

TRABAJO FIN DE MÁSTER
MÁSTER UNIVERSITARIO OFICIAL EN CIENCIA DE DATOS E
INGENIERÍA DE COMPUTADORES

Epidemiología y detección de biomarcadores en cáncer

Autor:

Daniel Redondo Sánchez

Tutores:

Ignacio Rojas

Luis Javier Herrera

Daniel Castillo

Granada, septiembre de 2020



**UNIVERSIDAD
DE GRANADA**

0. Índice general

| | |
|--|-----------|
| Abstract | 5 |
| 1. Introducción | 7 |
| 1.1. Objetivos del trabajo | 7 |
| 1.2. Cáncer: definición y factores de riesgo | 7 |
| 1.2.1. Cáncer de hígado | 8 |
| 1.2.2. Cáncer de colon-recto | 9 |
| 1.3. Transcriptómica | 9 |
| 1.3.1. Ciencias -ómicas | 9 |
| 1.3.2. Secuenciación del genoma? | 10 |
| 2. Epidemiología del cáncer | 11 |
| 2.1. Indicadores epidemiológicos | 11 |
| 2.2. Epidemiología del total del cáncer (¿excepto piel no melanoma?) | 11 |
| 2.3. Epidemiología del cáncer de hígado | 12 |
| 2.3.1. Incidencia de cáncer de hígado | 12 |
| 2.3.2. Mortalidad por cáncer de hígado | 12 |
| 2.3.3. Supervivencia de cáncer de hígado | 12 |
| 2.3.4. Prevalencia de cáncer de hígado | 12 |
| 2.4. Epidemiología del cáncer de colon-recto | 12 |
| 2.4.1. Incidencia de cáncer de hígado | 12 |
| 2.4.2. Mortalidad por cáncer de hígado | 12 |
| 2.4.3. Supervivencia de cáncer de hígado | 12 |
| 2.4.4. Prevalencia de cáncer de hígado | 12 |
| 3. <i>Machine learning</i> aplicado a transcriptómica | 13 |
| 3.1. Selección de biomarcadores | 13 |
| 3.2. Algoritmos de clasificación | 13 |

| | |
|---|-----------|
| 4. Detección de biomarcadores en cáncer de páncreas | 15 |
| 4.1. Introducción | 15 |
| 4.2. Metodología | 15 |
| 4.3. Resultados | 15 |
| 4.4. Conclusiones | 15 |
| 5. Detección de biomarcadores en cáncer de colon-recto | 17 |
| 5.1. Introducción | 17 |
| 5.2. Metodología | 17 |
| 5.3. Resultados | 17 |
| 5.4. Conclusiones | 17 |
| 6. Aplicación web para realizar detección de biomarcadores | 19 |
| 7. Conclusiones y líneas abiertas de trabajo | 21 |
| Bibliografía | 22 |
| AnexoI: Código de análisis en R | 23 |
| Anexo II: Código de aplicación web | 25 |

Abstract

Abstract en inglés

Resumen

Abstract en español

1. Introducción

1.1. Objetivos del trabajo

En el presente Trabajo Fin de Máster se analiza la epidemiología de los cánceres de hígado y colon-recto, detectando genes que permiten identificar tumores.

- En el capítulo 1,
- En el capítulo 2,

1.2. Cáncer: definición y factores de riesgo

El cáncer es una enfermedad en la que se produce una división incontrolada de las células [1]. Aunque generalmente se habla del cáncer como una única enfermedad se trata en realidad de un conjunto de enfermedades, existiendo más de 100 tipos distintos de cáncer [2].

El cáncer es una enfermedad genética, esto es, causada por cambios en los genes que controlan las funciones celulares [2]. En general, el proceso de creación del cáncer es complejo y multifactorial: a menudo el causante no es un solo elemento, sino la combinación e interacción de distintos factores ambientales y genéticos [3].

Los factores causantes del cáncer se pueden clasificar principalmente en tres categorías:

1. Factores no modificables. Son elementos que no se pueden cambiar, como la edad o la herencia genética.
2. Factores modificables o prevenibles, como el tabaco, el alcohol, la dieta o la exposición a distintos carcinógenos [4].
3. Otros factores. Algunas circunstancias no se corresponden a ninguna de las categorías anteriores ya que algunos de sus aspectos no se pueden cambiar. Es el caso de factores socioeconómicos (como cobertura sanitaria en el lugar

de residencia o privación económica) y factores reproductivos u hormonales (como toma de anticonceptivos, lactancia materna o terapia hormonal sustitutiva en mujeres menopáusicas).

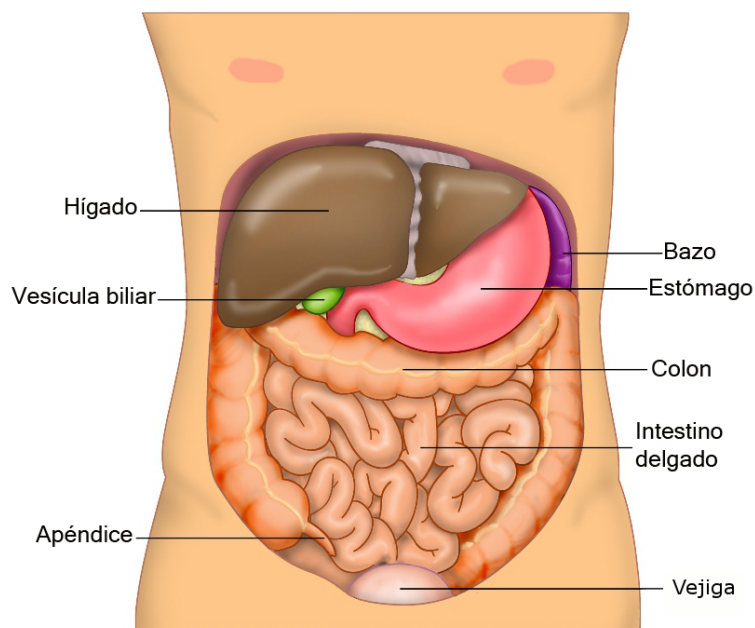
A continuación se introducen dos cánceres con los que se trabajará más adelante: el cáncer de hígado y el cáncer de colon-recto.

1.2.1. Cáncer de hígado

Anatomía y funciones del hígado

El hígado es el órgano interno más grande y pesado del cuerpo humano, está situado en el cuadrante superior derecho del abdomen, debajo de las costillas, y está compuesto principalmente por dos lóbulos. [5].

Figura 1. Anatomía del abdomen humano. Ilustración de Ties van Brussel.



Las funciones del hígado son múltiples y diversas. Las principales son procesar, particionar y metabolizar macronutrientes, regular el volumen de sangre, apoyar al sistema inmune, eliminar sustancias químicas como el alcohol y otras drogas y producir bilis para absorber grasas [6]. Es un órgano imprescindible para la vida.

Factores de riesgo

Uno de los factores de riesgo más comunes del cáncer de hígado es la presencia de cirrosis, o sustitución de células sanas de hígado por tejido cicatrizado. La cirrosis puede producirse por varias causas, siendo las más habituales el consumo excesivo de alcohol y la infección con el virus de la hepatitis B o C [7]. Otros factores de riesgo son el tabaco, la obesidad, padecer diabetes tipo II, consumir esteroides anabólicos [7, 8].

La prevención del cáncer de hígado se basa en reducir la exposición a factores de riesgo como el tabaco y el alcohol, y en vacunarse contra la hepatitis B [7].

1.2.2. Cáncer de colon-recto

Anatomía y funciones del colon-recto

El colon es

Los cánceres de colon, recto y unión rectosigmoidea ¿y ano? a menudo se estudian agrupados por tener características muy similares.

Factores de riesgo

[Prevención]

1.3. Transcriptómica

Daniel: La genómica estudia el genoma como tal (Cromosomas, mutaciones y variaciones tanto de nucleótidos concretos como de regiones del genoma), sin embargo la transcriptómica estudia las transcripciones de los genes, transcripciones que luego son convertidas a proteínas. Tanto RNA-seq como microRNA, se enmarcan en el ámbito de transcriptómica.

1.3.1. Ciencias -ómicas

Principales con descripción + mencionar otras

1.3.2. Secuenciación del genoma?

Human Genome Project, ...
DEGs,

2. Epidemiología del cáncer

Definición epidemiología [9].

2.1. Indicadores epidemiológicos

Para medir en la población el impacto del cáncer se utilizan principalmente cuatro indicadores:

- **Incidencia** (casos nuevos). Mide el riesgo de presentar cáncer.
- **Mortalidad** (defunciones). Mide el riesgo de morir por cáncer.
- **Supervivencia** (porcentaje de casos vivos). Mide la historia natural del cáncer y efectividad del tratamiento.
- **Prevalencia** (casos nuevos y antiguos, vivos). Mide la carga asistencial de la enfermedad.

AÑADIR TENDENCIAS!! DE TODO

2.2. Epidemiología del total del cáncer (¿excepto piel no melanoma?)

Medidas de incidencia

Cómo se obtiene la incidencia. Importancia de RCPoblacionales, estimaciones y proyecciones. Para incidencia, la unidad de análisis es del tumor, no la persona. Número de casos. Tiene problemas por tamaño de población. Tasa bruta. Tiene problemas por estructura de población. Tasas estandarizadas (mundiales, europeas viejas y nuevas). Se usan a veces otros indicadores como tasas acumulativas. Diagrama de Marimekko de incidencia de cáncer. Añadir categoría de Otros
Cómo se obtiene la mortalidad. Importancia de certif de defunción. También estimaciones y proyecciones.

Supervivencia se calcula principalmente a partir de inc, mort y tablas de vida población general

2.3. Epidemiología del cáncer de hígado

2.3.1. Incidencia de cáncer de hígado

2.3.2. Mortalidad por cáncer de hígado

2.3.3. Supervivencia de cáncer de hígado

2.3.4. Prevalencia de cáncer de hígado

2.4. Epidemiología del cáncer de colon-recto

2.4.1. Incidencia de cáncer de hígado

2.4.2. Mortalidad por cáncer de hígado

2.4.3. Supervivencia de cáncer de hígado

2.4.4. Prevalencia de cáncer de hígado

3. *Machine learning* aplicado a transcriptómica

3.1. Selección de biomarcadores

mRMR, RF, DA

3.2. Algoritmos de clasificación

SVM; KNN

4. Detección de biomarcadores en cáncer de páncreas

4.1. Introducción

4.2. Metodología

Fuente de los datos, etc

4.3. Resultados

4.4. Conclusiones

5. Detección de biomarcadores en cáncer de colon-recto

5.1. Introducción

5.2. Metodología

Fuente de los datos, etc

5.3. Resultados

5.4. Conclusiones

6. Aplicación web para realizar detección de biomarcadores

7. Conclusiones y líneas abiertas de trabajo

Bibliografia

- [1] American Cancer Society. What is Cancer?, 2015.
- [2] National Cancer Institute. What is Cancer?, 2015.
- [3] Lucia Migliore and Fabio Coppedè. Genetic and environmental factors in cancer pathogenesis. *Mutation Research*, 512:135–153, 2012.
- [4] V. J. Cogliano, R. Baan, K. Straif, Y. Grosse, B. Lauby-Secretan, F. El Ghissassi, V. Bouvard, L. Benbrahim-Tallaa, N. Guha, C. Freeman, L. Galichet, and C. P. Wild. Preventable Exposures Associated With Human Cancers. *JNCI Journal of the National Cancer Institute*, 103(24):1827–1839, dec 2011.
- [5] Sherif R. Z. Abdel-Misih and Mark Bloomston. Liver Anatomy. *Surgical Clinics of North America*, 90(4):643–653, aug 2010.
- [6] Elijah Trefts, Maureen Gannon, and David H. Wasserman. The liver. *Current Biology*, 27(21):R1147–R1151, nov 2017.
- [7] American Cancer Society. Liver Cancer Risk Factors, 2019.
- [8] Jorge A. Marrero, Robert J. Fontana, Sherry Fu, Hari S. Conjeevaram, Grace L. Su, and Anna S. Lok. Alcohol, tobacco and obesity are synergistic risk factors for hepatocellular carcinoma. *Journal of Hepatology*, 42(2):218–224, feb 2005.
- [9] Isabel dos Santos Silva. *Cancer Epidemiology: Principles and Methods*. International Agency for Research on Cancer, World Health Organization, 1999.

AnexoI: Código de análisis en R

Anexo I: Código de análisis en R

Funciones

Caja XX. Definición de funcion.

```
1 # Ejemplo de comentario
2 parametro <- 24000
3
4 texto <- "texto"
```


Anexo II: Código de aplicación web

Anexo II: Código de aplicación web

Funciones

Caja XX. Definición de funcion.

```
1 # Ejemplo de comentario
2 parametro <- 24000
3
4 texto <- "texto"
```