

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	7
1.2.1. Cáncer de hígado	8
1.2.2. Cáncer de colon-recto	9
1.3. Ciencias -ómicas	11
1.3.1. Secuenciación del genoma?	11
1.3.2. Transcriptómica	11
2. Epidemiología del cáncer	13
2.1. Indicadores epidemiológicos	13
2.2. Epidemiología del cáncer	13
2.2.1. Incidencia de cáncer	14
2.2.2. Mortalidad por cáncer	14
2.2.3. Supervivencia de cáncer	14
2.2.4. Prevalencia de cáncer	14
2.3. Epidemiología del cáncer de hígado	14
2.3.1. Incidencia de cáncer de hígado	14
2.3.2. Mortalidad por cáncer de hígado	14
2.3.3. Supervivencia de cáncer de hígado	14
2.3.4. Prevalencia de cáncer de hígado	14
2.4. Epidemiología del cáncer de colon-recto	14
2.4.1. Incidencia de cáncer de colon-recto	14
2.4.2. Mortalidad por cáncer de colon-recto	14
2.4.3. Supervivencia de cáncer de colon-recto	14
2.4.4. Prevalencia de cáncer de colon-recto	14

3. <i>Machine learning</i> aplicado a transcriptómica	15
3.1. Selección de biomarcadores	15
3.2. Algoritmos de clasificación	15
4. Detección de biomarcadores en cáncer de páncreas	17
4.1. Introducción	17
4.2. Metodología	17
4.3. Resultados	17
4.4. Conclusiones	17
5. Detección de biomarcadores en cáncer de colon-recto	19
5.1. Introducción	19
5.2. Metodología	19
5.3. Resultados	19
5.4. Conclusiones	19
6. Aplicación web para realizar detección de biomarcadores	21
7. Conclusiones y líneas abiertas de trabajo	23
Bibliografía	24
AnexoI: Código de análisis en R	27
Anexo II: Código de aplicación web	29

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

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 [4, 5].
2. Factores modificables o prevenibles, como el tabaco, el alcohol, la dieta o la exposición a distintos carcinógenos [6].
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) [5].

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

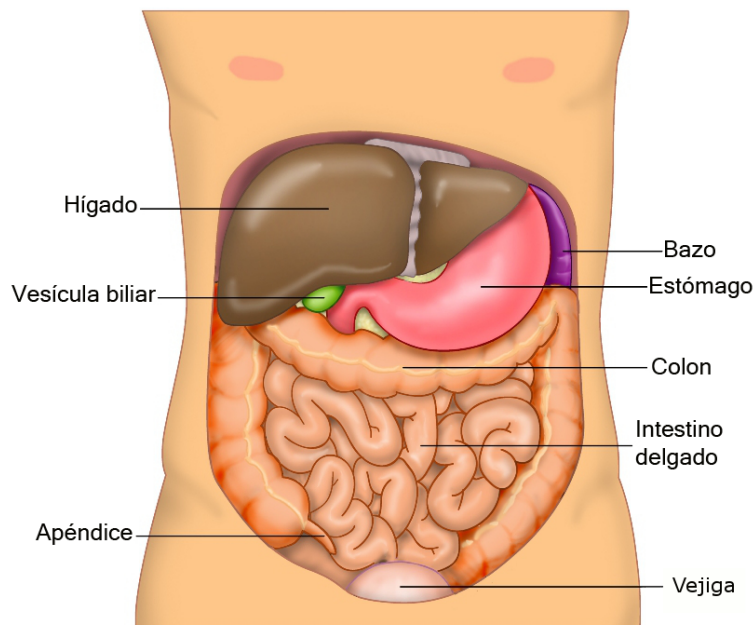
1.2.1. Cáncer de hígado

El cáncer de hígado se corresponde con el código C22 de la Clasificación Internacional de Enfermedades, Décima Revisión, integrando las neoplasias malignas de hígado y vías biliares intrahepáticas [7,8].

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 [9].

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 [10]. 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 [11]. Otros factores de riesgo son el tabaco, la obesidad, padecer diabetes tipo II y consumir esteroides anabólicos [11, 12].

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 [11].

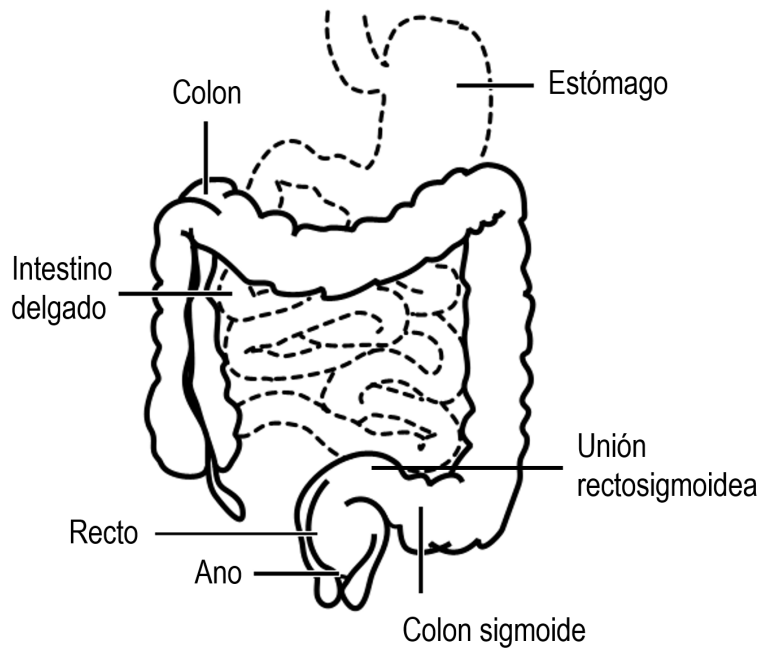
1.2.2. Cáncer de colon-recto

Las neoplasias malignas de colon, recto, unión rectosigmoidea, ano y canal anal (códigos C18-C21 según la Clasificación Internacional de Enfermedades, Décima Revisión [7, 8]) a menudo se estudian agrupadas por tener características muy similares.

Anatomía y funciones del colon-recto

El colon tiene 3 funciones principales: absorción de agua y electrolitos, producción y absorción de vitaminas y movimiento de heces hacia el recto para su eliminación por el ano [13].

Figura 2. Anatomía del intestino humano. Ilustración de Ties van Brussel.



Factores de riesgo

Entre los factores de riesgo del cáncer de colon-recto se puede distinguir entre factores modificables y no modificables.

Entre los factores de riesgo que son modificables destacan el sobrepeso, la inactividad física, las dietas con alto consumo de carnes rojas o procesadas, y el consumo de tabaco y alcohol [14].

Una edad superior a 50 años, padecer diabetes tipo 2 y tener antecedentes personales o familiares de cáncer de colon-recto, pólipos o enfermedad intestinal inflamatoria, como colitis ulcerosa y enfermedad de Crohn, son algunos de los factores de riesgo no modificables [14]. También existen algunos síndromes hereditarios como el síndrome de Lynch que aumentan las posibilidades de padecer cáncer de colon-recto [15].

Para intentar prevenir el cáncer de colon-recto se deben cambiar aquellos factores que son modificables: realizar ejercicio, mantener una dieta saludable y evitar

el consumo de tabaco y alcohol. Además, en los últimos años se están implementando programas de cribado de cáncer de colon-recto para detectar pólipos o diagnosticar el cáncer en etapas iniciales mediante análisis como pruebas de sangre oculta en heces o colonoscopias [16].

1.3. Ciencias -ómicas

Principales con descripción + mencionar otras

1.3.1. Secuenciación del genoma?

Human Genome Project, ...
DEGs,

1.3.2. 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.

2. Epidemiología del cáncer

Definición epidemiología [17].

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

2.2. Epidemiología del cáncer

Poca importancia de 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

2.2.1. Incidencia de cáncer**2.2.2. Mortalidad por cáncer****2.2.3. Supervivencia de cáncer****2.2.4. Prevalencia de cáncer****2.3. Epidemiología del cáncer de hígado****2.3.1. Incidencia de cáncer de hígado**

GLOBOCAN - [\[18,19\]](#)

ECIS - [\[20,21\]](#)

REDECAN - [\[22\]](#)

Población INE - [\[23\]](#)

Defunciones Ministerio - [\[24\]](#)

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 colon-recto****2.4.2. Mortalidad por cáncer de colon-recto****2.4.3. Supervivencia de cáncer de colon-recto****2.4.4. Prevalencia de cáncer de colon-recto**

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

Bibliografía

- [1] American Cancer Society. What is Cancer? Disponible en: <https://www.cancer.org/cancer/cancer-basics/what-is-cancer.html> [Consultado 18/06/2020].
- [2] National Cancer Institute. What is Cancer? Disponible en: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer> [Consultado 18/06/2020].
- [3] Lucia Migliore and Fabio Coppedè. Genetic and environmental factors in cancer pathogenesis. *Mutation Research*, 512:135–153, 2012.
- [4] World Health Organization. *World Cancer Report 2014*. 2014.
- [5] World Health Organization. *World Cancer Report. Cancer research for cancer prevention*. 2020.
- [6] V. J. Cogliano, R. Baan, K. Straif, Y. Grosse, B. Lauby-Secretan, F. El Ghis-sassi, 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, 2011.
- [7] World Health Organization (WHO). *ICD-10: International Statistical Classification of diseases and related health problems: 10th revision*. 1990.
- [8] Ministerio de Sanidad Consumo y Bienestar Social. Edición electrónica de la CIE-10-ES Diagnósticos. Disponible en: https://eciemaps.mscbs.gob.es/ecieMaps/browser/index_10_mc.html [Consultado 21/06/2020].
- [9] Sherif R. Z. Abdel-Misih and Mark Bloomston. Liver Anatomy. *Surgical Clinics of North America*, 90(4):643–653, 2010.
- [10] Elijah Trefth, Maureen Gannon, and David H. Wasserman. The liver. *Current Biology*, 27(21):R1147–R1151, 2017.
- [11] American Cancer Society. Liver Cancer Risk Factors. Disponible en: <https://www.cancer.org/cancer/liver-cancer/causes-risks-prevention/risk-factors.html> [Consultado 18/06/2020], 2019.

-
- [12] 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, 2005.
- [13] Laura L. Azzouz and Sandeep Sharma. *Physiology, Large Intestine*. 2020.
- [14] American Cancer Society. Colorectal Cancer Risk Factors. Disponible en: <https://www.cancer.org/cancer/colon-rectal-cancer/causes-risks-prevention/risk-factors.html> [Consultado 20/06/2020].
- [15] Henry T. Lynch and Albert de la Chapelle. Hereditary Colorectal Cancer. *New England Journal of Medicine*, 348(10):919–932, 2003.
- [16] B. Levin, D. A. Lieberman, B. McFarland, R. A. Smith, D. Brooks, K. S. Andrews, C. Dash, F. M. Giardiello, S. Glick, T. R. Levin, P. Pickhardt, D. K. Rex, A. Thorson, and S. J. Winawer. Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, 2008: A Joint Guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA: A Cancer Journal for Clinicians*, 58(3):130–160, 2008.
- [17] Isabel dos Santos Silva. *Cancer Epidemiology: Principles and Methods*. International Agency for Research on Cancer, World Health Organization, 1999.
- [18] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6):394–424, 2018.
- [19] International Agency for Research on Cancer and World Health Organization. Global Cancer Observatory, Cancer Today. Disponible en: <https://gco.iarc.fr/today/home> [Consultado 21/06/2020].
- [20] European Commission. ECIS - European Cancer Information System. Disponible en: <https://ecis.jrc.ec.europa.eu> [Consultado 21/06/2020].
- [21] J. Ferlay, M. Colombet, I. Soerjomataram, T. Dyba, G. Randi, M. Bettio, A. Gavin, O. Visser, and F. Bray. Cancer incidence and mortality patterns in

- Europe: Estimates for 40 countries and 25 major cancers in 2018. *European Journal of Cancer*, 103:356–387, nov 2018.
- [22] Red Española de Registros de Cáncer (REDECAN). Estimaciones de la incidencia del cáncer en España, 2020. Disponible en: https://funca.cat/redecán/redecán.org/es/Informe_incidenicia_REDECAN_2020.pdf [Consultado 18/06/2020].
- [23] Instituto Nacional de Estadística (INE). Estadísticas de cifras de población. Disponible en: <http://ine.es/> [Consultado 21/06/2020].
- [24] Ministerio de Sanidad Consumo y Bienestar Social. Estadísticas de defunciones según la causa de muerte. Disponible en: <https://pestadistico.inteligenciadegestion.mscbs.es/> [Consultado 21/06/2020].

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"
```