

Sergio Bianchi^{1,2}, Natalia Rego³, Pilar Moreno^{4,5}, Pablo Opezzo^{4,6}, Alvaro Pena³, Hugo Naya³, Guillermo Dighiero¹ y Otto Pritsch^{1,6}

¹Unidad de Biofísica de Proteínas, Institut Pasteur de Montevideo, Uruguay; ²Departamento de Fisiopatología, Hospital de Clínicas, Universidad de la República, Uruguay; ³Unidad de Bioinformática, Institut Pasteur de Montevideo, Uruguay; ⁴Unidad de Proteínas Recombinantes, Institut Pasteur de Montevideo, Uruguay; ⁵Laboratorio de Virología Molecular, Centro de Investigaciones Nucleares, Uruguay; ⁶Departamento de Inmunobiología, Facultad de Medicina, Universidad de la República, Uruguay.

Introducción

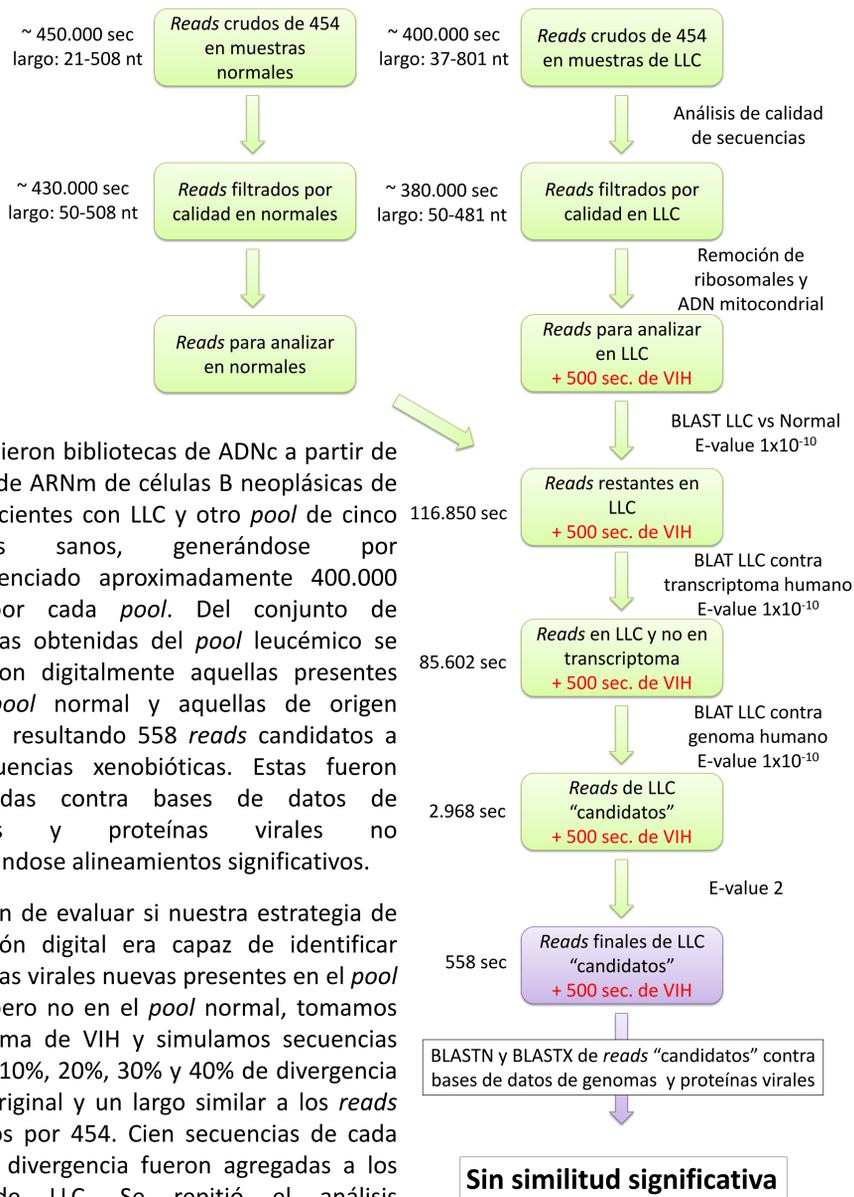
La Leucemia Linfocítica Crónica (LLC) es la leucemia más común en adultos de países occidentales. Su baja incidencia en personas de países orientales, incluyendo aquellos que han emigrado a occidente, sugiere que las influencias genéticas son más importantes que las ambientales en su patogenia, sin embargo su etiología es aún desconocida. Actualmente al menos seis virus diferentes han sido implicados en alrededor de 15% de los cánceres humanos. Por otro lado, el Virus de la Leucemia Bovina es un retrovirus identificado como el agente etiológico de la Leucemia Bovina Enzootica, una patología con características similares a la LLC humana. Estos y otros datos permiten plantear la hipótesis de un posible origen viral de la LLC. En este trabajo realizamos la búsqueda de un virus como posible agente etiológico de la LLC mediante análisis del transcriptoma de linfocitos B leucémicos utilizando dos estrategias de secuenciado masivo (pirosecuenciado y secuenciado por síntesis) combinadas con una sustractiva digital.

Pacientes, estrategias experimentales y resultados

Pirosecuenciado (454)

Paciente	Sexo	Edad	Binet	CD38	LPL	EM	Evolución
LLC 046	F	45	A	Neg	Pos	NM	Progresor
LLC 072	M	61	C	Neg	Pos	NM	Indolente
LLC 080	F	63	C	Neg	Pos	NM	Progresor
LLC 083	M	60	B	Pos	ND	NM	Progresor
LLC 096	M	70	C	Neg	Neg	NM	Progresor

Binet: estadio de Binet; LPL: lipoproteinlipasa; EM: estado mutacional de genes IgVH; ND: no determinado.



Se obtuvieron bibliotecas de ADNc a partir de un pool de ARNm de células B neoplásicas de cinco pacientes con LLC y otro pool de cinco donantes sanos, generándose por pirosecuenciado aproximadamente 400.000 reads por cada pool. Del conjunto de secuencias obtenidas del pool leucémico se sustrajeron digitalmente aquellas presentes en el pool normal y aquellas de origen humano, resultando 558 reads candidatos a ser secuencias xenobióticas. Estas fueron comparadas contra bases de datos de genomas y proteínas virales no encontrándose alineamientos significativos.

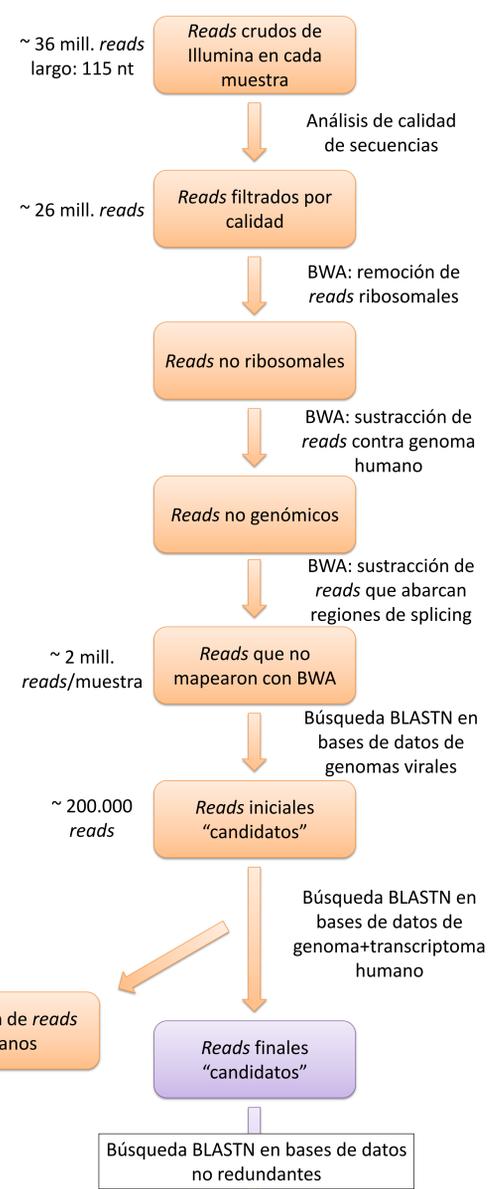
Con el fin de evaluar si nuestra estrategia de sustracción digital era capaz de identificar secuencias virales nuevas presentes en el pool de LLC pero no en el pool normal, tomamos un genoma de VIH y simulamos secuencias con 0%, 10%, 20%, 30% y 40% de divergencia con el original y un largo similar a los reads obtenidos por 454. Cien secuencias de cada nivel de divergencia fueron agregadas a los reads de LLC. Se repitió el análisis identificándose las secuencias con hasta 30% de divergencia como secuencias de VIH, mientras aquellas con un 40% de divergencia fueron identificadas como de origen viral.

Secuenciado por síntesis (Illumina)

Paciente	Sexo	Edad	Binet	CD38	LPL	EM	Evolución
LLC 238	M	76	C	Neg	Neg	M	Progresor
LLC 250	M	57	B	ND	Pos	UM	Progresor

Binet: estadio de Binet; LPL: lipoproteinlipasa; EM: estado mutacional de genes IgVH; ND: no determinado.

Con la finalidad de aumentar la profundidad del análisis, células B de dos pacientes con LLC fueron estimuladas *in vitro* con un potente agente mitogénico (éster de forbol, TPA), generándose dos muestras pareadas: una activada y otra no activada por cada uno de los pacientes. Las bibliotecas de ARN total fueron normalizadas y secuenciadas masivamente por síntesis (Illumina, Inc.), incrementándose alrededor de 90 veces la profundidad del secuenciado. Luego de ser filtrados, entre 22 y 28 millones de reads por muestra fueron mapeados utilizando la herramienta Burrows-Wheeler Alignment (BWA) al genoma y transcriptoma humano de referencia. Alrededor de 2 millones de reads no fueron asignados a un origen humano. Estos fueron cotejados sucesivamente con BLAST (siguiendo un protocolo de mayor sensibilidad) en base de datos de genomas virales, secuencias humanas y no redundante del NCBI. Si bien nueve reads de una de las muestras activadas fueron finalmente asignados a virus de Epstein-Barr (demostrando la eficiencia de la estrategia empleada), ninguna secuencia pudo ser asociada a un posible agente etiológico.



Reads	NM	NM_act	Mut	Mut_act
Mamíferos	176	162	198	455
MPMV-like ¹	336	193	139	92
Otros virus-like ²	0	1	13	6
Virus Epstein-Barr	0	0	0	9
Otros ³	146	121	161	176
Total	658	477	511	738

MPMV: Mason-Pfizer monkey virus.

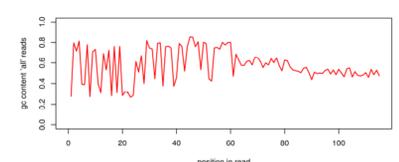
¹Los reads que mostraron alineamientos con el genoma de MPMV fueron finalmente asignados al tRNA-Lys de origen humano.

²Comprenden reads que alinean con genomas virales diferentes a MPMV o virus de Epstein-Barr. Estos alineamientos también fueron descartados.

³Alineamientos que involucraron especies imprevistas. Errores de secuenciación pueden ser la causa de estos resultados llamativos.

Contenido de GC según posición en el read de las secuencias que alinearon con MPMV

El análisis mostró secuencias símil quiméricas, en las cuales la primera mitad del read deriva del precursor de tRNA-Lys mientras la región siguiente es variable.



Conclusiones y discusión

- Luego de estos análisis transcriptómicos no se logró identificar ninguna secuencia que pueda ser relacionada con un agente etiológico de la LLC.

- Las estrategias de sustracción digital mostraron ser adecuadas y eficientes para la identificación de secuencias virales, no sólo de las simuladas (secuencias divergentes del VIH) sino también de las obtenidas a partir de las muestras biológicas (virus de Epstein-Barr).

- Estos resultados no permiten descartar un posible origen viral de la leucemia ya que la metodología empleada requiere un nivel mínimo de similitud del nuevo agente con los virus hoy conocidos. Tampoco podemos excluir la posibilidad de la existencia de un virus integrado al genoma celular con nula o indetectable tasa de transcripción, lo que podría ser estudiado con nuevas tecnologías de análisis a nivel genómico.