

Departamento de Medicina  
Centro de Investigación del Cáncer  
Instituto de Investigación Biomédica de Salamanca

TESIS DOCTORAL

**CARACTERIZACIÓN PROTEÓMICA DE LA CÉLULA B  
TUMORAL DE LA LEUCEMIA LINFOCÍTICA CRÓNICA  
Y SU CONTRAPARTIDA CELULAR NORMAL**

**Paula Díez García  
2017**

Directores:

Dr. Manuel Fuentes

Dr. Alberto Orfao





Esta Tesis Doctoral fue desarrollada siendo Paula Díez García beneficiaria de una “Ayuda Predoctoral de la Junta de Castilla y León” financiada por el Fondo Social Europeo (JCYL-EDU/346/2013).

---

Para el desarrollo de la misma contó también con el apoyo (financiación por proyectos, utilización de infraestructuras y/o cesión de muestras) de las siguientes instituciones:

- Fondo de Investigaciones Sanitarias (FIS) (PI14/01538, PI12/00905, PI12/00624, PI11/02114), Instituto de Salud Carlos III (ISCIII), Fondo Europeo de Desarrollo Regional (FEDER) “Una manera de hacer Europa”.
  - Junta Castilla-León (BIO/SA07/15, SA198A12-2, BIO/SA03/15).
  - Fundación Solórzano (FS/23-2015).
  - Unidad de Proteómica del Centro de Investigación del Cáncer (Salamanca, España) perteneciente a ProteoRed (PRB2-ISCIII, PT13/0001).
  - Banco Nacional de ADN – ISCIII (Salamanca, España).
-





## DEPARTAMENTO DE MEDICINA

**D. Manuel Fuentes, Doctor en Bioquímica y Biología Molecular y Profesor Contratado Doctor del Departamento de Medicina de la Universidad de Salamanca**

### **CERTIFICA:**

Que el trabajo doctoral realizado en formato de compendio de artículos/publicaciones bajo mi dirección por Dña. Paula Díez García y titulado **“Caracterización proteómica de la célula B tumoral de la leucemia linfocítica crónica y su contrapartida celular normal”**, reúne las condiciones de originalidad y calidad científicas requeridas para optar al grado de Doctor por la Universidad de Salamanca.

Y para que así conste, firmo el presente certificado en Salamanca a 19 de junio del año 2017.

**Fdo.: Prof. Dr. D. Manuel Fuentes**





## DEPARTAMENTO DE MEDICINA

**D. Alberto Orfao, Doctor en Medicina y Cirugía y Catedrático del Departamento de Medicina de la Universidad de Salamanca**

### **CERTIFICA:**

Que el trabajo doctoral realizado en formato de compendio de artículos/publicaciones bajo mi dirección por Dña. Paula Díez García y titulado **“Caracterización proteómica de la célula B tumoral de la leucemia linfocítica crónica y su contrapartida celular normal”**, reúne las condiciones de originalidad y calidad científicas requeridas para optar al grado de Doctor por la Universidad de Salamanca.

Y para que así conste, firmo el presente certificado en Salamanca a 19 de junio del año 2017.

**Fdo.: Prof. Dr. D. Alberto Orfao**





La presente tesis doctoral incluye cinco trabajos originales, tres de ellos ya publicados y dos en período de revisión en revistas indexadas en el *Science Citation Reports*, los cuales se detallan a continuación:

**1. Integrative proteomics and phosphoproteomics profile of chronic lymphocytic leukemia tumor B-cells and the signaling pathways involved.**

Paula Díez<sup>1,2</sup>, Conrad Droste<sup>3</sup>, Julia Almeida<sup>1</sup>, Marcos González<sup>4</sup>, Alberto Orfao<sup>1</sup>, Javier De Las Rivas<sup>3</sup>, Manuel Fuentes<sup>1,2</sup>

(en revisión)

**2. Multipronged functional proteomics approaches for global identification of altered cell signaling pathways in B-cell chronic lymphocytic leukemia.**

Paula Díez<sup>1,2</sup>, Seila Lorenzo<sup>1</sup>, Rosa M. Dégano<sup>2</sup>, Nieves Ibarrola<sup>2</sup>, María González-González<sup>1,2</sup>, Wendy Nieto<sup>1</sup>, Julia Almeida<sup>1</sup>, Marcos González<sup>4</sup>, Alberto Orfao<sup>1</sup>, Manuel Fuentes<sup>1,2</sup>

Proteomics. 2016 Abr; 16(8):1193-1203.

doi: 10.1002/pmic.201500372

**3. A systematic approach for peptide characterization of B-cell receptor in chronic lymphocytic leukemia cells.**

Paula Díez<sup>1,2</sup>, Nieves Ibarrola<sup>2</sup>, Rosa M. Dégano<sup>2</sup>, Quentin Lécrevisse<sup>1</sup>, Arancha Rodríguez-Caballero<sup>1</sup>, Ignacio Criado<sup>1</sup>, Wendy Nieto<sup>1</sup>, Rafael Góngora<sup>1</sup>, Marcos González<sup>4</sup>, Julia Almeida<sup>1</sup>, Alberto Orfao<sup>1</sup>, Manuel Fuentes<sup>1,2</sup>

Oncotarget. 2017 Abr.

doi: 10.18632/oncotarget.17076

**4. Integration of proteomics and transcriptomics data sets for the analysis of a lymphoma B-cell line in the context of the chromosome-centric human proteome project.**

Paula Díez<sup>1,2</sup>, Conrad Droste<sup>3</sup>, Rosa M. Dégano<sup>2</sup>, María González-Muñoz<sup>1</sup>, Nieves Ibarrola<sup>2</sup>, Martín Pérez-Andrés<sup>1</sup>, Alba Garin-Muga<sup>5</sup>, Víctor Segura<sup>5</sup>, György Marko-Varga<sup>6</sup>, Joshua LaBaer<sup>7</sup>, Alberto Orfao<sup>1</sup>, Fernando J. Corrales<sup>5</sup>, Javier De Las Rivas<sup>3</sup>, Manuel Fuentes<sup>1,2</sup>

Journal of Proteome Research. 2015 Sep; 14(9):3530-3540.

doi: 10.1021/acs.jproteome.5b00474

**5. A B-cell global landscape during antigen differentiation reveals cell signaling dynamics across B-cell subpopulations**

Paula Díez<sup>1,2</sup>, Martín Pérez-Andrés<sup>1</sup>, Martin Bøgsted<sup>8</sup>, Mikel Azkargorta<sup>9</sup>, Rodrigo García-Valiente<sup>2</sup>, Rosa M. Dégano<sup>2</sup>, Elena Blanco<sup>1</sup>, Santiago Santa Cruz<sup>10</sup>, Félix Elortza<sup>9</sup>, Karen Dybkaer<sup>8</sup>, Alberto Orfao<sup>1</sup>, Manuel Fuentes<sup>1,2</sup>

(en preparación)

---

<sup>1</sup> Servicio General de Citometría-NUCLEUS, Departamento de Medicina; Centro de Investigación del Cáncer (IBMCC-CSIC/USAL) e IBSAL, Universidad de Salamanca.

<sup>2</sup> Unidad de Proteómica; Centro de Investigación del Cáncer (IBMCC-CSIC/USAL) e IBSAL, Universidad de Salamanca.

<sup>3</sup> Grupo de Investigación de Bioinformática y Genómica Funcional; Centro de Investigación del Cáncer (IBMCC-CSIC/USAL) e IBSAL, Universidad de Salamanca.

<sup>4</sup> Servicio de Hematología (IBMCC) e IBSAL, Hospital Universitario de Salamanca.

<sup>5</sup> Unidad de Proteómica y Bioinformática; División de Hepatología y Terapia Génica; Centro de Investigación Médica Aplicada (CIMA), Universidad de Navarra.

<sup>6</sup> Clinical Protein Science and Imaging, Department of Biomedical Engineering; Lund University.

<sup>7</sup> Biodesign Institute; Arizona State University.

<sup>8</sup> Department of Haematology, Aalborg University Hospital.

<sup>9</sup> Plataforma Proteómica, CIC bioGUNE, Parque Tecnológico Bizkaia.

<sup>10</sup> Servicio de Otorrinolaringología y Patología Facial Cervical. Hospital Universitario de Salamanca.

*A mis padres.  
A mi hermano.*



*A Ibán.*



*“Lo importante es no dejar de hacerse preguntas”*

*Albert Einstein.*





# Índice



<b>GLOSARIO DE ABREVIATURAS</b>	1
<b>INTRODUCCIÓN</b>	7
<b>1. Linfocito B</b>	9
1.1. Ontogenia del linfocito B	9
1.1.1. Maduración B independiente de Ag	9
1.1.2. Maduración B dependiente de Ag	11
1.2. Principales rutas de señalización intracelular en los linfocitos B	13
1.2.1. Vías de señalización intracelular a través de BCR	13
1.2.2. Vías de señalización intracelular mediada por receptores de quimiocinas	15
1.2.3. Vías de señalización intracelular a través de receptores tipo Toll (TLR)...	16
<b>2. Leucemia linfocítica crónica de célula B</b>	16
2.1. Epidemiología y etiología de la LLC	16
2.2. Diagnóstico de LLC-B	17
2.3. Características clínico-biológicas y factores pronósticos en la LLC-B	18
2.3.1. Factores pronósticos clínicos	20
2.3.2. Factores pronósticos biológicos	21
2.3.2.1. Factores pronósticos séricos	21
2.3.2.2. Factores pronósticos genéticos	21
2.3.2.3. Factores pronósticos fenotípicos	23
2.3.2.4. Otros factores pronósticos	23
2.4. Origen celular de la LLC	24
2.5. Papel del microambiente tumoral en la LLC	26
2.6. Tratamiento de la LLC	27
<b>3. Técnicas proteómicas de alto rendimiento para el estudio del linfocito B normal y patológico</b>	29
3.1. Definición de proteoma	30
3.2. Complejidad del proteoma	34
3.3. Proteómica: áreas de estudio y técnicas de análisis	35
3.3.1. Técnicas de separación de proteínas	37
3.3.2. Identificación, cuantificación y caracterización de proteínas	38
3.3.3. Integración de tecnologías -ómicas	43
3.4. Ventajas y retos de la proteómica	43
<b>HIPÓTESIS DE TRABAJO Y OBJETIVOS</b>	47

<b>MATERIALES, MÉTODOS Y RESULTADOS</b>	53
1. Capítulo 1	57
Caracterización proteómica y fosfoproteómica integrada de las células B tumorales de la leucemia linfocítica crónica y de las rutas de señalización implicadas.	
2. Capítulo 2	89
Aproximaciones proteómicas funcionales para la identificación global de rutas de señalización celulares alteradas en la leucemia linfocítica crónica de célula B.	
3. Capítulo 3	107
Una aproximación sistemática para la caracterización peptídica del receptor de célula B en células de la leucemia linfocítica crónica.	
4. Capítulo 4	125
Integración de datos masivos proteómicos y transcriptómicos para el análisis de una línea celular de linfoma B en el contexto del Proyecto Proteoma Humano (C-HPP).	
5. Capítulo 5	143
La caracterización de distintos compartimentos madurativos de células B muestra un importante dinamismo en las vías de señalización durante la diferenciación B dependiente de antígeno	
<b>DISCUSIÓN</b>	187
<b>CONCLUSIONES</b>	219
<b>BIBLIOGRAFÍA</b>	225

## Glosario de abreviaturas

### a

- **Ac:** anticuerpo
- **ADN:** ácido desoxirribonucleico
- **Ag:** antígeno
- **Arg:** arginine
- **ARN:** ácido ribonucleico
- **ARNm:** ácido ribonucleico mensajero

### b

- **BCR:** receptor de célula B (del inglés, *B-cell receptor*)
- **B/D-HPP:** Proyecto Proteoma Humano basado en la biología/enfermedad (del inglés, *biology/disease Human Proteome Project*)

### C

- **CB:** centroblasto
- **CC:** centrocito
- **CDR:** región determinante de complementariedad (del inglés, *complementarity determining region*)
- **CDF:** célula dendrítica folicular
- **CG:** centro germinal
- **C-HPP:** Proyecto Proteoma Humano basado en cromosomas (del inglés, *chromosome-centric Human Proteome Project*)
- **CMH:** célula madre hematopoyética
- **CP:** célula plasmática
- **CPA:** célula presentadora de antígenos
- **CXCR4:** receptor de quimiocina C-X-C de tipo 4

- **Cy:** cianina

## e

- **ELISA:** ensayo por inmunoabsorción ligado a enzimas (del inglés, *Enzyme-Linked ImmunoSorbent Assay*)
- **emPAI:** índice de abundancia proteica exponencialmente modificado (del inglés, *exponentially modified Protein Abundance Index*)
- **ESI:** ionización por electrospray (del inglés, *electrospray ionization*)

## f

- **FCR:** fludarabina, ciclofosfamida y rituximab
- **FDR:** tasas de falsos descubrimientos (del inglés, *false discovery rate*)
- **FISH:** hibridación *in situ* fluorescente (del inglés, *fluorescence in situ hybridization*)
- **FR:** región marco (del inglés *framework region*)
- **FT-MS:** espectrometría de masas de resonancia ion-ciclotrón con transformada de Fourier (del inglés, *Fourier transform ion cyclotron resonance mass spectrometry*)

## g

- **GEO:** *Gene Expression Omnibus*
- **GPMDB:** base de datos sobre proteoma global (del inglés, *Global Proteome Machine and Database*)

## h

- **HPA:** atlas proteico humano (del inglés, *Human Protein Atlas*)
- **HPP:** Proyecto Proteoma Humano (del inglés, *Human Proteome Project*)
- **HUPO:** Organización del Proteoma Humano (del inglés, *Human Proteome Organization*)

## i

- **iCAT:** marcaje por afinidad basado en isótopos codificados (del inglés, *Isotope-Coded Affinity Tag*)
- **Ig:** inmunoglobulina
- **IGH:** del inglés *heavy*, refiriéndose a las cadenas pesadas de las inmunoglobulinas
- **IGHV:** región variable de la cadena pesada de la inmunoglobulina
- **IGK:** cadena kappa de las inmunoglobulinas
- **IGL:** cadena lambda de las inmunoglobulinas
- **IMAC:** cromatografía de afinidad por metal inmovilizado (del inglés, *immobilized metal affinity chromatography*)
- **ITAM:** motivo de activación de inmunoreceptor basado en tirosinas (del inglés, *immunoreceptor tyrosine-based activation motif*)
- **ITRAQ:** marcaje isobárico para la cuantificación absoluta y relativa (del inglés, *Isobaric Tag for Relative and Absolute Quantification*)

## l

- **L:** del inglés *light*, refiriéndose a las cadenas ligeras de las inmunoglobulinas
- **LBDCG:** linfoma B difuso de células grandes
- **LC-MS/MS:** cromatografía líquida acoplada a espectrometría de masas en tándem
- **LH:** linfoma de Hodgkin
- **LLA:** leucemia linfoblástica aguda
- **LLC:** leucemia linfocítica crónica
- **LLCP:** linfoma linfocítico de células pequeñas
- **LMB:** linfocitosis monoclonal de célula B
- **Lys:** lisina

## m

- **MALDI:** ionización/desorción por láser asistida por matriz (del inglés, *matrix-associated laser desorption/ionization*)
- **miARN:** micro ácido ribonucleico
- **MO:** médula ósea
- **MRM:** monitoreo de reacción múltiple (del inglés, *Multiple Reaction Monitoring*)
- **MS:** espectrometría de masas (del inglés, *mass spectrometry*)

## n

- **NGS:** secuenciación de próxima generación (del inglés, *Next-Generation Sequencing*)

## O

- **OMS:** Organización Mundial de la Salud

## p

- **PAGE:** electroforesis en gel de poliacrilamida (del inglés, *polyacrilamide gel electrophoresis*)
- **PAI:** índice de abundancia proteica (del inglés, *protein abundance index*)
- **PAMP:** patrones moleculares asociados a patógenos (del inglés, *pathogen-associated molecular patterns*)
- **PE:** pruebas proteicas (del inglés, *protein evidence*)
- **PKC:** proteína quinasa proteica C (del inglés, *protein kinase C*)
- **PRB2:** Plataforma de Recursos Biomoleculares y Bioinformáticos
- **PRIDE:** base de datos de identificadores proteómicos (del inglés, *PRoteomics IDentifications Database*)
- **PSM:** coincidencia del espectro peptídico (del inglés, *peptide spectrum match*)



- **PTM:** modificación post-traducciona (del inglés, *Post Translational Modification*)
- **pTyr:** fosfotirosina

## q

- **Q:** cuadrupolo (del inglés, *quadrupole*)
- **QI:** quimioinmunoterapia

## S

- **SCF:** factor de célula madre (del inglés *stem cell factor*)
- **Ser:** serina
- **SILAC:** marcaje isotópico estable de aminoácidos en cultivo celular (del inglés, *stable isotope labeling by amino acids in cell culture*)
- **SpHPP:** Proyecto Proteoma Humano español (del inglés, *Spanish Human Proteome Project*)
- **SRM:** monitoreo de reacción seleccionada (del inglés, *Selected Reaction Monitoring*)

## t

- **TDL:** tiempo de duplicación linfocitaria
- **Thr:** treonina
- **TLR:** receptores tipo Toll (del inglés, *Toll-like receptors*)
- **TK1:** timidina quinasa 1
- **TOF:** tiempo de vuelo (del inglés, *time of flight*)
- **Tyr:** tirosina

## X

- **XIC:** cromatograma de iones extraídos (del inglés, *extracted ion chromatogram*)

## Z

- **ZM:** zona marginal

# Introducción



## 1. Linfocito B

El linfocito B es una célula central de la inmunidad humoral cuyas principales funciones son la presentación de antígenos (Ag) a los linfocitos T CD4<sup>+</sup> (en tal caso actúan como células presentadoras de Ag, CPA), y la diferenciación a células plasmáticas (CP) con capacidad de sintetizar y secretar inmunoglobulinas (Ig)<sup>1</sup>.

### 1.1. Ontogenia del linfocito B

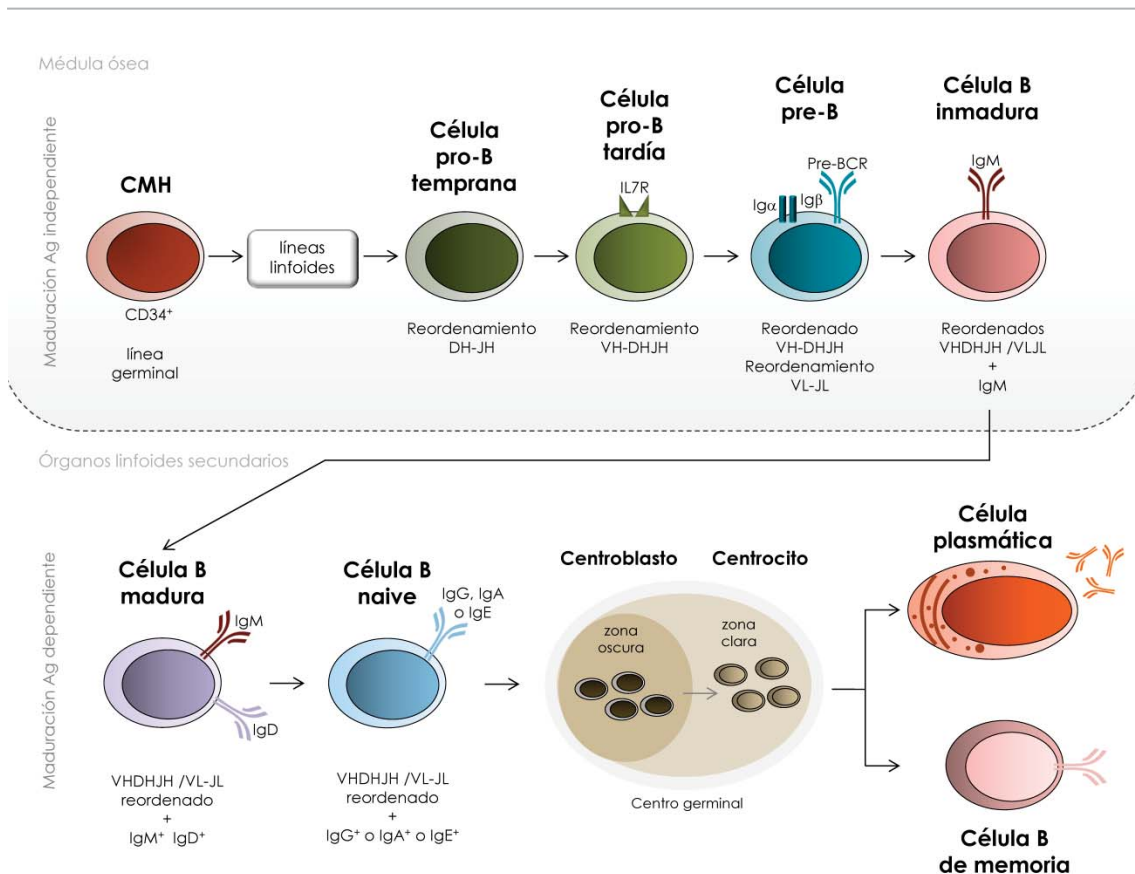
Durante el desarrollo embrionario, la ontogenia de los linfocitos B ocurre inicialmente en el saco vitelino desplazándose, a partir de la octava y novena semana de gestación, al hígado fetal. Tras el nacimiento, la diferenciación de estas células tiene lugar en la médula ósea (MO) donde se generan también otras células hematopoyéticas a través de un proceso continuo de formación de células sanguíneas maduras a partir de células madre indiferenciadas (*i.e.* hematopoyesis)<sup>1,2</sup>.

En la ontogenia de los linfocitos B se distinguen dos fases principales: la maduración independiente de Ag, que ocurre en la MO, y la maduración dependiente de Ag, típica de los órganos linfoides secundarios (Figura 1).

#### 1.1.1. Maduración B independiente de Ag

La maduración de células B independiente de Ag tiene lugar en la MO a partir de un precursor hematopoyético CD34<sup>+</sup> y dará lugar a la formación de linfocitos B maduros vírgenes e inmunocompetentes. Como en la MO no existe un contacto con el Ag, específico de cada linfocito B, la diferenciación depende, fundamentalmente, de la interacción de proteínas de membrana de las células del estroma de la MO (p.ej. adipocitos, fibroblastos, mastocitos, células mesenquimales) con determinadas proteínas de membrana de los linfocitos B, y señales derivadas de factores solubles inductores de diferenciación B que incluyen factores de crecimiento, quimiocinas y citocinas (p.ej. IL7). En las etapas iniciales de la linfopoyesis B, la molécula CD44 de la célula madre de la MO interacciona con el ácido hialurónico del estroma medular; esta señal, junto a la interacción del *stem cell factor* (SCF) con el receptor KIT (CD117),

induce la puesta en marcha de un proceso hematopoyético indiferenciado del programa de señalización dirigido hacia la diferenciación a célula B temprana<sup>3,4</sup>. Así, la cascada de señalización desencadenada por KIT induce la producción y expresión de receptores de citocinas (p.ej. el receptor de la interleucina 7, IL7R) formándose la célula pro-B tardía<sup>5</sup>. A su vez, la célula pro-B tardía interactúa con células estromales reticulares poniendo en marcha el reordenamiento, primero, de los segmentos génicos que codifican para la región variable de la cadena pesada de las Ig (genes IGH, del inglés *heavy*) (primero se reordenan los segmentos D y J, y después los segmentos V y DJ) y, posteriormente, de las cadenas ligeras de las Ig (IGL/IGK, cadenas ligeras lambda/kappa). Tras el reordenamiento inicial de los genes IGH, el precursor B produce cadenas pesadas  $\mu$  que se expresan en la membrana en forma de pre-BCR (del inglés *B-cell receptor*) dando lugar a la célula pre-B<sup>6-8</sup>. Posteriormente, esta célula pre-B reordena los genes IGL/IGK y, al expresar IgM de superficie, se transforma en una célula B inmadura. Así, el linfocito B inmaduro es la primera célula que presenta en su membrana un receptor de célula B completo (BCR) de tipo IgM específico de un Ag diferente en cada célula B. Aquellas células B inmaduras que expresan IgM autorreactiva pueden experimentar una segunda fase de reordenamiento de los segmentos génicos codificantes de cadena L con el fin de editar un BCR nuevo. En caso de no corregirse la autorreactividad, dichas células B mueren por apoptosis o bien permanecen en situación de anergia (selección negativa); en caso contrario, continúan con el proceso madurativo y terminan coexpresando IgM e IgD de superficie constituyendo el linfocito B maduro (naive). En la transición entre célula B inmadura y linfocitos B maduros, estas células abandonan la MO y emigran a los órganos linfoides secundarios donde potencialmente reconocen selectiva y específicamente Ag y se diferenciarán a células B de memoria y CPs<sup>7,9,10</sup>.



**Figura 1. Diferenciación linfocítica B.** Desarrollo y maduración de la célula B desde la célula madre hematopoyética (CMH) hasta las células B efectoras (célula plasmática y célula B de memoria). La diferenciación comienza en la médula ósea sin contacto con antígeno (Ag) hasta que en la transición entre la célula B inmadura y el linfocito B maduro, que ya expresa inmunoglobulina (Ig)-M e IgM/D en su superficie, este último migra a los órganos linfoides secundarios y se convierte en una célula B madura inmunocompetente. Esta célula B naiva/madura accede al centro germinal donde reconoce Ag y se transforma secuencialmente de centroblasto a centrocito y, tras sufrir el fenómeno de hipermutación somática y el cambio de isotipo, se diferencia a célula plasmática secretora de anticuerpos o a linfocito B de memoria.

### 1.1.2. Maduración B dependiente de Ag

El linfocito B maduro (o naivo), una vez abandona la MO y migra a los órganos linfoides secundarios, puede encontrarse (o no) con un Ag específico de su BCR presente en su superficie. En caso de no contactar con dicho Ag, pasados unos días, el linfocito B entra en apoptosis y muere. Por el contrario, si entra en contacto con el Ag, el linfocito B naivo se activa, prolifera y se diferencia a CP productora de anticuerpos (Ac) o a linfocito B de memoria<sup>11,12</sup>.

Las etapas principales de la diferenciación del linfocito B naivo ocurren en los órganos linfoides secundarios donde la célula naiva se diferencia a célula secretora de Ac (CP o célula de memoria)<sup>13</sup>. Este proceso ocurre tras el reconocimiento del Ag por

el BCR del linfocito B naive. En los órganos linfoides secundarios, este reconocimiento ocurre en el entorno de zonas ricas en linfocitos T, donde se localizan los centros germinales (CG), lugar preferente donde el linfocito B contacta con el Ag<sup>14,15</sup>. En dichos CGs, el linfocito B activado prolifera y se diferencia ampliando el repertorio de BCRs que expresa mediante el fenómeno de hipermutación somática del gen de las Igs, que da lugar a una elevada diversidad del receptor para Ag con el objetivo último de incrementar la afinidad del BCR por el mismo Ag que inicialmente indujo su activación<sup>15,16</sup>. Dentro del CG se distinguen dos zonas: una zona oscura, donde ocurre la proliferación del linfocito B y el fenómeno de hipermutación somática y donde se observa un incremento de tamaño de los linfocitos B activos asociados a una cromatina más laxa - centroblastos (CB) -, y una zona clara, donde los CB se diferencian a centrocitos (CC)<sup>16,17</sup>. Estas células B activadas, expandidas y con mayor o menor afinidad de su BCR por el Ag se seleccionan en base a su capacidad para i) reconocer selectiva y específicamente, a través de su BCR, Ags presentados por células dendríticas foliculares (CDF)<sup>18</sup> y ii) procesar y presentar los péptidos procedentes de dichos Ag a las células T de memoria de la zona del manto y paracortical mediante moléculas HLA de clase II<sup>19</sup>. La interacción entre el linfocito B activado presentador de Ag y la célula T desencadena el cambio de isotipo de cadena pesada de las Igs expresado por la célula B, seguido de la diferenciación a CP o célula de memoria<sup>20,21</sup>. La selección y generación de las células de memoria no se ha descrito de forma clara hasta el momento<sup>17,22</sup>, aunque parecen tener un papel determinante en su diferenciación tanto la afinidad del BCR por el Ag como la interacción de CD40 con CD40L<sup>22</sup>.

Las alteraciones genéticas relacionadas con el fenómeno de hipermutación somática de los genes de las Igs y el cambio de isotipo de la célula B activada pueden dar lugar a su transformación neoplásica y progresión a distintas neoplasias B como la leucemia linfoblástica aguda (LLA), el linfoma folicular y otros linfomas no hodgkinianos, y la leucemia linfocítica crónica (LLC), entre otros<sup>1,23</sup>. Asimismo, la



generación de células B autorreactivas, capaces de producir auto-Ac frente a proteínas propias, puede resultar también en enfermedades autoinmunes como la esclerosis múltiple, el lupus sistémico eritematoso, la artritis reumatoide y las citopenias autoinmunes, entre otras enfermedades<sup>24,25</sup>.

## **1.2. Principales rutas de señalización intracelular en los linfocitos B**

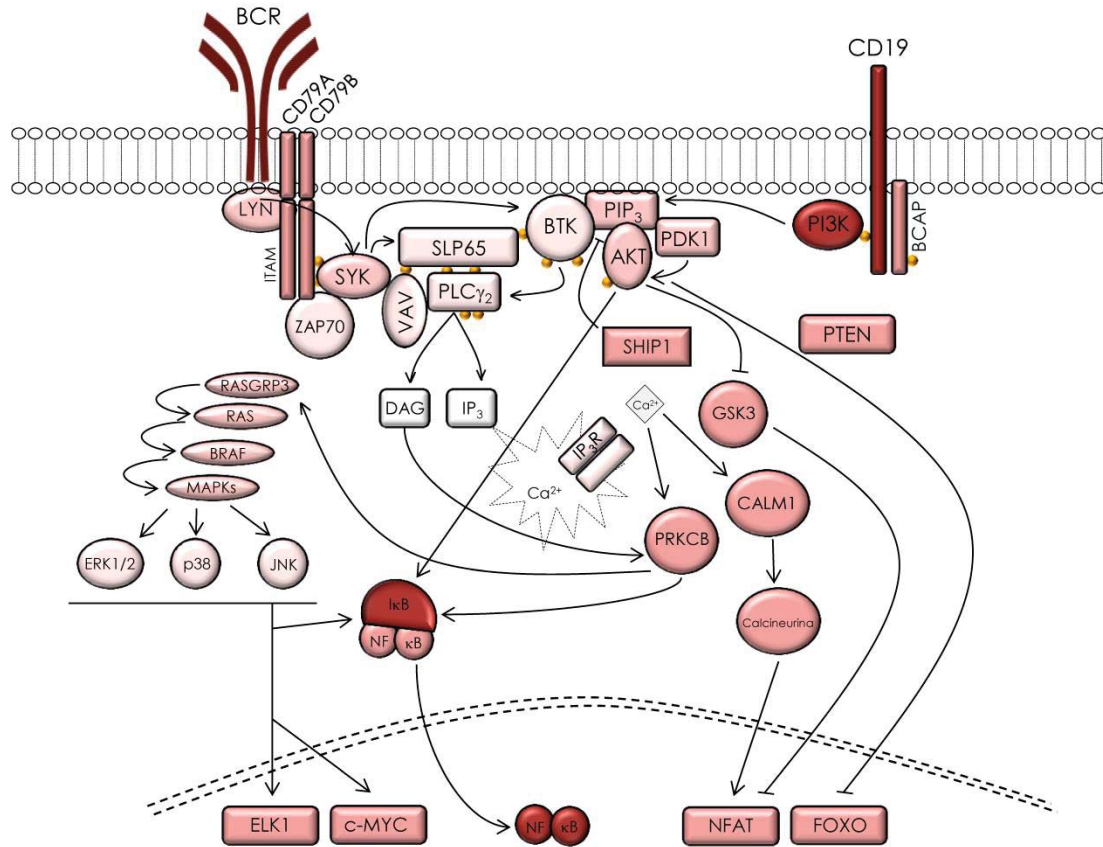
La activación del linfocito B para ejercer sus funciones en la respuesta inmune se realiza a través de distintas rutas de señalización intracelular cuya finalidad es inducir la proliferación, diferenciación, supervivencia o muerte de la célula B. Aunque el fin principal que persiguen las principales vías de señalización B es la respuesta inmune mediada por células B, esta activación puede resultar contraproducente en células B tumorales al contribuir, por ejemplo, a la supervivencia de células genéticamente alteradas y malignas. De todas las vías de señalización intracelular existentes en el linfocito B, las asociadas a BCR son, sin duda, las más relevantes desde el punto de vista fisiopatológico<sup>26-28</sup>.

### *1.2.1. Vías de señalización intracelular a través de BCR*

La activación del BCR puede ser inducida por el reconocimiento específico de Ag o puede ser ligando-independiente (conocida como señalización "tónica" del BCR)<sup>27,29</sup>. En ambos casos, la activación desencadena una cascada de señalización intracelular que presenta efectos directos en la selección de células específicas de Ag en su proliferación y diferenciación, así como en la producción y secreción de Acs<sup>1</sup> (Figura 2).

Las funciones principales del BCR incluyen: i) la transmisión de la señal intracelular producida tras la unión específica del Ag y ii) la mediación de la internalización del Ag para su posterior procesamiento y presentación a las células T CD4<sup>+</sup> <sup>30</sup>. En caso de que estas funciones se vean alteradas, se modificaría la producción de Acs conduciendo a estados de inmunodeficiencia y/u otras enfermedades con origen en la célula B<sup>26,27</sup>. Diversos estudios han demostrado cómo los patrones de fosforilación, entre otras

modificaciones post-traduccionales (PTM, del inglés *post translational modification*), de las proteínas implicadas en la señalización a través del BCR alteran la dinámica de señalización intracelular generando importantes consecuencias en la respuesta inmune que siguen a dicha activación/señalización<sup>31,32</sup>.



**Figura 2. Ruta de señalización intracelular mediada por el receptor de célula B (BCR) para antígeno.**

A nivel estructural, el BCR es un receptor transmembrana ubicado en la superficie exterior de la membrana plasmática de las células B, el cual consta de dos componentes principales: una Ig con un dominio de unión a membrana (no secretable) de un isotipo de cadena pesada concreto (IgD, IgM, IgA, IgG o IgE) con capacidad de reconocer un Ag específico del mismo, y una molécula asociada con capacidad de transmitir y amplificar señales intracelulares (CD79A + CD79B). De esta forma, cada célula B presenta en su superficie muchas copias de un solo tipo de BCR

(único en su secuencia peptídica) de manera que cada célula B tiene especificidad por un único Ag<sup>26</sup>. El reconocimiento específico del Ag a través del BCR promueve la re-localización en su proximidad de todos los componentes del BCR (incluyendo, además de CD79, las moléculas CD19, CD21 y CD81) que amplifican la fosforilación de los motivos de activación del inmunoreceptor basados en tirosinas (ITAM, del inglés *immunoreceptor tyrosine-based activation motif*) en los dominios intracitoplasmáticos de CD79A y CD79B. Posteriormente, las quinasas LYN, FYN y BLK fosforilan de forma secuencial SYK, BTK y PI3K, activando diversas rutas de señalización como la vía de RAS, receptores de citocinas y quimiocinas, la movilización de calcio, la activación de MAPK, PKC $\beta$ , CARD11 y la ruta de NF- $\kappa$ B<sup>28,33</sup>. En conjunto, estos eventos de la cascada de señalización intracelular en la célula B promueven distintos efectos que abarcan desde la activación y proliferación hasta la supervivencia<sup>23</sup>.

#### 1.2.2. Vías de señalización intracelular mediada por receptores de quimiocinas

Los receptores de quimiocinas, fundamentalmente los receptores de quimiocinas C-X-C de tipo 4 (CXCR4), intervienen en procesos de migración de las células implicadas en la respuesta inmune<sup>34</sup>. Además, también desempeñan un papel fundamental en la génesis tumoral, siendo claves en distintas hemopatías malignas<sup>35</sup>. En las células B normales, la estimulación del BCR promueve la internalización de CXCR4 a través de la vía de señalización intracelular mediada por SYK, BTK, PLC $\gamma$ 2 y PKC<sup>36</sup>. Las interacciones que se establecen entre CXCR4 y moléculas del estroma como CXCL12 son críticas en el proceso de desarrollo linfoide, la migración y la adhesión de células B a células del microambiente, la angiogénesis y el desarrollo celular<sup>37</sup>. Estos mecanismos son utilizados por las células tumorales para promover tanto su supervivencia como para favorecer la invasión que se puede desencadenar en el proceso metastásico<sup>38</sup>, estando su sobreexpresión en células de la LLC generalmente asociada a un peor pronóstico<sup>39</sup>.

### 1.2.3. Vías de señalización intracelular a través de receptores tipo Toll (TLR)

Los receptores tipo Toll (TLR, del inglés *Toll-like receptors*) son receptores transmembrana e intracelulares que forman parte del sistema inmune innato y reconocen distintos componentes de microorganismos que presentan motivos altamente conservados; su activación estimula una amplia variedad de respuestas celulares habitualmente asociadas a la inmunidad innata<sup>40,41</sup>. Así, los TLR reconocen patrones moleculares asociados a patógenos (PAMP, del inglés *pathogen-associated molecular patterns*) como los del lipopolisacárido bacteriano o el ácido ribonucleico (ARN) de doble cadena, entre otros. La activación de estos receptores desencadena la señalización intracelular relacionada con respuestas inflamatorias dependientes o no de la proteína adaptadora MyD88 (p.ej. señalización por la vía de NF- $\kappa$ B, activación de CD80, CD86 y CD40), las cuales median la respuesta inmune innata<sup>42</sup>.

## 2. Leucemia linfocítica crónica de célula B

La LLC-B es una neoplasia hematológica a expensas de linfocitos B maduros clonales que presentan una elevada heterogeneidad genómica<sup>43</sup>. La clasificación de neoplasias hematológicas de la Organización Mundial de la Salud (OMS) define la LLC-B como una neoplasia de células B maduras idénticas a las de linfoma linfocítico de células pequeñas (LLCP). Mientras que en la LLC-B las células clonales predominan en sangre periférica y MO, el LLCP se distingue por la presencia de células B clonales que infiltran de forma preferente los ganglios linfáticos y otras regiones linfáticas, asociándose así al desarrollo de adenopatías y/o esplenomegalia<sup>44</sup>.

### 2.1. Epidemiología y etiología de la LLC

La LLC-B es la neoplasia hematológica más prevalente en los países occidentales, suponiendo aproximadamente el 11% de todas las hemopatías malignas y el 25-30% de las leucemias en adultos. En Europa y América del Norte, la incidencia

es de alrededor de 4,2 casos por cada 100.000 habitantes y año, aumentando hasta >30:100.000/año en sujetos con edad superior a 80 años. En los países asiáticos y africanos se considera una enfermedad mucho más rara sin que exista una explicación para estas diferencias<sup>45,46</sup>.

Así, la LLC-B afecta de forma preferente a personas mayores de 70 años, siendo su incidencia superior en hombres que en mujeres (1,7-1)<sup>47-50</sup>. Presumiblemente, la prevalencia y mortalidad de la LLC-B se verán incrementadas en las próximas décadas debido a los cambios demográficos que se esperan en la sociedad<sup>48,51</sup>.

En la actualidad, la etiología de la LLC-B sigue siendo desconocida. La disparidad existente en la incidencia de la enfermedad en distintas áreas geográficas parece deberse a diferencias en factores genéticos y medioambientales, como el estado socio-económico, el acceso al sistema sanitario y la exposición continua y reiterada a agentes potencialmente asociados causantes de la leucemia<sup>52</sup>. Lo que parece claro es que existe una susceptibilidad genética inherente a padecer la LLC-B, con un índice de riesgo de padecer la leucemia de 6 a 9 veces superior entre los familiares de pacientes con LLC-B<sup>45,53-55</sup>.

## **2.2. Diagnóstico de LLC-B**

El diagnóstico de la LLC-B se establece en base a los siguientes criterios<sup>43,48</sup>: i) recuento leucocitario de >5.000 linfocitos B monoclonales por  $\mu\text{L}$  de sangre periférica durante al menos 3 meses (la clonalidad de la población linfocitaria ha de confirmarse mediante inmunofenotipado por citometría de flujo a través de la restricción de la expresión de cadenas IGK o IGL); ii) una morfología característica de células pequeñas, con escaso citoplasma y un núcleo denso con cromatina compacta, carente de nucleolo reconocible, asociado a sombras de Gumprecht y <55% de prolinfocitos<sup>56</sup>; iii) co-expresión del Ag de superficie pan-B, CD19, junto a las proteínas CD5, CD200 y CD23 y niveles bajos de expresión de CD20 e IGK o IGL de superficie. Además, en la revisión de la clasificación de la OMS de 2016<sup>44</sup>, se estableció un nuevo criterio por el cual la LLC-B no puede diagnosticarse como tal en caso de que existan

<5.000 linfocitos B clonales/ $\mu$ L en ausencia de enfermedad extramedular, incluso si se detectan citopenias o síntomas relacionados con la enfermedad.

La presentación clínica de la LLC-B al diagnóstico es muy variable; así, se observan desde formas asintomáticas durante largos periodos de tiempo en las que el diagnóstico se establece habitualmente tras la realización de un hemograma de rutina por otros motivos, hasta pacientes que muestran síntomas relacionados con la enfermedad (p.ej. fatiga, fiebre, pérdida de peso, sudoración nocturna, infecciones recurrentes, adenopatías, hepatomegalia, esplenomegalia y/o anemia hemolítica autoinmune)<sup>46,57,58</sup>.

La linfocitosis monoclonal de célula B (LMB) se considera un estadio asintomático previo a la LLC en el que se observan células habitualmente idénticas a las de la LLC-B en sangre periférica, pero en número <5.000/ $\mu$ L y sin que existan otros síntomas o signos de enfermedad tumoral<sup>59,60</sup>. El índice de progresión de LMB a LLC-B se sitúa en 1-2% de casos al año<sup>61,62</sup>, probablemente asociado a un acúmulo progresivo de lesiones genéticas y epigenéticas, a factores medioambientales o a una combinación de ambos<sup>63,64</sup>. En la actualidad, según el número de célula B clonales en sangre, se distinguen dos subtipos principales de LMB: de bajo y de alto recuento. En el primer caso, el número de linfocitos B monoclonales por  $\mu$ L de sangre es <500 y con una probabilidad de progresión a LLC-B a 5-10 años prácticamente nula. Sin embargo, la LMB de "alto recuento" presenta  $\geq$ 500-5.000 linfocitos B clonales/ $\mu$ L y requiere de un seguimiento rutinario anual; ambos subtipos de LMB presentan características muy similares a la LLC-B, tanto a nivel fenotípico como genético/molecular, especialmente a la LLC-B en estadio Rai 0<sup>65,66</sup>. En una pequeña proporción de casos de LMB, las células clonales muestran características diferentes de la LLC-B y se asemejan a otros síndromes linfoproliferativos crónicos B.

### **2.3. Características clínico-biológicas y factores pronósticos en la LLC-B**

Tal como hemos comentado anteriormente, la LLC es una enfermedad clínicamente muy heterogénea; así, mientras que algunos pacientes necesitan

tratamiento desde el momento del diagnóstico debido al carácter agresivo de la enfermedad que presentan asociada a una supervivencia significativamente más corta, otros padecen formas indolentes de la enfermedad que permanece estable durante largos períodos de tiempo<sup>46</sup>.

En la actualidad, para la evaluación del pronóstico de la LLC se tienen en cuenta, de forma conjunta, varios parámetros clínicos y biológicos (Tabla 1)<sup>67,68</sup>.

**Tabla 1. Parámetros clínico-biológicos asociados al pronóstico de la leucemia linfocítica crónica (LLC) B.**

Factor pronóstico	Condición mal pronóstico
<b>Parámetros clínicos</b>	
Estadio de Binet	Estadio clínico (C)
Estadio de Rai	Estadio clínico avanzado (III/IV)
Morfología linfocitaria	Atípica y presencia de prolinfocitos (>10%)
TDL	Corta (<6 meses)
Patrón de infiltración médula ósea	Patrón difuso
<b>Parámetros biológicos</b>	
<b>Séricos</b>	
β2-microglobulina	Niveles elevados (>3.5 mg/L)
TK1	Niveles elevados
<b>Genéticos</b>	
Estado mutacional de IGHV	No mutado
Reordenamiento IGHV	VH3-21
Alteraciones citogenéticas	17p-/11q-/6q-, alto % de células alteradas
Cariotipo	Complejo (≥3 aberraciones cromosómicas)
Mutaciones genéticas	<i>TP53</i> , <i>SF3B1</i> , <i>NOTCH1</i> , <i>BIRC3</i> , <i>ATM</i> mutados; elevado número de genes mutados
<b>Fenotípicos</b>	
ZAP70	Expresión elevada (≥30%)
CD38	Expresión elevada (≥20%)
CD49d	Expresión elevada (≥30%)
<b>Otros</b>	
Enfermedad mínima residual	Positiva
miARN	Sobre-expresión de miARN-155

TDL, tiempo de duplicación linfocitaria; TK1, timidina quinasa 1; IGHV, región variable de la cadena pesada de la inmunoglobulina; miARN, micro ácido ribonucleico.

### 2.3.1. Factores pronósticos clínicos

En la actualidad, existen dos sistemas de estratificación pronóstica de la LLC-B ampliamente aceptados y que permiten clasificar a los pacientes con LLC en tres grupos pronósticos principales (Tabla 2). El sistema de Rai<sup>69</sup> es quizás el más utilizado en Estados Unidos, mientras que el sistema Binet<sup>70</sup> se suele emplear más comúnmente en el resto del mundo. Ambos se basan en el recuento linfocitario y en exámenes físicos, resultando fáciles de realizar y económicos; todo ello permite que hoy en día sigan siendo ampliamente utilizados y aceptados. Sin embargo, no permiten una clasificación muy precisa de los pacientes con LLC debido a que no consideran el carácter dinámico de la enfermedad. Así, aunque se siguen utilizando ambos sistemas de clasificación, se requieren factores adicionales para una estratificación más precisa de los pacientes. Clásicamente se ha sugerido la posible utilidad de la presencia de linfocitos atípicos, el tiempo de duplicación linfocitaria (TDL)<sup>71</sup> y del patrón de infiltración de MO<sup>72,73</sup>. Sin embargo, otros factores biológicos han adquirido más peso a la hora de refinar la estratificación de los pacientes con LLC-B.

**Tabla 2. Sistemas de estratificación para la leucemia linfocítica crónica (LLC).**

Estadio	Características clínicas
<b>Sistema de Rai</b>	
Bajo riesgo	
Rai 0	Linfocitosis > 15 x 10 <sup>9</sup> linfocitos B monoclonales/L
Riesgo intermedio	
Rai I	Linfocitosis y adenopatías
Rai II	Linfocitosis y hepatomegalia y/o esplenomegalia con/sin adenopatías
Riesgo elevado	
Rai III	Linfocitosis con hemoglobina <11 g/dL con/sin adenopatías/organomegalias
Rai IV	Linfocitosis con <100x10 <sup>9</sup> plaquetas/L con/sin adenopatías/organomegalias
<b>Sistema de Binet</b>	
Binet A (bajo riesgo)	Hemoglobina ≥10.0 g/dL, ≥100x10 <sup>9</sup> plaquetas/L, <3 regiones ganglionares afectadas
Binet B (riesgo intermedio)	Hemoglobina ≥10.0 g/dL, ≥100x10 <sup>9</sup> plaquetas/L, ≥3 regiones ganglionares afectadas
Binet C (riesgo elevado)	Hemoglobina <10.0 g/dL y/o <100x10 <sup>9</sup> plaquetas/L



### 2.3.2. Factores pronósticos biológicos

Dentro de este grupo se incluyen marcadores séricos, genéticos y fenotípicos. A continuación, se comentan brevemente los más relevantes.

#### 2.3.2.1. Factores pronósticos séricos

Dentro de los factores pronósticos séricos se incluyen algunos marcadores presentes y detectables en el suero en forma soluble, como CD23, la enzima timidina quinasa 1 (TK1) y los niveles de  $\beta$ 2-microglobulina, parámetros que se asocian de forma muy significativa con la supervivencia de los pacientes con LLC<sup>74-77</sup>. De hecho, se ha demostrado su elevado potencial predictivo incluso cuando se consideran únicamente los pacientes en el estadio de Binet A<sup>78</sup>.

#### 2.3.2.2. Factores pronósticos genéticos

Dentro de los factores genéticos destacan: i) el estado mutacional y el tipo de reordenamiento de las regiones variables de las cadenas pesadas de las Ig (IGHV), ii) las alteraciones citogenéticas, incluida la presencia de un cariotipo complejo, y iv) las mutaciones genéticas adquiridas por la célula tumoral.

La expresión en las células leucémicas de una Ig que presente  $\geq 2\%$  o  $< 2\%$  de mutaciones somáticas a nivel de los genes IGHV es uno de los principales parámetros pronósticos que se tiene en cuenta en la actualidad en la LLC. Así, el estudio mutacional de los genes IGHV permite distinguir dos grupos de pacientes: aquellos que presentan secuencias IGHV no mutadas y las que tienen secuencias IGHV mutadas, en función de si presentan o no  $\geq 98\%$  de homología con la línea germinal, respectivamente<sup>79,80</sup>. Además, independientemente del estado mutacional, la selección de la versión VH3-21 del gen IGHV en las células clonales de la LLC constituye también un marcador de mal pronóstico<sup>81,82</sup>; al igual que ocurre con la versión VH3-48 en los pacientes no mutados<sup>83</sup>. Por el contrario, un mejor pronóstico se relaciona con las versiones VH4-34 y VH2-30<sup>84,83</sup>.

Además del estado mutacional y subtipo del gen IGHV, la presencia de distintas alteraciones genéticas se ha asociado también con el pronóstico de la LLC. En

términos generales, >80% de las LLC presentan una o más alteraciones genéticas detectables mediante hibridación *in situ* fluorescente (FISH, del inglés *fluorescence in situ hybridization*) o técnicas de análisis moleculares. La alteración más común (>50% de los casos) es la pérdida de la región terminal del brazo largo del cromosoma 13 - del(13q14.1) – que, junto con un cariotipo normal, se relaciona con un pronóstico favorable. Además, pueden observarse pérdidas del brazo largo de los cromosomas 11 [del(11q)] (18% de casos, asociado a delección del gen *ATM*)<sup>85</sup> y 6 [del(6q)], y del brazo corto del cromosoma 17 [del(17p)] (7% de los casos; asociado a la pérdida del supresor tumoral *TP53*)<sup>86,87</sup>, todas ellas – especialmente del(17p) – asociadas a un pronóstico adverso<sup>87</sup>. Asimismo, 16% de las LLC presentan trisomía del cromosoma 12, alteración que se ha asociado con un riesgo intermedio<sup>88</sup>.

Independientemente del tipo de alteraciones genéticas, la presencia simultánea de 3 o más alteraciones cromosómicas al diagnóstico se asocian a estadios más avanzados y peor pronóstico<sup>89-91</sup>.

Con la aplicación de las nuevas técnicas de secuenciación masiva, en los últimos años se ha incrementado notablemente el conocimiento acerca de la variabilidad genética presente en la LLC<sup>92-94</sup>. Así, además de las alteraciones cromosómicas antes referidas, se han detectado mutaciones somáticas en genes relacionados con el daño en el ácido desoxirribonucleico (ADN) (p.ej. *TP53* y *ATM*), el procesamiento de ARN mensajero (ARNm) (p.ej. *SF3B1* y *XPO1*), la modificación de cromatina (p.ej. *HIST1H1E*, *CHD2* y *ZMYM3*), la señalización intracelular mediada por Wnt y Notch (p.ej., *NOTCH1*) en vías inflamatorias (p.ej. *MYD88*) de señalización intracelular específica de célula B (p.ej. *EGR2*) y de la transcripción (p.ej. *BRAF*)<sup>95</sup>. Además, algunas mutaciones en genes conductores se han relacionado con la LLC, como los que involucran a distintos elementos de la ruta de Wnt y que afectan a la viabilidad celular<sup>96</sup>, a *POT1* que tiene un papel en la protección de los telómeros, a *SF3B1* relacionado con el *splicing* aberrante de ARN y una respuesta a daño del ADN alterado<sup>97,98</sup>, y a *SAMHD1* que está implicado en la regulación de deoxinucleótidos a

nivel intracelular<sup>99</sup>. Adicionalmente, se han descrito mutaciones puntuales en *PTPN11*, *PAX5*, *BIRC3*, *IKZF3* y *MGA* que se consideran posibles marcadores con relevancia clínica en LLC<sup>100,101</sup>.

### 2.3.2.3. Factores pronósticos fenotípicos

El auge de la proteómica durante los últimos años ha permitido la identificación y cuantificación de la expresión de algunas proteínas en la célula B de la LLC con un posible impacto pronóstico. Entre ellas, destacan CD38, ZAP70 y CD49d (subunidad  $\alpha$  de la integrina  $\alpha 4/\beta 7$ )<sup>102</sup>. En relación a ZAP70 y CD38, diferentes estudios han demostrado de forma recurrente su relación con un peor pronóstico, independientemente del estado mutacional de los genes *IGHV*<sup>79,103-106</sup>. A pesar de ello, su uso como factores predictivos sigue siendo controvertido ya que existen ciertas limitaciones técnicas para una identificación robusta y reproducible de las células positivas para ZAP70 y CD38<sup>107</sup>. Por otra parte, altos niveles de expresión de CD49d se han correlacionado con tiempos de supervivencia más cortos en la LLC-B, especialmente dentro de los casos que presentan mutaciones en su BCR<sup>108-110</sup>.

### 2.3.2.4. Otros factores pronósticos

Además de todos los factores pronósticos mencionados anteriormente, en los últimos años se han descrito asociaciones entre otros marcadores de LLC y del comportamiento de la enfermedad, incluyendo, por ejemplo, los relacionados con la expresión de algunos micro ARNs (miARNs) concretos. Así, la LLC fue la primera enfermedad humana en la que se identificó una asociación entre la expresión de determinados miARNs y el pronóstico. Específicamente, el miARN-15a (dirigido a *BCL2*) y el miARN-16-1 (dirigido a *MCL1*) se encontraron delecionados, alterados o subexpresados en ~60% de los pacientes con LLC<sup>111-113</sup>. Finalmente, se ha demostrado que la pérdida o expresión reducida de miR-29a/b, miR-29c, miR-34b, miR-181b y miR-3676 conduce a un incremento de la expresión de *TCL1A*, gen partícipe en la ontogenia de la LLC<sup>114,115</sup>. Por el contrario, niveles de expresión elevados de miR-155 se

han relacionado con un incremento en la señalización del BCR, una mayor proliferación de la célula B tumoral y la génesis de esta leucemia<sup>116,117</sup>.

### 2.4. Origen celular de la LLC

Inicialmente, la LLC-B se consideraba una enfermedad que resultaba de la acumulación de células B funcionalmente incompetentes e inactivas<sup>118</sup>. Esta hipótesis, basada en estudios morfológicos, fue rebatida tras la realización de estudios *in vivo* que demostraron un mayor dinamismo y proliferación de las células tumorales frente al carácter acumulativo descrito con anterioridad para esta enfermedad<sup>119</sup>. Por otra parte, las aproximaciones genómicas permitieron refutar la hipótesis de que las células de la LLC se originaban a partir de las células B del manto folicular, hipótesis basada en la expresión común de los marcadores de superficie de membrana CD5 y CD23<sup>120-123</sup>. En resumen, todos los avances tecnológicos y metodológicos ocurridos en las últimas décadas han permitido refutar muchas de las hipótesis planteadas hasta el momento acerca del origen de la LLC; sin embargo, aún no se ha llegado a dilucidar el verdadero origen/es de estas células tumorales.

El incremento en el conocimiento de las mutaciones en los genes IGHV (que permite estratificar a los pacientes en grupos con diferente pronóstico) junto con la identificación de BCRs polirreactivos y/o autorreactivos – asociados a alteraciones de las vías de señalización intracelular – han servido para postular que la LLC podría tener un origen diverso y proceder de más de una célula distinta, a diferencia de lo que inicialmente se había planteado<sup>123</sup>. De hecho, en relación al estado mutacional de los genes IGHV, se ha postulado un origen dual que permitiría distinguir dos tipos de células LLC (mutadas y no mutadas en genes IGHV) que presentarían distintas historias celulares de diferenciación y encuentro con Ag<sup>123</sup>. De tal forma, el repertorio de genes IGHV no sería generado de manera aleatoria pudiendo distinguir células de la LLC mutadas vs. no mutadas, y células de la LLC vs. células normales<sup>124,125</sup>. No obstante, el análisis de los perfiles de expresión génica mediante microarrays contradice esta

hipótesis sugiriendo un origen único en el que factores secundarios (genéticos y no genéticos) serían los causantes de la heterogeneidad de la enfermedad<sup>126,127</sup>.

Con el objetivo de dirimir estos modelos enfrentados, se propuso una nueva teoría unificadora, mediante la cual las células de la LLC (mutadas y no mutadas en los genes IGHV) derivarían de células B de la zona marginal (ZM)<sup>128</sup> en las que ocurriría el/los procesos de transformación maligna incluyendo las mutaciones en los genes IGHV, la recombinación VDJ y cambio de clase de Ig (con origen en el CG)<sup>129,130</sup> y la expresión de moléculas y genes similares a los expresados por las células B de memoria derivadas del CG<sup>126,131</sup> y que de forma global presentarían características fenotípicas y moleculares muy similares a las detectadas en las células de la LLC. Sin embargo, esta teoría no permitía explicar algunas diferencias entre esas células de la ZM y las células de la LLC, entre las que se encuentra la expresión de las proteínas de membrana CD5 y CD23 por parte de las células tumorales<sup>132</sup>; asimismo, tampoco permitía explicar la génesis de las células LLC no mutadas<sup>133</sup>.

A pesar de ello, parece evidente que las células de la LLC, independientemente del estado mutacional del gen IGHV, tienen su origen en una célula que probablemente haya reconocido Ag, existiendo diferencias evidentes en la activación del BCR entre los casos con IGHV mutado y no mutado. Mientras que en los casos no mutados las células leucémicas podrían estar estimuladas de forma crónica y persistente a través del BCR, las células LLC mutadas presentarían un comportamiento anérgico (funcionalmente inactivas)<sup>133,134</sup>. De hecho, se considera que la señalización a través del BCR constituiría la principal ruta de activación en la célula tumoral de la LLC<sup>28</sup>.

Con todo ello, el origen celular de la LLC sigue siendo incierto, aunque fenotípicamente cercano a una célula B inmadura/naive temprana, existiendo probablemente un modelo progresivo en el que diferentes eventos a la largo de la diferenciación de la célula B dan lugar a subpoblaciones celulares con presencia de ciertos rasgos patológicos, pudiendo existir un origen múltiple de la LLC que permitiría

explicar la elevada heterogeneidad de la enfermedad. Así, la subpoblación B donde ocurra la alteración final podría determinar las características de la leucemia al diagnóstico, sin que ello suponga que las alteraciones genéticas que han conducido a la transformación maligna de esa célula no hayan ocurrido en un punto anterior de la diferenciación (p.ej. en las células madre hematopoyéticas pluripotentes)<sup>123,135</sup>.

En resumen, parece evidente que la estimulación antigénica juega un papel clave en el origen de la LLC, por lo que profundizar en el estudio de la señalización mediada por el BCR, junto con la evaluación de los Ags más relevantes a la hora de inducirlo, podría tener un gran impacto en el conocimiento de la enfermedad y, por ende, en la identificación de nuevas dianas terapéuticas y el consecuente desarrollo de nuevos tratamientos.

### **2.5. Papel del microambiente tumoral en la LLC**

Los centros de proliferación de la LLC se encuentran en los ganglios linfáticos y en la MO y constituyen el lugar donde la célula leucémica interacciona y contacta con otras células próximas<sup>136</sup>. Entre éstas células presentes en el microambiente tumoral, se encuentran las células del estroma de la MO, las células nodrizas derivadas de monocito, las CDF, las células endoteliales y las células T<sup>137</sup> (Figura 3). En el contacto entre la célula tumoral de la LLC y las células del estroma interviene la unión de VCAM1 a la integrina CD49d, así como interacciones quimiocina-receptor de quimiocina (CXCL12-CXCR4), resultando de dichas interacciones un efecto protector de la célula de la LLC frente a la apoptosis (espontánea e inducida por fármacos), así como la inducción de la activación de la ruta de NF- $\kappa$ B<sup>138-140</sup>.

De forma similar, en la interacción de las células leucémicas con las células nodrizas derivadas de monocito intervendrían las quimiocinas CXCL12 y CXCL13, con efecto antiapoptótico<sup>141,142</sup>. De forma paralela, la supervivencia de la célula tumoral se ve favorecida también por los contactos celulares mediados por las interacciones entre BAFF y BCMA (=TNF17), TACI-BAFFR, CD38-CD31, CD100-plexina B1 y la vía de NOTCH<sup>143</sup>. Por otra parte, las CDF y las células endoteliales son fundamentales para la

migración de la célula leucémica, mediante distintos mecanismos de interacción: CXCR5-CXCL13, linfotoxina  $\alpha$ - receptor de linfotoxina  $\beta$  y/o CD100-plexina B<sup>137</sup>, entre otros. Finalmente, la interacción con los linfocitos T sucede principalmente a través de CD40-CD40L y CD100-plexina B<sup>144</sup>.

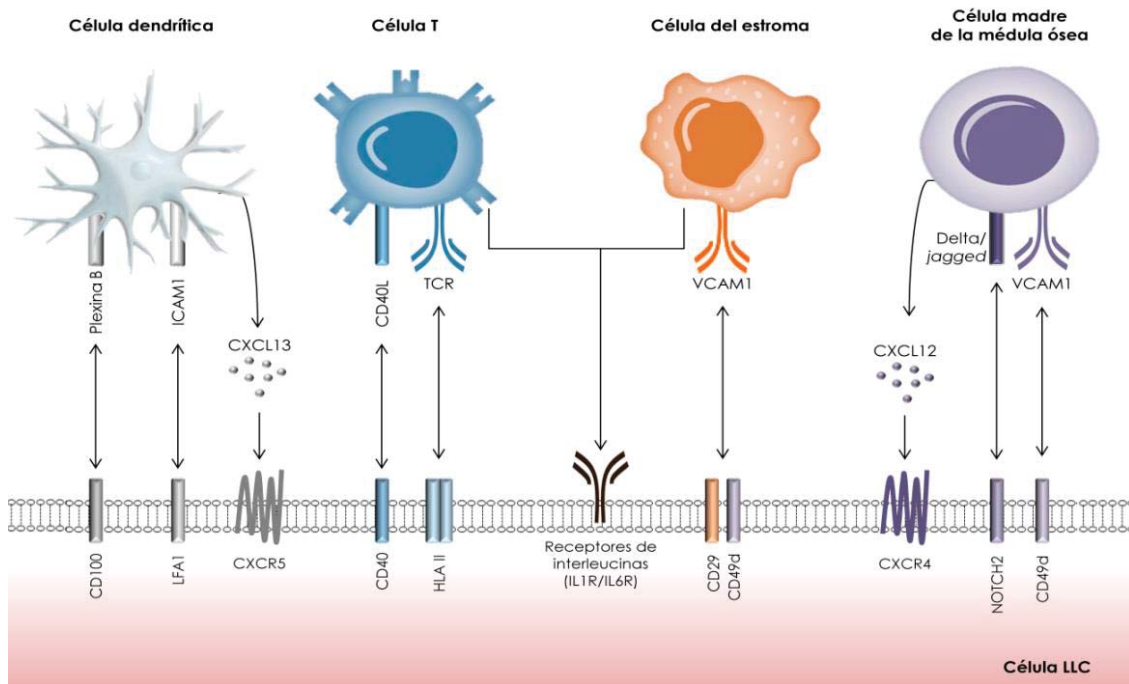


Figura 3. Elementos celulares más relevantes del microambiente de la célula tumoral de la leucemia linfocítica crónica (LLC).

## 2.6. Tratamiento de la LLC

En general, la presencia de anemia o trombocitopenia, de adenopatías y/o síntomas constitucionales asociados a LLC activa (p.ej. sudoración nocturna, fatiga, pérdida de peso o fiebre en ausencia de infección) determinan el inicio de la administración de la terapia antitumoral en la LLC<sup>43</sup>, siendo la presencia de las alteraciones citogenéticas del(17p) y/o la mutación de *TP53*, los marcadores que acaban por determinar la terapia más adecuada.

El tratamiento de los pacientes con LLC puede incluir quimioterapia, combinación de quimioterapia e inmunoterapia o, más recientemente, fármacos dirigidos a rutas de señalización relacionadas con el crecimiento y/o supervivencia de las células tumorales (Figura 4)<sup>45</sup>. Clásicamente, el tratamiento de la LLC se ha basado

en la administración de quimioterapia con análogos de las purinas (fludarabina, pentostatina o cladribina)<sup>145,146</sup> y agentes alquilantes (clorambucilo, ciclofosfamida o bendamustina)<sup>147</sup>. Estos tratamientos pueden causar mielosupresión e inmunosupresión asociándose a un riesgo incrementado de infecciones y/o neoplasias secundarias. Actualmente, los protocolos estándar para el tratamiento de la LLC, tanto en pacientes más jóvenes como en adultos de edad avanzada, se basan en el uso combinado de quimioterapia y de agentes inmunoterápicos, sobre todo Acs monoclonales (p.ej. frente a CD20 como rituximab, obinutuzumab y ofatumumab)<sup>148</sup>. Las combinaciones más empleadas suelen incluir fludarabina, ciclofosfamida y rituximab (FCR)<sup>149</sup>, bendamustina y rituximab<sup>150</sup> u obinutuzumab<sup>151</sup>, fludarabina junto a rituximab<sup>149</sup>, y clorambucilo con ofatumumab<sup>152</sup>.

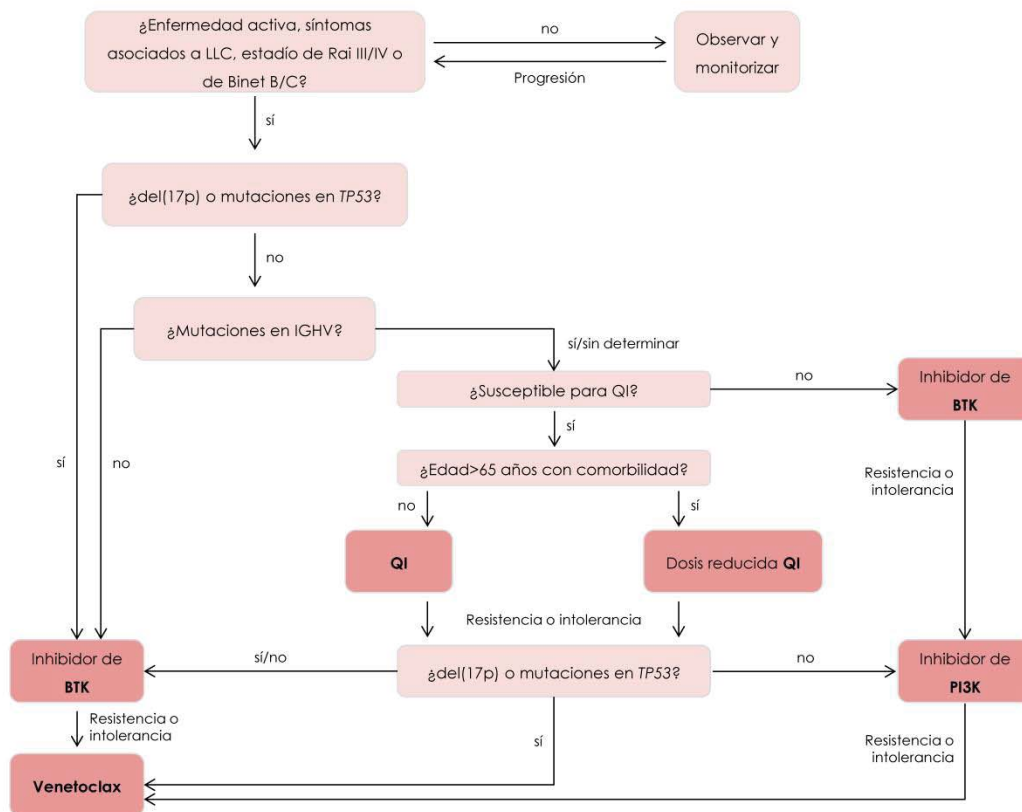


Figura 4. Algoritmo genérico y actual para la toma de decisiones terapéuticas en pacientes con leucemia linfocítica crónica (LLC). QI, quimioinmunoterapia.

Aunque con los protocolos de inmunquimioterapia se ha incrementado notablemente la tasa de respuestas al tratamiento, en la mayor parte de los pacientes



la enfermedad recae. Por ello, en los últimos años se siguen ensayando nuevas terapias, entre las que destacan aquellas que emplean inhibidores de la señalización intracelular mediada por el BCR. Hasta el momento, se han evaluado 3 clases de inhibidores de quinasas implicadas en la señalización intracelular del BCR: inhibidores frente a BTK (como ibrutinib y acalabrutinib)<sup>153-155</sup>, PI3K (como idelalisib, duvelisib, TGR-1022 y ACP-319)<sup>156,157</sup> y SYK (como fostamatinib)<sup>158,159</sup>. Además, merece destacar el uso de inhibidores de BCL-2, como venetoclax, que reducen la capacidad de BCL-2 para secuestrar la molécula pro-apoptótica BIM, siendo este fármaco especialmente efectivo en pacientes con enfermedad en recaída y/o refractaria, así como en pacientes con del(17p) y recidiva<sup>160</sup>.

A pesar de todos los tratamientos disponibles hoy en día, la LLC sigue siendo una enfermedad considerada como incurable, lo que hace necesario el seguimiento de los pacientes a lo largo de su vida aun cuando responden al tratamiento administrado. Además, dichos pacientes presentan un riesgo de desarrollar neoplasias secundarias de 2 a 7 veces superior al de la población sana de la misma edad. Además, la LLC puede evolucionar o sufrir una transformación a linfoma B difuso de células grandes (LBDCG) (también conocida como transformación a síndrome de Richter) o a linfoma de Hodgkin (LH), lo que ocurre en cerca del 2-7% de los casos a lo largo del curso de la enfermedad<sup>161,162</sup>. Asimismo, a lo largo de la vida, estos pacientes también presentan una incidencia más elevada de tumores sólidos, como el cáncer de piel, sarcomas, cáncer de pulmón, renal y de próstata<sup>163</sup>.

### **3. Técnicas proteómicas de alto rendimiento para el estudio del linfocito**

#### **B normal y patológico**

Las aproximaciones basadas en el estudio y caracterización de las proteínas – técnicas proteómicas - han experimentado un notable auge en los últimos años. A continuación, se describen algunos conceptos clave relacionados con la proteómica.

### 3.1. Definición de proteoma

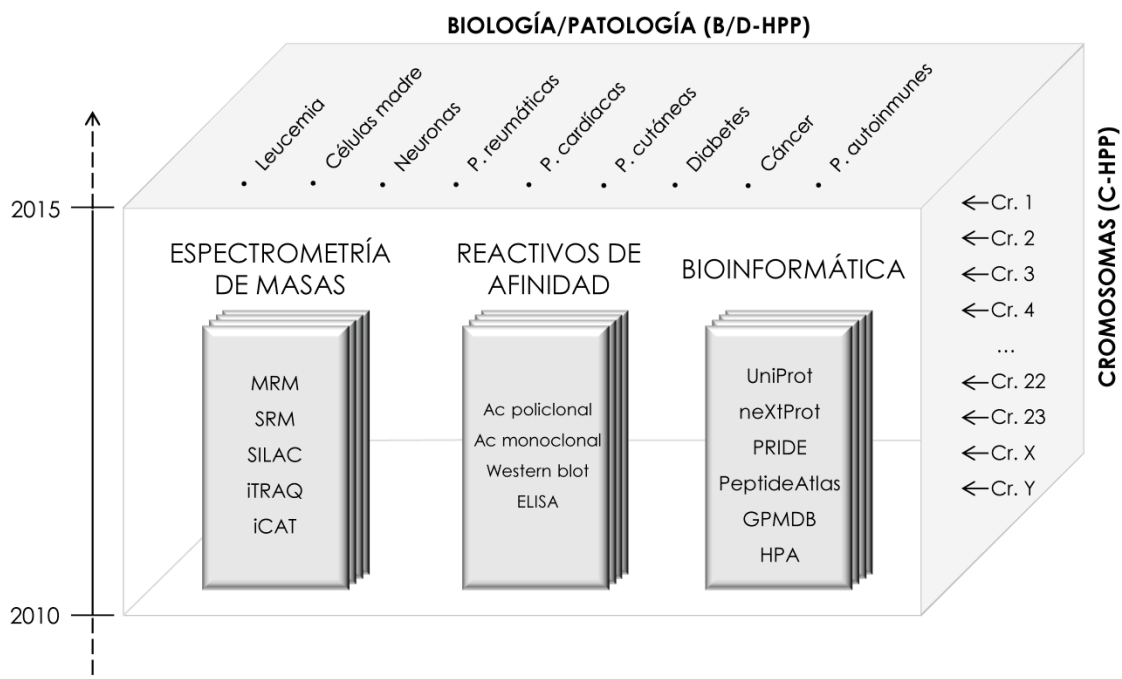
El término "proteoma", resultante de la fusión de "**proteína**" y "**genoma**", fue acuñado en 1995 por el genetista Marc Wilkins<sup>164</sup> y hace referencia al conjunto de proteínas codificadas por el genoma que se expresan en un determinado momento y condiciones en una célula, tejido u organismo<sup>165</sup>.

La "proteómica" constituye el estudio exhaustivo, a gran escala, del proteoma, permitiendo su definición mediante el análisis de su función, PTMs, localización y compartimentalización celular, interacciones proteína-proteína, niveles de expresión y estructura tridimensional<sup>165</sup>. Todos estos aspectos funcionales y estructurales de las proteínas, a su vez, conforman redes dinámicas que varían durante los procesos biológicos como la diferenciación, proliferación y migración celulares, la endocitosis y el desarrollo, entre otros procesos<sup>166</sup>.

El auge de la proteómica se produjo tras la conclusión del Proyecto Genoma Humano<sup>167</sup> en 2003, año en el que los 20.000 – 25.000 genes codificantes de proteínas que componen el genoma humano fueron identificados y cartografiados por primera vez de forma global<sup>168</sup>. Tras la era genómica, la comunidad científica fue consciente de que los genes, por sí mismos, no eran suficientes para explicar todas las funciones biológicas<sup>169</sup>. Si bien es cierto que la información genómica es esencial, las proteínas son los principales elementos moleculares efectores de las células y, por tanto, su estudio reviste mayor relevancia, aportando por ejemplo, información del estado celular en un momento concreto. Mientras que el genoma permanece prácticamente constante a lo largo de la vida de una célula y, salvo excepciones, no existen variaciones significativas entre distintos tipos celulares, el proteoma presenta un mayor dinamismo sufriendo notables cambios en el tiempo en relación directa con el estado funcional de la célula o en respuesta a señales intra- o extra-celulares<sup>170</sup>. De tal forma, la proteómica permite el estudio completo y exhaustivo de las proteínas proporcionando una lectura clara y precisa de los sistemas celulares y de sus interrelaciones<sup>171</sup>.

En este contexto y con esta finalidad nace en 2010 el Proyecto Proteoma Humano (HPP, de sus siglas en inglés *Human Proteome Project*) de la mano de la Organización del Proteoma Humano (HUPO, de sus siglas en inglés *Human Proteome Organization*)<sup>172</sup>. Tal y como HUPO define ([www.hupo.org/human-proteome-project](http://www.hupo.org/human-proteome-project)), la misión del HPP es "la caracterización sistemática y pormenorizada de los 20.300 genes conocidos del genoma humano para generar el mapa proteico humano que permita dilucidar las funciones biológicas y moleculares de las células para mejorar el diagnóstico y tratamiento de las enfermedades".

Dicho proyecto HPP engloba dos iniciativas: C-HPP y B/D-HPP (de sus siglas en inglés *chromosome-centric* y *biology/disease*, respectivamente) fundamentadas en tres estrategias metodológicas: espectrometría de masas (MS, de sus siglas en inglés *mass spectrometry*), reactivos de afinidad (p.ej. Ac) y bioinformática<sup>173</sup> (Figura 5).



**Figura 5. Esquema de la organización del Proyecto Proteoma Humano (HPP).** Tres pilares (espectrometría de masas, reactivos de afinidad y bioinformática) sustentan el HPP que se aborda desde dos iniciativas (B/D-HPP y C-HPP). [Adaptado de Legrain *et al.*<sup>173</sup>].

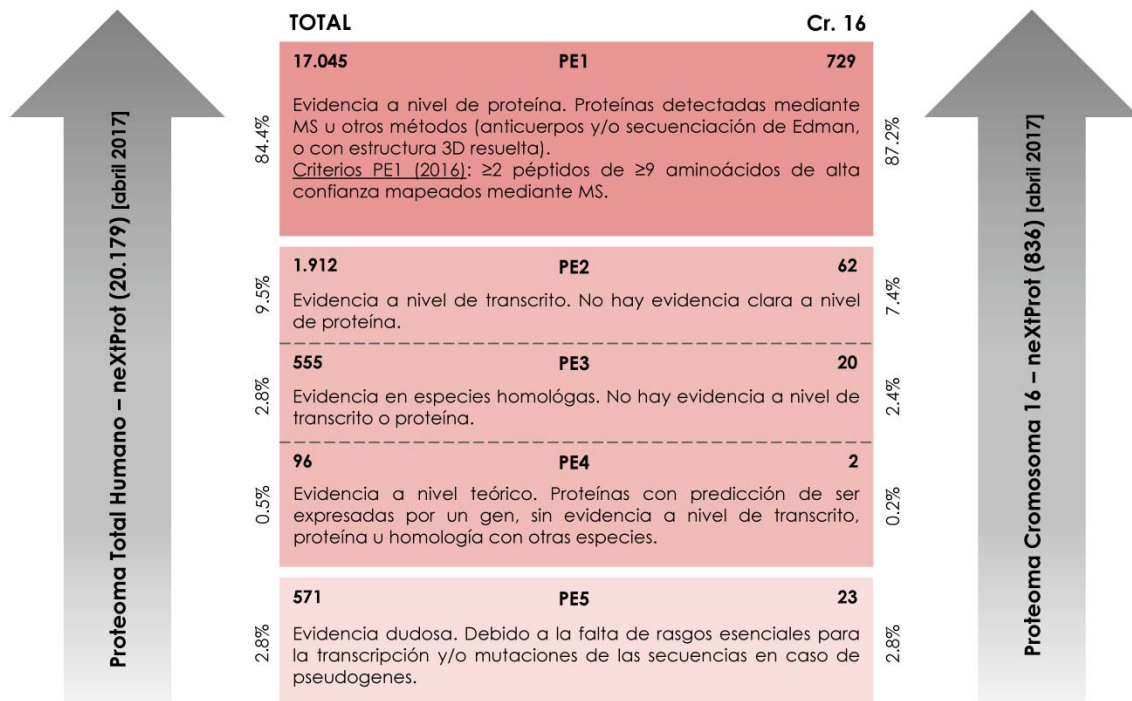
Ac, anticuerpo; P., patología; Cr., cromosoma; MRM, Multiple Reaction Monitoring; SRM, Selected Reaction Monitoring; SILAC, Stable Isotope Labeling by Amino acids in Cell culture; iTRAQ, Isobaric Tag for Relative and Absolute Quantitation; iCAT, Isotope-Coded Affinity Tag; ELISA, Enzyme-Linked ImmunoSorbent Assay; PRIDE, PRoteomics IDentifications Database; GPMDB, Global Proteome Machine and Database; HPA, Human Protein Atlas.

A continuación se describen algunas de las características más significativas de cada iniciativa.

• **Iniciativa C-HPP.** En primer lugar, la iniciativa basada en cromosomas (C-HPP)<sup>174-176</sup> se ha organizado asignando a cada consorcio participante (procedente de un total de 21 países distintos) el mapeo y anotación sistemáticos de todas las proteínas codificadas en los genes de cada cromosoma humano (incluyendo aquellas proteínas codificadas en el genoma mitocondrial). Se estima que alrededor de 20.179 genes codificantes de proteínas conforman el proteoma humano (fuente: base de datos neXtProt, versión de abril de 2017)<sup>177</sup>. Por su parte, el consorcio español del HPP (SpHPP) - englobado dentro del HPP, de la plataforma ProteoRed-ISCI y de la Plataforma de Recursos Biomoleculares y Bioinformáticos (PRB2) – tiene asignada la caracterización proteómica del cromosoma 16<sup>178,179</sup>. La iniciativa española está construida sobre una base multidisciplinar compuesta por 15 grupos científicos organizados en 5 secciones de trabajo denominadas i) microarrays de proteínas/Acs/proteínas recombinantes, ii) estandarización de péptidos y expresión de proteínas, iii) secuenciación y caracterización proteómica mediante *Selected Reaction Monitoring (SRM)* y *Multiple Reaction Monitoring (MRM)*, iv) salud, clínica y bioinformática, y v) biobancos<sup>179</sup>.

Dentro del HPP, el consorcio español se encarga de la caracterización sistemática de todas las proteínas codificadas en el cromosoma 16. Dicho cromosoma contiene alrededor de 90 millones de pares de bases, representando casi el 3% del ADN total presente en el genoma humano. De acuerdo a la información disponible en Ensembl (versión 88, [www.ensembl.org](http://www.ensembl.org)), el cromosoma 16 contiene 2.373 genes, de los cuales 865 están identificados como genes codificantes de proteínas con una estimación inicial de 305 proteínas *missing* (proteínas sin evidencia experimental)<sup>180</sup>. Actualmente, y según la base de datos neXtProt<sup>177</sup> (versión de enero de 2017), 836 proteínas están codificadas en el cromosoma 16: 729 validadas (*protein evidence 1*, PE1), 84 pendientes de validación (= proteínas *missing*) [62 con evidencias a nivel

transcriptómico; PE2; 20 inferidas mediante homología con otras especies, PE3; 2 proteínas con evidencia teórica, PE4], y 23 proteínas cuya evidencia es dudosa (PE5) (Figura 6).



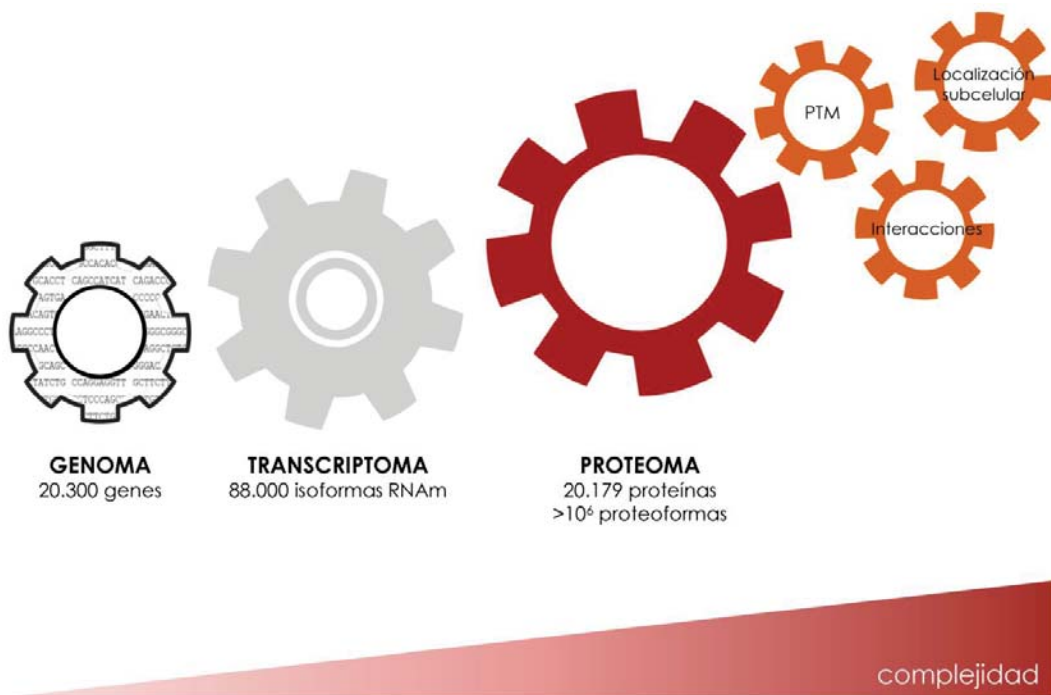
**Figura 6. Definiciones de evidencias de las proteínas – protein evidence (PE) - según la base de datos neXiProt.** [Adaptado de Baker *et al.*<sup>180</sup>]. MS, espectrometría de masas.

• **Iniciativa B/D-HPP.** En segundo lugar, la iniciativa enfocada a la descripción de los proteomas de muestras biológicas y/o estados patológicos concretos (B/D-HPP)<sup>181</sup> pretende i) la determinación de estrategias espectrométricas específicas y validadas para cuantificar la abundancia de todas las proteínas relacionadas o implicadas en una situación biológica concreta, ii) la identificación/diseño de reactivos de afinidad para la detección específica de proteínas, y iii) la mejora en el empleo de dichas estrategias y reactivos en un formato accesible para la comunidad científica. De tal forma, la iniciativa B/D-HPP persigue proporcionar las herramientas oportunas para el estudio de las proteínas relevantes en áreas concretas de la biología y/o enfermedades específicas<sup>182,183</sup>. Así, se han establecido equipos B/D-HPP enfocados al análisis proteómico de diferentes enfermedades, grupos de enfermedades o situaciones fisiológicas particulares, o tipos de muestras biológicas concretas como,

por ejemplo, la diabetes<sup>184,185</sup>, cáncer<sup>186-188</sup>, enfermedades infecciosas, enfermedades reumáticas<sup>189</sup>, hígado<sup>190,191</sup>, ojos<sup>192,193</sup>, cerebro<sup>191,194</sup>, glicoproteínas<sup>195</sup>, células madre<sup>183</sup>, plasma<sup>196,197</sup>, enfermedades cardiovasculares<sup>191,198,199</sup>, orina y riñón<sup>191,200,201</sup>, entre otros.

### 3.2. Complejidad del proteoma

Tal como hemos mencionado anteriormente, el estudio de los proteomas reviste mayor complejidad que el análisis genómico, debido principalmente a la mayor complejidad química de las proteínas, junto a la gran diversidad de modificaciones que pueden ocurrir a nivel de las secuencias aminoacídicas y/o la estructura de una proteína. Estas modificaciones pueden suceder en los ARNm que codifican la proteína (*splicing* alternativo)<sup>202</sup> o durante el procesamiento post-traducciona<sup>203</sup> que puede conducir a alteraciones de la función, localización o actividad de las proteínas. Tal es así que se estima que el genoma humano (20.300 genes) da lugar a ~88.000 isoformas de ARNm distintas (transcriptoma humano), las cuales pueden convertirse en más de 1 millón de proteoformas diferentes (proteoma humano)<sup>170</sup> (Figura 7).



**Figura 7. Complejidad de los sistemas biológicos.** Desde la información genómica y transcriptómica a la proteómica en la que se evalúan la localización subcelular, las modificaciones post-traduccionales (PTM) y las interacciones de las proteínas.

En relación a las PTM, éstas se producen por la modificación covalente o enzimática de las proteínas durante o tras la biosíntesis proteica. Como resultado, puede ocurrir la alteración de un grupo químico funcional existente o la introducción de un grupo químico nuevo (p.ej. fosforilación, acetilación, O- y N-glicosilación, deamidación, acetilación, metilación, ubiquitinación, adición de ácido pirrolidona carboxílico y sulfatación, entre otros)<sup>204</sup>. Asimismo, las PTMs pueden referirse a la formación de puentes disulfuro entre residuos de cisteína o a la fragmentación de uniones peptídicas y proteólisis o degradación peptídica<sup>203</sup>.

Estadísticamente, la PTM detectada más frecuentemente a nivel intracelular es la fosforilación<sup>205</sup>. Ésta consiste en la adición de un residuo fosfato ( $\text{PO}_3^-$ ) a un grupo químico de la cadena aminoacídica de la proteína - normalmente serina (Ser), treonina (Thr) y tirosina (Tyr) - y generalmente conduce a la activación de dicha proteína. Quinasas y fosfatasas son las responsables de fosforilar y defosforilar las proteínas, respectivamente. La formación de una fosfoproteína (proteína fosforilada) ocurre cuando el grupo fosfato reacciona con el grupo -OH de la Ser, Thr o Tyr durante una reacción de esterificación (generando enlaces fosfodiéster)<sup>206</sup>. Estas proteínas fosforiladas son de vital importancia en las rutas de señalización intracelular, donde existe un gran dinamismo en los procesos de fosforilación/defosforilación que resulta en redes de activación/desactivación complejas<sup>207</sup> donde es necesaria, no sólo la identificación del conjunto de fosfoproteínas presentes (fosfoproteoma), sino también su cuantificación y determinación del número de aminoácidos fosforilados, así como la estequiometría y cinética de fosforilación<sup>208</sup>.

### **3.3. Proteómica: áreas de estudio y técnicas de análisis**

Según las áreas de estudio<sup>165</sup>, la proteómica puede clasificarse en proteómica estructural, de expresión diferencial y funcional.

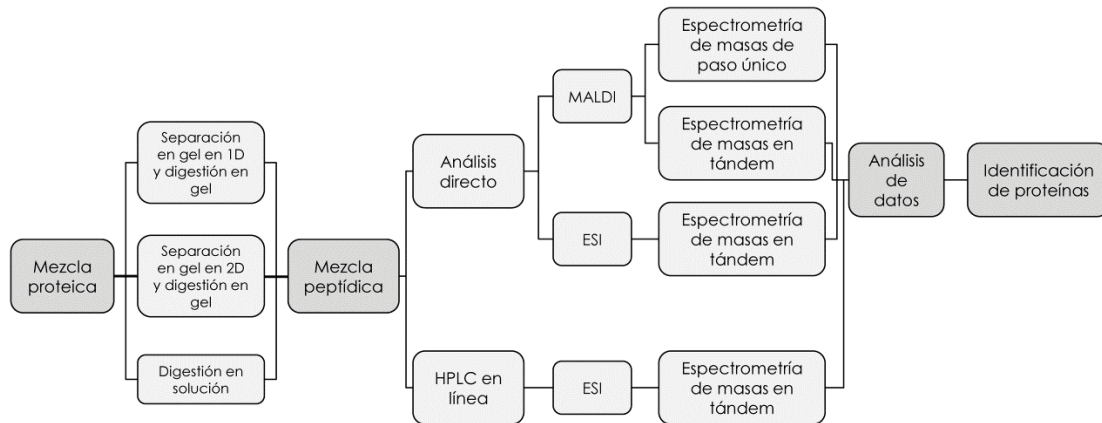
La proteómica estructural se encarga de la caracterización de la estructura tridimensional y la secuencia aminoacídica de las proteínas expresadas en un complejo proteico u orgánulo, determinando su localización subcelular y

caracterizando las interacciones proteína-proteína. A su vez, la proteómica de expresión diferencial compara la expresión de proteínas dentro de un proteoma completo o de subproteomas de diferentes muestras permitiendo la identificación de nuevas proteínas implicadas/relacionadas con la transducción de señales o bien que sean específicas de enfermedades (biomarcadores) o situaciones fisiológicas particulares. Finalmente, la proteómica funcional es la encargada de la evaluación de actividades proteicas tras el aislamiento de proteínas específicas, con el fin de determinar mecanismos de actuación en distintas situaciones patológicas, interacciones proteína-fármaco y señalización intracelular proteica.

Asimismo, dependiendo de cómo se aborde el estudio de las proteínas, la proteómica se clasifica en dos grupos principales: "top-down" y "bottom-up". La proteómica "top-down" caracteriza individualmente cada proteoforma. Tanto experimentalmente como computacionalmente es una estrategia compleja debido a la dificultad que supone analizar proteínas completas/enteras, en lugar de péptidos, y a la falta de métodos de aislamiento de proteínas completas para su posterior análisis por MS debido a la reducida capacidad de ionización y el alto peso molecular de las mismas<sup>209</sup>. A su vez, la proteómica "bottom-up" identifica las proteínas a través de la detección específica de los péptidos generados tras la digestión enzimática. La principal desventaja de esta estrategia es que los péptidos identificados y cuantificados pueden no corresponder a proteínas únicas, dificultando la identificación inequívoca del proteoma<sup>166,204</sup>.

Con el fin de abordar el estudio de estas áreas, en las últimas décadas se han desarrollado un conjunto de técnicas proteómicas, que se detallan brevemente a continuación<sup>210-212</sup>, para lograr la identificación de proteínas únicas partiendo de una mezcla proteica compleja (Figura 8).





**Figura 8. Flujo de trabajo en las técnicas proteómicas basadas en espectrometría de masas.** [Adaptado de Köcher *et al.*<sup>213</sup>].

MALDI, ionización/desorción por láser asistida por matriz (del inglés, *matrix-associated laser desorption/ionization*); ESI, ionización por electrospray (del inglés, *electrospray ionization*).

### 3.3.1. Técnicas de separación de proteínas

- Fraccionamiento subcelular

Los métodos de fraccionamiento subcelular más comúnmente usados en combinación con los análisis basados en MS incluyen la centrifugación diferencial y el gradiente basado en el uso de detergentes. El principal problema al que se enfrentan estas técnicas es la redistribución de proteínas a otras fracciones tras la lisis celular (contaminación cruzada)<sup>214</sup>. A pesar de ello, los métodos de subfraccionamiento celular tienen una gran importancia ya que permiten una mayor identificación proteica al reducir la complejidad de las muestras<sup>166</sup>, punto clave en el aislamiento de proteínas de membrana (tanto de membranas plasmáticas como de orgánulos).

- Electroforesis

La electroforesis en gel de poliacrilamida (PAGE, del inglés *polyacrilamide gel electrophoresis*) es una técnica ampliamente usada en genética, bioquímica, biología molecular y proteómica. La finalidad de esta técnica es separar macromoléculas (proteínas o ácidos nucleicos) en función de su movilidad electroforética (masa, conformación y carga). En proteómica, existen dos variantes en función del estado estructural de las proteínas a analizar: electroforesis desnaturizante (SDS-PAGE) o nativa (sin desnaturizar). El uso de agentes desnaturizantes (p.ej. SDS, urea o 2-mercaptoetanol) permite romper puentes disulfuro obteniendo una versión linealizada

de la proteína. Asimismo, dependiendo de las dimensiones que se utilicen para separar las proteínas, se dispone de dos tipos de electroforesis: 1D, en función de la masa proteica y 2D, en función del punto isoeléctrico y masa proteicas.

### 3.3.2. Identificación, cuantificación y caracterización de proteínas

Las técnicas proteómicas en general abordan la determinación a gran escala de las proteínas y sus funciones celulares convirtiéndose en una disciplina indispensable para interpretar la información codificada en los genomas.

- Espectrometría de masas (MS)

Básicamente, los espectrómetros de masas constan de una fuente de ionización, un analizador de masas que mide la relación masa/carga ( $m/z$ ) de los analitos ionizados, y un detector que registra el número de iones en cada valor  $m/z$ <sup>215</sup>.

En relación al primer componente, la ionización por electrospray (ESI, del inglés *ElectroSpray Ionization*) y la ionización/desorción por láser asistida por matriz (MALDI, del inglés *Matrix-Assisted Laser Desorption/Ionization*) son las dos técnicas más comúnmente usadas para ionizar proteínas y péptidos<sup>212</sup>.

En cuanto a los analizadores de masas – caracterizados por su sensibilidad, resolución y precisión en la medición de la relación carga/masa<sup>216,217</sup> – se clasifican en 4 tipos principales: i) trampa iónica, ii) tiempo de vuelo (TOF, del inglés *time of flight*), iii) cuadrupolo (Q, del inglés *quadrupole*) y iv) analizador de resonancia ciclotrónica de iones mediante transformada de Fourier (FT-MS, del inglés *Fourier transform ion cyclotron mass spectrometry*). Estos analizadores pueden ser empleados de forma individual o en tándem (tándem MS/MS).

Los analizadores de masas de trampa iónica se caracterizan por ser robustos, relativamente baratas y presentar una alta sensibilidad. Sin embargo, su principal inconveniente reside en la baja precisión en la determinación de masas<sup>218</sup>. Por su parte, los analizadores de masas TOF tienen una alta sensibilidad, resolución y precisión en la medida de masas proporcionando una información más extensa que la

generada por los de trampas iónicas. Suelen emplearse en tándem y acoplados a MALDI (MALDI-TOF/TOF)<sup>219,220</sup>. Por su parte, los Q suelen emplearse en tándem (triple Q, QqQ) y/o acoplados a TOF (QqTOF). Finalmente, el FT-MS presenta una elevada sensibilidad, precisión en la determinación de masas, resolución y rango dinámico que le confiere un gran potencial. A pesar de ello, su elevado coste, junto con la complejidad en el manejo y la baja eficiencia en fragmentación peptídica han limitado su uso<sup>221</sup>.

Mediante la MS es posible realizar un análisis cualitativo (detección de la presencia de determinados compuestos concretos en una muestra) y/o cuantitativo (determinación de su concentración)<sup>222</sup>. Esta última aproximación puede realizarse de forma absoluta (determinación de la cantidad total de proteína en la muestra, habitualmente expresada en ng/mL o número de copias de una proteína/célula) o relativa (variaciones en la cantidad de proteína entre muestras a comparar)<sup>223,224</sup>.

Uno de los métodos (sin necesidad de marcaje proteico previo) más usados para la cuantificación de proteínas en una muestra determinada se basa en parámetros observables con gran precisión que determinan “índices de abundancia proteica” (PAI, del inglés *protein abundance index*). Entre esos parámetros se incluye el número de péptidos identificados por proteína, que se incrementa conforme aumenta la cantidad de proteína presente en la muestra. Considerando el valor logarítmico del término PAI, se calcula el valor de emPAI (PAI exponencialmente modificado), valor que frecuentemente se utiliza para normalizar el número de péptidos observados de una proteína frente al número de péptidos observables para dicha proteína<sup>225,226</sup>. De forma similar, se ha acuñado el término XIC (del inglés *extracted ion chromatogram*) que evalúa el área bajo la curva correspondiente a los distintos péptidos de una proteína (relacionado directamente con su intensidad) para determinar la cantidad de cada péptido de la misma presente en la muestra. De forma habitual, suelen seleccionarse los valores correspondientes a los tres péptidos más intensos de una proteína para su cuantificación<sup>227</sup>.

Respecto a los métodos cuantitativos que emplean marcaje isotópico estable, destacan tres estrategias denominadas SILAC (del inglés *stable isotope labeling by amino acids in cell culture*), iTRAQ (del inglés *isobaric tag for relative and absolute quantitation*) y iCAT (del inglés *isotope-coded affinity tag*)<sup>228</sup>. La tecnología SILAC permite la comparación de dos poblaciones celulares distintas mediante el marcaje de cada una de ellas con aminoácidos ligeros (=normales) o bien aminoácidos pesados; de tal forma, tras el análisis por MS, la diferencia de masa detectada entre aminoácidos ligeros y pesados permite determinar el origen celular de la proteína correspondiente<sup>229</sup>. A su vez, en la tecnología iTRAQ, se marcan las proteínas con distintas moléculas isotópicas permitiendo la comparación de hasta 8 muestras distintas de forma simultánea<sup>230</sup>. Finalmente, la tecnología iCAT se basa en el empleo de tres elementos: un marcador de afinidad (generalmente biotina, que se usa para aislar los péptidos marcados), un conector isotópicamente codificado que permite la diferenciación de los péptidos durante el análisis por MS, y un grupo reactivo con especificidad por grupos tio|<sup>231,232</sup>.

- Fosfoproteómica

El estudio y caracterización de las PTMs es imprescindible dado éstas dan lugar a proteínas reguladoras de numerosas funciones biológicas determinando, asimismo, su localización subcelular, interacciones y tasa de renovación<sup>166</sup>. Con este propósito, la MS es clave para la determinación del tipo y los aminoácidos concretos que presentan dichas PTMs<sup>233</sup>. Las estrategias de análisis suelen incluir el mapeo peptídico mediante diferentes enzimas para permitir la máxima secuenciación posible. A pesar de ello, su estudio es complejo debido a la baja estequiometría de las PTMs, su tamaño y la capacidad de ionización de sus péptidos, así como el comportamiento en los espectrómetros de masas<sup>203,208</sup>. Con el fin de superar estos inconvenientes, recientemente se han desarrollado nuevas estrategias para la purificación de las PTMs en muestras complejas, que permitan un análisis más exhaustivo<sup>207,234,235</sup>. Entre ellos, destacan los métodos de enriquecimiento mediante inmunocaptura (p.ej. esferas

recubiertas de Ac) de secuencias peptídicas específicas de PTM para su posterior análisis mediante LC-MS/MS<sup>236</sup>. Asimismo, métodos basados en el uso de metales de afinidad (como TiO<sub>2</sub> o ZrO<sub>2</sub>) se emplean también comúnmente para enriquecer las muestras en péptidos fosforilados frente a aquellos no fosforilados<sup>237-239</sup>. Finalmente, la tecnología IMAC (del inglés *immobilized metal affinity chromatography*) se basa en la afinidad que presentan los grupos fosfato por iones metálicos de transición (Cu<sup>2+</sup>, Ni<sup>2+</sup>, Zn<sup>2+</sup>, Fe<sup>3+</sup>, Co<sup>2+</sup>, Al<sup>3+</sup> y Ga<sup>2+</sup>)<sup>240,241</sup>. La mayor limitación de esta técnica es la unión inespecífica de residuos de aminoácidos cargados negativamente<sup>242</sup>.

En resumen, el estudio de las PTMs es fundamental para comprender e interpretar las rutas de señalización activadas en los sistemas celulares que determinan, por ejemplo, los estados patológicos vs. sanos<sup>243,244</sup>. En relación a la LLC, la caracterización de la ruta de señalización intracelular del BCR tiene un gran interés debido al papel crítico que ésta juega en el desarrollo de dicha enfermedad, dando lugar a patrones de fosforilación específicos en proteínas relacionadas con la proliferación, diferenciación y supervivencia celular<sup>28,156</sup>. Tal es su relevancia que las terapias más recientes y prometedoras en la LLC se basan en la inhibición de la activación inducida por el BCR (p.ej. mediante inhibidores de BTK y PI3K<sup>245</sup>, entre otros), destacando también el papel que en ella realizan las quinasas LYN y SYK.

- Microarrays de Acs

Los microarrays de proteínas son plataformas de alto rendimiento para la caracterización funcional de las proteínas, permitiendo analizar desde la actividad hasta las interacciones con otras biomoléculas<sup>246-248</sup>. Entre sus principales ventajas destacan: i) la detección de proteínas de baja abundancia en mezclas complejas con mayor sensibilidad y especificidad; ii) la detección múltiple de proteínas en una única plataforma; y iii) la mínima cantidad de muestra requerida<sup>249,250</sup>. Hoy día existen muchas clasificaciones de microarrays de proteínas según se atienda al contenido, formato o aplicación final de la misma<sup>251,252</sup>. De acuerdo a su aplicación, se distinguen los microarrays analíticos de los funcionales. Los microarrays analíticos se emplean para

comparar la expresión cuantitativa de proteínas en diferentes muestras. Por el contrario, los microarrays funcionales<sup>253</sup>: están especialmente enfocados al estudio de interacciones proteína-proteína, proteína-ADN, proteína-ARN, proteína-fosfolípido, proteína-molécula pequeña, análisis de actividad enzimática y la detección de Ac.

El diseño y construcción de los microarrays requiere de superficies compatibles cuya señal de fondo sea baja, presente un bajo nivel de uniones inespecíficas, buena estabilidad y accesibilidad para las sondas. En general, el material empleado suele ser vidrio, plástico, dióxido de silíceo y oro. La superficie de estos materiales requiere ser funcionalizada químicamente con grupos reactivos, como por ejemplo grupos aldehído, epoxi, isotiocianato, amino o mercapto<sup>254,255</sup>.

Habitualmente, se emplean dos métodos estándar para el depósito de las moléculas de interés en la superficie del microarray: impresión por contacto y sin contacto. En el primer grupo se encuentran los equipos que utilizan *pins* sólidos para distribuir ordenadamente las biomoléculas en la superficie del microarray mediante contacto físico. Dichos sistemas son flexibles permitiendo variaciones en el tipo de sustrato, hidrofobicidad y viscosidad, sin comprometer el diseño del microarray. En cuanto al segundo grupo de métodos, la deposición de las sondas se realiza mediante eyección de gotas desde un atomizador hacia la superficie a través de señales piezo-eléctricas<sup>256</sup>. En comparación con el primer tipo de métodos de impresión, los sistemas de impresión sin contacto son menos flexibles en relación a la cantidad de muestra que puede ser depositada en la superficie, aunque su velocidad de impresión es claramente superior<sup>255</sup>.

Finalmente, y en relación a los formatos<sup>251,252</sup>, los microarrays pueden diseñarse en formato de superficie esférica<sup>257</sup> o de superficie plana<sup>258,259</sup>, presentando distintas particularidades que deben tenerse en consideración de cara a la aplicación final para la que se producen.

### 3.3.3. Integración de tecnologías –ómicas

La integración de la proteómica con otras aproximaciones “ómicas” (p.ej. genómica, transcriptómica, metabolómica, biología estructural) requiere notables avances en bioinformática y proteómica computacional<sup>204</sup>. Dicha integración supone un gran reto debido principalmente a la naturaleza dinámica y multidimensional de los proteomas junto con el gran tamaño de los conjuntos de datos generados. A pesar de estos inconvenientes, los análisis multidimensionales y multiparamétricos son necesarios para comprender la biología celular y tisular<sup>166</sup>.

Actualmente, existen numerosas herramientas informáticas para el análisis de datos de MS (p.ej. MaxQuant<sup>260,261</sup>, Skyline<sup>262</sup>, PeptideShaker<sup>263</sup> y COMPASS<sup>264</sup>) que permiten el análisis cualitativo y cuantitativo de los proteomas. Además, se han desarrollado también programas bioinformáticos para el estudio de las rutas de señalización intracelulares (p.ej. DAVID<sup>265</sup>) e interacciones célula-célula (p.ej. Cytoscape<sup>266</sup>).

Adicionalmente, son necesarios paquetes bioestadísticos robustos que permitan la evaluación de los resultados obtenidos para determinar el verdadero significado de los cambios en la abundancia proteica o de otros parámetros de estudio, así como plataformas gráficas que permitan una visualización integral y exhaustiva de la gran cantidad de información generada (p.ej. GSEA<sup>267</sup>).

Finalmente, es necesario que los datos generados sean accesibles a toda la comunidad científica. Con tal fin, se han creado repositorios públicos para el depósito de datos proteómicos con altos estándares de calidad; de ellos merece destacar los repositorios PRIDE (del inglés *Proteomics Identification Database*)<sup>268</sup>, GPMDB (del inglés, *Global Proteome Machine and Database*)<sup>269</sup>, HPA (del inglés, *Human Proteome Atlas*)<sup>270</sup>, Tranche<sup>271</sup> y *Human Proteinpedia*<sup>272</sup>.

### 3.4. Ventajas y retos de la proteómica

A pesar del desarrollo de técnicas innovadoras, como las mencionadas anteriormente, siguen existiendo una serie de limitaciones que dificultan una

caracterización completa exhaustiva y sistemática de los proteomas. Entre estas limitaciones pueden citarse: el análisis de mezclas biológicas complejas, la habilidad para cuantificar especies proteicas aisladas, la falta de sensibilidad suficiente para detectar proteínas poco abundantes en una muestra y la cuantificación masiva en rangos dinámicos amplios<sup>210,273</sup>.

Asimismo, la correcta identificación y cuantificación de proteínas de membrana (como los receptores de membrana de la célula B, incluidas las Ig de superficie) sigue constituyendo un reto importante, debido principalmente a la tendencia de estas proteínas a precipitar y agregarse debido a su hidrofobicidad, lo cual dificulta su análisis por MS. Además, los residuos aminoacídicos que son diana del procesamiento triptico - es decir, lisina (Lys) y arginina (Arg) - son bastante escasos en las regiones transmembrana, lo que dificulta aún más su estudio proteómico llegando a hacerlo imposible. Con el fin de mejorar la identificación de estas proteínas de membrana, se están implementando protocolos de subfraccionamiento celular y de marcaje proteico, junto con distintas combinaciones de enzimas para la digestión proteica, que contribuyen a conseguir un enriquecimiento previo en dichas proteínas<sup>274-276</sup>. Entre los métodos de fraccionamiento cromatográfico más empleados y/o efectivos, merece destacar el intercambio catiónico y el fraccionamiento en fase reversa a pH elevado<sup>277</sup>.

Adicionalmente, otro de los principales retos a los que se enfrenta actualmente la proteómica es el análisis de proteínas poco abundantes. Muchas clases de proteínas, como los factores de transcripción, las quinasas y las proteínas reguladoras, están presentes en los sistemas celulares en un número bajo de copias, dificultando su análisis directo en ausencia de procedimientos de enriquecimiento<sup>165</sup>.

Por otra parte, uno de los retos actuales más importantes que pretende abordar la proteómica consiste en analizar el proteoma completo de células individuales<sup>278</sup>. Hasta el momento, el fracaso en la consecución de este propósito se ha debido a la falta de una sensibilidad suficiente de las aproximaciones empleadas para el análisis



---

de cantidades mínimas de muestra, debido a que esto limita una producción eficiente de iones, genera pérdidas iónicas dentro del instrumento de medida y se ve afectada en gran medida por los contaminantes y solventes en los que está incluida la muestra. De esta forma, los avances tecnológicos han de encaminarse al desarrollo y fabricación de equipos más sensibles con niveles bajos de "ruido químico". En este sentido, los sistemas microfluídicos parecen tener un potencial importante que llegue a conferirles un papel fundamental en el alcance de dicha sensibilidad<sup>279</sup>. Sin duda, conseguir este objetivo conducirá a notables avances en el área permitiendo, entre otras aplicaciones, el desarrollo de fármacos dirigidos con mayor precisión a células concretas dentro de un tumor heterogéneo<sup>280,281</sup>.



# **Hipótesis de trabajo y objetivos**



La LLC-B constituye la neoplasia hematológica más común en los países occidentales, presentando una importante heterogeneidad en cuanto a su comportamiento y evolución clínica. Así, mientras que en muchos pacientes la enfermedad muestra un curso indolente durante años, otros requieren tratamiento desde las etapas más tempranas que siguen al diagnóstico de la enfermedad. En la actualidad, disponemos de un amplio espectro de factores pronósticos que contribuyen a predecir la evolución de la enfermedad de los que merece destacar las alteraciones citogenéticas [p.ej. del(13q) y cariotipo normal vs. del(17p) y del(11q)] y el estado mutacional de los genes que codifican para la región variable de los genes de IGHV. En los últimos años, el desarrollo de técnicas de secuenciación masiva ha permitido ampliar la caracterización de la célula tumoral de la LLC a nivel genómico y transcriptómico. Sin embargo, en el momento de iniciar el trabajo de esta tesis doctoral, no existían estudios que abordaran la caracterización masiva de la LLC-B desde una perspectiva proteómica que contribuyese a la identificación y caracterización de las rutas de señalización intracelular afectadas, los niveles de fosforilación proteico del proteoma de la célula leucémica (fosfoproteoma) y la secuenciación (peptídica) del BCR, especialmente de las Igs. Entendemos que este abordaje es esencial para poder comprender el proceso de transformación maligna que sufren las células de la LLC-B dado que, como la mayoría de los tumores malignos, ésta podría estar dirigida por la regulación/desregulación de redes proteicas que tienen su punto de partida en un receptor extracelular capaz de desencadenar rutas de señalización en las que tienen un papel clave los procesos de fosforilación/desfosforilación de distintos sustratos proteicos.

En la actualidad, existen diferentes modelos que tratan de explicar el origen de la célula tumoral de la LLC-B basados en un origen celular único o bien en la existencia de un origen en múltiples etapas que discurren a lo largo de la ontogenia de la célula B normal. En este sentido, resulta clave a la hora de descifrar los mecanismos de transformación maligna de la célula B de la LLC, establecer un modelo de referencia

de la contrapartida celular normal, en lo que a nivel proteómico se refiere. Asimismo, la integración de distintas estrategias genómicas y transcriptómicas con la información proteómica y fosfoproteómica permitiría avanzar, de forma significativa, en el conocimiento de la enfermedad y contribuir a determinar con mayor precisión la probabilidad de progresión celular y clínica de la enfermedad. Además, la optimización y adecuación de las metodologías proteómicas al estudio de la célula leucémica de la LLC-B serían de gran relevancia para poder caracterizar mejor esta enfermedad en la rutina del laboratorio clínico.

De forma indirecta, todo lo anterior podría contribuir también, no sólo a explicar mejor la heterogeneidad de la LLC-B desde un punto de vista clínico-biológico, sino que además podría proporcionar información clave para un mejor seguimiento, tratamiento y control de la enfermedad.

De acuerdo con estos antecedentes, en el presente trabajo doctoral nos propusimos como **objetivo general** caracterizar de forma masiva desde un punto de vista proteómico la célula B de la LLC-B y su contrapartida normal y diseñar estrategias para la integración de datos proteómicos con otro tipo de información molecular (p.ej. datos transcriptómicos). Para ello, planteamos cinco **objetivos específicos**:

1. Caracterizar, de forma sistemática y cuantitativa, el conjunto de proteínas (proteomas) y sus modificaciones post-traduccionales (fosfoproteomas) de las células B tumorales de pacientes con LLC-B y LMB.
2. Evaluar las posibles alteraciones proteómicas que involucran a vías de señalización intracelular de la célula B de la LLC-B con el fin de conocer el perfil proteico global de esta célula y su relación con las alteraciones citogenéticas y moleculares presentes en la misma.
3. Establecer una metodología óptima para la secuenciación peptídica del BCR en la célula B y comparar los resultados obtenidos con los derivados de técnicas estándar de secuenciación genómica a través de un algoritmo de integración de ambos tipos de información (proteómica y genómica).

4. Diseñar y desarrollar una estrategia para la integración de datos proteómicos y transcriptómicos, en un modelo de línea celular tumoral B, que contribuya a evaluar la complementariedad de ambas aproximaciones metodológicas aplicadas al conocimiento de las células B tumorales.
5. Caracterizar el proteoma de células B normales en distintas etapas de diferenciación (célula B naive, CB, CC, célula B de memoria y CP), con el fin de establecer una referencia de normalidad para el estudio de la célula B tumoral de LLC-B, e integrar dichos resultados proteómicos con los obtenidos mediante microarrays de expresión empleando la estrategia desarrollada anteriormente.





# **Materiales, métodos y resultados**



En esta sección de la memoria se describen las muestras, pacientes, materiales y métodos empleados, así como los resultados obtenidos en el presente trabajo doctoral, mediante la inclusión de los artículos originales publicados y en revisión que están directamente relacionados con la hipótesis de trabajo y los objetivos planteados. Cada uno de los artículos está precedido de un resumen redactado en castellano para facilitar una revisión rápida de la información contenida en los mismos.



# Caracterización proteómica y fosfoproteómica integrada de las células B tumorales de la leucemia linfocítica crónica y de las rutas de señalización implicadas



Paula Díez<sup>1,2</sup>, Conrad Droste<sup>3</sup>, Julia Almeida<sup>1</sup>,  
Marcos González<sup>4</sup>, Alberto Orfao<sup>1</sup>, Javier De Las Rivas<sup>3</sup>,  
Manuel Fuentes<sup>1,2</sup>

<sup>1</sup> Servicio General de Citometría-NUCLEUS, Departamento de Medicina;  
Centro de Investigación del Cáncer (IBMCC-CSIC/USAL) e IBSAL,  
Universidad de Salamanca.

<sup>2</sup> Unidad de Proteómica; Centro de Investigación del Cáncer (IBMCC-  
CSIC/USAL) e IBSAL, Universidad de Salamanca.

<sup>3</sup> Grupo de Bioinformática y Genómica Funcional; Centro de Investigación  
del Cáncer (IBMCC-CSIC/USAL) e IBSAL, Universidad de Salamanca.

<sup>4</sup> Servicio de Hematología (IBMCC) e IBSAL, Hospital Universitario de  
Salamanca.

*en revisión*

*La información suplementaria correspondiente a este manuscrito  
está disponible, en formato electrónico, en el CD-ROM anexo al final  
de la presente tesis doctoral.*



**Introducción.** La LLC representa el tumor hematológico más frecuente en el mundo occidental caracterizándose a nivel genómico por presentar una gran heterogeneidad. A pesar de que el genoma de las células de la LLC ha sido analizado en detalle en la última década, el perfil proteico global de la célula tumoral y las rutas de señalización implicadas aún no han sido completamente estudiadas y descifradas. Esto supone un punto crítico para conocer los procesos implicados en la transformación maligna de la célula leucémica, sobre todo si tenemos en cuenta que en la mayoría de los tumores estos procesos están controlados por redes proteicas alteradas, incluyendo aquellas que implican quinasas. Específicamente, en la LLC-B estas alteraciones afectan, de forma recurrente, a las vías de señalización Ag-específicas asociadas al BCR.

**Objetivo.** En este trabajo nos propusimos caracterizar de forma sistemática, tanto cualitativa como cuantitativamente, y por primera vez, el fosfoproteoma completo (proteínas fosforiladas en residuos Tyr, Ser y/o Thr) de la célula B tumoral de pacientes con LLC y LMB, mediante el empleo de técnicas de alta resolución basadas en MS. Asimismo, se investigó el proteoma global en estas células a nivel cuantitativo, con el fin de obtener una visión global de las rutas de señalización presentes en estas células, independientemente de que se encuentren activas (p.ej. fosforiladas) o no.

**Muestras, materiales y métodos.** Se analizaron linfocitos B purificados a partir de la sangre periférica de 4 pacientes con LLC y un paciente con LMB. La purificación de las células B tumorales se realizó mediante citometría de flujo. Para la extracción del contenido proteico global de las células tumorales se empleó una solución de lisis de urea. Dichas muestras proteicas fueron separadas mediante electroforesis unidimensional seguida de un proceso de digestión trípica. Los péptidos resultantes fueron analizados mediante dos procedimientos: para el análisis cuantitativo del proteoma completo se analizaron los péptidos obtenidos directamente en el equipo de MS, mientras que para el análisis fosfoproteómico se aplicó primero el método

PTMScan®, consistente en incubar los péptidos con una suspensión de microesferas cuya superficie estaba recubierta de un Ac dirigido específicamente frente a motivos fosfotirosinas (anti-p-Tyr) presentes en dichos péptidos, seguido del análisis de MS que se realizó en un equipo Orbitrap Q-Exactive. Los datos resultantes fueron evaluados con SEQUEST y la plataforma proteómica Core de la Universidad de Harvard (UniProt, versión abril 2015).

**Resultados.** El estudio del proteoma cuantitativo y del fosfoproteoma de las células B de pacientes con LLC y LMB tipo LLC permitieron identificar 2.970 proteínas y 327 fosfoproteínas (incluyendo la detección por primera vez de 329 fosfopéptidos no identificados previamente). Respecto al proteoma global de estas células, se observó un solapamiento significativo (73%) entre las 5 muestras analizadas. Por el contrario, el fosfoproteoma de esas muestras varió de forma muy significativa entre ellas, lo cual resalta la importancia de determinar los patrones de fosforilación de células B de LLC y LMB. A pesar de la variabilidad observada en el fosfoproteoma de las células tumorales estudiadas, éstas mostraron en común la expresión de un gran número de fosfoproteínas, de las que merece destacar aquellas que están implicadas en las vías de señalización del BCR, en interacciones célula-célula (sobre todo en las células del sistema inmune) y en la ruta de señalización NF- $\kappa$ B/STAT3. En relación con la vía de señalización a través del BCR, la fosforilación afectaba a un número importante de proteínas (LYN, SYK, PI3K, BTK, ZAP70, PLCG3, ERK, NFAT, CD19, PRKCB, NFKB, JNK y VAV), lo cual apoya la hipótesis de que esta ruta puede jugar un papel importante en el mantenimiento/supervivencia de las células tumorales de la LLC y la LMB. Cabe señalar además que estos perfiles fosfoproteómicos eran aparentemente independientes de las alteraciones citogenéticas y/o del estado mutacional de las cadenas IGHV de las células tumorales. De forma adicional, a pesar de la gran similitud observada entre las células tumorales de la LLC y de la LMB analizadas, se detectaron distintos niveles de fosforilación de ARIH2 y PTPN11 entre ellas, detectadas en LLC y ausentes en LMB, estando ambas proteínas implicadas en la inducción de la apoptosis



celular, lo que sugiere un papel importante de ambas moléculas en el desarrollo de la enfermedad. Asimismo, la expresión de algunas fosfoproteínas estaba limitada a solo una de las muestras estudiadas (p.ej. DYNL1, RFWD3, y PIK3R4).

**Conclusiones.** A través del conocimiento global del proteoma y del fosfoproteoma de las células leucémicas de pacientes con LLC y LMB, el presente trabajo ha proporcionado nuevas perspectivas en relación con las rutas de señalización inmune implicadas en el desarrollo y progresión de dichos tumores y potencialmente también en su transformación maligna.

\* Los resultados proteómicos fueron depositados en ProteomeXchange (PXD005997).



**Aproximaciones proteómicas funcionales para la  
identificación global de rutas de señalización  
celulares alteradas en la leucemia linfocítica  
crónica de célula B**

2

Paula Díez<sup>1,2</sup>, Seila Lorenzo<sup>1</sup>, Rosa M. Dégano<sup>2</sup>,  
Nieves Ibarrola<sup>2</sup>, María González-González<sup>1,2</sup>,  
Wendy Nieto<sup>1</sup>, Julia Almeida<sup>1</sup>, Marcos González<sup>3</sup>,  
Alberto Orfao<sup>1</sup>, Manuel Fuentes<sup>1,2</sup>

<sup>1</sup> Servicio General de Citometría-NUCLEUS, Departamento de Medicina;  
Centro de Investigación del Cáncer (IBMCC-CSIC/USAL) e IBSAL,  
Universidad de Salamanca.

<sup>2</sup> Unidad de Proteómica; Centro de Investigación del Cáncer (IBMCC-  
CSIC/USAL) e IBSAL, Universidad de Salamanca.

<sup>3</sup> Servicio de Hematología (IBMCC) e IBSAL,  
Hospital Universitario de Salamanca.

*Proteomics*. 2016 Abr; 16(8):1193-1203.

doi: 10.1002/pmic.201500372



**Introducción.** La LLC representa la neoplasia hematológica humana más común en los países occidentales, caracterizándose por una elevada heterogeneidad clínica y pronóstica. En la actualidad disponemos de un amplio abanico de factores pronósticos que contribuyen a predecir el curso clínico de la LLC y que, entre otros, incluyen parámetros citogenéticos (trisomía del cromosoma 12, deleciones de 13q14, 17p13 y 11q), y moleculares (mutaciones puntuales en *NOTCH1*), junto al estado mutacional de los genes de IGHV. Aunque las alteraciones genómicas de la LLC han sido analizadas ampliamente, los estudios dirigidos a la caracterización proteómica de la célula leucémica siguen siendo muy limitados.

**Objetivo.** En el presente trabajo hemos analizado la relación existente entre las alteraciones citogenéticas de la LLC y los niveles de expresión de distintas proteínas implicadas en diferentes rutas de señalización intracelular de células B.

**Muestras, materiales y métodos.** Se analizaron linfocitos B patológicos de 14 pacientes diagnosticados de LLC, los cuales se agruparon en función de las alteraciones citogenéticas que presentaban: 4 pacientes sin alteraciones cromosómicas; 3 con trisomía del cromosoma 12; 3 presentaban del(13q14.3); un caso tenía trisomía del cromosoma 12 + del(11q22.3); un caso de trisomía del cromosoma 12 y del(17p13.1); otro paciente con del(11q22.3) asociada a del(13q14) y del(13q14.3); y el caso restante mostraba del(13q14.3) asociada a del(17p13.1). De los 8 pacientes en los que se analizó el estado mutacional de *NOTCH1*: en 4 se detectó mutación de este gen. Como control, se analizaron linfocitos B purificados de sangre periférica de 63 individuos sanos. La purificación de los linfocitos B, tanto de los pacientes como de los individuos sanos, se realizó mediante citometría de flujo. Para la purificación de las proteínas de citoplasma se empleó un protocolo de subfraccionamiento celular seguido de depleción de IgG al ser proteínas muy abundantes que podrían enmascarar la señal de otras proteínas minoritarias. Dichas proteínas fueron marcadas con cianinas (Cy3 y Cy5) para su posterior incubación en microarrays de Acs (224 Acs

monoclonales humanos dirigidos frente a proteínas implicadas en rutas de señalización intracelular, como apoptosis, ciclo celular, citoesqueleto, transducción de señales y proteínas nucleares, entre otras. En cada microarray se incubaron simultáneamente dos muestras (cada una marcada con una Cy distinta), de tal forma que una de ellas se consideraba como referencia de la otra.

Como validación se realizó un análisis mediante LC-MS/MS en un subgrupo de las muestras estudiadas. Brevemente, las muestras proteicas marcadas fueron digeridas en solución con tripsina y analizadas mediante un equipo LTQ-Orbitrap-Velos. El análisis de los resultados se efectuó mediante PeptideShaker y la base de datos UniProt (versión de noviembre de 2015) con una tasa de falsos descubrimientos (FDR, del inglés *false discovery rate*) <1% a nivel de coincidencia del espectro peptídico (PSM, del inglés *peptide spectrum match*), péptidos y proteínas. Para determinar la abundancia proteica relativa, se determinaron los valores emPAI.

**Resultados.** El análisis de las células B de pacientes con LLC vs. individuos sanos empleando microarrays de expresión diferencial, basados en dos emisiones fluorescentes distintas, permitió la identificación y cuantificación relativa de un gran número de proteínas expresadas de forma diferencial en las células leucémicas respecto a las células B normales. No obstante, las células B de pacientes con LLC que no presentaban alteraciones citogenéticas o mostraban trisomía del cromosoma 12 como única alteración cromosómica, mostraron perfiles de expresión proteica mucho más parecidos a los de las células B de individuos sanos frente a las células B leucémicas de pacientes con LLC y otros perfiles de alteración genética.

En más detalle, cabe señalar que los resultados obtenidos mostraron una discriminación relativa de proteínas de la familia PKC (del inglés *protein kinase C*) en la mayoría de muestras de LLC analizadas (75%), lo que se relacionó con formas más agresivas de la enfermedad. Asimismo, la proteína MCL-1 se encontraba sobreexpresada entre los pacientes que presentaban *NOTCH1* mutado y/o del(17p13.1) (TP53), correlacionándose así con un fenotipo más maligno, lo cual podría

ser debido al papel inhibidor de la apoptosis que presenta dicha proteína. En términos globales, las muestras de pacientes que presentaban mutación en *NOTCH1* mostraron un perfil proteico caracterizado por un incremento relativo de proteínas implicadas en el ciclo celular (niveles elevados de CDK4 y CDK6 asociados a niveles más bajos de p21) y una disminución de proteínas implicadas en la respuesta celular a estrés. Por otra parte, las muestras con del(13q14) mostraron un incremento en la expresión de JUN y en varias ciclinas (ciclinas A, B1 y D1). El aumento de estas últimas podría contrarrestar el efecto negativo de RAF1 (proteína activadora de la señalización MEK/ERK), cuyos niveles se encontrarían bajos en la LLC. Finalmente, los pacientes con del(17p13.1) presentaron una disminución de los niveles de expresión de proteínas de la vía de señalización RAF/MEK y del inhibidor p21, junto con un aumento relativo de JNK y los complejos ciclina A/CDK2-1 y ciclina D/CDK4.

Los estudios de LC-MS/MS realizados permitieron validar los niveles de expresión y abundancia relativa detectados por los microarrays de proteínas.

**Conclusiones.** En resumen, nuestros resultados muestran la utilidad del empleo de estrategias proteómicas (y su combinación) para evaluar de forma global alteraciones en los patrones de expresión proteica de las células B leucémicas de pacientes con LLC frente a células B normales. En términos globales, la LLC-B presenta alteraciones en la expresión de proteínas asociadas a distintas vías de señalización intracelular dependiendo, al menos en parte, de las alteraciones citogenéticas y moleculares presentes en las células B patológicas.





## RESEARCH ARTICLE

# Multipronged functional proteomics approaches for global identification of altered cell signalling pathways in B-cell chronic lymphocytic leukaemia

Paula Díez<sup>1,2</sup>, Seila Lorenzo<sup>1</sup>, Rosa M<sup>a</sup> Dégano<sup>2</sup>, Nieves Ibarrola<sup>2</sup>, María González-González<sup>1,2</sup>, Wendy Nieto<sup>1</sup>, Julia Almeida<sup>1</sup>, Marcos González<sup>3</sup>, Alberto Orfao<sup>1</sup> and Manuel Fuentes<sup>1,2</sup>

<sup>1</sup> Department of Medicine and General Cytometry Service-Nucleus, Cancer Research Centre (IBMCC/CSIC/USAL/IBSAL), Salamanca, Spain

<sup>2</sup> Proteomics Unit, Cancer Research Centre (IBMCC/CSIC/USAL/IBSAL), Salamanca, Spain

<sup>3</sup> Service of Haematology, IBMCC, IBSAL, University Hospital of Salamanca, Salamanca, Spain

Chronic lymphocytic leukaemia (CLL) is a malignant B cell disorder characterized by its high heterogeneity. Although genomic alterations have been broadly reported, protein studies are still in their early stages. Herein, a 224-antibody microarray has been employed to study the intracellular signalling pathways in a cohort of 14 newly diagnosed B-CLL patients as a preliminary study for further investigations. Several protein profiles were differentially identified across the cytogenetic and molecular alterations presented in the samples (deletion 13q14 and 17p13.1, trisomy 12, and NOTCH1 mutations) by a combination of affinity and MS/MS proteomics approaches. Among others altered cell signalling pathways, PKC family members were identified as down-regulated in nearly 75% of the samples tested with the antibody arrays. This might explain the rapid progression of the disease when showing p53, Rb1, or NOTCH1 mutations due to PKC-proteins family plays a critical role favouring the slowly progressive indolent behaviour of CLL. Additionally, the antibody microarray results were validated by a LC-MS/MS quantification strategy and compared to a transcriptomic CLL database. In summary, this research displays the usefulness of proteomic strategies to globally evaluate the protein alterations in CLL cells and select the possible biomarkers to be further studied with larger sample sizes.

Received: September 14, 2015

Revised: December 4, 2015

Accepted: February 9, 2016

## Keywords:

Antibody microarray / Biomedicine / Cell signalling pathway / Chronic lymphocytic leukaemia / Cytoplasmic proteins / Proteome profiling



Additional supporting information may be found in the online version of this article at the publisher's web-site

**Correspondence:** Dr. Manuel Fuentes, Department of Medicine and General Cytometry Service-Nucleus, Cancer Research Centre (IBMCC/CSIC/USAL/IBSAL). Avda. Universidad de Coimbra, S/N, 37007 Salamanca, Spain  
**E-mail:** mfuentes@usal.es  
**Fax:** +34 923294743

**Abbreviations:** CLL, chronic lymphocytic leukaemia; **del(11q)**, deletion 11q; **del(13q14)**, deletion 13q14; **del(17p13.1)**, deletion 17p13.1; **emPAI**, exponentially modified protein abundance index; **FISH**, fluorescence in situ hybridization; **HPP**, Human Proteome Project; **IgHV**, immunoglobulin heavy chain variable

## 1 Introduction

Chronic lymphocytic leukaemia (CLL) is the most common human blood cancer in Western countries [1–3]. Typically, clinical diagnosis of CLL includes the accumulation of large numbers of mature monoclonal B cells in bone marrow, blood, and lymphoid tissues [4]. The disease is characterized by a high heterogeneity; some patients have an indolent

region; **KEGG**, Kyoto Encyclopaedia of Genes and Genomes; **RT**, room temperature

**Colour Online:** See the article online to view Figs. 1–3 and 5 in colour.

## Significance of the study

The characterization of chronic lymphocytic leukaemia (CLL) at genomic level has been performed throughout the last decades providing a broad knowledge of the disease. Nevertheless, studies concerning proteins have been limited in this cancer. Thanks to the development of high-throughput, specific and sensitive proteomic approaches, we have gained a deeper knowledge in this area allowing a better understand-

ing of this heterogeneous disease. In this study, affinity and MS/MS proteomic strategies have been employed to establish preliminary proteomic profiles of different CLL stages depending on cytogenetic and molecular alterations. This biological data might increase the understanding of the CLL by characterizing possible biomarkers that could be of high interest for the diagnosis and prognosis of the disease.

course while others require early treatment for disease progression [5]. The clinical course may be predicted by several cytogenetic factors such as trisomy 12 and deletions of 13q14 [del(13q14)] for Rb1, 17p13.1[del(17p13.1)] for p53, and 11q [del(11q)], as well as NOTCH1 mutations [2, 6]. Up to the moment, among cytogenetic analysis, the most robust approaches for predicting the disease outcome are based on sequencing of immunoglobulin heavy chain variable region (IgHV), in order to stratify patients since it is an important prognostic factor [7]. Therefore, it would be of great interest the application of new strategies for the analysis of CLL patient progression as well as for a better understanding of the pathology in order to provide an appropriate treatment and/or predict the clinical outcome.

Genomic studies have revealed the implication of alterations at gene expression level in CLL [8]. Chromosomal alterations, mainly at early stages of B cell development, promote B cell neoplasms. Specifically, CLL derives from germinal or post-germinal centre B cells. Alterations in RAS/ERK, BCR/ABL1 or PI3K pathways, among others, could lead to the progression of malignant B cells, resulting in malignant neoplasms [9].

In such a way, the recent rise of the combination of several proteomics approaches has allowed the discovery of proteome profiles in CLL patients by different methodological strategies. For instance, Alsagaby and collaborators have applied iTRAQ technology for analysing the CLL proteome [10]. In turn, Vogler and colleagues have employed SILAC approaches to study the resistance mechanism of the ABT-263 drug in CLL cells within lymph nodes [3].

As a consequence, the Human Proteome Project (HPP) is promoting the integration of genomics and proteomics information to better understand diseases mechanisms and to provide light inside pathologies by investigating which specific protein pathways contribute to disease pathogenesis [11]. The combination of both disciplines might lead to the identification of disease-altered proteins in a rapid manner for the discovery of novel diagnostic biomarkers as well as opening hallmarks for new treatments. In addition, HUPO is promoting the combination of both approaches to guide the genome-wide proteomics analyses.

In the present research, we aimed to describe a correlation between cytogenetic alterations identified in diagnosed

CLL patients with global differential protein expression levels (224 proteins involved in intracellular signalling pathways). To achieve this goal, a well-established and one-step commercial antibody array platform was employed for the screening of pathological B cells from 14 CLL samples, being equally represented all the common clinically described cytogenetic aberrations (del(13q14), del(17p13.1), trisomy 12, and/or NOTCH1 mutations). Moreover, the differential protein expression profiles have been compared with well-established transcriptomic databases of aberrant B cells obtained from CLL samples. Finally, a LC-MS/MS quantification strategy was employed for validation of the findings.

## 2 Materials and methods

A global view of the experimental workflow is shown in Fig. 1.

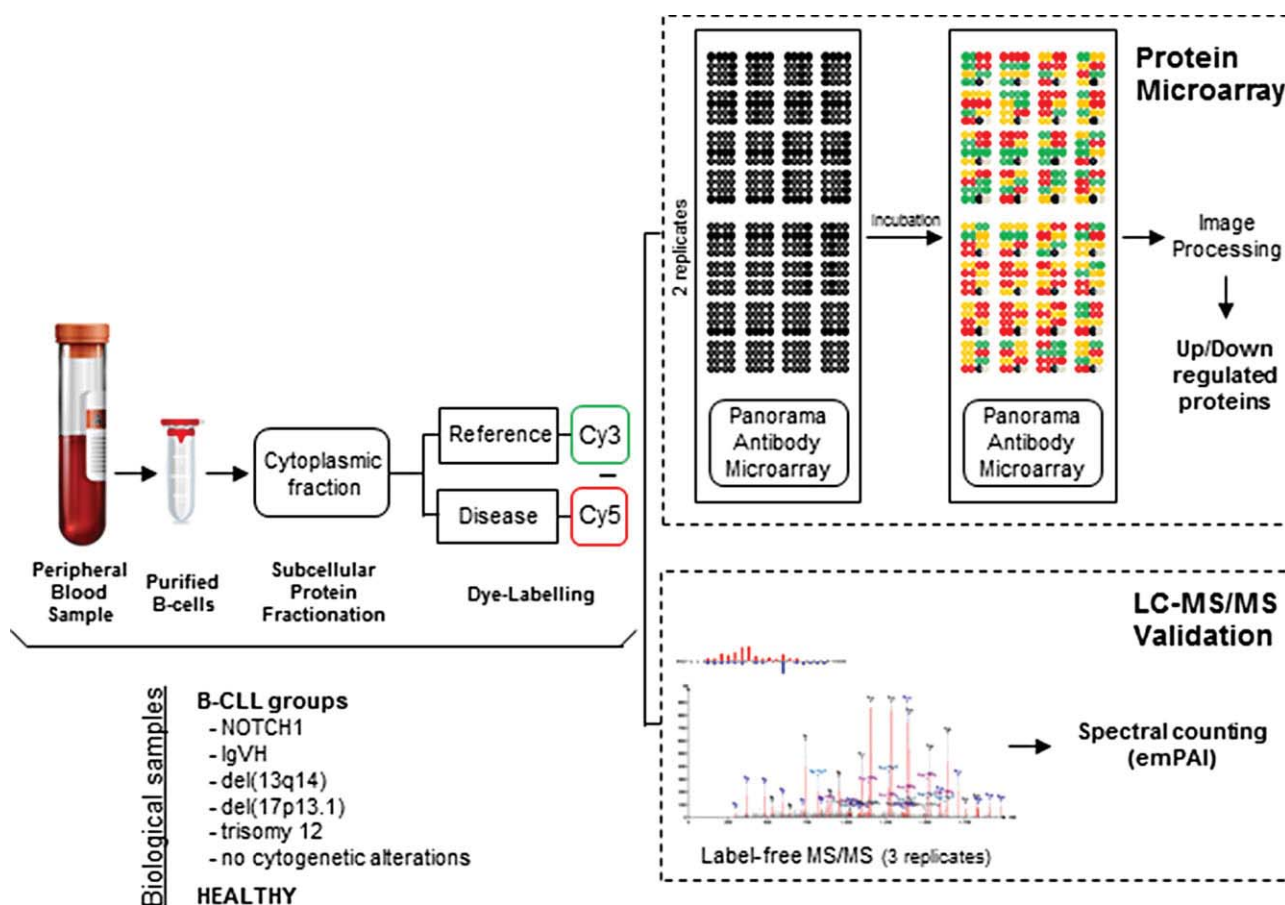
### 2.1 Sample collection

Peripheral blood B-lymphocytes (CD19<sup>+</sup> CD5<sup>+</sup>) from 14 newly diagnosed CLL patients (with representative cytogenetic alterations, such as trisomy 12, del(13q14), del(17p13.1), and NOTCH1 mutations [3, 12]), and 63 healthy subjects were systematically purified by FACSARIA (BD, San Jose/CA, USA) at General Cytometry Service of University of Salamanca [13].

Peripheral blood samples from patients and healthy donors were obtained with informed consent and according to the guidelines of the local ethics committees of the University Hospital of Salamanca and the National DNA Bank-ISCIII, in accordance with the Helsinki Declaration of 1975, as revised in 2008.

Immunophenotypic studies were performed by high-sensitive multi-parameter flow cytometry to screen for the presence and full characterization of clonal B cell populations, according to procedures previously described [14]. Cytogenetic and molecular analysis using fluorescence in situ hybridization (FISH) were performed on FACS-purified and fixed aberrant cells [15].

In Table 1, it is described the clinical information related to the 14 CLL samples included in the study.



**Figure 1.** Workflow overview. A brief description of steps followed since the obtaining of the sample to the image processing and data analysis and LC-MS/MS validation. After the purification of the B-cells from peripheral blood of B-CLL patients, cytoplasmic proteins are isolated by using a subcellular protein fractionation protocol. Then, proteins are Cy-labelled and incubated in the Panorama antibody microarray (reference sample vs cytogenetic-altered sample). After image processing, data analysis is performed by normalizing with housekeeping proteins to identify the up- and down-regulated proteins for each group of study. A LC-MS/MS quantification strategy has been used for the validation of the results.

## 2.2 Relative abundance assays by antibody arrays

### 2.2.1 Subcellular protein fractionation and dye-labelling

Cytoplasmic proteins were isolated from frozen FACS-purified B cells in 3 volumes of lysis buffer (5 mM HEPES, 10 mM MgCl<sub>2</sub>, 140 mM NaCl, 0.1% (v/v) Tween 20 at pH 8.0, and 1% (v/v) protease inhibitor mix (all purchased from Sigma, St. Louis/MO, USA) by a sub-fractionation protocol. Briefly, lysed-cells were centrifuged for 15 min at 15 000 × g and 4°C. Henceforth, the supernatant was collected and 10% (w/v) octyl-β-D-glucopyranoside (Sigma, St. Louis/MO, USA) was added followed by sonication. Samples were preserved on ice for 30 min and then centrifuged obtaining the cytoplasmic proteins.

Immediately after IgG depletion (by using protein-G-agarose beads from GE Healthcare, Buckinghamshire, UK), purified proteins were verified by SDS-PAGE (Supporting

Information Fig. S1) and quantified by Lowry-DC-Protein Assay (Bio-Rad Laboratories, CA, USA). Then, cytoplasmic proteins were labelled with Cy3 or Cy5 (GE Healthcare, Buckinghamshire, UK) at >2 dye/protein ratio as indicated by the manufacturer.

### 2.2.2 Panorama antibody microarray assays

Relative abundance assays were performed by using the Panorama Cell Signalling Kit (Sigma, St. Louis/MO, USA) following manufacturer's instructions. The antibody microarray consists of 224 human monoclonal antibodies targeting proteins involved in intracellular signalling pathways, spotted in duplicate onto a nitrocellulose-coated slide. Equal amounts of fluorescently labelled proteins of B cells from healthy controls and CLL patients (each labelled with a different cyanine) were incubated on the microarrays at room temperature (RT) for 1 h on an orbital shaker.

**Table 1.** Patient characteristics. The clinical characteristics of the 14 newly diagnosed B-CLL samples employed in this study are indicated including the age, gender, diagnosis and data related to cytogenetic alterations

Case ID	Age <sup>a)</sup> (yr)	Gender	Diagnosis <sup>b)</sup> % clonal population (from WBC) <sup>b)</sup> V(D)J rearrangement <sup>c)</sup>	Notch1 status <sup>c)</sup>	MS <sup>c)</sup>	CD38 status <sup>b)</sup>	ZAP70 status <sup>b)</sup>	BCL2 status <sup>b)</sup>	iFISH <sup>d)</sup>
1	72	F	Typical CLL 68,0 V1-69(D3-3)J6	M	UM	-	-	++	+12 (33%) del(11q22.3) (24%)
2	82	F	N.A. 88,0 V1-2(D6-19)J4	UM	UM	-	-	++	N.D.
3	57	F	Typical CLL 21,0 V3-7(D2-2)J4	N.A.	M	-	-	++	del(13q14.3) (38%)
4	61	F	Atypical CLL 73,0 V4-39(D3-16)J4	N.A.	UM	het	-	++	+12 (80%) del(17p13.1) (36%)
5	45	M	Typical CLL 90,0 V4-34(D2-15)J6	N.A.	M	-	-	++	N.D.
6	71	F	Atypical CLL 70,0 V3-7(D3-3)J6	N.A.	M	het	-	++	+12 (85%)
7	78	F	Atypical CLL 93,4 V4-39(D6-13)J4	UM	M	het	-	++	del(11q22.3) (68%) del(13q14) (47%) biallelic del(13q14.3) (48%)
8	73	M	Typical CLL 32,4 V4-61(D6-13)J4	UM	M	-	-	++	N.D.
9	59	M	Typical CLL 57,0 V3-21(D3-3)J6	UM	UM	het	-	++	N.D.
10	64	M	Atypical CLL 87,0 V4-59(D3-3)J4	M	UM	het	-	++	+12 (92%)
11	44	M	Typical CLL (bad prognosis) 50,0 V3-64(D6-19)J6	M	M	het	-	++	biallelic del(13q14.3) (88%) del(17p13.1) (88%) del(13q14.3) (97%)
12	52	F	Atypical CLL 42,0 V1-69(D3-16)J3	M	UM	het	-	++	del(13q14.3) (85%)
13	69	M	Typical CLL 64,9 V5(D6-19)J4	N.A.	UM	het	-	++	del(13q14.3) (85%)
14	55	M	N.A. 83,0 V1-69(D3-3)J6	N.A.	UM	het	-	++	+12 (73%)

a) Age at time of sample collection.

b) From phenotypic screening.

c) From molecular biology studies.

d) From FISH studies.

WBC, white blood cells. MS, mutational status (UM, unmutated; M, mutated). CLL, chronic lymphocytic leukaemia. N.A., information not available. -, absence of expression. ++, high antigen expression. het, heterogeneous antigen expression. N.D., no chromosomal alterations detected by iFISH for the probes studied.

### 2.3 Data analysis: image processing and normalization

Microarrays were scanned with Scanner GenePix 4000B (Axon, USA) using Cy3 and Cy5 settings. The images were analysed with GenePix Pro 4.0 software (Axon, USA). Firstly, background correction for each spot was locally made by applying standardized settings of GenePix software. Then, the median value of the set of negative control spots in each fluorescence channel was subtracted from the individual spot values. Among several described strategies for data analysis of these Cy5/Cy3 arrays [16], here the same normalization strategy, commonly extended, has been used for protein arrays and for LC-MS/MS assays. Briefly, the mean minus background results of the Cy3 and Cy5 channels were employed for the calculations and a mean value of the two replicates for each spot was calculated. Then, Cy3 and Cy5 values for housekeeping proteins ( $\beta$ -actins) were averaged and a normalization ratio was calculated by dividing the average of

Cy5 values by that of Cy3. Housekeeping proteins were employed as spiked in standards to correct the systematic error. All numerical values of Cy3 spots were normalized by multiplying them by the normalization ratio. Afterwards, a Cy ratio for each protein was calculated considering the value corresponding to the healthy or non-mutated sample as the reference value. Proteins of interest were those with a Cy ratio  $\geq 2$  or  $\leq 0.5$ .

### 2.4 Functional annotation of regulated/differentially expressed proteins

Functional Enrichment Analysis (FEA) and clustering of altered-expression proteins were done using the DAVID [17] tool. The main biological databases selected to find genes with annotated enriched terms were: (i) GO using annotations spaces GOTERM\_BP, GOTERM\_CC and GOTERM\_MF; (ii) the pathways database KEGG\_PATHWAY; (iii) the

INTERPRO protein structural domain database; and (iv) the UNIGENE\_EST and GNF\_U133A\_QUARTILE tissues-specific expression databases. To generate the functional clusters in DAVID we used classification stringency Medium [17].

In addition, Circos plots [18] (Circos Table Viewer, v0.63-9) were designed for up- and down-regulated proteins.

## 2.5 LC-MS/MS assays

### 2.5.1 In-solution digestion and LC-MS/MS analysis

A total of 10 µg of cytoplasmic proteins from each sample were dissolved in 50 mM ammonium bicarbonate/8M urea and reduced and alkylated with 10 mM DTT at 37°C for 1 h, and with 55 mM iodoacetamide at RT for 1 h, respectively. Proteins were digested for 18 h with trypsin 1:100 (w/w, enzyme to protein ratio) at 37°C. The peptide solution was acidified with formic acid and desalted by using C18-Stage-Tips columns [19]. The samples were partially dried and stored at -20°C until analysed by LC-MS/MS.

A nano-UPLC system (nanoAcquity, Waters Corp., Milford/MA, USA) coupled to a LTQ-Velos-Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose/CA, USA) via a nano-electrospray ion source (NanoSpray flex, Proxeon, Thermo) was used for reversed-phase LC-MS/MS analysis. Peptides were dissolved in 0.5% formic acid / 3% acetonitrile and loaded onto a trapping column (nanoACQUITY UPLC 2G-V/M Trap Symmetry 5 µm particle size, 180 µm × 20 mm C18 column, Waters Corp., Milford/MA, USA). Peptides were separated on a nanoACQUITY UPLC BEH 1.7 µm, 130 Å, 75 µm × 250 mm C18 column (Waters Corp., Milford/MA, USA) with a linear gradient from 3% to 35% solvent B (ACN/0.1% formic acid) at a flow rate of 250 nL/min over 120 minutes.

The LTQ-Orbitrap Velos was operated in the positive ion mode applying a data-dependent automatic switch between survey MS scan and tandem mass spectra (MS/MS) acquisition. Survey scans were acquired in the mass range of  $m/z$  400 to 1600 with 60 000 resolution at  $m/z$  400 with lock mass option enabled for the 445.120025 ion [20].

The 20 most intense peaks having  $\geq 2$  charge state and above the 500 intensity threshold were selected in the ion trap for fragmentation by collision-induced dissociation with 37% normalized energy, 10 ms activation time,  $q = 0.25$ ,  $\pm 2$   $m/z$  precursor isolation width and wideband activation. Maximum injection time was 250 and 50 ms for survey and MS/MS scans, respectively. AGC was  $1 \times 10^6$  for MS and  $5 \times 10^3$  for MS/MS scans. Dynamic exclusion was enabled for 180 s.

### 2.5.2 Database search

Raw data were translated to mascot general file (mgf) format using Thermo Scientific Proteome Discoverer software (v. 1.4.1.14). Peak lists obtained from MS/MS spectra were identified

using Comet version 2015.01 rev. 2 [21]. The search was conducted using SearchGUI version 1.30.1 [22]. Protein identification was conducted against a concatenated target/decoy version of all human-reviewed sequences downloaded from the UniProt database (version of November 2015, 20310 target sequences) and common contaminant sequences (e.g., human keratins, trypsin). The decoy sequences were created by reversing the target sequences in SearchGUI. The identification settings were as follows: trypsin with a maximum of two missed cleavages; 10.0 ppm as MS1 and 0.6 Da as MS2 tolerances; fixed modifications: carbamidomethylation of cysteine, variable modifications: acetylation of the protein n-terminus and oxidation of methionine. Peptides and proteins were inferred from the spectrum identification results using PeptideShaker version 0.41.1 [23]. Peptide Spectrum Matches (PSMs), peptides and proteins were validated at a 1% FDR estimated using the decoy hit distribution.

### 2.5.3 Protein abundance determination

A label-free technique based on spectral counting was selected to determine the abundance of the proteins identified by LC-MS/MS. Specifically, the exponentially modified protein abundance index (emPAI) was used to calculate the protein content in molar percentage [24] following the Eq. (1).

$$\text{Protein content (mol \%)} = \frac{\text{emPAI}}{\sum(\text{emPAI})} \times 100 \quad (1)$$

where  $\sum(\text{emPAI})$  is the summation of emPAI values for all identified proteins.

emPAI values were directly obtained from the report generated by the PeptideShaker software.

### 2.5.4 Statistical methods

A Pearson correlation coefficient was determined to assess the relationship between the healthy samples across the pooled microarray assays (SPSS software 18.0 package; SPSS, Inc., Chicago, IL). Correlation significance was set at a  $p$  value of  $<0.01$ .

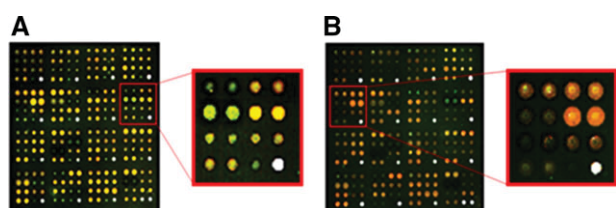
## 3 Results and discussion

A total of 14 clinical samples (Table 1) were screened with the antibody array platform under the same conditions as described in the Materials and Methods Section. For establishing global differential protein profiles, a total of eleven assays were performed (Table 2): assays #1–6 were comparisons carried out with pools of cytoplasmic proteins obtained from samples with the same cytogenetic alterations; whereas

**Table 2.** Relation of samples for Panorama antibody microarray assays

	Assay	Dye-labelling		Gene Alteration
		Cy5-labelled sample	Cy3-labelled sample	
pool sample	1	B-CLL	healthy	B-CLL
	2	healthy	B-CLL	B-CLL
	3	M <sub>Notch1</sub>	UM <sub>Notch1</sub>	Notch1 status
	4	UM	M	MS
	5	healthy	M <sub>Notch1</sub>	Notch1 status
	6	UM <sub>Notch1</sub>	healthy	Notch1 status
individual sample	7	case ID 14	healthy	+12
	8	healthy	case ID 7	del(13q14)
	9	case ID 3	healthy	del(13q14.3)
	10	healthy	case ID 11	del(17p13.1)
	11	case ID 2	healthy	N.D.

Pool samples are referred to the mixing of equal amounts of dye-labelled-proteins belonging to the same group characterized by the studied alteration. MS, mutational status (UM, unmutated; M, mutated); N.D., no chromosomal alterations detected by iFISH for the probes studied; B-CLL, chronic lymphocytic leukaemia of B-cells. Case ID referred to Table S-1. B-CLL pool includes all the patients (cases ID 1–14); MNotch1 pool (NOTCH1 mutated) includes ID 1, 10–12; UMNotch1 pool (NOTCH1 non-mutated) includes ID 2, 7–9; UM pool includes ID 1–2, 4, 9–10, 12–14; M pool includes ID 3, 5–8, 11.



**Figure 2.** Representative images of antibody microarrays. They were obtained with scanner GenePix 4000B and processed with GenePix Pro 4.0 software. Scanner settings were identical within microarrays. (A) The left image shows the upper section of the assay #6. The right image corresponds to a zoom of one block in which is highlighted the different colour expression between spots depending on the probed samples. (B) The left image shows the upper section of the assay #9. The right image is a zoom of the selected block.

assays #7–11 corresponded to comparisons across pathology (these assays were performed with side by side comparisons between healthy and CLL samples). In this study, for the relative protein abundance, cytoplasmic proteins from B cells obtained from healthy donors were used as reference (assays #1–2, #5–11); except particular assays #3–4 that corresponded to comparisons with the non-mutated counterparts. Representative images of performed experiments are shown in Fig. 2.

As it was described above, this study has been focused on an experimental plan for deciphering differential protein profiles in several pathological situations (by multipronged proteomics tools) in order to correlate them with cytogenetic and transcriptomics information. Afterwards, for further and deep screening, selected cell signalling proteins as possible biomarkers will require to be tested with a large cohort of CLL samples.

### 3.1 Differential protein expression profiles

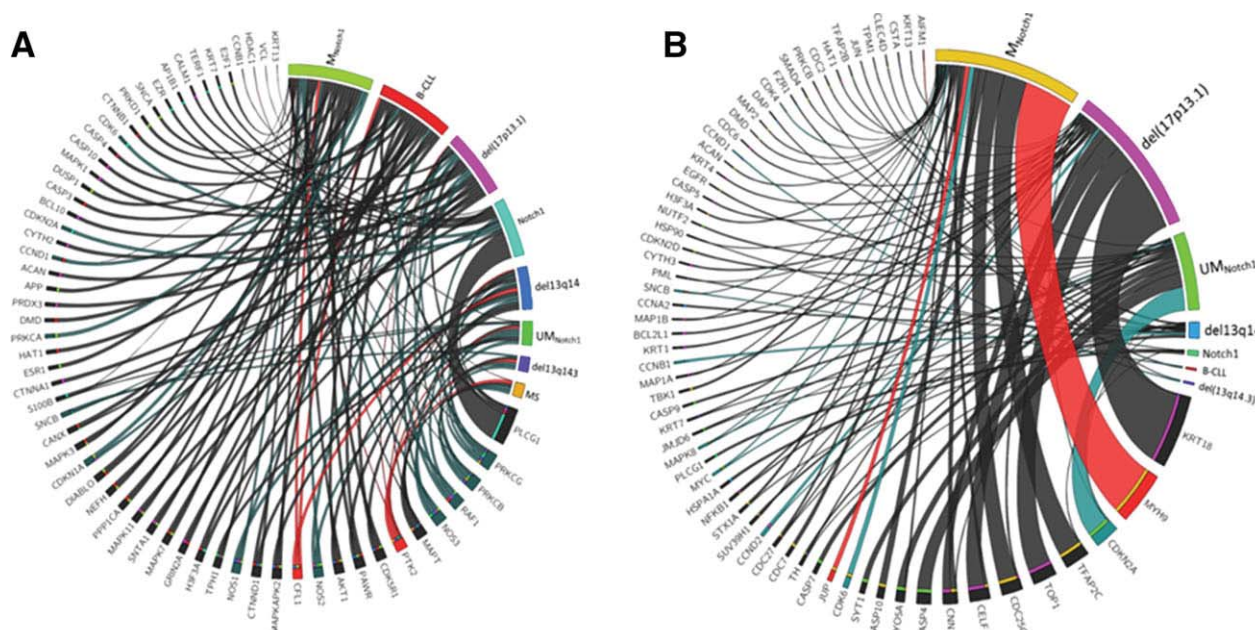
This experimental design was focused on the study and identification of global differential protein profiles in pathological conditions, together with the analysis of the possible correlation between the protein expression patterns and the cytogenetic alterations.

#### 3.1.1 Quality control assays

To evaluate the possible influence of dye labelling in the array platform, a quality control assay was performed with the same samples but swapping the dyes. Specifically, normalized data sets of Cy3 and Cy5 channels from healthy control were selected to test this effect. In Supporting Information Fig. S2, it is represented both data sets showing an  $r^2$  value equal to 0.93 indicating the results are dye-conjugation independent. In a similar manner, the correlation between the healthy samples in all pooled assays (assays #1, 2, 5, and 6) was evaluated to confirm the absence of processing variability. Pearson correlation coefficients (Supporting Information Table S1) were calculated confirming the high linear relationship.

#### 3.1.2 Global analysis of differential protein expression in CLL samples

The up- and down-regulated proteins identified beside the performed assays are shown in Supporting Information Tables S2–9. Representative Circos plots for up- and down-regulated proteins identified throughout the assays are shown in Fig. 3. The relation between the protein and the cytogenetic alteration is proportional to the ribbon size. Thus, it is possible to visualize the contribution of a particular protein depending on the genetic alteration present in the samples.



**Figure 3.** Graphical visualization of down- and up-regulated proteins (panel A and B, respectively) represented by Circos plot. The correlation between proteins and cytogenetic alterations are shown by ribbons. Thus, specific proteins are correlated with different alterations at gene level and the correlation degree is proportional to the width of the ribbon. In such a way, tight ribbons are referred to low expression values, and wide ribbons to high expression values. M<sub>Notch1</sub>, NOTCH1 mutated (assay #5); UM<sub>Notch1</sub>, NOTCH1 non-mutated (assay #6); B-CLL, chronic lymphocytic leukaemia of B-cells (assays #1 and 2); del(17p13.1), p53 deletion (assay #10); del(13q14.3), Rb1 deletion (assay #9); MS, mutational status (assay #4); Notch1, NOTCH1 status (assay #3); del(13q14) (assay #8). Red ribbons correspond to proteins also identified by LC-MS/MS (MS/MS quantification strategy for validation). Green ribbons correspond to proteins highlighted throughout the text.

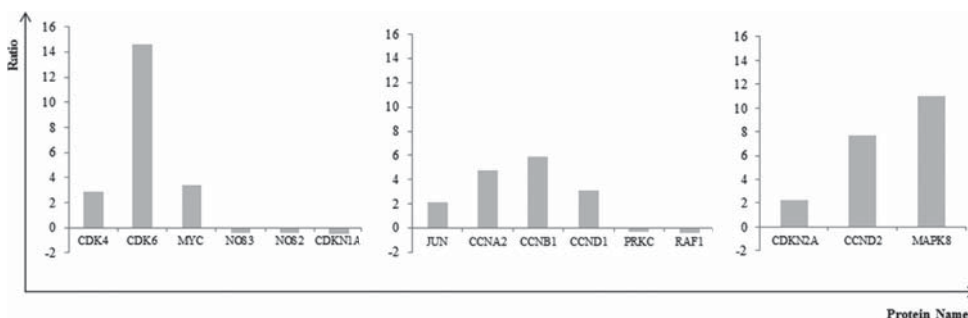
The global analysis of protein profiles suggests that more differential expressed proteins are presented in aberrant B cells (cytogenetically altered del(13q14), del(17p13.1) and NOTCH1) than in normal B cells (from healthy donors) (assays # 8,10, and 5–6, respectively). In turn, global protein profiles are quite similar between normal B cells and pathological B cells with no genetic alterations and only trisomy 12 as unique cytogenetic alteration (assays #11 and 7, respectively).

Additionally, there seem to be not significant differences at global protein profiles concerning the mutational status of IgVH genes (assay #4, Supporting Information Table S4). In turn, the absence of genetic alterations correlates with the

results at protein profile since there are not noteworthy detected changes in assay #11, in which a B-CLL sample without cytogenetic alterations or gene mutations was compared with the healthy control.

The results revealed an overall down-regulation trend concerning PKC family members in all the assays. These proteins are involved in the typical slowly progressive indolent behaviour in CLL cases, which are considered as important survival factors, as previously described by Alkan et al. [5].

Also, Mcl-1 (Bcl-2 family member) expression is differentially determined when NOTCH1 and p53 alterations are presented. As described by Liu et al. [25], the up-regulation of



**Figure 4.** Main altered proteins detected in each group of study. A) NOTCH1 group, including assays #3, 5 and 6; B) del(13q14), including assays #8 and 9; C) del(17p13.1), including assay #10.

Mcl-1 contributes to a malignant phenotype since this protein plays an important role in apoptosis inhibition.

### 3.2 Analysis of differentially expressed proteins in B-CLL cells including genetic alterations

Next step consisted of the characterization of the correlation between protein expression profiles, the cytogenetic alterations, the existing transcriptomic data and the cellular signaling pathways; which could be associated for a better understanding of the B-CLL pathology. In Fig. 4, it is represented the main altered proteins detected for each group of study (according to reported genetic alterations) and in Supporting Information Table S10, the significantly altered cell signalling pathways identified after a functional enrichment analysis.

#### 3.2.1 NOTCH 1 mutational status

Among other studied genetic alterations, NOTCH1 analysis revealed a protein profile containing the higher number of deregulated proteins. Firstly, the NOTCH1 mutated pool was compared with the non-mutated counterpart (assay #3) to establish the differences between them at protein expression level; and also to analysis the different protein profiles with regard to the normal B cells from healthy controls (assays #5 and 6).

For mutated versus non-mutated NOTCH1 assay, the greatest differences appeared in down-regulated proteins. Pathological B cells with mutations in NOTCH1 exhibited relative lower expression levels of cyclin B1, PKC $\delta$  and  $\beta$ -synuclein, among others.

For NOTCH1 non-mutated samples versus normal B-CLL cells (assay #6), a protein profile was identified containing a large number of down- and up-regulated proteins. This fact might demonstrate the significance of NOTCH1 mutations to vary the standard protein expression profiles in B cells, and it could be related as a progression or prognostic factor of the disease.

NOTCH1 mutated samples (assay #5) presented relative low abundance levels of p21 CDK inhibitor, whereas CDK4 and 6 seem to be up-regulated. These results might suggest that an increment in cell cycle activity may be expected in these pathological cells since the cyclin D-CDK4 complex is not inhibited by p21 CDK inhibitor. In addition, a deregulation in c-Myc was presented, which is responsible of poor prognostic. Also, this protein was altered in NOTCH1 non-mutated pool, possibly due to a combination effect with other co-existing genetic alterations presented in the samples. In turn, non-mutated status presented a higher relative abundance in proteins involved in cellular response to stress.

Five up-regulated and 8 down-regulated proteins were in common between mutated and non-mutated status. In both

cases, there was a decrease in relative abundance of NOS family protein members.

#### 3.2.2 Chromosome alteration: del(13q14)

Results derived from assays #8 and 9 depicted an increment in relative abundance of JUN, which is considered one of the leading proteins in the proliferation of B cells. In turn, low relative abundance of PRKC was identified, which it might be corresponded with a reduction of the apoptosis via p38 MAPK pathway. RAF1 is known as an activating protein of BCR signalling pathway; in this study, RAF1 could be down-regulated according to the relative abundance levels detected of this protein. This could lead to the inactivation of the cell cycle through MEK/ERK signalling. However, the significant increment (in relative abundance) of cyclins (A, B1, D1) may act as a counter-balance to the negative effect of RAF1.

#### 3.2.3 Chromosome alteration: del(17p13.1)

p53 tumour suppressor gene is affected by 17p deletions. Previous studies in this field determined a poor outcome for patients with deletions or mutations in p53 protein. Our results indicated a highly variation in relative protein expression patterns regarding to p53<sup>wt</sup> from normal B cells obtained from healthy donors. Bearing in mind the relative abundance detected levels, 22 proteins were considered down-regulated and 27 were up-regulated (Supporting Information Table S9). Cell signalling analysis revealed an alteration in phosphatidylinositol and ErbB signalling pathways, as well as in programmed cell death and apoptosis. Also, leukocyte trans-endothelial migration seems to be also down-regulated due to the low relative abundance.

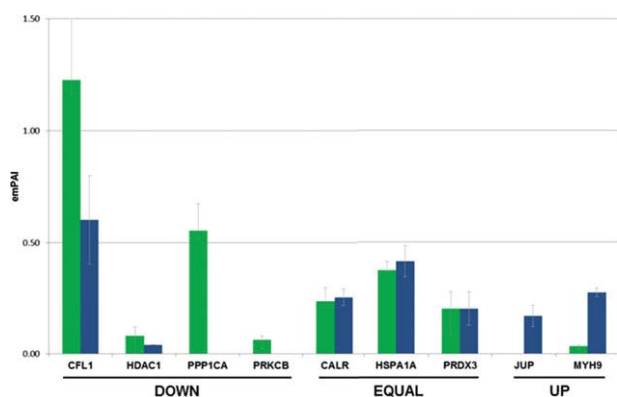
In this study, the relative abundance of CDKN2A, stabilizer of p53, was increased. This protein might act in response to p53 degradation, preventing its union to MDM2, the principal regulator of p53.

In the same way as in the cases that present del(13q14), RAF/MEK signalling was decreased in relative abundance, but JNK up-regulation could re-direct the intra-cellular signalling pathway to proliferation and differentiation, possibly due to the alteration of the remaining copy of p53. Also, the abundance level of p21 CDK inhibitor was down-regulated. Thus, cyclin A/CDK2-1 and cyclin D/CDK4 complexes were not inhibited, which could be traduced in an increment in cell proliferation, and avoiding G2 and G1 cell cycle arrest, respectively.

#### 3.2.4 Chromosome alteration: trisomy 12 and no-cytogenetic alterations

No significant differences in global protein expression profiles were observed in B-CLL cells from patients with trisomy





**Figure 5.** emPAI data for proteins included in the protein microarray and identified by LC-MS/MS. As described for the protein microarrays, there are down-regulated (DOWN), up-regulated (UP) and proteins without differences in expression (EQUAL) when comparing NOTCH1 pooled samples and healthy controls (reference). Green bars correspond to healthy pooled samples, blue bars correspond to mutated NOTCH1 pooled samples. Error bars were calculated as the standard deviations derived from the three replicates of each sample.

12 and/or absence of cytogenetic alterations. Although they are pathological cells, these genetics characteristics do not produce any modification at relative protein abundance when comparing with global protein profiles from normal B-cells.

### 3.3 Integration of LC-MS/MS quantification strategy

In order to validate the differentially expressed proteins identified by the antibody microarray platform, a LC-MS/MS quantification strategy was performed with healthy and NOTCH1 mutated samples (the identified proteins are indicated in Supporting Information Table S5). To keep the same conditions employed with microarrays, 10  $\mu$ g of cytoplasmic proteins extracted from B cells of healthy pooled control and 10  $\mu$ g from NOTCH1 mutated pool were analysed in triplicate by LC-MS/MS as described in the Materials and Methods section. The three runs performed for each sample resulted in a total of  $1000 \pm 30$  and  $862 \pm 41$  proteins identified in the healthy and NOTCH1 pooled samples, respectively (complete lists of identified proteins are shown in Supporting Information Tables S11 and S12). To validate if differential expression profiles of proteins up- and down- detected by microarrays were also identified by this complementary approach, a label-free strategy based on spectral counting was employed. emPAI values were obtained from the LC-MS/MS data and comparisons of these values between healthy and NOTCH1 mutated samples were carried out to estimate protein abundance and confirm protein microarray results. Since three technical replicates were performed for each sample, the emPAI values in Fig. 5 correspond to mean values (error bars were calculated as the standard deviation). Thus, Fig. 5

displays the emPAI values corresponding to several proteins identified as down- (CFL1, HDAC1, PPP1CA, PRKCB) and up-regulated (JUP, MYH9) by the microarray platform. Additionally, three proteins included in the microarray that were not differentially expressed in the healthy vs NOTCH1 mutated comparison (CALR, HSPA1A, PRDX3) have been represented in the graphic showing a similar protein abundance for both samples (detected by the LC-MS/MS strategy).

In summary, as the emPAI estimation offers a quantitative approach to evaluate proteins as increasing or decreasing in comparison to a reference (healthy controls), the LC-MS/MS data confirm the results obtained by the protein microarray and highlight the usefulness of combine different proteomics techniques. Moreover, independently of using targeted (protein microarray) or untargeted methods (label-free technique), there is a good correlation between the obtained results, which indicates the robustness, complementarity and accuracy of both approaches in deciphering alterations in cell signalling pathways.

### 3.4 Overview of transcriptomic and proteomic studies for B-CLL cells

To compare our antibody-microarray results with those obtained in Genomics field, a transcriptomic database of CLL studies from The European Bioinformatics Institute (EMBL-EBI) was employed. This database contains information about the expression value of genes in CLL cases. Results from three genomic assays (E-GEOD-10820, E-GEOD-3982, and E-Mtab-62) were employed for the comparison. Up to 60% of our proteomic results agreed with the transcriptomic database. Specifically, the assays #1-2, 8 and 10 (corresponding to B-CLL, del(13q14), and del(17p13.1), respectively) showed the highest concordance with the transcriptomics results.

## 4 Concluding remarks

Multipronged proteomic studies in combination with genetic information might improve the knowledge about B-CLL pathogenesis and the deciphering of the intracellular signalling pathways involved in the development of the disease. Specifically, affinity proteomics (i.e., antibody microarrays) are a useful methodology for the rapid and accurate screening of global protein expression profiles of B cells. Additionally, the requirement of minimal amount of sample constitutes a great advantage together with the high specificity and sensitivity. Herein, the usefulness of a well-characterized and established antibody microarray for proteomic analysis of B-CLL samples has been tested allowing the integration of proteomics and cytogenetic information with high clinical interest.

According to our knowledge, this is a comprehensive study in which cytoplasmic proteins derived from purified B cells

from CLL patients are analysed by using antibody microarrays. In addition, B-CLL samples are stratified in function of their cytogenetic alterations that allow not only the identification of general protein expression patterns, but also the specific alterations at protein level depending on the altered gene or NOTCH1 status of each sample. To obtain more information from our samples, the assays were based on two working directions: B-CLL samples (pool) comparing to a healthy reference and side by side comparisons (belonging to a specific cytogenetic group). Cell signalling pathways were identified as altered depending on the described alterations including, among others, regulation of apoptosis, proliferation and activation of lymphocytes, and cell cycle pathways.

Finally, a LC-MS/MS quantification strategy based on spectral counting (empAI) was used to validate the results obtaining a high correlation between both analysis strategies (protein microarrays and LC-MS/MS).

*We gratefully acknowledge financial support from the Carlos III Health Institute of Spain (ISCIII, FIS PI11/02114 and FIS PI14/01538), Fondos FEDER (EU), Junta Castilla-León SA198A12-2 and BIO/SA03/15. The proteomics analysis was performed in the Proteomics facility of Centro de Investigación del Cáncer (Salamanca, Spain) that belongs to ProteoRed, PRB2-ISCIII, supported by grant PT13/0001. P.D. is supported by a JCYL-EDU/346/2013 PhD scholarship. We are grateful to EMBL-EBI for providing us the CLL databases. We are also grateful to Andrea Mayado for the supports with the statistical analysis.*

*The authors have declared no conflict of interest.*

## 5 References

- [1] Chiorazzi, N., Rai, K. R., Ferrarini, M., Chronic lymphocytic leukemia. *N. Engl. J. Med.* 2005, **352**, 804–815.
- [2] Lopez, C., Delgado, J., Costa, D., Villamor, N. et al., Clonal evolution in chronic lymphocytic leukemia: analysis of correlations with IGHV mutational status, NOTCH1 mutations and clinical significance. *Genes. Chromosomes Cancer* 2013, **52**, 920–927.
- [3] Dasilva, N., Díez, P., Matarraz, S., González-González, M. et al., Biomarker discovery by novel sensors based on nanoproteomics approaches. *Sensors* 2012, **12**, 2284–2308.
- [4] Villamor, N., Conde, L., Martínez-Trillos, A., Cazorla, M. et al., NOTCH1 mutations identify a genetic subgroup of chronic lymphocytic leukemia patients with high risk of transformation and poor outcome. *Leukemia* 2013, **27**, 1100–1106.
- [5] Alkan, S., Huang, Q., Ergin, M., Denning, M. F. et al., Survival role of protein kinase C (PKC) in chronic lymphocytic leukemia and determination of isoform expression pattern and genes altered by PKC inhibition. *Am. J. Hematol.* 2005, **79**, 97–106.
- [6] Zanesi, N., Balatti, V., Bottoni, A., Croce, C. M., Pekarsky, Y., Novel insights in molecular mechanisms of CLL. *Curr. Pharm. Des.* 2012, **18**, 3363–3372.
- [7] Ghia, P., Stamatopoulos, K., Belessi, C., Moreno, C. et al., ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia* 2007, **21**, 1–3.
- [8] Calin, G. A., Ferracin, M., Cimmino, A., Di Leva, G. et al., A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N. Engl. J. Med.* 2005, **353**, 1793–1801.
- [9] Rickert, R. C., New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. *Nat. Rev.* 2013, **13**, 578–591.
- [10] Alsagaby, S. A., Khanna, S., Hart, K. W., Pratt, G. et al., Proteomics-Based Strategies To Identify Proteins Relevant to Chronic Lymphocytic Leukemia. *J. Proteome Res.* 2014, **13**, 5051–5062.
- [11] Végvári, Á., Welinder, C., Lindberg, H., Fehniger, T. E., Markovarga, G., Biobank resources for future patient care: developments, principles and concepts. *J. Clin. Bioinforma.* 2011, **1**, 24.
- [12] Haslinger, C., Schweifer, N., Stilgenbauer, S., Dohner, H. et al., Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status. *J. Clin. Oncol.* 2004, **22**, 3937–3949.
- [13] Nieto, W. G., Almeida, J., Romero, A., Teodosio, C. et al., Increased frequency (12%) of circulating chronic lymphocytic leukemia-like B-cell clones in healthy subjects using a highly sensitive multicolor flow cytometry approach. *Blood* 2009, **114**, 33–37.
- [14] Henriques, A., Rodríguez-Caballero, A., Criado, I., Langerak, A. W. et al., Molecular and cytogenetic characterization of expanded B-cell clones from multiclonal versus monoclonal B-cell chronic lymphoproliferative disorders. *Haematologica* 2014, **99**, 897–907.
- [15] Henriques, A., Rodríguez-Caballero, A., Nieto, W. G., Langerak, A. W. et al., Combined patterns of IGHV repertoire and cytogenetic/molecular alterations in monoclonal B lymphocytosis versus chronic lymphocytic leukemia. *PLoS One* 2013, **8**, e67751.
- [16] Ptacek, J., Devgan, G., Michaud, G., Zhu, H. et al., Global analysis of protein phosphorylation in yeast. *Nature* 2005, **438**, 679–684.
- [17] Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J. et al., DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 2003, **4**, P3.
- [18] Krzywinski, M., Schein, J., Birol, I., Connors, J. et al., Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009, **19**, 1639–1645.
- [19] Rappsilber, J., Mann, M., Ishihama, Y., Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2007, **2**, 1896–1906.
- [20] Olsen, J. V., de Godoy, L. M., Li, G., Macek, B. et al., Parts per million mass accuracy on an Orbitrap mass spectrometer via

- lock mass injection into a C-trap. *Mol. Cell. Proteomics* 2005, 4, 2010–2021.
- [21] Eng, J. K., Jahan, T. A., Hoopmann, M. R., Comet: an open-source MS/MS sequence database search tool. *Proteomics* 2013, 13, 22–24.
- [22] Vaudel, M., Barsnes, H., Berven, F. S., Sickmann, A. et al. SearchGUI: an open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics* 2011, 11, 996–999.
- [23] Vaudel, M., Burkhart, J. M., Zahedi, R. P., Oveland, E. et al. PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat. Biotechnol.* 2015, 33, 22–24.
- [24] Ishihama, Y., Oda, Y., Tabata, T., Sato, T. et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* 2005, 4, 1265–1272.
- [25] Liu, X. S., Jiang, J., Jiao, X. Y., Wu, Y. E. et al., Lycorine induces apoptosis and down-regulation of Mcl-1 in human leukemia cells. *Cancer Lett.* 2009, 274, 16–24.



**Una aproximación sistemática para la  
caracterización peptídica del receptor de célula  
B en células de la leucemia linfocítica crónica**

3

Paula Díez<sup>1,2</sup>, Nieves Ibarrola<sup>2</sup>, Rosa M. Dégano<sup>2</sup>,  
Quentin Lécrevisse<sup>1</sup>, Arancha Rodríguez-Caballero<sup>1</sup>,  
Ignacio Criado<sup>1</sup>, Wendy Nieto<sup>1</sup>, Rafael Góngora<sup>1</sup>,  
Marcos González<sup>3</sup>, Julia Almeida<sup>1</sup>, Alberto Orfao<sup>1</sup>,  
Manuel Fuentes<sup>1,2</sup>

<sup>1</sup> Servicio General de Citometría-NUCLEUS, Departamento de Medicina;  
Centro de Investigación del Cáncer (IBMCC-CSIC/USAL) e IBSAL,  
Universidad de Salamanca.

<sup>2</sup> Unidad de Proteómica; Centro de Investigación del Cáncer (IBMCC-  
CSIC/USAL) e IBSAL, Universidad de Salamanca.

<sup>3</sup> Servicio de Hematología (IBMCC) e IBSAL, Hospital Universitario de  
Salamanca.

*Oncotarget*. 2017 Abr.  
doi: 10.18632/oncotarget.17076



**Introducción.** En respuesta al gran abanico de antígenos existentes, el sistema inmune tiene la capacidad de producir un amplio repertorio de Ig mediante procesos como la recombinación VDJ, la hipermutación somática y la selección del receptor. Las Igs están presentes como proteínas solubles en distintos fluidos corporales (p.ej. plasma, líquido cefalorraquídeo, líquido sinovial, saliva) o como proteínas unidas a membranas en la superficie de las células B donde actúan como elemento efector y regulador de la respuesta inmune y receptores específicos de Ag en la célula B, respectivamente. En la actualidad, se cree que la activación celular descontrolada por la señalización intracelular mediada por el BCR podría tener un papel clave en la patogenia de la LLC.

La caracterización de la secuencia concreta de la Ig (tanto a nivel de ADN como peptídico) mediante secuenciación de nueva generación (NGS, del inglés *Next-Generation Sequencing*) y MS han facilitado notablemente la caracterización de las Ig al constituir ambas técnicas aproximaciones de alto rendimiento y sensibilidad. Mientras que la secuenciación a nivel de ADN se ha convertido en una técnica de rutina en muchos laboratorios, el análisis peptídico de las secuencias del BCR sigue suponiendo un reto debido principalmente a los problemas en el procesamiento y purificación de la proteína. Aunque la caracterización de las Igs a nivel genómico proporciona una información detallada de la secuencia de las mismas, la evaluación proteómica tiene valor adicional ya que pequeñas variaciones en la secuencia de las Ig, sólo detectables a nivel peptídico, pueden conducir a modificaciones en la selectividad, especificidad y afinidad en la unión del Ag. En la actualidad, la caracterización del BCR de la célula tumoral de la LLC-B supone un importante reto debido principalmente a la baja abundancia relativa del BCR y a la dificultad para aislar de forma eficaz proteínas ancladas a membranas celulares.

**Objetivo.** En este trabajo planteamos evaluar tres métodos distintos de procesamiento para la secuenciación peptídica del BCR purificado a partir de células B de pacientes con LLC mediante técnicas proteómicas de alta resolución, y determinar su

correlación con las secuencias génicas identificadas por técnicas convencionales de biología molecular.

**Muestras, materiales y métodos.** En conjunto analizamos 9 muestras de pacientes diagnosticados de LLC (8 presentaban una población clonal B slg kappa<sup>+</sup> y 1 presentaba una población clonal B slg lambda<sup>+</sup>). La purificación de los linfocitos B de los pacientes se llevó a cabo en paralelo con la de células B de individuos sanos mediante citometría de flujo. La secuenciación molecular de las secuencias génicas VDJ se realizó mediante PCR siguiendo las recomendaciones del Consorcio EuroClonality. Para la purificación de las proteínas de citoplasma y de membrana se empleó un protocolo de subfraccionamiento celular. Asimismo, se realizó un enriquecimiento en Ig usando microesferas de agarosa recubiertas de proteína G. A continuación, se evaluaron tres estrategias para la caracterización peptídica de la Ig de célula B: i) aislamiento de regiones Fab de Ig basado en la digestión en fase sólida con papaína y posterior enriquecimiento mediante proteína A; ii) tratamiento con papaína en solución, para la digestión de Ig inmovilizadas en microesferas de agarosa y la obtención de los fragmentos Fab en forma soluble; y iii) desnaturalización con DTT de Ig inmovilizadas en microesferas para la rotura de los puentes disulfuro y la obtención de las cadenas pesadas y ligeras en forma soluble.

Todas las muestras resultantes fueron analizadas mediante electroforesis unidimensional y procesadas para su posterior análisis mediante LC-MS/MS usando un equipo LTQ-Orbitrap-Velos y el programa informático PeptideShaker para el análisis de los datos con la base UniProt (versión de febrero de 2014) como referencia.

Con la finalidad de evaluar los resultados proteómicos obtenidos con aquellos derivados de la secuenciación molecular, se diseñó un algoritmo para la integración de ambos conjuntos de datos y el análisis de la homología entre datos de secuencias obtenidas con ambas estrategias.



**Resultados.** La evaluación de las tres estrategias de procesamiento propuestas mostró que el procedimiento basado en la desnaturalización mediante DTT era el más eficiente al permitir identificar 171 proteínas pertenecientes al sistema inmune (frente a las 40 y 54 proteínas identificadas con las otras dos aproximaciones), correspondiendo 21% de ellas a proteínas del sistema HLA.

Respecto a la caracterización proteómica del BCR, se identificaron hasta 426 secuencias peptídicas diferentes. Cabe mencionar que sólo 15 de los péptidos identificados eran comunes a las tres estrategias, lo que sugiere la necesidad e importancia de definir con detalle y precisión el protocolo de procesamiento de muestras complejas para la secuenciación peptídica del BCR con el fin de obtener resultados fiables y reproducibles. En este sentido, se establecieron las siguientes pautas: i) incluir un paso de enriquecimiento específico en Ig; ii) reducir el número de pasos del procedimiento tanto cuanto sea posible para evitar la pérdida de muestra, al tratarse de proteínas asociadas directamente a membranas plasmáticas; iii) emplear un motor de búsqueda proteómico que permita la identificación de péptidos no asignados a proteínas, debido a la alta variabilidad de las secuencias de Ig, lo cual dificulta la existencia de librerías peptídicas completas; y iv) integrar la información proteómica y genómica para obtener una visión más completa de las verdaderas secuencias de las Igs.

Finalmente, la integración de los datos proteómicos y de biología molecular reportó una coincidencia de identificación hasta 62% de la secuencia de Ig, lo cual supone un punto de partida adecuado para incentivar los estudios MS/MS de secuencias del BCR.

**Conclusiones.** En resumen, nuestros resultados documentan el potencial de la aplicación de una aproximación simple para llevar a cabo una caracterización proteogenómica compleja a través de la integración de datos genómicos y proteómicos que podría proporcionar una caracterización detallada y más amplia de

las secuencias de Ig, del BCR así como de proteínas relacionadas con el sistema inmune expresadas en momentos y condiciones celulares concretos.

\* Los resultados proteómicos fueron depositados en ProteomeXchange (PXD004466).

## A systematic approach for peptide characterization of B-cell receptor in chronic lymphocytic leukemia cells

Paula Díez<sup>1,2</sup>, Nieves Ibarrola<sup>2</sup>, Rosa M. Dégano<sup>2</sup>, Quentin Lécresse<sup>1</sup>, Arancha Rodríguez-Caballero<sup>1</sup>, Ignacio Criado<sup>1</sup>, Wendy G. Nieto<sup>1</sup>, Rafael Góngora<sup>1</sup>, Marcos González<sup>3</sup>, Julia Almeida<sup>1</sup>, Alberto Orfao<sup>1</sup> and Manuel Fuentes<sup>1,2</sup>

<sup>1</sup>Department of Medicine and General Cytometry Service-Nucleus, Cancer Research Centre (IBMCC/CSIC/USAL/IBSAL), 37007 Salamanca, Spain

<sup>2</sup>Proteomics Unit, Cancer Research Centre (IBMCC/CSIC/USAL/IBSAL), 37007 Salamanca, Spain

<sup>3</sup>Hematology Service, Institute of Biomedical Research of Salamanca, University Hospital of Salamanca, Cancer Research and Institute of Molecular Biology and Cellular Oncology, 37007 Salamanca, Spain

Correspondence to: Manuel Fuentes, email: mfuentes@usal.es

Keywords: B-cell receptor, chronic lymphocytic leukemia, immunoglobulin, mass spectrometry, peptide sequencing

Received: December 22, 2016

Accepted: March 22, 2017

Published: April 13, 2017

Copyright: Díez et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

A wide variety of immunoglobulins (Ig) is produced by the immune system thanks to different mechanisms (V(D)J recombination, somatic hypermutation, and antigen selection). The profiling of Ig sequences (at both DNA and peptide levels) are of great relevance to developing targeted vaccines or treatments for specific diseases or infections. Thus, genomics and proteomics techniques (such as Next-Generation Sequencing (NGS) and mass spectrometry (MS)) have notably increased the knowledge in Ig sequencing and serum Ig peptide profiling in a high-throughput manner. However, the peptide characterization of membrane-bound Ig (e.g., B-cell receptors, BCR) is still a challenge mainly due to the poor recovery of mentioned Ig.

Herein, we have evaluated three different sample processing methods for peptide sequencing of BCR belonging to chronic lymphocytic leukemia (CLL) B cells identifying up to 426 different peptide sequences (MS/MS data are available via ProteomeXchange with identifier PXD004466). Moreover, as a consequence of the results here obtained, recommended guidelines have been described for BCR-sequencing of B-CLL samples by MS approaches.

For this purpose, an in-house algorithm has been designed and developed to compare the MS/MS results with those obtained by molecular biology in order to integrate both proteomics and genomics results and establish the steps to follow when sequencing membrane-bound Ig by MS/MS.

### INTRODUCTION

The immune system has the capacity to produce a vast repertoire of immunoglobulins, (Ig) in response to the wide number of existing antigens, by processes such as V(D)J recombination (it occurs during B-cell maturation) [1, 2], somatic hypermutation (generated during B-cell affinity maturation) [3, 4] and antigen selection (during B-cell activation) [5].

The Ig structure comprises four chains: two identical heavy chains and two identical light chains linked by

disulfide bonds. Heavy chains are classified according to the heavy chain into 5 groups:  $\alpha$ ,  $\delta$ ,  $\epsilon$ ,  $\gamma$ , and  $\mu$  (for IgA, IgD, IgE, IgG, and IgM, respectively). Concerning light chains, there are two types: kappa ( $\kappa$ ) and lambda ( $\lambda$ ). In turn, each heavy and light chain presents a constant (C) and a variable (V) region. The effector function of the Ig is located in the C region of the heavy chain. The Igs also comprise fragment antigen binding (Fab) and fragment crystallizable (Fc) regions, where Fab region is the part which binds to antigens (including the variable domain of each heavy and light chain) and presents a

set of complementary determining regions (CDR), also known as hypervariable regions, located between stable regions named frameworks (FR). In turn, the Fc region is the binding site for endogenous receptors (presented in macrophages, dendritic cells, and other cells of the immune system) and complement system proteins [6].

IgG are present as soluble proteins in proximal body fluids (serum, synovial fluid, and saliva, among others) or membrane-bound proteins attached to the B-cells (similar structure to IgM) for regulating the immune system and performing the function of B-cell receptors (BCR), respectively. Specifically, the disrupted activation of the BCR appears as the responsible of the chronic lymphocytic leukemia (CLL) [7, 8]. This disease, characterized by its high heterogeneity, is the most common human blood cancer in Western countries and it usually shows a monoclonal expansion of an aberrant B-cell clone [9, 10].

Determining the protein profiles of B-CLL cells could have a great impact on disease knowledge, progression, origin, and identification of new drug targets; particularly referring to immune-system proteins, antibodies and immunoglobulins. In this sense, high-throughput DNA sequencing has been the approach of choice for extracting the most Ig biological information [11]. In fact, Next Generation Sequencing (NGS) allows the characterization of millions of BCR sequences in a single experiment [12]. However, despite the remarkable advances made in the field, the peptide sequence information is required for the complete description of the Ig sequences as the cellular state is finally defined by the translated genes (i.e., proteins) – everything that is transcribed (genomics) may not be translated (proteomics). Henceforth, the integration of genomics and proteomics data sets for Ig sequencing might provide complementary information about evidence of gene expression at the protein level [13, 14, 15].

The determination of the length and peptide sequence of the H-chain CDR-H3 region (related to antibody specificities) as well as the IGHV and IGHJ gene patterns and the VH:VL pairing [16, 17] are paramount for antibody clonotyping. In fact, the alterations in these sequences could be specific and characteristic of the pathology highlighting the importance of their determination as previously reported by Henriques and colleagues [18]. Moreover, the evaluation of immune repertoires is coming up as a useful approach for the identification of antigen-specific BCR, and evaluation of the efficacy of immune checkpoint blockers and cancer immune therapy, among others [19, 20].

Also, these described approaches are becoming quite critical in the characterization of pharmaceutical IgGs, such as anti-CD20, trastuzumab, and cetuximab [21]. Related to this, the production of biosimilars (biologic medical products which are almost identical copies of original products), which are considered as good alternatives to improve healthcare access and outcomes at reduced costs, could be greatly improved if disease-related

antibody sequences are determined (e.g., monoclonal antibodies, cytokines...) [22, 23].

Generally, Ig-sequeing (soluble Ig) studies have been performed by using MS/MS approaches in serum samples [24], whereas BCR-sequeing (membrane-bound Ig) has been addressed from a genomic point of view using B-cells as a sample [25]. Thus, their combination might have a tremendous importance to understand the humoral antibody responses.

Peptide analysis of BCR sequences remains a challenge mainly due to the drawbacks in its processing and purification (e.g., low BCR levels). Thus, NGS approaches have been the usually selected strategies for the amplification of the BCR and its sequencing. However, a library including the mass spectra for BCR-related peptides may be of great interest for the full characterization of BCR production, Ig secretion, and immune responses. Since little variations in antibody sequences (even one single amino acid) could lead to modified specific antigen bindings; then, MS/MS strategies seem a good alternative for providing confidence in peptide sequencing [25].

To this end, we have evaluated three approaches for peptide sequencing of BCR as well as B-cell proteins associated to BCR (MHC-I, MHC-II, CD20, CD79b, among others) and also related to the immune system using an MS/MS approach. Moreover, an in-house algorithm has been designed for the comparison of sequences obtained by MS/MS and molecular sequencing (Figure 1) to demonstrate the potential of MS/MS approaches in B cells to characterize Ig sequences (426 unique peptide sequences have been identified in this study) and correlate them with DNA sequence libraries.

## RESULTS

### Sample processing for BCR profiling by MS/MS assays

As mentioned in the Introduction section, cellular Igs use to be characterized by DNA-sequencing strategies whereas soluble Igs (due to their high concentration in serum) are easier to be characterized by MS approaches. The key point is the isolation and enrichment of BCR (membrane-bound Ig) in order to increase the peptide coverage that facilitates the peptide sequencing of the BCR. With this purpose, three different approaches (termed Fab preparation kit (a), papain (b), and DTT (c) in Materials and Methods section) were tested to evaluate the utility and robustness of MS/MS assays for Ig peptide identification.

While the first approach (Fab purification kit) uses an available commercial kit for the isolation of Fab fragments from protein samples employing affinity columns, the second and third approaches (papain and DTT, respectively) include an Ig-enrichment step – based on protein G microbeads - previous to the addition of papain or DTT for the isolation of Fab fragments or the

breakage of disulfide bonds, respectively. With all these strategies, it is expected to purify the B-cell Igs in order to increase the peptide coverage and the subsequent BCR-peptide characterization.

In Supplementary Information 1, the proteins and the corresponding peptides for each approach are collected. As it has been depicted in Table 1, CLL samples have been randomly distributed among the 3 suggested approaches. Additionally, human IgG protein (whole molecule) was processed using the Fab preparation kit to use it as a reference.

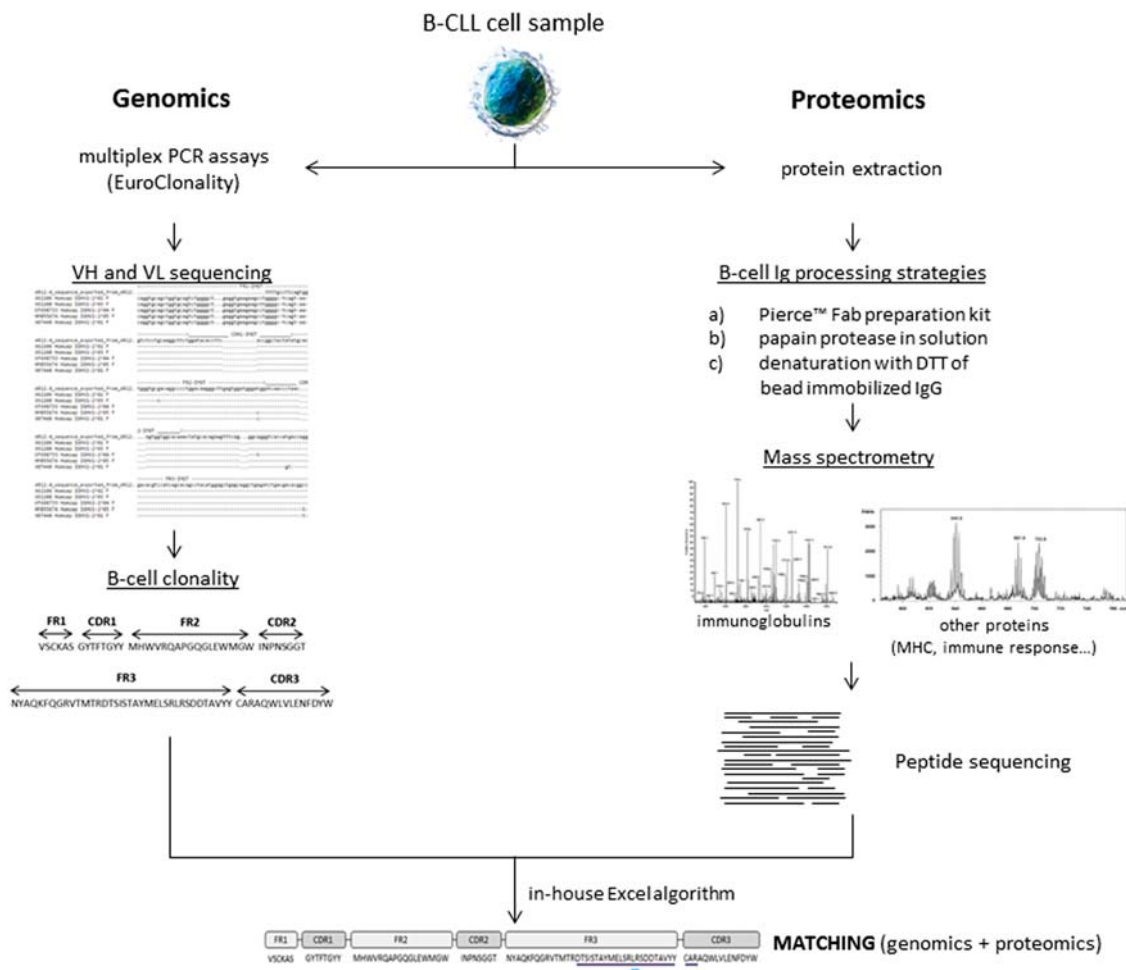
A total of 31 different peptides related to Ig were identified when processing the human IgG sample, corresponding to 20 different proteins. As expected, the approach concerning Fab preparation kit allowed the isolation of these peptides from a purified Ig sample. However, its application on a complex sample (purified B-CLL cells, sample 6) reported 22 different proteins (related to BCR and Ig) in comparison with the control sample. Nevertheless, when looking for the 20 human IgG proteins mentioned above in the samples processed with

papain and DTT, the overlapping was increased up to 40% (8/20) and 70% (14/20), respectively.

Considering as reference these results, the denaturation approach provides better results when dealing with high complex samples. Nonetheless, and due to this complexity, the purification and enrichment of the sample may become a key step during the processing.

Our results show that the number of identified proteins related to the immune system is increased as the processing strategy is simplified. Thus, 40, 54, and 171 immune system proteins were reported in CLL samples for Fab preparation, papain, and DTT approaches, respectively (Supplementary Information 1). Therefore, for MS/MS characterization of cellular Ig, two main considerations are needed: i) enrichment step for the isolation and purification of Ig-related proteins, and ii) simple post-processing for minimizing loss of sample.

Regarding the distribution of immune system related proteins reported by the three approaches (Figure 2), papain approach was the most efficient for the identification of specific Ig proteins (40.8%), followed



**Figure 1: General workflow of the study.**

**Table 1: Patient characteristics**

Sample ID	Age <sup>a</sup> (yr)	Gender	Diagnosis <sup>b</sup> % clonal population (from WBC) <sup>b</sup> V(D)J rearrangement <sup>c</sup>	MS <sup>c</sup> Rai stage	Light chain <sup>b</sup>	Molecular biology sequence	MS/MS processing approach <sup>d</sup> Gel section digested
1	83	F	CLL 88.1 VH4-34(D5-12)J4	UM-H III	kappa <sup>+dim</sup>	Avail.	DTT Whole lane
2	58	F	CLL 21.4 VH3-7(D2-15)	M-H 0	kappa <sup>+dim</sup>	N.A.	DTT Whole lane
3	62	F	CLL 22.1 VH4-39(D2-2)	UM-H I	kappa <sup>+dim</sup>	Avail.	DTT Whole lane
4	48	M	CLL 90.0 VH3-30(D2-15)J6	M-H I	kappa <sup>+dim</sup>	Avail.	DTT Whole lane
5	72	F	CLL 70.1 VH3-7(D2-2)	M-H III	kappa <sup>+dim</sup>	N.A.	DTT Whole lane
6	49	M	CLL 80.4 VH1-69(D3-10)J6	M-H IV	kappa <sup>+dim</sup>	N.A.	Fab preparation kit Gel band
7	65	M	CLL 87.5 VH1-69(D2-15)J3	UM-H N.A.	kappa <sup>+dim</sup>	N.A.	Papain Whole lane
8	70	M	CLL 64.9 VH5-a(D5-12)	M-H II-H	kappa <sup>+dim</sup>	N.A.	Papain Gel band
9	56	M	CLL 83.2 VH1-69(D2-2)J6	UM-H II	lambda <sup>+dim</sup>	N.A.	DTT Whole lane

Clinical characteristics of the samples used in this study, including the age, gender, diagnosis, V(D)J rearrangement, molecular biology features, and MS/MS processing approach applied to each sample.

<sup>a</sup> Age at time of sample collection. <sup>b</sup> From phenotypic studies. <sup>c</sup> From molecular biology studies. <sup>d</sup> DTT for denaturation of bead-immobilized immunoglobulins (Ig); *papain* for isolation of Fab fragments from bead-immobilized Ig; *Fab preparation kit* for isolation of Fab fragments using a commercial kit.

WBC, white blood cells. MS, mutational status (UM-H, unmutated; M-H, mutated). F, female. M, male. CLL, chronic lymphocytic leukemia. Avail., available. N.A., information not available.

by Fab preparation and DTT approaches. Furthermore, proteins belonging to the human leukocyte antigen (HLA) system were better detected by the DTT approach (20.5%) in comparison to papain and Fab preparation strategies (11.1 and 2.5%, respectively). The remaining protein groups (B-cell proteins, antigens, leukocyte/lymphocyte proteins, and others) presented a similar distribution.

### Characterization of Ig peptide sequences by MS/MS

The construction of peptide sequence libraries for soluble Igs and BCRs is still a challenge. As mentioned before, several improvements have been done in serum-Ig sequencing, but profiling cellular Ig without using DNA-

approaches is really difficult mainly due to the low amount of membrane-bound Ig present in the cell (compared to the rest of cell proteins).

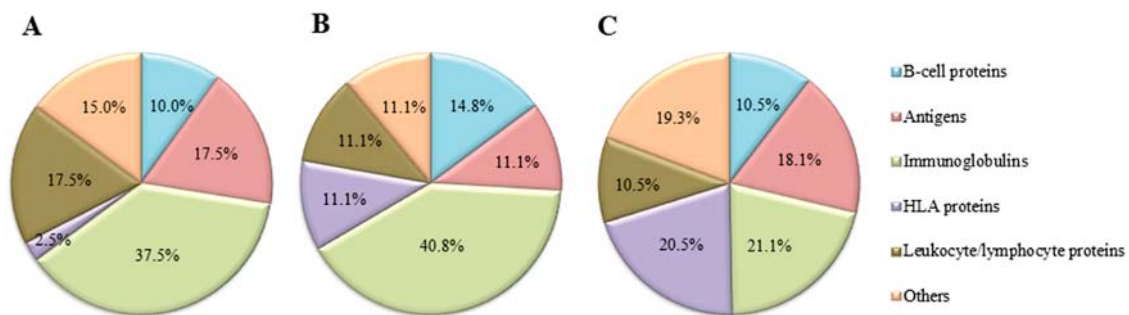
In our study, a total of 60, 98, and 426 different unique peptides were identified across the three approaches (Fab preparation, papain, and DTT, respectively) establishing the beginning of a CLL peptide sequence library exclusively obtained by MS/MS technique. Only 15 peptides out of the total identified were in common among the studied strategies, whereas the approach concerning denaturation of Ig enriched samples reported 362 unique peptides not identified by the other two techniques (data not shown). These results suggest the importance of perfectly well-defined workflow when processing complex samples for Ig sequencing.

As Figure 3 shows, the peptide sequence length is distributed from 6 to 23 amino acids in the samples. This distribution is more uniform for peptides obtained with the DTT approach. However, the Fab preparation kit has generated a low number of long unique peptides which could be a drawback for the identification of cellular Ig.

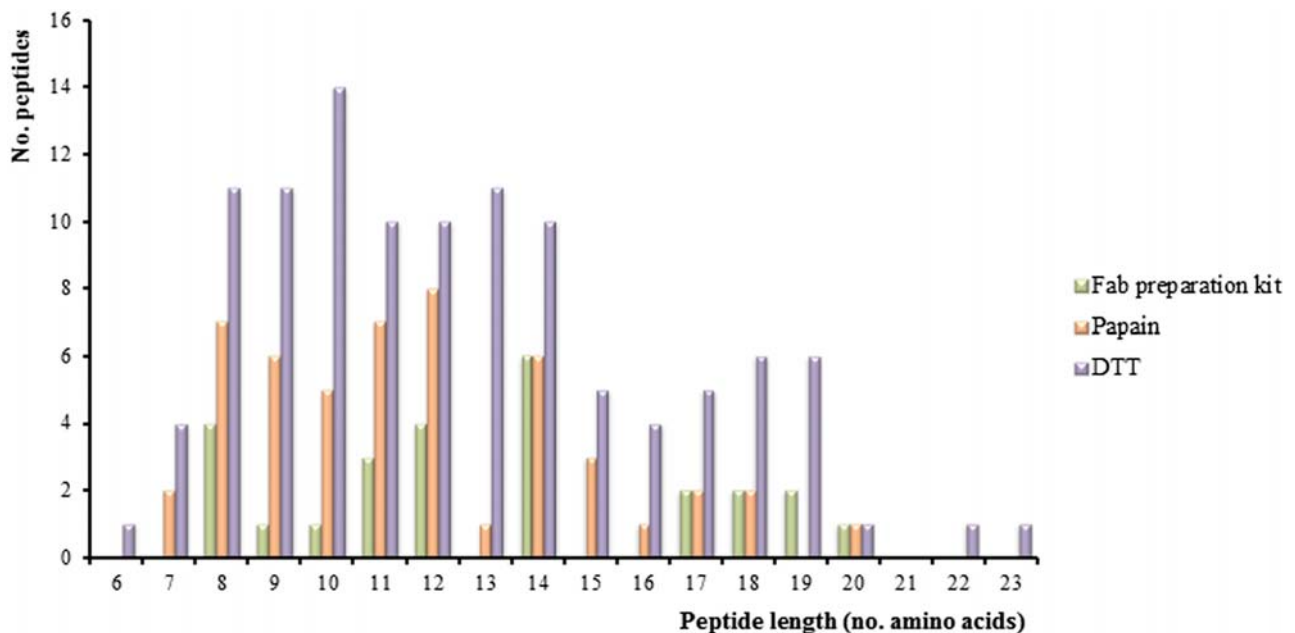
Regarding peptide sequence variability, the combination of different proteases (in addition to trypsin, which has been used in this study) might be an important improvement. In this way, peptide sequence libraries could increase in size providing a better Ig characterization.

### Matching genomic and proteomic data sets for BCR profiling at peptide level

Sequencing of membrane-bound Ig of clonal B-cell populations belonging to CLL patients is routinely done. This sequencing is performed by using molecular biology techniques and allows the stratification of the patients, being also crucial for determining the patient treatment. Despite multiplex PCR assays are performed to determine the B-cell clonality (following the guidelines of the EuroClonality European Consortium, <http://www.euroclonality.org/>); this approach presents a limitation because it is not possible to characterize the whole set of proteins related to the immune



**Figure 2: Distribution of proteins related to the immune system detected by shotgun MS/MS after processing the samples by three different approaches.** (A) *Fab preparation kit* for isolation of Fab fragments using a commercial kit; (B) *papain* for isolation of Fab fragments from bead-immobilized immunoglobulins (Ig); (C) *DTT* for denaturation of bead-immobilized Ig). The MS/MS analysis was performed using Peptide Shaker search engine (neXtProt database, 2016).



**Figure 3: Peptide length distribution.** Graphic showing the number of peptides corresponding to immunoglobulin (Ig) proteins detected by the three approaches (*Fab preparation kit* for the isolation of Fab fragments using a commercial kit, *papain* for isolation of Fab fragments from bead-immobilized Ig, and *DTT* for denaturation of bead-immobilized Ig) using Peptide Shaker as search engine and neXtProt (release February 2016) as database.

system in a single assay. In this regard, proteomics is able to achieve the Ig characterization at peptide level with the same processing speed and robustness as genomics, but with the gain of sequencing thousands of peptides at once (including Ig or proteins related to the immune response, among others). This is of great relevance given the importance of surrounding proteins that determine the final behavior of B-CLL cells.

In fact, both disciplines must be considered as complementary instead of competitive or encountered methodologies. With the purpose of integrating the results from genomics and proteomics, an in-house algorithm in Excel software (Supplementary Information 2) was designed. This tool allows the identification of sequences overlapping between the peptide and DNA-translated providing information about which specific Ig region (CDR or FR) presents the overlapping and the coverage percentage for each sample.

In this study, information about the Ig sequences from three samples (samples ID 1, 3, and 4) are reported and compared to our MS/MS peptide sequences (from the same samples) using the mentioned algorithm.

As Ig sequences are characterized by their high variability, the common databases used for MS/MS searching do not include all of them (hence specific databases only focused on Igs are dramatically needed). Thus, two different search engines (i.e., Mascot and PeptideShaker) were selected for evaluating their capability on detecting non-assigned peptides (i.e., amino acid sequences belonging to highly variable proteins). On one side, PeptideShaker only presents as result sequences matching with annotated unique peptides and their corresponding proteins; whereas the Mascot search engine reports well-defined peptides although they do not correspond to annotated proteins. This comparison will highlight the relevance of the appropriate search engine

(independently of the sample preparation) according to the sample features.

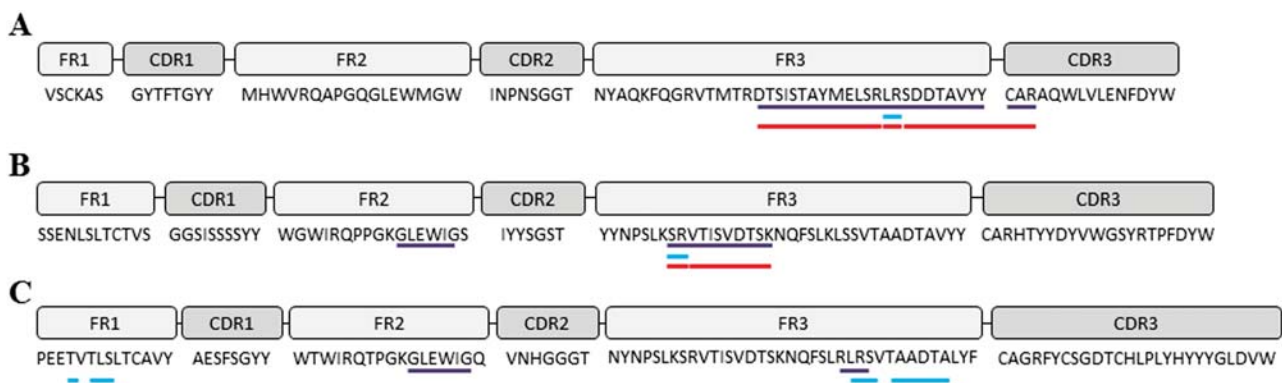
The peptides identified by both search engines are collected in Supplementary Information 3 for the three samples mentioned above (“List of peptides” sheet). For Mascot results, “annotated” and “unknown” proteins are displayed in different columns. These peptide sequences were compared to the corresponding molecular biology sequence by using the in-house Excel algorithm. The results of these comparisons are also depicted in Supplementary Information 3.

Figure 4 summarizes the overlapping between both genomics and proteomics and both Mascot and PeptideShaker. As indicated in this Figure 4, FR regions present the greater overlap between genomics and proteomics. Regarding the CDR3 region, 3 amino acids (CAR) have been properly detected (high ion score) within an 11-amino acids length peptide constituting a 20% coverage of the mentioned region.

Regarding the search engine employed, Mascot seems to be more efficient than PeptideShaker. In Table 2, it is reported the coverage percentage for each sample and Ig region. Although PeptideShaker coverage for sample 4 is good enough, Mascot is more homogeneous along the three samples, possibly due to the fact previously mentioned.

## DISCUSSION

The high diversity present in gene sequences requires tailor-made DNA sequence databases, particularly in encoding genes for Ig and histocompatibility complexes. Next-generation sequencing approaches (genomics) have allowed a significant improvement in this area (BCR- and Ig-sequencing), but the progress in proteomics for the



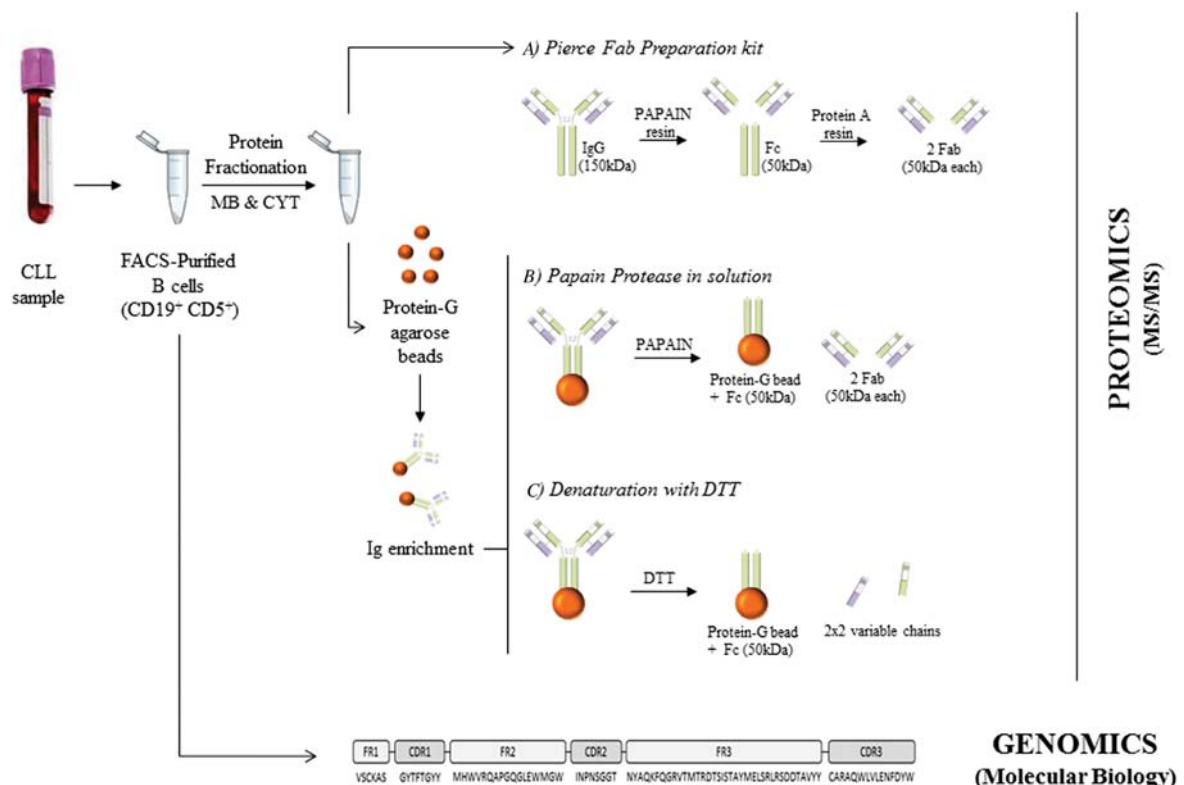
**Figure 4: IGH gene sequences.** The framework (FR) and complementary determining regions (CDR) of IGH genes were analyzed and translated into amino acids after DNA-sequencing using the IMGT tool. (A) sample ID 1; (B) sample ID 3; (C) sample ID 4. Violet lines depict the sequence coverage by both DNA-sequencing and MS/MS sequencing using Mascot search engine (UniProt database). Blue lines correspond to sequence coverage by both DNA-sequencing and MS/MS sequencing using Peptide Shaker search engine (neXtProt database). Red lines correspond to predicted tryptic peptides determined by PeptideCutter tool (ExPASy Bioinformatics Resource Portal).



**Table 2: Sequence coverage of IGH genes after DNA-sequencing and MS/MS sequencing and analysis by Mascot and Peptide Shaker search engines**

Sample ID <sup>a</sup>	Search engine <sup>b</sup>	Region <sup>c</sup>	Region coverage <sup>d</sup> (%)	Total coverage <sup>e</sup> (%)
1	Mascot	FR3	23/37 (62.1%)	26/91 (28.6%)
		CDR3	3/15 (20.0%)	
3	Peptide Shaker	FR3	2/37 (5.4%)	2/91 (2.2%)
	Mascot	FR2	6/17 (35.3%)	
	Mascot	FR3	11/37 (29.7%)	17/104 (16.3%)
	Peptide Shaker	FR3	2/37 (5.4%)	
4	Mascot	FR2	6/17 (35.3%)	9/109 (8.3%)
		FR3	3/37 (8.1%)	
	Peptide Shaker	FR1	4/14 (28.6%)	13/109 (11.9%)
		FR3	9/37 (24.3%)	

<sup>a</sup> Sample ID information is available at Table 1. <sup>b</sup> Mascot search engine was used with UniProt database, whereas Peptide Shaker was tested with the neXtProt database. <sup>c</sup> It is referred to the framework (FR) or complementary determining region (CDR) of the IGH gene sequence. <sup>d</sup> Region coverage was calculated considering the number of amino acids in common between DNA-sequencing and MS/MS sequencing in relation to the region (FR or CDR) length. <sup>e</sup> Total coverage was calculated considering the total number of amino acids in common between DNA-sequencing and MS/MS sequencing of the whole IGH gene in relation to the IGH gene length.



**Figure 5: Overview of the procedures performed in this study.** B cells were purified from CLL peripheral blood using FACSaria. These purified cells were directly employed for BCR-sequencing by molecular biology and also processed to obtain membrane (MB) and cytoplasmic (CYT) proteins. These isolated proteins were used for Fab purification using Pierce Fab preparation kit (A). Moreover, they were Ig-enriched for applying papain protease in solution (B) and denaturation with DTT (C) for the BCR-sequencing by LC-MS/MS.

BCR-sequencing is still in their beginnings. Whereas the sequencing of soluble Ig is perfectly addressed by MS/MS strategies, the peptide profiling of membrane-bound Ig is still a great challenge.

In this study, we have characterized Ig attached to the B-CLL cells surface at peptide level by performing three different sample preparation approaches (i.e., usage of a Fab preparation commercial kit, and usage of papain in solution and DTT in Ig-enriched agarose beads from CLL protein samples for the isolation of Fab and denaturation of the Ig, respectively). Our results have indicated that the strategy including denaturation with DTT is the most efficient approach. Bearing in mind these results, several guidelines could be described about peptide sequencing of membrane-bound Ig: i) including an Ig-enrichment step; ii) reducing the complexity of the procedure (with regard to the number of steps) to avoid the loss of the sample (membrane-bound Ig are less abundant than soluble ones); iii) using a suitable MS/MS search engine for the identification of peptides corresponding to uncharacterized or unassigned proteins; and iv) integrating MS/MS and molecular biology results to get a wide and comprehensive view of Ig and associated immune system proteins expressed in response to a specific disease or particular cellular condition.

In summary, all these results indicate that the integration of genomics and proteomics could provide a full characterization of Ig sequences, BCR and the proteins related to the immune system in response to a specific disease or particular cell situation. Nonetheless, further assays will be required to identify specific patterns in BCR sequences which could be of interest in pathological situations such as CLL, autoimmunity or immunodeficiency, among others.

## MATERIALS AND METHODS

Figure 5 depicts a general overview of the procedure described below.

### Sample collection

Peripheral blood (PB) B-lymphocytes (CD19<sup>+</sup> CD5<sup>+</sup>) from 9 CLL patients (Table 1) were systematically purified by FACSARIA II flow cytometer (BD, San Jose/CA, USA) (at a purity >95%) at General Service of Cytometry of University of Salamanca [26]. High-sensitive multi-parameter flow cytometry was performed for the immunophenotyping of clonal B-cell populations, according to previously described procedures [18]. All cases presented a clonal (8 kappa CLL clone, 1 lambda CLL clone) and aberrant CD5<sup>+</sup> CLL-like B-cell population.

PB samples were obtained with informed consent and according to the guidelines of the local ethics committees of the University Hospital of Salamanca, in accordance with the Helsinki Declaration of 1975, as revised in 2008.

### Molecular studies: sequencing of immunoglobulin human domains

Each clonal B-cell population – purified by FACSARIA – was screened for its patterns of rearrangement of the IGHV genes (Table 1) following genomic DNA preparation, PCR amplification, sequencing, and analysis of V, (D), J genes, as described elsewhere [27, 28]. Specifically, the B-cell clonality was determined following the guidelines described by the EuroClonality Consortium, <http://www.euroclonality.org/>, which combine the usage of IGH and IGK multiplex PCR assays for the assessment of Ig gene rearrangements [29, 30]. The sequences of the V(D)J fragments of these samples have been widely reported in a previous study performed by our group [31]. Resulting sequences were aligned with germline sequences using the IMGT database and tools (<http://www.imgt.org>). Sequences were considered as somatically mutated if they contained 2% deviation from the germline sequence.

### Subcellular protein fractionation

FACS-purified B cells were snap-frozen in liquid nitrogen after adding lysis solution (5 mM HEPES, 10 mM MgCl<sub>2</sub>, 140 mM NaCl, 0.1% Tween 20, and 1% protease inhibitor mix. All purchased from Sigma, St. Louis/MO, USA). Then, cells were thawed on ice and centrifuged at 15,000 g for 15 min; henceforth, supernatant and pellet were treated separately, although processing stepwise in an identical manner. 10 % (w/v) octyl-β-D-glucopyranoside (Sigma) was added followed by sonication (3 pulses for 3 seconds each). Samples were incubated for 30 min and centrifuged at 15,000 g for 15 min. Cytoplasmic proteins were collected from the supernatant, whereas membrane proteins were collected from the pellet.

### Immunoglobulin enrichment from cell extract

Samples were IgG-enriched by using protein-G-agarose beads (GE Healthcare, Buckinghamshire, UK), according to manufacturer's instructions, and stored at -80°C until used.

### B-Cells immunoglobulin processing strategies

Three different approaches were performed to achieve the peptide sequencing of B-cell Igs: a) Pierce™ Fab preparation kit (Thermo Scientific, Waltham/MA, USA), based on the usage of immobilized papain protease for IgG digestion and protein A agarose for purification; b) papain protease in solution (Worthington, Biochemical Corporation, Lakewood/NJ, USA), for the digestion of bead immobilized IgG and the obtaining of Fab fragments [32]; c) denaturation with DTT of bead immobilized IgG for the breakage of their disulfide bonds and the obtaining of heavy and light chains.

Approach a) was performed following the manufacturer's indications. For approach b), papain was activated with a digestion buffer (5 mM L-Cys-HCl, 1 mM EDTA and PBS; all from Sigma) at pH=7 for 30 min at 37°C. The activated papain was incubated with the sample at 1:50 ratio (papain:sample) for 2 h at 37°C with gently mixing. Afterward, samples were centrifuged at 12,000 g for 3 min and the supernatant was collected. The supernatant contained the Fab released from the IgG, whereas the pellet contained the beads and Fc remaining fractions. Papain digestion was stopped with SDS sample buffer. The second step of papain digestion was performed with the pellet at the same conditions indicated above and joined to the one obtained in the first digestion. Finally, approach c) was performed incubating the samples with 40 mM DTT for 5 min at 99°C.

In all these performed approaches (a, b, and c), human IgG protein (Abcam, Cambridge/UK) was used as a reference.

### SDS-PAGE separation

Each processed sample was separated on a 10% SDS-PAGE gel under non-reducing conditions for approaches a) and b), and reducing conditions for approach c). After electrophoresis, gels were stained in a solution of 2.5 g/L silver nitrate and stored at 4 °C in an aqueous solution containing 1% (v/v) acetic acid, until analysis (Supplementary Figure 1).

### nUPLC-MS/MS analysis

Each gel lane was cut into fragments and digested with trypsin following the method of Shevchenko *et al.* [33] with slight modifications. Briefly, gel pieces were destained with 15 mM potassium ferrocyanide and 50 mM sodium thiosulfate. Protein reduction and alkylation were performed with 10 mM DTT at 56 °C for 45 min, and with 55 mM IAA at room temperature for 30 min, respectively. Proteins were digested with trypsin (6.25 ng/mL) at 37 °C for 18 h. The peptide solution was acidified with FA and desalted by using C18-Stage-Tips columns [34]. The samples were partially dried and stored at -20 °C until they were analyzed by LC-MS/MS.

A nUPLC system (nanoAcquity, Waters Corp., Milford/MA, USA) coupled to a LTQ-Velos-Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose/CA, USA) via a nanoelectrospray ion source (NanoSpray flex, Proxeon, Thermo) was used for reversed-phase LC-MS/MS analysis. Peptides were dissolved in 0.5% FA/3% ACN and loaded onto a trapping column (nanoACQUITY UPLC 2G-V/M Trap Symmetry 5 µm particle size, 180 µm × 20 mm C18 column, Waters Corp., Milford/MA, USA). Peptides were separated on a nanoACQUITY UPLC BEH 1.7 µm, 130 Å, 75 µm × 250 mm C18 column (Waters Corp., Milford/MA, USA) with a linear gradient from 7% to 35% solvent B (ACN/0.1% FA) at a flow rate of 250 nL/min over 120 min.

The nUPLC- LTQ-Orbitrap Velos was operated in the positive ion mode by applying a data-dependent automatic switch between survey MS scan and tandem mass spectra (MS/MS) acquisition. Survey scans were acquired in the mass range of  $m/z$  400 to 1600 with a 60 000 resolution at  $m/z$  400 with lock mass option enabled for the 445.120025 ion [35].

The 20 most intense peaks having  $\geq 2$  charge state and above the 500 intensity threshold were selected in the ion trap for fragmentation by collision-induced dissociation with 35% normalized energy, 10 ms activation time,  $q = 0.25$ ,  $\pm 2$   $m/z$  precursor isolation width and wideband activation. Maximum injection time was 1000 and 50 ms for survey and MS/MS scans, respectively. AGC was  $1 \times 10^6$  for MS and  $5 \times 10^3$  for MS/MS scans. Dynamic exclusion was enabled for 90 s.

### Database search

Raw data were translated to mascot general file (.mgf) format using Thermo Scientific Proteome Discoverer software (v. 1.4.1.14). The searches were conducted using both Mascot and Peptide Shaker search engines. For the first approach, the MASCOT [36] algorithm was used to search for the acquired MS/MS spectra, using Thermo Scientific Proteome Discoverer software (v. 1.4.1.14) against a custom database of all human-reviewed sequences downloaded from the UniProt database (February 2014) and common contaminant sequences (e.g., human keratins, trypsin, BSA). Search parameters were as follows: fully-tryptic digestion with up to two missed cleavages; 10 ppm, and 0.8 Da mass tolerances for precursor and product ions, respectively; carbamidomethylation of cysteines as fixed modification and oxidation of methionine and n-terminus acetylation as variable modifications were considered. Peptides having MASCOT ion scores of  $< 20$  were not considered for analysis. A 1% false discovery rate (FDR) using Percolator [37] was employed for peptide validation. For Peptide Shaker approach, the search was conducted using SearchGUI version 1.30.1 [38]. Protein identification was conducted against a concatenated target/decoy version of all human-reviewed sequences downloaded from the neXtProt database (release of September 2015, 20066 protein entries) and common contaminant sequences (e.g., human keratins, trypsin). The decoy sequences were created by reversing the target sequences in SearchGUI. The identification settings were the same fixed for the MASCOT searching. Peptides and proteins were inferred from the spectrum identification results using PeptideShaker version 0.41.1 [39]. Peptide-Spectrum Matches (PSMs), peptides, and proteins were validated at a 1% FDR estimated using the decoy hit distribution.

The MS data along with the identification results have been deposited to the ProteomeXchange Consortium via the PRIDE [40] partner repository with the dataset identifier PXD004466.

## Peptide mass spectra and DNA sequences interpretation

An in-house Excel algorithm was designed for the integration of results obtained after DNA sequencing and MS/MS analysis. The basis of this Excel macro was the searching of homologies between the translated DNA sequences and the peptides identified by nUPLC-MS/MS. This tool looks for homology of 2 or more consecutive amino acids corresponding to the CDR or FR regions. For reviewing purposes, the Excel algorithm is provided as Supplementary Information 2.

### Abbreviations

BCR, B-cell receptor; CDR, complementary determining region; CLL, chronic lymphocytic leukemia; Fab, fragment antigen binding; FR, framework; HLA, human leukocyte antigen; Ig, immunoglobulin; MS, mass spectrometry; NGS, Next-Generation Sequencing; nUPLC-MS/MS, nano-ultra-performance liquid chromatography tandem MS; PB, peripheral blood; PSM, peptide-spectrum match.

### ACKNOWLEDGMENTS AND FUNDING

We gratefully acknowledge financial support from the Spanish Health Institute Carlos III (ISCIII) for the grants: FIS PI11/02114 and FIS PI114/01538. We also acknowledge Fondos FEDER (EU) and Junta Castilla-León (grant BIO/SA07/15). This work has been also sponsored by Fundación Solórzano (FS/23-2015). The Proteomics Unit belongs to ProteoRed, PRB2-ISCIII, supported by grant PT13/0001, of the PE I+D+I 2013-2016, funded by ISCIII and FEDER. The authors would like to thank all the clinicians and technicians in the Cytometry and Cell Purification Services of the University of Salamanca, the Spanish National DNA Bank (Banco Nacional de DNA Carlos III, University of Salamanca) and the Genomic Unit of Cancer Research Centre (IBMCC, USAL-CSIC) for their support in the data collection for the preparation of this manuscript. P.D. is supported by a JCYL-EDU/346/2013 Ph.D. scholarship.

### CONFLICTS OF INTEREST

The authors declare no competing financial interest.

### REFERENCES

- Schmidt S, Rainer J, Riml S, Ploner C, Jesacher S, Achmüller C, Presul E, Skvortsov S, Crazzolara R, Fiegl M, Raivio T, Jänne OA, Geley S, et al. Utilization of Ig heavy chain variable, diversity, and joining gene segments in children with B-lineage acute lymphoblastic leukemia: implications for the mechanisms of VDJ recombination and for pathogenesis. *Blood*. 2006; 91:4602-4609.
- Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D, Scharff MD. The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes and Development*. 2004; 18:1-11.
- Tonegawa S. Somatic generation of antibody diversity. *Nature*. 1983; 302:575-581.
- Odegard VH, Schatz DG. Targeting of somatic hypermutation. *Nat Rev Immunol*. 2006; 6:573-583.
- Jordan M, Baxter AG. Quantitative and qualitative approaches to GOD: the first 10 years of the clonal selection theory. *Immunol Cell Biol*. 2008; 86:72-79.
- Schroeder HW Jr, Cavacini L. Structure and function of immunoglobulins. *J Allergy Clin Immunol*. 2010; 125:S41-52.
- Burger JA, Chiorazzi N. B cell receptor signaling in chronic lymphocytic leukemia. *Trends in Immunology*. 2013; 34:592-601.
- Díez P, Góngora R, Orfao A, Fuentes M. Functional proteomics insights in B-cell chronic lymphocytic leukemia. *Expert Review of Proteomics*. 2016; 14:137-146.
- Chiorazzi N, Rai KR, Ferrarini M. Chronic lymphocytic leukemia. *N Engl J Med*. 2005; 352:804-815.
- Lopez C, Delgado J, Costa D, Villamor N, Navarro A, Cazorla M, Gomez C, Arias A, Munoz C, Cabezas S, Baumann T, Rozman M, Aymerich M, et al. Clonal evolution in chronic lymphocytic leukemia: analysis of correlations with IGHV mutational status, NOTCH1 mutations and clinical significance. *Genes Chromosomes Cancer*. 2013; 52:920-927.
- Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol*. 2014; 32:158-168.
- Brown JR, Fernandes SM, Kasar S, Hoang K, Moorhead M, Carlton V, Faham M, Miklos DB, Logan AC. Next-Generation Sequencing Reveals Clonal Evolution at the Immunoglobulin Loci in Chronic Lymphocytic Leukemia. *Blood*. 2014; 124:3302-3302.
- Woo S, Cha SW, Merrihew G, He Y, Castellana N, Guest C, Maccoss M, Bafna V. Proteogenomic database construction driven from large scale RNA-seq data. *J Proteome Res*. 2014; 13:21-28.
- Castellana N, Bafna V. Proteogenomics to discover the full coding content of genomes: A computational perspective. *Journal of Proteomics*. 2010; 73:2124-35.
- Díez P, Fuentes M. Proteogenomics for the comprehensive analysis of human cellular and serum antibody repertoires. *Proteogenomics, Advances in Experimental Medicine and Biology*. 2016; 926:153-162.
- Lavinder JJ, Horton AP, Georgiou G, Ippolito GC. Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires. *Current Opinion in Chemical Biology*. 2015; 24:112-120.

17. Xu JL, Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity*. 2000; 13:37-45.
18. Henriques A, Rodriguez-Caballero A, Criado I, Langerak AW, Nieto WG, Lecrevisse Q, Gonzalez M, Cortesao E, Paiva A, Almeida J, Orfao A. Molecular and cytogenetic characterization of expanded B-cell clones from multiclonal versus monoclonal B-cell chronic lymphoproliferative disorders. *Haematologica*. 2014; 99:897-907.
19. Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol*. 2014; 32:158-168.
20. Robins H. Immunosequencing: applications of immune repertoire deep sequencing. *Curr Opin Immunol*. 2013; 25:646-652.
21. Beck A, Sanglier-Cianféroni S, Van Dorsselaer A. Biosimilar, biobetter, and next generation antibody characterization by mass spectrometry. *Anal Chem*. 2012; 84:4637-4646.
22. Dorvignit D, Palacios JL, Merino M, Hernández T, Sosa K, Casacó A, López-Requena A, Mateo de Acosta C. Expression and biological characterization of an anti-CD20 biosimilar candidate antibody. *MAbs*. 2012; 4:488-496.
23. Wooding KM, Peng W, Mechref Y. Characterization of pharmaceutical IgG and biosimilars using miniaturized platforms and LC-MS/MS. *Curr Pharm Biotechnol*. 2016; 17:788-801.
24. Boutz DR, Horton AP, Wine Y, Lavinder JJ, Georgiou G, Marcotte EM. Proteomic identification of monoclonal antibodies from serum. *Anal Chem*. 2014; 86:4758-4766.
25. Cheung WC, Beausoleil SA, Zhang X, Sato S, Schieferl SM, Wieler JS, Beaudet JG, Ramenani RK, Popova L, Comb MJ, Rush J, Polakiewicz RD. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nat Biotechnol*. 2012; 1080:1-8.
26. Nieto WG, Almeida J, Romero A, Teodosio C, Lopez A, Henriques AF, Sanchez ML, Jara-Acevedo M, Rasillo A, Gonzalez M, Fernandez-Navarro P, Vega T, Orfao A, et al. Increased frequency (12%) of circulating chronic lymphocytic leukemia-like B-cell clones in healthy subjects using a highly sensitive multicolor flow cytometry approach. *Blood*. 2009; 114:33-37.
27. González D, Balanzategui A, García-Sanz R, Gutiérrez N, Seabra C, van Dongen JJM, González M, San Miguel JF. Incomplete DJH rearrangements of the IgH gene are frequent in multiple myeloma patients: immunobiological characteristics and clinical implications. *Leukemia*. 2003; 17:1398-1403.
28. van Dongen JJ, Langerak AW, Brüggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuurin E, Garcia-Sanz R, van Krieken JH, Droese J, Gonzalez D, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia*. 2003; 17:2257-2317.
29. Evans PA, Pott Ch, Groenen PJ, Salles G, Davi F, Berger F, Garcia JF, van Krieken JH, Pals S, Kluin P, Schuurin E, Spaargaren M, Boone E, et al. Significantly improved PCR-based clonality testing in B-cell malignancies by use of multiple immunoglobulin gene targets. Report of the BIOMED-2 Concerted Action BHM4-CT98-3936. *Leukemia*. 2007; 21:207-214.
30. Wren D, Walker BA, Brüggemann M, Catherwood MA, Pott C, Stamatopoulos K, Langerak AW, Gonzalez D. Comprehensive translocation and clonality detection in lymphoproliferative disorders by next-generation sequencing. *Haematologica*. 2017; 102:e57-e60.
31. Henriques A, Rodríguez-Caballero A, Nieto WG, Langerak AW, Criado I, Lécresse Q, González M, Pais ML, Paiva A, Almeida J, Orfao A. Combined Patterns of IGHV Repertoire and Cytogenetic/Molecular Alterations in Monoclonal B Lymphocytosis versus Chronic Lymphocytic Leukemia. *PLoS One*. 2013; 8:e67751.
32. Andrew SM, Titus JA. Fragmentation of immunoglobulin G. *Curr Protoc Immunol*. 2003; chapter 16:unit 16.4.
33. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc*. 2006; 1:2856-2560.
34. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc*. 2007; 2:1896-1906.
35. Olsen JV, de Godoy LM, Li G, Macek B, Mortensen P, Pesch R, Makarov A, Lange O, Horning S, Mann M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics*. 2005; 4:2010-2021.
36. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999; 20:3551-3567.
37. Brosch M, Yu L, Hubbard T, Choudhary J. Accurate and sensitive peptide identification with mascot percolator. *J Proteome Res*. 2009; 8:3176-3181.
38. Vaudel M, Barsnes H, Berven FS, Sickmann A, Martens L. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics*. 2011; 11:996-999.
39. Vaudel M, Burkhardt JM, Zahedi RP, Oveland E, Berven FS, Sickmann A, Martens L, Barsnes H. PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat Biotechnol*. 2015; 33:22-24.
40. Vizcaíno JA, Csordas A, Del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu QW, Wang R, Hermjakob H. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res*. 2016; 44:D447-456.



**Integración de datos masivos proteómicos y  
transcriptómicos para el análisis de una línea  
celular de linfoma B en el contexto del Proyecto  
Proteoma Humano (C-HPP)**

**4**

Paula Díez<sup>1,2</sup>, Conrad Droste<sup>3</sup>, Rosa M. Dégano<sup>2</sup>,  
María González-Muñoz<sup>1</sup>, Nieves Ibarrola<sup>2</sup>, Martín Pérez-Andrés<sup>1</sup>,  
Alba Garin-Muga<sup>4</sup>, Víctor Segura<sup>4</sup>, Gyorgy Marko-Varga<sup>5</sup>,  
Joshua LaBaer<sup>6</sup>, Alberto Orfao<sup>1</sup>, Fernando J. Corrales<sup>4</sup>,  
Javier De Las Rivas<sup>3</sup>, Manuel Fuentes<sup>1,2</sup>

<sup>1</sup> Servicio General de Citometría-NUCLEUS; Centro de Investigación del  
Cáncer (IBMCC-CSIC/USAL) e IBSAL, Universidad de Salamanca.

<sup>2</sup> Unidad de Proteómica; Centro de Investigación del Cáncer (IBMCC-  
CSIC/USAL) e IBSAL, Universidad de Salamanca.

<sup>3</sup> Grupo de Bioinformática y Genómica Funcional; Centro de Investigación  
del Cáncer (IBMCC-CSIC/USAL) e IBSAL, Universidad de Salamanca.

<sup>4</sup> Unidad de Proteómica y Bioinformática; Centro de Investigación Médica  
Aplicada (CIMA), Universidad de Navarra.

<sup>5</sup> Clinical Protein Science and Imaging, Department of Biomedical  
Engineering; Lund University.

<sup>6</sup> Biodesign Institute; Arizona State University.

*Journal of Proteome Research*. 2015 Sep; 14(9):3530-3540.  
doi: 10.1021/acs.jproteome.5b00474





**Introducción.** La caracterización exhaustiva del conjunto molecular activo de las células puede ser abordado mediante la integración de dos aproximaciones distintas pero complementarias: la transcriptómica y la proteómica. Tras la era genómica, la proteómica ha surgido como una potente herramienta para identificar y caracterizar simultáneamente el conjunto de proteínas activas en un momento/situación particular de una célula. En este sentido, el Proyecto Proteoma Humano – concretamente, la estrategia basada en cromosomas, C-HPP – promueve la caracterización completa y sistemática del proteoma humano mediante el empleo de estrategias proteómicas de alta resolución, en combinación con datos derivados del análisis de expresión de genes codificantes de proteínas.

**Objetivo.** En este apartado del trabajo nos propusimos diseñar, desarrollar e implementar un método de trabajo para la integración de datos transcriptómicos y proteómicos derivados de muestras biológicas concretas, y evaluar de forma comparativa, las principales bases de datos proteicas disponibles (UniProt y neXtProt). Asimismo, se valoraron dos procedimientos de subfraccionamiento celular para la extracción y enriquecimiento de proteínas.

**Muestras, materiales y métodos.** Como modelo de estudio seleccionamos la línea celular *Ramos*, una línea celular de linfocitos B derivados de un paciente con linfoma de Burkitt. Dichas células fueron cultivadas, para su posterior procesamiento proteómico (en triplicado), mediante dos métodos de subfraccionamiento celular distintos basados en el uso de detergentes y choque osmótico mediante distintas concentraciones salinas para la purificación secuencial de proteínas de citoplasma, membrana, núcleo y orgánulos. Tras la cuantificación proteica y su separación mediante electroforesis unidimensional, cada muestra fue preparada para su digestión triptica. Los péptidos resultantes fueron analizados mediante LC-MS/MS empleando un equipo LTQ-Orbitrap-Velos. Los datos obtenidos fueron analizados mediante los algoritmos MASCOT y SEQUEST y las bases de datos UniProt (versión de febrero de 2014)

y neXtProt (versión de septiembre de 2014) con un índice de falsos positivos <1% a nivel de PSM, péptido y proteína. Los datos transcriptómicos de la línea celular *Ramos* fueron obtenidos del repositorio GEO (del inglés *Gene Expression Omnibus*) y se correspondían con análisis realizados con los microarrays de Affymetrix, Human Gene ST 1.0. La integración de los datos transcriptómicos y proteómicos se efectuó empleando el código R y permitió la generación de tres conjuntos de datos (intersección, unión y máximo) en función del grado de confianza (número de péptidos/proteína y FDR) y el nivel de cobertura de las proteínas/péptidos analizados.

**Resultados.** La integración de los datos proteómicos y transcriptómicos reveló un 94% de solapamiento entre las proteínas y los transcritos identificados en términos cualitativos por ambas aproximaciones "ómicas". En total, se identificaron ~6.000 proteínas y ~30.000 péptidos mediante la aproximación proteómica empleada (incluyendo la identificación de hasta 315 proteínas *missing*, PE2-4). En este sentido, cabe señalar que con ello identificamos alrededor del 30% del proteoma humano (con un promedio relativamente estable del 30% a lo largo de cada uno de los cromosomas humanos) poniendo de manifiesto el tamaño del proteoma promedio presente en un único tipo celular humano concreto. Asimismo, se ha constatado que el método de subfraccionamiento celular más eficiente se corresponde con aquel donde las proteínas de membrana se purifican mediante un detergente no iónico y no desnaturizante (IGEPAL) y las proteínas de núcleo se aíslan con n-dodecil- $\beta$ -D-maltopiranosido. Finalmente, se alcanzó un mayor número de identificaciones cuando la búsqueda se realizó con la base de datos neXtProt (versión de septiembre de 2014) versus Uniprot (versión de febrero de 2014).

**Conclusiones.** En resumen, en este trabajo mostramos cómo la integración sistemática de los datos transcriptómicos y proteómicos proporcionó una elevada cobertura del proteoma expresado por las células B de linfoma de Burkitt *Ramos* evidenciando

además la utilidad de combinar tecnologías -ómicas para la caracterización exhaustiva de sistemas biológicos complejos en situaciones fisiológicas concretas.

\* Los resultados proteómicos fueron depositados en ProteomeXchange (PXD001933).



# Integration of Proteomics and Transcriptomics Data Sets for the Analysis of a Lymphoma B-Cell Line in the Context of the Chromosome-Centric Human Proteome Project

Paula Díez,<sup>†,‡</sup> Conrad Droste,<sup>§</sup> Rosa M. Dégano,<sup>‡</sup> María González-Muñoz,<sup>†</sup> Nieves Ibarrola,<sup>‡</sup> Martín Pérez-Andrés,<sup>†</sup> Alba Garin-Muga,<sup>||</sup> Víctor Segura,<sup>||</sup> Gyorgy Marko-Varga,<sup>⊥</sup> Joshua LaBaer,<sup>#</sup> Alberto Orfao,<sup>†</sup> Fernando J. Corrales,<sup>||</sup> Javier De Las Rivas,<sup>\*,§</sup> and Manuel Fuentes<sup>\*,†,‡</sup>

<sup>†</sup>Department of Medicine and General Cytometry Service-Nucleus, Cancer Research Centre (IBMCC/CSIC/USAL/IBSAL), 37007 Salamanca, Spain

<sup>‡</sup>Proteomics Unit, Cancer Research Centre (IBMCC/CSIC/USAL/IBSAL), 37007 Salamanca, Spain

<sup>§</sup>Bioinformatics and Functional Genomics Research Group, Cancer Research Centre (IBMCC/CSIC/USAL/IBSAL), 37007 Salamanca, Spain

<sup>||</sup>Division of Hepatology and Gene Therapy, Proteomics and Bioinformatics Unit, Centre for Applied Medical Research (CIMA), University of Navarra, 31008 Pamplona, Spain

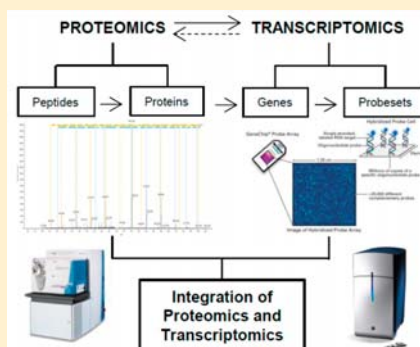
<sup>⊥</sup>Clinical Protein Science and Imaging, Biomedical Centre, Department of Biomedical Engineering, Lund University, BMC D13, 221 84 Lund, Sweden

<sup>#</sup>Biodesign Institute, Arizona State University, 1001 South McAllister Avenue, Tempe, Arizona 85287, United States

## S Supporting Information

**ABSTRACT:** A comprehensive study of the molecular active landscape of human cells can be undertaken to integrate two different but complementary perspectives: transcriptomics, and proteomics. After the genome era, proteomics has emerged as a powerful tool to simultaneously identify and characterize the compendium of thousands of different proteins active in a cell. Thus, the Chromosome-centric Human Proteome Project (C-HPP) is promoting a full characterization of the human proteome combining high-throughput proteomics with the data derived from genome-wide expression profiling of protein-coding genes. Here we present a full proteomic profiling of a human lymphoma B-cell line (*Ramos*) performed using a nanoUPLC-LTQ-Orbitrap Velos proteomic platform, combined to an in-depth transcriptomic profiling of the same cell type. Data are available via ProteomeXchange with identifier PXD001933. Integration of the proteomic and transcriptomic data sets revealed a 94% overlap in the proteins identified by both -omics approaches. Moreover, functional enrichment analysis of the proteomic profiles showed an enrichment of several functions directly related to the biological and morphological characteristics of B-cells. In turn, about 30% of all protein-coding genes present in the whole human genome were identified as being expressed by the *Ramos* cells (stable average of 30% genes along all the chromosomes), revealing the size of the protein expression-set present in one specific human cell type. Additionally, the identification of missing proteins in our data sets has been reported, highlighting the power of the approach. Also, a comparison between neXtProt and UniProt database searches has been performed. In summary, our transcriptomic and proteomic experimental profiling provided a high coverage report of the expressed proteome from a human lymphoma B-cell type with a clear insight into the biological processes that characterized these cells. In this way, we demonstrated the usefulness of combining -omics for a comprehensive characterization of specific biological systems.

**KEYWORDS:** C-HPP, lymphoma B-cell line, protein expression profile, transcriptomics, subcellular fractionation



## INTRODUCTION

Upon successful completion of the Human Genome Project in 2003—approximately 20 055 protein-coding genes have been reported according to neXtProt version of September 19, 2014—a major challenge remains as regards the understanding of how gene expression levels relate to the regulatory behavior

**Special Issue:** The Chromosome-Centric Human Proteome Project 2015

**Received:** May 28, 2015

**Published:** July 28, 2015

of cells.<sup>1,2</sup> After the genomics era, the scientific community realized that DNA coding sequences are not by themselves sufficient to provide an overview of cellular biological processes.<sup>3</sup> The genome remains nearly constant throughout the lifetime of a cell and there are no significant changes regarding the cell type once it achieves its differentiated specific state. However, both the transcriptome and the proteome are much more dynamic and they can vary with the functional state of the cell or in response to intra- and extra-cellular environmental signals. Henceforth, studying changes in mRNA and protein levels can provide a clear and accurate readout of the cell state.<sup>1</sup> Proteins are usually the final regulatory and effector molecules of cells coded by genes, and proteomics allows a comprehensive and integrative study of all proteins in a cellular system.<sup>4</sup> The main goal of proteomics is often to generate a complete and quantitative map of proteins, including cellular localization of proteins, identification of protein complexes, protein isoforms, and post-translational protein modifications (PTM). In fact, proteomes are characterized by large protein-abundance differences, cell-type, time-dependent expression patterns, and PTMs, all of which carry biological information that is not commonly accessible by genomics or transcriptomics data.

In turn, transcriptomics methodologies are the most used to determine the active expression of predicted protein-coding genes. In this respect, RNA deep sequencing technology has emerged as a promising strategy that provides a whole transcriptome shotgun approach to quantifying detailed genome-wide expression.<sup>5,6</sup> Previous to the development of high-throughput RNA and DNA sequencing technologies, microarray technologies have been considered powerful large-scale methods that have been applied for gene expression profiling of multiple samples in many different biological studies. Nevertheless, the detection and quantification of expressed mRNAs do not unequivocally determine the presence of the corresponding translated proteins. Regulatory mechanisms, such as PTMs or silencing processes, can result in an imbalance between the transcribed and the translated portions, as well as the half-live differences between transcripts and proteins.<sup>3,7–11</sup>

Proteomics measures proteins directly, providing information about the active genes at the translational level, and can be used as verification of gene expression.<sup>3,12</sup> Major advances which have occurred in proteomics over the last years have allowed detection and validation of putative genes, together with the added benefit for genome annotation.<sup>13,14</sup> However, many challenges still remain in these approaches due to proteome complexity.<sup>15</sup> Among others, these are related to (i) the sample preparation procedures, because to get maximum coverage of the proteome it is necessary to use multiple sample fractionation methods, either through protein extraction protocols or processing techniques; (ii) the peptide separation optimized via the usage of combinations of different proteases (e.g., trypsin, Lys-C, proteinase K, etc.) to increase peptide recovery leading to increase protein sequence coverage; (iii) the precision and accuracy of mass spectrometry (MS) measurements, because to reduce the occurrence of false positive hits it is recommended to use instruments with high resolution, high sensitivity, fast scanning speed, and high mass accuracy (ppm) such as the Fourier Transform Ion Cyclotron Resonance (FT-ICR) and the Linear Trap Quadrupole-Orbitrap (LTQ-Orbitrap) (also affected by a different ionization efficiency of different peptides); (iv) the data

processing methods and database search engines (i.e., Mascot, Sequest, OMSSA), as well as the scoring and validation parameters (false discovery rate – FDR, precursor mass tolerance), which determine the robustness of the results about the identified peptides and proteins<sup>9,13–13</sup>.

All these challenges are still open in proteomics; therefore, the integration of transcriptomics and proteomics data, working with adequate bioinformatics strategies, can offer new insights in the field<sup>1,16</sup> and provide reliable information about how genes and proteins are regulated and integrated at the molecular, cellular, and organismal levels to control a set of biological responses.<sup>17</sup>

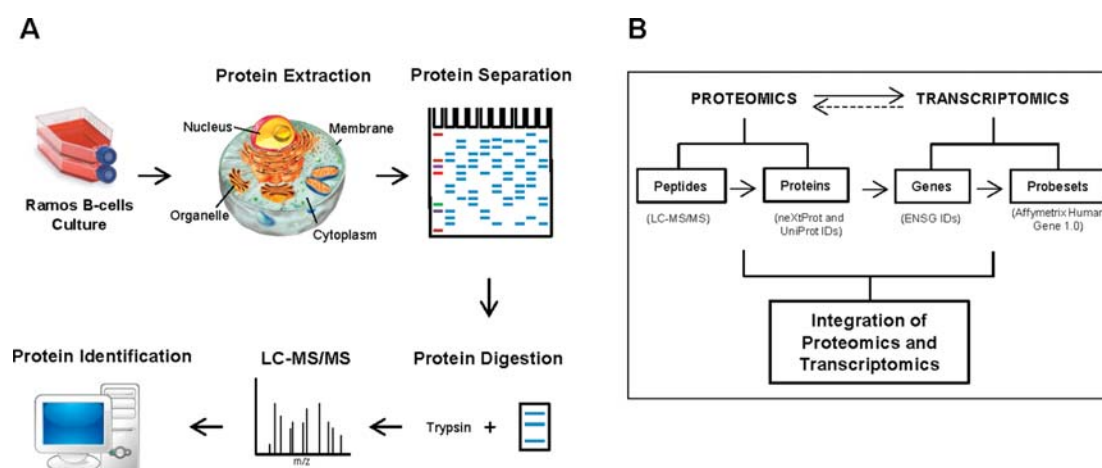
The Human Proteome Organization (HUPO) has coordinated the efforts of the international community promoting several initiatives to describe the human proteome through a well-planned working scheme. The project is partially organized according to a chromosome-based strategy (C-HPP) where scientific groups from different nationalities agreed to characterize the proteome of a specific chromosome.<sup>18</sup> All 24 chromosomes plus the mitochondrial DNA have been already assigned to many teams from 21 different countries. The vast heterogeneity, wide dynamic range, and different ionization efficiencies of peptides are causing a restriction in detection and quantification capacities of large-scale proteomics studies. Hence, C-HPP groups are now integrating transcriptomics and proteomics data sets in order to better guide the genome-wide proteomics analysis.<sup>19</sup> Specifically, the Spanish team of the Human Proteome Project (SpHPP) addresses the protein mapping of chromosome 16, and that it is the reason why a lymphoma cell line has been used in this research, because a great number of proteins from B-cells are encoded in chromosome 16.<sup>2,20</sup> The development of studies integrating proteomics and transcriptomics may lead to the full characterization of chromosomes and also to determine the relationship between the transcripts and their products (i.e., proteins). Additionally, the usage of shotgun MS approaches generates huge amounts of data providing lots of information in which it might be possible to detect missing proteins. Also, the C-HPP initiative is expected to improve the knowledge about diseases and their biology contributing to the Biology and Disease (B/D)-HPP initiative.

Here, we present a workflow integrating transcriptomics and proteomics data sets about a specific biological sample. As a model, we selected the *Ramos* human B-cell lymphoma cell line, which is well-characterized at the gene expression level. In addition, the proposed approach has reported the identification of missing proteins that correspond to certain subset of protein-coding genes well-annotated in the human genome that have not been yet detected by any proteomic MS-based experimental approach. Additionally, a further comparison between neXtProt and UniProt database searches has been accomplished. On balance, all these points may be of great interest for the scientific community and, specifically, for the C-HPP consortium.

## ■ MATERIALS AND METHODS

### Reagents

Protease inhibitor cocktail, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), Tween 20, tris (2-carboxyethyl) phosphine hydrochloride (TCEP), phenylmethanesulfonyl fluoride (PMSF), digitonin, octylphenoxypolyethoxyethanol (IGEPAL), RPMI-1640 media, potassium ferrocyanide, sodium



**Figure 1.** Schematic representation of the strategies followed for the integration of proteomics and transcriptomics analysis. (A) Overview of the experimental workflow. Subcellular protein extraction was performed from *Ramos* B-cells. After protein separation in an SDS-PAGE gradient gel (4–20%), proteins were digested with trypsin. The digests were analyzed using an nUPLC-LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). SEQUEST and MASCOT database search algorithms were used for protein identification. (B) Integration of proteomics and transcriptomics workflow. Comparison of proteomics and transcriptomics data was made via mapping from peptides (obtained by an LC-MS/MS strategy) to DNA probes (Affymetrix Human Gene 1.0 platform).

thiosulfate, dithiothreitol (DTT), iodoacetamide (IAA), formic acid (FA), and acetonitrile (ACN) were purchased from Sigma (St. Louis/MO, U.S.A.). Heat-inactivated fetal bovine serum (FBS), L-glutamine, penicillin, and streptomycin were obtained from Gibco (Scotland, U.K.). n-Dodecyl- $\beta$ -D-maltopyranoside was purchased from Affymetrix (Maumee, OH, U.S.A.), Coomassie Brilliant Blue was from Merck (Kenilworth, NJ, U.S.A.), and trypsin was from Promega (Madison, WI, U.S.A.). Monoclonal antihuman antibodies conjugated with phycoerythrin (PE): CD20 (clone L27), CD22 (clone HIB22), CD11a (clone HI111), CD45 (clone HI30), CD19 (clone HIB19), and CD3 (clone SK7) were from BD Biosciences (Pharmingen, San Diego, CA, U.S.A.); CD79b (clone CB3–1), and IgG1 mouse isotypic control antibody was from BioLegend (San Diego, CA, U.S.A.). Polyclonal antihuman antibodies conjugated with PE: kappa-Ig and lambda-Ig light chains were purchased from Cytognos SL (Salamanca, Spain).

The hypotonic lysis buffer used contained 30 mM HEPES pH = 8, 20% (v/v) glycerol, 15 mM KCl, 1 mM EDTA, 2 mM MgCl<sub>2</sub>, 1 mM PMSF, 1 mM TCEP, and 1% (v/v) protease inhibitor cocktail. The hypertonic lysis buffer contains the same components as the hypotonic lysis buffer, except glycerol.

### Cell Culture

The *Ramos* Burkitt's lymphoma-derived B-cell line (ATCC CRL 1596) was cultured at 37 °C in a humidified CO<sub>2</sub> incubator (5% CO<sub>2</sub>) in complete RPMI media (RPMI-1640 medium supplemented with 10% (v/v) FBS, 200 mM L-glutamine, 10 000U/mL penicillin, and 10 000  $\mu$ g/mL streptomycin).

### Cell Harvest and Cell Lysis Methods

Cellular proteins were harvested by washing the cells (30  $\times$  10<sup>6</sup> cells/experiment) twice with PBS, and cells were pelleted for 5 min at 1000g. The lysis buffer was then added at a volume equal to 5 times that of the cell pellet; all steps were performed at 4 °C (Figure 1A). The cell lysis methods A and B (Supplementary Figure 1) were performed in triplicate.

**Method A.** The hypotonic lysis buffer supplemented with 0.015% (w/v) digitonin was added to pelleted cells. After 30

min of rotation, the sample was centrifuged at 1500g for 5 min, and the cytoplasmic proteins (CYT) were collected in the supernatant. The remaining fractions were processed stepwise in an identical manner. For organelle proteins (ORG), the hypotonic lysis buffer with 0.5% (v/v) Tween 20 was used; for membrane proteins (MB), the hypotonic lysis buffer supplemented with 0.5% (v/v) IGEPAL detergent was employed, and lastly, nuclear proteins (NUC) were extracted by adding the hypertonic lysis buffer supplemented with 1% (w/v) n-dodecyl- $\beta$ -D-maltopyranoside after 10 min incubation. Two washing steps were performed with nonsupplemented hypotonic lysis buffer between the distinct fractionation steps.

**Method B.** The CYT and ORG fractions were obtained as described in Method A. The NUC fraction was extracted with hypertonic lysis buffer supplemented with 140 mM NaCl. Lastly, the MB fraction was obtained after incubating for 5 min with hypotonic lysis buffer plus 1% (w/v) n-dodecyl- $\beta$ -D-maltopyranoside.

### Protein Quantification and SDS-PAGE Separation

After protein quantification by the Lowry-DC-Protein Assay as recommended by the manufacturer (Bio-Rad Laboratories, CA, U.S.A.), each sample was separated in a 4–20% gradient SDS-PAGE gel under reducing conditions. The same amount of protein (15  $\mu$ g) was run for each fraction (CYT, ORG, NUC, MB). After electrophoresis, gels were stained in a solution of 0.5% (w/v) Coomassie Brilliant Blue. Gels were stored at 4 °C in an aqueous solution containing 1% (v/v) acetic acid, until analysis.

### In-Gel Digestion and LC-MS/MS Analysis

Each gel lane was cut into five fragments and digested with trypsin following the method of Shevchenko et al.<sup>21</sup> with slight modifications. Briefly, gel pieces were destained with 15 mM potassium ferrocyanide and 50 mM sodium thiosulfate. Protein reduction and alkylation were performed with 10 mM DTT at 56 °C for 45 min, and with 55 mM IAA at room temperature for 30 min, respectively. Proteins were digested with trypsin (6.25 ng/mL) at 37 °C for 18 h. The peptide solution was acidified with FA and desalted by using C18-Stage-Tips

columns.<sup>22</sup> The samples were partially dried and stored at  $-20^{\circ}\text{C}$  until they were analyzed by LC-MS/MS.

A nanoUPLC system (nanoAcquity, Waters Corp., Milford, MA, U.S.A.) coupled to a LTQ-Velos-Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA, U.S.A.) via a nanoelectrospray ion source (NanoSpray flex, Proxeon, Thermo) was used for reversed-phase LC-MS/MS analysis. Peptides were dissolved in 0.5% FA/3% ACN and loaded onto a trapping column (nanoACQUITY UPLC 2G-V/M Trap Symmetry 5  $\mu\text{m}$  particle size, 180  $\mu\text{m}$   $\times$  20 mm C18 column, Waters Corp., Milford, MA, U.S.A.). Peptides were separated on a nanoACQUITY UPLC BEH 1.7  $\mu\text{m}$ , 130  $\text{\AA}$ , 75  $\mu\text{m}$   $\times$  250 mm C18 column (Waters Corp., Milford, MA, U.S.A.) with a linear gradient from 7% to 35% solvent B (ACN/0.1% FA) at a flow rate of 250 nL/min over 120 min.

The nUPLC- LTQ-Orbitrap Velos was operated in the positive ion mode by applying a data-dependent automatic switch between survey MS scan and tandem mass spectra (MS/MS) acquisition. Survey scans were acquired in the mass range of  $m/z$  400 to 1600 with a 60 000 resolution at  $m/z$  400 with lock mass option enabled for the 445.120025 ion.<sup>23</sup>

The 20 most intense peaks having  $\geq 2$  charge state and above the 500 intensity threshold were selected in the ion trap for fragmentation by collision-induced dissociation with 35% normalized energy, 10 ms activation time,  $q = 0.25$ ,  $\pm 2$   $m/z$  precursor isolation width and wideband activation. Maximum injection time was 1000 and 50 ms for survey and MS/MS scans, respectively. AGC was  $1 \times 10^6$  for MS and  $5 \times 10^3$  for MS/MS scans. Dynamic exclusion was enabled for 90 s.

#### Database Search

Raw data were translated to mascot general file (mgf) format and searched against the neXtProt database (release September 19, 2014) using the target-decoy strategy with an in-house MASCOT Server v. 2.3 (Matrix Science, London, U.K.). Decoy database was created using the peptide pseudoreversed method, and separate searches were performed for target and decoy databases. Search parameters were set as follows: carbamidomethylation of cysteine as a fixed modification, oxidation of methionine and acetylation of the protein n-terminus as variable ones, precursor and fragment mass tolerance were set to 10 ppm and 0.8 Da, respectively, and fully tryptic digestion with up to two missed cleavages. FDR at PSM level (psmFDR) and protein level (protFDR) were calculated using MAYU.<sup>24</sup> Using C-HPP guidelines, protein identifications were obtained using the criteria psmFDR  $< 1\%$  and protFDR  $< 1\%$ . Lastly, protein inference was performed using the PAnalyzer algorithm,<sup>25</sup> and nonconclusive protein groups were discarded.

For the search against UniProt database, the MASCOT<sup>26</sup> and SEQUEST HT<sup>27</sup> algorithms were used to search for the acquired MS/MS spectra, using Thermo Scientific Proteome Discoverer software (v. 1.4.1.14) against a custom database of all human reviewed sequences downloaded from the UniProt database (February, 2014) and common contaminant sequences (e.g., human keratins, trypsin, BSA). Search parameters were the same as for the search against neXtProt database. Peptides having MASCOT ion scores of  $< 20$  were not considered for analysis. A 1% FDR using Percolator<sup>28</sup> was employed for peptide validation as well as for PSM level.

Supporting Information 1 contains raw data about the results obtained for representative samples.

#### Transcriptomic Analysis

To perform the gene expression profiling and analysis of the Ramos B-cells, we used a raw data set of three samples of mRNA from biological replicates of these cells hybridized on Affymetrix Human Gene ST 1.0 high-density oligonucleotide microarrays. The data are available at GEO (Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo/>) database: series number GSE40168; samples GSM987747, GSM987748, GSM987749; platform GPL6244 ([HuGene-1\_0-st]). The preprocessing, normalization, and signal calculation of these data were done using the Bioconductor packages oligo<sup>29</sup> and pd.hugene.1.0.st.v1.<sup>30</sup>

#### Integration of Transcriptomics and Proteomics Data Sets

In order to map the neXtProt IDs of the proteins identified in the LC-MS/MS proteomic assays to the corresponding genes and probesets detected by the gene expression transcriptomics assays, an ID-mapping procedure was run in two steps (Figure 1B): (1st) from protein neXtProt IDs to gene Ensembl IDs, using the mapping to genes provided by neXtProt database (release September 19, 2014); and (2nd) from the gene Ensembl IDs to gene Affymetrix probesets IDs, using the Brainarray tool<sup>31</sup> with the mapping table hugene10st\_Hs\_ENSG\_mapping 18.0.0 corresponding to the arrays used. Before this ID-mapping, we unified all the neXtProt IDs obtained from the different isolated subcellular fractions (i.e., cytoplasm, organelles, membranes, and nucleus). Following these unification and mapping procedures, the neXtProt IDs and UniProt ACs were used to create different data sets with different levels of coverage and confidence based on the combination of the results obtained with three replicates (i.e., all proteomic experiments consisted of three independent biological replicates of the Ramos B-cells). In this way, the following data sets were generated (Table 1): (i) an *intersection* data set, including those proteins which were systematically identified in the three replicated experiments, with at least 2 proteotypic peptides at protFDR  $< 0.01$ ; (ii) an *union* data set, including all proteins identified in any of the replicated experiments with at least 2 proteotypic peptides and protFDR  $< 0.01$ , and (iii) a *maximum* data set, including all proteins identified in any of the replicated experiments with at least 1 proteotypic peptide and protFDR  $< 0.01$ . These three data sets are included in each other, the third one being the one with the largest coverage (i.e., *maximum* data set) and therefore the one that provides the largest list of proteins identified.

To map proteins to genes in chromosomes, the Biomart<sup>32</sup> managed with R and Bioconductor tools were used. A brief R-script is provided as Supporting Information (Supporting Information 2) to show the details of the mapping protocol and the comparison of proteomic and transcriptomic data. Functional enrichment analysis (FEA) and clustering of the gene lists were done using the DAVID<sup>33</sup> and GeneTerm-Linker<sup>34</sup> tools. The main biological databases selected to find genes with annotated enriched terms were the following: (i) Gene Ontology (GO) using annotations spaces GOTERM\_BP, GOTERM\_CC and GOTERM\_MF; (ii) the pathways database KEGG\_PATHWAY; (iii) the INTERPRO and PFAM protein structural domain database; and (iv) the UNIGENE\_EST and GNF\_U133A\_QUARTILE tissues-specific expression databases. To generate the functional clusters in DAVID, we used classification stringency *medium*.<sup>33</sup> All statistical analyses of data distributions, the comparisons, and most of the mapping were done working in the R/Bioconductor environment.<sup>35</sup>



**Table 1. Number of Proteins and Genes Included in the Datasets Produced in the Analyses of Ramos B-Cells<sup>a</sup>**

	no. of proteins (neXtProt)		no. of genes (Ensembl)	
	total neXtProt IDs	ENSG IDs	Affymetrix probeset IDs	
intersection <sup>b,c</sup>	3383	3433	4088	
union <sup>b,d</sup>	5494	5540	6175	
maximum <sup>d,e</sup>	8931	8976	9494	
“exclusive identifications” in transcriptomics <sup>f</sup>	-	1290	-	
“exclusive identifications” in proteomics <sup>g</sup>	516	-	-	

<sup>a</sup>The table indicates the number of protein and gene distinct IDs found in the proteomic and genomic assays, respectively: Row 1, intersection dataset; row 2, union dataset; row 3, maximum dataset (these datasets are defined in Materials and Methods). Row 4 [“Exclusive identifications” in Transcriptomics] includes the proteins that were not detected in proteomics but detected in the genomic data in the 25% highest expression quartile of the Affymetrix Human Gene ST 1.0 microarrays (calculating the expression signal average for the 3 arrays). Row 5 [“Exclusive identifications” in Proteomics] includes the genes that were not detectable by the genomic platform (i.e. genes not present in the microarray) but were detected by the proteomic approach. The columns in the table correspond to (i) all the human neXtProt IDs detected, (ii) the mapped Ensembl IDs, and (iii) the mapped Affymetrix probeset IDs. <sup>b</sup>Proteins detected by at least 2 unique peptides in the MS proteomic experiments. <sup>c</sup>Proteins detected in all the 3 experimental biological replicates. <sup>d</sup>Proteins detected in any replicate. <sup>e</sup>Proteins detected by at least 1 unique peptide in the MS proteomic experiments. <sup>f</sup>Genes detected in the 25% higher expression quartile of the microarrays but not present in the MS data. <sup>g</sup>Proteins detected in the MS/MS data but not present in the expression microarrays.

The same analysis strategy was performed for mapping UniProt IDs (using UniProtKB database release February 2014).

### Immunophenotypic Analysis

Surface membrane expression of some proteins was validated by flow cytometry using a direct immunofluorescence technique with antihuman phycoerythrin (PE)-conjugated antibodies for the following proteins: CD3, CD11a, CD19, CD20, CD22, CD45, CD79b, kappa-Ig, and lambda-Ig light chains, together with a PE-conjugated isotype control antibody. All the antibodies were purchased from BDBiosciences (San José, CA, U.S.A.). Briefly, Ramos B-cells ( $0.5 \times 10^6$ ) were washed in 0.5% BSA in PBS and incubated with the antibodies for 15 min at room temperature, washed, and acquired on a FACSCanto II flow cytometer (BD Biosciences, San José, CA, U.S.A.) using the FACSDiva software (version 6.1, BD Biosciences). For data analysis, the Infinicyt software (Cytognos SL, Salamanca, Spain) was used.

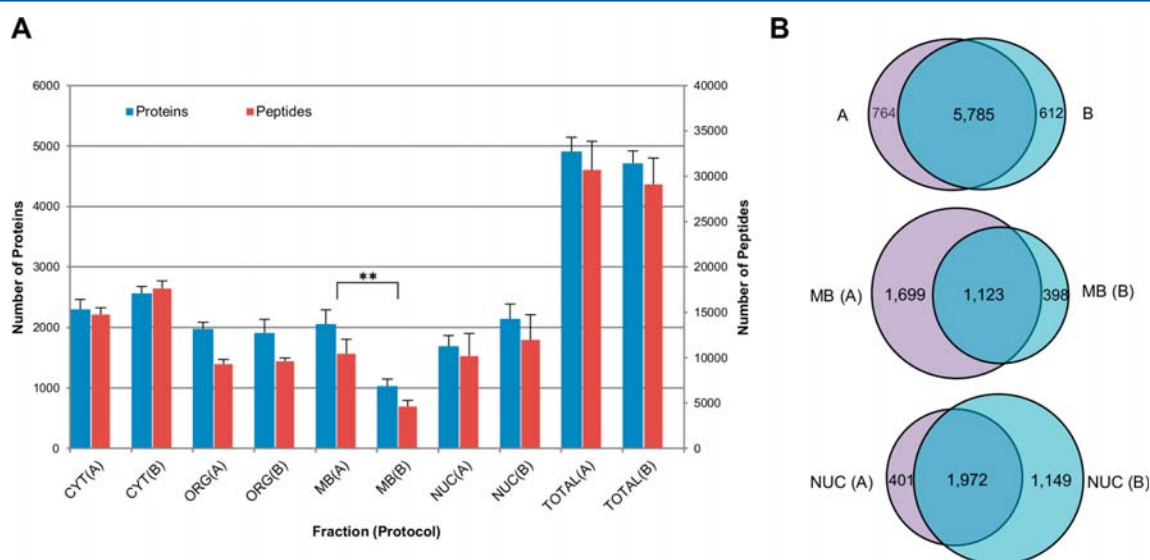
### Statistical Methods

For all continuous variables, mean values and their standard deviation (SD) were calculated. To evaluate the statistical significance of differences observed between groups, the two independent sample Student's *t* test was used for continuous variables displaying a normal distribution (SPSS software 18.0 package; SPSS, Inc., Chicago, IL). Statistical significance was set at a *P* value of <0.01.

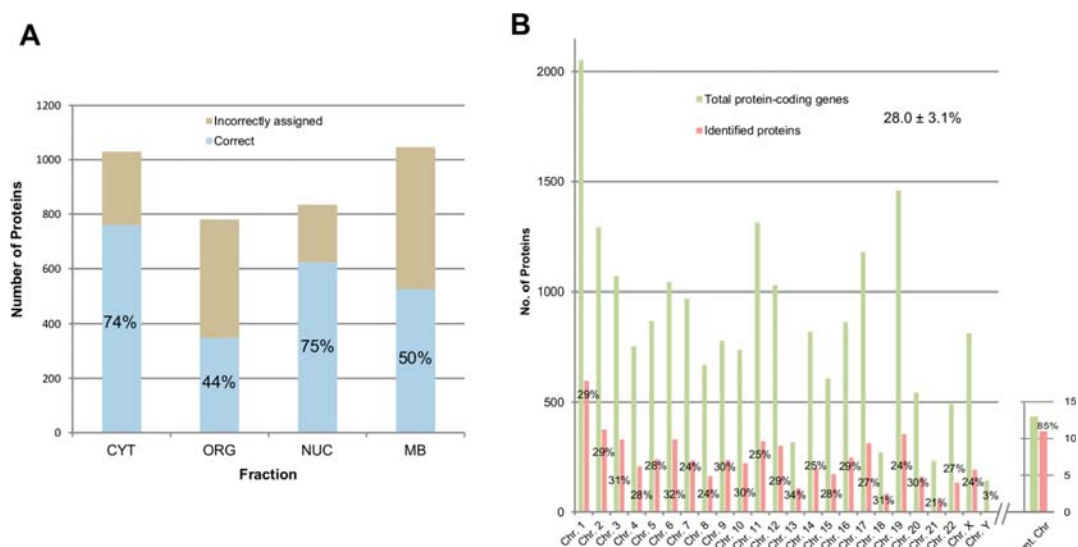
## RESULTS AND DISCUSSION

### Subcellular Fractionation, Key To Increase Proteome Coverage

Mass-spectrometry techniques can increase the coverage of the proteome by using subcellular fractionation strategies for protein extraction.<sup>20</sup> This approach enables the in-depth analysis of biomolecules by reducing the sample complexity through isolation of different subcellular fractions. Here, we performed sequential extraction of proteins (from three biological replicates) by using a combination of different detergents specific for protein profiling from distinct subcellular



**Figure 2.** Comparison of protein extraction methods. (A) Total number of proteins and peptides (without duplicates) identified by LC-MS/MS assays for each subcellular fraction (CYT, cytoplasmic; ORG, organelle; MB, membrane; NUC, nuclear; and total protein extract, TOTAL) and protein extraction method (A or B). Each value was calculated as an average using the three replicates. (B) Robustness of LC-MS/MS assays for protein extraction methods A and B. Union of proteins from the three replicates (without duplicates) were considered for this comparison. Upper panel: comparison of total protein extracts. Middle and lower panels: comparisons of the membrane and nuclear proteins, respectively. \*\* *p* < 0.01.



**Figure 3.** Subcellular protein localization. (A) Considering protein extraction method A and proteins identified in common in all three replicates (*intersection*), we validated our protein identifications for each subcellular fraction comparing with the subcellular location information given in the protein database. Correctly assigned proteins (in light blue) are the proteins whose subcellular location matches both in our MS/MS results and in the literature. Incorrectly assigned (in light brown) proteins are referred to not matching. (B) Chromosome mapping of proteins identified by at least one unique peptide (*maximum*). Green bars show the total protein-coding genes identified per chromosome (data from Ensembl release 78), red bars show the number of proteins identified by our MS/MS strategy. Corresponding percentages are noted within each bar.

compartments of *Ramos* lymphoma B-cells. The efficiency of solubilization and the maintenance of protein structure are directly dependent on the detergent choice, salt concentration, and pH. Thus, two similar approaches (methods A and B) were tested in parallel to isolate four different subcellular fractions of proteins (cytoplasmic, CYT; organelle, ORG; nuclear, NUC; and membrane, MB). These methods only differ in the extraction of membrane and nuclear proteins, which can be isolated by using IGEPAL and n-dodecyl- $\beta$ -D-maltopyranoside (for method A) or NaCl and n-dodecyl- $\beta$ -D-maltopyranoside (for method B), respectively. However, in both methods, the cytoplasmic and organelle fractions were similarly isolated using digitonin and Tween 20 (both are nonionic detergents), respectively. Digitonin effectively water-solubilizes membrane proteins, but not the nuclear ones which remain structurally intact. Thus, cytosolic proteins can be recovered. Second, we used Tween 20, which does not affect protein activity, to extract organelle-related proteins.

Supplementary Figure 2A displays the total amount of protein extracted from each subcellular fraction. By SDS-PAGE analysis (Supplementary Figure 2B), a homogeneous distribution of proteins independently of the molecular weight is observed.

The MS/MS data analysis of the four subcellular fractions described revealed a high number of identifications at the protein and peptide levels. In these assays, a very tight precursor mass tolerance (0.8 Da) and an FDR lower than 1% for both PSM, peptide and protein, were considered to reduce the chance of false positive identifications. The results obtained suggest that the differences in the number of identified proteins, between the protein extraction methods A and B, correlate with the detergent selected/used (Figure 2A). Specifically, the difference between the number of membrane proteins recovered with IGEPAL (method A) and n-dodecyl- $\beta$ -D-maltopyranoside (method B) is statistically significant (*p*-value of 0.007). In case of method A, IGEPAL solubilizes membrane

proteins, whereas it is not strong enough to lyse the nuclear membrane, which allows this subcellular compartment to remain intact for the effects of the n-dodecyl- $\beta$ -D-maltopyranoside detergent (for extraction of nuclear proteins). In turn, in method B, the membrane fraction was isolated in the last step by using n-dodecyl- $\beta$ -D-maltopyranoside.

Regarding nuclear proteins, we detected that usage of high salt concentrations (method B) was more efficient than n-dodecyl- $\beta$ -D-maltopyranoside (method A) for nuclear protein extraction. In summary, these preliminary analyses indicate a slight increase in proteome coverage with method A versus method B. Therefore, the differences between both methods provide an explanation to the variations observed in protein recovery. In fact, with method A, a better recovery of membrane proteins was achieved, and we were particularly interested in the membrane fraction because it is the cellular compartment where more missing proteins are estimated to be located in this specific cell line.

For a more in-depth comparison between the two protein extraction methods, the robustness through the 3 replicates was analyzed with an overall overlap of 81%, 35%, and 56% between methods A and B for total, membrane, and nuclear proteins, respectively (Figure 2B). Of note, again method A identified a greater number of proteins than method B (up to 4 times). Therefore, on the basis of these preliminary analyses, we selected method A for further transcriptomics–proteomics comparisons owing to the higher proteome coverage not only at the total protein level but also at the membrane fraction of proteins.

#### Characterization of the Proteome of *Ramos* B-cells

As described above, *Ramos* B-cells derive from a Burkitt lymphoma carrying the *MYC* gene rearrangements (i.e., t(8;14)<sup>36</sup>), and they are often used as a model for proteomics of B lymphocytes. *MYC* gene was first discovered in Burkitt lymphoma patients as a proto-oncogene whose activation leads to the induction of cellular proliferation. Although our

proteomics strategy has not detected the Myc protein—mainly due to the fact that this protein is located on the nucleus, as it is further explained in the section entitled Exclusive Identifications from Comparison of both Proteomics and Transcriptomics—a large number of proteins interacting with Myc have been identified (Actl6a, Bcl2, Chd8, Gtf2l, Mapk1, Max, Mlh1, Mycbp2, Mycbp, Nmi, Nyfc, Pfdn5, Ruvb, Sap130, Smad2, Smad3, Smarca4, Smarcb1, Taf9, Wdr5, Yyi, among others). Thus, this -omic platform allows the characterization of cell signaling pathways by identifying the components of the interactions leading to changes in cellular responses.

In order to evaluate the robustness of the MS/MS data set and its reliability for integration with transcriptomics data sets, the presence of cross-contamination between the subcellular fractions obtained with method A was determined according to the protein database. Thus, the percentage of the proteins identified in each subcellular fraction that were correctly isolated by the protein extraction method is presented in Figure 3A. In total, 74–75% was the proportion found for the fraction of proteins assigned by the protein database to cytoplasm (CYT) and nucleus (NUC). The lowest percentages for correct protein location corresponded to the organelle (ORG) and the membrane protein (MB) fractions. Both fractions were sequentially and consecutively extracted, and the nature of their protein mixture is highly related, hindering their correct localization. In addition, the literature-based protein database assignment can be ambiguous because many times there is more than one unique location for each protein, and additionally, there are membrane-associated organelle proteins that could be included in both fractions. Because our main goal was to profile the *Ramos* cell proteome, all proteins identified were grouped into a unique data set independently of their correct or incorrect location. In this regard, it is important to keep in mind that subcellular fractionation was specifically used to increase protein recovery, not for the independent analysis of the fractions.

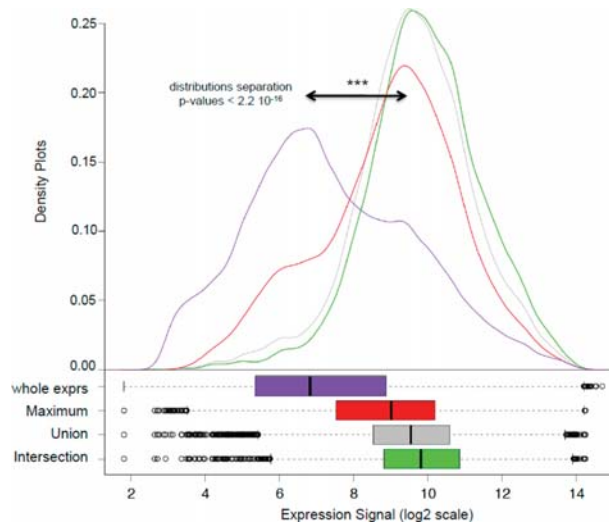
On the basis of chromosome mapping (Figure 3B) of the proteins identified, overall we identified about 30% of all protein-coding genes (~5000–6000) present in the human genome (without considering the X and Y chromosomes, and mitochondrial DNA). This confirms that this approach is proteome-wide and unbiased, proving that selecting for a specific cell type is a useful approach for total proteome coverage. Interestingly, the proteins identified also covered about an 85% of the total protein-coding genes of the mitochondrial DNA, using the LC-MS/MS approach here employed. This good coverage of the mitochondrial proteins indicates a great improvement with respect to other proteomic procedures and platforms (e.g., GFP-tagging analysis<sup>37</sup>). This is mainly due to the development of proteomic strategies with high sensitivity and accuracy in combination with the usage of subcellular fractionation approaches that allow the enrichment of the organelle proteomes such as the mitochondrial one.<sup>38</sup>

Other studies have recently published proteomic data related to B-cells lymphomas by using SILAC approaches. Mann's lab has developed a super-SILAC to classify large B-cell lymphomas subtypes by their protein profiles.<sup>39</sup> In turn, Rüetschi and colleagues studied the same disease by a SILAC-based quantitative approach.<sup>40</sup> A comparison of our results revealed an overlap with 1696 and 2141 proteins identified in Mann's and Rüetschi's studies, respectively, constituting 54% and 60% of proteins in common (data not shown). The differences in identification might be due to the absence of subcellular

fractionation steps in these studies that improves the recovery and protein identification. Even so, B-cell specific proteins have been detected by both approaches, demonstrating the feasibility of proteomics strategies to characterize and model the disease.

### Integration of Proteomics and Transcriptomics Data Sets

The characterized MS/MS proteomic data was compared to transcriptomic measurements performed on *Ramos* B-cells (3 biological replicates) with genome-wide expression high-density oligonucleotide microarrays (Figure 4). To accomplish the



**Figure 4.** Density plots and boxplots showing the distributions of the whole gene expression signal versus the signal corresponding to the genes detected in the MS proteomic profiles for *Ramos* B-cells. The expression was measured with *Affymetrix* high-density oligonucleotide expression arrays type Human Gene 1.0. (Purple) Whole expression signal corresponding to all the 33 297 probesets included in the arrays [whole exprs]. (Red) 9494 probesets corresponding to the 8976 distinct human genes (ENSG IDs) and 8391 proteins (neXtProt IDs), which showed at least one unique identifying peptide in the MS proteomic assays [maximum]. (Gray) 6175 probesets corresponding to the 5540 distinct human genes (ENSG IDs) and 5494 proteins (neXtProt IDs), which showed at least two identifying peptides in the proteomic assays [union]. (Green) 4088 probesets corresponding to the 3433 distinct human genes (ENSG IDs) and 3383 proteins (neXtProt IDs), which had at least two identifying peptides in the proteomic assays and were found in all the replicates of the proteomic isolations [intersection].

proteomics-transcriptomics integration, the neXtProt IDs (corresponding to identified proteins by the LC-MS/MS approach) were mapped into the Ensembl IDs (for the genes), and these were mapped into the *Affymetrix* probesets IDs that identify the gene-specific DNA oligo probes which are present in the microarrays (Supporting Information 3,4 and 5). Such comparison was addressed from different ways termed *intersection*, *union*, and *maximum*, as described above in the Materials and Methods section. On the basis of this strategy, the following goals were pursued: (i) to identify as many proteins as possible using this approach present in *Ramos* B-cells (*maximum* and *union*), and (ii) to characterize those proteins that are unequivocally identified with this proteomic approach (*intersection*). Briefly, 3383, 5494, and 8931 known human proteins (i.e., neXtProt IDs) were found within the *intersection*, *union*, and *maximum* data sets respectively (Table 1). These data sets corresponded to 3433, 5540, and 8976

genes (i.e., Ensembl IDs). These results indicate that proteomics and transcriptomics studies, once applied to the global molecular characterization of a cell type, display a high accuracy and overlap within each analytical level. Such overlap can be estimated considering the proportion of proteins that are detected by proteomics and transcriptomics, because—as we explain below—there were only 516 proteins exclusive of proteomics (i.e., not detected by the transcriptomic platform) out of a total of 8931, and this corresponds to a 94% of overlap of the expression data over the proteomic data.

FEA of those 3383 proteins detected in a coherent and steady manner (*intersection* data set) by the proteomics assays (Supporting Information 6) revealed the expression of many essential proteins for general cell functions and house-keeping processes (e.g., anabolism and synthesis processes, together with catabolism and cellular respiration) as well as the activity and regulation of major biological macromolecules (DNA, RNA, and proteins) involved in key maintenance processes like cell cycle, cell growth, cell proliferation, and so forth.

On the other side, mapping of proteins to tissue-type and cell-type databases provided a clear enrichment in B-cell specific genes and proteins that was in agreement with the character and properties of a lymphocytic cell type. For example, regarding B-cell receptor (BCR) cross-linking, many intracellular signaling cascades appended to be activated leading to regulation of gene expression. Moreover, synthesis initiation proteins (eIF3a/h proteins), proteins for cellular adhesion and costimulatory signaling (CD11a), accessory signal transduction components (such as CD20, CD19, CD79a or CD79b), protein tyrosine kinases (Fyn, Lyn, Syk and Btk), and proteins related to the activation of Ras signaling pathway (N-, K-, and H-Ras) were identified. In turn, proteins belonging to the B-cell receptor signaling pathways have been widely detected including those related to BCR internalization (SHP-1, CD19, Bam32), cytoskeletal rearrangements and integrin activation (Bam32, PLC $\gamma$ 2, DAG), transcription (Blk, Grb2, SHP-1, Erk1/2, Raf), proteasomal degradation (PKC, Carma1, Bcl10, MALT1, IKK, NF- $\kappa$ B) and growth arrest and apoptosis (Akt, FoxO), among others pathways that follow BCR activation.

### Exclusive Identifications from Comparison of both Proteomics and Transcriptomics

As mentioned before, it is notorious that the overlap observed between both the proteomic and the transcriptomic methodologies applied to the same cell type, although these two experimental technologies had also a complementary part. In fact, comparing the results of both approaches, each one enables to cover the failures in identification (“exclusive identifications”) due to its technical limits with respect to the other strategy.

Performing the FEA for the exclusively identified proteins in proteomics (Supporting Information 7), we identified a gain of proteins related to the mitochondrial and ribosomal organelles, as well as cytoplasmic ones. This is a quite expected result because the genomic microarrays employed do not include probes for mitochondrial DNA. In addition, we could infer that our subcellular extraction method in the proteomic procedure was effective on isolating organelle proteins, as it is shown for mitochondria. A loss of identifications associated with immunoglobulins (Ig) and major histocompatibility complex (MHC) proteins was also detected in the transcriptomic data probably due to Ig gene rearrangements and hypersomatic

mutations of Ig genes that hamper the design of adequate array probes for these genes.

Regarding exclusively identified proteins in transcriptomics, the functional enrichment analysis (Supporting Information 7) revealed identifications related to nuclear and DNA-binding proteins. This is probably due to the fact that isolating the nuclear fraction in the last step of the proteomics approach decreases the recovery for nuclear proteins. Upon comparing the method chosen for this study (method A) with method B (Figure 2B), it seems clear that extracting the nuclear fraction in a previous step and with another detergent improves isolation of these proteins. Therefore, it could be appropriate to previously establish the protein fraction of interest to perform the best protein extraction method for such fraction. In our case, we select method A as the overall most suitable approach combining all fractions and also because we were interested in enriching for membrane proteins since these are the most commonly lost in many proteome-wide studies. Additionally, around 300 of these transcripts exclusively identified in transcriptomics correspond to noncoding RNAs. Thus, it is obvious that their corresponding proteins have not been detected by the proteomics platform.

### Missing Proteins

Since thousands of human proteins have not been detected yet (as it is noted in the last release of missing proteins from the neXtProt database), the exploration of proteome data sets has become an indispensable exercise that may be accomplished to reduce the number of existing missing proteins. With this purpose, we have performed the mapping of our results into the neXtProt database for missing proteins (current release of April 28, 2015) obtaining the results shown in Table 2. Specifically,

**Table 2. Missing Proteins Identified Across the Three Datasets Obtained after Searching against the neXtProt Database<sup>a</sup>**

data set	PE group				total
	PE2	PE3	PE4	PE5	
<i>maximum</i>	273	37	5	55	370
<i>union</i>	18	3	-	11	32
<i>intersection</i>	-	-	-	4	4

<sup>a</sup>Proteins from each dataset (*maximum*, *union*, and *intersection*) were mapped into the neXtProt missing proteins database (April 28, 2015)—used as reference of missing proteins—to identify missing proteins. The missing proteins have been classified accordingly to their protein existence (PE) level (i.e., PE2 for experimental evidence at transcript level; PE3 for protein inferred from homology; PE4 for predicted protein; and PE5 for uncertain proteins).

the searching has been carried out across the three data sets generated (*maximum*, *union*, and *intersection*) identifying up to 370 missing proteins from our *maximum* data set (containing 8931 neXtProt identifications). As the number of neXtProt IDs decreases from one data set to another (from *maximum* to *union* and *intersection*), the number of identified missing proteins decreases as expected (from 370 to 32 and 4, respectively). However, this reduction in number involves an increase in quality because these missing proteins have been identified in 3 different biological replicates and with, at least, 2 peptides per protein. In Supporting Information 8 is shown the information related to missing proteins identified for each data set and the related data (PE group, chromosome location).

### Comparative Analysis of the Results Obtained with neXtProt and UniProt Database Searches

In a further analysis, the effect of database search on identification was evaluated. With this purpose, we additionally performed a search against the UniProt database and compared the results (Supporting Information 9, 10, 11, 12) with those obtained with neXtProt. In general, neXtProt database search generated a greater number of identified proteins what determined, consequently, an increased number of missing protein identifications compared to UniProt (Supporting Information 13 and 14). The integration of information from different databases (Swiss-Prot, Ensembl, Human Protein Atlas, PeptideAtlas....) makes possible a better characterization of the proteomes and justifies this increase in identifications (more spectrum and peptide information in the database leads to a possible increasing in identifications). In addition, this supposes an improvement in the characterization of the specific functions of identified proteins. In Supporting Information 15 are reported the FEA of proteins exclusively identified in proteomics and transcriptomics for searches performed against neXtProt and UniProt, respectively, reporting an enriched annotation of the functions for neXtProt search.

### Immunophenotypic Characterization of the Ramos B-cells

In order to give support and provide some external validation of the characterization of the cell type studied in our proteomic and transcriptomic studies, we use an independent cell-oriented platform to characterize the immunophenotype of the Ramos B-cells. With this purpose, several proteins were selected and screened by multiparametric flow cytometry (FCM) (Supplementary Figure 3) showing high expression levels of the B-cell associated antigen—CD19, CD20, CD22 and CD45—as well as the coreceptor of the B-cell receptor, CD79b. Expression of the lambda-Ig light chain was high as well, confirming the available data for the Ramos cell line (ATCC CRL 1596, [www.atcc.org](http://www.atcc.org)). Additionally, CD11a, an integrin involved in cellular adhesion and costimulatory signaling was evaluated showing a dim expression. As negative controls for FCM assays, T-cell marker CD3 and kappa-Ig light chain were also tested.

### PERSPECTIVES AND CONCLUSIONS

Integration of proteomics and transcriptomics technologies has emerged as a potent strategy for the mapping of the biomolecules that define a cell type or a cellular state. The searches derived from both -omics methodologies are complementary, and once performed in conjunction one with the other, they allow a mutual validation and increase the total coverage of genome-wide protein-coding genes identified. With this goal in mind, the C-HPP initiative has promoted an effective combination of proteomics data into a genomic framework to achieve a full map of the human proteome that will provide a better understanding the studied biological systems.

Based on deep characterization of the proteome of the Ramos lymphoma B-cells by LC-MS/MS, a total of up to 6000 proteins and ~30 000 different peptides were identified. This data resulted in about 30% coverage of all human gene-coded proteins per chromosome found in the proteome of these lymphoma B-cells (such percentage being as high as 85% for the mitochondrial DNA). Integration of this proteomics data set with transcriptomics data sets (derived from high-density oligonucleotide microarrays technology) showed an 82% overlap between both technologies. Despite this, several gaps

were identified with each of the two technologies. Regarding proteomics, special attention must be given to the protein extraction method that could affect correct extraction of proteins from specific subcellular compartments; moreover, proteins expressed at very low concentrations and in highly complex multimers could also have problems in being detected. For this reason, extensive fractionation of the proteome of interest might contribute to improve the resolution of proteomics via detection of thousands of proteins simultaneously, improving current knowledge about the functional and biological cellular processes via higher coverage of the proteins involved in these mechanisms.

Finally, the characterization of PTMs within the lymphoma proteome would be of great interest to model the disease. In this sense, further studies will be performed to determine the PTMs that may influence the normal response behavior of these cells.

### ASSOCIATED CONTENT

#### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.5b00474.

Additional data as noted in the text (ZIP)

### AUTHOR INFORMATION

#### Corresponding Authors

\*E-mail: [jrivas@usal.es](mailto:jrivas@usal.es). Phone: +34 923294819. Fax: +34923294743.

\*E-mail: [mfuentes@usal.es](mailto:mfuentes@usal.es). Phone: +34 923294811. Fax: +34 923294743.

#### Notes

The authors declare no competing financial interest.

### ACKNOWLEDGMENTS

We gratefully acknowledge financial support from the Carlos III Health Institute of Spain (ISCIII, FIS PI11/02114, FIS PI14/01538, and FIS PI12/00624), Fondos FEDER (EU) and Junta Castilla-León SA198A12-2. The Proteomics Unit belongs to ProteoRed, PRB2-ISCIII, supported by grant PT13/0001. P.D. and C.D. are supported by a JCYL-EDU/346/2013 Ph.D. scholarship. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium<sup>41</sup> via the PRIDE partner repository with the dataset identifier PXD001933. We thank Peter Horvatovich for providing us the list of missing proteins (release 2015-04-28).

### ABBREVIATIONS

ACN, acetonitrile; BCR, B-cell receptor; C-HPP, Chromosome-human Proteome Project; CYT, cytoplasmic proteins; DTT, dithiothreitol; FA, formic acid; FBS, fetal bovine serum; FCM, flow cytometry; FDR, False Discovery Rate; FEA, Functional Enrichment Analysis; FT-ICR, Fourier Transform-Ion Cyclotron Resonance; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; HUPO, Human Proteome Organization; IAA, iodoacetamide; Ig, immunoglobulin; IGEPAL, octylphenoxypolyethoxyethanol; LTQ, Linear Trap Quadrupole; MB, membrane proteins; MHC, major histocompatibility complex; MS, mass spectrometry; NUC, nuclear proteins; ORG, organelle proteins; protFDR, FDR at protein level; psmFDR, FDR at PSM level; PMSF, phenylmethane-

sulfonyl fluoride; PTM, post-translational modification; TCEP, tris (2-carboxyethyl) phosphine hydrochloride

## REFERENCES

- (1) Muñoz, J.; Heck, A. J. R. From the human genome to the human proteome. *Angew. Chem., Int. Ed.* **2014**, *53*, 10864–10866.
- (2) Segura, V.; Medina-Aunon, J. A.; Mora, M. I.; Martínez-Bartolomé, S.; Abian, J.; Aloria, K.; Antúnez, O.; Arizmendi, J. M.; Azkargorta, M.; Barceló-Batlloir, S.; et al. Surfing transcriptomic landscapes. A step beyond the annotation of chromosome 16 proteome. *J. Proteome Res.* **2014**, *13*, 158–172.
- (3) Ansong, C.; Purvine, S. O.; Adkins, J. N.; Lipton, M. S.; Smith, R. D. Proteogenomics: Needs and roles to be filled by proteomics in genome annotation. *Briefings Funct. Genomics Proteomics* **2008**, *7*, 50–62.
- (4) Jensen, O. N. Interpreting the protein language using proteomics. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 391–403.
- (5) Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63.
- (6) Toung, J. M.; Morley, M.; Li, M.; Cheung, V. G. RNA-sequence analysis of human B-cells. *Genome Res.* **2011**, *21*, 991–998.
- (7) Clamp, M.; Fry, B.; Kamal, M.; Xie, X.; Cuff, J.; Lin, M. F.; Kellis, M.; Lindblad-Toh, K.; Lander, E. S. Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 19428–19433.
- (8) Woo, S.; Cha, S. W.; Merrihew, G.; He, Y.; Castellana, N.; Guest, C.; Maccoss, M.; Bafna, V. Proteogenomic database construction driven from large scale RNA-seq data. *J. Proteome Res.* **2014**, *13*, 21–28.
- (9) Renuse, S.; Chaerkady, R.; Pandey, A. Proteogenomics. *Proteomics* **2011**, *11*, 620–630.
- (10) Nilsson, C. L.; Berven, F.; Selheim, F.; Liu, H.; Moskal, J. R.; Kroes, R. A.; Sulman, E. P.; Conrad, C. A.; Lang, F. F.; Andrés, P. E.; et al. Chromosome 19 annotations with disease speciation: A first report from the global research consortium. *J. Proteome Res.* **2013**, *12*, 135–150.
- (11) Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global quantification of mammalian gene expression control. *Nature* **2011**, *473*, 337–342.
- (12) Nesvizhskii, A. I. Proteogenomics: concepts, applications and computational strategies. *Nat. Methods* **2014**, *11*, 1114–1125.
- (13) Castellana, N.; Bafna, V. Proteogenomics to discover the full coding content of genomes: A computational perspective. *J. Proteomics* **2010**, *73*, 2124–2135.
- (14) Krug, K.; Nahnsen, S.; Macek, B. Mass spectrometry at the interface of proteomics and genomics. *Mol. Biosyst.* **2011**, *7*, 284–291.
- (15) Jacob, F.; Goldstein, D. R.; Fink, D.; Heinzelmann-Schwarz, V. Proteogenomic studies in epithelial ovarian cancer: established knowledge and future needs. *Biomarkers Med.* **2009**, *3*, 743–756.
- (16) Haider, S.; Pal, R. Integrated analysis of transcriptomic and proteomic data. *Curr. Genomics* **2013**, *14*, 91–110.
- (17) Cox, B.; Kislinger, T.; Emili, A. Integrating gene and protein expression data: Pattern analysis and profile mining. *Methods* **2005**, *35*, 303–314.
- (18) Legrain, P.; Aebersold, R.; Archakov, A.; Bairoch, A.; Bala, K.; Beretta, L.; Bergeron, J.; Borchers, C. H.; Corthals, G. L.; Costello, C. E. et al. The human proteome project: current state and future direction. *Mol. Cell. Proteomics* **2011**, *10*, 10.1074/mcp.M111.009993.
- (19) Paik, Y. K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Marko-Varga, G.; Aebersold, R.; Bairoch, A.; Yamamoto, T.; Legrain, P.; Lee, H. J.; et al. Standard guidelines for the chromosome-centric human proteome project. *J. Proteome Res.* **2012**, *11*, 2005–2013.
- (20) Segura, V.; Medina-Aunon, J. A.; Guruceaga, E.; Gharbi, S. I.; González-Tejedo, C.; San Chez Del Pino, M. M.; Canals, F.; Fuentes, M.; Casal, J. I.; Martínez-Bartolomé, S.; Elortza, F.; Mato, J. M.; Arizmendi, J. M.; Abian, J.; Oliveira, E.; Gil, C.; Vivanco, F.; Blanco, F.; Albar, J. P.; Corrales, F. J.; et al. Spanish human proteome project: Dissection of chromosome 16. *J. Proteome Res.* **2013**, *12*, 112–122.
- (21) Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006**, *1*, 2856–2860.
- (22) Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2007**, *2*, 1896–1906.
- (23) Olsen, J. V.; de Godoy, L. M. F.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4*, 2010–2021.
- (24) Reiter, L.; Claassen, M.; Schrimpf, S. P.; Jovanovic, M.; Schmidt, A.; Buhmann, J. M.; Hengartner, M. O.; Aebersold, R. Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol. Cell. Proteomics* **2009**, *8*, 2405–2417.
- (25) Prieto, G.; Aloria, K.; Osinalde, N.; Fullaondo, A.; Arizmendi, J. M.; Matthiesen, R. PAnalyzer: a software tool for protein inference in shotgun proteomics. *BMC Bioinf.* **2012**, *13*, 288.
- (26) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551–3567.
- (27) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
- (28) Brosch, M.; Yu, L.; Hubbard, T.; Choudhary, J. Accurate and sensitive peptide identification with mascot percolator. *J. Proteome Res.* **2009**, *8*, 3176–3181.
- (29) Carvalho, B. S.; Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **2010**, *26*, 2363–2367.
- (30) Carvalho, B. Platform design info for Affymetrix HuGene-1\_0-st-v1. R package version 3.8.0.
- (31) Dai, M.; Wang, P.; Boyd, A. D.; Kostov, G.; Athey, B.; Jones, E. G.; Bunney, W. E.; Myers, R. M.; Speed, T. P.; Akil, H.; Watson, S. J.; Meng, F. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* **2005**, *33*, E175.
- (32) Durinck, S.; Spellman, P. T.; Birney, E.; Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **2009**, *4*, 1184–1191.
- (33) Dennis, G.; Sherman, B. T.; Hosack, D. A.; Yang, J.; Gao, W.; Lane, H. C.; Lempicki, R. A. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **2003**, *4*, P3.
- (34) Fontanillo, C.; Nogales-Cadenas, R.; Pascual-Montano, A.; de Las Rivas, J. Functional analysis beyond enrichment: Non-redundant reciprocal linkage of genes and biological terms. *PLoS One* **2011**, *6*, e24289.
- (35) Gentleman, R. C.; Carey, V. J.; Bates, D. M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **2004**, *5*, R80.
- (36) Williams, S. C.; Winter, G. Cloning and sequencing of human immunoglobulin V lambda gene segments. *Eur. J. Immunol.* **1993**, *23*, 1456–1461.
- (37) Gabaldón, T.; Huynen, M. A. Shaping the mitochondrial proteome. *Biochim. Biophys. Acta, Bioenerg.* **2004**, *1659*, 212–220.
- (38) Huber, L. A.; Pfaller, K.; Vietor, I. Organelle proteomics: Implications for subcellular fractionation in proteomics. *Circ. Res.* **2003**, *92*, 962–968.
- (39) Deeb, S. J.; D'Souza, R. C. J.; Cox, J.; Schmidt-Supprian, M.; Mann, M. Super-SILAC Allows Classification of Diffuse Large B-cell Lymphoma Subtypes by Their Protein Expression Profiles. *Mol. Cell. Proteomics* **2012**, *11*, 77–89.
- (40) Coiffier, B.; Lepage, E.; Briere, J.; Herbrecht, R.; Tilly, H.; Bouabdallah, R.; Morel, P.; Van Den Neste, E.; Salles, G.; Gaulard, P. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **2002**, *346*, 235–242.

(41) Vizcaíno, J.; Deutsch, E.; Wang, R.; Csordas, F.; Reisinger, F.; et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **2014**, *32*, 223–226.





**La caracterización de distintos  
compartimentos madurativos de células B  
muestra un importante dinamismo en las  
vías de señalización durante la  
diferenciación B dependiente de antígeno**

5

Paula Díez<sup>1,2</sup>, Martín Pérez-Andrés<sup>1</sup>, Martin Bøgsted<sup>3</sup>,  
Mikel Azkargorta<sup>4</sup>, Rodrigo García-Valiente<sup>2</sup>, Rosa M. Dégano<sup>2</sup>,  
Elena Blanco<sup>1</sup>, Santiago Santa Cruz<sup>5</sup>, Félix Elortza<sup>4</sup>,  
Karen Dybkaer<sup>3</sup>, Alberto Orfao<sup>1</sup> and Manuel Fuentes<sup>1,2</sup>

<sup>1</sup> Servicio General de Citometría-NUCLEUS; Centro de Investigación del  
Cáncer (IBMCC-CSIC/USAL) e IBSAL, Universidad de Salamanca.

<sup>2</sup> Unidad de Proteómica; Centro de Investigación del Cáncer (IBMCC-  
CSIC/USAL) e IBSAL, Universidad de Salamanca.

<sup>3</sup> Department of Haematology, Aalborg University Hospital, Denmark.

<sup>4</sup> Proteomics Platform, CIC bioGUNE, Bizkaia Technology Park, Bizkaia, Spain.

<sup>5</sup> Servicio de Otorrinolaringología y Patología Facial Cervical.  
Hospital Universitario de Salamanca, Salamanca, España.

*en preparación*

*La información suplementaria correspondiente a este manuscrito  
está disponible, en formato electrónico, en el CD-ROM anexo al final  
de la presente tesis doctoral.*



**Introducción.** Las etapas finales de la diferenciación de células B dependiente de Ag tienen lugar en los órganos linfoides secundarios. Allí, las células B naive que han salido de la MO migran hacia los CG donde tras el reconocimiento antigénico se convierten en CB y CC y sufren los procesos de hipermutación somática y cambio de isotipo de IGH. Como resultado de estos mecanismos, se generan CP y linfocitos B de memoria, siendo las células centrales y efectoras de la respuesta humoral. Hasta el momento, existen numerosos estudios genómicos y transcriptómicos que han evaluado los cambios que ocurren durante la diferenciación de células B, desde su origen en la MO hasta la formación de células B secretoras de Ig tras los procesos de maduración de afinidad. Sin embargo, la información derivada de la caracterización proteómica cuantitativa de la diferenciación B dependiente de Ag sigue siendo muy limitada, especialmente en lo que se refiere a cambios en los perfiles proteicos intracelulares que pueden resultar claves en el desarrollo del proceso de diferenciación de una célula B inmadura a una célula B terminalmente diferenciada.

**Objetivo.** En este estudio nos propusimos evaluar el proteoma cuantitativo global de distintas poblaciones celulares B en distintos estadios de diferenciación Ag-dependiente (células naive, CB, CC, células B de memoria y CP). Asimismo, de acuerdo con el resto de trabajos incluidos en la presente tesis doctoral, se integraron los resultados proteómicos obtenidos con datos transcriptómicos derivados del estudio de dichas poblaciones celulares B y obtenidos previamente.

**Muestras, materiales y métodos.** En conjunto, en este trabajo se incluyeron 5 muestras de amígdalas humanas. En cada una de ellas, se purificaron 5 poblaciones de células B en distinto estadio de diferenciación - células naive, CB, CC, células B de memoria y CP - empleando citometría de flujo. Cada población celular fue procesada para la purificación de su contenido (total) proteico mediante incubación con una solución de lisis celular rica en urea. Dichas muestras proteicas fueron separadas mediante electroforesis unidimensional (a excepción de las CP purificadas que fueron

procesadas en solución). A continuación, se llevó a cabo la digestión proteica con tripsina y los péptidos resultantes, una vez purificados, fueron analizados mediante una aproximación LC-MS/MS libre de marcaje, realizada con un equipo Orbitrap XL. Los resultados obtenidos fueron evaluados cuantitativamente usando el programa informático Progenesis (versión 3.0). Los datos transcriptómicos, resultantes de un análisis mediante microarrays de expresión en las mismas poblaciones de células B purificadas, fueron obtenidos de la base de datos GEO (GSE69033).

**Resultados.** El análisis cualitativo de las proteínas contenidas en cada una de las 5 poblaciones subcelulares B analizadas (células naive, CB, CC, células B de memoria y CP) en cada una de las 5 muestras estudiadas mostró la existencia de un elevado grado de solapamiento de los perfiles proteicos de cada una de las citadas poblaciones celulares. De tal forma, alrededor del 95% de las proteínas totales identificadas en las CP y ~75% de las detectadas en CB, CC, células naive y células B de memoria fueron simultáneamente descritas en los 5 tipos celulares. A pesar de ello, se detectaron para cada subpoblación de células B proteínas exclusivas que no se expresaron en ningún otro tipo celular estudiado. De estas, cabe señalar, entre otras proteínas: i) VAV1 y VAV2, expresadas en CB, y que constituyen proteínas con un rol inductor de la formación de CP, lo que podría corroborar la formación de dichas CP a partir de células del CG; ii) RGS13, proteína exclusivamente identificada en CC que está relacionada con la quimiotaxis y la limitación en la expansión de las células del CG; iii) IRF4, característica de las CP y que determina el carácter secretor (de Ig) de dichas células; iv) ABL1, típica también de las CP, y que tiene un papel relacionado con la supervivencia de dichas células; y v) LRBA, presente en células B de memoria y que está directamente relacionada con la promoción de la proliferación, la expansión clonal y la supervivencia de células secretoras de Ac.

Por otra parte, la evaluación cuantitativa del proteoma presente en cada población celular B analizada (por limitaciones técnicas las CP no pudieron ser evaluadas cuantitativamente) mostró variaciones significativas ( $p$ -valor $<0.05$ ) en 753

proteínas, distinguiéndose dos grupos de células B con perfiles proteicos similares: CB y CC por una parte, y células B de memoria y naive por otra. De entre ellas, las células naive fueron las que presentaron niveles de expresión proteica más bajos de forma global, exhibiendo únicamente niveles de expresión más elevados con respecto a los otros tipos celulares en proteínas relacionadas con la replicación de ADN. Por otra parte, los análisis funcionales de los perfiles proteicos correspondientes a las distintas células estudiadas mostraron que las células de memoria presentaban un enriquecimiento en proteínas asociadas al procesamiento y presentación de Ag, mientras que en las CP predominaban perfiles de proteínas asociados a rutas metabólicas, lo que podría relacionarse con el elevado gasto energético realizado por estas células a causa de la producción y secreción de Ig. Además, los perfiles proteicos de las células del CG (CB y CC) incluían proteínas de la ruta de señalización de las quimiocinas y de los procesos de fagocitosis mediada por receptor, entre otras.

Finalmente, el análisis comparativo del proteoma y los datos transcriptómicos mostró un incremento progresivo en la correlación existente entre la expresión génica y la contrapartida proteica conforme avanzaba la diferenciación de las células B estudiadas. De tal forma, una correlación relativamente baja entre ambos sets de datos fue detectada en las células naive, mientras que en el resto de poblaciones dicha correlación fue, progresivamente, más elevada desde las células del CG a las células B de memoria.

**Conclusiones.** En resumen, nuestros resultados establecen por primera vez los perfiles proteicos de células B a lo largo de distintas fases de la maduración dependiente de Ag en órganos linfoides secundarios. El análisis cuantitativo de la expresión diferencial de dichas proteínas en las distintas subpoblaciones B analizadas muestra que las mayores diferencias existentes entre dichas subpoblaciones son debidas, principalmente, a diferencias en los niveles de expresión de proteínas, más que a la presencia/ausencia de las mismas.

\* Los resultados proteómicos están depositados en ProteomeXchange (PXD006191).

# Discusión





La LLC es una enfermedad clínicamente muy heterogénea en la que mientras algunos pacientes presentan un curso agresivo que requiere tratamiento, en otros la enfermedad tiene un comportamiento indolente y estable durante décadas<sup>46</sup>. Esta heterogeneidad clínica de la LLC se ve reflejada también en una elevada diversidad genética y molecular, sin que se haya podido identificar hasta la fecha una alteración común capaz de explicar la transformación neoplásica y/o el comportamiento maligno de la enfermedad. Aun así, en la actualidad disponemos de un amplio abanico de marcadores clínicos<sup>69-73</sup> y biológicos<sup>67,68</sup> útiles a la hora de realizar una estratificación pronóstica de los pacientes ya desde el momento del diagnóstico, contribuyendo a definir en cada caso la estrategia terapéutica más adecuada<sup>45</sup>.

Dentro de los factores biológicos, merecen especial atención los marcadores genéticos, principalmente los derivados de la evaluación del estado mutacional de los genes IGHV<sup>79,80</sup> y de las alteraciones citogenéticas<sup>87</sup> - sobre todo aquellas que afectan a los cromosomas 13 [como la del(13q) con o sin delección del gen *RB1*]<sup>282,283</sup>, 11 [del(11q), gen *ATM*]<sup>284-286</sup>, 17 [del(17p), gen *TP53*]<sup>287,288</sup> y 12 (trisomía)<sup>289-291</sup> -, además de las mutaciones descritas más recientemente, en especial las mutaciones del gen *NOTCH1*<sup>92,94</sup>. En cierta medida, los avances recientes en el conocimiento de la célula tumoral de la LLC a nivel molecular son reflejo del desarrollo de las técnicas de secuenciación genómica masiva y de microarrays de expresión<sup>93,292</sup>, cuya aplicación en el estudio de la LLC ha permitido una caracterización extensa y pormenorizada, especialmente a nivel génico, de las células B tumorales; con ello, se han logrado avances en el diagnóstico, la estratificación pronóstica y el seguimiento de los pacientes con LLC. En paralelo, en la LLC se han identificado también algunos factores fenotípicos de pronóstico adverso, como la expresión de las proteínas CD49d, ZAP70 y CD38<sup>102-106</sup>. Estos últimos hallazgos, junto con los estudios inmunofenotípicos requeridos para el correcto diagnóstico de la enfermedad, suponen un punto de partida para la incorporación del análisis de proteínas concretas en la estratificación pronóstica de la LLC. Sin embargo, en el momento de iniciar este trabajo doctoral no existían estudios

pormenorizados que abordasen la caracterización masiva y simultánea del conjunto de todas las proteínas – proteoma - expresadas en las células tumorales de la LLC.

Tratándose la LLC de una enfermedad heterogénea tanto a nivel clínico como molecular, la valoración global (y conjunta) de los múltiples componentes moleculares que integran la célula tumoral adquiere especial relevancia. Esto es debido, en parte, a que es raro que una única proteína (al igual que una única mutación) pueda ser considerada como la causante única y directa de la enfermedad, y a que habitualmente se requiere de la alteración de un (extenso) entramado de conexiones moleculares intra- y extra-celulares para el desarrollo de un proceso neoplásico. Asimismo, a pesar del poder predictivo que presentan los marcadores génicos, los efectores finales responsables de las funciones celulares alteradas/patogénicas son las proteínas, lo que les convierte en elementos imprescindibles para conocer con precisión el papel que realmente juegan las alteraciones genéticas. En este sentido, es importante conocer no sólo qué proteínas se expresan en una célula y su cantidad, sino también su estado (p.ej. fosforilado/desfosforilado, u otras PTMs) que puede modificar e influir de forma decisiva en su comportamiento y/o localización, al desencadenar la activación o inhibición de la activación de vías de señalización concretas que determinan la actividad final de la célula. El conocimiento proteómico de la célula tumoral de la LLC no solamente permite avanzar en el conocimiento de su patogenia (incluyendo el conocimiento también del microambiente tumoral), sino que además puede abrir nuevas perspectivas terapéuticas al proporcionar una visión más amplia del proteoma celular de la LLC, que puede facilitar el diseño y desarrollo de tratamientos de acción múltiple dirigidos a proteínas/vías de señalización alteradas.

En la actualidad, el origen celular de la LLC sigue siendo, en gran medida, desconocido, así como los cambios que se suceden a lo largo de la ontogenia de la célula B normal que llevan a su transformación maligna<sup>122</sup>. Si bien se han planteado varios modelos teóricos, hasta la fecha<sup>123,126,127</sup> ninguno de ellos ha sido capaz de esclarecer por completo el origen de la enfermedad y su elevada heterogeneidad. En

este campo, el estudio proteómico detallado de la célula tumoral podría proporcionar también información útil a la hora de determinar con mayor precisión las etapas claves en los procesos de transformación neoplásica y progresión maligna de la LLC.

Por todo lo anteriormente expuesto, entendemos que la metodología proteómica actual podría complementar la caracterización genética/genómica de la LLC, pudiendo la integración de la información derivada de ambos tipos de estudios ofrecer una visión más fidedigna de la enfermedad que pudiera traducirse en un abordaje más adecuado del diagnóstico, estratificación pronóstica, tratamiento y monitorización de la LLC. Así, en este trabajo doctoral nos planteamos como objetivo central la caracterización del proteoma de la célula tumoral de la LLC frente al de las célula B normales, integrando además esa información proteómica con información derivada de los estudios transcriptómicos.

**Caracterización sistemática y cuantitativa del proteoma y fosfoproteoma de la célula B tumoral de la LLC y de la LMB: determinación del papel de la vía de señalización a través del receptor de la célula B específico de Ag.**

Tal como hemos comentado anteriormente en esta discusión, entendemos que la caracterización proteómica global y sistemática de la célula B de la LLC podría tener una elevada relevancia en el mejor conocimiento de los procesos de transformación neoplásica y progresión maligna de la enfermedad, al asociarse la mayoría de los tumores humanos con alteraciones de la regulación de redes proteicas, fundamentalmente de aquellas que están relacionadas con vías de señalización intracelular mediada por quinasas<sup>293,294</sup>, y que se considera podrían constituir también dianas terapéuticas<sup>48,295,296</sup>. Asimismo, los procesos celulares de diferenciación, reconocimiento antigénico, proliferación y maduración celular B se encuentran regulados de forma estrecha por los niveles de fosforilación de distintas proteínas implicadas en vías de señalización intracelular relacionadas con el microambiente

tumoral, las interacciones célula-célula y célula-matriz extracelular, y la dinámica del citoesqueleto, entre otras funciones<sup>31,297</sup>. Por tanto, para ahondar en el conocimiento de los mecanismos patogénicos tumorales sería fundamental, además del análisis (simultáneo y masivo) del proteoma global, el estudio de los perfiles de expresión de fosfoproteínas de la célula tumoral de la LLC. Esto permitiría generar un mapa global de las vías de señalización alteradas y, por lo tanto, crear una referencia útil para clasificar la enfermedad de una forma más precisa y efectiva<sup>298</sup>. Sin embargo, en la actualidad, la caracterización fosfoproteómica de las células tumorales constituye una tarea difícil debido a la complejidad inherente de las redes proteicas, el amplio rango dinámico de los fosfopéptidos, su estequiometría variable y la elevada heterogeneidad de sus niveles en distintas células<sup>207,298</sup>. Con la finalidad de realizar una caracterización proteómica global de la célula tumoral de la LLC, en este primer apartado del trabajo doctoral empleamos una aproximación basada en MS, combinada con un paso de pre-enriquecimiento selectivo y específico mediante Acs anti-fosfotirosina (p-Tyr) conjugados en superficies esféricas; con ello obtuvimos, por primera vez, un mapa cuantitativo de las fosfoproteínas (y de las proteínas totales) expresadas en dichas células tumorales, sin utilizar marcajes isotópicos<sup>236</sup>. Merece destacar que en este capítulo del trabajo incluimos también el estudio de las células de LMB con el fin de evaluar de forma comparativa y preliminar su proteoma y fosfoproteoma, respecto al de la LLC.

En conjunto, el análisis realizado a partir del contenido proteico global de células tumorales purificadas de LLC y LMB permitió identificar 13.504 péptidos únicos, correspondientes a 2.970 proteínas distintas. De todas las proteínas identificadas, sólo 2% correspondieron a proteínas sin evidencia previa a nivel proteómico (PE 2-4), mientras que 10% se encontraron fosforiladas. Aunque las muestras estudiadas eran heterogéneas en cuanto a las alteraciones citogenéticas, el estado mutacional de los genes IGHV y el estadio clínico de la enfermedad, observamos un solapamiento del 73% en el conjunto de proteínas presentes en las distintas muestras analizadas. Dichas

proteínas estaban relacionadas principalmente con funciones celulares genéricas entre las que se incluían las funciones asociadas a procesos metabólicos y catabólicos, el ciclo celular, o la adhesión célula-célula, además de funciones específicas de la célula B (p.ej. presentación y procesamiento de Ags, inmunidad humoral, sistema del complemento y señalización mediada por el BCR). A pesar de ello, observamos también como algunos grupos concretos de proteínas permitían diferenciar y subclasificar distintos subgrupos de muestras, independientemente del estadio clínico de la enfermedad o las alteraciones genéticas y el estado mutacional de los genes IGHV de la célula tumoral. Entre otras proteínas, estos grupos incluían proteínas relacionadas con la actividad del proteasoma (PSMD1, PSMD9, PSMA1), mecanismos de regulación inmune (IgHM, CD5, CD47, CD48, CD53, CD79B, B2M, ZAP70), moléculas HLA (-A, -C, -DMB, -DRB1, -E), y proteínas asociadas a señalización vía BCR (SYK, LSP1, MYCBP, BLNK, ATM), la presentación antigénica, y la síntesis y expresión proteica. Aunque no se observaron patrones específicos de expresión de proteínas para las distintas alteraciones genéticas presentes en las células analizadas, las proteínas codificadas en las regiones delecionadas en cada muestra estudiada se encontraban cuantitativamente disminuidas. Tal fue el caso de los niveles de expresión de ATM y CUL5, cuantitativamente menores en la muestra que presentaba del(11q) vs. el resto de muestras, o de la proteína RB1, con niveles disminuidos en la muestra que presentaba del(13q14).

Otras asociaciones que nos parece importante mencionar incluyen la expresión de DYNL1 (una proteína con capacidad de neutralizar la actividad anti-apoptótica de BCL2 y favorecer la expansión de las células B leucémicas<sup>299,300</sup>) y los estadios intermedios de la enfermedad, la expresión de RFWD3 (un regulador positivo de la estabilidad de TP53) restringida a células tumorales con alteraciones genéticas de riesgo bajo/intermedio – p.ej. del(13q14), del(13q14.3), del(14q32) y trisomía 12 -, y la expresión diferencial de PIK3R4 en las células de pacientes con LLC de mal pronóstico (estadio C/IV, IGHV no mutado). Esta última proteína es un elemento clave en la

iniciación del proceso de autofagia y ha sido descrita por Kristensen *et al.*<sup>301</sup> como un biomarcador de pronóstico adverso en la LLC.

A su vez, las células de la LMB presentaron niveles de expresión elevados de la proteína supresora tumoral ARIH2. Dicha proteína es bloqueada por PTPN11, la cual presentó en el caso de la LMB niveles de expresión 20 veces inferiores a los detectados en las células de la LLC, en ausencia de fosforilación. Estos hallazgos sugieren que la forma fosforilada de PTPN11 podría bloquear la expresión de ARIH2 y, por ende, anular su efecto supresor leucémico en las células de la LLC<sup>302</sup>. Por el contrario, en la LMB, los niveles elevados de ARIH2 permitirían bloquear el desarrollo y progresión de la LLC, permaneciendo las células sensibles a la apoptosis. Además, identificamos un total de 293 proteínas que, pese a estar presentes de forma recurrente en las células tumorales de la LLC, no fueron detectadas en las células de la LMB; entre otras, estas proteínas incluían VAV1, BCL2, ZAP70 y CD53, proteínas clave en la adhesión de las células tumorales, la señalización vía BCR, y la supervivencia de la célula tumoral<sup>303-305</sup>.

En paralelo, el análisis del fosfoproteoma (proteínas fosforiladas en Tyr, Thr y Ser) de las células tumorales de la LLC/LMB permitió identificar, en el conjunto de las muestras analizadas, 594 fosfopéptidos correspondientes a un total de 327 fosfoproteínas distintas. Entre estos 594 fosfopéptidos, se incluyeron un total de 329 fosfopéptidos descritos por primera vez en este estudio y que no habían sido referidas en la última actualización disponible de la base de datos neXtProt (versión de febrero de 2017). Por otra parte, 69 de las 327 fosfoproteínas identificadas no habían sido detectadas en el estudio proteómico convencional realizado en paralelo sobre las mismas muestras, lo que apoya la utilidad de las técnicas de inmunoenriquecimiento para incrementar la sensibilidad analítica de las técnicas proteómicas. A diferencia de lo observado en el proteoma global, las células tumorales únicamente compartían una fracción relativamente pequeña (<20%) del fosfoproteoma, lo que posiblemente refleja un mayor dinamismo del fosfoproteoma y las fosfoproteínas. Dentro del fosfoproteoma común a todas las muestras analizadas se incluían proteínas implicadas

en la fosforilación (PRKCB, LYN, SYK, ATM, BLK, JAK2), el transporte intracelular de proteínas (NSF, CUL3) y la transducción de señales a través del BCR, receptores de quimiocinas y TLRs<sup>36,306,307</sup>. Merece destacar que la expresión de fosfoproteínas implicadas en la señalización de receptores de quimiocinas y de TLRs estaba restringida a CXCR4, STAT3 y TLR1. En este sentido, podríamos concluir que la representación de estas vías de señalización en el proteoma de las células leucémicas es claramente inferior a la de la vía de señalización de BCR, para la que se detectaron un gran número de proteínas fosforiladas (LYN, SYK, PI3K, BTK, ZAP70, PLCG3, ERK, NFAT, CD19, PRKCB, NFKB, JNK y VAV)<sup>158,308-310</sup>. Resulta interesante también destacar que células de pacientes en el mismo estadio clínico (Binet/Rai) de la enfermedad (*i.e.* C/IV), aunque con características citogenéticas [a excepción de la del(13q14)] y estados mutacionales distintos, compartieron alrededor del 50% de sus proteínas fosforiladas (p.ej. BLK, CD19, PLCG2, SCIMP), siendo además éstas detectadas únicamente en células de LLC y no en LMB. A su vez, las células de pacientes con LLC (en estadio avanzado C/IV) y múltiples alteraciones citogenéticas [del(13q14), del(14q32), del(11q22.3), del(11q23.3)] en ausencia de mutaciones en los genes IGHV, expresaban de forma exclusiva 10 fosfoproteínas (p.ej. JAK1, PPA2, SRRT), mientras que células con IGHV mutado y del(13q14) se observaba expresión exclusiva de otras 49 fosfoproteínas distintas (p.ej. DOCK11, DOK2, ILF3, LSP1, PTPN18). Finalmente, las células de LLC no mutadas de riesgo intermedio y del(13q14.3) expresaron de forma exclusiva las proteínas fosforiladas HERC1, NUDT3 y PSMD9.

Tal como ocurrió con el proteoma global, el análisis del fosfoproteoma de las células tumorales mostró la existencia de fosforilación incrementada en la LLC vs. LMB de proteínas implicadas en las vías de señalización de NF- $\kappa$ B y STAT3, relacionadas con la activación, proliferación, supervivencia, adhesión y migración celulares<sup>311</sup>, sugiriendo la implicación de las vías de NF- $\kappa$ B y STAT3 en la transformación maligna de la enfermedad.

En conjunto, nuestros resultados han servido para establecer el primer mapa global del proteoma y del fosfoproteoma de la LLC, contribuyendo a la identificación de vías de señalización candidatas a estar implicadas en los mecanismos moleculares asociados al origen y progresión de la enfermedad, así como a determinar el comportamiento heterogéneo de la misma. No obstante, debido al número limitado de muestras analizadas hacen falta estudios futuros en series más amplias de pacientes que confirmen los hallazgos preliminares de este trabajo.

### **Evaluación global de los perfiles proteicos y vías de señalización intracelular alteradas en la célula B de la LLC y su relación con las alteraciones citogenéticas y moleculares típicas de la enfermedad.**

Como resultado del análisis global de los proteomas y fosfoproteomas de la LLC reflejados en el apartado anterior, quedó en evidencia la existencia de numerosas alteraciones en la expresión de múltiples proteínas y fosfoproteínas en las células B de pacientes con LLC y LMB, y potencialmente también de las vías de señalización intracelular asociadas a las mismas. Estudios genómicos y transcriptómicos previos han referido de forma recurrente distintas alteraciones genéticas típicas de la LLC<sup>93,94,111,292,312</sup>. Entre otros genes, los que habitualmente muestran con mayor frecuencia mutaciones somáticas en la LLC son: *NOTCH1*, *XPO1*, *MYD88*, *KLHL6*, *SF3B1*, *BIRC3*, *POT1* y *TP53*<sup>93,97,98,313–315</sup>. Dichas mutaciones generan modificaciones en las cascadas de señalización implicadas en la formación de neoplasias y su transformación maligna, afectando especialmente a los mecanismos de supervivencia y proliferación celulares. Así, estudios recientes en muestras de pacientes con LLC y mutación de *NOTCH1* demuestran la existencia de alteraciones, tanto en la vía de señalización de *NOTCH1* como en rutas metabólicas de fosforilación oxidativa y glicolisis/gluconeogénesis<sup>316</sup>. De forma análoga, las alteraciones en el gen *MYD88* podrían resultar en modificaciones de las vías de señalización relacionadas con la IL-1



y los TLRs<sup>317</sup>. En un estudio reciente de Landau *et al.*<sup>94</sup>, se identifican como las principales vías de señalización afectadas por genes mutados en la LLC las vías de señalización de NOTCH1, vías inflamatorias, la vía de señalización de Wnt, las vías de control del ciclo celular y de la respuesta a daño en el ADN, rutas de procesamiento de ARN, modificación de la cromatina y de señalización mediada por el BCR. Precisamente, las alteraciones en esta última vía pueden desencadenar también alteraciones importantes en otras cascadas asociadas a la misma, como la vía de señalización de NF- $\kappa$ B<sup>318</sup>, la ruta de las MAPK (asociadas a la activación de PKC)<sup>319</sup>, de PI3K<sup>320</sup> y de c-MYC<sup>321</sup>. No obstante, todos estos estudios, centrados en la información génica y transcriptómica de la LLC, únicamente reflejan una parte de las alteraciones presentes en las células patológicas al no incidir sobre el impacto de dichas mutaciones a nivel proteico<sup>322</sup>. Así, en la actualidad es bien conocido que la relación existente entre los niveles de ARNm y proteína es, para muchos genes, prácticamente nula, tal como se ha demostrado de forma específica en algunos trabajos<sup>323,324</sup>. Teniendo en consideración estos hallazgos, y una vez descritos el proteoma y fosfoproteoma de la célula tumoral de la LLC, nuestro objetivo se centró en la identificación y análisis de los posibles patrones de expresión diferencial (global) de un amplio conjunto de 224 proteínas, caracterizadas por estar implicadas en las principales vías de señalización intracelular afectadas en el desarrollo de distintas neoplasias malignas, incluyendo la LLC. Con este fin, analizamos una serie más amplia, aunque ciertamente aún limitada, de pacientes con LLC que incluía una representación de casos con diferente estado mutacional para *NOTCH1* y los genes IGHV, y con distintas alteraciones citogenéticas – p.ej. del(13q), del(17p) y trisomía 12 –; para su estudio proteómico empleamos una aproximación metodológica que combinaba el aislamiento selectivo de las proteínas citoplasmáticas de las células B de pacientes con LLC junto con el análisis de la posible expresión diferencial de las mismas entre distintos grupos de pacientes y/o sujetos sanos, basada en el uso de microarrays de Acs que permiten la determinación de la abundancia relativa de cada proteína.

Además, utilizamos ensayos LC-MS/MS para la validación, mediante contaje de espectros (valores emPAI), de los resultados obtenidos mediante los microarrays de Acs.

Globalmente, nuestros resultados mostraron la existencia de diferencias significativas en la abundancia relativa de un número importante de proteínas incluidas en el microarray de Acs en células B tumorales de pacientes con LLC con respecto a las células B normales de sujetos sanos. A modo de ejemplo, la proteína TRF1 - perteneciente al complejo enzimático encargado del mantenimiento de la longitud de los telómeros - se encontraba sobreexpresada en muestras de LLC vs. linfocitos B normales. Sin embargo, en los estudios transcriptómicos publicados previamente existe cierta controversia respecto a la posible sobreexpresión de esta proteína en LLC. Así, mientras Hoxha *et al.*<sup>325</sup> han detectado niveles elevados de transcrito de TRF1 en muestras de LLC, Poncet *et al.*<sup>326</sup> y Guièze *et al.*<sup>326</sup> han referido una expresión significativamente disminuida del mismo. Asimismo, la proteína  $\beta$ -sinucleína (conocida por su rol en el desarrollo de la enfermedad de Parkinson) se encontraba sobreexpresada en muestras de LLC vs. linfocitos B normales. Estudios previos han señalado la implicación de dicha proteína (junto con la versión  $\alpha$ -sinucleína) en la leucemia aguda de línea eritroide y megacariocítica<sup>327</sup>, por lo que nuestros resultados podrían evidenciar también la importancia de esta proteína en el desarrollo de la LLC. Otro caso controvertido afectaba a la ciclina D1, subexpresada en nuestro análisis en la LLC y con resultados contradictorios en estudios previos<sup>328</sup>. Estos hallazgos, asociados a una expresión significativamente reducida de proteínas relacionadas/implicadas en la apoptosis (p.ej. DIABLO, CASP4, CASP3, PAR4) en pacientes con LLC, podría reflejar (e incluso determinar) una mayor capacidad de supervivencia de las células tumorales<sup>329-331</sup> que se vería también apoyada por la sobreexpresión de JUN. De forma similar, nuestros resultados muestran también una expresión significativamente reducida de proteínas de la familia de PKC en la gran mayoría (~75%) de las muestras tumorales analizadas, independientemente de las

alteraciones moleculares, el estado mutacional y/o citogenético de las mismas. Estas proteínas se han relacionado con una progresión más lenta y un comportamiento indolente de la enfermedad<sup>332</sup>. Además, algunas proteínas asociadas a la señalización vía BCR, como la chaperona calnexina<sup>333</sup>, que se asocia con las proteínas CD79A y CD79B, y RAF1 fueron identificadas en niveles inferiores en la LLC respecto a las células B normales. En este sentido, la detección de niveles bajos de expresión relativa de dichas proteínas podría constituir uno de los posibles mecanismos asociados a un comportamiento evolutivo más agresivo de la enfermedad, asociadas o no a alteraciones génicas concretas. De acuerdo con esta hipótesis, merece destacar que los patrones de expresión proteica, en base a la abundancia relativa de múltiples proteínas, fueron relativamente similares entre células B normales y células de LLC que no presentaban ninguna alteración citogenética o que tenían trisomía 12 como única alteración, lo cual apoyaría la hipótesis que adelantábamos en el apartado anterior de esta discusión de que, en general, distintas alteraciones genéticas (p.ej. la trisomía 12) no confieren perfiles proteicos diferenciales claros en la LLC. En relación al estado mutacional de los genes IGHV, la comparación de perfiles proteicos de pacientes con genes IGHV mutados vs. no mutados tampoco mostró diferencias significativas en los perfiles proteicos globales de ambos grupos de células, a excepción de una disminución de la abundancia relativa de las proteínas PKC (como se mencionó anteriormente para la mayoría de las muestras de la LLC) y PTK2.

Pese a lo expuesto anteriormente, la expresión de algunas proteínas, como la proteína MCL-1, presentó asociación (niveles elevados en células de LLC) con la presencia de algunas alteraciones genéticas/moleculares – la del(17p) y mutación en *NOTCH1* – (vs. células B normales); de acuerdo con estos hallazgos, el incremento de MCL-1 podría estar relacionado con un fenotipo maligno debido al papel que juega esta proteína de la familia BCL-2 en la inhibición de la apoptosis y la estimulación de la supervivencia de las células tumorales<sup>334</sup>. En este mismo sentido, los pacientes que presentaron *NOTCH1* mutado se caracterizaron por mostrar disminución de la

expresión de un mayor número de proteínas que las demás LLC, entre las que se incluían la ciclina B1, PKC $\gamma$ ,  $\beta$ -sinucleína, calnexina, S100-B, CDK6, CDKN2A y TRF-1. Por otra parte, AP2 $\gamma$ , HSP90, fosfolipasa C $\gamma$ 1 y el receptor de fosfatidilserina presentaban niveles más elevados de expresión entre los subgrupos de LLC con alteraciones génicas de peor pronóstico. Asimismo, en estos casos, los procesos apoptóticos parecen estar balanceados al detectarse bajos niveles de CASP3 y DIABLO y niveles elevados de CASP10. Por otra parte, al comparar pacientes con *NOTCH1* germinal vs. las células B normales de individuos sanos observamos mayor expresión de CDK4 y CDK6 junto con niveles inferiores del inhibidor de CDKs p21 entre los pacientes, lo cual sugiere que pudiera existir un incremento en la actividad asociada a ciclo celular entre las células tumorales de este subgrupo de LLC<sup>335,336</sup>. En este sentido, cabe destacar también la regulación positiva de proteínas relacionadas con la apoptosis (CASP4, CASP7, CASP9), la vía de señalización de NF- $\kappa$ B y el receptor de EGF (EGFR) en pacientes con LLC sin mutación en *NOTCH1*, lo que pese al incremento del ciclo celular, sugiere mayor predisposición a la apoptosis y podría conferirles un mejor pronóstico. Dado que la señalización por la vía de NOTCH1 incrementa la activación de las vías de señalización de NF- $\kappa$ B y PI3K/AKT, la detección de niveles elevados de NF- $\kappa$ B era un hecho esperable<sup>337,338</sup>. De acuerdo con ello, existía también sobreexpresión de c-MYC en estos pacientes con mutación en *NOTCH1*, de forma que todo ello podría contribuir a explicar la asociación descrita previamente por otros autores entre la mutación de *NOTCH1* en LLC y un mayor riesgo de progresión, por ejemplo, a síndrome de Richter<sup>161,313,339</sup>.

Respecto a los pacientes que mostraban del(13q), nuestros resultados mostraron una asociación entre esta alteración genética y un incremento en la abundancia relativa de la proteína JUN; esta asociación podría determinar una mayor proliferación y grado de expansión del clon de células B tumorales en estos pacientes (vs. células B normales de sangre periférica) dirigida por JUN<sup>340</sup> y facilitada por la pérdida de RB1 y el aumento asociado a ésta de las ciclinas A, B1 y D1. En paralelo, estos pacientes

mostraron niveles de PRKC y de las proteínas pro-apoptóticas PAR4 y CASP3 notablemente reducidos, lo que podría traducirse en un mayor grado de inhibición de la apoptosis vía p38 MAPK, facilitando así también la expresión del clon B tumoral. Finalmente, otro hallazgo interesante fue la regulación negativa de RAF1 (proteína de señalización de la vía del BCR), la cual podría interpretarse como un mecanismo compensatorio de inactivación del ciclo celular a través de la vía de señalización de MEK/ERK, contrarrestada por la pérdida de RB1, el aumento de JUN y de las ciclinas A, B1 y D1.

La del(17p) constituye la alteración citogenética asociada a un peor pronóstico en la LLC. Dicha alteración, que afecta al gen supresor de tumores *TP53*, se asoció con alteraciones en las rutas de señalización de ERBB, fosfatidilinositol, apoptosis, migración leucocitaria trans-endorotelial y niveles elevados de *CDKN2A*, una proteína supresora tumoral (también conocida como p14arf) cuyo papel en la regulación del ciclo celular es clave al participar en las vías de señalización RB1-E2F1 y ARF-MDM2-TP53 induciendo la parada del ciclo celular en fase G2 y la subsecuente muerte por apoptosis (normalmente mediada por la vía de TP53), que por lo tanto, estarían inhibidas en pacientes que presentaban del(17q)<sup>341</sup>. Al igual que en los pacientes que presentaban del(13q), la presencia de del(17p) se asoció a: 1) regulación negativa de la señalización RAF/MEK que podría ser contrarrestada por la regulación positiva de JNK; 2) aumento del ciclo celular asociado a un incremento de los complejos ciclina A/CDK2-1 y ciclina D/CDK4 y la disminución del inhibidor de CDK, p21<sup>335</sup>; y 3) disminución de la apoptosis asociada a niveles anormalmente bajos de RAF1 y BCL-10 e incremento de las proteínas anti-apoptóticas BCL-XL y NF-κB. Todo ello contribuiría a explicar el pronóstico especialmente adverso de este subgrupo de pacientes con LLC.

Finalmente, merece destacar que el empleo de una metodología basada en MS con el fin de validar los resultados obtenidos mediante los microarrays de Acs demostró un elevado grado de correlación entre ambas aproximaciones para las proteínas y vías de señalización intracelular representadas en los microarrays de Acs, quedando

por lo tanto validada la tecnología de microarrays de Acs para el estudio de los niveles relativos de expresión de las proteínas analizadas.

En conjunto, los resultados obtenidos con los microarrays de Acs apoyan la importancia del estudio de los perfiles proteómicos de las células tumorales, a la hora de obtener una visión más fidedigna de los procesos que se encuentran activados/reprimidos en las células tumorales de pacientes con LLC, especialmente cuando se emplean simultáneamente métodos de evaluación cualitativos y cuantitativos. Si bien los microarrays de Acs ofrecen una visión dirigida (sólo se analizan las proteínas contenidas en la plataforma) y altamente específica del proteoma, la MS podría considerarse como una estrategia de análisis masivo que, aplicada de forma previa en una fase de investigación abierta (no dirigida), proporciona una herramienta complementaria a los microarrays de Acs con un elevado potencial para la caracterización del proteoma de sistemas celulares complejos.

### **Establecimiento de una metodología proteómica óptima para la secuenciación peptídica del BCR de células tumorales de pacientes con LLC y su relación con la información proporcionada por técnicas estándar de secuenciación genómica.**

El análisis global del proteoma y del fosfoproteoma de la LLC, así como de las vías de señalización alteradas, han puesto de manifiesto que, a pesar de la gran heterogeneidad encontrada, existe en común una activación de distintas proteínas/fosfoproteínas asociadas al BCR.

La señalización a través del BCR se considera una de las principales rutas de activación de las células B tanto normales como tumorales de la LLC<sup>28</sup>. Dicho BCR está configurado como una Ig anclada a membrana cuya capacidad de reconocimiento antigénico se forja a través de procesos de i) la recombinación de los segmentos VDJ de los genes de las Ig durante la maduración de la célula B<sup>342,343</sup>, ii) los mecanismos de hipermutación somática que ocurren durante la maduración de afinidad del

receptor<sup>344,345</sup> y iii) la selección específica de Ag que ocurre durante la fase de activación y tras la maduración de afinidad<sup>346</sup>. La versión soluble de las Ig puede explorarse, de forma sencilla, a través de técnicas proteómicas gracias a su fácil accesibilidad, al estar presente en cantidades significativas en diferentes líquidos corporales (p. ej. suero, saliva, fluido sinovial)<sup>347</sup>, su elevada estequiometría y su fácil aislamiento. Sin embargo, la contrapartida de las Igs ancladas a la membrana celular (i.e. los BCR) presenta una mayor dificultad para su análisis peptídico, debido principalmente a los problemas existentes durante el procesamiento y purificación requeridos para el análisis de proteínas de membrana, amplificado aún más cuando dichas proteínas (en este caso, el BCR) presentan una elevada variabilidad lo que dificulta encontrar bases de datos peptídicas completas y detalladas que permitan la identificación de todas las secuencias analizadas. Por estos motivos, se emplean tradicionalmente las técnicas de secuenciación genética NGS<sup>348</sup> como técnicas de referencia para caracterizar las secuencias genómicas de los BCR de células B normales y patológicas. Sin embargo, la descripción completa del receptor requiere de la información peptídica ya que ello permite, entre otras cosas, identificar con mayor precisión la especificidad antigénica del mismo, a la vez que establecer estrategias eficaces de bloqueo o expresión dirigida de determinados BCR<sup>349,350</sup>.

Dada la relevancia del BCR en la LLC, en este estudio nos propusimos evaluar tres estrategias proteómicas para la preparación y caracterización de las secuencias peptídicas del BCR mediante MS. Nuestro objetivo se centraba de forma preferente, en determinar qué aproximación sería la más adecuada a la hora de analizar una muestra compleja de proteínas derivadas de células B para identificar las secuencias de la Ig dentro del BCR, mediante un proceso relativamente sencillo y reproducible. Asimismo, diseñamos y desarrollamos un algoritmo sencillo para poder comparar de forma rápida y fácil la información genómica obtenida mediante secuenciación de ácidos nucleicos por técnicas de biología molecular, y los datos proteómicos resultantes del análisis de la Ig mediante MS. Con tal propósito, investigamos el uso de

(i) una técnica de purificación por afinidad de fragmentos Fab de Ig, (ii) una estrategia basada en la digestión de las Igs con papaína para la obtención de los fragmentos Fab, y (iii) un método basado en el uso de DTT para romper los puentes disulfuro que unen las dos cadenas ligeras y pesadas de las Ig presentes en las células B. Para los métodos (ii) y (iii), se realizó un paso previo de purificación de Igs mediante microesferas recubiertas de proteína G. Todos los péptidos resultantes de las tres aproximaciones anteriores se analizaron mediante MS.

En un primer momento, nuestros análisis se centraron en determinar el perfil proteico de referencia, por lo que procesamos una muestra de IgG humana (técnicamente pura) identificando 31 péptidos diferentes correspondientes a 20 proteínas distintas, es decir, que correspondían a distintas regiones de las Igs (diversas zonas constantes), las cadenas kappa, lambda, gamma y subtipos de cadenas gamma, pero también a proteínas diferentes de IgG como cadena IgH $\mu$ , IgA $\alpha_1$ , IgH $\alpha_2$ , IgH $\delta$ , IgH $\epsilon$ , proteína J y otras proteínas del sistema inmune presentes en la muestra. La identificación de estas mismas proteínas a partir de células B humanas mediante las 3 estrategias descritas reveló un 10%, 40% y 70% de cobertura por parte de los métodos (i), (ii) y (iii), respectivamente. Globalmente, nuestros resultados mostraron que el número de proteínas identificadas se incrementaba a medida que la estrategia de purificación se simplificaba. Así, identificamos 22, 54 y 171 proteínas mediante los métodos (i), (ii) y (iii), respectivamente; sin embargo, la capacidad de cada estrategia para permitir la identificación de proteínas específicas del BCR se distribuía de tal forma que los métodos (i) y (ii) presentaban los valores más elevados (6 y 16 proteínas, respectivamente, que suponen ~30% del total de proteínas identificadas relacionadas con el sistema inmune por cada uno de las estrategias), mientras que en el método (iii) dichas identificaciones se correspondían con 28 proteínas (de 171 proteínas totales, lo que supone solo el 16%). Por tanto, en base a lo anteriormente expuesto, deducimos que: a) es necesario un paso de enriquecimiento en Ig previo al análisis de la secuencia peptídica de las mismas, b) cuanto más simple es el procesamiento post-



enriquecimiento, menor es la pérdida de muestra y, por lo tanto, mayor es la tasa de éxito en la identificación proteica, pero c) mayor es también la contaminación por proteínas que no forman parte del BCR.

En relación a la distribución de las proteínas identificadas, cada aproximación presentó un patrón distinto. En general con las tres aproximaciones pudimos identificar de forma mayoritaria Ig (38%, 41%, y 21% del total de las proteínas identificadas para los métodos (i), (ii) y (iii), respectivamente). Esto se asoció a mayor contaminación progresiva por proteínas HLA (3%, 11% y 21% con los métodos (i), (ii) y (iii), respectivamente). El resto de proteínas y grupos de proteínas, constituían fracciones menos representadas y con una distribución similar dentro de cada una de las tres aproximaciones de enriquecimiento de Igs empleadas.

Asimismo, la identificación de péptidos (y de sus proteínas) comunes a las tres metodologías fue muy bajo; así sólo 15 péptidos de un total de 60, 98 y 426 péptidos identificados con los métodos (i), (ii) y (iii), respectivamente, eran comunes a las 3 estrategias. Como cabría esperar, la estrategia más sencilla (iii) que permitía identificar más proteínas, fue también la que permitió identificar un mayor número de péptidos únicos (*i.e.* 362 péptidos) asociados a proteínas de Ig, al BCR y a otros componentes del sistema inmune; sin embargo, tal y como se indicó anteriormente, el grado de especificidad de este método para proteínas del BCR fue más limitado que en el caso de los métodos (i) y (ii), que incluían un paso previo de enriquecimiento específico de Ig. Con estos resultados podemos concluir que la definición precisa del método de trabajo es fundamental al procesar muestras complejas, para obtener el mayor número de identificaciones peptídicas posibles correspondientes a las proteínas diana que pretendemos caracterizar.

Por otra parte, en este apartado, evaluamos también la longitud de los péptidos aislados e identificados mediante cada una de las tres estrategias descritas anteriormente, dada su gran importancia a la hora de trabajar con secuencias peptídicas no descritas ni incluidas en bases de datos de referencia (p.ej. IGMT,

neXtProt y UniProt) como son los péptidos de las regiones determinantes de la complementariedad 1-3 (CDR, del inglés *complementarity determining region*) de las Igs. Así, la existencia de péptidos más largos favorece la identificación inequívoca de la proteína correspondiente ya que, a mayor longitud, menor es la probabilidad de coincidencia aleatoria y, por tanto, la asignación incorrecta de la proteína a la que corresponde un determinado péptido o conjunto de péptidos. Así, la aproximación (iii), basada en la ruptura de puentes disulfuro, permitió identificar tanto péptidos cortos (de 6 aminoácidos) como otros más largos (de hasta 23 aminoácidos) generando una distribución homogénea del tamaño de los péptidos identificados. Sin embargo, los métodos (i) y (ii) presentaron una distribución más heterogénea identificando además péptidos con una longitud casi siempre inferior a 15 aminoácidos.

Finalmente, la integración de los hallazgos proteómicos con los datos genómicos resultantes de los análisis moleculares de la secuencia de nucleótidos de los genes de las Igs permitió asignar de forma inequívoca regiones variables específicas de distintos BCRs (CDR y región marco, FR, del inglés *framework region*) y secuenciadas en paralelo por ambas tecnologías a nivel de proteína y ARNm. Para ello, se desarrolló un algoritmo simple que permitía comparar las secuencias peptídicas obtenidas en cada muestra con la secuencia genómica de referencia (traducida a aminoácidos), distinguiendo de forma específica las regiones CDR (1-3) y FR (1-3), y que calcula la coincidencia de cobertura. Así, hasta un 30% de la secuencia génica completa de la Ig pudo ser identificada mediante MS, alcanzando dicha identificación un 62% de la secuencia en el caso de las regiones FR. Teniendo en cuenta la elevada variabilidad de las zonas CDR1-3, la ausencia de librerías peptídicas extensas de Ig y el número de péptidos tripsinizables (la tripsina sólo genera cortes en proteínas cuando existen en la secuencia los aminoácidos Lys y Arg), los resultados obtenidos resultan muy prometedores, estableciendo un punto de partida para el abordaje de las limitaciones que presentan este tipo de ensayos. Como ejemplo, y teniendo en cuenta los parámetros estándar de péptidos identificables, por cada secuencia genómica había

alrededor de 8 péptidos posibles, de los cuales detectamos 3 con alta fiabilidad. Así, estos resultados nos permitieron concluir además que el desarrollo de librerías peptídicas de referencia para las secuencias de las Igs es clave para la identificación de las mismas cuando éstas están presentes en muestras complejas con mezclas de múltiples proteínas, además de las Igs.

En conjunto, nuestros resultados en relación con las técnicas de secuenciación peptídica del BCR demuestran la importancia de un desarrollo metodológico adecuado para una caracterización correcta e inequívoca de Igs unidas a membrana (*i.e.* BCR) mediante estrategias de MS. Dicho análisis a partir de muestras complejas supone un gran reto para la proteómica actual, debido principalmente a dos factores: i) la dificultad que entraña aislar y purificar proteínas de membrana y ii) la baja estequiometría del BCR en relación al resto de proteínas contenidas en la célula (*p.ej.* bajo rango dinámico y limitaciones en la digestión enzimática). Por este motivo, entendemos que los mecanismos de secuenciación peptídica del BCR deben enfocarse en el enriquecimiento de las muestras en Ig, reduciendo los pasos de procesamiento para minimizar las pérdidas de muestra durante dicho proceso, sin afectar, siempre que sea posible, la pureza alcanzada. Asimismo, la integración de los datos proteómicos con los derivados de estudios genómicos convencionales supondrá un gran avance a la hora de posibilitar una mejor caracterización del BCR, y potencialmente también, de otras proteínas con secuencias variables del sistema inmune, como las proteínas HLA.

### **Desarrollo de una estrategia para la integración de datos proteómicos y transcriptómicos para la evaluación de la complementariedad de ambas aproximaciones en el contexto biológico de células B tumorales.**

En los apartados anteriores de este trabajo doctoral, nuestro interés se centró, de forma preferente y específica, en el análisis a nivel proteico de los mecanismos

moleculares que ocurren en las células tumorales de la LLC-B y el desarrollo de herramientas dirigidas a la secuenciación peptídica de su BCR. Sin embargo, somos conscientes de que el análisis proteómico por sí solo, aunque fundamental, no es suficiente para la caracterización completa de las células patológicas, debiendo integrarse la información proteómica generada con datos derivados del análisis masivo de otros elementos y componentes celulares (p.ej. de los ácidos nucleicos). En este sentido, la genómica, aunque permite la determinación de las secuencias génicas contenidas en cada individuo, no advierte acerca de los cambios que ocurren en la expresión génica que afectan de forma distinta a las proteínas finales generadas en cada estirpe celular y que terminan por determinar comportamientos y estados celulares concretos. Por el contrario, la transcriptómica, tal como la proteómica, abordan estos cambios, al evaluar moléculas más dinámicas (transcritos y proteínas) que permiten una lectura más precisa del estado celular en un momento concreto<sup>170</sup>. La transcripción de los genes responde a las necesidades de la célula, condicionadas por su entorno concreto, lo que convierte a dichos transcritos en indicadores del comportamiento celular y de la expresión activa de genes codificantes de proteínas<sup>351</sup>. Aun así, dichos transcritos pueden ser degradados o modificados durante el procesamiento post-traducciona, convirtiendo a las proteínas en las verdaderas efectoras celulares<sup>171,352,353</sup>. Además, los avances en el campo de la proteómica han permitido mejorar la anotación del genoma<sup>354,355</sup>, a pesar de las múltiples e importantes limitaciones que siguen presentando las técnicas proteómicas en relación, por ejemplo, con la preparación específica de muestras, el subfraccionamiento celular, la separación peptídica, la precisión y exactitud técnicas, la generación y procesamiento de datos o los parámetros de puntuación y validación peptídica/proteica<sup>354</sup>. Por todo lo anteriormente expuesto, resulta evidente que la integración de los datos transcriptómicos y proteómicos, junto con el uso de estrategias bioinformáticas adecuadas, permitiría adquirir información pormenorizada y fidedigna acerca de cómo se regula la expresión de los genes y las proteínas y cómo

ambos elementos se integran a nivel molecular, celular y del propio organismo, para controlar las respuestas biológicas<sup>356,357</sup>.

Con tal propósito, en este capítulo de esta tesis diseñamos y desarrollamos un flujo de trabajo para integrar conjuntos de datos derivados de análisis transcriptómicos y proteómicos, tomando como modelo una línea celular concreta (*i.e.* Ramos, línea celular B derivada de un paciente con linfoma de Burkitt). De forma paralela, y respecto a los datos proteómicos, estos se generaron empleando dos métodos de aislamiento de proteínas mediante fraccionamiento subcelular y una estrategia basada en MS, y se analizaron frente a las dos bases de datos más empleadas en el área de la proteómica (*i.e.* UniProt y neXtProt). A su vez, los datos transcriptómicos se obtuvieron directamente de la base de datos GEO. Para la integración, se desarrolló un código bioinformático en lenguaje de programación R.

Mediante el empleo de procedimientos de subfraccionamiento celular logramos reducir la complejidad de las muestras y con ello incrementar la cobertura proteómica. En concreto, empleamos secuencialmente distintos detergentes para la extracción selectiva de las proteínas del citoplasma, las membranas, los orgánulos y el núcleo, resultando en la identificación de alrededor de 6.000 proteínas y 30.000 péptidos mediante ambos métodos. Sin embargo, la identificación de las proteínas de las membranas (sobre las que ya hemos referido la dificultad que entraña su análisis) resultó incrementada de forma significativa ( $p$ -valor=0,007) al emplear IGEPAL en comparación con el uso de *n*-dodecil- $\beta$ -D-maltopiranósido. De forma análoga, la realización de choques osmóticos, empleando altas concentraciones salinas, permitió una mayor identificación de proteínas nucleares respecto al uso de detergentes no iónicos.

Aunque el objetivo principal de este estudio fue el abordaje de la integración de datos proteómicos y transcriptómicos, también analizamos los datos proteómicos desde la perspectiva de su significado biológico, lo que nos permitió obtener información relevante en relación con la línea celular analizada. Así, la línea celular

*Ramos* empleada como modelo de célula B tumoral en este estudio (linfoma de Burkitt) fue la línea celular donde primero se describió el gen *MYC* debido a la sobreexpresión de este oncogen, cuya activación por translocación a la proximidad de los genes *IGH* o *IGL* induce una proliferación celular descontrolada<sup>358,359</sup>. El análisis de la vía de señalización asociada a *MYC* mostró la presencia en estas células B de un gran número de proteínas que interaccionan con *MYC* como *ACTL6A*, *BCL2*, *CHD8*, *GTF2I*, *MAPK1*, *MAX*, *MYCBP2*, *SMAD2* y *SMAD3*, reflejando la existencia de una selección positiva de las mismas.

Además, evaluamos la robustez del método de extracción subcelular mediante la determinación de la contaminación cruzada entre proteínas detectadas en distintos comportamientos subcelulares, que se ve reflejada por la presencia de proteínas foráneas en cada una de las distintas fracciones aisladas. Así, el método de subfraccionamiento permitió el correcto aislamiento del 75% de las proteínas del citoplasma y del núcleo, viéndose reducido este porcentaje hasta un 50% en la fracción de membranas y 44% en la fracción de los orgánulos, lo que probablemente es debido a la gran heterogeneidad de estas dos últimas fracciones subcelulares y a la mayor dificultad de la extracción completa de sus proteínas. En base al mapeo por cromosomas según la estrategia C-HPP, identificamos aproximadamente 30% de los genes codificantes de proteínas presentes en el genoma humano en la línea celular *Ramos* y en un ensayo único. Asimismo, este porcentaje se mantuvo relativamente constante cuando evaluamos de forma independiente cada cromosoma, destacando la cobertura del ADN mitocondrial (85%). Cabe señalar además que al evaluar el grado de identificación de proteínas *missing* en el conjunto de datos proteómicos, en el momento de realizar este estudio logramos identificar hasta 315 de dichas proteínas *missing* (grupos PE 2-4).

Además de los resultados anteriores, el objetivo principal de este capítulo de nuestro trabajo se centró en la integración de datos proteómicos, dirigida en la práctica a través de 3 grupos de criterios distintos de asignación proteica,

denominados *intersección*, *unión* y *máximo*. Estos criterios hacen referencia al conjunto global de proteínas i) identificadas en todas las réplicas biológicas analizadas con  $\geq 2$  péptidos proteotípicos únicos por proteína, ii) identificadas en al menos una réplica biológica con al menos 2 péptidos proteotípicos únicos por proteína, y iii) identificadas en al menos una réplica biológica con 1 solo péptido proteotípico, respectivamente. Con este enfoque pretendíamos generar conjuntos de datos que permitieran, por un lado, identificar el mayor número de proteínas posibles (mediante *unión* y *máximo*) y, por otra parte, determinar aquellas proteínas que se identificaban de forma inequívoca y restrictiva con la estrategia proteómica (*intersección*). Así, en conjunto se determinó la presencia de 3.383, 5.494 y 8.931 proteínas en la línea celular Ramos en los conjuntos de *intersección*, *unión* y *máximo*, respectivamente. El mapeo genómico correspondiente resultó en 3.433, 5.540 y 8.976 genes. La integración de los datos, con la consiguiente identificación de aquellos genes transcritos pero no traducidos a proteínas, permitió constatar que existía un elevado grado de solapamiento (94%) entre ambas técnicas. Las proteínas identificadas de forma exclusiva a nivel proteómico estaban relacionadas con proteínas mitocondriales y ribosomales fundamentalmente, lo que podría explicarse al no estar incluidas en los microarrays de expresión sondas específicas para segmentos génicos presentes en el ADN mitocondrial. Además, las pérdidas específicas de identificación en la plataforma transcriptómica, asociadas a las Igs y a las proteínas del complejo mayor de histocompatibilidad, fue probablemente debida al elevado número y variabilidad de ambas proteínas, y en especial de las Igs, por la gran diversidad de reordenamientos génicos y mutaciones hipersomáticas que existen/ocurren en estos genes, lo cual dificulta el diseño de sondas adecuadas para una completa cobertura de los mismos. Asimismo, detectamos también transcritos mediante la plataforma transcriptómica que no estaban representados en el proteoma de la célula (sobre todo transcritos relacionados con proteínas nucleares y de unión a ADN). Esta discrepancia, probablemente sea debida a que el método de extracción de los diferentes

compartimentos subcelulares elegido para la integración permitía un mejor rendimiento a la hora de aislar proteínas de membrana, en detrimento de las proteínas nucleares, contribuyendo así a una reducción del número de estas últimas en relación a la identificación transcriptómica del mismo tipo de muestras. Además, alrededor de 300 transcritos identificados se correspondían con ARN no codificante de proteínas, lo que resultaba en identificaciones imposibles de realizar a nivel proteómico.

Finalmente, la comparación de los datos proteómicos con las bases de datos UniProt y neXtProt mostró que esta última permitía generar un mayor número de identificaciones proteicas (y, por ende, también de proteínas *missing*: 54 vs. 315 (PE1-4) para UniProt y neXtProt, respectivamente) durante los análisis descritos anteriormente. La integración en neXtProt de información procedente de diferentes bases de datos (p.ej. Swiss-Prot, Ensembl, HPA y PeptideAtlas)<sup>177</sup> hizo posible una mejor caracterización de los proteomas y confirmó el incremento en el número de identificaciones con esta base de datos respecto a UniProt.

En resumen, nuestros resultados han permitido confirmar la posibilidad de integrar datos derivados de tecnologías proteómicas y transcriptómicas demostrando con ello la posibilidad de lograr un mapeo más detallado y masivo de las biomoléculas que definen cada tipo y/o estado celular. Además, dado que las búsquedas derivadas de cada metodología -ómica son complementarias, una vez desarrolladas en conjunto, permiten la validación mutua y el incremento de la cobertura total de los genes codificantes y no codificantes de proteínas presentes en el genoma completo. Adicionalmente, la elección del método de extracción proteica subcelular, así como de la base de datos a emplear, demostraron ser claves a la hora de identificar el mayor número posible de proteínas, siendo muy relevante en esta área la calidad de los datos a integrar, dependiente directamente de los criterios establecidos para la identificación de proteínas según el número de péptidos proteotípicos y de su presencia en todas las réplicas biológicas evaluadas.



---

**Caracterización proteómica cuantitativa de la célula B normal de órganos linfoides secundarios en distintas etapas de la diferenciación Ag-dependiente con el fin de establecer una referencia de normalidad para el estudio de la célula B tumoral de la LLC.**

El abordaje del análisis de la célula de la LLC realizado en los estudios previos permitió una caracterización muy detallada de la célula B patológica desde distintas perspectivas proteómicas. Sin embargo, el estudio de las células B normales en distintos estadios madurativos (vs. el conjunto de las células B totales) es clave a la hora de definir con mayor precisión las alteraciones concretas presentes no solo en células B tumorales, sino también en células B de pacientes con inmunodeficiencias primarias, enfermedades autoinmunes o durante el envejecimiento celular. En el momento en el que diseñamos este capítulo de nuestro trabajo de investigación, se había definido ya de forma detallada las diferentes etapas de diferenciación de la célula B empleando aproximaciones genómicas y transcriptómicas; con ellos, se había obtenido información clave sobre los principales genes expresados (y reguladores) en cada etapa de la maduración B<sup>360-364</sup>. A pesar de la importancia de esta información, la comprensión de los mecanismos (tan) dinámicos que suceden durante el proceso de diferenciación celular B requiere de la exploración de los proteomas expresados en cada una de las etapas madurativas<sup>324,365-367</sup>, debido principalmente a que las proteínas constituyen los elementos funcionales efectores de distintos procesos celulares<sup>368</sup>. No obstante, hasta la fecha ningún estudio había abordado este tipo de análisis proteómico. Con estos antecedentes, el último objetivo de este trabajo se centró en el análisis de los perfiles proteómicos cualitativos y cuantitativos de poblaciones de células B en distintas etapas de la maduración Ag-dependiente, derivadas de un órgano linfoide secundario como es la amígdala. En conjunto purificamos 5 subpoblaciones B (células naive, CB, CC, células B de memoria y CP) para analizar posteriormente en cada una de ellas su proteoma mediante una estrategia analítica basada en MS libre de marcaje con isotopos. Con la finalidad de

comparar los datos obtenidos con su contrapartida transcriptómica, se seleccionaron resultados de la base de datos GEO, derivados del análisis mediante microarrays de expresión de las mismas poblaciones de células B de amígdala procesadas mediante el mismo método de purificación. Para la integración de los datos derivados de los análisis con ambas estrategias -ómicas, se empleó una versión adaptada del código informático desarrollado y descrito en el apartado anterior de los resultados de esta tesis.

En general, nuestros resultados mostraron un elevado grado de solapamiento - media de 75% - entre el proteoma de las 5 poblaciones celulares B estudiadas, alcanzando éste niveles de 95% para las proteínas presentes en las CP. Estos hallazgos generales probablemente reflejan el origen B común de los 5 tipos celulares estudiados. De acuerdo con ello, el proteoma común incluía proteínas relacionadas con funciones celulares generales como: unión de actina y tubulina, biosíntesis, proteólisis, adhesión celular, procesamiento de ARN y metabolismo de la glucosa. Sin embargo, a pesar del elevado grado de solapamiento observado, también existían proteínas específicas de cada población. Por ejemplo, las proteínas MED21, DDX47, GINS3, EIF4G2, NUP62 e HIST2HBC, que participan en procesos de fagocitosis mediada por receptor y en la señalización mediada por quimiocinas, únicamente se encontraban presentes en células del CG; dentro de estas células, los CB expresaron de forma exclusiva VAV1 y VAV2, moléculas que son clave en el proceso de la producción de Ig secretables<sup>369</sup>, mientras que los CC fueron las únicas células B positivas para RGS13, una molécula que juega un papel importante en la quimiotaxis y la limitación de la expansión de las células del CG<sup>370</sup>. Por su parte, las células naive eran la subpoblación celular que presentaba un menor número de proteínas específicas (7), relacionándose su expresión con diversas rutas metabólicas, los ribosomas y la replicación del ADN. Respecto a las células B de memoria y CP, cabe señalar que aunque ambas derivan habitualmente de células B naive, sus vías de diferenciación específicas a partir del CG son distintas, siendo aún su conocimiento

relativamente limitado<sup>371</sup>. En este contexto observamos que, mientras las CP mostraban expresión de 29 proteínas relacionadas con Ig (IGKV3-11, IGKV3-15, IGHD, IGLC3, JCHAIN), citocromo C oxidasas (COX6B1, COX7C), factores de transcripción y reguladores (IRF4, GTF2A1, LZTFL1) y ABL1, entre otras, ninguna de estas proteínas estaba presente en células B de memoria. De todas ellas, merece destacar IRF4, una proteína que se asocia a la condición de célula secretora de Ig<sup>16,372</sup> característica de las CP y que induce expresión de BLIMP1 y XBPI, potenciando a su vez aún más la secreción de las Igs. Asimismo, cabe señalar que la expresión en CP de citocromo C oxidasas y otras proteínas implicadas en el ciclo del ácido cítrico, y que revelan un elevado gasto energético de estas células, podrían estar asociadas a la elevada actividad de síntesis proteica que ocurre específicamente en la CP. Asimismo, la presencia de ABL1 en CP podría resultar ser un marcador determinante de la vía de diferenciación B a CP de larga vida, no sólo por el papel que esta quinasa tiene en la supervivencia de las CP<sup>373</sup>, sino también por el rol que su alteración puede tener en el desarrollo de tumores hematológicos. A diferencia de las CP, las células B de memoria mostraron expresión diferencial de las proteínas LRBA<sup>1,374</sup> (relacionadas con la promoción de la proliferación, la expansión clonal y la supervivencia de células secretoras de Ac); CD29<sup>375</sup>; y THEMIS2<sup>376</sup>, entre otras proteínas.

En conjunto, puede considerarse que las diferencias referidas hasta el momento serían insuficientes para explicar todos los cambios celulares que ocurren entre las diferentes poblaciones analizadas a lo largo de la diferenciación B. Probablemente, esto sea debido a que más que la presencia vs. ausencia de una proteína, lo que realmente determina las variaciones observadas en el distinto comportamiento celular de las subpoblaciones B analizadas son los niveles distintos de expresión de cada una de las proteínas y/o su estado post-traducciona. En este sentido, la evaluación cuantitativa de los proteomas de las células naive, CB, CC y células B de memoria mostró un total de 753 proteínas expresadas en niveles significativamente diferentes entre ellas. Un análisis más pormenorizado permitió distinguir dos grupos principales de

poblaciones celulares en función de las diferencias cuantitativas y cualitativas en sus perfiles de expresión proteica: CB y CC por una parte, y células naive y de memoria por la otra. De todas las poblaciones celulares B analizadas, las células naive presentaron en general los niveles de expresión más bajos para el conjunto de proteínas analizadas. Por el contrario, las células del CG (*i.e.* CB y CC) presentaron los niveles de expresión más elevados, lo cual podría reflejar el elevado grado de renovación proteica que sucede en dichas células<sup>14</sup>.

Un análisis más detallado de las proteínas que mostraban distintos niveles de expresión en diferentes poblaciones de células B mostró que la proteína HLA-DO se expresaba en niveles significativamente superiores en células naive y de memoria, mientras que estaba prácticamente ausente en las células del CG, de acuerdo con estudios genómicos previos<sup>14,19</sup>. Por el contrario, la expresión de HLA-DR y CD74 fue superior en CB y CC. En relación a la expresión de Ig, los niveles de IGHM y de IGHV en general mostraron gran similitud entre todas las poblaciones B analizadas, al igual que CD22, CD44 y CD79B; sin embargo, en el CG se detectó mayor expresión de IGHA, IGHG, IG $\lambda$  e IG $\kappa$ <sup>377</sup>. Además, las proteínas 14-3-3, relacionadas con la recombinación del gen IGH que ocurre durante el cambio de isotipo de Ig<sup>378</sup>, estaban sobreexpresadas también en CB y CC frente a linfocitos naive y de memoria. Otra proteína que mostró niveles de expresión significativamente diferentes fue CARD11, que incrementó su expresión conforme avanzaba la maduración, estando su función asociada a apoptosis y proliferación de células B maduras estimuladas por Ag<sup>379</sup>. En un sentido opuesto, la oncoproteína TCL1A mostró una expresión progresivamente menor según avanzaba la maduración de las células B<sup>380,381</sup>. Por otra parte, las proteínas NFKB1 y NFKB2, que poseen un rol importante en la activación de células B y en la formación del CG<sup>16</sup>, estaban presentes en niveles más elevados en CB y CC. Asimismo, los CC mostraron expresión elevada de la proteína IRF8 cuyo incremento se ha asociado a los procesos de hipermutación somática y al cambio de isotipo de cadena pesada de las Ig<sup>382</sup>.

---

Finalmente, la integración de los datos proteómicos con la información transcriptómica de la base de datos GEO permitió corroborar los resultados proteómicos que indicaban una mayor similitud entre las poblaciones de células naive y de memoria por un lado, y CB y CC por otra parte. Asimismo, se constató que a pesar del elevado número de transcritos identificados en las células naive, la contrapartida proteómica era relativamente más escasa, evidenciando que no todo lo que se transcribe en estas células se traduce finalmente en niveles detectables de proteína.

En resumen, estos resultados demuestran la necesidad de evaluar de forma independiente el proteoma de células B en distintos estadios madurativos, no sólo desde una perspectiva proteómica cualitativa, sino también desde un punto de vista cuantitativo, que pueda aportar más información sobre el dinamismo que existe en la maduración celular B, más allá de la simple presencia/ausencia de proteínas concretas. Estos hallazgos son claves para poder definir con mayor precisión aquellas proteínas concretas cuyos niveles varían en células B tumorales respecto a su contrapartida madurativa normal, y así poder entender mejor los mecanismos patogénicos inherentes a distintas neoplasias de células B.

Estos hallazgos, asociados a los resultados descritos anteriormente en los demás apartados de esta tesis doctoral, demuestran la posibilidad que existe hoy día de analizar de forma exhaustiva y global el proteoma de la célula B patológica y compararlo al de su contrapartida normal. No obstante, ello requiere no sólo de una caracterización cualitativa del proteoma, sino también cuantitativa, incluyendo además el análisis del fosfoproteoma como información clave a la hora de definir con más precisión las vías de señalización celular alteradas, con las potenciales implicaciones diagnósticas, pronósticas y terapéuticas que tal conocimiento puede aportar. Además, el abordaje de la secuenciación peptídica del BCR de los linfocitos de la LLC ha supuesto un paso clave para impulsar futuros estudios en los que se caracterice a nivel proteómico el repertorio de receptores de célula B tumorales, lo

que facilitaría la identificación de las regiones peptídicas específicas de unión a epítomos antigénicos determinados. Todo ello, asociado a la integración de la información proteómica generada con los resultados de análisis genómicos/transcriptómicos, permitirá sin duda en el futuro alcanzar una visión más amplia de las características e interacciones moleculares que definen a las células B patológicas y las diferencian de su contrapartida normal y de sus respectivos estados funcionales.

# Conclusiones





---

**En relación a la caracterización sistemática y cuantitativa del proteoma y fosfoproteoma de las células B de pacientes con LLC y LMB:**

1. Los estudios de proteómica realizados muestran un elevado grado de solapamiento en el proteoma global de las células tumorales de LLC y LMB independientemente del tipo de alteraciones citogenéticas, del estado mutacional de los genes IGHV y del estadio clínico de la enfermedad, lo que probablemente es debido a que la mayoría de las proteínas detectadas en común están relacionadas con funciones celulares genéricas, incluidas funciones comunes a las células B.

2. La combinación de una aproximación basada en MS, con un pre-enriquecimiento en fosfopéptidos empleando Acs anti-fosfotirosina (p-Tyr) conjugados en superficies esféricas, permitió caracterizar cualitativa y cuantitativamente el fosfoproteoma de la LLC consistente en 594 fosfopéptidos correspondientes a un total de 327 fosfoproteínas distintas, de los que 329 fosfopéptidos fueron descritos por primera vez y anotados en la base de datos neXtProt.

3. A diferencia de lo observado en el proteoma global, las células tumorales de la LLC únicamente comparten entre sí una fracción relativamente pequeña (<20%) del fosfoproteoma, lo que posiblemente refleja el mayor dinamismo de las proteínas modificadas post-traduccionalmente (fosfoproteoma), respecto al conjunto de todas las proteínas expresadas en estas células.

4. El análisis fosfoproteómico de las posibles vías de señalización activadas en las células leucémicas de la LLC revela una mayor representación en los distintos pacientes analizados de la vía de señalización del BCR, respecto a otras vías de activación, lo que sugiere la implicación de esta ruta en la ontogenia y/o la supervivencia del clon leucémico. Además, nuestros resultados mostraron la existencia de una fosforilación incrementada en la LLC (vs. LMB) de proteínas de las vías de señalización de NF- $\kappa$ B y STAT3, sugiriendo la posible implicación de las mismas en la transformación maligna de la enfermedad.

**En relación a las posibles alteraciones proteómicas que involucran a vías de señalización intracelular de la célula B de la LLC y su relación con las alteraciones citogenéticas y moleculares presentes en la misma:**

5. El empleo de una aproximación metodológica basada en microarrays de Acs permitió identificar y evaluar la abundancia relativa de cientos de proteínas simultáneamente en células B tumorales de LLC frente a linfocitos B normales de sujetos sanos, observándose importantes diferencias entre ambos grupos de células, reflejando una mayor activación y supervivencia de las células tumorales.

6. La evaluación cualitativa de los perfiles de expresión de proteínas en células B tumorales de pacientes con LLC mostró diferencias significativas entre células portadoras de distintas alteraciones genéticas y moleculares, confirmando el posible impacto de estas últimas en los niveles de expresión de un número variable de proteínas asociadas a diferentes grupos funcionales. Así, mientras que las células tumorales que no presentaban alteraciones genéticas, o tenían únicamente trisomía 12, mostraban perfiles similares a las células B normales, i) la presencia de del(13q) se asoció a una mayor abundancia relativa de la proteína JUN y las ciclinas A, B1 y D1 y una reducción de los niveles de PRKC y de proteínas pro-apoptóticas, reflejando una mayor capacidad de proliferación y expansión clonal; y ii) la positividad para la del(17p) se caracterizó por mostrar alteraciones en las rutas de señalización de ERBB, una mayor disminución de la expresión de proteínas apoptóticas con incremento de los niveles de proteínas anti-apoptóticas y de los complejos ciclina-CDK, reflejando características celulares típicamente asociadas a un comportamiento tumoral más agresivo.

**En relación al establecimiento de una metodología óptima para la secuenciación peptídica del BCR y la integración de dichas secuencias con las secuencias derivadas de técnicas estándar de secuenciación genómica:**

7. El empleo de métodos que incluyan la purificación por afinidad de fragmentos Fab de Ig, o bien la digestión con papaína de Igs purificadas, permite incrementar la eficiencia de la secuenciación peptídica de proteínas del BCR, respecto a técnicas basadas en la simple rotura de los puentes disulfuro que unen las dos cadenas ligeras y pesadas de Ig purificadas. No obstante, este último procedimiento se asoció a una menor pérdida de muestra, a expensas de mayor contaminación cruzada por otras proteínas no asociadas de forma específica al BCR.

8. La integración de los hallazgos proteómicos con los datos genómicos resultantes de los análisis moleculares de la secuencia de nucleótidos de los genes de las Igs permite asignar de forma inequívoca un mayor número de péptidos a regiones variables específicas de distintos BCRs, incrementando así notablemente la eficiencia de la secuenciación peptídica del BCR.

**En relación al desarrollo de una estrategia para la integración de datos proteómicos y transcriptómicos, en el modelo de una línea celular B de linfoma de Burkitt:**

9. La integración de datos proteómicos y transcriptómicos derivados del estudio de la línea celular B *Ramos* demuestra una elevada eficiencia y complementariedad en la información derivada de ambos tipos de análisis con un grado de solapamiento >90% entre ambas técnicas, asociándose las discrepancias observadas sobre todo con la identificación de genes transcritos pero no traducidos a proteínas.

**En relación a la caracterización del proteoma de células B normales de amígdala en distintas etapas de diferenciación y su relación con el correspondiente transcriptoma:**

10. El proteoma de las distintas poblaciones de células B de amígdala analizadas mostró un elevado grado de solapamiento entre ellas, aunque existen proteínas que se expresan de forma específica en cada una de las distintas poblaciones B estudiadas. Estas diferencias se incrementan notablemente cuando, además de la presencia/ausencia de una proteína, se consideran los niveles de expresión de las mismas, lo que indica que en la progresión entre los distintos estadios de la maduración B existe un gran dinamismo en la regulación de los niveles de expresión de muchas proteínas, más que cambios radicales en la presencia vs. ausencia de las mismas.

11. En términos globales, las diferencias cuantitativas y cualitativas del proteoma de células B de amígdalas en distintos estadios madurativos permite diferenciar, además de las CP, dos grupos principales de poblaciones de linfocitos B: CB y CC por una parte, y células naive y de memoria. De ellas, las células naive presentaban, en general, los niveles de expresión más bajos de proteínas, mientras que las células del CG (*i.e.* CB y CC) mostraban los niveles más elevados, reflejo del distinto dinamismo de estas poblaciones celulares B.

12. La integración de datos de expresión proteica con la información transcriptómica permitió validar los datos proteómicos, al identificarse también con ello, una mayor similitud entre las poblaciones CB y CC por un lado, y naive y memoria por el otro. Además, se constata que, a pesar del elevado número de transcritos identificados en las células naive, su contrapartida proteómica es relativamente más escasa, evidenciando que no todo lo que se transcribe en esta población celular se traduce finalmente en niveles detectables de proteínas.

# Bibliografía



- 1 Lebien TW, Tedder TF. B lymphocytes: How they develop and function. *Blood* 2008; **112**: 1570–1580.
- 2 Kondo M. Lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. *Immunol Rev* 2010; **238**: 37–46.
- 3 Li YS, Wasserman R, Hayakawa K, Hardy RR. Identification of the earliest B lineage stage in mouse bone marrow. *Immunity* 1996; **5**: 527–535.
- 4 Allman D, Li J, Hardy RR. Commitment to the B Lymphoid Lineage Occurs before DH-JH Recombination. *J Exp Med* 1999; **189**: 735–740.
- 5 Namen AE, Lupton S, Hjerrild K, Wignall J, Mochizuki DY, Schmierer A *et al*. Stimulation of B-cell progenitors by cloned murine interleukin-7. *Nature* 1988; **333**: 571–3.
- 6 Korsmeyer SJ, Hieter PA, Ravetch J V, Poplack DG, Waldmann TA, Leder P. Developmental hierarchy of immunoglobulin gene rearrangements in human leukemic pre-B-cells. *Proc Natl Acad Sci U S A* 1981; **78**: 7096–7100.
- 7 Pieper K, Grimbacher B, Eibel H. B-cell biology and development. *J. Allergy Clin. Immunol.* 2013; **131**: 959–971.
- 8 Hardy RR, Hayakawa K. B Cell Development Pathways. *Annu Rev Immunol* 2001; **19**: 595–621.
- 9 Pillai S. The chosen few? Positive selection and the generation of naive B lymphocytes. *Immunity* 1999; **10**: 493–502.
- 10 Nagasawa T. Microenvironmental niches in the bone marrow required for B-cell development. *Nat Rev Immunol* 2006; **6**: 107–116.
- 11 Cariappa A, Pillai S. Antigen-dependent B-cell development. *Curr. Opin. Immunol.* 2002; **14**: 241–249.
- 12 McHeyzer-Williams LJ, McHeyzer-Williams MG. Antigen-specific memory B cell development. *Annu Rev Immunol* 2005; **23**: 487–513.
- 13 Pérez ME, Billordo LA, Baz P, Fainboim L, Arana E. Human memory B cells isolated from blood and tonsils are functionally distinctive. *Immunol Cell Biol* 2014; **92**: 882–887.
- 14 Allen CDC, Okada T, Cyster JG. Germinal-Center Organization and Cellular

- Dynamics. *Immunity*. 2007; **27**: 190–202.
- 15 Kjeldsen MK, Perez-Andres M, Schmitz A, Johansen P, Boegsted M, Nyegaard M *et al*. Multiparametric flow cytometry for identification and fluorescence activated cell sorting of five distinct B-cell subpopulations in normal tonsil tissue. *Am J Clin Pathol* 2011; **136**: 960–969.
- 16 De Silva NS, Klein U. Dynamics of B cells in germinal centres. *Nat Rev Immunol* 2015; **15**: 137–148.
- 17 Shinnakasu R, Inoue T, Kometani K, Moriyama S, Adachi Y, Nakayama M *et al*. Regulated selection of germinal-center cells into the memory B cell compartment. *Nat Immunol* 2016; : 1–11.
- 18 Tew JG, Wu J, Fakher M, Szakal AK, Qin D. Follicular dendritic cells: Beyond the necessity of T-cell help. *Trends Immunol*. 2001; **22**: 361–367.
- 19 Chalouni C, Banchereau J, Vogt AB, Pascual V, Davoust J. Human germinal center B cells differ from naive and memory B cells by their aggregated MHC class II-rich compartments lacking HLA-DO. *Int. Immunol*. 2003; **15**: 457–466.
- 20 Arpin C, Déchanet J, Van Kooten C, Merville P, Grouard G, Brière F *et al*. Generation of memory B cells and plasma cells in vitro. *Science* 1995; **268**: 720–722.
- 21 Liu YJ, Joshua DE, Williams GT, Smith CA, Gordon J, MacLennan IC. Mechanism of antigen-driven selection in germinal centres. *Nature* 1989; **342**: 929–31.
- 22 Scheeren FA, Naspetti M, Diehl S, Schotte R, Nagasawa M, Wijnands E *et al*. STAT5 regulates the self-renewal capacity and differentiation of human memory B cells and controls Bcl-6 expression. *Nat Immunol* 2005; **6**: 303–13.
- 23 Shaffer AL, Young RM, Staudt LM. Pathogenesis of human B cell lymphomas. *Annu Rev Immunol* 2012; **30**: 565–610.
- 24 Pillai S, Mattoo H, Cariappa A. B cells and autoimmunity. *Curr. Opin. Immunol*. 2011; **23**: 721–731.
- 25 Lipsky PE. Systemic lupus erythematosus: an autoimmune disease of B cell hyperactivity. *Nat. Immunol*. 2001; **2**: 764–6.
- 26 Seda V, Mraz M. B-cell receptor signalling and its crosstalk with other pathways in normal and malignant cells. *Eur. J. Haematol*. 2015; **94**: 193–205.



- 27 Gauld SB, Dal Porto JM, Cambier JC. B cell antigen receptor signaling: roles in cell development and disease. *Science* 2002; **296**: 1641–1642.
- 28 Burger JA, Chiorazzi N. B cell receptor signaling in chronic lymphocytic leukemia. *Trends Immunol.* 2013; **34**: 592–601.
- 29 Monroe JG. ITAM-mediated tonic signalling through pre-BCR and BCR complexes. *Nat Rev Immunol* 2006; **6**: 283–294.
- 30 Hou P, Araujo E, Zhao T, Zhang M, Massenburg D, Veselits M *et al.* B cell antigen receptor signaling and internalization are mutually exclusive events. *PLoS Biol* 2006; **4**: 1147–1158.
- 31 Purroy N, Carabia J, Abrisqueta P, Egia L, Aguiló M, Carpio C *et al.* Inhibition of BCR signaling using the Syk inhibitor TAK-659 prevents stroma-mediated signaling in chronic lymphocytic leukemia cells. *Oncotarget* 2017; **8**: 742–756.
- 32 Robinson R. Distinct B cell receptor functions are determined by phosphorylation. *PLoS Biol* 2006; **4**: e231.
- 33 Stevenson FK, Krysov S, Davies AJ, Steele AJ, Packham G. B-cell receptor signaling in chronic lymphocytic leukemia. *Blood.* 2011; **118**: 4313–4320.
- 34 O'Hayre M, Salanga CL, Kipps TJ, Messmer D, Dorrestein PC, Handel TM. Elucidating the CXCL12/CXCR4 signaling network in chronic lymphocytic leukemia through phosphoproteomics analysis. *PLoS One* 2010; **5**. doi:10.1371/journal.pone.0011716.
- 35 Li SH, Dong WC, Fan L, Wang GS. Suppression of chronic lymphocytic leukemia progression by CXCR4 inhibitor WZ811. *Am J Transl Res* 2016; **8**: 3812–3821.
- 36 Chen S-S, Chang BY, Chang S, Tong T, Ham S, Sherry B *et al.* BTK inhibition results in impaired CXCR4 chemokine receptor surface expression, signaling and function in chronic lymphocytic leukemia. *Leukemia* 2016; **30**: 833–43.
- 37 O'Hayre M, Salanga CL, Handel TM, Allen SJ. Chemokines and cancer: migration, intracellular signalling and intercellular communication in the microenvironment. *Biochem J* 2008; **409**: 635–649.
- 38 Ten Hacken E, Burger JA. Molecular pathways: Targeting the microenvironment in chronic lymphocytic leukemia-focus on the B-cell receptor. *Clin Cancer Res* 2014; **20**: 548–556.

- 39 Nannini PR, Borge M, Mikolaitis VC, Abreu C, Morande PE, Zanetti SR *et al.* CCR4 expression in a case of cutaneous Richter's transformation of chronic lymphocytic leukemia (CLL) to diffuse large B-cell lymphoma (DLBCL) and in CLL patients with no skin manifestations. *Eur J Haematol* 2011; **87**: 80–86.
- 40 Kumar H, Kawai T, Akira S. Toll-like receptors and innate immunity. *Biochem. Biophys. Res. Commun.* 2009; **388**: 621–625.
- 41 Gerondakis S, Grumont RJ, Banerjee A. Regulating B-cell activation and survival in response to TLR signals. *Immunol Cell Biol* 2007; **85**: 471–475.
- 42 Yamamoto M, Sato S, Hemmi H, Hoshino K, Kaisho T, Sanjo H *et al.* Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway. *Science* 2003; **301**: 640–643.
- 43 Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Dohner H *et al.* Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood* 2008; **111**: 5446–5456.
- 44 Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R *et al.* The 2016 revision of the World Health Organization (WHO) classification of lymphoid neoplasms. *Blood* 2016; **127**: 2375–2391.
- 45 Eichhorst B, Robak T, Montserrat E, Ghia P, Hillmen P, Hallek M *et al.* Chronic lymphocytic leukaemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann Oncol* 2015; **26**: v78.
- 46 Kipps TJ, Stevenson FK, Wu CJ, Croce CM, Packham G, Wierda WG, O'Brien S, Gribben J RK. Chronic lymphocytic leukaemia. *Nat Rev Dis Prim* 2017; **3**: 16099.
- 47 Rai KR, Jain P. Chronic lymphocytic leukemia (CLL)-Then and now. *Am J Hematol* 2016; **91**: 330–340.
- 48 Hallek M. Chronic lymphocytic leukemia: 2015 Update on diagnosis, risk stratification, and treatment. *Am J Hematol* 2015; **90**: 446–60.
- 49 Molica S. Sex differences in incidence and outcome of chronic lymphocytic leukemia patients. *Leuk Lymphoma* 2006; **47**: 1477–1480.
- 50 Nabhan C, Aschebrook-Kilfoy B, Chiu BC, Smith SM, Shanafelt TD, Evens AM *et al.* The impact of race, ethnicity, age and sex on clinical outcome in chronic

- lymphocytic leukemia: a comprehensive Surveillance, Epidemiology, and End Results analysis in the modern era. *Leuk Lymphoma* 2014; **55**: 2778–2784.
- 51 Pulte D, Redaniel MT, Bird J, Jeffreys M. Survival for patients with chronic leukemias in the US and Britain: Age-related disparities and changes in the early 21st century. *Eur J Haematol* 2015; **94**: 540–545.
- 52 Yang S-M, Li J-Y, Gale RP, Huang X-J. The mystery of chronic lymphocytic leukemia (CLL): Why is it absent in Asians and what does this tell us about etiology, pathogenesis and biology? *Blood Rev* 2015; **29**: 205–213.
- 53 Slager SL, Caporaso NE, de Sanjose S, Goldin LR. Genetic susceptibility to chronic lymphocytic leukemia. *Semin Hematol* 2013; **50**: 296–302.
- 54 Linet MS, Van Natta ML, Brookmeyer R, Khoury MJ, McCaffrey LD, Humphrey RL *et al.* Familial cancer history and chronic lymphocytic leukemia. A case-control study. *Am J Epidemiol* 1989; **130**: 655–664.
- 55 Cerhan JR, Slager SL. Familial predisposition and genetic risk factors for lymphoma. *Blood*. 2015; **126**: 2265–2273.
- 56 Dearden C. B- and T-cell prolymphocytic leukemia: antibody approaches. *Hematol Am Soc Hematol Educ Progr* 2012; **2012**: 645–651.
- 57 Scarfo L, Ferreri A, Ghia P. Chronic lymphocytic leukaemia. *Crit Rev Oncol Hematol* 2016; **104**: 169–182.
- 58 Bachow SH, Lamanna N. Evolving Strategies for the Treatment of Chronic Lymphocytic Leukemia in the Upfront Setting. *Curr. Hematol. Malig. Rep.* 2016; **11**: 61–70.
- 59 Strati P, Shanafelt T. Monoclonal B-cell lymphocytosis and early stage CLL: diagnosis, natural history, and risk stratification. *Blood* 2015; **126**: 454–462.
- 60 Landgren O, Albitar M, Ma W, Abbasi F, Hayes RB, Ghia P *et al.* B-cell clones as early markers for chronic lymphocytic leukemia. *N Engl J Med* 2009; **360**: 659–67.
- 61 Rawstron AC, Bennett FL, O'Connor SJ, Kwok M, Fenton JA, Plummer M *et al.* Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. *N Engl J Med* 2008; **359**: 575–583.
- 62 Vardi A, Dagklis A, Scarfo L, Jelinek D, Newton D, Bennett F *et al.* Immunogenetics shows that not all MBL are equal: the larger the clone, the more

- similar to CLL. *Blood* 2013; **121**: 4521–4528.
- 63 Barrio S, Shanafelt T, Ojha J, Chaffee K, Secreto C, Kortüm K *et al*. Genomic characterization of high-count MBL cases indicates that early detection of driver mutations and subclonal expansion are predictors of adverse clinical outcome. *Leukemia* 2016; : 1–7.
- 64 Ojha J, Secreto C, Rabe K, Ayres-Silva J, Tschumper R, Dyke D V *et al*. Monoclonal B-cell lymphocytosis is characterized by mutations in CLL putative driver genes and clonal heterogeneity many years before disease progression. *Leukemia*. 2014; **28**: 2395–2398.
- 65 Swerdlow SH, Campo E, Pileri SA, Lee Harris N, Stein H, Siebert R *et al*. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016; **127**: 2375–2390.
- 66 Morabito F, Mosca L, Cutrona G, Agnelli L, Tuana G, Ferracin M *et al*. Clinical monoclonal B lymphocytosis versus Rai 0 chronic lymphocytic leukemia: A comparison of cellular, cytogenetic, molecular, and clinical features. *Clin Cancer Res* 2013; **19**: 5890–5900.
- 67 Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C *et al*. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood* 2013; **121**: 1403–1412.
- 68 Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T *et al*. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood* 2014; **124**: 49–62.
- 69 Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. *Blood* 1975; **46**: 219–34.
- 70 Binet JL, Lepage M, Dighiero G, Charron D, D'Athis P, Vaugier G *et al*. A clinical staging system for chronic lymphocytic leukemia: prognostic significance. *Cancer* 1977; **40**: 855–864.
- 71 Molica S, Alberti A. Prognostic value of the lymphocyte doubling time in chronic lymphocytic leukemia. *Cancer* 1987; **60**: 2712–2716.
- 72 Hallek M. Prognostic factors in chronic lymphocytic leukemia. *Ann Oncol Off J Eur Soc Med Oncol* 2008; **19 Suppl 4**: iv51-3.
- 73 Cramer P, Hallek M. Prognostic factors in chronic lymphocytic leukemia-what do

- we need to know? *Nat Rev Clin Oncol* 2011; **8**: 38–47.
- 74 Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do KA *et al.* Prognostic nomogram and index for overall survival in previously untreated patients with chronic lymphocytic leukemia. *Blood* 2007; **109**: 4679–4685.
- 75 Sarfati M, Chevret S, Chastang C, Biron G, Stryckmans P, Delespesse G *et al.* Prognostic importance of serum soluble CD23 level in chronic lymphocytic leukemia. *Blood* 1996; **88**: 4259–4264.
- 76 Matthews C, Catherwood MA, Morris TCM, Kettle PJ, Drake MB, Gilmore WS *et al.* Serum TK levels in CLL identify Binet stage A patients within biologically defined prognostic subgroups most likely to undergo disease progression. *Eur J Haematol* 2006; **77**: 309–317.
- 77 Hallek M, Wanders L, Ostwald M, Busch R, Senekowitsch R, Stern S *et al.* Serum beta(2)-microglobulin and serum thymidine kinase are independent predictors of progression-free survival in chronic lymphocytic leukemia and immunocytoma. *Leuk Lymphoma* 1996; **22**: 439–447.
- 78 Bergmann MA, Eichhorst BF, Busch R, Adorf D, Stilgenbauer S, Eckart MJ *et al.* Prospective Evaluation of Prognostic Parameters in Early Stage Chronic Lymphocytic Leukemia (CLL): Results of the CLL1-Protocol of the German CLL Study Group (GCLLSG). *Blood* 2015; **110**: 625 LP-625.
- 79 Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL *et al.* Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* 1999; **94**: 1840–1847.
- 80 Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V Genes Are Associated With a More Aggressive Form of Chronic Lymphocytic Leukemia. *Blood* 1999; **94**: 1848–1854.
- 81 Thorselius M, Krober A, Murray F, Thunberg U, Tobin G, Buhler A *et al.* Strikingly homologous immunoglobulin gene rearrangements and poor outcome inV(H)3-21-using chronic lymphocytic leukemia patients independent of geographic origin and mutational status. *Blood* 2006; **107**: 2889–2894.
- 82 Tobin G, Thunberg U, Johnson A, Thörn I, Söderberg O, Hultdin M *et al.* Somatically mutated Ig VH3-21 genes characterize a new subset of chronic lymphocytic leukemia. *Blood* 2002; **99**: 2262–2264.

- 83 Baliakas P, Hadzidimitriou A, Sutton L-A, Minga E, Agathangelidis A, Nichelatti M *et al.* Clinical effect of stereotyped B-cell receptor immunoglobulins in chronic lymphocytic leukaemia: a retrospective multicentre study. *Lancet Haematol* 2014; **1**: e74–e84.
- 84 Murray F, Darzentas N, Hadzidimitriou A, Tobin G, Boudjogra M, Scielzo C *et al.* Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis. *Blood* 2008; **111**: 1524–1533.
- 85 Kröber A, Seiler T, Benner A, Bullinger L, Brückle E, Lichter P *et al.* V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood* 2002; **100**: 1410–1416.
- 86 Zenz T, Mertens D, Küppers R, Döhner H, Stilgenbauer S. From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nat Rev Cancer* 2010; **10**: 37–50.
- 87 Döhner H, Stilgenbauer S, Benner A, Leupolt E, Kröber A, Bullinger L *et al.* Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* 2000; **343**: 1910–1916.
- 88 Kriston C, Bődör C, Szenthe K, Bánáti F, Bánkúti B, Csernus B *et al.* Low CD23 expression correlates with high CD38 expression and the presence of trisomy 12 in CLL. *Hematol Oncol* 2015. doi:10.1002/hon.2244.
- 89 Mayr C, Speicher MR, Kofler DM, Buhmann R, Strehl J, Busch R *et al.* Chromosomal translocations are associated with poor prognosis in chronic lymphocytic leukemia. *Blood* 2006; **107**: 742–751.
- 90 Haferlach C, Dicker F, Schnittger S, Kern W, Haferlach T. Comprehensive genetic characterization of CLL: a study on 506 cases analysed with chromosome banding analysis, interphase FISH, IgV(H) status and immunophenotyping. *Leukemia* 2007; **21**: 2442–51.
- 91 Dicker F, Herholz H, Schnittger S, Nakao A, Patten N, Wu L *et al.* The detection of TP53 mutations in chronic lymphocytic leukemia independently predicts rapid disease progression and is highly correlated with a complex aberrant karyotype. *Leukemia* 2009; **23274**: 117–124.
- 92 Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 2010; **463**: 191–196.

- 93 Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* 2011; **475**: 101–105.
- 94 Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* 2013; **152**: 714–726.
- 95 Damm F, Mylonas E, Cosson A, Yoshida K, Della Valle V, Mouly E *et al.* Acquired initiating mutations in early hematopoietic cells of CLL patients. *Cancer Discov* 2014; **4**: 1088–1101.
- 96 Wang L, Shalek AK, Lawrence M, Ding R, Gaublot JM, Pochet N *et al.* Somatic mutation as a mechanism of Wnt/beta-catenin pathway activation in CLL. *Blood* 2014; **124**: 1089–1098.
- 97 Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K *et al.* SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med* 2011; **365**: 2497–506.
- 98 Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L *et al.* Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet* 2012; **44**: 47–52.
- 99 Raa GD te, Derks I a. M, Navrkalova V, Skowronska A, Moerland PD, Laar J van *et al.* The impact of SF3B1 mutations in CLL on the DNA-damage response. *Leukemia* 2015; **29**: 1133–1142.
- 100 Ljungström V, Cortese D, Young E, Pandzic T, Mansouri L, Plevova K *et al.* Whole-exome sequencing in relapsing chronic lymphocytic leukemia: Clinical impact of recurrent RPS15 mutations. *Blood* 2016; **127**: 1007–1016.
- 101 O'Brien P, Morin P, Ouellette RJ, Robichaud GA. The Pax-5 gene: A pluripotent regulator of B-cell differentiation and cancer disease. *Cancer Res.* 2011; **71**: 7345–7350.
- 102 Chiorazzi N. Implications of new prognostic markers in chronic lymphocytic leukemia. *Hematol Am Soc Hematol Educ Progr* 2012; **2012**: 76–87.
- 103 Malavasi F, Deaglio S, Damle R, Cutrona G, Ferrarini M, Chiorazzi N. CD38 and chronic lymphocytic leukemia: A decade later. *Blood.* 2011; **118**: 3470–3478.
- 104 Rassenti LZ, Huynh L, Toy TL, Chen L, Keating MJ, Gribben JG *et al.* ZAP-70

- compared with immunoglobulin heavy-chain gene mutation status as a predictor of disease progression in chronic lymphocytic leukemia. *N Engl J Med* 2004; **351**: 893–901.
- 105 Claus R, Lucas DM, Ruppert AS, Williams KE, Weng D, Patterson K *et al*. Validation of ZAP-70 methylation and its relative significance in predicting outcome in chronic lymphocytic leukemia. *Blood* 2014; **124**: 42–48.
- 106 Orchard JA, Ibbotson RE, Davis Z, Wiestner A, Rosenwald A, Thomas PW *et al*. ZAP-70 expression and prognosis in chronic lymphocytic leukaemia. *Lancet (London, England)* 2004; **363**: 105–111.
- 107 Rassenti LZ, Jain S, Keating MJ, Wierda WG, Grever MR, Byrd JC *et al*. Relative value of ZAP-70, CD38, and immunoglobulin mutation status in predicting aggressive disease in chronic lymphocytic leukemia. *Blood* 2008; **112**: 1923–1930.
- 108 Gattei V, Bulian P, Del Principe MI, Zucchetto A, Maurillo L, Buccisano F *et al*. Relevance of CD49d protein expression as overall survival and progressive disease prognosticator in chronic lymphocytic leukemia. *Blood* 2008; **111**: 865–873.
- 109 Bulian P, Shanafelt TD, Fegan C, Zucchetto A, Cro L, Nuckel H *et al*. CD49d Is The Strongest Flow Cytometry-Based Predictor Of Overall Survival In Chronic Lymphocytic Leukemia. *Blood* 2013; **122**: 672 LP-672.
- 110 Brachtl G, Piñón Hofbauer J, Greil R, Hartmann TN. The pathogenic relevance of the prognostic markers CD38 and CD49d in chronic lymphocytic leukemia. *Ann. Hematol.* 2014; **93**: 361–374.
- 111 Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE *et al*. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med* 2005; **353**: 1793–1801.
- 112 Cimmino A, Calin GA, Fabbri M, Iorio M V, Ferracin M, Shimizu M *et al*. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A* 2005; **102**: 13944–13949.
- 113 Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E *et al*. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 2002; **99**: 15524–9.
- 114 Balatti V, Rizzotto L, Miller C, Palamarchuk A, Fadda P, Pandolfo R *et al*. TCL1



- targeting miR-3676 is codeleted with tumor protein p53 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 2015; **112**: 2169–74.
- 115 Bichi R, Shinton SA, Martin ES, Koval A, Calin GA, Cesari R *et al.* Human chronic lymphocytic leukemia modeled in mouse by targeted TCL1 expression. *Proc Natl Acad Sci* 2002; **99**: 6955–6960.
- 116 Costinean S, Zanesi N, Pekarsky Y, Tili E, Volinia S, Heerema N *et al.* Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc Natl Acad Sci U S A* 2006; **103**: 7024–9.
- 117 Cui B, Chen L, Zhang S, Mraz M, Fecteau J-F, Yu J *et al.* MicroRNA-155 influences B-cell receptor signaling and associates with aggressive disease in chronic lymphocytic leukemia. *Blood* 2014; **124**: 546–54.
- 118 Dameshek W. Chronic lymphocytic leukemia—an accumulative disease of immunologically incompetent lymphocytes. *Blood* 1967; **29**: Suppl:566-84.
- 119 Messmer BT, Messmer D, Allen SL, Kolitz JE, Kudalkar P, Cesar D *et al.* In vivo measurements document the dynamic cellular kinetics of chronic lymphocytic leukemia B cells. *J Clin Invest* 2005; **115**: 755–764.
- 120 Pascual V. Analysis of somatic mutation in five B cell subsets of human tonsil. *J Exp Med* 1994; **180**: 329–339.
- 121 Dono M, Zupo S, Augliera A, Burgio VL, Massara R, Melagrana A *et al.* Subepithelial B cells in the human palatine tonsil. II. Functional characterization. *Eur J Immunol* 1996; **26**: 2043–2049.
- 122 Seifert M, Sellmann L, Bloehdorn J, Wein F, Stilgenbauer S, Dürig J *et al.* Cellular origin and pathophysiology of chronic lymphocytic leukemia. *J Exp Med* 2012; **209**: 2183–98.
- 123 Chiorazzi N, Ferrarini M. Cellular origin(s) of chronic lymphocytic leukemia: cautionary notes and additional considerations and possibilities. *Blood* 2011; **117**: 1781–1791.
- 124 Schroeder HW, Dighiero G. The pathogenesis of chronic lymphocytic leukemia: Analysis of the antibody repertoire. *Immunol. Today*. 1994; **15**: 288–294.
- 125 Fais F, Ghiotto F, Hashimoto S, Sellars B, Valetto A, Allen SL *et al.* Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest* 1998; **102**: 1515–1525.

- 126 Klein U, Tu Y, Stolovitzky G a, Mattioli M, Cattoretti G, Husson H *et al.* Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J Exp Med* 2001; **194**: 1625–1638.
- 127 Rosenwald A, Alizadeh AA, Widhopf G, Simon R, Davis RE, Yu X *et al.* Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med* 2001; **194**: 1639–47.
- 128 Chiorazzi N, Rai KR, Ferrarini M. Chronic lymphocytic leukemia. *N Engl J Med* 2005; **352**: 804–815.
- 129 Dono M, Zupo S, Leanza N, Melioli G, Fogli M, Melagrana A *et al.* Heterogeneity of tonsillar subepithelial B lymphocytes, the splenic marginal zone equivalents. *J Immunol* 2000; **164**: 5596–5604.
- 130 Dono M, Burgio VL, Colombo M, Sciacchitano S, Reverberi D, Tarantino V *et al.* CD5+ B cells with the features of subepithelial B cells found in human tonsils. *Eur J Immunol* 2007; **37**: 2138–2147.
- 131 Damle RN, Ghiotto F, Valetto A, Albesiano E, Fais F, Yan XJ *et al.* B-cell chronic lymphocytic leukemia cells express a surface membrane phenotype of activated, antigen-experienced B lymphocytes. *Blood* 2002; **99**: 4087–4093.
- 132 Weill J-C, Weller S, Reynaud C-A. Human marginal zone B cells. *Annu Rev Immunol* 2009; **27**: 267–85.
- 133 Ghia P, Caligaris-Cappio F. The origin of B-cell chronic lymphocytic leukemia. *Semin Oncol* 2006; **33**: 150–156.
- 134 Herishanu Y, Pérez-Galán P, Liu D, Biancotto A, Pittaluga S, Vire B *et al.* The lymph node microenvironment promotes B-cell receptor signaling, NF-kappaB activation, and tumor proliferation in chronic lymphocytic leukemia. *Blood* 2011; **117**: 563–74.
- 135 Fabbri G, Dalla-Favera R. The molecular pathogenesis of chronic lymphocytic leukaemia. *Nat Rev Cancer* 2016; **16**: 145–162.
- 136 Stein H, Bonk A, Tolksdorf G, Lennert K, Rodt H, Gerdes J. Immunohistologic analysis of the organization of normal lymphoid tissue and non-Hodgkin's lymphomas. *J Histochem Cytochem* 1980; **28**: 746–760.
- 137 Ten Hacken E, Burger JA. Microenvironment interactions and B-cell receptor

- signaling in Chronic Lymphocytic Leukemia: Implications for disease pathogenesis and treatment. *Biochim Biophys Acta* 2016; **1863**: 401–413.
- 138 Burger JA, Burger M, Kipps TJ. Chronic Lymphocytic Leukemia B Cells Express Functional CXCR4 Chemokine Receptors that Mediate Spontaneous Migration Beneath Bone Marrow Stromal Cells. *Blood* 1999; **94**: 3658–3667.
- 139 Kurtova A V., Balakrishnan K, Chen R, Ding W, Schnabl S, Quiroga MP *et al.* Diverse marrow stromal cells protect CLL cells from spontaneous and drug-induced apoptosis: Development of a reliable and reproducible system to assess stromal cell adhesion-mediated drug resistance. *Blood* 2009; **114**: 4441–4450.
- 140 Lutzny G, Kocher T, Schmidt-Supprian M, Rudelius M, Klein-Hitpass L, Finch AJ *et al.* Protein kinase c-beta-dependent activation of {NF}-{\kappa}B in stromal cells is indispensable for the survival of chronic lymphocytic leukemia B cells in vivo. *Cancer Cell* 2013; **23**. doi:10.1016/j.ccr.2012.12.003.
- 141 Whiteside TL. Immune cells in the tumor microenvironment. Mechanisms responsible for functional and signaling defects. *Adv Exp Med Biol* 1998; **451**: 167–171.
- 142 Burger J a. Nurture versus Nature: The Microenvironment in Chronic Lymphocytic Leukemia. *Hematology* 2011; **2011**: 96–103.
- 143 Arruga F, Gizdic B, Serra S, Vaisitti T, Ciardullo C, Coscia M *et al.* Functional impact of NOTCH1 mutations in chronic lymphocytic leukemia. *Leukemia* 2014; **28**: 1060–70.
- 144 Os A, Bürgler S, Ribes AP, Funderud A, Wang D, Thompson KM *et al.* Chronic lymphocytic leukemia cells are activated and proliferate in response to specific T helper cells. *Cell Rep* 2013; **4**: 566–77.
- 145 Robak T. Therapy of chronic lymphocytic leukaemia with purine nucleoside analogues: facts and controversies. *Drugs Aging* 2005; **22**: 983–1012.
- 146 Lukenbill J, Kalaycio M. Fludarabine: A review of the clear benefits and potential harms. *Leuk. Res.* 2013; **37**: 986–994.
- 147 Chang JE, Kahl BS. Bendamustine for treatment of chronic lymphocytic leukemia. *Expert Opin Pharmacother* 2012; **13**: 1495–1505.
- 148 Wierda WG, Kipps TJ, Mayer J, Stilgenbauer S, Williams CD, Hellmann A *et al.* Ofatumumab as single-agent CD20 immunotherapy in fludarabine-refractory

- chronic lymphocytic leukemia. *J Clin Oncol* 2010; **28**: 1749–1755.
- 149 Hallek M, Fischer K, Fingerle-Rowson G, Fink A-M, Busch R, Mayer J *et al.* Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet* 2010; **376**: 1164–1174.
- 150 Gentile M, Zirlik K, Ciolli S, Mauro FR, Di Renzo N, Mastrullo L *et al.* Combination of bendamustine and rituximab as front-line therapy for patients with chronic lymphocytic leukaemia: multicenter, retrospective clinical practice experience with 279 cases outside of controlled clinical trials. *Eur J Cancer* 2016; **60**: 154–165.
- 151 Brown JR, O'Brien S, Kingsley CD, Eradat H, Pagel JM, Lymph J *et al.* Obinutuzumab plus fludarabine/cyclophosphamide or bendamustine in the initial therapy of CLL patients: The phase 1b GALTON trial. In: *Blood*. 2015, pp 2779–2785.
- 152 Hillmen P, Robak T, Janssens A, Babu KG, Kloczko J, Grosicki S *et al.* Chlorambucil plus ofatumumab versus chlorambucil alone in previously untreated patients with chronic lymphocytic leukaemia (COMPLEMENT 1): A randomised, multicentre, open-label phase 3 trial. *Lancet* 2015; **385**: 1873–1883.
- 153 Burger JA, Tedeschi A, Barr PM, Robak T, Owen C, Ghia P *et al.* Ibrutinib as initial therapy for patients with chronic lymphocytic leukemia. *N Engl J Med* 2015; **373**: 2425–2437.
- 154 Byrd JC, Harrington B, O'Brien S, Jones J a., Schuh A, Devereux S *et al.* Acalabrutinib (ACP-196) in Relapsed Chronic Lymphocytic Leukemia. *N Engl J Med* 2015; : 151207133057003.
- 155 De Rooij MFM, Kuil A, Geest CR, Eldering E, Chang BY, Buggy JJ *et al.* The clinically active BTK inhibitor PCI-32765 targets B-cell receptor- and chemokine-controlled adhesion and migration in chronic lymphocytic leukemia. *Blood* 2012; **119**: 2590–2594.
- 156 Herman SEM, Gordon AL, Wagner AJ, Heerema NA, Zhao W, Flynn JM *et al.* Phosphatidylinositol 3-kinase- $\delta$  inhibitor CAL-101 shows promising preclinical activity in chronic lymphocytic leukemia by antagonizing intrinsic and extrinsic cellular survival signals. *Blood* 2010; **116**: 2078–2088.
- 157 Blunt MD, Steele AJ. Pharmacological targeting of PI3K isoforms as a therapeutic strategy in chronic lymphocytic leukaemia. *Leuk. Res. Reports*. 2015; **4**: 60–63.

- 158 Friedberg JW, Sharman J, Sweetenham J, Johnston PB, Vose JM, LaCasce A *et al.* Inhibition of Syk with fostamatinib disodium has significant clinical activity in non-Hodgkin lymphoma and chronic lymphocytic leukemia. *Blood* 2010; **115**: 2578–2585.
- 159 Buchner M, Baer C, Prinz G, Dierks C, Burger M, Zenz T *et al.* Spleen tyrosine kinase inhibition prevents chemokine- and integrin-mediated stromal protective effects in chronic lymphocytic leukemia. *Blood* 2010; **115**: 4497–4506.
- 160 Roberts AW, Davids MS, Pagel JM, Kahl BS, Puvvada SD, Gerecitano JF *et al.* Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia. *N Engl J Med* 2016; **374**: 311–322.
- 161 Rossi D, Gaidano G. Richter syndrome: Pathogenesis and management. *Semin Oncol* 2016; **43**: 311–319.
- 162 Jain P, Young KH. Haematological cancer: Richter's transformation in CLL-a distinct lymphoma. *Nat Rev Clin Oncol* 2014; **11**: 6–8.
- 163 Hisada M, Biggar RJ, Greene MH, Fraumeni JF, Travis LB. Solid tumors after chronic lymphocytic leukemia. *Blood* 2001; **98**: 1979–1981.
- 164 Wilkins MR, Sanchez J-C, Gooley A a., Appel RD, Humphery-Smith I, Hochstrasser DF *et al.* Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 1996; **13**: 19–50.
- 165 Graves PR, Haystead TAJ. Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev* 2002; **66**: 39–+.
- 166 Larance M, Lamond AI. Multidimensional proteomics for cell biology. *Nat Rev Mol Cell Biol* 2015; **16**: 269–280.
- 167 Consortium IHGS. Initial sequencing and analysis of the human genome. *Nature* 2001; **409**: 860–921.
- 168 Hattori M. Finishing the euchromatic sequence of the human genome. *Tanpakushitsu Kakusan Koso* 2005; **50**: 162–168.
- 169 Ansong C, Purvine SO, Adkins JN, Lipton MS, Smith RD. Proteogenomics: Needs and roles to be filled by proteomics in genome annotation. *Briefings Funct Genomics Proteomics* 2008; **7**: 50–62.

- 170 Muñoz J, Heck AJR. From the Human Genome to the Human Proteome. *Angew Chem Int Ed Engl* 2014; : 10864–10866.
- 171 Jensen ON. Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol* 2006; **7**: 391–403.
- 172 Rabilloud T, Hochstrasser D, Simpson RJ. A gene-centric human proteome project: HUPO--the Human Proteome organization. *Mol Cell Proteomics* 2010; **9**: 427–9.
- 173 Legrain P, Aebersold R, Archakov A, Bairoch A, Bala K, Beretta L *et al.* The human proteome project: current state and future direction. *Mol Cell Proteomics* 2011; **10**: M111.009993.
- 174 Paik Y-K, Jeong S-K, Omenn GS, Uhlen M, Hanash S, Cho SY *et al.* The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat Biotechnol* 2012; **30**: 221–223.
- 175 Paik YK, Omenn GS, Uhlen M, Hanash S, Marko-Varga G, Aebersold R *et al.* Standard guidelines for the chromosome-centric human proteome project. *J. Proteome Res.* 2012; **11**: 2005–2013.
- 176 Hühmer AFR, Paulus A, Martin LB, Millis K, Agreste T, Saba J *et al.* The chromosome-centric human proteome project: A call to action. *J. Proteome Res.* 2013; **12**: 28–32.
- 177 Gaudet P, Michel PA, Zahn-Zabal M, Britan A, Cusin I, Domagalski M *et al.* The neXtProt knowledgebase on human proteins: 2017 update. *Nucleic Acids Res* 2017; **45**: D177–D182.
- 178 Segura V, Medina-Aunon JA, Gुरुceaga E, Gharbi SI, González-Tejedo C, San Chez Del Pino MM *et al.* Spanish human proteome project: Dissection of chromosome 16. *J Proteome Res* 2013; **12**: 112–122.
- 179 Segura V, Medina-Aunon JA, Mora MI, Martínez-Bartolomé S, Abian J, Aloria K *et al.* Surfing transcriptomic landscapes. A step beyond the annotation of chromosome 16 proteome. *J Proteome Res* 2014; **13**: 158–172.
- 180 Baker MS, Ahn SB, Mohamedali A, Islam MT, Cantor D, Verhaert PD *et al.* Accelerating the search for the missing proteins in the human proteome. *Nat Commun* 2017; **8**: 14271.
- 181 Van Eyk JE, Corrales FJ, Aebersold R, Cerciello F, Deutsch EW, Roncada P *et al.*

- Highlights of the Biology and Disease-driven Human Proteome Project, 2015-2016. *J. Proteome Res.* 2016; **15**: 3979–3987.
- 182 Aebersold R, Bader GD, Edwards AM, Van Eyk JE, Kussmann M, Qin J *et al.* The biology/disease-driven human proteome project (B/D-HPP): Enabling protein research for the life sciences community. *J. Proteome Res.* 2013; **12**: 23–27.
- 183 Aebersold R, Bader GD, Edwards AM, van Eyk J, Kussman M, Qin J *et al.* Highlights of B/D-HPP and HPP Resource Pillar Workshops at 12th Annual HUPO World Congress of Proteomics: September 14-18, 2013, Yokohama, Japan R. Aebersold *et al.* *Proteomics* 2014; **14**: 975–988.
- 184 Topf F, Schwartz D, Gaudet P, Priego-Capote F, Zufferey A, Turck N *et al.* The Human Diabetes Proteome Project (HDPP): From network biology to targets for therapies and prevention. *Transl. Proteomics.* 2013; **1**: 3–11.
- 185 Schwartz D, Bergsten P, Baek KH, De La Rosa AB, Cantley J, Dayon L *et al.* The Human Diabetes Proteome Project (HDPP): The 2014 update. *Transl Proteomics* 2015; **8–9**: 1–7.
- 186 Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE *et al.* Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* 2016; **166**: 755–765.
- 187 Mertins P, Clauser KR, Mani DR, Gillette M, Fenyo D, Wang P *et al.* 36: Proffered Paper: Proteogenomic analysis of human breast cancer connects genetic alterations to phosphorylation networks. *Eur J Cancer* 2017; **50**: S10.
- 188 Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z *et al.* Proteogenomic characterization of human colon and rectal cancer. *Nature* 2014; **513**: 382–7.
- 189 Ruiz-Romero C, Calamia V, Albar JP, Casal JI, Corrales FJ, Fernández-Puente P *et al.* The Spanish biology/disease initiative within the human proteome project: Application to rheumatic diseases. *J. Proteomics.* 2015; **127**: 406–413.
- 190 He F. Human liver proteome project: plan, progress, and perspectives. *Mol Cell Proteomics* 2005; **4**: 1841–8.
- 191 Lam MPY, Venkatraman V, Xing Y, Lau E, Cao Q, Ng DCM *et al.* Data-Driven Approach to Determine Popular Proteins for Targeted Proteomics Translation of Six Organ Systems. *J Proteome Res* 2016; **15**: 4126–4134.
- 192 Semba RD, Enghild JJ, Venkatraman V, Dyrland TF, Van Eyk JE. The Human Eye

- Proteome Project: Perspectives on an emerging proteome. *Proteomics*. 2013; **13**: 2500–2511.
- 193 Omenn GS. The human eye proteome project. *Proteomics*. 2013; **13**: 2375–2376.
- 194 Hamacher M, Marcus K, Stephan C, Klose J, Park YM, Meyer HE. HUPO Brain Proteome Project: toward a code of conduct. *Mol Cell Proteomics* 2008; **7**: 457.
- 195 Ito H, Kaji H, Togayachi A, Azadi P, Ishihara M, Geyer R *et al*. Comparison of analytical methods for profiling N- and O-linked glycans from cultured cell lines: HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycoconj J* 2016; **33**: 405–415.
- 196 Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H *et al*. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* 2005; **5**: 3226–3245.
- 197 Cominetti O, Nájera Galindo A, Corthesy J, Oller Moreno S, Irincheeva I, Valsesia A *et al*. Proteomic Biomarker Discovery in 1000 Human Plasma Samples with Mass Spectrometry. *J Proteome Res* 2016; **15**: 389–399.
- 198 Fert-Bober J, Giles JT, Holewinski RJ, Kirk JA, Uhrigshardt H, Crowgey EL *et al*. Citrullination of myofilament proteins in heart failure. *Cardiovasc Res* 2015; **108**: 232–242.
- 199 Berhane BT, Zong C, Liem DA, Huang A, Le S, Edmondson RD *et al*. Cardiovascular-related proteins identified in human plasma by the HUPO plasma proteome project pilot phase. *Proteomics* 2005; **5**: 3520–3530.
- 200 Yamamoto T, Langham RG, Ronco P, Knepper MA, Thongboonkerd V. Towards standard protocols and guidelines for urine proteomics: a report on the Human Kidney and Urine Proteome Project (HKUPP) symposium and workshop, 6 October 2007, Seoul, Korea and 1 November 2007, San Francisco, CA, USA. *Proteomics*. 2008; **8**: 2156–2159.
- 201 Farrah T, Deutsch EW, Omenn GS, Sun Z, Watts JD, Yamamoto T *et al*. State of the human proteome in 2013 as viewed through peptideatlas: Comparing the kidney, urine, and plasma proteomes for the biology- and disease-driven human proteome project. *J Proteome Res* 2014; **13**: 60–75.



- 
- 202 Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* 2010; **11**: 345–355.
- 203 Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol* 2003; **21**: 255–61.
- 204 Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. *Nature* 2016; **537**: 347–355.
- 205 Khoury GA, Baliban RC, Floudas CA. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep* 2011; **1**: 90.
- 206 Garcia-Garcia T, Poncet S, Derouiche A, Shi L, Mijakovic I, Noirot-Gros MF. Role of protein phosphorylation in the regulation of cell cycle and DNA-related processes in bacteria. *Front. Microbiol.* 2016; **7**. doi:10.3389/fmicb.2016.00184.
- 207 Lawrence RT, Searle BC, Llovet A, Villen J. Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. *Nat Methods* 2016; **13**: 431–434.
- 208 Mann M, Ong S, Gr M, Steen H, Jensen ON, Pandey A. Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol* 2002; **20**: 261–8.
- 209 Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR *et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* 2011; **480**: 254–8.
- 210 Chandramouli K, Qian P-Y. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Hum Genomics Proteomics* 2009; **2009**: 22.
- 211 Cristea IM, Gaskell SJ, Whetton AD. Proteomics techniques and their application to hematology. *Blood.* 2004; **103**: 3624–3634.
- 212 Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003; **422**: 198–207.
- 213 Köcher T, Superti-Furga G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* 2007; **4**: 807–15.
- 214 Tinnikov AA, Samuels HH. A Novel Cell Lysis Approach Reveals That Caspase-2

- Rapidly Translocates from the Nucleus to the Cytoplasm in Response to Apoptotic Stimuli. *PLoS One* 2013; **8**. doi:10.1371/journal.pone.0061085.
- 215 Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science* 2006; **312**: 212–7.
- 216 Pandey a, Mann M. Proteomics to study genes and genomes. *Nature* 2000; **405**: 837–846.
- 217 Aebersold R, Goodlett DR. Mass spectrometry in proteomics. *Chem Rev* 2001; **101**: 269–295.
- 218 Hager JW. A new linear ion trap mass spectrometer. *Rapid Commun Mass Spectrom* 2002; **16**: 512–526.
- 219 Mamyrin BA. Time-of-flight mass spectrometry (concepts, achievements, and prospects). *Int J Mass Spectrom* 2001; **206**: 251–266.
- 220 Clark AE, Kaleta EJ, Arora A, Wolk DM. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clin Microbiol Rev* 2013; **26**: 547–603.
- 221 Marshall a G, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* 1998; **17**: 1–35.
- 222 Walther TC, Mann M. Mass spectrometry-based proteomics in cell biology. *J Cell Biol* 2010; **190**: 491–500.
- 223 Ong S-E, Mann M. Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* 2005; **1**: 252–62.
- 224 Zhu W, Smith JW, Huang C-M. Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* 2010; **2010**: 840518.
- 225 Shinoda K, Tomita M, Ishihama Y. emPAI Calc-for the estimation of protein abundance from large-scale identification data by liquid chromatography-tandem mass spectrometry. *Bioinformatics* 2009; **26**: 576–577.
- 226 Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J *et al*. Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Mol Cell Proteomics* 2005; **4**: 1265–1272.

- 227 Bubis JA, Levitsky LI, Ivanov M V, Tarasova IA, Gorshkov M V. Comparative evaluation of label-free quantification methods for shotgun proteomics. *Rapid Commun Mass Spectrom* 2017; **31**: 606–612.
- 228 Chahrour O, Cobice D, Malone J. Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *J. Pharm. Biomed. Anal.* 2015; **113**: 2–20.
- 229 Mann M. Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* 2006; **7**: 952–958.
- 230 Burkhardt JM, Vaudel M, Zahedi RP, Martens L, Sickmann A. iTRAQ protein quantification: A quality-controlled workflow. *Proteomics* 2011; **11**: 1125–1134.
- 231 Shiiro Y, Aebersold R. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. *Nat Protoc* 2006; **1**: 139–145.
- 232 Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999; **17**: 994–9.
- 233 Lim YP. Mining the tumor phosphoproteome for cancer markers. *Clin. Cancer Res.* 2005; **11**: 3163–3169.
- 234 Riley NM, Coon JJ. Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal Chem* 2016; **88**: 74–94.
- 235 Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* 2006; **24**: 1285–1292.
- 236 Stokes MP, Farnsworth CL, Moritz a., Silva JC, Jia X, Lee K a. et al. PTMScan Direct: Identification and Quantification of Peptides from Critical Signaling Proteins by Immunoaffinity Enrichment Coupled with LC-MS/MS. *Mol Cell Proteomics* 2012; **11**: 187–201.
- 237 Aryal UK, Ross ARS. Enrichment and analysis of phosphopeptides under different experimental conditions using titanium dioxide affinity chromatography and mass spectrometry. *Rapid Commun Mass Spectrom* 2010; **24**: 219–231.
- 238 Cutillas PR, Timms JF. Approaches and applications of quantitative LC-MS for proteomics and activitomics. *Methods Mol Biol* 2010; **658**: 3–17.

- 239 Olsen J V, Vermeulen M, Santamaria A, Kumar C, Miller ML, Jensen LJ *et al.* Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* 2010; **3**: ra3.
- 240 Ye J, Zhang X, Young C, Zhao X, Hao Q, Cheng L *et al.* Optimized IMAC-IMAC protocol for phosphopeptide recovery from complex biological samples. *J Proteome Res* 2010; **9**: 3561–3573.
- 241 Thingholm TE, Jensen ON. Enrichment and Characterization of Phosphopeptides by Immobilized Metal Affinity Chromatography (IMAC) and Mass Spectrometry. In: Grauw M de (ed). *Phospho-Proteomics: Methods and Protocols*. Humana Press: Totowa, NJ, 2009, pp 47–56.
- 242 Reinders J, Lewandrowski U, Moebius J, Wagner Y, Sickmann A. Challenges in mass spectrometry-based proteomics. *Proteomics*. 2004; **4**: 3686–3703.
- 243 Choudhary C, Mann M. Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol* 2010; **11**: 427–439.
- 244 Fleuren EDG, Zhang L, Wu J, Daly RJ. The kinome 'at large' in cancer. *Nat Rev Cancer* 2016; **16**: 83–98.
- 245 Wiestner a. Emerging role of kinase targeted strategies in chronic lymphocytic leukemia. *Blood* 2012; **120**: 88–96.
- 246 Chandra H, Reddy PJ, Srivastava S. Protein microarrays and novel detection platforms. *Expert Rev Proteomics* 2011; **8**: 61–79.
- 247 Dasilva N, Díez P, González-González M, Matarraz S, Sayagués JM, Orfao A *et al.* Protein microarrays: technological aspects, applications and intellectual property. *Recent Pat Biotechnol* 2013; **7**: 142–52.
- 248 Melton L. Protein arrays: proteomics in multiplex. *Nature* 2004; **429**: 101–107.
- 249 Gonzalez-Gonzalez M, Jara-Acevedo R, Matarraz S, Jara-Acevedo M, Paradinas S, Sayagües JM *et al.* Nanotechniques in proteomics: Protein microarrays and novel detection platforms. In: *European Journal of Pharmaceutical Sciences*. 2012, pp 499–506.
- 250 Romanov V, Davidoff SN, Miles AR, Grainger DW, Gale BK, Brooks BD. A critical comparison of protein microarray fabrication technologies. *Analyst* 2014; **139**: 1303–1326.

- 
- 251 Díez P, Dasilva N, González-González M, Matarraz S, Casado-Vela J, Orfao A *et al.* Data Analysis Strategies for Protein Microarrays. *Microarrays* 2012; **1**: 64–83.
- 252 Zhu H, Snyder M. Protein arrays and microarrays. *Curr. Opin. Chem. Biol.* 2001; **5**: 40–45.
- 253 Moore CD, Ajala OZ, Zhu H. Applications in high-content functional protein microarrays. *Curr Opin Chem Biol* 2016; **30**: 21–27.
- 254 Gonzalez-Gonzalez M, Bartolome R, Jara-Acevedo R, Casado-Vela J, Dasilva N, Matarraz S *et al.* Evaluation of homo- and hetero-functionally activated glass surfaces for optimized antibody arrays. *Anal Biochem* 2014; **450**: 37–45.
- 255 Sauer U. Analytical Protein Microarrays: Advancements Towards Clinical Applications. *Sensors (Basel)* 2017; **17**. doi:10.3390/s17020256.
- 256 McWilliam I, Chong Kwan M, Hall D. Inkjet printing for the production of protein microarrays. *Methods Mol Biol* 2011; **785**: 345–61.
- 257 Chen R, Yu X-L, Gao X-B, Xue C-Y, Song C-X, Li Y *et al.* Bead-based suspension array for simultaneous differential detection of five major swine viruses. *Appl Microbiol Biotechnol* 2015; **99**: 919–928.
- 258 Wingren C. Planar Antibody Arrays for Biomarkers in Nephritis. In: Patel VB (ed). *Biomarkers in Kidney Disease*. Springer Netherlands: Dordrecht, 2015, pp 1–17.
- 259 Petersson L, Coen M, Amro N a, Truedsson L, Borrebaeck C a K, Wingren C. Miniaturization of multiplexed planar recombinant antibody arrays for serum protein profiling. *Bioanalysis* 2014; **6**: 1175–85.
- 260 Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 2008; **26**: 1367–72.
- 261 Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen J V., Mann M. Andromeda: A peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 2011; **10**: 1794–1805.
- 262 MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B *et al.* Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010; **26**: 966–968.
- 263 Vaudel M, Burkhardt JM, Zahedi RP, Oveland E, Berven FS, Sickmann A *et al.*

- PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat Biotechnol* 2015; **33**: 22–24.
- 264 Wenger CD, Phanstiel DH, Lee MV, Bailey DJ, Coon JJ. COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics* 2011; **11**: 1064–1074.
- 265 Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC *et al*. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003; **4**: P3.
- 266 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al*. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13**: 2498–2504.
- 267 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette M *a et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; **102**: 15545–50.
- 268 Deutsch EW, Csordas A, Sun Z, Jarnuczak A, Perez-Riverol Y, Ternent T *et al*. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res* 2016; : gkw936.
- 269 Beavis RC. Using the global proteome machine for protein identification. *Methods Mol Biol* 2006; **328**: 217–228.
- 270 Desiere F. The PeptideAtlas project. *Nucleic Acids Res* 2006; **34**: D655–D658.
- 271 Smith BE, Hill JA, Gjukich MA, Andrews PC. Tranche Distributed Repository and ProteomeCommons.org. In: *Methods in molecular biology (Clifton, N.J.)*. 2011, pp 123–145.
- 272 Kandasamy K, Keerthikumar S, Goel R, Mathivanan S, Patankar N, Shafreen B *et al*. Human Proteinpedia: A unified discovery resource for proteomics research. *Nucleic Acids Res* 2009; **37**. doi:10.1093/nar/gkn701.
- 273 Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng* 2009; **11**: 49–79.
- 274 Gatto L, Vizcaíno JA, Hermjakob H, Huber W, Lilley KS. Organelle proteomics experimental designs and analysis. *Proteomics*. 2010; **10**: 3957–3969.
- 275 Drissi R, Dubois ML, Boisvert FM. Proteomics methods for subcellular proteome

- analysis. *FEBS J* 2013; **280**: 5626–5634.
- 276 Lee YH, Tan HT, Chung MCM. Subcellular fractionation methods and strategies for proteomics. *Proteomics*. 2010; **10**: 3935–3956.
- 277 Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T *et al.* Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* 2011; **11**: 2019–2026.
- 278 Wang D, Bodovitz S. Single cell analysis: The new frontier in 'omics'. *Trends Biotechnol.* 2010; **28**: 281–290.
- 279 Salehi-Reyhani A, Kaplinsky J, Burgin E, Novakova M, DeMello AJ, Templer RH *et al.* A first step towards practical single cell proteomics: a microfluidic antibody capture chip with TIRF detection. *Lab Chip* 2011; **11**: 1256–1261.
- 280 Irish JM, Kotecha N, Nolan GP. Mapping normal and cancer cell signalling networks: towards single-cell proteomics. *Nat Rev Cancer* 2006; **6**: 146–155.
- 281 Angel TE, Aryal UK, Hengel SM, Baker ES, Kelly RT, Robinson EW *et al.* Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem Soc Rev* 2012; **41**: 3912–28.
- 282 Parker H, Rose-Zerilli MJJ, Parker A, Chaplin T, Wade R, Gardiner A *et al.* 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. *Leukemia* 2011; **25**: 489–497.
- 283 Van Dyke DL, Shanafelt TD, Call TG, Zent CS, Smoley S a, Rabe KG *et al.* A comprehensive evaluation of the prognostic significance of 13q deletions in patients with B-chronic lymphocytic leukaemia. *Br J Haematol* 2009; **148**: 544–550.
- 284 Döhner H, Stilgenbauer S, James MR, Benner A, Weilguni T, Bentz M *et al.* 11q deletions identify a new subset of B-cell chronic lymphocytic leukemia characterized by extensive nodal involvement and inferior prognosis. *Blood* 1997; **89**: 2516–22.
- 285 Neilson JR, Auer R, White D, Bienz N, Waters JJ, Whittaker JA *et al.* Deletions at 11q identify a subset of patients with typical CLL who show consistent disease progression and reduced survival. *Leukemia* 1997; **11**: 1929–1932.
- 286 Rose-Zerilli MJJ, Forster J, Parker H, Parker A, Rodriguez AE, Chaplin T *et al.* ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in

- 11q-deleted chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. *Haematologica* 2014; **99**: 736–742.
- 287 Zenz T, Eichhorst B, Busch R, Denzel T, Häbe S, Winkler D *et al.* TP53 mutation and survival in chronic lymphocytic leukemia. *J Clin Oncol* 2010; **28**: 4473–4479.
- 288 Rossi D, Cerri M, Deambrogi C, Sozzi E, Cresta S, Rasi S *et al.* The prognostic value of TP53 mutations in chronic lymphocytic leukemia is independent of Del17p13: Implications for overall survival and chemorefractoriness. *Clin Cancer Res* 2009; **15**: 995–1004.
- 289 Riches JC, O'Donovan CJ, Kingdon SJ, McClanahan F, Clear AJ, Neuberg DS *et al.* Trisomy 12 chronic lymphocytic leukemia cells exhibit upregulation of integrin signaling that is modulated by NOTCH1 mutations. *Blood* 2014; **123**: 4101–4110.
- 290 Quijano S, Lopez A, Rasillo A, Sayagues JM, Barrena S, Sanchez ML *et al.* Impact of trisomy 12, del(13q), del(17p), and del(11q) on the immunophenotype, DNA ploidy status, and proliferative rate of leukemic B-cells in chronic lymphocytic leukemia. *Cytometry B Clin Cytom* 2008; **74**: 139–149.
- 291 Losada AP, Wessman M, Tiainen M, Hopman AH, Willard HF, Sole F *et al.* Trisomy 12 in chronic lymphocytic leukemia: an interphase cytogenetic study. *Blood* 1991; **78**: 775–779.
- 292 Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI *et al.* Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* 2015; **526**: 519–524.
- 293 Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000; **100**: 57–70.
- 294 Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell* 2010; **141**: 1117–1134.
- 295 Casado P, Alcolea MP, Iorio F, Rodríguez-Prados J-C, Vanhaesebroeck B, Saez-Rodriguez J *et al.* Phosphoproteomics data classify hematological cancer cell lines according to tumor type and sensitivity to kinase inhibitors. *Genome Biol* 2013; **14**: R37.
- 296 Wiestner A. Emerging role of kinase-targeted strategies in chronic lymphocytic leukemia. *Blood*. 2012; **120**: 4684–4691.
- 297 Caligaris-Cappio F, Ghia P. Novel insights in chronic lymphocytic leukemia: are we getting closer to understanding the pathogenesis of the disease? *J Clin*



- Oncol* 2008; **26**: 4497–503.
- 298 Rush J, Moritz A, Lee KA, Guo A, Goss VL, Spek EJ *et al.* Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol* 2005; **23**: 94–101.
- 299 Wong DM, Li L, Jurado S, King A, Bamford R, Wall M *et al.* The Transcription Factor ASCIZ and Its Target DYNLL1 Are Essential for the Development and Expansion of MYC-Driven B Cell Lymphoma. *Cell Rep* 2016; **14**: 1488–1499.
- 300 Renault TT, Chipuk JE. Getting away with murder: how does the BCL-2 family of proteins kill with immunity? *Ann N Y Acad Sci* 2013; **1285**: 59–79.
- 301 Kristensen L, Kristensen T, Abildgaard N, Thomassen M, Frederiksen M, Mourits-Andersen T *et al.* High expression of PI3K core complex genes is associated with poor prognosis in chronic lymphocytic leukemia. *Leuk Res* 2015; **39**: 555–560.
- 302 Sakamoto KM, Grant S, Saleiro D, Crispino JD, Hijiya N, Giles F *et al.* Targeting novel signaling pathways for resistant acute myeloid leukemia. *Mol. Genet. Metab.* 2015; **114**: 397–402.
- 303 Oberley MJ, Wang D-S, Yang DT. Vav1 in hematologic neoplasms, a mini review. *Am J Blood Res* 2012; **2**: 1–8.
- 304 Chen L, Widhopf G, Huynh L, Rassenti L, Rai KR, Weiss A *et al.* Expression of ZAP-70 is associated with increased B-cell receptor signaling in chronic lymphocytic leukemia. *Blood* 2002; **100**: 4609–4614.
- 305 Kelly PN, Strasser A. The role of Bcl-2 and its pro-survival relatives in tumourigenesis and cancer therapy. *Cell Death Differ* 2011; **18**: 1414–1424.
- 306 Ganghammer S, Gutjahr J, Hutterer E, Krenn PW, Pucher S, Zelle-Rieser C *et al.* Combined CXCR3/CXCR4 measurements are of high prognostic value in chronic lymphocytic leukemia due to negative co-operativity of the receptors. *Haematologica* 2016; **101**: e99-102.
- 307 Mongini PKA, Gupta R, Boyle E, Nieto J, Lee H, Stein J *et al.* TLR-9 and IL-15 Synergy Promotes the In Vitro Clonal Expansion of Chronic Lymphocytic Leukemia B Cells. *J Immunol* 2015; **195**: 901–23.
- 308 Davids MS, Brown JR. Targeting the B Cell Receptor Pathway in Chronic Lymphocytic Leukemia. *Leuk Lymphoma* 2012; **8194**: 1–23.
- 309 Herman SE, Gordon AL, Hertlein E, Ramanunni A, Zhang X, Jaglowski S *et al.*

- Bruton tyrosine kinase represents a promising therapeutic target for treatment of chronic lymphocytic leukemia and is effectively targeted by PCI-32765. *Blood* 2011; **117**: 6287–6296.
- 310 Shinohara H, Kurosaki T. Comprehending the complex connection between PKC $\zeta$ , TAK1, and IKK in BCR signaling. *Immunol. Rev.* 2009; **232**: 300–318.
- 311 Bruno S, Ledda B, Tenca C, Ravera S, Orengo AM, Mazzarello AN *et al.* Metformin inhibits cell cycle progression of B-cell chronic lymphocytic leukemia cells. *Oncotarget* 2015; **6**: 22624–22640.
- 312 Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R *et al.* Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* 2012; **120**: 4191–4196.
- 313 Rossi D, Rasi S, Fabbri G, Spina V, Fangazio M, Forconi F *et al.* Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood* 2011; **119**: 521–529.
- 314 Ramsay AJ, Quesada V, Foronda M, Conde L, Martínez-Trillos A, Villamor N *et al.* POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat Genet* 2013; **45**: 526–530.
- 315 Valdes-Mas R, Gutierrez-Abril J, Puente XS, Lopez-Otin C. Chronic lymphocytic leukemia: looking into the dark side of the genome. *Cell Death Differ* 2016; **23**: 7–9.
- 316 Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2009; **38**. doi:10.1093/nar/gkp896.
- 317 O'Neill LAJ, Bowie AG. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat Rev Immunol* 2007; **7**: 353–364.
- 318 Ruland J, Mak TW. Transducing signals from antigen receptors to nuclear factor kappaB. *Immunol Rev* 2003; **193**: 93–100.
- 319 Crassini K, Stevenson W s, Mulligan SP, FRCPA. MAPK-Erk1/2 Pathway Activation and Survival Of Chronic Lymphocytic Leukemia (CLL) Cells Is Induced By &lt;em>In Vitro</em> Modeling Of The Tumor Microenvironment and Can Be Effectively Inhibited Using The MEK1 Inhibitor MEK162. *Blood* 2013; **122**: 4192 LP-4192.

- 320 Cuni S, Perez-Aciego P, Perez-Chacon G, Vargas JA, Sanchez A, Martin-Saavedra FM *et al.* A sustained activation of PI3K/NF-kappaB pathway is critical for the survival of chronic lymphocytic leukemia B cells. *Leukemia* 2004; **18**: 1391–1400.
- 321 Murn J, Mlinaric-Rascan I, Vaigot P, Alibert O, Frouin V, Gidrol X. A Myc-regulated transcriptional network controls B-cell fate in response to BCR triggering. *BMC Genomics* 2009; **10**: 323.
- 322 Thurgood LA, Chataway TK, Lower KM, Kuss BJ. From genome to proteome: Looking beyond DNA and RNA in chronic lymphocytic leukemia. *J Proteomics* 2017; **155**: 73–84.
- 323 Pradet-Balade B, Boulmé F, Beug H, Müllner EW, Garcia-Sanz JA. Translation control: Bridging the gap between genomics and proteomics? *Trends Biochem. Sci.* 2001; **26**: 225–229.
- 324 Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 2012; **13**: 227–232.
- 325 Hoxha M, Fabris S, Agnelli L, Bollati V, Cutrona G, Matis S *et al.* Relevance of Telomere/Telomerase System Impairment in Early Stage Chronic Lymphocytic Leukemia. *Genes Chromosomes Cancer* 2014; **53**: 612–621.
- 326 Poncet D, Belleville A, De Roodenbeke CTK, De Climens AR, Simon E Ben, Merle-Beral H *et al.* Changes in the expression of telomere maintenance genes suggest global telomere dysfunction in B-chronic lymphocytic leukemia. *Blood* 2008; **111**: 2388–2391.
- 327 Maitta RW, Wolgast LR, Wang Q, Zhang H, Bhattacharyya P, Gong JZ *et al.* Alpha- and beta-synucleins are new diagnostic tools for acute erythroid leukemia and acute megakaryoblastic leukemia. *Am. J. Hematol.* 2011; **86**: 230–234.
- 328 Gradowski JF, Sargent RL, Craig FE, Cieply K, Fuhrer K, Sherer C *et al.* Chronic lymphocytic leukemia/small lymphocytic lymphoma with cyclin D1 positive proliferation centers do not have CCND1 translocations or gains and lack SOX11 expression. *Am J Clin Pathol* 2012; **138**: 132–139.
- 329 Li J, Yuan J. Caspases in apoptosis and beyond. *Oncogene* 2008; **27**: 6194–6206.
- 330 Hengartner MO. The biochemistry of apoptosis. *Nature* 2000; **407**: 770–6.

- 331 Nooti SK, McKenna MK, Frizzera FW, Gachuki BW, Alhakeem SS, Greene JT *et al.* Role of Prostate Apoptosis Response-4 Tumor Suppressor in the Survival and Growth of Chronic Lymphocytic Leukemia. *Blood* 2015; **126**: 1731 LP-1731.
- 332 Alkan S, Huang Q, Ergin M, Denning MF, Nand S, Maududi T *et al.* Survival role of protein kinase C (PKC) in chronic lymphocytic leukemia and determination of isoform expression pattern and genes altered by PKC inhibition. *Am J Hematol* 2005; **79**: 97–106.
- 333 Foy SP, Matsuuchi L. Association of B lymphocyte antigen receptor polypeptides with multiple chaperone proteins. *Immunol Lett* 2001; **78**: 149–160.
- 334 Liu X shan, Jiang J, Jiao X yang, Wu Y e., Lin J hua, Cai Y mu. Lycorine induces apoptosis and down-regulation of Mcl-1 in human leukemia cells. *Cancer Lett* 2009; **274**: 16–24.
- 335 Malumbres M, Sotillo R, Santamaría D, Galán J, Cerezo A, Ortega S *et al.* Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. *Cell* 2004; **118**: 493–504.
- 336 Stein GH, Drullinger LF, Soulard A, Dulić V. Differential roles for cyclin-dependent kinase inhibitors p21 and p16 in the mechanisms of senescence and differentiation in human fibroblasts. *Mol Cell Biol* 1999; **19**: 2109–17.
- 337 Bellavia D, Campese A F, Alesse E, Vacca A, Felli MP, Balestri A *et al.* Constitutive activation of NF-kappaB and T-cell leukemia/lymphoma in Notch3 transgenic mice. *EMBO J* 2000; **19**: 3337–48.
- 338 Rosati E, Sabatini R, Rampino G, Tabilio A, Di Ianni M, Fettucciari K *et al.* Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells. *Blood* 2009; **113**: 856–865.
- 339 Parikh SA, Shanafelt TD. Risk Factors for Richter Syndrome in Chronic Lymphocytic Leukemia. *Curr. Hematol. Malig. Rep.* 2014; **9**: 294–299.
- 340 Shaulian E, Karin M. AP-1 in cell proliferation and survival. *Oncogene* 2001; **20**: 2390–2400.
- 341 Weber HO, Samuel T, Rauch P, Funk JO. Human p14(ARF)-mediated cell cycle arrest strictly depends on intact p53 signaling pathways. *Oncogene* 2002; **21**: 3207–12.
- 342 Li A, Rue M, Zhou J, Wang H, Goldwasser MA, Neuberg D *et al.* Utilization of Ig

- heavy chain variable, diversity, and joining gene segments in children with B-lineage acute lymphoblastic leukemia: implications for the mechanisms of VDJ recombination and for pathogenesis. *Blood* 2004; **103**: 4602–4609.
- 343 Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D, Scharff MD. The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev* 2004; **18**: 1–11.
- 344 Tonegawa S. Somatic generation of antibody diversity. *Nature* 1983; **302**: 575–581.
- 345 Odegard VH, Schatz DG. Targeting of somatic hypermutation. *Nat Rev Immunol* 2006; **6**: 573–83.
- 346 Jordan M a, Baxter AG. Quantitative and qualitative approaches to GOD: the first 10 years of the clonal selection theory. *Immunol Cell Biol* 2008; **86**: 72–9.
- 347 Boutz DR, Horton AP, Wine Y, Lavinder JJ, Georgiou G, Marcotte EM. Proteomic identification of monoclonal antibodies from serum. *Anal Chem* 2014; **86**: 4758–4766.
- 348 Brown JR, Fernandes SM, Kasar S, Hoang K, Moorhead M, Carlton V *et al*. Next-Generation Sequencing Reveals Clonal Evolution at the Immunoglobulin Loci in Chronic Lymphocytic Leukemia. *Blood* 2014; **124**: 3302–3302.
- 349 Robins H. Immunosequencing: applications of immune repertoire deep sequencing. *Curr. Opin. Immunol.* 2013; **25**: 646–652.
- 350 Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* 2014; **32**: 158–168.
- 351 Toung JM, Morley M, Li M, Cheung VG. RNA-sequence analysis of human B-cells. *Genome Res* 2011; **21**: 991–998.
- 352 Clamp M, Fry B, Kamal M, Xie X, Cuff J, Lin MF KM, Lindblad-Toh K SLE. Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A* 2007; **104**: 19428–433.
- 353 Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J *et al*. Global quantification of mammalian gene expression control. *Nature* 2011; **473**: 337–342.
- 354 Castellana N, Bafna V. Proteogenomics to discover the full coding content of

- genomes: A computational perspective. *J. Proteomics*. 2010; **73**: 2124–2135.
- 355 Krug K, Nahnsen S, Macek B. Mass spectrometry at the interface of proteomics and genomics. *Mol Biosyst* 2011; **7**: 284–291.
- 356 Haider S, Pal R. Integrated analysis of transcriptomic and proteomic data. *Curr Genomics* 2013; **14**: 91–110.
- 357 Cox B, Kislinger T, Emili A. Integrating gene and protein expression data: Pattern analysis and profile mining. *Methods* 2005; **35**: 303–314.
- 358 Bemark M, Neuberger MS. The c-MYC allele that is translocated into the IgH locus undergoes constitutive hypermutation in a Burkitt's lymphoma line. *Oncogene* 2000; **19**: 3404–3410.
- 359 Otto C, Scholtysik R, Schmitz R, Kreuz M, Becher C, Hummel M *et al.* Novel IGH and MYC Translocation Partners in Diffuse Large B-Cell Lymphomas. *Genes Chromosom Cancer* 2016; **55**: 932–943.
- 360 Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME *et al.* Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 2011; **144**: 296–309.
- 361 Matthias P, Rolink AG. Transcriptional networks in developing and mature B cells. *Nat Rev Immunol* 2005; **5**: 497–508.
- 362 de Yébenes VG, Bartolomé-Izquierdo N, Ramiro AR. Regulation of B-cell development and function by microRNAs. *Immunol. Rev.* 2013; **253**: 25–39.
- 363 Petri A, Dybkær K, Bøgsted M, Thruø CA, Hagedorn PH, Schmitz A *et al.* Long noncoding RNA expression during human B-cell development. *PLoS One* 2015; **10**. doi:10.1371/journal.pone.0138236.
- 364 Kulis M, Merkel A, Heath S, Queiros AC, Schuyler RP, Castellano G *et al.* Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat Genet* 2015; **47**: 746–756.
- 365 Altelaar AFM, Munoz J, Heck AJR. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* 2013; **14**: 35–48.
- 366 Frenkel-Morgenstern M, Lacroix V, Ezkurdia I, Levin Y, Gabashvili A, Prilusky J *et al.* Chimeras taking shape: potential functions of proteins encoded by chimeric RNA transcripts. *Genome Res* 2012; **22**: 1231–1242.

- 367 Baker M. Proteomics: The interaction map. *Nature* 2012; **484**: 271–275.
- 368 Cox J, Mann M. Is Proteomics the New Genomics? *Cell*. 2007; **130**: 395–398.
- 369 Stephenson LM, Miletic A V, Kloeppe T, Kusin S, Swat W. Vav proteins regulate the plasma cell program and secretory Ig production. *J Immunol* 2006; **177**: 8620–8625.
- 370 Hwang IY, Hwang KS, Park C, Harrison KA, Kehrl JH. Rgs13 Constrains Early B Cell Responses and Limits Germinal Center Sizes. *PLoS One* 2013; **8**. doi:10.1371/journal.pone.0060139.
- 371 Suan D, Sundling C, Brink R. Plasma cell and memory B cell differentiation from the germinal center. *Curr Opin Immunol* 2017; **45**: 97–102.
- 372 Klein U, Casola S, Cattoretti G, Shen Q, Lia M, Mo T *et al*. Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination. *Nat Immunol* 2006; **7**: 773–782.
- 373 Li Y-F, Xu S, Huang Y, Ou X, Lam K-P. Tyrosine kinase c-Abl regulates the survival of plasma cells. *Sci Rep* 2017; **7**: 40133.
- 374 Lopez-Herrera G, Tampella G, Pan-Hammarström Q, Herholz P, Trujillo-Vargas CM, Phadwal K *et al*. Deleterious mutations in LRBA are associated with a syndrome of immune deficiency and autoimmunity. *Am J Hum Genet* 2012; **90**: 986–1001.
- 375 Berkowska MA, Schickel JN, Grosserichter-Wagener C, de Ridder D, Ng YS, van Dongen JJ *et al*. Circulating Human CD27-IgA+ Memory B Cells Recognize Bacteria with Polyreactive Igs. *J Immunol* 2015; **195**: 1417–1426.
- 376 Hartweger H, Schweighoffer E, Davidson S, Peirce MJ, Wack A, Tybulewicz VLJ. Themis2 Is Not Required for B Cell Development, Activation, and Antibody Responses. *J Immunol* 2014; **193**: 700–707.
- 377 Casola S, Cattoretti G, Uyttersprot N, Koralov SB, Seagal J, Segal J *et al*. Tracking germinal center B cells expressing germ-line immunoglobulin gamma1 transcripts by conditional gene targeting. *Proc Natl Acad Sci U S A* 2006; **103**: 7396–401.
- 378 Xu Z, Fulop Z, Wu G, Pone EJ, Zhang J, Mai T *et al*. 14-3-3 adaptor proteins recruit AID to 5'-AGCT-3'-rich switch regions for class switch recombination. *Nat Struct Mol Biol* 2010; **17**: 1124–1135.
- 379 Jeelall YS, Wang JQ, Law H-D, Domaschensch H, Fung HKH, Kallies A *et al*. Human

- lymphoma mutations reveal CARD11 as the switch between self-antigen-induced B cell death or proliferation and autoantibody production. *J Exp Med* 2012; **209**: 1907–1917.
- 380 Teitel M a. The TCL1 family of oncoproteins: co-activators of transformation. *Nat Rev Cancer* 2005; **5**: 640–648.
- 381 Müschen M, Lee S, Zhou G, Feldhahn N, Barath VS, Chen J *et al.* Molecular portraits of B cell lineage commitment. *Proc Natl Acad Sci U S A* 2002; **99**: 10014–10019.
- 382 Shukla V, Lu R. IRF4 and IRF8: Governing the virtues of B lymphocytes. *Front. Biol. (Beijing)*. 2014; **9**: 269–282.