Trabajo de fin de grado realizado en el centro Pfizer de Genómica e Investigación Oncológica (GENyO)

Grado en Ingeniería Electrónica Industrial

# Descubrimiento de Marcadores en Enfermedades Complejas a partir de Datos Genómicos mediante Técnicas de Entrenamiento Supervisado

**Autor**
Luis Martín Gámez

**Directores**
Juan Manuel Górriz Sáez
Pedro Carmona Sáez

Granada, 1 de septiembre de 2018

# Descubrimiento de Marcadores en Enfermedades Complejas a partir de Datos Genómicos mediante Técnicas de Entrenamiento Supervisado

Luis Martín Gámez

**Palabras clave**: Validación cruzada, matriz de confusión, error OOB, gen, expresión, metilación, integración, base de datos, random forest, máquina de soporte vectorial, SVM, clasificación, lupus, SLE, transcriptoma, metiloma, técnicas de clasificación, entrenamiento supervisado.

## Resumen

La presente investigación, realizada por el autor en el Centro Pfizer de Genómica e Investigación Oncológica (GENyO), tiene por objeto la búsqueda de marcadores en enfermedades complejas, en particular de la enfermedad del Lupus Eritematoso Sistémico, a partir de datos genómicos heterogéneos de transcriptoma y metiloma. Mediante técnicas de entrenamiento supervisado se construyen diferentes modelos de clasificación binaria, realizando un estudio comparativo entre técnicas lineales y no lineales. Los principales métodos utilizados son SVM (máquina de soporte vectorial) y Random Forest. Este estudio presenta un análisis comparativo de diferentes técnicas de clasificación aplicadas a tres conjuntos o bases de datos: en la primera se utilizan como predictores datos de transcriptoma o expresión de genes; en la segunda, sitios CpG de metilación y, en la tercera, se integran las variables de expresión y metilación con el objeto de descubrir si se reduce el error de clasificación.