Escola Tècnica Superior d'Enginyeria ETSE-UV

MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA

TRABAJO DE FIN DE MÁSTER

METAANÁLISIS FUNCIONAL DE ESTUDIOS DE CÁNCER COLORRECTAL

AUTORA:
ALICIA GUILLÓ RECUERDA

TUTORES:
FRANCISCO GARCÍA GARCÍA
VICENTE ARNAU LLOMBART

JULIO 2017

VNIVERSITAT DE VALÈNCIA

Escola Tècnica Superior d'Enginyeria ETSE-UV

MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA

TRABAJO DE FIN DE MÁSTER

METAANÁLISIS FUNCIONAL DE ESTUDIOS DE CÁNCER COLORRECTAL

AUTORA:
ALICIA GUILLÓ RECUERDA

TUTORES
FRANCISCO GARCÍA GARCÍA
VICENTE ARNAU LLOMBART

TRIBUNAL:

PRESIDENTE/A:

VOCAL 1:

VOCAL 2:

FECHA DE DEFENSA:

CALIFICACIÓN:

RESUMEN

El cáncer colorrectal (CCR) es considerado uno de los cánceres más frecuentes en todas las regiones del mundo, lo cual implica un importante problema de salud pública, ocupando una de las primeras posiciones entre todos los cánceres en cuanto a incidencia y mortalidad causada por el mismo. Existen múltiples factores que se encuentran implicados en su desarrollo, como los genéticos, ambientales y dietético, sin embargo, aún existe un desconocimiento sobre algunos de los mecanismos de funcionamiento de esta enfermedad, que justifica la necesidad de realizar un trabajo como el presente. Por otro lado, el análisis de los perfiles moleculares ha permitido comprender mejor la patogénesis de la enfermedad e identificar subgrupos de pacientes en los que las terapias actuales tienen más probabilidades de ser eficaces para el desarrollo de nuevos tratamientos.

Este trabajo trata de caracterizar molecularmente el CCR, mediante la aplicación de una metodología de enriquecimiento funcional basada en técnicas de metaanálisis (MA), que proporcionan una síntesis cualitativa y cuantitativa de los resultados de diferentes estudios independientes que tratan de un mismo tema. Para ello se llevó a cabo la revisión sistemática y selección de estudios, análisis primario de cada uno de ellos y MA a nivel funcional del conjunto de los estudios. Se detectaron funciones significativas de interés presentes en los distintos estudios evaluados. En el caso de los *controles*, aquellas funciones y rutas que aparecen principalmente sobrerrepresentadas son aquellas relacionadas con procesos metabólicos, biosíntesis de nucleótidos y aminoácidos, replicación del ADN, señalización y regulación del ciclo celular, mantenimiento enzimático y estructural de la expresión génica, junto con mecanismos de reparación. Mientras que en el caso de *enfermos*, aquellas funciones y rutas que aparecen principalmente sobrerrepresentadas son aquellas relacionadas con pH celular, procesos de señalización y transducción de señales, inflamación y vasodilatación. Estos resultados favorecen y potencian la medicina personalizada, la cual jugará un papel importante en la terapia de las enfermedades, y permitirá identificar a aquellos pacientes que se beneficien de un tratamiento específico según su perfil molecular, y por tanto ajustar las pautas de tratamiento de forma individualizada.

Palabras clave: cáncer colorrectal, metaanálisis, bioinformática, perfil molecular, biomarcadores, genómica funcional

ABSTRACT

Colorectal cancer is considered one of the most frequent cancers in all regions of the world, which implies an important public health problem, occupying one of the first positions among all cancers in incidence and mortality caused by it. There are many factors that are involved in its development, such as genetic, environmental and dietary, however, there is still a lack of knowledge about some of the mechanisms of this disease, which justifies the need to perform a job like this. On the other hand, the analysis of molecular profiles has made it possible to a better understanding of the pathogenesis of the disease in order to identify subgroups of patients in which current therapies are more likely to be effective for the development of new treatments.

This paper tries to characterize molecularly colorectal cancer by applying a methodology of functional enrichment based on meta-analysis techniques, which try to make a qualitative and quantitative synthesis of the results of different independent studies dealing with the same subject. For this reason, the systematic review and selection of studies, primary analysis and meta-analysis at functional level were carried out. Significant functions of interest were detected in the different studies evaluated. In the case of controls, those functions and pathways that appear mainly overrepresented are those related to metabolic processes, nucleotide and amino acid biosynthesis, DNA replication, signaling and regulation of the cell cycle, enzymatic and structural maintenance of gene expression, together with mechanisms of repair. While in the case of patients, those functions and pathways that appear mainly overrepresented are those related to cellular pH, signaling and signal transduction processes, inflammation and vasodilation. These results favor and enhance the personalized medicine, which will play an important role in the therapy of diseases, and will allow to identify those patients who benefit from a specific treatment according to their molecular profile, and therefore to adjust the treatment guidelines in an individualized way.

Keywords: colorectal cancer, meta-analysis, bioinformatic, molecular profile, biomarker, functional genomics

Índice

ÍNDICE

ABREVIATURAS.....**1**

INTRODUCCIÓN..... **3**

1. EPIDEMIOLOGÍA DEL CÁNCER COLORRECTAL..... **4**

1.1. Nivel mundial..... **4**

1.2. Europa..... **7**

1.3. España..... **8**

2. BASES BIOLÓGICAS DEL CÁNCER COLORRECTAL..... **9**

3. ANÁLISIS BIOINFORMÁTICO DE PERFILES DE EXPRESIÓN GÉNICA..... **10**

3.1. La Bioinformática y el descubrimiento de biomarcadores de cáncer..... **10**

4. METAANÁLISIS FUNCIONAL..... **11**

HIPÓTESIS Y OBJETIVOS..... **13**

MATERIAL Y MÉTODOS..... **17**

1. REVISIÓN SISTEMÁTICA Y SELECCIÓN DE ESTUDIOS..... **19**

2. ANÁLISIS PRIMARIO..... **21**

2.1. Procesamiento de los datos..... **21**

2.2. Análisis de expresión diferencial..... **21**

3.3 Análisis de enriquecimiento de grupos de genes..... **22**

3. METAANÁLISIS A NIVEL FUNCIONAL..... **23**

RESULTADOS..... **26**

1. ANÁLISIS PRIMARIO..... **28**

1.1. Procesamiento de los datos..... **28**

1.2. Análisis de la expresión diferencial..... **30**

1.3. Análisis de enriquecimiento de grupos de genes..... **30**

2. METAANÁLISIS A NIVEL FUNCIONAL..... **33**

DISCUSIÓN..... **47**

CONCLUSIONES..... **53**

BIBLIOGRAFÍA..... **57**

ABREVIATURAS

CCR: Cáncer colorrectal
MA: Metaanálisis
CIN: Vía de inestabilidad cromosómica
MSI: Vía de inestabilidad de microsatélites
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
GEO: Gene Expresion Omnibus
PCA: Análisis de Componentes Principales
GO: Gene Ontology
KEGG: Kyoto Encyclopedia of Genes and Genomes
MF: Función molecular
BP: Proceso biológico
CC: Componente celular
LOR: Logaritmo del odds ratio
SE: Error estándar
DL: DerSimonian-Laird
ADN: Ácido desoxirribonucleico
DAG: Diacilglicerol
AMPc: AMP cíclico
GMPc: GMP cíclico
NO: Óxido nítrico
NOS: Sintasa del óxido nítrico
AC: Acetilcolina
IP3: Inositol trifosfato
TNFα: Factor de necrosis tumoral alfa

Introducción

1. Epidemiología del cáncer colorrectal

El CCR se considera uno de los tipos de cánceres más frecuentes en todas las regiones del mundo. Éste supone un grave problema de salud pública en España, a nivel europeo y mundial, ocupando una de las primeras posiciones entre todos los cánceres en cuanto a incidencia y mortalidad causada por el mismo (FERLAY *et al.*, 2013). A continuación se proporcionará la información y datos epidemiológicos de dicho cáncer (incidencia y mortalidad) a nivel mundial, en Europa y en España.

1.1. Nivel mundial

El CCR es considerado uno de los principales problemas de salud, ya que en la actualidad, representa el tercer tumor más frecuente en hombres (después del cáncer de pulmón y el de próstata) y el segundo en mujeres (tras el cáncer de mama) (FERLAY *et al.*, 2013) (Figura 1).

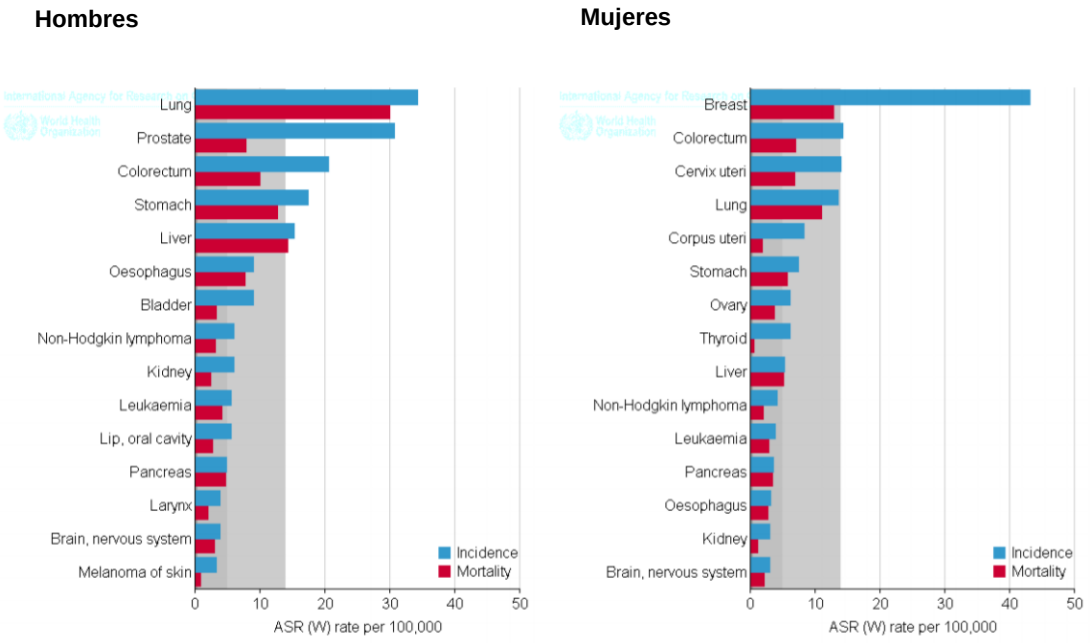

Figura 1. Incidencia y mortalidad del CCR a nivel mundial en hombres y mujeres.

Más de la mitad de los casos de CCR se encuentran registrados en países desarrollados, existiendo una incidencia creciente en la mayoría de áreas geográficas (FERLAY *et al.*, 2013).

Asimismo, se ha comprobado que existe una gran variabilidad en la incidencia del CCR en las distintas regiones del mundo. Por ejemplo, las tasas más altas de incidencia en el mundo se encuentran localizadas en Australia/Nueva Zelanda (44.8 por 100000 en hombres y 33.2 por 100000 en mujeres), mientras que las tasas más bajas se localizan en África (4.5 por 100000 en varones y 3.8 por 100000 en mujeres) (FERLAY *et al.*, 2013) (Figura 2).

No obstante, se observa una variabilidad menor en las tasas de mortalidad entre los distintos países a nivel mundial, siendo Europa Central y Oriental las que cuentan con las tasas estimadas más altas de mortalidad en ambos sexos (20.3 por 100000 en hombres y 11.7 por 100000 en mujeres), y las más bajas en el África Occidental (3.5 y 3.0 respectivamente) (Figura 2).

El CCR representa el tercer tumor más frecuente del mundo en ambos sexos (9.7% del total de tumores detectados), después del cáncer de pulmón y el de mama, teniendo una tasa de incidencia ajustada por edad de 17.2 casos por 100000 personas (Tabla 1).

Haciendo referencia a la mortalidad, el CCR representa el cuarto tumor con mayor mortalidad (8.5% de la mortalidad entre el total de los cánceres) en ambos sexos con una tasa ajustada de mortalidad de 8.4 casos por 100000 personas (Tabla 1).

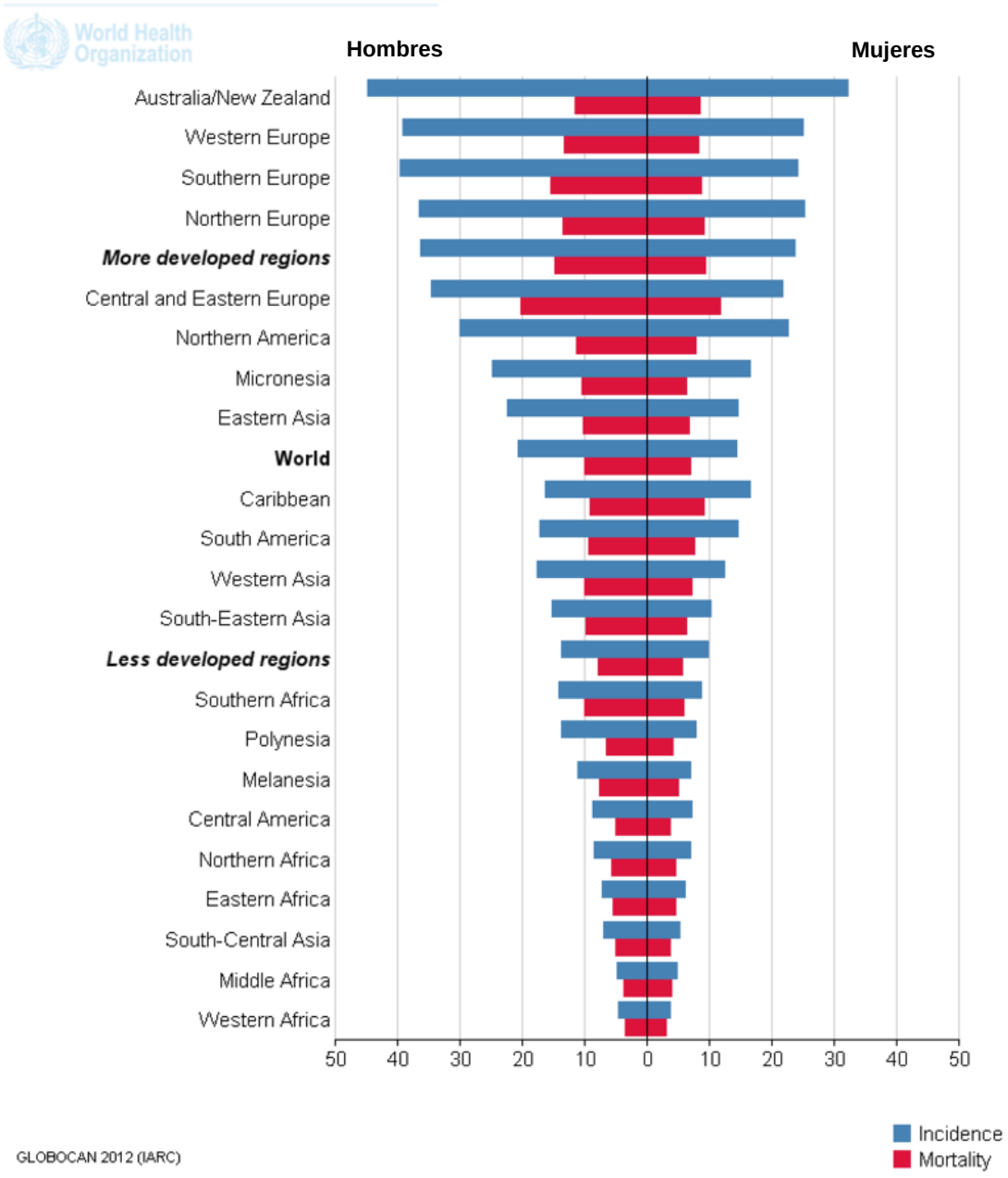

Figura 2. Incidencia y mortalidad en hombres y mujeres del CCR en diferentes regiones mundiales.

Tabla 1. Tasas de incidencia y mortalidad (brutas y ajustadas por edad) de los diferentes tipos de cáncer a nivel mundial (FERLAY *et al*., 2013).

Cáncer	Incidencia			Mortalidad		
	Número	(%)	TA	Número	(%)	TA
Labios, cavidad oral	300373	2,1	4,0	145353	1,8	1,9
Nasofaringe	86691	0,6	1,2	50831	0,6	0,7
Faringe	142387	1,0	1,9	96105	1,2	1,3
Esófago	455784	3,2	5,9	400169	4,9	5,0
Estomago	951594	6,8	12,1	723073	8,8	8,9
Colorrectal	1360602	9,7	17,2	693933	8,5	8,4
Hígado	782451	5,6	10,1	745533	9,1	9,5
Vesícula Biliar	178101	1,3	2,2	142823	1,7	1,7
Páncreas	337872	2,4	4,2	330391	4,0	4,1
Laringe	156877	1,1	2,1	83376	1,0	1,1
Pulmón	1824701	13,0	23,1	1589925	19,4	19,7
Melanoma de piel	232130	1,7	3,0	55488	0,7	0,7
Sarcoma Kaposi	44247	0,3	0,6	26974	0,3	0,3
Mama	1671149	11,9	43,1	521907	6,4	12,9
Cérvix	527624	3,8	14,0	265672	3,2	6,8
Cuerpo útero	319605	2,3	8,3	76160	0,9	1,8
Ovarios	238719	1,7	6,1	151917	1,9	3,8
Próstata	1094916	7,8	30,7	307481	3,7	7,8
Testículos	55266	0,4	1,5	10351	0,1	0,3
Riñón	337860	2,4	4,4	143406	1,7	1,8
Vejiga	429793	3,1	5,3	165084	2,0	1,9
Cerebro, sistema nervioso	256213	1,8	3,4	189382	2,3	2,5
Tiroides	298102	2,1	4,0	39771	0,5	0,5
Linfoma Hodgkin	65950	0,5	0,9	25469	0,3	0,3
Linfoma No-Hodgkin	385741	2,7	5,1	199670	2,4	2,5
Mieloma múltiple	114251	0,8	1,5	80019	1,0	1,0
Leucemia	351965	2,5	4,7	265471	3,2	3,4
Total tumores malignos	14067894	100,0	182,0	8201575	100,0	102,4

%: Porcentaje entre el total de los canceres
TA: Tasas ajustadas por edad (utilizando la población mundial como estándar)

1.2. Europa

El CCR representa el segundo tumor más frecuente en ambos sexos (12.7% del total de tumores detectados), después del cáncer de pulmón y el de mama, teniendo una tasa de incidencia ajustada por edad de 28.2 casos por 100000 personas (Tabla 2).

Haciendo referencia a la mortalidad, el CCR representa el tercer tumor con mayor mortalidad (11.8% de la mortalidad entre el total de los cánceres) en ambos sexos, con una tasa ajustada de mortalidad de 12.3 casos por 100000 personas (Tabla 2).

Tabla 2. Tasas de Incidencia y mortalidad (brutas y ajustadas por edad) de los diferentes tipos de cáncer en Europa (FERLAY *et al*., 2013).

Cáncer	Incidencia			Mortalidad		
	Número	(%)	TA	Número	(%)	TA
Labios, cavidad oral	65933	1,8	4,6	25202	1,3	1,7
Nasofaringe	5307	0,1	0,4	2697	0,1	0,2
Faringe	35800	1,0	2,7	19160	1,0	1,4
Esófago	53457	1,4	3,4	46512	2,4	2,9
Estomago	161846	4,4	10,0	126315	6,5	7,4
Colorrectal	471240	12,7	28,2	228275	11,8	12,3
Hígado	70576	1,9	4,3	69046	3,6	4,0
Vesícula Biliar	31409	0,8	1,7	22352	1,2	1,2
Páncreas	110499	3,0	6,5	111029	5,7	6,4
Laringe	46168	1,2	3,3	22329	1,2	1,5
Pulmón	448618	12,1	28,8	388203	20,1	24,0
Melanoma de piel	104192	2,8	7,7	23508	1,2	1,5
Sarcoma Kaposi	2642	0,1	0,2	479	0,0	0,0
Mama	494076	13,3	66,5	142979	7,4	16,0
Cérvix	67355	1,8	11,2	28003	1,4	3,8
Cuerpo útero	107496	2,9	13,6	25878	1,3	2,6
Ovarios	70320	1,9	9,4	45945	2,4	5,2
Próstata	419915	11,3	58,5	101419	5,2	11,5
Testículos	23560	0,6	5,0	2302	0,1	0,4
Riñón	121629	3,3	8,3	52816	2,7	3,1
Vejiga	166583	4,5	9,7	58758	3,0	2,9
Cerebro, sistema nervioso	66487	1,8	5,5	50744	2,6	3,8
Tiroides	62811	1,7	5,4	7469	0,4	0,4
Linfoma Hodgkin	20410	0,5	2,0	5887	0,3	0,5
Linfoma No-Hodgkin	101940	2,7	7,0	42632	2,2	2,5
Mieloma múltiple	41719	1,1	2,5	26342	1,4	1,4
Leucemia	90391	2,4	6,8	60055	3,1	3,7
Total tumores malignos	3714707	100,0	247,0	1932760	100,0	114,0

%: Porcentaje entre el total de los canceres
TA: Tasas ajustadas por edad (utilizando la población mundial como estándar)

1.3. España

En este caso, el CCR es el tumor más frecuente en ambos sexos (15% del total de tumores incidentes) (Tabla 3). Al ajustar la tasa de incidencia por edad, es el tercer cáncer más incidente después del cáncer de pulmón y el de mama, teniendo una tasa de incidencia ajustada por edad de 33.1 casos por 100000 personas (Tabla 3).

Haciendo referencia a la mortalidad, el CCR representa el segundo tumor con mayor mortalidad (14.3% de la mortalidad entre el total de los cánceres) en ambos sexos, con una tasa ajustada de mortalidad de 12.3 casos por 100000 personas (Tabla 3).

Tabla 3. Tasas de Incidencia y mortalidad (brutas y ajustadas por edad) de los diferentes tipos de cáncer en España (FERLAY *et al.*, 2013).

Cáncer	Incidencia			Mortalidad		
	Número	(%)	TA	Número	(%)	TA
Labios, cavidad oral	4098	1,9	4,7	1117	1,1	1,2
Nasofaringe	350	0,2	0,5	188	0,2	0,2
Faringe	1530	0,7	2,1	765	0,7	1,0
Esófago	2090	1,0	2,5	1728	1,7	1,9
Estomago	7810	3,6	7,8	5389	5,2	4,9
Colorrectal	32240	15,0	33,1	14700	14,3	12,3
Hígado	5522	2,6	6,0	4536	4,4	4,3
Vesícula Biliar	2002	0,9	1,7	1174	1,1	0,9
Páncreas	6367	3,0	6,3	5720	5,6	5,5
Laringe	3182	1,5	4,1	1321	1,3	1,5
Pulmón	26715	12,4	30,3	21118	20,6	22,8
Melanoma de piel	5004	2,3	6,9	967	0,9	1,0
Sarcoma Kaposi	316	0,1	0,5	24	0,0	0,0
Mama	25215	11,7	67,3	6075	5,9	11,9
Cérvix	2511	1,2	7,8	848	0,8	2,1
Cuerpo útero	5121	2,4	11,6	1211	1,2	1,9
Ovarios	3236	1,5	7,7	1878	1,8	3,7
Próstata	27853	12,9	65,2	5481	5,3	8,6
Testículos	823	0,4	3,5	42	0,0	0,1
Riñón	6474	3,0	7,8	2295	2,2	2,2
Vejiga	13789	6,4	13,9	5007	4,9	4,0
Cerebro, sistema nervioso	3717	1,7	5,1	2668	2,6	3,3
Tiroides	2059	1,0	3,4	286	0,3	0,3
Linfoma Hodgkin	1150	0,5	2,3	212	0,2	0,3
Linfoma No-Hodgkin	6130	2,8	7,5	2337	2,3	2,2
Mieloma múltiple	2420	1,1	2,3	1675	1,6	1,4
Leucemia	5190	2,4	6,5	3212	3,1	3,0
Total tumores malignos	215534	100,0	249,1	102762	100,0	98,1

%: Porcentaje entre el total de los canceres
TA: Tasas ajustadas por edad (utilizando la población mundial como estándar).

2. Bases biológicas del cáncer colorrectal

La patogenia del CCR es muy compleja y diversa. Existen múltiples factores que se encuentran implicados en su desarrollo. Estudios epidemiológicos han demostrado la influencia de factores ambientales y dietéticos como el sedentarismo, la obesidad, una dieta con exceso de proteínas o grasas animales o baja en fibra, entre otros, en el desarrollo de esta enfermedad (LE MARCHAND *et al.,* 1997; SLATTERY, 2004).

Por otro lado, el estudio de los síndromes de predisposición hereditaria al CCR ha permitido identificar algunas alteraciones moleculares, genéticas y epigenéticas, también relevantes en la patogénesis del CCR esporádico.

En 1990 Fearon y Vogelstein describieron por primera vez las bases moleculares del desarrollo del CCR como un proceso con múltiples pasos en el que la acumulación de mutaciones llevaría a la transformación de una célula del epitelio colónico normal hasta su conversión en una célula tumoral (FEARON & VOGELSTEIN 1990). Las mutaciones en el gen APC iniciarían la secuencia de transformación adenoma-carcinoma, mientras que otras como p53 ocurrirían de forma más tardía (LYNCH & HOOPS 2002) (Figura 3).

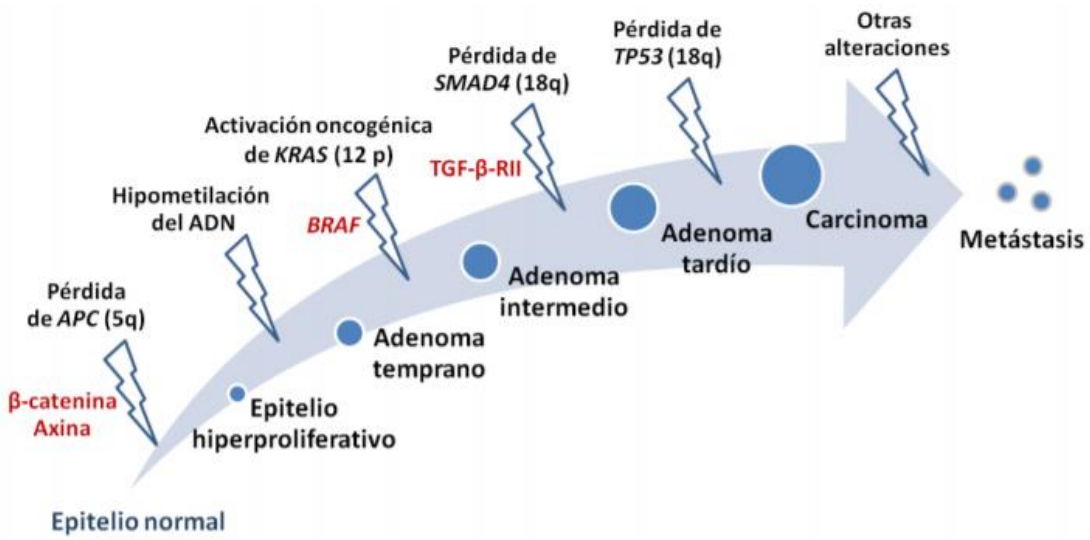

Figura 3. Secuencia adenoma-carcinoma en el CCR. Las alteraciones indicadas en negro se asocian a etapas determinadas de la transformación del tumor colorrectal. En rojo, se indican algunos genes o proteínas que también aparecen alterados con frecuencia en esa etapa de transformación. Adaptado de FEARON & VOGELSTEIN, 1990.

Actualmente se sabe que las alteraciones genéticas en el CCR se pueden producir principalmente mediante tres vías caracterizadas por distintas manifestaciones clínico-patológicas y moleculares (BURREL *et al.,* 2013; COLUSSI *et al.,* 2013): 1) la vía de inestabilidad cromosómica (CIN) o vía supresora, 2) la vía de inestabilidad de microsatélites (MSI) o vía mutadora y 3) la vía de fenotipo metilador o vía serrada.

Sin embargo, todavía existe un desconocimiento sobre algunos de los mecanismos de funcionamiento de esta enfermedad, lo cual justifica la necesidad de realizar un trabajo como el presente.

3. Análisis bioinformático de perfiles de expresión génica

El cáncer es una de las patologías con las mayores tasas de mortalidad en el mundo, con una incidencia aun creciente (PARKIN *et al.*, 2005). Los avances recientes en las estrategias terapéuticas, incluyendo la quimioterapia adyuvante y las terapias biológicas dirigidas, han dado lugar a progresos modestos en la supervivencia de pequeños subgrupos de pacientes, sin embargo, son necesarios nuevos enfoques de tratamiento para ampliar sustancialmente los resultados.

Como ha sucedido para otros tipos tumorales, el análisis de los perfiles moleculares ha permitido comprender mejor la patogénesis de la enfermedad e identificar subgrupos de pacientes en los que las terapias actuales tienen más probabilidades de ser eficaces para el desarrollo de nuevos tratamientos. A través del análisis de los perfiles de expresión génica (BRAMBILLA *et al.*, 2001; KHUDER *et al.*, 2001) se han rastreado muchos biomarcadores de cáncer que han aportado información acerca del diagnóstico, monitoreo, seguimiento y pronóstico de los pacientes. Además, han contribuido a reducir la mortalidad (SPIRA *et al.*, 2007; HIRSCH *et al.*, 2002). Por esto, la importancia de los estudios de perfiles moleculares es identificar los genes cuya expresión se altera por cambios heredables en la función del gen y aquellos en que los cambios son una consecuencia inevitable de la patogénesis de la enfermedad.

La correlación de los resultados derivados de la transcriptómica y de la genómica permitirá realizar análisis más específicos de la gran cantidad de alteraciones genéticas identificadas en los perfiles moleculares.

3.1. La Bioinformática y el descubrimiento de biomarcadores de cáncer

Los perfiles de expresión génica utilizando *microarrays* (ADN y ARN) surgieron como una tecnología eficaz para evaluar tanto el genoma como el transcriptoma de las células tumorales. La utilización de esta tecnología ha generado una cantidad considerable de información relacionada con la identificación de grupos de genes involucrados en la modulación de vías de señalización diferencialmente afectadas en los tejidos tumorales con respecto a los tejidos sanos.

La convergencia de estas tecnologías genómicas permitió el diseño y desarrollo de fármacos contra dianas moleculares específicas, mediante la integración de las herramientas moleculares y los grandes conjuntos de datos generados. Con ello se han logrado rastrear genes que representan los diferentes estados biológicos presentes en la patología, los cuales podrían ser aplicados como potenciales blancos moleculares para su posterior validación experimental.

En general, la búsqueda, verificación, interpretación biológica, bioquímica, fisiológica y la validación de marcadores de enfermedad requiere de la innovación en las tecnologías de alto rendimiento, la Bioestadística y la Bioinformática, además del trabajo interdisciplinario de los clínicos, biólogos, bioquímicos y bioinformáticos para llevar a cabo todos los pasos en el estudio de seguimiento de los biomarcadores, su implementación y control (BAUMGARTNER *et al.*, 2010).

4. Metaanálisis funcional

El MA, definido inicialmente por Glass en 1976, es un conjunto de técnicas estadísticas cuyo objetivo es la realización de una síntesis cualitativa y cuantitativa de los resultados de diferentes estudios independientes que tratan de un mismo tema (HUNTER *et al.*, 1982). Actualmente en los MA se utilizan ensayos clínicos, estudios observacionales, estudios de dosis-respuesta y estudios de evaluación de pruebas diagnósticas.

El MA en general se realiza para dar respuesta a una de las siguientes preguntas:(GISBERT *et al.*, 2004)

- Obtener un estimador promedio ideal a partir de las estimaciones cuantitativas de los estudios individuales comparables que intentan responder a una misma pregunta científica, con el objetivo de aumentar la precisión de la estimación y por tanto la potencia (poder estadístico) en la evaluación de las hipótesis.
- Aclarar incertidumbres cuando las diversas investigaciones disponibles difieren en sus resultados y proporcionar respuesta a cuestiones no abordables desde la perspectiva de estudios aislados, pero que pueden examinarse en el contexto de la comparación de estudios en grupos diversos.

Los criterios más frecuentemente utilizados para seleccionar los estudios en MA son:

- Comparabilidad de exposiciones y variables de evaluación. Un dato a tener en cuenta en el que pueden diferir los estudios es la definición de los factores de exposición, de las intervenciones y de las enfermedades estudiadas.
- Diseño de los estudios.
- Calidad de los estudios. Se suelen excluir aquellos estudios que no alcancen un mínimo de calidad establecido previamente.
- Exhaustividad de la información presentada en el artículo original. Los estudios originales que no presenten datos sobre el diseño o los resultados, se eliminarán.
- Año de publicación. Es necesario establecer un límite en las búsquedas en cuanto a las publicaciones en las que se van a realizar las búsquedas bibliográficas y se van a seleccionar los estudios.
- Idioma de publicación del artículo. Generalmente se incluyen solo artículos en inglés debido a que la mayoría de las publicaciones están escritas por investigadores de Estados Unidos, Canadá y Europa.

Los MA valoran los estudios de acuerdo, entre otras cosas, a su tamaño, concediendo a los más grandes un mayor peso; por este motivo, los resultados globales representan un promedio ponderado de los resultados de los estudios individuales. Por supuesto, el uso de métodos estadísticos, no garantiza que los resultados de una revisión sean válidos, como ocurre también en un estudio primario. Así, como cualquier herramienta, matemática o no, los métodos estadísticos pueden utilizarse de modo inapropiado.

Hipótesis y objetivos

HIPÓTESIS

El CCR es una de las neoplasias con mayor incidencia y mayor tasa de mortalidad en los países desarrollados. En la actualidad los sistemas de clasificación, de pronóstico y de tratamiento de estos pacientes están única y exclusivamente basados en criterios clínicos y anatomopatológicos, los cuales no son capaces de reflejar toda la heterogeneidad de una enfermedad tan compleja. Este hecho provoca que sujetos que se encuentran clasificados en un mismo estadio, evolucionen en la enfermedad o respondan a un mismo tratamiento de manera diferente.

La secuenciación del genoma humano y la aparición de técnicas de análisis masivo, como los *microarrays*, han permitido abordar de manera global y efectiva la complejidad molecular de esta enfermedad, aportando, en algunos casos, como en el de cáncer de mama, herramientas efectivas para el manejo clínico de los pacientes. Sin embargo, existe una gran variabilidad en los resultados obtenidos en distintos estudios de *microarrays* sobre una misma patología. Esto puede ser debido a diferentes factores como el diseño específico de cada estudio y el reducido tamaño muestral utilizado en cada uno de ellos. Teniendo en cuenta estos antecedentes, nos planteamos la realización del presente estudio cuyos objetivos se detallan a continuación:

OBJETIVOS

OBJETIVO GENERAL

La caracterización del CCR a nivel molecular, mediante la aplicación de una metodología de enriquecimiento funcional basada en técnicas de MA.

OBJETIVOS ESPECÍFICOS

- Revisión sistemática y selección de los estudios de CCR con datos de *microarrays* disponibles en el repositorio Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/).
- Desarrollo de un análisis inicial que incluirá el procesamiento de los datos, el análisis de la expresión diferencial y un análisis de enriquecimiento funcional para cada uno de los estudios primarios.
- Integración de todos los resultados funcionales obtenidos del análisis de enriquecimiento funcional mediante técnicas de MA.

Material y métodos

Para llevar a cabo la caracterización del CCR a nivel molecular se propone la siguiente estrategia de análisis, compuesta por los siguientes pasos:

1. Revisión sistemática y selección de estudios.

2. Análisis primario:
    - Procesamiento de los datos.
    - Análisis de la expresión diferencial.
    - Análisis de enriquecimiento de grupos de genes.

3. MA a nivel funcional.

1. Revisión sistemática y selección de estudios

Las revisiones sistemáticas y, en especial, los MA, son un tipo de investigación científica que tiene como propósito integrar de forma objetiva y sistemática los resultados de los estudios empíricos sobre un determinado problema de investigación, con objeto de determinar el "estado del arte" en ese campo de estudio. Por ello es importante tener presente una serie de elementos necesarios para la obtención de indicadores interpretables del efecto de interés:

- Diseño del estudio.
- Criterios de inclusión/exclusión de los estudios.
- Variables de interés a evaluar.
- Búsqueda bibliográfica.
- Selección de los estudios.
- Evaluación de la calidad de los estudios.
- Extracción de la información.
- Análisis estadístico.

El análisis y presentación de resultados se realizaron de acuerdo con la guía *PRISMA* (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, http://www.prisma-statement.org).

Los datos corresponden a una selección de siete estudios de CCR del repositorio de acceso público *GEO*, una base de datos de expresión génica en la que es posible encontrar estudios primarios que incorporen datos ómicos, los cuales utilizan diferentes *arrays* de la plataforma Affymetrix:

- GSE44076
- GSE33113
- GSE47074
- GSE37364
- GSE41258
- GSE32323
- GSE24550

La selección de muestras fue realizada por los investigadores en función de criterios clínicos y biológicos de interés:

- Tumores colorrectales primarios.
- Estadio II-III, que representan tumores no metastásicos.
- Expresión medida con *arrays* de Affymetrix (https://www.thermofisher.com/es/es/home/life-science/microarray-analysis/affymetrix.html#).

De forma que contaban con la siguiente información:

- Listado de todos los estudios disponibles.
- Términos utilizados para la búsqueda en GEO.
- Un primer filtrado utilizando estudios pareados.
- Filtrado definitivo de estudios con tumores en estadios tempranos: no fueron seleccionadas todas las muestras, sino las que cumplían los criterios de selección anteriormente mencionados. Hay estudios que contienen varios tipos de muestras y únicamente han sido seleccionados uno a uno los que interesaba incluir.
- Muestras seleccionadas con su ID correspondiente en la base GEO y su clasificación en tumoral vs mucosa normal.

A continuación se detalla una descripción de cada uno de los estudios:

Tabla 4. Descripción de los estudios. En la columna 4 se indica el filtrado definitivo de estudios.

Estudio	Normalización	Chip	Número de muestras
GSE44076	RMA	[HG-U219] Affymetrix Human Genome U219 Array	196
GSE33113	GAPDH	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	96
GSE47074	GAPDH	[HuGene-2_0-st] Affymetrix Human Gene 2.0 ST Array [transcript (gene) version]	8
GSE37364	-	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	14
GSE41258	MAS5 & RMA	[HG-U133A] Affymetrix Human Genome U133A Array	240
GSE32323	RMA	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	18
GSE24550	RMA	[HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version] & [HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [HuEx-1_0-st v2,coreR3,A20071112,EP.cdf]	44

De los siete estudios seleccionados se han descartado GSE37364 y GSE24550, ya que únicamente incluían muestras correspondientes a tumor, siendo necesario para el MA la existencia de muestras normales y de tumor.

2. Análisis primario

2.1. Procesamiento de los datos

Tras la revisión sistemática y selección de estudios de interés, descargamos los datos normalizados de expresión del repositorio *GEO*. Este procesamiento de los datos incluyó la exploración de las matrices de niveles de expresión, la selección de las muestras indicadas por el investigador, el cambio de los identificadores originales de cada *dataset* (ids de las sondas del chip) a una misma referencia (Ensembl ID) y el promedio de las medidas de expresión a nivel de gen. De esta forma se pretende homogeneizar y facilitar todos aquellos análisis que se realicen posteriormente sobre el conjunto de estudios.

Asimismo, la tecnología de *microarrays* y la preparación experimental pueden introducir ciertos artefactos en la medida de la expresión genética: debidos a la fluorescencia, a la impresión o al experimento biológico; por ello, una vez realizado el procesamiento de los datos, se llevó a cabo un análisis exploratorio de cada uno de los estudios, el cual nos permitió la evaluación de la distribución de la expresión en las distintas muestras, identificando características tales como valores atípicos o *outliers*, concentraciones de valores, dispersión y forma de la distribución; mediante la representación gráfica de diagrama de cajas, análisis de *clustering* y de componentes principales (*PCA*)

Tras la exploración de los datos, se realizaron ciertas correcciones en algunos estudios para completar el control de calidad y preparar los elementos de entrada de la siguiente fase del análisis (eliminación de muestras con un comportamiento anómalo y transformación logarítmica en base 2 para aquellos grupos de datos que no presentaban esta escala).

2.2. Análisis de la expresión diferencial

Este análisis nos permite localizar genes diferencialmente expresados (activados/reprimidos) entre los experimentos, y además ver en qué se diferencian entre sí las diferentes muestras. En los cinco estudios se realizó la comparación de la condición enfermedad frente el control.

Para llevar a cabo la evaluación del nivel de expresión entre los grupos, se utilizó un modelo lineal que se encuentra implementado en el paquete *limma* (Smith 2005) de *Bioconductor*, que nos proporciona la diferencia de expresión de todos los genes evaluados en el experimento.

2.3. Análisis de enriquecimiento de grupos de genes

Tras realizar el análisis anterior, hemos obtenido una lista de todos los genes ordenados según su patrón de expresión diferencial entre la condición enfermedad y control. Para identificar las posibles implicaciones asociadas a estos genes, realizamos una caracterización funcional que nos permite detectar los procesos biológicos en los que los genes se encuentran involucrados. De esta forma, no sólo se mejora la comprensión de la lista de genes obtenida, sino que además, -si muchos de ellos están colaborando en un mismo proceso- se podría reducir la complejidad de los resultados (pasando de decenas, cientos o incluso miles de genes individuales, a sólo unas pocas decenas de funciones o procesos biológicos).

Una vez disponemos de la lista de genes, es necesario investigar los "genes relevantes" por separado. Parece obvio que estos genes interaccionen juntos en procesos biológicos, por lo que, será informativo intentar entender la lista como un conjunto, es decir, analizar el significado biológico, lo cual nos ayudará a responder preguntas tales como si los genes que aparecen en la lista tienen funcionalidades similares o si están involucrados en los mismos procesos, y también, por supuesto, encontrar cuales son estos procesos y como se relacionan con el problema biológico de interés. De esta forma, el segundo *input* que necesitamos para llevar a cabo un análisis de enriquecimiento de grupo de genes, es la información contenida en las bases de datos biológicos. Nos basamos principalmente en dos:

- Gene Ontology (GO): Es una base de datos que contiene anotaciones genéricas (especie independiente) que describen procesos biológicos (*bp*) en los que el gen o producto génico está involucrado, funciones moleculares (*mf*), y la ubicación del producto génico en la célula (componentes celulares) (*cc*) asociados con cada gen. Se organiza en una jerarquía que relaciona todos los términos en refinamientos sucesivos.

- KEGG (Kyoto Encyclopedia of Genes and Genomes): Se trata de una base de datos sobre vías celulares o metabólicas.

A partir de estas bases de datos, obtuvimos la anotación funcional necesaria de los genes, y junto con los genes ordenados por un criterio de interés, se llevó a cabo un método de enriquecimiento funcional de grupos de genes basado en modelos de regresión logística (MONTANER *et al.,* 2009; SARTOR *et al.*, 2009; MONTANER & DOPAZO, 2010).

3. Metaanálisis a nivel funcional

A partir del análisis de enriquecimiento funcional, hemos obtenido una serie de resultados que forman parte del experimento evaluado. Sin embargo, el reducido tamaño muestral de la mayor parte de estos experimentos junto con su carácter individual, representan factores limitantes al evaluar dichos estudios.

Por lo tanto, para mejorar la integración de diversos experimentos en el contexto funcional y proporcionar claridad en la interpretación de los resultados, vamos a realizar un MA para detectar resultados funcionales de interés global, reduciendo así el efecto de los experimentos específicos.

Esta metodología también comprueba la consistencia de dichos experimentos y genera una estimación del efecto teniendo una mayor potencia estadística que el obtenido por cada uno de los experimentos por separado.

Realizamos varios MA funcionales para cada anotación (GO y KEGG). Con respecto a la base de datos GO, tuvimos en cuenta las anotaciones *bp*, *mf* y *cc* por separado, al igual que llevamos a cabo en el apartado anterior de enriquecimiento funcional. En este caso es importante hacer referencia a los métodos de MA existentes: *Modelo de efectos fijos* y *modelo de efectos aleatorios*. El modelo de efectos fijos asume la existencia de un único efecto en la población y no tiene en cuenta la variabilidad de los resultados entre los estudios primarios, mientras que el modelo de efectos aleatorios incluye la posible heterogeneidad de los efectos entre los distintos estudios. En este último caso, la ponderación en la determinación de un efecto combinado incorpora tanto la variabilidad entre-estudios como la variabilidad intra-estudios. En el contexto genómico, la heterogeneidad entre estudios suele ser bastante habitual debido al empleo de diferentes plataformas de tecnologías de alto rendimiento, tamaños muestrales y tipos de contrastes entre grupos experimentales. Por lo tanto, un modelo de efectos aleatorios se ajustaría mejor a las características de estos estudios. En el MA de cada término funcional se utilizaron las funciones incluidas en el paquete *metafor* (Viechtbauer 2010) de R, que incorpora la implementación del modelo de efectos fijos (*FE*) y distintos modelos de efectos aleatorios:

- DL, DerSimonian & Laird (1986): Es el método más ampliamente utilizado.
- HE, Hedges *et al.* (2008).
- HS, Schmidt & Hunter (2014).

Se utilizó un conjunto de representaciones gráficas para resumir los resultados del MA: gráficos de árbol, embudo y radial.

Para cada término funcional se realizó un MA para combinar las medidas de los efectos de todos los estudios (logaritmo *odds ratios* entre casos y controles). La siguiente tabla (Tabla 5) es un ejemplo en el que podemos observar varios indicadores para evaluar los resultados del MA para una función específica. Para el resto de funciones, la interpretación es análoga:

Tabla 5. Ejemplo de indicadores para la evaluación de los resultados del MA para una función específica.

ID	name	LB	LOR	UB	pvalue
hsa00010	L-valine transaminase activity	0.51	0.939	1.367	0

ID	p.adjust	QE	QEp	SE	tau2	I2	H2
hsa00010	0.018	6.496	0.483	0.218	0	0	1

Para una mejor comprensión de los procedimientos utilizados, revisamos cada uno de los indicadores obtenidos en el MA.

IDENTIFICACIÓN Y DESCRIPCCIÓN DEL TÉRMINO FUNCIONAL

- *ID*: Identificador de cada término funcional.
- *name*: Descripción de cada término funcional.

EVALUACIÓN DE LA HETEROGENEIDAD

- *QE y QEp*: Representan el estadístico de contraste y el valor *p* respectivamente de la prueba de DL (DERSIMONIAN & LAIRD, 1986), empleadas para detectar la presencia de heterogeneidad entre los estudios. La hipótesis nula apunta a la presencia de homogeneidad entre los estudios.

MEDIDA DEL EFECTO ENTRE CASOS Y CONTROLES

- *LOR:* Logaritmo del odds ratio. Es la estimación del efecto combinado de todos los estudios. La magnitud de este indicador cuantifica la sobrerrepresentación de la función en un grupo frente el otro. Por ejemplo, si tiene signo negativo, nos está indicando mayor presencia de genes con nivel alto de expresión en la segunda clase experimental respecto la primera clase, en la comparación valorada (*enfermo* frente a *control*); mientras que si tiene signo positivo, nos está indicando mayor presencia de genes con nivel alto de expresión en la primera clase experimental respecto la primera clase, en la comparación valorada.

- *LB y UB*: La estimación del efecto estudiado viene acompañado de su *intervalo de confianza al 95%* construido con la variabilidad estimada en el modelo seleccionado. La no inclusión del valor 0 en el intervalo, confirmaría la significatividad del *LOR*

- *pvalue*: Informa del nivel de significación de un efecto combinado nulo. *pvalues* < 0.05 indican que *LOR* no es 0 (efectos significativos entre casos y controles).

- *p.adjust*: p-valor ajustado obtenido a partir del *pvalue* el cual ha sido corregido para reducir el número de falsos positivos.

- *SE*: Error estándar para *LOR* (variabilidad estimada en el modelo).

ESTIMACIÓN DE LA HETEROGENEIDAD

- *tau2:* Es la estimación de la heterogeneidad entre los estudios. Será 0 cuando utilicemos un modelo de efectos fijos.

- *I2*: Estima (en porcentaje) la relación entre la variabilidad entre estudios y el total de la variabilidad.

- *H2*: Es el cociente entre la variabilidad total y la variabilidad en el muestreo. *I2* y *H2* son medidas para facilitar la interpretación de la heterogeneidad estimada (HIGGINS & THOMPSON, 2002). Sin embargo, es importante destacar que *tau2*, *I2*, y *H2* son estimados frecuentemente de forma imprecisa, especialmente cuando el número de estudios es reducido.

El análisis exploratorio, la expresión diferencial, el enriquecimiento funcional de cada estudio y el MA funcional fueron realizados utilizando el software de R (R CORE TEAM, 2016).

Resultados

1. Análisis primario

1.1. Procesamiento de los datos

Una vez realizado el procesamiento de los datos, se llevó a cabo un análisis exploratorio, tal y como se menciona en el apartado anterior. En este caso se van a mostrar los resultados de diagrama de cajas, *PCA* y *clustering* referentes a uno de los estudios, ya que los cuatro restantes muestran un comportamiento análogo.

Se examinó la distribución global de las muestras del estudio GSE32323 mediante un diagrama de cajas, el cual nos informa si la normalización se ha realizado correctamente. Esto será así, cuando todas las cajas tengan la misma distribución, tal como observamos en la figura 4.

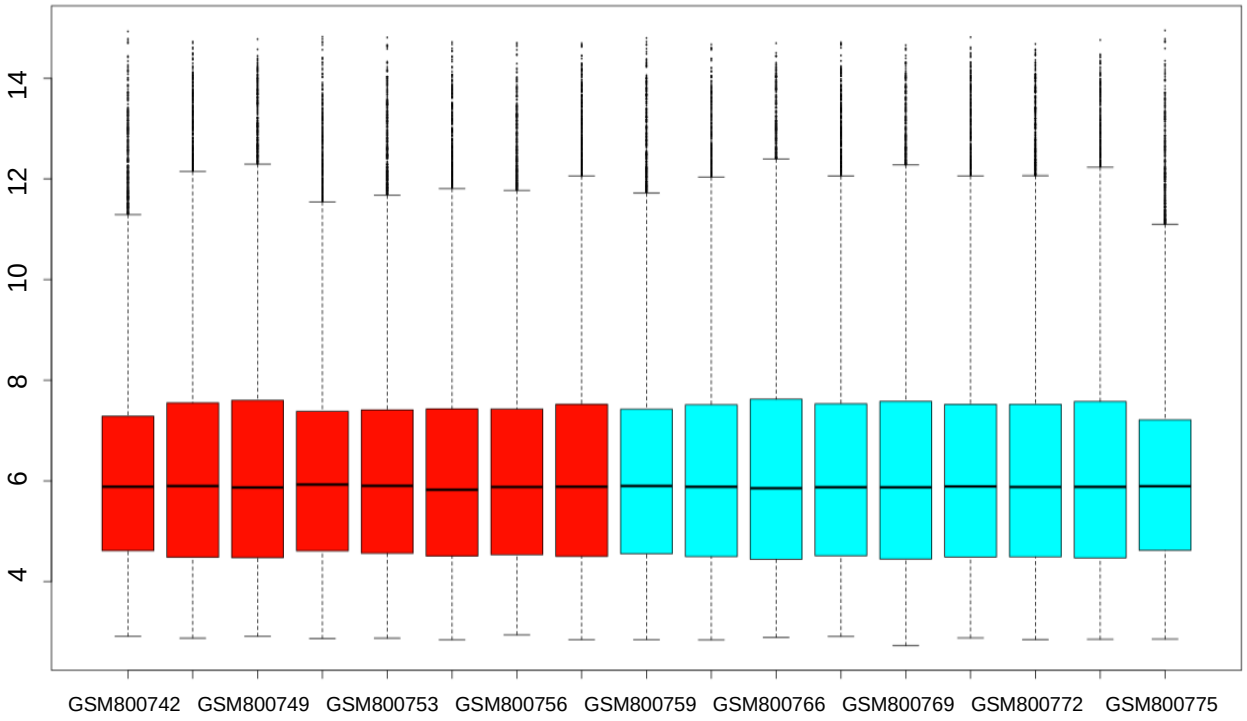

Figura 4. Diagrama de cajas de resultados del análisis exploratorio del estudio GSE32323. En el eje de las X se muestran los valores de expresión en base 2. En el eje Y aparecen las muestras con las que trabajaremos, sin embargo son 17 y no 18 (tal y como aparece en la tabla 4) ya que eliminamos una muestra que presentaba un comportamiento anómalo.

Las relaciones entre las muestras del estudio fueron evaluadas mediante *PCA*. Ésta es una técnica estadística de síntesis de la información o reducción de la dimensión. Su exploración permitió detectar grupos de tumores que mantienen un comportamiento similar, al igual que ocurre con los grupos de muestras normales. Nos centraremos en el primero de los dos gráficos (Figura 5): observamos que hay una clara diferenciación entre los tumores y las muestras normales. Esto nos avanza que encontraremos una clara diferencia de expresión entre los grupos experimentales.

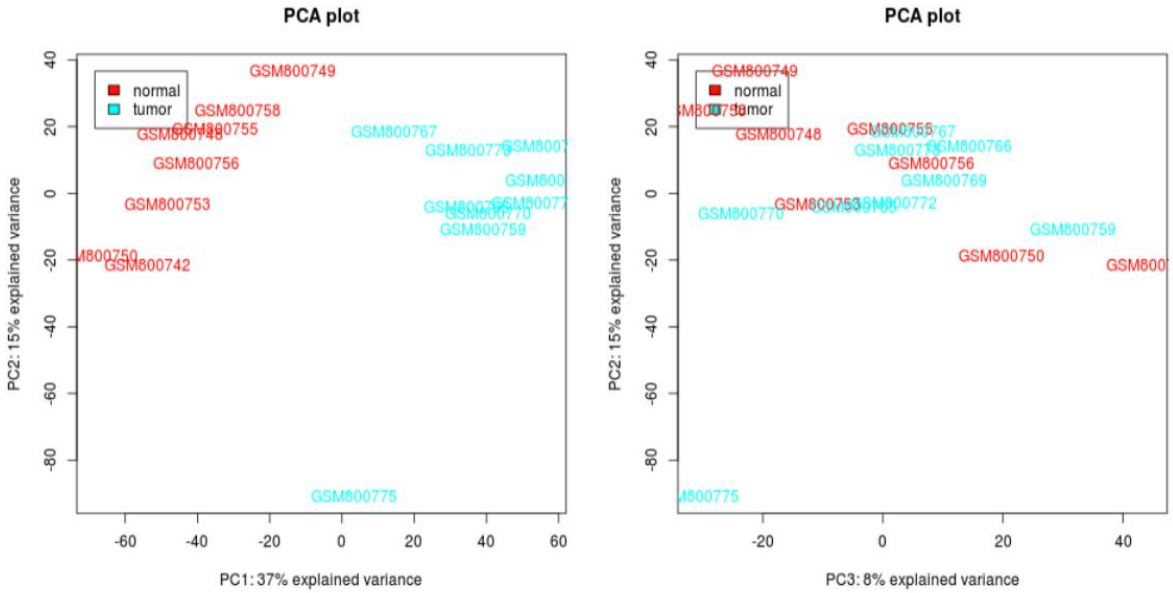

Figura 5. PCA de los resultados del análisis exploratorio del estudio GSE32323. Tanto el eje X como el Y, son los "loadings" del PCA. Estos son las puntuaciones que tienen cada una de las muestras en términos de componentes principales. Informan de la relación existente entre las variables.

Para identificar grupos de muestras con un patrón de comportamiento similar se aplicó un procedimiento de *clustering*, utilizando para ello la medida de distancia (o semejanza) de correlación, la cual tiende a agrupar las muestras de acuerdo a la tendencia de las mismas. En la figura 6 también apreciamos el mismo comportamiento que en el *PCA*: los grupos de muestras casos y control están agrupados por separado y ello también confirma que habrá un gran número de genes diferencialmente expresados.

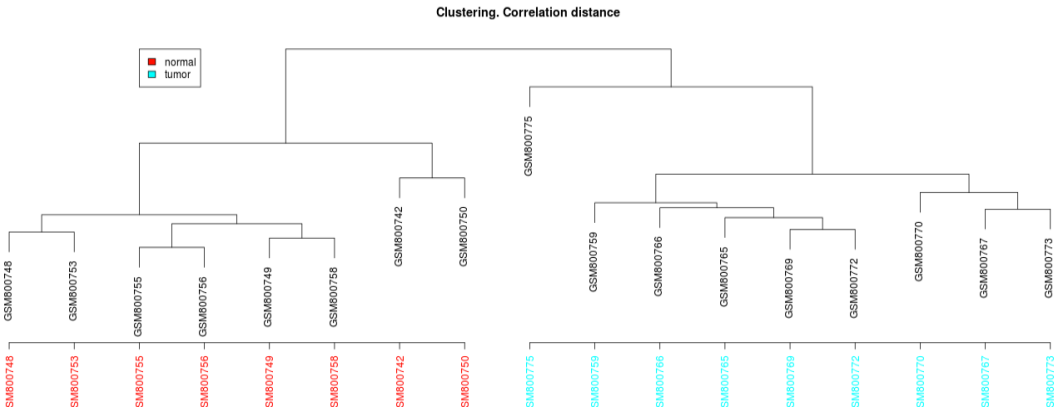

Figura 6. Análisis de *clustering* de resultados del análisis exploratorio del estudio GSE32323.

1.2. Análisis de expresión diferencial

Para tener una visión general de los resultados correspondientes al análisis de expresión diferencial, vamos a mostrar una tabla que incluye los cinco estudios y el número de genes que se encuentran *sobre-expresados*, *infra-expresados* y *no expresados* en cada uno de ellos.

Tabla 6. Número de genes *sobre-expresados*, *infra-expresados* y *no expresados* en cada uno de los estudios.

Estudio	***Genes sobre-expresados***	***Genes infra-expresados***	***Genes no expresados***
GSE32323	2596	2764	17459
GSE33113	8980	2324	11515
GSE41258	3819	4782	4610
GSE44076	6647	7075	5668
GSE47074	6	5	34622

1.3. Análisis de enriquecimiento funcional de grupos de genes

En este caso se evaluaron las tres ontologías de la GO referidas en el análisis de enriquecimiento funcional: *bp*, *mf*, y *cc*, junto con la base de datos de rutas KEGG. A continuación se muestran tres tablas correspondientes a cada una de las ontologías, y una correspondiente a KEGG. En cada una de ellas se observa, para cada uno de los estudios, los siguientes elementos (en la comparación valorada: *enfermo* frente a *control*).

- **Funciones significativas** (*p-valor* < 0.05).

- **Funciones significativas y sobrerrepresentadas en controles** (p-valor < 0.05 y *LOR* < 0). Esto nos indica que una determinada función se encuentra presente en un grupo de genes que tienen un nivel de expresión alto en controles.

- **Funciones significativas y sobrerrepresentadas en enfermos** (p-valor < 0.05 y *LOR* > 0). Esto nos indica que una determinada función se encuentra presente en un grupo de genes que tienen un nivel de expresión alto en enfermos.

PROCESOS BIOLÓGICOS (*bp*)

Tabla 7. Número de funciones significativas (*p-valor* < 0.05), en ambos grupos experimentales, significativas y sobrerrepresentadas en controles (*p-valor* < 0.05 y *LOR* < 0) y significativas y sobrerrepresentadas en enfermos (*p-valor* < 0.05 y *LOR* > 0) de cada uno de los estudios tras realizar el análisis de enriquecimiento. Ontología *bp*.

Estudio	**Funciones significativas (*p-valor* < 0.05)**	**Funciones significativas sobrerrepresentadas (*p-valor* < 0.05 *LOR* < 0)**	**Funciones significativas sobrerrepresentadas (*p-valor* < 0.05 *LOR* > 0)**
GSE32323	1202	525	677
GSE33113	902	788	114
GSE41258	592	310	282
GSE44076	964	529	435
GSE47074	995	690	305

FUNCIONES MOLECULARES (*mf*)

Tabla 8. Número de funciones significativas (*p-valor* < 0.05), en ambos grupos experimentales, significativas y sobrerrepresentadas en controles (*p-valor* < 0.05 y *LOR* < 0) y significativas y sobrerrepresentadas en enfermos (*p-valor* < 0.05 y *LOR* > 0) de cada uno de los estudios tras realizar el análisis de enriquecimiento. Ontología *mf*.

Estudio	**Funciones significativas (*p-valor* < 0.05)**	**Funciones significativas sobrerrepresentadas (*p-valor* < 0.05 *LOR* < 0)**	**Funciones significativas sobrerrepresentadas (*p-valor* < 0.05 *LOR* > 0)**
GSE32323	215	111	104
GSE33113	190	136	54
GSE41258	98	56	42
GSE44076	214	121	93
GSE47074	150	93	57

COMPONENTES CELULARES (*cc*)

Tabla 9. Número de funciones significativas (*p-valor* < 0.05), en ambos grupos experimentales, significativas y sobrerrepresentadas en controles (*p-valor* < 0.05 y *LOR* < 0) y significativas y sobrerrepresentadas en enfermos (*p-valor* < 0.05 y *LOR* > 0) de cada uno de los estudios tras realizar el análisis de enriquecimiento. Ontología *cc*.

Estudio	Funciones significativas (*p-valor* < 0.05)	Funciones significativas sobrerrepresentadas (*p-valor* < 0.05 *LOR* < 0)	Funciones significativas sobrerrepresentadas (*p-valor* < 0.05 *LOR* > 0)
GSE32323	343	187	156
GSE33113	250	216	34
GSE41258	164	95	69
GSE44076	287	182	105
GSE47074	228	173	55

KEGG

Tabla 10. Número de funciones significativas (*p-valor* < 0.05), en ambos grupos experimentales, significativas y sobrerrepresentadas en controles (*p-valor* < 0.05 y *LOR* < 0) y significativas y sobrerrepresentadas en enfermos (*p-valor* < 0.05 y *LOR* > 0) de cada uno de los estudios tras realizar el análisis de enriquecimiento. KEGG.

Estudio	Funciones significativas (*p-valor* < 0.05)	Funciones significativas sobrerrepresentadas (*p-valor* < 0.05 *LOR* < 0)	Funciones significativas sobrerrepresentadas (*p-valor* < 0.05 *LOR* > 0)
GSE32323	28	4	24
GSE33113	197	10	187
GSE41258	20	6	14
GSE44076	117	11	106
GSE47074	212	13	199

2. Metaanálisis a nivel funcional

Los resultados del MA de las funciones incluidas para cada base de datos seleccionada (GO y KEGG), se han organizado en dos niveles:

- Resultados globales del MA para el conjunto de funciones incluidas en la base de datos seleccionada.
- Resultados específicos del MA de cada una de las funciones.

RESULTADOS GLOBALES

Nos proporcionan indicadores y representaciones gráficas para conocer cómo es la medida combinada del efecto en el conjunto de las funciones de las bases de datos de interés.

Todos los modelos presentados son de efectos aleatorios (*DL*, *HE*, *HS*) excepto *FE* que corresponde a un modelo de efectos fijos. En esta tabla, observaremos los términos GO o rutas sobrerrepresentadas en grupos de genes que tienen un nivel de expresión alto en el grupo de los enfermos y de los controles (*E* y *C* respectivamente), además esta estimación incluye las funciones que se consideran significativas en el grupo de enfermos y controles (columnas *Sig.E* y *Sig.C* respectivamente). Las columnas 6 y 7 mostrarán el número de términos GO o rutas KEGG significativas y que además presentan una magnitud del *LOR* > 1 para cada grupo experimental (*Sig.LOR.E y Sig.LOR.C*).

Es importante entender que, cuando indicamos que existe sobrerrepresentación de una determinada función, por ejemplo en los controles, queremos decir que estamos detectando un conjunto de genes que participan en un determinado proceso o ruta de señalización y que además tienen un patrón de expresión alto en el grupo control, o lo que sería lo mismo, un conjunto de genes con un patrón de expresión bajo en el grupo enfermo. Esta información referente a la alteración de una vía asociada a un grupo experimental de interés puede tener gran importancia en distintas situaciones, como por ejemplo la planificación de estudios posteriores, la detección de dianas terapéuticas o la confirmación de hipótesis iniciales.

RESULTADOS ESPECÍFICOS

Tras conocer un escenario global de resultados, la metodología propuesta proporciona un grupo de resultados específicos de cada función que permite llevar a cabo una revisión exhaustiva del peso de cada estudio, la presencia de variabilidad y sesgos, así como la interpretación de la información obtenida.

En esta tabla se muestra el nivel de información de los resultados. Se presentan los estimadores de la medida del efecto y los indicadores de heterogeneidad en el MA con el estimador *DL*, para aquellas funciones que presentan una mayor significación y magnitud del efecto.

La disposición de distintos niveles de información de los resultados del MA contribuye a la revisión desde lo más general a lo más específico, facilitando la interpretación de los resultados.

PROCESOS BIOLÓGICOS *(bp)*

Tabla 11. Resultados globales del MA funcional con *bp*. Modelos de efectos aleatorios y fijos para la estimación de la variabilidad del efecto medido.

Métodos	***E***	***F***	***Sig.E***	***Sig.C***	***Sig.LOR.E***	***Sig.LOR.C***
DL	4510	3406	1698	1496	16	160
HE	4509	3407	1718	1505	16	161
HS	4508	3410	1821	1558	16	160
FE	4509	3409	2192	1703	20	161

En la tabla 12 se muestran 5 de las 176 funciones significativas de *bp* con mayor magnitud de sobrerrepresentación en los grupos experimentales evaluados. La exploración del conjunto de procesos significativos proporciona una interpretación de los resultados de los estudios transcriptómicos considerados.

Tabla 12. Estimadores de la medida del efecto e indicadores de heterogeneidad en el MA funcional con *bp*.

ID	**Nombre**
GO:0010649	Regulation of cell communication by electrical coupling
GO:0019934	cGMP-mediated signaling
GO:0023041	Neuronal signal transduction
GO:0086023	Adrenergic receptor signaling pathway involved in heart process
GO:1901844	Regulation of cell communication by electrical coupling involved in cardiac conduction

ID	***LB***	***LOR***	***UB***
GO:0010649	0.851	1.211	1.570
GO:0019934	0.847	1.213	1.579
GO:0023041	0.799	1.289	1.779
GO:0086023	1.251	1.539	1.827
GO:1901844	0.760	1.203	1.646

ID	*pvalue*	*p.adjust*	*QE*	*QEp*	*SE*	*tau2*	*I2*	*H2*
GO:0010649	0	0	10.326	0.035	0.183	0.102	61.262	2.581
GO:0019934	0	0	8.634	0.071	0.187	0.091	53.672	2.159
GO:0023041	0	0	5.709	0.127	0.250	0.117	47.452	1.903
GO:0086023	0	0	4.943	0.293	0.147	0.021	19.083	1.236
GO:1901844	0	0	11.852	0.018	0.226	0.167	66.249	2.963

FUNCIONES MOLECULARES (*mf*)

Tabla 13. Resultados globales del MA funcional con *mf*. Modelos de efectos aleatorios y fijos para la estimación de la variabilidad del efecto medido.

Métodos	*E*	*F*	*Sig.E*	*Sig.C*	*Sig.LOR.E*	*Sig.LOR.C*
DL	896	752	352	343	11	58
HE	897	751	355	343	11	60
HS	897	751	371	346	11	60
FE	900	748	444	377	12	61

En la tabla 14 se muestran 5 de las 69 funciones significativas de *mf* con mayor magnitud de sobrerrepresentación en los grupos experimentales evaluados. La exploración del conjunto de funciones significativas proporciona una interpretación de los resultados de los estudios transcriptómicos considerados.

Tabla 14. Estimadores de la medida del efecto e indicadores de heterogeneidad en el MA funcional con *mf*.

ID	Nombre
GO:0004022	Alcohol dehydrogenase (NAD) activity
GO:0004957	Prostaglandin E receptor activity
GO:0008048	Calcium sensitive guanylate cyclase activator activity
GO:0008131	Primary amine oxidase activity
GO:0008179	Adenylate cyclase binding

ID	*LB*	*LOR*	*UB*
GO:0004022	0.625	1.258	1.890
GO:0004957	0.803	1.153	1.503
GO:0008048	0.602	1.100	1.598
GO:0008131	0.813	1.135	1.456
GO:0008179	0.904	1.212	1.519

ID	*pvalue*	*p.adjust*	*QE*	*QEp*	*SE*	*tau2*	*I2*	*H2*
GO:0004022	0	0	17.634	0.001	0.323	0.400	77.316	4.408
GO:0004957	0	0	4.311	0.366	0.178	0.012	7.207	1.078
GO:0008048	0	0	0.519	0.471	0.254	0.000	0.000	1.000
GO:0008131	0	0	3.596	0.463	0.164	0.000	0.000	1.000
GO:0008179	0	0	6.973	0.137	0.157	0.052	42.633	1.743

COMPONENTES CELULARES (*cc*)

Tabla 15. Resultados globales del MA funcional con *cc*. Modelos de efectos aleatorios y fijos para la estimación de la variabilidad del efecto medido.

Métodos	*E*	*F*	*Sig.E*	*Sig.C*	*Sig.LOR.E*	*Sig.LOR.C*
DL	453	498	183	318	1	79
HE	454	497	185	319	1	79
HS	453	498	196	323	0	79
FE	450	503	263	343	1	81

En la tabla 16 se muestra 1 de las 80 funciones significativas de *cc* con mayor magnitud de sobrerrepresentación en los grupos experimentales evaluados. La exploración del conjunto de componentes significativos proporciona una interpretación de los resultados de los estudios transcriptómicos considerados.

Tabla 16. Estimadores de la medida del efecto e indicadores de heterogeneidad en el MA funcional con *cc*.

ID	Nombre
GO:0042583	Chromaffin granule

ID	LB	LOR	UB
GO:0042583	0.632	1.001	1.371

ID	pvalue	p.adjust	QE	QEp	SE	tau2	I2	H2
GO:0042583	0	0	5.551	0.235	0.189	0.05	27.935	1.388

KEGG

Tabla 17. Resultados globales del MA funcional con rutas de señalización KEGG. Modelos de efectos aleatorios y fijos para la estimación de la variabilidad del efecto medido.

Métodos	*E*	*F*	*Sig.E*	*Sig.C*	*Sig.LOR.E*	*Sig.LOR.C*
DL	262	40	51	14	1	3
HE	262	40	59	14	1	3
HS	262	40	158	16	1	3
FE	265	37	236	19	1	3

En la tabla 18 se muestra 1 de las 4 rutas significativas de *KEGG* con mayor magnitud de sobrerrepresentación en los grupos experimentales evaluados. La exploración del conjunto de las rutas significativas proporciona una interpretación de los resultados de los estudios transcriptómicos considerados.

Tabla 18. Estimadores de la medida del efecto e indicadores de heterogeneidad en el MA funcional con rutas de señalización *KEGG*.

ID	Nombre
hsa00472	D-Arginine and D-ornithine metabolism

ID	LB	LOR	UB
hsa00472	0.497	1.002	1.196

ID	pvalue	p.adjust	QE	QEp	SE	tau2	I2	H2
hsa00472	0	0	1.105	0.893	0.178	0	0	1

Para completar la comprensión de los resultados obtenidos, a continuación se detalla la información generada en el MA para dos funciones significativas (*p-valor* < 0.05) y que además presentan una magnitud del *LOR* > 1 para cada grupo experimental, es decir, una función sobrerrepresentada en enfermos y otra función sobrerrepresentada en controles, utilizando para ello distintos gráficos:

- ***Gráficos de bosque:*** Para visualizar el efecto aportado de cada estudio en la estimación del efecto global en una función determinada, utilizamos los *gráficos de bosque*. Este gráfico representa un bosque donde los árboles serían los estudios primarios del MA y donde se resumen todos los resultados relevantes de la síntesis cuantitativa. Elementos de interés en la representación gráfica:

    - A la izquierda del gráfico se enumeran los estudios incluidos en el MA.
    - En la parte derecha se incluye la estimación de la medida resumen individual de cada estudio y su intervalo de confianza (95 %).

- En el centro del gráfico se visualiza la medida del efecto (cuadrado en color negro) cuyo tamaño es proporcional a la precisión de las estimaciones, de modo que una mayor variabilidad se visualizaría con una figura de menor tamaño. El cuadrado está ubicado dentro de un segmento que representa los extremos de su intervalo de confianza.
- En la parte inferior, el resultado global del MA se representa con un rombo en color rojo. Su posición respecto a la línea de efecto nulo nos informa sobre la significación estadística del resultado global, mientras que su anchura nos proporciona una idea de su precisión (su intervalo de confianza).
- También en la parte inferior derecha de esta figura se indica el valor de significación de los intervalos de confianza (habitualmente 95 %) y a la izquierda el modelo de análisis de datos que se ha utilizado. En este caso se indica que el modelo incluyó el estimador de DL (modelo de efectos aleatorios).

- ***Gráficos de embudo***: Mediante los *gráficos de embudo* se evalúa la variabilidad de los distintos estudios, así como la presencia de sesgos. En este tipo de representaciones se muestra la magnitud del efecto medido (eje X) frente a una medida de precisión (eje Y), como la desviación estándar o el inverso de la varianza. Cada punto representa un estudio primario y el diagnóstico del gráfico se realiza tras la valoración de la nube de puntos (STERNE & EGGER, 2001).

  En ausencia de sesgos y heterogeneidad, se espera que los puntos se distribuyan en forma de embudo, donde la mayoría de los puntos caerían dentro de la región de confianza de la estimación del efecto.

- ***Gráficos radial:*** Los *gráficos radial* (GALBRAITH, 1988b; GALBRAITH, 1988a; GALBRAITH, 1994) se utilizan para valorar la consistencia de los efectos según su nivel de precisión. En el eje X se puede observar la inversa de los errores estándares (precisión) y en el eje Y el tamaño de los efectos observados estandarizados por su correspondiente error estándar. En la derecha del gráfico, se incluye un arco donde se representan los efectos observados sin estandarizar, de modo que se determina la magnitud del efecto observado para un estudio específico, siguiendo la proyección de la línea con origen en (0,0), pasando por el estudio de interés y llegando a la proyección en el arco.

FUNCIÓN SOBRERREPRESENTADA EN ENFERMOS: Adrenergic receptor signaling pathway involved in heart process.

En la figura 7 se muestra el *gráfico de bosque* con los resultados del MA para la función *GO: 0086023* (*adrenergic receptor signaling pathway involved in heart process*) donde la medida resumen del efecto es positiva.

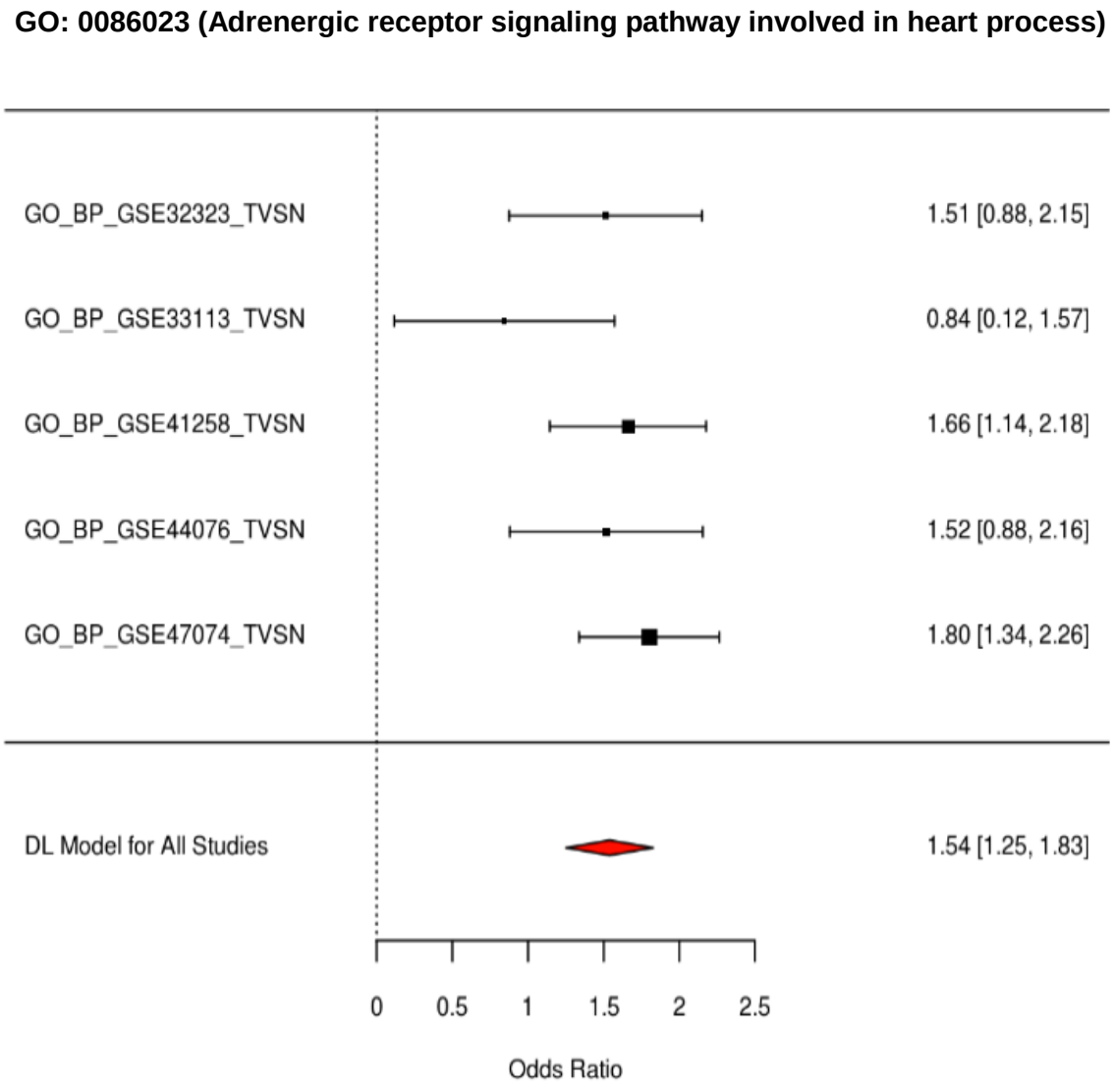

Figura 7. Distribución del efecto para la función GO: 0086023.

En la figura 8 se muestra el *gráfico de embudo*, el cual presenta la relación entre el efecto estudiado y distintos indicadores de la variabilidad para la función GO:0086023. En el eje X se muestran los valores del logaritmo del *odds ratio* y en el eje Y: el error estándar, la varianza en el muestreo y sus respectivos valores inversos (medidas de precisión). En todos los gráficos se repite le mismo patrón:

- En los dos primeros gráficos de la figura 6, se observan cambios en la distribución de la variabilidad (error estándar o varianza) en función del tamaño del efecto. Los estudios con *LOR* igual o superior a 1.5, muestran un ligero incremento de su error estándar y cuando son evaluados mediante los gráficos que incluyen los valores inversos del error estándar o varianza, se aprecia con mayor claridad, el despunte de dos estudios.

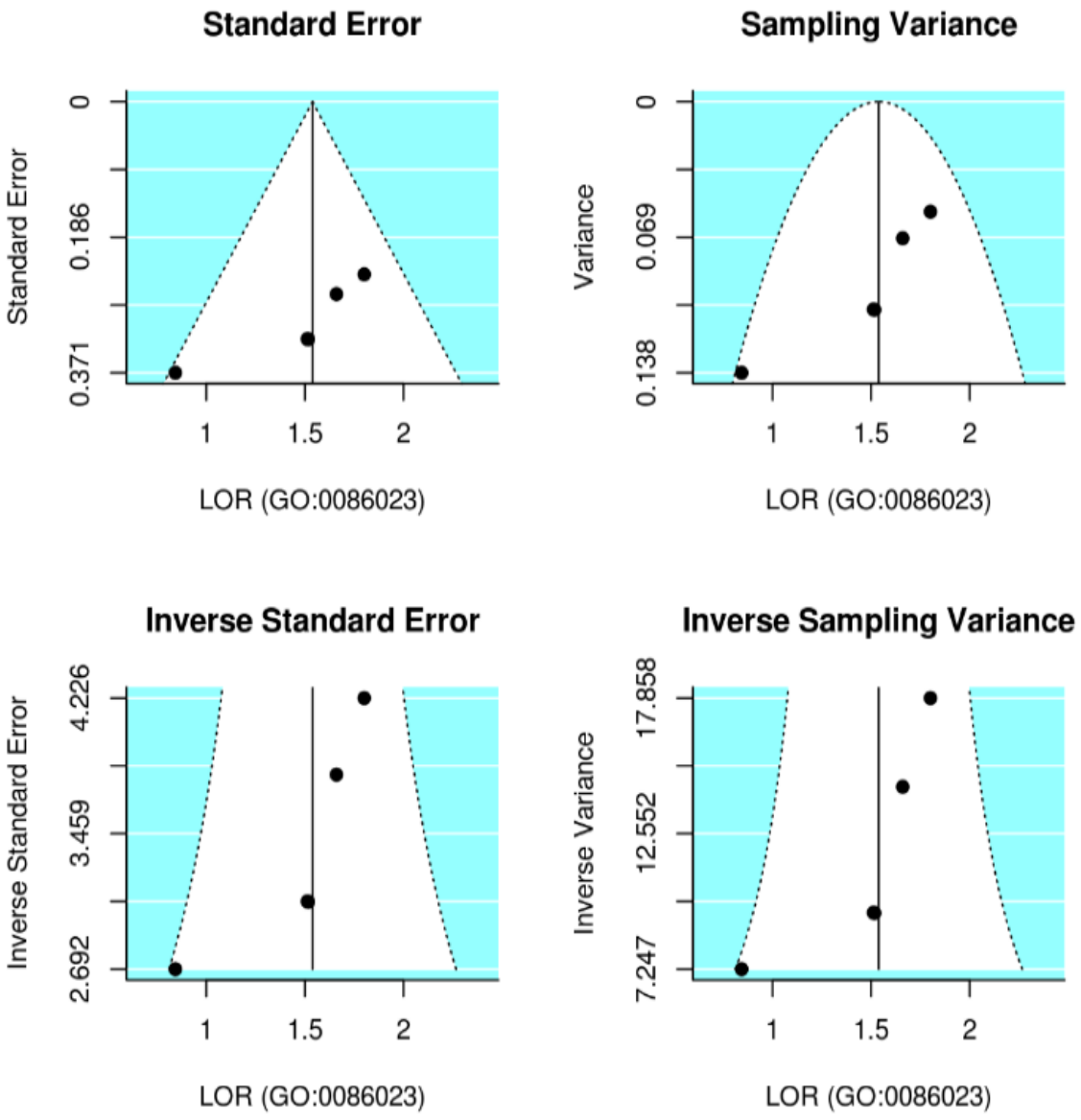

Figura 8. Variabilidad del efecto estudiado en la función GO:0086023.

En la figura 9 se observa el *gráfico radial* en el que se comprueba que los grupos de estudios muestran una precisión media elevada, y además dichos estudios quedan dentro del intervalo de confianza.

GO: 0086023 (Adrenergic receptor signaling pathway involved in heart process)

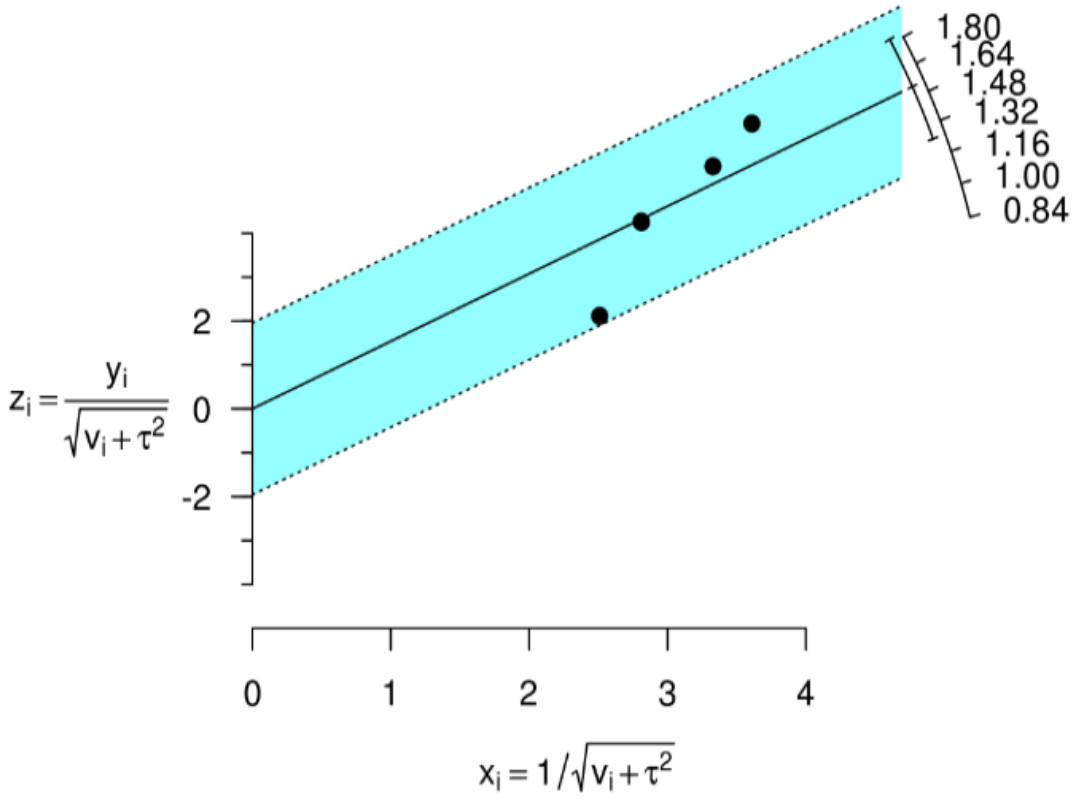

Figura 9. Variabilidad del efecto estudiado en la función GO:0086023.

FUNCIÓN SOBRERREPRESENTADA EN CONTROLES: Ctf18 RFC-like complex.

En la figura 10 se muestran el *gráfico de bosque* con los resultados del MA para la función *GO: 0031390* (*Ctf18 RFC-like complex*) donde la medida resumen del efecto es negativa.

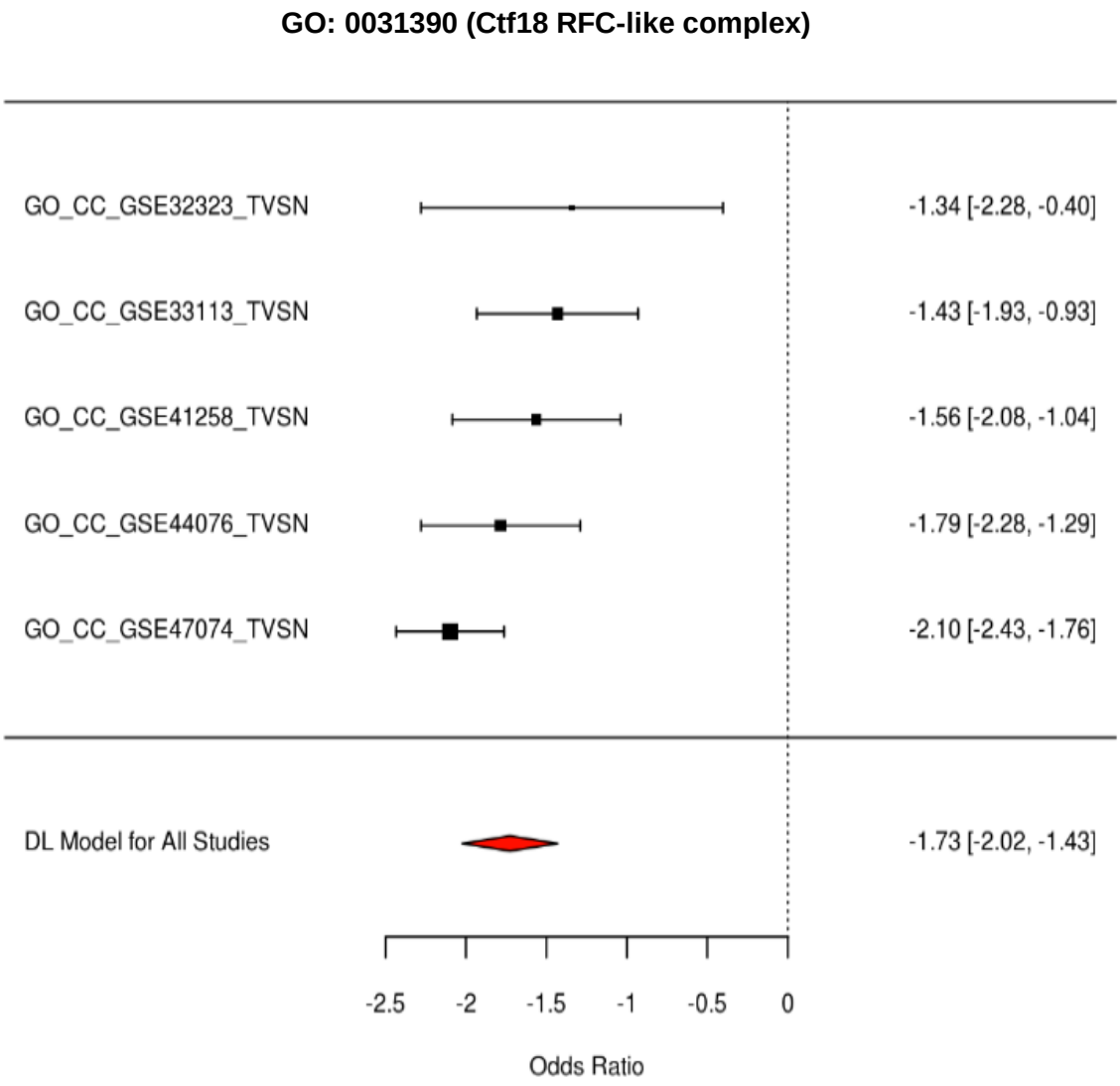

Figura 10. Distribución del efecto para la función GO:0031390.

En la figura 11 se muestra el *gráfico de embudo* para la función GO:0031390. En todos los gráficos se repite el mismo patrón:

- Hay un estudio que se separa de la región de confianza.
- En los dos primeros gráficos se observan grandes cambios en la distribución de la variabilidad (error estándar o varianza) en función del tamaño del efecto. También se comprueba que hay tres estudios con un error estándar mayor, el cual sigue apreciándose claramente en los gráficos que incorporan los valores inversos del error estándar o varianza.

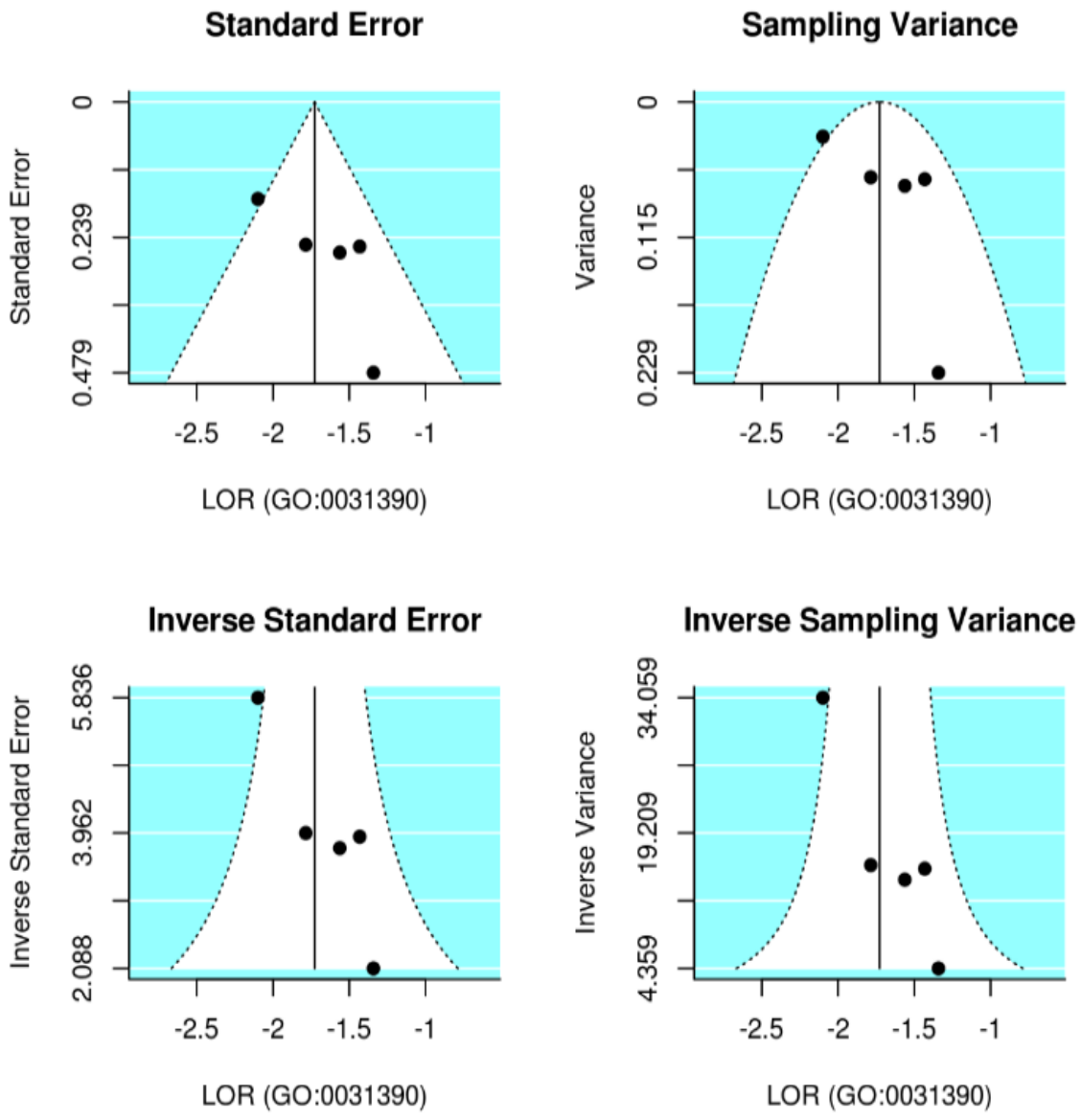

Figura 11. Variabilidad del efecto estudiado en la función GO:0031390.

En la figura 12 se observa el *gráfico radial* en el que se comprueba que los grupos de estudios muestran una precisión media elevada, y además dichos estudios quedan dentro del intervalo de confianza.

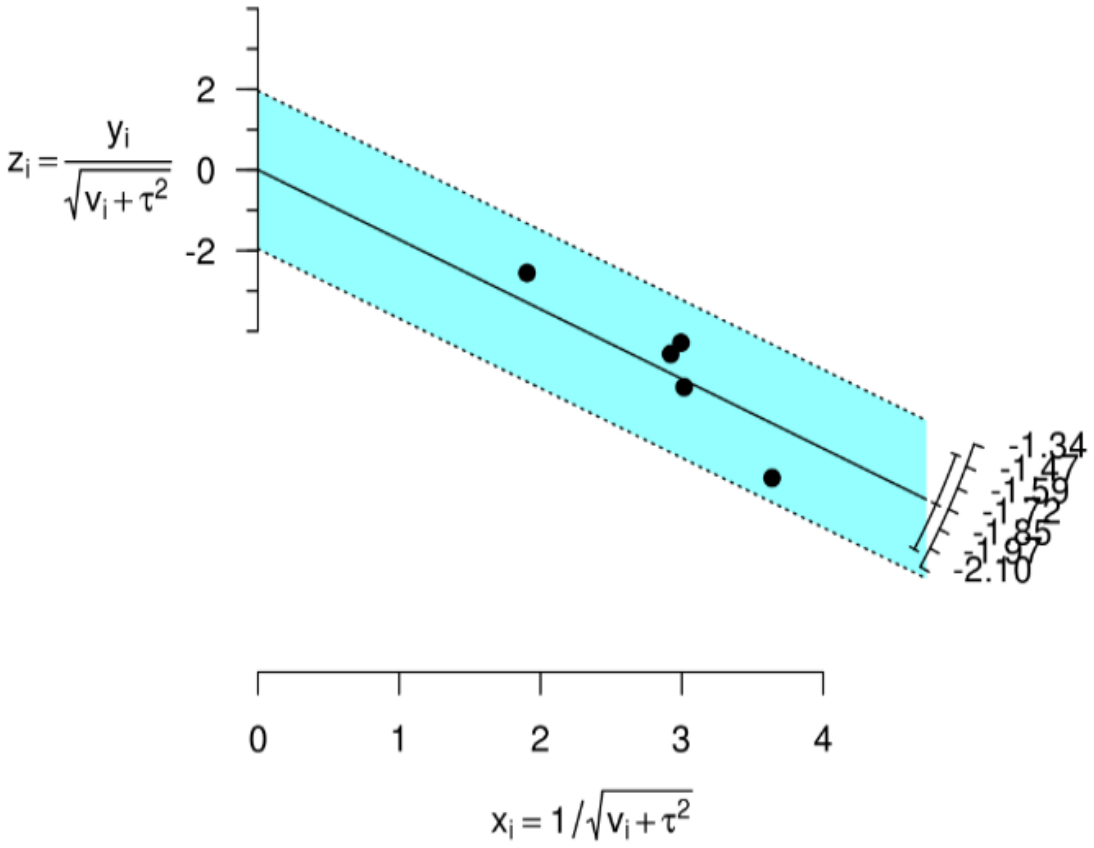

Figura 12. Variabilidad del efecto estudiado en la función GO:0031390.

Discusión

Ante el reto de la búsqueda de biomarcadores de diagnóstico CCR, se ha demostrado que la Bionformática contribuye en gran medida a facilitar la difícil tarea de buscar potenciales biomarcadores en un gran volumen de datos y seleccionar los mejores.

Hoy en día no se podría entender la oncología molecular sin la participación y contribución decisiva de abordajes computacionales, ambos van íntimamente ligados. La Bioinformática cuenta con muchos campos de aplicación y uno de los más importantes y prácticos, sin duda, es el de la oncología o estudio del cáncer. Gracias a la Bioinformática se están acelerando los avances científicos en la investigación del cáncer y se están reduciendo mucho los plazos de obtención de resultados. Hemos observado como ante la gran heterogeneidad de genotipos y fenotipos de CCR, existe consecuentemente una gran heterogeneidad de posibles biomarcadores a ensayar.

Hemos detectado un interesante grupo de funciones significativas y con una magnitud alta de sobrerrepresentación en los grupos experimentales descritos. A continuación vamos a detallar y a caracterizar el CCR a nivel molecular a partir de estos resultados, de forma que tendremos términos GO y rutas KEGG que estén sobrerrepresentadas en el grupo de *enfermos* (términos GO y KEGG sobrerrepresentados en grupos de genes que tienen un nivel de expresión alto en el grupo de enfermos y bajo en el grupo de controles), y términos que estén sobrerrepresentados en el grupo de *controles* (términos GO y KEGG sobrerrepresentados en grupos de genes que tienen un nivel de expresión alto en el grupo control y bajo en el grupo de en enfermos).

Con respecto a *controles*, encontramos términos GO que forman parte de las condiciones fisiológicas del ser humano: aquellos involucrados en procesos metabólicos, biosíntesis de nucleótidos y aminoácidos, replicación del ADN, señalización y regulación del ciclo celular, mantenimiento enzimático y estructural de la expresión génica, mecanismos de reparación, y también mantenimiento de la metilación, un fenómeno epigenético de importancia médica ya que se encarga de mantener el silenciamiento génico en el desarrollo normal, la impronta genómica y la inactivación del cromosoma X. No obstante, alteraciones de este proceso se encuentran implicadas en procesos cancerígenos, donde las células malignas van acompañadas de una hipometilación global que está asociada a la activación de los genes necesarios para la invasión y metástasis, y de una hipermetilación local relacionada con la represión de los genes supresores de tumores (RODRIGUEZ-DORANTES *et al*., 2004).

Asimismo también encontramos procesos de organización de componentes citoplasmáticos, replicación de ADN y modificación de ARN mitocondrial, proceso de gran importancia, ya que se encarga de optimizar la síntesis proteica y actúa como regulador de la expresión génica. Defectos en este proceso se encuentran asociados con enfermedades neuromusculares (MESENGUER *et al*., 2015).

Y por último se destacan procesos y componentes relacionados con el mantenimiento de la actividad del espliceosoma, encargado de la eliminación de los intrones que no van a ser traducidos a proteínas, permitiendo generar variantes proteicas a partir de un solo gen, cada una con funciones diversas. Actualmente se sabe que este proceso, conocido como *splicing* es la principal fuente de diversidad proteica ya que el 70 % de los genes humanos lo sufren, y defectos en este procesos originan hasta el 50 % de las enfermedades genéticas, incluido el cáncer (MARTÍNEZ-MONTIEL *et al.,* 2015).

En el caso de las rutas KEGG, únicamente hemos obtenido tres, dos de las cuales se encuentran relacionadas con los términos GO que acabamos de mencionar: replicación y reparación del ADN, y una hace referencia al sistema ubiquitina-proteosoma, principal encargado de la depuración celular de las proteínas anormales, mutantes, dañadas por lesión oxidativa o ubicadas incorrectamente (MICHELI, 2006).

A continuación nos centraremos en los *enfermos*, en los que encontramos un grupo muy reducido de términos GO, relacionados principalmente con condiciones anormales y no fisiológicas: entre ella la transducción de señales, siendo posible que se encuentre alterado este mecanismo, ya que frecuentemente, estas vías son activadas de manera inadecuada en las células cancerosas, ya sea por la expresión inadecuada de un oncogen que especifica un factor de crecimiento, un receptor para un factor de crecimiento, o parte de las vías de señalización intracelular (SOTO, 2003).

También se destacan procesos de regulación de comunicación celular, en este caso es interesante destacar los resultados de un estudio Español (BAZELLIÈRES *et* al., 2015) en el que se ha descubierto un nuevo mecanismo de comunicación física celular que conduce a la metástasis en el cáncer. Dicha comunicación es importante para el funcionamiento coordinado de los órganos del cuerpo y su pérdida se considera uno de los aspectos característicos de diversas enfermedades como el cáncer o las enfermedades inflamatorias crónicas. Tradicionalmente esta pérdida de comunicación entre células se ha entendido como una alteración de señales puramente bioquímica, como las hormonas, sin embargo el grupo dirigido por Xavier Trepat, investigador ICREA en el IBEC, ha cuestionado esta visión tradicional, y ha considerado que la comunicación física entre células es tan importante como la química. En este estudio el equipo ha identificado las moléculas que se encuentran involucradas en esta comunicación física, algunas de las cuales están alteradas en cáncer.

Asimismo, en *enfermos* también aparecen procesos de reducción del pH celular. Se ha establecido recientemente (HARGUINDEY *et al.*, 2017) que el desarrollo del cáncer es debido principalmente a la pérdida del equilibrio natural ácido-base de la célula, aunque ello no excluya otros factores importantes que además de éste puedan jugar un papel importante en el proceso cancerígeno. De esta forma se denota que todas las personas con cáncer sufren una "alcalosis celular maligna" en las células tumorales, causada principalmente por una continua e incontrolada extracción de iones de hidrógeno del interior de la célula. Una vez puesto en marcha el mecanismo canceroso, para que tenga lugar la replicación celular debe mantenerse un cierto pH intracelular elevado inhibiéndose así todo intento de inducir la apoptosis selectiva. Para ello, las células tumorales inician una serie de mecanismo antiacidificantes destinados a mantener el pH lo más alcalino posible. Con esto se puede intuir que este proceso que se encuentra sobrerrepresentado en *enfermos* pueda estar contribuyendo a la reducción del pH ante una situación de alcalinidad producida por el cáncer.

Finalmente se destacan procesos relacionados con la señalización y transducción de señales. En este aspecto se sabe que existen diferentes moléculas que se encuentran en el medio, las cuales actúan sobre receptores de membrana o citoplasmáticos, induciendo un conjunto de fenómenos que dirigen el crecimiento, diferenciación o muerte celular, sin embargo, en ciertas ocasiones, estos receptores pueden mostrar un comportamiento anómalo, produciendo señales equivocadas, que en última instancia inducen la transformación maligna. Muchos oncogenes están relacionados con esta disfunción, ya sea en los receptores mismos o en el sistema de mediadores intracelulares (PENCHASZADEH, 2002).

Ciertos receptores generan segundos mensajeros como calcio, diacilglicerol (DAG), AMP cíclico (AMPc) o GMP cíclico (GMPc). Estas sustancias son producidas mediante un sistema complejo de reacciones de fosforilación y desfosforilación, desencadenadas por las proteínas G.

De este modo, el haber obtenido en *enfermos* procesos biológicos y componentes relacionados con la señalización de GMPc, receptores adrenérgicos, secreción de acetilcolina (AC) y receptor acoplado a proteínas G, justificaría que en este grupo estas vías de transducción de señales puedan estar alteradas, generando así un comportamiento anormal celular.

Con respecto a las rutas KEGG, detectamos una vía significativa que hace referencia al metabolismo de la arginina y ornitina. Esta ruta está relacionada con los términos GO obtenidos para el grupo de los *enfermos*. La arginina es un aminoácido semiesencial que en ciclo de la urea se descompone en urea y ornitina, siendo además precursora del óxido nítrico (NO), una importante molécula en el organismo que actúa como vasodilatador. En esta formación del NO interviene la enzima sintasa del óxido nítrico (NOS), la cual es activada por la acetilcolina (AC). La AC se une a su receptor, y la señal es transmitida mediante el inositol trifosfato (IP3), produciendo la liberación de iones calcio, que a su vez activan a la NOS. Por otro lado interviene el factor de necrosis tumoral alfa (TNFα), una proteína perteneciente al grupo de las citosinas que participa en la infamación y apóptosis de células tumorales. Este factor incrementa la actividad de la NOS, así, el NO producido en una célula, se difunde al músculo liso adyacente y activa la guanilato ciclasa, incrementándose la concentración de GMPc. De todo esto se deduce que, en estas personas *enfermas* está teniendo lugar un proceso inflamatorio que puede estar asociado al CCR, de forma que uno de los mecanismos que utiliza el cuerpo humano para hacer frente a esta situación es la vasodilatación, un aumento del calibre de las arteriolas que produce un incremento de la cantidad de sangre que llega al lugar inflamado, lo cual permite el aporte de moléculas y células sanguíneas.

Conclusiones

1. La aplicación de métodos y técnicas bioinformáticas a datos transcriptómicos de muestras de pacientes con CCR ha resultado eficaz para conseguir una mejor caracterización de los perfiles moleculares de la enfermedad estudiada.

2. Todas aquellas funciones y rutas comunes obtenidas (en *enfermos* y *controles*) permiten obtener una visión global de la enfermedad, y una mejor comprensión de su desarrollo y progresión.

3. La integración y análisis bioinformático de estos datos ómicos, no sólo permite obtener un mejor conocimiento de la enfermedad, sino que también puede ayudar a descubrir posibles causas.

4. La aplicación de una metodología de enriquecimiento funcional basada en técnicas de MA ha permitido la confirmación y descubrimiento de relaciones funcionales, además ha ofrecido una gran flexibilidad en el manejo de información funcional y ha permitido la integración de información biológica de interés.

5. En el caso de los *controles*, aquellas funciones y rutas que aparecen principalmente sobrerrepresentadas son aquellas relacionadas con procesos metabólicos, biosíntesis de nucleótidos y aminoácidos, replicación del ADN, señalización y regulación del ciclo celular, mantenimiento enzimático y estructural de la expresión génica, junto con mecanismos de reparación.

6. En el caso de *enfermos* aquellas funciones y rutas que aparecen principalmente sobrerrepresentadas son aquellas relacionadas con pH celular, procesos de señalización y transducción de señales, inflamación y vasodilatación.

7. La Bioinformática se ha convertido en una disciplina necesaria para el manejo de grandes volúmenes de datos y su transformación en información biológica y clínicamente comprensible. La integración de modelos computacionales con los datos clínicos y las evidencias previas descritas en la literatura científica, trata de identificar nuevos biomarcadores de diagnóstico y pronóstico, que tras su validación en ensayos clínicos serán incorporados a la práctica clínica. Por todo ello, los abordajes descritos contribuyen a la aplicación y potenciación de la medicina personalizada, que jugará un papel importante en la terapia de las enfermedades y permitirá identificar a aquellos pacientes que se beneficien de un tratamiento específico según su perfil molecular, y por tanto ajustar las pautas de tratamiento de forma individualizada.

Bibliografía

BAUMGARTNER, C., GREGORY, D. L., MICHAEL, N., BERNHARD, P. & GERSZTEN, R.E. (2010). A new data mining approach for profiling and categorizing kinetic patterns of metabolic biomarkers after myocardial injury. Bioinformatics. 26: 1745-1751.

BRAMBILLA, C. & S. SPIRO (2001). Highlights in lung cancer. *Eur. Respir. J.* 18(4): 617-618.

BURREL, R.A., MCGRANAHAN, N., BASTEK, J. & C. SWANTON (2013). The causes and consequences of genetic heterogeneity in cáncer evolution. *Nature* 501(7467): 338-345.

COLUSSI, D., BRANDI, G., BAZZOLI, F. & L. RICCIARDIELLO (2013). Molecular pathways involved in colorectal cáncer: implications for disease behaviour and prevention. *Int. J. Mol. Sci.* 14(8): 16365-16385.

DERSIMONIAN, R. & N. LAIRD (1986). Meta-analysis in clinical trial. *Controlled clinical trials* 7(3): 177-188.

BAZELLIÈRES, E., CONTE, V., ELOSEGUI-ARTOLA, A., SERRA-PICAMAL, X., BINTANEL-MORCILLO, M., ROCA-CUSACHS, P., MUÑOZ, J.J., SALES-PARDO, M., GUIMERÀ, R. & X. TREPAT (2015). Control of cell-cell forces and collective cell dynamics by intercelular adhesome. *Nature Cell Biology* 17: 409-420.

FEARON, E.R. & B. VOGELSTEIN (1990). A genetic model for colorectal tumorigenesis. *Cell* 61(5): 759-767.

FERLAY, J., SOERJOMATARAM, I., ERVIK, M., DIKSHIT, R., ESER, S., MATHERS, C., REBELO, M., PARKIN, D.M., FORMAN, D. & F. BRAY (2013). GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase N0, 11 [Internet]. Lyon, France: International Agency for Research on Cancer. Available from: http://globocan.iarc.fr, accessed on 09/05/2017.

GALBRAITH, R., (1988a). A note on graphical presentation of estimated odds ratios from several clinical trials. *Statistics in medicine* 7(8): 889–894.

GALBRAITH, R., (1988b). Graphical display of estimates having differing standard errors. *Technometrics*, 30(3): 271–281.

GALBRAITH, R.F., (1994). Some applications of radial plots. *Journal of the American Statistical Association* 89(428): 1232–1242.

GISBERT, J.P. & X. BONFIL (2004). Systematic reviews and meta-analyses: how should they be performed, evaluated and used?. *Gastroenterol. Hepatol.* 27(3): 129-149.

HARGUINDEY, S., STANCIU, D., DEVESA, J., ALFAROUK, K., CARDONE, R.A., POLO-OROZCO, J.D., DEVESA, P., RAUCH, C., ORIVE, G., ANITUA, E., ROGER, S. & S.J. RESHKIN (2017). Cellular acidification as a new approach to cancer treatment and to the understanding and therapeutics of neurodegenerative diseaeses. *Semin. Cancer Biol.* 43: 157-179.

HIGGINNS, J & S.G. THOMPSON (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine* 21(11): 1539-1558.

HIRSCH, F.R., MERRICK, D.T. & W.A. FRANKLIN (2002). Role of biomarkers for early detection of lung cancer and chemoprevention. *Eur. Respir. J.* 19(6):1151-1158.

HUNTER, J.E., SCHMIDT, F.L. & G.B. JACKSON (1982). Meta-analysis: cumulating research findings across studies. *Beverly Hills, Ca: Sage Publications*.

KHUDER, S.A., MUTGI, A.B. & S. NUGENT (2001). Smoking and breast cancer: a meta-analysis. *Environ Health*. 16(4): 253-261.

LE MARCHAND, L., WILKENS, LR., HANKIN, J.H., KOLONEL, L.N. & L.C. LYU (1997). A case-control study of diet and colorectal cáncer in a multiethnic population in Hawaii (United States): lipids and foods of animal origin. *Cancer Causes Control* 8(4): 637-648.

LYNCH, J.P. & T.C. HOOPS (2002). The genetic pathogenesis of colorectal cancer. *Hematol Oncol. Clin. North Am.* 16(4): 775-810.

MESEGUER, S., MARTÍNEZ-ZAMORA, A., GARCÍA-ARUMÍ, E., ANDREU, AL. & M.E. ARMENGOD (2015). The ROS-sensitive microRNA-9/9* controls the expression of mitochondrial tRNA-modifying enzymes and is involved in the molecular mechanism of MELAS syndrome. *Human molecular genetics* 24: 167-184.

MICHELI, F (2006). Enfermedad de Parkinson y transtornos relacionados. 2. Ed. Buenos Aires: Médica Panamericana, 68 pp.

MONTANER, D., MINGUEZ, P., AL-SHAHROUR, F. & J. DOPAZO (2009). Gene set internal coherence in the context of functional profiling. *BMC Genomics* 10:197

MONTANER, D. & J. DOPAZO (2010). Multidimensional gene set analysis of genomic data. *PLoS ONE* 5(4), e10348.

MARTÍNEZ MONTIEL, N., ROSAS-MURRIETA, N. & R. MARTÍNEZ-CONTRERAS (2015). Alternative splicing regulation. Implication in cancer diagnosis and treatment. *Medicina Clínica* 144(7):317-323.

PARKIN, D.M., BRAY, F., FERLAY, J. & P. PISANI (2005). Global cancer statistics, 2002. *C.A. Cancer J. Clin.*; 55:74-108.

PENCHASZADEH, V.B. (2002). Bioética y genética médica en América Latina. *Braz. J. Genetics* 20(1): 163-170.

PÉREZ, J., JIMÉNEZ, R., LIE, A.E. & GONZÁLEZ, B.Y. (2010). Comportamiento bioquímico del cáncer. *Medimay* 16(1).

R CORE TEAM (2016). *R: A lenguaje and environment for statical computing*, Vienna, Austria: R Foundation for Statical Computing. Available at: https://www.R-project.org/.

RODRIGUEZ-DORANTES, M., TELLEZ-ASCENCIO, N., CERBÓN, M.A., LÓPEZ, M. & A. CERVANTES (2004). Metilación del ADN: un fenómeno epigenético de importancia médica. *Rev. Invest. Clin.* 56: 56-71.

SARTOR, M.A., LEIKAUF, G.D. & M. MEDVEDOVIC (2009). LRpath: a logistic regression approach for identifying enriched biological groups in gene expression data. *Bioinformatics* 25(2): 211-217.

SLATTERY, M.L. (2004). Physical activity and colorectal cáncer (2004). *Sports Med* 34(4): 239-252.

SOTO I. (2003). Transducción de señales y cáncer. *Revista Especializada en Ciencias de la Salud* 6(1): 45-50.

SPIRA, A., BEANE, J.E., SHAH, V., STEILING, K., LIU, G., SCHEMBRI, F., GILMAN, S., DUMAS, Y.M., CALNER, P., SEBASTIANI, P., SRIDHAR, S., BEAMIS, J., LAMB, C., ANDERSON, T., GERRY, N., KEANE, J., LENBURG, M.E. & J.S. BRODY (2007). Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer, *Nat. Med.* 13(3):361-6.

STERNE, J.A. & M. EGGER (2001). Funnel plots for detecting bias in metaanalysis: Guidelines on choice of axis. *Journal of clinical epidemiology* 54(10): 1046–1055.