



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE MEDICINA

BIOLOGÍA EXPERIMENTAL

CARACTERIZACIÓN DE LOS SEGMENTOS ÚNICOS DE CITOMEGALOVIRUS HUMANO (HCMV) Y HERPESVIRUS HUMANO 6 (HHV6) EN PERSONAS CON VIH POR SECUENCIACIÓN DE SIGUIENTE GENERACIÓN

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

CERVANTES ROBLES ALFONSO DE JESÚS

**TUTOR PRINCIPAL DE TESIS: DR. EDGAR ENRIQUE SEVILLA REYES
FACULTAD DE MEDICINA, UNAM**

**COMITÉ TUTOR: DRA. LORENA SOFÍA OROZCO OROZCO
FACULTAD DE MEDICINA, UNAM**

DR. JOSÉ IVAN SÁNCHEZ BETANCOURT

FACULTAD DE MEDICINA VETERINARIA Y ZOOTECNIA, UNAM

MÉXICO, CD. MX., MAYO, 2016



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE MEDICINA

BIOLOGÍA EXPERIMENTAL

**CARACTERIZACIÓN DE LOS SEGMENTOS ÚNICOS DE CITOMEGALOVIRUS
HUMANO (HCMV) Y HERPESVIRUS HUMANO 6 (HHV6) EN PERSONAS
CON VIH POR SECUENCIACIÓN DE SIGUIENTE GENERACIÓN**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

CERVANTES ROBLES ALFONSO DE JESÚS

TUTOR PRINCIPAL DE TESIS: DR. EDGAR ENRIQUE SEVILLA REYES

FACULTAD DE MEDICINA, UNAM

COMITÉ TUTOR: DRA. LORENA SOFÍA OROZCO OROZCO

FACULTAD DE MEDICINA, UNAM

DR. JOSÉ IVAN SÁNCHEZ BETANCOURT

FACULTAD DE MEDICINA VETERINARIA Y ZOOTECNIA, UNAM

MÉXICO, CD. MX., MAYO, 2016

Dr. Isidro Ávila Martínez
Director General de Administración Escolar, UNAM
Presente

Me permito informar a usted que el Subcomité de Biología Experimental y Biomedicina del Posgrado en Ciencias Biológicas, en su sesión ordinaria del día 28 de septiembre de 2015, aprobó el grado para la presentación del examen para obtener el grado de **MAESTRO EN CIENCIAS BIOLÓGICAS** del alumno **CERVANTES ROBLES ALFONSO DE JESÚS** con número de cuenta 514012472 con la tesis titulada **"CARACTERIZACIÓN DE LOS SEGMENTOS ÚNICOS DE CITOMEGALOVIRUS HUMANO (HCMV) Y HERPES VIRUS 6 (HHV6) EN PERSONAS CON VIH POR SECUENCIACIÓN DE SIGUIENTE GENERACIÓN"**, realizada bajo la dirección del **DR. EDGAR ENRIQUE SEVILLA REYES**:

Presidente: DR. ARTURO CARLOS II BECERRA BRACHO
Vocal: DRA. ROSA ELENA SARMIENTO SILVA
Secretario: DRA. LORENA SOFÍA OROZCO OROZCO
Supiente: DR. JOEL ARMANDO VÁZQUEZ PÉREZ
Supiente: DR. JOSÉ IVÁN SÁNCHEZ BETANCOURT

Sin otro particular, me es grato enviarle un cordial saludo.

ATENTAMENTE
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Cd. Universitaria, Cd. Mx., a 04 de mayo de 2016



DRA. MARÍA DEL CORO ARIZMENDI ARRIAGA
COORDINADORA DEL PROGRAMA



Agradecimientos.

A la Universidad Nacional Autónoma de México (UNAM) y a su posgrado de Ciencias biológicas en el área de biología experimental, por darme la oportunidad de crecer académicamente y como persona, por las nuevas experiencias y el conocimiento que me permitió adquirir.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) y a su Programa Nacional de Posgrados de Calidad (PNPC), por apoyarme con la beca de manutención durante el periodo 2013-2015 a través de la beca No. 363077.

A mi director de tesis, el Dr. Edgar E. Sevilla Reyes por permitirme ser parte de su equipo, por guiarme en toda esta nueva experiencia, por proporcionarme todas las facilidades para realizar el presente trabajo y por toda la paciencia que me tuvo cada día y en cada pasa de este proyecto. Por resolver todas mis dudas, por más simples que estas fueran las dudas más simples, pero sobre todo por confiar en mí.

A los miembros de mi comité, la Dra. Lorena S. Orozco Orozco y el Dr. José I. Sánchez Betancourt, por todos sus consejos, ideas y contribuciones a mi trabajo, por guiarme hacia las metas establecidas, pero sobre todo por ayudarme a desarrollar la confianza para defender este trabajo y los posteriores.

Agradecimientos personales.

A la Q.F.B. Elvira Piten Isidro, por el apoyo y la ayuda brindada para realizar y analizar los diferentes experimentos, por las nuevas herramientas en las que me instruyo y en general por su colaboración durante toda la maestría.

Al Dr. Gustavo Reyes Terán por proporcionarme un espacio en el CIENI (Departamento de Investigación de Enfermedades Infecciosas), y al Dr. Santiago Ávila Ríos por el material e instrumental proporcionado durante el proyecto.

A la Q.B.P Margarita Matías Florentino (Magui) por ayudarme a dar los primeros pasos dentro de la secuenciación masiva, pero sobre todo le agradezco por su paciencia y todas el tiempo que le brindo a este proyecto, ya que ella fue uno de los principales elementos que permitió a este proyecto alcanzar sus objetivos.

Al Instituto Nacional de Enfermedades Respiratorias (INER) y en concreto al Departamento de Investigación en Enfermedades Infecciosas (CIENI), por permitirme utilizar sus instalaciones y por proporcionarme todas las herramientas para realizar mi maestría.

A la Red Temática de Inmunología en Cancer y Enfermedades Infecciosas (INMUNOCANEI), por el patrocinio proporcionado para asistir al IX Congreso Nacional de Virología, en el que presente el poster titulado Prevalencia de Material Genético de Betaherpesvirus en Sangre de Personas que Viven con VIH en México.

Al personal técnico y administrativo del posgrado en Ciencias Biológicas, por apoyarme con todos los tramites implicados durante mi permanencia en el posgrado. Agradezco por su esfuerzo y paciencia para aclarar todas nuestras dudas.

Quedo muy agradecido con la Q.F.B Elvira Piten Isidro, la M.C. María de los Ángeles Mondragón, con Christopher Ormsby (Keo), y el equipo de informática dirigido por el Ing. Pablo Cesar, no solo por el apoyo académico y laboral que me proporcionaron durante la

maestría, también por sus consejos, las nuevas experiencias y aventuras que compartimos, pero sobre todo por su amistad.

Nuevamente a mi tutor, el Dr. Edgar E. Sevilla, por ser mi guía durante estos años, por todos los consejos y el apoyo brindado durante este tiempo, por su gran paciencia y por las múltiples destrezas que me propició a desarrollar. Siempre recordaré afectuosamente la forma en la que nos enseñaba no solo en el campo académico, sino también en el de la vida, pero sobre todo su gran pasión por los chocolates y los momentos durante la comida, en los que comparaba la composición nutricional de mis bocadillos con los de él 😊.

A mis padres, quienes me apoyaron en todo momento y en sentido, quienes a pesar de no estar cerca y tener múltiples diversas tareas buscaban la forma de estar al pendiente de mí, sin importar lo complicado que pueda llegar a ser comunicarse con migo (quienes me conocen saben a qué me refiero jeje!) y porque sé que mientras puedan, siempre estarán ahí para mí, sin importar el día, hora o lugar.

A la Dra. Patricia Ramos Morales, por todo el apoyo brindado durante este tiempo y su gran sentido del humor.

A mi novia Citlalli Andrade Ramos, por tenerme paciencia y por estar ahí en los buenos momentos, pero sobre todo en los difíciles, cuando las cosas no salían como uno las planeaba o el estrés me agobiaba, por apoyarme en esta nueva aventura y en las que están por venir.

Y finalmente quiero dedicar esta tesis a mis abuelos María Agudo Guzmán y Julián Robles Méndez, quienes me enseñaron a nunca darme por vencido a pesar de las adversidades, y a luchar hasta el último momento sin importar la situación; a enfrentar los problemas con actitud y humor, pero sobre todo a nunca perder la esperanza, ya que al final pase lo que pase todo estará bien.

ÍNDICE.

1.	ÍNDICE DE TABLAS.	i
2.	ÍNDICE DE FIGURAS.	ii
3.	RESUMEN.	1
4.	ABSTRACT.	2
5.	INTRODUCCIÓN.	3
5.1	Citomegalovirus Humano (HCMV).	3
5.1.1	Epidemiología.	3
5.2	Herpesvirus humano 6 (HHV6).	6
5.2.1	Epidemiología.	7
6.	OBJETIVOS.	9
7.	ANTECEDENTES.	9
7.1	Secuenciación de HCMV	9
7.2	Secuenciación de HHV6	11
7.2.1	Integración cromosomal de HHV6.	12
7.2.1.1	Patologías asociadas a la integración cromosomal de HHV6.	14
7.3	Secuenciación masiva del DNA.	15
7.3.1	Antecedentes de la secuenciación masiva.	15
7.3.2	Método de degradación química (Maxam and Gilbert).	16
7.3.3	Método de secuenciación Sanger.	16
7.3.4	Secuenciación de siguiente generación (NGS).	17
7.3.5	Ion Torrent o PGM.	18
7.3.5.1	Construcción de librerías.	18
7.3.5.2	Preparación del templado positivo y su enriquecimiento.	18
7.3.5.3	Secuenciación.	19
7.3.6	MiSeq.	19
7.3.6.1	Construcción de la librería y grupos de secuenciación.	19
7.3.6.2	Secuenciación.	20

7.3.7	Ion Torrent vs MiSeq.....	20
8.	MATERIAL Y MÉTODOS.....	22
8.1	Cultivo viral de HHV 6A cepa GS en células HSB-2.....	22
8.1.1	Cultivo de células HSB-2.....	22
8.1.2	Infección de las células HSB-2 con HHV 6A cepa GS.....	23
8.2	Cultivo viral de HCMV cepa Merlin.....	24
8.2.1	Cultivo de células adherentes HFFF-2.....	24
8.2.1.1	Tripsinización.....	24
8.2.2	Cultivo viral de HCMV cepa Merlin (infección de las células HFFF-2).....	25
8.3	Extracción del DNA genómico de los cultivos virales.....	25
8.4	Prueba de <i>Mycoplasma</i>	26
8.5	Muestras Clínicas.....	26
8.5.1	Cuantificación de la carga viral de los cultivos virales y de algunas muestras clínicas. 27	
8.6	Construcción de las librerías genómicas de secuenciación para la plataforma Ion Torrent. 28	
8.6.1	Fragmentación por sonicación.....	28
8.6.2	Fragmentación enzimática.....	29
8.7	Preparación del Templado Positivo.....	30
8.8	Enriquecimiento del Templado Positivo.....	30
8.9	Cargado del chip Ion 316 y secuenciación de los cultivos virales.....	31
8.10	Construcción de librerías y cargado de la celda para secuenciación de la plataforma MiSeq. 31	
8.10.1	Fragmentación y etiquetado del DNA.....	31
8.10.2	PCR de Amplificación y purificación de la PCR.....	31
8.10.3	Normalización y cargado de la celda de secuenciación.....	32
8.11	Análisis de los datos de secuenciación.....	33
8.11.1	Análisis de los datos de secuenciación del Ion Torrent.....	33
8.11.2	Análisis de los datos de secuenciación del MiSeq.....	34
8.11.3	Comparación y análisis de los polimorfismos de los cultivos virales respecto a las secuencias de referencia.....	34

8.12	Prevalencia de los herpesvirus HCMV (citomegalovirus) y HHV6 (herpesvirus humano 6) en la población mexicana de VIH.	38
9.	RESULTADOS.	39
9.1	Cultivo de HHV6A cepa GS.	39
9.1.1	Cosecha del cultivo de HHV6A cepa GS.	39
9.2	Cultivo de HCMV cepa Merlin.	40
9.2.1	Infección de las células HFFF-2 con la cepa Merlin de HCMV.	40
9.2.2	Cosecha de los cultivos virales de HCMV-Merlin.	40
9.3	Extracción del DNA genómico de los cultivos virales.	40
9.4	Cuantificación y diagnóstico de las muestras clínicas.	41
9.5	Prueba de Mycoplasma.	44
9.6	Cuantificación de la carga viral de los cultivos virales y de las muestras clínicas CAS y RCMV107s.	45
9.7	Construcción de librerías.	48
9.7.1	Fragmentación por Sonicación.	48
9.7.2	Fragmentación Enzimática.	50
9.8	Secuenciación de los cultivos virales.	53
9.8.1	Construcción de las librerías de secuenciación para Ion Torrent.	53
9.8.2	Secuenciación de los cultivos virales en el Ion Torrent.	54
9.8.3	Cobertura y profundidad de la secuenciación obtenida respecto al contenido de G+C.	61
9.8.4	Librerías de secuenciación para MiSeq.	63
9.8.5	Secuenciación de los cultivos virales en plataforma MiSeq.	64
9.9	Análisis de polimorfismos.	69
9.9.1	Polimorfismos encontrados en el cultivo de HCMV.	70
9.9.2	Polimorfismos encontrados en el cultivo de HHV6A cepa GS.	71
9.9.3	Confirmación de polimorfismos a través de secuenciación Sanger.	73
9.10	Prevalencia de HCMV y HHV6 en la población mexicana con VIH.	75
9.11	Secuenciación y análisis de muestras clínicas.	77
9.11.1	Muestras clínicas secuenciadas en el Ion Torrent.	77
9.11.2	Muestras clínicas secuenciadas en el MiSeq.	81

10.	DISCUSIÓN.....	84
10.1	Cultivos virales.	84
10.2	Fragmentación por sonicación vs. Fragmentación enzimática.	84
10.3	Secuenciación en plataformas Ion Torrent y MiSeq.	85
10.3.1	Profundidad y cobertura obtenida en la secuenciación de las regiones únicas de los cultivos virales respecto al contenido de G+C.	85
10.3.2	Polimorfismos de los cultivos virales.	87
10.4	Análisis de las muestras clínicas de personas con VIH.	89
10.4.1	Prevalencia de betaherpesvirus (HCMV, HHV6 y HHV7) en la población mexicana con VIH.	89
10.4.2	Secuenciación de las muestras clínicas.	90
11.	CONCLUSIONES.	91
12.	LITERATURA CITADA.....	93

1. ÍNDICE DE TABLAS.

TABLA 1. COMPOSICIÓN DEL MEDIO DE PROPAGACIÓN PARA EL CULTIVO DE CÉLULAS HSB-2	22
TABLA 2. COMPOSICIÓN DEL MEDIO DE PROPAGACIÓN PARA EL CULTIVO DE CÉLULAS HFFF-2.....	24
TABLA 3. REACCIÓN HERPLEX.	27
TABLA 4. REACCIÓN PARA DETERMINAR LA CARGA VIRAL EN EL MATERIAL BIOLÓGICO.....	28
TABLA 5. CONDICIONES DE FRAGMENTACIÓN USADAS EN EL COVARIS S2.	29
TABLA 6. PRIMERS DISEÑADOS PARA EL ANÁLISIS DE LOS POLIMORFISMOS DE LOS CULTIVOS VIRALES.	36
TABLA 7. REACCIÓN DE PCR PARA AMPLIFICAR LAS REGIONES DE INTERÉS.....	36
TABLA 8. REACCIÓN DE SECUENCIACIÓN.	37
TABLA 9. REACCIÓN HERPLEX PARA DETERMINAR LA PREVALENCIA DE LOS HERPESVIRUS HCMV Y HHV6.	38
TABLA 10. DENSIDAD Y VIABILIDAD DE LOS CULTIVOS DE HSB-2 INFECTADOS CON HHV6A CEPA GS (4 DPI).	39
TABLA 11. CANTIDAD Y PUREZA DEL DNA EXTRAÍDO DE LOS CULTIVOS VIRALES (DNA ELUIDO EN AGUA).	41
TABLA 12. CONCENTRACIÓN Y PUREZA DEL DNA DE LAS MUESTRAS CLÍNICAS (DIL. 1:5).....	41
TABLA 13. COMPARACIÓN DEL DIAGNÓSTICO PREVIO Y ACTUAL DE LAS MUESTRAS CLÍNICAS.....	43
TABLA 14. RESULTADO DEL HERPLEX DE LOS CULTIVOS VIRALES.....	44
TABLA 15. CARGA VIRAL DE LAS MUESTRAS CLÍNICAS Y LOS CULTIVOS VIRALES.....	48
TABLA 16. NUMERO DE LECTURAS OBTENIDAS DE LOS CULTIVOS VIRALES DE ACUERDO A LA PLATAFORMA UTILIZADA.	65
TABLA 17. COMPARACIÓN DE LOS POLIMORFISMOS ENCONTRADOS POR EL ION TORRENT Y MISEQ EN EL CULTIVO DE HCMV CEPA MERLÍN.	71
TABLA 18. COMPARACIÓN DE LOS POLIMORFISMOS ENCONTRADOS POR EL ION TORRENT Y MISEQ EN EL CULTIVO DE HHV6A CEPA GS.	72
TABLA 19: PREVALENCIA DEL DNA DE HCMV Y HHV6 EN MUESTRAS CLÍNICAS.....	77
TABLA 20. LECTURAS MAPEADAS EN LAS MUESTRAS CLÍNICAS SECUENCIADAS.	82
TABLA 21. LECTURAS DE LA MUESTRA CLÍNICA E0804 MAPEADAS CONTRA LA SECUENCIA DE REFERENCIA DE HHV6A CEPA GS.....	83
TABLA 22. LECTURAS DE LA MUESTRA CLÍNICA E0804 MAPEADAS CONTRA LA SECUENCIA DE REFERENCIA DE HCMV CEPA MERLIN.	83

2. ÍNDICE DE FIGURAS.

FIGURA 1. ALGORITMO EMPLEADO PARA EL ANÁLISIS DE POLIMORFISMOS.	35
FIGURA 2. CURVAS DE DISOCIACIÓN (HRM) DE LAS MUESTRAS CLÍNICAS (CAS, AHG, GGD, 108.2S Y JAGG) Y CULTIVOS VIRALES (HHV6A-A, HHV6A-B, HCMV-1, HCMV-2 Y HCMV-3).....	42
FIGURA 3. HRM DE LAS MUESTRAS CLÍNICAS LV001, RCMVOR Y RCMVs.....	43
FIGURA 4. PRUEBA DE MYCOPLASMA.	45
FIGURA 5. CURVA ESTÁNDAR PARA CUANTIFICAR HCMV.	46
FIGURA 6. DETERMINACIÓN DE LA CARGA VIRAL DE LOS CULTIVOS DE HCMV.	47
FIGURA 7. DETERMINACIÓN DE LA CARGA VIRAL EN LOS CULTIVOS DE HHV6A.....	47
FIGURA 8. ELECTROFEROGRAMA DEL CULTIVO DE HHV6A SONICADO (UTILIZANDO 80NG DE DNA).	49
FIGURA 9. ELECTROFEROGRAMA DE LA PURIFICACIÓN MANUAL (EMPLEANDO E-GELS) DEL SONICADO DEL CULTIVO DE HHV6A.	50
FIGURA 10. ELECTROFEROGRAMA DEL CULTIVO DE HHV6A FRAGMENTADO CON ENZIMAS (UTILIZANDO 80NG DE DNA).	51
FIGURA 11. ELECTROFORESIS DEL DNA GENÓMICO DE LOS CULTIVOS VIRALES.	52
FIGURA 12. ELECTROFEROGRAMA DE LA LIBRERÍA DE HHV6A (FRAGMENTACIÓN ENZIMÁTICA) A PARTIR DE 500NG DE DNA.	53
FIGURA 13. ELECTROFEROGRAMA DE LA LIBRERÍA DE HCMV A PARTIR DE 500NG DE DNA.	53
FIGURA 14. CARGADO DEL CHIP 316 (MAPA DE CALOR).	54
FIGURA 15. RESUMEN DE LA PRIMER CORRIDA EN IONTORRENT 316V2.	55
FIGURA 16. DISTRIBUCIÓN DEL TAMAÑO DE LAS LECTURAS.	56
FIGURA 17. ALINEACIÓN DE LAS LECTURAS A LAS SECUENCIAS DE REFERENCIA Y EXACTITUD DEL LLAMADO DE LAS BASES.	57
FIGURA 18. CALIDAD PHRED DE LAS BASES Y LECTURAS ALINEADAS.....	58
FIGURA 19. CALIDAD DE LAS LECTURAS DE LA LIBRERÍA DE HHV6A HECHA CON 80NG.	58
FIGURA 20. CALIDADES DE LA LIBRERÍA DE HHV6A HECHA CON 500NG.	59
FIGURA 21. CALIDAD DE LA LIBRERÍA DE HCMV HECHA CON 500NG.....	59
FIGURA 22. SECUENCIACIÓN MASIVA DEL CULTIVO DE HCMV ALINEADO CON SU REFERENCIA.....	61
FIGURA 23. COMPARACIÓN DE LA COBERTURA Y PROFUNDIDAD OBTENIDA POR LA SECUENCIACIÓN, RESPECTO AL CONTENIDO DE G+C DEL GENOMA DE REFERENCIA (ION TORRENT).....	62
FIGURA 24. COBERTURA Y PROFUNDIDAD OBTENIDA EN LA SECUENCIACIÓN DE LOS CULTIVOS DE HHV6A, RESPECTO AL CONTENIDO DE G+C DEL GENOMA DE REFERENCIA (ION TORRENT).	63
FIGURA 25. ELECTROFEROGRAMA DE UN CHIP DNA HIGH SENSITIVITY DE LAS LIBRERÍAS POST-LIMPIEZA DE PRODUCTOS DE PCR.....	64
FIGURA 26. CALIDAD DE LAS LECTURAS R1 DE LA LIBRERÍA DE HHV6A CEPA GS (LECTURAS FORWARD).	66
FIGURA 27. CALIDAD DE LAS LECTURAS R2 DE LA LIBRERÍA HHV6A CEPA GS (LECTURAS REVERSE).....	66
FIGURA 28. CALIDAD DE LAS LECTURAS R1 DE LA LIBRERÍA DE HCMV CEPA MERLIN (LECTURAS FORWARD).....	67
FIGURA 29. CALIDAD DE LAS LECTURAS R2 DE LA LIBRERÍA DE HCMV CEPA MERLIN (LECTURAS FORWARD).....	67

FIGURA 30. COBERTURA Y PROFUNDIDAD DEL CULTIVO DE HCMV, RESPECTO AL CONTENIDO DE G+C (MISEQ). ..	68
FIGURA 31. COBERTURA Y PROFUNDIDAD DEL CULTIVO DE HHV6A, RESPECTO AL CONTENIDO DE G+C (MISEQ)..	69
FIGURA 32. EJEMPLO DE POLIMORFISMO EN ARTEMIS.	70
FIGURA 33. EJEMPLO DE POLIMORFISMO EN IGV.....	70
FIGURA 34. POLIMORFISMOS SECUENCIADOS VÍA SANGER.	74
FIGURA 35. QPCR DE LOS BETAHERPESVIRUS PRESENTES EN LAS MUESTRAS CLÍNICAS DEL PROTOCOLO E10-10. ...	76
FIGURA 36. HRM PARA LA IDENTIFICACIÓN DE LOS VIRUS PRESENTES.....	76
FIGURA 37. MAPA DE CALOR DE CHIP 316, CON LAS MUESTRAS CLÍNICAS CAS Y RCMV017s.....	78
FIGURA 38. RESUMEN DE LAS PARTÍCULAS CARGADAS.	78
FIGURA 39. PERFIL DE CALIDAD DE LAS LECTURAS DE LA LIBRERÍA DE CAS.	79
FIGURA 40. ALINEACIÓN DE LAS LECTURAS DE CAS Y RCMV107S CONTRA LA SECUENCIA DE REFERENCIA DE HCMV.	79
FIGURA 41. LECTURAS DE LAS MUESTRAS CLÍNICAS CAS Y RCMV107S ALINEADAS A LA SECUENCIA DE REFERENCIA.	81

3. RESUMEN.

El citomegalovirus humano (HCMV) y el herpesvirus humano 6 (HHV6) son herpesvirus cosmopolitas cuyas seroprevalencias van del 70%-90% en países desarrollados y del 90-99% en países subdesarrollados. El genoma de HCMV está formado por 2 regiones únicas (una corta [US] y una larga [UL] respectivamente) unidas covalentemente, cada región se encuentra flanqueada por sus inversiones repetitivas respectivas (TRL, IRL, TRS e IRs), HCMV presenta un contenido de G+C aproximado del 57%. El genoma de HHV6 está conformado por una región única (U), flanqueada por repeticiones invertidas (DRL y DRR), con un contenido de G+C del 43% aproximadamente.

Hoy en día, las plataformas de secuenciación de siguiente generación (NGS) han revolucionado los campos de investigación, diagnóstico clínico, entre otros. Sin embargo estas plataformas presentan algunos problemas para secuenciar regiones con un bajo contenido de G+C (Quail et al., 2012), entre las que se incluyen las plataformas MiSeq (Illumina), pero sobre todo el Ion Torrent (Life Technologies). En este estudio se retó la capacidad de las plataformas NGS Ion Torrent y MiSeq, para secuenciar las regiones únicas de los genomas de HCMV y HHV6A a partir del cultivo viral de las cepas de referencia Merlin y GS respectivamente. Posteriormente se compararon las secuencias obtenidas con las referenciadas mediante un análisis informático, con el fin de analizar y comparar la capacidad de secuenciación de ambas plataformas.

Aunado a esto se intentó secuenciar directamente los genomas de HCMV y HHV6 presentes en muestras clínicas de pacientes con VIH, con el propósito de identificar variables regionales respecto a las secuencias referenciadas, simultáneamente se determinó la prevalencia de estos herpesvirus en la población mexicana con VIH, observando una prevalencia de HCMV y HHV6 del 99% y 71% respectivamente.

4. ABSTRACT.

The human cytomegalovirus (HCMV) and human herpesvirus 6 (HHV6) are cosmopolitan herpesvirus whose seroprevalence go from 70%-90% in developed countries and 90-99% in developing countries. The genome of HCMV consists in two unique regions (a short [US] and a long [UL] respectively) covalently attached, each region is flanked by their respective repetitive investment (TRL, IRL, TRS and IRs), HCMV has a G + C content of 57% approximately. HHV6 genome consists in a single region (U), flanked by inverted repeats (DRL and DRR), with a G + C of 43% approximately.

Nowadays, the next generation sequencing platforms (NGS) have revolutionized the fields of research, clinical diagnosis, among others. However these platforms have some problems to sequencing regions with a low content of G + C (Quail et al., 2012), including MiSeq platform (Illumina), but especially the Ion Torrent (Life Technologies). In this study the capacity of NGS platforms (Ion Torrent and MiSeq) were challenged, to sequence the unique regions of the genomes of HCMV and HHV6A from viral culture of the reference strains Merlin and GS respectively. Subsequently the obtained sequences were compared with the sequences referenced by a computer analysis in order to analyze and compare the ability of both platforms.

Added to this the directly sequencing genomes of HCMV and HHV-6 present in clinical samples from patients with HIV was attempted, with the aim of identifying regional variables regarding the referenced sequences, simultaneously the prevalence of these herpesviruses were determined in the Mexican population with HIV, observing a HHV6 and HCMV prevalence of 99% and 71% respectively.

5. INTRODUCCIÓN.

Actualmente se reconocen 9 herpesvirus humanos de importancia clínica, estos se encuentran clasificados dentro de la familia *Herpesviridae*, la cual se divide a su vez en 3 subfamilias: en la *Alphaherpesvirinae* encontramos a los virus herpes simplex 1 (HSV1), herpes simplex 2 (HSV2) y varicela zóster (VZV); la subfamilia *Betaherpesvirinae* incluye al citomegalovirus humano (HCMV), a los herpesvirus humanos 6, especies A y B (HHV6-A Y HHV6-B respectivamente) y al herpesvirus humano 7 (HHV7); la subfamilia *Gammaherpervirinae* está compuesta por el virus de Epstein-Barr (HHV4), y el herpesvirus humano 8 (HHV8). Los herpesvirus se caracterizan por producir infecciones que alternan entre la latencia y la recurrencia, causando una amplia variedad de cuadros clínicos.

5.1 Citomegalovirus Humano (HCMV).

HCMV se aisló por primera vez a mediados de la década de los 50s (1955-1956) por tres grupos independientes a partir de muestras de pacientes con citomegalia. Primeramente por el grupo del Dra. Margaret G. Smith de San Luis, Missouri, EUA; posteriormente fue aislado por otros dos grupos, el del Dr. Thomas H. Weller en Boston, Massachusetts quienes aislaron la cepa Davis y el grupo del Dr. Rowe en conjunto con el grupo del Dr. Weller en Bethesda, Maryland aislando la cepa AD169 (Weller, 2000).

HCMV es el herpesvirus con el genoma más grande de los 9 herpesvirus de importancia clínica para los humanos. Su genoma está conformado por DNA lineal de doble hebra que mide 235 kpb aproximadamente y codifica para 230 proteínas aproximadamente (Crumpacker, 2010a). Este está formado por 2 regiones únicas: una larga y una corta (U_L y U_S respectivamente por sus siglas en inglés), cada una flanqueada por inversiones repetitivas (TR_L , IR_L , TR_S e IR_S) respectivamente, que se caracterizan por ser ricas en A+T (Davison et al., 2003).

5.1.1 Epidemiología.

HCMV se considera como un virus cosmopolita, ya que su prevalencia en la población mundial es muy alta: en los países desarrollados alcanza una seroprevalencia de 70%-90%,

mientras que en países subdesarrollados puede alcanzar una prevalencia del 90-100% (Gutierrez and Cruz, 2008); recientemente se ha estimado la prevalencia de HCMV en la población mexicana, estimándose una prevalencia del 85%, siendo que el 58% pertenecen a mujeres en edad reproductiva (Salud, 2012). También se ha observado que las poblaciones que presentan una menor tasa de hacinamiento y mejor acceso a servicios en general presentan una incidencia menor por HCMV, contrario a las que poseen un grado de hacinamiento alto y tienen un menor número de servicios, con lo que se observa cierta correlación entre el nivel socio-económico de las poblaciones y la prevalencia de HCMV.

HCMV es capaz de infectar un amplio rango de células humanas dentro de las que sobresalen las células epiteliales, células endoteliales, células neuronales, células del músculo liso, así como fibroblastos, monocitos y macrófagos.

La habilidad de HCMV para mantenerse latente después de la infección es lo que impide eliminar esta infección. En la mayoría de los casos, las primoinfecciones por HCMV son asintomáticas y solo un pequeño porcentaje de las personas infectadas (el 10% aproximadamente) presentarán síntomas, entre los que se incluye fiebre prolongada, cansancio, esplenomegalia, hepatomegalia, infección aguda de vías respiratorias y urinarias, linfadenopatía, linfocitos atípicos (monocitosis), neutropenia leve entre otros. HCMV se transmite fácilmente a través del contacto directo con diversos fluidos corporales de personas infectadas, tales como saliva; a través del contacto sexual o por medio de transfusiones de sangre y trasplantes de órganos (Chen et al., 2015).

Aunque la mayoría de las infecciones por HCMV pasan desapercibidas, estas constituyen un riesgo elevado en las personas trasplantadas, ya que el receptor del órgano queda expuesto a una infección severa de HCMV, debido a que los tratamientos de inmunosupresión a los que es sometido para evitar el rechazo del órgano juegan un rol importante en la reactivación de HCMV. También se ha observado que la infección por HCMV es significativamente más severa en individuos que son seronegativos antes del trasplante y que adquieren la infección del órgano trasplantado (Crumpacker, 2010b).

Entre las complicaciones más graves por una infección o reinfección con HCMV que se pueden presentar después de un trasplante se encuentran: la hepatomegalia y esplenomegalia, así como la hepatitis y neumonía que ponen en riesgo la vida del individuo, HCMV también puede propagarse al sistema nervioso central induciendo meningoencefalitis.

HCMV también tiene un rol importante como un factor potencial para el desarrollo de enfermedades cardiovasculares como miocarditis, adjudicándosele participación en infartos al miocardio debido al incremento de citosinas pro-inflamatorias como el factor de necrosis tumoral α (TNF- α) e interferón γ (INF- γ) así como interleucina 10 (IL-10) para contrarrestar la infección, lo que produce lesiones en las paredes de los vasos sanguíneos (Roubille et al., 2010).

HCMV tiene la capacidad de traspasar la barrera placentaria, lo que le da la capacidad de producir infecciones congénitas; la infección intrauterina por HCMV ocurre del 0.5% al 22%, la infección congénita en la mayoría de los casos es asintomática y es menos frecuente que la perinatal (Crumpacker, 2010a). Las infecciones congénitas sintomáticas ocurren con mayor frecuencia cuando las mujeres embarazadas adquieren una infección primaria durante su embarazo. Durante el primer trimestre la infección por HCMV puede ser mortal para el feto ya que este no llega a término. Las infecciones posteriores al primer trimestre de embarazo también resultan de un riesgo muy elevado para los neonatos, ya que estos pueden desarrollar una inclusión citomegálica fulminante, que consiste en hepatosplenomegalia, petequias, y fallas en múltiples órganos. Durante la infección congénita HCMV afecta seriamente el sistema nervioso central (CNS) produciendo microcefalia, distrofia muscular, coriorretinitis, e incluso calcificaciones cerebrales. Al nacer o poco tiempo después los infantes desarrollan letargo, dificultades respiratorias y convulsiones, complicaciones que pueden finalmente acabar con la vida de los infantes en cuestión de días o algunas semanas. Algunos síntomas que presentaron infantes que sobreviven a una infección congénita pueden cesar como la ictericia y

hepatosplenomegalia, no obstante, las secuelas neurológicas como la microcefalia, el retraso mental y la pérdida de la audición persistirán (Buonsenso et al., 2012).

Las infecciones perinatales de los infantes, más frecuentes que las congénitas, ocurren cuando estos entran en contacto con secreciones infectadas como las del tracto uterino o a través de la leche materna (Crumpacker, 2010a).

HCMV es un serio problema para las personas inmunosuprimidas pero sobre todo para los individuos con VIH ya que constituye una importante causa de morbilidad en estos, propiciando una retinitis severa (Bowen et al., 1997). Entre otras complicaciones que HCMV puede producir en individuos inmunocomprometidos se encuentran episodios de colitis o neumonía severa, gastritis, radiculopatía y esofagitis. HCMV tiene la capacidad de infectar el SNC produciendo meningitis, encefalitis, demencia, fiebre, fotofobia, letargia, problemas de memoria y en casos más severos el daño producido en el sistema nervioso es fatal. De forma similar la infección prolongada de HCMV puede llegar a afectar múltiples órganos incrementando el riesgo de muerte de los afectados (Crumpacker, 2010b; Goodman et al., 2015).

5.2 Herpesvirus humano 6 (HHV6).

HHV6 antes conocido como virus humano linfotrópico de células B (HBLV), fue aislado por primera vez en 1986 de pacientes con desórdenes linfoproliferativos (Salahuddin et al., 1986). Actualmente se reconocen 2 especies de HHV6, A y B (Adams and Carstens, 2012), ya que aunque estos virus son relativamente cercanos y comparten el 90% de su secuencia de nucleótidos, presentan patologías características, distintos tropismos celulares, y además reaccionan ante paneles de anticuerpos diferentes (Adams and Carstens, 2012; Dominguez et al., 1999).

Estos virus tienen una organización genética similar con otros betaherpesvirus como HHV7 (Dominguez et al., 1999; Gompels et al., 1995). Su genoma está formado por una región única (U) rica en A+T (40.8% de G+C) y está flanqueada por fragmentos repetitivos (DR_L y DR_R) ricos en G+C (riqueza de 59.1%) lo que le da un 43% de G+C en promedio. Los

tamaños genómicos de HHV6-A y HHV6-B van de 159 a 162 kpb respectivamente (Dominguez et al., 1999; Gompels et al., 1995).

5.2.1 Epidemiología.

HHV6 al igual que HCMV produce infecciones latentes, aunque existen pocos estudios de la prevalencia, se ha llegado a estimar que un 90% de la población mundial posee anticuerpos contra HHV6 (Pellett et al., 2012).

HHV6-A es detectado más frecuentemente en líquido cefalorraquídeo que HHV6-B lo que sugiere un mayor grado de neurotropismo, sin embargo HHV6-B también tiene la capacidad de establecer latencia en el sistema nervioso central (SNC), por lo que actualmente existen múltiples reportes que sugirieren una relación entre HHV6 y meningoencefalitis (MS). Por otra parte, HHV6-B se encuentra íntimamente relacionado con el exantema súbito o roséola infantum que afecta a niños de entre 2 meses y 4 años de edad (Yamanishi, 2000).

La incidencia por infección de HHV6 alcanza su pico entre los infantes de 6 a 9 meses de edad, indicando una elevada seroprevalencia de la infección en adultos. Los virus de HHV6 al igual que HCMV se consideran cosmopolitas. No obstante, la mayoría de las infecciones clínicas en personas inmunocompetentes parecen asociarse a infecciones por HHV6-B, mientras que HHV6-A afecta principalmente a personas inmunocomprometidas. HHV6 se transmite principalmente entre los infantes a través de la saliva, aunque también existe el riesgo del contagio prenatal de madre a hijo. La roséola o exantema súbito es uno de los síntomas que se presentan con mayor frecuencia, además de picos febriles prolongados, erupciones en la piel (rash), e inclusive linfocitosis. Entre las principales complicaciones que puede presentar la infección están encefalopatía, hepatitis y artritis. HHV6 puede reactivarse durante el embarazo o durante infecciones o enfermedades críticas. Las mayores complicaciones por HHV6 se dan por infecciones del SNC en niños menores de 2 años, que presentan convulsiones febriles o ataques, a pesar de ello el pronóstico

generalmente es bueno, sin embargo meningitis, encefalitis y déficits en el SNC también han sido asociados a infecciones por HHV6 (Yamanishi, 2000).

La reactivación de HHV6 en personas con trasplantes de órganos y otras condiciones inmunosupresoras, como VIH, se asocia más frecuentemente a enfermedades como neumonías y encefalitis, aunque también puede afectar a diversos tipos de tejidos y órganos como pulmones, bazo, hígado, riñones y nódulos linfáticos: contribuyendo con el rol de inmunodeficiencia e incrementando el riesgo de mortalidad en personas coinfectadas por HHV6 y HIV. La relación de HHV6 y VIH ha sido muy discutida, ya que HHV6 incrementa la expresión del receptor CD4 en los linfocitos T, y aumenta el rango de células que expresan el receptor CD4 incluyendo a los linfocitos T CD8, los linfocitos T $\gamma\delta$ y las células natural killer (NK), lo cual puede potencialmente incrementar el rango de células susceptibles a la infección por VIH (Dockrell, 2003).

Entre las secuelas clínicas indirectas producidas por la infección de HHV6 pos-trasplante se incluyen alteraciones en el estado neto de inmunosupresión. Las infecciones por HHV6 en personas con trasplantes de órganos pueden presentar síntomas severos cuando se asocia a una infección conjunta con HCMV, en particular, este último representa un factor de riesgo más elevado (Dockrell, 2003).

En México existen escasos estudios epidemiológicos y de prevalencia sobre HHV6 a pesar de tratarse de un virus ampliamente distribuido dentro de la población mundial en general, por lo que se deberían implementarse programas que aborden su estudio para obtener una mayor información y comprensión sobre estos patógenos que facilite su prevención y tratamiento.

6. OBJETIVOS.

- Secuenciar los segmentos únicos de citomegalovirus humano (HCMV) y herpesvirus humano (HHV6) a partir de cultivos virales (de las cepas de referencia HCMV-Merlin y HHV6A-GS) y de muestras clínicas de pacientes con VIH.
- Analizar y evaluar la fidelidad, cobertura y profundidad obtenida por la secuenciación realizada por las plataformas NGS (Ion Torrent y Miseq) en relación con la oscilación de la cantidad de G+C en cepas de referencia de HCMV y HHV6.
- Analizar las diferencias entre las secuencias obtenidas de muestras clínicas y las secuencias de referencia (HCMV-Merlin y HHV6A-GS) a través del análisis bioinformático.

7. ANTECEDENTES.

7.1 Secuenciación de HCMV

HCMV (cepa AD169) fue secuenciado por primera vez por Chee et al. (1990), quienes realizaron la descripción más precisa del genoma hasta ese momento, señalando que había tenido problemas para secuenciar las regiones homopoliméricas y ricas en A+T; Chee previó que los resultados de secuenciación podrían y serían modificados con el tiempo.

Posteriormente se secuenciaron otras cepas de HCMV. En 1996 (Cha et al., 1996) publicó la secuencia de la cepa Toledo, encontrando 15 kpb extras en el extremo derecho de U_L; los autores también notaron que la cepa Towne tenía un menor número de deleciones que AD169 en la misma región (extremo derecho de U_L), lo que sugirió la capacidad de HCMV para adaptarse rápidamente a las células hospederas y que este perdía material genético durante los procesos extensivos de cultivo en líneas celulares (como era el caso

de AD169). Por otro lado, era bien sabido que la cepa Toledo también tenía numerosos pases de cultivo, aunque estos eran más limitados que los de AD169 (Murphy et al., 2003) por lo que en el 2004 se buscó secuenciar a HCMV presente en muestras clínicas, pero cultivado únicamente durante 3 pases en células HFFF2 (Fibroblastos fetales humanos primarios) (Dolan et al., 2004), para posteriormente secuenciar el genoma viral. Se obtuvo el genoma de la cepa Merlin de HCMV (el genoma más cercano al virus silvestre) cuyo tamaño genómico es de 235,645 pb. La cepa Merlin no presenta pérdida de secuencias en los extremos de UL, sin embargo presentaba una sustitución nucleotídica en el gen UL128 que introduce un codón de paro prematuro en el tercer exón y en consecuencia el gen se trunca (Dolan et al., 2004), aunque no se observaron otros cambios significativos.

HCMV fue re-secuenciado por Cunningham en 2010 empleando la plataforma NGS Illumina Genome Analyzer (IGA), a partir de cultivos virales con un bajo número de pases (menos de 5) con la finalidad de evitar procesos de presión de selección y las adaptaciones génicas que modifican el genoma viral, además también secuenciaron muestras clínicas directamente con la finalidad de evitar el cultivo viral, la clonación en plásmidos bacterianos o BACs y procesos de mutagénesis que pudieran alterar al genoma original (Cunningham et al., 2010). Al comparar los genomas de HCMV provenientes de las muestras contra genomas provenientes de cultivos de cepas como AD169 y Towne, entre estos sobresale que a pesar de que las muestras clínicas tenían menor carga viral de HCMV que las provenientes de cultivo, conformando menos del 3% del DNA total, fueron capaces de obtener una cobertura del 99.41-99.97% del genoma de HCMV a pesar de obtener una cantidad de lecturas inferior a las provenientes de los cultivos. Se demostró que AD169 además de carecer de grandes regiones genómicas, numerosos genes contenían deleciones (UL42 y UL43, RL5A, RL13 y UL131A) o sustituciones deletéreas (UL36). Aunado a esto (Dolan et al., 2004) comparó las secuencias de varias cepas sin pases (secuencias provenientes directamente de muestras clínicas) observando un rango de variación considerable, como el gen RL12 en el que sólo el 49% de la secuencia de amino ácidos era idéntica entre estas.

En contraste con las cepas sin pases de cultivos, las que tenían varios pases presentaban múltiples mutaciones en una o varias familias de genes que podrían modificar su tropismo celular, como la familia de genes de glicoproteínas RL11 (RL6, RL12, RL13, UL1, UL6, UL9 Y UL11) o los genes UL128, UL130 y UL131 que codifican para proteínas secretadas.

7.2 Secuenciación de HHV6

En 1995 se determinó la secuencia genómica de la cepa U1102 aislada en Uganda de una paciente inmunodeficiente (Gompels et al., 1995). Para la secuenciación del genoma de HHV6-A cepa U1102 se realizaron cultivos virales para incrementar la cantidad de DNA disponible, posteriormente se fragmentó y clonó el genoma de HHV6-A en plásmidos, cósmidos y bacteriófagos. Las secuencias clonadas se superponían a otros que correspondían a las secuencias adyacentes de un fragmento determinado y el DNA viral fue amplificado por PCR, procedimiento que podría haber modificado la secuencia de nucleótidos original. Se realizó una secuenciación adicionando dideoxinucleótidos marcados radiactivamente, de la cual se obtuvo una secuencia de 159 kpb, de la que resultó una estructura genómica compuesta por una región única rica en A+T con un G+C del 41% aproximadamente, flanqueada por repeticiones terminales (DR) ricas en C+G con una media de 58% (Gompels et al., 1995).

Por otro lado, HHV6-B cepa Z29, fue secuenciado en 1999 (Dominguez et al., 1999). Para realizar la construcción de las librerías genómicas, el genoma de HHV6-B se fragmentó en grandes segmentos por medio de endonucleasas, posteriormente estos segmentos fueron clonados para construir la librería genómica (al igual que con HHV6-A), aunque cabe mencionar que existieron varias regiones que no fue posible clonar. Los nucleótidos fueron determinados con una redundancia de 4 veces o más. Un punto interesante es que no se logró la cobertura de todo el genoma, debido a la presencia de elementos repetitivos y regiones homopoliméricas, que incluían 454 pb del segmento DR_L y 694 pb del segmento DR_R.

Al describir y comparar la organización genética de HHV6 A y B, se observó que aunque estos virus son relativamente cercanos, existe una diferencia cercana al 5% entre las secuencias de aminoácidos para las que codifican. Además no existe un gradiente genético entre HHV6 A y B, es decir, no se han detectado virus recombinantes entre estos virus. Domínguez también describe el contenido de G+C en HHV6-B, encontrando un contenido del 40.8% en la región única (U) y 59.1% de G+C en las regiones repetitivas (DR)(Dominguez et al., 1999), de forma similar a lo observado en HHV6-A (Gompels and Macaulay, 1995).

Por otro lado, Isegawa y colaboradores secuenciaron la cepa HST de HHV6-B aislada en Japón de un paciente con exantema súbito y la compararon con la secuencia de HHV6-A de Gompels, observando que HST es 2,244 pb más grande que U1102 y un contenido de GC similar en ambos (43%). En las regiones DR de ambos, los elementos básicos fueron repeticiones teloméricas humanas (TAACCC), una de estas regiones teloméricas (T1) consistía en 54 copias de elementos teloméricos básicos, sin embargo en HST esta región estaba compuesta por 26 copias de elementos teloméricos básicos y 28 copias relacionadas con elementos teloméricos humanos, mientras que en U1102 esta misma región (compuesta también por 54 copias) se componía solo por 14 copias de elementos teloméricos básicos y 40 copias relacionadas a elementos teloméricos humanos (Isegawa et al., 1999).

7.2.1 Integración cromosomal de HHV6.

Algunos tipos de virus necesitan integrar su genoma viral al genoma celular para poder llevar al cabo su replicación tal es el caso de los retrovirus y adenovirus asociados (AAV), contrario a estos, los herpesvirus mantienen sus genomas como episomas extracromosomales circulares, sin embargo existe evidencia de integración cromosomal del DNA de HHV6 al DNA celular, lo que parece ocurrir en aproximadamente el 1% de la población mundial (Dhepakson et al., 2002; Lower et al., 1996). A mediados de la década de los 90s el grupo de Luppi M. y colaboradores (Morissette and Flamand, 2010), aportaron el primer reporte demostrando la existencia de una posible integración parcial

o total del genoma de HHV6 en DNA aislado proveniente de células mononucleares en sangre.

Los individuos con integración cromosomal de HHV6 (CIHHV-6) se caracterizan por tener el genoma de HHV6 integrado en cada célula de su cuerpo, desde los folículos del cabello hasta en células mononucleares de sangre (Hall et al., 2008). Las consecuencias clínicas de poseer CIHHV-6 no se encuentran muy claras, por lo que las células, órganos y tejidos de estos individuos aún son considerados médicamente seguros para trasplantes (Potenza et al., 2009).

Aún se desconoce el mecanismo implicado en el proceso de integración cromosomal de HHV6, sin embargo, una de las características del genoma de HHV6 consiste en que su secuencia única se encuentra flanqueada por secuencias repetitivas de aproximadamente 9 kpb, que a su vez son flanqueadas por **pac1** y **pac2** que son secuencias aparentemente involucradas en el corte y empaquetamiento del genoma de HHV6. Adyacentes a **pac2**, se encuentra una serie de repeticiones nucleotídicas (TTAGGG) que son idénticas a la secuencia de repeticiones teloméricas (TRS) humanas (Gompels and Macaulay, 1995), mientras que las adyacentes a **pac1** son repeticiones teloméricas imperfectas a las que se les conoce como **het** (TAACCCn) por lo que presumiblemente podría existir una recombinación homóloga entre los TRS presentes en el genomas de HHV6 y los telómeros, quedando integrado el genoma de HHV6 en los cromosomas humanos. Hasta ahora todos los sitios de integración que han sido localizados se encuentran en regiones teloméricas.

Por otro lado, el hecho de poseer secuencias teloméricas en las extremidades de su genoma no parece ser el aspecto más importante o fundamental para que se lleve a cabo la integración genómica, un ejemplo de esto es HHV7 que es otro herpesvirus relativamente cercano a HHV6 y al igual que este posee secuencias teloméricas, sin embargo no existe algún reporte en el que HHV7 se encuentre integrado a genoma de la célula hospedera.

Por otra parte, HHV6 presenta un marco de lectura abierto (ORF) en la región U94, este ORF codifica para una proteína que presenta similitud (24% de los aminoácidos son idénticos) con las proteínas REP68/78 de los adenovirus asociados tipo 2 (AAV-2) lo que le confiere una característica que no se presenta en ningún otro herpesvirus humano (Dhepakson et al., 2002). Rep68 y Rep78 no son proteínas estructurales en los AAV-2, sin embargo son esenciales para la integración cromosómica. Estas proteínas desempeñan 3 funciones básicas en los AAV-2: poseen actividad endonucleasa, actividad ATPasa y helicasa, además de que llevan a cabo la unión del DNA exógeno al DNA hospedero. Se ha demostrado que la proteína expresada por U94 tiene actividad de unión al DNA, no obstante se desconoce si tiene alguna actividad de endonucleasa. Se ha reportado que la expresión de U94 inhibe la replicación viral por lo que se ha sugerido que podría estar implicada en el establecimiento de latencia (Dhepakson et al., 2002).

Los niños que son concebidos de gametos que poseen HHV6 integrado de forma cromosomal se caracterizan por tener al menos una copia del genoma de HHV6 en cada núcleo celular de su cuerpo. HHV6 puede transmitirse de padre a hijos lo que sugiere que este es capaz de integrarse a las células germinales en algún punto, dando origen a gameto infectados con HHV6 o bien, existe la posibilidad de que los óvulos y espermatozoides puedan infectarse de forma directa en los momentos previos a la fertilización (Morissette and Flamand, 2010).

7.2.1.1 Patologías asociadas a la integración cromosomal de HHV6.

No se han definido patologías asociadas a individuos con HHV6 integrado de forma cromosomal, no obstante existen algunas patologías que parecen sobre representadas en estos individuos como neoplasia hematológica y el síndrome de Miller-Dieker, este último se ha asociado a las regiones teloméricas en las que ocurren las inserciones de HHV6 (Morissette and Flamand, 2010). Aún existen varias preguntas a responder como si grandes inserciones de genomas en los telómeros celulares podrían afectar la integridad de estos e inducir senescencia prematura, si se desarrolla algún tipo de tolerancia

antigénica durante la embriogénesis o incluso si se podría desarrollar una reacción autoinmune.

7.3 Secuenciación masiva del DNA.

Se le conoce como secuenciación al conjunto de métodos y técnicas bioquímicas empleadas para determinar el orden de los nucleótidos que conforma el material genético de un organismo.

7.3.1 Antecedentes de la secuenciación masiva.

En un principio se pensó que la secuenciación de los ácidos nucleicos era mucho más difícil que la de las proteínas y la del RNA, por lo que hasta la década de los 60s el progreso para ello fue mínimo.

Los principales problemas que se presentaron en un inicio para la secuenciación del DNA fue la carencia de sustratos o sustancias puras para realizar la secuenciación, la falta de enzimas de restricción específicas equivalentes a la ribonucleasa T1 ya conocida para la época por su capacidad para fragmentar cadenas de RNA simple en los residuos de guanina en su extremo 3', pero principalmente por el gran tamaño que caracteriza a las moléculas de DNA.

La secuenciación del DNA tuvo su origen en la década de los 70s: en un principio sólo era posible determinar la secuencia de pequeños fragmentos de DNA de aproximadamente 50 nucleótidos. A partir de 1975, los métodos de secuenciación comenzaron a progresar aceleradamente con el descubrimiento de las endonucleasas de restricción y el desarrollo de técnicas de clonación del DNA, que permitieron el desarrollo de mejores técnicas de secuenciación del DNA.

En 1977 se reportan 2 métodos para la secuenciación del DNA de forma independiente: El primero fue el método de *degradación química* diseñado por Allan Maxam y Waler

Gilbert, y el método de secuenciación *Sanger* desarrollado por Frederick Sanger, ambas técnicas permitían determinar las secuencias de DNA mediante una autoradiografía.

7.3.2 Método de degradación química (Maxam and Gilbert).

En esta técnica una muestra de DNA era marcada en sus extremos 5' o 3' primeramente, luego se hacían 4 alícuotas que eran fragmentadas en reacciones químicas diferentes (corte de purinas, corte de pirimidinas, corte de citosinas y corte de adeninas. Finalmente las 4 alícuotas sometidas a distintos tratamientos eran corridas en una electroforesis (cada reacción en un carril distinto) que permitía deducir el nucleótido que se encontraba al final de cada corte, permitiendo conocer la secuencia de los fragmentos de DNA. Una de las grandes limitantes de este método fue la baja resolución obtenida por los geles de poliacrilamida, ya que este no proporcionaba una buena separación de las bandas de gran tamaño por lo que eran necesario correr geles muy delgados y de gran tamaño (Maxam and Gilbert, 1977).

7.3.3 Método de secuenciación Sanger.

Este método se basa en el uso de la DNA polimerasa, en el que partiendo de una hebra de DNA molde se sintetizaba la hebra complementaria. En esta técnica (al igual que en la degradación química) originalmente se realizaban 4 reacciones independientes en las que la hebra de DNA complementaria era sintetizada, a cada reacción se le agregaban los elementos necesarios para la síntesis de la hebra incluyendo los nucleótidos. Una variante importante en cada reacción era la adición de los dideoxinucleótidos (ddNTPs), estos nucleótidos carecen de un grupo hidroxilo en su extremo 3', lo que finalizaba la síntesis de la hebra complementaria, además estos ddNTPs eran marcados radiactivamente; a cada una de las 4 reacciones de síntesis se agregaba un solo tipo de ddNTP (ddATP, ddCTP, ddGTP o ddTTP) en bajas concentraciones junto con los nucleótidos normales. Finalmente las 4 reacciones eran corridas en una electroforesis (en carriles separados), que permitía observar fragmentos de distintos tamaños en las distintas reacciones, esto permitía deducir el nucleótido en el que finalizaba cada banda dependiendo del tipo de ddNTP

adicionado en el tratamiento observado, con lo que era posible determinar la secuencia de la cadena molde (Sanger et al., 1977).

Desde un principio la secuenciación Sanger fue más rápida que la secuenciación por fragmentación química, confiriéndole una gran ventaja sobre esta.

En un principio la secuenciación Sanger usaba ddNTPs marcados radiactivamente, lo que lo volvía un método riesgoso para los investigadores. Posteriormente con el desarrollo tecnológico, los marcadores radiactivos fueron sustituidos por fluoróforos que contribuyeron de forma relevante a la seguridad del método, y a que las 4 reacciones pudieran hacerse en un solo tubo aprovechando las diferentes longitudes de onda de emisión de cada uno. En los años posteriores el método Sanger se volvería la técnica de secuenciación más popular sobre todo gracias a la automatización y seguridad de la técnica.

7.3.4 Secuenciación de siguiente generación (NGS).

En los últimos años se han realizado grandes avances tecnológicos, un ejemplo de esto es la carrera por obtener la secuencia del genoma humano a un precio cada vez más bajo utilizando plataformas cada vez más eficientes.

Las plataformas 454 de Roche, Solexa de Illumina y Solid de Applied Biosystems fueron algunas de las primeras plataformas de secuenciación masiva en desarrollarse. Posteriormente se desarrollaron otras plataformas NGS que se basan en la lectura de las moléculas de DNA a través de su amplificación, como las desarrolladas por Illumina destacando el *Miseq*, el *HiSeq 2000/2500*, el *GA IIX* (Genome Analyzer IIX) y las diseñadas por Life Technologies, como el Ion Torrent y Ion proton. En cambio, existen otras plataformas también llamadas de tercera generación que involucran la lectura de moléculas individuales sin amplificación (PacBio de Pacific Biosciences, Helicos tSMS de Helicos Bioscience, GridION y MinION de Oxford Nanopore y 3G de ZS Genetics).

A continuación se profundizará en el mecanismo de secuenciación de las plataformas Ion Torrent y MiSeq que han sido utilizadas para el desarrollo de esta tesis.

7.3.5 Ion Torrent o PGM.

7.3.5.1 Construcción de librerías.

Esta secuenciación comienza con la fragmentación del DNA, sea a través de un método físico o enzimático. Estos fragmentos son flanqueados por una serie de adaptadores (fragmentos de nucleótidos cuya secuencia se conoce) que permiten la adhesión del fragmento de DNA a la perla o placa de secuenciación biotinilada. También se pueden adicionar “códigos de barra” (secuencia de nucleótidos conocidos que etiquetan a los fragmentos de DNA) específicos a una determinada librería, para su posterior identificación. Para la construcción de las librerías se puede utilizar el AB Library Builder™ System, que consiste en un sistema semi-automatizado que simplifica la fragmentación, construcción y purificación de librerías.

7.3.5.2 Preparación del templado positivo y su enriquecimiento.

El templado positivo consiste en la construcción de nano-esferas cubiertas por un solo tipo de templado (partículas monoclonales), para ello se prepara en una serie de nano-reactores (nano-esferas) a las que se une hipotéticamente un solo fragmento de DNA por medio de los adaptadores. Cada nano-esfera y su fragmento de DNA es inmerso en una solución de emulsión que contiene todos los elementos necesarios (incluyendo la DNA polimerasa) para llevar al cabo la de síntesis del DNA, y una vez que las partículas están inmersas en el medio de emulsión se lleva a cabo una la reacción de PCR para dar origen a la partículas monoclonales que conformaran los templados positivos.

Posteriormente se realiza el enriquecimiento del templado positivo, este consiste en eliminar las partículas libres (partículas sin templado) con el fin de evitar cargar el chip con partículas vacías. Por otro lado el enriquecimiento no nos permite eliminar las partículas con varios tipos de templado (partículas policlonales), estas serán eliminadas durante el filtrado de calidad de las lecturas obtenidas por la secuenciación.

7.3.5.3 Secuenciación.

Finalmente la emulsión enriquecida es introducida al chip de secuenciación, que contiene millones de pozos dentro de los que deberá caer un reactor de secuenciación por pozo.

Este método de secuenciación se basa en la detección de los protones (iones H^+), que son liberados por la unión de los nucleótidos a la hebra complementaria, produciendo un cambio de pH en el pozo en el que se lleva a cabo la reacción. Cada nucleótido es agregado de forma específica e independiente por el equipo, de esta forma al detectarse una variación en el pH es posible determinar el nucleótido que acaba de incorporarse, y así conocer la secuencia de nucleótidos de los fragmentos de DNA.

Al finalizar la secuenciación, los cambios en el voltaje son traducidos a lecturas de secuencia (*reads*) con los cuales se realiza el ensamblado. El ensamblaje en general consiste en superponer las secuencias de las lecturas que posean extremos iguales, por los que puedan unirse y obtener una lectura más grande hasta a completar la secuencia.

Existen 2 tipos de ensamblados: el de novo y el de referencia. El *ensamble de novo* consiste en reconstruir la secuencia de DNA sin tener ningún conocimiento previo, este busca lecturas con extremos iguales que se puedan unir. Mientras que en el *ensamble con referencia* utiliza un genoma previamente secuenciado para colocar cada lectura en la posición adecuada, utilizando la referencia como guía.

7.3.6 MiSeq.

7.3.6.1 Construcción de la librería y grupos de secuenciación.

De forma similar al Ion Torrent, las librerías del MiSeq se construyen con los fragmentos del DNA de interés. A estas librerías se agregan 2 adaptadores diferentes que flanquean uno de los extremos del cada fragmento y códigos de barras o índices específicos para cada librería. Posteriormente las librerías son colocadas en un chip dentro del cual se encuentra una placa con adaptadores adheridos a la superficie, estos serán complementarios a uno de los adaptadores de las librerías permitiendo la unión de las librerías a la placa de secuenciación, dejando uno de los extremos libre.

En seguida se realiza una PCR puente, que consiste en la formación de un puente de amplificación debido a la complementariedad entre los adaptadores de la placa y los extremos que se encuentran libres. Esto genera grupos de secuencias (cluster) a partir un solo fragmento, aislados espacialmente de otros. Finalmente los extremos 3' de los fragmentos son bloqueados para evitar que se adhieran fragmentos inespecíficos u otras hebras complementarias.

7.3.6.2 *Secuenciación.*

El MiSeq se basa en la secuenciación por terminadores reversibles. Durante la reacción de secuenciación los 4 nucleótidos (A, C, G y T) son agregados de forma simultánea, cada nucleótido posee un residuo (terminador) que interrumpe temporalmente la síntesis de la hebra complementaria, cada residuo se encuentra unido a un determinado fluoróforo dependiendo del nucleótido del que se trate. La DNA polimerasa tiene la capacidad de escindir los terminadores reiniciando la síntesis de la hebra, cuando los terminadores son separados de los nucleótidos los fluoróforos fluorescen y su luminiscencia es captada por sensores ópticos de la plataforma identificando el nucleótido que fue insertado en la hebra complementaria. Este procedimiento se repite durante varios ciclos hasta finalizar los ciclos de secuenciación. Posteriormente las lecturas son ensambladas para determinar la secuencia génica de la o las muestras de interés.

7.3.7 **Ion Torrent vs MiSeq.**

Aunque los requerimientos necesarios para la preparación de las librerías son similares en ambas plataformas, pueden identificarse diferencias significativas en los resultados reportados. En el 2012 (Quail et al., 2012) compararon las características de varias plataformas NGS (incluyendo el MiSeq y Ion Torrent) al secuenciar los genomas de *B. pertussis*, *S. aureus* y *P. falciparum*, una de las características comparadas fue la velocidad y tasa de error de secuenciación, aunque la velocidad de secuenciación del Ion Torrent es mayor a la del MiSeq, la tasa de error en este último (0.80%) fue mucho menor a la generada por el Ion Torrent (1.71%). Otro de los problemas fueron los errores producidos en las regiones repetitivas y de homopolímeros (2 o más bases representadas por un

mismo nucleótido), que fueron mayores en el Ion Torrent ya que las lecturas proporcionadas por este pueden ser confusas o inconclusas debido a que se observa una señal (un pico en el cambio del pH) por la sumatoria de los protones liberados por la adición de los nucleótidos, lo que vuelve complicado determinar el número de nucleótidos presentes en dicha región.

Sin embargo, quizás el principal problema al que se enfrentan las NGS pueda ser la distribución de G y C en los diversos genomas. Al secuenciar el genoma de *P. falciparum* (contenido de G+C de 19.3% y regiones con un contenido de 0% de G+C) (Quail et al., 2012) observaron una caída drástica en la profundidad y cobertura de las lecturas obtenidas por el Ion Torrent (sobre todo en regiones pobres en G+C), mientras que las lecturas del MiSeq caen ligeramente en las mismas regiones, sobrellevando mucho mejor los problemas de secuenciación en las regiones con bajo contenido de G+C.

Dado que existe escasa información sobre los genomas de HCMV y HHV6 en México, en este estudio nos propusimos secuenciar a los virus circulantes en pacientes VIH positivos, aunque primero se comparó el rendimiento de 2 plataformas NGS para secuenciar los genomas de HCMV y HHV6-A, que por sus características ya antes mencionadas podrían representar un reto para ambas plataformas.

8. MATERIAL Y MÉTODOS.

8.1 Cultivo viral de HHV 6A cepa GS en células HSB-2.

8.1.1 Cultivo de células HSB-2.

El cultivo de las células HSB-2 fué obtenido a través del NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: HSB-2 from Electro-Nucleonics, Inc.(Salahuddin et al., 1986), que son células linfoblastoides humanas, se realizó un medio de propagación libre de antibiótico compuesto de la siguiente forma:

Tabla 1. Composición del medio de propagación para el cultivo de células HSB-2

Reactivo.	Concentración Final.
RPMI 1640 (cat.12-167F, Lonza)	90%
Suero bovino fetal (Lonza)	10%
Amino ácidos no esenciales, 10mM [100x] (cat. 13-114E, Lonza)	1%
L-glutamina [200mM] (cat. 17-605E, Lonza)	2mM
Hepes [1M] (cat. 17-737F, Lonza)	10mM
Piruvato [100mM] (cat. 13-115E, Lonza)	1mM

Inicialmente se prepararon frascos T-25 con 8mL de medio de propagación calentado a 37°C en un baño de agua. Las células HSB-2 fueron descongeladas en un baño de agua a 37°C y se inoculó un vial de células HSB-2 por cada frasco T-25. Se dejó sin inocular un frasco T-25 y se utilizó como frasco testigo (control negativo), finalmente los frascos fueron puestos a incubar toda la noche a 37°C en una incubadora con CO₂ (de Thermo Scientific). Para eliminar el dimetilsulfóxido (DMSO) utilizado como criopreservador, se cambió el medio de cultivo al día siguiente de la descongelación.

Diariamente se calculó la densidad y viabilidad de los frascos por medio de cámaras Glasstic slide 10 with grids (Kova International) utilizando 4 cuadrantes por cámara, y para calcular la viabilidad se utilizó azul de bromofenol con la finalidad de diferenciar células vivas y muertas.

Los frascos T-25 fueron expandidos a frascos T-75 cuando estos alcanzaron una densidad de 1-2 millones de células por mililitro; cada frasco T-75 se inoculó con 2 millones de células y un volumen final de 20mL de medio de propagación. Finalmente los frascos T-75 fueron expandidos a frascos T-175 cuando estos alcanzaron una densidad celular de 1.5-2 millones de células/mL.

8.1.2 Infección de las células HSB-2 con HHV 6A cepa GS.

Frascos T-175 con una concentración de 1 millón de células HSB-2/mL fueron infectados con células HSB-2 previamente infectadas con HHV6A cepa GS este reactivo se obtuvo a través del NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: HSB-2/HHV-6 from Dr. Robert Gallo (Ablashi et al., 1988; Ablashi et al., 1987; Lusso et al., 1988; Salahuddin et al., 1986). Las células infectadas sólo tenían un pase previo, ocurrido cuando el material llegó al laboratorio.

El vial de células infectadas se descongeló a 37°C en un baño de agua, y cada frasco T-175 se inoculó con 0.5mL de células infectadas (cabe señalar que no se encontraba señalada la concentración celular de los viales de las células HSB-2 infectadas). Adicional a esto se inoculó un frasco T-75 únicamente con células infectadas a manera de control positivo, de forma similar se colocaron células no infectadas en otro frasco T-75 a manera de control negativo. Cuando la viabilidad celular bajó hasta el 70-75% se procedió a congelar las células en 95% del medio de propagación y 5% de DMSO. Las células infectadas fueron colocadas en viales criogénicos, preparando un stock de congelación en nitrógeno líquido con el medio de congelación y un stock de células a secuenciar utilizando medio de propagación nuevo.

8.2 Cultivo viral de HCMV cepa Merlin.

8.2.1 Cultivo de células adherentes HFFF-2.

El medio de propagación empleado para cultivar los fibroblastos HFFF-2 (human fetal foreskin fibroblast, amablemente donados por Derrick Dargan, MRC Virology Unit, Reino Unido) se compuso de acuerdo a la siguiente tabla (Tabla 2).

Tabla 2. Composición del medio de propagación para el cultivo de células HFFF-2

Reactivo.	Concentración final.
DMEM, 4500mg de glucosa. (cat. 12-604F, Lonza)	90%
Suero bovino fetal. (Lonza)	10%
Amino ácidos no esenciales, 10mM [100x] (cat. 13-114E, Lonza)	1%
L-glutamina [200mM] (cat. 17-605E, Lonza)	4mM

Se preparó medio de propagación (libre de antibióticos) de acuerdo a lo indicado en la Tabla 2 y se llevó a 37°C en un baño de agua. Cada frasco T-75 se inoculó con 2×10^6 células HFFF-2/mL previamente calentadas a 37°C en un baño de agua, e incubados a 37°C y 5% de CO₂ (Thermo Fisher Scientific).

Con la finalidad de eliminar el DMSO utilizado para congelar las células HFFF2 inoculadas, el día posterior a la inoculación se cambió el medio de propagación cuidando de no dañar la monocapa de células. La confluencia y viabilidad de los cultivos se monitorearon diariamente, cuando las caras de adhesión de los frascos T-75 alcanzaron una confluencia del 95-100%, se procedió a propagar a frascos T-175.

8.2.1.1 Tripsinización.

Con el fin de propagar los cultivos celulares era necesario desprender la monocapa celular de la cara adherente, para ello la monocapa era primeramente lavada con Versene (cat. 17-711E, Lonza) para remover proteínas y medio de propagación que pudiera interferir con la tripsina. La monocapa de células era incubada brevemente con 2 mL de tripsina-EDTA (cat. CC-05012, Lonza) diluida 1:10 en Versene (0.25 g/100mL). Las células eran desprendidas mediante un golpe firme en uno de los costados del frasco. Finalmente la

tripsina era diluida con medio de propagación en proporción 10:1 con el fin de neutralizarla.

8.2.2 Cultivo viral de HCMV cepa Merlin (infección de las células HFFF-2).

Cuando se observó una confluencia del 80-90% en los frascos T-175 se procedió a realizar la infección de las células, para ello se tomaron viales criogénicos de 1mL, cada vial contenía 15 millones de células HFFF-2 previamente infectadas con HCMV cepa Merlin, los viales fueron llevados a 37°C en un baño de agua y posteriormente se inoculó un vial en cada frasco T-175.

Para cosechar los cultivos de HCMV, la monocapa de células infectadas fue desprendida cuidadosamente con tripsina y simultáneamente se preparó el medio de congelación que consistió en 95% del medio de propagación y 5% de DMSO. Para preparar el stock de congelación en nitrógeno líquido se colocó el medio de congelación en relación a la concentración de células deseada y posteriormente las células fueron colocadas en nitrógeno líquido. Por otro lado las células para el stock de secuenciación se centrifugaron a 3000 rpm por 10 min y se resuspendieron en PBS de acuerdo a la concentración celular deseada y congeladas -20°C.

8.3 Extracción del DNA genómico de los cultivos virales.

Las plataformas de secuenciación masiva son equipos de alta sensibilidad que requieren un DNA con un alto grado de pureza, de lo contrario la construcción de las librerías de secuenciación o el proceso de secuenciación en sí podrían inhibirse. Para ello es importante verificar el grado de pureza del DNA empleado para secuenciar.

Se utilizó el kit QIAamp DNA Mini Kit de Qiagen y se siguió el protocolo de extracción para sangre y fluidos corporales para extraer el DNA genómico (DNAg) de 3 viales del stock de secuenciación de Merlin (marcados como HCMV: 1, 2, y 3) y 2 viales del stock de secuenciación de HHV6A-GS (marcados como HHV6A: A y B). El DNA fue finalmente eluído

en agua libre de nucleasas. Se tomaron alícuotas de todos los tubos extraídos y se diluyeron en proporción 1:10 (en agua), para su posterior cuantificación en el NanoDrop 1000 y en el Nanodrop 8000 (NanoDrop Technologies).

8.4 Prueba de *Mycoplasma*.

Los cultivos celulares son susceptibles a contaminarse con diversos microorganismos al ser ricos en nutrientes, favoreciendo el crecimiento de diversos microorganismos. La contaminación de los cultivos celulares con diferentes especies de *Mycoplasma* es frecuente, lo que puede alterar el crecimiento celular e inducir anomalías en los cultivos. Con la finalidad de constatar que los cultivos virales se encontraran libres de mycoplasma se empleó el Kit Universal Mycoplasma Detection (ATCC) de acuerdo al protocolo indicado por el fabricante.

8.5 Muestras Clínicas.

Con la finalidad de secuenciar los genomas de HCMV y HHV6 presentes en la población mexicana y determinar si existe alguna diferencia entre estos herpesvirus y los ya secuenciados, se tomaron 8 muestras clínicas (CAS, AHG, GGD, 108.2S, JAGG, LV001, RCMV107or y RCMV107s) que se fueron tomadas entre 2008 y 2011 para protocolos de investigación previos y se encontraban almacenadas a -80°C y que ya poseían un diagnóstico previo de HCMV y HHV6. Primero se determinó la concentración y pureza del DNA de las muestras con el NanoDrop 1000. Cabe señalar que estas habían sido eluidas en el buffer AE (del kit de extracción de DNA de Qiagen), que contiene EDTA (0.5mM) que podría inhibir la reacción de secuenciación, debido a esto las muestras fueron diluidas en agua en una proporción 1:5 como lo recomienda el protocolo Ion Xpress Plus Library (Life Technologies) para realizar la construcción de las librerías.

Para corroborar el estado del DNA, tanto las muestras clínicas como los cultivos se sometieron al ensayo **Herplex** (método de diagnóstico desarrollado en el laboratorio) con

el fin de verificar la presencia de los virus de interés debido al tiempo prologado de almacenamiento que presentaban. El Herplex es una técnica de diagnóstico de herpesvirus que emplea la PCR de tiempo real (qPCR) acoplada a curvas de disociación de alta resolución (HRM), esta última nos proporciona una curva de disociación específica para cada herpesvirus, permitiéndonos realizar el diagnóstico para la muestra en cuestión.

Para realizar la qPCR y el HRM se empleó la plataforma StepOne Real-Time PCR System (Life Technologies); el mix de PCR utilizado se constituyó de acuerdo a la tabla (Tabla 3), empleando los primers universales diseñados en el laboratorio:

Para los herpesvirus 1, 2 y 5. Forward: TCATCTACGGGGACACGGAC; Reverse: TCGCACCCAGATCCACG.

Para los herpesvirus 6 y 7. Forward: GATGGCGCATAATCTGTGT; Reverse: TCCAAGACACGTTACAGAAGC

Tabla 3. Reacción Herplex.

Mix de PCR	Concentración final.	Vol. por reacción (µL)
Qiagen Multiplex PCR Kit (Qiagen)		12.5
Primer mix F/R 2 µM	0.2 µM	2.5
EvaGreen 20X	1X	1.25
Agua		6.75
Volumen de DNA.		2
Volumen Total.		25 µL
Programa de ciclado: 95°C, 10min; 45 ciclos [95°C, 30s; 60.0°C, 90s]		
Curva HRM: 70-99° C		

8.5.1 Cuantificación de la carga viral de los cultivos virales y de algunas muestras clínicas.

Con el fin de cuantificar la carga viral de los cultivos virales y de las muestras clínicas se realizó una qPCR para construir una curva estándar empleando diferentes concentraciones de los controles positivos de HCMV y HHV6 (de 1×10^0 - 1×10^6 copias/µL), la reacción de PCR se preparó acorde a la siguiente tabla (Tabla 4) y se montó en la plataforma StepOne.

Para cuantificar la carga viral de las muestras clínicas y de los cultivos virales, se emplearon sondas de hidrólisis específicas para HCMV y HHV6 diseñadas en el laboratorio.

Tabla 4. Reacción para determinar la carga viral en el material biológico.

Mix de PCR.	Concentración final.	Vol. por reacción (µL).
Quantitect Probe PCR 2X (Qiagen)	1X	12.5
Primer mix F/R 5 µM	0.4 µM	2
Sonda 5 µM	0.2 µM	1
Agua		7.5
Volumen de DNA.		2
Volumen Total.		25 µL
Programa de ciclado:		
95°C, 15min; 45 ciclos [94°C, 15s; 60.0°C, 60s]		

8.6 Construcción de las librerías genómicas de secuenciación para la plataforma Ion Torrent.

De acuerdo con el manual para la construcción de librerías Ion Xpress Plus del Ion Torrent, la cantidad recomendada de DNA para realizar la construcción de las librerías va de 50ng hasta 1µg. Con la finalidad de simular la limitada cantidad de DNA de las muestras clínicas previamente obtenidas, se utilizaron 80ng de DNA de los cultivos virales para realizar la construcción de las primeras librerías de secuenciación. Para segmentar el DNA se utilizaron 2 métodos de fragmentación diferentes: La fragmentación por sonicación y la fragmentación enzimática.

8.6.1 Fragmentación por sonicación.

La fragmentación por sonicación consiste en la aplicación in vitro de sonidos de alta frecuencia (ultrasonidos) para fragmentar el DNA de forma aleatoria. El tamaño de los fragmentos de DNA generados por la sonicación depende de la frecuencia y el tiempo de sonificado al que fue expuesto el DNA. Se utilizó el sonicador S2 (Covaris) y 80ng de DNA de ambos cultivos virales (HCMV y HHV6), los parámetros de fragmentación aplicados al sonicador fueron los establecidos en el manual de construcción de librerías Ion Xpress Plus

Library (Tabla 5).

Tabla 5. Condiciones de fragmentación usadas en el Covaris S2.

Condición	Fragmentos de 100-250pb
Número de ciclos	6
Temperatura del baño	5 °C
Temperatura límite del baño	12 °C
Modo	Frecuencia de barrido
Función de prueba de calidad del agua	Apagado
Ciclo de trabajo	10%
Intensidad	5
Ciclos/ráfaga	100
Tiempo	60 segundos

Es importante señalar que este método solo realiza la fragmentación del DNA, debido a esto el tamaño de los fragmentos esperados estará alrededor de las 200pb, la unión de los adaptadores y código de barras se deberá realizar posteriormente. Con la finalidad de cerciorarnos que la fragmentación se realizará correctamente se emplearon chips de DNA High Sensitivity de Agilent Technologies y se siguió el protocolo correspondiente.

8.6.2 Fragmentación enzimática.

Esta es la técnica más común empleada para la construcción de librerías de secuenciación para el Ion Torrent, ya que se trata de un método automatizable que emplea diversas enzimas de restricción para realizar la fragmentación del DNA. Para realizar la fragmentación se utilizó el kit Ion Xpress Pluss Fragment Library y el sistema semi-

automatizado AB Library Builder (Life Technologies). La construcción de las librerías se realizó de acuerdo a los parámetros establecidos en el manual Ion Xpress Plus Library.

Cabe mencionar que este método da como resultado la construcción completa de las librerías, ya que en este se incluye la unión de los adaptadores y códigos de barra, por lo cual el tamaño esperado de los fragmentos es de 200-260pb. Con la finalidad de corroborar el tamaño de los fragmentos de las librerías se emplearon los chips de DNA High Sensitivity (Agilent Technologies) siguiendo su protocolo correspondiente.

8.7 Preparación del Templado Positivo.

Para la aplicación de este tratamiento primeramente es necesario normalizar las librerías, es decir crear una mezcla equimolar (preparar una mezcla con todas las librerías en la misma concentración) de las librerías a secuenciar, a *26 pM* de concentración, no obstante nosotros la preparamos a *40 pM* por las recomendaciones brindadas por el técnico de la plataforma de secuenciación, quien contaba con una amplia experiencia en dicho procedimiento. El templado positivo se realizó de acuerdo al protocolo del kit Ion PGM™ Template OT2 200 y se empleó la plataforma Ion OneTouch™ 2 Instrument.

8.8 Enriquecimiento del Templado Positivo.

Una vez obtenidos los templados positivos fueron enriquecidos utilizando el Ion OneTouch™ ES (sistema de enriquecimiento), los kits Ion PGM™ Template OT2 solutions 200, Ion PGM™ Template OT2 Supplies 200 y Ion PGM™ Enrichment Beads. Este proceso consiste en eliminar todas las nanopartículas a las que no se les haya adherido un templado, con la finalidad de evitar cargar el chip de secuenciación con nanopartículas “vacías”, lo que nos reduciría la cantidad de nanopartículas viables dentro del chip y en consecuencia se obtendría un cargado ineficiente del chip y una menor cantidad de lecturas. El enriquecimiento del templado positivo se realizó de acuerdo al protocolo Ion PGM™ Template OT2 200.

Cabe señalar que este proceso no nos permite eliminar las partículas policlonales (partículas con más de un templado diferente), estas son eliminadas por el software de la plataforma en los análisis previos a la alineación de la corrida de secuenciación.

8.9 Cargado del chip Ion 316 y secuenciación de los cultivos virales.

La inicialización de la plataforma Ion PGM (Ion Torrent), el cargado del chip 316 y la corrida de secuenciación de la plataforma se realizaron de acuerdo al protocolo Ion PGM Sequencing 200 Kit v2. Cabe mencionar que los pasos para realizar el cargado del chip se repitieron 2 veces más a las indicadas en el protocolo con la finalidad de maximizar el cargado lo mejor posible; antes de iniciar la corrida de secuenciación se cargaron las secuencias de referencia de los herpes virus HCMV cepa Merlin (ID GenBak: AY446894.2) y HHV6A cepa GS (ID GenBank: KC465951.1), además del genoma humano Hg19 (ID GenBank: NC_000001.11) con la finalidad de realizar un primer análisis de las lecturas obtenidas.

8.10 Construcción de librerías y cargado de la celda para secuenciación de la plataforma MiSeq.

8.10.1 Fragmentación y etiquetado del DNA.

Para realizar la fragmentación del DNA de los cultivos virales y muestras clínicas para la secuenciación en MiSeq se empleó protocolo Nextera XT DNA Library (Illumina), el cual consiste en un proceso de fragmentación enzimática que a diferencia del protocolo Ion Xpress Plus del Ion Torrent que emplea enzimas de restricción, este emplea transposasas cargadas con fragmentos de DNA (los adaptadores) que serán insertados al DNA etiquetándolo y fragmentándolo en pedazos de 300-500pb. Subsecuentemente se verificó el tamaño de los fragmentos generados utilizando los chips High Sensitivity (Agilent).

8.10.2 PCR de Amplificación y purificación de la PCR.

Una vez que el DNA de las muestras ha sido fragmentado y los adaptadores han sido adheridos a los fragmentos de DNA, se realiza la selección y adición de los índices

(secuencia de nucleótidos con los que se identificará cada librería) además de la amplificación de los fragmentado de DNA a través de una reacción de PCR (el programa de ciclado utilizado fue proporcionado por el protocolo Nextera XT DNA) con la que finaliza la construcción de las librerías. Al finalizar la reacción de PCR se realizó otro punto del control con el fin de observar la amplificación de las librerías, para lo cual se tomaron pequeñas alícuotas de las librerías y se analizaron en chips de DNA High Sensitivity para verificar la amplificación de las librerías.

Al finalizar la reacción de PCR y constatar que las librerías amplificaron correctamente se realizó la limpieza de la PCR, esta consiste en eliminar los productos inespecíficos generados durante el ciclado y en seleccionar los fragmentos con un tamaño determinado (300-500pb), una vez finalizada la limpieza de la PCR se tomaron pequeñas alícuotas de las librerías y se realizó nuevamente un punto de control con los chips de DNA High Sensitivity. Tanto la PCR de Amplificación como la purificación de esta se realizaron de acuerdo al manual de secuenciación Nextera XT DNA Sample Prep (Illumina).

8.10.3 Normalización y cargado de la celda de secuenciación.

Al finalizar la purificación de la PCR se realizó la normalización de las librerías, con la finalidad de procurar obtener la misma cantidad de lecturas durante la secuenciación para cada librería. El procedimiento se realizó de acuerdo a las indicaciones del protocolo Nextera XT DNA Sample Prep.

La celda de secuenciación de MiSeq utilizada (flowcell) para la corrida procede del kit MiSeq® Reagent Kit v3 (150 cycle), y se estableció como *Paired End* con la finalidad de obtener la mayor cantidad de información de las librerías, ya que nos permite leer hasta 150 bases de ambos extremos de los templados de DNA (por la química de este kit). Finalmente se realizó la secuenciación de las librerías, cabe señalar que a diferencia de los chips de secuenciación del Ion Torrent que son cargados manualmente por los usuarios de esta plataforma, el cargado de la celda de secuenciación de MiSeq se realiza de forma

automática por el instrumental de la plataforma. Cabe mencionar que el tiempo de la corrida de secuenciación del MiSeq, mucho mayor al del Ion Torrent, puede ser de hasta 48hrs.

8.11 Análisis de los datos de secuenciación.

8.11.1 Análisis de los datos de secuenciación del Ion Torrent.

En una primera instancia los cultivos secuenciados fueron analizados y ensamblados, utilizando las herramientas precargadas en el software de secuenciación de la plataforma Ion Torrent y empleando sus respectivas secuencias de referencia. Posteriormente los archivos (BAM y FASTQ) fueron re-analizados en la plataforma Galaxy-Server (<https://usegalaxy.org/>). Cabe señalar que las lecturas generadas por el Ion Torrent se denominan *Single End*, lo que quiere decir que las lecturas de los fragmentos de DNA se hicieron solo en un sentido.

Para el procesamiento *in situ*, se implementó el uso del sistema operativo **Bio-Linux** 7 y 8 (Field et al., 2006) en un servidor con un alto poder de procesamiento, mismo al que se le instalaron diversas herramientas para el procesamiento de datos de plataformas NGS, además de las herramientas preinstaladas en el sistema operativo.

Se desarrollaron 2 algoritmos para el análisis de los datos, en el primero se tomaron los archivos FASTQ pre-ensamblados con las secuencias de referencia por el software de la plataforma de secuenciación, las lecturas resultantes fueron filtradas de acuerdo a su tamaño y a la calidad Phred de estas, siendo la mínima establecida de Q20 en todos los casos y el tamaño mínimo establecido fue de 50pb. Para el segundo algoritmo se tomaron los archivos BAM del Ion Torrent (archivos en los que solo se le habían eliminado las secuencias de los barcodes en las lecturas) y se ensamblaron las lecturas contra la secuencia de referencia de humano empleando el software Samtools y sus herramientas con la finalidad de eliminar las lecturas que correspondieran a humano. Las lecturas

restantes fueron posteriormente filtradas por tamaño y calidad de lectura, para finalmente ser mapeadas contra las secuencias de referencia de HCMV cepa Merlin y HHV6A cepa GS. Aunado a esto, se utilizó la herramienta coverage del programa samtools con el fin de determinar la profundidad de mapeo promedio alcanzada por ambas plataformas, cabe señalar que la herramienta coverage no considera las regiones que no fueron mapeadas, por lo que la cobertura promedio proporcionada por la herramienta es un aproximado.

8.11.2 Análisis de los datos de secuenciación del MiSeq.

Debido a que la plataforma de secuenciación MiSeq no dispone en si misma de herramientas para el análisis de datos de secuenciación, y a que el tipo de lecturas proporcionadas por la plataforma son de diferente carácter informativo (Paired End) a las del Ion Torrent (Single End), las lecturas fueron procesadas de diferente forma.

Los archivos FASTQ de MiSeq se procesaron en la plataforma virtual Galaxy-Server. Primero se realizó el mapeo de todas las lecturas que correspondieran al genoma humano utilizando la referencia correspondiente (Hg19, ID GenBank: NC_000001.11), con la finalidad de eliminar todas las lecturas que correspondieran a este y obtener las lecturas no mapeadas (todas las lecturas que no pertenecieran a humano). Posteriormente las lecturas no mapeadas fueron filtradas utilizando tamaño mínimo de 90 bases y una calidad mínima de Q20. Finalmente las lecturas fueron mapeadas sobre los genomas de referencia de los herpesvirus HCMV-Merlin y HHV6A-GS.

8.11.3 Comparación y análisis de los polimorfismos de los cultivos virales respecto a las secuencias de referencia.

Para realizar el análisis de los polimorfismos de las secuencias de los cultivos virales, se utilizaron los archivos BAM mapeados contra las respectivas secuencias de referencia para generar archivos VCF (Figura 1), estos archivos contienen los posibles polimorfismos de las secuencias mapeadas respecto a las secuencias de referencia. Los archivos VCF fueron

generados con las herramientas Mpileup y bcftools view (Li, 2011; Minevich et al., 2012) de la plataforma Galaxy-Server.

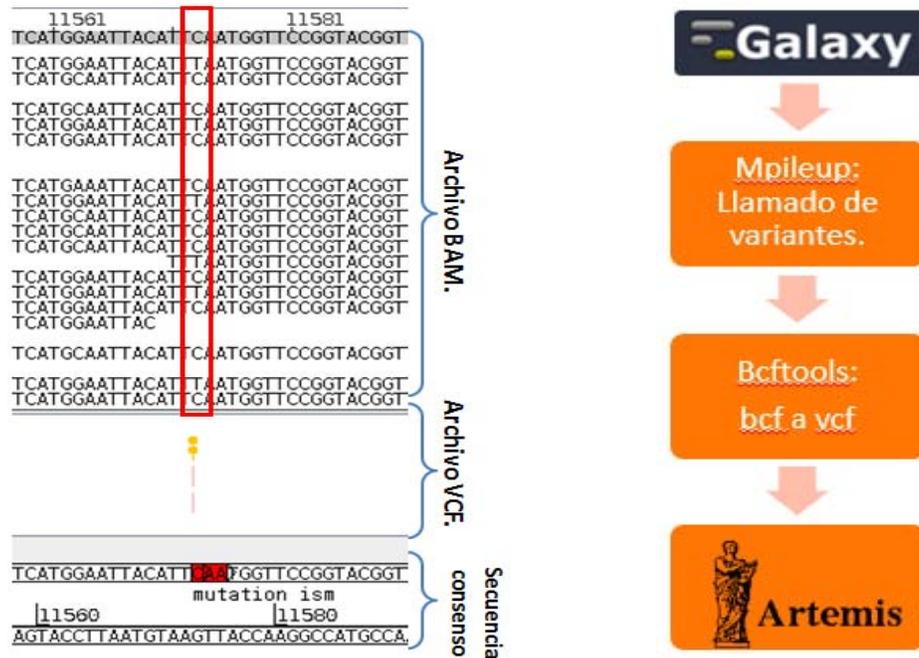


Figura 1. Algoritmo empleado para el análisis de polimorfismos. Visualización y análisis (en Artemis) de los polimorfismos de la secuenciación del cultivo viral de HCMV.

Posteriormente se visualizaron los archivos BAM y VCF junto con sus respectivas secuencias de referencia en los navegadores genómicos Artemis (Wellcome Trust Sanger Institute) (Rutherford et al., 2000) e IGV (Broad Institute) (Robinson et al., 2011; Thorvaldsdottir et al., 2013). Se analizaron detalladamente todos los posibles polimorfismos descartando aquellos que se localizaran en regiones repetitivas. Algunos de los polimorfismos encontrados fueron corroborados por medio de secuenciación Sanger convencional.

Para corroborar la presencia de algunos de los polimorfismos encontrados se diseñaron 5 pares de primers: 4 para HCMV y 1 para HHV6 (Tabla 6). Cada mix de PCR para amplificar

los diferentes segmentos de interés se preparó de acuerdo a lo indicado en la Tabla 7, por triplicado en un termociclador tipo Veriti (Life Technologies), con gradiente de temperatura con el fin de distinguir la mejor temperatura de amplificación.

Tabla 6. Primers diseñados para el análisis de los polimorfismos de los cultivos virales.

Primers	Secuencia
HCMV	
RL13sF	TCTCAAACCAACCGCAGTCATA
RL13sR	GGTAACGGTGCATATTGGTTCC
UL24AHF	GTTTGAAGTCCTCGTCCGTCT
UL24AHR	GGTGAGGACAAGAAGACGGAG
UL70ACF	GCGGAAAGAGGCTGCGAT
UL70ACR	CAGCTCGAACGCATCAACAC
UL112AF	AACAATACCCGAGGAGGCG
UL112AR	TTCCGGTCTCGGCTCTG
HHV6	
HHV6AgF	ATTGAGGACGGGAGAACGAG
HHV6AgR	ACCTCGAAGAAGCTTGCCCTAA

Tabla 7. Reacción de PCR para amplificar las regiones de interés.

Mix de PCR	Concentración final.	Vol. por reacción (µL)
Agua		15.5
PCR buffer 2SA 10X (Advantage)	1 X	2.5
dNTPs mix 10mM	0.2 mM	0.5
Primer Forward 10 µM	0.8 µM	2.0
Primer Reverse 10 µM	0.8 µM	2.0
Mix de polimerasas (Advantage 2)		0.5
Volumen de DNA.		2.0
Volumen Total.		25 µL
Programa de ciclado:		
95°C, 1min; 35 ciclos [95°C, 30s; gradiente (58°C/60°C/62°C), 30s; 68°C, 30s] + 68°C, 1min + 10°C, ∞		

Posteriormente se purificaron los productos de PCR por una electroforesis en un gel de agarosa al 2.5% y se cortaron las bandas con la mayor cantidad de producto, los productos de PCR se extrajeron y purificaron con el kit QIAquick Gel Extraction de Qiagen de acuerdo a su protocolo. La secuenciación de Sanger se preparó acorde a lo establecido en la Tabla 8.

Tabla 8. Reacción de secuenciación.

Mix de secuenciación	Vol. por reacción (µL)
Big Dye 3.1	0.5
Agua	2.5
Primer Forwar o Reverse 2 µM	1
DNA.	2
Volumen Total	6 µL
Programa de ciclado: 95°C, 1min; 25 ciclos [96°C, 10s; 50.0°C, 0.05s; 60°C, 4min] + 60°C, 4min	

Al finalizar se precipitaron las reacciones de secuenciación agregando 12.5 µl de etanol al 100% y 0.5 µL de acetato de sodio 3M en cada reacción. Se centrifugaron a 2000 rpm por 30 min y se decantó el sobrenadante, enseguida los productos se centrifugaron nuevamente durante 1 min a 150 rpm (boca abajo) para retirar el sobrenadante que no se hubiera desechado. Posteriormente se agregaron 75 µL de etanol al 70% a cada producto y se centrifugaron nuevamente a 2000 rpm durante 5 min para lavar los productos de PCR, el sobrenadante se decantó nuevamente y se centrifugó a 150 rpm durante 1 min. Finalmente los productos se incubaron durante 10 min para permitir que todo el etanol se evaporara y se resuspendió el DNA en 20 µL de formamida. Para realizar la secuenciación se utilizó un secuenciador ABI Prism 3130 Genetic Analyzer (Applied Biosystems).

8.12 Prevalencia de los herpesvirus HCMV (citomegalovirus) y HHV6 (herpesvirus humano 6) en la población mexicana de VIH.

Con el objetivo de determinar la prevalencia de DNA de HCMV y HHV6 en la población mexicana que vive con VIH, junto con el de encontrar muestras clínicas con una carga viral alta y cuya secuenciación sea viable, se utilizaron 270 muestras de diferentes estados de la república (Yucatán, Puebla, Baja California, Quintana Roo, Estado de México y Distrito Federal) participantes en el protocolo E1010 del INER.

Las muestras se dividieron en 3 lotes de 90 muestras, cada lote fue sometido al ensayo Herplex (qPCR acoplada a HRM) utilizando placas de PCR MicroAmp de 0.1mL. Para realizar el mix de PCR se utilizó el kit SsoFast EvaGreen Supermix with Low Rox (Bio-Rad), de acuerdo a la tabla (Tabla 9).

Tabla 9. Reacción Herplex para determinar la prevalencia de los herpesvirus HCMV y HHV6.

Mix de PCR	Concentración final.	Vol. por reacción (µL)
SsoFast with Low Rox (Bio Rad)		7.5
Primer mix F/R 2 µM	0.3 µM	2.25
Agua		3.25
Volumen de DNA.		2
Volumen Total.		15 µL
Programa de ciclado: 95°C, 3min; 45 ciclos [95°C, 30s; 60.0°C, 60s]		
Curva HRM: 70-99° C		

9. RESULTADOS.

9.1 Cultivo de HHV6A cepa GS.

Se infectaron células HSB-2 en 4 frascos T-175 con células HSB-2 previamente infectadas con HHV6A cepa GS. Las células HSB-2 infectadas con HHV6A se identificaban fácilmente, ya que mostraban una morfología muy distinta a las no infectadas: la principal característica era el notable incremento en su volumen celular y la formación de agregados celulares. La densidad celular de los cultivos virales se vigiló durante los días posteriores a la infección únicamente con el microscopio (durante estos días no se determinó concentración celular ni viabilidad). Al día 4 después de la infección (DPI), cuando el efecto citopático era extenso, se revisó la densidad y viabilidad celular de estos y se procedió a cosechar y congelar las células.

Tabla 10. Densidad y viabilidad de los cultivos de HSB-2 infectados con HHV6A cepa GS (4 DPI).

Cuento y viabilidad de los cultivos.					
	T175-1	T175-2	T175-3	T175-4	Control Negativo
Num. células/mL	.99x10 ⁶	.99x10 ⁶	1x10 ⁶	.93x10 ⁶	> 1x10 ⁶
Porcentaje de Viabilidad	77.7%	78.0%	76.1%	77.2%	98%

9.1.1 Cosecha del cultivo de HHV6A cepa GS.

Los 4 frascos se transfirieron a tubos de 50mL, se centrifugaron a 1500 rpm por 10 minutos, se decantó el sobrenadante y las células se resuspendieron en medio de congelación. Se prepararon 11 viales con una concentración de 2.5x10⁶ células/mL para congelación en nitrógeno líquido y también se prepararon 14 viales de 1mL con 3x10⁶ células/mL que se almacenaron a -20°C para la posterior extracción y secuenciación de su DNA.

9.2 Cultivo de HCMV cepa Merlin.

9.2.1 Infección de las células HFFF-2 con la cepa Merlin de HCMV

Cuando se observó una confluencia >80% en los frascos T-175 de los cultivos celulares se procedió a inocularlos con 3 millones de células HFFF-2 previamente infectadas con la cepa Merlin de HCMV (pase 6). Para ello se descongeló en un baño de agua a 37°C un vial criogénico con 1mL de células HFFF-2 infectadas. Luego de 7 DPI, las células mostraron un fenotipo característico de la infección por HCMV: células redondas de gran citoplasma que eventualmente se desprendían de la monocapa.

9.2.2 Cosecha de los cultivos virales de HCMV-Merlin.

Para preparar el stock de secuenciación (muestras a secuenciar) las células se desprendieron utilizando tripsina. Las células se lavaron en PBS, centrifugándolas a 1500 rpm por 10 minutos y se resuspendieron a una concentración de 1×10^6 células/mL. De tal manera se prepararon 45 viales de 1mL que fueron centrifugados a 3000 rpm por 10 minutos, se retiró el sobrenadante y fueron almacenados a -20°C.

9.3 Extracción del DNA genómico de los cultivos virales.

Se extrajo el DNA a partir de los pellets de células cultivadas utilizando extracción genómica de DNA (QIAamp DNA Mini Kit, Qiagen), dado el tamaño del genoma herpesviral. El DNA se eluyó en 200µL de buffer AE (Qiagen) de acuerdo al protocolo de extracción del fabricante. En una primera instancia se cuantificó el DNA extraído de 3 viales del stock de secuenciación de Merlin y 3 viales del stock de secuenciación de HHV6A-GS.

Alícuotas de los 6 tubos extraídos, se diluyeron 1:10 con el buffer AE, y se cuantificaron en el NanoDrop 8000 pero aunque los valores de $A_{260/280}$ eran altos (>2.0), los valores en $A_{260/230}$ eran negativos. Luego de observar que el propio buffer AE generaba una lectura de 1ng/µL, se determinó que el blanqueo del equipo era incorrecto, y además al verificar el

contenido del buffer, se encontró que contiene 0.5mM de EDTA, el cual absorbe en 230nm y podría estar afectando la cuantificación. Por lo anterior, se decidió extraer nuevamente DNA a partir de viales de cultivo viral nuevos.

Se tomaron 3 viales de células infectadas con HCMV y 2 viales de HHV6A nuevos y se marcaron como HCMV 1, HCMV 2, HCMV 3, HHV6A-A y HHV6A-B respectivamente. Las extracciones se realizaron nuevamente de acuerdo al protocolo, pero esta vez el DNA fue eluido en 150µL de agua libre de nucleasas. El DNA extraído fue diluido 1:10 para su cuantificación en NanoDrop 1000 (Tabla 11). Este material fue utilizado posteriormente para estudios de NGS.

Tabla 11. Cantidad y pureza del DNA extraído de los cultivos virales (DNA eluido en agua).

Muestras.	Concentración (ng/µL).	A₂₆₀	A₂₈₀	A₂₆₀/A₂₈₀	A₂₆₀/A₂₃₀
HHV6A A	18.77	0.375	0.183	2.05	2.03
HHV6A B	14.25	0.285	0.129	2.21	1.00
HCMV 1	12.38	0.248	0.108	2.29	1.93
HCMV 2	11.39	0.228	0.108	2.10	1.85
HCMV 3	11.59	0.232	0.110	2.11	2.00

9.4 Cuantificación y diagnóstico de las muestras clínicas.

Se cuantificó la concentración y pureza del DNA de las muestras clínicas diagnosticadas previamente con HCMV o HHV6 (diluidas en proporción 1:5) obteniendo los siguientes resultados (Tabla 12).

Tabla 12. Concentración y pureza del DNA de las muestras clínicas (dil. 1:5).

Muestras.	Concentración (ng/µL)	A₂₆₀	A₂₈₀	A₂₆₀/A₂₈₀	A₂₆₀/A₂₃₀
CAS	4.54	0.091	0.049	1.85	0.11
AHG	9.36	0.187	0.084	2.22	1.33

GGD	5.25	0.105	0.053	2.00	0.75
108.2S	2.85	0.057	0.015	3.69	0.76
JAGG	6.22	0.124	0.087	1.43	0.81
LV001*	1.73	0.035	0.021	1.67	0.37
RCMV107 [■]	3.06	0.061	0.052	1.17	0.08
RCMV107s	7.86	0.157	0.086	1.82	0.77

*: Exudado vesicular; ■: orina; el resto de las muestras eran DNA extraído de sangre total.

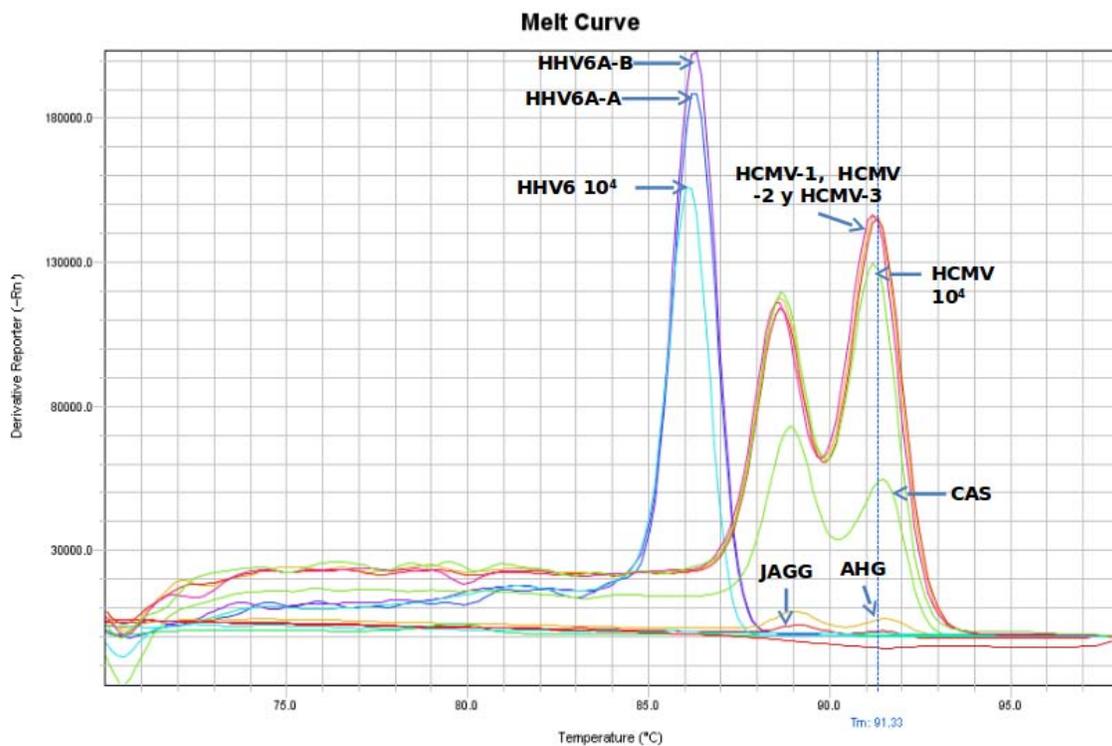


Figura 2. Curvas de disociación (HRM) de las muestras clínicas (CAS, AHG, GGD, 108.2S Y JAGG) y cultivos virales (HHV6A-A, HHV6A-B, HCMV-1, HCMV-2 Y HCMV-3).

Se comparó el perfil HRM de los controles positivos HCMV y HHV6 (con una concentración de 1×10^4 copias/ μ L) con el HRM de los cultivos (HCMV 1, HCMV-2, HCMV-3, HHV6A-A y HHV6A-B) y las muestras clínicas (CAS, JAGG y AHG) con el fin de obtener el diagnóstico de cada uno.

Las muestras clínicas y los cultivos virales se re-analizaron mediante la técnica Herplex donde se comparó el perfil del HRM de los controles positivos de HCMV y HHV6 (con carga viral de 1×10^4 copias/ μ L) con el de las muestras clínicas y los cultivos virales (HHV6A-A,

HHV6A-B, HCMV-1, HCMV-2 y HCMV-3) diluidos en proporción 1:10 (Figura 2 y Figura 3).

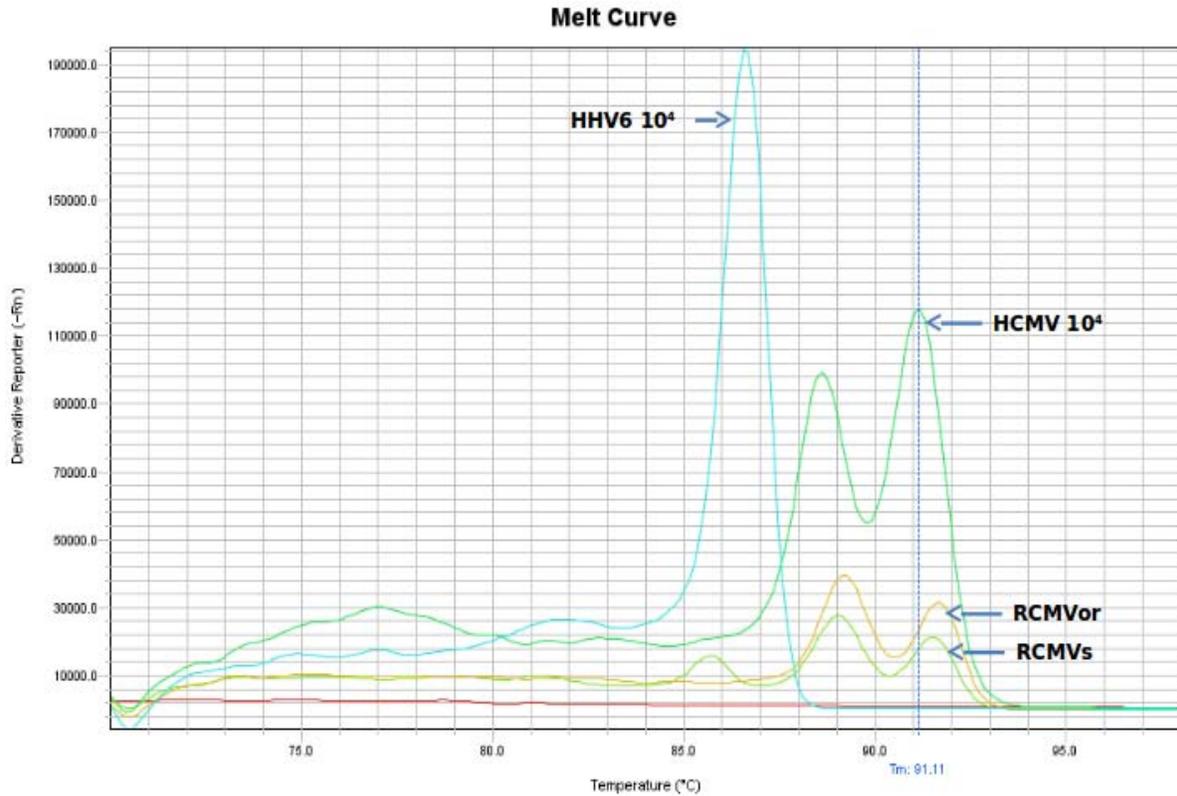


Figura 3. HRM de las muestras clínicas LV001, RCMVor y RCMVs.

Se comparó el HRM de los controles positivos HCMV y HHV6 (con concentración de 1×10^4 copias/ μ L) con el de las muestras clínicas RCMVor y RCMVs para realizar el diagnóstico de cada una.

Tabla 13. Comparación del diagnóstico previo y actual de las muestras clínicas.

Muestras Clínicas	Diagnóstico Previo (2008-2011)	Diagnóstico (Herplex)
CAS	HCMV y HHV6	HCMV
AHG	HCMV y HHV6	HCMV
GGD	HCMV	negativo
108.2S	HHV6 y HHV7	negativo
JAGG	HCMV	HCMV
LV001	HHV6 y HHV7	negativo

RCMV 107or	HCMV y HHV6	HCMV
RCMV 107s	HCMV y HHV6	HCMV y HHV6

En varias muestras clínicas no fue posible detectar uno o ambos virus con los que habían sido previamente diagnosticadas (Tabla 13), esto podría deberse a que parte del DNA viral se encontraba degradado debido a la manipulación de la que han sido objeto las muestras y por los periodos de congelación y descongelación a los que se sometieron en periodos pasados. Por otro lado, se constató la presencia de los virus de interés en los respectivos cultivos virales (Tabla 14).

Tabla 14. Resultado del Herplex de los cultivos virales.

Cultivos virales	Diagnóstico (Herplex)
HCMV-1	HCMV
HCMV-2	HCMV
HCMV-3	HCMV
HHV6A-A	HHV6
HHV6A-B	HHV6

9.5 Prueba de *Mycoplasma*.

Los cultivos virales fueron sometidos a una prueba de PCR y electroforesis para determinar si estos estaban libres de *Mycoplasma*, cabe señalar que el tamaño de los productos positivos era de 234-268pb; el ensayo demostró que los cultivos se encontraban libres de *Mycoplasma* (Figura 4).

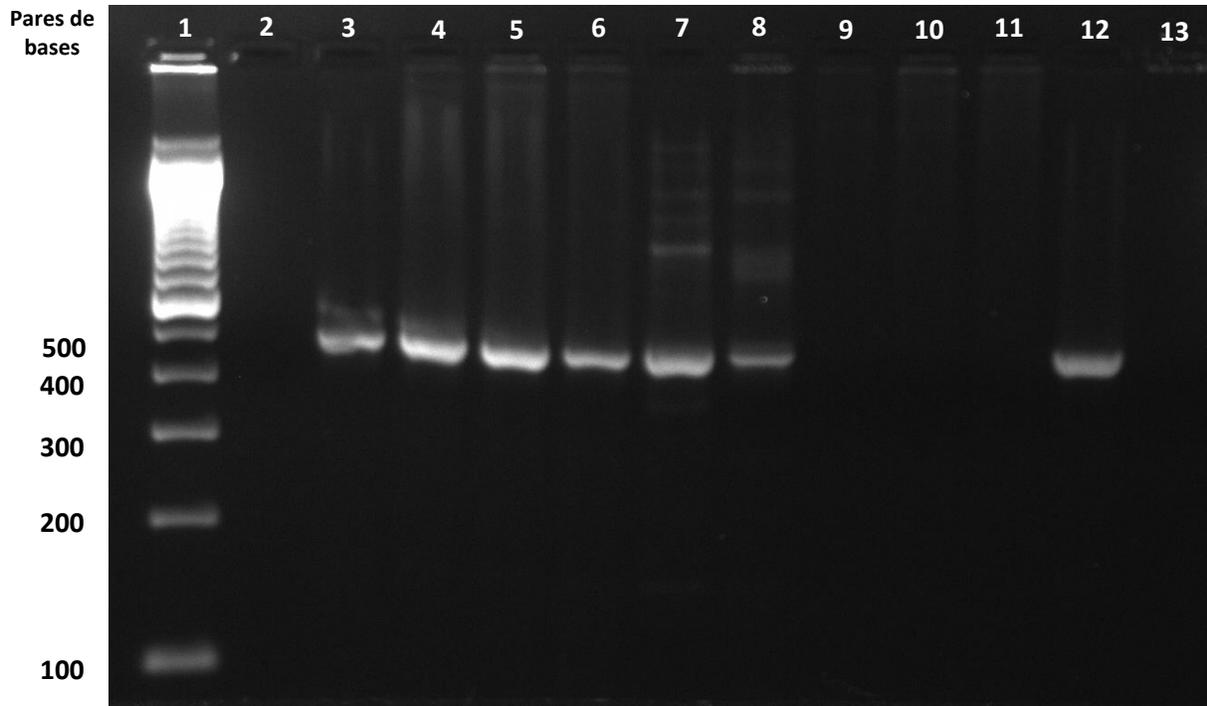


Figura 4. Prueba de mycoplasma.

Electroforesis de los productos de PCR de mycoplasma. Marcador de peso molecular de 100pb (1), control negativo (2), serie de controles positivos de PCR (3-8); Control positivo de mycoplasma (3), cultivos virales de HHV6A + control positivo de mycoplasma (4 y 5), cultivo viral HCMV + control positivo de mycoplasma (6), muestras no relacionadas + control positivo (7 y 8), cultivos virales de HHV6A (9 y 10), cultivo viral de HCMV (11) y muestras no relacionadas (12 y 13).

9.6 Cuantificación de la carga viral de los cultivos virales y de las muestras clínicas CAS y RCMV107s.

Para determinar la carga viral de las diluciones 1:10 de los cultivos virales y de las muestras clínicas CAS y RCMV107s (diluidas en agua en relación 1:5), se construyó una curva estándar mediante diluciones seriadas de plásmidos con insertos correspondientes a las regiones de interés de HCMV y HHV6 (1×10^6 a 1×10^0 copias/ μ L) (Figura 5).

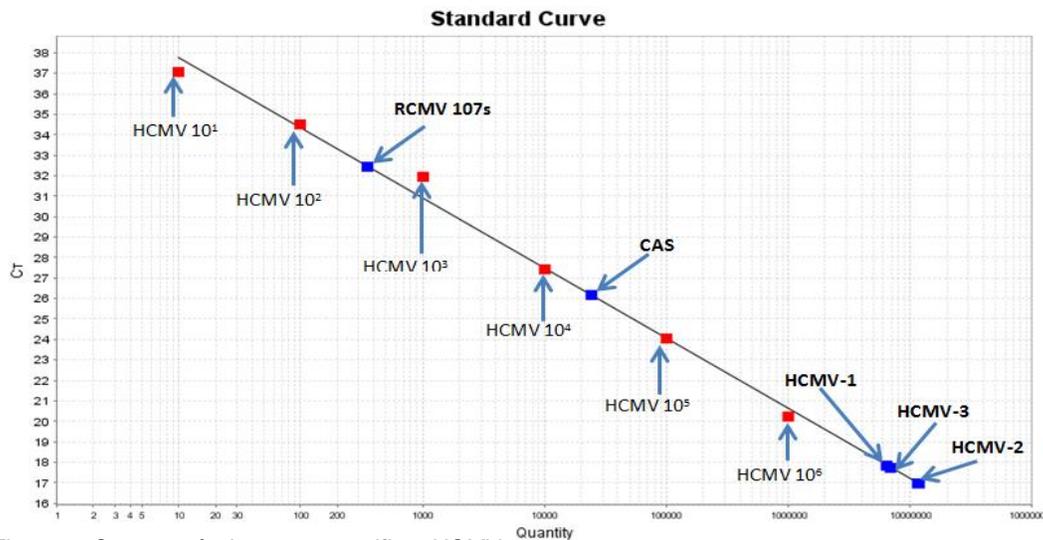
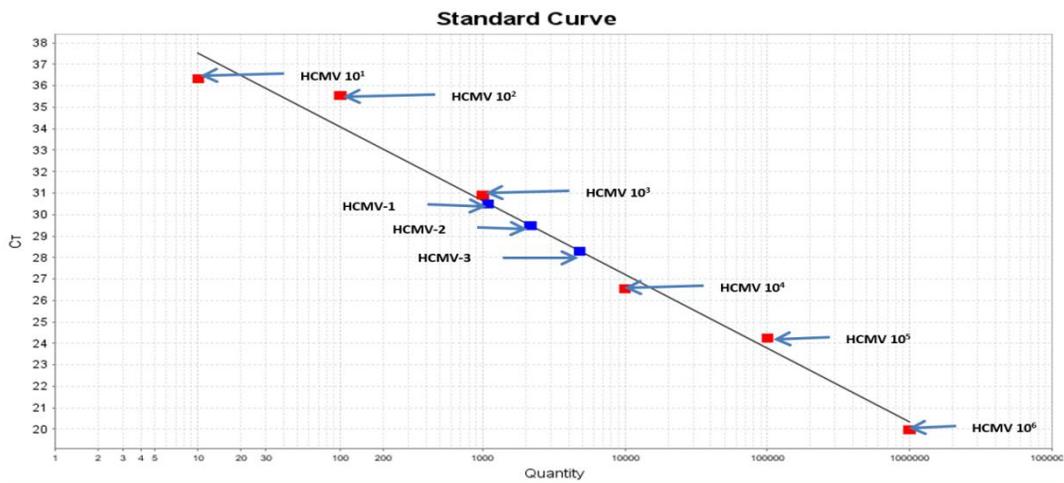


Figura 5. Curva estándar para cuantificar HCMV.

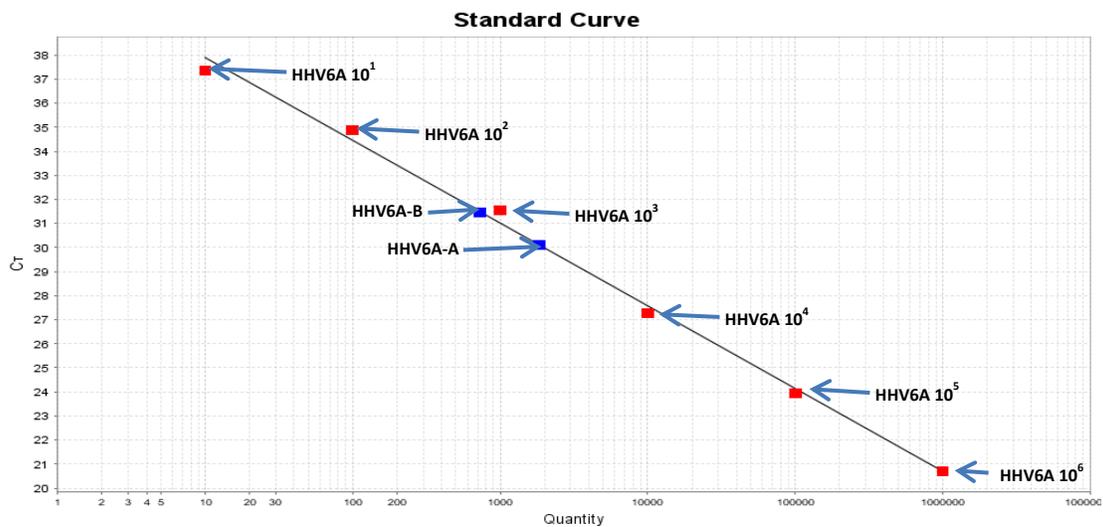
Se realizó una curva estándar con diferentes concentraciones del control positivo de HCMV para cuantificar la carga viral de las muestras clínicas CAS y RCMVs, y los cultivos virales de HCMV. Cabe notar que la carga viral de los cultivos se encuentra por encima de la curva estándar.

La carga viral de HCMV en las muestras clínicas RCMV107s y CAS obtenida fue de 354 y 24,080 copias/ μ L respectivamente; cabe señalar que esta cuantificación corresponde a la dilución 1:5 de las muestras clínicas, por lo cual la carga viral de las muestras clínicas originales se estimó en 1, 770 y 120, 400 copias/ μ L respectivamente. La carga viral de las diluciones 1:10 de los cultivos sobrepasó el umbral de cuantificación de las curvas estándares (Figura 5). Para cuantificar la carga viral de los cultivos virales se incrementó el número de diluciones (Figura 6 y Figura 7). No fue posible cuantificar la carga viral de HHV6 para la muestra RCMV107s, es posible que ésta se encontrara por debajo del umbral de detección.



Target: H5POL1 Slope: -3.434 Y-Inter: 40.948 R²: 0.979 Eff%: 95.526

Figura 6. Determinación de la carga viral de los cultivos de HCMV.
 Cuantificación de la carga viral de las diluciones 1:100 mil de los cultivos virales de HCMV.



Target: H6POL_1 Slope: -3.44 Y-Inter: 41.326 R²: 0.996 Eff%: 95.309

Figura 7. Determinación de la carga viral en los cultivos de HHV6A.
 Cuantificación de la carga viral de las diluciones 1:1 millón de los cultivos virales de HHV6A-A y HHV6A-B.

Tabla 15. Carga viral de las muestras clínicas y los cultivos virales.

Muestra	Dilución	Carga viral del material genético (copias/ μ L)
HCMV-1	1:100,000	109,000,000
HCMV-2	1:100,000	216,500,000
HCMV-3	1:100,000	481,600,000
HHV6A-A	1:1,000,000	1,832,000,000
HHV6A-B	1:1,000,000	728,000,000
RCMV107s	1:5	1,770
CAS	1:5	120,400

9.7 Construcción de librerías.

Con el fin de seleccionar el método de construcción de librerías más adecuado, se realizó la fragmentación del DNA genómico por un método físico (sonicación) y un enzimático siguiendo las recomendaciones para cada método indicadas en el manual del Ion Xpress Plus de construcción de librerías para la plataforma Ion Torrent.

9.7.1 Fragmentación por Sonicación.

Se tomaron 80ng de DNA de los cultivos virales de HCMV y HHV6A y se fragmentó a 5°C en el DNA shearing S2 (Covaris) bajo las condiciones descritas en la metodología (Tabla 5). Al finalizar la fragmentación se tomaron alícuotas y se analizó el perfil de fragmentación de estas en un chip de DNA high sensitivity (Agilent) (Figura 8).

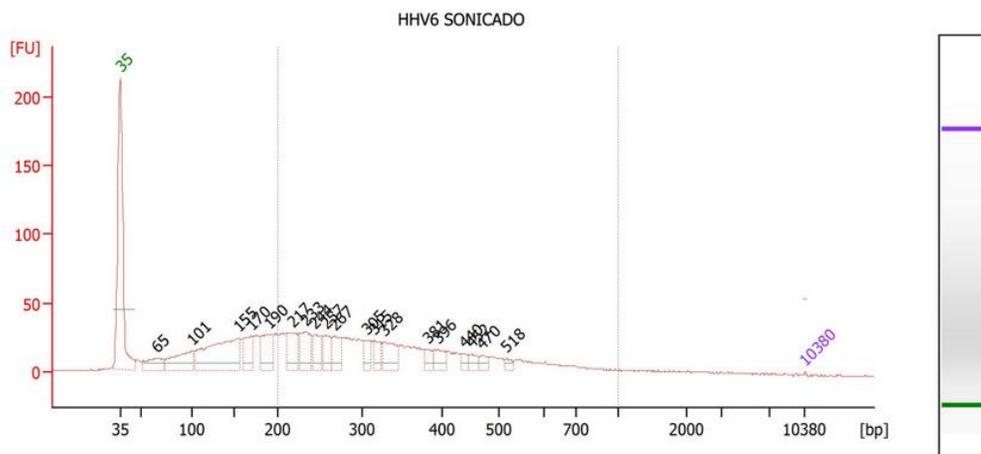


Figura 8. Electroferograma del cultivo de HHV6A sonicado (utilizando 80ng de DNA). Perfil de fragmentación (gama de fragmentos obtenidos) del DNA del cultivo de HHV6A siguiendo los parámetros establecidos en el protocolo Ion Xpress Pluss del Ion Torrent, utilizando el sonicador Covaris-S2. Aunque el equipo estableció el marcador terminal (10380) se observa que no logró migrar en el tiempo de corrida, por lo cual es posible que los valores de tamaño estén sobre estimados.

Al visualizar el electroferograma de HHV6A se observa una amplia distribución en el tamaño de los fragmentos; con el fin de construir adecuadamente la librería era necesario obtener los fragmentos con un tamaño determinado (200 bases), para ello se realizó una serie de electroforesis en geles de agarosa, seleccionando manualmente los fragmentos de 200 pares de bases del DNA sonicado de HHV6A-A a través de E-Gel SizeSelect Agarose Gel 2% (Invitrogen) sistema de electroforesis en gel de agarosa.

Una vez purificadas las bandas con los fragmentos de interés se tomó de nueva cuenta una alícuota y se analizó nuevamente en un chip de DNA high sensitivity para corroborar que el tamaño de los fragmento purificados sea el correcto (Figura 9).

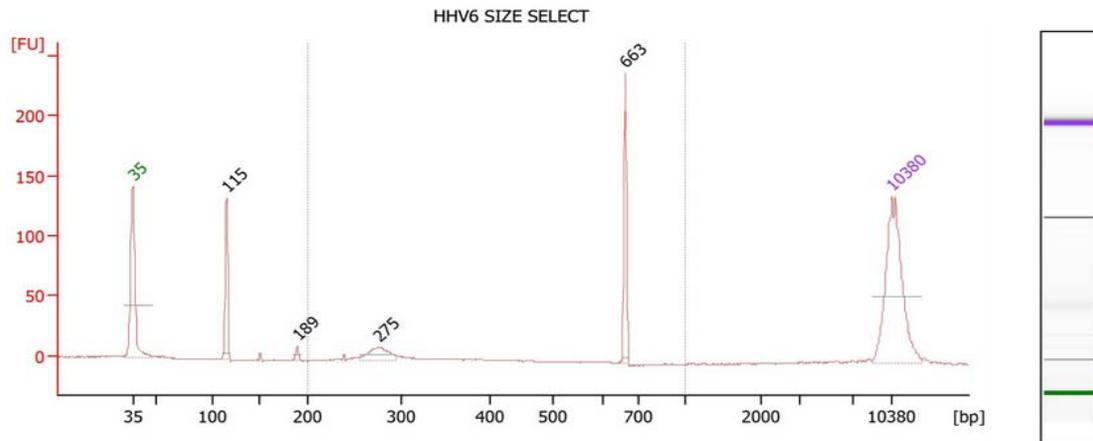


Figura 9. Electroferograma de la purificación manual (empleando E-Gels) del sonicatedo del cultivo de HHV6A.

Se seleccionaron y purificaron manualmente los fragmentos de 200pb del DNA del cultivo de HHV6A sonicatedo, sin obtener los resultados esperados.

Al analizar el electroferograma pudimos constatar que las bandas purificadas no correspondían al tamaño de banda deseado, aún cuando el protocolo se siguió cuidadosamente, por lo que no se continuó con la construcción de la librería manualmente. Cabe señalar que además se observó una pérdida significativa en la cantidad de DNA entre el proceso de fragmentación y la selección manual de los fragmentos con un tamaño determinado.

9.7.2 Fragmentación Enzimática.

Se tomaron 80ng de DNA de los cultivos virales y se fragmentaron de acuerdo al protocolo de fragmentación enzimática y selección automática del tamaño de los fragmentos en el equipo AB Library Builder (LifeTechnologies) según el manual del Ion Xpress Plus de construcción de librerías para el Ion Torrent. Se utilizaron los códigos de barra 005 y 006 para los cultivos de HCMV y HHV6A respectivamente. Cabe recordar que este procedimiento da como resultado la construcción de la librería, con tamaño esperado de los fragmentos de 200-260 pb. Al finalizar la construcción de las librerías se analizó el perfil electroforético de los fragmentos en un chip DNA High Sensitivity (Figura 10) sin observar

algún resultado para la librería de HCMV. Por otro lado en el electroferograma de la librería de HHV6A se observó una menor dispersión en el tamaño de los fragmentos respecto a los obtenidos por sonicación.

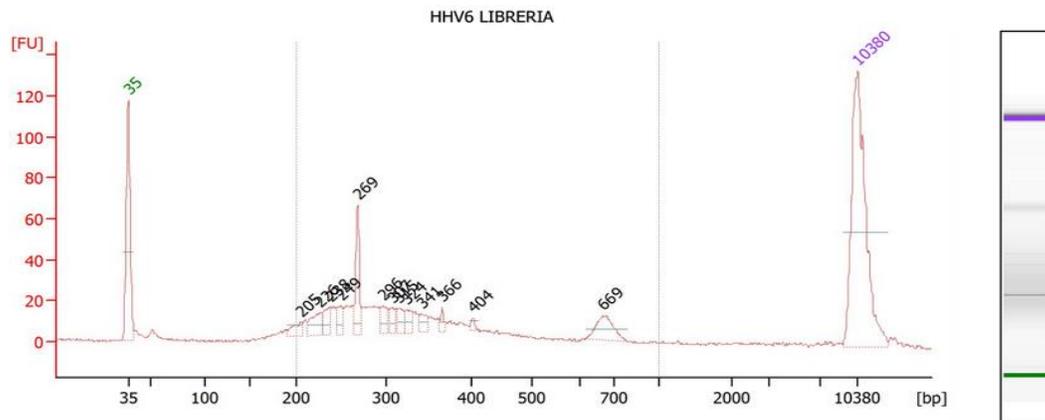


Figura 10. Electroferograma del cultivo de HHV6A fragmentado con enzimas (utilizando 80ng de DNA).
Perfil de la fragmentación enzimática del DNA del cultivo de HHV6A. En la figura se observa que la mayoría de los fragmentos obtenidos se encuentran entre los 200-260pb.

La prueba se repitió nuevamente para ambos cultivos, sin obtener algún resultado para HCMV, por lo que se realizó una electroforesis en gel de agarosa al 2% con el fin de descartar algún indicio de degradación del DNA genómico de los cultivos. Al observar la electroforesis (Figura 11) no se encontró ningún tipo de bandeo que indicara degradación del DNA y el barrido observado concordaba con el perfil de una corrida de DNA genómico, por lo que se concluyó que el DNA de los cultivos virales se encuentra en buen estado.

Otra posibilidad es que la cantidad de DNA utilizada haya sido mínima o bien que se haya calculado erróneamente el volumen de DNA a tomar, colocando una cantidad de DNA menor a la mínima establecida por el protocolo.

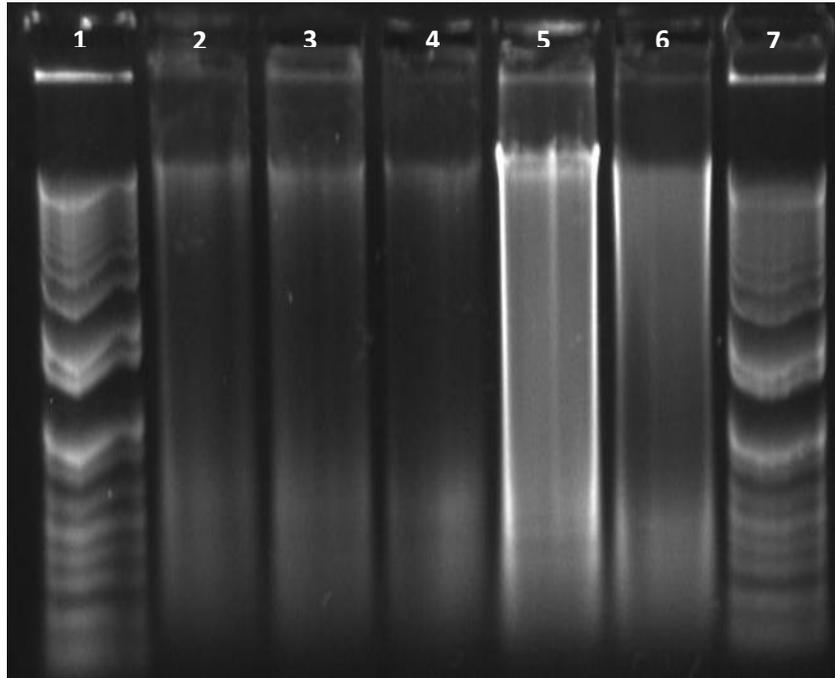


Figura 11. Electroforesis del DNA genómico de los cultivos virales.

Marcador de peso molecular de 1kb (1), Cultivo de HCMV-1 (2), cultivo de HCMV-2 (3), cultivo de HCMV-3 (4), cultivo de HHV6A-A (5) y cultivo de HHV6A-B (6), marcador de peso molecular de 1kb (7). No se observó algún tipo de bandeo en ninguno de los cultivos virales, con lo cual quedo verificado que el DNA de los cultivos no se encuentra degradado.

Debido a que la fragmentación enzimática fue el método que nos proporcionó los mejores resultados, se seleccionó como la técnica a implementar en la construcción de las librerías genómicas posteriores.

9.8 Secuenciación de los cultivos virales.

9.8.1 Construcción de las librerías de secuenciación para Ion Torrent.

Dado los resultados anteriores, se utilizaron 500ng de DNA de cada cultivo para la elaboración de librerías y se revisó su perfil (Figura 12 y Figura 13).

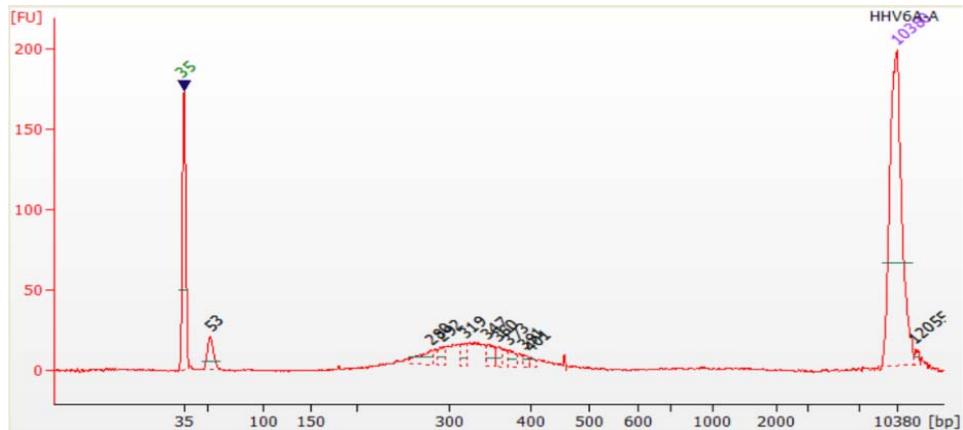


Figura 12. Electroferograma de la librería de HHV6A (fragmentación enzimática) a partir de 500ng de DNA.

En el electroferograma se observa que el tamaño de los fragmentos obtenidos se encuentra por encima del tamaño esperado (librería a secuenciar).

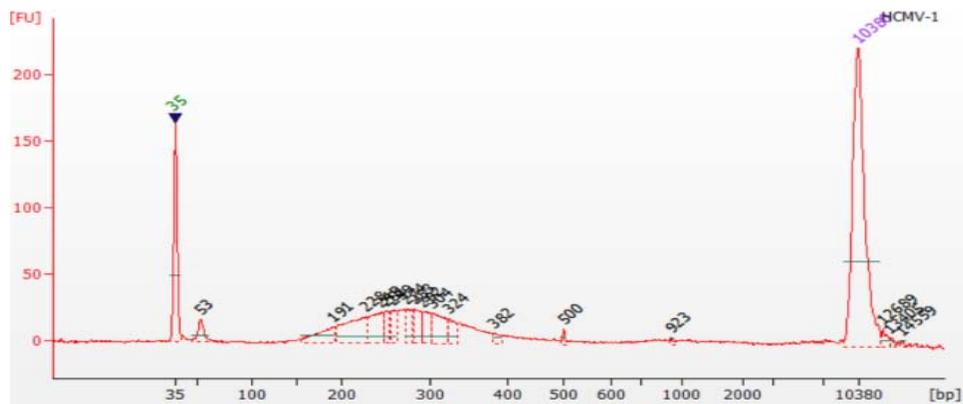


Figura 13. Electroferograma de la librería de HCMV a partir de 500ng de DNA.

Perfil de fragmentación enzimática de la librería de HCMV a secuenciar.

Se observó el perfil de fragmentación de la librería de HHV6A observando que los

fragmentos eran de mayor tamaño al esperado (260pb), a pesar de ello decidimos emplearla en la secuenciación, con el fin de conocer cómo se comportaban las librerías cuyo tamaño era mayor al esperado.

Por otro lado, al observar la distribución del tamaño de los fragmentos de la librería de HCMV en el electroferograma correspondiente (Figura 13) se aprecia que la mayoría de los fragmentos se encuentran dentro del rango de tamaño esperado (200-260pb).

9.8.2 Secuenciación de los cultivos virales en el Ion Torrent.

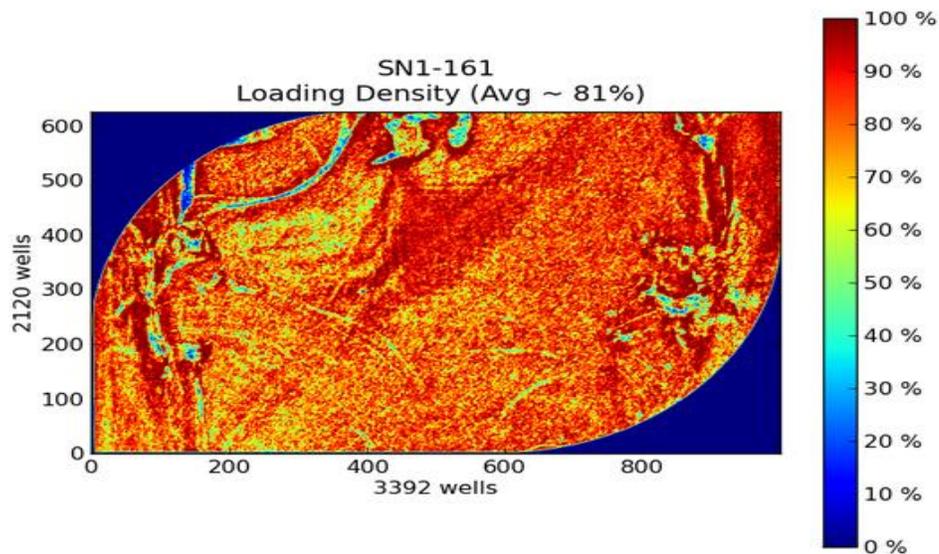


Figura 14. Cargado del chip 316 (mapa de calor).

Mapa de calor del cargado del primer chip 316 corrido, correspondiente a la corrida de las librerías HCMV-1 (código de barra 30), HHV6A-A (código de barra 29) y HHV6A-A (código de barra 6).

En la primera corrida de NGS se emplearon 3 librerías procedentes de los cultivos virales, de las cuales 2 procedían del cultivo de HHV6A marcadas con los códigos de barra 006 (80ng de DNA iniciales) y 029 (500ng) respectivamente y una de HCMV marcada con el

código de barra 030 (500ng), obteniendo un excelente cargado del chip 316 v2 (LifeTechnologies) (Figura 14), cabe señalar que una de las librerías de HHV6A se construyó con 80ng de DNA y las 2 restantes con 500ng de DNA. El software de secuenciación del Ion Torrent realizó un primer análisis observando que de los 5, 147, 086 de pozos cargados (81% del chip), el 22% correspondían a partículas policlonales (partículas con más de un templado diferente), por lo que estas fueron descartadas por el software de la plataforma, y posteriormente el 50% de las lecturas fueron descartadas principalmente al no pasar el filtro de calidad en la secuencia. Finalmente se obtuvieron poco más de 2 millones de pozos (*reads*) de carácter informativo (Figura 15).

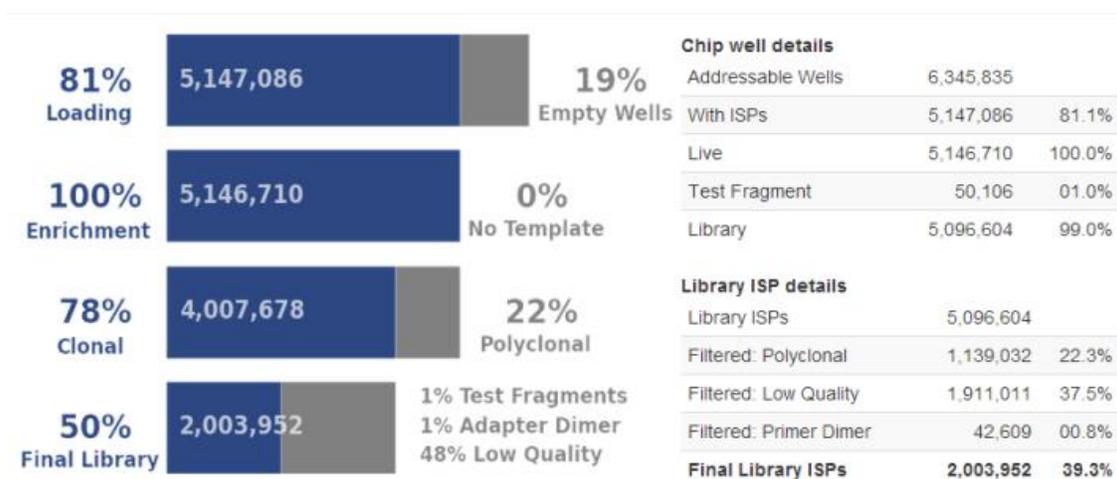


Figura 15. Resumen de la primer corrida en IonTorrent 316v2.

Muestra el total de lecturas obtenidas, el porcentaje de partículas con templado que fueron enriquecidas, la cantidad de partículas policlonales y las partículas que finalmente fueron de carácter informativo.

Figura 4. Características generales del cargado del chip.

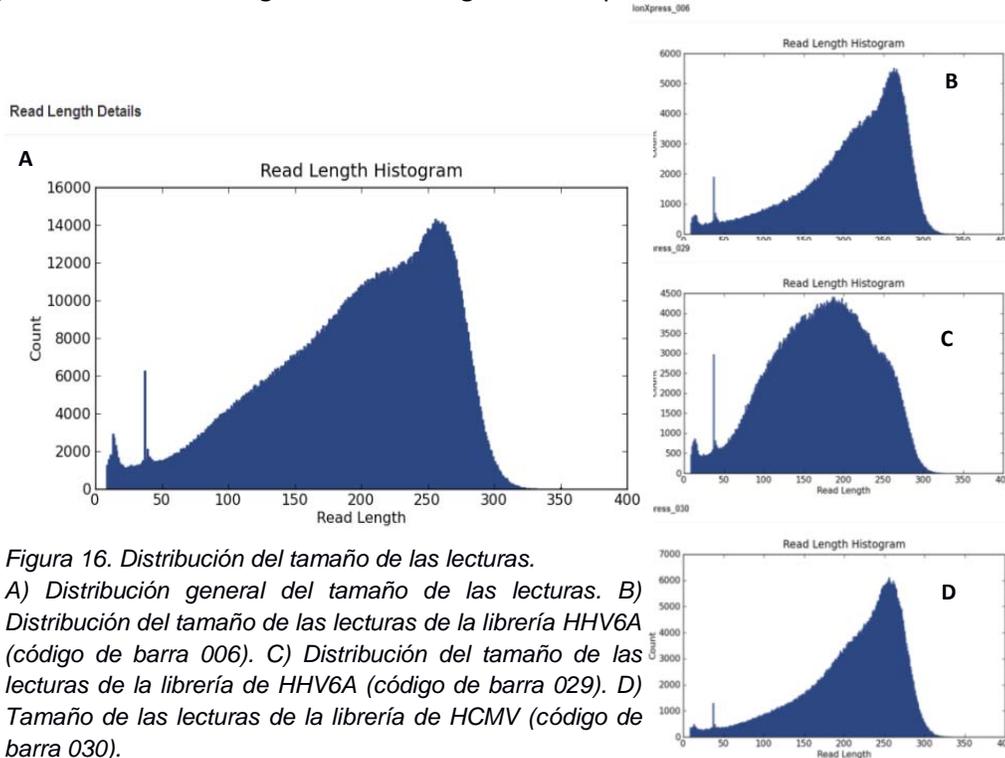


Figura 16. Distribución del tamaño de las lecturas.

A) Distribución general del tamaño de las lecturas. B) Distribución del tamaño de las lecturas de la librería HHV6A (código de barra 006). C) Distribución del tamaño de las lecturas de la librería de HHV6A (código de barra 029). D) Tamaño de las lecturas de la librería de HCMV (código de barra 030).

La distribución del tamaño de las lecturas de las librerías de HHV6A (codigo de barra 006) y HCMV (codigo de barra 030) era de acuerdo al esperado, a diferencia de la librería de HHV6A (codigo de barra 029), (Figura 16), esta presentaba una distribución más amplia en el tamaño de sus lecturas, obteniendo principalmente lecturas de menor tamaño al esperado.

El 16% de las lecturas fue exitosamente ensamblado contra las secuencias de referencia de HCMV cepa Merlin y HHV6A cepa GS con el software de la plataforma; el resto posiblemente pertenecen a DNA humano en su mayoría (Figura 17A), pero todo el material tuvo buenas calidades en las lecturas incluso en 250pb (Figura 17B). La Figura 18A nos muestra que 52 Mpb poseen una calidad igual o mayor a Q20 dentro de la escala de phred. Posteriormente se examinó de forma independiente la calidad de cada librería

(Figura 19 - Figura 21), observando que en todas la calidad se mantuvo por encima de Q30 hasta la posición 150 y luego esta comienza a decaer, sin embargo este comportamiento es normal en la calidad de las librerías del IonTorrent.

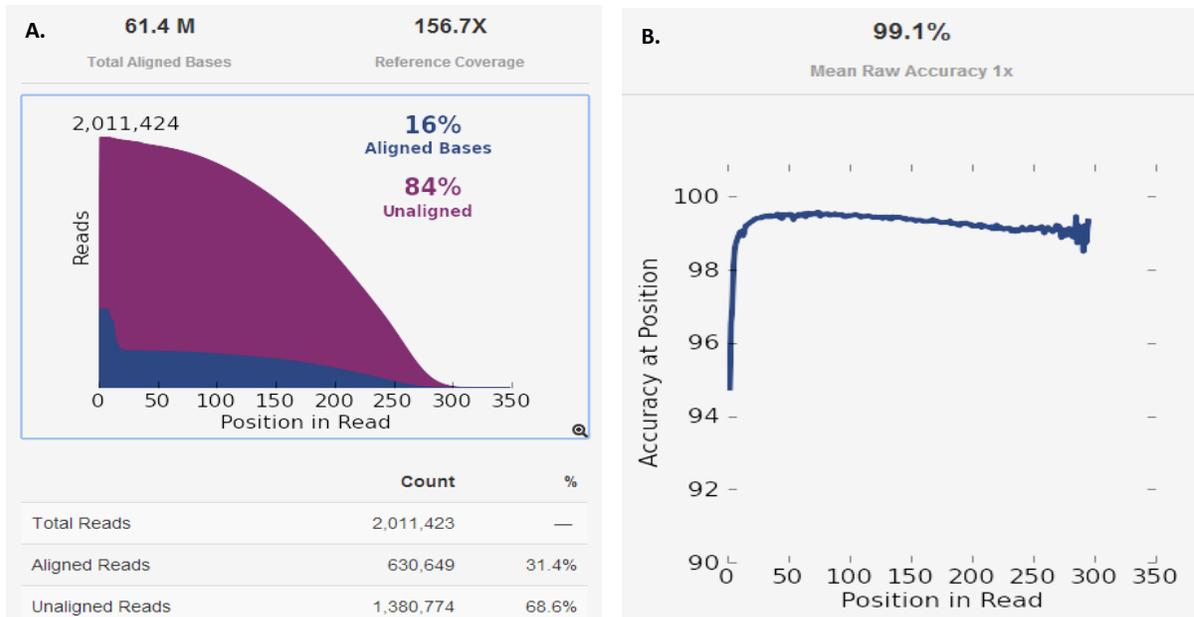


Figura 17. Alineación de las lecturas a las secuencias de referencia y exactitud del llamado de las bases.

A) Porcentaje de lecturas alineadas contra las secuencias de referencia (de HCMV-Merlín y HHV6A- GS). B) Exactitud de los nucleótidos ensamblados de acuerdo a su posición en las lecturas.

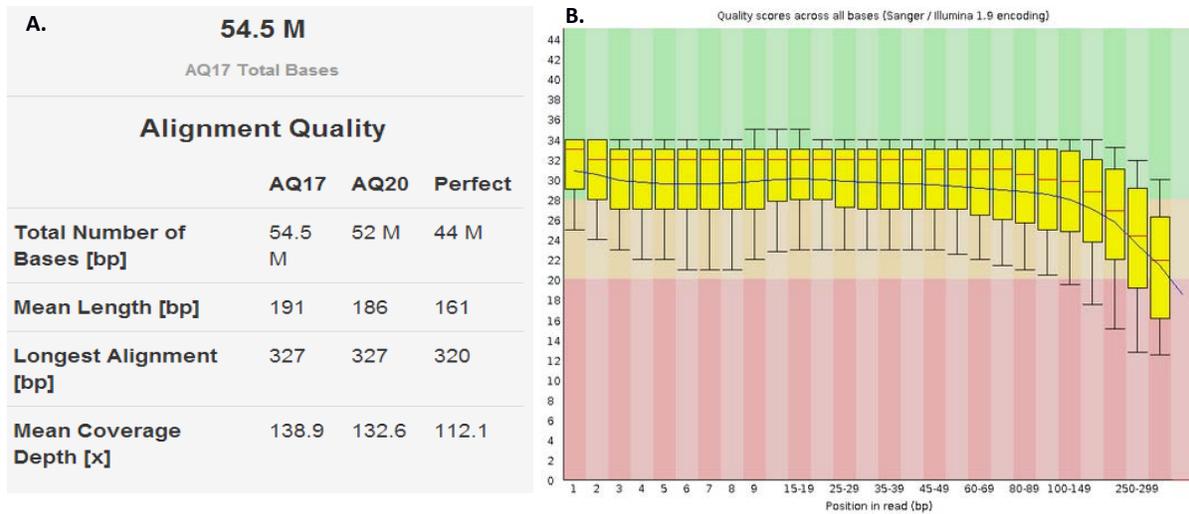


Figura 18. Calidad phred de las bases y lecturas alineadas.

A) Número de bases con calidad Q17 o superior, longitud y calidad promedio de las lecturas alineadas, profundidad promedio de la cobertura. B) Calidad phred general de los nucleótidos de acuerdo a su posición en las lecturas.

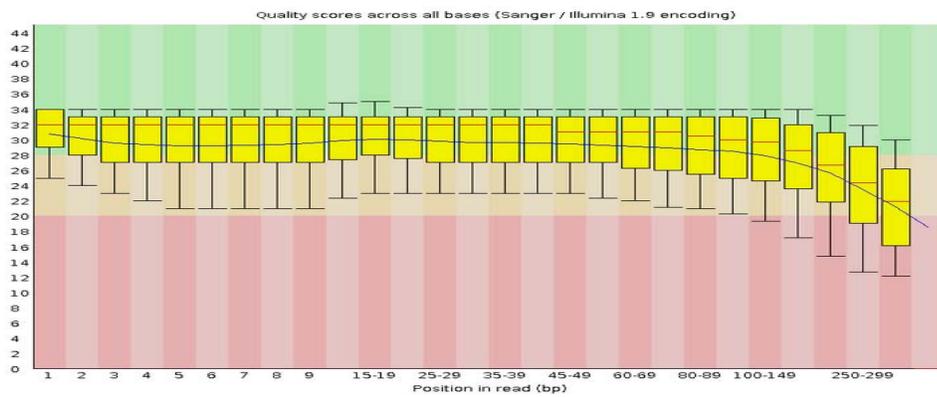


Figura 19. Calidad de las lecturas de la librería de HHV6A hecha con 80ng.

Calidad promedio de cada base por posición en las lecturas de la librería de HHV6A de 80ng.



Figura 20. Calidades de la librería de HHV6A hecha con 500ng.
 Calidad promedio de cada base por posición en las lecturas de la librería de HHV6A de 500ng.



Figura 21. Calidad de la librería de HCMV hecha con 500ng.
 Calidad de cada base de acuerdo a su posición en las lecturas de la librería de HCMV.

Después de obtener los resultados del Ion Torrent, se extrajeron los archivos FASTQ y se procesaron con la herramienta Filter FaSTQ de la plataforma online Galaxy (Blankenberg et al., 2010) filtrando las lecturas con una calidad menor a Q20 y tamaño mínimo de 50pb, para finalmente realinear las lecturas con sus respectivas secuencias de referencia se utilizó el programa BWA (Li and Durbin, 2009).

Se comparó el mapeo de las lecturas según el algoritmo estándar del Ion Torrent contra el procesamiento manual mencionado. Los archivos se visualizaron en Artemis respecto a las secuencias de referencia (Figura 22) observando que el procesamiento manual mostraba una mejor calidad de lecturas y una menor cantidad de polimorfismos que el procesamiento automático del Ion Torrent además de la conducente exclusión de los fragmentos cortos. Cabe recalcar brevemente que el procesamiento manual consiste en alinear las lecturas contra el genoma humano y posteriormente filtrar las lecturas restantes (lecturas no alineadas a humano) por tamaño y calidad para finalmente mapearlas contra los genomas de referencia viral.

Aunado a esto se determinó un aproximado de la profundidad promedio alcanzada para cada genoma, obteniendo un profundidad aproximada de 100 lecturas para el cultivo de HCMV cepa Merlin, mientras que la del cultivo de HHV6 cepa GS fue de apenas 28 lecturas de profundidad en promedio.



Figura 22. Secuenciación masiva del cultivo de HCMV alineado con su referencia. Se visualizó la alineación de las lecturas de la librería del cultivo de HCMV con el visualizador Artemis, las líneas negras y verdes representan las lecturas obtenidas por la secuenciación, donde el color verde representa las lecturas con una calidad promedio de $Q < 20$ y las líneas negras las lecturas con una calidad de $Q > 30$. Visualización del archivo BAM del cultivo de HCMV analizadas automáticamente (A) y después del filtrado manual de las lecturas en Galaxy (B).

9.8.3 Cobertura y profundidad de la secuenciación obtenida respecto al contenido de G+C.

Aunque no se realizó ningún análisis estadístico, se observó un fenómeno particular: Las regiones con una menor profundidad de mapeo (áreas en rojo) corresponden a áreas con un contenido de G+C relativamente alto. De forma inversa, las áreas que presentan una

mayor profundidad (áreas en azul) corresponden a zonas en las que el contenido de G+C decae drásticamente (Figura 23).

Las regiones con repeticiones terminales invertidas, que se caracterizan por presentar un alto contenido de A+T en el genoma de HCMV, ocasionan una calidad de mapeo bajo por lo que estas áreas no se tomaron en cuenta para el análisis.

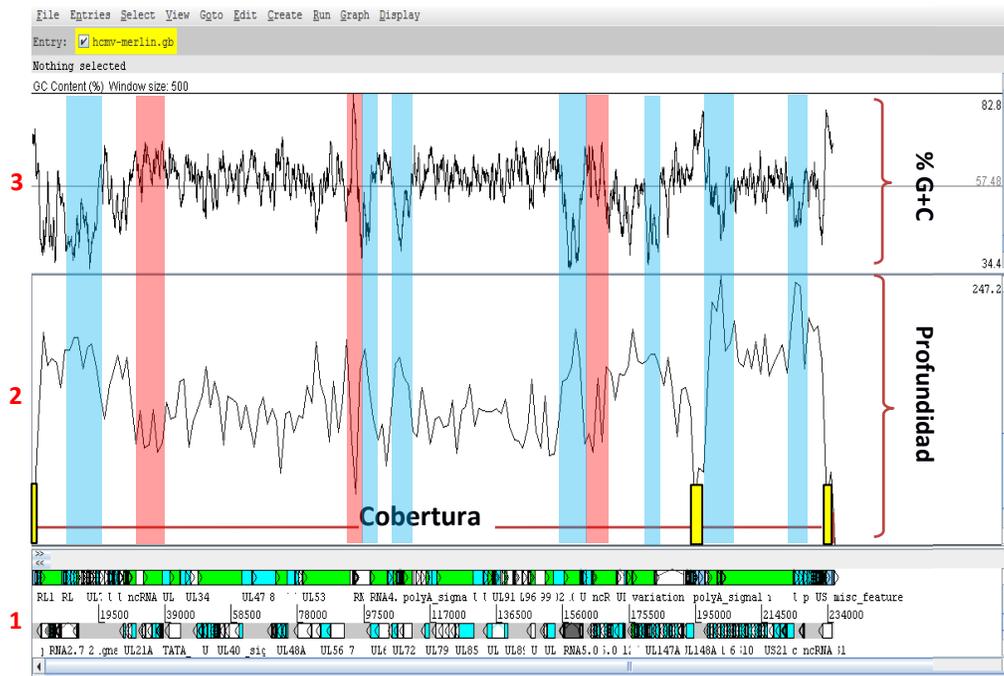


Figura 23. Comparación de la cobertura y profundidad obtenida por la secuenciación, respecto al contenido de G+C del genoma de referencia (Ion torrent).

Se comparó la alineación del archivo BAM del cultivo de HCMV cepa Merlin contra su respectiva secuencia de referencia, en relación al contenido de G+C. 1) Genoma de referencia (HCMV cepa Merlin). 2) Cobertura y profundidad alcanzada por la alineación de las lecturas del archivo BAM. 3) Contenido de G+C de la secuencia de referencia. Las áreas en azul corresponden a zonas con un bajo contenido de G+C, mientras que las áreas en rojo a zonas con un contenido de G+C alto. En amarillo se señalan las repeticiones terminales invertidas que no se tomaron en cuenta para el análisis.

De forma similar las lecturas de las librerías de HHV6A tuvieron el mismo comportamiento en la profundidad y cobertura de las lecturas respecto al contenido de G+C del genoma

(Figura 24), aunque se observó un exceso de cobertura en un sitio cercano al origen de replicación del virus.

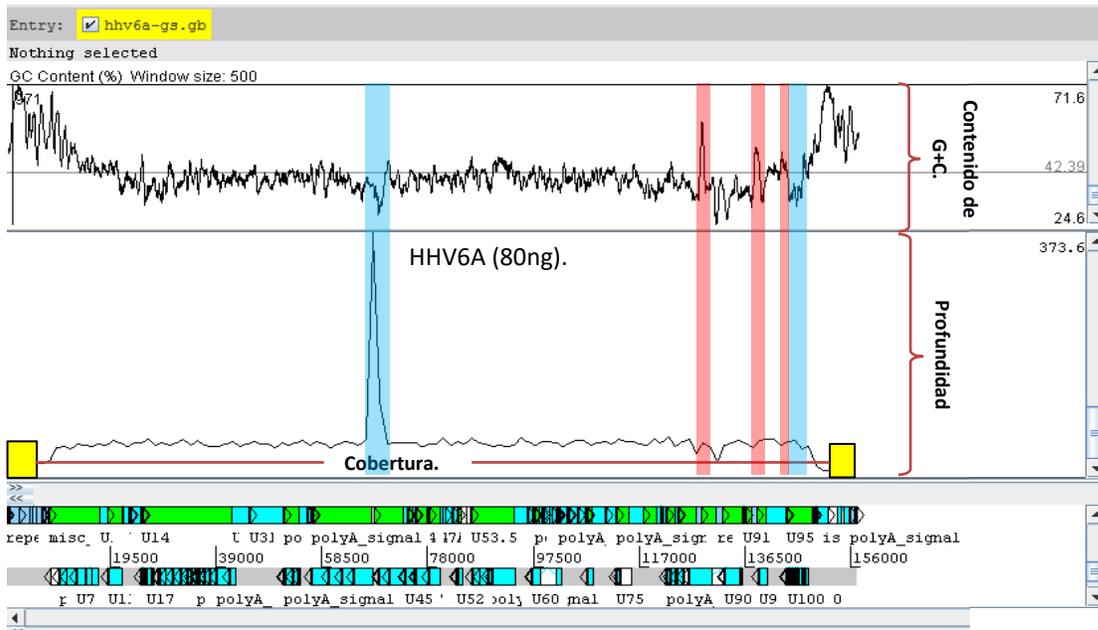


Figura 24. Cobertura y profundidad obtenida en la secuenciación de los cultivos de HHV6A, respecto al contenido de G+C del genoma de referencia (Ion Torrent).

Se comparó la alineación del archivo BAM del cultivo de HCMV cepa Merlin contra su respectiva secuencia de referencia, en relación al contenido de G+C. 1) Genoma de referencia (HCMV cepa Merlin). 2) Cobertura y profundidad alcanzada por la alineación de las lecturas del archivo BAM. 3) Contenido de G+C de la secuencia de referencia. Las áreas en azul corresponden a zonas con un bajo contenido de G+C, mientras que las áreas en rojo a zonas con un contenido de G+C alto. En amarillo se señalan las repeticiones terminales que no se tomaron en cuenta para el análisis.

9.8.4 Librerías de secuenciación para MiSeq.

Para construir las librerías de secuenciación para la plataforma MiSeq, se empleó el kit Nextera XT DNA Library Prep y siguiendo los protocolos de fragmentación, etiquetado (inserción de los índices), PCR de amplificación y limpieza de la PCR del mismo. Para este último se emplearon las perlas Agencourt AMPure XP (Beckman Coulter), de acuerdo al protocolo del kit Nextera.

Al finalizar la construcción y amplificación de las librerías se tomaron alícuotas de estas y

se analizaron en un chip High Sensitivity DNA (Agilent Technologies). El electroferograma (Figura 25) mostró que la selección de los fragmentos no fue precisa, ya que se observa una amplia distribución en el tamaño de los fragmentos purificados (fragmentos de 200pb a más de 1000pb), a pesar de ello una amplia proporción de los fragmentos purificados se encontraban dentro del tamaño deseado por lo que se siguió adelante con la secuenciación.

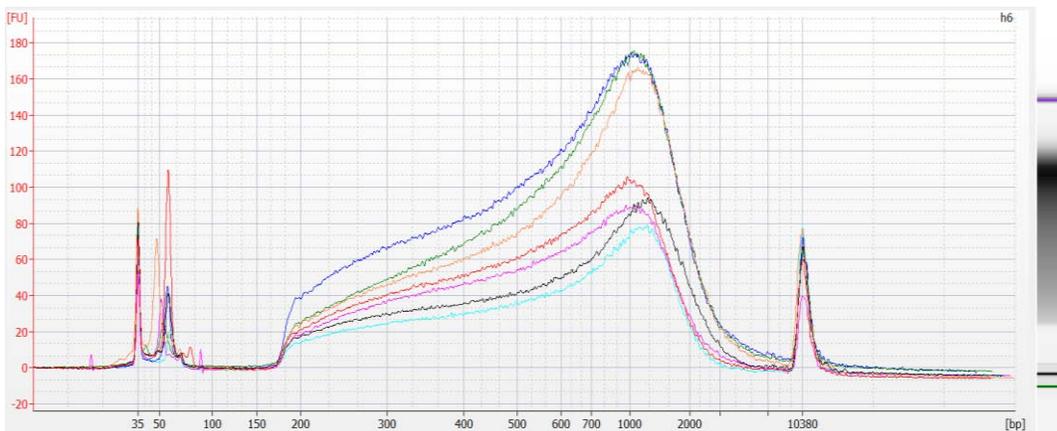


Figura 25. Electroferograma de un chip DNA High Sensitivity de las librerías post-limpieza de productos de PCR.

Limpieza de la PCR de amplificación y selección de los fragmentos con el tamaño deseado (300pb-500pb). Código de colores de las muestras clínicas y de cultivo: M1 (E0804), M2 (E0811), M3 (E0955), M4 (E1023), M5 (E1206), H5 (HCMV-Merlín), H6 (HHV6A-GS).

9.8.5 Secuenciación de los cultivos virales en plataforma MiSeq.

Al finalizar la secuenciación de los cultivos virales no fue posible realiza un pre-análisis de las lecturas obtenidas como en el Ion Torrent, por lo que los archivos FASTQ de secuenciación fueron extraídos y analizados con herramientas especializadas desde el principio (todo el procesamiento se realizó en la plataforma en línea Galaxy). Primeramente se alinearon los archivos FASTQ contra la secuencia de referencia del

genoma humano empleando BWA con la finalidad de eliminar todas las lecturas que pudieran pertenecer a alinear a este. Posteriormente se filtraron las lecturas de acuerdo al tamaño y calidad deseada, para ello se utilizó la herramienta Filter FASTQ de Galaxy (Blankenberg et al., 2010) indicando que se trataba de un archivo pair-end y estableciendo como calidad mínima de filtrado Q20 y un tamaño de lectura mínimo de 90pb, posteriormente se extrajeron las lecturas no alineadas del archivo BAM, y se verificó el perfil de calidad de cada archivo forward y reverse (Figura 26 - Figura 29) con la herramienta FastQC de BioLinux (Andrews, 2010). En todos los casos la calidad de las lecturas MiSeq superó a las obtenidas en Ion Torrent (Tabla 16). Aunado a esto se determinó un aproximado de la profundidad de mapeo mediante la herramienta coverage de samtools, obteniendo una profundidad promedio de 196 lecturas para el cultivo de HCMV cepa Merlin y 295 lecturas de profundidad para el cultivo de HHV6 cepa GS.

Tabla 16. Numero de lecturas obtenidas de los cultivos virales de acuerdo a la plataforma utilizada.

Lecturas de los cultivos virales obtenidas.		
Cultivo	Lecturas Alineadas del MiSeq	Lecturas Alineadas del Ion Torrent
HCMV cepa Merlin	518,602	115,011
HHV6A cepa GS	287,348	28,758

Después de verificar la calidad de las lecturas restantes, las lecturas fueron alineadas contra sus respectivas secuencias de referencia, generando un archivo BAM para cada cultivo secuenciado, estos fueron analizados en el visualizador en Artemis (Figura 30. Cobertura y profundidad del cultivo de HCMV, respecto al contenido de G+C (MiSeq).

Se visualizó la cobertura y profundidad de las lecturas del cultivo de HCMV en la plataforma MiSeq, respecto al contenido de G+C. Las regiones en azul presentan un contenido de G+C bajo, mientras que las de rojo presentan un contenido de G+C alto. En rojo áreas con bajo contenido de G+C. En azul áreas con alto contenido de G+C. y Figura 31). En el caso de la secuenciación por MiSeq no se observó una relación entre contenido de G+C y las coberturas de mapeo obtenidas como la observada en Ion Torrent.

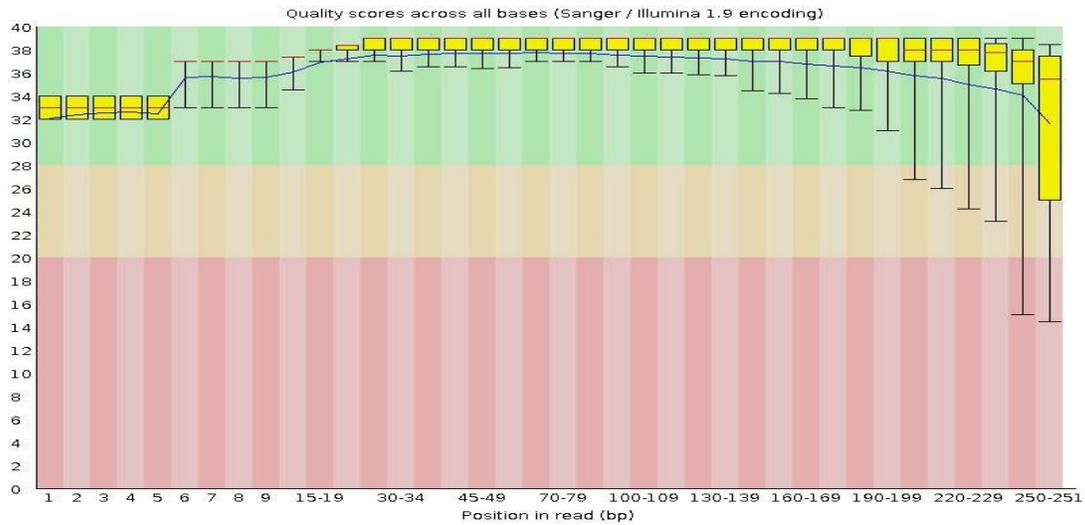


Figura 26. Calidad de las lecturas R1 de la librería de HHV6A cepa GS (lecturas forward). Calidad phred promedio de cada base de acuerdo a su posición en las lecturas forward de la librería de HHV6A cepa GS.

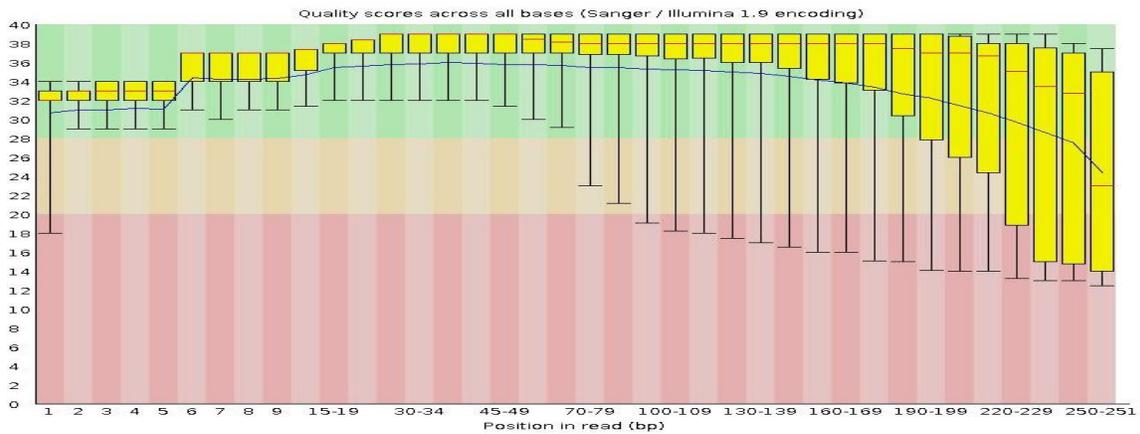


Figura 27. Calidad de las lecturas R2 de la librería HHV6A cepa GS (lecturas reverse). Calidad phred promedio de cada base de acuerdo a su posición en las lecturas reverse de la librería de HHV6A cepa GS.

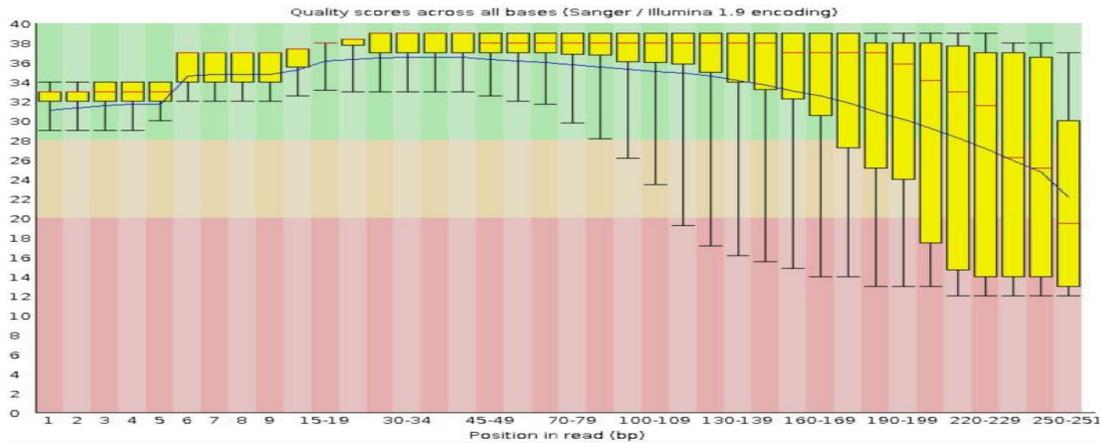


Figura 28. Calidad de las lecturas R1 de la librería de HCMV cepa Merlin (lecturas forward). Calidad phred promedio de cada base de acuerdo a su posición en las lecturas forward de la librería de HCMV cepa Merlin.

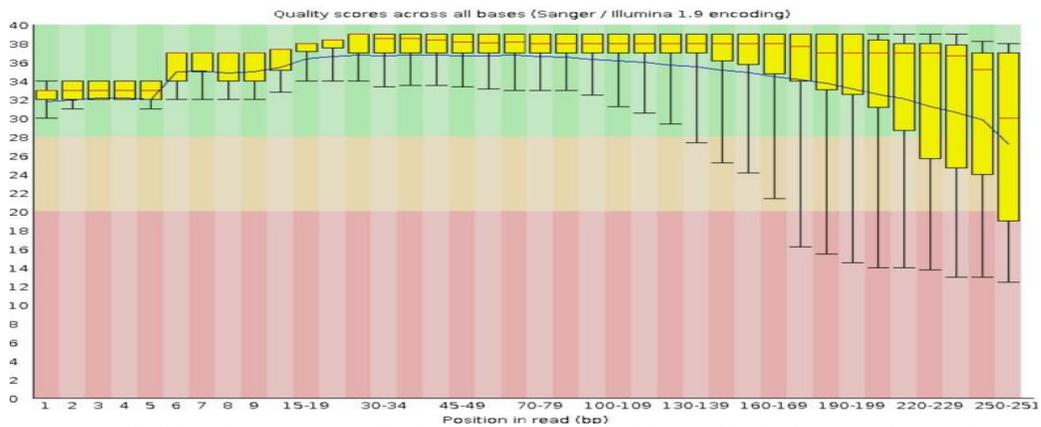


Figura 29. Calidad de las lecturas R2 de la librería de HCMV cepa Merlin (lecturas reverse). Calidad phred promedio de las bases de acuerdo a su posición en las lecturas reverse de la librería de HCMV cepa Merlin.

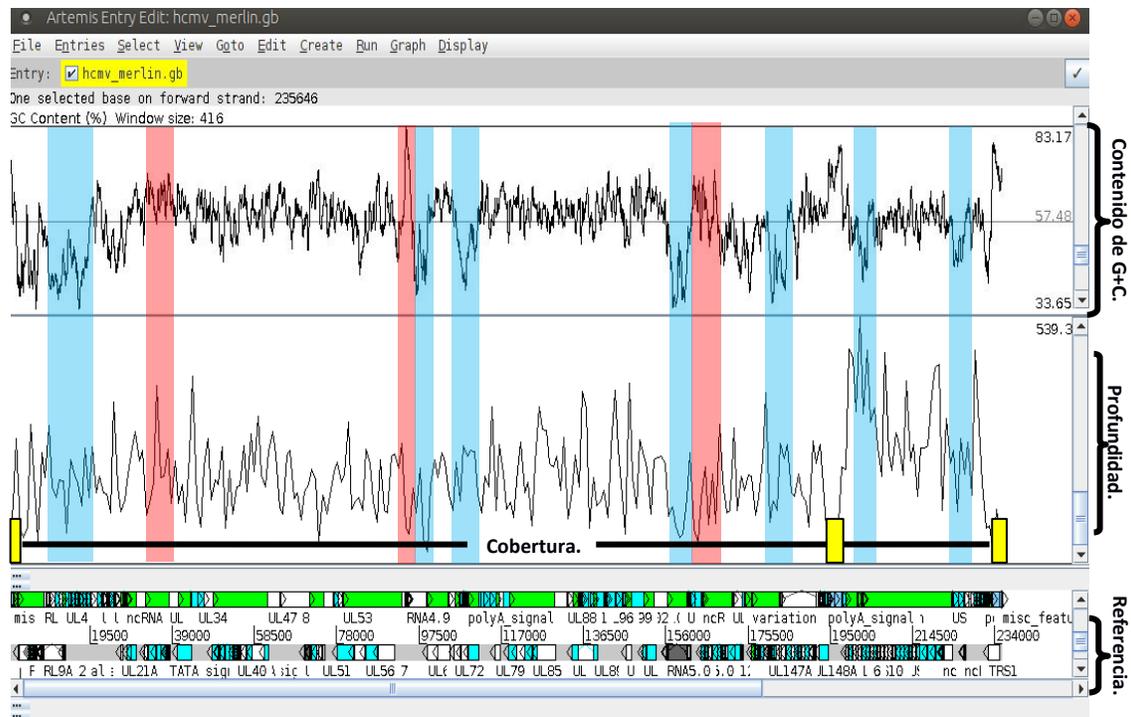


Figura 30. Cobertura y profundidad del cultivo de HCMV, respecto al contenido de G+C (MiSeq).

Se visualizó la cobertura y profundidad de las lecturas del cultivo de HCMV en la plataforma MiSeq, respecto al contenido de G+C. Las regiones en azul presentan un contenido de G+C bajo, mientras que las de rojo presentan un contenido de G+C alto. En rojo áreas con bajo contenido de G+C. En azul áreas con alto contenido de G+C.

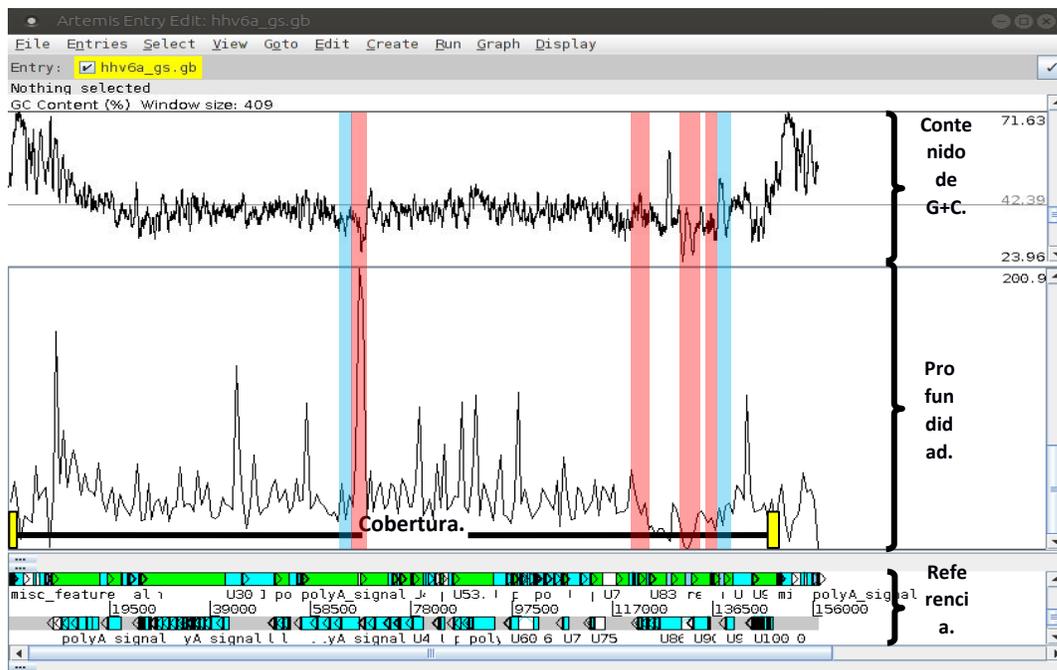


Figura 31. Cobertura y profundidad del cultivo de HHV6A, respecto al contenido de G+C (MiSeq).

Se analizó la cobertura y profundidad obtenidas por las lecturas del cultivo de HHV6A en le MiSeq, respecto al contenido de G+C. Las regiones en azul presentan un contenido de G+C bajo, mientras que las de rojo presentan un contenido de G+C alto.

9.9 Análisis de polimorfismos.

Para observar los polimorfismos de los cultivos virales se emplearon 2 visualizadores de forma simultánea: Artemis e IGV. Se construyeron archivos VCF con las herramientas MPileup y BCFtools (Li, 2011; Minevich et al., 2012) de la plataforma Galaxy con una calidad de mapeo $\geq Q20$ como valor de corte para definir los posibles polimorfismos de entre las lecturas existentes. Con el fin de corroborar la existencia de cada SNP, estos se verificaron individualmente en cada visualizador advirtiendo que fueran visibles en ambos. Es importante señalar que las lecturas correspondientes al polimorfismo tenían que presentar una proporción mayor o igual al 20% de las lecturas totales en el sitio en cuestión, para ser considerados polimorfismos verdaderos y presentados en los visualizadores.

cultivo de HCMV es el de la posición 11573, ya que produce un codón de paro prematuro en el gen RL13.

Tabla 17. Comparación de los polimorfismos encontrados por el Ion Torrent y MiSeq en el cultivo de HCMV cepa Merlin.

En color negro se muestran los porcentajes correspondientes al nucleótido presente en la secuencia de referencia en NCBI, y en color azul los porcentajes correspondientes a las posibles variantes. El asterisco (*) muestra un posible polimorfismo identificado en la plataforma Ion Torrent mediante los archivos vcf, cuya proporción de lecturas polimórficas era inferior al 20% de las lecturas totales.

Frecuencia de bases en sitios polimórficos encontrados en el cultivo viral de HCMV cepa Merlin											
Núm	Posición	Plataforma MiSeq				Plataforma Ion Torrent				Cambio	Gen
		A	C	G	T	A	C	G	T		
1	11563	6%	34%	60%	0%	5%	12%	83%	0%	Trp -> Cys (G->C)	RL13
2	11573	0%	81%	0%	19%	0%	66%	0%	34%	Gln -> codón paro (C->T)	RL13
3	*30113	0%	81%	0%	19%	0%	82%	0%	18%	Arg -> His (C->T)	UL24
4	40231	0%	83%	17%	0%	0%	67%	33%	0%	Sustitución sinónima (G->C)	UL32
5	103487	9%	91%	0%	0%	16%	84%	0%	0%	Ala -> Cys (C->A)	UL70
6	164018	72%	0%	28%	0%	100%	0%	0%	0%	Ser -> Gly (A->G)	UL112
7	164021	0%	24%	76%	0%	0%	0%	100%	0%	Gly -> Arg (G->C)	UL112
8	164024	0%	73%	3%	24%	0%	100%	0%	0%	Arg -> Cys (C->T)	UL112

9.9.2 Polimorfismos encontrados en el cultivo de HHV6A cepa GS.

De forma similar, se comparó la secuencia genómica obtenida por el cultivo viral de HHV6A cepa GS con su respectiva secuencia de referencia. En este caso la cantidad de polimorfismos detectados fue mucho mayor a la de HCMV, sin embargo a diferencia de HCMV la mayoría de los polimorfismos encontrados se encuentran en regiones repetitivas internas y no codificantes, además de las múltiples sustituciones sinónimas.

Tabla 18. Comparación de los Polimorfismos encontrados por el Ion Torrent y MiSeq en el cultivo de HHV6A cepa GS.

En color negro se observan los porcentajes correspondientes a los nucleótidos presente en la secuencia de referencia, y en color azul los porcentajes correspondientes a variantes. Los asteriscos (*) señalan los polimorfismos ubicados en un locus con secuencias repetitivas.

Frecuencia de bases en sitios polimórficos encontrados en el cultivo viral de HHV6A cepa GS											
Núm	Posición	Plataforma MiSeq				Plataforma Ion Torrent				Mutación	Gen
		A	C	G	T	A	C	G	T		
1	30139	0%	100%	0%	0%	0%	100%	0%	0%	Lys -> Glu (T->C)	U19
2	32204	21%	0%	79%	0%	21%	0%	79%	0%	Sustitución sinónima (G->A)	U20
3	56376	0%	100%	0%	0%	0%	100%	0%	0%	Glu -> Gln (G->C)	U37
4	65808	0%	3%	0%	97%	4%	4%	0%	92%	Ala -> Thr (C->T)	U41
5	67896	0%	18%	0%	82%	0%	21%	0%	79%	(C->T)	no codifica
6	68289-68326	Delección de 37 bases								Delección	no codifica
7	81602	33%	0%	67%	0%	26%	0%	71%	3%	Thr -> Ala (A->G)	U50
8	107164	0%	62%	38%	0%	0%	50%	50%	0%	Val -> Leu (G->C)	U71
9	107802	0%	0%	0%	100%	0%	0%	0%	100%	Gly -> Asp (C->T)	U72
10	108884	0%	86%	0%	14%	0%	91%	0%	9%	His -> Tyr (C->T)	U73
11	123014	0%	88%	0%	13%	0%	64%	0%	36%	Val -> Ile (C->T)	U82
12	123058	74%	0%	0%	26%	46%	0%	0%	54%	Ile -> Asn (A->T)	U82
13	124425	0%	69%	0%	31%	0%	48%	0%	52%	Met -> Ile (C->T)	U84
14	128139	25%	0%	75%	0%	25%	0%	75%	0%	Sustitución sinónima (A->G)	U86
15	128145	0%	17%	0%	83%	0%	0%	0%	100%	Sustitución sinónima (C->T)	U86
16	129231	0%	0%	0%	100%	0%	0%	0%	100%	Sustitución sinónima (C->T)	U86
17	133591	0%	65%	0%	35%	0%	89%	0%	11%	Cys -> Codon stop (A->T)	U90
18	135595	0%	100%	0%	0%	0%	78%	0%	22%	Asp -> Asn (C->T)	U90
19	135906	0%	0%	0%	100%	0%	0%	0%	100%	(A->T)	no codifica
20	*137663	47%	0%	53%	0%	4%	0%	96%	0%	(G->A) región repeat	no codifica
21	*137729	0%	19%	0%	81%	3%	79%	0%	18%	(C->T) región repeat, no codifica	no codifica
22	*137730	23%	0%	77%	0%	82%	0%	18%	0%	(A->G) región repeat	no codifica
23	*137782	0%	85%	0%	15%	0%	21%	0%	79%	(T->C) región repeat	no codifica
24	*137819	0%	62%	38%	0%	0%	11%	89%	0%	(G->C) región repeat	no codifica

25	*137829	53%	47%	0%	0%	a10%	90%	0%	0%	(C->A) región repeat	no codifica
26	*137833	0%	14%	0%	86%	0%	42%	3%	55%	(C->T) región repeat	no codifica
27	*137887	0%	81%	0%	19%	0%	45%	0%	55%	(T->C) región repeat	no codifica
28	*137892	71%	0%	0%	29%	68%	2%	0%	30%	(G->C) región repeat	no codifica
29	*137924	0%	59%	41%	0%	0%	25%	76%	0%	(C->A) región repeat	no codifica
30	*137934	71%	29%	0%	0%	25%	73%	2%	0%	(A->G) región repeat	no codifica
31	*138053	41%	0%	59%	0%	92%	0%	8%	0%	(A->G) región repeat	no codifica
32	*138158	56%	0%	44%	0%	98%	0%	2%	0%	(T->C) región repeat	no codifica
33	*138331	0%	32%	0%	68%	0%	14%	0%	86%	(C->A) región repeat	no codifica
34	*138352	33%	67%	0%	0%	10%	90%	0%	0%	(A->T) región repeat	no codifica
35	142413	0%	0%	0%	100%	0%	0%	0%	100%	(T->C)	no codifica
36	144039	0%	100%	0%	0%	0%	100%	0%	0%	(T->C)	no codifica
37	144249	0%	100%	0%	0%	0%	100%	0%	0%	Met -> Val (T-C)	U100
38	144476	0%	100%	0%	0%	0%	100%	0%	0%	Lys -> Glu (T->C)	U100
39	145928	0%	100%	0%	0%	0%	100%	0%	0%	Sustitución sinónima (T->C)	U100
40	148674	0%	100%	0%	0%	0%	100%	0%	0%	(G->C)	no codifica
41	148694	0%	100%	0%	0%	0%	100%	0%	0%	(C->T) región repeat	no codifica

9.9.3 Confirmación de polimorfismos a través de secuenciación Sanger.

Con el fin de corroborar la existencia de algunos polimorfismos se diseñaron 5 pares de primers (4 para polimorfismos de HCMV y 1 para HHV6A) y se realizó la secuenciación Sanger de estos, utilizando el mismo material genético que también se usó para las librerías de secuenciación. Mediante PCR punto final y purificación del producto de PCR se realizó la reacción de Sanger por métodos convencionales. Se verificó la existencia de todos los polimorfismos que fueron secuenciados por el método Sanger (Figura 34). Cabe señalar que en imagen D de la Figura 34 solo se observa una sola curva en los 3 polimorfismos señalados, debido a que estos han sustituido completamente (al 100%) a las bases que se encontraban originalmente de acuerdo a la secuencia de referencia.

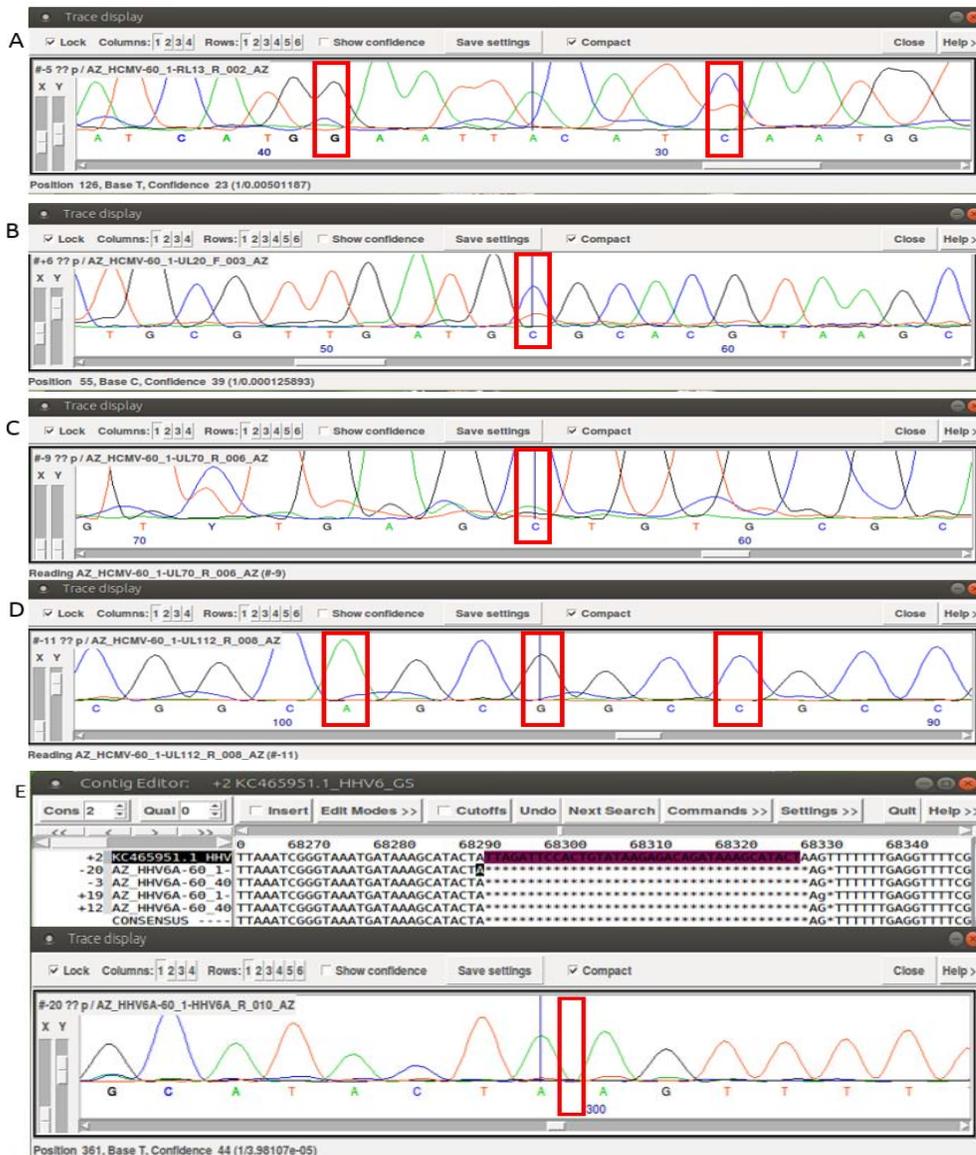


Figura 34. Polimorfismos secuenciados vía Sanger.

Se corroboró la existencia de polimorfismos en cultivo de HCMV, y una delección en HHV6. A) Corresponde a los polimorfismos del gen RL13 del cultivo de HCMV, posiciones 11563 (G→C) y 11573 (C→T) respectivamente, esta ultima genera un codón stop. B) Polimorfismo del gen UL24 del cultivo de HCMV, posición 30113 (C→T). C) Polimorfismo del gen UL70 del cultivo de HCMV, posición 103487 (C→A). D) Polimorfismos del gen UL112 del cultivo de HCMV, posiciones 164018 (A→G), 164021 (G→C) y 164024 (C→T). E) Delección de 37pb del cultivo de HHV6A, de la posición 68289-68326.

9.10 Prevalencia de HCMV y HHV6 en la población mexicana con VIH.

Con la finalidad de estudiar la prevalencia de estos herpesvirus en la población mexicana con VIH y encontrar nuevas muestras para su posible secuenciación, se seleccionaron 270 muestras clínicas pertenecientes al protocolo E10-10 que corresponde a pacientes con VIH con una carga viral >10,000 copias de VIH/mL y vírgenes al tratamiento antirretroviral. Tales muestras se colectaron entre los años 2013-2014. Cuando los pacientes ingresan a este protocolo se extrae rutinariamente el DNA de células mononucleares de sangre periférica y se almacena a -80°C para diversos ensayos genómicos. A ese material se realizó la detección de herpesvirus mediante el ensayo Herplex para los betaherpesvirus humanos únicamente. En los ensayos se observó la curva de amplificación y la curva de disociación (HRM).

Se identificaron las muestras con la mayor carga viral aparente (Figura 35) y se visualizó la curva HRM de los controles positivos que se comparó con la de cada muestra con el fin de determinar el o los herpesvirus presentes en estas (Figura 36). Se observó que el HRM del control negativo presenta una curva de gran tamaño, sin embargo el perfil de esta curva es irregular y su temperatura de disociación es mucho menor al de los controles positivos, lo cual es indicativo de producto inespecífico o dimerización de primers.

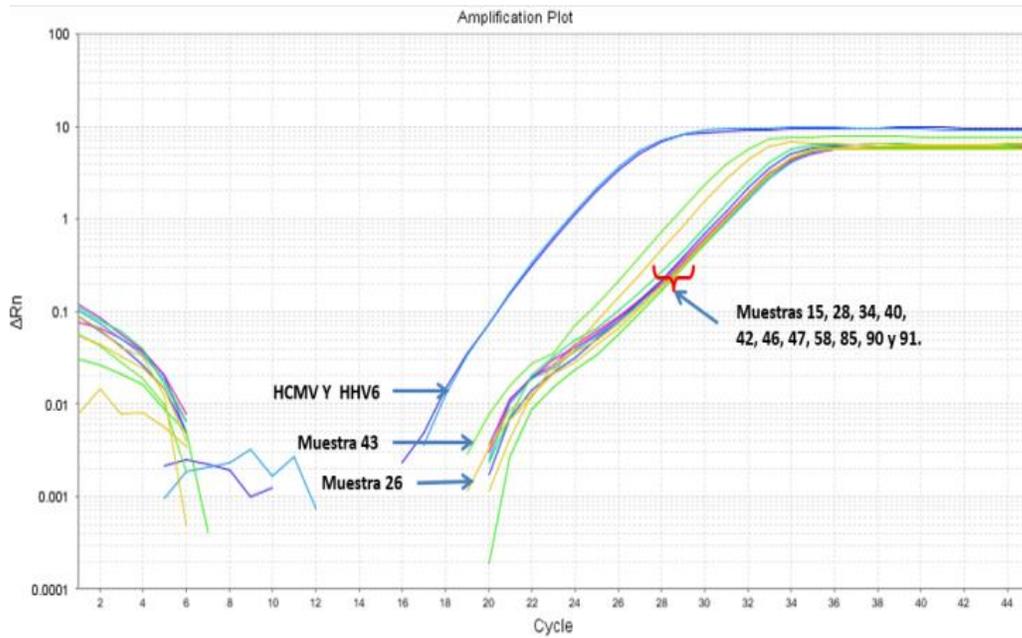


Figura 35. qPCR de los betaherpesvirus presentes en las muestras clínicas del protocolo E10-10.

Grafica de amplificación de los controles positivos (HCMV y HHV6) y de las muestra clínicas con la mayor carga viral.

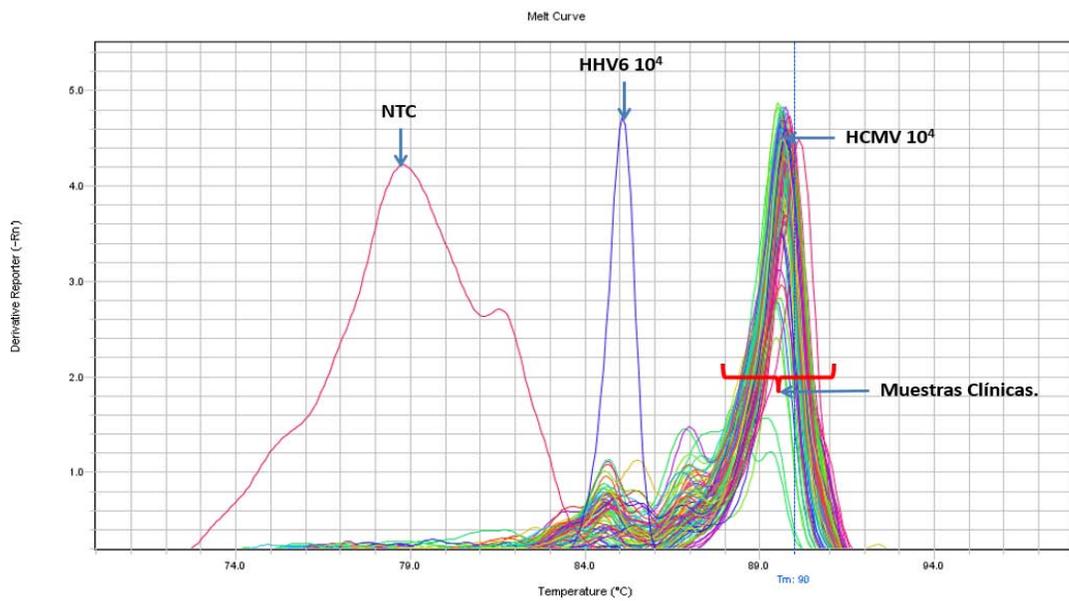


Figura 36. HRM para la identificación de los virus presentes.

HRM de los controles positivos (HHV5 y HHV6) y de las muestra clínicas.

Se determinó la prevalencia del DNA de HCMV y HHV6 presentes en las muestras analizadas observando que HCMV presenta una prevalencia del 99% en los pacientes VIH positivo, mientras que para HHV6 es del 68%, mientras que ninguna muestra fue positiva a HHV7.

Lotes	Porcentaje de muestras con HCMV	Porcentaje de muestras con HHV6
A	97.72%	50%
B	100%	72.22
C	100%	84.61%

Tabla 19: Prevalencia del DNA de HCMV y HHV6 en muestras clínicas

9.11 Secuenciación y análisis de muestras clínicas.

Para evitar los sesgos inherentes a los cultivos y los cambios en el genoma viral, se extrajo el DNA directamente de las muestras clínicas y el procedimiento de secuenciación se realizó de forma directa al DNA extraído.

9.11.1 Muestras clínicas secuenciadas en el Ion Torrent.

En primera instancia se construyeron las librerías de 2 muestra clínicas (CAS y RCMV107s) diagnosticadas con HCMV, para posteriormente realizar la secuenciación de estas. Apenas se logró el cargado de más de la mitad de los pozos (Figura 37) y poco más de 1.45 millones de lecturas con carácter informativo (Figura 38). El tamaño de los fragmentos obtenidos fue el adecuado (200-260pb) (Figura 39), con lecturas de buena calidad hasta la posición 150 para ambas librerías (solo se muestra las calidades de la librería de CAS, Figura 39).

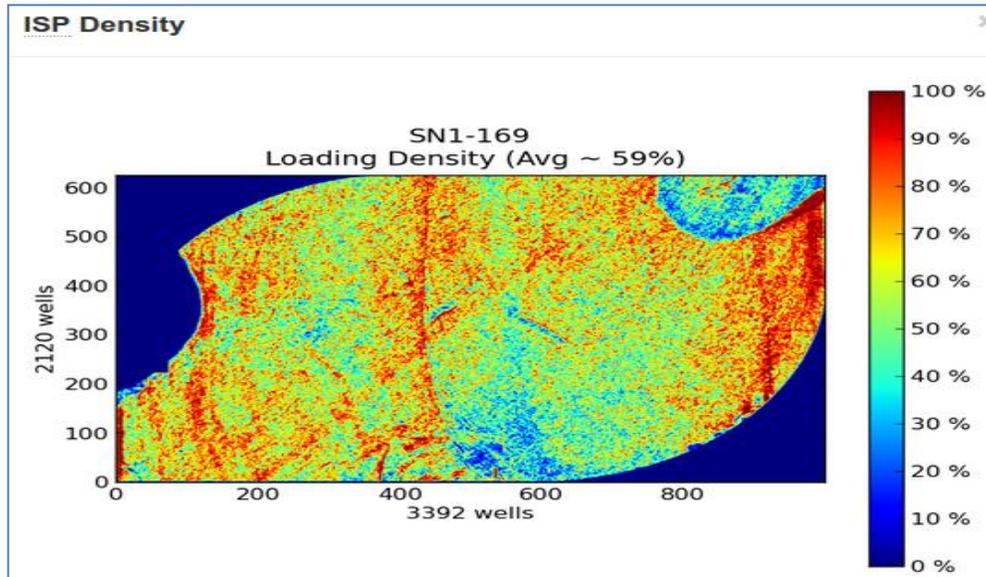


Figura 37. Mapa de calor de chip 316, con las muestras clínicas CAS y RCMV017s. Mapa de calor del chip 316, cargado con las librerías de las muestras clínicas CAS y RCMV017s, el mapa nos muestra la densidad y porcentaje de pozos cargados en el chip.

ISP Summary

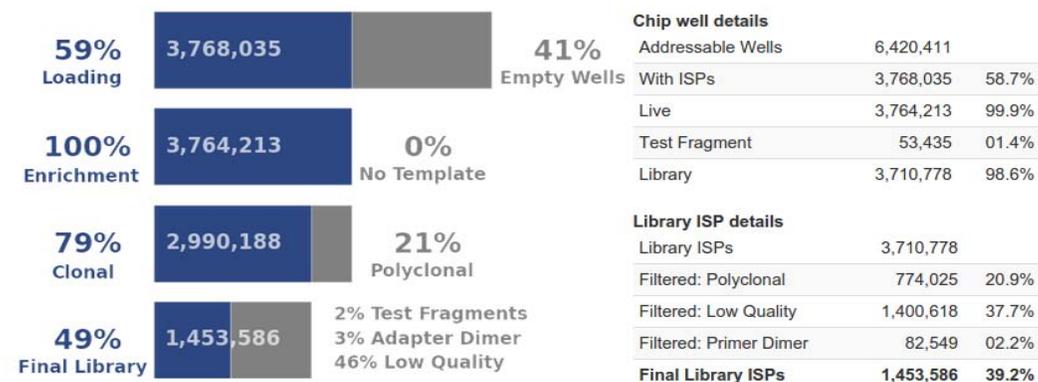


Figura 38. Resumen de las partículas cargadas. Muestra el total de lecturas obtenidas con el 59% del chip cargado, así como el porcentaje de las partículas con templado que fueron enriquecidas, la cantidad de partículas policlonales y las partículas que finalmente fueron de carácter informativo.

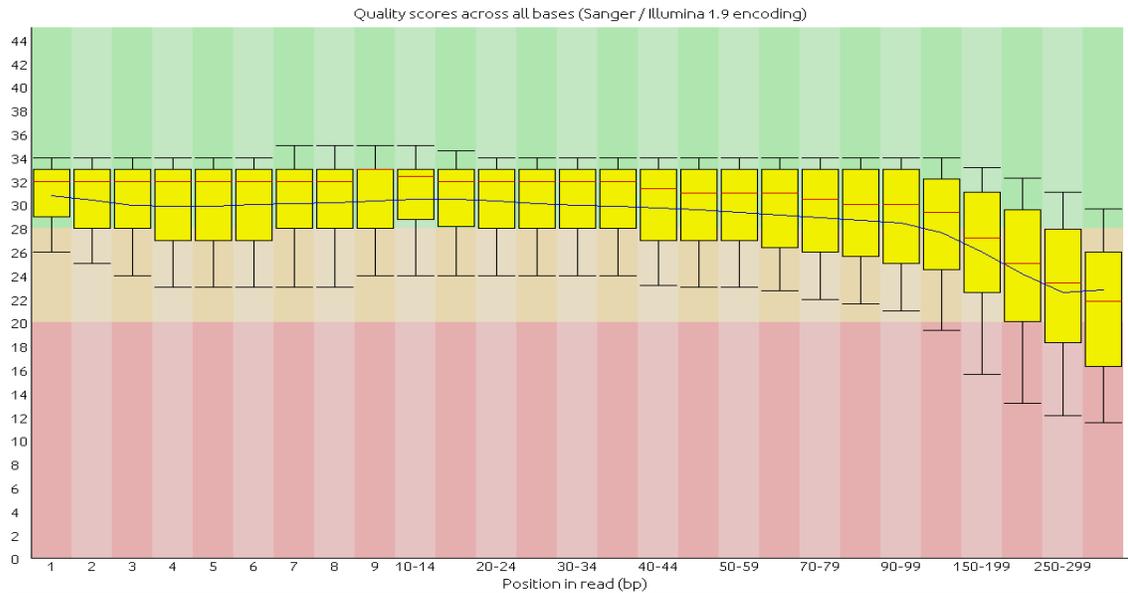


Figura 39. Perfil de calidad de las lecturas de la librería de CAS.

Calidad promedio de cada base de acuerdo a su posición en las lecturas de la librería de CAS.

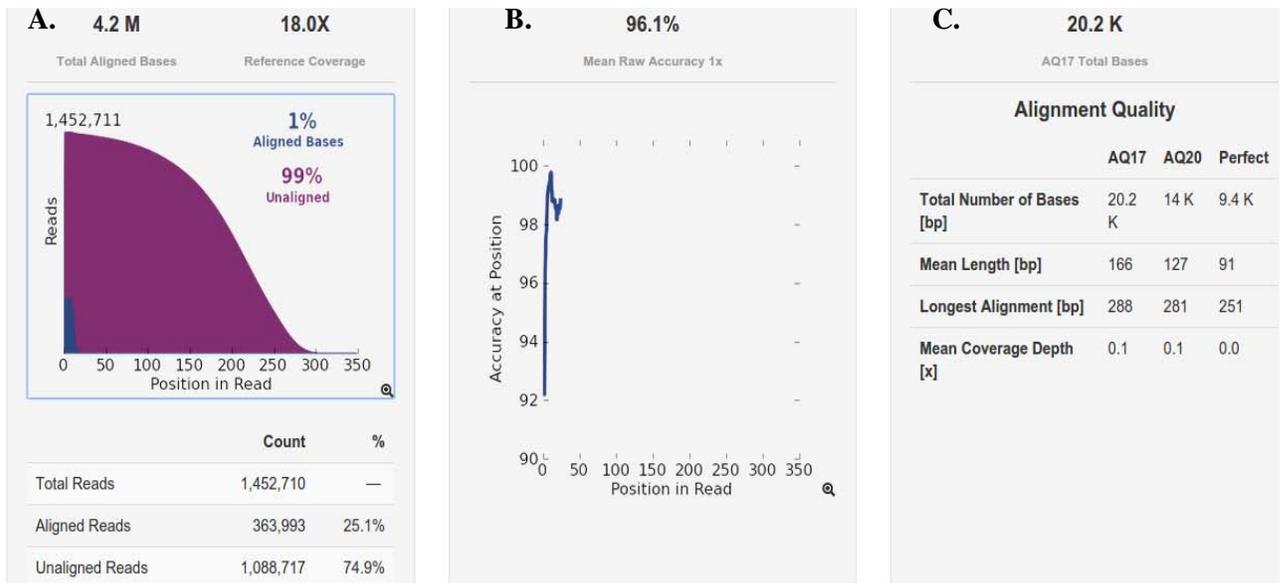


Figura 40. Alineación de las lecturas de CAS y RCMV107s contra la secuencia de referencia de HCMV.

A). El porcentaje de lecturas alineadas contra la secuencia de referencia fue mínimo; B). Las lecturas alineadas a la secuencia de referencia son muy pequeñas; C). A pesar que las lecturas alineadas contra la secuencia de referencia presenta una buena calidad, se obtuvo una cobertura y profundidad mapeo mínimo.

A pesar de la buena secuenciación realizada, una mínima cantidad de lecturas fueron mapeadas contra la referencia (Figura 40, inciso A), pero casi todas las secuencias alineadas eran menores a 50 pb (Figura 40, inciso B) por lo que posiblemente correspondían a fragmentos inespecíficos.

Con el fin de realizar un análisis más riguroso, se descargaron los archivos FASTQ de secuenciación y se re-analizaron con programas especializados montados en un servidor del laboratorio con el sistema operativo BioLinux (Field et al., 2006). Primero se retiraron las lecturas pertenecientes a humano, para ello se alinearon las lecturas de los archivos FASTQ contra el genoma de referencia Hg19 de humano (ID GenBank: NC_000001.11) utilizando BWA. Posteriormente se utilizó el programa Samtools (Li, 2011) para extraer las lecturas no alineadas contra humano, aunado a esto las lecturas no alineadas fueron filtradas de acuerdo al tamaño y calidad deseadas (calidad mínima de Q20 y tamaño mínimo de 50 pb), para ello se empleó la herramienta Filter FASTQ de la plataforma Galaxy. Finalmente las lecturas de los archivos FASTQ filtrados fueron alineados contra sus respectivas secuencias de referencia (Merlin de HCMV y GS de HHV6A). En los archivos BAM resultantes (Figura 41) se observan múltiples lecturas de la librería de CAS alineadas a la referencia, no obstante estas se encuentran muy dispersas y la profundidad alcanzada es mínima. Por otro lado solo se alinearon 3 lecturas correspondientes a RCMV107s.

Debido a la pequeña cantidad de lecturas alineadas no fue posible realizar el análisis de la cobertura y profundidad de secuenciación respecto al contenido de G+C, ni buscar posibles polimorfismos. Esfuerzos subsecuentes con otras muestras arrojaron resultados similares, sugiriendo que la calidad del material es inadecuada para el estudio o que posiblemente sea necesario utilizar otros protocolos de secuenciación.

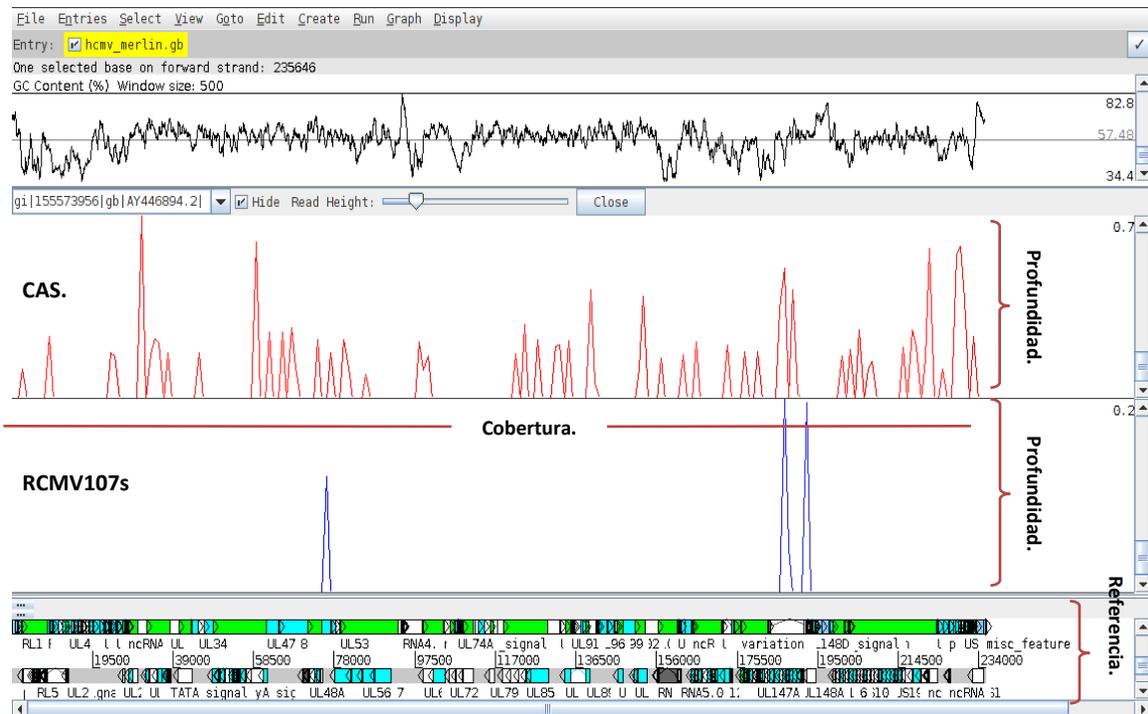


Figura 41. Lecturas de las muestras clínicas CAS y RCMV107s alineadas a la secuencia de referencia. Lecturas de las muestras clínicas alineadas contra la secuencia de referencia de HCMV cepa Merlin.

9.11.2 Muestras clínicas secuenciadas en el MiSeq.

Con el fin de comparar la capacidad de secuenciación del MiSeq con la del Ion Torrent y obtener la secuencia genómica de un herpesvirus local, se secuenciaron varias muestras clínicas en la plataforma MiSeq, sin embargo a pesar de la mayor capacidad de la celda de secuenciación del MiSeq, y que además se agregaron muestras clínicas colectadas más recientes (2013-2014) a las usadas en el Ion Torrent, los resultados obtenidos en todos los casos fueron similares a los observado en a la secuenciación de las muestras clínicas con el Ion Torrent, obteniendo lecturas dispersas a lo largo de los genomas de referencia con una profundidad mínima (Tabla 20). Es relevante señalar que la muestra E0804 correspondía a una muestra positiva para HCMV únicamente por qPCR, sin embargo, se identificaron múltiples lecturas que correspondían a HHV6.

Tabla 20. Lecturas mapeadas en las muestras clínicas secuenciadas.

Numero de lecturas (de los virus de interés) de las muestras clínicas mapeadas contras las secuencias de referencia, de acuerdo a la plataforma de secuenciación utilizada. NA: No aplica para esta muestra.

Muestras Clínicas	Herpesvirus presente	Lecturas de HCMV	Lecturas de HHV6	Plataforma utilizada.
RCMV107s	HCMV	3	NA	Ion Torrent
CAS	HCMV	74	NA	Ion Torrent
E1090	HCMV	0	NA	Ion Torrent
E1339	HCMV y HHV6	0	0	Ion Torrent
E1628	HCMV	0	NA	Ion Torrent
E0804	HCMV y HHV6	806	542	MiSeq
E0811	HCMV	388	NA	MiSeq
E0955	HCMV y HHV6	90	2	MiSeq
E1023	HCMV y HHV6	392	4	MiSeq
E1206	HCMV y HHV6	90	2	MiSeq

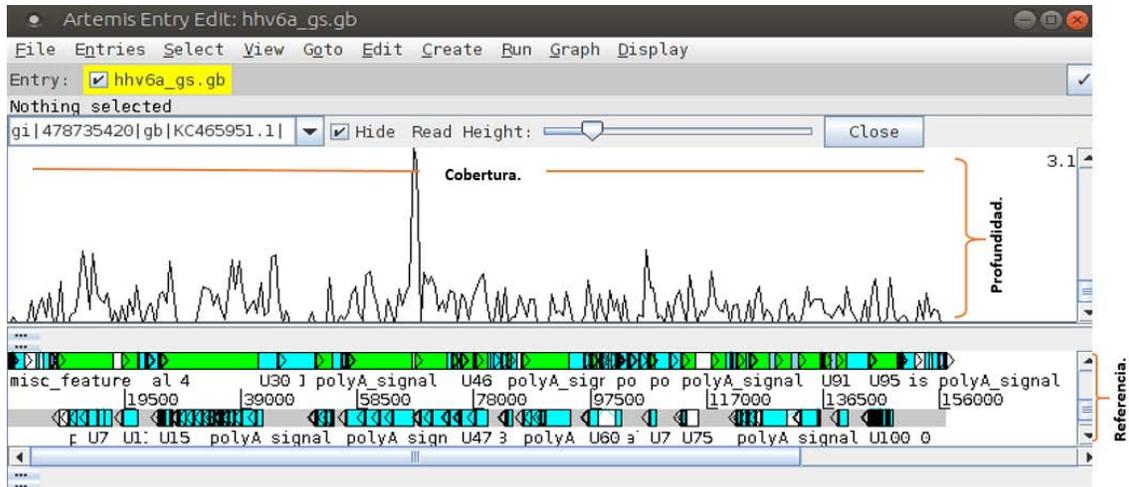


Tabla 21. Lecturas de la muestra clínica E0804 mapeadas contra la secuencia de referencia de HHV6A cepa GS.

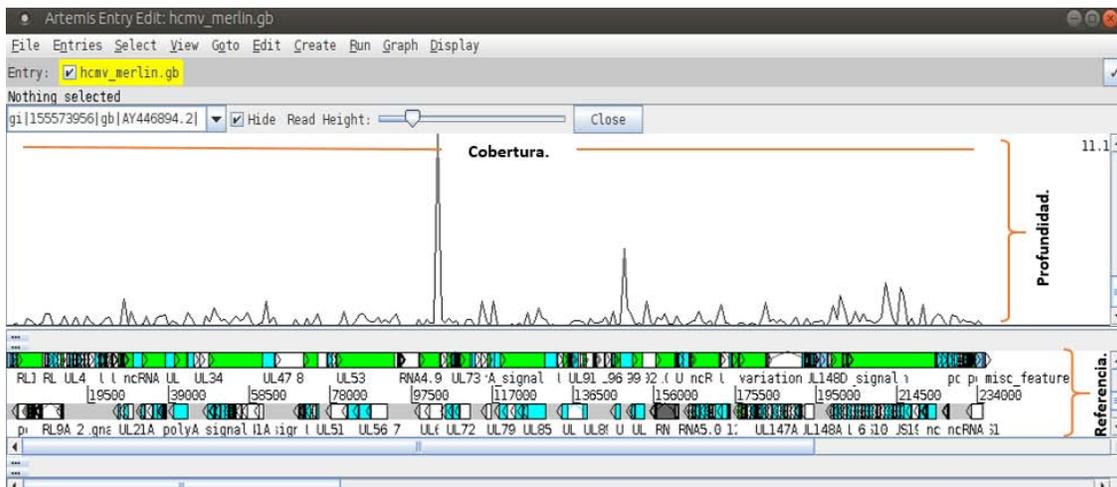


Tabla 22. Lecturas de la muestra clínica E0804 mapeadas contra la secuencia de referencia de HCMV cepa Merlin.

10. DISCUSIÓN.

10.1 Cultivos virales.

La cuantificación viral de los cultivos se estimó por medio de la carga viral y no por titulación. En una primera instancia por la sensibilidad del método y por fines prácticos, ya que se contaba con controles positivos cuyas cargas virales eran conocidas y el método de qPCR para cuantificar la carga viral de estos se tenía bien estandarizado. Por otro lado no hubiera sido posible cuantificar la carga viral de las muestras clínicas por el mismo método (por titulación) de ser el caso, ya que no se contaba con material clínico para realizar este ensayo, este había sido utilizado completamente en la extracción de DNA.

10.2 Fragmentación por sonicación vs. Fragmentación enzimática.

Al comparar dos métodos de fragmentación se constataron ventajas y desventajas de estos.

La fragmentación por sonicación nos ofrece la posibilidad de obtener fragmentos con un tamaño específico al seleccionar y purificar los fragmentos con el tamaño deseado a partir del DNA total fragmentado, evitando la obtención de fragmentos que sería perjudiciales durante para la secuenciación. Sin embargo, durante la purificación de los fragmentos de tamaño deseado es posible perder una cantidad importante de material, por lo que el método requiere partir de la suficiente cantidad de DNA (mayor a la de la fragmentación enzimática) para compensar el DNA perdido durante la selección de los fragmentos de tamaño deseado.

Debido a que la capacidad para muestras del sonicador es limitada (a una sola muestra), esto conlleva un incremento en el tiempo de construcción de las librerías dependiendo del número de muestras, siendo el tiempo de fragmentación de cada muestra de 5 minutos aproximadamente. Aunado a esto, los fragmentos con el tamaño de interés deben ser purificados (además de verificar la correcta selección de los fragmentos), y posteriormente

finalizar la construcción de las librerías (unir los adaptadores y códigos de barras a los fragmento de DNA) incrementando sustancialmente el tiempo de construcción en varias horas o días (de acuerdo las condiciones de trabajo).

Por otro lado, una de las ventajas que nos ofrece la fragmentación enzimática (en el protocolo del Ion Torrent) es el poco tiempo (3 horas aproximadamente) en el que se puede obtener la construcción de las librerías debido a su automatización, así como el número de muestras (15 muestras) que pueden procesarse simultáneamente. Además este método nos generó una menor distribución en el tamaño de los fragmentos, lo que se traduce en un mayor rendimiento del DNA empleado para la construcción de las librerías.

10.3 Secuenciación en plataformas Ion Torrent y MiSeq.

10.3.1 Profundidad y cobertura obtenida en la secuenciación de las regiones únicas de los cultivos virales respecto al contenido de G+C.

La cobertura de los genomas de referencia realizada por las secuencias obtenidas de los cultivos virales fue similar entre ambas plataformas (Ion Torrent y MiSeq), por otro lado ambas plataformas presentaron problemas para mapear las regiones repetitivas (cabe señalar que estas regiones fueron excluidas del estudio). Por otro lado la profundidad de secuenciación fue diferente entre ambas, observando que la plataforma MiSeq generó una mayor profundidad de mapeo que el Ion Torrent, con una profundidad aproximada de 196 lecturas en promedio para el cultivo de HCMV cepa Merlin y de 295 lecturas para el cultivo de HHV6 cepa GS, mientras que el Ion Torrent presentó una profundidad aproximada de 100 lecturas en promedio para el cultivo de HCMV (Merlin) y poco más de 28 lecturas de profundidad para el cultivo de HHV6A (GS) (cabe señalar que las regiones no mapeadas no son tomadas en cuenta por el algoritmo empleado para determinar la profundidad promedio), esto se debe sobre todo a la mayor capacidad de lecturas de la celda de secuenciación del MiSeq (25Mb) respecto a la capacidad del chip 316 del Ion Torrent (2-3Mb).

Las secuenciaciones de HHV6A de ambas plataformas muestran una región con una profundidad de mapeo sobrerrepresentada (principalmente en el Ion Torrent), dado que esta región comprende un origen de replicación y al momento de cosechar los cultivos estos aún presentaban una alta tasa de replicación, es probable que esta zona se encontrara altamente expresada y que en consecuencia tuviera una elevada tasa de secuenciación.

Al analizar los archivos de secuenciación de algunas muestras cuyos tamaños de fragmentación fueron menores a los deseados (en ambas plataformas), constatamos que la longitud de las lecturas es un factor que influye de forma importante en la calidad y ensamblado de las secuencias: lecturas muy cortas podrían interferir con la cobertura del genoma al ensamblar en múltiples sitios del genoma (incluso en diferentes genomas), además de presentar una baja calidad de mapeo provocando que no pasen los filtros de calidad y que en consecuencia sean descartados, por otro lado las librerías con lecturas de un tamaño mayor al deseado presentarían extremos con un llamado de base (*base calling*) de baja calidad, lo cual es común en cualquier tipo de secuenciación, pero se maximiza en este tipo de librerías.

Al alinear las secuenciaciones de los cultivos virales de ambas plataformas (Ion Torrent y MiSeq) con sus respectivas secuencias de referencia, se observó una relación entre el contenido de G+C de las lecturas del Ion Torrent mapeadas contra sus respectivas secuencias de referencia, notando un incremento en la profundidad de las lecturas mapeadas en las áreas con un menor contenido de G+C, y una disminución en la profundidad de las áreas con un contenido de G+C bajo. Por el contrario, en la plataforma MiSeq no se observó ninguna relación entre la profundidad y cobertura de mapeo respecto al contenido de G+C.

De acuerdo con lo reportado previamente (Quail et al., 2012), al comparar la profundidad y cobertura de las lecturas de secuenciación de varias plataformas NGS (incluidas el Ion

Torrent y MiSeq), esta se reducía en la regiones con bajo contenido de G+C, principalmente en la plataforma Ion Torrent. Contrario a esto, en este trabajo se observó que al secuenciar los cultivos virales, el Ion Torrent no sólo fue capaz de secuenciar las regiones con un bajo contenido de G+C, sino que además estas tuvieron la mayor profundidad, mientras que las regiones que presentaban un contenido de G+C alto obtuvieron un menor profundidad, a pesar de esto las regiones únicas de los genomas de HCMV cepa Merlin y HHV6A cepa GS fueron secuenciadas casi en su totalidad. Es importante tomar en cuenta que tanto los kits de secuenciación como las plataformas NGS son modificadas constantemente con el fin de maximizar la cantidad y calidad de las lecturas obtenidas, y el de mejorar la secuenciación de áreas problemáticas (áreas homopoliméricas, con alto o bajo contenido de G+C, repeticiones, etc.). Además, hay una gran diferencia en el contenido de G+C en los genomas estudiados: el contenido de G+C del cromosoma 11 de *P. falciparum* estudiado por Quail et al., oscila entre el 50% y 0%, mientras que los contenidos de G+C de HCMV cepa Merlin y HHV6A cepa GS van del 85% al 30% y del 62% al 26% respectivamente. El uso de kits de secuenciación más desarrollados y organismos con una menor variación en el contenido de G+C pudo contribuir de forma sustancial a observar una mejor secuenciación en los virus estudiados que en el estudio de inicial (Quail et al., 2012).

10.3.2 Polimorfismos de los cultivos virales.

Con polimorfismo nos referimos a cualquier cambio nucleotídico presente en la secuencias de DNA de una población (en este caso la población viral de nuestros cultivos), respecto a una secuencia de referencia. Por otro lado, considerando que todas las plataformas de secuenciación masiva por amplificación introducen errores en lecturas (artefactos), es importante señalar que para considerar un polimorfismo como “verdadero” utilizamos el criterio de que tal cambio debe estar en el 20% o más de las lecturas totales correspondientes a tal posición.

Al comparar los polimorfismos detectados por ambas plataformas, nos percatamos que tuvimos varias situaciones: polimorfismos que se confirmaban en ambas plataformas y que estaban en proporciones similares y otros en proporciones disímiles.

Algunos polimorfismos fueron confirmados de forma contundente por el Ion Torrent y no en el MiSeq (polimorfismos en las posiciones 11573 y 103487 de HCMV; polimorfismos 108884, 123014 y 135595 de HHV6A), pero que también se encontraban presentes en las lecturas de este, sin embargo la proporción de las lecturas polimórficas para el nucleótido en cuestión era inferior al 20% (19% en algunos casos) requerido por las herramientas utilizadas. De forma similar, existieron polimorfismos en el MiSeq y no en el Ion Torrent (polimorfismos 11563, 164018, 164021 y 164024 de HCMV; polimorfismos 5, 8, 17, 18-36, 40 y 41 de HHV6A) al encontrarse en una proporción menor al 20% en esta última plataforma (Tabla 17 y Tabla 18). Además, se identificó una posición potencialmente polimórfica (posición 30113) en el cultivo de HCMV cepa Merlin, cuyo porcentaje en las lecturas totales fue inferior al 20% (Tabla 17) en ambas plataformas (Ion Torrent y MiSeq) e incluso confirmado mediante secuenciación Sanger (Figura 34B). En primera instancia parecería que con el límite de 20% se asignaría un falso negativo, y esto debe ser evaluado con cuidado pues podría representar la aparición de una nueva variante.

A pesar de que se observa una gran cantidad de polimorfismos en el cultivo de HHV6A (cepa GS), es importante señalar que estos se conforman en su mayoría por sustituciones sinónimas y polimorfismos conservativos, y solo una pequeña proporción de los polimorfismos representan un cambio significativo para el genoma de HHV6A. Por otro lado los polimorfismos detectados en el cultivo de HCMV (cepa Merlin) dictan cambios importantes en este (tal es el caso del polimorfismo 11573 que da origen a un codón de paro), sin embargo existe múltiples reportes que demuestran que el genoma de HCMV sufre importantes cambios desde los primeros pases de cultivo.

La mayor capacidad de la celda de secuenciación MiSeq nos permitió obtener una mayor profundidad de cobertura en ambos cultivos, y en consecuencia fue posible identificar una mayor cantidad de posibles polimorfismos en el MiSeq que en el Ion Torrent, el ejemplo más claro se observa en los 18 polimorfismos observados en las repeticiones internas del genoma de HHV6A (Tabla 18, Polimorfismos: 18-36), estos no fueron observados con el Ion Torrent ya que la región presentaba una baja profundidad de mapeo.

10.4 Análisis de las muestras clínicas de personas con VIH.

10.4.1 Prevalencia de betaherpesvirus (HCMV, HHV6 y HHV7) en la población mexicana con VIH.

De las 270 muestras clínicas con VIH positivo analizadas, el 99% presentaba HCMV y el 71% HHV6, ambas cifras son mucho mayores a la seroprevalencia reportada para cada virus respectivamente, sin embargo esto es factible ya que al ser personas con sistemas inmunes deficientes son más susceptibles a las infecciones que no se presentan de forma recurrente en personas inmunocompetentes.

Todos los individuos infectados con HHV6 se encontraban coinfectados con HCMV, esto se debe posiblemente a la elevada prevalencia de HCMV (casi el 100%) en estos pacientes y no a un evento de asociación.

Aunque este estudio no se enfocó a HHV7, es importante señalar que a pesar de la existencia de reportes en los que se observa una alta seroprevalencia de HHV7 en personas sanas (Rojo Medina et al., 1995), y que en las muestras estudiadas no fue posible detectar individuos infectados con HHV7, a pesar de que estos no presentaban ningún tipo de terapia antiviral al momento de tomar las muestras clínicas estudiadas y de la alta susceptibilidad de estos. Una posible explicación es que la carga viral de HHV7 se encontrara por debajo del umbral de detección de la técnica, por otro lado también es posible que la epidemiología de HHV7 en los individuos con VIH se vea modificada

respecto a las personas sanas, será necesario realizar estudios enfocados a conocer la epidemiología de HHV7 en estos pacientes.

10.4.2 Secuenciación de las muestras clínicas.

Desafortunadamente no fue posible secuenciar el genoma de alguno de los betaherpesvirus procedente de las muestras clínicas debido (sobre todo) a la baja cantidad de lecturas virales obtenidas. El principal problema al que nos enfrentamos fue la baja carga viral de las muestras clínicas en comparación a la de los cultivos virales, recordando cerca del 15-16% de las lecturas obtenidas de los cultivos pertenecían a los virus de interés, mientras que menos del 1% de las lecturas de las muestras clínicas pertenecían a estos. Al no utilizar ningún método de enriquecimiento, la mayor parte del material genómico secuenciado correspondía a DNA humano. Es importante señalar que el uso de la PCR y el cultivo viral para el enriquecimiento de las muestras se descartó por una parte porque estos genomas son muy largos y requerirían decenas de primers y que no se contaba con material cultivable, además que se busca evitar la alteración genética de la población viral.

Debido a que nosotros no nos encontrábamos a cargo de la extracción del DNA de las muestras clínicas, estas eran extraídas en el buffer de elución AE (del kit de extracción de DNA para sangre y fluidos corporales de Qiagen) que contiene EDTA al 0.5 mM, por lo que era necesario diluir las muestras y en consecuencia la carga viral de estas se diluía. En varias ocasiones en las que las muestras no fueron diluidas durante la construcción de librerías genómicas (en el Ion Torrent), fue posible observar indirectamente la inhibición de las enzimas al analizar estas librerías con los chips de DNA High Sensitivity (datos no mostrados). Una alternativa sería el uso de buffers libres de EDTA o en su defecto agua libre de nucleasas (en caso de que no se requiera almacenar el DNA por un periodo largo).

Otro factor crítico en la obtención de una baja cantidad de lecturas, fue el cargado deficiente de los chips de secuenciación para las muestras clínicas del Ion Torrent. Es

posible que el problema se produjera durante mezcla equimolar de las librerías al colocar una menor cantidad de DNA a la requerida, la concentración de DNA colocada siempre fue de 40pM sin que esta fuera previamente estandarizada, la concentración de la mezcla equimolar deberá ser estandarizada en trabajos posteriores.

Para lograr secuenciar los genomas de estos herpesvirus directamente de las muestras clínicas, sin el empleo de técnicas de amplificación de DNA o del cultivo viral de estas muestras que pueda modificar la secuencia de estos, sería necesario explorar alternativas que nos permitan eliminar el DNA humano y aumentar la proporción de DNA viral. Otra alternativa sería el empleo de chips o celdas de secuenciación con mayor capacidad, teniendo en cuenta de que la mayor proporción de las lecturas obtenidas posiblemente sean humanas.

11. CONCLUSIONES.

- Se secuenciaron los segmentos únicos de los genomas de HCMV cepa Merlin y HHV6A cepa GS a partir de cultivos virales empleando las plataformas de secuenciación de siguiente generación Ion Torrent (Life Technologies) y MiSeq (Illumina) e identificando varios polimorfismos propios de estos cultivos.
- La cobertura y profundidad de la secuenciación de HCMV y HHV6 obtenidas a partir de muestras clínicas de personas con VIH fue escasa y posiblemente asociada a la baja carga herpesviral. Para lograr caracterizar los genomas de HCMV o HHV6 será necesario emplear alternativas de enriquecimiento viral.

- En la plataforma Ion Torrent la profundidad de las lecturas de secuenciación obtenidas tuvo una relación inversa con el contenido de G+C, lo cual parece un fenómeno exclusivo a esta plataforma.
- Se observó una prevalencia de HCMV del 99% en pacientes con VIH, mientras que la prevalencia de HHV6 en estos fue del 71%. Existe una elevada prevalencia de betaherpesvirus en la población con VIH del país (exceptuando la presencia de HHV7).

12. LITERATURA CITADA.

1. Ablashi, D.V., Lusso, P., Hung, C.L., Salahuddin, S.Z., Josephs, S.F., Llana, T., Kramarsky, B., Biberfeld, P., Markham, P.D., Gallo, R.C., 1988. Utilization of human hematopoietic cell lines for the propagation and characterization of HBLV (human herpesvirus 6). *International journal of cancer. Journal international du cancer* 42, 787-791.
2. Ablashi, D.V., Salahuddin, S.Z., Josephs, S.F., Imam, F., Lusso, P., Gallo, R.C., Hung, C., Lemp, J., Markham, P.D., 1987. HBLV (or HHV-6) in human cell lines. *Nature* 329, 207.
3. Adams, M.J., Carstens, E.B., 2012. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2012). *Archives of virology* 157, 1411-1422.
4. Andrews, S., 2010. A quality control tool for high throughput sequence data, *Babraham Bioinform.*
5. Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., 2010. Manipulation of FASTQ data with Galaxy. *Bioinformatics* 26, 1783-1785.
6. Bowen, E.F., Sabin, C.A., Wilson, P., Griffiths, P.D., Davey, C.C., Johnson, M.A., Emery, V.C., 1997. Cytomegalovirus (CMV) viraemia detected by polymerase chain reaction identifies a group of HIV-positive patients at high risk of CMV disease. *AIDS* 11, 889-893.
7. Buonsenso, D., Serranti, D., Gargiullo, L., Ceccarelli, M., Ranno, O., Valentini, P., 2012. Congenital cytomegalovirus infection: current strategies and future perspectives. *European review for medical and pharmacological sciences* 16, 919-935.
8. Crumpacker, C.S., 2010a. Cytomegalovirus, in: Inc., E. (Ed.), *Principles and Practice of Infectious Diseases*, 7th ed, pp. 1971-1984
9. Crumpacker, C.S., 2010b. Invited commentary: human cytomegalovirus, inflammation, cardiovascular disease, and mortality. *American journal of epidemiology* 172, 372-374.
10. Cunningham, C., Gatherer, D., Hilfrich, B., Baluchova, K., Dargan, D.J., Thomson, M., Griffiths, P.D., Wilkinson, G.W., Schulz, T.F., Davison, A.J., 2010. Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. *The Journal of general virology* 91, 605-615.
11. Cha, T.A., Tom, E., Kemble, G.W., Duke, G.M., Mocarski, E.S., Spaete, R.R., 1996. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *Journal of virology* 70, 78-83.
12. Chen, J.J., Chang, H.F., Liu, C.Y., Chen, D.L., 2015. Infectious Mononucleosis Complicated with Acute Cerebral Infarction: A Case Report. *Acta neurologica Taiwanica* 24, 25-29.
13. Davison, A.J., Dolan, A., Akter, P., Addison, C., Dargan, D.J., Alcendor, D.J., McGeoch, D.J., Hayward, G.S., 2003. The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. *The Journal of general virology* 84, 17-28.
14. Dhepakson, P., Mori, Y., Jiang, Y.B., Huang, H.L., Akkapaiboon, P., Okuno, T., Yamanishi, K., 2002. Human herpesvirus-6 rep/U94 gene product has single-stranded DNA-binding activity. *The Journal of general virology* 83, 847-854.
15. Dockrell, D.H., 2003. Human herpesvirus 6: molecular biology and clinical features. *Journal of medical microbiology* 52, 5-18.

16. Dolan, A., Cunningham, C., Hector, R.D., Hassan-Walker, A.F., Lee, L., Addison, C., Dargan, D.J., McGeoch, D.J., Gatherer, D., Emery, V.C., Griffiths, P.D., Sinzger, C., McSharry, B.P., Wilkinson, G.W., Davison, A.J., 2004. Genetic content of wild-type human cytomegalovirus. *The Journal of general virology* 85, 1301-1312.
17. Dominguez, G., Dambaugh, T.R., Stamey, F.R., Dewhurst, S., Inoue, N., Pellett, P.E., 1999. Human herpesvirus 6B genome sequence: coding content and comparison with human herpesvirus 6A. *Journal of virology* 73, 8040-8052.
18. Field, D., Tiwari, B., Booth, T., Houten, S., Swan, D., Bertrand, N., Thurston, M., 2006. Open software for biologists: from famine to feast. *Nature biotechnology* 24, 801-803.
19. Gompels, U.A., Macaulay, H.A., 1995. Characterization of human telomeric repeat sequences from human herpesvirus 6 and relationship to replication. *The Journal of general virology* 76 (Pt 2), 451-458.
20. Gompels, U.A., Nicholas, J., Lawrence, G., Jones, M., Thomson, B.J., Martin, M.E., Efstathiou, S., Craxton, M., Macaulay, H.A., 1995. The DNA sequence of human herpesvirus-6: structure, coding content, and genome evolution. *Virology* 209, 29-51.
21. Goodman, A.L., Murray, C.D., Watkins, J., Griffiths, P.D., Webster, D.P., 2015. CMV in the gut: a critical review of CMV detection in the immunocompetent host with colitis. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* 34, 13-18.
22. Gutierrez, J., Cruz, L., 2008. Estudio de la seroprevalencia de la infección por citomegalovirus. *Revista mexicana de patología clínica* 55, 175-186.
23. Hall, C.B., Caserta, M.T., Schnabel, K., Shelley, L.M., Marino, A.S., Carnahan, J.A., Yoo, C., Lofthus, G.K., McDermott, M.P., 2008. Chromosomal integration of human herpesvirus 6 is the major mode of congenital human herpesvirus 6 infection. *Pediatrics* 122, 513-520.
24. Isegawa, Y., Mukai, T., Nakano, K., Kagawa, M., Chen, J., Mori, Y., Sunagawa, T., Kawanishi, K., Sashihara, J., Hata, A., Zou, P., Kosuge, H., Yamanishi, K., 1999. Comparison of the complete DNA sequences of human herpesvirus 6 variants A and B. *Journal of virology* 73, 8053-8063.
25. Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987-2993.
26. Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.
27. Lower, R., Lower, J., Kurth, R., 1996. The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proceedings of the National Academy of Sciences of the United States of America* 93, 5177-5184.
28. Lusso, P., Markham, P.D., Tschachler, E., di Marzo Veronese, F., Salahuddin, S.Z., Ablashi, D.V., Pahwa, S., Krohn, K., Gallo, R.C., 1988. In vitro cellular tropism of human B-lymphotropic virus (human herpesvirus-6). *The Journal of experimental medicine* 167, 1659-1670.
29. Maxam, A.M., Gilbert, W., 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* 74, 560-564.
30. Minevich, G., Park, D.S., Blankenberg, D., Poole, R.J., Hobert, O., 2012. CloudMap: a cloud-based pipeline for analysis of mutant genome sequences. *Genetics* 192, 1249-1269.
31. Morissette, G., Flamand, L., 2010. Herpesviruses and chromosomal integration. *Journal of virology* 84, 12100-12109.

32. Murphy, E., Rigoutsos, I., Shibuya, T., Shenk, T.E., 2003. Reevaluation of human cytomegalovirus coding potential. *Proceedings of the National Academy of Sciences of the United States of America* 100, 13585-13590.
33. Pellett, P.E., Ablashi, D.V., Ambros, P.F., Agut, H., Caserta, M.T., Descamps, V., Flamand, L., Gautheret-Dejean, A., Hall, C.B., Kamble, R.T., Kuehl, U., Lassner, D., Lautenschlager, I., Loomis, K.S., Luppi, M., Lusso, P., Medveczky, P.G., Montoya, J.G., Mori, Y., Ogata, M., Pritchett, J.C., Rogez, S., Seto, E., Ward, K.N., Yoshikawa, T., Razonable, R.R., 2012. Chromosomally integrated human herpesvirus 6: questions and answers. *Reviews in medical virology* 22, 144-155.
34. Potenza, L., Barozzi, P., Masetti, M., Pecorari, M., Bresciani, P., Gautheret-Dejean, A., Riva, G., Vallerini, D., Tagliazucchi, S., Codeluppi, M., Di Benedetto, F., Gerunda, G.E., Narni, F., Torelli, G., Luppi, M., 2009. Prevalence of human herpesvirus-6 chromosomal integration (CIHHV-6) in Italian solid organ and allogeneic stem cell transplant patients. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 9, 1690-1697.
35. Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics* 13, 341.
36. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. *Nature biotechnology* 29, 24-26.
37. Rojo Medina, J., Krueger, G.R., Bonifaz Gracias, R., Berneman, Z., Koch, B., 1995. [Prevalence of human herpesvirus 7 in Mexican blood donors]. *Revista de investigacion clinica; organo del Hospital de Enfermedades de la Nutricion* 47, 467-471.
38. Roubille, C., Brunel, A.S., Gahide, G., Vernhet Kovacsik, H., Le Quellec, A., 2010. Cytomegalovirus (CMV) and acute myocarditis in an immunocompetent patient. *Intern Med* 49, 131-133.
39. Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944-945.
40. Salahuddin, S.Z., Ablashi, D.V., Markham, P.D., Josephs, S.F., Sturzenegger, S., Kaplan, M., Halligan, G., Biberfeld, P., Wong-Staal, F., Kramarsky, B., et al., 1986. Isolation of a new virus, HBLV, in patients with lymphoproliferative disorders. *Science* 234, 596-601.
41. Salud, S.d., 2012. Prevención, diagnóstico y tratamiento de la infección por citomegalovirus en edad pediátrica. *México*, pp. 13-21.
42. Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74, 5463-5467.
43. Thorvaldsdottir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* 14, 178-192.
44. Weller, T.H., 2000. Cytomegaloviruses: a Historical Perspective. *Herpes : the journal of the IHMF* 7, 66-69.
45. Yamanishi, K., 2000. Human Herpesvirus 6: an Evolving Story. *Herpes : the journal of the IHMF* 7, 70-75.