

## REFERENCIAS

Brunner, R. J., Djorgovski, S. G., Prince, T. A., and Szalay, A. S.: 2002, in J. S. Mulchaey and J. T. Stocke (eds.), *Extragalactic Gas at Low Redshift*, Vol. 254 of *Astronomical Society of the Pacific Conference Series*, p. 383.

Hassan, A. and Fluke, C. J.: 2011, *Publications of the Astronomical Society of Australia*. 28, 150.

McCord, T. B. and Bosel, J. P.: 1975, in *Charge-Coupled Device Technology for Scientific Imaging Applications*, pp 65-69.

Parimucha, S. and Vanko, M.: 2005, *Contributions of the Astronomical Observatory Skalnaté Pleso* 35, 35.

Samuelsson, H.: 1975, *ESA Scientific Technical Review* 1, 219.

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., and Pearce, F.: 2005, *Nature* 435, 629.

# MANEJO DE DATOS MASIVOS EN BIOMEDICINA COMPUTACIONAL



**VÍCTOR  
CASTAÑEDA**

Ingeniero eléctrico, Universidad de Chile (2005) y Doctor en Ciencias de la Computación, TUM-Alemania (2012), especializado en el procesamiento de imágenes médicas. Su Tesis Doctoral estuvo enfocada en el procesamiento de imágenes endoscópicas y sensores 3D tales como cámaras Time-Of-Flight y Kinect. Actualmente hace su postdoctorado (Proyecto FONDECYT 3140444) en el seguimiento de núcleos y segmentación de membranas celulares provenientes de microscopía Light Sheet, en SCIAN Lab del Prof. Härtel, ICBM, Facultad de Medicina, U. de Chile.

[vcastaneda@med.uchile.cl](mailto:vcastaneda@med.uchile.cl)

**A nivel mundial, la investigación de excelencia en el ámbito biológico, clínico, médico y biomédico depende en forma crucial de la capacidad de análisis de los datos recolectados por experimentos que generan crecientes volúmenes de datos. Como ejemplo, la microscopía confocal *in vivo* es capaz de generar cientos de gigabytes (GB) de imágenes tridimensionales en una sola captura. En estos casos los investigadores de BioMedicina Computacional recurren a información cuantitativa, mediante modelamiento matemático y computacional para entender y predecir procesos biológicos con relevancia en medicina y ciencia básica.**

Las aplicaciones involucradas entrelazan a los mundos de la Salud Pública (bases de datos), Clínico/Hospitalario (sistemas de información), (Neuro) Ciencias Biomédicas (imágenes, bioinformática y biología computacional), Ciencias de la Computación/Ingeniería (algoritmos), Física y Matemática (herramientas y modelos). Se reconoce que la creación del campo de la BioMedicina Computacional requiere un esfuerzo mayor a través de los años para generar equipos multidisciplinares y una nueva cultura de trabajo desde las ciencias básicas hasta la investigación clínica, salud pública, y la introducción de nuevos servicios en sistemas de salud. Hasta la fecha, Chile y la mayoría de los países latinoamericanos no cuentan con respuestas adecuadas en este tema, principalmente por la falta inversión de fondos estratégicos con visión de mediano y largo plazo. La obtención de esta capacidad conlleva a la creación de alianzas estratégicas entre las disciplinas involucradas para desarrollar y acceder a nuevas tecnologías necesarias para el análisis de datos masivos. Instituciones emblemáticas en todo el mundo han respondido a este desafío a través de la creación de centros o institutos que persiguen misiones afines: (i) Johns Hopkins University [1], (ii) U-Michigan [2], (iii) U-Cincinnati [3], (iv) Janelia Farm [HHMI] [4], (v) BioQuant, DKFZ, Uni-Heidelberg [5], y (vi) Mt. Sinai [6], por nombrar solo algunos.

El análisis de datos en BioMedicina Computacional requiere de una

alta capacidad de procesamiento, siendo necesaria la interacción con arquitecturas de computación de alto rendimiento (high performance computing, HPC). Como ejemplo, un equipo para la digitalización de muestras de tejido (como biopsias) puede escanear una serie de portaobjetos completos, generando varios GB de imágenes en una captura (aproximadamente 10 terápíxeles). Esta imagen de gran tamaño debe ser procesada para extraer características que permitan reconocer distintos tipos de células características de diversas patologías como el cáncer. Su procesamiento requiere de muchos procesadores (como cluster de CPU o GPU, comunes en HPC) para disminuir los tiempos de respuesta a niveles que permitan el diagnóstico y tratamiento oportuno. Otro ejemplo está en el ámbito de la genómica, donde un gran número de algoritmos matemáticos son aplicados para extraer características de los genomas recolectados provenientes de distintos individuos, generando varios GB de datos que deben ser almacenados en una memoria principal para su procesamiento, lo que también genera la necesidad de acceder a arquitecturas HPC.

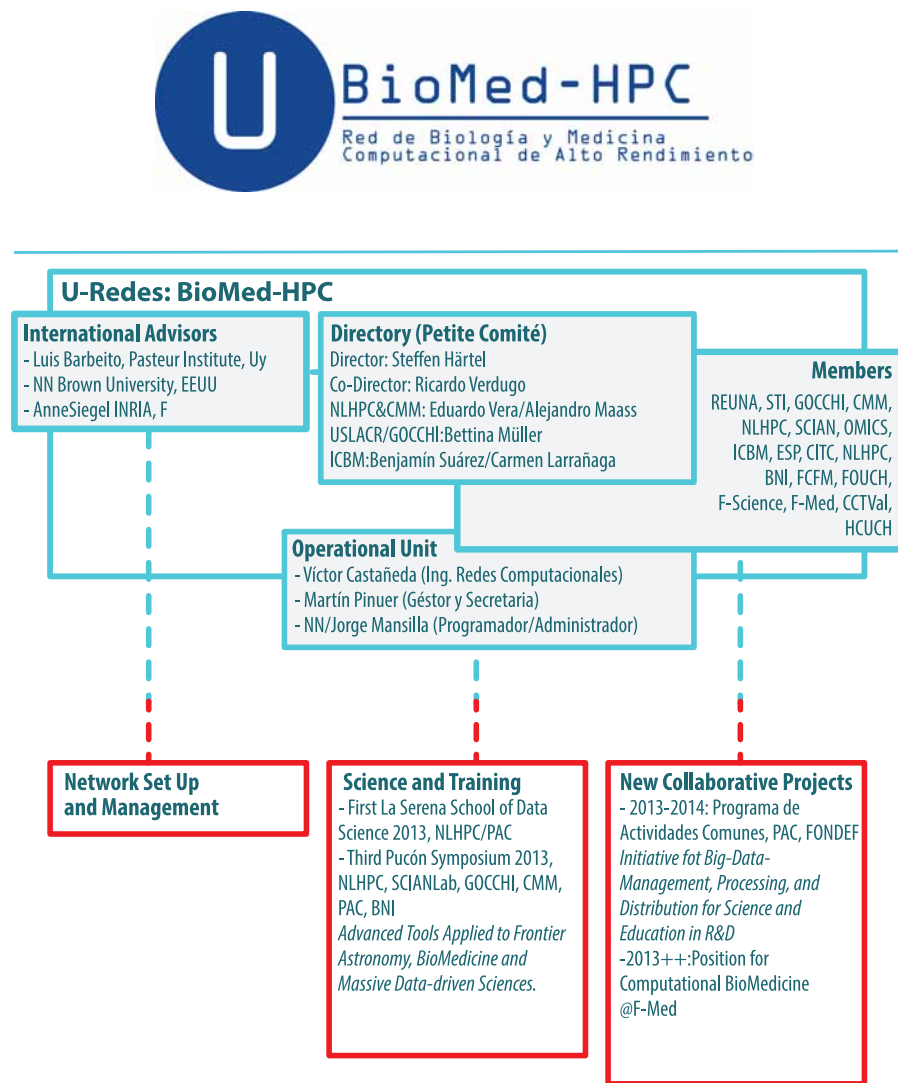
El manejo de datos masivos en el ámbito de la BioMedicina Computacional genera grandes desafíos, desde luego en el análisis pero también problemas de almacenamiento, manipulación, confidencialidad, seguridad y transmisión de datos. En este ámbito reviste especial importancia la seguridad en el almacenamiento y trans-

misión de datos relacionados a pacientes, que deben estar salvaguardados ante cualquier evento indeseado. Es por esto que se requieren diseños para sistemas con altos niveles de seguridad y confidencialidad, que garanticen su correcto uso y manipulación en caso de ser utilizados en esce-

narios de investigación, planificación o reportes. Se deriva también un desafío importante para generar herramientas de cómputo con HPC accesibles a usuarios sin conocimientos relacionados con Ciencia de la Computación, como por ejemplo investigadores del área de la Biología.

## BIOMED-HPC: UNA INICIATIVA PIONERA EN SUDAMÉRICA

Dentro de los ejemplos regionales está la reciente creación del Insti-



**FIGURA 1.**  
ORGANIGRAMA DE U-REDES.

tuto de Investigación en Biomedicina de Buenos Aires CONICET, relacionado con la Sociedad Max Planck en Buenos Aires, dedicado a temas actuales de las biociencias, especialmente investigación en biomedicina [7], y la reciente creación de la red interna BioMed-HPC en la Universidad de Chile, que interconecta las facultades de Medicina, Odontología, Química y Farmacia, Ciencias y Ciencias Físicas y Matemáticas, en una amplia e interdisciplinaria red de colaboración (Figura 1).

El proyecto BioMed-HPC está dentro del contexto de los proyectos internos de la Universidad de Chile, llamados U-Redes Domeyko II. BioMed-HPC busca fortalecer el desarrollo de investigación básica y aplicada en todas las áreas abarcadas por los miembros de la red, desde la física a estudios en el cáncer y los sistemas de información hospitalarios. Para alcanzar el objetivo, la estrategia de la Red es: (i) facilitar la gestión de proyectos colaborativos, sobre todo en aspectos financieros y organizacionales, y (ii) asegurar el financiamiento necesario para mejorar la infraestructura de Intranet en la U. de Chile que permita el flujo de grandes cantidades de datos entre unidades académicas ubicadas en distintas facultades. Los tres pilares que sustentan esta Red son una unidad de operaciones, la organización de eventos formativos y la implementación de infraestructura. El

primer pilar es una unidad de operaciones, la cual es financiada por U-Redes y contribuciones del Instituto Ciencias Biomedicas - ICBM, National Laboratory for High Performance Computing - NLHPC, Red Universitaria Nacional - REUNA, Dirección de Servicios de Tecnologías de Información y Comunicaciones - STI U-Chile, United States-Latin America Cancer Research Network - USLACRN. El segundo pilar es la formación avanzada de científicos mediante la organización de simposios y escuelas de verano, en que especialistas internacionales del amplio campo de la informática médica y HPC compartirán eventos con los integrantes de BioMed-HPC. Por último, el tercer pilar es la creación de una nueva red física de investigación de alta velocidad (10 Gbps) que permitirá: (i) transmisión de los datos entre el Campus Norte (en un comienzo la Facultad de Medicina) y el clúster HPC del proyecto NLHPC en el CMM, en la Facultad de Ciencias Físicas y Matemáticas, (ii) transmisión segura de datos masivos, (iii) análisis de los datos en el clúster antes mencionado, (iv) acceso a almacenamiento seguro y confidencial, y (v) creación de herramientas remotas para la utilización de HPC para usuarios no especializados en supercómputo. Esta Red se habilitará a mediados de 2014, comenzando con la marcha blanca de la red física exclusiva de investigación. ■

## REFERENCIAS

- [1] Development of quantitative approaches for understanding the mechanisms and treatment of human disease through applications of mathematics, engineering and computational science (R. Winslow, Dir. of the Institute for Computational Medicine, J Hopkins University, [www.icm.jhu.edu](http://www.icm.jhu.edu)).
- [2] Creation of novel and impactful informatics and computationally-based methods, tools, algorithms, and resources to extend basic and clinical research capabilities and results (Department of Computational Medicine and Bioinformatics, U-Michigan).
- [3] Usage of data and computational systems to make disease more preventable, illness more predictive and treatment more personalized (Computational Medicine Center, U-Cincinnati).
- [4] Understanding the core principles that give rise to the developmental building plans of animals from single-cell to whole-organism level, Janelia Farm, Howard Hughes Medical Institute (HHMI), (<http://janelia.org/lab/keller-lab>).
- [5] The Center for "Quantitative Analysis of Molecular and Cellular Biosystems", which functions as a platform for the development and constant refinement of mathematical models of complex biological systems as well as the swift validation of scientific hypotheses via experimental data, BioQuant, Heidelberg University (<http://www.bioquant.uni-heidelberg.de/>).
- [6] Icahn School of Medicine at Mount Sinai - New York City. <http://www.mssm.edu/>
- [7] Polo Científico Tecnológico, Agencia Nacional de Promoción Científica Tecnológica y CONICET, Argentina. <http://www.mincyt.gob.ar/polo>