

Estudio bioinformático para la identificación de potenciales modelos de enfermedades genéticas usando el panel NCI-60.

Bioinformatic study for identification of potential genetic disease models using the NCI-60 panel.

Trabajo Fin de Grado en Bioquímica
Universidad de Málaga

Autor: María Sánchez Martín.

Área de conocimiento y Departamento: Bioquímica y Biología Molecular; Biología Molecular y Bioquímica.

Fecha de presentación: Junio de 2023.

Tema: Estudio de mecanismos de adaptación y resistencia celular en cáncer.

Tipo: Experimental.

Modalidad: Individual.

Nº de páginas de la memoria: 40.

Nº de páginas de los anexos: 10.

DECLARACIÓN DE ORIGINALIDAD DEL TFG

D./Dña. María Sánchez Martín, con DNI 77241688Y, estudiante del Grado en Bioquímica de la Facultad de Ciencias de la Universidad de Málaga,

DECLARO:

Que he realizado el Trabajo Fin de Grado titulado “Estudio bioinformático para la identificación de potenciales modelos de enfermedades genéticas usando el panel NCI-60” y que lo presento para su evaluación. Dicho trabajo es original y todas las fuentes bibliográficas utilizadas para su realización han sido debidamente citadas en el mismo.

De no cumplir con este compromiso, soy consciente de que, de acuerdo con la Normativa reguladora de los procesos de evaluación de los aprendizajes del estudiantado de la Universidad de Málaga de 23 de julio de 2019, esto podrá conllevar la calificación de suspenso en la asignatura, sin perjuicio de las responsabilidades disciplinarias en las que pudiera incurrir en caso de plagio.

Para que así conste, firmo a presente en Málaga, el 26 de mayo 2023

Fdo:.....

ÍNDICE

Resumen

Abstract

1. Introducción	1
1.1. ¿Qué es el cáncer?	1
1.1.1. Inestabilidad genética y mutacional.....	2
1.2. Enfermedades de base genética	2
1.2.1. ¿Qué son las enfermedades raras?	3
1.3. ¿Qué son los genes esenciales?.....	3
1.3.1. Índices de esencialidad	4
1.4. Empleo de métodos computacionales en la investigación de cáncer y enfermedades raras	4
2. Hipótesis y objetivos.....	6
3. Materiales y metodología.....	6
3.1. Materiales	6
3.2. Metodología	9
3.2.1. Búsqueda bibliográfica	9
3.2.2. Primer filtrado de datos: determinación de las líneas celulares	9
3.2.3. Segundo filtrado de datos: determinación de los genes esenciales	10
3.2.4. Determinación de la distribución de estos genes	10
3.2.5. “Cross-indexado” de genes DepMap/Monarch	11
3.2.6. Genoma humano	11
3.2.7. Estudio de enriquecimiento funcional	11
3.2.8. Estudio de anotaciones GO.....	12
3.2.9. Pruebas estadísticas	13
3.2.9.1. Test de Fisher.....	13
3.2.9.2. <i>t</i> de Student	13
4. Resultados.....	14
4.1. Búsqueda bibliográfica	14
4.2. Filtrado de datos.....	14
4.3. Distribución de datos	15

4.4. “Cross-indexado” de genes DepMap/Monarch	16
4.5. Estudios de enriquecimiento.....	17
4.6. Estudio de anotaciones GO.....	27
4.7. Pruebas estadísticas	28
4.7.1. Test de Fisher	28
4.7.2. <i>t</i> de Student	29
5. Discusión	32
6. Conclusiones	37
7. Perspectivas futuras y limitaciones.....	38
Bibliografía	
Webgrafía	
Anexo I	

Estudio bioinformático para la identificación de potenciales modelos de enfermedades genéticas usando el panel NCI-60

RESUMEN

Las enfermedades genéticas son un grupo heterogéneo de enfermedades hereditarias con baja prevalencia en la población. Por ello, los datos disponibles para la investigación suelen ser limitados. El cáncer se caracteriza por ser un grupo heterogéneo de enfermedades que aparecen por la acumulación de mutaciones. El Instituto Nacional del Cáncer ha desarrollado el panel NCI-60, un cribado de líneas celulares de cáncer, con el fin de avanzar en la investigación del cáncer y la enfermedad. En este TFG investigamos el panel NCI-60 en cuanto a su uso potencial como modelo complementario para el estudio de enfermedades genéticas. Se utiliza la base de datos DepMap para obtener datos relacionados con el panel e identificar genes esenciales para diferentes líneas celulares. A continuación, utilizando el lenguaje de programación R, comparamos estos genes esenciales con Monarch, una base de datos que contiene información sobre enfermedades genéticas. La selección de los genes esenciales se realiza utilizando una puntuación de fitness establecida por DepMap. Así, se obtiene un conjunto de datos que relaciona cada línea celular con una serie de enfermedades genéticas, en función de los genes esenciales. Por último, se realizan análisis de enriquecimiento funcional para determinar las vías en las que están implicados los genes esenciales. Los resultados indican que existe una relación directa entre la esencialidad de un gen y la probabilidad de estar relacionado con una enfermedad genética. Además, se ha podido determinar que existen líneas celulares específicas para el estudio de una determinada enfermedad.

Palabras clave:

Cáncer, enfermedades genéticas, genes esenciales, índices de esencialidad, rutas de enriquecimiento

Bioinformatic study for identification of potential genetic disease models using the NCI-60 panel

ABSTRACT

Genetic diseases are a heterogeneous group of inherited diseases with low prevalence in the population. As such, there is often limited data available for research. Cancer is characterized as a heterogeneous group of diseases that appear because of the accumulation of mutations. The National Cancer Institute has developed the NCI-60 panel, a screen of cancer cell lines, in order to advance research of cancer and disease. In this TFG we investigate the NCI-60 panel in terms of its potential use as a complementary model for the study of genetic diseases. The DepMap database is used to obtain data related to the panel to identify essential genes for different cell lines. Then, using the R programming language, we compare these essential genes to Monarch, a database containing information on genetic diseases. The selection of essential genes is performed using a fitness score established by DepMap. Thus, a dataset is obtained, which relates each cell line to a series of genetic diseases, based on the essential genes. Finally, functional enrichment analyses are performed to determine the pathways in which the essential genes are involved. The results indicate that there is a direct relationship between the essentiality of a gene and the probability of being related to a genetic disease. In addition, it has been possible to determine that there are specific cell lines for the study of a particular disease.

Keywords:

Cancer, genetic diseases, essential genes, fitness score, enrichment pathways