



Instituto de Investigación Sanitaria Gregorio Marañón

## Servicios de Apoyo a la Investigación –

### Unidad de Bioinformática

### Cartera de Servicios y Tarifas

#### 1. Justificación

La Bioinformática y su aplicación en el procesado y análisis de datos se ha convertido en un pilar indispensable no sólo de la investigación básica centrada en el procesado de grandes volúmenes de datos biológicos, sino también en una herramienta de ayuda a la toma de decisiones en la investigación y la medicina traslacional. Desde hace casi dos décadas la biología molecular y la bioinformática han encontrado una vía de desarrollo y expansión en el área sanitaria como herramientas para el diagnóstico y pronóstico de numerosas enfermedades con componente genético, en la estratificación de los pacientes en función de su trasfondo genético, en el seguimiento de la enfermedad y en la potencial identificación de dianas terapéuticas en el escenario de la medicina personalizada.

Con estos antecedentes, la Unidad de Bioinformática se constituye con la intención de ofrecer soporte para el diseño y el análisis de datos genómicos procedentes de proyectos financiados de los grupos de investigación de dentro y fuera del Instituto de Investigación Sanitaria Gregorio Marañón.

#### 2. Descripción de las instalaciones

El *core* principal de la Unidad de Bioinformática se encuentra situado en la Planta Baja del Edificio de Medicina y Cirugía Experimental del HGUGM. Para la realización de estudios en los que sea necesaria una gran carga computacional, como tareas de *Next Generation Sequencing* o modelizado estadístico avanzado, se cuenta con dos *Workstation* de alta gama y dos servidores NAS, adquiridos por el Grupo de investigación de oncología médica traslacional y terapia individualizada del cáncer del IISGM. Se trata de dos *Workstation* de Hewlett-Packard de la serie Z, un servidor Proliant ML y una NAS QNAP TS-EC2480U, además de equipos de sobremesa y la infraestructura de redes necesaria.

La Unidad de Bioinformática trabaja de manera coordinada con el Laboratorio de Oncología Traslacional (LAOT) y está iniciando su colaboración con el SAI de la Unidad de Genómica, siendo frecuentes las tareas de procesado de datos procedentes de las plataformas de cuantificación de NanoString y de Illumina MiSeq.



### 3. Cartera de Servicios

#### **3.1 Análisis de Genómica**

##### a.- Ensamblado y alineamiento de secuencias de DNA

Tras un paso en el que se hace un control de calidad tanto de la calidad de la muestra como de la presencia de potencial contaminación cruzada, se procede al ensamblado y alineamiento a partir de datos crudos procedentes de plataformas de secuenciación.

##### b.- Análisis de exoma humano

Se procede a la anotación y llamada a variantes y CNVs.

##### c.- Genotipado

Análisis de genotipado a todos los niveles, siendo posible el abordaje de análisis de GWAS sencillos. En el caso de GWAS con muestras procedentes de muestras de varias poblaciones se puede añadir un paso de análisis destinado a la caracterización y corrección de la posible presencia de estratificación poblacional en las muestras.

#### **3.2 Análisis de Transcriptómica**

##### a.- Ensamblado y alineamiento de secuencias de mRNA

Tras un paso en el que se hace un control de calidad tanto de la calidad de la muestra como de la presencia de potencial contaminación cruzada, se procede al ensamblado y alineamiento a partir de datos crudos procedentes de plataformas de secuenciación.

##### b.- Análisis de microarrays

Tanto procedentes de plataformas comerciales o de plataformas personalizadas, seguido de un análisis simple de expresión diferencial entre grupos de muestras.

##### c.- Análisis funcional de los biomarcadores

Se establece la potencial función biológica del conjunto de biomarcadores diferencialmente expresados del estudio mediante aproximaciones basadas en Gene Ontology y análisis de enriquecimiento de *pathways*.

##### d.- Cuantificación de transcritos de mRNA-Seq.

Se pueden utilizar diversos algoritmos en función del objetivo del estudio. Dependiendo de cuáles vayan a ser los análisis posteriores se normalizan los datos

mediante el procedimiento más adecuado y se procede al análisis de expresión diferencial entre grupos de muestras.

e.- Estudios de *splicing* alternativo e isoformas usando datos de mRNA-Seq.

f.- Estudios de expresión usando la plataforma nCounter de NanoString.

### **3.3 Programación y apoyo estadístico**

#### a.- Análisis estadístico

Basados en su mayor parte en el uso de R/Bioconductor. Abarca el desarrollo de *scripts* en R, estadística descriptiva y análisis de asociación, análisis de *clustering* y componentes principales (PCA); además de cálculo de tamaños muestrales y potencia estadística.

#### b.- Evaluación de efectos de *batch*

Evaluación de efectos de *batch* en los datos experimentales mediante *clustering*, PCA, *multidimensional scaling* (MDS) o programas *ad hoc*.

c.- Asistencia para la preparación de manuscritos (métodos y resultados) y aplicación a convocatorias de ayudas competitivas (diseño experimental/bioinformático del estudio, material y métodos, resultados preliminares o estudio piloto).

## **4. Procedimiento de solicitud de un servicio determinado**

Se utilizará el Impreso *Hoja de Encargo IM-SPE-07* del IISGM.

La Hoja de Encargo consta de 5 partes:

- 1- Identificación SAI
- 2- Identificación Solicitante\*
- 3- Datos del servicio requerido y presupuesto\*\*
- 4- Observaciones: Indicar si es auto-factura o no\*\*\*
- 5- Firmas

\* Identificación del solicitante, en el caso en el que el investigador a imputar el cargo del servicio sea un nuevo usuario, deberá cumplimentar y adjuntar con la primera hoja de encargo el formulario de datos de facturación firmado.

\*\*Datos de los servicios prestados:

- [ ] Ensamblado y alineamiento
- [ ] Análisis de exoma
- [ ] Análisis de genotipado
- [ ] Análisis de microarrays
- [ ] Análisis de mRNA-Seq humano



Instituto de Investigación Sanitaria Gregorio Marañón

- [ ] Análisis estadístico
- [ ] Asistencia en escritura de manuscritos y proyectos

\*\*\*En caso de tener que justificar en proyectos públicos, se aplicará el procedimiento PNT Autofacturación PR-SPE-06 del IISGM.

*IMPORTANTE: el gasto debe haber sido aprobado por la agencia financiadora, bien porque estaba contemplado en el presupuesto de la solicitud inicial (situación ideal), o bien porque se haya aprobado posteriormente el gasto después de solicitud formal por escrito.*

El procedimiento a seguir es el siguiente:

- Solicitud del estudio que se desea realizar mediante una memoria de extensión breve en la que se especifiquen claramente el objetivo primario del estudio y los secundarios, si los hubiera.
- Reunión con los miembros del SAI de Bioinformática del área de análisis bajo el que se engloba el estudio para poner en común cuestiones como el tamaño muestral con el que se va a trabajar, la potencia estadística que se desea obtener y la variables que se desean analizar.
- Preparación y entrega de los datos a ser analizados en el formato previamente acordado entre ambas partes.
- Realización del análisis.
- Entrega de resultados en los formatos pertinentes.
- Los datos crudos de los que parte el análisis se archivan en los servidores de Unidad de Bioinformática durante 2 años para posibles (re)análisis sucesivos, siempre y cuando ambas partes lo acuerden.

## 5. Personas de contacto

Antonio Clemente Picornell Company  
Responsable de la Unidad de Bioinformática  
Instituto de Investigación Sanitario Gregorio Marañón (IISGM)  
Calle Doctor Esquerdo 46  
28007 Madrid  
E-mail: antonio.picornell@iisgm.com  
Telf. 915290724

Enrique Luis Álvarez Castillo  
Master Predoctoral de la Unidad de Bioinformática