

# Propuesta de estancia formativa

## SOPHiA GENETICS

### VCF format check

*Con una perspectiva transfronteriza, el proyecto europeo Refbio II tiene como fines reforzar la colaboración entre instituciones y mejorar la permeabilidad y sinergias entre el medio académico y el empresarial en el ámbito de la biotecnología.*

*En este contexto, se proponen misiones de investigadores en empresas por un periodo de 3 meses, siendo la financiación del doctorando a cargo de REF BIO II.*

SOPHiA GENETICS se fundó en 2011 para dar forma al futuro de la genómica clínica. Con la adopción de tecnologías digitales, como Next Generation DNA Sequencing (NGS), la industria ha entrado en el mundo de Big Data, planteando nuevos retos en el análisis de datos y la seguridad. Para resolver estos retos, un equipo de expertos de renombre mundial ha desarrollado la Inteligencia Artificial SOPHiA (IA). Al igual que los relojeros suizos, la tecnología de SOPHiA GENETICS es reconocida por su calidad, precisión y fiabilidad. Esto ha resultado en las certificaciones ISO 13485 (Sistema de Gestión de Calidad de Dispositivos Médicos) e ISO 27001 (Sistema de Gestión de Seguridad de la Información).

SOPHiA GENETICS proporciona tecnología avanzada y universal para el análisis de datos de NGS.

SOPHiA, la tecnología universal para la medicina basada en datos (disponible exclusivamente en SOPHiA DDM) detecta, anota y preclasifica variantes genómicas para ayudar a los médicos a utilizarlas para diagnosticar y tratar mejor a sus pacientes.

SOPHiA DDM es una plataforma SaaS (software-as-a-service) que facilita la visualización e interpretación de variantes, gracias a su intuitiva interfaz de usuario y a las funciones integradas. Además, protege todos los datos cifrándolos según los más altos estándares de la industria antes de almacenarlos de forma redundante en centros de datos privados y seguros.

Permite el análisis y protección de los datos clínicos de NGS. Nuestras tecnologías principales procesan y analizan datos genómicos en bruto para ayudar a los hospitales a diagnosticar con precisión y rapidez a los pacientes en áreas como oncología, metabolismo, pediatría, cardiología y cáncer hereditario.

### **Actividad prevista del doctorando en la entidad:**

**Objeto de investigación :** Proyecto 1: VCF format check

El formato de archivo VCF es un formato estándar para describir las diferencias entre una muestra y una referencia. Estas diferencias pueden ser pequeñas variantes como SNP e Indels pero también grandes modificaciones como CNVs.

En SOPHiA GENETICS, los archivos VCF se manejan en muchas etapas internas del proceso. El archivo VCF final se pone a disposición del cliente a través de nuestra plataforma SOPHiA DDM.

Dado que muchos de nuestros clientes utilizan el archivo VCF para el análisis posterior con la ayuda de otros programas informáticos de terceros, es esencial que el archivo VCF cumpla con las especificaciones oficiales del formato VCF.

Por lo tanto, nos gustaría implementar un verificador de formatos VCF que asegure que los archivos VCF son compatibles y que proporcionará mensajes de error significativos si este no es el caso.

En el futuro, cada script que manipule archivos VCF deberá ejecutar el checker para asegurarse de que la salida producida por el script es correcta.

Programa financiado por el FEDER

### Resultados esperados:

Una herramienta para comprobar que el formato se ajusta a la especificación del formato VCF y genera mensajes de error significativos en caso contrario. Debe configurarse de forma eficiente para que un archivo VCF grande con 100.000 variantes pueda comprobarse en cuestión de segundos.

De este modo, SOPHiA GENETICS podrá garantizar que todas las herramientas desarrolladas internamente y que produzcan archivos VCF cumplan con las especificaciones del formato VCF. También garantizará que el archivo VCF que puede ser descargado por el cliente SOPHiA GENETICS a través de la plataforma SOPHiA GENETICS DDM sea un VCF válido y pueda ser utilizado en software de terceros para su posterior análisis si fuera necesario.

### Perfil y competencias requeridas:

Buscamos un estudiante con habilidades de programación, idealmente con experiencia en Bio-IT, que se sienta cómodo en programación R, Python o C++. El candidato debe saber cómo gestionar un proyecto desde el principio (recopilación de las necesidades del usuario) hasta la finalización (implementación que corresponda a las necesidades). También deben ser capaces de comunicar y documentar correctamente su proyecto para que pueda ser mantenido después de su partida. El conocimiento de las tecnologías de análisis de datos de las NGS es una ventaja, pero no un requisito para la pasantía.

### Área de estudios :

Informática, bioinformática o biotecnología

### Área de investigación:

Desarrollo informático, análisis de datos, visualización de datos; la experiencia con datos biológicos sería una ventaja, pero no es obligatoria.

### Competencias deseadas:

Motivado, comunicación fácil, bien organizado, independiente.

### Experiencia profesional:

No se requiere nivel de experiencia.

### Idiomas deseados:

Inglés obligatorio, idioma utilizado principalmente en SOPHiA GENETICS.

**Material y procesos implicados :** Se suministra con ordenador.

El ordenador estará equipado con todos los programas y aplicaciones necesarios.

**Fecha inicio de la actividad:** LO ANTES POSIBLE

**Duración de las prácticas:** 3 meses

**Sitio de prácticas:**

SOPHiA GENETICS

Technopole IZARBEL

64210 BIDART – Francia

### Contacto:

Las solicitudes deberán enviarse por correo electrónico junto con el CV y el formulario de solicitud cumplimentado, especificando la oferta de prácticas, a Cluster Santé Osasuna, antes del 31/10/2018

E-mail : [clusterosasuna@gmail.com](mailto:clusterosasuna@gmail.com)

Téléphone : 05 59 43 54 51

Programa financiado por el FEDER