

CONFERENCIAS INVITADAS

ESTUDIO DE LA PRESENTACIÓN DE PÉPTIDOS POR HLA DE CLASE II MEDIANTE TÉCNICAS PROTEÓMICAS

Montserrat Carrascal, Laia Muixi¹, Dolores Jaraquemada¹ y Joaquín Abián

Unidad de Espectrometría de Masas Estructural y Biológica y de Servicio de Proteómica. Instituto de Investigaciones Biomédicas de Barcelona (IIBB), CSIC-IDIBAPS. Roselló 161, 7 Planta, 08036 Barcelona.

(1) Immunology Unit, Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Barcelona, Spain

La presentación de péptidos por moléculas de MHC de clase II a los linfocitos T es el proceso iniciador de la respuesta inmune adaptativa. En las enfermedades autoinmunes órgano-específicas, las moléculas de clase II se expresan en las células diana del ataque autoinmune, habitualmente células epiteliales endocrinas. Esta expresión capacita a las células epiteliales para la presentación de péptidos autólogos. El conocimiento de la naturaleza de los péptidos presentados así como de las proteínas implicadas es básico para la comprensión de mecanismos relevantes en el origen y el mantenimiento de la patología autoinmune.

En este sentido, hemos analizado los repertorios peptídicos asociados a clase II presentados en células epiteliales y se ha comparado con el repertorio correspondiente el mismo alelo en células linfoides, demostrando que los repertorios asociados a las células epiteliales son diferentes tanto en el tipo de péptidos presentados como en las proteínas que dan origen a dichos péptidos. Se ha analizado también la influencia de las chaperonas de la vía de clase II, Ii y HLA-DM, en estos repertorios y se han caracterizado, por primera vez, péptidos asociados a clase II en tejido epitelial humano (tiroides) afectado por autoinmunidad *ex vivo*, identificando un total de 16 péptidos provenientes de proteínas autólogas, uno de ellos proveniente de la tiroglobulina, una proteína descrita clásicamente como un autoantígeno en la enfermedad de Graves-Basedow.

Por otra parte se han identificado diversas proteínas asociadas a la molécula de clase II en las células epiteliales mediante su coinmunoprecipitación, separación mediante 2DE e identificación por MALDI-TOF y nESI-ITMS/MS. De esta forma se han identificado más de 15 proteínas asociadas a la molécula de clase II que podrían jugar un papel importante en la presentación antigénica en células presentadoras no profesionales.

GENÓMICA EN ARABIDOPSIS. RECURSOS Y HERRAMIENTAS

Javier Paz-Ares.

Centro Nacional de Biotecnología. CSIC. Campus de la Universidad Autónoma. 28049-Madrid

Estamos asistiendo a una revolución en el campo de la Biología, representado por la genómica, resultado del rápido desarrollo de los métodos de análisis a gran escala de la estructura y función de los genes. Así, en los últimos años se han producido avances espectaculares en el desarrollo de técnicas experimentales de gran eficacia para el análisis global de transcritos (transcriptómica), proteínas (proteómica) y metabolitos (metabolómica), así como de herramientas bioinformáticas de almacenamiento, análisis y procesamiento de datos a gran escala. Esta metodología genómica permite por primera vez el estudio de cualquier proceso biológico en términos de la contribución de al mismo de cada gen de un organismo. Esta renovación metodológica y conceptual requiere una profunda transformación en la forma de organizar la actividad investigadora en Biología, ya que no es posible que cada laboratorio o instituto pueda dotarse de la infraestructura necesaria. En España, hemos promovido una iniciativa encaminada a implantar estructuras, servicios y recursos que posibiliten a los grupos de este país la aplicación de enfoques genómicos en las investigaciones sobre *Arabidopsis*, especie modelo vegetal por excelencia, primera para la que se conoció la secuencia de su genoma y para la que se dispone de un número mayor de recursos y herramientas genómicas. Esta iniciativa incluye las siguientes actividades: 1) Análisis del transcriptoma, mediante la producción y utilización de "chips" de DNA; 2) Análisis del proteoma; 3)) Identificación y clonación de las pautas de lectura abierta del genoma (ORFoma), específicamente los ORF no clonados del cromosoma 4; 4) Obtención de mutantes de inserción, y secuenciación de los sitios de inserción; 5) Obtención de series alélicas de mutaciones puntuales; 6) Cartografía génica automatizada, para facilitar la clonación posicional y 7) Explotación de la variabilidad natural, mediante la recolección y caracterización de estirpes silvestres de *Arabidopsis* y de la crucífera *Carrichtera* en España y la realización de estudios filogenómicos para la identificación de regiones *cis*-reguladoras de la expresión génica. En esta comunicación se discute este proyecto en el contexto de la situación actual de la genómica de *Arabidopsis*.

PLEUROGENE: EL PRIMER PROYECTO DE GENOMICA DE PECES EN ESPAÑA

J. Cerdà¹, C. Buesa², P. Cañavate³, S. Douglas⁴, J. López Barea⁵, T. Maes², M. Manchado³, D. Martín-Robichaud⁶, G. Martínez⁷, J. I. Navas⁸, F. Piferrer⁹, J. Planas¹⁰, F. Prat^{11}, M. Reith⁴, M. Ruiz Rejón¹², M. Yúfera⁷.*

¹Centre d'Aqüicultura, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), CIMIMA (CSIC), 08003 Barcelona; ²Oryzon Genomics S.A., Parc Científic de Barcelona, 08028 Barcelona; ³C.I.C.E.M "El Toruño", 11500 El Puerto Santa María, Cádiz; ⁴Institute of Marine Biosciences, Halifax, Canadá; ⁵Departamento de Bioquímica y Biología Molecular, Universidad de Córdoba, 14071 Córdoba; ⁶Department of Fisheries and Oceans, St Andrews, New Brunswick, E5B 2L9 Canadá; ⁷Instituto de Ciencias Marinas de Andalucía, CSIC, 11510 Puerto Real, Cádiz; ⁸C.I.C.E.M "Agua del Pino", 21071 Huelva; ⁹Institut de Ciències del Mar, CSIC, 08003 Barcelona; ¹⁰Departament de Fisiologia, Facultat de Biologia, Universidad de Barcelona, 08028 Barcelona; ¹¹Instituto de Acuicultura de Torre de la Sal, C.S.I.C, 12595 Torre de la Sal, Castellón; ¹²Departamento de Genética, Universidad de Granada, 18071 Granada.

El proyecto "Genómica de peces planos: optimización del cultivo del fletán y el lenguado senegalés (PLEUROGENE)" es el primero de genómica de peces en España y está financiado por Genoma España en colaboración con Genoma Canadá a través de un grupo inversor compuesto por las CC.AA. de Andalucía y Cataluña, la Diputación de Cádiz por medio del Instituto de Empleo y Desarrollo Tecnológico y el Cluster de Empresas Gallegas. El proyecto en el que participan tanto grupos españoles como canadienses, está coordinado por la parte española por el IRTA y participan 9 grupos de investigación procedentes de la Consejería de Andalucía, CSIC, Universidad de Córdoba, Universidad de Granada, Universidad de Barcelona y de la empresa biotecnológica Oryzon Genomics S.A. Por parte canadiense, los grupos participantes provienen del Institute of Marine Bioscience, Department of Fisheries and Oceans, y la empresa Scotian Halibut Ltd.

Las especies escogidas para este proyecto son peces planos marinos con gran interés en acuicultura tanto en España (el lenguado) como Canadá (el fletán), pero que su producción industrial encuentra dificultades en temas relacionados con la reproducción en cautividad, el cultivo larvario, la nutrición y la respuesta inmunitaria. Para superar estos obstáculos en el cultivo de peces planos se ha optado por una aproximación genómica y proteómica con la finalidad de estudiar y caracterizar la regulación génica en estos procesos biológicos. Para ello se ha planteado los siguientes objetivos: 1) 1. Desarrollo de una base de datos integrada, constituida por aproximadamente 20.000 ESTs y de proteínas ("shot-gun proteomics") procedentes de diversos tejidos del adulto y larvas de lenguado y fletán; 2) Diseño y construcción de chips de DNA (microarrays) para el estudio de la expresión génica global (genómica funcional) durante la gametogénesis, diferenciación sexual, desarrollo y nutrición larvaria e inmunidad innata; 3) Búsqueda de marcadores moleculares y desarrollo de un mapa genético de ligamiento (genómica comparada); 4) Análisis proteómico de diversos estadios gonadales y larvarios; 5) Integración de la información genética, genómica y proteómica en una plataforma bioinformática interactiva.

*En representación del consorcio de participantes en el proyecto PLEUROGENE.

MICROARRAYS GENÓMICOS EN EL ANÁLISIS MOLECULAR DEL CÁNCER: CONEXIÓN CON FACTORES CLÍNICOS Y TERAPÉUTICOS

José Angel Martínez-Climent

Servicio de Hematología y Oncología Médica. Hospital Clínico Universitario. Universidad de Valencia

Los linfomas no Hodgkin de células B (LNH-B), una forma común de cáncer humano, comprenden un grupo heterogéneo de entidades que a nivel genético se distinguen por presentar una traslocación cromosómica característica que afecta a los genes reguladores de las inmunoglobulinas. Esta traslocación resulta en sobre-expresión de un gen (*BCL1*, *BCL2*, *BCL10*) que contribuye directamente al inicio de la enfermedad neoplásica. Sin embargo, el acumulo de otras alteraciones genéticas de forma secuencial es necesario para el desarrollo del tumor. Estas anomalías genómicas afectan dos tipos de genes: oncogenes y genes supresores. La naturaleza y la secuencia temporal de las alteraciones de estos genes contribuye definitivamente al fenotipo del tumor y al comportamiento clínico del paciente con linfoma.

El análisis global del genoma tumoral empleando la tecnología de microarrays de ADN ha supuesto un gran avance en el conocimiento de los linfomas. Estos análisis nos han mostrado el patrón de expresión génica aberrante característico de cada subgrupo de LNH-B, confirmando las clasificaciones anatómo-clínicas previas con gran precisión. Más aún, los microarrays han identificado entidades no reconocibles *a priori*, tales como los subgrupos del linfoma difuso de célula grande B (originados de células del centro germinal, o de linfocitos B activados), o el linfoma del manto sin sobre-expresión de ciclina D1. Por otro lado, los perfiles de expresión determinan subgrupos de pacientes con respuesta al tratamiento y supervivencia diferentes. Entre todos estos genes con expresión aberrante, se han identificado algunos particularmente relevantes en la patogenia de estas enfermedades. En la leucemia linfática crónica (LLC-B), el perfil de expresión ha definido a la enfermedad como una única entidad derivada de una célula común. Sin embargo, este patrón de expresión distingue dos subgrupos de LLC-B clínicamente diferentes en base a la presencia de mutaciones del gen regulador de las inmunoglobulinas. Uno de los genes identificados en estos estudios con microarrays fue *ZAP-70*, que ha resultado ser un marcador distintivo de la LLC-B con *IGH* no mutada y clínicamente agresiva. En la transformación del linfoma folicular a linfoma difuso de célula grande, el estudio combinado con microarrays de ADN ha identificado dos tipos de transformación a nivel molecular que se distinguen por presentar expresión baja o alta de *MYC* y de otros genes regulados por *MYC*. Por ello, *MYC* parece ser un gen clave en este proceso.

Otra de las nuevas tecnologías de microchips es la hibridación genómica comparada en microarrays (array CGH), basada en la CGH sobre cromosomas. Esta técnica ha facilitado la localización de oncogenes en regiones de amplificación, tales como *MALT1* en 18q21 en el linfoma MALT y *Glypican 5 (GPC5)* en 13q31-q32 en el linfoma folicular. La técnica de array CGH puede aplicarse también para delimitar del mismo modo las regiones de delección cromosómica en tumores, donde residen hipotéticamente genes supresores. De esta forma, nuestro grupo ha delimitado la región de delección común en el cromosoma 8p a unos 600 kb de secuencia en 8p21.3 en LNH-B, habiendo identificado a *TRAIL-DR4* y *DR5* como genes supresores candidatos en dicha región. Por último, presentaremos datos sobre la aplicación clínica que el análisis genómico podría tener en un futuro inmediato empleando un modelo de linfoma del manto.

Comunicaciones libres.

Sesión I: PROTEÓMICA.

Moderador *Ismael Mingarro (UVEG).*

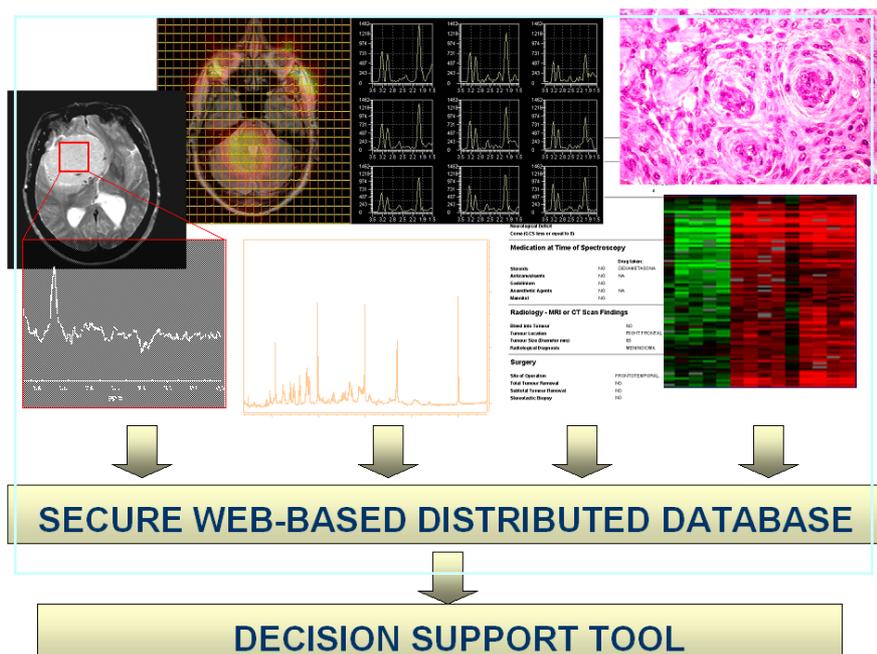
WEB ACCESSIBLE MR DECISION SUPPORT SYSTEM FOR BRAIN TUMOUR DIAGNOSIS AND PROGNOSIS, INCORPORATING IN VIVO AND EX VIVO GENOMIC AND METABOLOMIC DATA (FP6-2002-LIFESCIHEALTH 503094); Acronym: eTUMOUR.

Bernardo Celda.

Departamento de Química-Física. Universitat de València.

Diagnosis and treatment of brain tumours is based on clinical symptoms, radiological appearance, and often a histopathological diagnosis of a biopsy. However, treatment response of histological or radiologically similar tumours can vary widely, particularly for childhood tumours. New technologies are available that may improve tumour classification in terms of diagnosis and prognosis, and may allow individually optimized treatments. ^1H Magnetic Resonance Spectroscopy (MRS) is a non-invasive technique for determining tissue biochemicals (the metabolomic profile). ^1H MRS can be performed along with clinical MR Imaging but widespread use is hampered by specialised analysis requirements and poor dissemination of the skills needed to interpret the data. The genomic profile of tumours can be determined with DNA microarrays. Early studies have demonstrated differences in gene expression between tumour grades and between tumour types not easily distinguished by morphologic appearance.

We will bring together the expertise required to study the genomic and metabolomic characteristics of brain tumours, with a multi-centre collaboration to acquire statistically significant data, particularly for rare tumour types. Clinical MRS, high-resolution ^1H MRS and gene array analysis of biopsies, will be used to investigate how metabolomic and genomic profiles relate to clinically relevant factors such as survival time and treatment response. As well as providing new scientific data on tumour biology, we will develop the technology for this information to be readily and easily used to help radiologists and neurosurgeons in the management and treatment of brain tumour patients. We will build upon expertise obtained with INTERPRET EU project IST-1999-10310, which created a MRS decision support tool (DSS) for tumour diagnosis. A new web-accessible DSS will be developed, incorporating genomic and metabolomic data, and its diagnostic performance validated in a clinical demonstration of added value.



TÉCNICAS DE PROTEÓMICA PARA LA IDENTIFICACIÓN DE NUEVAS DIANAS TERAPÉUTICAS Y BIOMARCADORES DE ENFERMEDAD VASCULAR

Markus Kubicek¹, Francisco Verdeguer¹, Felipe Javier Chaves³, Josep Redón³, Juan José Calvete² y Vicente Andrés¹

Laboratorios de ¹Biología Vascular y ²Proteómica Estructural, Instituto de Biomedicina de Valencia (IBV-CSIC) y ³Hospital Clínico Universitario de Valencia

La arteriosclerosis y enfermedades asociadas ocupan el primer lugar de mortalidad y morbilidad en los países industrializados. Sin embargo, por falta de marcadores diagnósticos fiables, la enfermedad no se puede diagnosticar antes de manifestaciones graves como infartos de miocardio o cerebrovasculares. Por lo tanto sería deseable disponer de marcadores que indiquen estados tempranos de la enfermedad. La formación de lesiones ateromatosas es un complejo proceso multifactorial. La acción de diversos factores de riesgo cardiovascular, tanto modificables como genéticos (hiperlipemia, hipertensión, tabaquismo, diabetes, etc), provoca la disfunción endotelial y el subsiguiente reclutamiento de células inmunitarias, cuya activación da lugar a una respuesta inflamatoria crónica de la pared arterial. En estas condiciones se produce una elevada proliferación y migración celular en la pared vascular. Además hay una liberación abundante de citoquinas y moléculas señalizadoras que provocan una cascada de eventos pro-aterogénicos que contribuyen al crecimiento del ateroma.

La formación de la placa es un proceso muy complejo en el que intervienen distintos tipos celulares y diferentes redes de señalización. Un abordaje de genómica y proteómica permitiría analizar masivamente los factores más relevantes de este proceso patológico. En concreto, la electroforesis bidimensional permite obtener una imagen de las proteínas expresadas en un tejido concreto. El estudio de los tejidos afectados en la arteriosclerosis como la arteria, las células mononucleares circulantes o el plasma nos puede permitir encontrar marcadores con valor diagnóstico y/o posibles dianas terapéuticas.

El modelo murino deficiente en apolipoproteína E utilizado en nuestros estudios reproduce adecuadamente la patología humana. Al igual que los humanos, los ratones alimentados con una dieta elevado en colesterol y grasas generan una hipercolesterolemia que desencadena la formación de placas de ateroma.

Nuestra estrategia básica consiste en la comparación proteómica de muestras procedentes de ratones apoE KO alimentados con dieta aterogénica versus ratones alimentados con dieta control . En función de la duración de la dieta aterogénica se consiguen diferentes niveles de hipercolesterolemia así como de la progresión de las placas de ateroma. En estados tempranos (3-5 días) aunque no hay placas de ateroma los niveles de colesterol son ya elevados (1500-2000 mg/dL), por lo que es esperable encontrar cambios en el proteoma. A los 7-10 días los niveles de colesterol son muy elevados y saturantes (3500 mg/dL) y se observan unas lesiones incipientes. Finalmente en un estado avanzado (30 días) los niveles de colesterol no cambian con respecto a los de 7-10 días pero el estado y el número de placas de ateroma es muy avanzado.

Nuestra estrategia de seguimiento secuencial analizando tejidos y plasma sanguíneo nos ha permitido identificar cambios en la expresión de un conjunto de proteínas y modificaciones post-traduccionales representativas de cada estadio.

Entre las proteínas reguladas en el plasma algunas se encontraron también reguladas de igual forma en la arteria. Además, gran parte de estos cambios en el plasma de ratón se vieron también en plasma de pacientes humanos afectados de enfermedad cardiovascular. Por lo tanto pensamos que las modificaciones proteómicas encontradas en el modelo de ratón se pueden extrapolar mayoritariamente a la patología humana ofreciendo marcadores de diagnóstico y/o dianas terapéuticas. Además nuestros resultados profundizarán en la comprensión de la patogénesis de las enfermedades cardiovasculares.

PERFILES PROTEICOS CON EXPRESION DIFERENCIAL EN PACIENTES DE LLC-B CON LOS GENES IGVH Y BCL6 MUTADOS O NO MUTADOS

Marín M¹⁻², Sanz L², Sarsotti E¹, Benet I¹, Terol MJ¹, García-Conde J¹, Calvete JJ².

1.Hospital Clínico Universitario de Valencia, Universidad de Valencia.

2.Instituto de Biomedicina de Valencia, CSIC, Valencia.

La leucemia linfática crónica B (LLC-B) es el síndrome linfoproliferativo crónico más frecuente en los países occidentales. La mayoría de los pacientes se diagnostican en un estadio inicial. Se han identificado diversos factores pronóstico, como mutaciones en la región variable del gen de las inmunoglobulinas (IgVH). De forma complementaria se han descrito mutaciones somáticas del gen BCL6 en un 20-25% de la LLC-B, aunque su significado clínico es incierto. Con el objetivo de identificar marcadores moleculares de los diferentes grupos de síndrome linfoproliferativo, hemos iniciado la caracterización del perfil proteómico de diversos grupos de pacientes con LLC-B.

Se analizaron 15 muestras de linfocitos B aislados de sangre periférica de pacientes con LLC-B diagnosticados en estadio A, 5 con mutaciones en IgVH(M) y BCL6(M), 5 sin mutaciones, IgVH, BCL6 (no M), y 5 con mutaciones en IgVH pero no en BCL6, IgVH (M), BCL6 (no M). Se utilizaron 100 µg de proteína total para el isoelectroenfoque, con tiras de 13 cm con rangos de pI 4-7 y 6-11. Para la segunda dimensión, se corrieron geles SDS-PAGE al 15% y posterior tinción con azul Coomassie. La captura y análisis de imágenes se realizó utilizando el software Image Master Platinum (Amersham Biosciences). Las bandas de interés se sometieron a digestión automática con tripsina bovina. Las proteínas se identificaron por su huella peptídica obtenida mediante espectrometría de masas MALDI-TOF utilizando el programa ProteinProspector. Para confirmar la identificación inicial, se procedió a secuenciar alguno de estos péptidos mediante fragmentación inducida por colisión (CID) utilizando un espectrómetro QTrap.

El mayor número de bandas proteicas se detectó en el rango de pI 4-7. El número medio de manchas por gel fue de 214, y el porcentaje medio de solapamiento entre grupos fue del 77%. Se establecieron tres grupos de comparación entre los diferentes pacientes atendiendo a los estados mutacionales de IgVH y BCL6, obteniéndose entre 2 y 6 manchas significativas de grupo. Las manchas cuya variación entre grupos mostró significación estadística se sometieron a MALDI-TOF y CID, obteniéndose las identificaciones mostradas en la Tabla 1.

Tabla 1. Identificación por huella peptídica, mediante **MALDI-TOF** y/o **ESI-QTrap MS** de las manchas con expresión diferencial en cada grupo de pacientes, y su nivel de expresión en cada situación mutacional (Valor central del porcentaje de volumen en cada uno de los cinco geles).

Protein	IgVH, BCL6(M)	IgVH(M), BCL6(no M)	IgVH, BCL6(no M)
14-3-3 Protein Z	0,82	0,30	ns
Proteína heterocromátina específica de no-histonas(*)	0,29	0,02	ns
NADH-Ubiquinona reductasa(*)	0.12	0.02	ns
Tubulin α	0.54	1.28	ns
Vimentin	0.27	0.98	ns
Similar a Beta Tubulina	0.40	1.85	ns
Proteína 1 asociada a citoesqueleto(**)	ns	0.03	0.26
Proteína 28 del lumen del reticulo endoplasmático (**)	ns	0.55	0.21
Chaperonin TCP-1	0.005	0.05	0.18
Nucleofosmina	0.17	1.51	1.22

El perfil proteómico de linfocitos B de tres grupos de pacientes, revela al menos 10 proteínas con expresión diferencial, que se pueden dividir en las siguientes categorías:
 La expresión de seis proteínas(*) cambia con el estado mutacional de BCL6.
 La expresión de dos proteínas(**) se correlaciona con el estado mutacional de IgVH.
 La Nucleofosmina exhibe un incremento de expresión en los pacientes con el gen BCL6 no mutado.
 La expresión de la Chaperonina TCP-1 es mas alta en los pacientes con IgVH no mutado.

PRESENTACIÓN DE LA SOCIEDAD ESPAÑOLA DE PROTEÓMICA (SEProt): OBJETIVOS Y ACTIVIDADES

Juan J. Calvete

Instituto de Biomedicina de Valencia, C.S.I.C.
Presidente en funciones de la SEProt

La **Sociedad Española de Proteómica** (SEProt; página web provisional: <http://www2.cbm.uam.es/seprot>) ya es una realidad! La idea de crear esta nueva asociación científica sin ánimo de lucro surgió durante las Jornadas de Proteómica celebradas en la Universidad de Córdoba en Febrero del 2003 y culminó en Valencia el 13 de Abril del presente año con la firma del Acta Fundacional por delegados de Unidades de Investigación y Servicios de Proteómica que representan a un amplio espectro del panorama proteómico español.

La **SEProt** nace con el objetivo de impulsar el desarrollo de la Proteómica en nuestro país, agrupando para ello a investigadores del campo de la Proteómica, usuarios de Servicios de Proteómica, Empresas del sector, y estudiantes interesados en el estudio de los proteomas. Para el cumplimiento de estos fines, la **SEProt** va a organizar reuniones periódicas que catalicen la discusión y el intercambio de nuevas ideas entre investigadores tanto del ámbito académico como industrial. El primer Congreso de la **SEProt** se celebrará en Córdoba durante los días 14-17 de Febrero del 2005. Toda la información relacionada con este evento estará accesible en la página web <http://www.uco.es/servicios/scai/seprot2005>.

Sesión II: BIOINFORMÁTICA Y GENÓMICA COMPARADA.

Moderador *Fernando González (UVEG).*

EPIDEMIOLOGÍA GENÓMICA Y BIOINFORMÁTICA DE SOPORTE

O. COLTELL^{1*}, M. ARREGUI¹, O. PORTOLÉS², C. PÉREZ¹, P. CARRASCO², D. CORELLA², R. CHALMETA¹

¹Grupo de Integración y Re-Ingeniería de Sistemas (IRIS). Departamento de Lenguajes y Sistemas Informáticos. Universitat Jaume I. Avda. Vicent Sos Baynat, s/n. 12071-Castellón. España.

²Unidad de Investigación en Epidemiología Genética y Molecular (EPIGEM). Departamento de Medicina Preventiva y Salud Pública. Facultad de Medicina y Odontología. Universitat de València. Avda. Blasco Ibáñez, 15. 46009-Valencia. España.

1. Resumen

La Epidemiología Genómica es una disciplina que estudia la contribución relativa y posible interacción de los factores genéticos y ambientales en la etiología, distribución y prevención de la enfermedad en los humanos. Uno de sus axiomas principales establece que las enfermedades humanas están causadas por la interacción entre factores genéticos (variaciones en el genoma) y ambientales (estilos de vida), y por tanto, dicha interacción debe tenerse en cuenta explícitamente en el diseño y análisis de los estudios epidemiológicos (Coltell, 2004).

El grupo EPIGEM y el grupo IRIS, en el marco de la red de Informática Biomédica INBIOMED (INBIOMED, 2004), están manteniendo una estrecha colaboración, en el campo de la Epidemiología Genómica de las Enfermedades Cardiovasculares (EGECV), para obtener evidencias acerca de los factores de riesgo. Los diseños de estudios en esta disciplina incluyen una parte muy importante que se basa en el trabajo en laboratorio de Biología Molecular: tratamiento de muestras de sangre de los participantes, extracción de ADN de las mismas, amplificación y digestión del ADN para obtener la expresión génica, análisis de polimorfismos, etc. Por otra parte, la situación actual en la Genómica Cardiovascular es que se han descrito miles de genes posiblemente implicados en el riesgo de ECV actuando a distintos niveles: metabolismo lipídico, tensión arterial, estrés oxidativo, intolerancia a la glucosa, obesidad, inflamación, hiperhomocisteinemia, etc. Por lo tanto, la información generada en el laboratorio de EPIGEM sobre el análisis de los genes incluidos en los estudios epidemiológicos no puede tratarse ya manualmente como hace unos pocos años. Ha sido necesario incorporar recursos bioinformáticos y de Informática Biomédica para poder llevar a cabo los análisis de datos genómicos. Esta ha sido la misión del grupo IRIS.

En el caso concreto de los recursos bioinformáticos, el grupo IRIS se ha responsabilizado de proporcionar el soporte bioinformático necesario para la investigación genómica cardiovascular mediante el desarrollo de varios proyectos de distinta envergadura. Algunos de ellos están finalizados y en explotación, y otros siguen todavía en fase de construcción. Entre los proyectos más destacables se tienen los siguientes (Coltell, 2004):

- **SeqPacker**: editor visual de secuencias (ADN y ARN) que permite trabajar con cadenas directas e inversas, y cambiar la presentación en varios formatos y puede leer y guardar datos en formato FASTA, Genbank y ABI, incluso de cromosomas enteros.
- **Squeezer**: compresor de secuencias (ADN y ARN) que admite bases indefinidas provenientes de trazas de secuenciación en formato ABI.
- **Estudio de ficheros de trazas de secuenciación en formato ABI**: dados los resultados de secuenciación de ADN de pacientes obesos, se comprobó la fiabilidad de los datos del cromatograma (valores umbrales y base asignada), se alinearon las

* Autor de contacto: Oscar Coltell. Departamento de Lenguajes y Sistemas Informáticos. Universitat Jaume I. Avda. Vicent Sos Baynat, s/n. 12071 - Castellón. España. e-mail: coltell@lsi.uji.es. Tel. +34-964-728314, fax. +34-964-728435.

secuencias y se detectaron algunas mutaciones en parte de las muestras. Queda pendiente la obtención de evidencia biológica sobre dichas mutaciones en este tipo de pacientes.

- **Análisis de haplotipos de individuos para un determinado genotipo:** Obtención de la combinación de haplotipos de un individuo participante utilizando una interfaz computacional que permita, además, el intercambio de datos con herramientas de análisis estadístico como SPSS. Este proyecto está en fase de desarrollo.

En resumen, la colaboración entre ambos grupos con respecto al soporte bioinformático ha producido dos tipos de resultados: por una parte, se han detectado varios y variados problemas de manejo de información genómica que no se podían resolver (o con un coste razonable) con los recursos bioinformáticos estándar (académicos y comerciales), esto ha enriquecido la casuística y experiencia del grupo IRIS; y por la otra parte, se han desarrollado, o están en curso, varios proyectos bioinformáticos específicos para los problemas planteados, esto ha proporcionado más capacidad de investigación al grupo EPIGEM. La colaboración sigue en marcha, en principio dentro del ámbito de INBIOMED (2003-2005), pero se plantea la continuidad ya que EPIGEM está ampliando su parque instrumental y su capacidad de investigación en la parte genómica de los estudios epidemiológicos.

2. Referencias

Coltell O. Integración de la Bioinformática en la Investigación Genómica Cardiovascular: Aplicaciones en el Framingham Heart Study. Tesis doctoral. Departamento de Lenguajes y Sistemas Informáticos. Universitat Jaume I, 2004.

INBIOMED. Red Temática de Investigación Cooperativa de Informática Biomédica. "INBIOMED. Plataforma de almacenamiento, integración y análisis de datos clínicos, genéticos, epidemiológicos e imágenes orientada a la investigación sobre patologías". <http://www.inbiomed.retics.net>. Accedido en 2 de septiembre de 2004.

METODOLOGÍA PARA LA EVALUACIÓN DE LA CALIDAD DE LOS ALGORITMOS DE CLASIFICACIÓN

José Antonio Heredia , Antonio Estruch y Pau Lluís Gozalbo.*
Dpto. Tecnología. Universitat Jaume I de Castellón.

Durante los últimos años los experimentos con microarrays de DNA han permitido la generación de grandes repositorios de conjuntos de datos caracterizando sistemas biológicos complejos. Los prometedores resultados iniciales del análisis matemático con la finalidad de predecir el comportamiento de las células y la explosión en el número de algoritmos desarrollados para la identificación y/o clasificación en una amplia variedad de áreas a partir de los patrones de expresión con microarrays, hace surgir la necesidad de validar de forma rigurosa los resultados obtenidos y establecer criterios para comparar la calidad de los algoritmos propuestos.

Con esta finalidad proponemos una metodología aplicable a cualquier algoritmo de clasificación que incluye métodos para validar los resultados de la predicción, estimaciones del tamaño de muestra necesario en los experimentos, y criterios para comparar de forma sistemática la adecuación de los distintos algoritmos al fin perseguido.

Las técnicas estadísticas que se proponen no provienen de la estadística teórica sino del enfoque moderno de la estadística computacional: permutaciones, validación cruzada, simulación de Monte Carlo y ajuste de curvas de aprendizaje.

En estos momentos nos encontramos desarrollando las aplicaciones informáticas que implantan esta metodología.

INBIOMED: PLATAFORMA DE ALMACENAMIENTO, INTEGRACIÓN Y ANÁLISIS DE DATOS CLÍNICOS, GENÉTICOS, EPIDEMIOLÓGICOS E IMÁGENES ORIENTADA A LA INVESTIGACIÓN SOBRE PATOLOGÍAS

O. Coltell^{1*}, F. Martín-Sánchez², G. Bueno³, D. Corella⁴, J. Díaz⁵, J. A. Heredia⁶, V. LOPEZ², M^a I. Loza⁷, N. Malats⁸, V. Maojo⁹, X. Pastor¹⁰, A. Pazos¹¹, M. Robles¹², F. Sanz¹³

¹Grupo de Integración y Re-Ingeniería de Sistemas (IRIS). Departamento de Lenguajes y Sistemas Informáticos. Universitat Jaume I. Castelló.

²Area de Bioinformática y Salud Pública. Instituto de Salud Carlos III. 28220-Majadahonda.

³Departamento de Ingeniería Eléctrica, Electrónica y Automática. Universidad Castilla-La Mancha. Ciudad Real.

⁴Unidad de investigación en Epidemiología Genómica y Molecular. Departamento de Medicina preventiva Facultad de Medicina. Universidad de Valencia.

⁵Servicio de Informática. Hospital Universitario Virgen de las Nieves. Granada.

⁶Departamento de Tecnología Escuela de Ingeniería Informática. Universidad Jaume I. Castellón.

⁷Unidad de Farmacogenómica Aplicada a I+D de Fármacos. Departamento de Farmacología. Facultad de Farmacia. Universidad de Santiago de Compostela.

⁸Unitat de Recerca Respiratòria i Ambiental. Instituto Municipal de Investigación Médica. Barcelona.

⁹Departamento de Inteligencia Artificial. Facultad de Informática. Universidad Politécnica de Madrid.

¹⁰Unitat d'Informàtica Mèdica. Corporació Sanitària Clinic. Barcelona.

¹¹Laboratorio de Redes de Neuronas Artificiales y Sistemas Adaptativos/ Informática Médica y Diagnóstico Radiológico. Universidade da Coruña.

¹²Grupo BET-Informática Médica. Facultad de Informática. Universidad Politécnica de Valencia.

¹³Unitat d'Investigació en Informàtica Biomèdica (GRIB). Institut Municipal d'Investigació Mèdica(IMIM). UPF. Barcelona.

3. Resumen

INBIOMED es una red temática de investigación cooperativa (INBIOMED, 2004) surgida de la convocatoria de redes temáticas del Ministerio de Sanidad, dentro del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I+D+I) para el periodo 2003-2006, en el área de Biomedicina (BOE, 2002). El campo de actuación de la Red es la Informática Biomédica y empezó a desarrollar sus actividades en enero de 2003 financiada por el Instituto de Salud "Carlos III" (ISCIII, 2003), el organismo del Ministerio de Sanidad encargado de la gestión administrativa, supervisión y control de las redes temáticas en Biomedicina.

El objetivo principal de la Red está definido en términos del área de la Informática Biomédica, es decir, definir y desarrollar una infraestructura informática distribuida para almacenar y relacionar todos los tipos de datos relevantes para la investigación clínica, epidemiológica y genómica de enfermedades complejas (cáncer, cardiovasculares, psiquiátricas, neurodegenerativas, etc.). Sin embargo, parte de los subobjetivos se han supeditado al ámbito concreto de la Bioinformática, porque se ha establecido que se debe proceder a la definición, el desarrollo y la integración de herramientas de análisis que permitan procesar, modelar, visualizar y predecir conocimiento biomédico, entre otros, en los niveles de información molecular y genético.

INBIOMED está compuesta por trece grupos de investigación españoles encuadrados en las áreas de Bioinformática, Informática Médica-clínica, Procesado de imágenes, Epidemiología y Farmacogenómica. Y por ello cuenta con la capacidad, preparación y recursos necesarios para abordar, dentro de los objetivos generales, los subobjetivos relacionados con la Bioinformática.

Otro de los objetivos principales de INBIOMED es la aplicación y el pilotaje de la estructura de datos y de las herramientas de análisis en los ámbitos de las patologías de cáncer, cardiovascular y neurológica. Esto, en primer lugar, se está aplicando solamente en el ámbito de la Red. Sin embargo, se ha previsto que los resultados obtenidos proporcionen soporte de investigación a las actividades de otras redes de patologías. De

* Autor de contacto: Oscar Coltell. Departamento de Lenguajes y Sistemas Informáticos. Universitat Jaume I. Avda. Vicent Sos Baynat, s/n. 12071 - Castellón. España. e-mail: coltell@lsi.uji.es. Tel. +34-964-728314, fax. +34-964-728435.

hecho, ya existe una colaboración con la red de Intervención en Dieta denominada PREDIMED (PREDIMED, 2004), surgida de la misma convocatoria (ISCIII, 2003).

Entre los subproyectos bioinformáticos realizados o en curso se pueden destacar los siguientes agrupados por campos científicos:

- **Epidemiología Genómica de las Enfermedades Cardiovasculares:** Herramientas bioinformáticas para la medida de calidad y fiabilidad de las secuencias de trazas de secuenciación, edición, búsqueda y compresión de secuencias (ADN y ARN), trabajos de alineamiento múltiple de secuencias (ADN y ARN), detección de nuevas mutaciones en ADN de los participantes, análisis de haplotipos basados en el genotipo individual, etc.
- **Epidemiología Genómica del Cáncer:** Herramientas bioinformáticas para modelar, visualizar y predecir conocimiento biomédico en el estudio epidemiológico ambiental, clínico, genético y molecular del cáncer de vejiga urinaria.
- **Epidemiología Genómica de las Enfermedades Psiquiátricas y Neurodegenerativas:** Herramientas bioinformáticas para definición de marcadores clínicos farmacogenómicos y farmacoproteómicos que aporten información útil al psiquiatra clínico para un mejor diagnóstico y tratamiento de la esquizofrenia.
- **Farmacogenómica:** Herramientas bioinformáticas para estudiar los aspectos genéticos diferenciales entre las personas a la hora de estudiar el desarrollo de nuevos fármacos y de analizar como afecta la diversidad genética en la respuesta frente a la administración de un fármaco.
- **Análisis genético en general:** Herramientas bioinformáticas para la predicción de sitios de unión de factores de transcripción en secuencias de ADN, la superposición molecular en el alineamiento de conjuntos de moléculas en virtud del potencial de interacción molecular que generen, la comparación entre secuencias genómicas, la predicción de genes sobre secuencias de ADN, y la gestión de experimentos de microarrays.
- **Integración:** Herramientas para la gestión de conocimiento científico, y para la gestión de la colaboración remota y sincrónica entre científicos que trabajan en proyectos de diseño de fármacos.

En resumen, INBIOMED, como red temática en Informática Biomédica, ha desarrollado (y está desarrollando otras adicionales) arquitecturas y herramientas de integración de datos biomédicos. Algunas de estas arquitecturas y herramientas son específicamente bioinformáticas y no sólo están disponibles para los investigadores de la Red, sino que se ofrecen a otros grupos y otras redes con el objetivo de establecer las colaboraciones científicas externas que sean posibles.

4. Referencias

INBIOMED. Red Temática de Investigación Cooperativa de Informática Biomédica. "INBIOMED. Plataforma de almacenamiento, integración y análisis de datos clínicos, genéticos, epidemiológicos e imágenes orientada a la investigación sobre patologías". <http://www.inbiomed.retics.net>. Accedido en 2 de septiembre de 2004.

Instituto de Salud Carlos III. Resolución de 30 de diciembre de 2002 de la Dirección del Instituto de Salud «Carlos III» por la que se aprueba la concesión y denegación de ayudas para el desarrollo de redes temáticas de investigación cooperativa. 3/01/2003.

Orden SCO/709/2002 de 22 de Marzo, el Ministerio de Sanidad y Consumo convocó la concesión de ayudas para el desarrollo de Redes Temáticas de Investigación Cooperativa. BOE núm. 80, de 2 de abril de 2002.

PREDIMED. Red Temática de Investigación Cooperativa de Intervención Dietética. "PREDIMED. Efectos de la dieta mediterránea en la prevención primaria de las enfermedades cardiovasculares". Web de trabajo de PREDIMED: <http://www.predimed.uji.es/drupal>. Accedido en 2 de septiembre de 2004.

HERRAMIENTAS BIOINFORMÁTICAS EN GENÓMICA FUNCIONAL DE CÍTRICOS

Ana Conesa¹, Juan Miguel García², Stefan Goetz², Montserrat Robles², Maria José Nueda³, Alberto Ferrer³, Manuel Talón¹.

¹ Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias

² ITACA-BET, Universidad Politécnica de Valencia

³ Departamento de Estadística e Investigación Operativa, UPV.

El abaratamiento en la implementación de las tecnologías genómicas ha hecho posible el planteamiento en muchos laboratorios de aproximaciones genómico-funcionales para la caracterización de la biología de los organismos y sistemas que en ellos se estudian. Un ejemplo de este tipo de aproximaciones es el desarrollo del "Citrus Chip" por el Consorcio Nacional de Genómica de Cítricos. Esta plataforma, que en la actualidad cuenta con más de 13000 secuencias expresadas que representan una amplia variedad de procesos biológicos, se utiliza en el IVIA para el estudio de la biología relacionada con los procesos de relevancia agronómica en cítricos. Muchas son las herramientas, comerciales o de libre acceso, que se encuentran actualmente disponibles para el análisis de la ingente cantidad de datos que las tecnologías "ómicas" típicamente produce. Sin embargo, estas herramientas han sido concebidas y desarrolladas, a menudo, en un contexto biológico concreto y o bien no son directamente aplicables o bien no se adecuan a la realidad específica de otros organismos. En el Centro de Genómica del IVIA se ha iniciado una línea de investigación bioinformática que pretende desarrollar herramientas de análisis centradas en el tipo de realidad biológica que nos ocupa, el estudio de los cítricos. En esta comunicación se mostrarán los avances en el desarrollo de dos de tales herramientas: FELIZ, un método de análisis estadístico para identificar genes de expresión diferencial en experimentos de series temporales con múltiples factores, y Blast2GO, una herramienta de anotación y visualización de categorías génicas funcionales aplicable en la fase de interpretación biológica en los experimentos de microarrays.

ANÁLISIS DEL FILOMA DE *BLOCHMANNIA FLORIDANUS*: DE LA FILOGENÉTICA A LA FILOGENÓMICA.

Iñaki Comas, Andrés Moya, Fernando González-Candelas.

Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València.

Las fuerzas evolutivas que actúan sobre los genomas bacterianos hacen de la reconstrucción de sus relaciones filogenéticas un problema. La transferencia génica horizontal, las paralogías, la pérdida y ganancia génicas o los diferentes sesgos en el uso de codones pueden influir en la correcta inferencia del árbol de especies. En este trabajo se han utilizado diferentes técnicas, desde las más puramente filogenéticas a las más recientes filogenómicas, para la correcta inferencia de la topología que relaciona 21 proteobacterias. Entre estas proteobacterias se encuentran tres β -proteobacterias (*Ralstonia solanacearum* y dos *Neisseria*), una α -proteobacteria (*Rickettsia prowazekii*) así como diferentes γ -proteobacterias entre las que destacan cinco endosimbiontes (las tres *Buchnera* secuenciadas hasta el momento, *Wigglesworthia glossinidia* y *Blochmannia floridanus*). Partiendo del genoma de esta última, realizaremos una aproximación filogenética consistente en la obtención de su filoma o conjunto de árboles génicos derivados del proteoma de la bacteria, mientras que la aproximación filogenómica consistirá en el uso de técnicas de consenso y en el análisis de concatenados. El objetivo de este proyecto no es sólo inferir una filogenia correcta de estas especies, sino también analizar las ventajas e inconvenientes inherentes a cada una de las metodologías usadas para dicha inferencia.

BUCHNERA APHIDICOLA, ENDOSIMBIONTE DEL PULGÓN CINARA CEDRI, TIENE UN GENOMA DE 450KB

Amparo Latorre, Vicente Pérez-Brocal, Laura Gómez-Valero, Rosario Gil y Andrés Moya.
Institut Cavanilles de Biodiversitat i Biologia Evolutiva y Departament de Genètica
Universitat de València.

Los pulgones mantienen una relación de simbiosis obligatoria con *Buchnera aphidicola*, una γ -proteobacteria próxima a enterobacterias de vida libre, como *Escherichia coli*. Los pulgones se alimentan de la savia elaborada de las plantas, que es deficiente entre otros productos, en aminoácidos esenciales, que son suministrados por la bacteria. La relación pulgón-*B. aphidicola* es un caso de mutualismo pues el pulgón depende de la bacteria para su crecimiento y fecundidad y, por otra parte, el pulgón le suministra a ésta un ambiente estable. Son diversos los cambios sufridos en el genoma de *B. aphidicola* en su adaptación a la vida endosimbiótica, como son drástica reducción en el tamaño del genoma por pérdida de genes, elevado contenido en A+T, aceleración de las tasas evolutivas, pérdida en el sesgo de uso de codones y la amplificación de genes esenciales en la síntesis de leucina y triptófano mediante su traslocación a plásmidos.

Hasta el momento se han secuenciado tres genomas de *B. aphidicola* pertenecientes a tres especies distintas de pulgones, dos de ellas muy emparentadas, con tamaños de 640 y 641 Kb y la tercera de 616 Kb perteneciente a un linaje muy alejado de las dos primeras. La comparación de los tres genomas muestra que el orden génico es prácticamente idéntico en los tres linajes, por lo que se ha postulado que existe un "orden génico fósil" de todas las Buchneras. Sin embargo, el número de genes es variable (609, 597 y 545), lo que sugiere que en los diferentes linajes se siguen produciendo pérdidas selectivas de genes, posiblemente relacionados con los distintos estilos de vida de cada pulgón. Puesto que el tamaño cromosómico es similar en las tres cepas de *B. aphidicola* secuenciadas se postuló que la bacteria estaría cercana al número mínimo de genes necesarios para la vida intracelular en el pulgón. La determinación mediante PFGE del tamaño del cromosoma de *B. aphidicola* de pulgones de distintas subfamilias demostró que existen genomas hasta 200 Kb más pequeños que los ya conocidos. En concreto el tamaño del genoma de *B. aphidicola* del pulgón *C. cedri* (Lachninae) es de 450 Kb, siendo el genoma bacteriano más pequeño descrito hasta el momento. Actualmente estamos terminando la secuenciación del mismo y hemos analizado el proceso de reducción en dos regiones seleccionadas. Es importante comentar que en el pulgón *C. cedri* coexisten, además de *B. aphidicola*, otra γ -proteobacteria llamada endosimbiote secundario y *Wolbachia*, una α -proteobacteria que infecta diversos grupos de artrópodos y que se ha encontrado por primera vez en pulgones. La masiva presencia de simbioses secundarias junto con su localización y morfología nos hace postular un posible reemplazamiento de, al menos, algunas funciones perdidas en *B. aphidicola*.

EN BUSCA DEL GENOMA MÍNIMO

Rosario Gil^{1,2}, Francisco J. Silva^{1,2}, Juli Peretó^{1,3}, Amparo Latorre^{1,2} y Andrés Moya^{1,2}

¹Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València

²Departament de Genètica, Universitat de València

³Departament de Bioquímica i Biologia Molecular, Universitat de València

Gracias a los avances de la genómica se conocen muchos genomas completos y crece la curiosidad sobre cuántos genes son esenciales para la vida. Tratando de acercarnos a este número mágico, nuestro grupo ha realizado un análisis comparativo de ocho genomas bacterianos completamente secuenciados. Seis de los genomas utilizados en este estudio son de pequeño tamaño, debido a un largo proceso de reducción genómica asociado a su forma de vida endosimbionte o parásita, mientras que los otros dos corresponden a bacterias de vida libre. La unificación de los datos sobre genes compartidos en todos los genomas reducidos, datos experimentales sobre genes esenciales en las distintas especies bacterianas analizadas, más la reconstrucción de la maquinaria metabólica esencial para sostener la vida, nos ha permitido proponer un hipotético genoma mínimo compuesto por sólo 206 genes. Este genoma mínimo contiene toda la información genética necesaria para el automantenimiento y reproducción de una hipotética célula bacteriana en un ambiente protegido y en presencia de todos los nutrientes esenciales.

ANÁLISIS GENÓMICO COMPARADO DE CEPAS HÍBRIDAS NATURALES DEL COMPLEJO *SACCHAROMYCES* 'SENSU STRICTO'

S. S. González^{1, 2}, A. Quero², J. García Martínez³, J.E. Pérez Ortín³ y E. Barrio¹

¹Institut 'Cavanilles' de Biodiversitat i Biologia Evolutiva, Universitat de València.

²Instituto de Agroquímica y Tecnología de los Alimentos, CSIC, Valencia.

³Servei de Chips de DNA i Dept. de Bioquímica i Biologia Molecular, Universitat de València.

Saccharomyces cerevisiae es la principal levadura responsable de la fermentación alcohólica. Otras especies del género comparten su mismo nicho ecológico y pueden formar híbridos interespecíficos viables, aunque estériles, por conjugación entre células haploides o entre ascosporas. La esterilidad de los híbridos es principalmente debida a incompatibilidad genómica por divergencia nucleotídica o por la presencia de reordenaciones cromosómicas [1] y, aunque sean estériles, estos híbridos se pueden mantener mediante reproducción asexual (gemación), o dar lugar a individuos alotetraploides o homoploides que recuperan su fertilidad [2].

En los últimos años, se ha demostrado que la presencia de híbridos naturales es relativamente frecuente en procesos de fermentación. Por ejemplo, *S. pastorianus* (antes *S. carlsbergensis*), responsable de la producción de cerveza 'lager', es un alotetraploide parcial proveniente de la hibridación entre *S. cerevisiae* y *S. bayanus*. Otras levaduras híbridas se han descrito también en vinos y sidra [3, 4]. Nosotros mismos hemos demostrado la presencia en vinos centroeuropeos de híbridos entre *S. cerevisiae* y *S. kudriavzevii*, una especie descrita originalmente a partir de dos cepas aisladas en ambientes naturales del Japón.

La presencia de estos híbridos naturales en procesos de fermentación resulta de interés ya que, aunque son menos eficaces que las levaduras parentales en condiciones ambientales específicas, están mejor adaptados a las condiciones fluctuantes o intermedias [2]. En el presente estudio, presentamos los resultados de la caracterización genómica comparada de híbridos naturales con la finalidad de determinar los mecanismos implicados en su origen y evolución.

[1] Grieg, D., Borts, R.H., Louis, E.J., Travisano, M. Proc. R. Soc. Lond. B 269: 1167-1171. 2002.

[2] Grieg, D., Louis, E.J., Borts, R.H., Travisano, M. Science 298: 1773-1775. 2002.

[3] Masneuf, I., Hansen, J., Groth, C., Piškur, J., Dubourdieu, D. Appl. Environ. Microbiol. 64: 3887-3892. 1998.

[4] Groth, C., Hansen, J., Piškur, J. Int. J. Syst. Bacteriol. 49: 1933-1938. 1999.

Sesión III: GENÓMICA DE PLANTAS.

Moderador *Vicente Pallás (IBMCP).*

DESARROLLO DE RECURSOS PARA ESTUDIOS GENÓMICOS EN CITRICOS: CREACION DE UNA COLECCIÓN DE ESTs Y UNA MICROMATRIZ DE cDNA

Vicente Conejero

(en representación del Proyecto de Genómica Funcional de Cítricos)

El Proyecto de Genómica Funcional de Cítricos se inició en el año 2001 con el objeto de desarrollar herramientas que permitan aproximar desde un punto de vista genómico los problemas más relevantes de la citricultura en España. En esta charla, se resumirán los resultados obtenidos hasta la fecha en este proyecto.

Hemos construido 25 genotecas de cDNA que cubren un amplio rango de situaciones biológicas y ambientales en cítricos, y hemos generado a partir de ellas una colección de 22635 ESTs de alta calidad, que representan un total de 11836 genes únicos, un tercio del número de genes estimado en el genoma de los cítricos. La colección incluye tanto genotecas estándar como de sustracción, 6 de ellas de longitud completa. Se ha prestado especial atención a los diferentes aspectos implicados en el proceso de producción y calidad del fruto, como crecimiento vegetativo, floración, *fruit set* y maduración, respuesta a estreses bióticos y abióticos y procesos post-cosecha.

La anotación funcional de los genes obtenidos revela que hemos obtenido representación en todas las categorías funcionales, garantizando el enfoque general de nuestra aproximación. Además, revela la existencia de elevado número de genes sin homología conocida, algunos de los cuales podrían ser genes específicos de cítricos de algún valor biotecnológico.

Como primera aproximación al estudio del transcriptoma, construimos una micromatriz de cDNA de primera generación que contiene 6782 unigenes. Esta micromatriz está siendo utilizada por los diferentes grupos de investigación del proyecto para caracterizar el transcriptoma de los cítricos en diferentes situaciones de interés. Se presentará de modo general la caracterización técnica del microarray, así como los resultados obtenidos en experimentos de expresión en tejidos vegetativos, en frutos sometidos a estrés por frío, y en plantas sometidas a estrés hídrico. Igualmente, se discutirá el posible uso de estas micromatrices para diferenciar variedades de cítricos.

DESARROLLO DE HERRAMIENTAS PARA EL ESTUDIO DE LA GENÓMICA DE CÍTRICOS.

Talón, M., Cercós, M., Colmenero-Flores JM., Conesa, A., Iglesias, D.J., Tadeo, F.R., Terol, J.

Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias, E-46113, Moncada, Valencia, España.

El Centro de Genómica del Instituto Valenciano de Investigaciones Agrarias está desarrollando diversas herramientas para el estudio de la genómica de los cítricos, con objeto de identificar y caracterizar genes de interés agronómico. Las líneas principales de trabajo se relacionan por un lado con la calidad, maduración y abscisión de los frutos, y por otro con la tolerancia a las principales fisiopatías del cultivo, que son la salinidad, la sequía y la deficiencia en hierro o clorosis férrica. El estudio del transcriptoma de cada una de estas condiciones se está efectuando mediante macro- y microarrays desarrollados, por el Consorcio Nacional de Genómica de Cítricos, y posterior análisis por PCR en “tiempo real”. El análisis de expresión génica de los distintos tipos celulares de un tejido especializado se está llevando a cabo mediante microdissección por láser. Por otro lado, se han construido 17 genotecas estándar de cDNA y 1 genoteca normalizada de clones de cDNA de longitud completa. El RNA utilizado para la generación de las colecciones ESTs de longitud completa se extrajo a partir de hojas, frutos y raíces de Clementina a distintos estadios de desarrollo. El material vegetal también se sometió a distintos estreses abióticos como salinidad, sequía y deficiencias nutricionales. Por el contrario, las librerías estándar de cDNA se construyeron con RNA extraído de tejidos específicos. Hasta la fecha se han secuenciado unos 37.000 ESTs de la librería de clones de longitud completa y unos 20.000 del resto de librerías. En relación con el material vegetal se han generado diversas colecciones de mutantes de cítricos de los cultivares de naranja Washington y mandarina Clementina. Los resultados de la caracterización fisiológica y agronómica indican que algunos mutantes han mantenido diferencias constantes respecto de la media durante los años de estudio, en relación a parámetros de producción, tamaño y calidad de la fruta, y maduración. También se realizan en estos mutantes estudios sobre la tolerancia a la salinidad y la carencia hídrica. La caracterización molecular de estos mutantes, que es muy complicada por la alta heterocigosis de los cítricos, se está abordando mediante el estudio de la expresión génica en microarrays y técnicas de Tilling. Con esta técnica también se pretende identificar SNPs en los grupos principales de cítricos.

UNA APROXIMACIÓN GENÓMICA AL ESTUDIO DE LAS RESPUESTAS DE LOS FRUTOS CÍTRICOS A ESTRESSES BIÓTICOS Y ABIÓTICOS DURANTE LA POSTCOSECHA

González-Candelas L.^{1*}, Sánchez-Torres P.¹, Alamar S.¹, Establés B.¹, Forment, J.², Ballester A.R.¹, Marcos J.F.¹, Lafuente M.T.¹, Zacarías L.¹

¹Instituto de Agroquímica y Tecnología de Alimentos (IATA-CSIC), Valencia, Spain

²Instituto de Biología Molecular y Celular de Plantas (IBMCP), CSIC-UPV, Valencia, Spain

*Presentador de la comunicación (lgonzalez@iata.csic.es)

Los frutos cítricos están sometidos a distintos tipos de estrés, tanto bióticos como abióticos, durante la manipulación y conservación postcosecha que afectan a su calidad y comercialización. El principal estrés biótico al que están sometidos los frutos cítricos es el ataque por hongos del género *Penicillium*, siendo *P. digitatum* el patógeno con mayor incidencia. Por otro lado, entre los distintos estreses abióticos, nosotros estamos interesados en los daños de frío y en la tolerancia adquirida a los mismos mediante tratamientos de calor previos a la frigoconservación. Con el fin de profundizar en los mecanismos moleculares subyacentes en las respuestas de defensa frente al ataque por patógenos y la adquisición de tolerancia a los daños de frío, hemos seguido dos aproximaciones complementarias. En el marco del proyecto español de genómica funcional de cítricos (CFGP), hemos generado una genoteca de cDNA parara cada una de las condiciones anteriores que nos permitirá disponer de una amplia colección de los genes que se están transcribiendo bajo estas situaciones. Una segunda aproximación ha consistido en la obtención de una genoteca substraída de cDNA, mediante la técnica de SSH (Suppression Subtractive Hybridization), que está enriquecida en fragmentos de cDNAs de genes de cítricos que se expresan diferencialmente en respuesta a la infección por *P. digitatum*. Después de realizar un escrutinio diferencial sobre un total de 1536 clones, seleccionamos un amplio número de los mismos para su caracterización. En esta comunicación presentaremos el análisis de los resultados de ambas aproximaciones, con un énfasis especial en las respuestas comunes entre los distintos tipos de estrés y, dentro de una misma situación, entre la genoteca no substraída frente a la genoteca de substracción, así como en la comparación con los resultados globales generados dentro del Proyecto de Genómica Funcional de Cítricos.

HACIA EL ESTABLECIMIENTO DE UNA CONEXIÓN FENOTIPO-GENOTIPO EN TOMATE

Irene Villalta¹, José María Jiménez², José Miguel Martínez-Zapater², Emilio Carbonell¹ & María J. Asins¹

1.IVIA, Apdo. Oficial, 46113 Moncada (Valencia)

2.CNB-UAM, Cantoblanco, 28049 Madrid

Las poblaciones de líneas recombinantes puras (RILs) constituyen un material experimental idóneo en genómica, proteómica y metabolómica ya que permiten la realización de estudios potentes de asociación entre la variación agronómica o fisiológica y la molecular a todos los niveles (ADN, ARN, proteínas...). Con este objetivo hemos obtenido y caracterizado mediante marcadores moleculares (SSR y SCAR) dos poblaciones de líneas F₇ derivadas, una de *Lycopersicon pimpinellifolium* (población P, 142 líneas) y la otra de *L. chesmannii* (población C, 115 líneas) que comparten el parental femenino del cruzamiento inicial, *L. esculentum* var. cerasiforme. Los mapas genéticos obtenidos para cada población cubren los 12 cromosomas y coinciden en el orden lineal de marcadores aunque difieren en la distribución de la frecuencia del alelo silvestre.

En una primera aproximación se ha utilizado la población C para seleccionar genes candidatos al control de "precocidad en la producción de frutos". De 4 genes que fueron seleccionados mediante análisis de segregación masal de líneas, dos de ellos muestran polimorfismos asociados significativamente ($p < 0.007$) con "tiempo de floración". También se ha detectado asociación entre algunos polimorfismos en estos genes y caracteres relacionados con el vigor y la productividad de frutos lo cual explica las relaciones encontradas entre estos caracteres y "tiempo de floración". Esta información será utilizada para enfocar el análisis molecular de expresión en un número mínimo de secuencias.

BIOCHIP ELECTRÓNICO PARA LA DETECCIÓN Y CARACTERIZACIÓN SIMULTÁNEA DE LOS PRINCIPALES VIRUS Y BACTERIAS PATÓGENOS DE LA PATATA

Ruiz-García, A.B.¹; Olmos, A.²; Arahal, D.R.¹; Antúnez, O.¹; Martínez-Bono, B.¹; Llop, P.²; Pérez-Ortín, J.E.¹; López, M.M.²; Cambra, M.².

¹Universidad de Valencia. Departamento de Bioquímica y Biología Molecular. 46100 Burjassot, Valencia. Ana.Belen.Ruiz@uv.es. ²Instituto Valenciano de Investigaciones Agrarias (IVIA). Dpto. de Protección Vegetal y Biotecnología. Ctra. Moncada-Náquera km 5. 46113 Moncada, Valencia.

Las enfermedades causadas por virus y bacterias son uno de los principales factores limitantes del cultivo de la patata. Estos agentes fitopatógenos provocan disminuciones en el rendimiento de la cosecha e importantes pérdidas económicas. Los análisis de selección de lotes de patata de siembra se realizan mediante la técnica ELISA que posee muchas ventajas por su fiabilidad, facilidad de uso y capacidad de análisis de muchas muestras en un solo ensayo. Los chips de DNA suponen una nueva y potente tecnología que puede ser de gran interés para detectar simultáneamente un gran número de patógenos en una única prueba. Uno de los soportes más novedosos dentro de este campo son los chips electrónicos, que presentan una mayor versatilidad en el análisis simultáneo de muestras que requieren diferentes condiciones de hibridación y por su capacidad de diferenciar fácilmente SNPs. En este trabajo se han desarrollado y evaluado chips electrónicos de DNA basados en la tecnología de Nanogen, capaces de detectar y caracterizar los principales virus (PVY, PVX y PLRV) y bacterias (*Ralstonia solanacearum*, *Pectobacterium carotovorum subsp. carotovorum*, *P. carotovorum subsp. atrosepticum*, *P. chrysanthemi* y *Clavibacter michiganensis subsp. sepedonicus*) que infectan la patata. La detección de los virus analizados se llevó a cabo mediante iniciadores capaces de amplificar por RT-PCR regiones específicas de cada virus. Las secuencias de DNA obtenidas fueron analizadas mediante sondas específicas para cada uno de los virus estudiados. En el caso de PVY también se ha podido diferenciar entre diferentes razas del virus (PVYN, PVY0, PVYNTN), mediante sondas específicas diseñadas dentro de una región conservada entre ellas.

Para la detección simultánea de todas las bacterias se han diseñado iniciadores presentes en dos regiones conservadas del RNA ribosómico 16S. Una vez amplificadas por una única PCR estas dos regiones, se procedió a la detección específica de cada una de las bacterias de interés, mediante sondas específicas. El biochip desarrollado ha permitido la detección específica de cualquiera de estos agentes fitopatógenos en muestras de patata de campo y podría ser una alternativa a los métodos hasta ahora utilizados.

Sesión IV: GENÓMICA FUNCIONAL.

Moderador José García-Martínez (UVEG).

IMPACT OF CONTROLLED OVARIAN HYPERSTIMULATION ON ENDOMETRIAL GENE EXPRESSION PROFILES DURING THE WINDOW OF IMPLANTATION.

*Horcajadas JA**, *Riesewijk A‡*, *Polman J‡*, *van Os R‡*, *Mosselman S‡*, *Pellicer A*†* and *Simón C*†§*

Foundation Instituto Valenciano de Infertilidad, and †Department of Pediatrics, Obstetrics and Gynecology, Valencia University School of Medicine, Valencia, Spain ‡NV Organon, Departments of Target Discovery & Pharmacology, Oss, The Netherlands.

Objective: Gene expression profiling of human endometrial receptivity has recently started and demonstrated that endometrial receptivity is an active process involving hundreds of up-/down-regulated genes. In the present work, we have investigated the impact of controlled ovarian hyperstimulation (COH) on endometrial gene expression in human endometrial receptivity.

Design: Endometrial samples were collected from ovum donors at days LH+2 (n=5) and LH+7 (n=5) as determined by urinary LH surge during a single natural cycle. Also, endometrial samples were obtained from the same patient at day LH+7 (n=5) of the natural cycle and at day hCG+7 (n=5) of the next cycle during COH used for IVF treatment. Genomic analysis of endometrial samples was performed using microarray technology.

Material and Methods: The protocol for ovarian stimulation consisted of a long protocol with leuprolide acetate with FSH/HP, hMG and hCG. Leuprolide acetate and gonadotropin injections were discontinued the day of hCG administration, and progesterone supplementation was not administered. Samples were hybridized onto the GeneChip HG_U133A (Affymetrix) encompassing more than 22,000 human DNA fragments, and the results were validated by RT-PCR and quantitative RT-PCR .

Results: COH induces large differences in endometrial gene expression as compared to the previous natural cycle from the same patient. In total 558 genes on the Affymetrix chip showed differential expression, of which 281 were up-regulated as compared to the natural cycle (166 genes between 2 to 3 fold, 74 genes 3 to 5 fold and 41 >5 fold-increase). Down-regulation was identified for 277 genes (of which 161 were regulated between 2 to 3 fold, 72 between 3 to 5 fold and 44 >5 fold). The LH+2 and LH+7 comparison revealed similar gene expression profiles as previously obtained with the HG_U95A chip.

Conclusions: Our study demonstrates that the gene expression in the COH hCG+7 samples is deregulated as compared to the natural cycle at LH+7. No luteal support was administered during COH for ovum donation, in order to get a clear comparison between natural and COH cycles. Genes regulated during the window of implantation in the natural cycle are more comparable with the LH+2 than with the LH+7 patterns in the COH cycle. This suggests a delay in the gene expression regulation that is required for the formation of a receptive endometrium in COH cycles. This remarkable finding justifies further research to identify COH protocols with minimal impact on endometrial development.

CONSTRUCCIÓN DE LA PARED CELULAR DE *CANDIDA ALBICANS*: ANÁLISIS DEL TRANSCRIPTOMA DURANTE LA REGENERACIÓN DE LA PARED CELULAR POR PROTOPLASTOS.

Castillo L., Garcerá A., Martínez A.I., Elorza M.V., Moreno I., Valentín E. Y Sentandreu R.
Departamento de Microbiología y Ecología, Facultad de Farmacia, Universitat de València.
Avda Vicent Andrés Estellés s/n, Burjassot-València.

El mecanismo de la construcción de la pared celular de *C. albicans* ha sido estudiado en protoplastos (células desprovistas de esta estructura) determinándose la cinética del transcriptoma durante su regeneración (5 hr a 28 °C). Se han utilizado microarrays de *C. albicans* que incorporan 6039 genes y el análisis de los datos muestra que 1128 genes se encuentran sobre-expresados en algún momento del proceso. Los valores de expresión fueron validados mediante RT-PCR semi-cuantitativa. Un grupo de 82 genes que codifican proteínas estructurales de pared o que participan en su síntesis presenta un patrón de expresión diferente del de las células de *C. albicans* CAI-4 en crecimiento exponencial. En protoplastos a 4 °C se observa la activación de 22 genes potencialmente de pared y los primeros que se expresan al iniciarse la regeneración (30 min a 28 °C) se encuentran relacionados con los choques térmico y osmótico o implicados en transporte y energía, resultado esperado ya que los protoplastos deben reiniciar los procesos biológicos básicos. Los genes de pared que primero se expresan codifican una proteína Pir (protein with internal repeats) y tres aglutininas.

La cinética de la expresión de determinados genes demuestra ser secuencial durante la regeneración sugiriendo que determinados productos génicos son necesarios previamente a que otros puedan realizar su función y demostrando que la construcción de la pared celular es un proceso altamente ordenado tanto en tiempo como en el espacio. Los genes sobre-expresados fueron agrupados en 15 categorías funcionales, destacando que cerca de un 50% de los genes activados durante la regeneración de la pared celular son de función desconocida.