

PROPUESTA DE TRABAJO FIN DE MÁSTER (Modalidad B)
Curso 2022-2023
MÁSTER EN TÉCNICAS ESTADÍSTICAS

Título	Comparación de técnicas estadísticas para evaluar el parentesco, estructura poblacional y ancestralidad genética
Nombre de la Empresa	CIMUS (Centro de Investigación en Medicina Molecular y Enfermedades Crónicas de la Universidad de Santiago)
Tutor/a en la empresa	Raquel Cruz Guerrero
Director/a	
Descripción del trabajo a realizar	<p>En los estudios de asociación de genoma completo (<i>genome wide association studies</i>, GWAS) normalmente se realiza la comparación de las frecuencias alélicas entre un grupo de casos (individuos con una patología o rasgo fenotípico de interés) y un grupo de población control (individuos sanos o que no muestran dicho rasgo).</p> <p>Una parte importante del proceso de control de calidad que hay que realizar antes del análisis estadístico de las diferencias entre las variables consiste en evaluar 1) la existencia de parentesco, 2) la posible estratificación (diferencias genéticas de fondo entre casos y controles, no relacionadas con la patología o fenotipo de estudio) y 3) el origen ancestral común de los individuos del estudio.</p> <p>En la literatura se han utilizado diversas técnicas estadísticas para valorar cada uno de estos aspectos (análisis de componentes principales, PCA, por ejemplo), pero todas ellas se basan en cierta forma en caracterizar la distancia o similitud genética entre individuos, por lo que con frecuencia los resultados son confusos: individuos de ancestralidad semejante son interpretados como parientes o distorsionan la distribución observada en el PCA con el que normalmente se caracteriza la existencia de estratificación.</p> <p>El problema se agrava en los estudios en los que hay población de diferentes orígenes ancestrales, pues en ese caso conviene realizar los análisis por separado. En el marco del proyecto SCOURGE, en el que se analiza la base genética de la severidad del COVID19, se dispone de una población numerosa de casos de origen mezclado y gran número de individuos emparentados. Los datos de dicho</p>

	proyecto, genotipados con un <i>array</i> de alta densidad e imputados a más de seis millones de posiciones del genoma, permitirían poder evaluar y comparar las diversas técnicas existentes en un caso especialmente complejo y definir la correcta <i>pipeline</i> como parte del proceso de control de calidad.
Recomendaciones	
Fechas de las practicas	A convenir
Lugar de trabajo y horario	A convenir
Convenio	
Otras observaciones	