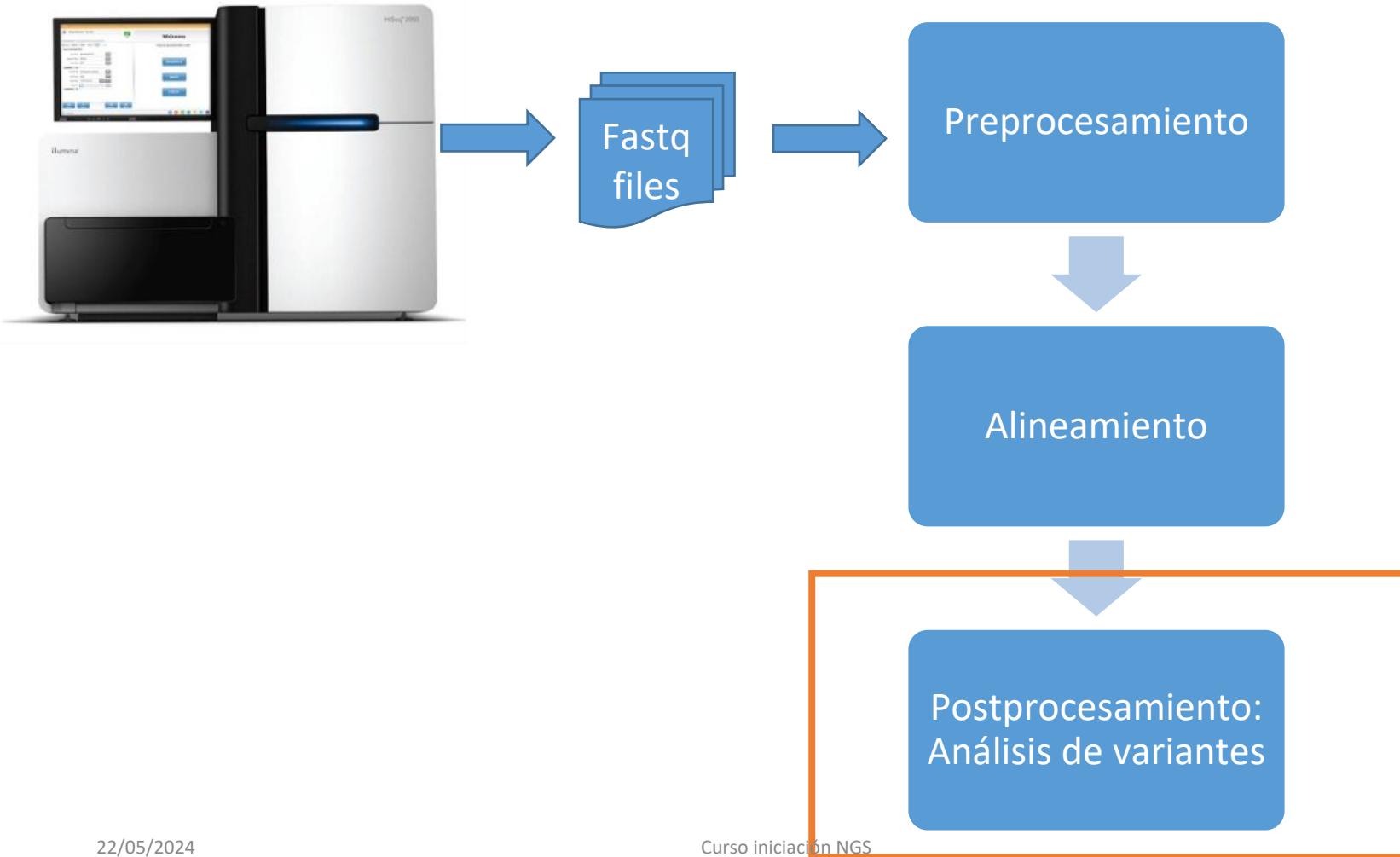
Unidades Científico Técnicas – SGAFI-ISCIII

20-24 Mayo 2024, 11^a Edición
Programa Formación Continua, ISCIII

Índice

- Dónde estamos
- ¿Qué es llamada a variantes?
- Problemas que nos encontramos
- Software de variant calling
- Formatos: vcf y bed
- Anotación y filtrado
- Ejemplos de llamada a variantes:
 - Cáncer
 - Trío
 - Bacterias

Dónde estamos

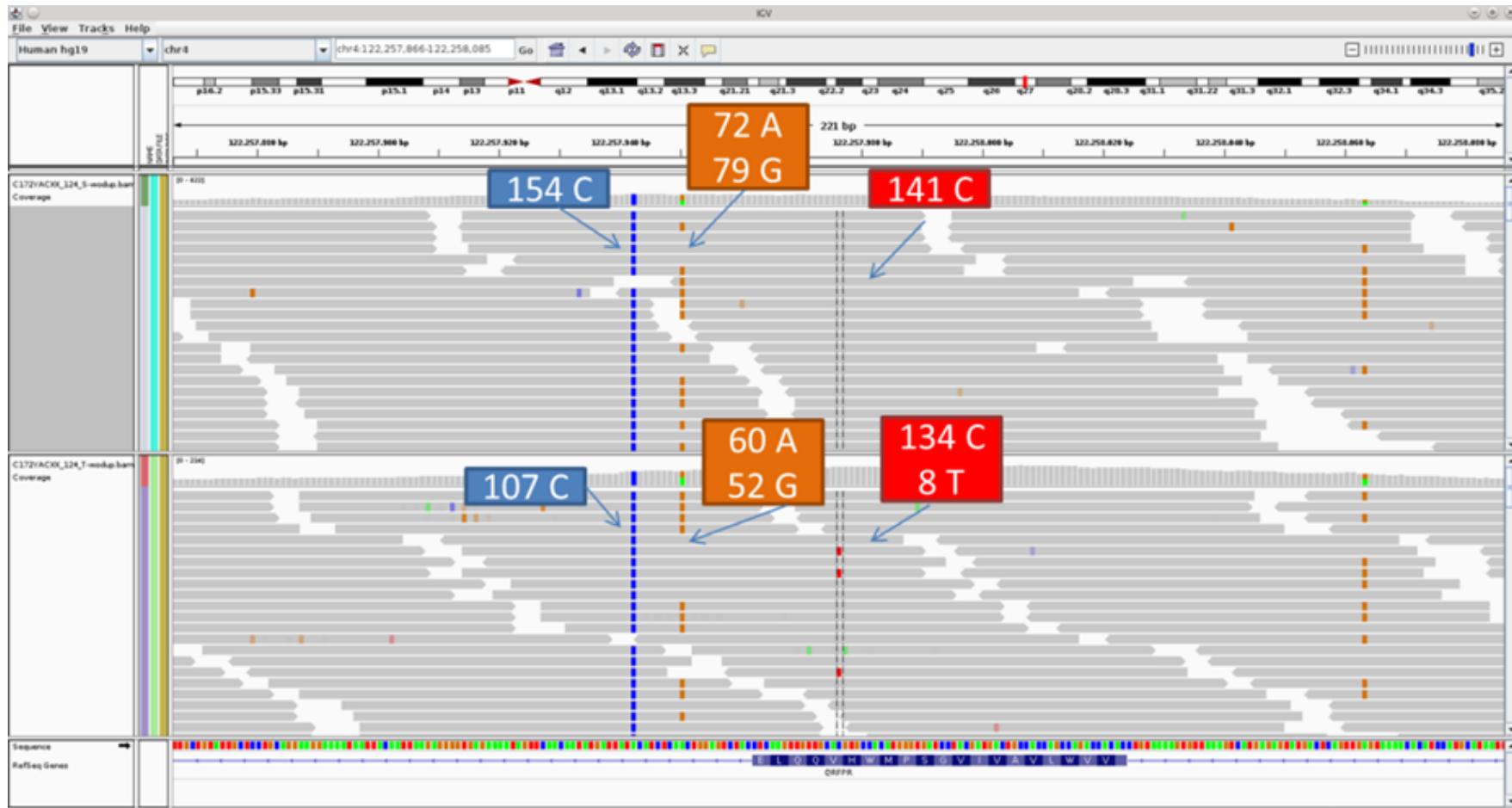


¿Qué es llamada a variantes?

- El concepto de la llamada a variantes es sencillo:
 - Encontrar posiciones en nuestras secuencias que sean diferentes a referencia -
- A partir de nuestras secuencias mapeadas en el genoma, se recorre cada columna del alineamiento y se cuentan cuántos alelos se encuentran y se comparan con la referencia.



¿Qué es la llamada a variantes?



Problemas que nos encontramos

- Preparación de la librería
- Errores en la secuenciación
- Errores de alineamiento
- Fiabilidad de la referencia

Problemas que nos encontramos

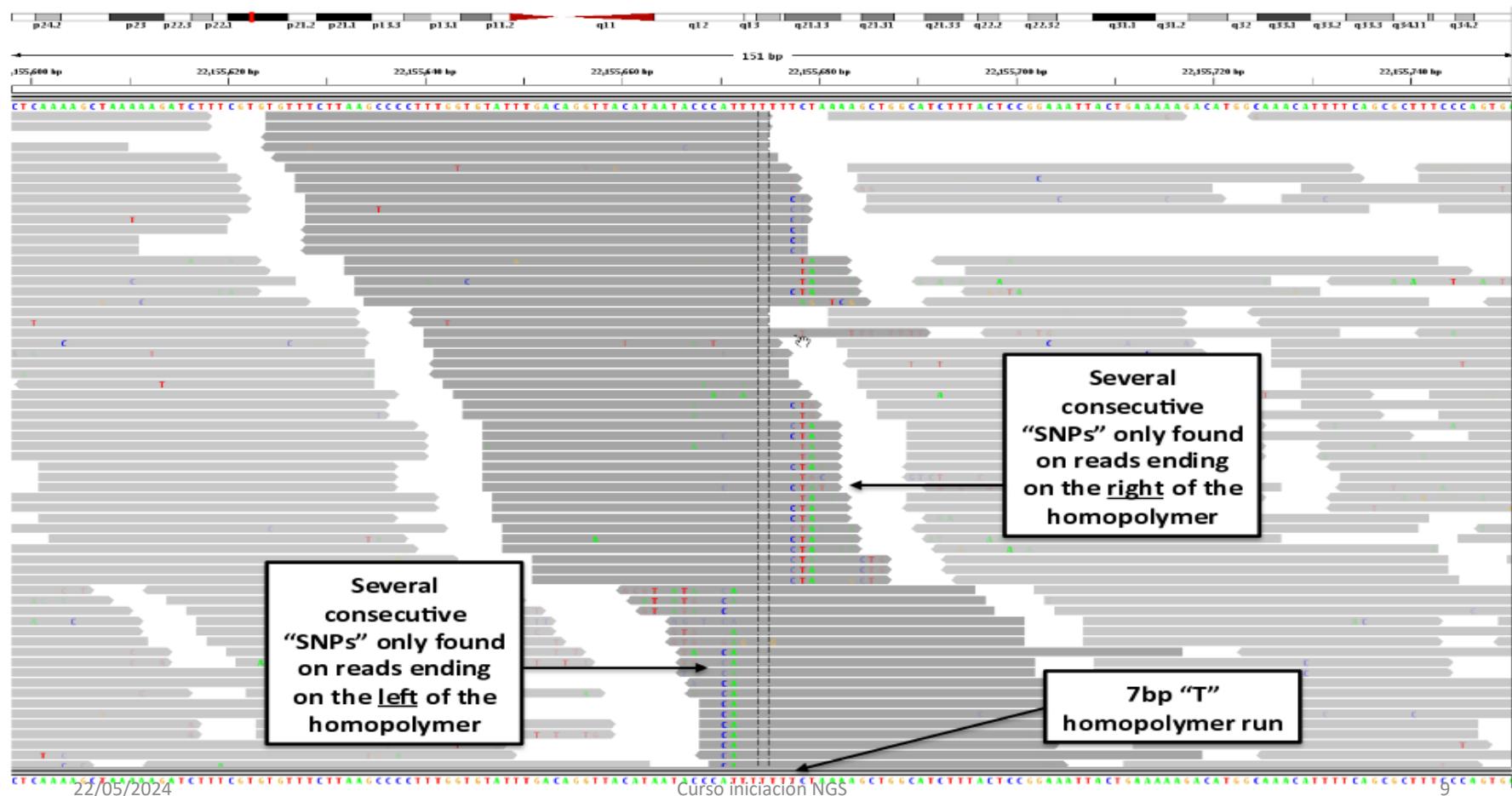
- Artefactos en la preparación de la librería
 - Mutaciones inducidas por PCR
 - Duplicados
 - Errores a final de la lectura
 - Contaminaciones

Problemas que nos encontramos

- Ratio de error asociado con la secuenciación.
- Soluciones:
 - Evaluación de Phred
 - Strand bias

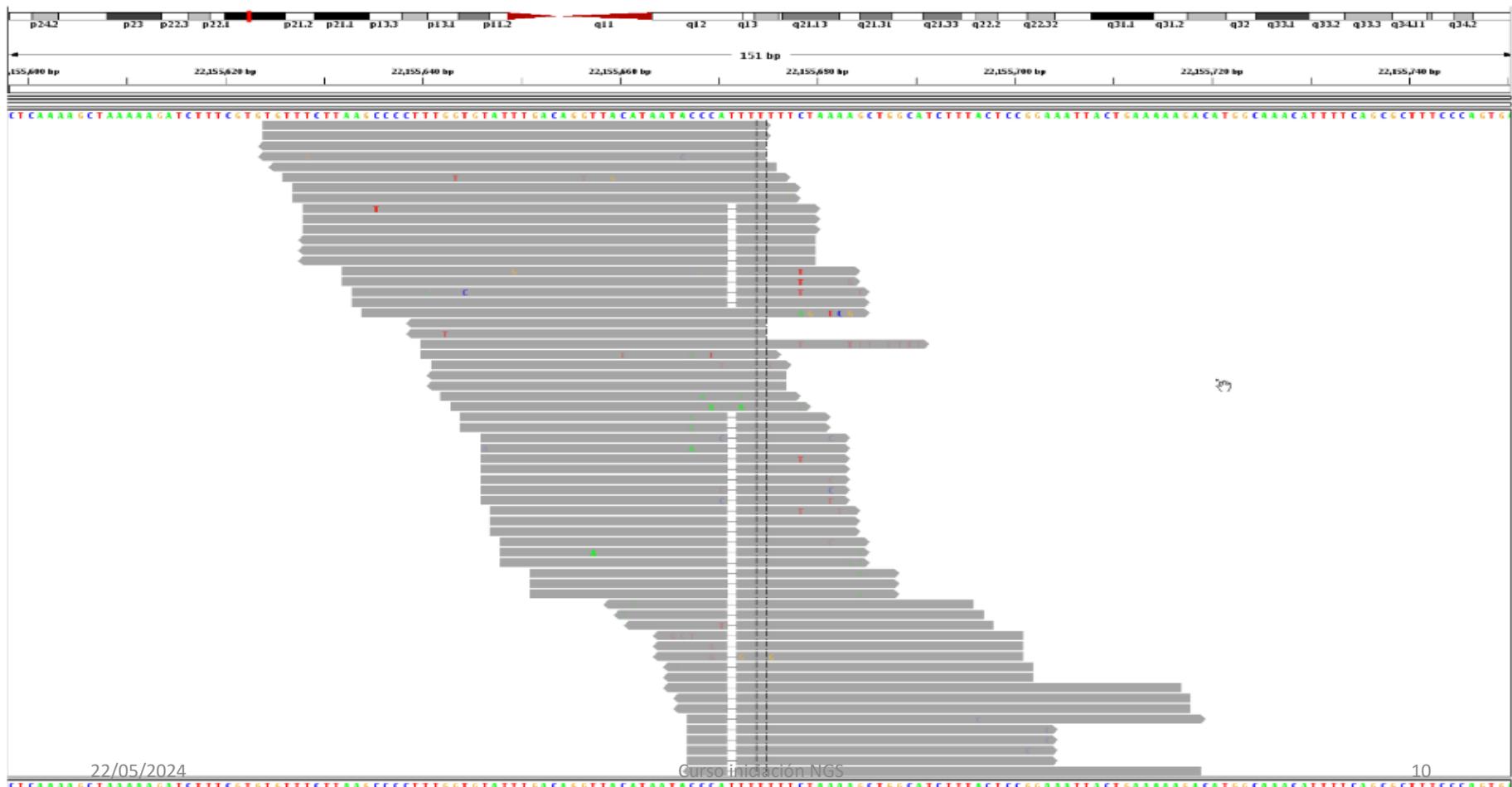
Problemas que nos encontramos

- Problemas de alineamiento



Problemas que nos encontramos

- Problemas de alineamiento



Problemas que nos encontramos

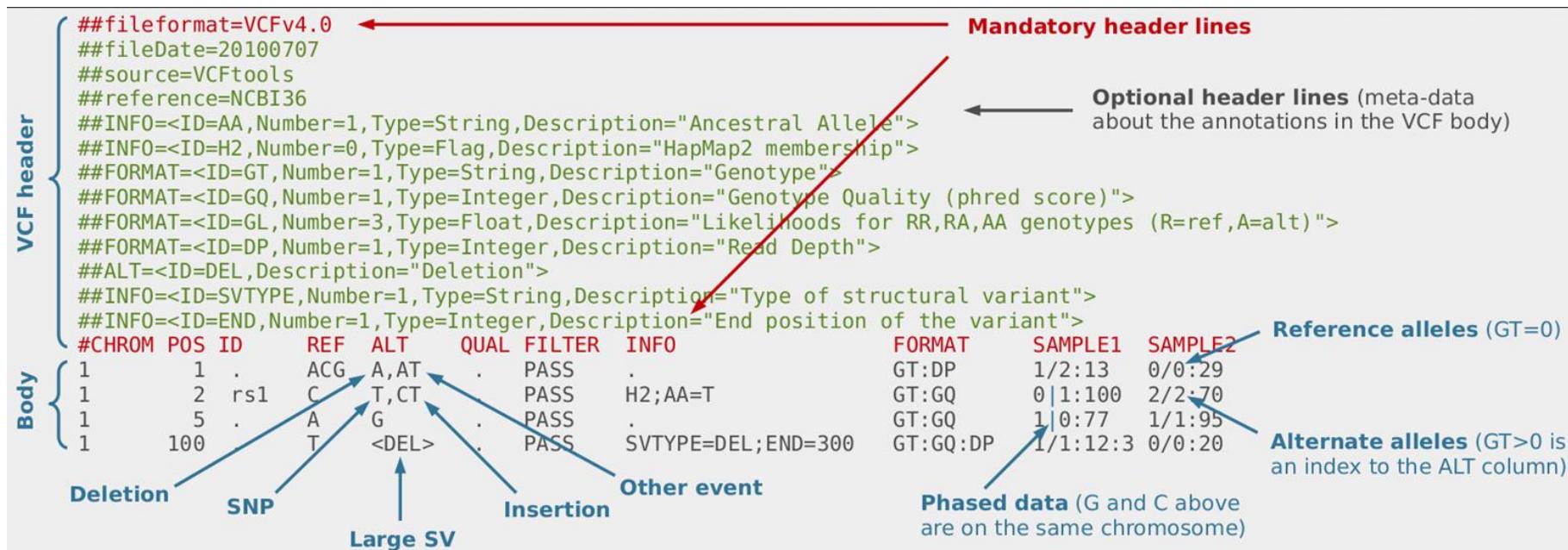
- **Fiabilidad del genoma de referencia.**
 - Ejemplo genoma humano:
 - Genoma obtenido de mezcla de 8 personas diferentes (Watson entre ellos)
 - Genoma haploide para individuo diploide.
 - Zonas de baja complejidad
 - Incompleto

Principales software de variant calling

- GATK
 - Samtools
 - Strelka
 - MuTect
 - JointSNVMix
 - VarScan
 - SolSNP
 - Largo etc.
- 
- Llamada germinal en organismos diploides/haploides
- Llamada somática en pares de control - tumor
- Otros

Formatos: vcf y bed

- Formato vcf



Formatos: vcf y bed

- Formato bed
 - Se utiliza para representar regiones y/o posiciones

chromosom	start	en	score	strand	thickstart	thickend	RGB	
chr7	127471196	127472363	Pos1	0	+	127471196	127472363	255,0,0
chr7	127472363	127473530	Pos2	0	+	127472363	127473530	255,0,0
chr7	127473530	127474697	Pos3	0	+	127473530	127474697	255,0,0
chr7	127474697	127475864	Pos4	0	+	127474697	127475864	255,0,0
chr7	127475864	127477031	Neg1	0	-	127475864	127477031	0,0,255
chr7	127477031	127478198	Neg2	0	-	127477031	127478198	0,0,255
chr7	127478198	127479365	Neg3	0	-	127478198	127479365	0,0,255
chr7	127479365	127480532	Pos5	0	+	127479365	127480532	255,0,0
chr7	127480532	127481699	Neg4	0	-	127480532	127481699	0,0,255

Formatos: vcf y bed

- Formato bed
 - Se utiliza para representar regiones y/o posiciones
 - Consideraciones para representar variantes.
 - Utiliza coordenadas 0-based para el inicio y 1-based para el final
 - De manera que la primera base del cromosoma 1 sería:

```
chr1    0    1    first_base
```

Formatos: vcf y bed

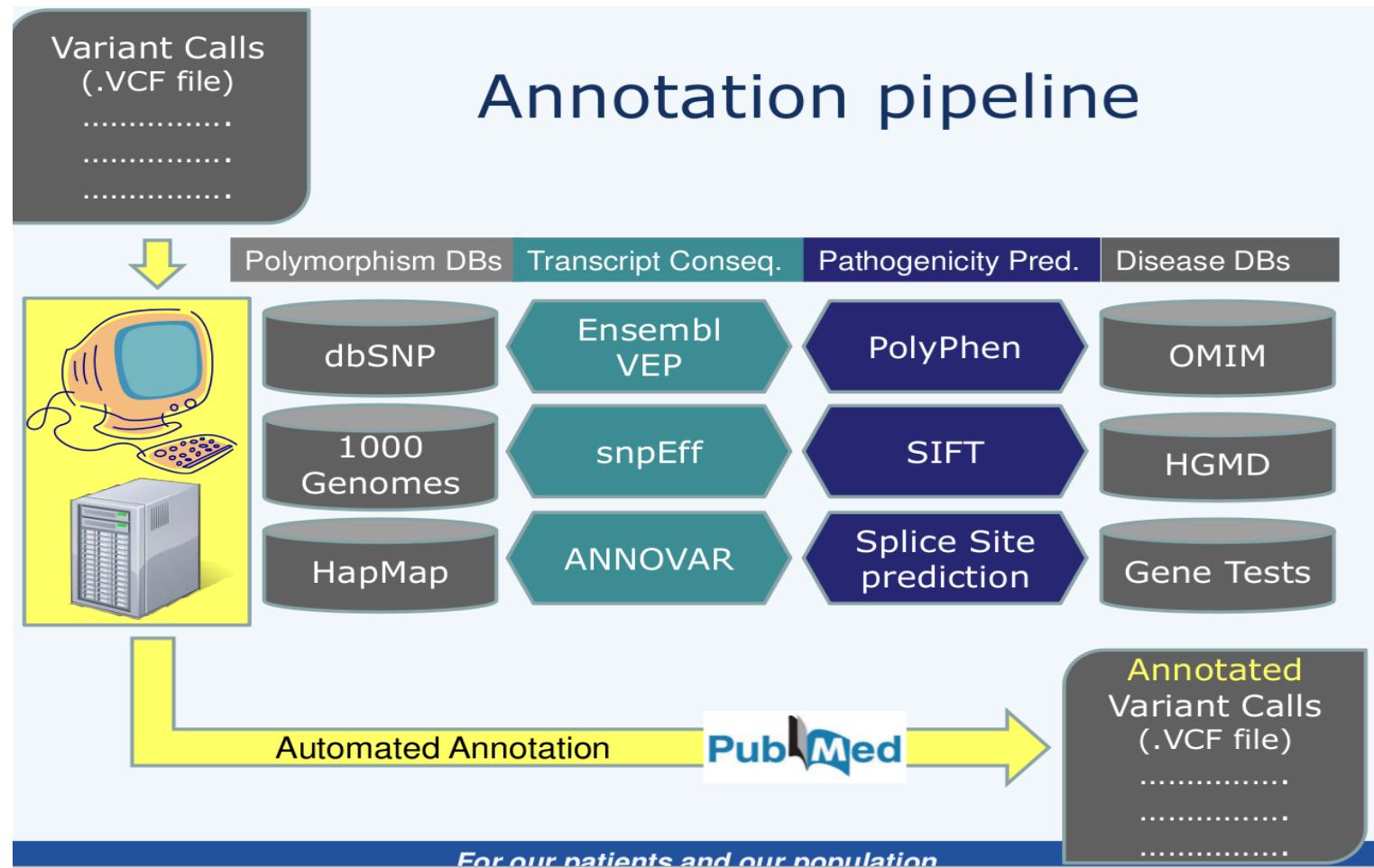
- Ejemplo de formato bed de variantes:

chr1	100154496	100154497	A	G
chr1	100182982	100182983	C	T
chr1	100195206	100195207	C	A
chr1	1002596	1002597	C	A
chr1	100343384	100343385	G	T
chr1	10041131	10041132	C	A
chr1	100575981	100575982	G	T
chr1	100621863	100621864	G	T
chr1	100672062	100672063	C	T
chr1	10067673	10067674	G	T
chr1	100733834	100733835	G	T
chr1	101007160	101007161	G	T
chr1	101186145	101186146	G	T
chr1	101376658	101376659	C	A
chr1	101379322	101379323	C	A
chr1	101490740	101490741	G	T
chr1	10161234	10161235	G	A
chr1	101705323	101705324	C	A
chr1	101705774	101705775	C	A
chr1	10179467	10179468	C	A
chr1	10197177	10197178	C	G
chr1	10197185	10197186	G	T

Esta es la posición donde se encuentra la variante

Admite prácticamente de todo
En este caso Alelo alternativo y Alelo referencia

Anotación y Filtrado



For our patients and our population

Anotación y filtrado

- Anotación:
 - A nivel de gen: se anota gen y “feature” según la base de datos refgene (variante tipo missense, frameshit, intron, etc.)
 - Anotación de variantes no sinónimas: dbNSFP
 - SLR
 - SIFT
 - Polyphen2_HDIV
 - Polyphen2_HVAR
 - LRT
 - Mutation Taster
 - Mutation Assesor
 - FATHMM_score
 - CADD_score
 - GERP++_NR
 - GERP++_RS
 - PhyloP100way_vertebrate
 - 29way_logOdds
- A nivel funcional: pseudogenes, UniprotFeature, etc.
- A nivel de enfermedad: anotación de enfermedad asociada con ese gen en OMIM

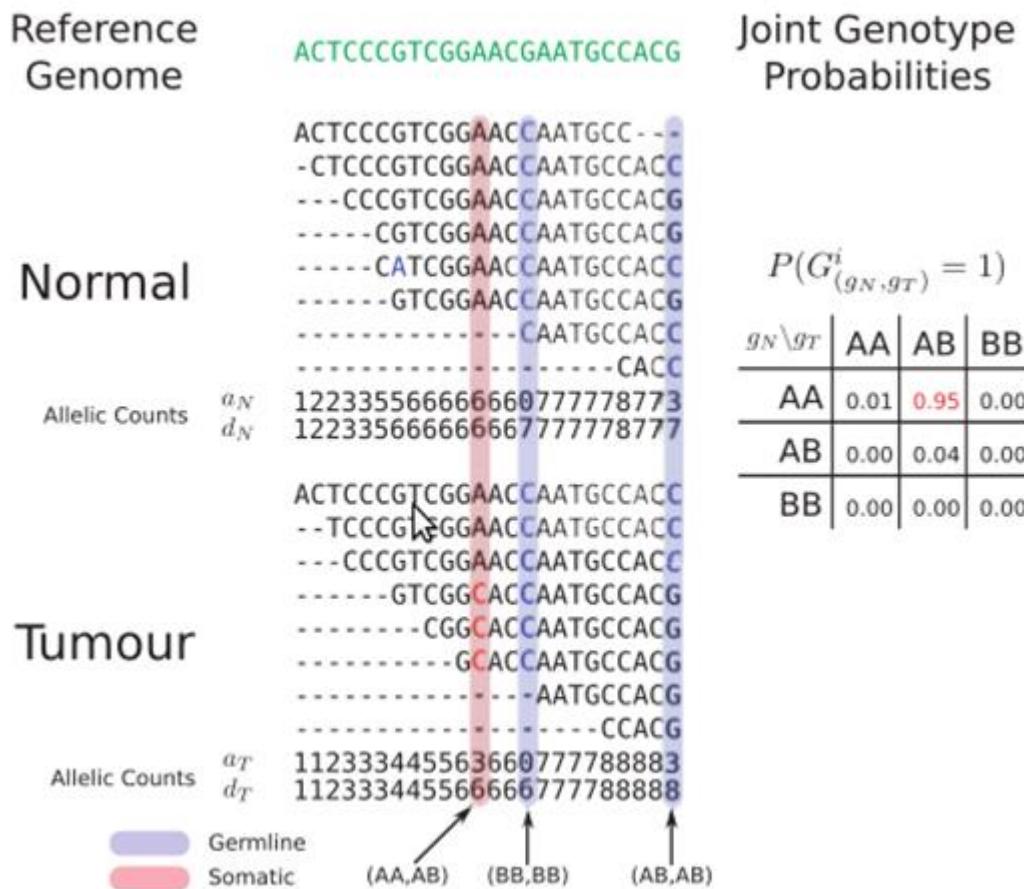
Ejemplo de variant calling: Cáncer

- Software específico para comparaciones tumor-control

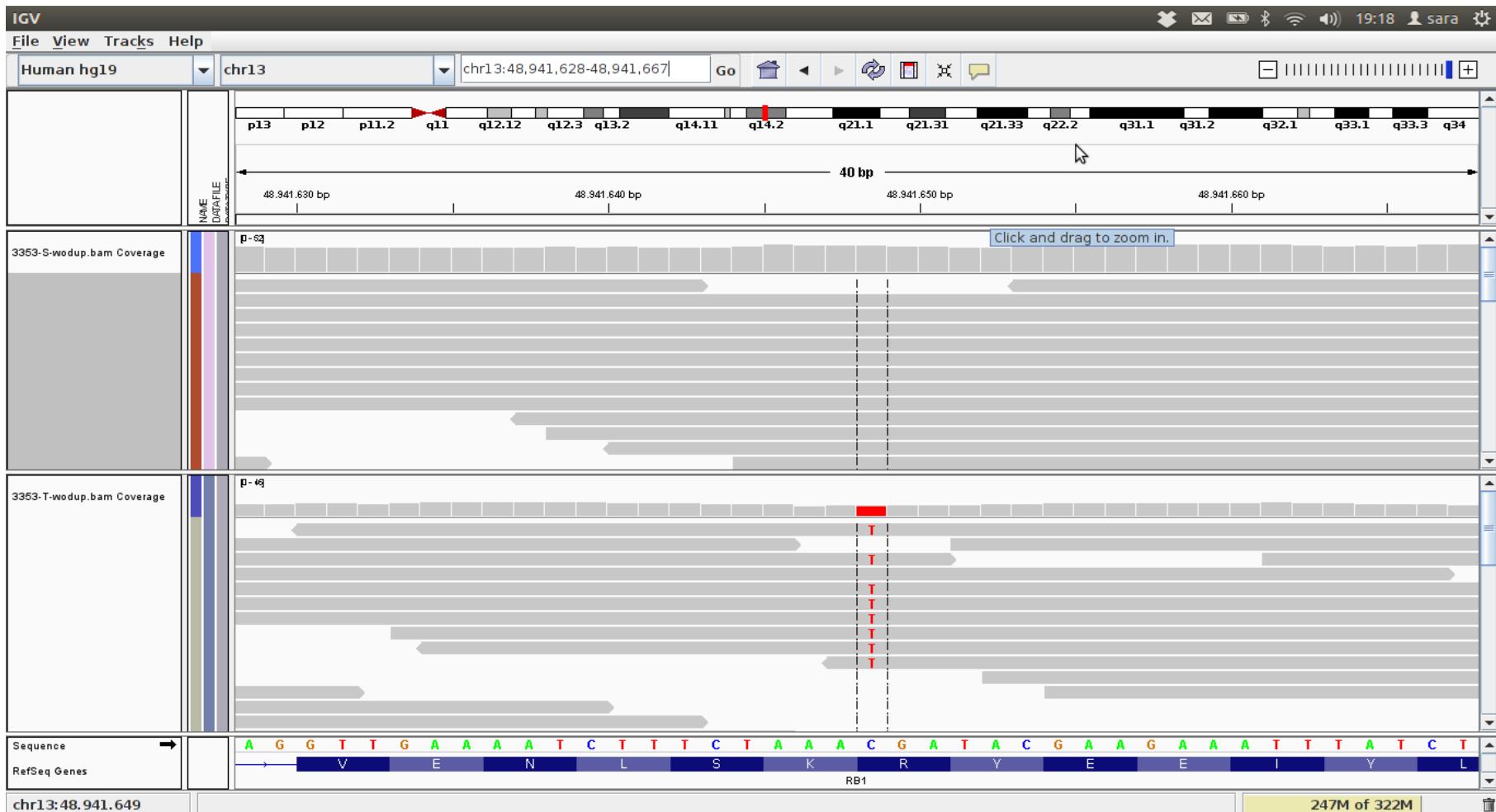
	Samtools	GATK	VarScan2	Somatic Sniper	JointSNVMix	Strelka	LoFree	MuTeC	Shimmer	EBCalling	Virmed
Publication	Li et al	McKenna et al	Koboldt et al	Larson et al	Roth et al	Sauders et al	Wilm et al	Cibulski et al	Hansen et al	Shiraishi et al	Kim et al
Year	2009	2010	2012	2012	2012	2012	2012	2013	2013	2013	2013
Model	Bayesian	Bayesian	Fisher test		Prob	Bayesian	Binomial	Bayesian		Bayesian	Prob
Programming language	C	java	Java,perl	C	python	perl	C,python	java	perl	Perl,c,R	Bayesian
Paired sample	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Realignment	No	Yes	No	No	No	Yes	No	Yes	No	No	No

Ejemplo de variant calling: Cáncer

- Teoría de la llamada a variantes en cáncer

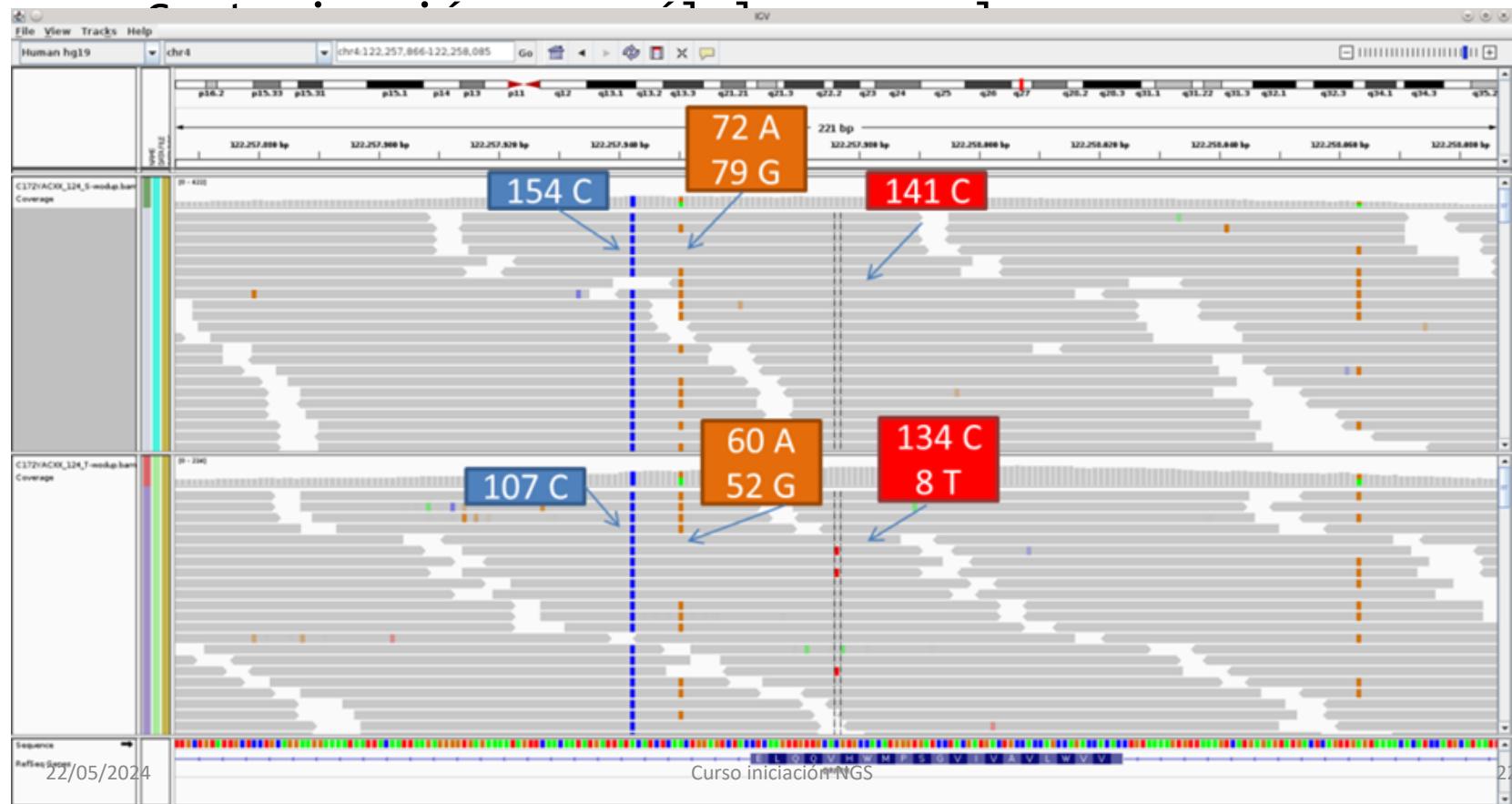


Ejemplo de variant calling: Cáncer



Ejemplo de variant calling: Cáncer

- Problemas adicionados a la llamada a variantes
 - Heterogeneidad tumoral



Ejemplo de variant calling: Cáncer

- Comparativa de distintos software de llamada a variantes en cáncer
 - Evaluar la mejor opción
 - Plantearse seleccionar la intersección de varios.
 - Caracterizar su comportamiento frente a cobertura y frecuencia del alelo alternativo.

Ejemplo de variant calling: Cáncer

- Problemas de no detectar verdaderas mutaciones somáticas





GOBIERNO
DE ESPAÑA

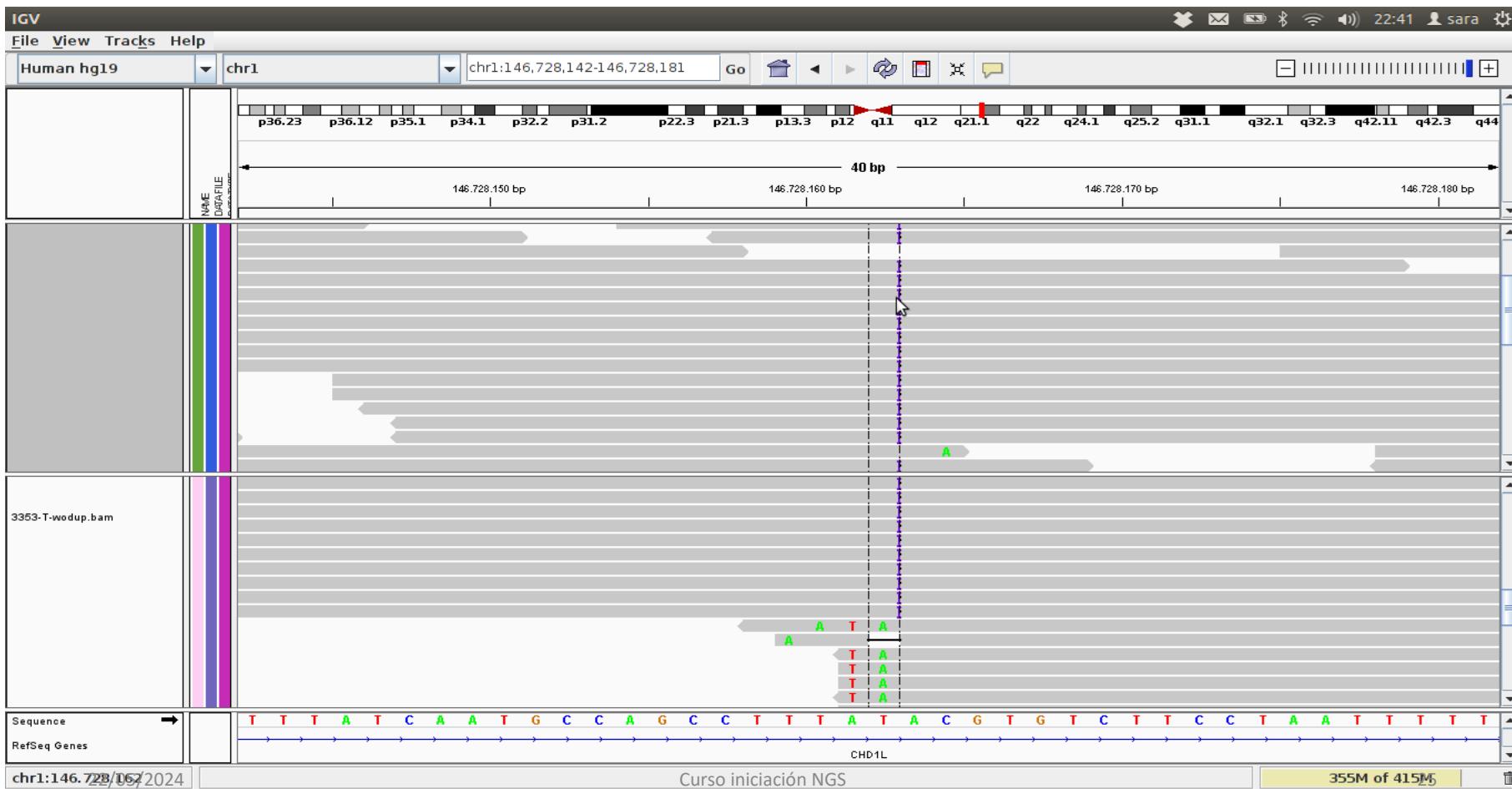
MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES



BU-ISCIII

Ejemplo de variant calling: Cáncer

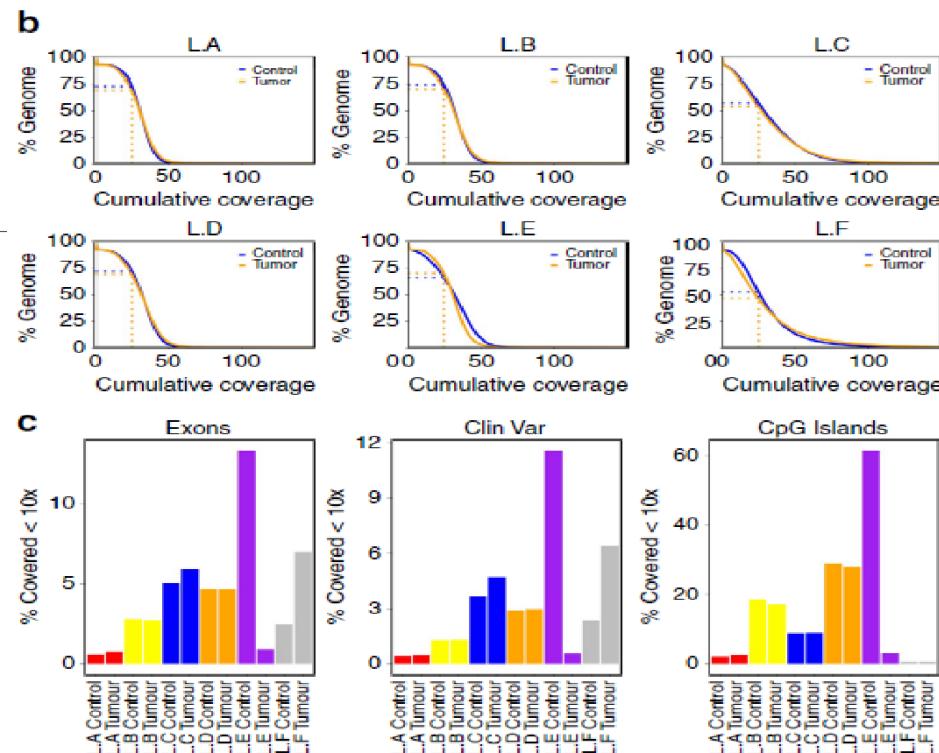
- Problemas de realineamiento



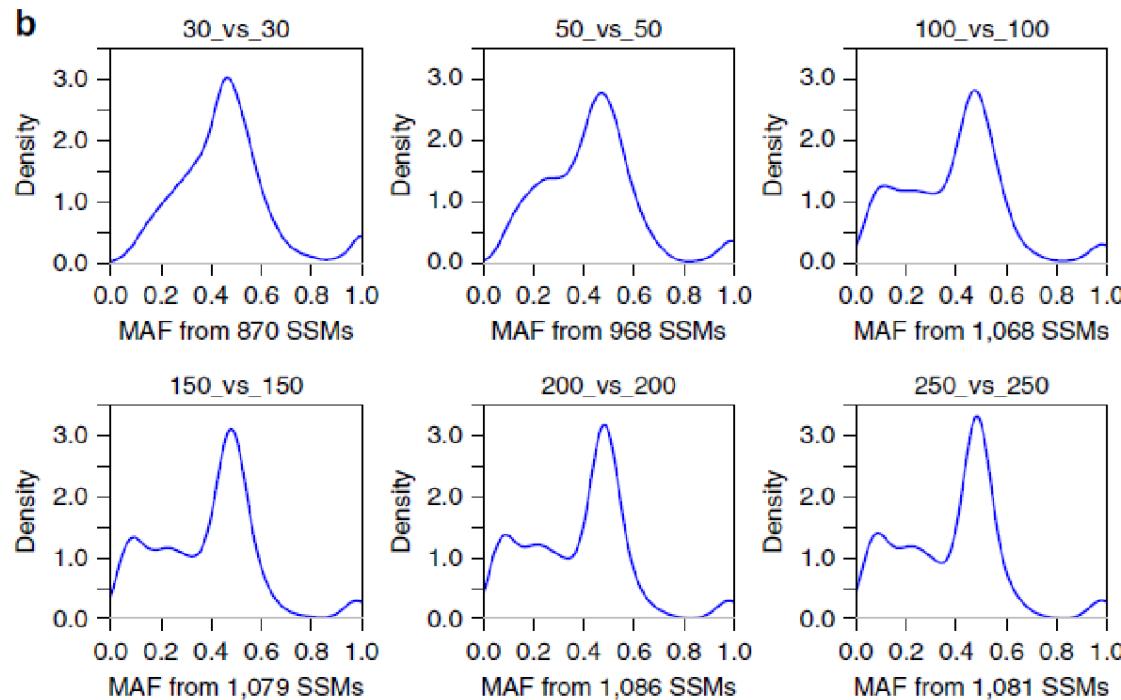
Ejemplo de variant calling: Cáncer

Table 1 | Summary of medulloblastoma tumour-normal pair library construction and sequencing.

Library	Starting DNA (μg)	Fragment Size (bp)	Size selection	Library protocol	PCR cycles	Sequencing machine	Chemistry (Illumina)	Depth (×) control:tumour
L.A	4	~400	2% Agarose gel	KapaBio	0	HiSeq 2500 HiSeq 2000 MiSeq	V1 (RR) V3 V2	29.6 : 40.5
L.B	1	~400	2% Agarose gel,	TrueSeq DNA	10			
L.C	2.5	~500	Invitrogen E-gel, 2% Agarose gel	NEBNEXT	12			
L.D	1	~550	Agarose gel	TrueSeq DNA	10			
L.E	2.8	~620	1.5% Agarose gel	NEBNEXT	0			
L.F	1	~400	pippin	NEBDNA	10			
L.G	1	~350	AMPureXP beads	TrueSeq DNA	0			
L.H	0.5	~175	AMPureXP beads	PCR-Free				
				SureSelect WGS	10			



Ejemplo de variant calling: Cáncer



Contrariamente a lo que se piensa identificar variantes somáticas de datos WGS es todavía un gran reto.

Ejemplo de variant calling: Cáncer

Table 3 | Summary of accuracy measures.

SSM calls	Aligner	SSM Detection Software	TP	FP	FN	P	R	F1
MB.GOLD	BWA, GEM	Curated	1,255 (8)	0	0	1.00	1.00	1.00
MB.A	BWA	In-house	775 (0)	147	480	0.84	0.62	0.71
MB.B	BWA	samtools, Varscan	788 (1)	12	467	0.99	0.63	0.77
MB.C	GEM	samtools, bcftools	766 (3)	1,025	489	0.43	0.61	0.50
MB.D	n.a.	SMuFin	737 (4)	1,086	518	0.41	0.59	0.48
MB.E	BWA	SomaticSniper	750 (4)	229	505	0.77	0.60	0.67
MB.F	BWA	Strelka	884 (2)	165	371	0.84	0.70	0.77
MB.G	BWA	Caveman, Picnic	899 (3)	140	356	0.87	0.72	0.78
MB.H	Novoalign	MuTect	947 (3)	6,296	308	0.13	0.76	0.22
MB.I	BWA	samtools	879 (7)	129	376	0.87	0.70	0.78
MB.J	None, BWA	SGA + freebayes	856 (1)	62	399	0.93	0.68	0.79
MB.K	BWA	Atlas2-snp	945 (8)	7,923	310	0.11	0.75	0.19
MB.L1	BWA	MuTect, Strelka	385 (0)	3	870	0.99	0.31	0.47
MB.L2	BWA	MuTect, Strelka	900 (1)	253	355	0.78	0.72	0.75
MB.M	BWA mem	samtools, GATK + MuTect	937 (4)	1,695	318	0.36	0.75	0.48
MB.N	BWA	Strelka	847 (1)	289	408	0.75	0.68	0.71
MB.O	BWA	MuTect	944 (3)	272	311	0.78	0.75	0.76
MB.P	BWA	Sidron	833 (3)	256	422	0.77	0.66	0.71
MB.Q	BWA	qSNP + GATK	842 (2)	25	413	0.97	0.67	0.79
<hr/>								
SIM calls								
MB.GOLD	BWA, GEM	Curated	337 (10)	0	0	1.00	1.00	1.00
MB.A	BWA	In-house	16 (0)	63	321	0.20	0.05	0.08
MB.B	BWA	GATK SomaticIndelDetector, Varscan	167 (0)	20	173	0.89	0.49	0.63
MB.C	GEM	samtools, bcftools	103 (0)	26	236	0.80	0.30	0.44
MB.D	none	SMuFin	29 (0)	25	308	0.54	0.09	0.15
MB.F	BWA	Strelka	147 (8)	12	193	0.93	0.43	0.58
MB.G	BWA	Pindel	189 (2)	82	152	0.70	0.55	0.61
MB.H	Novoalign	VarScan2	55 (0)	248	282	0.18	0.16	0.17
MB.I	BWA	Platypus	271 (7)	224	70	0.55	0.79	0.65
MB.J	None	SGA	90 (1)	34	249	0.72	0.26	0.38
MB.K	BWA	Atlas2-indel	268 (6)	444	72	0.38	0.79	0.51
MB.L1	BWA	Strelka	64 (1)	3	273	0.96	0.19	0.32
MB.L2	BWA	Strelka	130 (3)	13	210	0.91	0.38	0.53
MB.N	BWA	Strelka	128 (6)	16	209	0.89	0.38	0.53
MB.O	BWA	GATK SomaticIndelDetector	140 (1)	47	197	0.75	0.42	0.53
MB.P	BWA	bcftools, PolyFilter	37 (0)	57	301	0.39	0.11	0.17
MB.Q	BWA	Pindel	100 (2)	61	237	0.63	0.30	0.40

F1, F1 score; FN, false negative; FP, false positives; P, precision; R, recall; TP, true positives.

Shown are the evaluation results with respect to the medulloblastoma Gold Set (Tier 3). Shown are the number of true calls (TP) with additional Tier 4 calls in parentheses, the number of FP, the number of FN, P, R and F1. The submissions with the best precision, recall and F1 score are in bold.

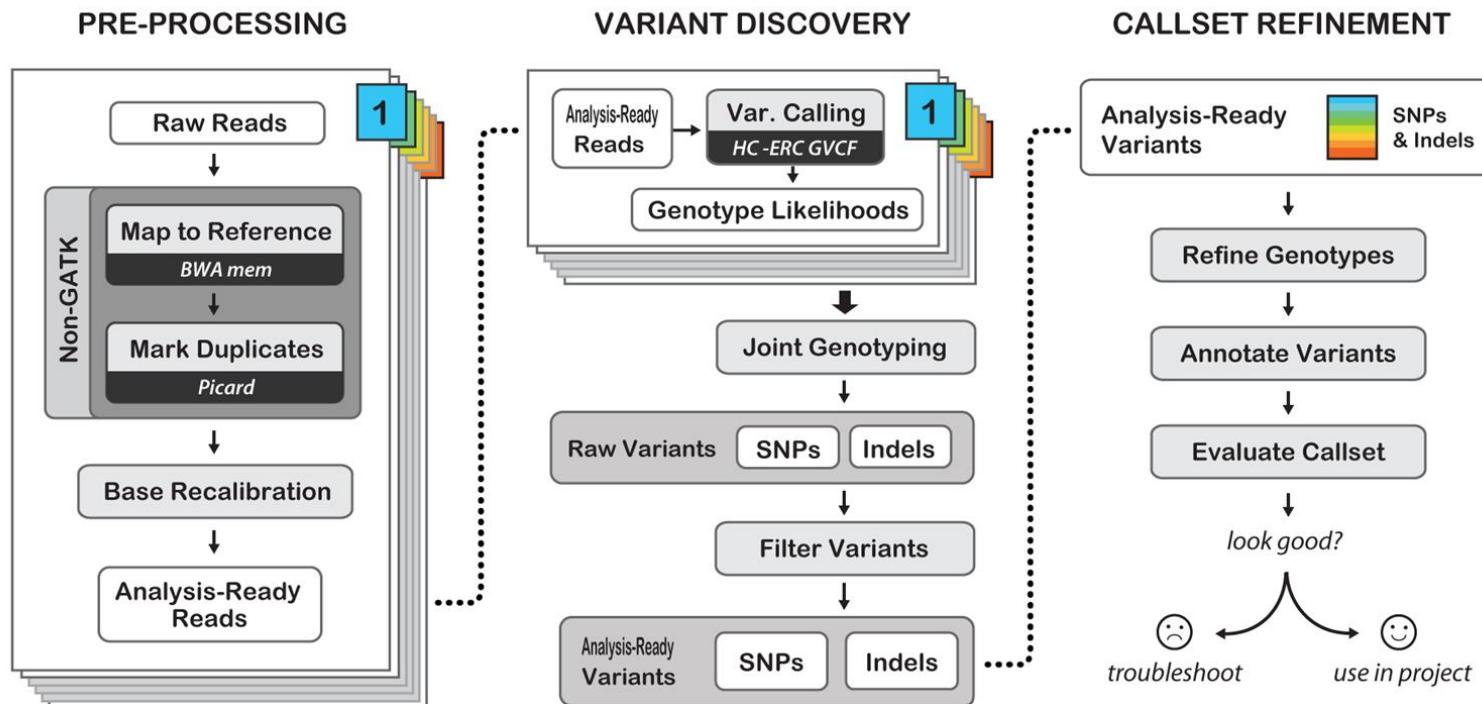
- Llamada a variantes con diferentes pipelines y datos de diferentes librerías da lugar a un bajo consenso.
- Checklist para estudios WGS en cáncer:
 - Preparar librería PCR-free
 - Tumor coverage 100x
 - Control coverage close to tumor coverage (+/-10%)
 - Reference genome hs37d5 o GRCh38
 - Combinación de alineador/variant caller óptimo
 - Combinar varios llamadores de mutaciones
 - Permitir mutaciones en zonas repetidas o cerca de repeticiones.
 - Filtrado por calidad de mapado, strand bias, positional bias, presencia de soft-clipping

Ejemplo de variant calling: TRIOS

- Formato: fastq
- Plataforma: HiSeq Illumina, 2x101
- Carreras: 1
- Lanes: 2 y 8
- Kit Enriquecimiento: TruSeq Exome Enrichment Kit (2011)

Pedigrí	Sexo	Afectado
Padre	V	N
Madre	M	N
Hijo	V	S

Variant Calling



Postprocesamiento y QC alineamiento

- Filtrado de duplicados: Picard
- Análisis de la calidad del BAM
- Recalibración de variantes
- Realineamiento

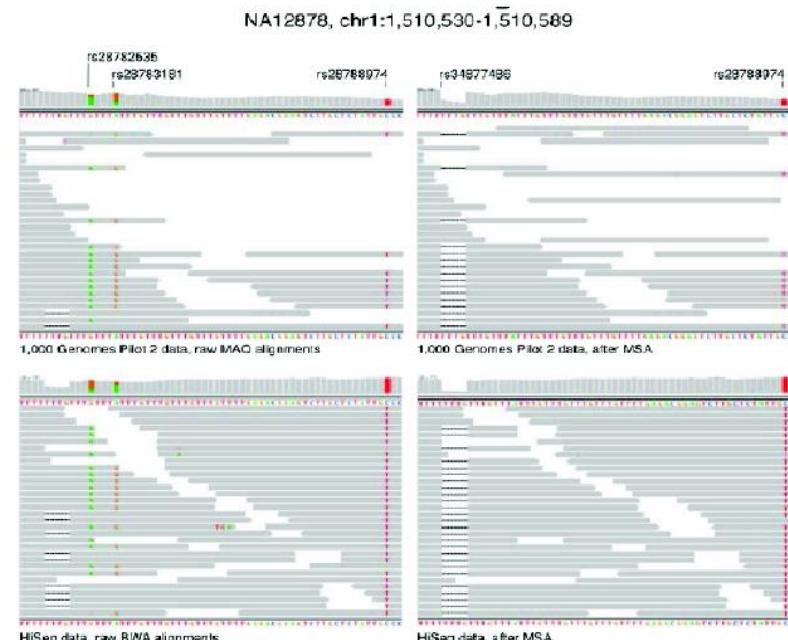
Sample	Target Specificity	Target Enrichment	Mean Coverage	SD Coverage	5X	10X	20X	30X
Padre	0.77	39.40	90.92	79.87	95%	93%	90%	85%
Madre	0.78	39.85	72.68	64.91	93%	92%	87%	81%
Hijo	0.78	39.78	60.46	53.12	93%	90%	84%	75%

Variant Calling: Realineamiento

Realineamiento local de múltiples secuencias

Proporciona un alineamiento consistente entre todas las lecturas. Se identifican las regiones susceptibles de realineamiento, si:

- Al menos una lectura contiene un indel
- Existe un *cluster* de bases *mismatch*
- Existe un indel conocido

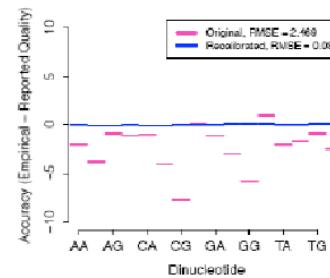
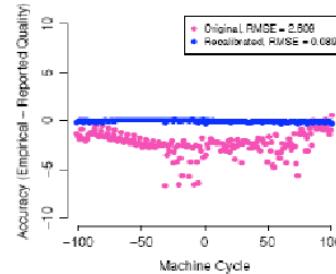
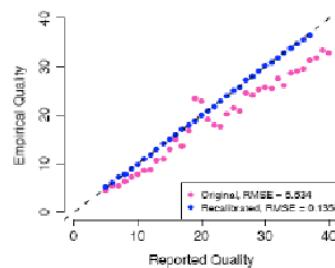


Variant Calling: Realineamiento

Score de calidad de una base

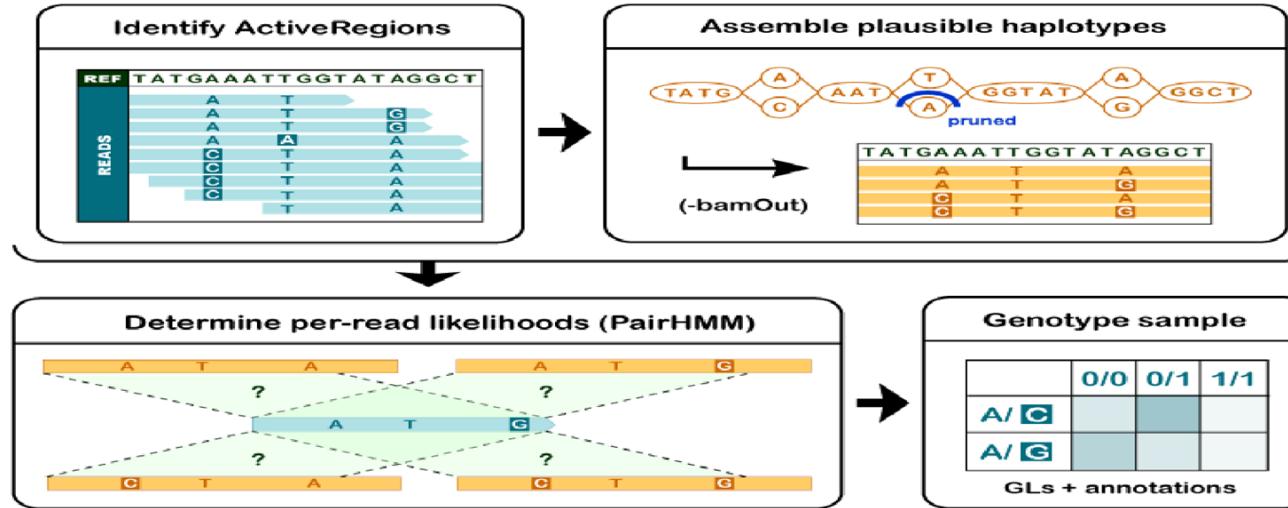
Probabilidad de que la base determinada sea verdadera (y no un error de secuenciación).

- En escala *Phred*. $Q = -10 \cdot \log_{10} P$
- Se codifica en ASCII (normalmente Q+33)
- Se estima de un modo muy inexacto porque sigue un esquema de correlación complejo entre la tecnología de secuenciación, el ciclo de máquina y el contexto de secuencia.



Los errores de mapeado y la inexactitud de los scores de calidad se propagan a la etapa de identificación de variantes y genotipado

Variant Calling: HaplotypeCaller



- Determina si una región es potencialmente variable
- Construye un ensamblado de Brujin de la región.
- Los “paths” en el grafo son haplotipos potenciales que tienen que ser evaluados.
- Se calcula los likelihoods de los haplotipos dados los datos usando un modelo PairHMM.
- Determina si hay alguna variante entre los haplotipos más probables.
- Calcula la distribución de la frecuencia alélica para determinar el conteo de alelos más probables y emite una variante si se da el caso.
- Si se emite una variante se calcula el genotipo para cada muestra.

Estadísticas del filtrado

Variants	Raw	HardFiltering*	GenotypeRefinement	Quality Filtering
SNPs	294018	189031	188910	177660
INDELs	40677	27695	26646	

*Siendo este número aquellas variantes marcadas como PASS después del filtrado.

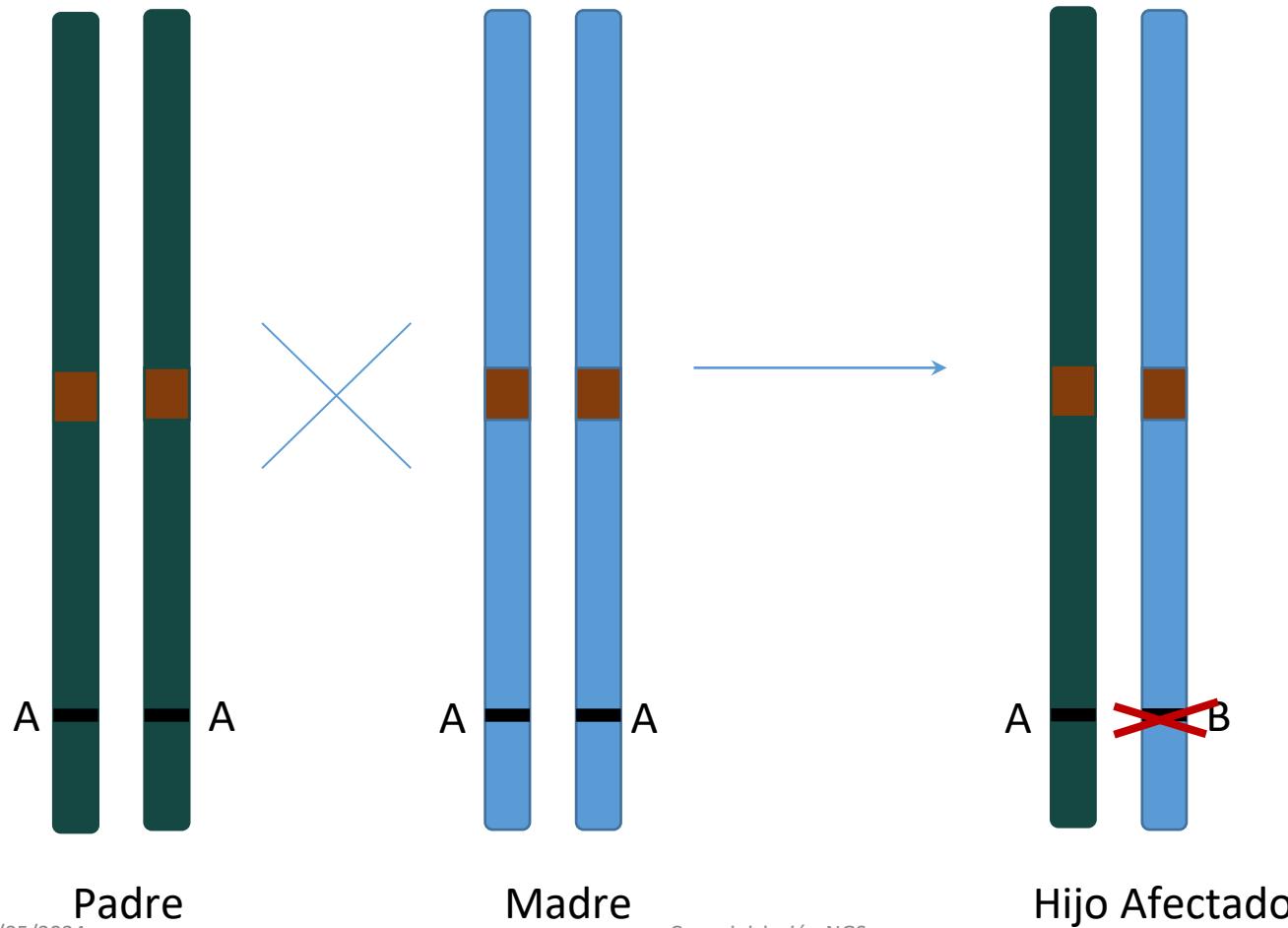
Modelos de Enfermedad

- Modelo de novo
 - Modelo double-hit gene
 - Modelo Recesivo
-
- Modelo dominante

Seleccionamos estos dos como los más probables en nuestro caso.

Modelos de enfermedad

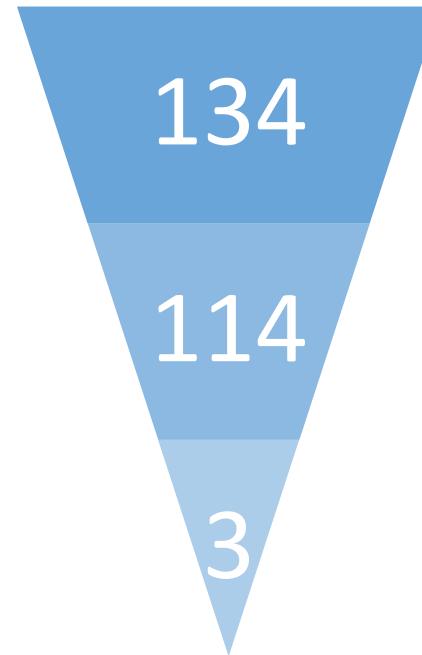
- Modelo de novo



Modelo de novo

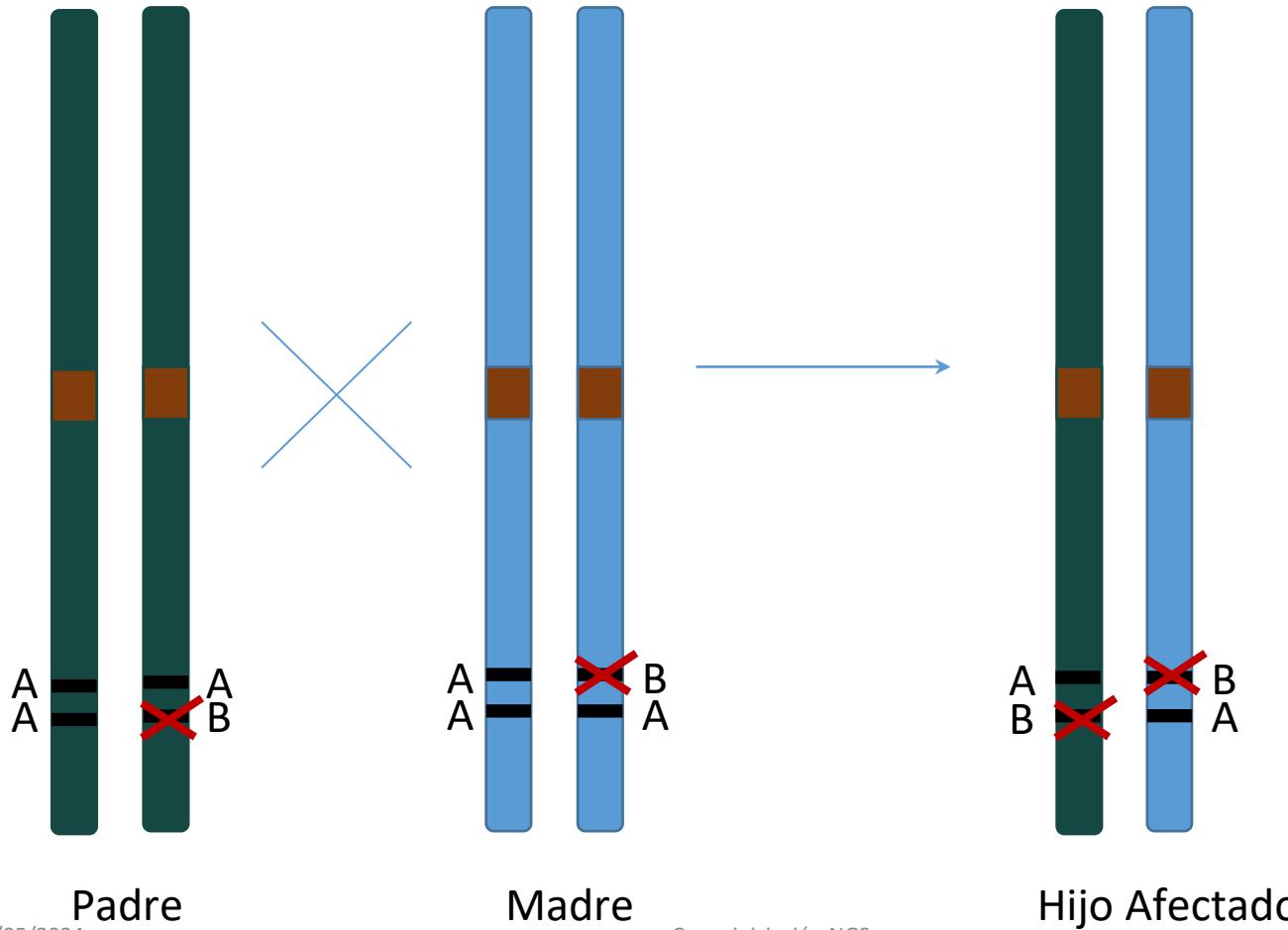
- Filtros ad-hoc
 - Primera aproximación:
 - FILTER = PASS
 - Genotipo =

Padre	0/0
Madre	0/0
Hijo	0/1 o 1/0



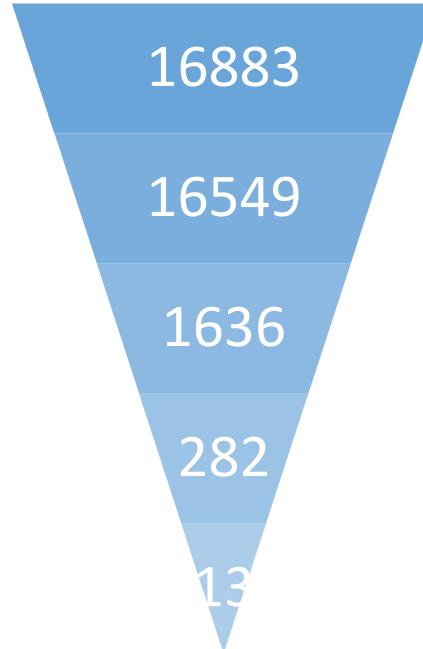
Modelos de enfermedad

- Modelo double-hit gene

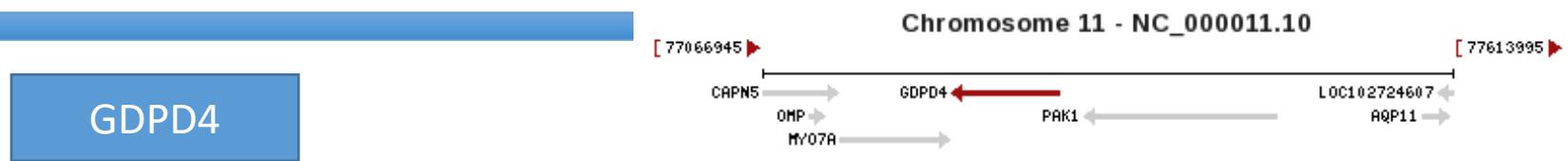


Modelo double-hit gene

- Filtros ad-hoc
 - Primera aproximación:
 - FILTER = PASS
 - Filtro por missense, frameshift, splicing o stop-gain
 - Mutation taster: A o D
 - Genotipo =



Modelo double-hit gene



- Glycerophosphodiester Phosphodiesterase domain-containing 4
- Proteína de membrana
- Relacionada con el metabolismo de glicerofosfolípidos.
- Relacionado mutaciones en este gen con el síndrome del shock tóxico (TSS).
- Variantes vistas en el gen:
 - Delección patogénica en el cromosoma 11 71680927-7794394
 - Relacionado con retraso en el desarrollo y fenotipos morfológicos significativos.



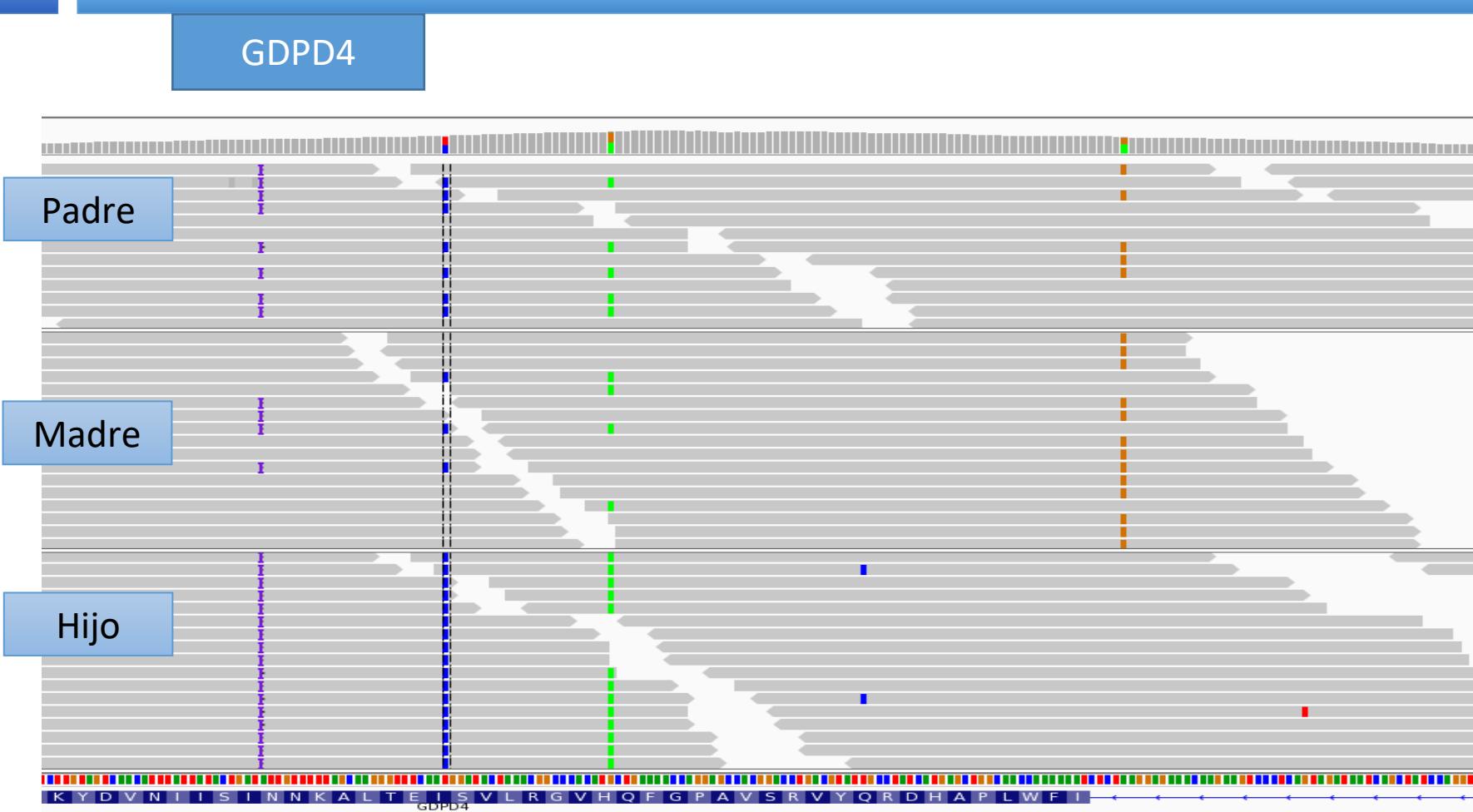
GOBIERNO
DE ESPAÑA

MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES



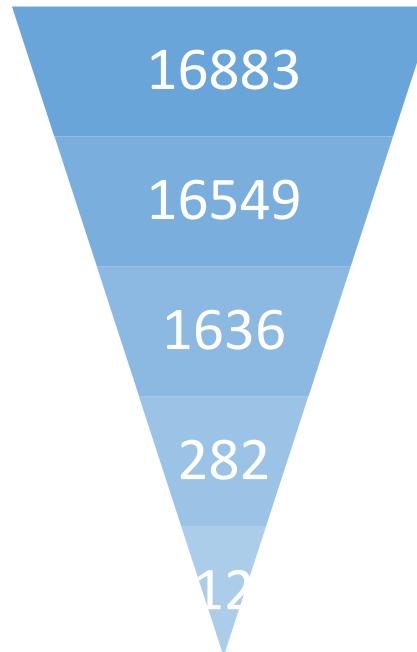
>X_BU-ISCIII

Modelo double-hit gene



Modelo double-hit gene

- Filtros ad-hoc
 - Segunda aproximación:
 - FILTER = PASS
 - Filtro por missense, frameshift, splicing o stop-gain
 - Mutation taster: A o D
 - Genotipo =



Padre	1/0	0/0
Madre	0/0	1/0
Hijo	0/1	0/1

Modelo double-hit gene

NBPF1

Chromosome	StartPosition	Reference AlternativeAllele	rsID	MostImportantFeatureGene	MostImportantGeneFeature	GeneDescription
1	16890484	G/C	rs12117084 (suspected)	NBPF1	missense (cys -> ser)	neuroblastoma breakpoint family, member 1 (Approved)
1	16909134	G/A	.	NBPF1	missense	

Mutación en el 5% de los reads

En madre e hijo. Posibles errores de alineamiento por tratarse de genes duplicados.

- Gen de la familia “breakpoint” de neuroblastoma. Docenas de genes duplicados localizados en duplicaciones segmentales en el cromosoma 1.
- Cambios en el número de copia se ha relacionado con enfermedades del desarrollo y neurogenéticas como microcefalia, macrocefalia, autismo, retraso mental, neuroblastoma, enfermedades del corazón congénitas, etc.

Creación de estándares

College of American Pathologists' Laboratory Standards for Next-Generation Sequencing Clinical Tests

Nazneen Aziz, PhD; Qin Zhao, PhD; Lynn Bry, MD, PhD; Denise K. Driscoll, MS, MT(ASCP)SBB; Birgit Funke, PhD; Jane S. Gibson, PhD; Wayne W. Grody, MD; Madhuri R. Hegde, PhD; Gerald A. Hoeltge, MD; Debra G. B. Leonard, MD, PhD; Jason D. Merker, MD, PhD; Rakesh Nagarajan, MD, PhD; Linda A. Palicki, MT(ASCP); Ryan S. Robetorye, MD; Iris Schrijver, MD; Karen E. Weck, MD; Karl V. Voelkerding, MD

- Recomendaciones en documentación, trazabilidad, validación, almacenamiento,...
 - Extracción de ADN
 - Preparación de librerías
 - Referencias y versiones
 - Pipeline bioinformático de análisis

Creación de estándares

Table 5

Rules for Combining Criteria to Classify Sequence Variants

Pathogenic

- 1 1 Very Strong (PVS1) *AND*
 - a. ≥ 1 Strong (PS1–PS4) *OR*
 - b. ≥ 2 Moderate (PM1–PM6) *OR*
 - c. 1 Moderate (PM1–PM6) and 1 Supporting (PP1–PP5) *OR*
 - d. ≥ 2 Supporting (PP1–PP5)
- 2 ≥ 2 Strong (PS1–PS4) *OR*
- 3 1 Strong (PS1–PS4) *AND*
 - a. ≥ 3 Moderate (PM1–PM6) *OR*
 - b. 2 Moderate (PM1–PM6) *AND* ≥ 2 Supporting (PP1–PP5) *OR*
 - c. 1 Moderate (PM1–PM6) *AND* ≥ 4 Supporting (PP1–PP5)

Likely Pathogenic

- 1 1 Very Strong (PVS1) *AND* 1 Moderate (PM1–PM6) *OR*
- 2 1 Strong (PS1–PS4) *AND* 1–2 Moderate (PM1–PM6) *OR*
- 3 1 Strong (PS1–PS4) *AND* ≥ 2 Supporting (PP1–PP5) *OR*
- 4 ≥ 3 Moderate (PM1–PM6) *OR*
- 5 2 Moderate (PM1–PM6) *AND* ≥ 2 Supporting (PP1–PP5) *OR*
- 6 1 Moderate (PM1–PM6) *AND* ≥ 4 Supporting (PP1–PP5)

Benign

- 1 1 Stand-Alone (BA1) *OR*
- 2 ≥ 2 Strong (BS1–BS4)

Likely Benign

- 1 1 Strong (BS1–BS4) and 1 Supporting (BP1–BP7) *OR*
- 2 ≥ 2 Supporting (BP1–BP7)

Interpretación de variantes.

* Variantes should be classified as Uncertain Significance if other criteria are unmet or the criteria for benign and pathogenic are contradictory.

Standards and Guidelines for the Interpretation of sequence variants. American College of Medical Genetics and Genomics. Association for Molecular Pathology. 2015

Ejemplo de variant calling: Bacterias

- Identificación de Brotes de origen alimentario, “Crisis del Pepino”

2011

Mayo

- 24 Primera muerte en Alemania
- 26 Alemania acusa a los pepinos españoles
- 30 Prohibición de importaciones de verduras de España y Alemania
- 31 Laboratorios alemanes desmienten oficialmente que los pepinos españoles sean el foco de infección

Junio

- 10 Resolución de la crisis

Causado por la toxi-infección de Escherichia coli enterohemorrágica (EHEC) (*Escherichia coli* O104:H4)

Muerte: 32 personas en Alemania, 1 Suecia y 1 Francia y 2263 infectados en 12 países de Europa.

Crisis Política y Económica Europa:
Alto impacto en la Economía Europea, mayor afectación en la Española



Secuenciación
Genoma

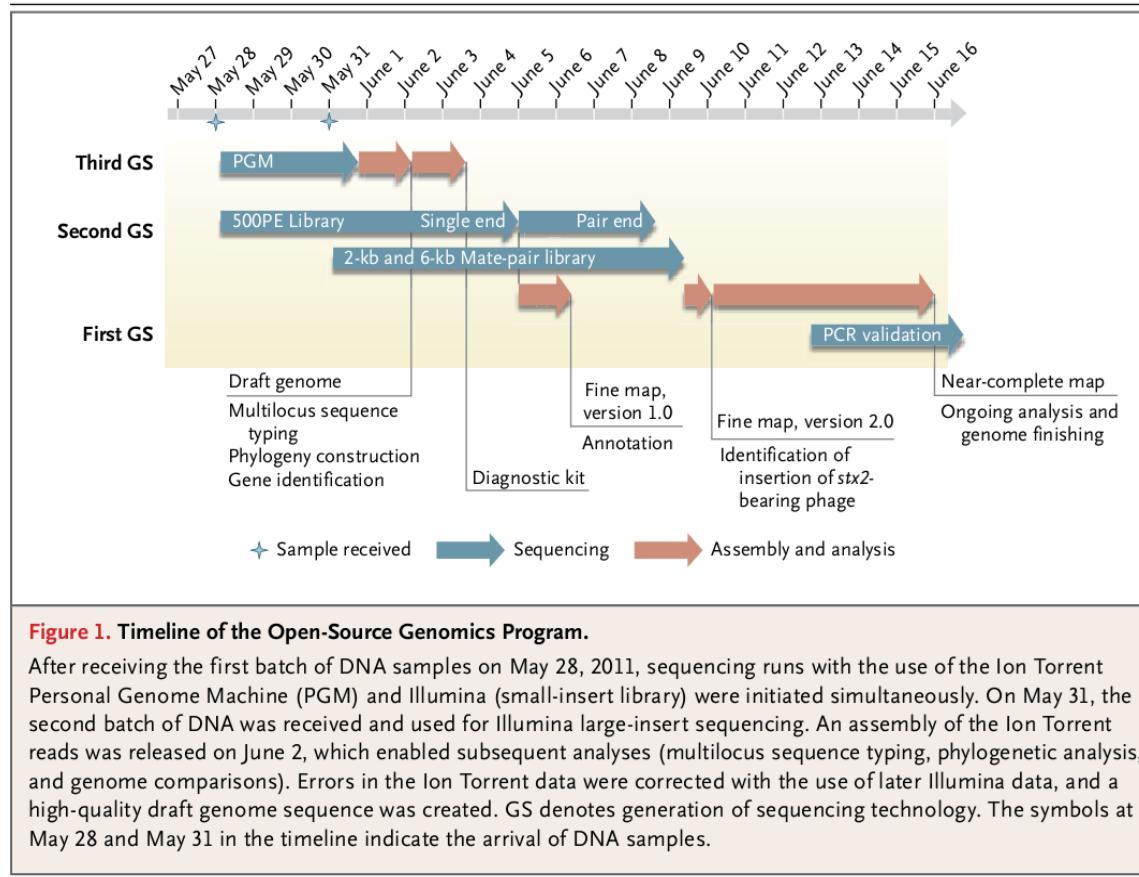
华大基因
BGI



Universitätsklinikum
Hamburg-Eppendorf

Ejemplo de variant calling: Bacterias

- Identificación de Brotes de origen alimentario, “Crisis del Pepino”



Ejemplo de variant calling: Bacterias

- Identificación de Brotes de origen alimentario, “Crisis del Pepino”

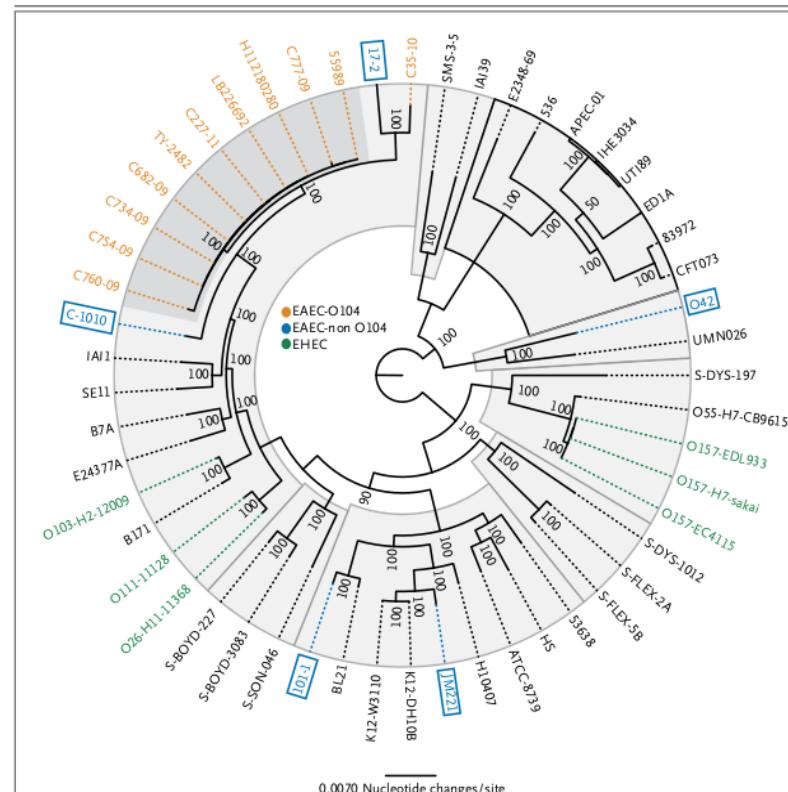


Figure 2. Phylogenetic Comparisons of 53 *Escherichia coli* and *Shigella* Isolates.

Genomic sequences were compared with the use of 100 bootstrap calculations, as described by Sahl et al.³⁵ The species-based phylogeny was inferred with the use of 2.56 Mbp of the conserved core genome. The O104:H4 isolates are shown in orange, the reference enteropathogenic *E. coli* (EAEC) isolates in blue, and the enterohemorrhagic *E. coli* isolates in green. (The classification of the other strains is shown in Fig. 4 and Table 4 in the Supplementary Appendix.) The O104:H4 isolates cluster into a single clade (dark gray); in contrast, the reference EAEC isolates are extremely divergent and are represented throughout the phylogeny.

Pipeline variant calling: Bacterias

Fast
Preprocessing

- Quality Control
- Low quality read filtering and trimming

Reference genome selection
&
Mapping

Variant
calling

High quality SNP selection
&
Build phylip matrix

- site coverage above 10x
- supported by at least 90% of the reads

Phylogenetic tree

- Evolution model selected by AIC criterium
- Maximum Likelihood tree

Cluster definition

Software disponible

- CFSAN SNP Pipeline

Extracción de SNPs de alta calidad de aislados relacionados

<http://snppipeline.readthedocs.io/en/latest/>

- GATK, modo haploide
- Samtools
- Varscan
- Snippy

Identificación de variantes haploides y construcción de filogenia usando core genome SNPs

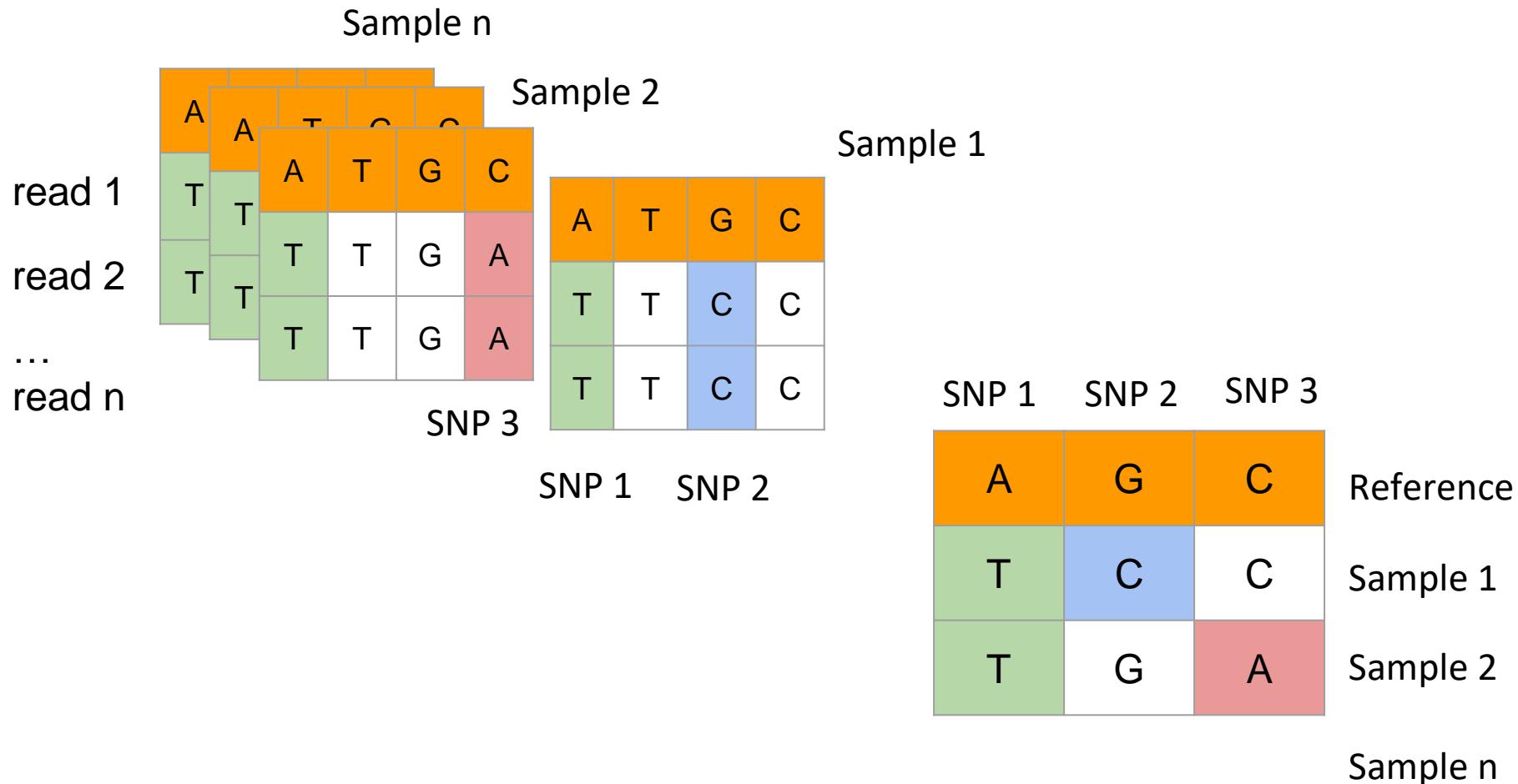
<http://github.com/tseemann/snippy>

- Live-SET

High-quality SNPs para crear filogenia para investigación de brotes

<https://github.com/lskatz/lyve-SET>

Generación de matriz de SNPs



Generación de matriz de SNPs

SNP matrix

SNP 1 SNP 2 SNP 3

	SNP 1	SNP 2	SNP 3
Reference	A	G	C
Sample 1	T	C	C
Sample 2	T	G	A

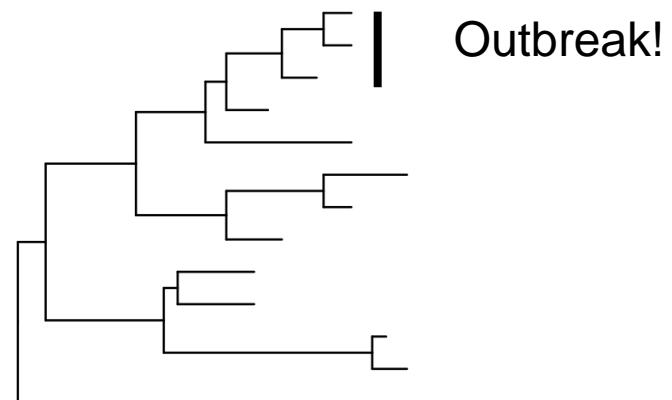
Sample n

Reference

Sample 1

Sample 2

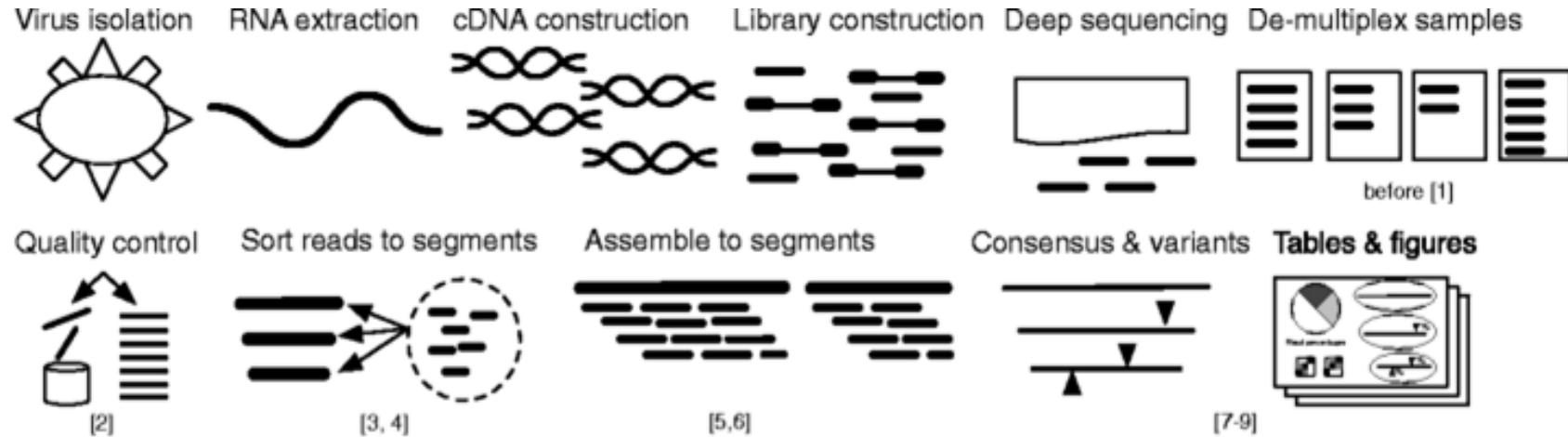
Phylogeny



Ejemplo de llamada a variantes: Virus

IRMA: Iterative Refinement Meta-Assembler

A



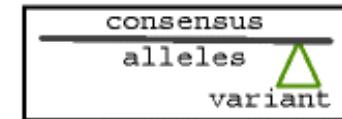
Shepard et al BMC Genomics 2016, 17:708

Ejemplo de llamada a variantes: Virus

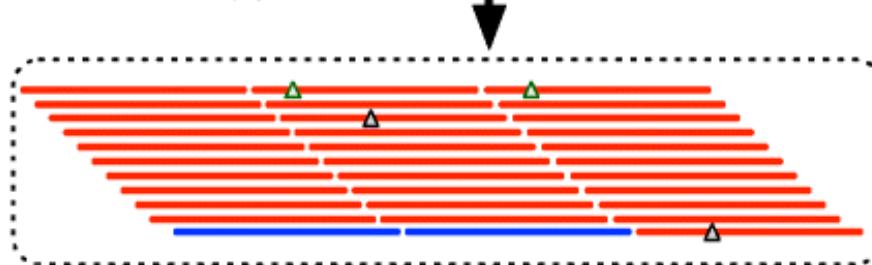
(1) donor virus 1



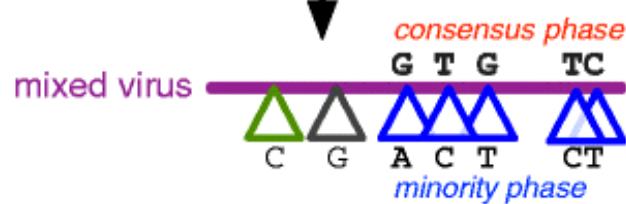
(1) donor virus 2



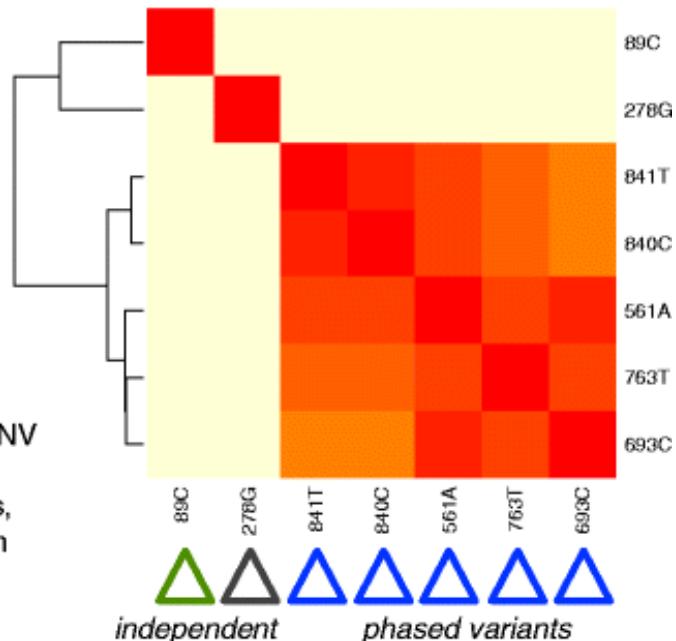
(2) 99% mixed with 1%



(3) assemble, call variants



(4) pairwise SNV associations, visualization



Shepard et al BMC Genomics 2016, 17:708



MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES



>X_BU-ISCIII

¿Preguntas?