Entregable: E10-E41
Valoración A14-A41

1.- Objetivos

Esta tarea ha consistido en reproducir un análisis de exoma de cáncer humano con llamada de variantes de tipo SNP/Indel usando el pipeline BWA-GATK-MUTECT (Li and Durbin 2009; McKenna, 2010; DePristo, 2011; Cibulskis et al 2013). Para la anotación de las variantes se usó la herramienta VEP de Ensembl (McLaren et al. 2016). En paralelo se usó dicho material para hacer un testado completo de todas las funciones de VariantSeq incluyendo las capacidades del sistema experto.

Este reporte de valoración de la actividad A14-A41 es parte material del entregable E10-E41.

2.- Material y métodos.

Los métodos empleados en esta prueba de concepto son los mismos usados en el artículo de investigación relacionado con este material (Trilla-Fuertes, et al. 2020).

Para llevar a cabo el análisis las muestras de cáncer se obtuvieron 5 muestras de cáncer a partir del SRA archive del NCBI que se detallan en tabla 1.

Concretamente se uso las siguientes librerías descargadas a partir del Bioproject accesible en esta URL https://www.ncbi.nlm.nih.gov/bioproject/PRJNA573670

Tabla 1: Muestras de exoma

Nombre de la librería	SRA Accessions
CAN2	SRR10164002
CAN3	SRR10163991
CAN4	SRR10163980
CAN5	SRR10163969
CAN12	SRR10163960

También se requirió el Resource Bundle de GATK para el release Hg19 que está disponible aquí https://gatk.broadinstitute.org/hc/en-us/articles/360035890811-Resource-bundle. También se requirió un fichero de intervalos basados en el sistema de captura seqCap VCRome V2 para exoma humano y un panel de normales (PON).

El fichero de intervalos puede ser descargado desde la siguiente URL:

https://sequencing.roche.com/en/support-resources/discontinued-products/seqcap-ez-hgsc-vcrome.html

El fichero PON se puede descargar desde este link:

https://ecampus.biotechvana.com/pluginfile.php/988/mod_folder/content/0/HPON.vcf?forcedownload=1

3.- Resultados de la valoración y material resultante

Todas las pruebas se realizaron tanto sobre las versiones RAP y RCP de la aplicación VariantSeq. Para facilitar de la realización de la prueba de concepto y dado que este material es de naturaleza big data, se ha habilitado un acceso FTP para acceder mediante el siguiente usuario anónimo y password y una carpeta con los entregables de esta prueba de concepto A14-A41 junto con otras asociadas al entregable E10-E41. Para acceder a dicho FTP recomendamos Filezilla que puede descargarse gratuitamente en https://filezilla-project.org. Las credenciales para acceder son concretamente las siguientes:

Servidor FTP: biotechvana.uv.es
Usuario: DIGITAL
Password: DiGi_19_21*

En concreto se debe acceder a la carpeta **01_valoracion_actividad_A14_A41_VariantSeq** donde se podrá encontrar:

- Carpeta step-by-step.
- Carpeta pipeline.

Para poder visualizar correctamente los resultados deben de descargarse al escritorio. Nótese que se ha creado una carpeta por modo de ejecución debido a que los resultados obtenidos en ambas versiones de la aplicación (RAP y RCP) son exactamente iguales y evitamos de esta forma la duplicidad de resultados. Este material se estructura de la siguiente manera:

En la carpeta step-by-step que contiene los resultados de la ejecución del protocolo SNP/Indels, se pueden encontrar las siguientes subcarpetas:

- **00_raw_data**: carpeta donde se depositan los archivos fastq sin procesar.
- **01_quality_analysis:** carpeta donde se depositan los resultados del análisis de calidad.
- **02_preprocessed_reads**: carpeta donde se depositan los resultados del pre-procesado
- **03_refseq:** carpeta donde se depositan los siguientes ficheros correspondientes al genoma humano (hg19), archivos de *training*, listado de intervalos y panel de genes normales (PON).
- **04_ mapping:** carpeta donde se depositan los resultados del mapeo.
- **05_addreplacegroups:** carpeta donde se depositan los resultados del post-procesado correspondientes a: AddReplaceGroups
- **06_MarkDuplicates:** carpeta donde se depositan los resultados del post-procesado correspondientes a: Markduplicates.
- **07_BSQR:** carpeta donde se depositan los resultados del post-procesado correspondientes a: BSQR.
- **08_variant_calling:** carpeta donde se depositan los archivos de resultados vcf de la llamada de variantes.
- **09_Variant_filtering:** carpeta donde se depositan los archivos de resultados vcf del filtrado de variantes.
- **10_annotation:** carpeta donde se depositan los archivos de resultados html y archivos de texto de la anotación de variantes.

En la carpeta pipeline: contiene los resultados de la ejecución del modo pipeline , se pueden encontrar las siguientes subcarpetas:

- **01_FASTCQ** carpeta donde se depositan los resultados del análisis de calidad.
- **02_CUTADAPT**: carpeta donde se depositan los resultados del pre-procesado procedentes de CUTADAPT
- **03_PRINSEQ**: carpeta donde se depositan los resultados del pre-procesado procedentes de PRINSEQ.
- **04_FASTQC**: carpeta donde se depositan los resultados del segundo análisis de calidad tras el pre-procesado.
- **05_Bwa:** carpeta donde se depositan los resultados del mapeo.
- **05_addreplacegroups:** carpeta donde se depositan los resultados del post-procesado correspondientes a: AddReplaceGroups
- **06_MarkDuplicates:** carpeta donde se depositan los resultados del post-procesado correspondientes a: Markduplicates.
- **07_BSQR:** carpeta donde se depositan los resultados del post-procesado correspondientes a: BSQR.
- **08_variant_calling:** carpeta donde se depositan los archivos de resultados vcf procedentes de la llamada de variantes.

4.- Testado de las funciones de VariantSeq

Los pasos reproducidos para realizar el análisis de exoma fueron los siguientes:

- Modo de ejecución: STEP-BY-STEP

1. Quality analysis: FASTQC (Andrews 2016)
2. Preprocessing: PRINSEQ (Schmieder and Edwards 2011)
3. Mapping: Mapping DNA → Bwa (Li and Durbin 2009)
4. Postprocessing: Picard tools → AddReplaceReadGroups (Wysoker, et al. 2011)
5. Postprocessing: Picard tools → MarkDuplicates (Wysoker, et al. 2011)
6. Postprocessing: GATK tools → BQSR (McKenna, et al. 2010; DePristo, et al. 2011; Cibulskis, et al. 2013)
7. Variant Calling: Mutect2 (McKenna, et al. 2010; DePristo, et al. 2011; Cibulskis, et al. 2013)
8. Variants Filtering: GATK - Cross-Sample contamination (McKenna, et al. 2010; DePristo, et al. 2011; Cibulskis, et al. 2013)
9. Variants Filtering: GATK – FilterMutectCalls (McKenna, et al. 2010; DePristo, et al. 2011; Cibulskis, et al. 2013
10. Annotation: VEP – Variant Effect Prediction (McLaren, et al. 2016)

- Modo de ejecución: PIPELINE

  Muestras: Pair-End

1. Quality analysis: FASTQC (Andrews 2016)
2. Preprocessing: PRINSEQ (Schmieder and Edwards 2011)
3. Mapping: Bwa (Li and Durbin 2009)

   Variant type: Somatic

4. Postprocessing: Picard tools → AddReplaceReadGroups (Wysoker, et al. 2011)
5. Postprocessing: Picard tools → MarkDuplicates (Wysoker, et al. 2011)
6. Postprocessing: GATK tools → BQSR (McKenna, et al. 2010; DePristo, et al. 2011; Cibulskis, et al. 2013)
7. Variant Caller: GATK - Mutect2 (McKenna, et al. 2010; DePristo, et al. 2011; Cibulskis, et al. 2013)

De forma adicional y aunque no forma parte de este entregable, hemos aprovechado esta prueba de concepto para crear un tutorial de uso en el análisis de variantes con VariantSeq. Pueden acceder al tutorial de VariantSeq en el siguiente enlace https://ecampus.biotechvana.com/course/view.php?id=19

A continuación, se presentan dos tablas detalladas con las pruebas realizadas a la aplicación VariantSeq en los dos modos de ejecución (step-by-step y pipeline) tanto en versión RAP como versión RCP. Por simplicidad se añade una tabla común a ambas versiones disponibles de la aplicación ya que están compuestos por las mismas herramientas.

Tabla 2. Step-by-step mode

Versión	Modo de ejecución	Herramienta	Descripción	Cumple Requisitos
RAP y RCP	STEP-BY-STEP: SNP/Indels	Preprocessing: Quality analysis FASTQC	Se realiza un análisis de calidad de los archivos fastq sin procesar.	Si Como resultado se obtiene un informe en el que se muestran los parámetros analizados en las muestras.
		Preprocessing: Demultiplex FastqMidCleaner	Clasifica y divide las lecturas de secuenciación de los archivos fastq en archivos separados de acuerdo con identificadores moleculares predefinidos (MID).	Si Como resultado se obtienen nuevos archivos fastq.
		Preprocessing: Trimming and cleaning CUTADAPT	Encuentra y elimina secuencias de adaptadores, primers, colas poli-A y otros tipos de artefactos de secuenciación presentes en los archivos sin procesar fastq.	Si Como resultado se obtienen nuevos archivos fastq de los cuales se han eliminado las secuencias de adaptadores u otros artefactos de secuenciación.
		Preprocessing: Trimming and cleaning PRINSEQ	Filtra, corta o reformatea los archivos sin procesar fastq añadiendo una serie de parámetros basados en el análisis de calidad.	Si Como resultado se obtienen nuevos archivos fastq modificados según los parámetros introducidos.
		Preprocessing: Trimming and cleaning Trimommatic	Es una herramienta de recorte específica para muestras tanto pair-end como single-end obtenidas a través de secuenciación NGS de Illumina.	Si Como resultado se obtienen nuevos archivos fastq modificados según los parámetros introducidos
		Preprocessing: Trimming and cleaning Fastx Toolkit	Esta función acoge a un conjunto de herramientas destinadas al pre-procesado de las muestras fastq.	Si Según la herramienta usada se pueden obtener nuevos archivos fastq, resultados estadísticos, gráficos…
		Preprocessing: Prepseq FastqCollapser	Elimina las lecturas duplicadas en las muestras fastq. Esta herramienta se basa en el análisis de calidad según el contenido de secuencia.	Si Como resultados se obtienen nuevos archivos fastq.
		Preprocessing: Prepseq FastqIntersect	Compara la información de dos archivos pair-end que han sido pre-procesados de forma independiente y la información de ambos archivos para editarlos manteniendo solo aquellas lecturas, y en el mismo orden, que están presentes en ambos archivos	Si Como resultados se obtienen nuevos archivos fastq.
		Preprocessing: Picard Picard - CreateSequenceDictionary	Esta herramienta crea un archivo de diccionario de secuencia a partir del genoma de referencia de la especie	Si Como resultado se obtiene un archivo con extensión .dict
		Mapping: DNAseq mappers Bowtie2	Bowtie2 es una herramienta de alineamiento de secuencias de largo tamaño (entre 50-100pb).	Si Como resultado se obtienen archivos de mapeo con extensión .bam.
		Mapping: DNAseq mappers Bwa	BWA es un paquete de software para mapear secuencias de baja divergencia contra grandes genomas de referencia	Si Como resultado se obtienen archivos de mapeo con extensión .bam.

RAP y RCP	STEP-BY-STEP: SNP/Indels	Mapping: RNAseq mappers Tophat	Tophat alinea lecturas de RNA-seq con un genoma de referencia que identifica las uniones de empalme exón-exón.	Si Como resultado se obtienen archivos de mapeo con extensión .bam.
		Mapping: RNAseq mappers STAR	STAR es un alineador universal para mapear lecturas y transcripciones empalmadas de RNAseq	Si Como resultado se obtienen archivos de mapeo con extensión .bam.
		Mapping: RNAseq mappers Hisat2	Hisat2 es un sistema altamente eficiente para alinear lecturas de experimentos de secuenciación de RNA.	Si Como resultado se obtienen archivos de mapeo con extensión .bam.
		Training Sets: Generate training set	Esta herramienta permite crear un training set utilizando el genoma de referencia y los archivos bam procedentes del mapeo.	Si Como resultado se obtienen un archivo con extensión .vcf en el que se almacenan un listado de variantes y un índice con extensión vcf.idx.
		Training Sets: GATK – CreateSomaticPanelOfNormals	Genera un panel de genes normales. Este panel de genes es un recurso usado en el análisis de variantes somáticas.	Si Como resultado se obtiene un archivo con extensión .vcf y un índice con extensión .vcf.idx
		Postprocessing: GATK tools BQSR	Esta función detecta errores sistemáticos generados durante la secuenciación cuando estima la precisión de cada llamada base.	Si Como resultado se obtienen archivos con extensión .bam
		Postprocessing: GATK tools SplitNCigarReads	SplitNCigars divide las lecturas que contienen Ns	Si Como resultado se obtienen archivos con extensión .bam
		Postprocessing: GATK tools LeftAlignIndels	Esta herramienta alinea a la izquierda cualquier indel (inserción/delección) contenida en los archivos bam.	Si Como resultado se obtienen archivos con extensión .bam
		Postprocessing: GATK tools Indel Local Realignment	Se lleva a cabo un realineamiento usando RealignerTargetCreator and IndelRealigner de GATK.	Si Como resultado se obtienen archivos con extensión .bam
		Postprocessing: Picard tools Picard - AddReplaceReadGroups	Reemplaza grupos de lectura en los archivos bam asignando un único grupo a todas las lecturas.	Si Como resultado se obtienen archivos con extensión .bam
		Postprocessing: Picard tools Picard - CreateSequenceDictionary	Esta herramienta crea un archivo de diccionario de secuencia a partir del genoma de referencia de la especie	Si Como resultado se obtiene un archivo con extensión .dict
		Postprocessing: Picard tools Picard - MarkDuplicates	Esta herramienta localiza y etiqueta las lecturas duplicadas en las cuales las lecturas duplicadas se definen como originadas en un solo fragmento de ADN	Si Como resultado se obtienen archivos con extensión .bam
		Postprocessing: SAM tools Samtools: Index	La herramienta Index de Samtools crea índices con extensión .bai o .csi procedentes de un archivo bam.	Si Como resultado se obtienen archivos con extensión .bam

RAP y RCP	STEP-BY-STEP: SNP/Indels	Postprocessing: SAM tools Samtools: Sort	La herramienta Sort de Samtools ordena alineamientos de los archivos bam por coordenadas más a la izquierda o por nombre de lectura cuando se utiliza la opción -n	Si Como resultado se obtienen archivos con extensión .bam
		Postprocessing: SAM tools Samtools: Merge	La herramienta Merge de Samtools fusiona varios archivos de alineación ordenados que contiene todos los registros de entrada mientras mantiene el orden de clasificación	Si Como resultado se obtienen archivos con extensión .bam
		Coverage analysis	Realiza un análisis de cobertura y amplitud de las lecturas en el alineamiento	Si Como resultado se obtienen archivos con extensión .vcf
		Variant Calling: GATK Based Haplotype caller GATK - HaplotypeCaller	HaplotypeCaller realiza la llamada de SNPs de la línea germinal a través de re-ensamblaje local de haplotipos	Si Como resultado se obtienen archivos con extensión .vcf
		Variant Calling: GATK Based Haplotype caller GATK – CombineGVCFs	Esta herramienta fusiona uno o más archivos procedentes de HaplotypeCaller GVCF en un único GVCF con las anotaciones apropiadas	Si Como resultado se obtienen archivos con extensión .vcf
		Variant Calling: GATK Based Haplotype caller GATK - GenotypeGVCFs	Realiza un genotipado conjunto de una o mas muestras pre-llamadas de HaplotypeCaller	Si Como resultado se obtienen archivos con extensión .vcf
		Variant Calling: GATK Based Mutect2	Realiza una llamada de variantes somáticas e indels a través local del ensamblaje local de haplotipos.	Si Como resultado se obtienen archivos con extensión .vcf
		Variant Calling: VarScan2 Based VarScan2 Germline variants	Realiza la llamada de variantes germinales (SNPs e indels) usando un método heurístico y test estadístico basado en el número de alineamientos de las lecturas respaldado por cada alelo.	Si Como resultado se obtienen archivos con extensión .vcf
		Variant Calling: VarScan2 Based VarScan2 Somatic variants	Realiza la llamada de variantes somáticas de multimuestras de SNP e Indels utilizando pares de tumores normales (pares de muestras de muestras tumorales de un paciente y otra muestra de un tejido del mismo paciente, no afectado por el tumor)	Si Como resultado se obtienen dos tipos de archivos. Uno de ellos con extensión .snp (contiene los resultados de SNVs) y otro con extensión .indel (contiene los resultados de indels)
		Variant Calling: VarScan2 Based VarScan2 trio calling	Se usa para analizar pedigríes familiares (padre-madre-hijo) con el objetivo de identificar alelos transmitidos y/o mutaciones de novo que puedan ser responsables de la enfermedad.	Si Como resultado se obtienen dos tipos de archivos. Uno de ellos con extensión .snp.vcf (contiene los resultados de los SNVs) y otro con extensión .indel.vcf (contiene los resultados de indels)
		Variants Filtering: GATK – Variant Quality Score Recalibration	Se basa en el comando VariantRecalibrator de GATK para construir un modelo de recalibración para puntuar la calidad de la variante con fines de filtrado	Si Como resultado se obtienen archivos con extensión .vcf
		Variants Filtering: GATK – VariantFiltration	Filtra las llamadas de variantes basadas en el campo de anotación FORMAT y/o INFO	Si Como resultado se obtienen archivos con extensión .vcf
		Variants Filtering: GATK – Cross-Sample Contamination	Calcula la fracción de lecturas procedentes de la contaminación entre muestras	Si

				Como resultado se obtienen archivos con extensión .contamination.table
		Variants Filtering: GATK – FilterMutectCalls	Filtra variantes somáticas SNVs e indels procedentes de Mutect2	Si Como resultado se obtienen archivos con extensión .vcf
		Annotation: VEP – Variant Effect Prediction	Determina el efecto de las variantes (SNPs, inserciones, delecciones, CVNs o variantes estructurales) en genes, transcritos y secuencias de proteína así como regiones reguladoras.	Si Como resultado se obtiene dos archivos con distintos formatos uno de ellos en formato html y el segundo resultado se obtiene en un archivo de texto
RAP/RCP	Sistema experto	Recomendaciones y soluciones automatizadas	Según el panel de reportes el sistema experto permite dar una solución a un problema dado, en la forma de recomendación o aplicación directa de la re-ejecución del proceso.	La aplicación acierta en un 50% con la resolución de recomendación a aplicar. La aplicación de las mismas es 100% correcta si bien es un elemento prototípico que necesita ser depurado y sometido a m´ss entrenamiento. El que se ha aplicado aquí es básico, si bien suficiente para que podamos integrarlo en las aplicaciones de GPRO operativo y funcionando, pero destacando que es una aplicación en fase beta.

Tabla 3. Pipeline mode

Versión	Modo de ejecución	Herramienta	Descripción	Cumple Requisitos
RAP y RCP	PIPELINE MODE	Preprocessing: Quality analysis FASTQC	Se realiza un análisis de calidad de los archivos fastq sin procesar.	Si Como resultado se obtiene un informe en el que se muestran los parámetros analizados en las
		Preprocessing: Trimming and cleaning CUTADAPT	Encuentra y elimina secuencias de adaptadores, primers, colas poli-A y otros tipos de artefactos de secuenciación presentes en los archivos sin procesar fastq.	Si Como resultado se obtienen nuevos archivos fastq de los cuales se han eliminado las secuencias de adaptadores u otros artefactos de secuenciación.
		Preprocessing: Trimming and cleaning PRINSEQ	Filtra, corta o reformatea los archivos sin procesar fastq añadiendo una serie de parámetros basados en el análisis de calidad.	Si Como resultado se obtienen nuevos archivos fastq modificados según los parámetros introducidos.
		Preprocessing: Trimming and cleaning Trimommatic	Es una herramienta de recorte específica para muestras tanto pair-end como single-end obtenidas a través de secuenciación NGS de Illumina.	Si Como resultado se obtienen nuevos archivos fastq modificados según los parámetros introducidos

RAP y RCP	PIPELINE MODE	Mapping: DNAseq mappers Bowtie2	Bowtie2 es una herramienta de alineamiento de secuencias de largo tamaño (entre 50-100pb).	Si Como resultado se obtienen archivos de mapeo con extensión .bam.
		Mapping: DNAseq mappers Bwa	BWA es un paquete de software para mapear secuencias de baja divergencia contra grandes genomas de referencia	Si Como resultado se obtienen archivos de mapeo con extensión .bam.
		Mapping: RNAseq mappers Tophat	Tophat alinea lecturas de RNA-seq con un genoma de referencia que identifica las uniones de empalme exón-exón.	Si Como resultado se obtienen archivos de mapeo con extensión .bam.
		Mapping: RNAseq mappers STAR	STAR es un alineador universal para mapear lecturas y transcripciones empalmadas de RNAseq	Si Como resultado se obtienen archivos de mapeo con extensión .bam.
		Mapping: RNAseq mappers Hisat2	Hisat2 es un sistema altamente eficiente para alinear lecturas de experimentos de secuenciación de RNA.	Si Como resultado se obtienen archivos de mapeo con extensión .bam.
		Postprocessing: GATK tools BQSR	Esta función detecta errores sistemáticos generados durante la secuenciación cuando estima la precisión de cada llamada base.	Si Como resultado se obtienen archivos con extensión .bam
		Postprocessing: GATK tools SplitNCigarReads	SplitNCigars divide las lecturas que contienen Ns	Si Como resultado se obtienen archivos con extensión .bam
		Postprocessing: GATK tools LeftAlignIndels	Esta herramienta alinea a la izquierda cualquier indel (inserción/deleccion) contenida en los archivos bam.	Si Como resultado se obtienen archivos con extensión .bam
		Postprocessing: GATK tools Indel Local Realignment	Se lleva a cabo un realineamiento usando RealignerTargetCreator and IndelRealigner de GATK.	Si Como resultado se obtienen archivos con extensión .bam
		Postprocessing: Picard tools Picard - AddReplaceReadGroups	Reemplaza grupos de lectura en los archivos bam asignando un único grupo a todas las lecturas.	Si Como resultado se obtienen archivos con extensión .bam
		Postprocessing: Picard tools Picard - CreateSequenceDictionary	Esta herramienta crea un archivo de diccionario de secuencia a partir del genoma de referencia de la especie	Si Como resultado se obtiene un archivo con extensión .dict
		Postprocessing: Picard tools Picard - MarkDuplicates	Esta herramienta localiza y etiqueta las lecturas duplicadas en las cuales las lecturas duplicadas se definen como originadas en un solo fragmento de ADN	Si Como resultado se obtienen archivos con extensión .bam

		Variant Calling: GATK Based Haplotype caller GATK - HaplotypeCaller	HaplotypeCaller realiza la llamada de SNPs de la línea germinal a través de re-ensamblaje local de haplotipos	Si Como resultado se obtienen archivos con extensión .vcf
		Variant Calling: GATK Based Haplotype caller GATK – CombineGVCFs	Esta herramienta fusiona uno o más archivos procedentes de HaplotypeCaller GVCF en un único GVCF con las anotaciones apropiadas	Si Como resultado se obtienen archivos con extensión .vcf
		Variant Calling: GATK Based Haplotype caller GATK - GenotypeGVCFs	Realiza un genotipado conjunto de una o mas muestras pre-llamadas de HaplotypeCaller	Si Como resultado se obtienen archivos con extensión .vcf
		Variant Calling: GATK Based Mutect2	Realiza una llamada de variantes somáticas e indels a través local del ensamblaje local de haplotipos.	Si Como resultado se obtienen archivos con extensión .vcf
		Variant Calling: VarScan2 Based VarScan2 Germline variants	Realiza la llamada de variantes germinales (SNPs e indels) usando un método heurístico y test estadístico basado en el número de alineamientos de las lecturas respaldado por cada alelo.	Si Como resultado se obtienen archivos con extensión .vcf
		Variant Calling: VarScan2 Based VarScan2 Somatic variants	Realiza la llamada de variantes somáticas de multimuestras de SNP e Indels utilizando pares de tumores normales (pares de muestras de muestras tumorales de un paciente y otra muestra de un tejido del mismo paciente, no afectado por el tumor)	Si Como resultado se obtienen dos tipos de archivos. Uno de ellos con extensión .snp (contiene los resultados de SNVs) y otro con extensión .indel (contiene los resultados de indels)
RAP/RCP	Sistema experto	Recomendaciones y soluciones automatizadas	Según el panel de reportes el sistema experto permite dar una solución a un problema dado, en la forma de recomendación o aplicación directa de la re-ejecución del proceso.	La aplicación acierta en un 50% con la resolución de recomendación aplicar. La aplicación de las mismas es 100% correcta si bien es un elemento prototípico que necesita ser depurado y sometido a más entrenamiento. El que se ha aplicado aquí es básico,

5.- Conclusiones

Todos los análisis de reprodujeron con éxito tanto usando el modo step-by-step como el modo pipeline tanto en la versión RCP como la versión RAP de la aplicación VariantSeq. Se verifica que todas las herramientas comprobadas funcionan correctamente y la aplicación está operativa y correctamente funcionando para su uso.

6.- Bibliografía

- Andrews S. 2016. FastQC: a quality control tool for high throughput sequence data.
- Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. 2013. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 31:213-219.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491-498.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297-1303.
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. Genome Biology 17:122.
- Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863-864.
- Trilla-Fuertes L, Ghanem I, Maurel J, L GP, Mendiola M, Pena C, Lopez-Vacas R, Prado-Vazquez G, Lopez-Camacho E, Zapater-Moros A, et al. 2020. Comprehensive Characterization of the Mutational Landscape in Localized Anal Squamous Cell Carcinoma. Translational oncology 13:100778.
- Wysoker A, Tibbetts K, Fennell T. 2011. Picard.