VNIVERSITAT DE VALÈNCIA

Programa de doctorado en Biomedicina y Biotecnología

Epidemiological and molecular associations between central nervous system disorders and cancer

TESIS DOCTORAL

Jaume Forés Martos

Directores: Dr. Rafael Tabarés Seisdedos
Dr. Ferrán Catalá López
Dr. Joan Climent Bataller

Tutor: Dr. Juli Peretó Magraner

Valencia, enero 2021

VNIVERSITAT DE VALÈNCIA

Programa de doctorado en Biomedicina y Biotecnología

Epidemiological and molecular associations between central nervous system disorders and cancer

TESIS DOCTORAL

Jaume Forés Martos

Directores: Dr. Rafael Tabarés Seisdedos
Dr. Ferrán Catalá López
Dr. Joan Climent Bataller

Tutor: Dr. Juli Peretó Magraner

Valencia, enero 2021

VNIVERSITAT DE VALÈNCIA

INFORME DIRECTORS/AS, TUTOR/A PER A DIPÒSIT DE TESI

Director (es) / Codirector (es):

1.- Cognoms, Nom. Tabarés Seisdedos, Rafael. N.I.F: 19097070D, Departament/Institut: Dept. Medicina Centre: Universitat de València

2.- Cognoms, Nom. Catalá-López, Ferrán. N.I.F: 24388411Q, Departament/Institut: Escuela Nacional de Sanidad, Centre: Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación

3.- Cognoms, Nom. Climent Bataller, Joan. N.I.F: 20426545S, Departament/Institut: PASAPTA, Centre: Facultad de Veterinaria. Universitat CEU Cardenal Herrera

Tutor o tutora (si pertoca)

Cognoms, Nom. Peretó Magraner, Juli. N.I.F: 20776408W, Departament/Institut: Dept. Bioquímica i Biologia Molecular Centre: Facultat de Ciències Biològiques. Universitat de València

Com a director (a) /codirector (a), tutor(a) de la tesi doctoral: "*Epidemiological and molecular associations between central nervous system disorders and cancer*"

de D/Dª **Jaume Forés Martos**,

estudiant del programa de doctorat **en Biomedicina i Biotecnologia**, de la Universitat de València, emeten informe *favorable* per a la realització del dipòsit i defensa de la tesi doctoral.

Data: 14/01/2021

Signat:	Signat:	Signat:	Signat:
RAFAEL| TABARES| SEISDEDOS Firmado digitalmente por RAFAEL|TABARES|SEISDEDOS Fecha: 2021.01.14 23:31:14 +01'00'	Ferrán Catalá-López Firmado digitalmente por Ferrán Catalá-López Fecha: 2021.01.14 22:13:19 +01'00'	JUAN| CLIMENT| BATALLER Firmado digitalmente por JUAN|CLIMENT|BATALLER Fecha: 2021.01.14 22:19:48 +01'00'	JULI| PERETO| MAGRANER Firmado digitalmente por JULI|PERETO|MAGRANER Fecha: 2021.01.15 09:24:24 +01'00'
Director/a	Director/a	Director/a	Tutor/a

Acknowledgments

I would first like to thank my family for always being supportive and patient. To my mom and dad, who know what it's like to raise a turtle. To my brothers, with whom I would like to spend more time and score many more three-pointers, and yes, why not, get some more double fines from time to time. And especially to Laura, who is always there for me with her hot kettle of tea and her funny Italian accent. She is the living proof that one can always have cold feet but a warm heart. I would also like to thank the recently imported international branch of the family, especially the two little monkeys. They are new in the hood, and they are growing so fast. Thanks for your laughter. It makes everything else fade away. I would also like to thank all the lovely people I met while doing my Ph.D., especially the Joanets. All the best to you guys. Finally, I would also like to thank my advisors. I have learned valuable lessons from them all, both personal and professional. In particular, I would like to thank my principal advisor Rafael Tabarés Seisdedos, for his support and kindness during these years.

Contents

Abbreviations

List of Figures

List of Tables

Appendices

Supplementary Appendices

The following appendices include additional information and analyses, that, due to their extension, could not be included in the main text.They are deposited in the Open Science Fremework (available at: https://osf.io/wzprg/)

Supplementary Appendix 1: Supplementary Figures and Tables.

Supplementary Appendix 2: Differential gene expression meta-analysis results for all the included disorders.

Supplementary Appendix 3: Complete Hallmarks GSEA results for all the included disorders.

Supplementary Appendix 4: Complete Canonical Pathways GSEA results for all the included disorders.

Supplementary Appendix 5: Complete Gene Ontology GSEA results for all the included disorders.

Supplementary Appendix 6: Disease-associated co-expression modules and enrichment in biological processes and cell type-specific markers.

Abstract

The study of comorbidity is becoming a key topic in biomedical research, which is especially relevant in the context of population ageing. Comorbidity has profound implications for individuals, practitioners, and health care systems. As a consequence, increasing efforts are being made by the scientific community to characterize better how disorders relate to each other and to identify the factors producing these associations.

Cancer and central nervous system (CNS) disorders are among the top leading causes of death and disease burden worldwide. In recent decades direct and inverse patterns of association between CNS disorders and cancer have been reported. However, observational studies have often found contrasting results. Consequently, evidence synthesis methods such as systematic reviews and meta-analysis have emerged as a critical tool to synthesize and evaluate the quality of the evidence regarding a specific research question.

In addition, in the course of the Omics era, an unprecedented amount of information regarding the molecular bases of individual disorders has been produced, opening the door to the study of comorbidity from a molecular perspective through the identification of joint alterations in variants, genes, and biological processes.

In the present thesis, we aimed to characterize the epidemiological and molecular associations between CNS disorders and cancer and to identify the potential role of their indicated medications. To this end, we first determined if CNS disorder patients presented an altered risk of subsequent cancer incidence and mortality by conducting systematic reviews and meta-analyses of observational studies. Second, we investigated if CNS disorders and cancers presented joint patterns of transcriptomic dysregulation using differential gene expression meta-analysis and weighed co-expression network analysis methods. Third, interactome-based methods and genetic correlations were employed to study the involvement of disease-associated genes and shared genetic variability. Finally, the impact of the medications indicated for the treatment of both sets of disorders in the reported comorbidities was assessed by the analysis of a large repository including information of cell lines treated with the indicated drugs.

Our results suggest that patients suffering from neurodegenerative disorders are at a reduced risk of subsequent cancer incidence and mortality compared to controls. Autism spectrum disorder, bipolar disorder, and schizophrenia (SCZ) patients are at an increased risk of cancer mortality but not cancer incidence, whereas major depression patients presented an increased risk of cancer incidence and mortality. Several associations between CNS disorders and site-specific cancers were also identified. Significant direct and inverse patterns of transcriptomic dysregulation between CNS disorders and cancers were observed in our transcriptomic analyses, as well as the

presence of joint alterations in several biological processes (i.e., cell cycle, apoptosis, immune system, and oxidative phosphorylation). Significant genetic correlations were also identified between CNS disorders and cancers, including those observed between Parkinson's disease and melanoma and SCZ and breast cancer. Finally, several drugs indicated for the treatment of CNS disorders, such as antipsychotics, antidepressants, and acetyl-cholinesterase inhibitors were found to produce transcriptomic alterations that mimicked or reversed those found in some cancer types, indicating their potential role in the CNS and cancer comorbidity.

Keywords: comorbidity, central nervous system disorders, cancer, systematic reviews, meta-analysis, transcriptomics, interactome, genetic correlations, medications

Resumen

Introducción

Alvan R. Feinstein acuñó el término comorbilidad en 1970 para referirse a "Cualquier entidad adicional que haya existido o pueda ocurrir en el transcurso clínico de un paciente que presente una enfermedad índice". Durante las décadas que siguieron a esta definición emergieron nuevos conceptos relacionados con la idea inicial de Feinstein, como por ejemplo los términos multimorbilidad, carga de enfermedad y fragilidad. La comorbilidad se define en términos epidemiológicos como la presencia de una probabilidad mayor que la esperada en la coocurrencia de dos patologías. Además, en los últimos años un número creciente de estudios ha puesto de manifiesto la importancia de la comorbilidad inversa, que se define como la presencia de una probabilidad menor que la esperada en la coocurrencia de dos enfermedades.

El estudio de la comorbilidad se está convirtiendo en un tema clave en la investigación biomédica, que es especialmente relevante en el contexto de los países con poblaciones envejecidas. La comorbilidad tiene implicaciones importantes para los pacientes, los profesionales sanitarios y los sistemas de salud. En este contexto, el esfuerzo conjunto de la comunidad científica está ayudando a caracterizar los patrones de asociación existentes entre los distintos trastornos, así como a elucidar los mecanismos y factores implicados en su origen.

El cáncer y las enfermedades del sistema nervioso central (SNC) se encuentran entre las principales causas de muerte y carga de enfermedad a nivel mundial. En las últimas décadas un número creciente de estudios ha sugerido la existencia de asociaciones epidemiológicas directas e inversas entre estos dos grupos de enfermedades. Sin embargo, estos han arrojado con frecuencia resultados discrepantes. En este sentido, los métodos de síntesis de le evidencia, tales como las revisiones sistemáticas y metaanálisis proporcionan una herramienta fundamental para evaluar la calidad de la evidencia disponible con respecto a una pregunta de investigación concreta.

El estudio más exhaustivo dedicado a analizar las asociaciones epidemiológicas entre el cáncer y las enfermedades del SNC publicado hasta la fecha sugiere que estos pacientes presentan un riesgo reducido en la incidencia de cáncer. En particular, los pacientes con enfermedades neurodegenerativas presentaron una reducción del riesgo más pronunciada. Además, también se identificaron incrementos en el riesgo de desarrollar tipos tumorales específicos. Por ejemplo, los pacientes con enfermedad de Párkinson (EP) están sujetos a una mayor incidencia de melanoma y las pacientes diagnosticadas con esquizofrenia (EZF) presentan una incidencia elevada de cáncer de mama en comparación con los grupos control.

Se han propuesto una amplia variedad de mecanismos para tratar de explicar las causas de las asociaciones epidemiológicas observadas. Estos incluyen la presencia de alteraciones

compartidas en genes y procesos biológicos, la influencia conjunta de variantes genéticas, la exposición a factores de riesgo comunes, así como la presencia de sesgos sistemáticos en los estudios observacionales, tales como el sesgo muestral o la falta de consideración de variables de confusión relevantes como los hábitos de consumo de tabaco. Además del rol potencial de determinados tratamientos farmacológicos.

Durante el transcurso de la era ómica se ha producido una acumulación de información sin precedentes sobre las bases moleculares de las enfermedades individuales. Este hecho ha abierto la puerta al estudio de la comorbilidad desde una perspectiva molecular, a través de la identificación de alteraciones compartidas en variantes, genes y procesos biológicos. Algunos estudios han intentado arrojar luz sobre la existencia de dichas alteraciones y sobre su implicación conjunta en la fisiopatología del cáncer y las enfermedades del SNC, señalando que las mismas podrían tener un papel modulador en las asociaciones epidemiológicas observadas. Por ejemplo, se ha sugerido que determinadas parejas de enfermedades del SNC y ciertos tipos tumorales, que presentan relaciones de comorbilidad inversa a nivel epidemiológico, exhiben también patrones de expresión diferencial opuestos. Es decir, que aquellos genes que tienden a estar sobreexpresados en un miembro de la pareja tienden a estar infraexpresados en el otro miembro y viceversa. A este respecto, se ha observado que los pacientes con EZF, enfermedad de Alzheimer (EA) y EP presentan patrones de expresión diferencial opuestos con los cánceres colorrectales, de pulmón y de próstata y se ha sugerido que genes como *PIN1* y *ATP13A2* y rutas biológicas como la señalización a través de P53 y Wnt podrían tener un papel clave en estos procesos.

Además de las aproximaciones basadas en métodos transcriptómicos, otros estudios han iniciado a explorar el efecto conjunto de variantes genéticas en parejas de enfermedades del SNC y tumores específicos. Por ejemplo, empleando un método denominado cross-trait LD score regression, un estudio reciente encontró correlaciones genéticas significativas entre EA y cánceres de mama y pulmón, sugiriendo que determinadas variantes genéticas podrían estar asociadas a la modulación conjunta del riesgo de ambas enfermedades.

Finalmente, el uso de determinados tratamientos farmacológicos se ha propuesto como un elemento de potencial relevancia en la modulación de las relaciones de comorbilidad entre las enfermedades del SNC y el cáncer. Por ejemplo, el disulfiram (DSF), un fármaco empleado durante décadas para el tratamiento de la dependencia al alcohol, se ha mostrado capaz de reducir la tasa de mortalidad por cáncer en pacientes que mantuvieron su uso durante el transcurso de la enfermedad en comparación con aquellos pacientes que fueron usuarios previos de DSF, pero que abandonaron el tratamiento al menos un año antes de ser diagnosticados de cáncer. Otros ejemplos de fármacos que han sido señalados como potenciales moduladores de las asociaciones entre las enfermedades del SNC y el cáncer incluyen los antipsicóticos, que han sido involucrados

en el aumento en el riesgo de cáncer de mama observado en pacientes con esquizofrenia o el uso de levodopa, que ha sido vinculado al aumento de riesgo de melanoma observado en pacientes con EP.

Objetivos

El objetivo general de la presente tesis es el de evaluar las relaciones de comorbilidad entre una selección de enfermedades del SNC que incluyen la EA, los trastornos del espectro autista (TEA), la EP, la enfermedad de Huntington (EH), la depresión mayor (DM), el trastorno bipolar (TB), la EZF, y el cáncer desde una perspectiva epidemiológica y molecular, además de examinar el posible papel de los fármacos de uso frecuente en dichas asociaciones. Para ello proponemos una serie de objetivos específicos.

Objetivo 1: Sintetizar la evidencia epidemiológica existente sobre las asociaciones entre las enfermedades del SNC seleccionadas y el cáncer a través de la elaboración de revisiones sistemáticas y metaanálisis de estudios observacionales. Estos estudios deben incluir como exposición el diagnóstico de alguna de las enfermedades del SNC y como outcome la incidencia o mortalidad por cualquier tipo de cáncer o por tipos tumorales específicos.

Objetivo 2: Analizar las asociaciones transcriptómicas entre las siete enfermedades del sistema nervioso central incluidas y 22 tipos tumorales específicos. Los tipos tumorales seleccionados son: leucemia linfoblástica aguda, leucemia mieloide aguda, cáncer de vejiga, cáncer de mama, cánceres cerebrales, cáncer de cérvix, colangiocarcinoma, leucemia linfocítica crónica, leucemia mieloide crónica, cáncer colorrectal, linfoma difuso de células B grandes, linfoma folicular, carcinoma de cabeza y cuello, cáncer de riñón, cáncer de pulmón, cáncer hepático, cáncer de ovario, cáncer de páncreas, melanoma, cáncer gástrico y cáncer de tiroides. El cálculo de las asociaciones transcriptómicas se basa en la comparación de las firmas de expresión obtenidas mediante el metaanálisis de expresión diferencial de múltiples estudios transcriptómicos para cada enfermedad y en la identificación de módulos de coexpesión asociados a cada una de las enfermedades y su comparación entre todas las posibles parejas formadas por cada enfermedad del SNC y cada tipo tumoral.

Objetivo 3: Evaluar la presencia de asociaciones a nivel del interactoma humano y de correlaciones genéticas entre las enfermedades del SNC y el cáncer. Los genes y proteínas relacionados con una enfermedad concreta tienden a interaccionar entre ellos y a formar subgrafos conexos dentro del interactoma humano, denominados módulos de enfermedad. Desde esta perspectiva, las enfermedades son producidas por perturbaciones locales en estos módulos y las comorbilidades podrían surgir como consecuencia de la perturbación de módulos de enfermedad

superpuestos. En otras palabras, si los módulos de enfermedad de dos desórdenes se superponen a nivel del interactoma, las perturbaciones locales causantes de la primea enfermedad alterarían también las rutas biológicas relacionadas con la segunda. Por tanto, la presencia de superposiciones significativas a nivel del interactoma humano entre las enfermedades del SNC y el cáncer podría constituir un indicio en favor del potencial sustrato biológico de las relaciones de comorbilidad observadas entre ambos grupos de enfermedades. Asimismo, la presencia de correlaciones genéticas significativas entre parejas de enfermedades, calculadas a partir de estudios de asociación de genoma completo (GWAS), por sus siglas en inglés, podría ser indicativa de la existencia de variabilidad genética compartida con la capacidad de modular el riesgo de ambas enfermedades.

Objetivo 4: Estudiar el posible impacto de las indicaciones farmacológicas empleadas para el tratamiento de las enfermedades del sistema nervioso central y el cáncer en las relaciones de comorbilidad observadas entre ambos grupos de enfermedades. El efecto de la medicación se estudió empleando dos aproximaciones complementarias. La primera basada en el cómputo de distancias entre las dianas moleculares de los fármacos y los módulos de cada una de las enfermedades a nivel del interactoma humano y la segunda mediante el análisis de las firmas de expresión genética producidas al tratar líneas celulares con dichos fármacos y el cómputo de correlaciones entre los perfiles obtenidos y las firmas de expresión de cada una de las enfermedades del SNC y cánceres seleccionados.

Material y métodos

Material y métodos empleados en el desarrollo del objetivo 1:

Con el objetivo de sintetizar la evidencia disponible sobre las asociaciones entre las enfermedades del SNC y el cáncer desde una perspectiva epidemiológica, se efectuaron revisiones sistemáticas y metaanálisis de estudios observacionales previos siguiendo las guías de publicación proporcionadas por PRISMA-P y MOOSE. Se llevaron a cabo búsquedas sistemáticas en cuatro bases de datos de literatura científica (MEDLINE, Scopus, Embase y Web of Science) y se seleccionaron aquellos estudios en los que la exposición fuese el diagnóstico de alguna de las enfermedades del SNC seleccionadas y el outcome, la incidencia o mortalidad subsiguiente por cualquier tipo de cáncer o por tipos tumorales específicos, a través de un proceso de cribado basado en la lectura de títulos, resúmenes y textos completos. Se seleccionaron aquellos estudios observacionales (caso-control o de cohorte) que incluyesen medidas cuantificando el grado de asociación entre la exposición y el outcome en forma de riesgos relativo, razón de probabilidades y razones de tasas de incidencia o mortalidad. Se extrajo la información relevante de aquellos artículos que

cumplieron con los criterios de inclusión. Los elementos extraídos fueron, entre otros, el año de publicación, el país, el diseño del estudio, la duración del seguimiento de las cohortes, el número de participantes diagnosticados con la enfermedad del SNC, la presencia de ajuste por variables de confusión y los tamaños de efecto de las asociaciones, así como sus intervalos de confianza al 95%. La calidad de cada uno de los estudios observacionales seleccionados fue evaluada empleando la escala Newcastle-Ottawa. Para cada uno de los posibles outcomes se llevaron a cabo metaanálisis empleando el método inverso de la varianza y un modelo de efectos aleatorios debido a la heterogeneidad esperada en los datos. Finalmente, se evaluó el sesgo de publicación mediante la visualización de gráficos de embudo y mediante el cómputo de los tests de Eggers y Beggs.

Material y métodos empleados en el desarrollo del objetivo 2:

Con la finalidad de evaluar la presencia de asociaciones transcritómicas entre las enfermedades del SNC y el cáncer y estudiar el papel potencial de las alteraciones conjuntas en rutas y procesos biológicos se siguió el siguiente esquema de trabajo: En primer lugar, se efectuaron búsquedas de conjuntos de datos de expresión de RNA en Gene Expression Omnibus (GEO), Array Express (AE), Stanley Medical Research Institute y The Cancer Genome Atlas (TCGA) para cada una de las enfermedades consideradas. Únicamente se seleccionaron aquellos conjuntos de datos con al menos tres muestras derivadas de casos y tres muestras derivadas de controles procedentes del mismo tejido. Para garantizar la aplicación de métodos de normalización compatibles, solo se eligieron aquellos estudios generados empleando las plataformas de un solo canal más populares de Affymetrix, Illumina y Agilent. Para cada una de las enfermedades del SNC, con el objetivo de reducir la heterogeneidad de los datos, se seleccionaron estudios y muestras de una sola región cerebral para la que existiesen evidencias sólidas de su implicación en la fisiopatología de la misma. En el caso de los cánceres, se incluyeron estudios con muestras de tumores primarios y sus respectivos controles. Los estudios y muestras obtenidos a partir de líneas celulares y muestras derivadas de metástasis fueron descartados. Los estudios fueron preprocesados de manera individual y sometidos a un proceso de detección de valores atípicos a nivel de muestra y de conjunto de datos. Para cada una de las enfermedades incluidas se efectuó un meta-análisis de expresión diferencial integrando la información de los distintos estudios disponibles. Aquellos genes que presentaron p-valores ajustados por comparaciones múltiples menores de 0.05 (FDR < 0.05) se consideraron genes diferencialmente expresados. Posteriormente, los perfiles de cada una de las enfermedades del SNC se compararon con los del todos los tipos tumorales. En breve, la probabilidad de encontrar un número mayor de genes que esperado por azar en cada una de las cuatro intersecciones formadas por los genes sobreexpresados e infraexpresados en cada pareja de enfermedades fue evaluada mediante el uso de tests hipergeométricos. El enriquecimiento en

categorías funcionales de los genes ubicados en las intersecciones se evaluó mediante análisis de sobrerrepresentación. Los perfiles de expresión diferencial de cada una de las enfermedades fueron empleados como input para efectuar un enriquecimiento en conjuntos de genes (Gene Set Enrichment Analysis, GSEA). Los resultados fueron comparados entre las distintas parejas de enfermedades del SNC y los diferentes tipos tumorales con el objetivo de encontrar rutas y procesos biológicos alterados de manera conjunta en ambos grupos de enfermedades. Un conjunto alternativo de estudios transcriptómicos con información sobre 15 tipos tumorales, que fué generado empleando técnicas de secuenciación de RNA, se utilizó para validar las asociaciones transcriptócmicas observadas en los análisis basados en arrays de expresión. Finalmente, se generaron módulos consenso de coexpresión usando, para cada una de las enfermedades, todos los estudios transcriptómicos disponibles y el método de análisis ponderado de redes de coexpresión, WGCNA por sus siglas en inglés. Se computaron mediadas de asociación de cada uno de los módulos de coexpresión detectados con el estatus de enfermedad de las muestras, de modo que se identificaron aquellos módulos consenso de genes que tienden a estar sobreexpresados en los casos respecto a los controles y aquellos módulos que tienden a estar infraexprados en los casos con respecto a los controles. Se efectuó un análisis de enriquecimiento en procesos biológicos y marcadores genéticos vinculados a tipos celulares específicos de los genes contenidos en cada módulo de coexpresión. Por último, la superposición en el contenido de genes de los módulos asociados cada pareja de enfermedades se evaluó mediante el cómputo de tests hipergeométricos.

Material y métodos empleados en el desarrollo del objetivo 3:

En primer lugar, se construyó un interactoma humano mediante la integración de diferentes fuentes de información sobre interacciones entre genes y proteínas humanas. Estas incluyeron interacciones proteína-proteína derivadas de *The Human Reference Interactome* (HuRI), parejas de genes coexpresados obtenidas a partir del análisis de los datos de *the Genotype-Tissue Expression* (GTEx), información sobre complejos proteicos derivada de *the comprehensive resource of mammalian protein complexes* (CORUM), información sobre factores de transcripción y sus dianas obtenida a partir de TRANSFAC, interacciones quinasa-sustrato derivadas de PhosphoSitePlus y asociaciones metabólicas obtenidas a partir de *the Kyoto Encyclopedia of Genes and Genomes* (KEGG). Dos interactomas adicionales previamente publicados derivados de STRING y BIOGRID fueron seleccionados con el objetivo de evaluar el impacto de los distintos parámetros asociados con cada interactoma (número de genes incluidos, número de conexiones entre genes, grado promedio, etc.) en los resultados. En segundo lugar, se efectuaron búsquedas de genes y genes asociados a variantes implicados en cada una de las enfermedades incluidas en los análisis relativos al objetivo dos. Estas búsquedas se llevaron a cabo en las bases de datos *Disease Gene*

Network DisGeNet, *Phenotype-Genotype Integrator* (PheGenI) y la base de datos de asociaciones gen-enfermedad eDGAR. Se emplearon dos umbrales distintos a la hora de seleccionar genes y variantes asociados a enfermedad en DisGeNET y PhenGeI, de modo que para cada enfermedad se generaron dos conjuntos de genes asociados, uno obtenido aplicando criterios de selección relajados y otro obtenido aplicando criterios de selección estrictos. Para cada enfermedad se calcularon dos medidas de localización a nivel del interactoma de sus genes asociados, con el objetivo de determinar si estos tienden a concentrarse en un vecindario específico del interactoma humano. Para ello se emplearon dos métricas previamente descritas y tests de permutaciones para el cómputo de la significatividad de los valores observados. Posteriormente el grado de superposición de los módulos de enfermedad de todas las posibles parejas de enfermedades fue computado mediante una métrica adicional denominada separación entre enfermedades (SEE). El grado de significatividad de esta medida de asociación, así como de las anteriores, fue evaluado mediante tests de permutaciones en los que se tuvo en cuenta la distribución de los grados de los genes asociados a enfermedad de los conjuntos de genes iniciales. Por último, se efectuaron búsquedas de estudios de GWAS previos dedicados a la identificación de variantes genéticas asociadas con una modulación del riesgo de desarrollo de las enfermedades bajo análisis. Se obtuvieron los datos resumidos de cada estudio de GWAS a partir de varios repositorios públicos y en algunos casos, a través del contacto directo con los autores de los mismos. Empleando estos datos y un método denominado *cross-trait LD-score regression* se computaron las correlaciones genéticas entre todas las posibles parejas de enfermedades disponibles.

Material y métodos empleados en el desarrollo del objetivo 4:

Se evaluó el papel de las medicaciones indicadas para el tratamiento de los desórdenes del SNC y el cáncer en sus relaciones de comorbilidad mediante el uso de dos metodologías distintas. La primera está basada en el cómputo de distancias a nivel del interactoma entre las dianas de los fármacos indicados y los módulos de enfermedad de cada una de las enfermedades estudiadas. La segunda implicó la obtención de firmas de expresión de líneas celulares tratadas con los fármacos indicados y su comparación con los perfiles de expresión diferencial de cada una de las enfermedades del SNC y cánceres generadas en el objetivo 2 mediante el cómputo de correlaciones. Las indicaciones para cada una de las enfermedades fueron obtenidas a partir de MEDI-an, un repositorio de indicaciones que almacena información de diferentes fuentes, mientras que los datos sobre las dianas moleculares conocidas de cada fármaco fueron obtenidos a partir de DruBank. Se utilizó una métrica similar a las descritas en el objetivo dos con la finalidad de medir la cercanía de cada fármaco al módulo de enfermedad característico de cada uno de los desórdenes incluidos y se siguió la misma estrategia a la hora de calcular la significatividad de las asociaciones

descritas. Distancias significativamente menores a las observadas por azar indicarían el impacto de un fármaco concreto en el módulo de enfermedad característico de un desorden específico. Los perfiles de expresión diferencial de líneas celulares tratadas con las indicaciones identificadas fueros obtenidos a partir de LINCS L1000, un repositorio público que contiene información sobre los perfiles de expresión diferencial generados por miles de perturbaciones (moléculas pequeñas, ligados, *kncokdowns*, etc.). Se calcularon las correlaciones entre los perfiles producidos por estas moléculas y los perfiles de expresión diferencial idiosincráticos de cada una de las enfermedades obtenidos en desarrollo del objetivo 2. Las correlaciones positivas sugieren que un fármaco concreto produce firmas de expresión que asemejan a las firmas observadas en una enfermedad determinada, mientras que las correlaciones negativas sugieren que un determinado fármaco tendría el potencial de revertir las alteraciones transcriptómicas observadas en una enfermedad concreta.

Resultados y conclusiones

Resultados y conclusiones del objetivo 1:

Las búsquedas sistemáticas en las bases de datos de literatura científica permitieron la identificación de un total de 8749 referencias. Tras los cribados basados en la lectura de títulos, resúmenes y textos completos, 192 artículos cumplieron con los criterios de inclusión y fueron empleados en los análisis sucesivos. En el caso de los resultados primarios (incidencia y mortalidad por cualquier tipo de cáncer) nuestros análisis incluyeron información sobre 1075159 y 2325378 individuos diagnosticados con alguna de las enfermedades del SNC y sus respectivas tasas de incidencia o mortalidad por cáncer. En general, los resultados de los metaanálisis efectuados sugieren una disminución significativa en la incidencia y mortalidad por cáncer en los pacientes con cualquiera de las enfermedades neurodegenerativas estudiadas. Se identificaron 7 estudios con información relativa a la incidencia de cáncer en pacientes previamente diagnosticados con EA. El metaanálisis de estos estudios sugiere que los pacientes con EA presentan una disminución del 31% en el riesgo de cáncer en comparación con los grupos control (RR = 0.69; 95% IC: 0.59-0.81). Seis estudios presentaron información sobre las asociaciones entre EA y la mortalidad por cáncer. Su análisis conjunto indica una reducción del 55% en la mortalidad por cáncer en pacientes previamente diagnosticados con EA (RR = 0.45; 95% IC: 0.33-0.61). En esta misma línea, los resultados de EP indican una reducción tanto de la incidencia (n = 17; RR = 0.85; 95% IC: 0.76-0.96) como de la mortalidad (n = 15; RR = 0.69; 95% IC: 0.54-0.87) por cualquier tipo de cáncer. Los pacientes con EH presentaron también un riesgo significativamente reducido en la incidencia y mortalidad subsiguientes por cualquier tipo de cáncer, aunque en este caso el número de estudios

identificado fue escaso, con 4 estudios disponibles en el metaanálisis de incidencia y 2 en el de mortalidad.

No se observaron alteraciones en el riesgo de incidencia de cualquier tipo de cáncer para TB, EZF o TEA. En cambio, se observó un incremento significativo en el caso de DM (RR = 1.17; 95% IC: 1.05-1.29) en el metaanálisis efectuado a partir de la información de 22 estudios observacionales. La mortalidad por cualquier tipo de cáncer presenta un incremento significativo en TB (n = 9; RR = 1.09; 95% IC: 1.03-1.15), EZF (n = 37; RR = 1.34; 95% CI: 1.23-1.46; I2 = 94.53%), TEA (n = 2; RR = 1.92; 95% IC: 1.58-2.32) y DM (n = 21; RR = 1.24; 95% IC: 1.15-1.34).

Varios tipos tumorales específicos presentaron patrones asociación significativos con distintas enfermedades del SNC. Entre ellos cabe destacar la reducción en el riesgo de cáncer de pulmón observada en pacientes con EA (n = 4; RR = 0.81; 95% IC: 0.7-0.94) y EP (n = 13; RR = 0.6; 95% IC: 0.47-0.76), las reducciones en el riesgo en cánceres hepáticos (n = 3; RR = 0.72; 95% IC: 0.61-0.85) y melanomas (n = 3; RR = 0.81; 95% IC: 0.69-0.94) observadas en EA, así como la disminución en la incidencia de cánceres de vejiga (n = 8; RR = 0.73; 95% IC: 0.57-0.93) y colorrectales (n = 15; RR = 0.8; 95% IC: 0.7-0.91) presentes en pacientes con EP. Además, los pacientes con EH presentaron una reducción significativa en el riesgo de cánceres colorrectales (n = 3; RR = 0.5; 95% IC: 0.27-0.93) y de próstata (n = 3; RR = 0.36; 95% IC: 0.25-0.5). Finalmente, los hombres con EZF presentaron una menor incidencia de cáncer de próstata (n = 14; RR = 0.56; 95% IC: 0.47-0.65).

Nuestros resultados también sugieren la presencia de incrementos en la incidencia de determinados tipos tumorales en pacientes con enfermedades del SNC. Entre los que cabe destacar aquellos observados entre EP y los cánceres cerebrales (n = 7; RR = 1.5; 95% IC: 1.11-2.04; I2 = 64.49%) y melanomas (n = 13; RR = 1.49; 95% IC: 1.17-1.89), DM y cánceres de mama (n = 15; RR = 1.24; 95% IC: 1.02-1.52) y pulmón (n = 9; RR = 1.31; 95% IC: 1.18-1.44), así como los incrementos en el riesgo de cánceres de mama (n = 18; RR = 1.37; 95% IC: 1.23-1.53) y útero presentes en mujeres con EZF (n = 11; RR = 1.35; 95% IC, 1.07-1.7).

La interpretación de estos resultados debe de tener en consideración las limitaciones inherentes a este estudio. En primer lugar, el número de trabajos dedicados al análisis de cada una de las asociaciones es altamente heterogéneo con una carencia evidente de estudios dedicados al análisis de los pacientes de TEA y EH. Estas carencias son especialmente relevantes en el caso del estudio de los resultados secundarios (incidencia y mortalidad por tipos tumorales específicos). En segundo lugar, el efecto de determinadas variables de confusión, tales como los hábitos de consumo de tabaco no son consideradas en la inmensa mayoría de los estudios observacionales identificados. Finalmente, la presencia de otras fuentes de sesgo, tales como el diagnóstico eclipsado y el sesgo maestral tampoco pueden ser excluidas.

Resultados y conclusiones del objetivo 2:

Ciento noventa y dos conjuntos de datos trasctriptómicos únicos, obtenidos mediante arrays de expresión, con muestras derivadas de tejidos procedentes de individuos con alguna de las 29 enfermedades estudiadas y sus respectivos controles sanos, fueron identificados tras efectuar las búsquedas pertinentes en repositorios públicos de datos ómicos. Después de aplicar los criterios de inclusión y la detección de muestras y estudios atípicos, 160 estudios incluyendo un total de 16132 muestras fueron seleccionados para su análisis subsiguiente. Además, diecisiete conjuntos de datos derivados de estudios efectuados mediante técnicas de secuenciación de RNA con muestras de 15 de los tipos tumorales estudiados mediante arrays de expresión, fueron empleados como cohorte de validación. Estos incluyeron un total de 7361 muestras. Los metaanálisis de expresión diferencial llevados a cabo arrojaron un número variable de genes diferencialmente expresados (GDE) en los distintos desórdenes estudiados (FDR < 0.05). Las enfermedades neurodegenerativas y el TEA presentaron un elevado número de GDEs que oscilaron entre los 1003 identificados en metaanálisis de TEA y los 4504 observados en EH. Por el contrario, las enfermedades neuropsiquiatrías (TB, DM, EZF) presentaron un número reducido de GDEs que oscilaron entre los 3 observados en EZF y los 15 identificados en DM. En el caso de los distintos tipos tumorales incluidos, el número de GDEs varió entre los 581 observados en AML y los 9757 detectados en tumores cerebrales.

Se observaron distintos patrones de asociación transcriptómica, tanto directos como inversos entre las enfermedades del SNC y los distintos tipos tumorales estudiados. En concreto, entre las asociaciones observadas en los análisis efectuados mediante arrays de expresión que fueron posteriormente validadas empleando estudios de RNA-seq, encontramos asociaciones directas entre EA, EP y EH, TEA y cánceres cerebrales, así como asociaciones directas entre EP, HD, TEA y cáncer de riñón. En estas asociaciones los genes tienden a estar desregulados en la misma dirección en cada pareja de desórdenes. Además, también se observó un número importante de patrones de desregulación en direcciones opuestas, que incluye las asociaciones transcriptómicas inversas observadas entre EA, EP, y EH y los cánceres de mama y pulmón, las asociaciones inversas entre EA, EP y cáncer de próstata y las asociaciones inversas identificadas entre EA y los cánceres hepáticos y de vejiga. En estos casos, aquellos genes sobreexpresados o infraexpesados en una de las enfermedades tienden a estar desregulados en direcciones opuestas en la otra enfermedad. Los análisis de enriquecimiento en categorías funcionales y los análisis de módulos de coexpresión consenso apuntan a la implicación conjunta de varios procesos biológicos en las enfermedades del SNC y el cáncer que incluyen alteraciones en el sistema inmune, señalización a través de P53, respuesta a proteínas no plegadas, señalización a través de MTOR C1, reparación del DNA, ciclo celular, apoptosis y síntesis de energía mediante la fosforilación oxidativa, entre otros.

Resultados y conclusiones del objetivo 3:

No se obtuvieron solapamientos significativos a nivel del interactoma en los módulos de enfermedad entre las enfermedades del SNC y el cáncer en ninguno de los análisis efectuados en los que se emplearon tres interactomas distintos y listas de genes y variantes asociados a enfermedad, construidas empleando criterios de inclusión restrictivos y relajados. Por el contrario, sí que se detectaron solapamientos significativos entre parejas de enfermedades del SNC y parejas de distiontos tipos tumorales que incluyeron, entre otros, aquellos observados entre TB y DM, DM y EZF, cánceres cerebrales y de pulmón y cáncer de pulmón y de ovario. Por su parte, el cómputo de correlaciones genéticas basadas en estudios de GWAS produjo correlaciones genéticas significativas entre enfermedades del SNC y tipos tumorales específicos, que incluyeron las correlaciones negativas observadas entre TEA y los cánceres de próstata (r_g = -0.16, p-val = 7.30e-03) y mama (r_g = -0.1, p-val = 1.60e-02), además de las correlaciones genéticas positivas observadas entre TB y cáncer de mama (r_g = 0.11, p-val = 6.00e-03), DM y cánceres de mama (r_g= 0.09, p-val = 4.02e-05) y pulmón (r_g = 0.28, p-val = 5.00e-04). EP presentó correlaciones genéticas positivas con melanoma (r_g = 0.14, p-val = 4.00e-02) y cáncer de próstata (r_g = 0.09, p-val = 3.16e-02). Finalmente, se observaron correlaciones genéticas positivas entre EZF cánceres de mama (r_g = 0.14, p-val = 1.75e-08) y ováricos (r_g = 0.12, p-val = 0.04). Estos resultados sugieren la presencia de variabilidad genética compartida asociada a una modulación conjunta del riesgo en determinadas parejas de enfermedades del SNC y tipos tumorales específicos. No obstante, la mayor parte de las correlaciones genéticas observadas son tenues y las repercusiones potenciales de estos hallazgos a la hora de modular las asociaciones epidemiológicas descritas deberá ser objeto de futuras investigaciones.

Resultados y conclusiones del objetivo 4:

El estudio del impacto de los fármacos indicados para el tratamiento de las enfermedades del SNC y el cáncer en la modulación de las relaciones de comorbilidad observadas entre ambos grupos de enfermedades sugiere que miembros de las familias de los inhibidores selectivos de la acetilcolinesterasa, anticonvulsivos, antipsicóticos, antidepresivos y benzodiacepinas, entre otros, tienen la capacidad de producir cambios transcripcionales que asemejan o podrían revertir aquellos observados en distintos tipos tumorales. Por ejemplo, los perfiles de expresión generados mediante el tratamiento de líneas celulares con dos bloqueadores de los canales de sodio, lamotrigina y carbamazepina, presentaron correlaciones negativas con los perfiles de expresión diferencial de los cánceres de próstata e hígado. Es tratamiento con galantamina, un inhibidor selectivo de la acetilcolinesterasa empleado en el tratamiento de EA produce alteraciones transcriptómicas en líneas celulares opuestas a las observadas en tumores colorrectales, estomacales y pancréaticos.

Por el contrario, el tratamiento con otros fármacos como el ácido valproico produce cambios de expresión en líneas celulares semejantes a aquellos observados en distintos tipos tumorales incluyendo los cánceres cervicales, hepáticos y pancreáticos.

Palabras clave: comorbilidad, enfermedades del sistema nervioso central, cancer, revisiones sistemáticas, metaanálisis, transcriptómica, interactoma, correlaciones genéticas, medicacion

Chapter 1. General introduction and objectives of this thesis

The study of disease-disease associations is becoming an increasingly important topic in biomedical research. The impact of comorbidity, especially in the context of population ageing, and the challenge it represents for practitioners and healthcare systems, are compelling the scientific community to deeper characterize this phenomenon from multiple perspectives [1]. Epidemiology provides insights into how disorders relate to each other, estimates their association's magnitude, and furnishes information about the risk factors contributing to them. Complementary, biomedical and molecular research have the potential to explore the underlying biological causes of comorbidity, which comprise the presence of shared genetic variability and joint alterations in genes and pathways, as well as the possible role of specific medications.

Cancer and central nervous system (CNS) disorders are among the top leading causes of death and disease burden worldwide. The last decades have witnessed the accumulation of observational studies addressed to explore CNS and cancer associations, which have often reported inconsistent findings. Consequently, evidence synthesis methods such as systematic reviews and meta-analysis have started to be applied to summarize the available data and appraise its quality. These ideas have been integrated into the theoretical framework of evidence-based medicine.

Furthermore, the emergence of the omics era has led to the unprecedented accumulation of data regarding disease's molecular bases, opening the door to the study of comorbidity from a molecular perspective through the identification of joint alterations in variants, genes, and biological processes.

In the following sections, we introduce the concept of comorbidity and its causal mechanisms and describe the tools used to study it from both an epidemiological and a molecular perspective.

1.1 The concept of comorbidity and its related constructs.

Alvan Feinstein first coined the term comorbidity in an article published in 1970 [2]. He initially formalized it as *"Any distinct additional entity that has existed or may occur during the clinical course of a patient who has the index disease under study"***.** Feinstein's seminal definition implied a reference disorder linked to other secondary conditions that could take place at different points in the temporal dimension of the patient's life. In his view, at that time, clinical sciences were paying little attention to disease-disease inter-relationships and their effect on individuals. He also expressed his preoccupation with the potential impact that comorbidity could produce in clinical trials' outcomes if information regarding it was neglected.

Feinstein classified comorbidities in the following groups: *ancestral* if a preexisting disease was converted into the index disease, *supervening* if the secondary condition was caused by a new pathologic process that was predisposed by anatomic effects of the index disease at the primary site, and *derivative* if the comorbid disorder was produced by the dissemination of the same pathologic process present in the index disease.

Fifty years later, new concepts linked to the idea of comorbidity have emerged, including *multimorbidity*, the *burden of disease*, and *frailty*, among others. Each term provides a slightly different conceptualization of the general idea of disease-disease association that differs in the description of some of its features (i.e., the nature of the health conditions, their relative importance, and their chronology) and is used preferentially by different sets of professionals, including researchers, physicians, and policymakers.

Expanded conceptualizations have also been created. An important example is the one provided by the *comorbidity burden* [3], which is defined as the total amount of disease load present on an individual. Different measures to quantify it have been developed, including the Charlson's Index [4], the Cumulative Illness Rating Scale (CIRS) [5], and the Index of Coexisting Disease (ICED) [6]. Finally, the term *patient's complexity* [7] adds to the comorbidity burden idea the effects of several factors (i.e., socioeconomic, cultural, environmental, behavioral).

The comorbidity picture has been completed in recent years with the introduction of the term *inverse comorbidity*. Contrary to *direct comorbidity*, which is defined as a higher-than-expected probability of disease occurring in individuals who have been diagnosed with other medical conditions, *Inverse comorbidity* is defined as a lower-than-expected probability of disease occurring in individuals who have been diagnosed with other medical conditions [8]. A close examination of of both direct and inverse comorbid associations could help to better understand the interplay between disorders and to gain insight into each disorder's physiopathological processes [9].

Some authors have pointed out that the heterogeneity of the available concepts is confusing and have advocated for a systematic reformulation of these ideas. Interesting reviews of this topic are available in references [10] and [3]. One of the causes of ambiguity is that most of the available terms can be seen from two different perspectives (i.e., the individual and the population). For instance, in some disciplines such as epidemiology, comorbidity is observed from a population rather than a patient's perspective. In this context, the critical element is to determine if a given pair of disorders co-occur in populations in frequencies that deviate from what is expected by chance. In contrast, from the medical practitioner's point of view, comorbidity is a patient-centered phenomenon in which chronology is essential and directly influences patient management, treatment selection, and prognosis. Linked to the problem of conceptualization, it is also important

to note that epidemiological studies can present different designs. In this context, there is a clear distinction between *prevalence* and *incidence* studies. In prevalence studies, exposures and outcomes are measured simultaneously, disregarding aspects, such as the chronology in which conditions appear. In contrast, incidence studies focus on new cases of a particular outcome in a specific population allowing researchers to study causal associations.

1.2 Potential causes of direct and inverse comorbidity

Comorbidities can arise as a consequence of diverse causes [3]. First, two or more disorders can be observed in the same patient simply by *chance*. Comorbid association due to chance can be identified at a population level when the frequency of their joint manifestation does not deviate from the product of the individual frequencies. The second mechanism by which comorbidities can be observed is the presence of *systematic biases* in epidemiological studies. For instance, selection bias arises because patients seeking medical attention present an increased likelihood of receiving a diagnosis of additional conditions. In other instances, a particular disorder's specific features could make patients less prone to seek medical care. This could result in the observation of a negative association between a given pair of diseases. Besides, systematic biases also include the lack of account for important confounding factors in epidemiological research, which could distort the results of observational studies and meta-analysis. In the cases of chance and systematic bias, the observed comorbidities do not indicate shared etiological factors.

Comorbidity can also emerge due to *causal associations* that can be divided into different categories. First, *direct causation* implies that disease A is responsible for increasing or decreasing the likelihood of developing disease B. Mechanisms involved in direct causation include shared genetic variability, such as in the case of the direct association found between type I diabetes and celiac disease [11], joint alterations in higher levels of biological organization, such as pathways**,** cell types**,** tissues, or organs, and pharmacological treatment's potential effects. The decreased risk of colorectal cancer observed in patients treated with high doses of selective serotonin reuptake inhibitors and [12] and the reduced cancer mortalities reported in patients with alcohol addiction treated with disulfiram [13] provide two examples of the latter. In the second place, causal links can also be due to the presence of *associated risk factors*. Environmental or behavioral correlated risk factors can be the cause of disease-disease associations. In this model, risk factors linked to disease A are correlated to disease B's risk factors, increasing the risk of both disorders. For instance, smoking and alcohol consumption are correlated risk factors that increase the risk of chronic obstructive pulmonary disease and liver disease, respectively. In other instances, risk factors are not correlated to each other, but each one can modulate the risk of specific disorders linked to the other. For instance, smoking and age are not correlated but are independent risk

factors for both cardiovascular disease and cancer. Besides, the presence of disease A could modulate the exposure of risk factors for disease B. For instance, neuropsychiatric patients are at a higher risk of being heavy smokers, increasing the likelihood of certain types of cancer, such as lung cancer. Finally, it is also possible that a given pair of associated diseases are not directly linked but are both due to the presence of a third disease.

1.3 The study of comorbidity using systematic reviews and meta-analyses of epidemiological data

The first known reported observation about the association of two conditions is due to Hippocrates, who noted that fever alleviated psychotic disorders [14]. Since this early finding, the study of comorbidity gradually developed with significant advances concentrating in the second half of the twentieth century. In 1909 the Commissioners of Lunacy for England and Wales reported that psychiatric patients appeared to be relatively immune to cancer, which constituted the first known association between a central nervous system (CNS) disorder and cancer [15]. A gap of more than 70 years separates this early event from the first epidemiological study on the topic in which Shekelle and co-workers observed that depressed patients presented a twofold increase of death from cancer [16].

Posterior decades have witnessed a gradual increase in the number of published studies aiming to characterize the associations between CNS disorders and cancer and the improvement of their sample sizes and statistical methods. Nevertheless, results have often led to contradictory conclusions, probably, as a consequence of the heterogeneity present in study designs and the limitations imposed by this particular research question, which relies on observational studies and cannot be addressed using more sophisticated forms of medical research such as randomized controlled trials [17].

The inconsistencies mentioned above claimed for methods that allowed researchers to synthesize and appraise the available evidence. In this context, systematic reviews and meta-analyses were introduced as analytical tools that allow to summarize the available data from a particular research question.

British statistician Karl Pearson published the first report including methods for combining the outcomes from different studies. In his work about enteric fever inoculation, he synthesized correlation estimates derived from different studies and provided combined significance measures. In the 20's Ronald Fisher, who was working on agricultural research, further developed methods for study combination and introduced the ideas of publication bias and heterogeneity. Later, one of Fisher's co-workers, William Cochran, created the theoretical framework of random-effect models. Gene Glass coined the term meta-analysis in 1976 to define the integrative statistical

analysis of an extensive collection of results derived from various individual studies. Soon after, meta-analytical methods started to be applied in clinical research. For instance, through a quantitative combination of the results of previous trials Peter Elwood and Archie Cochrane observed that patients treated with aspirin were at a reduced risk of heart attack [18].

Since then, evidence synthesis methods have continued to evolve and have been integrated into an epistemological framework known as evidence-based medicine, emphasizing the use of evidence derived from well-designed research to inform and optimize decision-making. One core element of the field is the classification of evidence according to its epistemological strength and the requirement that only the most substantial types, which are those derived from systematic reviews, meta-analysis, and randomized control trials, are employed to guide the clinical decision.

A systematic review can be defined as "the application of strategies that limit bias in the assembly, critical appraisal, and synthesis of all relevant studies on a specific topic" [19]. The concept meta-analysis refers to the use of statistical methods that combine the results of multiple scientific studies. Since meta-analysis methods applied to the integration of observational and transcriptomic data are a core part of chapters 2 and 3 of this thesis, we will briefly review them. Therefore, we will present the two main statistical models employed in a meta-analysis, Fixed Effect Models (FEM) and Random Effect Models (REM). The variations of these methods developed to analyze transcriptomic data will be described in **Section 1.5**.

Fixed effect models (FEM)

Fixed effects models assume that the studies' effect sizes included in a particular meta-analysis are drawn from a distribution with true effect size μ and variance σ^2. Therefore, each effect size O_i can be decomposed as the sum of two terms: $O_i = \mu + \epsilon_i$, where μ is the true underlying effect, an ϵ_i is an error term which measures how much O_i deviates from μ (**Figure 1A**). The pooled effect computed through FEM is an estimate of this true underlying effect μ. Under FEM weights are assigned to each study based on the inverse of the variance method. Therefore, the weight of study i, denoted W_i, is defined as the reciprocal of its variance (**Figure 1C eq.2**). The pooled effect size estimate is then computed, as shown in **Figure 1C eq. 3**, which involves a quotient between the summation of the effect sizes multiplied by their respective weights the weights' summation. The pooled effect size variance is computed as the reciprocal of the summation of weights (**Figure 1C eq. 4**), and the standard error as the square root of the variance **Figure 1C eq.5**. The confidence intervals for the estimated pooled effect size are finally computed from the standard errors (**Figure 1C eq. 6 and 7**).

Random effects model (REM)

Contrary to the fixed effects model, the random effects model allows the true underlying effect to vary between studies. In other words, instead of a single true effect size, we start from a distribution of true effect sizes. This model better captures the inherent heterogeneity of observational studies, which do not present identical designs and often differ on the disposal of different covariates (i.e., the age and gender proportion of the studied population, the instruments employed to carry out diagnosis, the smoking status). Those covariates could influence the magnitude of the observed effect sizes. Therefore, the combined effect obtained when applying REM represents the mean of the distribution of true effects. Let's assume O_i to be the observed effect size derived from a particular study. O_i is determined by the true effect θ_j plus the within-study error ε_i, and θ_j is, in turn, determined by the mean of all true effects μ, and the between-study error ζ_j (**See Figure 1B**). Thus, any observed effect O_i can be decomposed as depicted in **Figure 1D eq.1**.

As we have seen, under the random-effects model, there are two levels of sampling and two levels of error. The true effect sizes θ are distributed arround μ with variance τ^2 and the observed effect O_i for a given θ_j is distributed around θ_j with variance σ^2. σ^2 depends on the sample size of that particular study. In the process of estimating μ, we will need to deal with both sources of sampling error. Therefore, an important aspect of REM is to decompose the observed variance into its two components, within-studies, and between-studies variances. This is achieved by computing the total variance (observed variance) and the within-studies variance (σ^2). Then, the between-studies variance (τ^2) is computed as the difference between the two previous values.

The total variance, denoted as $Q$is computed according to the formula displayed in **Figure 1D eq.2**. In short, the sum of the squared deviations of each study O_i to the combined mean $\bar{O}$**.** is computed and weighted by the inverse of the variance (W_i) of each study. According to this equation, larger studies deviating from the mean will have more impact on Q than smaller studies.

If the only source of variance contributing to the total observed variance was within-study error, then the expected value of Q would be equivalent to the degrees of freedom for the meta-analysis, which is defined to be the number of included studies minus one. $df = (Number\ of\ Studies) - 1$. Therefore, we can compute the between-study variance τ^2, as depicted in **Figunre 1. D) eq.3.** The numerator $Q - df$, represents the excess variance, and the denominator C is a scaling factor that deals with the fact of Q being a weighted sum of squares and ensures that τ^2 is computed in the same metric as the within-study variance (**See Figure. 1D eq. 4**).

Next, each study is weighted based on the inverse of its variance. As we have already seen, under REM, the variance will include the terms linked to both within-studies and between-studies

variances. **Figure 1D eqs. 5 6, and 7** are used in order to calculate the combined effects under random-effects models weighted by their variance.

Finally, the variance and the standard error of the combined effect are computed, as shown in **Figure 1D eqs. 8 and 9,** which will be in turn used to compute the confidence intervals as depicted in **Figure 1D eqs.10 and 11**.

In the present thesis, random effect models were selected *a priori* to synthesize the epidemiological evidence of the associations between CNS disorders and cancer since high levels of heterogeneity were expected due to the presence of differences in study designs, diagnosis methods, and adjustment for confounding variables, among other factors.

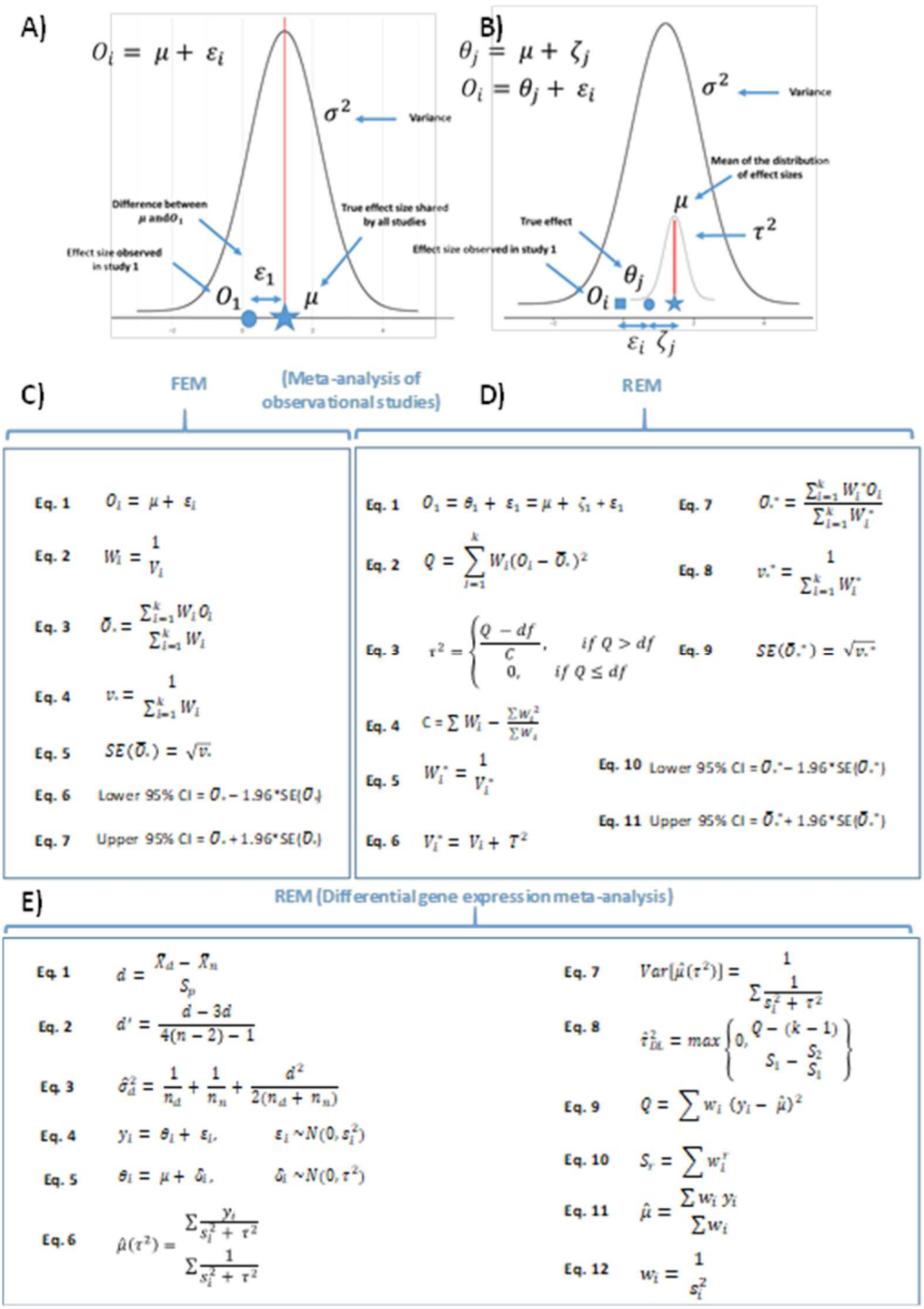

Figure 1: Equations used to carry out fixed and random effects model meta-analyses in Chapters 2 and 3. A) Fixed effects model design. B) Random effects model design. C) Equations used for FEM meta-analysis of observational studies. D) Equations used REM meta-analysis of observational studies E) Equations used for REM differential gene expression meta-analysis.

In 1994, McGee and co-workers reported the first systematic review and meta-analysis aimed to summarize the available evidence of the associations between a CNS disorder and cancer. They described a positive but not significant link between depression and later development of cancer [20]. However, it was not until the late 2000's when most systematic reviews and meta-analyses focusing on this topic started to appear.

To date, the work of Catalá *et al.* [21] constitutes the most exhaustive systematic review analyzing the epidemiological associations between CNS disorders and cancer. In summary, the authors observed that CNS disorders were associated with a decrease in overall cancer risk, which was especially evident in the case of patients with neurodegenerative disorders. In addition, some CNS diseases were also found to be associated with an increased risk of several site-specific cancers (**See Section 2.1**). This fact, together with the lack of sufficient data characterizing all possible CNS and site-specific cancer associations, has compelled the scientific community to keep characterizing this phenomenon with the publication of new observational studies that should be integrated with previous data through systematic reviews and meta-analysis.

1.4 The study of comorbidity from a molecular perspective

The history of the molecular study of disease starts in 1949 with the identification of Sickle cell anemia as a molecular disorder by Linus Pauling and collaborators [22]. They observed that the structural features of sickle cells were due to the presence of an abnormal form of hemoglobin. Soon after their discovery, Pauling claimed:

'…our structural chemistry and understanding of molecules is getting to the point in which it should be of assistance in converting medicine into a real science'.

He was right. His work served to inaugurate a new research field that was the germen of what is currently known as biomedical research. After his discovery, unprecedented advances in the area of molecular biology propelled the molecular study of disease. DNA structure was solved in 1953 by Watson and Crick thanks to the DNA crystal structure obtained by Rosalind Franklin, Arthur Kornberg isolated DNA polymerase in 1955, and Crick proposed the central dogma of molecular biology in 1958. Soon after, in 1961, Crick, working with Sydney Brenner, suggested that each amino acid was encoded by a group of three nucleotide bases that were called a codon. Messenger RNA (mRNA) was discovered in 1961, as well as the function of the ribosome as a protein building machine.

Technical advances played a pivotal role in the discoveries that took place in further decades. Sanger and co-workers developed the chain termination sequencing method between 1975 and 1977. In 1985 Kray B. and Mullis created the polymerase chain reaction technique (PCR). Things scaled up from the nineties onwards after the completion of the human genome project in

2004, and the emergence of several analysis methods such as the quantitative PCR (qPCR) and the irruption of techniques that produced genome-scale information such as gene expression microarrays, comparative genomic hybridization (CGH) arrays, methylation arrays, and the more recent next-generation sequencing techniques.

Those breakthroughs have profoundly impacted the understanding of biology and disease and are at the core of the transition of the study of diseases as single entities to the study of the molecular bases of comorbidity [23].

The first genomic regions linked to disease date from the 80s. The application of linkage analysis allowed researchers to map the gene responsible for Huntington's disease in 1983 [24], and the Duchenne muscular dystrophy and cystic fibrosis genomic locations were identified in 1987 and 1989, respectively [24, 25]. In the case of Huntington's disease, the transition from linkage studies to the effective identification of the underlying pathogenic mutations took about a decade [26].

It has been estimated that in the period comprehended between 1980 and 2000, the molecular bases of about 1000 Mendelian disorders were identified. These numbers were multiplied by three in the first decade after the completion of the Human Genome Project [23].

The introduction of next-generation sequencing techniques also greatly impacted the research of the molecular bases of disease. For instance, only one year after the introduction of the exome sequencing techniques in 2009, the genetic causes of Miller's syndrome were established [27].

The study of the molecular bases of complex disorders has proven more challenging. Advances in this field have been driven by the technical and conceptual developments in high-throughput genotyping technologies and improvements in the knowledge of the patterns of linkage disequilibrium (LD) present in human populations [28]

At the present time, the GWAS catalog, a database of published genome-wide association studies, contains 138312 phenotype variant associations retrieved from 3989 publications. However, in the case of complex diseases, the discovered variants usually confer only a small increase in risk. In general, GWAS has failed in identifying disease genes presenting large effects. The concept of *missing heritability* describes the fact that single genetic variation only accounts for a small fraction of the heritability of complex disorders [29]. Some potential causes have been invoked to explain this phenomenon, including the role of rare and structural variants that are not represented in GWAS studies [29] and the inaccuracy of heritability estimates due to the effect of gene-gene interactions [30].

As we have seen, the development of molecular biology and the omics methods have entailed the description of genotypic and molecular features of diseases. In the last fifteen years, a

shift from the study of single diseases to disease-disease associations from the molecular point of view has also started to take place powered by data accumulation regarding the molecular bases of individual disorders. Many open-access platforms that store gene-disease association, variant-disease association, and omic data repositories are now available online, including the Online Mendelian Inheritance in Man (OMIM), The Genome-Wide Association Studies Catalog (GWAS catalog), the Gene Expression Omnibus (GEO), Array Express (AE) and The Cancer Genome Atlas (TCGA).

In 2007 Goh and co-workers presented the first large-scale study design to uncover disease-disease associations based on molecular data [31]. Using the information about disease-associated genes available in OMIM, they created a bipartite graph containing two kinds of nodes (diseases and genes). The resulting graph was called human diseaseasome. They observed that it was highly interconnected, with most disorders having a link to at least another disease. In addition, links between diseases were not random, and diseases with similar pathophysiology tended to form clusters. Genes associated with the same disease were found to encode proteins that were more likely to interact with one another, shared more gene ontology terms, and were prone to be expressed in the same set of human tissues.

Several exiting extensions of the human disease network have been developed. Based on the disease module hypothesis, which models a disease as a result of the perturbation in a specific functional module, and using interactome-based distances and sets of genes linked to disease, Barabási and co-workers determined that diseases with overlapping network modules presented significant co-expression patterns, symptoms similarity, and comorbidity values [32]. Another interesting instance is the creation of the drugome, which consists of a bipartite network combining the set of available drug chemicals and their target genes [33]. The study of the drugome, in combination with the diseasome, provided some interesting insights. The authors noted that available drugs tend to concentrate in specific regions of the diseasome and that the majority of existing drugs target genes far away from the disease-causing genes, which suggests that most of them were developed for palliative treatment rather than curative. This approach also opened a door for interesting practical applications such as drug repurposing.

In a follow-up article published in 2012, Goh and co-workers proposed a roadmap to the complete diseasome that should integrate several types of data, including information derived from transcriptomic studies, disease-drug interactions, the influence of environmental factors, metabolic links, interactome-based measures of association, and genetic links [30].

Despite the importance of the studies mentioned above, most molecular approaches to the study of disease-disease associations have neglected the idea of inverse comorbidity.

Only a handful of works addressing the topic of CNS and cancer associations from the molecular perspective have been published to date. In 2014 Ibañez and co-workers reported that epidemiological data describing inverse comorbid patterns of association between three CNS disorders (schizophrenia, Alzheimer's disease, and Parkinson's disease) and three cancer types (colorectal cancer, lung cancer, and prostate cancer) translated into opposite patterns of differentially expressed genes [34]. In other words, according to the authors, genes that tended to be upregulated in CNS disorders were found to be downregulated in cancer and vice versa. Altered expression in genes such as *PIN1* and *ATP13A2* and pathways, such as P53 signaling, Wnt, and folding and protein degradation pathways were identified as potential agents modulating some of the reported associations.

More recently, in 2017, using a similar methodology, Sanchez and co-workers identified potential molecular substrates that could underlie both the direct and the inverse epidemiological associations found between AD and glioblastoma and AD and lung cancer, respectively [35]. Their results suggested that immune system-related processes and mitochondrial metabolism could constitute the molecular substrates of the direct epidemiological associations observed between AD and glioblastoma and the inverse epidemiological associations between AD and lung cancer, respectively.

As Ibañez and Sanchez work, **Chapter 3** of the present thesis is also based on differential gene expression meta-analysis methods. Therefore, we will briefly introduce them.

Differential gene expression meta-analyses

The differential gene expression analyses carried out in this thesis are based on Choi's method [36]. For each available transcriptomic dataset, the differences in gene expression between cases and controls are computed using an estimator for the standardized mean denoted as d, which is computed as shown in **Figure 1E eq. 1**, where $\bar{X}_d$ and $\bar{X}_n$ represent the gene expression means of the diseased tissue and the healthy tissue, respectively, and S_p is an estimated of the pooled standard deviation. An unbiased estimate of d, denoted as d' is obtained using the following **Figure 1E eq.2** formula and its variance computed as shown in **Figure 1E eq.3**. Where n_d and n_n are the disease and control group sample sizes, respectively. As in the case of the meta-analysis of observational studies, a random-effects model was selected to combine the outputs of the available study since high between-study heterogeneity was expected. Let y_i be the observed effect size for independent studies $i = 1,2, \dots, k$. The observed y_i is derived from the model presented **in Figure 1E eqs. 4 and 5**. Where μ is the overall mean (i.e., the average measure of differential expression across the datasets for each gene), τ^2 represents the between-study variability, and s_i^2 is the sampling error associated with the ith study. In this context, y_i and s_i^2 are given by d' and $\hat{\sigma}_d^2$. As

we have seen under FEM, the differences in effect sizes observed between different studies are due to sampling error alone. Therefore, under FEM, τ^2 = 0. In consequence, $y_i \sim N(\mu, s_i^2)$. On the other hand, under REM, each effect size is supposed to be drawn from a distribution with mean θ_i and variance s_i^2, in turn, each θ_i is assumed to be drawn from a superpopulation with overall mean μ and variance τ^2. **Figure 1E eqs. 6 and 7 show** the equations used for estimating $\hat{\mu}(\tau^2)$ and its variance ,$Var[\hat{\mu}(\tau^2)]$, respectively. Both equations depend upon the computation of τ^2. τ^2 is estimated by applying the method of moments developed by DerSimonian and Laird presented in **Figure 1E eq. 8** and is denoted as $\hat{\tau}_{DL}^2$ which in tur is based on Q, a widely used statistic for between-study homogeneity computation proposed by Cochran in 1954 (**See Fiugre 1E eq. 9**) and S_r (**Figure 1E eq. 10**). The computation of Q and S_r depend upon the values of w_i and $\hat{\mu}$. $\hat{\mu}$ is the the weighted least squares estimator, which ignores between-study variance (**Figure 1E eq. 11**) and w_i is defined as depicted in **Figure 1E eq. 12**. Finally, the Z statistic is computed as the ratio of $\hat{\mu}(\tau^2)$ over its standard error and the significance of the Z statistics is computed by the generation of empirical distributions by random permutations.

Finally, other approaches have also been developed to explore CNS and cancer associations from the molecular perspective. An interesting instance is the one provided by the work of Jane Driver and co-workers [37]. Making use of GWAS summary statistics and a method called cross-trait LD score regression, they found significant genetic correlations between AD and breast and lung cancer [38], suggesting that shared genetic variability could be involved in the modulation of their comorbid associations.

1.5 Characteristics of the CNS disorders and cancers studied in this thesis

Epidemiological and molecular associations between seven CNS disorders (Alzheimer's disease (AD), autism spectrum disorders (ASD), bipolar disorder (BD), Huntington's disease (HD), major depression (MD), Parkinson's disease (PD), and schizophrenia (SCZ)) and cancer were investigated and integrated into the present thesis. This section provides a description of the included disorders. The selection of the diseases was based on their relevance as medical entities and the feasibility to perform the differential gene expression meta-analyses counterpart of the study presented in **Chapter 3** due to data availability constraints.

Alzheimer's disease

AD is the most prevalent neurodegenerative disorder of aging people and constitutes almost 70% of cases of dementia, with over 50 million affected patients worldwide [39]. Core AD symptoms include memory loss and cognitive impairment, but additional symptoms such as problems with language, disorientation, and mood changes are also present. Genetically, AD is divided into familial and sporadic cases. Familial AD represents less than 1%, and it has been linked

to autosomal dominant mutations in the amyloid precursor protein (*APP*), presenilin 1, (*PSEN1*), or presenilin 2 (*PSEN2*) [40]. By its side, sporadic AD constitutes more than 90% of cases. *APOE* has been identified as a susceptibility gene for AD. In particular, the presence of one of its three common polymorphisms (epsilon 4) is linked to an increased risk of AD. Both types of AD share a set of neuropathological manifestations, including the accumulation of misfolded and aggregated proteins, the deposition of senile plaques in the extracellular space formed by β-amyloid peptides and the intracellular formation of neurofibrillary tangles of which hyperphosphorylated tau protein is a major constituent [40]. The accumulation of Aβ aggregates eventually triggers a cascade of cellular changes, including mitochondrial oxidative damage, tau hyperphosphorylation, synaptic failure, and inflammation, which is associated with the loss of synapses and neuronal death, initially in focal areas including the entorhinal cortex and hippocampus and ultimately more broadly in the cortex [41].

Autism spectrum disorders

The estimated prevalence of ASD in developed countries is 1.5% [42]. ASD affects more than 24 million individuals worldwide [43]. Disease onset is situated around the second or third year of life, and it is four times more frequent in males than in females [42]. The heritability of the disease based on twin studies is 0.7 [44]. ASD symptoms include impairment in social interactions, communication deficits, patterns of restricted interests, and repetitive behaviors. In some cases, episodes of self-injury are also present [45]. ASD is often associated with other conditions such as intellectual disability, attention-deficit/hyperactivity disorder (ADHD), anxiety disorder, and epilepsy [45].

Two forms of ASD (syndromic and idiopathic) have been described. Syndromic ASD, which represents between 3% and 5% of ASD cases, is regarded as the behavioral manifestation of known monogenic disorders, including type 1 neurofibromatosis, fragile X syndrome, tuberous sclerosis syndrome, and Rett syndrome, among others, which are driven by mutations in the *NF1*, *FMR1*, *TSC1*, and *MECP2* genes respectively [46]. Idiopathic ASD represents more than 80% of cases and is linked to other genetic variability sources such as common variation and recurrent copy number variants [47]. Several cellular and tissular alterations have been observed in ASD patients, including an excess of neurons that lead to over connection of specific brain areas, impaired neuronal migration, unbalance in excitatory and inhibitory networks, and alterations in synapses and dendritic spines due to changes in adhesion molecules. Other biological processes such as peripheral and central nervous system inflammation with increased pro-inflammatory cytokines and microglial activation [48-50] have been linked to ASD. A proportion of ASD patients present macrocephaly and faster brain growth than control children. This event is often followed by regular or slower brain growth during childhood.

Bipolar disorder

Bipolar disorder is a neuropsychiatric disease characterized by the alternation of periods of depression with periods of mania or hypomania, which are defined as situations of abnormally elevated mood [51, 52]. Changes in energy and activity usually follow these mood transitions. The presence of mania or hypomania divides BD diagnosis in BD type I or BD type II, respectively. BD prevalence is around 0.6% for BD type I, and 0.4% for BD type II, and the median age of disease onset is 25 years [53]. The overall heritability of the disease has been estimated to be 0.71 [54]. Imbalances in neurotransmitters concentrations have been suggested to be an important factor in the disease, including alterations in dopamine and glutamatergic transmission [55-57]. Bipolar disorder is regarded to be a complex disease for which GWAS studies have only found variants presenting small effect sizes. In addition, BD GWAS studies have often shown inconsistent outcomes [58-60]. Alterations in several biological pathways have been linked to the disease, including glutamate receptor signaling and Wnt and Notch signaling, among others [61].

Major depression

According to DSM-VI [62], MD diagnosis requires the presence of at least five symptoms of the following list for a two weeks period: 1. Depressed mood most of the day (i.e., feels sad, empty, hopeless), 2. Markedly diminished interest or pleasure in almost all activities nearly every day, 3. Significant appetite changes or significant weight loss or gain, 4. Insomnia or hypersomnia nearly every day, 5. Psychomotor agitation or retardation, 6. Fatigue or loss of energy, 7. Feelings of worthlessness or excessive guilt, 8. Diminished ability to think or concentrate or indecisiveness. In the USA, the lifetime prevalence of MD has been estimated to be 16% [63], whereas worldwide estimates suggest that it affects more than 350 million people [64]. Classical hypotheses have attributed depression to functional imbalances and deficiencies in monoamine-series neurotransmitters, including dopamine, serotonin (5-HT), and norepinephrine (NE). According to it, the depletion of serotonin and norepinephrine in the synapse triggers the development of depressive symptoms. These views have been supported by the outcomes of treatments with selective serotonin reuptake inhibitors (SSRI) [65]. Several alternative and complementary hypotheses have also been proposed in order to explain some phenomena linked to depression that cannot be explained by the monoamine hypothesis [66]. In general, studies have reported limited success in finding MD-associated variants, although some genes have been linked to the disease, including serotonin transporter *SLC6A4*, Piccolo presynaptic cytomatrix protein (*PCLO*), and the 5-hydroxytryptamine (serotonin) receptor 2A (*HTR2A*), among others.

Huntington's disease

Historically known as Huntington's Chorea, HD is characterized by early symptoms such as mood problems and cognitive dysfunction, followed by a general lack of coordination and a random and uncontrollable movement at later stages [67]. As the disease progresses to its most advanced phases, it also leads to dementia [68]. HD prevalence has been estimated to be 2.71 per 100,000 (95% CI: 1.55-4.72) worldwide [69]. Disease onset takes place between the third and the fifth decade of life [68], with about 10% of the patients showing disease onset before the age of 20 [70]. HD is a monogenic disease caused by an autosomal dominant mutation in the Huntingtin gene (*HTT*) located on chromosome 4, which consists of an expansion of a cytosine-guanine-adenine triplet that results in an abnormal protein that triggers the cascade of events that result in neuronal damage. The repeated codon encodes for the amino acid glutamine, and its expansion generates a polyglutamine tract in the resulting protein. Sequences containing more than 36 glutamines produce proteins with altered properties. Thirty-six to thirty-nine repeats result in a reduced-penetrance form of the disease, whereas very large repeat counts are associated with a full penetrance form of HD. In addition, the length of the polyQ expansion is inversely correlated with the age of disease onset. The mutated version of the *HTT* gene first produces damage on the striatum and other cerebral areas, including the substantia nigra, layers 3, 5, and 6 of the cerebral cortex, and the hippocampus [67], and it is linked to the presence of alterations in several biological processes such as the mitochondrial function, immune system-related processes, high astrocyte counts and microglia activation [71, 72].

Parkinson's disease

PD is characterized by the presence of motor and non-motor symptoms such as resting tremor, postural instability, rigidity, bradykinesia, slight depression, constipation, fatigue, sleep disturbances, and hyposmia. Cognitive impairment increases as the disease progress [73]. It is the second more prevalent neurodegenerative disorder affecting more than 2% of the population older than 65 years [74]. Two types of PD, hereditary, and idiopathic have been identified. The first represents 15% of PD cases, and it is caused by mutations in specific genes such as *SNCA* (*PARK1*), *LRRK2* (*PARK8*), and Parkin (*PARK2*), among others [75, 76]. The second comprises most PD cases, and it is thought to be prompted by interactions between genetic and environmental risk factors. PD neuropathological hallmarks include dopamine reduction in the basal ganglia due to neuronal death in the Substantia Nigra pars compacta (SNpc) and the presence of abnormal intra-cytoplasmic deposits called Lewy bodies (LB), which are protein aggregates where the major component is alpha-synuclein associated with other proteins [77, 78]. Alterations in several biological processes in PD have also been reported, including mitochondrial dysfunction, oxidative stress, calcium and

protein homeostasis dysregulation, impairments in the ubiquitin-proteasome system, and lysosomal-mediated autophagy, and neuroinflammation, among others [73].

Schizophrenia

Schizophrenia is a neuropsychiatric illness defined by three groups of symptoms, positive, negative, and cognitive. Positive symptoms include delusion, disorganized thoughts and speech, and hallucinations. Negative symptoms are deficits in the normal emotional response, such as little emotion, poverty of speech, anhedonia, and lack of motivation. Cognitive abilities are also impaired in SCZ patients, with deficient working and long-term, verbal declarative memory, semantic processing, and attention insufficiencies. SCZ affects 1% of the population, and it is linked to poor vital outcomes. It presents a heritability estimate of 0.81 [79]. The dopaminergic hypothesis states that SCZ is characterized by alterations of the dopamine neurotransmission in the mesolimbic system, which would be responsible for the positive symptoms and in the mesocortical pathway that would cause negative symptoms. Structural brain alterations such as the enlargement of the third and lateral ventricles and the reduction in total brain and grey matter volume have also been reported in SCZ patients. GWAS studies have identified several common SCZ risk alleles presenting weak individual effects in SCZ risk (OR < 1.2). Based on these studies, SCZ has been established to be a polygenic disease involving a thousand common alleles [80]. Advances in the last decade have helped in the identification of de novo mutations, rare copy number variation (CNV), rare single point nucleotide variants, and small insertion/deletion and single nucleotide polymorphisms [81].

Cancer

Only in 2017, 24.5 million cases of cancer incidence and 9.6 cancer deaths were observed worldwide. Cancer incidence was found to be on the rise in the period comprehended between 2007 and 2017. The top five site-specific cancers ranked by incidence were nonmelanoma skin cancer, tracheal, bronchus, and lung, breast, colorectal, and prostate cancers [82].

Cancer is a heterogeneous disease, and a detailed description of the specific characteristics of the complete set of cancer types included in this thesis is out of the scope of this introduction. Thus, we will summarize the main characteristics of cancer pathogenesis.

In 2000, Douglas Hanahan and Robert A. Weinberg gathered the available evidence about the mechanisms involved in cancer pathogenesis and hypothesized that the variety of cancer manifestations would be due to six essential alterations in cell physiology, termed “The Hallmarks of Cancer” [83]. The *first hallmark* is based on the acquisition of cancer cells of self-sufficiency in growth signals. In healthy cells, growth is modulated by the presence or absence of exogenous growth factors (GFs). Cancer cells display different strategies to overcome this limitation. For instance, some cancer cells are able to synthesize the very growth factors to which they are responsive (e.g., glioblastomas produce platelet-derived growth factor PDGF). Cell surface

receptors, which are responsible for transducing the signals delivered by GFs, are also modified in cancer. For example, the HER2/neu receptor is overexpressed in a subset of mammary carcinomas. Finally, alterations in intracellular transducer systems, such as SOS-Ras-Raf-MAPK are also found in cancer. The *second hallmark* consists of if the acquisition of insensitivity to growth-inhibitory signals. Healthy cells are responsive to antiproliferative signals which function is to maintain cellular quiescence by arresting them in the G0 phase of the cell cycle or by inducing them to enter postmitotic differentiated states. Examples of these hallmarks include tumor suppressor genes such as *RB* and *TP53*, which are responsible for the transduction of extracellular and intracellular growth-inhibitory signals, respectively, which are frequently disrupted in cancer cells. The *third hallmark* implies the development of mechanisms that allow cancer cells to evade programmed cell death (i.e., apoptosis). Intracellular and extracellular conditions cell conditions are monitored by sensors that regulate apoptotic effectors that can trigger programmed cell death when conditions require it. When pro-apoptotic signals are produced, the mitochondria respond by liberating the cytochrome C, a process mediated by the Bcl-2 protein family. P53 mediates apoptosis in response to DNA damage through the upregulation of the Bax pro-apoptotic factor, and it is found frequently mutated in most cancers representing a common mechanism of apoptosis resistance acquisition. The *fourth hallmark* involves the acquisition of limitless replicative potential. A cell-autonomous program limits the number of divisions that a cell can undergo in many normal mammalian cell types. This finite replicative potential ends with a phase called senescence, which can be circumvented by pRb and p53 inactivation. If those genes are inactivated, cells continue replicating until they reach a stage called crisis characterized by massive cell death, karyotic disarray, and chromosomes end-to-end fusions. This second stage is thought to be controlled by telomere length. Progressive telomere shortening takes place in each replication round, which implies a loss of 50 to 100 bp. Eventually, this progressive erosion causes the loss of telomere capacity to protect the chromosome ends, leading to the chromosome fusion associated with the crisis stage. Cancer cells upregulate telomerase, an enzyme that catalyzes telomere extension to circumvent this problem. The *fifth hallmark* implies the acquisition of sustained angiogenesis. Cells must be placed within 100 μm of a capillary blood vessel in order to be supplied with the appropriate amount of oxygen and nutrients. Tumors activate the angiogenic switch by changing the balance between angiogenic inducers and inhibitors. For instance, increased expression of VEGF or EGFs is observed in many tumors. The sixth hallmark implies tissue invasion and metastatic potential, which enables cancer cells to abandon the primary tumor mass and colonize new tissues. Alterations in cell-cell adhesion molecules (CAMs), including immunoglobulin, calcium-dependent cadherin-families, and integrins, which link cells to extracellular matrix substrates, are often found in cancer. For instance, E-cadherin function is lost in most epithelial cancers by mutational inactivation of the E-cadherin or

β-catenin genes. N-CAM switches from an isoform with highly adhesive properties to a poorly adhesive isoform in several cancer types. Besides, changes in integrin expression are also observed in metastatic cells.

In 2011 Harahan and Weinberg [84] updated their previous list, including two additional items (the metabolic reprogramming of tumor cells and the development of mechanisms to evade immune surveillance). The *seventh hallmark* regards the ability of cancer cells to reprogram metabolism. Otto Warburg observed that even in the presence of oxygen, cancer cells were able to reprogram their glucose metabolism by limiting their energy metabolism to glycolysis. Paradoxically glycolysis generates a low ATP yield compared to oxidative phosphorylation but allows the diversion of glycolytic intermediates to biosynthetic pathways. Cancer cells circumvent this problem by upregulating glucose transporters such as GLUT1, which increases the flux of glucose to the cytoplasm. Finally, *the eighth* hallmark is based on the ability of cancer cells to escape from immune surveillance.

1.6 Proposed mechanisms to explain the comorbidity patterns observed between CNS diseases and cancer.

Several mechanisms have been proposed to explain the disease-disease associations observed between CNS disorders and cancer, including biases in the epidemiological studies, behavioral changes, and alterations in the exposure to risk factors, biological factors, effects derived from pharmacological treatment. Some instances are provided in the following paragraph. In addition, a more exhaustive description can be found in the discussion sections of the following chapters.

Biases and flaws in the observational studies may occur as a consequence of the use of inadequate study designs. For instance, it has been suggested that the use of hospital-based samples could result in artificially inflated estimates of CNS disorders and cancer associations. The study of community-based samples has been proposed in order to avoid this kind of bias. The diversity of the psychometric scales used to diagnose some CNS disorders has also been reported as a potential confounding factor in the case of depression [85]. In addition, the lack of control for cancer-promoting lifestyle confounding factors in observational studies has also been reported to be a potential factor distorting observational studies results [86]. Behavioral changes in CNS disorders patients are also thought to modulate changes in cancer risk. It is known that patients with severe mental illness present high rates of morbidity and mortality and are more prone to exhibit unhealthy behaviors [87], including increased risks of obesity [88], heavy smoking habits[88], alcohol abuse, lack of physical activity, and poor diet [88-90]. In addition, it has also been reported that they are less likely to receive preventive care and access medical care [91, 92] and

tend to revive cancer diagnoses significantly later than their control counterparts, which decreases the chances of effective treatment [93, 94]. Several biological hypotheses have also been proposed to explain the direct and inverse associations observed between CNS disorders and cancer. They include alterations at different molecular levels, including changes in patterns of methylation, the presence of shared genetic variability, the existence of mutations on overlapping sets of genes, the joint alteration of biological processes and pathways, and changes in cellular activity and system-level processes they will be covered in **Chapters 3 and 4**. Finally, pharmacological treatment has also been invoked as a relevant factor that could modulate CNS disorders and cancer associations. Treatment with antipsychotics has been linked to the increase in breast cancer risk observed in SCZ patients [95], whereas the use of levodopa to treat PD patients has been implicated in the increased melanoma risk observed in those patients [96]. The role of drugs in these associations will be further discussed in **Chapter 5**.

1.7 Thesis objectives:

Main objective:

The present thesis aims to evaluate the comorbidities between CNS disorders and cancer from both an epidemiological and a molecular perspective and examine the potential effect of frequently prescribed medications on the associations.

Specific objectives:

Objective 1: Synthesize the available evidence regarding the epidemiological associations between CNS disorders and cancer

We aimed to synthesize the current knowledge about the epidemiological associations between seven CNS disorders (AD, ASD, PD, HD, MD, BD, and SCZ) and both all-cancer and site-specific cancer incidence and mortality through systematic reviews and meta-analyses. To this end, we searched scientific literature repositories for previous systematic reviews and observational studies reporting measures of association between the diagnosis of any of the included CNS disorders and subsequent all-cancer and site-specific cancer incidence or mortality and combined them by means of random effects models meta-analyses (**See Chapter 2**).

Objective 2: Analyze the transcriptomic associations between CNS disorders and cancers

We identified transcriptomic associations between seven CNS disorders and 22 site-specific cancers (acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), bladder cancer (BLCA), breast cancer (BRCA), brain cancers (BRNCA), cervical cancer (CERV), cholangiocarcinoma (CHLCA), chronic lymphocytic leukemia (CLL), chronic myeloid leukemia (CML), colorectal cancer (CRCA), diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FLYMPH), head and neck carcinomas (HANC), kidney cancer (KDNCA), lung cancer (LGCA), liver cancer (LIVCA), ovarian cancer (OVCA), pancreatic cancer (PACA), prostate cancer (PRCA), melanoma (SKCM), gastric cancer (STCA), and thyroid cancer(THCA). We carried out differential gene expression meta-analyses of micro-array studies for each disorder and then assessed the presence of shared deregulated genes and pathways between them. In addition, we presented a network-based approach to explore the overlaps between the gene co-expression modules significantly associated with each disorder. Finally, we partially validated our observations using an independent cohort of cancers derived from RNA-seq experiments (**See Chapter 3**).

Objective 3: Assess the interactome-based overlap of genes associated with CNS disorders and cancers and study their genetic correlations

We determined whether disease-associated genes linked to CNS disorders and cancer are placed in overlapping regions of the human interactome. The disease module hypothesis states that the components associated with a particular disease tend to be distributed in the same

neighborhood of the human interactome called the disease module. According to it, a particular disorder could arise as a consequence of the disruption of the function of genes placed at the disease module. In addition, comorbidity patterns between a given pair of disorders could be due to the existence of overlapping disease modules between both disorders. If disease modules overlap, it can be expected that the disruption of their genes jointly impacts the likelihood of developing both disorders. Therefore, we determined if the set of disease-associated genes of CNS disorders and cancers present interactome based overlaps (**See Chapter 4)**.

In addition, we determined whether genetic relationships between CNS diseases and cancer exist. Genome-Wide Association Studies (GWAS) allow us to obtain associations between a particular trait and millions of genetic variants. We used a recently developed method (cross-trait LD-score regression), which allows the computation of genetic correlations between pairs of disorders using GWAS summary statistics as an input. This provides information about the role that shared genetic variability could play in the observed comorbidity patterns (**See Chapter 4**).

Objective 4: Study the potential impact of frequently used CNS and cancer medications in the observed comorbidities.

Medications administered to treat CNS disorders or cancer could act as modulating agents that shape the observed population patterns of association between them. Furthermore, understanding the transcriptomic impact of particular drugs employed in both sets of diseases opens the door to drug repositioning, an essential topic in translational medicine and a pressing issue given its potentially beneficial consequences for patients and healthcare systems. To determine the potential impact of frequently used medications in the observed associations between CNS disorders and cancer, we used two complementary approaches. The first was a network approach based on the computations of the distances of drug targets to disease modules in the human interactome, whereas the second was based on the comparison of thousands of transcriptomic profiles generated by treating cell lines with the drugs indicated for the treatment of CNS disorders (LINCS L1000 repository) and cancers with the differential expression profiles of each disorder obtained by means of differential gene expression meta-analyses in chapter 2. We seek to understand if the transcriptomic alterations produced by these drugs were able to mimic or revert the expression profiles of a given (**See Chapter 5**).

Chapter 2. Epidemiological associations between CNS disorders and cancer

2.1 Introduction

Cancer and CNS disorders are among the top leading causes of death and disease burden internationally. On the one hand, only in 2017, 24.5 million incident cancer cases and 9.6 million cancer deaths were registered worldwide [82], with estimates suggesting that one-third of the men and one-fourth of the women would develop cancer during their lifetime. Besides, far from having reached a steady-state, cancer incidence dynamics continue on the rise, as it is suggested by the fact that age-standardized averages for annual overall-cancer incidence have increased in 123 out of 195 countries in the ten-year period comprehended between 2007 and 2017 [82]. Although the average death rates for all cancers have decreased in most countries in the same timeframe, cancer still is the second cause of death worldwide surpassed only by cardiovascular diseases [97]. On the other hand, CNS disorders have been identified as one of the leading causes of disease burden in the Global Burden of Disease study [43, 98, 99]. In the period comprehended between 2005 to 2015, almost all neurological disorders presented an increase in disease burden or disability indicators, such as the total disability-adjusted life years (DALYs).

The demographic shift towards larger and older populations is generating an epidemiological transition from communicable to chronic non-communicable diseases and from premature mortality to morbidity and a rise in the global burden of chronic conditions such as CNS disorders and cancer [43, 99, 100]. Also, this transition has propelled a passage from a paradigm focused on the study and treatment of single diseases to one characterized by the presence of multiple co-occurring conditions known as comorbidity or multimorbidity [100, 101]. Therefore, it is becoming critical to understand how diseases interact from multiple perspectives ranging from epidemiology to the study of shared molecular traits and the analysis of the exposure to environmental factors and drugs. The epidemiological study of disease associations can stimulate the development of specific screening and prevention programs, which could, in turn, translate into better clinical outcomes for the patients and more efficient use of the available resources.

An increasing body of knowledge regarding the CNS disorders and cancer associations has been produced during the last decades, with dozens of observational studies available in the literature. In some instances, authors have concluded that CNS disorders are associated with an increased risk of specific cancer types (direct comorbidity). For example, PD patients have been documented to be at a higher risk of melanoma than controls, whereas SCZ patients have been observed to present a higher probability of developing breast cancer [86, 102-106]. Other studies

have reported that specific CNS disorders are linked to decreased cancer risk (inverse comorbidity). Some representative examples include the decreased overall cancer risk observed in AD and PD patients [107-114]. However, the available observational studies have often produced heterogeneous and contrasting results, which highlighted the need for the application of methods of evidence synthesis, such as systematic reviews and meta-analyses.

To date, the most comprehensive systematic review and meta-analysis addressing the topic of CNS disorders and cancer associations is the work carried out by Catalá and co-workers in 2014 [21]. In their study, information regarding 577,013 participants derived from 50 observational studies involving eight CNS disorders (AD, amyotrophic lateral sclerosis, ASD, Down's syndrome, HD, multiple sclerosis (MS), PD, and SCZ) and several site-specific cancers was evaluated through meta-analyses. Their results suggested that CNS disorders are associated with a decrease in overall cancer risk. This risk reduction was found to be stronger in neurodegenerative disorders patients. The general trend towards overall cancer risk reduction was accompanied by some specific instances in which a higher co-occurrence of some site-specific cancers was observed. For instance, PD was found to be associated with an increased risk of melanoma, MS was linked to an increased risk of brain cancer risk, and SCZ patients were found to present a higher incidence of breast cancer compared to controls.

Since the presentation of Catalá's work in 2014, several observational studies reporting CNS and cancer associations have been published. Thus, in the context of an ongoing project [1] we considered timely to update previous results with the newly released data and expand them by exploring the associations between CNS disorder diagnosis and subsequent cancer mortality. Therefore, this chapter aims to evaluate the associations of seven CNS disorders (AD, ASD, BD, HD, PD, MD, and SCZ) with the risk of developing or dying from overall and site-specific cancers.

2.2 Material and methods

The present analysis was guided by a study protocol and previous research carried out by our group. The protocol was registered in PROSPERO (registration number CRD42016052762) and subsequently published [1]. The complete study aims to conduct an umbrella review with multiple meta-analyses addressing cancer and CNS associations. This thesis will present a preliminary update of the meta-analyses of observational studies of a subset of CNS disorders included in the ongoing project. The analyses have been reported in accordance with the reporting guidance provided in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) statement [115, 116] and the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) reporting guideline [117].

2.2.1 Inclusion criteria

2.2.1.1 Participants and exposure

This study included human participants of any age and sex. The exposure was defined as the incidence of a specific CNS disorder. CNS definitions should be based on accepted diagnostic criteria, including the International Classification of Disease (ICD) codes and Diagnosis and Statistical Manual of Mental Disorders (DSM). The central nervous system disorders included in our analysis were Alzheimer's disease (AD: ICD-9: 331.0, 290.1; ICD-10: F00, G30), autism spectrum disorders (ASD: ICD-9: 299.0, 299.8; ICD-10: F84), bipolar disorder (BD: ICD-9: 296-296.16, 296.4-296.99, 301.1-301.13; ICD-10: F06.3-F06.34, F30-F31.9, F34.0), major depression (MD: ICD-9: 296.2-296.36, 311-311.9, V11.1, V11.2; ICD-10: F32-F33.9), Huntington's disease (HD: ICD-9: 294.1, 333.4; ICD-10: F02.2, G10), Parkinson's disease (PD: ICD-9: 332-332.9; ICD-10: G20-G21.0, G21.2-G22.0), and schizophrenia (SCZ: ICD-9: 295-295.95, 301.0, 301.2-301.22, V11.0; ICD-10: F06.2, F20-F23.9, F25-F29.9). Exclusion criteria were animal studies and *in vitro* or *in vivo* experiments. The selection of the CNS disorders included in this thesis was based on two criteria. First, we intended to capture some of the variability existent in CNS disorders, which are characterized by the presence of diverse etiological origins and pathological manifestations. Therefore, we included instances of neurodegenerative (AD, HD, and PD), neuropsychiatric (BD, MD, and SCZ), and neurodevelopmental (ASD) conditions. The second involved the availability of enough data to carry out the transcriptomic analyses presented in **Chapter 3**.

2.2.1.2 Outcomes

The primary outcomes of interest were all-cancer incidence and mortality (all malignant neoplasms; ICD-9 codes 140-209; IDC-10 codes C00-C97). Secondary outcome measures were site-specific cancer incidence and mortality. **Appendix I** shows the ICD-9 and ICD-10 codes of the primary and secondary outcomes under consideration.

2.2.1.3 Study design

To be included, primary studies had to be observational studies, either case-control or cohort. Randomized control trials are not available given the nature of our research question. Studies could present any setting (inpatient, outpatient, or mixed) and coverage (single-center, multi-centric, or population-based). The articles must include a quantification of the degree of association between the exposure and the outcome in the form of relative risks (RR) or odds ratio (OR) for case-control studies or Standardized Incidence Ratios (SIR) and Standardized Mortality Ratios (SMR) in the case of cohort studies. Articles where the CNS disorder was not the exposure of interest and cancer incidence or mortality were not the outcomes of interest were excluded. No year-of-publication restrictions were applied.

2.2.2 Search strategy

One of us (J.F.M) queried MEDLINE, Scopus, Embase, and the Web of Science using the following search scheme from their inception up to November 2018. First, searches of previously published systematic reviews and meta-analyses examining the associations between a specific CNS disorder and the risk of developing or dying of cancer were carried out. The main search strategy for autism spectrum disorders in MEDLINE is presented in **Table 1**. This search strategy was adapted, when necessary, to fit the other databases. If previous systematic reviews and meta-analyses were available and included information about both incidence and mortality, the relevant observational studies references were selected, and their information extracted. The list of PubMed IDs identified in the previous step was used as an input to perform a related article search query in PubMed, which aimed to find observational studies published out of the time frame of the detected systematic reviews. This search effectively identifies relevant studies in the presence of an already large evidence base [118]. In the case in which no systematic reviews reporting CNS and cancer incidence or CNS and cancer mortality associations were available systematic searches of observational studies were carried out from scratch following the query structures depicted in **Table 2** (MEDLINE query structure). Two of us (F.C. L and J.F.M) independently screened the titles and abstracts of the retrieved studies. Full texts of potentially eligible studies were then assessed to evaluate their final inclusion.

Search number	Example
#1 Exposure	CNSd* (e.g. Autism OR ASD)
#2 Outcome	cancer OR carcinoma OR neoplasia OR tumor OR neoplasm OR maligna*
#3 Systematic review	systematic review* OR systematic overview* OR evidence based review* OR evidence-based overview* OR meta-review* OR meta-analy* OR metaanaly* OR matanaly* OR research overview* OR collaborative review*
# Final search	#1 AND #2 AND #3

Table 1: MEDLINE query structure for systematic reviews and meta-analyses (AD example).

Search number	Example
#1 Exposure	CNSd* (e.g. Autism OR ASD)
#2 Outcome	cancer OR carcinoma OR neoplasia OR tumor OR neoplasm OR maligna*
#3 Observational studies	incidence OR comorbidity OR multimorbidity OR mortality OR death OR epidemiologic* OR cohort stud* OR longitudinal stud* OR case-control stud*
#4 Final search	#1 AND #2 AND #3

Table 2: MEDLINE query structure example for observational studies (AD example).

2.2.3 Data extraction

Information regarding the following items was extracted from each article: Year of publication, country, study design (prospective, retrospective, cohort or nested case-control), setting (population-based or hospital-based), coverage (single-center, multicenter, population-based), study years, mean follow-up, number of participants with the exposure of interest, characteristics of participants (sex and age), number of incident cases of cancer deaths in the exposed individuals, outcome definitions (i.e., ICD codes, DSM codes), endpoint measures (OR, RR, SIR, SMR), and the adjustment for confounding factors. In addition, effect sizes with their correspondent, 95% confidence intervals were also extracted from the original articles. For those studies in which both raw and adjusted effect sizes were reported, the effect sizes adjusted by the maximum number of confounding variables were retrieved.

Studies that independently reported measures for colon and rectum cancers were combined under FEM in a shared category (colorectal cancer). Studies including disaggregated information about myeloid and lymphoid leukemias were also combined by FEM meta-analyses in a unique category called leukemia.

2.2.4 Appraisal of the quality of observational studies.

One of us (J.F.M) evaluated the quality and the risk of bias of each observational study using the Newcastle-Ottawa scale (NOS)[119]. NOS evaluates different aspects of the study, including how participants are selected, the comparability between cases and controls or cohorts and the reference population, and how exposures and outcomes are defined. Stars are allocated based on the adherence to pre-specified criteria. The scale ranges from 0 (lowest quality) to 9 (highest quality). In our analysis, we divided studies into three groups according to their risk of bias. High risk (0 to 3 stars), moderate risk (4 to 6 stars), and low risk (7 to 9 stars). The complete list of items included in NOS can be checked in **Appendix 2**, which contains the templates of the scale for both case-control and cohort studies.

2.2.5 Data synthesis

Following Cochranes' Handbook recommendations, all effect sizes retrieved from observational studies were subjected to the log transformation prior to meta-analytic integration, and standard errors were computed using the logarithms of the 95% confidence interval, as shown by the next equation:

$$SE = \frac{log(Upper\ CI) - log(Lower\ CI)}{3.92}$$

Pooled estimates for each CNS disorder and cancer associations were computed through meta-analysis using the inverse-variance method. Random-effects models were selected *a priori*, given that between-study heterogeneity was expected [120]. Cochran Q tests [121] and the I^2 statistic were used to assess heterogeneity between studies. Cochran Q tests p-values lower than 0.05 suggest that heterogeneity is present. The I^2 statistic is expressed as a percentage. I^2 values of 0%-25%, 25%-50%, 50%-75%, and 75%-100% indicate low, moderate, substantial, and considerable heterogeneity respectively. All analyses were carried out using the R statistical programming language and the metafor package [122].

2.2.6 Additional analyses

Potential sources of heterogeneity were studied when feasible. Subgroup analysis according to sex (male or female), study design (cohort or case-control), and risk of bias (high or low-moderate risk of bias) were conducted for overall cancer incidence and mortality. For cancer incidence, subgroup analysis was carried out comparing the site-specific cancers linked to smoking against those with no reported association with smoking status. Information regarding which site-specific cancers are associated with smoking can be found in **Appendix 1**. We also carried out random-effects meta-analyses for the cluster of neurodegenerative disorders (including

Alzheimer's disease, Parkinson's disease, and Huntington's disease) and their association with primary outcomes.

The potential publication bias was assessed for those meta-analyses including ten or more studies. Publication bias relies on the idea that significant findings or studies presenting large effect sizes are more likely to be published [123]. This implies that moderate and small-sized studies reporting non-significant results would be more likely to be missing and, therefore, would not be integrated into meta-analyses. Publication bias was assessed by funnel plot visualization and the computation of Egger's and Begg's test statistics [124, 125], which are designed to measure funnel plot asymmetries.

2.3. Results

2.3.1 Relevant literature identified.

Searches for previously published systematic reviews and meta-analyses yielded a total of 9374 article references. After title, abstract, and full-text inspection, 28 previous systematic reviews were identified. (**See Appendix 3**) from which 487 references to observational studies were extracted. In the case of ASD, BD, HD, and PD, the searches for observational studies were carried out from scratch due to the lack of previous systematic reviews examining incidence, mortality, or none of them. In total, sixteen thousand one-hundred and twenty-one article references were yielded by these searches. In the case of AD, MD, and SCZ, a related article search was performed, and 8749 article references were recovered. Overall, 801, 1097, 2935, 3628, 4370, 8854, 3619 unique references were obtained for AD, ASD, BD, HD, MD, PD, and SCZ, respectively. After screening titles and abstracts 30, 13, 55, 9, 423, 112, and 122, references for AD, ASD, BD, HD, MD, PD, and SCZ were selected for full-text review. Two-hundred and eighteen articles (192 unique articles) met inclusion criteria and were used in the downstream analysis, and 546 were excluded due to different reasons, which are summarized in **Figure 2**. **Appendix 4** shows the characteristics of all the studies that met the inclusion criteria.

2.3.2 Description of the studies and participants

Data from one hundred and seventy-four cohorts reported in a total of 192 unique article references met inclusion criteria and were selected for downstream analysis. One hundred and seventy-two articles included information about one CNS disorder, whereas twenty-one studies incorporated information for associations between cancer incidence or mortality and more than one CNS disorder. Most of them presented data regarding different neuropsychiatric conditions. Six studies included associations between cancer and BD, MD, and SCZ [126-131], two studies reported information regarding associations between cancer and bipolar disorder or major depression [132, 133]. Eight showed results for BD and SCZ [134-141]. Three works had data on MD or SCZ and cancer associations [142-144]. Finally, only two articles presented data about two neurodegenerative disorders, Parkinson's disease and Alzheimer's disease [145, 146].

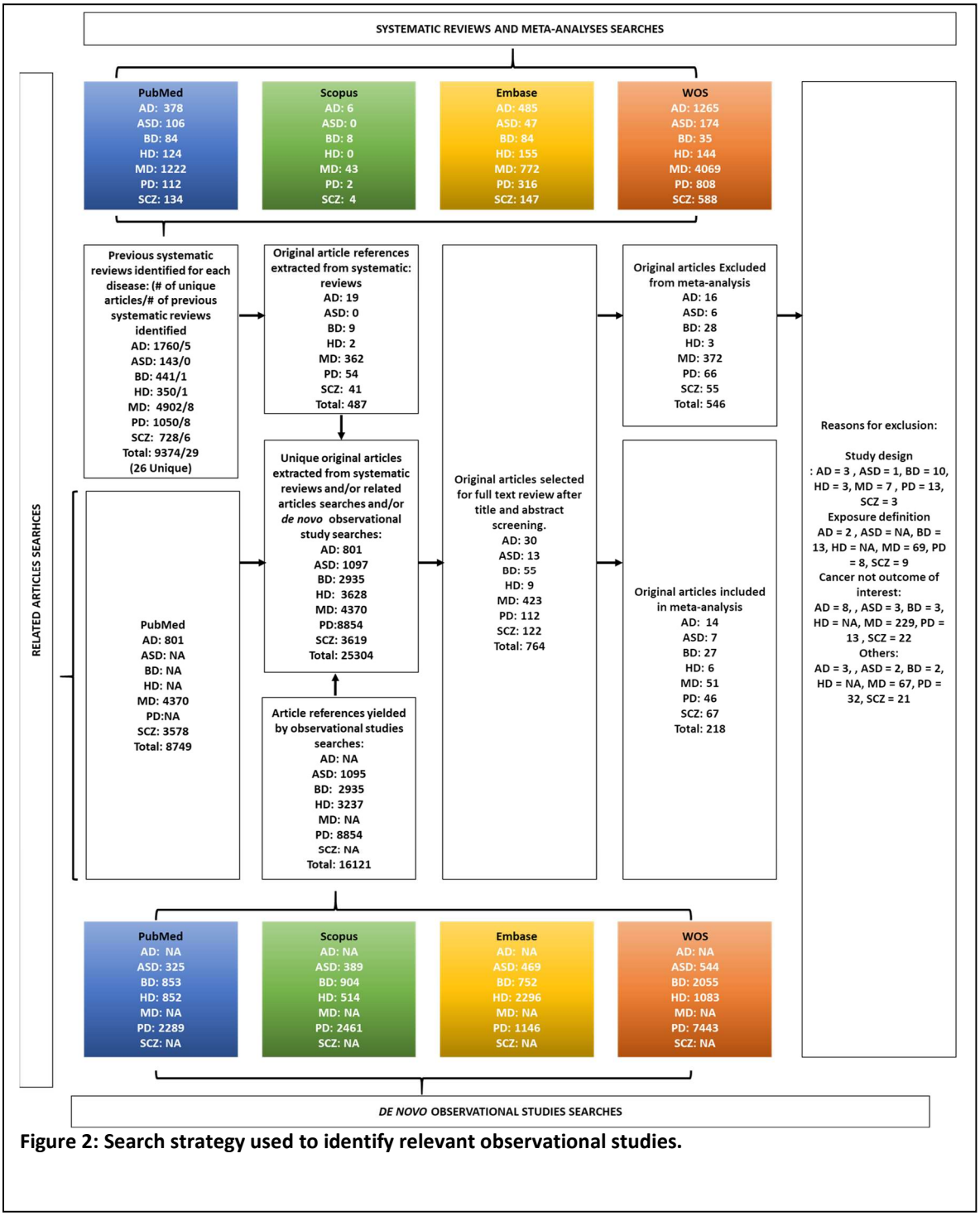

Figure 2: Search strategy used to identify relevant observational studies.

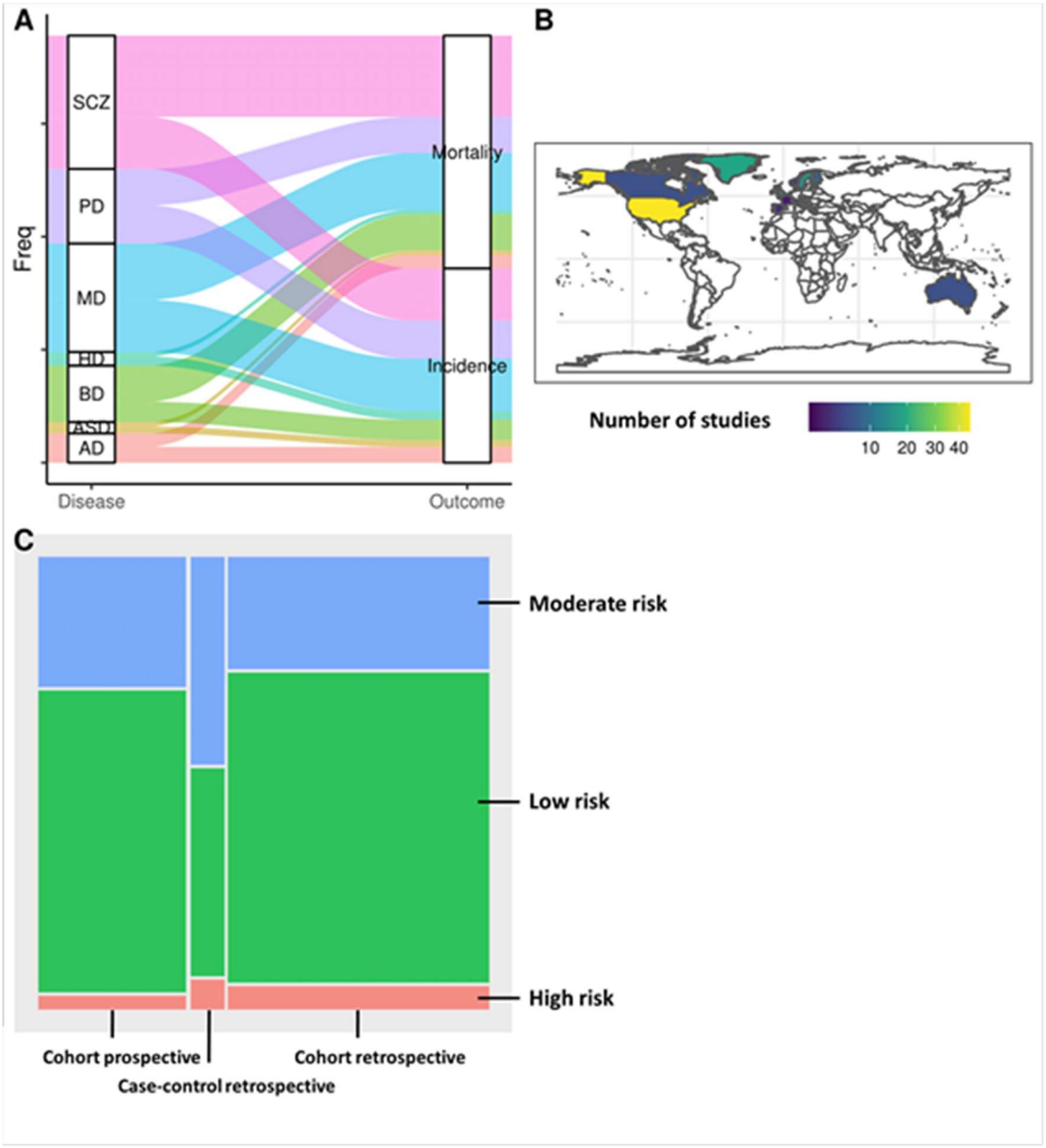

Figure 3: Characteristics of the studies that met inclusion criteria. A) Proportion of the studies reporting associations for each CNS disorder and main outcome. B) Geographic distribution of the identified studies. C) Prpotrtion of the NOS study qualities based on their design.

The number of identified studies devoted to each disease and outcome was found to be highly heterogeneous. For instance, cancer incidence following schizophrenia diagnosis was reported in 26 studies, whereas cancer incidence following ASD diagnosis was reported only in four. In the case of cancer mortality, the number of available studies ranged between 37 in SCZ and 2 in Huntington's disease. **Figure 3 A** graphically depicts the volume of studies available for each CNS disorder and each primary outcome. SCZ, MD, and PD were the most extensively studied CNS disorders, with more than ten studies available for both major outcomes. On the other hand, AD, ASD, BD, and HD presented less than ten studies reporting cancer incidence or mortality. The less extensively studied diseases were HD and ASD.

Regarding the geographical distribution of the research, fifty-one studies were based in North America (45 in the United States and 6 in Canada). Nordic countries produced a total of 45 studies (19 in Denmark, 14 in Sweden, 8 in Finland, and 4 in Norway). East Asia contributed with 25 studies (17 in Taiwan and 6 in Japan). The UK was another major contributor, with 19 studies. In the Middle East, only Israel contributed with seven studies. Works produced in nine other countries were available, with each country contributing with five or fewer studies. For studies analyzed data from cohorts derived from more than one country. There was a complete lack of research coming from Central and South America, Africa, and most of Asia.

More than 1075159 participants with a CNS disorder and more than 69359 incident cancers were included in the analyses of CNS and cancer incidence associations. CNS disorder patients raged between 31 and 219194, whereas the number of incident cancers ranged between 8 and 17524. In the case of mortality, information regarding more than 2325378 individuals (range: 59-1138853) with a particular CNS disorder and more than 22410 deaths due to cancer (range: 2-9638) in the CNS groups were included. For all-cancer incidence studies, female participant's proportions ranged between 19.3% and 85.5%, whereas for all-cancer mortality, it ranged between 20.6% and 85.6%.

One hundred and sixty-one studies (89.5%) presented a cohort design, whereas 19 (10.5%) were case-control studies. Ninety-seven cohort studies (60.6%) were retrospective, and 63 (39.4%) were prospective studies. One hundred and sixty-five studies included information regarding their follow-up duration. One hundred and thirty-seven (83%) presented follow-up durations longer than five years, whereas 28 (17%) presented follow-up durations between 1 and 5 years. After the Newcastle Ottawa Scale quality appraisal, most studies 115 (63.9%) were placed in the low risk of bias category. Thirty-one percent were (56) were classified as a moderate risk of bias studies, and only 9 (0.5%) were considered to be a high risk of bias studies.

Regarding the publication dates, 74.7% of the studies were published after the year 2000, whereas 25.3% were published before. Site-specific cancer data was also highly heterogeneous in terms of the number of reports including information about each cancer type.

The best-studied site-specific cancers in terms of incidence were breast, lung, prostate, colorectal, brain and central nervous, stomach, and malignant skin melanoma, with 57, 54, 52, 51, 35, 34, and 31 studies reporting measures of associations between them and any of the included CNS disorders. Breast, Lung, colorectal, pancreatic, and prostate cancers with 15, 10, 9, 7, and 7 studies reporting site-specific mortality rates were the most extensively studied cancers in the context of mortality.

2.3.3 Meta-analyses results

2.3.3.1 All-cancer incidence and mortality in patients with CNS disorders.

This section presents the meta-analysis results for the primary outcomes (all-cancer incidence and mortality) for all the CNS disorders included in our study. **Figures 4 and 5** show the pooled estimates and their 95% confidence intervals of all-cancer incidence and mortality of all the studied CNS disorders. **Table 3** includes the results stratified by gender and additional information regarding the number of available studies and participants included in each meta-analysis and their heterogeneity estimates.

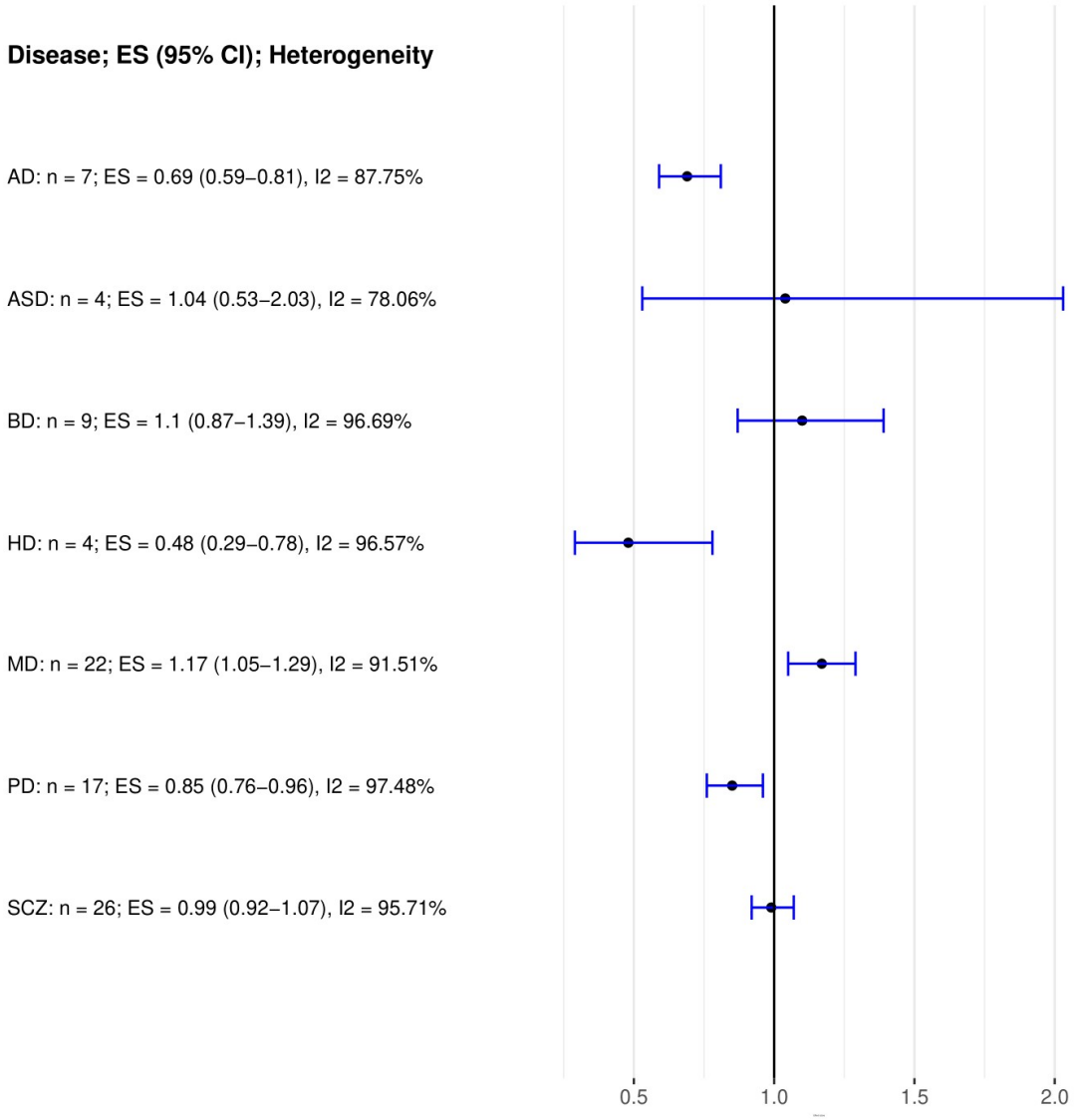

Figure 4: Forest plot depicting the meta-analyses results for the associations between each CNS disorder and all-cancer incidence. Points represent the pooled effect sizes for each meta-analysis, whereas whiskers show their 95% confidence intervals. Points situated to the left of the black line placed at 1 with non-intersecting upper confidence intervals indicate significantly reduced risks, whereas points situated to the right of the line with non-intersecting lower confidence intervals indicate significantly higher risks.

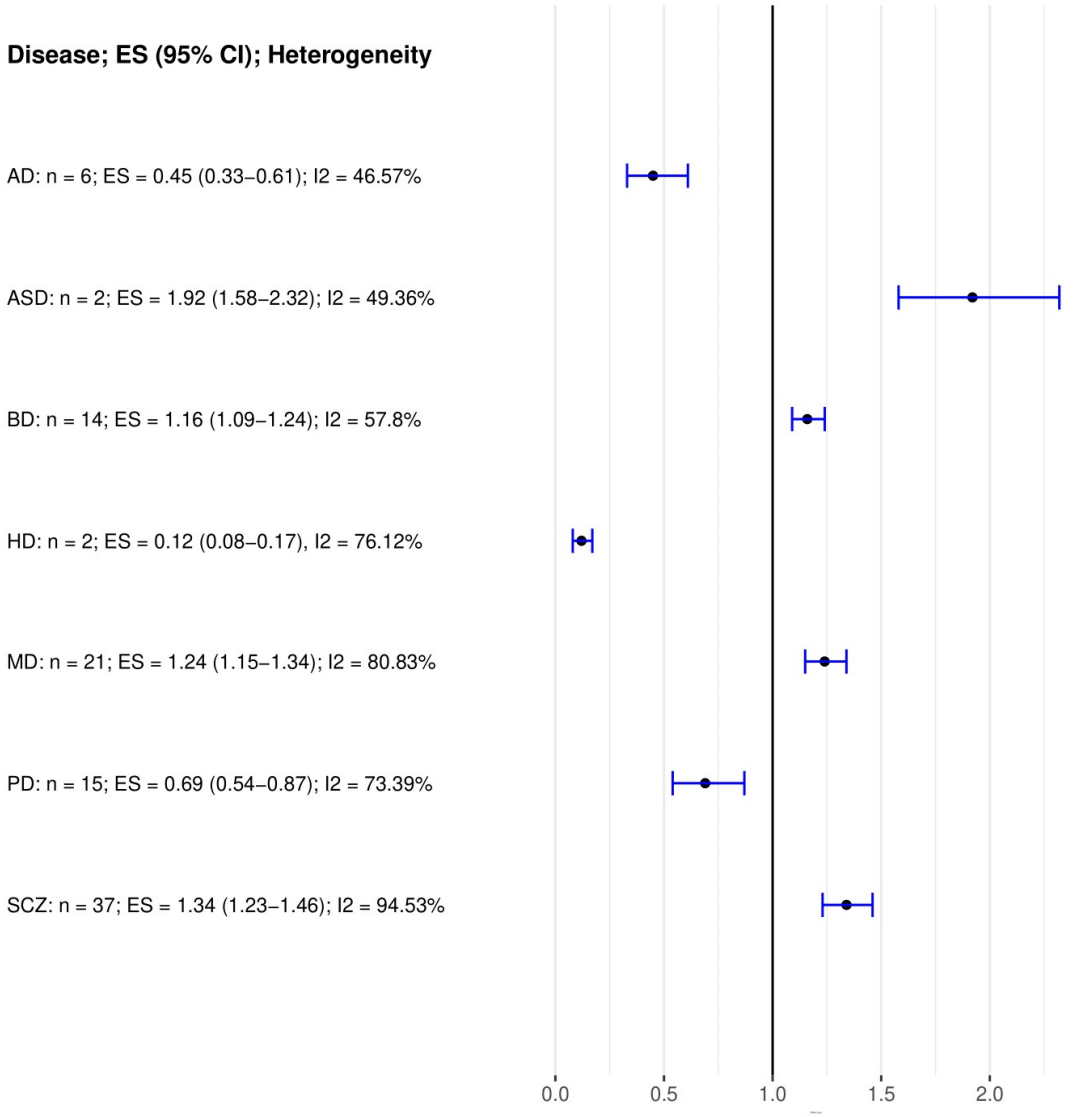

Figure 5: Forest plot depicting the meta-analyses results for the associations between each CNS disorder and all-cancer mortality. Points represent the pooled effect sizes for each meta-analysis, whereas whiskers show their 95% confidence intervals. Points situated to the left of the black line placed at 1 with non-intersecting upper confidence intervals indicate significantly reduced risks, whereas points situated to the right of the line with non-intersecting lower confidence intervals indicate significantly higher risks.

AD all cancer incidence and mortality

Alzheimer's disease analyses included data from 43463 and 2278 AD patients for incidence and mortality divided into 7 and 6 studies, respectively. In both cases, AD patients presented significantly reduced overall cancer risks compared to controls. (Incidence: n = 7; RR = 0.69; 95% CI: 0.59-0.81; I2 = 87.75%), (Mortality: n = 6; RR = 0.45; 95% CI: 0.33-0.61; I2 = 46.57%). Three studies included gender-specific incidence data and showed a significant reduction in the risk of cancer incidence in both sexes (Men incidence, n = 3; RR = 0.8; 95% CI: 0.66-0.97; I2 = 80.31%), (Women incidence, n = 3; RR = 0.74; 95% CI: 0.57-0.96; I2 = 89.53%). Gender-specific associations between AD and subsequent cancer mortality were not computed out due to the lack of data.

HD all cancer incidence and mortality

HD patients were also found to be at a significantly lower risk of all-cancer incidence (n = 4; RR = 0.48; 95% CI: 0.29-0.78; I2 = 96.57%) and mortality (n = 2; RR = 0.12; 95% CI: 0.08-0.17; I2 = 76.12%) using data derived from four and two studies, that included information regarding 13609 and 954 HD patients, respectively. There were no studies reporting sex-specific associations.

PD all cancer incidence and mortality

PD patients presented a significantly decreased risk of subsequent cancer incidence (n = 17; RR = 0.85; 95% CI: 0.76-0.96; I2 = 97.48%) and mortality (n = 15; RR = 0.69; 95% CI: 0.54-0.87; I2 = 73.39%). The pooled estimates were computed using 17 incidence studies including 354971 PD patients and 15 mortality studies gathering information about 22245 PD patients. In the case of overall-cancer incidence, sex-specific analysis for men (n = 12; RR = 0.88; 95% CI: 0.74-1.05; I2 = 97.56%) and women (n = 11; RR = 0.93; 95% CI: 0.81-1.08; I2 = 94.63%) did not show a significant reduction in cancer risk. However, a trend towards a negative association was observed in both cases. In contrast, both men (n = 3; RR = 0.56; 95% CI: 0.4-0.77; I2 = 0%) and women (n = 2; RR = 0.62; 95% CI: 0.4-0.97; I2 = 19.42%) were found to be at a lower risk of all-cancer mortality than controls, however, these meta-analyses included only 3 and 2 observational studies, respectively.

MD all cancer incidence and mortality

All-cancer incidence was found to be significantly increased in MD patients compared to controls (n = 22; RR = 1.17; 95% CI: 1.05-1.29; I2 = 91.51%). The analysis included more than 109742 MD patients from 22 different studies. Seven studies included gender-specific estimates. Pooled estimates suggested a non-significant trend towards increased cancer risk in depressed men (n = 7; RR = 1.2; 95% CI: 0.99-1.45; I2 = 92.96%) and women (n = 7; RR = 1.1; 95% CI: 0.84-1.43; I2 = 96.62%), respectively. Associations between all-cancer mortality and MD were reported in 21 studies. Pooled estimates indicate that MD patients are at higher risk of dying due to cancer than controls (n = 21; RR = 1.24; 95% CI: 1.15-1.34; I2 = 80.83%). A similar significant association was also found for men

(n = 12; RR = 1.42; 95% CI: 1.18-1.71; I2 = 88.73%) and women (n = 8; RR = 1.24; 95% CI: 1.12-1.38; I2 = 58.45%).

BD all-cancer incidence and mortality

BD patients were not found to be at higher risk of cancer incidence than controls (n = 9; RR = 1.1; 95% CI: 0.87-1.39; I2 = 96.69%) but presented a significantly increased probability of dying of cancer (n = 9; RR = 1.09; 95% CI: 1.03-1.15; I2 = 0%). Nine studies comprehending 59105 BD patients and 14 studies including information regarding 46406 BD patients were available for the incidence and mortality meta-analyses, respectively. The gender-specific analyses yielded no significant associations between previous exposure to BP and subsequent all-cancer incidence either for men (n = 7; RR = 0.98; 95% CI: 0.67-1.43; I2 = 94.01%) or women (n = 7; RR = 1.21; 95% CI: 0.84-1.73; I2 = 95.3%). In contrast, both men (n = 9; RR = 1.09; 95% CI: 1.03-1.15; I2 = 0%) and women (n = 7; RR = 1.11; 95% CI, 1.04-1.19; I2 = 32.06%) were found to be at an increased risk of cancer mortality after BD diagnosis.

SCZ all-cancer incidence and mortality

Twenty-six studies, including 505132 SCZ patients, reported associations between SCZ and subsequent cancer incidence. No differences in the risk of developing cancer were observed between SCZ patients and controls (n = 26; RR = 0.99; 95% CI: 0.92-1.07; I2 = 95.71%). A significant reduction of overall-cancer risk was found in males (n = 22; RR = 0.87; 95% CI: 0.78-0.96; I2 = 95.05%) but not in female patients (n = 20; RR = 1.08; 95% CI, 0.99-1.18; I2 = 93.84%). Regarding cancer mortality, the meta-analysis of thirty-seven studies including 1978502 SCZ participants showed that SCZ patients are at an increased risk of cancer mortality compared to controls (n = 37; RR = 1.34; 95% CI: 1.23-1.46; I2 = 94.53%). Sex-specific analyses showed an increased risk in cancer mortality in both men (n = 22; RR = 1.28; 95% CI: 1.12-1.46; I2 = 94.17%) and women (n = 19; RR = 1.43; 95% CI: 1.29-1.6; I2 = 93.11%) with SCZ.

ASD and all-cancer incidence and mortality

All-cancer incidence was not found to be different in patients with ASD and controls (n = 4; RR = 1.04; 95% CI: 0.53-2.03; I2 = 78.06%). However, as in the case of HD, only four studies were available, which gathered data derived from 15388 participants. Gender-specific analyses based on three studies did not show significant differences. In the case of cancer mortality, ASD patients were found to be at a higher risk than controls (n = 2; RR = 1.92; 95% CI: 1.58-2.32; I2 = 49.36%). However, this meta-analysis was carried out only with two studies. The risk of mortality due to cancer was significantly increased for both men (RR = 1.79, 95% CI, 1.34-2.38) and women (RR = 1.83, 95% CI, 1.33-2.5) in the unique observational study reporting these associations.

Considerable between-study heterogeneity was present in most of the meta-analyses (24 out of 36) carried out in this section, as suggested by the presence of I^2 values higher than 75% and

the results of the Q-tests. In addition, those meta-analyses where considerable heterogeneity was not detected were characterized by the presence of a reduced number of studies. The only exceptions to this trend were the BD sex-specific cancer mortality analyses for men and women, which presented moderate and low heterogeneity and pooled data from 9 and 7 studies, respectively. **Table 3** summarizes all the meta-analysis results obtained in this section, including the number of studies and participants, the pooled estimates, and heterogeneity measures. **Appendix 1 Figures 1 to 27** include the forest and funnel plots of the primary outcomes.

Outcome	Nº Studies	Nº of participants with CNS	Pooled RR (95% CI)	P-value	I^2	Q (P-value)
AD All-cancer incidence both	7	43463	0.69 (0.59-0.81)	5.60e-06	87.75%	48.99 (7.5e-09)
AD All-cancer incidence men	3	6839	0.8 (0.66-0.97)	2.10e-02	80.31%	10.16 (6.2e-03)
AD All-cancer incidence women	3	10283	0.74 (0.57-0.96)	2.20e-02	89.53%	19.11 (7.1e-05)
AD All-cancer mortality	6	2278	0.45 (0.33-0.61)	3.00e-07	46.57%	9.36 (9.6e-02)
ASD All-cancer incidence both	4	15388	1.04 (0.53-2.03)	9.20e-01	78.06%	13.67 (3.4e-03)
ASD All-cancer incidence men	3	8886	1.01 (0.37-2.76)	9.80e-01	85.43%	13.73 (1e-03)
ASD All-cancer incidence women	3	3462	1.16 (0.39-3.46)	7.90e-01	47.19%	3.79 (1.5e-01)
ASD All-cancer mortality	2	40233	1.92 (1.58-2.32)	3.30e-11	49.36%	1.97 (1.6e-01)
ASD All-cancer mortality men	1	18693	1.79 (1.34-2.38)	-	-	-
ASD All-cancer mortality women	1	8429	1.83 (1.33-2.5)	-	-	-
BD All-cancer incidence both	9	59105	1.1 (0.87-1.39)	4.40e-01	96.69%	241.89 (9e-48)
BD All-cancer incidence men	7	21977	0.98 (0.67-1.43)	9.00e-01	94.01%	100.16 (2.3e-19)
BD All-cancer incidence women	7	25248	1.21 (0.84-1.73)	3.00e-01	95.30%	127.61 (4.1e-25)
BD All-cancer mortality	14	46406	1.16 (1.09-1.24)	7.50e-06	57.80%	30.8 (3.6e-03)
BD All-cancer mortality men	9	9636	1.09 (1.03-1.15)	1.50e-03	0%	6.95 (5.4e-01)
BD All-cancer mortality women	7	12927	1.11 (1.04-1.19)	3.40e-03	32%	8.83 (1.8e-01)
HD All-cancer incidence both	4	13609	0.48 (0.29-0.78)	3.40e-03	96.57%	87.47 (7.6e-19)
HD All-cancer mortality	2	954	0.12 (0.08-0.17)	2.70e-29	76.12%	4.19 (4.1e-02)

MD All-cancer incidence both	22	109742	1.17 (1.05-1.29)	3.10e-03	91.51%	247.34 (1.4e-40)
MD All-cancer incidence men	7	45076	1.2 (0.99-1.45)	6.00e-02	92.96%	85.22 (3e-16)
MD All-cancer incidence women	7	21607	1.1 (0.84-1.43)	5.10e-01	96.62%	177.63 (1.1e-35)
MD All-cancer mortality	21	252573	1.24 (1.15-1.34)	1.00e-08	80.83%	104.33 (2.1e-13)
MD All-cancer mortality men	12	10078	1.42 (1.18-1.71)	2.50e-04	88.73%	97.61 (5.3e-16)
MD All-cancer mortality women	8	10140	1.24 (1.12-1.38)	2.70e-05	58.45%	16.85 (1.8e-02)
PD All-cancer incidence both	17	354971	0.85 (0.76-0.96)	9.10e-03	97.48%	634.69 (9.9e-125)
PD All-cancer incidence men	12	179269	0.88 (0.74-1.05)	1.60e-01	97.56%	450.71 (1e-89)
PD All-cancer incidence women	11	145229	0.93 (0.81-1.08)	3.70e-01	94.63%	186.27 (1.2e-34)
PD All-cancer mortality	15	22245	0.69 (0.54-0.87)	2.00e-03	73.39%	52.61 (2.2e-06)
PD All-cancer mortality men	3	1207	0.56 (0.4-0.77)	5.00e-04	0.00%	1.95 (3.8e-01)
PD All-cancer mortality women	2	493	0.62 (0.4-0.97)	3.80e-02	19.42%	1.24 (2.7e-01)
SCZ All-cancer incidence both	26	505132	0.99 (0.92-1.07)	7.60e-01	95.71%	582.44 (5.6e-107)
SCZ All-cancer incidence men	22	257654	0.87 (0.78-0.96)	7.70e-03	95.05%	423.96 (1e-76)
SCZ All-cancer incidence women	20	196272	1.08 (0.99-1.18)	8.10e-02	93.84%	308.32 (3.9e-54)
SCZ All-cancer mortality	37	1978502	1.34 (1.23-1.46)	6.10e-12	94.53%	657.57 (2.9e-115)
SCZ All-cancer mortality men	22	837914	1.28 (1.12-1.46)	2.90e-04	94.17%	360.3 (1.4e-63)
SCZ All-cancer mortality women	19	706815	1.43 (1.29-1.6)	1.00e-10	93.11%	261.27 (4.1e-45)

Table 3: All-cancer incidence and mortality meta-analysis results for the primary outcomes. Results stratified by gender are also reported.

2.3.3.2 Site-specific cancer incidence and mortality.

In this section, we present the results regarding secondary outcomes (site-specific cancer incidence and mortality). Overall, neurodegenerative disorders were characterized by the presence of negative patterns of associations with site-specific cancers. From the 27 significant associations reported in this section, 24 were found to point in the direction of cancer risk reduction, and only three suggested an increase in cancer risk. The significant associations observed between MD, BP,

and SCZ and site-specific cancers were pointing to the direction of an increase of risk (18/22), whereas almost no data was available for ASD.

AD site-specific cancer incidence and mortality

AD patients were found to present a significantly reduced risk of several site-specific cancers. Three of them were derived from meta-analyses carried out using three or more studies. Those were liver (n = 3; RR = 0.72; 95% CI: 0.61-0.85; I2 = 0%) and lung (n = 4; RR = 0.81; 95% CI: 0.7-0.94; I2 = 59.23%) cancers and malignant skin melanoma (n = 3; RR = 0.81; 95% CI: 0.69-0.94; I2 = 0%). A significant risk reduction was also observed for two site-specific cancers for which only two studies were available which were kidney (n = 2; RR = 0.77; 95% CI: 0.63-0.93; I2 = 0%) and uterus (n = 2; RR = 0.64; 95% CI: 0.52-0.78; I2 = 0%) cancers. In addition, AD patients were found to be at a reduced risk of cancers of the larynx (n = 1; RR = 0.52; 95% CI: 0.34-0.8), and thyroid (n = 1; RR = 0.58; 95% CI: 0.36-0.94) with only one available observational study reporting these associations. **Table 4** summarizes the significant results. Associations between AD status and site-specific cancer mortality could not be carried out due to the lack of data.

Outcome	Nº Studies	Nº of participants with CNS	Pooled RR (95% CI)	P-value	I^2	Q (P-value)
Kidney	2	14281	0.77 (0.63-0.93)	7.20E-03	0%	0.43 (5.1e-01)
Larynx	1	7321	0.52 (0.34-0.8)	-	-	-
Liver	3	32577	0.72 (0.61-0.85)	1.10E-04	0%	1.95 (3.8e-01)
Lung	4	42670	0.81 (0.7-0.94)	6.40E-03	59.23%	7.36 (6.1e-02)
Malignant skin melanoma	3	39838	0.81 (0.69-0.94)	7.20E-03	0.00%	1.06 (5.9e-01)
Thyroid	1	7321	0.58 (0.36-0.94)	-	-	-
Uterus (Women)	2	8398	0.64 (0.52-0.78)	1.60E-05	0%	0.04 (8.3e-01)

Table 4: Significant associations between AD and site-specific cancer incidence and mortality.

HD site-specific cancer incidence and mortality

HD patients were found to be at a reduced incidence risk of colorectal (n = 3; RR = 0.5; 95% CI: 0.27-0.93; I2 = 75.89%), prostate (n = 3; RR = 0.36; 95% CI: 0.25-0.5; I2 = 6.96%), and stomach (n = 3; RR = 0.52; 95% CI: 0.36-0.74; p = 3.5e-04; I2 = 0%) cancers. An additional cancer type, breast cancer (n = 2; RR = 0.41; 95% CI: 0.31-0.53; I2 = 33.32%), showed a decreased incidence risk in HD patients compared to controls in a meta-analysis carried out using data from two studies. Finally, a single observational study reported a significant risk reduction of uterine cancer (n = 1; RR, 0.66; 95% CI, 0.36-0.96) in HD patients. No site-specific cancer mortality data were available for HD (**See Table 5**).

Outcome	Nº Studies	Nº of participants with CNS	Pooled RR (95% CI)	P-value	I^2	Q (P-value)
Breast (Women)	2	8050	0.41 (0.31-0.53)	2.90E-11	33.32%	1.5 (2.2e-01)
Colorectal	3	12915	0.5 (0.27-0.93)	2.80E-02	75.89%	8.29 (1.6e-02)
Prostate (Men)	3	12915	0.36 (0.25-0.5)	4.80E-09	6.96%	2.15 (3.4e-01)
Uterus (Women)	1	6540	0.66 (0.36 - 0.96)	-	-	-

Table 5: Significant associations between HD and site-specific cancer incidence and mortality.

PD site-specific cancer incidence and mortality

PD patients presented a lower risk for several site-specific cancers compared to controls, including bladder (n = 8; RR = 0.73; 95% CI: 0.57-0.93; I2 = 89.6%), colorectal (n = 15; RR = 0.8; 95% CI: 0.7-0.91; I2 = 88.51%), larynx (n = 5; RR = 0.57; 95% CI: 0.37-0.87; I2 = 57.45%), lip and oral cavity (n = 6; RR = 0.75; 95% CI: 0.62-0.89; I2 = 0%), and lung (n = 13; RR = 0.6; 95% CI: 0.47-0.76; I2 = 94.56%). Contrarily, the incidences of brain cancer (n = 7; RR = 1.5; 95% CI: 1.11-2.04; I2 = 64.49%) and malignant skin melanoma (n = 13; RR = 1.49; 95% CI: 1.17-1.89; I2 = 76.21%) were found to be significantly increased in PD patients compared to controls. The risk of testicular cancer (n = 2; RR = 1.67; 95% CI: 1.03-2.68; I2 = 0%) was also found to be increased in PD patients. However, only two studies were available in this case. Regarding cancer mortality, bladder (n = 2; RR = 0.55; 95% CI: 0.35-0.85; I2 = 0%), colorectal (n = 3; RR = 0.62; 95% CI, 0.5-0.76; I2 = 0%), and lung (n = 4; RR = 0.3; 95% CI: 0.17-0.52; I2 = 48.11%) cancers followed the same trend of risk reduction found in the incidence analyses, as well as pancreatic (n = 3; RR = 0.56; 95% CI: 0.32-0.97; I2 = 8.95%) and stomach (n = 3; RR = 0.53; 95% CI: 0.3-0.95; I2 = 13.78%; Q = 2.32) cancers.

The probabilities of dying from breast (n = 2; RR = 0.61; 95% CI: 0.41-0.89; I2 = 0%) and liver (n = 1; RR = 0.31; 95% CI: 0.17-0.51) cancer were also found to be lower for PD patients than in individuals without PD; however, only two and one studies reported these associations, respectively. Finally, the risk of dying by malignant skin melanoma (n = 2; RR = 1.86; 95% CI: 1.4-2.48; I2 = 53.36%) was found to be increased in PD patients compared to controls, which was in agreement with the results of the incidence meta-analysis. **Table 6** summarizes the results obtained for Parkinson's disease.

Outcome	Nº Studies	Nº of participants with CNS	Pooled RR (95% CI)	P-value	I^2	Q (P-value)
Bladder	8	326585	0.73 (0.57-0.93)	1.20E-02	89.60%	67.28 (5.2e-12)

Bladder (death)	2	10790	0.55 (0.35-0.85)	6.6e-03	0%	0.05 (8.2e-01)
Brain and CNS	7	327824	1.5 (1.11-2.04)	9.30E-03	64.49%	16.9 (9.7e-03)
Breast (Women death)	2	5888	0.61 (0.41-0.89)	1.1e-02	0%	0.03 (8.7e-01)
Colorectal	15	350055	0.8 (0.7-0.91)	8.90E-04	88.51%	121.83 (2.8e-19)
Colorectal (death)	3	13788	0.62 (0.5-0.76)	4.2e-06	0%	0.59 (7.5e-01)
Larynx	5	256871	0.57 (0.37-0.87)	9.60E-03	57.45%	9.4 (5.2e-02)
Lip and oral cavity	6	40278	0.75 (0.62-0.89)	1.30E-03	0%	1.34 (9.3e-01)
Liver (death)	1	10322	0.31 (0.17-0.51)	-	-	-
Lung	13	349150	0.6 (0.47-0.76)	3.00E-05	94.56%	220.66 (1.7e-40)
Lung (death)	4	13958	0.3 (0.17-0.52)	2.5e-05	48.11%	5.78 (1.2e-01)
Malignant skin melanoma	13	335222	1.49 (1.17-1.89)	1.20E-03	76.21%	50.45 (1.2e-06)
Malignant skin melanoma (deaths)	2	13320	1.86 (1.4-2.48)	2.2e-05	53.36%	2.14 (1.4e-01)
Pancreas (death)	3	13788	0.56 (0.32-0.97)	3.8e-02	8.95%	2.2 (3.3e-01)
Stomach (deaths)	3	13788	0.53 (0.3-0.95)	3.2e-02	13.78%	2.32 (3.1e-01)
Testicular	2	132076	1.67 (1.03-2.68)	3.60E-02	0%	0.62 (4.3e-01)

Table 6: Significant associations between PD and site-specific cancer incidence and mortality.

MD site-specific cancer incidence and mortality

MD was found to be associated with an increased risk of brain (n = 5; RR = 2.91; 95% CI: 1.55-5.44; I2 = 90.05%), lung (n = 9; RR = 1.31; 95% CI: 1.18-1.44; I2 = 1.4%), and pancreatic (n = 3; RR = 2; 95% CI:1.11-3.59; I2 = 96.02%) cancer incidence. In addition, for three cancer types, esophagus (n = 2; RR = 1.81; 95% CI: 1.38-2.38; I2 = 94.49%), larynx (n = 2; RR = 2.02; 95% CI: 1.3-3.11; I2 = 90.93%), and leukemia (n = 2; RR = 1.48; 95% CI: 1.17-1.87; I2 = 33.46%), significantly increased risk were also observed in meta-analyses carried out using only two observational studies. Depressed women were at a higher risk of both breast cancer incidence (n = 15; RR = 1.24; 95% CI: 1.02-1.52; I2 = 81.61%) and mortality (n = 5; RR = 1.28; 95% CI: 1.15-1.43; I2 = 0%) compared to controls. (**See Table 7**)

Outcome	Nº Studies	Nº of participants with CNS	Pooled RR (95% CI)	P-value	I^2	Q (P-value)
Brain and CNS	5	55147	2.91 (1.55-5.44)	8.70E-04	90.05%	40.19 (4e-08)
Breast (Women)	15	51645	1.24 (1.02-1.52)	3.30E-02	81.61%	76.14 (1.5e-10)
Breast (Women death)	5	3619	1.28 (1.15-1.43)	1.1e-05	0%	1.25 (8.7e-01)
Esophagus	2	37012	1.81 (1.38-2.38)	1.70E-05	94.49%	18.15 (2e-05)
Larynx	2	37012	2.02 (1.3-3.11)	1.60E-03	90.93%	11.02 (9e-04)
Leukaemia	2	35979	1.48 (1.17-1.87)	9.90E-04	33.46%	1.5 (2.2e-01)
Pancreas	3	45805	2 (1.11-3.59)	2.10E-02	96%	50.27 (1.2e-11)
Lung	9	55964	1.31 (1.18-1.44)	9.60E-08	1.40%	8.11 (4.2e-01)

Table 7: Significant associations between MD and site-specific cancer incidence and mortality.

BD and site-specific cancer incidence and mortality

BP patients showed higher stomach cancer incidence than controls (n = 5; RR = 1.37; 95% CI: 1.15-1.64; I2 = 0.53%) in a meta-analysis of five independent studies. Esophagus cancer risk was also found to be increased in BD patients after combing the information of two studies. (n = 2; RR = 2.56; 95% CI: 1.6-4.09; I2 = 0%). One study reported significant direct associations between BD and colorectal cancer mortality (n = 1; RR = 1.99; 95% CI: 1.29-3.05).

Outcome	Nº Studies	Nº of participants with CNS	Pooled RR (95% CI)	P-value	I^2	Q (P-value)
Colorectal (death)	1	6618	1.99 (1.29-3.05)	-	-	-
Esophagus	2	11845	2.56 (1.6-4.09)	8.40E-05	0%	0.83 (3.6e-01)
Stomach	5	38468	1.37 (1.15-1.64)	4.90E-04	0.53%	4.02 (4e-01)

Table 8: Significant associations between BD and site-specific cancer incidence and mortality.

SCZ site-specific cancer incidence and mortality

Breast (n = 18; RR = 1.37; 95% CI: 1.23-1.53; I2 = 85.06%) and uterine (n = 11; RR = 1.35; 95% CI, 1.07-1.7; I2 = 83.57%) cancer incidences were observed to be increased in women with schizophrenia compared to controls, as well as breast cancer mortality (n = 7; RR = 1.71; 95% CI: 1.32-2.22; I2 = 66.12%). In contrast, schizophrenic men were found to be at a reduced risk of prostate cancer (n = 14; RR = 0.56; 95% CI: 0.47-0.65; I2 = 53.46%). The risks of mortality due to colorectal (n = 5; RR = 1.71; 95% CI: 1.59-1.85; I2 = 0%), liver (n = 3; RR = 1.41; 95% CI: 1.26-1.57; I2 = 0%), lung (n = 5; RR = 2.23; 95% CI: 1.85-2.69; I2 = 42.6%) and pancreatic cancer (n = 4; RR = 1.4; 95% CI: 1.27-1.54; I2 = 0%) were also found to be increased in SCZ patients compared to controls. On the other hand, melanoma (n = 7; RR = 0.74; 95% CI: 0.64-0.86; I2 = 8.24%), nasopharynx (n = 7; RR = 0.74; 95% CI, 0.64-0.86; I2 = 8.24%) and thyroid (n = 6; RR = 0.69; 95% CI: 0.58-0.82; I2 = 0%) cancers presented a reduced incidence risk in the SCZ group.

Outcome	Nº Studies	Nº of participants with CNS	Pooled RR (95% CI)	P-value	I^2	Q (P-value)
Breast (Women)	18	189804	1.37 (1.23-1.53)	1.70E-08	85.06%	113.78 (2.3e-16)
Breast (Women deaths)	7	534375	1.71 (1.32-2.22)	6e-05	66.12%	17.71 (7e-03)
Colorectal (Deaths)	5	1160743	1.71 (1.59-1.85)	1.1e-44	0%	3.1 (5.4e-01)
Liver (Deaths)	3	1150198	1.41 (1.26-1.57)	1.4e-09	0%	1.68 (4.3e-01)
Lung (Deaths)	5	1158845	2.23 (1.85-2.69)	5.4e-17	42.6%	6.97 (1.4e-01)
Malignant skin melanoma	7	150049	0.74 (0.64-0.86)	9.50E-05	8.24%	6.54 (3.7e-01)
Nasopharynx	7	308465	0.74 (0.64-0.86)	9.50E-05	8.24%	6.54 (3.7e-01)

Pancreas (Deaths)	4	1152466	1.4 (1.27-1.54)	5.3e-12	0%	0.14 (9.9e-01)
Prostate (Men)	14	212109	0.56 (0.47-0.65)	7.10E-13	53.46%	27.94 (9.2e-03)
Thyroid	6	281469	0.69 (0.58-0.82)	1.70E-05	0%	4.66 (4.6e-01)
Uterus (Women)	11	163062	1.35 (1.07-1.7)	1.20E-02	83.57%	60.88 (2.5e-09)

Table 9: Significant associations between Schizophrenia patients and site-specific cancer incidence and mortality.

ASD site-specific cancer incidence and mortality

ASD studies reporting site-specific cancers were scarce. Only two significant results derived from the combination of two studies each were observed. The incidence risk of both tumor types, ovarian cancer (n = 2; RR = 10.26; 95% CI: 2.29-45.95; I2 = 0%) and eye cancer (n = 2; RR = 9.1; 95% CI: 1.16-71.67; I2 = 0%) was found to be increased in ASD patients compared to controls.

Outcome	Nº Studies	Nº of participants with CNS	Pooled RR (95% CI)	P-value	I^2	Q (P-value)
Eye	2	8556	9.1 (1.16-71.62)	3.60E-02	0%	0 (9.7e-01)
Ovary (Women)	2	3344	10.26 (2.29-45.95)	2.30E-03	0%	0.07 (7.9e-01)

Table 10: Significant associations between ASD and site-specific cancer incidence and mortality.

Supplementary Appendix 1 Figures 29 to 76 show the forest and funnel plots of the secondary outcomes for which ten or more observational studies were available, whereas **Supplementary Appendix 1 Tables 1 to 7** present the complete meta-analysis results for all the CNS site-specific cancer analyses, both significant and non-significant.

2.3.4 Additional analyses results

In this section, we describe the additional analysis results. We start by reporting the pooled meta-analysis results of two subgroups of CNS disorders. Those whose main underlying processes are neurodegeneration (AD, HD, and PD) and those considered to be neuropsychiatric disorders (MD, BD, SCZ). The motivation to group neurodegenerative disorders together is that the underlying biological processes linked to neurodegeneration could be responsible for the reduced cancer incidence observed. Besides, neuropsychiatric disorders have been studied in combination in a number of previous publications, and patients suffering from them are known to be exposed to several shared risks, including unhealthy behaviors and impaired access to health care systems.

Second, we will present the sensitivity analysis results. The potential influence of several factors (Ethnicity, follow-up duration, number of participants with a specific CNS disorder included in the study, design, study quality, and smoking-related cancers versus non-smoking related cancers) on the meta-analyses results was evaluated when feasible. Those analyses were carried out only for the study's primary outcomes (all-cancer incidence and mortality).

Finally, we assess the risk of publication bias for those meta-analyses, including information from ten or more observational studies are presented by means of Begg's and Egger's tests.

2.3.4.1 Associations between neurodegenerative and neuropsychiatric disorders and all-cancer incidence and mortality

Patients with neurodegenerative disorders presented a lower risk of all-cancer incidence (n = 28; RR = 0.73; 95% CI: 0.66-0.81; I2 = 97.4%) and mortality (n = 23; RR = 0.53; 95% CI: 0.42-0.67; I2 = 83.92%) compared to controls. A significant reduction in cancer incidence was observed in both men (n = 15; RR = 0.86; 95% CI: 0.74-1; I2 = 97.02%) and women (n = 14; RR = 0.88; 95% CI: 0.78-0.99; I2 = 94.09%). Gender-specific analyses also showed a reduced mortality risk in men (n = 3; RR = 0.56; 95% CI: 0.4-0.77; I2 = 0%) and women (n = 2; RR = 0.62; 95% CI: 0.4-0.97; I2 = 19.42%). However, a small number of studies (3 and 2) were available in this case.

A slightly increased all-cancer incidence was observed in neuropsychiatric patients compared to controls (n = 56; RR = 1.06; 95% CI: 1-1.13; I2 = 95.36%). This trend was not observed in men (n = 35; RR = 0.94; 95% CI: 0.85-1.04; I2 = 96.39%), but it was also present in women (n = 34; RR = 1.11; 95% CI: 1.03-1.21; I2 = 94.71%). In the case of mortality due to cancer, neuropsychiatric patients were found to be at an increased risk compared to controls (n = 71; RR = 1.27; 95% CI: 1.19-1.35; I2 = 95.01%). This trend was found to be present in both men (n = 42; RR = 1.25; 95% CI: 1.13-1.38; I2 = 93.91%) and women (n = 33; RR = 1.48; 95% CI: 1.46-1.5; I2 = 96.15%). However, all-cancer mortality was found to be significantly higher in women than in men with neuropsychiatric conditions.

2.3.4.2 Additional/Sensitivity analyses.

Additional analyses were carried out for potential confounding factors, including ethnicity, follow-up duration, the total number of patients with a particular CNS disorder, study design, and the study quality. A comparison between smoking-related cancers and cancers that have not been linked to cigarette consumption was also carried out. Meta-analyses in which only two data points were available are marked with an asterisk symbol, whereas items for which only one study was available are identified using a plus symbol in **Tables 11 and 12**. All other results are derived from meta-analyses for which information regarding three or more studies was available.

In the case of AD, we observed significant differences between the studies carried out in non-Asian (RR = 0.5, 95% CI: 0.34-0.73) and Asian (RR = 0.85, 95% CI: 0.8-0.9) populations. Significant differences were also observed between those studies including less than 500 AD patients (RR = 0.33, 95%: 0.22-0.47) and those studies reporting to include more than 1000 (RR = 0.79, 95% CI: 0.69-0.89). PD patients were at a significantly lower risk for smoking-related cancers (RR = 0.81, 95% CI: 0.75-0.88) compared to those cancers that have not been linked to tobacco

consumption (RR = 1.04, 95% CI: 0.91-1.2). The opposite pattern was observed in MD patients who were found to be at a higher risk of developing smoking-related cancers (RR = 1.36, 95% CI: 1.12-1.65) than controls but were not found to be associated with non-smoking related cancers (RR = 0.99, 95% CI: 0.72-1.35).

Studies including less than 500 PD participants showed significantly higher all-cancer mortality rates (RR = 0.82, 95% CI: 0.65-1.04) than those in which the number of participants was higher than 1000 (RR = 0.52, 95% CI: 0.37-0.74). However, in both cases, the reported associations showed a trend towards a reduction in all-cancer mortality, with only the latter reaching significant values.

Factor		AD all-cancer incidence	ASD all-cancer incidence	BD all-cancer incidence	HD all-cancer incidence	PD all-cancer incidence	MD all-cancer incidence	SCZ all-cancer incidence
Race	**Asian**	0.85 (0.8-0.9)	1.91 (0.52-4.88)+	1.36 (1.25-1.48)*	-	0.98 (0.57-1.68)	1.35 (0.91-2)	1.08 (0.88-1.33)
	Non-Asian	0.5 (0.34-0.73)	0.8 (0.43-1.49)	1.04 (0.76-1.41)	0.48 (0.29-0.78)	0.82 (0.76-0.89)	1.06 (1.01-1.11)	0.97 (0.89-1.07)
Follow-up	**>1-5**	0.72 (0.58-0.88)	-	1.03 (0.93-1.15)+	-	0.87 (0.46-1.63)	1.17 (0.94-1.44)	1.1 (1.01-1.21)
	>5	0.29 (0.18-0.47)*	1.41 (0.88-2.26)	1.29 (1.03-1.61)	0.59 (0.46-0.76)	0.87 (0.81-0.95)	1.16 (1.02-1.31)	0.98 (0.91-1.06)
N total	**<500**	0.33 (0.22-0.47)	1.36 (0.54-3.21)+	-	-	1.12 (0.88-1.43)	1.25 (1.03-1.5)	1.27 (0.78-2.06)+
	>500-1000	-	-	-	0.6 (0.5-0.8)	0.51 (0.25-1.06)	1.21 (1.03-1.44)	0.76 (0.54-1.06)*
	>1000	0.79 (0.69-0.89)	0.96 (0.41-2.24)	1.1 (0.87-1.39)	0.62 (0.54-0.7)	0.85 (0.74-0.97)	1.11 (0.96-1.28)	0.99 (0.92-1.07)
Design	**Case-control**	0.85 (0.79-0.9)*	0.65 (0.39-1.1)*	-	0.6 (0.5-0.8)+	0.77 (0.72-0.83)*	1.21 (1.02-1.42)*	-
	Cohort	0.69 (0.56-0.85)	1.49 (1.03-2.15)*	1.1 (0.87-1.39)	0.44 (0.24-0.83)	0.86 (0.75-0.97)	1.16 (1.04-1.29)	0.99 (0.92-1.07)
Quality group	**Low-risk**	0.61 (0.49-0.77)	1.49 (1.03-2.15)*	1.12 (0.97-1.28)	0.52 (0.45-0.61)*	0.93 (0.82-1.07)*	1.18 (1.03-1.35)	0.99 (0.91-1.08)
	Moderate-risk	0.86 (0.81- 0.92)+	0.65 (0.39-1.1)*	2.06 (1.76-2.41)*	0.43 (0.39-0.48)*	0.58 (0.44-0.76)	1.09 (1.02-1.16)	0.96 (0.92-1.01)
	High-risk	-	-	0.27 (0.21-0.34)+	-	-	1.35 (0.62-2.93)+	-
Smoking related cancers	**Yes**	0.88 (0.82-0.94)	3.61 (1.12-11.68)	1.14 (0.94-1.39)	0.58 (0.44-0.77)	0.81 (0.75-0.88)	1.36 (1.12-1.65)	0.96 (0.89-1.03)
	No	0.69 (0.52-0.91)	2.48 (0.83-7.4)	1.08 (0.84-1.4)	0.84 (0.6-1.18)	1.04 (0.91-1.2)	0.99 (0.72-1.35)	0.73 (0.65-0.81)

Table 11: Sensitivity analysis results for all-cancer incidence. Results marked with an asterisk (*) indicate results derived from the combination of only two studies. Results marked with a plus symbol (+) indicate data derived from a single available study.

Factor		AD all-cancer mortality	ASD all-cancer mortality	BD all-cancer mortality	HD all-cancer mortality	PD all-cancer mortality	MD all-cancer mortality	SCZ all-cancer mortality
Race	**Asian**	-	-	1.35 (0.58-3.15)+	-	0.53 (0.23-1.23)+	1.64 (1.05-2.58)	1.14 (0.87-1.48)*
	Non-Asian	0.45 (0.33-0.61)	1.92 (1.58-2.32)	1.16 (1.09-1.24)	0.12 (0.08-0.17)	0.7 (0.55-0.9)	1.23 (1.14-1.33)	1.35 (1.24-1.47)
Follow-up	**>1-5**	0.83 (0.4-1.53)	-	-	-	0.52 (0.37-0.74)	1.47 (1.21-1.78)	0.99 (0.54-1.82)
	>5	0.48 (0.35-0.66)	2.59 (1.63-4.1)+	1.17 (1.09-1.26)	0.05 (0.02-0.12)+	0.69 (0.52-0.91)	1.19 (1.1-1.29)	1.34 (1.22-1.48)
N total	**<500**	0.52 (0.4-0.69)	-	0.97 (0.71-1.35)	0.05 (0.02-0.12)	0.82 (0.65-1.04)	1.4 (1-1.97)	1.42 (1.09-1.85)
	>500-1000	0.35 (0.26-0.48)	-	2.05 (0.65-4.95)+	-	0.58 (0.4-0.85)+	0.93 (0.55-1.56)	0.88 (0.63-1.23)*
	>1000	-	1.92 (1.58-2.32)*	1.17 (1.09-1.25)	-	0.53 (0.49-0.58)	1.27 (1.17-1.37)	1.31 (1.18-1.44)
Design	**Cohort**	0.51 (0.4-0.64)	2.59 (1.63-4.1)+	1.16 (1.09-1.24)	0.12 (0.08-0.17)*	0.69 (0.54-0.87)	1.24 (1.15-1.34)	1.34 (1.23-1.46)
	Case-control	0.25 (0.16-0.41)+	1.8 (1.46-2.23)+	-	-	-	-	-
Quality group	**Low-risk**	0.63 (0.4-0.99)	1.92 (1.58-2.32)*	1.15 (1.07-1.24)	0.14 (0.093-0.212)	0.69 (0.54-0.87)	1.32 (1.22-1.44)	1.41 (1.28-1.56)
	Moderate-risk	0.39 (0.26-0.57)	-	1.21 (0.99-1.49)	-	0.67 (0.42-1.06)	1.15 (0.98-1.35)	1.2 (0.99-1.46)
	High-risk	0.44 (0.14-1.38)+	-	2.05 (0.65-4.95)+	0.05 (0.02-0.12)	0.98 (0.22-4.35)+	0.91 (0.76-1.09)	0.35 (0.23-0.55)+

Table 12: Sensitivity analysis results for all-cancer mortality. Results marked with an asterisk (*) indicate results derived from the combination of only two studies. Results marked with a plus symbol (+) indicate data derived from a single available study.

2.3.4.3 Assessment of the risk of publication bias using Egger's and Begg's tests.

For those meta-analyses including information regarding ten or more different observational studies, we investigated if publication bias was present by means of visual inspection of the resulting funnel plots (**Supplementary Appendix 1 Figures 29 to 76**) and by means of two statistical tests design to detect asymmetries in the funnel plots (Egger's and Begg's tests). In general, we did not found evidence of publication bias. Only the SCZ all-cancer mortality meta-analyses, including the overall analysis and the analyses stratified by sex, presented p-values < 0.05 in the Egger's test, as well as the cervical cancer incidence analysis in women with SCZ.

Meta-analysis	Egger's test results	Begg's test results.
BD (All cancer mortality)	(n = 14; t = 1.15; p-val = 2.73e-01)	(n = 14; t = 0.13; p-val = 5.11e-01)
MD (All cancer incidence)	(n = 22; t = 1.08; p-val = 2.92e-01)	(n = 22; t = 0.29; p-val = 6.22e-02)
MD (All cancer mortality men)	(n = 12; t = 1.48; p-val = 1.69e-01)	(n = 12; t = 0.06; p-val = 8.41e-01)
MD (Breast cancer women)	(n = 15; t = 1.53; p-val = 1.49e-01)	(n = 15; t = 0.33; p-val = 9.26e-02)
MD (Prostate cancer men)	(n = 12; t = 0.35; p-val = 7.31e-01)	(n = 12; t = 0.21; p-val = 3.81e-01)
PD (All cancer incidence)	(n = 17; t = -0.32; p-val = 7.54e-01)	(n = 17; t = 0.1; p-val = 5.98e-01)
PD (All cancer incidence men)	(n = 12; t = -0.13; p-val = 9.01e-01)	(n = 12; t = 0.21; p-val = 3.81e-01)
PD (All cancer incidence women)	(n = 11; t = 0.14; p-val = 8.93e-01)	(n = 11; t = 0.13; p-val = 6.48e-01)
PD (All cancer mortality)	(n = 15; t = 1.53; p-val = 1.51e-01)	(n = 15; t = 0.2; p-val = 3.28e-01)
PD (Breast cancer women)	(n = 11; t = -1.08; p-val = 3.07e-01)	(n = 11; t = -0.2; p-val = 4.45e-01)
PD (Colorectal cancer)	(n = 15; t = -0.91; p-val = 3.78e-01)	(n = 15; t = 0.07; p-val = 7.70e-01)
PD (Lung cancer)	(n = 13; t = -0.98; p-val = 3.48e-01)	(n = 13; t = 0.08; p-val = 7.65e-01)
PD(Malignant skin melanoma)	(n = 13;; t = 1.13; p-val = 2.82e-01)	(n = 13; t = 0.08; p-val = 7.65e-01)
PD (Prostate cancer men)	(n = 15; t = -0.85; p-val = 4.10e-01)	(n = 15; t = 0.18; p-val = 3.79e-01)
SCZ (All cancer incidence)	(n = 26; t = 0.6; p-val = 5.54e-01)	(n = 26; t = 0.05; p-val = 7.60e-01)
SCZ (All cancer incidence men)	(n = 22; t = 1.84; p-val = 8.05e-02)	(n = 22; t = 0.13; p-val = 4.34e-01)
SCZ (All cancer incidence women)	(n = 20; t = -0.74; p-val = 4.71e-01)	(n = 20; t = 0.04; p-val = 8.23e-01)
SCZ (All cancer mortality)	(n = 37; t = -3.09; p-val = 3.92e-03)	(n = 37; t = 0.04; p-val = 7.65e-01)
SCZ (All cancer mortality men)	(n = 22; t = -2.3; p-val = 3.24e-02)	(n = 22; t = 0.03; p-val = 8.67e-01)
SCZ (All cancer mortality women)	(n = 19; t = -2.57; p-val = 1.99e-02)	(n = 19; t = 0.12; p-val = 4.89e-01)
SCZ (Bladder cancer)	(n = 10;; t = 2.16; p-val = 6.25e-02)	(n = 10; t = 0.2; p-val = 4.84e-01)
SCZ (Brain and CNS)	(n = 11; t = -0.19; p-val = 8.51e-01)	(n = 11; t = -0.02; p-val = 1.00e+00)
SCZ (Breast cancer women)	(n = 18; t = -0.25; p-val = 8.06e-01)	(n = 18; t = 0.1; p-val = 6.01e-01)
SCZ (Cervix cancer women)	(n = 10; t = -2.62; p-val = 3.06e-02)	(n = 10; t = -0.2; p-val = 4.84e-01)
SCZ (Colorectal cancer)	(n = 14; t = 0.89; p-val = 3.90e-01)	(n = 14; t = 0.05; p-val = 8.30e-01)
SCZ (Kidney cancer)	(n = 10; t = 1.13; p-val = 2.90e-01)	(n = 10; t = 0.16; p-val = 6.01e-01)
SCZ (Lung cancer)	(n = 16; t = -0.41; p-val = 6.92e-01)	(n = 16; t = -0.17; p-val = 3.98e-01)
SCZ (Prostate cancer men)	(n = 14; t = 0.72; p-val = 4.83e-01)	(n = 14; t = 0.14; p-val = 5.18e-01)
SCZ (Stomach cancer)	(n = 10; t = 0; p-val = 9.99e-01)	(n = 10; t = -0.07; p-val = 8.62e-01)
SCZ (Uterus cancer women)	(n = 11; t = -1.32; p-val = 2.19e-01)	(n = 11; t = -0.02; p-val = 1.00e+00)

Table 13: Tests for funnel plot asymmetry for those meta-analyses including ten on more studies.

2.4. Discussion

Our results suggest that CNS disorders present diverse patterns of associations with cancer. On the one hand, all three neurodegenerative disorders (AD, HD, and PD) were significantly linked to a reduction in the risk of subsequent overall cancer incidence and mortality. In the case of cancer incidence, HD presented the sharpest decrease in RR (52%), followed by AD (31%) and PD (15%). The same ordering was found in the case of cancer mortality with HD, AD, and PD, showing relative risk reductions of 78%, 55%, and 31%, respectively. However, the associations regarding HD should be taken with special caution due to the small number of studies included in the analyses. Gender-specific analyses could not be performed for mortality in the case of AD nor for incidence or mortality in the case of HD due to the lack of data. Women and men with AD presented significant reductions in the relative risks of cancer incidence. In the case of PD patients, we did not observe significant reductions in cancer incidence, neither for men nor for women. However, a trend towards a risk reduction was observed in both cases (12% and 7% of relative risk reduction, respectively).

In addition to the body of research examining cancer incidence following the diagnosis of neurodegenerative disorders, several studies have investigated the reverse association. In the case of AD, at least seven observational studies have reported significantly reduced AD incidence after cancer diagnosis [108-110, 147-150]. However, previous data have pointed towards the possibility that cognitive impairment could arise as a result of cancer therapy. For instance, meta-analyses have provided some evidence for a possible link between chemotherapy and impairments in the domains of memory and executive function [151], and cranial irradiation could induce neuronal damage and loss [152]. In the case of PD, at least two works have suggested reduced PD risks after cancer diagnosis [153, 154], whereas some others have shown non-significant trends towards risk reduction or no associations [155-159]. Finally, one study found low cancer prevalence in Huntington's disease patients compared to controls [160].

Different factors could account for the inverse patterns of comorbidity observed between neurodegenerative disorders and cancer.

First, both neurodegenerative disorders and cancer are characterized by the presence of abnormal cell behaviors. While a progressive neural loss is a hallmark of neurodegeneration, cancer is characterized by uncontrolled proliferation and survival of cells. The reported inverse comorbid associations between both sets of conditions indicate that neurodegeneration mechanisms could have a protective effect against cancer development, which translates into a reduced cancer incidence and mortalities.

Other biological factors could also be the potential underlying causes of the observed inverse comorbidities. For instance, *PIN1* is thought to play a role in the pathogenesis of both cancer and AD [161]. Deletions of this gene produce AD-like pathologic changes in mice [162], and its expression levels have been found to be downregulated in AD brain tissues and upregulated in several cancer types tissues [148, 163, 164]. Alterations in the PI3K/Akt/mTOR axis are a shared feature of both AD and cancer [165]. Estrogen has been found to reduce the risk of AD [166] and to increase the risk of certain cancers such as endometrial and breast cancer [167]. Given the cell proliferation stimulating effect of acetylcholine, the cholinergic system deficits found in AD patients have also been proposed as a potential candidate to explain the inverse patterns of comorbidity between AD and cancer [108, 168]. AD is associated with the degeneration of cholinergic neurons and reduced nicotinic acetylcholine receptor (nAChRs) signaling. In addition, acetylcholine and nAChRs can induce the synthesis and release of neurotrophins, growth factors, and angiogenic factors that can stimulate cancer proliferation. The inhibition of nAChRs causes apoptosis in cancerous cells [169]. Finally, Aβ, which plays a prominent role in AD pathogenesis, can activate the tumor suppressor p53, resulting in p53-dependent apoptosis. Several studies have found p53 to be upregulated in AD brains compared to controls [170].

In the case of PD, some genes linked to its familial forms (*INK1*, *LRRK2*, C-terminal hydroxylase) have been found to be desregulated in tumor tissues [171]. The joint involvement of many biological processes has also been reported in both PD and cancer, including DNA damage, response to oxidative stress, metabolic dysregulation, and alterations in the ubiquitin-proteasome system. In addition, high levels of cholesterol and fatty acids have been linked to lower risks of PD and cancer-promoting effects [172].

Regarding HD, mutant huntingtin has been shown to present a pro-apoptotic effect [173] based on the toxicity of the glutamine expansions (polyQ) present on the HTT protein. An alternative mechanism involving the role of the CAG triplet expansion at an RNA level has been shown to slow down tumor growth progression [174].

Second, the effect of some potential confounding covariates, including age, life expectancies, and smoking status, is thought to modulate the observed associations.

Aging is known to be the main risk factor for both AD and cancer [175, 176]. AD patients often present reduced life expectancies [177-182]**.** This could be associated with the reduced cancer risk observed in AD patients. However, all the observational studies reporting AD and cancer incidence associations used in our analyses accounted for age either by design or by adjustment.

Differences in patients' smoking habits could also constitute a confounding factor. For instance, AD patients are thought to smoke less than the general population, which could explain the observed reduced cancer rates. In fact, lung cancer was found to be significantly reduced in our

site-specific cancer analyses (with a relative risk reduction of 19%). However, our results suggest that AD is associated with a risk reduction of smoking-related (12%) and smoking-unrelated cancers (31%). Therefore, in the light of our data, it seems unlikely that differences in the smoking habits account for all the cancer risk reductions observed in AD patients.

A different conclusion can be drawn in the case of PD. PD patients have been shown to be lifelong non-smokers [183]. Our subgroup analyses showed that PD patients presented a significant reduction in the risk of smoking-related cancers (19%) but not in the risk of the smoking unrelated group of site-specific cancers. However, these conclusions are difficult to generalize to the overall cancer analyses since the vast majority of observational studies reporting PD and all-cancer associations did not account for the smoking status by means of study design or adjustment. If smoking habits in PD patients can account for all the differences in cancer incidence and mortality observed could be only determined by further epidemiological research.

Third, some authors have suggested that the associations arise as a consequence of the intrinsic limitations of the epidemiologic studies or systematic biases.

Cancer screening and diagnosis could be compromised in the case of patients with dementia. A handful of epidemiologic studies have also acknowledged the possible existence of ascertainment bias in cognitively impaired persons [184-187], and at least one study has found that AD patients present stronger inverse associations in older age groups (higher neurocognitive impairment) [147]. Patients with dementia have been found to present difficult issues of communication and consent to testing [188] and lower rates of cancer screening[189]. It has also been suggested that AD patients tend to be diagnosed with cancers at more advanced stages [190].

Cancer is also thought to be underdiagnosed in HD patients because its signs and symptoms may not be noticed or covered by HD symptoms. One interesting example in this regard is cachexia, which is a symptom common in both cancers and the latter stages of HD progression. In contrast, some authors have found an increased cancer diagnosis rate in the first year after admission for HD, suggesting that overdiagnosis rather than underdiagnoses are more likely in this population [191]. Finally, claims about the role of specific medications as modulating agents in the inverse comorbid association found between CNS and cancer have also been made. In this regard, some interesting hypotheses include the observations that the use of levodopa, dopamine agonists, monoamine oxygenase-B inhibitors, and anticholinergics possibly affect tumor growth and progression [192]. In addition, two kinds of cancer chemotherapy (bexarotene and carmustine) could induce the clearance of physiological Aβ and reduce the cognitive deficits produced by its accumulation in mice models. A more exhaustive description of the role of medications as comorbidity modulators will be provided in **Chapter 5**.

BD, SCZ, and ASD patients did not present altered all-cancer incidence rates, whereas MD patients were found to be at higher risk of subsequent overall cancer incidence (17%). In contrast, BD, SCZ, ASD, and MD presented a significant increase in cancer mortality (16%, 34%, 92%, and 24%).

In principle, neuropsychiatric patients are exposed to several factors that would make them prone to cancer. Stress, a common trait found in different mental disorders, is associated with physiological changes that promote cancer [193]. Mental disorders are also linked to poor health behaviors, which are associated with increases in cancer risk, including smoking, alcohol consumption, and obesity, among others [194, 195].

Those factors could partially account for the increased risk of cancer in MD patients observed in our analysis. Our results also provide additional evidence in this direction since the risk of smoking-related cancers was found to be significantly increased in MD patients (36% increase), whereas the risk of smoking unrelated cancers did not. Furthermore, depression has been linked to other cancer-promoting factors such as disturbances in the autonomic nervous system, increased inflammation, and the reduction in circulating endothelial progenitor cells [196-198].

In the case of schizophrenia, gender-based differences were observed for overall cancer incidence, with men presenting a significant reduction in cancer risk (13% relative risk reduction) and women showing a non-significant trend towards an increase in cancer risk (8%). This disagreement could be due to the significant increase in breast cancer incidence and mortality risks observed in schizophrenic women (37% and 71% increase, respectively) and the reduction of prostate cancer risk observed in schizophrenic men (44% reduction in prostate cancer). According to the World Health Organization, breast cancer and prostate cancer are the second and fourth most incident cancers and probably significantly impact the sex-specific overall cancer incidence associations. In the case of MD, gender-specific analyses for incidence did not yield significant results, but a trend towards an increase in the relative risks was observed. This is probably due to the inclusion of a reduced number of studies compared with the joint analysis of both genders.

. Some authors have proposed that SCZ would present a protective effect against cancer. Catts and co-workers examined the results of three studies of cancer incidence in first degree relatives of SCZ patients and found a reduced cancer incidence [86]. However, according to our results, SCZ patients do not present a lower-than-expected risk of overall cancer. In this context, it has been argued that the lifestyle differences observed in SCZ patients compared to controls (high smoke and alcohol consumption) would compensate for the underlying genetic protective effect [199].

Patients with severe mental illnesses such as SCZ and BP present higher overall mortality rates than controls [200, 201]**,** and SCZ is more common in young adults, whereas most cancers are

diagnosed in patients older than 60 years, which makes it important to control for age in the studies. All the studies included in the meta-analysis of cancer incidence in SCZ patients accounted for age as a confounding factor, reducing the likelihood for this factor to distort the study outcomes.

A number of different but not mutually-exclusive explanations could account for the increased cancer mortality rates observed in patients with neuropsychiatric disorders. First, the mental disorder itself could complicate the disease's treatment, including the radiotherapy administration or the adherence to the pharmacological treatment. Second, severe mental disorders have been associated with alterations in cancer screening, which could explain the higher cancer death rates observed in the case of BD, MD, and SCZ. Howard and colleagues [199] gathered data from 12 observational studies investigating cancer screening services' uptake by patients with mental illness. Six of the available studies suggested that patients with severe mental illness presented lower cancer screening rates than controls, whereas the rest of them reported no associations. The authors also noted that the studies reporting no associations included fewer patients with psychotic disorders. Third, the inequity of access to care and treatment has also been observed in people with psychiatric disorders [199]. A variety of factors can explain unequal access. Delays in help-seeking have been found to be longer in psychiatric patients [202]. General physicians tend to underestimate the importance of neuropsychiatric patients' physical symptoms and attribute them to psychological or psychiatric causes, a term known as diagnostic overshadowing [203]. Some studies report that the cancer treatment of patients with severe mental illnesses presents specific challenges due to the frequent co-occurrence of cardiovascular and respiratory disease, obesity, and malnourishment [199]. Finally, specific oncological treatments can interact with psychotropic medications producing adverse-effects. For instance, clozapine used in conjunction with cytotoxic drugs can result in agranulocytosis, increasing infection risk. Several psychotropic drugs are known to inhibit the metabolism of chemotherapy, including clomipramine, duloxetine, haloperidol, paroxetine, sertraline, and fluoxetine, which can increase the side-effects of chemotherapy, whereas other psychotropic drugs such as carbamazepine have been shown to increase the metabolism of chemotherapy reducing their efficacy [199]. All these factors could account for the higher-than-expected probability of cancer mortality yielded by our analyses.

Different patterns of association were observed between CNS disorders and site-specific cancers. We will limit the discussion to those site-specific cancers for which significant associations were found in meta-analyses including three or more studies.

The incidence risks of lung cancer and malignant skin melanoma were reduced in AD, colorectal and prostate cancer risks were found to be lower in patients with HD, whereas PD patients presented a reduced risk in bladder, colorectal, larynx, lip and oral cavity, and lung cancers. Colorectal and lung cancer mortality were also found to be reduced in PD patients, as well as

pancreatic and stomach cancer. CNS and site-specific cancer associations which presented an opposed comorbidity pattern than the one observed in all-cancer analyses are particularly interesting. For instance, PD patients were found to be at an increased risk of brain and malignant skin melanoma. PD is characterized by the loss of dopaminergic neurons in the substantia nigra pars compacta (SNpc) and a depletion of dopamine in the striatum [204, 205]**.** Pigmentation has been found to play a crucial role in both PD and melanoma. Melanin is responsible for the hair and skin color pigmentation in humans. In the brain, pigmented neurons produce a pigment known as neuromelanin. Abnormalities in both melanin and neuromelanin have been associated with both cancer and PD, providing a potential link between both disorders. Besides, light skin populations present higher incidences of melanoma than darker skin populations [206, 207]**.** This protective effect is thought to be due to melanin's capacity to scatter or absorb UVR light, preventing DNA damage, which leads to skin cancers. The lack of neuromelanin in human neurons makes them more susceptible to oxidative stress [208, 209]**.** Other mechanisms have been proposed to explain the co-occurrence of PD and melanoma. First, the treatment with levodopa was found to increase the risk of melanoma in some studies [210, 211]. However, this hypothesis remains controversial, and at least one epidemiological study found no associations between levodopa treatment and the incidence of melanoma [212]. Shared genetic alterations, including genes with roles in cell detoxification such as *CYPD6* and *GSTM1*, and genes linked to some forms of PD such as Parkin are other potential modulating agents.

Major depression was associated with an increased risk of brain, lung, and pancreatic cancer. MD women were found to be at a higher risk of both breast cancer incidence and mortality. Some hypotheses state that depression influences the neuroendocrine axis and the neuro-immunological function. This CNS dysregulation would put the organism at greater risk of morbidity and mortality [213]. Moreover, depression renders breast cancer survivors less capable of functioning successfully in modern society. People with depression are less likely to receive appropriate and correct health screening [214]. Some studies suggest that cancer survivors may be faced with adherence difficulties to chemotherapy and other complementary treatment, leading to faster cancer progression [213] . Finally, the use of tricyclic anti-depressants has been associated with a higher risk of breast cancer [215, 216].

Women with SCZ presented an increased risk of breast and uterine cancer incidence. Breast cancer mortality was also found to be increased in SCZ patients. The higher incidence of breast cancer found in SCZ women could be due to the lower parity observed in women with SCZ [217, 218]**.** The use of antipsychotic drugs, such as risperidone to treat these patients, has also been suggested to increase breast risk through hyperprolactinemia [219]. In contrast, the use of some neuroleptic medications has been proposed to reduce the risk of subsequent cancer. Women with

schizophrenia have a higher prevalence of other breast cancer risk factors, including unhealthy lifestyle, obesity, and diabetes [220-222]. SCZ men were found to be at a reduced risk of prostate cancer. The risks of mortality due to colorectal, liver, lung, and pancreatic cancer were also found to be increased in SCZ patients compared to controls.

In summary, the landscape of the associations between CNS disorders and cancer emerges as a complex picture with different associations pointing towards different directions. A large set of possible confounding factors and systematic biases difficult the interpretation of the results. Moreover, our study presents some limitations that will be summarized next. First, although the body of evidence devoted to the study of CNS and cancer associations has grown in the last decades, the number of available studies was insufficient to draw robust conclusions in many of the performed analyses. In the case of primary outcomes, there was an uneven distribution of studies available for the different disorders, which was especially low in the case of HD and ASD. This issue was worse in the case of secondary outcomes in which only a handful of site-specific versus CNS cancer associations could be investigated a sufficient number of studies. The lack of data was critical in the case of the site-specific cancer mortality analyses. Second, the effect of potential confounding factors cannot be properly appraised. Besides age and gender, which are usually accounted for in observational studies, other factors are rarely adjusted for or marched by design. An important example in this regard is the smoking status of the patients. Smoking is a major risk factor for cancer, and despite this fact, we observed that the inclusion of information regarding it in observational studies is the exception rather than the norm. We carried out a stratified analysis of smoking-related versus smoking-unrelated cancers. However, those analyses were based on data regarding site-specific cancers. Further epidemiological research will be needed to clarify the effect of smoking as a modulating agent in the comorbidity patterns observed between CNS disorders and overall-cancer.

This limitation also applies to other confounding factors such as the family history of cancer, the body mass index, the indicators of physical activity or alcohol consumption, as well as data regarding the parity status of women hormonal related cancers such as breast cancer.

Furthermore, some intrinsic limitations of meta-analyses should also be taken into account (e.g., the inclusion of studies with different designs, settings, populations, treatment strategies, and diagnostic tools).

In the following chapters, we will try to identify some of the potential modulating agents involved in CNS and cancer comorbidities, such as the presence of joint alterations in genes and pathways and the roles of the genetic variability and medications.

Chapter 3: Transcriptomic associations between CNS disorders and cancer

3.1 Introduction

In the previous chapter, we synthesized the epidemiological evidence regarding the associations between a selection of CNS disorders and cancer incidence and mortality, which led to the identification of diverse comorbidity patterns between them. Several factors have been proposed to account for the observed associations, including the presence of shared biological alterations such as the altered expression of genes and pathways.

Since its emergence in the early two-thousands, high throughput technologies have been applied to query the transcriptomes of human disorders. In particular array-based methods have been widely employed to compare the gene expression levels of healthy versus diseased tissues by means of differential gene expression analysis. The evolution of the field in the subsequent decades involving the improvement of methodologies and the reduction of their cost has led to the placement of many of this data in public repositories freely accessible to the scientific community.

Differential gene expression profiles provide information about genes and biological pathways that are deregulated in a particular disorder, and they can be regarded as a phenotypic manifestation of the disease. Therefore, for a given pair of disorders, the analysis of their similarities and differences can be used as an instrument to shed light upon the genetic factors and biological processes underlying the comorbid associations observed at a population level.

The availability of multiple studies comparing the changes in gene expression in specific disorders carried out using different cohorts of patients and controls by diverse research groups opens the door to the application of differential gene expression meta-analysis methods to combine the results. Meta-analysis methods in the context of gene expression have been shown to increase the statistical power, reduce the noise of gene expression measurements and the presence of false positives, and improve the estimates of effect sizes [223, 224].

Two major meta-analysis methods are available to synthesize the evidence from individual differential gene expression studies, those based on the combination of p-values, such as Fisher's method, and those based on the combination of effect sizes. We selected the method proposed by Choi et al. [36], which is based on the combination of the effect sizes of individual studies under a random effects model. It provides information about the strength and direction of gene expression changes and allows the downstream pathway-level analysis using methods like Gene Set Enrichment Analysis (GSEA).

It has been previously hypothesized that population-level direct and inverse comorbidity patterns between a given pair of disorders could translate into the presence of significant patterns of deregulation of genes and pathways in the same or opposite directions, respectively [30, 31]. Accordingly, for a given pair of directly comorbid disorders, a significant overlap is expected in the number of jointly upregulated and downregulated genes. In contrast, in the case of a pair of diseases exhibiting epidemiological patterns of inverse comorbidity (lower-than-expected probability of co-occurrence), genes and pathways upregulated in one disorder are expected to be downregulated in the other and vice versa.

Following this principle, Ibañez and co-workers [34] observed significant opposite patterns of transcriptomic deregulation between a set of three CNS disorders (Alzheimer's disease, Parkinson's disease, and schizophrenia) and three tumor types (colorectal, lung, and prostate cancer), which had been previously reported to be inversely associated by observational studies. The authors also provided functional insight, identifying specific genes and pathways deregulated in opposite directions, including *PIN1* and *ATP13A2*, and p53 signaling. Along the same lines, Sanchez and co-workers [35] identified potential molecular substrates that could underlie both the direct epidemiological association between Alzheimer's disease and glioblastoma and the inverse associations observed between Alzheimer's disease and lung cancer. Results pointed towards immune system-related processes and alterations in the mitochondrial metabolism as potential modulating agents.

This chapter aims to explore the transcriptomic associations between the set of seven CNS disorders included in the previous chapter (AD, ASD, BD, PD, MD, HD, and SCZ) and a large collection of twenty-two cancer types (i.e., acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), bladder cancer (BLCA), breast cancer (BRCA), brain cancers (BRNCA), cervical cancer (CERV), cholangiocarcinoma (CHLCA), chronic lymphocytic leukemia (CLL), chronic myeloid leukemia (CML), colorectal cancer (CRCA), diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FLYMPH), head and neck carcinomas (HANC), kidney cancer (KDNCA), lung cancer (LGCA), liver cancer (LIVCA), ovarian cancer (OVCA), pancreatic cancer (PACA), prostate cancer (PRCA), melanoma (SKCM), gastric cancer (STCA), and thyroid cancer (THCA)) through the joint analysis of their differential gene expression profiles. For each disease, differential expression meta-analyses were carried using studies available in public repositories. First, we tested if CNS disorders were significantly associated with cancers by analyzing the intersections formed by the upregulated and downregulated genes in each given pair of disorders. Second, we characterized the functions of potential molecular substrates (genes and pathways) that could be involved in the disease-disease association modulation using gene set enrichment analysis methods. Finally, we introduced a systems biology approach by generating gene co-expression networks for each individual disease. The resulting co-

expression modules were tested for their association with disease status. Then, co-expression modules significantly associated with each disorder were compared between all possible disease pairs by means of Fisher's exact tests.

Additional analyses were carried out to test the potential impact of unknown confounding variables, and the observed associations were further validated using an alternative cancer cohort derived from The Cancer Genome Atlas project (TCGA), which was profiled using RNA sequencing.

3.2 Material and methods

3.2.1 Chapter workflow

The workflow of this chapter is summarized in **Figure 6**. For a given disorder, we first searched for available transcriptomic data in public omics repositories. Each study was independently preprocessed and subjected to outlier detection analysis and removal. A study-level quality control step was carried out, and studies surpassing a defined quality threshold were excluded from downstream analyses. Differentially expressed genes (DEGs) were identified for each disorder by means of random effects model differential gene expression meta-analysis. The outputs of the analysis were used to carry out gene-set enrichment analysis. For a given disease pair, the four possible intersections formed by the jointly upregulated and downregulated genes were tested for significance using Fisher's exact tests and, the joint deregulation of biological processes and pathways was explored by comparing gene-set enrichment analyses results. Finally, gene co-expression networks were constructed for each disease, and the resulting modules were tested for association with disease status. Finally, modules significantly associated with each disorder were tested for overlap in gene content using hypergeometric tests.

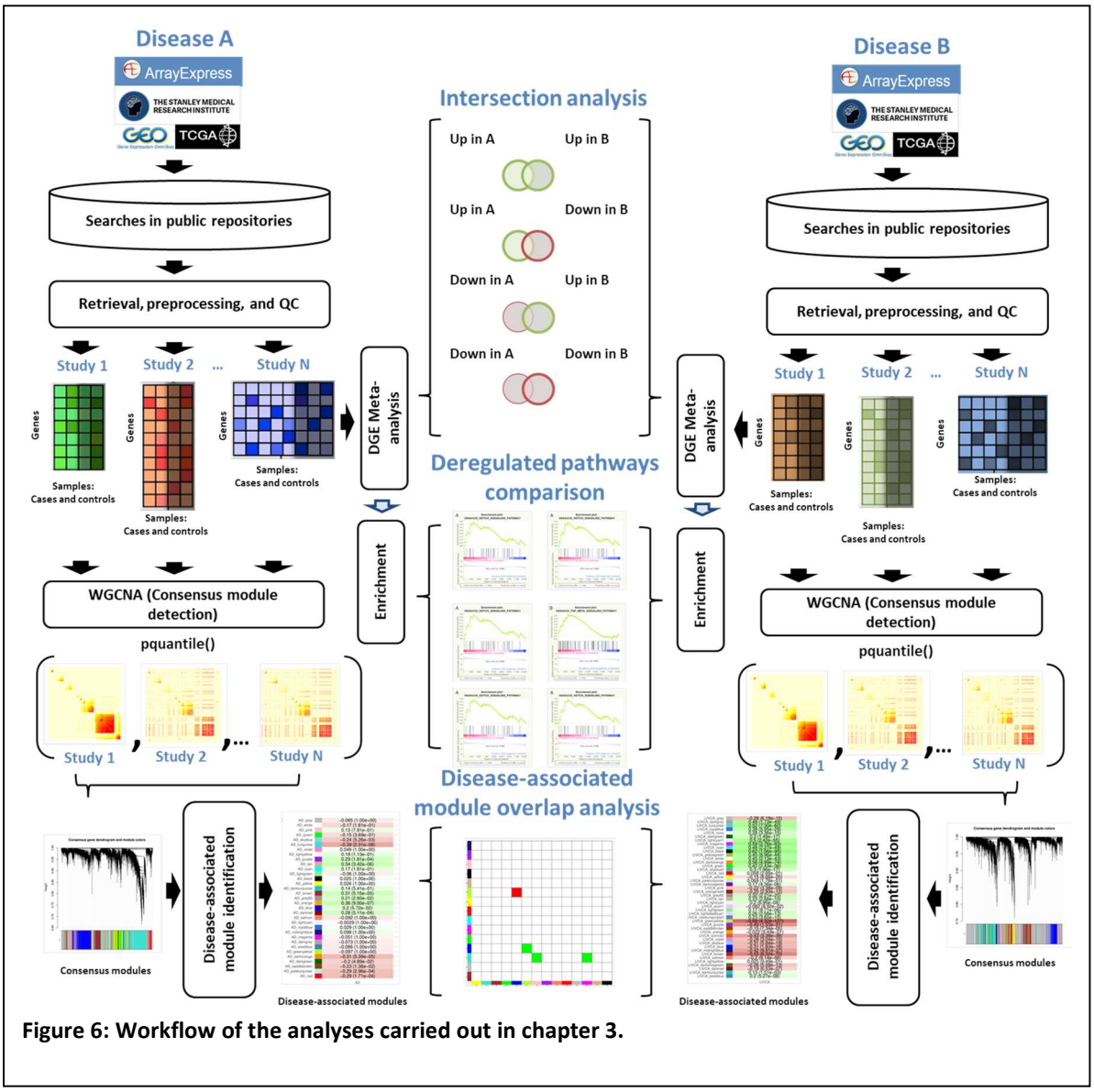

Figure 6: Workflow of the analyses carried out in chapter 3.

3.2.2 Search strategy

Searches for available datasets were performed in the following genomic data repositories: Gene Expression Omnibus (GEO), Array Express (AE), the Stanley medical research institute. (SMRI), and the cancer genome atlas (TCGA). In the case of CNS disorders, given the scarcity of studies including brain samples, additional non-systematic searches for previously published gene expression datasets were undertaken. Raw array data for such studies was directly requested to the authors when possible.

3.2.3 Inclusion and exclusion criteria

Transcriptomic datasets should include at least 3 cases and three control samples derived from matching tissues. To ensure that compatible preprocessing strategies could be applied to each study, we only selected one channel microarray data derived from the most popular Affymetrix,

Illumina, and Agilent platforms. See **Appendix 5** for a complete list of the included studies and their platforms.

In the case of CNS disorders, we focused on studies carried out using samples obtained from post-mortem brains and excluded datasets derived from blood samples. To reduce the heterogeneity that would be introduced by the inclusion of datasets from different brain regions, we selected one specific brain region for each disorder. The choice was based on two criteria, data availability and strong evidence linking the specific brain region with the physiopathology of the disease under consideration. Information regarding the brain regions selected for each CNS disorder can be found in **Table 14**.

Cancer studies derived from primary tumor samples and matched tissue control samples were selected for inclusion, whereas datasets based on blood, metastatic, or cell line samples, were excluded. Since the presence of studies derived from the same cohort could artificially inflate the differential gene expression meta-analysis results, efforts were made to exclude datasets derived from potentially redundant cohorts of individuals.

Appendix 5 shows all the identified studies that met inclusion criteria, as well as information regarding the micro-array platform and the number of samples (cases and controls) included before and after outlier sample filtering. The summary values of the study-level quality control and each study's inclusion status based on these values are also provided in the same appendix. **Appendix 6** shows the number of samples included in the TCGA cancer validation datasets.

3.2.4 Preprocessing steps for array and RNA-seq data

Datasets generated using Affymetrix platforms were preprocessed as follows: CEL files were retrieved from GEO, AE or directly from the study authors, and the R packages oligo [225] and affy [226] were used to read them and to perform RMA normalization and summarization, which was followed by quantile between-sample normalization and log2 transformation.

For Illumina platforms, non-normalized data was loaded to the R's environment using the limma package [227] and a set of custom functions. The Lumi package [228] was used to perform background correction using a normal exponential model fitting with the normexp RMA option selected, followed by quantile normalization and log2 transformation. Agilent data was preprocessed using the limma package [227] following the same preprocessing steps. In the case of the RNA-Seq dataset used in the ASD meta-analysis, raw counts obtained from GEO were loaded in the R's environment. The Rlog function from the DESeq2 package [229] was then used to transform the RNA-Seq count distribution to a continuous distribution suitable for integration with the array data. In short, the Rlog function transforms count data into a continuous log2 scale distribution,

minimizing the differences between samples for rows with small counts and normalizing the data with respect to the library size. Further details about the ASD analysis, which include the addition of control samples from other studies and the elimination of samples derived from the same patients, can be consulted in reference [230].

We transformed dataset-specific IDs into ENTREZ IDs using annotation packages to harmonize probe annotations between different dataset platforms. Probes targeting the same gene were collapsed using the collapseRows function from the WGCNA [231, 232] package using the MaxMean method.

RNAseq data derived from TCGA was preprocessed as follows: The TCGAbiolinks package [233] was used to download RNAseq counts derived from the HTSeq workflow. The TCGAanalyze_Preprocessing function was then applied with a correlation threshold of 0.6. Next, the TCGAAanalyze_Normalization function was employed to carry out normalization using the gcContent method. TCGAanalyze_Filtering was finally employed to remove mRNA transcripts with low expression values through the quantile method with a cutoff value of 0.30.

3.2.5 Quality control and outlier samples removal

We performed quality control and outlier sample detection using two different approaches for CNS and cancer-related datasets, respectively. Given the low amount of datasets and samples available for the analysis of CNS disorders, we selected a conservative and more computationally intensive approach to detect and remove outlier samples. We used three array quality measures described in detail below. Each one is aimed to evaluate a specific trait related to array quality in each sample. Measure 1 computes distances between arrays. The distance between two arrays a and b (d_{ab}), is the mean absolute difference between the expression values of all probes. $d_{ab} = mean\,|M_{ai} - M_{bi}|$, where M_{ai} is the value of the i-th probe of array a and M_{bi} is the is the value of the i-th probe of array b. Measure 1 tags a particular sample as an outlier when the sum of the distances to all other arrays $S_a = \sum_b d_{ab}$ is large. Measure 2 examines the array intensity distributions of the arrays. It is expected that the intensity distribution of the arrays has similar positions and widths. Intensity distributions of specific arrays that are very different from the rest of arrays distributions may indicate experimental problems. The Kolmogorov-Smirnov statistic K_a is used to measure the level of agreement between each array's distribution and the distribution of the pooled data. Arrays presenting a large deviation from the pooled intensities distribution are tagged as potential outliers by this measure. Finally, Measure 3 appraises the individual array quality by examining MA plots. where M is defined as $M = log_2(I_1) - log_2(I_2)$ and A is defined as $A = \frac{1}{2}\,(log_2(I_1) + log_2(I_2))$ where I_1 is the intensity of the studied array and $I_2$is the intensity of a "pseudo"-array that consists of the median across arrays. It is expected that the mass of the

distribution concentrates around the M = 0 axis. Outlier detection was performed by computing Hoedffding's statistic D_a on the joint distribution of A and M for each array. Further information about the sample-level quality assessment can be found in the documentation of the ArrayQualityMetrics package [234]. Samples were removed from datasets only when the three measures tagged them as a potential outlier. To prevent that outlier samples were removed preferentially from cases or controls due to unbalanced study designs, the quality control and outlier detection procedures were carried out independently in each subgroup of samples.

For cancer data, given the high amount of available studies and samples, we applied a less conservative outlier detection method, which had the advantage of being less computationally intensive. For each dataset, we computed array-array correlations. Then, the mean inter-array correlation was calculated. If the mean inter-array correlation was higher than 0.9, samples were not removed, and the complete dataset was included for downstream analyses. On the contrary, when the mean inter-array correlation was lower than the selected threshold, samples presenting more than two standard deviations in mean correlation measures with all other arrays were tagged as outliers and removed from the dataset. This procedure was repeated iteratively until a global mean inter-array correlation higher than 0.9 was reached. As in the instance of CNS-derived diseases, to prevent preferential removal of the case or controls samples due to unbalanced study designs, the method was applied independently to cases and controls.

3.2.6 Study-level quality assessment

Study-level quality control was carried out using the MetaQC package [235]. MetaQC computes six different quality control measures. IQC evaluates the homogeneity of the co-expression structure across studies, and it is based on the comparison of the co-expression structure of study k to the co-expression structure of all other studies. EQC appraises the consistency of the co-expression information with a pathway database. AQCg assesses the accuracy of biomarker detection by comparing the list of differentially expressed genes derived from study k to the list of differentially expressed genes obtained by performing meta-analysis using all studies except study k. AQCp represents an extension of AQCq where enriched pathways substitute genes**.** CQCg evaluates the consistency of the gene differential expression ranking from single study to the rank of differentially expressed genes obtained by performing a meta-analysis with all studies except study k, and CQCp assesses the consistency of the enriched pathway ranking. In general low values of each quality measure suggest poor agreement of study k to the rest of the studies indicating that it is a potential outlier, whereas high values indicate good agreement of a specific study with all other studies. Finally, MetaQC employs principal component analysis (PCA) biplots and a standardized mean rank (SMR) summary score to assist in identifying problematic

studies. High SMR values indicate potential outlier studies. In the present study, we choose an SMR threshold of 7 as an exclusion criterion for potentially problematic studies.

3.2.7 Differential gene expression meta-analyses

Differential gene expression meta-analyses are known to increase the statistical power and reduce the noise of gene expression measurements [223]. For each disease, meta-analyses were carried out using Choi's et al. method [36] implemented in the MetaDE package [235]. All meta-analyses were performed using random effect models since high heterogeneity was expected, given our data's biological and technical variability. Meta-analysis methods based on effect size combinations have been described in detail in this manuscript's introductory section. Genes showing a false discovery rate (FDR) adjusted p-value lower than 0.05 were considered to be differentially expressed.

The potential impact of confounding factors was assessed as follows. For each study, surrogate variables (SV) were identified and computed using the SVA package's sva algorithm [236]. The algorithm is divided into two parts. The first part is oriented to the detection of the number of surrogate variables. We restricted the number of surrogate variables to 3 since the inclusion of a large number could remove true biological signal. The second part of the algorithm is dedicated to the computation of the values of the SVs. For each study, differential gene expression analysis was carried out using limma [227] in two different settings. In setting one, only the predictive variable of disease status was introduced in the model. In setting two, the model was constructed using the disease status, and the detected SVs were introduced as adjustment factors. Then for each disorder, differential gene expression meta-analyses were carried out under the two different settings (i.e., including the individual study differential gene expression result with and without adjustment for surrogate variables). Since MetaDE does not allow covariate adjustment, an alternative differential gene expression meta-analysis method based on the combination of the log-fold changes of individual differential gene expression analyses was used. This method is implemented in the MetaVolcanoR package [237], which internally makes use of the metafor package [122] to computed pooled estimates of the log-fold changes of each gene under a random effects model using the inverse of the variance method. Then the results (i.e., the differentially expressed genes) obtained under both settings were compared using Jaccard indexes.

3.2.8 Leave-one-out and cumulative analyses

We applied a leave-one-out strategy to detect individual studies' impact in our differential gene expression meta-analysis results. Therefore, we repeated each differential gene expression meta-analysis using Choi's method for each disease by removing one of the studies at a time. Thus, for a particular disease with n available studies, n meta-analysis using n-1 studies were carried out.

Results were inspected to detect datasets displaying an abnormally large impact in the differential gene expression meta-analysis results. The significance threshold was set to a conventional level of 0.05, and genes with a false discovery rate (FDR)-corrected p-value lower than 0.05 were considered to be differentially expressed.

In addition, we analyzed the impact of the stepwise inclusion of studies in the meta-analysis. In order to do it, we sorted the studies based on their SMR quality score (Computed by MetaQC) from lowest to the highest quality. We first performed the meta-analysis using the two studies showing the lowest quality scores, and then we repeated it iteratively by adding one study at a time based on the cited order.

3.2.9 Measures of disease-disease transcriptomic associations

The expression profiles of each CNS disorder and all the studied cancer types were compared to evaluate the significance of the overlaps between the differentially expressed genes, as previously described in [34] and [35]. For each CNS disorder and-cancer pair, the significance of the four possible intersections formed by the upregulated and downregulated genes was evaluated by means of one-tailed Fisher's exact tests. The intersections were:

- 1) Genes upregulated in both a specific CNS and the selected cancer type (Intersection A),
- 2) Genes downregulated in both a specific CNS and the selected cancer type (Intersection B),
- 3) Genes upregulated in a specific CNS and downregulated in the selected cancer type (Intersection C), and
- 4) Genes downregulated in a specific CNS and upregulated in the selected cancer type (Intersection D).

Fisher's test p-values were corrected by multiple testing using the false discovery rate method (FDR). Overlaps showing adjusted p-values lower than 0.05 were considered significant. The background number of genes was set as the number of genes jointly included in the two meta-analyses under consideration, which depended on the platforms included in each meta-analysis. A cancer type was considered to be deregulated in the same direction as a CNS disorder when Intersections A and B were significant, and Intersections C and D were not. These cancer types were referred to as same direction deregulated cancers (SDDCs) and could be candidates for direct comorbidity with the specific CNS disorder. Conversely, a cancer type was considered to be deregulated in the opposite direction from a particular CNS disorder when intersections C and D were significant, but intersections A and B were not. These cancer types were referred to as

opposite direction deregulated cancers (ODDCs) and could be candidates for inverse comorbidity with a specific CNS disorder.

Additionally, to determine the strength of the overall associations between differential expression profiles, Pearson's correlations were computed using the Z-values obtained from each differential gene expression meta-analysis. Positive correlations suggest similar patterns of differential expression, while negative correlations would indicate opposite patterns.

3.2.10 Validation of the CNS and cancer associations using an independent cohort of cancers.

The cancer genome atlas was queried for RNAseq-based experiments interrogating tumor samples for the tumor types included in our array-based analysis. Data was downloaded and preprocessed using the R package TCGAbiolinks [233]. DEseq2 [229] was used in order to compute differential gene expression between cases and controls. Results were used to perform intersection analysis following the same methodology explained for arrays in the previous section.

3.2.11 Weighted co-expression network analyses

We carried out consensus module detection for each disorder using all the datasets employed in the differential gene expression meta-analysis step and the WGCNA package [231]. Consensus modules are clusters of densely interconnected genes present across datasets. First, Bi-correlation networks were constructed for each study resulting in a set of gene-gene correlation matrices. The parameter choices for consensus module detection analyses were based on the parameters selected by the authors of the package in their analysis of a set of eight lung cancer datasets reported in [238]. Bi-correlation matrices were transformed into signed-hybrid adjacency matrices as described in the following equations.

$$a_{ij} = \left[cor(x_i, x_j)\right]^{\beta} \; for \; cor(x_i, x_j) > 0$$

$$a_{ij} = 0 \; for \; cor(x_i, x_j) \leq 0$$

Adjacency matrices are called "hybrid-signed" because they use a combination of hard and soft thresholding. A hard threshold is used for correlation values ≤ 0 and a soft threshold for values above 0. The parameter β is selected such that the resulting adjacency matrix fits a power-law distribution by using the pickSoftThreshold function implemented in the package. Adjacency matrices. $A = \{a_{ij}\}$ have the following properties:

$$a_{ij} = a_{ji}$$

$$0 \leq a_{ij} \leq 1$$

$$a_{ii} = 1$$

The next step involves the transformation of adjacency matrices into topological overlap matrices using the following equation:

$$TOM_{ij}(A) = \frac{\sum_{k \neq i,j} a_{ik} a_{kj} + a_{ij}}{min\left(\sum_{k \neq i} a_{ik}, \sum_{k \neq j} a_{jk}\right) + 1 - a_{ij}}$$

The $TOM_{ij}(A)$ matrix is also an adjacency matrix. The topological overlap of two genes reflects their similarity in terms of the commonality of the genes to which they are connected. The TOM leads to a more robust network and larger modules which satisfies the same properties than A. Using the resulting topological overlap matrices derived from each individual dataset, a consensus TOM matrix was generated following the quantile method. For a set of K matrices $TOM^{(1)}, TOM^{(2)}, \ldots TOM^{(k)}$. A consensus adjacency matrix is constructed using the quantile method.

$$Quantile_{q,ij}\left(TOM^{(1)}, TOM^{(2)}, \ldots TOM^{(k)}\right) = Quantile_q\left(tom_{ij}^{(1)}, tom_{ij}^{(2)} \ldots tom_{ij}^{(k)}\right)$$

The consensus TOM was defined as the consensus of the individual TOM matrices with percentile q = 0.25. In the consensus TOM, two variables are connected with the strength that is common to all input networks. Finally, a dissimilarity matrix based on the consensus TOM matrix is fed as an input for the dynamic tree-cutting algorithm, which identifies co-expressed gene clusters (consensus modules). The steps described above were carried out using the blockwiseConsensusModules with the following parameters set to maxBlockSize = 30000, corType = "bicor", networkType = "signed hybrid", deepSplit = 3, mergeCutHeight = 0.25, and consensusQuantile = 0.25).

Once the consensus modules were identified, each module's eigengenes were computed, and correlations between them and disease status were calculated. A module eigengene is defined as the first principal component of the expression matrix of the genes included in a specific module. It typically explains more than 50% of the module expressions' variance [239]. Significant positive correlations between disease status and a particular module eigengene suggest that genes placed at that particular module tend to be upregulated in cases compared to controls. In contrast, significant negative correlations have the opposite implication (the genes places in the module tend to be downregulated in disease samples compared to controls).

3.2.12 Module-module overlap analyses

Genes contained in all modules significantly correlated to disease status were identified for all the included disorders. Then, associations between modules of co-expressed genes significantly correlated with disease status were computed for all possible pairwise comparisons through Fisher's exact tests. The p-values of all pairwise module-gene overlap analyses were adjusted for multiple comparisons by FDR.

3.2.13 Functional and cell type-specific enrichment analyses

Diverse functional analysis methods and sources for sets of genes were used in the different sections. Classic overrepresentation analysis was carried out to compute enrichment in gene ontology (GO) terms for the genes placed at the intersections observed in the intersection analyses and for genes placed in the detected co-expression modules using the clusterProfiler [240] and anRichment [241] R packages. Gene Set Enrichment Analysis (GSEA) was used to identified upregulated and downregulated pathways in each disorder using as an input the list of genes ordered by its Z-score obtained by Choi's differential gene expression meta-analysis. GSEA analyses were carried out using the fGSEA package implementation [242]. The sets of genes were retrieved from the molecular signatures database (MSigDB) [243]and included the hallmarks (H) set of genes, the canonical pathways subset (C2:CP) of the C2 curated set of genes, and the Gene Ontology (C5) set of genes.

Finally, enrichment in cell type-specific genetic markers was carried out for each detected consensus module using a collection of cell-specific gene markers derived from PanglaoDB database and hypergeometric tests. This database contains a list of gene expression markers used to define 154 cell types. The obtained p-values were corrected for multiple comparisons using the false discovery rate (FRD) method.

3.3 Results

3.3.1 Tissues selected for each CNS disorder

Tissues included to carry out the differential gene expression meta-analyses of CNS disorders were selected based on data availability and strong support linking the selected tissue with the disorder's physiopathology. In the case of AD, we only selected studies and samples derived from hippocampal tissues. For HD, studies based on caudate nucleus samples were selected, whereas, in the case of PD, we only picked datasets and samples derived from the substantia nigra. Reduced hippocampal volume, neural loss, and the presence of neurofibrillary tangles and beta-amyloid deposition are a core feature of AD [244]. Substantia nigra, placed in the ventral midbrain, contains groups of nerve cells that play a pivotal role in movement control. Those neurons communicate with the neurons in the basal ganglia by liberating the neurotransmitter dopamine. PD is characterized by the progressive degeneration and loss of neurons of the substantia nigra [245]. The striatum, which comprises the caudate nucleus and the putamen, is the HD's primarily affected region, where up to 90% of neurons have been found to be lost [246]. In the cases of ASD, BD, MD, and SCZ, we included datasets derived from the prefrontal cortex of disease patients and controls. One of SCZ patients' characteristic features is their difficulty in prioritizing, processing, and responding to information. Some studies have reported modified levels of activity

and structural and metabolic abnormalities in frontal regions in SCZ patients compared to controls. Alterations in the prefrontal cortex have also been observed in MD, BD, and ASD [247, 248]. Neuron overabundance has been found in the dorsolateral prefrontal cortex of ASD patients [249]. **Table 14** shows the selected tissues included in each differential gene expression meta-analysis for each CNS disorder.

CNS disorder	Studied tissue
Alzheimer's disease	Hippocampus
Autism spectrum disorders	Frontal cortex
Bipolar disorder	Frontal cortex
Huntington's disease	Caudate Nucleus
Major depression	Frontal cortex
Parkinson's disease	Substantia nigra
Schizophrenia	Frontal cortex

Table 14: Brain tissues studied in each CNS disorder.

3.3.2 Number of included studies and samples

In total, 192 unique datasets were identified and selected for potential inclusion in our analyses. After study-level quality control, 160 individual studies gathering 16132 samples remained. Note that several datasets included information for more than one disorder. **Appendix 5** shows information regarding the identified studies, including the initial number of samples in each dataset, the final number of samples after removal due to exclusion criteria application and outlier removal, and the study level quality control summary value (SMR), as well as the status of final inclusion in our analysis based on the reported SMR.

Five studies from CNS disorders, all belonging to Parkinson's disease, were excluded from subsequent analysis because they presented SMR values >= 7. A total of 29 CNS-related datasets gathering 1371 samples, divided into 618 cases and 753 controls, were selected after study-level quality control. The number of samples available for each CNS disorder ranged from 89 (47 cases and 42 controls) in the case of HD to 347 (166 cases and 181 controls) in the case of SCZ. Twenty-seven cancer datasets were excluded after study-level quality control (AML = 1, BRNCA = 3, CRCA = 8, KDNCA = 2, LGCA = 9, LIVCA = 4) because they presented SMR values higher than the selected threshold. One hundred and thirty-one datasets gathering 14761 samples (11152 cases and 3609 controls) were included in cancer analyses. The total number of samples available for each site-specific cancer ranged from 139 (82 cases and 57 controls) in the case of cervical cancer to 1500 (1406 cases and 94 controls) in the case of ALL.

The cancer validation dataset consisted of 7361 samples, including 6717 cases and 644 controls from 17 tumor types corresponding to 15 of the cancer types included in our array-based analyses were obtained from TCGA. Not all cancer types included in the array-based analyses were

available. Data for leukemias, lymphomas, melanoma, and ovarian cancer did not include matched normal tissue. Therefore, differential gene expression analysis could not be carried out. In addition, some tumor types were represented by more than one dataset. For instance, lung cancer was divided into two datasets corresponding to lung adenocarcinomas and lung squamous cell carcinomas. **Appendix 6** shows the TCGA datasets used in the validation step, along with the correspondence between the array-based analysis and the number of samples and differentially expressed genes for each cancer type.

3.3.3 AD differential gene expression meta-analyses results

In this section, we present the complete differential gene expression analysis results for AD. A summarized version of the results shown in this section for all the studied disorders will be presented in **Section 3.3.4**.

Seven studies including samples derived from the hippocampus of AD patients and controls, were identified in public repositories. Individual datasets were preprocessed as described in methods and subjected to outlier sample detection and removal. Study-level quality control was then undertaken. All seven datasets presented SMR values under the selected threshold of 7 and were therefore selected for downstream analysis (**See Table 15**)

Study	IQC	EQC	CQCg	CQCp	AQCg	AQCp	SMR
GSE48350	8.56	4.3	302.79	6.51	186.95	2.9	1.5
E_MEXP_2280	3.48	4.3	103.81	1.35*	44.47	2*	3
GSE36980	2.01*	4.3	127.17	0.63*	74.98	0.46*	3.88
GSE5281	3.02	4.3	66.65	0.24*	42.57	1.3*	4.25
GSE84422	3.85	2.34	0.76*	3.02	0.03*	0.52*	4.75
GSE1297	5.59	1.95*	74.19	0.46*	36.33	0.39*	5.12
GSE29378	0.44*	4.3	4.47	1.13*	3.3	0.01*	5.5

Table 15: Study-level quality control of the datasets included hippocampal samples for AD and control subjects.

In total, 226 samples derived in 124 cases and 102 controls distributed among the seven studies were available for downstream analysis.

Differential gene expression analysis of individual studies yielded heterogeneous results with three studies (GSE84422, GSE1297, E_MEXP_2280) presenting no significantly deregulated genes, one (GSE29378) showing only 2 DEGs, and three studies (GSE48350, GSE5281, GSE36980) showing 3275, 3667, and 1115 DEGs respectively (**Figure 7 A**).

Random effects models differential gene expression yielded 3341 DEGs, of which 1364 were found to be significantly upregulated, whereas 1977 were found to be significantly downregulated after correction for multiple comparisons (FDR < 0.05) (**Figure 7 B**). The top 10 upregulated genes in AD hippocampal samples compared to controls were *PLEC*, *ITPKB*, *MTSS2*,

TRIP10, *PPP1R13L*, *KCNN3*, *PSD4*, *MXRA8*, *LPAR4*, and *PSTPIP1*, whereas the top 10 downregulated genes were *GAD1*, *MAK16*, *POLB*, *RPP40*, *NREP*, *TXNDC9*, *TNFRSF21*, *LRRC8B*, *MRPS22*, *PEX11B*. **Supplementary Appendix 1 Figure 77** shows the expression values of the top upregulated gene (*PLEC*) across studies, whereas **Supplementary Appendix 1 Figure 78** does it for the top downregulated gene (*GAD1*) to exemplify how differential gene expression meta-analyses are able to capture DEGs that are consistently deregulated between studies even in the cases in which they are not found to be significantly deregulated in a subset of individual studies.

The mean Q value of the genes included in the analysis was Q = 10.59 and the mean τ^2 was 033. Cumulative meta-analysis showed how the number of DEGs increases as higher-quality studies are introduced in the meta-analysis sequentially (**Figure 7 C**).

The leave-one-out analysis showed that the study with the highest impact on the final outcomes was GSE4835 since its exclusion from the meta-analysis produces the output with the lowest number of DEGs (**Figure 7 D**).

Surrogate variable detection identified three significant surrogate variables for GSE84422 and E_MEXP_2280, one for datasets GSE5281 and GSE29378, and none for the rest of the included studies. The alternative meta-analysis method using the raw data yielded 1856 up- and 2232 down-regulated genes. In contrast, when the meta-analysis was carried out after adjustment for surrogate variables, only 963 up- and 1256 down-regulated genes were found. This indicates the presence in the data of unknown sources of variation, which could be linked to covariates not included in the analysis (i.e., age, sex, post mortem interval) (**Figure 7 G**). Most of the genes found to be differentially expressed in the surrogate variable adjusted analysis were also found to be differentially expressed in the raw dataset. Besides, the correlation of the log-fold changes of both analyses is was found to be large (r = 0.93).

GSEA analysis showed that upregulated pathways in AD samples compared to controls included instances of immune-related pathways such as cytokine-cytokine receptor interaction (p-adj: 4.85e-09), interleukin 4 and interleukin 13 signaling (p-adj: 1.77e-08), and complement cascade (p-adj: 3.54e-06), among others. Pathways downregulated in AD patients compared to controls were found to be linked to mitochondrial energy production and included instances, such as oxidative phosphorylation (p-adj: 4.85e-09) and respiratory electron transport and ATP synthesis by chemiosmotic coupling and heat production by uncoupling proteins (p-adj: 4.85e-09) alongside with other processes such as degradation of AXIN (p-adj: 7.34e-09), and regulation of PTEN stability and activity (p-adj: 4.85e-09) as well as processes linked to the neuronal system (p-adj: 6.03e-09), among others.

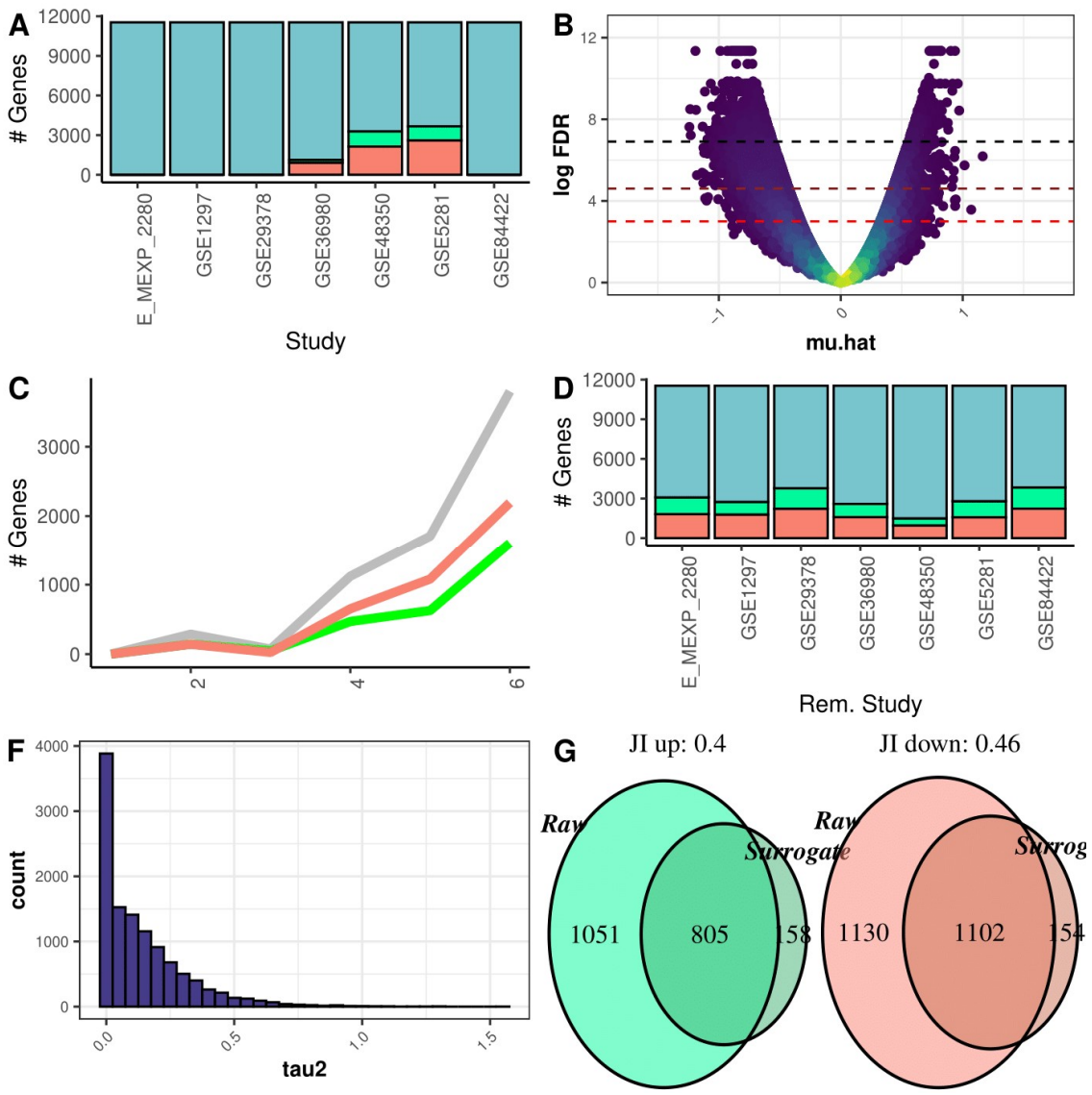

Figure 7: AD differential gene expression meta-analysis results. A) Differential expression results of the individual studies included in the meta-analysis. For each study, the bar's blue portion shows the number of genes for which no gene expression change was identified. Green and red segments depict the number of unregulated and downregulated genes, respectively. B) Volcano plot depicting the $\hat{\mu}$ values obtained in MetaDE meta-analysis (x-axis) against the negative logarithm of the false discovery rate adjusted p-values (y-axis). Horizontal dashed lines represent FDR thresholds of 0.05, 0.01, and 0.001 from bottom to top. C) Cumulative meta-analysis results. The plot shows the number of total deregulated (blue), upregulated (green), and downregulated (red) genes obtained by performing the meta-analysis with an increasing number of studies. The addition of studies proceeded from those showing higher SMR values (lowest quality) to those presenting the lowest SMR values (highest quality). D) Leave-one-out meta-analysis. Each bar represents the meta-analysis results after removing only one study (the removed study is indicated in the x-axis labels). F) Histogram showing the distributions of heterogeneity τ^2 values for all genes included in the meta-analysis. G) Results for the alternative meta-analysis method. The green Venn diagram represents the number of upregulated genes found in the meta-analysis performed with the raw expression values (lighter), whereas the right one (darker) shows the number of upregulated genes after performing the meta-analysis using adjustment for surrogate variables. The red Venn diagram follows the same scheme for downregulated genes. On top of each Venn diagram, the Jaccard indexes (JI) for the overlaps are shown.

3.3.4 Summary of the differential gene expression meta-analysis results for all the included disorders

Differential gene expression meta-analyses were carried out for the seven CNS disorders and the 22 tumor types under consideration. The number of DEGs identified through meta-analysis varied greatly among the different disorders ranging from the three DEGs found in schizophrenia (representing 0.02% of the tested genes) to the 9575 DEGs (69.24% of the tested genes) observed in kidney cancer samples. BD, MD, and SCZ patients' brain transcriptomes did not differ substantially from their tissue-matched controls, with only 5, 15, and 3 DEGs identified, respectively. This fact prevented us from including those disorders in subsequent intersection analysis. Neurodegenerative disorders, including AD, PD, and HD, presented a large number of differentially expressed genes with 3341, 3473, and 4504 DEGs detected, respectively, under an FDR threshold of 0.05. One-thousand and three genes were found to be deregulated in ASD (FDR < 0.05). Except for AML, cancers tended to present a large number of DEGs, which were found to be exceptionally high in the cases of BRNCA, CRCA, KDNCA, LGCA, and LIVCA (**See table 16**). In general, higher mean Q values and mean tau values were observed in cancer meta-analyses compared to CNS, suggesting the presence of higher heterogeneity among studies. The alternative meta-analysis method using both surrogate variable adjusted and un-adjusted datasets showed a lower level of agreement in CNS disorders than in cancer, indicating the unknown co-variates could have a higher impact in the differential gene expression meta-analyses results of CNS disorders than cancers (**See Table 17**). Note that nominal p-values instead of FDR adjusted p-values are used to call a gene differentially expressed when using the alternative meta-analysis method, partially explaining the disagreement between these results and those derived from Choi's method. **Supplementary Appendix 2** contains the differential gene expression meta-analysis results for all the included disorders.

Disease	Included studies	Nº samples (cases/control)	DEGs 0.05 / Tested genes / %	Up 0.05 / Down 0.05	DEGs 0.01 / Tested genes / %	Up 0.01 / Down 0.05	Mean Q	Mean tau
AD	7	226 (124/102)	3341/11536 (28.96%)	1364/1977	1641/11536 (14.23%)	563/1078	10.59	0.33
ASD	3	164 (34/ 130)	1003/13669 (7.34%)	465/538	291/13669 (2.13%)	170/121	2.5	0.2
BD	4	173 (66/107)	5/12436 (0.04%)	3/2	0/12436 (0%)	0/0	9.34	0.4
MD	6	223 (98/125)	15/12436 (0.12%)	5/10	2/12436 (0.02%)	0/2	10.57	0.28
HD	2	89 (47/42)	4504/11714 (38.45%)	2496/2008	3447/11714 (29.43%)	1842/1605	1.69	0.33
PD	6	149 (83/66)	3473/11714 (29.65%)	1626/1847	2267/11714 (19.35%)	949/1318	7.24	0.31
SCZ	8	347 (166/181)	3/12436 (0.02%)	1/2	1/12436 (0.01%)	0/1	8.83	0.16
ALL	5	1500 (1406/94)	2315/12436 (18.62%)	1373/942	1387/12436 (11.15%)	920/467	16.67	0.63
AML	8	1162 (949/213)	581/16271 (3.57%)	415/166	217/16271 (1.33%)	163/54	39.52	0.47
BCLA	5	228 (199/29)	5277/16054 (32.87%)	3209/2068	3069/16054 (19.12%)	1983/1086	10.41	0.85
BRCA	8	1396 (1164/232)	8362/20536 (40.72%)	4700/3662	6662/20536 (32.44%)	3781/2881	33.97	0.46
BRNCA	7	1218 (1124/94)	9757/15401 (63.35%)	5161/4596	8228/15401 (53.43%)	4437/3791	19.89	0.35
CERV	4	139 (82/57)	3612/15683 (23.03%)	2215/1397	2344/15683 (14.95%)	1517/827	11.73	0.64
CHLCA	4	152 (130/22)	1234/16111 (7.66%)	766/468	535/16111 (3.32%)	346/189	15.6	1.55
CLL	6	1021 (884/137)	2547/16634 (15.31%)	1670/877	1488/16634 (8.95%)	996/492	62.94	1.24
CML	3	226 (128/98)	1593/15125 (10.53%)	1115/478	740/15125 (4.89%)	461/279	6.88	0.33
CRCA	7	1261 (968/293)	9041/13595 (66.5%)	4793/4248	7872/13595 (57.9%)	4235/3637	27.98	0.3
DLBCL	3	246 (173/73)	4673/20503 (22.79%)	3284/1389	3184/20503 (15.53%)	2339/845	32.7	1.67
FLYMPH	6	214 (147/67)	1945/15456 (12.58%)	1195/750	854/15456 (5.53%)	542/312	26.71	1.22
HANC	9	560 (436/124)	6964/14760 (47.18%)	3617/3347	5280/14760 (35.77%)	2874/2406	24.94	0.5
KDNCA	8	650 (396/254)	9575/13828 (69.24%)	4974/4601	8193/13828 (59.25%)	4426/3767	31.65	0.38
LGCA	5	850 (522/328)	9468/14718 (64.33%)	5637/3831	7954/14718 (54.04%)	4695/3259	26.39	0.23
LIVCA	8	839 (383/456)	6624/10648 (62.21%)	3567/3057	5401/10648 (50.72%)	2996/2405	26.91	0.23
OVCA	6	222 (154/68)	5241/18544 (28.26%)	3704/1537	3415/18544 (18.42%)	2496/919	21.1	0.9
PACA	9	463 (318/145)	8306/14257 (58.26%)	3938/4368	6270/14257 (43.98%)	3061/3209	39.27	0.79
PRCA	6	1019 (648/371)	4663/15169 (30.74%)	2001/2662	3253/15169 (21.45%)	1378/1875	24.06	0.17
SKCM	5	486 (409/77)	4158/11651 (35.69%)	2271/1887	2752/11651 (23.62%)	1453/1299	13.43	0.38
STCA	9	664 (393/271)	7047/15066 (46.77%)	3766/3281	5243/15066 (34.8%)	2920/2323	44.96	0..42
THCA	5	245 (139/106)	8662/20536 (42.18%)	4271/4391	6356/20536 (30.95%)	2975/3381	10.02	0.39

Table 16: Number of differentially expressed genes found in each meta-analysis. The results for two different FDR thresholds (FDR < 0.05 and FDR < 0.01) are provided. The number of DEGs and the percentage with respect to the total number of tested genes can be found in columns three and five, whereas the number of up and downregulated genes can be found in columns four and six for FDR thresholds of 0.05 and 0.01. Columns seven and eight show the mean Q and tau values, respectively.

Disease	UP raw	Down Raw	Up Surrogate	Down Surrogate	Jaccard Up	Jaccard Down
AD	1856	2232	963	1256	0.4	0.46
ASD	1166	1468	845	1126	0.59	0.55
BD	813	313	712	236	0.62	0.54
MD	631	619	615	598	0.63	0.61
HD	2463	1923	2483	2108	0.72	0.75
PD	2016	2019	1786	1739	0.78	0.79
SCZ	603	522	519	533	0.44	0.44
ALL	1682	1392	1360	1336	0.7	0.7
AML	1400	631	1440	676	0.78	0.77
BLCA	3621	2226	3621	2226	1	1
BRCA	4951	3819	4996	3848	0.97	0.98
BRNCA	5045	4593	4897	4397	0.91	0.89
CERV	2378	1737	2322	1915	0.84	0.77
CHLCA	1572	737	1473	814	0.69	0.66
CLL	2151	1463	2060	1454	0.75	0.73
CML	1777	602	2008	535	0.73	0.71
CRCA	4662	4181	4654	4101	0.97	0.96
DLBCL	3916	1796	3940	1765	0.98	0.96
FLYMPH	1896	1298	1948	1279	0.88	0.85
HANC	3703	3296	3673	3253	0.96	0.92
KDNCA	4974	4663	4950	4598	0.98	0.97
LGCA	5574	3853	5590	3860	0.99	0.99
LIVCA	3628	3133	3496	3061	0.95	0.95
OVCA	4218	1930	4204	1963	0.92	0.89
PACA	3817	4293	3717	4172	0.95	0.94
PRCA	2204	3018	2233	2842	0.86	0.84
SKCM	2387	2025	2500	1980	0.82	0.83
STCA	3666	3204	3614	3120	0.96	0.93
THCA	4573	4683	4340	4535	0.89	0.9

Table 17: Jaccard indexes for the alternative meta-analysis method based on MetaVulcanoR. Comparison of the differentially expressed genes between the meta-analysis carried out without surrogate variables correction and with surrogate variables correction.

3.3.5 Intersection analysis results for each CNS disorder and the panel of 22 site-specific cancer

This section compares the differential gene expression meta-analysis results of each CNS disorder with those derived from the panes of 22 site-specific cancers. The lack of DEGs observed in BD, MD, and SCZ prevented us from carrying out these analyses in those disorders. Tables showing the FDR adjusted p-values of the intersections formed by genes jointly upregulated, jointly downregulated, and upregulated in one disorder and downregulated in the other are shown. Besides, to determine if specific biological processes were overrepresented in the genes placed at the intersections, classic overrepresentation analysis (ORA) was carried out using the biological process BP branch of gene ontology. Those tumors presenting direct transcriptomic associations with cancer will be referred to as Same Direction Deregulated Cancers (SDDC), whereas those presenting inverse patterns of transcriptomic associations will be called Opposite Direction Derregulated Cancers (ODDC).

AD and cancer

AD was found to be directly associated with brain cancer and thyroid cancer since the intersections formed by the genes jointly upregulated and down-regulated were found to be significant after adjustment by multiple comparisons. Several cancer types were inversely associated with AD, including bladder, breast, cervical, colorectal, head and neck, lung, liver, pancreatic, and prostate cancers. In those cases, the intersections formed by genes upregulated in one disorder and down-regulated in the other were found to be significant. Melanoma also presented a significant inverse pattern of association with AD**. Figure 8** shows the intersection analysis results. Genes jointly upregulated between AD and BRNCA were enriched in functions linked to blood vessel morphogenesis (GO:0048514, p-adj = 6.96e-10), extracellular matrix organization (GO:0030198, p-adj = 6.96e-10), and several immune-related processes such as, lymphocyte activation (GO:0046649, p-adj = 2.64e-06), T cell activation (GO:0042110, p-adj = 4.33e-05), and adaptive immune response (GO:0002250, p-adj = 1.58e-04), among others, whereas jointly downregulated genes, were found to be enriched in neural-related processes including vesicle-mediated transport in synapse (GO:0099003, p-adj = 1.83e-13), neurotransmitter transport (GO:0006836, p-adj = 2.69e-08), and processes linked to the mitochondrial energy production such as respiratory electron transport chain (GO:0022904, p-adj = 4.08e-06), and the oxidative phosphorylation (GO:0006119, p-adj = 2.19e-05). In the case of thyroid cancer, jointly upregulated genes were enriched in the extracellular matrix organization (GO:0030198, p-adj = 6.73e-10), myeloid cell activation involved in immune response (GO:0002275, p-adj = 2.12e-03), and positive regulation of I-kappaB kinase/NF-kappaB signaling (GO:0043123, p-adj = 4.54e-03). Jointly

downregulated genes were found to be also enriched in neuronal-related processes such as vesicle-mediated transport in synapse (GO: 0099003, p-adj = 1.83e-13), and bioenergetics processes like the electron transport chain (GO:0022904, p-adj = 4.08e-06), and oxidative phosphorylation (GO:0006119, p-adj = 2.19e-05)

Some biological processes were found to be enriched in genes jointly up-regulated in AD and downregulated in different negatively associated cancers. They were found to be linked to immune-related processes, i.e. , regulation of cell activation (GO:0050865; BLCA: p-adj = 1.49e-04, CRCA: p-adj = 5.64e-08, LGCA: p-adj = 4.68e-06, LIVCA: p-adj = 6.65e-08), adaptive immune response (GO:0002250; BLCA: p-adj = 1.37e-03, CRCA: p-adj = 6.22e-07, LGCA: p-adj = 5.45e-06, LIVCA: p-adj = 4.18e-07), and to other processes such as, blood vessel morphogenesis (GO:0048514; BLCA: p-adj = 1.11e-04, CERV: p-adj = 2.07e-02, LGCA: p-adj = 2.15e-08, PRCA: p-adj = 4.68e-07, SKCM: p-adj = 6.36e-03), and extracellular matrix organization (GO:0030198; BLCA: p-adj = 1.07e-03, BRCA: p-adj = 3.76e-03, LGCA: p-adj = 1.15e-05, LIVCA: p-adj = 7.92e-03, PRCA: p-adj = 5.18e-06).

Genes downregulated in AD and upregulated in inversely associated cancers tended to be enriched in biological processes linked the G2/M transition of mitotic cell cycle (GO:0000086; BLCA: p-adj = 2.10e-04, BRCA: p-adj = 2.72e-05, CERV: p-adj = 1.22e-02, CRCA: p-adj = 5.21e-06, HANC: p-adj = 2.43e-06, LGCA: p-adj = 2.38e-03, LIVCA: p-adj = 3.62e-12, PACA: p-adj = 8.01e-07), mitochondrial functions such as mitochondrial gene expression (GO:0140053; BLCA: p-adj = 8.84e-07, BRCA: p-adj = 2.38e-06, CERV: p-adj = 1.32e-04, CRCA: p-adj = 2.30e-14, HANC: p-adj = 6.67e-11, LGCA: p-adj = 2.69e-13, LIVCA: p-adj = 2.94e-09, PACA: p-adj = 2.46e-04, PRCA: p-adj = 5.95e-06) and bioenergetics processes such as oxidative phosphorylation (GO:0006119; BLCA: p-adj = 3.43e-02 , LGCA: p-adj = 2.96e-05, LIVCA: p-adj = 4.41e-07, PACA: p-adj = 4.27e-03), ATP synthesis coupled electron transport (GO:0042773; BLCA: p-adj = 1.25e-02, BRCA: p-adj = 3.35e-02, LGCA: p-adj = 6.03e-05, LIVCA: p-adj = 1.09e-06, PACA: p-adj = 1.29e-03) as well as processes linked to proteasomal function such as, SCF-dependent proteasomal ubiquitin-dependent protein catabolic process (GO:0031146; BLCA: p-adj = 9.25e-04, BRCA: p-adj = 1.10e-04, CERV: p-adj = 1.24e-02, CRCA: p-adj = 2.07e-09, HANC: p-adj = 1.16e-10, LGCA: p-adj = 4.95e-04, LIVCA: p-adj = 4.07e-12, PACA: p-adj = 1.43e-11). The differential gene expression profiles of brain cancer and thyroid cancer were positively correlated with the AD profile (BRNCA r = 0.32, THCA r = 0.11). Negatively associated cancers presented Pearson's correlaions of: -0.08, -0.09, -0.17, -0.11, -0.12, -

0.22, -0.09, and -0.1 for bladder, breast, cervical, colorectal, head and neck, lung, liver, pancreatic, and prostate cancers, respectively (**See Figure 9**).

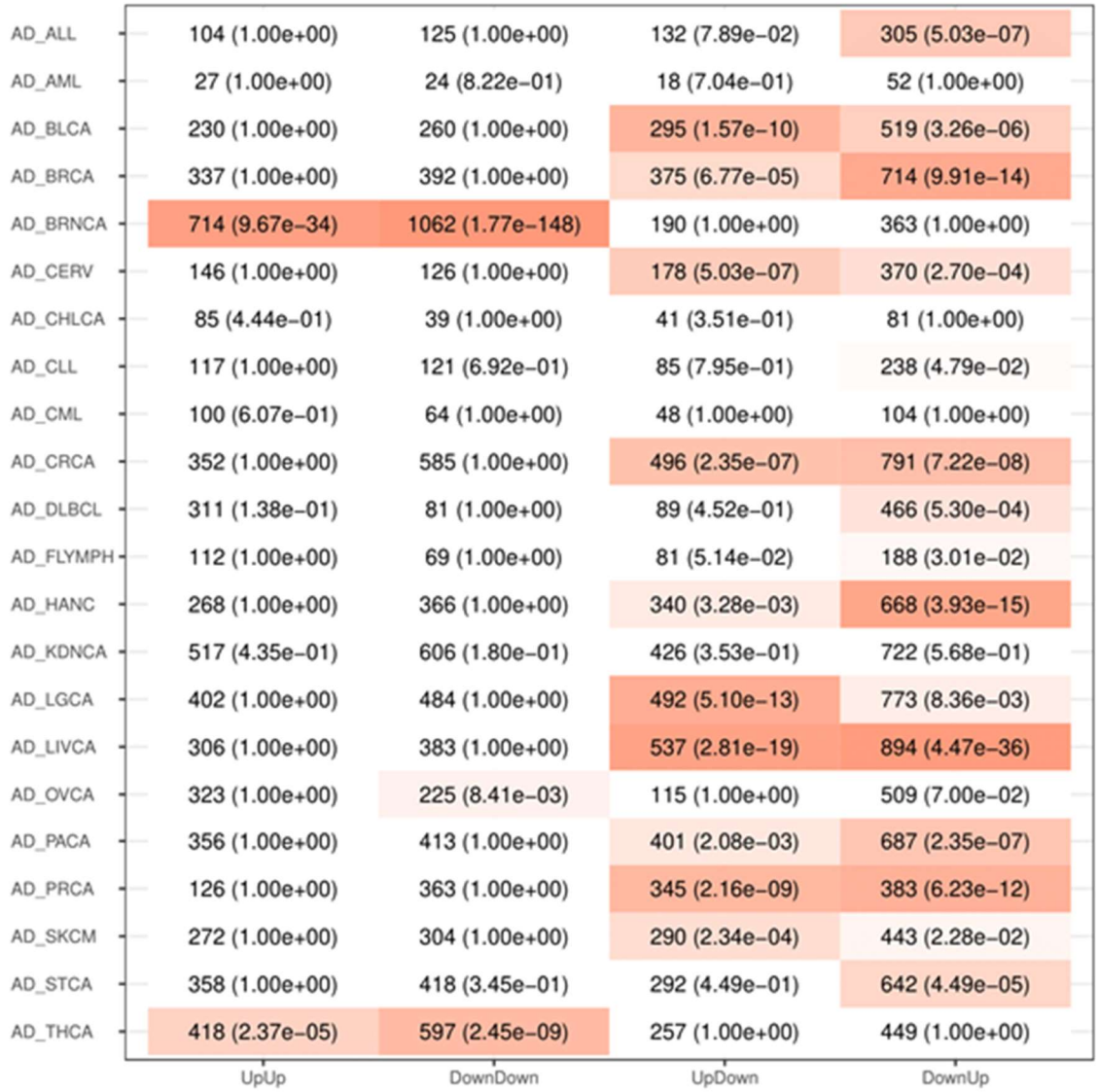

Figure 8: AD and cancer intersection analysis results. Each column shows the number of genes placed at the four possible intersections formed by the genes up and downregulated in each AD and cancer pair, as well as the FDR adjusted p-values of the exact Fisher's tests. Column 1 shows the genes jointly upregulated, column 2 shows the genes jointly downregulated, column 3 shows the genes upregulated in AD and downregulated in each cancer, and column 4 shows the genes downregulated in AD and upregulated in each cancer.

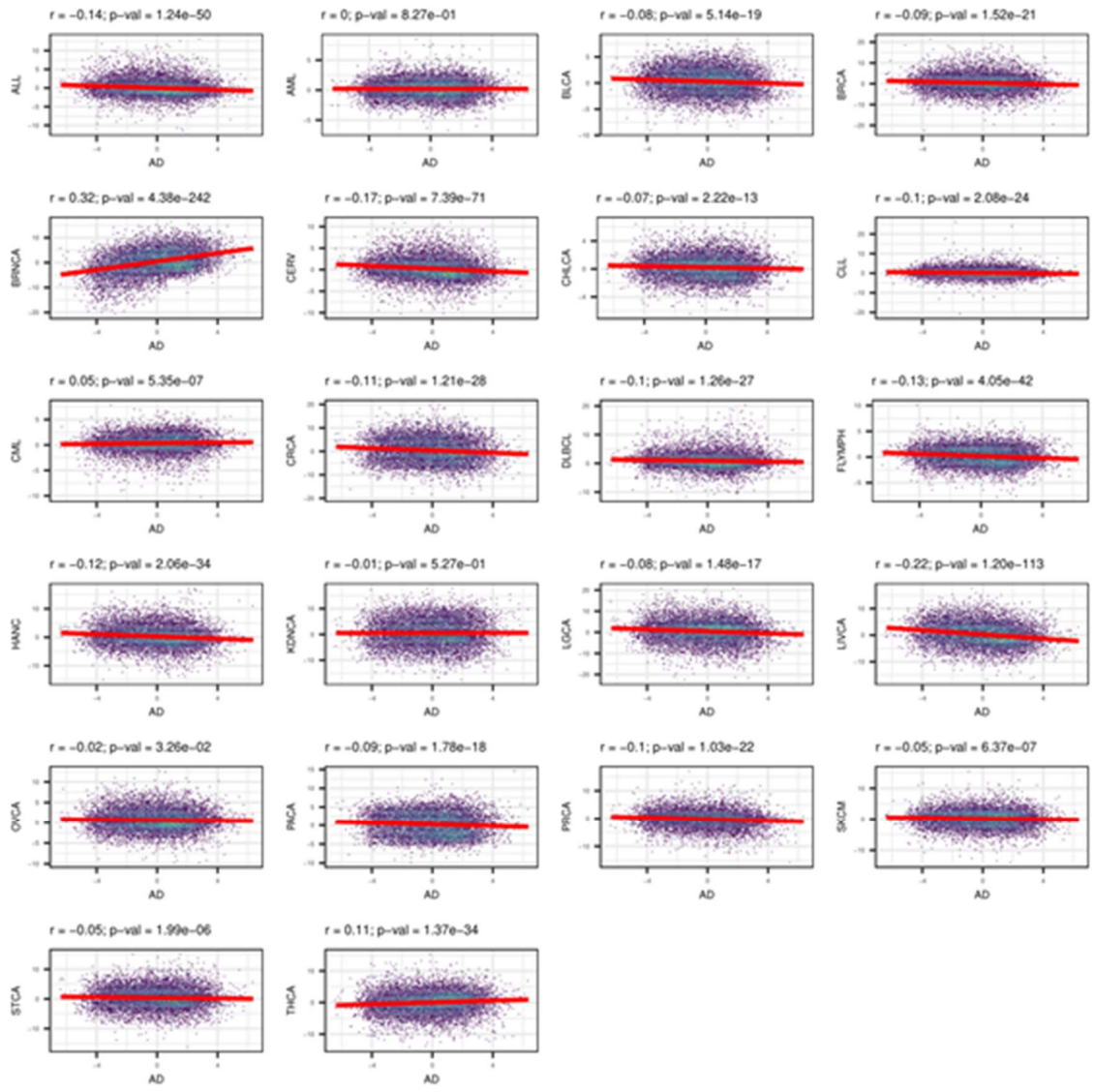

Figure 9: Pearson's correlations of the $\hat{\mu}$ values derived from the differential expression profiles of AD and all the included cancers.

PD and cancer

PD was found to present direct transcriptomic associations with thyroid, kidney, brain, and stomach cancers and inverse transcriptomic associations with breast, lung, and prostate cancers, and also CLL to a minor extent.

Jointly upregulated genes found in the intersections formed by PD and the three directly associated cancer types were mainly enriched in immune related processes including, leukocyte degranulation (GO:0043299; BRNCA: p-adj = 4.53e-06, KDNCA: p-adj = 2.40e-06, THCA: p-adj = 8.29e-06), myeloid leukocyte activation (GO:0002274; BRNCA: p-adj = 1.50e-06, KDNCA: p-adj = 4.11e-09), macrophage activation (GO:0042116; BRNCA: p-adj = 7.01e-03, KDNCA: p-adj = 4.94e-05, THCA: p-adj = 7.33e-05), and toll-like receptor signaling pathway (GO:0002224; BRNCA: p-adj = 2.28e-03, KDNCA: p-adj = 2.41e-04, THCA: p-adj = 2.47e-05), among others.

Genes placed at the intersections formed by jointly down-regulated genes in PD and each of the four directly associated cancer types were enriched in processes linked to oxidative phosphorylation (GO:0006119; BRNCA: p-adj = 2.17e-03, KDNCA: p-adj = 8.31e-21, THCA: p-adj = 1.18e-25), ATP synthesis coupled electron transport (GO:0042773; BRNCA: p-adj = 5.53e-04, KDNCA: p-adj = 4.29e-19, THCA: p-adj = 1.24e-19). Jointly downregulated PD and BRNA genes were also heavily enriched in related neuronal processes, such as synaptic vesicle cycle (GO:0099504; BRNCA: p-adj = 1.59e-15) and neuron projection morphogenesis (GO:0048812; BRNCA: p-adj = 1.05e-08).

In the case of inversely associated cancers, genes jointly upregulated in PD and downregulated in lung cancer were found to be enriched in positive regulation of cell migration (LGCA: p-adj = 1.50e-08) and regulation of vasculature development (LGCA: p-adj = 7.18e-08), whereas poor overrepresentation results were found for the rest of inversely associated cancers.

Genes jointly downregulated in PD and upregulated in inversely associated cancer types were enriched in mitochondrial processes such as, mitochondrial translation (GO:0032543; BRCA: p-adj = 3.79e-04, LGCA: p-adj = 2.60e-05, PRCA: p-adj = 5.49e-04), mitochondrial gene expression (GO:0140053; BRCA: p-adj = 7.01e-04, LGCA: p-adj = 9.16e-06, PRCA: p-adj = 5.49e-04). In the case of lung cancer enrichment in oxidative phosphorylation (GO:0006119; LGCA: p-adj = 1.02e-04) was also observed. Finally, genes linked to the regulation of cell cycle G2/M phase transition (GO:1902749; BRCA: p-adj = 6.31e-04, LGCA: p-adj = 3.58e-02) were also found to be downregulated in PD and upregulated in lung and breast cancers.

The differential gene expression profiles of brain cancer and kidney cancer were found to be positively correlated with the PD profile (BRNCA r = 0.33, KDNCA r = 0.17), whereas the correlation between the Parkinson's and thyroid cancer gene expression signatures was found to be negligible (THCA r = 0.03). Among the negatively associated cancers, only lung cancer presented a negative

correlation with an absolute value higher than 0.1 (BRCA r = -0.06, LGCA r = -0.13, PRCA r = -0.04, CLL r = -0.1) (**See Figure 11**).

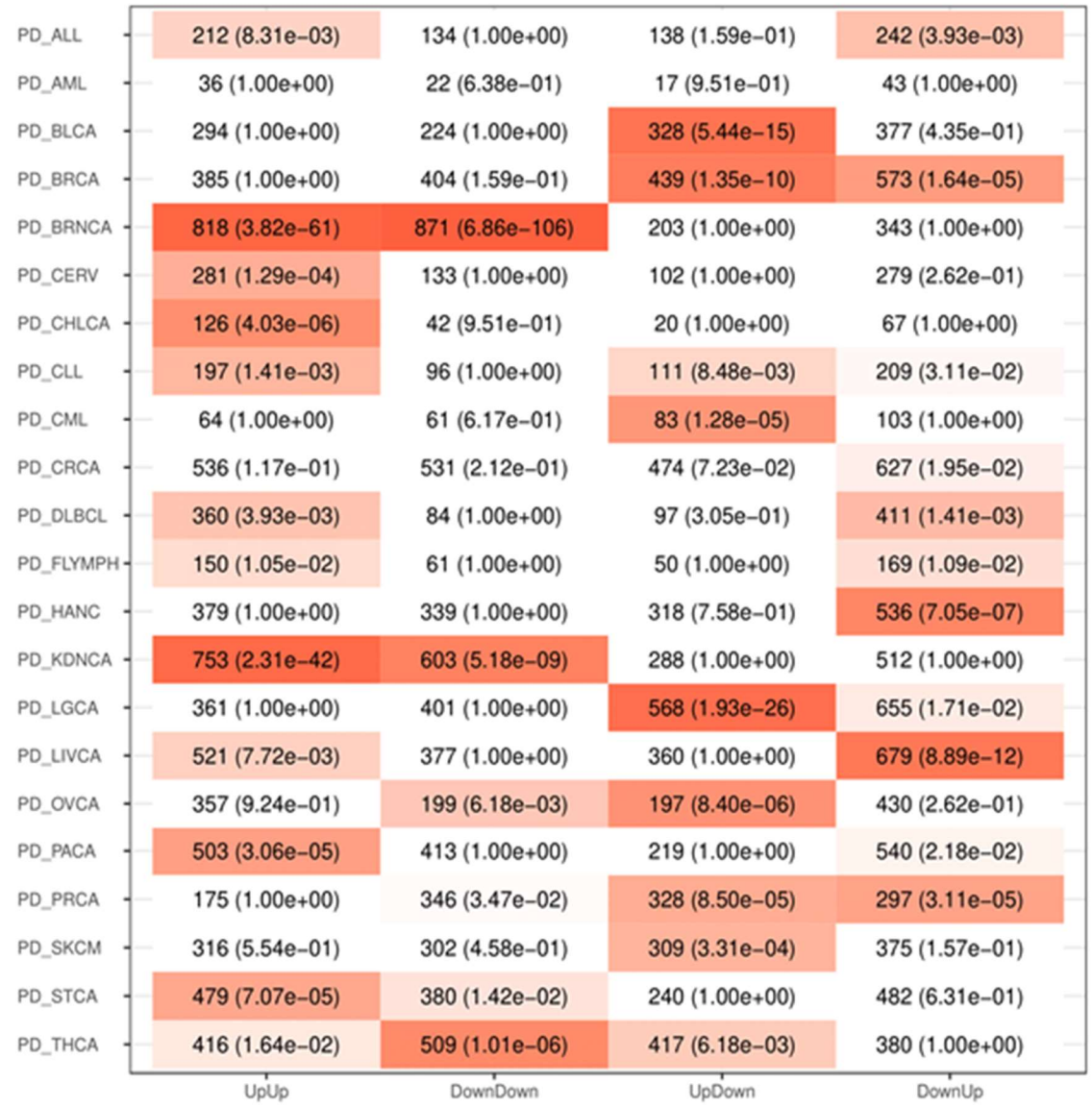

Figure 10: PD and cancer intersection analysis results. Each column shows the number of genes placed at the four possible intersections formed by the genes up and downregulated in each PD and cancer pair, as well as the FDR adjusted p-values of the exact Fisher's tests. Column 1 shows the genes jointly upregulated, column 2 shows the genes jointly downregulated, column 3 shows the genes upregulated in PD and downregulated in each cancer, and column 4 shows the genes downregulated in PD and upregulated in each cancer.

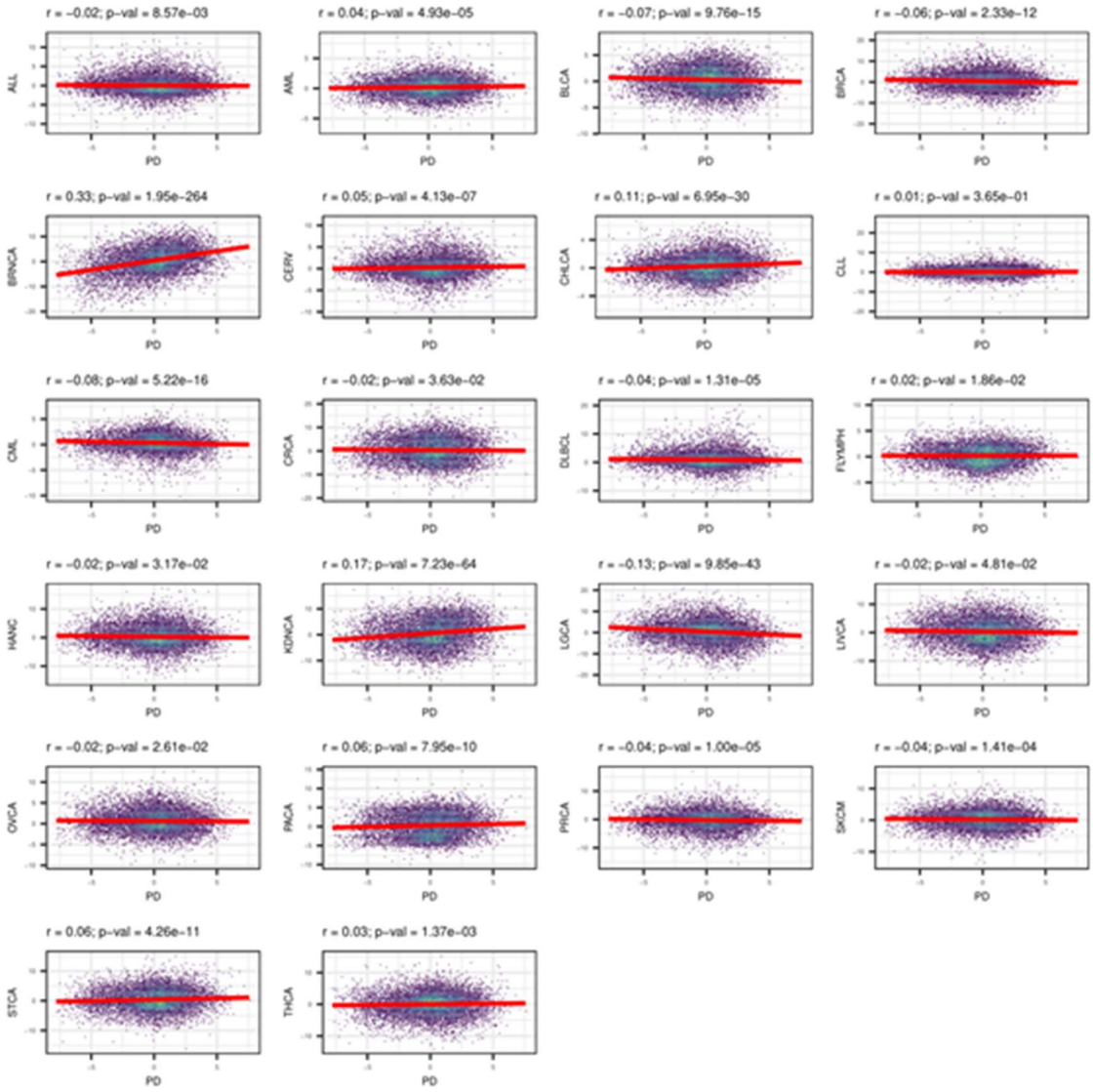

Figure 11: Pearson's correlations of the $\hat{\mu}$ values derived from the differential expression profiles of PD and all the included cancers.

HD and cancer

Huntington's disease was found to be directly associated with brain, kidney, pancreatic cancer, and to a minor extent with CLL, whereas negative associations between HD and breast, lung, and prostate cancers were also observed (**See Figure 12**). Genes jointly upregulated in HD and in the directly associated cancers were enriched in I-kappaB kinase/NF-kappaB signaling (GO:0007249; BRNCA: p-adj = 2.53e-03, KDNCA: p-adj = 8.48e-04, PACA: p-adj = 2.74e-03), defense response to virus (GO:0051607; BRNCA: p-adj = 6.12e-04, KDNCA: p-adj = 3.38e-05, PACA: p-adj = 6.11e-04), and regulation of cytokine production (GO:0001817; BRNCA: p-adj = 1.53e-03, KDNCA: p-adj = 8.48e-04, PACA: p-adj = 4.36e-03), among others. On the other hand, jointly downregulated genes were found to be enriched in processes linked to synapse and neurons, such as modulation of chemical synaptic transmission (GO:0050804; BRNCA: p-adj = 1.75e-26, KDNCA: p-adj = 1.02e-11, PACA: p-adj = 2.61e-12) and vesicle-mediated transport in synapse (GO:0099003; BRNCA: p-adj = 2.80e-18, KDNCA: p-adj = 2.60e-07, PACA: p-adj = 1.29e-08) among others. Genes linked to the ATP synthesis were also found to be jointly downregulated in HD and KDNCA ATP synthesis coupled electron transport (GO:0042773; KDNCA: p-adj = 1.12e-09), oxidative phosphorylation (KDNCA: p-adj = 1.61e-10). For those cancers inversely associated with HD, jointly HD upregulated genes and cancer downregulated genes were enriched in angiogenesis (GO:0001525; BRCA: p-adj = 1.15e-04, LGCA: p-adj = 2.22e-06, PRCA: p-adj = 7.27e-03) and blood vessel morphogenesis (LGCA: p-adj = 7.34e-08, PRCA: p-adj = 6.97e-03) whereas genes downregulated in HD and upregulated in inversely associated cancers were found to be poorly enriched in functional categories. Pearsons correlations between HD and SDDC were, r = 0.48, 0.16, 0.13, and 0.02 for bran, kidney, and pancreatic cancer, and CLL, respectively, whereas for ODDC were r = -0.09, -0.13, and -0.07 for breast, lung, and prostate cancers, respectively (**Figure 13**).

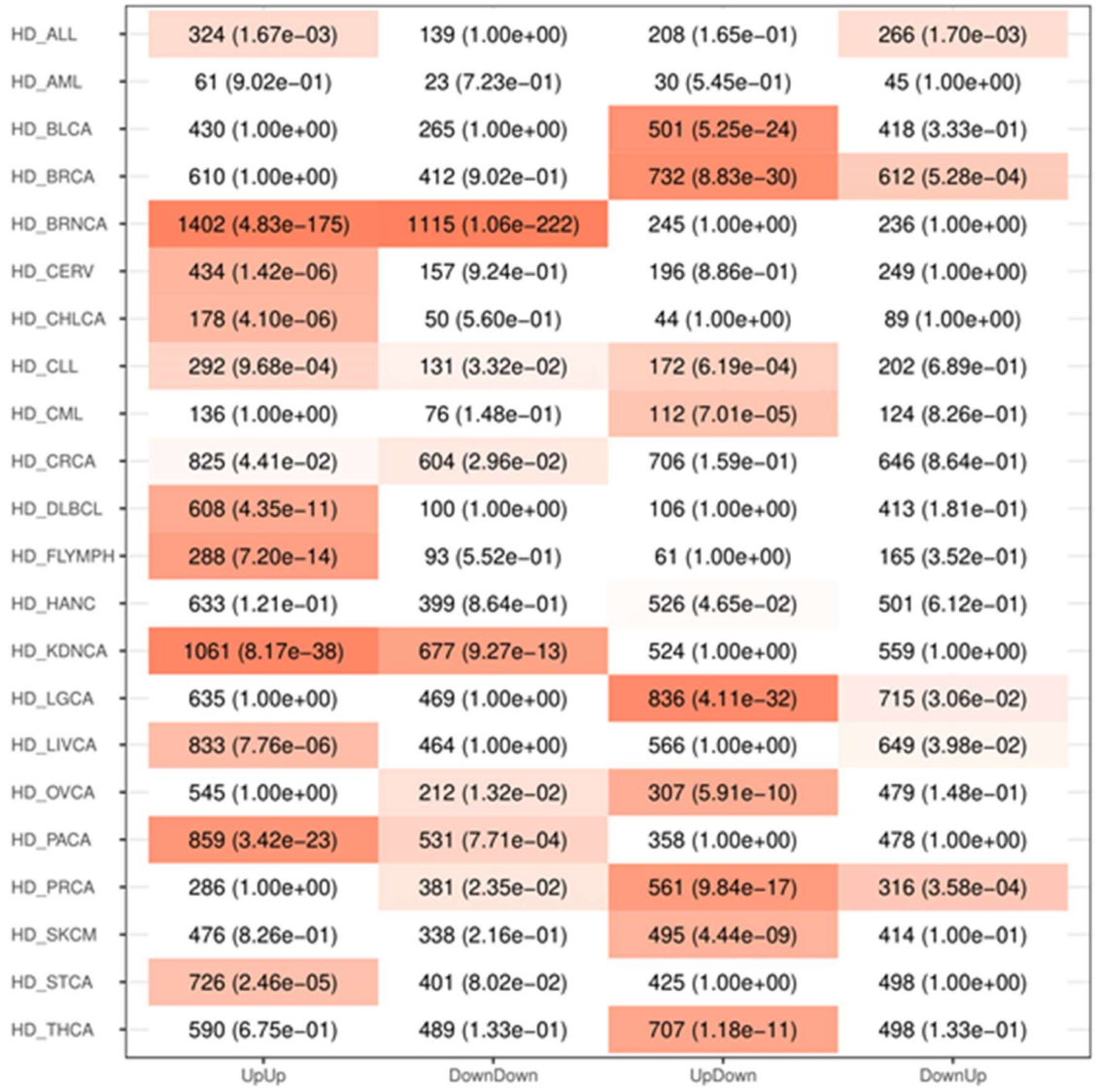

Figure 12 HD and cancer intersection analysis results. Each column shows the number of genes placed at the four possible intersections formed by the genes up and downregulated in each HD and cancer pair, as well as the FDR adjusted p-values of the exact Fisher's tests. Column 1 shows the genes jointly upregulated, column 2 shows the genes jointly downregulated, column 3 shows the genes upregulated in HD and downregulated in each cancer, and column 4 shows the genes downregulated in HD and upregulated in each cancer.

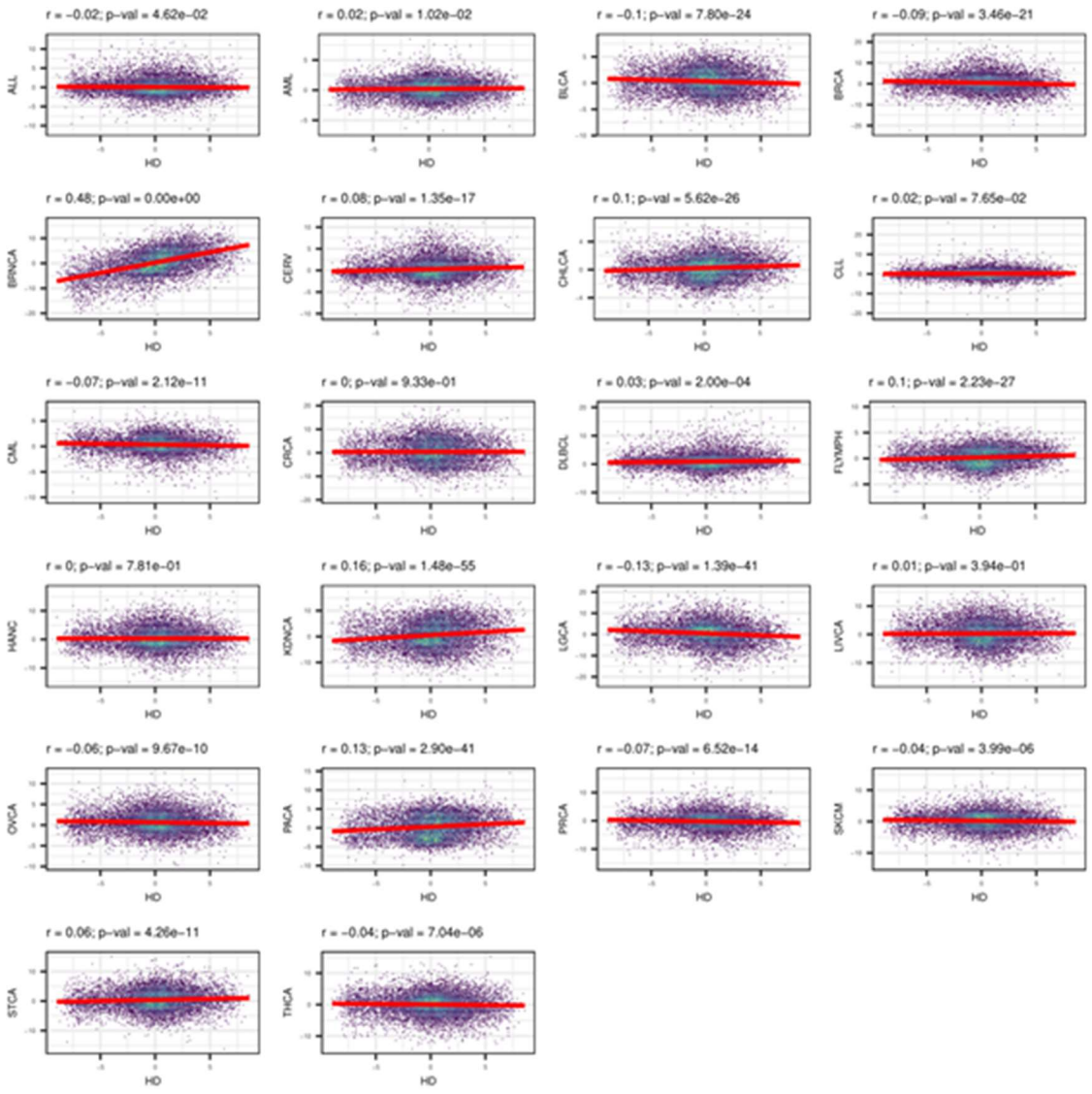

Figure 13: Pearson's correlations of the $\hat{\mu}$ values derived from the differential expression profiles of HD and all the included cancers.

ASD and cancer

ASD presented significant direct associations with brain, kidney, thyroid cancer, and cholangiocarcinoma, whereas no inverse associations were found. Genes jointly upregulated in ASD and in the directly associated cancers were found to be enriched in innate immune response (GO:0045087; BRNCA: p-adj = 3.28e-14, KDNCA: p-adj = 7.81e-18), cytokine production (GO:0001816; BRNCA: p-adj = 1.17e-10 , KDNCA: p-adj = 1.10e-14), wound healing (GO:0042060; BRNCA: p-adj = 1.83e-08, KDNCA: p-adj = 3.49e-06, THCA: p-adj = 1.11e-07), myeloid leukocyte mediated immunity (GO:0002444; BRNCA: p-adj = 8.52e-08, CHLCA: p-adj = 3.19e-03, KDNCA: p-adj = 5.39e-08, THCA: p-adj = 7.26e-11), T cell activation (GO:0042110; BRNCA: p-adj = 8.56e-08, CHLCA: p-adj = 4.08e-02, KDNCA: p-adj = 5.39e-08, THCA: p-adj = 3.43e-09), inflammatory response (GO:0006954; BRNCA: p-adj = 1.17e-10, CHLCA: p-adj = 1.41e-05, KDNCA: p-adj = 1.87e-13, THCA: p-adj = 6.32e-11), whereas genes jointly downregulated were enriched in biological processes linked to oxidative phosphorylation (GO:0006119; KDNCA: p-adj = 7.56e-06, THCA: p-adj = 1.10e-07), ATP synthesis coupled electron transport (GO:0042773; KDNCA: p-adj = 7.56e-06, THCA: p-adj = 4.69e-07). Brain, kidney, tyroid cancers, and cholangiocarcinoma presented Pearson's correlations of 0.38, 0.11, 0.11, and 0.05, respectively.

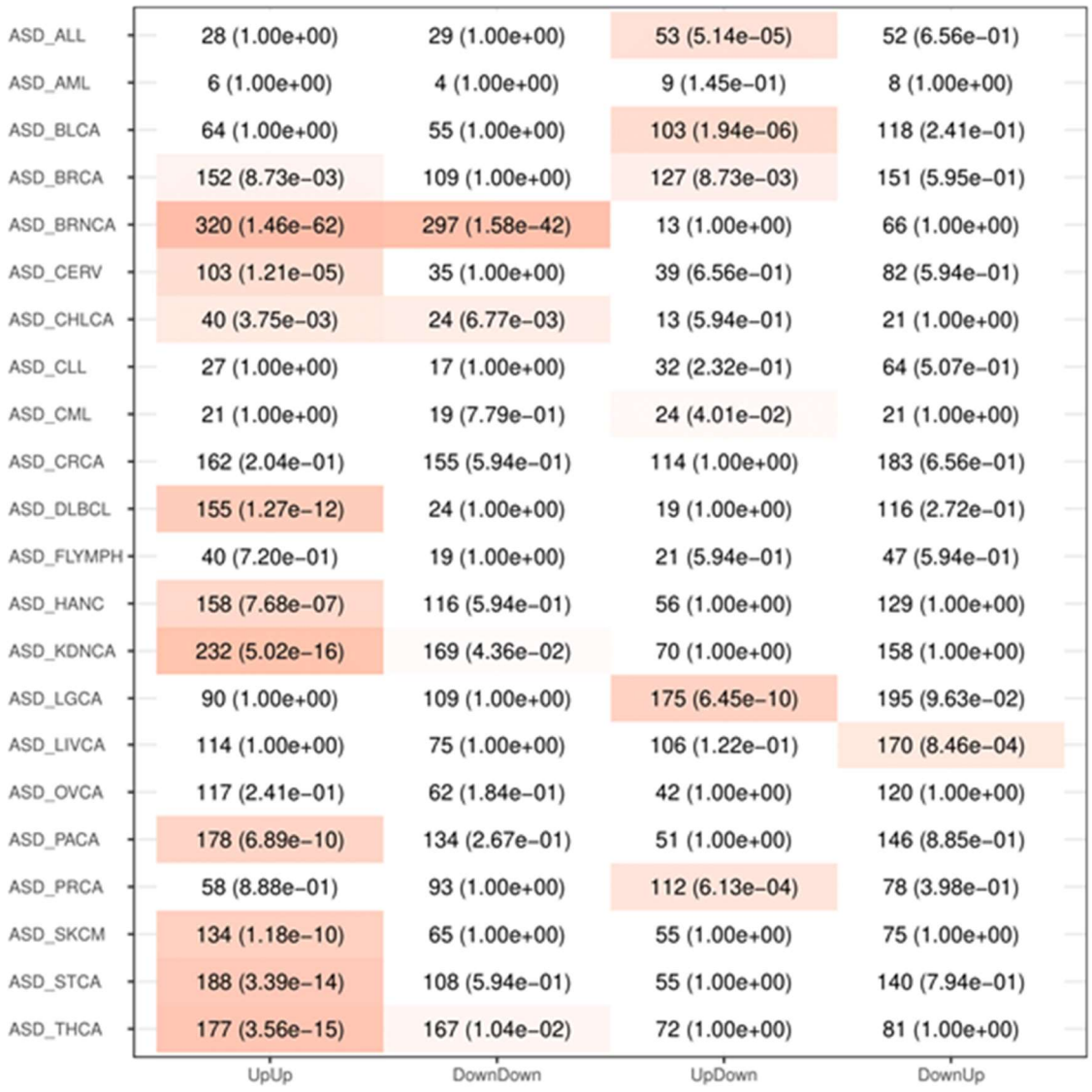

Figure 14: ASD and cancer intersection analysis results. Each column shows the number of genes placed at the four possible intersections formed by the genes up and downregulated in each ASD and cancer pair, as well as the FDR adjusted p-values of the exact Fisher's tests. Column 1 shows the genes jointly upregulated, column 2 shows the genes jointly downregulated, column 3 shows the genes upregulated in ASD and downregulated in each cancer, and column 4 shows the genes downregulated in ASD and upregulated in each cancer.

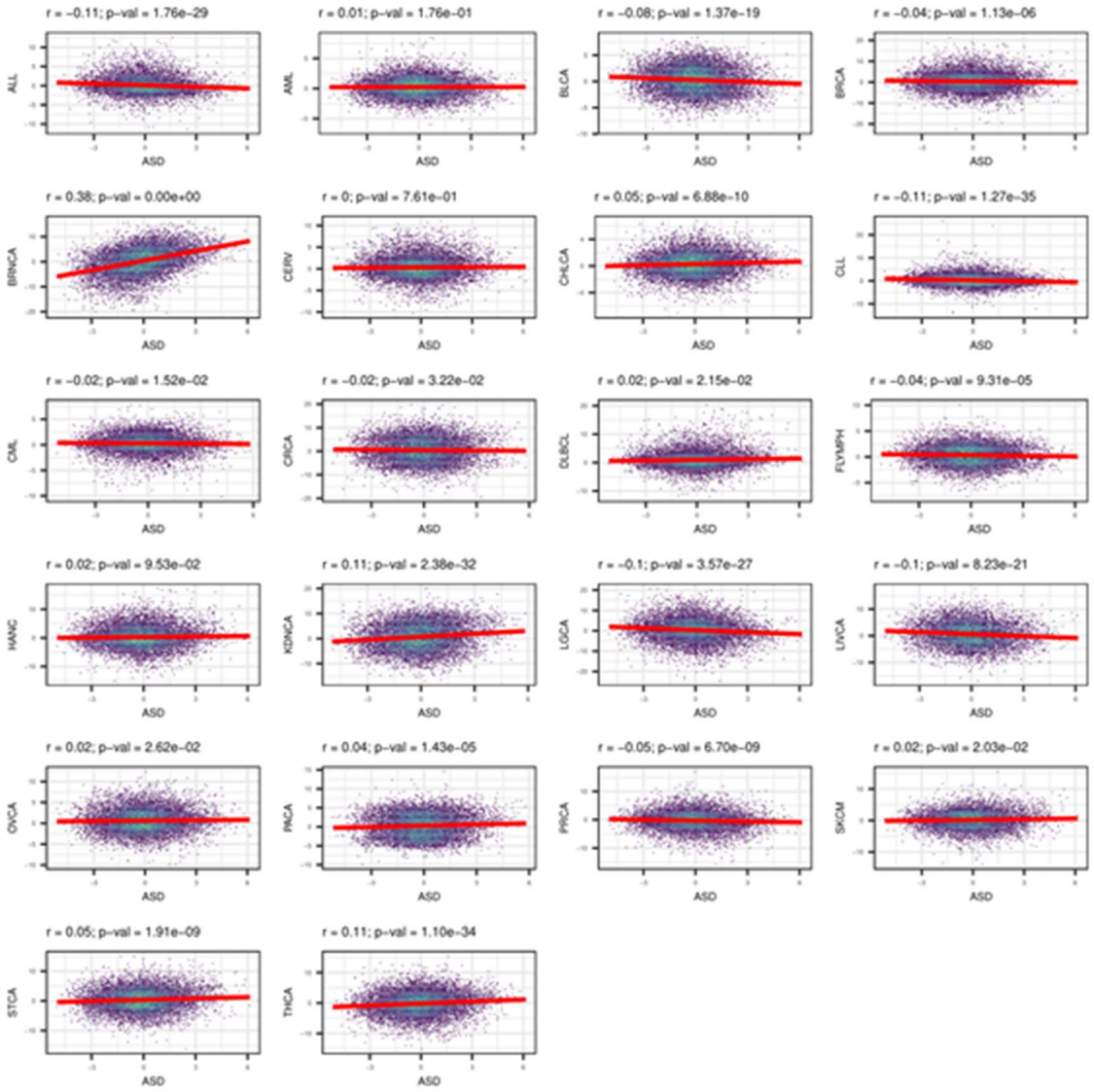

Figure 15: Pearson's correlations of the $\hat{\mu}$ values derived from the differential expression profiles of ASD and all the included cancers.

3.3.6 Gene set enrichment analyses comparison

In this section, we present the gene set enrichment analysis results. We carried out GSEA analysis using three molecular signature sets derived from the Molecular Signatures Database (MSigDB). The hallmarks gene sets (H, n = 50), the curated gene sets (C2, n =2868), which gathers information from canonical pathways derived from KEGG, REACTOME, and BIOCARTA, and the ontology gene sets (C5, n = 10271), which includes gene sets from the three Gene Ontology branches (BP, MF, and CC). These analyses serve to complement the overrepresentation analysis results of the genes placed at the intersections. In contrast to overrepresentation analysis, which works only with a list of genes of interest, Gene Set Enrichment Analysis (GSEA) uses the output of each differential expression meta-analysis to identify coordinated changes of expression in pathways even in the absence of significantly differentially expressed individual genes. Therefore, those disorders which were not included in the intersection analysis due to the lack of differentially expressed genes will be included in the GSEA enrichment analysis. For the sake of simplicity and interpretability, we will briefly summarize the number of pathways deregulated in the same and opposite directions on the of the three gene sets sources employed, and then we will focus on the results of the hallmarks gene sets, which include 50 specific well defined biological states which display coherent expression. The full GSEA enrichment analysis results for all the tested molecular signatures are provided in **Supplementary Appendices 3, 4, and 5**.

3.3.6.1 Pathways deregulated in the same and opposite directions in all pairwise CNS and cancer comparisons

Cancers that were found to present significant directly transcriptomic associations with CNS disorders had more pathways deregulated in the same direction than in the opposite direction. For instance, 11, 156, 648, H, C2, and GO gene sets presented significant joint deregulation in the same direction in AD and BRCNA (instance of direct transcriptomic association), whereas 6, 116, and 161 did it in opposite directions. The same trend was observed for AD and KDNCA, showing 15, 111, and 386 pathways deregulated in the same direction for H, C2, and GO, respectively and 6, 126, and 166 pathways deregulated in opposite directions. In contrast, cancers found to present opposite transcriptomic association patterns with a particular CNS disorder had more pathways deregulated in opposite directions. AD and LGCA presented the following number of jointly deregulated pathways (Same direction: H = 1, C2 = 29, GO = 69, Opposite direction: H = 18, C2 = 174, GO = 547) whereas AD and LIVCA had (Same direction: H = 2, C2 = 21, GO = 86, Opposite direction: H = 19, C2 = 233, GO = 514). **Supplementary Figures 79, 80, and 81** show the number of jointly deregulated pathways in the same and opposite directions for every CNS disorder and cancer pair and each source of gene sets (H, C2, and GO).

3.3.6.2 Detailed hallmarks (H) gene set results

Overall results

The most frequent upregulated hallmark gene sets in CNS disorders were found to be epithelial-mesenchymal transition, which was observed to be upregulated in all CNS disorders with the exception of MD, followed by IL6 JAK STAT3 signaling, which was found to be upregulated in AD, ASD, BD, PD, and SCZ but not in HD, and MD. Other immune-related sets of genes such as inflammatory response were found to be upregulated in AD, ASD, BD, and PD. In contrast, the most common downregulated hallmark gene sets in CNS disorders were oxidative phosphorylation, which was downregulated in all CNS disorders, MYC targets v1, and protein secretion. Those pathways were observed to be downregulated in all CNS disorders except in HD. Other important downregulated pathways in CNS disorders included MTORC1 signaling, and the unfolded protein response, which were downregulated in AD, BD, MD, PD, and SCZ and in AD, BD, MD, and PD, respectively. In contrast, MTORC1 signaling was found to be upregulated in ASD. For those CNS disorders that presented only a small amount of differentially expressed genes in the differential gene expression meta-analyses (BD, MD, and SCZ), gene set enrichment analysis also identified enriched pathways. This suggests the presence of small amounts of coordinated gene expression changes affecting specific pathways in those disorders.

For cancers, the most frequently upregulated hallmark gene sets were found to be MYC targets v1, E2F targets, and MYC targets v2. These pathways were found to be upregulated in 17, 17, and 16 site-specific cancers, respectively. MTORC1 signaling and G2M checkpoints were upregulated in 16 cancers, whereas DNA repair, glycolysis, unfolded protein response, and E2F targets did it in 15 cancer types. The most common downregulated pathways in cancer were bile acid metabolism, which was found to be downregulated in 13 site-specific cancers, and KRAS signaling dn. **Figure 21** summarizes the GSEA enrichment analysis results up- and downregulation status for all the studied disorders and the Hallmarks gene sets.

Specific analysis results for each CNS disorder and their SDDC and ODDC

Here we focus on the jointly deregulated pathways observed in the CNS and cancer pairs that presented significant direct (SDDCs) or inverse (ODDCs) transcriptomic associations in the intersection analyses. All the pathways reported had adjusted p-values < 0.05 in the CNS disorder and cancer under consideration.

AD and SDDC

For AD and its SDDCs, jointly upregulated pathways were mainly linked to immune function and included inflammatory response, allograft rejection, IL6 JAK STAT3 signaling, complement, TNFA signaling via NFKB, IL2 STAT5 signaling, and interferon-gamma response. Other pathways jointly upregulated in AD and its SDDCs were epithelial-mesenchymal transition, hypoxia, coagulation, and the P53 pathway, which were found to be jointly upregulated in AD and in BRNCA and THCA. Some instances of pathways jointly upregulated in AD and THCA were KRAS signaling and apical junction. Jointly downregulated pathways in AD and its SDDCs included oxidative phosphorylation, fatty acid metabolism, unfolded protein response, and adipogenesis.

AD and ODDC

Pathways upregulated in AD and downregulated in its ODDCs were found to be linked to immune-related processes, such as inflammatory response, IL6 JAK STAT3 signaling, allograft rejection, TNFA signaling via NFKB, interferon-gamma response, complement, IL2 STAT5, and KRAS signaling. Following the same trend, other gene set hallmarks included epithelial-mesenchymal transition, myogenesis, coagulation, hypoxia, and estrogen response early. Note that not all the listed pathways were found to be downregulated in all AD's ODDCs. The exact tumor types enriched in the reported processes can be consulted in **Figure16**.

Gene sets jointly downregulated in AD and upregulated in its ODDCs included MYC targets, the P53 pathway, MTORC1 signaling, DNA repair, unfolded protein response, UV response, MYC targets v2, and protein. Finally, oxidative phosphorylation was found to be downregulated in AD and upregulated in LGCA.

PD and SDDC

Hallmark gene sets jointly upregulated in PD, and its SDDCs also included many instances of immune system-related processes, including inflammatory response, allograft rejection, IL6 JAK STAT3 signaling, TNFA signaling via NFKB, among many others. In addition, other processes such as epithelial-mesenchymal transition, coagulation, hypoxia, and the P53 pathway were found to be upregulated in PD and its SDDCs. Apoptosis was jointly upregulated in PD and its SDDCs, including BRNCA, KDNCA, STCA, and THCA. The G2M checkpoint pathway was slightly upregulated in PD and heavily upregulated in its SDDCs. Instances of jointly downregulated processes between PD and its

SDDCs were oxidative phosphorylation, which was found to be strongly downregulated in PD and KDNCA, STCA, and THCA and adipogenesis.

PD and ODDC

Many processes upregulated in PD and downregulated in ODDCs were related to the immune system. Instances include inflammatory response, TNFA signaling via NFKB, IL2 STAT5 signaling, allograft rejection, and IL6 JAK STAT3 signaling, among many others, as well as other pathways, such as epithelial-mesenchymal transition and P53. Apoptosis was found to be jointly upregulated in HD and BRNCA, KDNCA, and PACA, as well as epithelial-mesenchymal transition. Other jointly upregulated pathways included coagulation and the P53 pathway. In contrast, hallmarks jointly downregulated included spermatogenesis and oxidative phosphorylation in BRNCA and KDNCA.

HD and SDDC

Like in the AD and PD cases, HD and its SDDCs presented a joint upregulation of immune system-related pathways, which involved TNFA signaling via NFKB, IL2 STAT5 signaling, and interferon-gamma and alpha response, among others. Apoptosis was found to be jointly upregulated in HD and BRNCA, KDNCA, and PACA, as well as epithelial-mesenchymal transition. Other jointly upregulated pathways included coagulation and the P53 pathway. In contrast, hallmarks jointly downregulated included spermatogenesis and oxidative phosphorylation in the cases of BRNCA, and KDNCA.

HD and ODDC

Several processes were upregulated in HD downregulated in ODDCs including, TNFA signaling via NFKB, IL2 STAT5 signaling, interferon-alpha and gamma response, the P53 pathway, epithelial-mesenchymal transition, and apoptosis, whereas pathways downregulated in PD and upregulated in ODDCs included spermatogenesis and oxidative phosphorylation in the case of LGCA.

ASD and SDDC

Hallmark gene sets jointly upregulated in ASD and its SDDCs were also linked to immune system-related functions, such as inflammatory response, allograft, IL6 JAK STAT3, TNFA signaling via NFKB, and IL2 STAT5 signaling, among many others, as well as other processes such as MTORC1 signaling, KRAS signaling, epithelial-mesenchymal transition, coagulation, hypoxia, and the P53 pathway. The apoptosis was also found to be jointly upregulated in ASD and BRNCA, KDNCA, and THCA. Jointly downregulated pathways included oxidative phosphorylation in the case of KDNCA and THCA and spermatogenesis.

Supplementary Appendix 1 Tables 8 to 37 show all jointly deregulated pathways found in CNS disorders and their SDDCs and ODDCs displaying adjusted FDR values lower than 0.001 in all the tested molecular signatures.

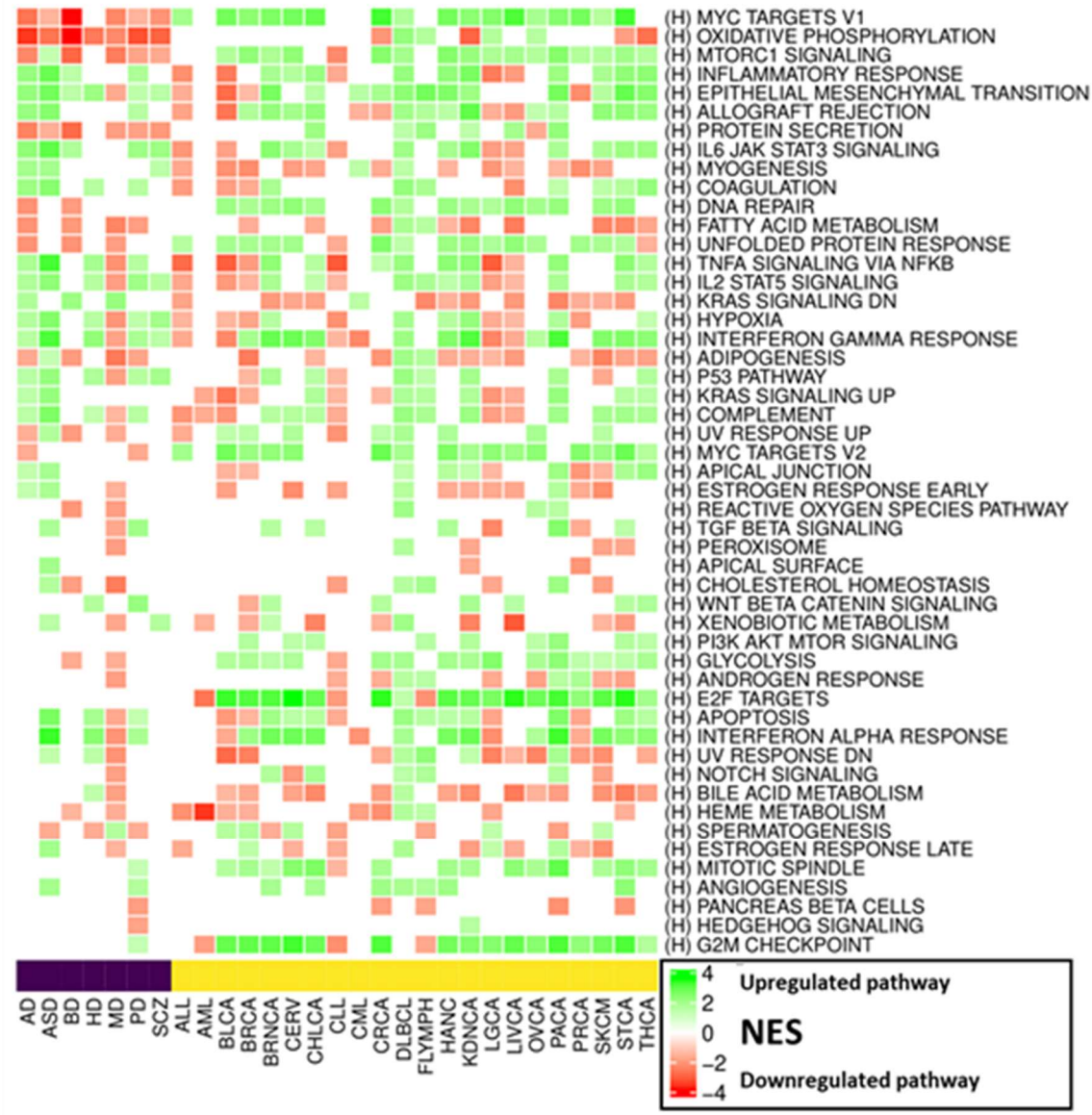

Figure 16: Heatmap representation of the Gene Set Enrichment Analysis (GSEA) results. The Hallmarks molecular signatures database was used as the source of gene set information. Green colors indicate that a pathway was found to be significantly upregulated in a particular disorder, whereas red colors indicate that the pathway was downregulated. White tiles are reserved for no significant associations. The intensity of the color is proportional to the value of NES. CNS disorders are annotated in purple, whereas cancers are annotated in yellow.

3.3.7 Validation of the intersection analysis results using an independent cohort of cancers

We carried out a validation step of the reported transcriptomic associations by performing differential gene expression analysis of independent cohorts of cancer samples derived from TCGA.

The validation step using independent cancer data derived from TCGA confirmed the direct associations found between AD and brain cancer but not between AD and thyroid cancer, which was found to be slightly negatively associated with AD in the TCGA datasets. Regarding the negatively associated cancers, the intersection analysis carried out using TCGA data confirmed the inverse associations found between AD and BLCA, BRCA, LGCA, LIVCA, and PRCA. The significant negative associations observed between AD and CERV, CRCA, HANC, and PACA could not be validated using TCGA data. In the case of PD, the direct transcriptomic associations observed with BRNCA and KDNCA in the array analyses were validated using TCGA data, as well as the negative associations identified between PD and BRCA, LGCA, and PRCA. However, the direct transcriptomic associations observed between PD and THCA or STCA could not be validated. In the case of HD, the significant direct transcriptomic associations with BRNCA and KDNCA were validated, as well as the negative associations with BRCA and LGCA, whereas the rest of the significant associations observed in the array-based analyses could not be validated. Finally, the transcriptome associations observed between ASD and BRNCA, and KDNCA were validated using the TCGA data. Additional significant associations were observed in the TCGA analysis, including inverse associations between ASD and BRCA, LGCA, and PRCA. **Appendix 6** includes the TCGA datasets employed in the validation analyses and a summary of the differential gene expression analysis results carried out for each, as well as the number of included samples and the correspondence with the respective array-based analysis. **Supplementary Appendix 1 Figures 82 to 85** show the intersection analysis results between CNS disorders and the TCGA cancer cohorts.

3.3.8 Consensus module detection and module overlap analysis results

In the next paragraphs, we report the consensus module detection and the module overlaps analysis results. In **3.3.8.1**, we provide a full description of the analyses for AD as an example. A summary of the results for all included disorders is available in **Section 3.3.8.2**. Finally, the module overlap analysis results are introduced **in Section 3.3.8.3**

3.3.8.1 Consensus gene co-expression module detection in Alzheimer's disease

Hybrid-signed weighted co-expression networks were constructed for AD. Guided by the **picSoftThreshold** function, the correlation matrices of each included dataset (GSE84422, GSE48350, GSE5281, GSE36980, GSE1297, E_MEXP_2280, and GSE29378) were raised to the 12, 11, 3, 13, 10, 15, and 10 exponents to fit each individual adjacency matrix to a power-law

distribution (**See methods section 3.2.11)**. **Figures 17 A and B** show the power-selection plots for the power-law distribution fit and the mean connectivity under the different tested powers for dataset GSE884422.

Thirty-two consensus modules of co-expressed genes (excluding the grey module) were identified across the seven AD datasets included in the analysis (**Figure 22 C and D**). Thirteen modules were found to be significantly associated with disease status. On the one hand, six of them (AD_orange, r = 0.36, p-adj = 9.00e-07; AD_tan, r = 0.34, p-adj = 3.42e-06; AD_brown, r = 0.31, p-adj = 5.15e-05; AD_purple, r = 0.29, p-adj = 1.81e-04; AD_darkred, r = 0.28, p-adj = 5.11e-04; AD_grey60, r = 0.21, p-adj = 2.60e-02) presented positive correlations with AD status, indicating that co-expressed genes included in these modules tend to be up-regulated in AD cases compared to controls. On the other hand, seven modules (AD_turquoise, r = -0.39, p-adj = 2.31e-08; AD_darkorange, r = -0.31 p-adj = 5.39e-05; AD_red, r = -0.29, p-adj = 1.71e-04; AD_paleturquoise, r = -0.29, p-adj = 2.96e-04; AD_skyblue, r = -0.24, p-adj = 5.26e-03; AD_saddlebrown, r = -0.23, p-adj = 1.38e-02; AD_darkgreen, r = -0.2, p-adj = 4.89e-02) were found to be negatively correlated with AD status. A bar plot depicting the number of genes included in each AD consensus module can be found in **Figure 17 E**. Among the positively associated modules AD_brown was found to be enriched in genes related to biological adhesion (GO:0022610; p-adj: 2.11e-12), immune response (GO:0006955; p-adj: 6.31e-09), and vasculature development (GO:0001944; p-adj: 7.50e-08), among others, as well as and in macrophages (p-adj: 1.80e-06) and microglia (p-adj: 7.68e-05) cell type-specific markers. AD_turquoise was the most significant negatively associated module and included genes that were found to be heavily enriched in processes linked to mitochondrial inner membrane (GO:0005743; p-adj: 8.89e-34), mitochondrial protein complex (GO:0098798; p-adj: 1.22e-33), and ATP synthesis coupled electron transport (GO:0042773; p-adj: 1.57e-21), and oxidative phosphorylation (GO:0006119; p-adj: 5.85e-21). Enrichment in cell type-specific markers related to neurons (p-adj: 3.07e-09), interneurons (p-adj: 1.08e-07), and purkinje neurons (p-adj: 2.68e-03) was also observed in this module.

AD_skyblue was enriched in protein processes linked to protein folding such as, unfolded protein binding (GO:0051082; p-adj: 7.77e-09), response to topologically incorrect protein (GO:0035966; p-adj: 9.13e-08), chaperone complex (GO:0101031; p-adj: 1.13e-07), response to unfolded protein (GO:0006986; p-adj: 4.82e-07).

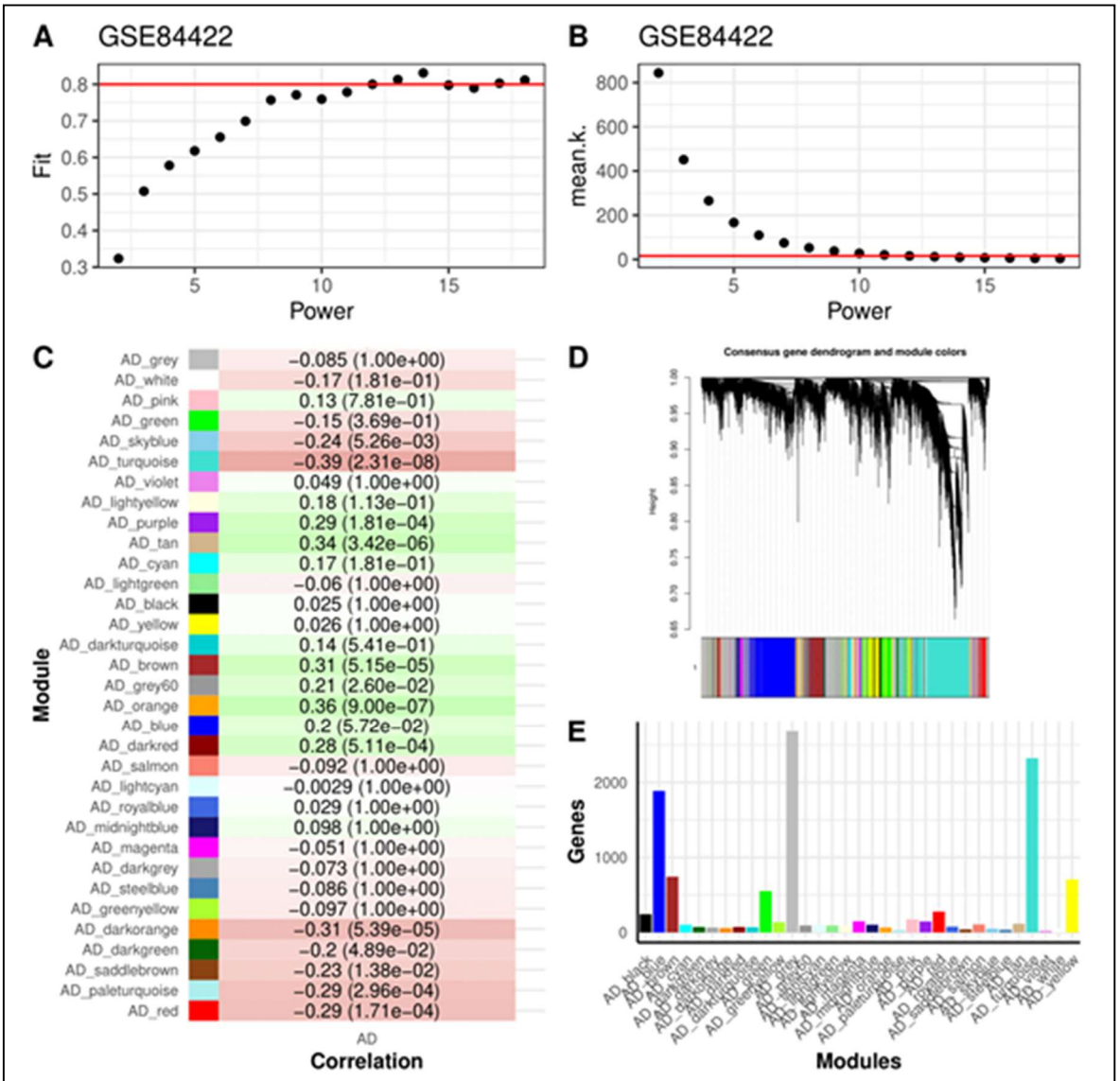

Figure 17: AD consensus co-expression module analysis results. A) Example of power selection based on the fit to a power law-distribution for dataset GSE88422 (Power selection is carried out independently for each dataset). B) Mean node connectivity base on the selected power. C) Module-trait correlation and their FDR adjusted p-values (Computed as explained in method section….). D) Dendrogram and color tags of the consensus modules identified across datasets. E) Gene content of each consensus module.

3.3.8.2 Consensus module analysis summary results for all the included disorders

In the previous section, we detailed the workflow and results of the co-expression analyses in the case of AD. For brevity, we present a summarized version of the results of the full set of included disorders. **Supplementary Appendix 6** contains information regarding all the modules significantly associated with disease status and their enrichment in biological processes and cell type-specific markers.

In total, 489 modules were found to be significantly associated with disease status after adjustment for multiple comparisons, of which 275 and 214 were found to present positive and negative correlations with disease status, respectively. Sixty-two modules belonged to CNS disorders, whereas 427 cancers were cancer-associated modules. In the case of major depression, only two gene co-expression modules were found to be significantly correlated with disease status, whereas no significantly associated modules were observed in the SCZ analyses.

In neurodegenerative disorders and ASD, several instances of co-expression modules enriched in immune system-related processes and specific cell-type markers were found to be positively correlated with disease status, including AD_brown (r = 0.31, p-adj: 5.15e-05), which was found to be enriched in immune response (GO:0006955; p-adj = 6.31e-09), and macrophages (p-adj: 1.80e-06) and microglia (p-adj: 7.68e-05) cell type makers, ASD_midnightblue (r = 0.26, p-adj: 2.52e-02), which was also enriched in immune system process (GO:0002376; p-adj = 6.50e-26), macrophages (p-adj: 1.25e-12), and microglia (p-adj: 2.67e-12), HD_magenta (r = 0.46, p-adj = 1.85e-04), which presented enrichment in immune response (GO:0006955; p-adj = 2.96e-10) and inflammatory response (GO:0006954; p-adj = 3.13e-10), and PD_black (r = 0.3, p-adj: 2.93e-03), which was enriched in cytokine-mediated signaling pathway (GO:0019221; p-adj = 1.58e-05) and response to cytokines (GO:0034097; p-adj = 1.65e-05).

In contrast, immune system-related co-expression modules were found to be negatively correlated with disease status in the MD_royalblue module (r = -0.25, p-adj: 3.36e-03), immune response (GO:0006955; p-adj = 3.53e-19), and in the BD_orange (r = -0.27, p-adj: 6.72e-03) module, immune response (GO:0006955; p-adj = 2.26e-14). In the cases of HD and PD, instances of modules positively correlated to disease status which where enriched in oligodendrocyte cell type markers HD_black (r = 0.62, p-adj: 1.16e-09), oligodendrocytes (p-adj: 1.39e-05) and PD_yellow (r = 0.35, p-adj: 1.64e-04), myelination (GO:0042552; p-adj = 1.98e-06), oligodendrocytes (p-adj: 2.56e-27) were observed.

Neurodegenerative disorders and ASD presented multiple instances of gene co-expression modules, which were negatively correlated to disease status and heavily enriched in mitochondrial activity, ATP synthesis functions and neural cell type-specific markers. AD_turquoise (r = -0.39, p-

adj: 2.31e-08) was found to be enriched in the mitochondrial inner membrane (GO:0005743; p-adj = 8.89e-34), ATP synthesis coupled electron transport (GO:0042775; p-adj = 1.57e-21), and synapse (GO:0045202; p-adj = 1.40e-23) functions, and in interneuron (p-adj: 1.08e-07) and neuron (p-adj: 3.07e-09) cell-type markers. ASD_pink (r = -0.25, p-adj: 3.25e-02) was found to be enriched in mitochondrial ATP synthesis coupled to electron transport (GO:0042775; p-adj = 5.18e-08) and ASD_red (r = -0.26, p-adj: 2.38e-02) was enriched in presynapse (GO:0098793; p-adj = 5.01e-05) and synapse (GO:0045202; p-adj = 7.47e-04) gene ontology processes, and in Interneuron (p-adj: 1.83e-03) cell type markers, whereas HD_yellow (r = -0.56, p-adj: 1.61e-07) presented enrichment in oxidative phosphorylation (GO:0006119; p-adj = 2.50e-26), and HD_turquoise (r = -0.7, p-adj: 3.40e-14) was found to be enriched in synapse-related genes (GO:0045202; p-adj = 3.71e-51) and interneuron markers (p-adj: 1.52e-15). PD_turquoise (r = -0.41, p-adj: 1.49e-06), was enriched in mitochondrial ATP synthesis coupled electron transport (GO: 0042775; p-adj = 3.28e-25), dopaminergic neurons (p-adj: 6.11e-03) and interneurons (p-adj: 1.57e-04).

Most cancers presented gene co-expression modules positively correlated with disease status, which were enriched in cell cycle-related functions. A non-comprehensive list of instances includes BLCA_turquoise (r = 0.53, p-adj: 1.10e-17), cell cycle (GO:0007049; p-adj = 9.88e-69), BRCA_green (r = 0.43, p-adj: 2.03e-64), cell cycle (GO:0007049; p-adj = 3.98e-115), BRNCA_brown (r = 0.35, p-adj: 8.62e-37), cell cycle (GO:0007049; p-adj = 3.41e-97) or CHLCA_brown (r = 0.52,,p-adj: 1.13e-11), mitotic cell cycle process (GO:1903047; p-adj = 1.81e-64).

Many of the studied cancer modules presenting significant negative correlations with disease status were enriched in biological processes and cell type markers characteristic from healthy tissues, suggesting that a process of dedifferentiation or tissue substitution has been taking place.

To cite some instances, the genes placed at the CERV_yellow module (r = -0.58, p-adj: 3.14e-13), the top negative correlated co-expression module found in the cervical cancer analysis, was found to be enriched in biological processes linked to cornification (GO:0070268; p-adj = 4.56e-28), keratinocyte differentiation (GO:0030216; p-adj = 3.12e-26), and epidermis development (GO:0008544; p-adj = 1.17e-25), as well as in keratinocyte (p-adj: 2.39e-13), epithelial (p-adj: 1.03e-03) , and basal (p-adj: 3.11e-03) cell type-specific markers. The BRNCA_blue module (r = -0.55, p-adj: 1.79e-100) was found to be enriched in biological process linked to synapse (GO:0045202; p-adj = 2.59e-36), axon (GO:0030424; p-adj = 6.09e-33), and neuron projection (GO:0043005; p-adj = 1.71e-27) and to Interneural (p-adj: 4.59e-10) and neural (p-adj: 6.24e-07) markers. CRCA_blue (r = -0.56, p-adj: 8.54e-109) was found to be enriched in lipid metabolic process (GO:0006629; p-adj = 5.78e-20), extracellular exosome (GO:0070062; p-adj = 5.72e-18), and in cell type-specific markers of enterocytes (p-adj: 7.55e-11), paneth cells (p-adj: 1.03e-12), and goblet cells (p-adj: 1.00e-04).

Finally, LGCA_turquoise (r = -0.74, p-adj: 1.98e-169) and LGCA_pink (r = -0.63, p-adj: 1.73e-100) were found to be enriched in pulmonary alveolar type I cells (p-adj: 2.54e-10) and pulmonary alveolar type II cells (p-adj: 8.14e-05), respectively.

Immune system-related gene co-expression modules were observed to be both positive and negatively correlated to disease status in different cancer types. For instance, BRCA_darkgreen (r = 0.27, p-adj: 1.52e-24) was found to be enriched in type I interferon signaling pathway (GO:0060337; p-adj = 4.44e-35), and innate immune response (GO:0045087; p-adj = 2.80e-31) biological processes, and in monocytes (p-adj: 1.33e-11) specific cell type markers, whereas BRCA_tan (r = -0.31, p-adj: 2.11e-31) was enriched in immune system process (GO:0002376; p-adj = 1.86e-16) and macrophage (p-adj: 5.31e-10) markers. Colorectal cancer status was also found to be both positively and negatively correlated with co-expression modules linked to immune system functions. CRCA_grey60 (r = 0.12, p-adj: 2.25e-04) was enriched in defense response (GO:0006952; p-adj = 4.02e-32), response to type I interferon (GO:0034340; p-adj = 2.67e-29), and monocytes (p-adj: 2.22e-07), whereas CRCA_black (r = -0.34, p-adj: 1.02e-34) was enriched in immune system process (GO:0002376; p-adj = 3.71e-71), lymphocyte activation (GO:0046649; p-adj = 2.06e-58), and the following specific cell type markers, B cells (p-adj: 1.17e-23), NK cells (p-adj: 1.43e-14), and T cells (p-adj: 3.20e-26). Lung cancer status was positively correlated to LGCA_grey60 (r = 0.37, p-adj: 1.53e-28), a module which was found to be enriched in B cell activation (GO:0042113; p-adj = 3.75e-13) and lymphocyte activation (GO:0046649; p-adj = 2.28e-11), as well as, in B cell (p-adj: 7.04e-28) markers, whereas it was also found to be negatively correlated to LGCA_green (r = -0.39, p-adj: 2.33e-32) which was enriched in immune response (GO:0006955; p-adj = 1.23e-62), and macrophage (p-adj: 3.84e-24), and dendritic cell (p-adj: 4.72e-12) markers. These findings underlie the complexity of the alterations of the immune function present in cancer.

Modules enriched in processes linked to the mitochondrial function were found to be correlated to the disease status of several cancer types. Both positive and negative correlations were observed. Instances of co-expression modules negatively correlated to disease status and enriched in mitochondrial function and ATP synthesis related processes are THCA_salmon (r = -0.32, p-adj: 7.71e-06), mitochondrial ATP synthesis coupled electron transport (GO:0042775; p-adj = 2.47e-50), KDNCA_turquoise (r = -0.78, p-adj: 4.95e-155), mitochondrial ATP synthesis coupled electron transport, (GO:0042775, p-adj = 3.53e-15), CRCA_magenta (r = -0.44, p-adj: 1.61e-60), ATP synthesis coupled electron transport (GO:0042773; p-adj = 6.49e-53). Whereas instances of positively correlated modules include PACA_white (r = 0.23, p-adj: 8.77e-06), ATP synthesis coupled electron transport (GO:0042773; p-adj = 5.70e-16), LIVCA_darkmagenta (r = 0.17, p-adj: 4.36e-06), oxidative phosphorylation (GO:0006119; p-adj = 4.36e-09), LGCA_brown (r = 0.44, p-adj: 1.00e-41), ATP synthesis coupled electron transport (GO:0042773; p-adj = 5.38e-10), BRCA_red (r = 0.28, p-

adj: 3.06e-25), oxidative phosphorylation, (GO:0006119, p-adj = 3.08e-09), BLCA_black (r = 0.45, p-adj: 3.17e-12), oxidative phosphorylation (GO:0006119, p-adj = 1.56e-07). These results suggest that different site-specific tumors employ diverse strategies regarding energy metabolism regulation.

Finally, many instances of modules enriched in extracellular matrix organization-related processes were also found to be positively and negatively correlated with disease status. A non-exhaustive list of instances is provided next. BRCA_cyan (r = 0.3 ,p-adj: 9.04e-30), extracellular matrix organization (GO:0030198; p-adj = 1.73e-30), CRCA_pink (r = 0.45, p-adj: 2.14e-64), extracellular matrix organization (GO:0030198; p-adj = 2.66e-35), HANC_green (r = 0.44, p-adj: 1.08e-25), extracellular matrix organization (GO:0030198; p-adj = 3.10e-42), CERV_lightgreen (r = -0.5, p-adj: 5.71e-09), extracellular matrix organization (GO:0030198; p-adj = 3.50e-07), LIVCA_purple (r = -0.49, p-adj: 1.53e-51), cell adhesion (GO:0007155; p-adj = 2.04e-11), SKCM_purple (r = -0.43, p-adj: 2.26e-22), extracellular matrix organization (GO:0030198; p-adj = 1.28e-08).

3.3.8.3 Overlap analysis of the modules significantly associated with disease

Overall, six-thousand eight-hundred and ninety significant overlaps between modules significantly associated with disease status were observed after adjustment for multiple comparisons. One hundred and seventeen were pairs of CNS modules, 5566 were produced between pairs of cancer-associated modules, and 1207 were observed between CNS related and cancer-associated modules. Among the overlapping pairs of modules found between CNS disorders and cancers, 275 belonged to AD, 173 to ASD, 100 to BD, 400 to HD, 72 to MD, and 181 to PD. No significant module overlaps were observed between SCZ and cancer modules since no gene co-expression modules significantly correlated with SCZ status were observed in our analyses.

In the case of AD, the AD_brown module (r = 0.31, p-adj: 5.15e-05), which was found to be enriched in biological processes linked to cell adhesion (GO:0007155; p-adj = 6.62e-12) and immune response (GO:0006955; p-adj = 6.31e-09), and in macrophage (p-adj: 1.80e-06) and microglia (p-adj: 7.68e-05) cell markers, presented significant overlaps with 35 co-expression modules from 15 site-specific cancers which were found to be positively correlated with cancer status and which were mainly enriched in immune system and extracellular matrix organization related processes. Instances include but are not limited to modules linked to cancers which were found to present direct transcriptomic associations with AD, BRNCA_yellow (r = 0.12, p-adj: 9.43e-05), immune response (GO:0006955; p-adj = 3.21e-147), macrophages (p-adj: 6.19e-24), and microglia (p-adj: 1.56e-13) or THCA_blue (r = 0.42, p-adj: 1.26e-10), immune system process (GO:0002376; p-adj = 1.96e-104), macrophages (p-adj: 1.79e-36).

Examples of overlapping modules which were found to be jointly negatively correlated to AD and several site-specific cancers were also observed. AD_turquoise (r = -0.39, p-adj: 2.31e-08) which was eniched in ATP synthesis coupled electron transport (GO:0042773; p-adj = 1.57e-21) and in Interneurons (p-adj: 1.08e-07) and neuron (p-adj: 3.07e-09) markers presented significant overlaps with 25 cancer related modules derived from 12 site-specific tumors. For instance significant overlaps were found between AD_turquoise and BRNCA_blue (r = -0.55, p-adj: 1.79e-100), which was found to be enriched in neuronal related processes and makers, synapse (GO:0045202; p-adj = 2.59e-36), interneurons (p-adj: 4.59e-10) and neurons (p-adj: 6.24e-07), CRCA_magenta (r= -0.44, p-adj: 1.61e-60), which was enriced in oxidative phosphorylation (GO:0006119; p-adj = 4.91e-60), mitochondrial ATP synthesis coupled electron transport (GO:0042775; p-adj = 2.99e-53), KDNCA_turquoise (r = -0.78, p-adj: 4.95e-155), with contained genes linked to mitochondrial ATP synthesis coupled electron transport (GO:0042775; p-adj = 3.530167e-15) or THCA_salmon (r = -0.32, p-adj: 7.71e-06), which was also found to be enriched in mitochondrial ATP synthesis coupled electron transport (GO:0042775; p-adj = 2.47e-50).

Gene co-expression modules correlated in different directions with the disease status were observed between AD and cancer. AD_turquoise (r = -0.39, p-adj: 2.31e-08) presented significant overlaps with gene co-expression modules positively correlated with many cancers including BLCA_black (r = 0.45, p-adj: 3.17e-12), oxidative phosphorylation, (GO:0006119, p-adj = 1.55e-07), BRCA_red (r = 0.28. p-adj: 3.06e-25), oxidative phosphorylation, (GO:0006119; p-adj = 3.08e-09), LGCA_brown (r = 0.44, p-adj: 1.00e-41), ATP synthesis coupled electron transport, (GO:0042773, p-adj = 5.38e-10), LIVCA_darkmagenta (r = 0.17, p-adj: 4.36e-06), oxidative phosphorylation (GO:0006119; p-adj = 4.36e-09), or PACA_white (r = 0.23, p-adj: 8.77e-06), ATP synthesis coupled electron transport (GO:0042773; p-adj = 5.70e-16). These observations highlight the importance of mitochondrial genes and energy synthesis related pathways in both AD and cancer and suggest that this could be an important axis in the modulation of their comorbidities. In addition, AD_skyblue (r = -0.24, p-adj: 5.26e-03), which was found to be enriched in protein folding (GO:0006457; p-adj = 1.58e-11), unfolded protein binding (GO:0051082; p-adj = 7.77e-09), response to topologically incorrect protein (GO:0035966; p-adj = 9.13e-08), and chaperone complex (GO:0101031; p-adj = 1.13e-07) biological processes, presented significant overlaps with 17 up-regulated cancer related co-expression modules.

AD_orange, AD_brown, and AD_purple are instances of modules that were found to be positively correlated to AD status and for which significant overlap with many instances of downregulated cancer-related co-expression modules was observed. As discussed previously, AD_brown was enriched in biological processes linked to adhesion and immune response. The

presence of overlaps of this AD module with both upregulated and downregulated cancer modules indicates the complexity of cancer's immune system scenario.

ASD, PD, and HD followed a similar pattern of module overlap with cancer than AD. For instance, two ASD upregualted co-expression modules, ASD_salmon and ASD_midnightblue, which were found to be enriched in response to stress (GO:0006950; p-adj = 1.84e-14) and immune system-related processes (GO:0002376; p-adj = 6.50e-26) presented significant overlaps with several cancer modules, both up and downregulated, whereas ASD_pink, HD_yelow, and PD_turquoise, three modules found to be negatively correlated with disease status and heavily enriched in genes linked to the oxidative phosphorylation presented significant overlaps with many cancer consensus modules, which were, both positively and negatively correlated with disease status. In general, these results highlight the potential importance of immuneand mitochondrial-related processes in the comorbid associations between CNS disorders and cancers.

A small number of co-expression modules were found to be associated with neuropsychiatric disorders. No disease module association was observed in SCZ, whereas only two co-expression modules were significantly correlated to MD. Both, MD_black (r = -0.23, p-adj: 1.29e-02) and MD_royalblue (r= -0.25, p-adj: 3.36e-03) were found to be negatively correlated to disease status and were enriched in anatomical structure morphogenesis (GO:0009653; p-adj = 1.35e-09) and immune response (GO:0006955; p-adj = 3.53e-19), respectively. Both modules presented overlaps with modules significantly associated with cancer in both directions. BD status was found to be significantly correlated to 6 co-expression consensus modules, which also presented significant overlaps with cancer positively and negatively correlated modules.

3.4 Discussion

We observed significant differences in the number of differentially expressed genes among the tested CNS disorders. A substantial amount of transcriptomic alteration was present in the meta-analyses of the neurodegenerative disorders (AD, HD, PD) and ASD suggesting that they are characterized by important changes in gene expression patterns and tissue composition. On the contrary, in the cases of BD, MD, and SCZ, the analyses yielded a low number of DEGs, indicating that the brain transcriptomes of patients and controls do not differ considerably, at least in the brain regions selected for the study. Enrichment analysis results and consensus module analyses pointed towards the presence of aberrant expression of several pathways and biological processes. Immune system-related pathways were found to be upregulated in most CNS disorders, whereas ATP synthesis and oxidative phosphorylation-related genes were found to be downregulated in all of them. Interestingly, gene set enrichment analysis showed that pathways alterations were also present in BD, MD, and SCZ. This suggests that even if differential gene expression lacks the power to identify individually deregulated genes, slight coordinated changes of gene expression of specific pathways occur in these disorders. In addition, these changes resemble those observed in the set of neurodegenerative diseases and in ASD in terms of the affected pathways. mTORC1 signaling was another instance of a pathway that was found to be downregulated in all CNS disorders, with the exception of ASD, where it was found to be upregulated.

For cancers, differential gene expression analyses yielded large amounts of DEGs, ranging from 3.57% of the tested genes observed in AML to 69.24% of tested genes observed in KDNCA. These observations are in agreement with previous differential gene expression studies. In general, pathways and modules upregulated in cancers were linked to well-known cancer-related processes, such as cell cycle, DNA repair, and EF2 targets. mTORC1 signaling was also found to be upregulated in most site-specific cancers. More heterogeneous results were observed in pathways linked to the immune function. For instance, the inflammatory response was found to be downregulated in 5 cancer types, whereas it was found to be upregulated in ten cancer types showing the complex interplay that takes place between the immune system and cancer.

The intersection analysis yielded several significant associations between CNS disorders and cancers. BRNCA was found to present significant direct patterns of transcriptomic deregulation with AD, PD, HD, and ASD. Thyroid cancer was found to be positively associated with AD, PD, and ASD, whereas kidney cancer was found to be directly linked to PD, HD, and ASD. In addition, PD was found to be directly linked to stomach cancer, ASD to cholangiocarcinoma, and HD to both pancreatic cancer and CLL.

Opposite patterns of transcriptomic deregulation were also observed between CNS disorders and cancers. Breast, lung, and prostate cancers were negatively associated with the three neurodegenerative disorders (AD, PD, and HD). AD was also found to be negatively associated with four additional cancer types, including (BLCA, CERV, CRCA, HANC, LIVCA, and PACA). Finally, PD was also observed to be inversely linked to CLL. Neuropsychiatric disorders were not included in the intersection analyses due to the lack of differential gene expression observed in meta-analyses.

An important number of the associations observed in the array-based analyses were validated using an independent cohort of cancers derived from TCGA, including the direct associations observed between AD and BRNCA, the negative associations between AD and BLCA, BRCA, LGCA, LIVCA, and PRCA. The positive transcriptomic associations observed between PD and BRNCA and KDNCA were also validated, as well as the opposite patterns of transcriptomic deregulation observed between PD and BRCA, LGCA, and PRCA. In the case of HD, both the direct associations observed with BRCNA and KDNCA and the inverse associations observed with BRCA and LGCA were replicated using the independent cancer cohort. Finally, the transcriptome associations observed between ASD and BRNCA and KDNCA were also observed when using TCGA data.

Differences with previously published studies

Previous studies have reported inverse transcriptomic associations between AD, PD, and SCZ and three tumor types (LGCA, PRCA, CRCA) and direct transcriptomic associations between AD and Glioblastomas [34, 250]. The inverse associations between LGCA, PRCA, and AD and PD were also observed in our analyses, as well as the inverse transcriptomic associations between AD and CRCA and the direct transcriptomic associations between AD and brain cancer. However, we could not reproduce the inverse transcriptomic associations between PD and CRCA, neither the associations between SCZ and LGCA, PRCA, or CRCA.

Several factors have the potential to account for these discrepancies. First, some methodological aspects are different in our study. The micro-array normalization method employed by the authors of the previous studies was fRMA [251], which utilizes information from publicly available Affymetrics microarray databases to precompute probe-specific effects and variances, whereas we used classic RMA. In addition, the authors limited the included array platforms to the two most popular Affymetrix microarray platforms. hgu133 plus2 and hgu133a, whereas we did integrate data from multiple single-channel microarray platforms from Affymetrix, Illumina, and Agilent. Second, the criteria for the cohort selection in CNS studies were different. We integrated data from only one specific brain region for each CNS disorder and included only datasets with samples not derived from the same subset of patients. In contrast, previous studies have included data from multiple brain regions, which in some instances were derived from the same subset of

patients. We tried to limit this since we consider that the inclusion of several samples derived from the same subset of patients could artificially inflate the differential gene expression meta-analyses results. In fact, the number of DEGs obtained in our analysis was comparatively lower than those observed both in [34, 35]. That is especially evident in the case of SCZ, for which we only observed 3 DEGs compared to the 1619 DEGs observed by Ibañez and co-workers [34]. In addition, and probably as a consequence of the previous comments, the p-values of the intersections reported in our work are comparatively higher than those observed previously. There is a trade-off between the number of disorders for which data is available and the amount of heterogeneity introduced by the use of multiple platforms.

The overrepresentation analyses of the genes placed at the intersections, the gene set enrichment analyses, and the module overlap comparisons pointed towards several pathways that could be important in modulating CNS and cancer comorbidities.

Cell cycle

Cell cycle alterations constitute one of the hallmarks of cancer [83]. Cell-cycle genes were found to be upregulated in the majority of our gene set enrichment analysis results of most cancers. In addition, the vast majority of cancer types were found to be positively correlated to gene co-expression modules that were heavily enriched in cell-cycle related genes. In principle, cancer and neurodegeneration are placed at two opposite ends in terms of cell behavior since cancer is due to uncontrolled cell growth, and neurodegeneration involves post-synaptic cells' death. However, an increasing body of evidence suggests that a dysfunctional neuronal cell cycle re-entry could precede neurodegeneration. For instance, significantly elevated levels of cyclin D, Cdk4, and Ki67 have been observed in AD neurons, suggesting their exit from the G0 phase and their progression through G1. M-phase markers such as MPM2, Cdc25 A and B, and phosphatases have also been reported to be elevated in AD neurons [252]. In addition, the presence of cyclin E-cdk2 has also been reported, indicating that neurons have passed G1 and are committed to cell division or death. However, no mitotic structures have been observed in AD neurons re-entering into the cell cycle. This fact suggests that the aberrant cell cycle re-entry in susceptible neurons could contribute to their cell death [253]. The cell cycle alterations do not limit to AD. Immortalized lymphocytes derived from PD patients have been found to present increased cell cycle activity [254], and cell cycle re-entry has also been observed in post-mitotic neurons of HD [255]. Furthermore, cell cycle alteration has been reported to play a critical role in ASD [256]. Our data suggest the presence of cell cycle alterations in some CNS disorders. Several cell cycle-related pathways were found to be downregulated in our gene set enrichment analysis of AD and PD. The overrepresentation analysis of the genes jointly downregulated in AD and PD and upregulated in their respective opposite direction deregulated cancers were also found to be enriched in cell-cycle genes. Altogether this

evidence point towards cell cycle alterations as a potential modulator of the comorbidities observed between CNS disorders and cancer. However, the mechanism by which the cell cycle is altered in CNS disorders and how it could modulate comorbid associations with cancer remains to be elucidated.

Myc signaling

Connecting with the previous section, we observed that MYC targets were downregulated in all the studied CNS disorders except for HD and upregulated in most cancers. MYC constitutes a family of transcription factors classified as proto-oncogenes, which regulate the expression of genes that participate in cell proliferation (cyclins and p21). They also promote cell growth and modulate apoptotic processes by downregulation of Bcl-2. A member of this family (c-MYC) has been found to be constitutively expressed in many cancer types, including cervix, colon, breast, lung, and stomach cancers. However, MYC dysregulation alone has been found not to be sufficient to induce cellular proliferation or neoplastic transformation, and other genetic events (loss of p53) are necessary to allow MYC to exert its influence in cellular proliferation and neoplastic transformation [257]. MYC has also been described to play a role in CNS disorders. As we have seen, in the context of AD, the cell cycle hypothesis tries to explain the presence of classic pathological hallmarks of AD and the expression of markers of cell proliferation. According to it, some markers linked to cell cycle progression in specific neuronal populations are dysregulated in AD despite the fact that neurons are post-mitotic cells incapable of division. Reactivation of the cell cycle in adult neurons *in vivo* and *in vitro* has been shown to produce AD-like changes, including neurodegeneration [258]. It has been observed that the neural expression of MYC causes neurodegenerative phenotypes in transgenic mice [258]. Many different proteins associated with the cell cycle, including cyclins, cyclin-dependent kinases, and proto-oncogenes such as c-MYC, are increased in degenerating neurons. This contrasts with our observations. A close examination of the differentially expressed pathways found in CNS disorders indicates that cell cycle-related genes are significantly downregulated rather than upregulated. However, the neurodegenerative process's dynamic nature makes it plausible that different pathway activation states are present as the disease progresses. The evidence of the involvement of the MYC family in neurodegeneration does not limit to AD. A member of the family of MYC (N-myc) has been found to regulate Parkin gene expression. Parkin expression was found to be inversely correlated with N-MYC levels in the developing mouse and human brains. In addition, tissue-specific induced expression of human c-MYC has been found to suppressed poly(Q)-mediated neurotoxicity in Drosophila.

The PI3K/AKT/MTOR axis

Genes belonging to the mTORC1 signaling pathway were found to be downregulated in five out of the seven CNS disorders included in our analysis. The exceptions were HD, for which no

association was found, and ASD, which was associated with a slight upregulation of this pathway. In contrast, sixteen of the 22 site-specific cancers, excluding ALL, AML, CHLCA, CLL, CML, and FLYMPH, were linked to an increase in the mRNA levels of mTORC1 signaling related genes. The PI3K/AKT/mTOR axis has a well-documented role in the nervous system where it participates in several processes, including neurogenesis, axonal sprouting, dendritic spine growth, myelination, axonal regeneration, and receptor channel expression [259]. In addition, a substantial amount of works have reported the involvement of the PI3K/AKT/mTOR axis in the pathology of central nervous system disorders using patients samples and animal models. AKT has been found to mediate the phosphorylation of the tau protein in AD through PI3K and GSK3β. GSK3β is inhibited by phosphorylation, which is produced as a consequence of PI3K/AKT activation. Therefore, the attenuation of PI3K/AKT/mTOR signaling would increase the unphosphorylated concentrations of GSK3β and, in turn, increase the levels of phosphorylated tau, inducing the formation of neurofibrillary tangles [260]. Moreover, it has been observed that increased concentrations of Aβ peptides enhance mTOR signaling, whereas further augments of the Aβ levels reaching cytotoxic concentrations produce decreases of the mTOR signaling. AD has also been linked to impaired insulin signaling [261]. And inhibition of PTEN has been found to recover synaptic function and cognition in AD animal models [262]. PI3K/AKT/mTOR signaling alterations have also been observed in PD and HD. In PD, the dysregulation of this axis has been linked to the loss of dopaminergic neurons [263] by a mechanism related to the regulation of apoptosis molecules [260] and has been observed in postmortem samples derived from PD patients [264]. Animal models of HD treated with insulin or insulin-like growth factor (IGF) presented improvements in microtubule transport, metabolic function, and autophagy, which resulted in the clearance of huntingtin aggregates and in the restoration of mitochondrial functions, and the amelioration of motor abnormalities [261]. Alteration of PI3K/AKT/mTOR signaling also seems to play an important role in bipolar disorder, major depression, and schizophrenia [259, 265-267]. In ASD, PTEN mutations (a negative regulator of the axis) have been found in patients with macrocephaly, and mutations in other genes encoding components of the PI3K/AKT/mTOR axis are responsible for disorders associated with syndromic cases of ASD, including fragile X syndrome, tuberous sclerosis, and type 1 fibromatosis [268-271], which are typically linked to an increase in the signaling of the pathway. Surprisingly, contrary to syndromic cases, idiopathic autism has been linked to a reduction of signaling through PI3K/AKT/mTOR in both patients and animal models [272]. Finally, PI3K/AKT/mTOR signaling plays a pivotal role in human tumors, in which it is involved in different processes, such as the regulation of cell proliferation, survival, metabolic reprogramming, invasion and metastasis, and the suppression of autophagy and apoptosis. A variety of the positive regulators of the axis, including the p110α and p85 α catalytic and regulatory subunits of class IA PI3K, as well as AKT, RHEB, mTOR,

and eIF4E present oncogenic potential and present activating mutations. The activation of tyrosine kinases and RAS (upstream modulators of the pathway activity) and the loss of function of negative regulators such as PTEN and TCS1 and 2, among others, are also usually found in cancer [273].

Mitochondrial dysfunction

Our results point towards mitochondrial dysfunction and alterations in the ATP synthesis process through oxidative phosphorylation as a shared feature of all the studied CNS disorders. These processes were found to be heavily enriched in genes that presented coordinated patterns of downregulation in our data. This was true even for the set of neuropsychiatric disorders for which no significant differentially expressed individual genes were identified. Impaired bioenergetic processes have been observed to be a common feature of neurodegenerative disorders [274]. Some authors have suggested that neuronal susceptibility to mitochondrial dysfunction might be explained by the fact neurons are highly dependent on oxidative phosphorylation [275]. Mounting evidence has been gathered for mitochondrial dysfunction in AD and includes the altered glucose metabolism observed in living AD subjects, the increased oxidative stress, and ROS damage observed in AD brains due to faulty electron transport chain, which results in DNA, lipids, and protein damage [274]. In addition, it has been recently demonstrated that mitochondrial dysfunction is already present in the parahippocampal region of early-stage AD patients [276]. Mitochondrial dysfunction has also been previously observed in the rest of the CNS disorders included in our analysis. Several genes linked to familial PD, such as *SNCZ*, *LRRK2*, *Parkin*, *PINK1*, and *ATP13A2*, have been associated with mitochondrial function. For instance, mutated α-Syn induces mitochondrial fragmentation, and Parkin has diverse functions in maintaining healthy mitochondria by regulating their biogenesis. Mutations in *PINK1* , which encodes for a mitochondrial serine/threonine kinase involved in the maintenance of mitochondria homeostasis, constitute the second most common cause of autosomal recessive early-onset PD [263]. Altered mitochondrial morphology is a common feature of HD, and this diverse morphology can be observed in tissues beyond the central nervous system in HD patients [277]. In addition, the altered mitochondrial structure has been shown to correlate with decreased electron transport chain activity, oxygen consumption, and calcium buffering [277]. Several lines of evidence have also pointed towards the role of mitochondrial dysfunction in bipolar disorder, schizophrenia, and major depression, which include functional assays, magnetic resonance spectroscopy studies, gene and protein expression studies, the observation of structural abnormalities, and the fact that *bona fide* mitochondrial disease patients often present psychotic symptoms that are often misdiagnosed as cases of BD and SCZ [278, 279]. Mitochondrial dysfunction is also an important feature in ASD, as suggested by several lines of evidence, including the presence of elevated levels of lactate in the plasma of ASD patients, the decreased brain glucose utilization and ATP levels, as well as evidence

of abnormal electron transport chain function in peripheral tissues and brain tissues [280]. Higher rates of glycolysis and suppression of mitochondrial function, even in the presence of oxygen, are traits commonly observed in cancer cells. This phenomenon, known as the Warburg effect, has been found to be extended among cancers. Since the ATP yield obtained by glycolysis is very low compared to respiration, multiple hypotheses have been generated in order to explain its preponderance, including the presence of dysfunctional mitochondria in cancer cells, the rapid pace of the glycolytic pathway, which would allow meeting the fast rates of cell proliferation found in cancer, the presence of reduced mitochondrial activities in cancer cells due to hypoxia and the generation of reactive oxygen species, or the increased demand of NAD+. Some of these hypotheses are under discussion or have been refuted. In addition, It has been observed that the Warburg effect is a common characteristic of all proliferating cells, both tumoral and normal. In summary, advances in the understanding of cancer metabolism depict alterations in the oxidative phosphorylation and the metabolism of glucose as a more complex phenomenon than previously thought. In fact, our data suggest that at least seven tumor types (BLCA, DLBCL, FLYMPH, LGCA, LIVCA, and OVCA) present upregulation of genes linked to the oxidative phosphorylation and the electron transport chain, whereas five of them (BRNCA, CRCA, KDNCA, STCA, and THCA) present downregulation in genes linked to these processes, highlighting the heterogeneity present in cancer energy metabolism. In summary, given the joint involvement of the alterations of oxidative phosphorylation in both CNS disorders and cancer, this pathway represents a good candidate for modulating the diverse comorbidity patterns observed between both sets of disorders.

p53 signaling, DNA damage, and apoptosis

The TP53 gene encodes for the P53 transcription factor, an extensively characterized tumor suppressor called “the guardian of the genome” that presents missense mutations in many human malignancies. The loss of function of p53 is linked to many cancer-related processes, including the cell cycle activation, the survival and avoidance of cell death, as well as the promotion of angiogenesis, anchorage-independent growth, cell migration and invasion capacities, and genomic instability [281]. The levels and activity of p53 have been previously found to be increased in neurodegenerative disorders. Accordingly, we observed that the p53 pathway was upregulated in AD, PD, and HD as well as in ASD and SCZ. AD and PD patients and mouse models present increased levels of p53 and apoptosis [281]. Moreover, increased levels of p53 have also been detected in patients and animal models of HD [282]. The increased p53 activity observed in CNS disorders could potentially be linked to a reduction in cancer risk. p53 can induce apoptosis by several mechanisms, the most frequent is by transcriptional induction of the intrinsic apoptotic pathway's genes, including the proteins of the Bcl2 family and caspases. In addition, proteins commonly associated with the physiopathology of CNS disorders have been shown to play roles in the regulation of p53.

Proper physiological APP processing has been found to be linked to p53 activity. Wild type α-synuclein has also been found to protect cells from pro-apoptotic stimuli by reducing the transcriptional activity of p53. Besides, mutations in Parkin produce an increase in p53 mRNA levels, and its transcriptional activity and DJ-1 weakens the DNA binding affinity of p53. Overexpression of mutant huntingtin also increases p53 levels. Also, Aβ42 is transported to the cell under oxidative stress conditions and activates BACE1 and APP, intensifying erroneous APP cleavage [282]. Apoptotic pathways were found slightly upregulated in our data in the case of AD, in which the Reactome apoptotic execution phase pathway was found to be upregulated and in PD. Apoptosis was also found to be upregulated in HD hallmarks analysis. Upregulation of the cell death pathway was observed in AD. Genes belonging to this pathway, including *NFKB1*, *CASP6*, and *BCL2*, *RELA*, *LMNA*, *APAF1*, and *CASP8*, were found to be upregulated in AD samples compared to controls. Pathways associated with cell death were not found to be upregulated in HD. However, several caspase genes, such as *CASP3*, *CASP6*, *CASP7*, *CASP9*, *FAS*, were found to be significantly upregulated in HD brain tissues compared to controls. The death pathway was also found to be upregulated in PD brains. Caspases 1, 4, 6, and 3 were found to be upregulated in PD brains, as well as members of the apoptosome such as APAF1. Apoptosis is a known mechanism taking place in neurodegenerative disorders. For instance, in AD, neural apoptosis has been found to be induced by the Aβ [283]. Apoptosis is also the main mechanism of neural loss in Parkinson's disease and HD [284]. In contrast, one of the hallmarks of cancers is the evasion of cell death mechanisms. Therefore, if trends towards pro-apoptotic mechanisms are a widespread phenomenon in neurodegenerative disorders not limited to the central nervous system, it could constitute a potential additional barrier to tumor development in those patients.

Proteasome and autophagocytosis

One important hallmark of neurodegenerative disorders is the impairment of systems linked to the maintenance of proteín homeostasis, including the ubiquitin-proteasome system (UPS) and the antiphagocytic mechanisms. Our data showed heavy downregulation of proteasome and autophagy-related pathways in the transcriptomes of AD and PD but not in HD. Impairments in the proteasome's function were also observed in ASD patients, as well as in SCZ, MD, and BD. Neurodegenerative disorders are characterized by deficits in the function of these systems and the accumulation of their idiosyncratic proteopaties. This is very clear in the case of PD and AD, which have been previously linked to the downregulation of the UPS, the impairment of autophagy, and the aggregation of α–synuclein and amyloid–β, respectively. In contrast, tumor cells have often been found to upregulate the USP and heat-shock protein systems since increased protein synthesis is required for the high rates of cell division observed in cancer cells [285]. Our analyses showed upregulation of unfolded protein response pathways in most of the studied cancer types in

concordance with this view. Therefore, the impaired proteostasis mechanism found in neurodegenerative disorder patients could constitute an impediment to tumor development.

The role of the immune system

Increasing evidence has linked inflammation to neurodegenerative disorders. Inflammation in neurodegenerative disorders may be due to alterations in the functions of endogenous or exogenous cells. Endogenous CNS cells driving inflammation include astrocytes and mononuclear phagocytes (microglia and perivascular macrophages). Astrocytes provide trophic support for neurons and facilitate synapse formation and synaptic pruning by phagocytosis. In addition, they also participate in the modulation of ion and neurotransmitter concentrations and in the blood-brain barrier maintenance. Inflamed astrocytes can produce proinflammatory cytokines and chemokines. Microglia are the resident macrophages of the central nervous system. Activated microglia has been found to be present in almost all neurodegenerative disorders. In addition, astrocyte activation and peripheral monocytes and lymphocytes can be detected in CNS disorders in certain cases [286]. It has been suggested that all neurodegenerative disorders present an inflammatory phenotype. Pro-inflammatory cytokines have been found to be upregulated in the brains of AD patients. Amyloid-β aggregates activate microglia, which transitions from its ramified form to its activated form and produces reactive oxygen species, IL-1β, and TNF-α or to its pro-phagocytic form. Activated pro-phagocytic microglia helps to clear amyloid, whereas neurotoxic microglia secrete autocrine factors that reinforce the inflammatory phenotype. Factors secreted by microglia stimulate the astrocyte response, which could stimulate microglia back and drive the recruitment of neuroinflammatory modulators or act on neurons producing neurotoxic or neurotrophic effects [286]. Besides IL-1β other cytokines and chemokines have been detected in samples derived from AD patients, including (TNF-α), IL-6, IL-8, IL-12, IL-23, CCL2, CCL3, CCL4, CXCL2, among others. The presence of activated microglia is also a characteristic feature of PD brains. Besides, susceptibility genes linked to the immune function have been associated with AD and PD in GWAS studies [286]. HD has also been linked to microglia activation, deposition of complement factors, the release of proinflammatory cytokines, and impaired macrophage migration [287]. Immune system alterations have been found to have a dual role in ASD. First, prenatal insults, such as maternal immunological activation due to infection, have been found to increase the risk of ASD. Second, in the postnatal environment, ASD pathology is characterized by immune dysregulation, inflammation, and the presence of endogenous antibodies [288]. Neuropsychiatric disorders, including BD, MD, and SCZ, have also presented significant immune function alterations [289]. Finally, cancer's capacity to evade the host's immune system is one of the hallmarks of cancer and the cornerstone of immunotherapy [290]. The presented evidence added to our transcriptomic meta-analyses results suggests that immune alterations are processes

jointly altered in CNS disorders and cancer. However, the complexities of the immune function alterations in both sets of disorders prevent us from hypothesizing further the potential influence that this joint dysregulation could have in the comorbidities between CNS disorders and cancers.

Concluding remarks

The biological processes described in the previous sections represent instances of functions jointly altered in CNS disorders and cancers. In addition to them, our analyses also highlighted the presence of joint alterations of genes linked to other mechanisms such as the extracellular matrix or the epidermal to mesenchymal transition, which won't be covered in detail. Besides, it is worth noticing that the reported pathways' alterations do not operate in isolation and often have a complex interplay. For instance, PI3K signaling has been implicated in the regulation of inflammatory response [291] and in the regulation of autophagocityosis [292], and p53 has been found to regulated proteins involved in the mitochondrial function and respiration [293].

Chapter 4: Genetic associations between CNS disorders and cancer

4.1 Introduction

In the previous chapters, we studied CNS and cancer associations at two different levels. In chapter two, we assessed the population-based patterns of comorbidity through meta-analyses of observational studies. In chapter three, their phenotypic similarities were appraised by comparing the outputs of disease-specific differential gene expression profiles and by the identification of gene co-expression modules jointly linked to each disorder. These analyses allowed us to observe diverse patterns of comorbid associations between CNS disorders and cancers and identify transcriptomic changes of genes and pathways involved in the process.

Alterations in joint sets of genes or genes linked to disease-associated variants (henceforth variants-genes) could contribute to the direct and inverse comorbidity patterns observed at a population level in chapter two and also could be involved in the emergence of the shared (direct and inverse) transcriptomic alterations described in chapter three.

Therefore, this chapter is devoted to the identification of shared genetic factors between disease pairs. Two complementary approaches were used for this purpose. The first approach relies on the measurement of the interactome-based overlap of lists of disease-associated genes. It has been shown that disease proteins (the products of disease genes) are not randomly scattered in the interactome [32]. Instead, they tend to interact with each other more than expected by chance forming connected subgraphs that are known as disease modules. According to this view, a disease would represent a local perturbation of the underlying disease module, and comorbidity would arise as a consequence of perturbations in overlapping disease modules. In other words, if two disease modules overlap, local perturbations leading to one disease will likely disrupt pathways involved in the other disease module, as well, resulting in shared clinical characteristics.

This method has been previously applied to uncover disease-disease associations [32]. In their work, Meanche and co-workers compiled 141296 interactions between 13460 proteins, which included protein-protein and regulatory interactions, metabolic pathways, and kinase-substrate interactions. They primarily focused on identifying the conditions by which disease modules could be detected in the incomplete human interactome and established that at least 20 disease-associated genes would be necessary for disease-module detection in their interactome. Their work focused on the detection of direct comorbid associations between disorders. However, if inverse comorbid associations could also arise due to the involvement of sets of genes, which overlap in the interactome still is an open question. As we have seen, CNS and cancer comorbidities provide interesting examples of both direct and inverse associations that we can study using interactome-based approaches.

To this aim, we constructed a human interactome using multiple sources of protein-protein and gene-gene associations. We gathered data from protein-protein interactions from a state-of-the-art human interaction dataset (HuRI), co-expression partners from GTEx, Transcription factor targets from Transfact, human protein complexes from CORUM, Kinase-substrate interactions from phosphosite plus, and metabolic association from KEGG.

Disease-associated genes and variants-genes were obtained for each disorder from different databases (eDgar, DisGeNet, and PheGenI) and tested for overlap in the interactome. Our analyses were repeated in two previously compiled human interactomes derived from STRING and BioGRID to test the results' stability.

The second approach was based on the computation of genetic correlations. In particular, a recent methodological development called cross-trait LD score regression was used to this aim [294]. There is evidence that most complex traits, including complex disorders, are influenced by hundreds of genetic loci with small effects. In recent years, high-throughput technologies have been applied to investigate complex disorders' underlying genetic structure through GWAS studies. The intuition behind LD score regression is that if a trait is genetically influenced, then variants with higher LD-scores (i.e., variants that tag more of the genome) should have a greater opportunity to tag causal variants and therefore to have higher tests statistics on average than variants that have low LD-scores. Cross-trait LD score regression uses GWAS summary statistics to compute genetic correlations between a given pair of traits. It is based on the fact that the product of the z-statistics involving each variant for a given pair of disorders is proportional to their linkage disequilibrium scores.

Access to patient-level data from GWAS studies is often restricted due to anonymity issues making impractical the computation of genetic associations between disorders. In contrast, GWAS summary statistics can be increasingly found in public repositories such as the GWAS catalog. In addition, cross-trait LD score regression is not biased when sample overlap is present in a given pair of studies, and results derived from previous meta-analyses of GWAS data can be used as an input.

This methodology has been previously applied to investigate the correlation between AD and a set of cancer types [38], which revealed a positive genetic correlation between AD and breast cancer. To generate a more complete picture of the genetic associations between CNS disorders and cancers, we queried public repositories for GWAS summary statistics involving the seven CNS disorders included in previous chapters and the 22 tumor types under consideration. With the exception of Huntington's disease (which is a monogenic or mendelian disorder), we were able to obtain GWAS summary statistics for all the CNS disorders included in the previous chapter. GWAS summary statistics were also retrieved for a subset of the studied site-specific cancers, including breast, prostate, colorectal, ovarian, lung, and digestive cancers, as well as skin cancer melanoma.

Cross-trait LD score regression was carried out for all disease pairs. Only European ancestry studies were included in the analysis.

4.2 Material and methods

4.2.1 Sources of human protein-protein/gene-gene interaction data.

4.2.1.1 Interactome 1:

We constructed a human interactome by integrating different sources of protein-protein and gene-gene interactions, including protein-protein interactions (PPI), gene co-expression data, information regarding protein complexes, data from transcription factors and their targets, and associations between genes participating in successive steps of metabolic pathways.

- Binary PPIs: PPIs data was derived from the Human Reference Interactome (HuRI) [295]. HuRI is an initiative of the Center for Cancer System Biology at Dana-Farber Cancer Institute, which interrogated all pairwise combinations of human protein-coding genes to identify which are involved in binary protein-protein interactions [296] using yeast two-hybrid (Y2H) screenings.
- Co-expressed genes identification: Co-expression partners were identified using data derived from the Genotype-Tissue Expression (GTEx) [297], which contains gene expression datasets derived from multiple healthy human tissues. Twenty-six of them (adipose tissue, adrenal gland, blood, blood vessel, brain, breast, colon, esophagus, heart, liver, lung, muscle, nerve, ovary, pancreas, pituitary, prostate, salivary gland, skin, small intestine, spleen, stomach, testis, thyroid, uterus, and vagina) were represented by more than one hundred samples. Expression data (Transcripts Per Million Reads, TPM) and annotation files were downloaded for those tissues. Then we randomly selected 100 samples from each tissue and computed Pearson's correlations between all pairwise gene combinations'. We considered that two genes were co-expression partners if they presented values of correlation higher than 0.7 in at least 16 tissue types.
- Human protein complexes: Data from human protein complexes was downloaded from The comprehensive resource of mammalian protein complexes (CORUM) [298]. Human complexes were selected, and self-interacting genes were removed from the data.
- Transcription factor targets data: The transcription factor targets dataset (TRANSFAC) was selected [299, 300] and retrieved from the Harmonizome database [301].
- Kinase-substrate interactions: Kinase substrate interactions were retrieved from PhosphoSitePlus [302], which gathers curated data from low and high-throughput phosphoproteomic studies.
- Metabolic links: Data regarding metabolic links was obtained from The Kyoto Encyclopedia of Genes and Genomes (KEGG)[303]. In particular, the KEGG PATHWAYS database that includes information regarding genes and their associated pathways and cellular processes

was employed. In this database, pathways are represented using graphs where nodes are molecules (protein and compounds), and edges represent nodes' relations. First, the list of KEGG's human metabolic pathways was retrieved, and KGML files, including them, were obtained using the retrieveKGML function from KEGGgraph package [304]. Then the files were parsed using parseKGML2Graph function included in the same package. Graphs were transformed to adjacency matrices, from which and all pairwise gene-gene associations were extracted.

4.2.1.2 Interactome 2:

The second interactome was obtained from STRING [305]. STRING gathers information about known and predicted protein-protein interactions, including direct or physical interactions and indirect or functional associations. In short, we downloaded the subset of human interactions subset of STRING. The database provides a score ranging from 0 to 1 that informs about the strength of each association's evidence. Only the high-quality associations (those presenting a quality score higher than 0.7) were selected.

4.2.1.3 Interactome 3:

The third interactome was derived from BioGriD [306]. BioGRID includes interaction data compiled through comprehensive curation efforts. We selected all human interactions included in the BioGRID database.

4.2.2 Sources of disease-associated genes and variants

Sets of genes and variant-genes linked to each studied disorder were extracted from the following repositories. Disease-associated genes were retrieved from DisGeNet [307]. This database gathers information from different sources, including curated data from UniProt, the Comparative Toxicogenomics Database (CTD), Orphanet, the Clinical Genome Resource (CLINGEN), Genomics England, The Cancer Genome Interpreter (CGI), and the Psychiatric disorders Gene association Network (PsyGeNET), as well as, sources of non-curated data. For each disease-gene association, DisGeNet provides a quality score that informs about the evidence supporting it. We excluded not-curated, literature-derived, and animal model-obtained association. Variants associated with each disease were retrieved from The Phenotype Genotype Integrator (PheGenI) [308], which contains information derived from the GWAS catalog, dbGaP, and dbSNP, among others. Genes mapped to each variant (henceforth variant genes) were selected and used for subsequent analysis. Finally, the eDGAR database [309], which collects information from OMIM, Humsvar, and ClinVar was also queried. Each included database employs its own dictionaries to identify phenotypes. **Appendix 7** shows the identifiers employed to select disease-associated genes and variant genes and the list of disease-associated variants and variant-genes for each disorder

under the stringent setting. **Supplementary Appendix 1 Table 38** provides the list of disease-associated genes and gene-variants obtained in the relaxed selection setting.

For each disorder, two lists of disease-associated genes (stringent and relaxed) were created based on the use of two different thresholds for both the DisGeNET association quality score and the p-value associated with the variants linked to disease. The stringent sets were selected using a threshold of 0.6 for DisGeNET quality score and p-values lower than 1e-10 for the disease variant associations reported in PheGenI. In the case of the relaxed sets, the thresholds were 0.3 and 1e-08. The threshold selection was found to be a large but uneven impact on the number of included disease-associated genes.

4.2.3 Measures of network localization and statistical significance

For each set of disease-related genes and gene-variants, we computed two network localization measures to determine the degree to which disease proteins aggregate in specific areas of the interactome.

The first measure, termed as the intra-disease average distance $\langle d_{AA} \rangle$, was computed as follows. Let *N* be a set containing the disease-associated genes and variant genes for a particular disorder. For each element *n* in *N* the interactome distance to the next closest element of the set is determined. Then, the average of all shortest distances computed for all the elements of *N* is computed. Note that the shortest distance is defined as the shortest path between a given pair of nodes (i.e., the lowest number of edges that separate a pair of specific nodes in the network).

To assess the computed intra-disease average distances' significance, we randomly selected 10000 sets of genes of the same size as the original set. The degree distribution of the genes placed at a given disease-associated gene list impacts the computed distances' values. Therefore, we should account for the degrees observed in the original gene set's genes for the randomization process to be accurate.

Given the scale-free topology of human interactomes, low degree nodes are much more abundant than high degree nodes. We grouped nodes within a certain degree interval together to avoid choosing the same high degree nodes repeatedly in the degree-preserving random sampling analysis. Let V be the set of nodes in the interactome and u a particular element of V such that $u \in V$ and let k_u be the set of all nodes with degree u and $B_{i,j}$ the bin containing the nodes with degree equal or higher than i and lower than j. Nodes, with an increasing degree value are included in each bin such that it has at least 100 elements. Therefore, each bin will contain nodes such that, $B_{i,j} = \{u \in V \mid i \leq k_u < j\}$ such that $\left|\left|B_{i,j}\right|\right| \geq 100$. Once the bins have been created, the random sampling strategy proceeds as follows. The degree in the reference interactome of each gene in the original gene set is determined. Then, for each gene, a random gene is selected from the degree-

matched bin. This is repeated for all genes in the studied set. The result is a randomly selected set of genes presenting the same size than the original set and an approximately equivalent degree distribution.

The original intra-disease average distance is then compared to the resulting random distribution of intra-disease average distances. d^{rand} as follows:

$$z_{score}(d) \equiv \frac{d_{AA} - \langle d^{rand} \rangle}{\sigma(d^{tand})}$$

Finally, one-tailed p-values are computed by comparing the obtained z-scores to the standard normal distribution.

The second measure, termed module size *S* represents the largest connected component produced by a particular set of disease-associated proteins in the human interactome (i.e., the highest number of disease proteins directly connected to one another). A random distribution of module sizes S^{rand} was then generated using the same random degree-preserving strategy described for the intra-disease average measures. Z-scores and p-values were also computed using the same methodology.

4.2.4 Network-based separation measures between disease pairs.

The following equation shows the network-based overlap measure proposed by Menche and co-workers [32].

$$S_{AB} = \langle d_{AB} \rangle - \frac{\langle d_{AA} \rangle + \langle d_{BB} \rangle}{2}$$

S_{AB} compares the shortest distances between proteins within each disease $\langle d_{AA} \rangle$ and $\langle d_{BB} \rangle$, to the shortest distances $\langle d_{AB} \rangle$ between A-B protein pairs. Proteins associated to both A and B have d_{AB} = 0. Positive values of S_{AB} indicated separation between disease modules, whereas negative values suggest disease-disease overlaps. A null distribution of S_{AB} values ${S_{AB}}^{rand}$was used to compute the significance of the observed overlaps following the same degree preserving strategy described in the previous section. Z-scores and p-values were computed likewise.

4.2.5 Cross-trait LD-score regression

GWAS summary statistics for the studied disorders were searched in public repositories (GWAS catalog) or directly requested to the authors. We were able to retrieve GWAS summary statistics for all the CNS disorders except for HD. Since HD is a monogenic disease, no genome-wide association studies have been carried out for it. GWAS summary statistics for a subset of the studied cancers were also obtained, including prostate cancer, breast cancer, cervical cancer, colorectal cancer, ovarian cancer, endometrial cancer, thyroid cancer, skin cancer melanoma, and lung cancer.

In order to generate the appropriate files to compute genetic correlations, GWAS summary statistic files were preprocessed using custom scripts in R and the munge_sumstats.py function from the ldsc python package. Note that ldsc interprets the A1 field as the effect allele. In other words, it assumes that the sign of the reported summary statistics (betta values, odds ratios, or log-odds ratios) are oriented in reference to the A1 allele.

Heritability estimates were computed for all disorders. Then disease-disease correlations were computed for those disease pairs that showed individually significant heritabilities.

Cross-trait LD score is a method for estimating genetic correlations from GWAS summary statistics data. It is based on the idea that GWAS effect-sizes estimates for a given SNP partially incorporate the effects of all SNPs in linkage disequilibrium (LD) with that SNP. For a polygenic trait, SNPs with high LD will have higher X^2 statistics on average than SNPs with low LD. A similar relationship holds if we replace the X^2 statistics of a single study with the product of z-scores from two studies of traits with non-zero genetic correlation.

Under a polygenic model, the expected value of $z_{1j}z_{2j}$ is defined by the following equation:

$$\mathrm{E}[z_{1j}z_{2j}] = \frac{\sqrt{N_1N_2}\rho_g}{M} l_j + \frac{\rho N_s}{\sqrt{N_1N_2}}$$

Were N_1 is the sample size of study one, $N_2$is the sample size of study two, ρ_g is the genetic covariance. l_j is the LD score of a particular SNP, N_s is the number of included in both studies, and ρ is the phenotypic correlation among the N_s overlapping samples. This setting allows estimating genetic covariance between two traits by identifying the slope from the regression of $z_{1j}z_{2j}$ on LD score. Sample overlap only affects the intercept of the regression because sample overlap creates spurious correlations between z_{1j} and z_{2j} but does not depend on LD score.

Then normalizing the genetic covariance by the SNP-heritability yields genetic correlation.

$$r_g := \frac{\rho_g}{\sqrt{h_1^2 h_2^2}}$$

Where $h_1^2$the SNP-heritability of is study one and h_2^2 is the study heritability from study two. Genetic correlations range between -1 and 1.

4.2.6 Sources of GWAS summary statistics

GWAS summary statistics were obtained from different sources, including the GWAS summary statistics section of the GWAS catalog [310] and the and UK biobank [311]. GWAS summary statistics for PD and SKCM were directly requested to the 23andME and Genomel consortia. Only studies that tested more than 450000 variants and included at least 5000 individuals between cases and controls derived from European ancestry populations were selected. Precomputed LD scores of individuals with European ancestry derived from the 1000 Genomes project were used in the analyses.

4.3 Results

4.3.1 Characteristics of the interactomes and the disease-associated gene lists used in this chapter

We computed measures of network localization ($\langle d_{AA} \rangle$ and S) and disease separation (S_{AB}) using the human interactome constructed by us (hereafter I1) and two additional interactomes, derived from STRING (I2) and BIOGRID (I3). Two different sets of disease-associated genes were used for each disorder. They were termed as the relaxed (R) and stringent (S) set and differed in gene selection thresholds applied when retrieving disease-associated genes from DisGeNet and PheGenId.

Table 18 shows the sources of protein-protein gene-gene association included in I1. I1 covered 324377 unique interactions between 16954 genes or proteins. I1, I2, and I3 had increasing numbers of interactions and edge densities. They ranged from the 324377 interactions and 0.22% edge density found in I1 to the 800238 interactions and the 0.59% edge density observed in I3. **Table 19** shows information regarding the basic parameters of I1, I2, and I3. The number of disease-associated genes included in I1 ranged from 2 in CML to 126 BRCA. Different disease identifiers were used to retrieve disease-associated genes in the different consulted databases (DisGeNet, PhenGenI, and Edgar). The identifiers used to extract the disease-associated genes and variant-genes in the stringent setting can be consulted in **Appendix 7**.

Interaction source	Type of interaction	Nº of interactions	Nº of unique genes
HURI	Binary protein-protein interactions	52248	8230
GTEx	Co-expression in multiple tissues	125803	6702
CORUM	Human protein complexes	39110	3432
TRANSFACT	Transcription factors targets	100560	13238
PhosphoSitePlus	Kinase-substrate interactions	6013	2527
KEGG	Metabolic interactions	15180	1400
Total	All interactions	324377	16954

Table 18: Sources of protein-protein/gene-gene interaction data included in I1.

Interactome	Nº Genes	Nº Edges	Av. degree	Diameter	Av. path length	Edge density
I1	16954	324377	38.26	9	2.93	0.22%
I2	18870	445420	47.21	8	2.91	0.25
I3	16381	800238	97.70	13	3.77	0.59%

Table 19 Parameters of the interactomes.

4.3.2 Network localization measures of the genes associated with each disorder.

Network localization measures of the disease-associated gene lists showed mixed results, which depended upon the employed interactome and the type of disease-associated gene-set used in the analysis (stringent or relaxed). Average intra-disease distances were found to be lower than

expected by chance in 13, 13, 26, 22, 19, and 28 of the included disorders in the I1S, I2S, I3S, I1R, I2R, and I3R settings, respectively. Similar results were observed in the case of the size of the largest connected components, which were found to be larger than expected by chance in 14, 13, 26, 22, 22, and 28 of the included disorders in the I1S, I2S, I3S, I1R, I2R, and I3R settings respectively. These results indicate that, in general, genes associated with a particular disorder tend to be more interconnected and to be placed in the same neighborhood of the human interactome than expected by chance. The number of disorders yielding significant values of their network localization measures increased as the employed interactomes' size and interconnectedness increased. The use of larger sets of disease-associated genes (relaxed sets) allowed the identification of a higher number of significant network localization measures. This suggests that we would be working under the detection level thresholds in certain settings given the incomplete nature of or interactomes and disease-associated gene sets. **Tables 20 and 21** show the average intra-disease distance and the largest connected component analysis results under the six tested settings. The genes that conform each disorder's largest connected component under all settings can be consulted in **Supplementary Appendix 1 Tables 39 and 40** include the genes placed in the largest connected component of each disorder in all tested settings.

	Interactome 1 stringent		Interactome 2 stringent		Interactome 3 stringent		Interactome 1 relaxed		Interactome 2 relaxed		Interactome 3 relaxed	
Disease	**Nº of genes**	**d_{AA} (p-val)**	**Nº of genes**	**d_{AA} (p-val)**	**Nº of genes**	**d_{AA} (p-val)**	**Nº of genes**	**d_{AA} (p-val)**	**Nº of genes**	**d_{AA} (p-val)**	**Nº of genes**	**d_{AA} (p-val)**
AD	55	1.8 (2.35e-07)	57	1.54 (1.02e-04)	55	1.35 (3.86e-09)	138	1.7 (4.50e-04)	142	1.43 (3.88e-06)	137	1.3 (1.31e-15)
ASD	19	2.05 (2.42e-01)	19	1.68 (1.15e-02)	17	1.59 (7.86e-07)	93	1.81 (3.51e-01)	98	1.68 (6.48e-04)	96	1.54 (7.34e-11)
BD	22	1.82 (9.09e-02)	24	2.12 (9.49e-01)	23	1.48 (1.11e-05)	454	1.45 (9.27e-03)	465	1.29 (6.90e-11)	465	1.18 (8.75e-12)
HD	NA	NA	NA	NA	NA	NA	15	2.13 (2.57e-01)	17	2.06 (3.92e-01)	17	1.88 (3.52e-03)
MD	23	1.83 (8.51e-02)	26	1.96 (5.91e-01)	24	1.5 (1.95e-05)	432	1.38 (4.65e-03)	436	1.24 (1.74e-11)	440	1.14 (2.82e-11)
PD	69	1.7 (1.24e-05)	77	1.69 (1.05e-01)	67	1.34 (4.46e-14)	138	1.72 (4.70e-03)	148	1.57 (4.44e-02)	137	1.24 (5.90e-13)
SCZ	38	2 (5.05e-01)	43	1.74 (3.67e-03)	41	1.44 (7.77e-10)	888	1.31 (5.28e-03)	923	1.23 (6.04e-05)	890	1.15 (5.05e-12)
ALL	16	1.56 (4.94e-03)	17	1.71 (5.86e-02)	18	1.39 (8.90e-11)	71	1.39 (1.03e-03)	74	1.31 (5.42e-05)	73	1.3 (4.09e-10)
AML	34	1.47 (2.17e-02)	34	1.24 (3.27e-03)	31	1.13 (1.17e-09)	167	1.3 (7.39e-03)	172	1.22 (2.20e-02)	163	1.23 (5.65e-11)
BLCA	22	2.05 (9.08e-01)	24	1.42 (3.85e-02)	23	1.48 (1.53e-03)	134	1.54 (2.48e-01)	142	1.3 (5.94e-03)	136	1.31 (1.04e-05)
BRCA	126	1.47 (1.47e-02)	135	1.28 (4.91e-03)	125	1.34 (8.60e-05)	1044	1.15 (1.57e-03)	1094	1.16 (6.70e-02)	1029	1.24 (2.01e-03)
BRNCA	17	1.82 (1.27e-01)	19	1.26 (6.97e-03)	17	1.41 (4.02e-04)	52	1.69 (1.75e-02)	57	1.28 (4.57e-02)	53	1.3 (7.93e-06)
CERV	3	1.67 (2.08e-02)	4	2.5 (7.48e-01)	4	1.75 (3.33e-02)	5	1.8 (2.42e-02)	7	2.71 (9.99e-01)	7	1.57 (2.86e-05)
CHLCA	3	2 (2.71e-01)	3	2 (8.45e-01)	3	1 (1.39e-02)	26	1.65 (1.16e-01)	26	1.31 (2.11e-02)	26	1.15 (9.74e-06)
CLL	32	1.56 (3.60e-05)	33	1.73 (2.93e-02)	29	1.79 (1.22e-03)	85	1.45 (4.56e-05)	90	1.4 (1.45e-03)	82	1.48 (5.16e-04)
CML	2	3 (5.47e-01)	2	2 (8.94e-02)	NA	NA	2	3 (5.49e-01)	2	2 (8.72e-02)	NA	NA
CRCA	67	1.54 (2.08e-02)	68	1.18 (1.21e-04)	68	1.26 (1.10e-03)	672	1.27 (3.48e-04)	700	1.25 (6.90e-03)	680	1.36 (2.22e-05)
DLBCL	7	2.43 (5.42e-01)	8	1.75 (1.16e-01)	7	1.29 (1.86e-04)	57	1.44 (8.87e-04)	59	1.24 (1.68e-03)	55	1.22 (5.61e-06)
FLYMPH	7	2.29 (2.11e-01)	7	2.29 (7.97e-01)	7	1.86 (8.09e-02)	26	2 (8.78e-02)	28	1.82 (3.27e-01)	23	1.3 (2.09e-07)
HANC	7	1.86 (5.21e-02)	7	1.57 (1.05e-01)	7	1.71 (1.95e-02)	51	1.39 (3.82e-04)	52	1.25 (3.28e-03)	51	1.24 (3.11e-05)
KDNCA	17	1.71 (1.17e-01)	19	1.47 (1.46e-03)	17	1.41 (1.57e-05)	135	1.59 (1.33e-02)	146	1.4 (2.87e-02)	141	1.47 (4.63e-04)
LGCA	33	1.64 (8.21e-02)	37	1.38 (8.68e-02)	34	1.35 (1.27e-04)	245	1.36 (1.91e-03)	261	1.27 (3.16e-02)	241	1.24 (3.40e-07)
LIVCA	35	1.46 (6.05e-04)	35	1.34 (1.56e-02)	35	1.29 (1.93e-03)	469	1.29 (5.52e-02)	486	1.23 (2.38e-01)	480	1.21 (9.31e-07)
OVCA	34	1.56 (3.51e-03)	35	1.17 (2.51e-03)	31	1.23 (2.75e-04)	144	1.38 (5.05e-03)	147	1.19 (3.62e-04)	141	1.3 (7.51e-04)
PACA	19	1.63 (8.37e-03)	20	1.55 (1.16e-01)	19	1.26 (6.79e-09)	103	1.54 (4.57e-03)	110	1.32 (6.77e-03)	104	1.18 (8.12e-11)
PRCA	75	1.83 (4.89e-02)	79	1.53 (4.77e-02)	73	1.73 (2.41e-05)	654	1.26 (3.45e-02)	679	1.15 (1.57e-03)	649	1.25 (5.69e-07)

SKCM	32	1.44 (1.07e-04)	34	1.38 (7.64e-02)	32	1.19 (6.06e-07)	234	1.47 (1.74e-01)	249	1.33 (1.95e-01)	240	1.4 (1.62e-02)
STCA	19	1.79 (1.92e-01)	20	1.6 (1.79e-01)	20	1.45 (1.04e-03)	289	1.33 (3.39e-03)	299	1.16 (5.51e-03)	287	1.18 (5.23e-07)
THCA	13	1.69 (7.99e-02)	14	1.86 (7.17e-01)	11	1.27 (4.23e-07)	53	1.7 (1.74e-01)	55	1.42 (1.13e-01)	52	1.13 (5.47e-09)

Table 20: Interactome average intra-disease distance localization measure under the six settings tested in the analysis. The first column of each group of two columns shows the number of genes associated with a particular disorder represented in the interactome under analysis. The second column provides the intra-disease distance values for each list of disease-associated genes and gene variants and their associated p-values.

	Interactome 1 stringent		Interactome 2 stringent		Interactome 3 stringent		Interactome 1 relaxed		Interactome 2 relaxed		Interactome 3 relaxed	
Disease	**Nº of genes**	**S (p-val)**	**Nº of genes**	**S (p-val)**	**Nº of genes**	**S (p-val)**	**Nº of genes**	**S (p-val)**	**Nº of genes**	**S (p-val)**	**Nº of genes**	**S (p-val)**
AD	55	6 (3.27e-11)	57	28 (7.50e-09)	55	37 (0.00e+00)	138	15 (1.71e-01)	142	75 (6.06e-05)	137	94 (6.36e-13)
ASD	19	1 (7.12e-01)	19	4 (8.85e-02)	17	6 (1.98e-12)	93	10 (4.06e-01)	98	17 (2.06e-02)	96	29 (0.00e+00)
BD	22	5 (1.41e-02)	24	1 (8.56e-01)	23	12 (0.00e+00)	454	228 (1.37e-02)	465	316 (3.06e-08)	465	358 (9.76e-09)
HD	NA	NA	NA	NA	NA	NA	15	1 (6.10e-01)	17	2 (1.23e-01)	17	3 (4.37e-03)
MD	23	2 (5.60e-01)	26	3 (3.47e-01)	24	15 (0.00e+00)	432	260 (5.33e-03)	436	317 (5.05e-09)	440	376 (2.01e-11)
PD	69	5 (2.58e-01)	77	24 (1.31e-02)	67	40 (0.00e+00)	138	12 (1.10e-01)	148	59 (2.77e-02)	137	101 (1.24e-09)
SCZ	38	2 (7.35e-01)	43	5 (1.31e-01)	41	28 (0.00e+00)	888	594 (1.20e-02)	923	684 (3.08e-03)	890	759 (3.11e-14)
ALL	16	7 (3.47e-06)	17	3 (8.11e-02)	18	14 (0.00e+00)	71	40 (1.18e-03)	74	51 (5.86e-05)	73	53 (9.54e-13)
AML	34	16 (2.18e-02)	34	21 (5.04e-02)	31	22 (8.31e-13)	167	113 (9.89e-03)	172	133 (1.97e-02)	163	121 (2.56e-05)
BLCA	22	2 (7.63e-01)	24	14 (2.69e-02)	23	14 (3.15e-09)	134	61 (1.07e-01)	142	99 (4.82e-03)	136	96 (5.68e-04)
BRCA	126	61 (2.21e-02)	135	95 (1.04e-02)	125	86 (1.27e-03)	1044	882 (1.53e-03)	1094	922 (2.98e-02)	1029	777 (8.34e-04)
BRNCA	17	4 (4.14e-02)	19	14 (8.51e-03)	17	12 (5.59e-10)	52	17 (1.36e-05)	57	42 (1.89e-02)	53	39 (7.80e-08)
CERV	3	2 (0.00e+00)	4	1 (5.45e-01)	4	2 (4.69e-04)	5	2 (0.00e+00)	7	1 (5.79e-01)	7	2 (1.36e-03)
CHLCA	3	1 (6.03e-01)	3	1 (8.29e-01)	3	3 (6.99e-04)	26	11 (3.46e-03)	26	18 (1.47e-02)	26	22 (6.46e-10)
CLL	32	5 (1.89e-01)	33	11 (2.26e-05)	29	9 (2.92e-07)	85	42 (1.31e-03)	90	53 (1.33e-03)	82	41 (2.77e-04)
CML	2	1 (5.06e-01)	2	1 (5.17e-01)	NA	NA	2	1 (5.04e-01)	2	1 (5.20e-01)	NA	NA
CRCA	67	26 (4.04e-03)	68	56 (1.11e-04)	68	51 (1.47e-03)	672	489 (2.27e-04)	700	518 (7.43e-03)	680	415 (3.31e-04)
DLBCL	7	1 (5.30e-01)	8	2 (2.02e-01)	7	3 (2.35e-04)	57	30 (1.44e-03)	59	41 (1.04e-02)	55	45 (1.18e-07)
FLYMPH	7	1 (5.46e-01)	7	1 (6.87e-01)	7	2 (2.26e-01)	26	3 (2.71e-03)	28	3 (5.97e-01)	23	6 (5.00e-03)
HANC	7	2 (8.07e-04)	7	3 (1.01e-01)	7	2 (8.68e-02)	51	34 (2.80e-06)	52	38 (4.60e-03)	51	40 (4.35e-09)
KDNCA	17	5 (2.90e-02)	19	9 (5.55e-05)	17	5 (1.96e-05)	135	42 (2.49e-02)	146	85 (2.04e-02)	141	77 (1.68e-05)
LGCA	33	12 (2.95e-02)	37	24 (5.48e-02)	34	24 (8.91e-11)	245	146 (6.02e-03)	261	191 (1.44e-02)	241	184 (1.35e-05)
LIVCA	35	15 (6.34e-03)	35	21 (4.41e-02)	35	28 (1.69e-05)	469	329 (4.62e-02)	486	371 (2.28e-01)	480	378 (6.14e-07)
OVCA	34	15 (1.29e-04)	35	29 (3.01e-03)	31	24 (8.87e-07)	144	86 (5.07e-03)	147	117 (8.41e-04)	141	97 (4.94e-03)
PACA	19	4 (2.17e-01)	20	10 (3.98e-02)	19	13 (0.00e+00)	103	42 (2.09e-02)	110	70 (4.10e-02)	104	86 (1.88e-05)
PRCA	75	7 (3.57e-01)	79	38 (1.08e-02)	73	27 (0.00e+00)	654	480 (3.17e-02)	679	574 (2.92e-03)	649	491 (2.71e-09)

SKCM	32	14 (1.84e-04)	34	22 (2.65e-02)	32	24 (3.11e-11)	234	121 (1.02e-01)	249	163 (2.15e-01)	240	142 (9.98e-03)
STCA	19	3 (4.40e-01)	20	8 (9.28e-02)	20	11 (3.06e-11)	289	185 (4.79e-03)	299	248 (6.99e-03)	287	233 (7.93e-07)
THCA	13	4 (8.82e-02)	14	2 (7.93e-01)	11	8 (0.00e+00)	53	7 (6.22e-01)	55	32 (6.99e-02)	52	43 (2.30e-06)

Table 21: Sizes of the largest connected components under the six tested settings. The first column of each group of two columns shows the number of genes associated with a particular disorder represented in the interactome under analysis. The second column provides the size of the largest connected component and its associated p-value.

4.3.3 Interactome-based overlaps between disease pairs

In interactome one stringent analysis, one hundred and fourteen overlaps were found to present values lower than 0. Of those, only eight presented FDR adjusted p-values under 0.05, and thus, separation values lower than expected by chance. The significant overlaps included one formed by a pair of CNS disorders, BD and MD (S_{AB} = -0.40, p-adj = 1.46e-03) and seven between cancer pairs, which included BRCA and LGCA (S_{AB}= -0.17, p-adj = 4.31e-02), BRCA and PRCA (S_{AB} = -0.18, p-adj = 8.70e-03), BRNCA and LGCA (S_{AB} = -0.35, p-adj = 2.27e-02), CRCA and LGCA (S_{AB} = -0.28, p-adj = 6.30e-03), CRCA and OVCA (S_{AB} = -0.28, p-adj = 7.00e-03), LGCA and OVCA (S_{AB} = -0.54, p-adj = 8.98e-09), and LGCA and STCA (S_{AB} = -0.41, p-adj = 3.02e-03). Interactome 2 stringent analysis yielded one hundred and twenty-one interactome overlaps with values lower than cero. Only four presented adjusted p-values lower than 0.05. Including BD and MD (S_{AB} = -0.68, p-adj = 3.30e-14) as well as three instances of cancer pairs. BRCA and LGCA (S_{AB}= -0.13, p-adj = 1.06e-02), CRCA and OVCA (S_{AB} = -0.23, p-adj = 1.40e-02), and LGCA and OVCA (S_{AB} = -0.39, p-adj = 1.13e-05). Finally, Interactome three analysis returned forty-four overlaps lower than cero for which only one was found to be significant, LGCA and OVCA (S_{AB} = -0.4, p-adj = 1.68e-02).

In the case of the interactome one relaxed analysis, 114 overlaps presented values lower than 0, from which nineteen had adjusted p-values inferior to 0.05. Significant overlaps were observed between pairs of CNS disorders, such as BD and MD (S_{AB} = -0.59, p-adj = 5.24e-107), BD and SCZ (S_{AB} = -0.44, p-adj = 8.20e-79), and MD and SCZ (S_{AB} = -0.35, p-adj = 3.14e-43), whereas the rest of significant overlaps belonged to cancer pairs, and included ALL and AML (S_{AB} = -0.15, p-adj = 4.06e-02) and BLCA and LGCA (S_{AB} = -0.16, p-adj = 1.93e-03), among others. The relaxed analysis of the interactome two also yielded 114 negative overlaps, of which nineteen were found to be significant, which included the same significant overlaps between CNS disorders and different instances of significant overlaps between diverse cancer types. Finally, in the case of the interactome three relaxed analysis, sixty overlap measures presented values lower than cero, but only eleven were found to be significant and included disease pairs similar to those reported under the rest of the settings.

In general, we did not observe significant negative separation values between CNS disorders and cancers' disease modules under any of the tested settings. Despite it, some instances of non- significant negative values were present in our results, including, ALL and ASD (S_{AB} = -0.18), ASD and HANC (S_{AB} = -0.11), ASD and KDNCA (S_{AB} = -0.24), and ASD and STCA (S_{AB} = -0.13) in the interactome one stringent analyses, or those observed between ASD and CHLCA (S_{AB}= -0.16), ASD and THCA (S_{AB}= -0.13), BD and CHLCA (S_{AB}= -0.14), CHLCA and MD (S_{AB}= -0.12) in the interactome two stringent analysis. Additional examples of non-signicant negative separation values between

CNS disorders and cancers were observed under the rest of the settings. **Figure 18** shows a heatmap representation of the disease-disease overlap analyses carried out under the six different settings. The analysis results obtained in the different tested settings presented a variety of correlation strengths when compared to each other ranging from r = 0.29 in the comparison between I2S and I3R and r = 0.92 in the comparison between I1R and I3R (**See Figure 19**). The complete list of significant disease-disease interactome-based overlaps is provided in **Supplementary Appendix 1 Table 41**.

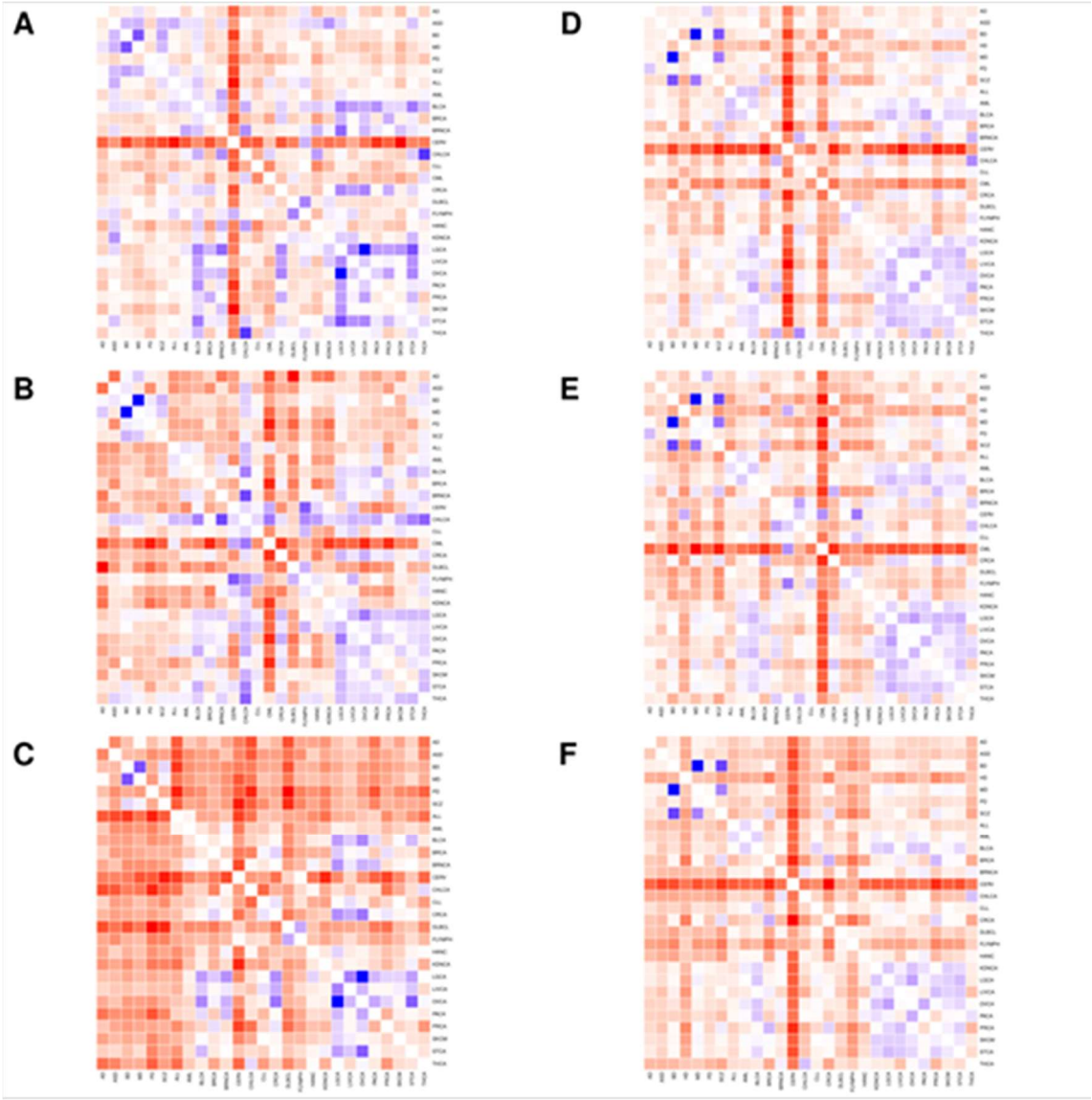

Figure 18: S_{AB} values under the six different tested settings. A) Interactome 1 stringent, B) Interactome 2 stringent, C) Interactome 3 stringent, D) Interactome 1 relaxed, E) Interactome 2 relaxed, F) Interactome 3 relaxed. For subfigures A, B, and C the order of the disorders from left to right and from top to bottom is. AD, ASD, BD, MD, PD, SCZ, ALL, AML, BLCA, BRCA, BRNCA, CERV, CHLCA, CLL, CML, CRCA, DLBCL, FLYMPH, HANC, KDNCA, LGCA, LIVCA, OVCA, PACA, PRCA, SKCM, STCA, THCA. In the case of subfigures D, E, and F, HD is placed between BD and MD. Blue colors indicate negative S_{AB} values, whereas red colors depict positive S_{AB}values. White colors symbolize S_{AB} values close to cero.

Figure 19: Correlations between The Sab values obtained under each setting. I1_St = Interactome 1 stringent, I2_St = Interactome 2 stringent, I3_St = Interactome 3 stringent, I1_Rel = Interactome 1 relaxed, I2_Rel = Interactome 2 relaxed, I3_Rel = Interactome 3 relaxed. Blue colors indicate positive correlations. The size and color intensity of the circles depicted in the upper-right corner is proportional to the correlation values, with higher values presenting bigger circles of more intense blue color. The lower-left corner contains the numeric values of the computed Pearson's correlations.

4.3.4 Genetic correlation analyses results

We were able to identify and retrieve GWAS summary statistics from 20 studies. Three of them examined the implication of genetic variability in AD's risk, whereas two did it for ASD and PD. BD, MD, and SCZ were represented with one GWAS study each. No GWAS summary statistics data was available for HD, given its monogenetic nature. We could only obtain information for six site-specific cancers, which included BRCA (3 studies), PRCA (2 studies), CRCA (1 study), OVCA (1 study), SKCM (2 studies), and LGCA (1 study). The observed scale estimated heritabilities ranged from 0.0058 in the case of CRCA to 0.45 in the case of the ASD 2 datasets. **Table 22** shows information regarding the dataset's sample sizes, the number of cases and controls, and the parameters linked to the SNP heritability estimation using LD-score regression. All pairwise genetic correlations were computed between the selected studies. **Table 23** displays the significant genetic correlations detected, presenting nominal p-values under 0.05. The genetic correlations computed between pairs of studies targeting the same disorder were found to present the maximum correlation values as well as the lower p-values, and included the positive genetic correlations found between AD 2 and AD 3 (r_g= 0.92, p-val = 1.60e-10), ASD 1 and ASD 2 (r_g = 0.84, p-val = 4.99e-87) or BRCA 1 and BRCA 3 (r_g = 0.92, p-val = 1.26e-32), among others (**See Table 23 A**). Significant positive genetic correlations were also observed between several pairs of CNS disorders. The most significant were those found between ASD 1 and MD (r_g = 0.37 , p-val = 9.34e-37), ASD 1 and SCZ (r_g= 0 .21, p-val = 1.15e-08), BD and MD (r_g= 0.33, p-val = 3.42e-34), BD and SCZ (r_g = 0.68, p-val = 8.82e-229), as well as SCZ and MD (r_g= 0.31 p-val = 6.33E-38). More modest significant genetic correlations were also observed between other pairs of CNS disorders, including AD 3 and PD 1 (r_g = 0.20, p-val = 0.012), among others (**See Table 23 B**). Genetic correlations between cancer pairs tended to be lower than those observed between CNS pairs. Instances of those include the positive association observed between BRCA 2 and OVCA (r_g = 0.23, p-val = 2.00e-04), or PRCA 2 and BRCA 2 (r_g = 0.11, p-val = 0.01) (**Table 23 C**). Finally, a number of significant genetic correlations were also observed between CNS disorders and cancers (**Table 23 D**). Those included the negative correlations between ASD 1 and PRCA 2 (r_g = -0.16, p-val = 7.30e-03), ASD 2 and BRCA 2 (r_g = -0.1, p-adj = 1.60e-02), as well as the positive correlations identified between BD and BRCA 1 (r_g = 0.11, p-val = 6.00e-03), BD and BRCA 2 (r_g = 0.07, p-val = 1.32e-02), MD and BRCA 1 (r_g = 0.08, p-val = 9.00e-03), MD and BRCA 2 (r_g= 0.085, p-val = 4.02e-05), and MD and LGCA (r_g = 0.28, p-val = 5.00e-04). Slight but significant positive correlations were also observed between PD and both PRCA and SKCM. PD 1 and PRCA 1 (r_g = 0.09, p-val = 3.16e-02) and PD 1 and SKCM 2 (r_g = 0.14, p-val = 4.00e-02). Finally, Schizophrenia was found to be positively correlated with breast cancer, BRCA 1 (r_g = 0.11, p-val = 5.00e-04), BRCA 2 (r_g = 0.14, p-val = 1.75e-08) and ovarian cancer OVCA (r_g = 0.12, p-val = 0.04).

Study	Source	N (Cases/ Controls)	Tested SNPs	Mean Chi^2	Lambda GC	Max Chi^2	Genome-Wide significant SNPs	SNP heritability
AD 1	IGAP [312]	54162 (17008/37154)	1150200	1.11	1.09	565.21	165	0.07
AD 2	GR@ACE project [313]	21235 (11999/9236)	1204123	1.09	1.068	1123.06	59	0.13
AD 3	GWAS catalog [314]	455258 (71880AD*/383378)	1203908	1.12	1.08	1009.11	320	0.01
ASD 1	iPSYCH project [315]	46350 (18381/27969)	1161575	1.186	1.16	35.93	28	0.20
ASD 2	PGC	10610 (5305/5305)	1189408	1.05	1.04	24.09	0	0.45
BD	PGC [316]	46582 (20352/31358)	1103145	1.39	1.33	56.53	74	0.35
SCZ	PGC [58]	150064 (36989/113075)	1140241	1.79	1.58	120.76	1725	0.24
MD	PGC [317]	500199 (170756/329443	1217311	1.60	1.45	77.81	943	0.06
PD 1	Nalls et al [318]	482730 (33674/449056)	1137530	1.14	1.09	180.42	276	0.02
PD 2	23 And Me [319]	308557 (6477/302080)	1211658	1.10	1.08	164.95	142	0.02
PRCA 1	GWAS catalog [320]	140254 (79148/61106)	1206082	1.51	1.23	846.34	2733	0.16
PRCA 2	UK Biobank	206770 (6879/199891)	1211361	1.13	1.09	181.15	397	0.03
BRCA 1	BCAC [321]	247173 (133384/113789)	519352	1.72	1.38	481.43	1389	0.12
BRCA 2	GWAS catalog [322]	139274 (76192/63082)	1128758	1.68	1.36	1424.99	2832	0.22
BRCA 3	UK Biobank	245494 (10478/235016)	1211361	1.11	1.08	314.15	276	0.02
CRCA	UK Biobank	387318 (4562/382756)	1215182	1.06	1.052	51.05	22	0.01
OVCA	GWAS catalog [323]	85426 (16924/68502)	1149515	1.09	1.06	169.17	209	0.04
SKCM 1	UK Biobank	452264 (2465/449799)	1211361	1.04	1.03	121.09	144	0.00
SKCM 2	Genomel [324]	32383 (11523/20860)	1100284	1.12	1.08	372.76	561	0.17
LGCA	UK Biobank	452264 (1655/450609)	1211361	1.02	1.01	26.03	0	0.00

Table 22: Included studies and characteristics in the genetic correlation analyses. *This dataset includes as cases AD patients, as well as individuals with a family history of AD. PGC = Psychiatric Genomic Consortium. BCAC = Breast cancer association consortium.

A) Significant genetic correlations between pairs of studies targeting the same disorder					
Disease 1	**Disease 2**	**r_g**	**r_g SE**	**z**	**p**
AD 1	AD 2	1.8	0.33	5.54	3.09e-08
AD 2	AD 3	0.92	0.14	6.4	1.60e-10
ASD 1	ASD 2	0.84	0.04	19.77	4.99e-87
PD 1	PD 2	0.86	0.06	13.44	3.51e-41
BRCA 1	BRCA 2	1.02	0	233.42	0.00e+00
BRCA 1	BRCA 3	0.93	0.08	11.89	1.26e-32
BRCA 2	BRCA 3	1	0.06	17.17	4.47e-66
PRCA 1	PRCA 2	1.04	0.05	20.58	4.59e-94
SKCM 1	SKCM 2	1.24	0.28	4.48	7.49e-06
B) Significant genetic correlations between pairs of CNS disorders					
Disease 1	**Disease 2**	**r_g**	**r_g SE**	**z**	**p**
AD 3	BD	0.15	0.06	2.42	1.54e-02
AD 3	MD	0.17	0.07	2.65	8.00e-03
AD 3	PD 1	0.21	0.08	2.51	1.21e-02
AD 3	SCZ	0.1	0.05	1.98	4.76e-02
ASD 1	BD	0.15	0.04	3.52	4.00e-04
ASD 1	MD	0.37	0.03	12.66	9.34e-37
ASD 1	PD 2	-0.18	0.07	-2.64	8.20e-03
ASD 1	SCZ	0.21	0.04	5.71	1.15e-08
ASD 2	MD	0.1	0.04	2.43	1.49e-02
ASD 2	PD 2	-0.2	0.07	-2.87	4.10e-03
ASD 2	SCZ	0.18	0.05	3.82	1.00e-04
BD	MD	0.33	0.03	12.19	3.42e-34
BD	SCZ	0.68	0.02	32.29	8.82e-229
SCZ	MD	0.31	0.02	12.87	6.33e-38
C) Significant genetic correlations between pairs of cancers					
Disease 1	**Disease 2**	**r_g**	**r_g SE**	**z**	**p**
BRCA 2	OVCA	0.23	0.06	3.68	2.00e-04
BRCA 2	SKCM 2	0.12	0.04	2.68	7.40e-03
BRCA 3	CRCA	-0.22	0.13	-1.74	8.11e-02
PRCA 1	BRCA 1	0.11	0.05	2.32	2.01e-02
PRCA 1	BRCA 2	0.07	0.03	2.41	1.59e-02
PRCA 2	BRCA 2	0.11	0.04	2.56	1.05e-02
D) Significant genetic correlations between CNS disorders and cancers					
Disease 1	**Disease 2**	**r_g**	**r_g SE**	**z**	**p**
ASD 1	PRCA 2	-0.16	0.06	-2.68	7.30e-03
ASD 2	BRCA 2	-0.1	0.04	-2.41	1.59e-02
BD	BRCA 1	0.11	0.04	2.71	6.70e-03
BD	BRCA 2	0.07	0.03	2.48	1.32e-02
MD	BRCA 1	0.08	0.03	2.63	8.60e-03
MD	BRCA 2	0.09	0.02	4.11	4.02e-05
MD	LGCA	0.28	0.08	3.46	5.00e-04
MD	OVCA	0.09	0.05	1.88	5.95e-02

PD 1	PRCA 1	0.09	0.04	2.15	3.16e-02
PD 1	SKCM 2	0.14	0.07	2.01	4.41e-02
PD 2	PRCA 2	0.16	0.08	2.01	4.44e-02
SCZ	BRCA 1	0.11	0.03	3.47	5.00e-04
SCZ	BRCA 2	0.14	0.02	5.63	1.75e-08
SCZ	OVCA	0.12	0.06	2.02	4.31e-02

Table 23: Significant genetic correlations between dataset pairs. The first and the second column shows the involved datasets. Column three presents the measured genetic correlation between each pair. Column four shows the standard error of the computed genetic correlations. Finally, columns five and six indicate the z-scores and the raw p-values linked to each association.

4.4 Discussion

This chapter investigated the potential implication of interactome-based overlaps and genetic correlations in the CNS and cancer comorbidities.

First, our results suggest that disease-associated genes and variant-genes tend to be placed in nearby regions of the human interactome. This trend was observed in CNS disorders and cancers for both measures of network localization (the average intra-disease distance and the size of the largest connected components). The use of larger disease-associated sets of genes and more interconnected interactomes was linked to an increase in the number of disorders displaying significant closer distances and larger connected components.

Second, we observed significant interactome-based overlaps between pairs of neuropsychiatric conditions, suggesting that MD, BD, and SCZ constitute a cluster of genetically linked disorders. Genetic correlation analyses based on GWAS summary statistics strengthened this idea since significant genetic correlations were observed between BD and MD, BD and SCZ, and SCZ and MD. These associations are interesting on their own and could arise as a consequence of many different factors. Some systematic artifacts could confuse studies trying to associated uncover genetic associations in neuropsychiatric disorders. For instance, some authors have stated that BD is often misdiagnosed as unipolar depression [325]. In psychiatry, nosology (the branch of medical sciences dealing with the classification of diseases), traditionally represented by the ICD and DSM classification systems, is based on a categorical classification of disorders. However, an open debate is still ongoing about such classification systems' adequacy in psychiatric nosology [326]. Transdiagnostic approaches represent an alternative to traditional classification systems that are more focused on identifying processes that underlie a given class of disorders. Although it is out of this thesis's scope, the interactome-based associations and genetic correlations observed between neuropsychiatric conditions add arguments to this debate.

Third, significant overlaps were observed between the disease modules of many cancer pairs. This observation was expected since recurrent somatic mutations of many genes play a pivotal role in different cancer types. The pan-cancer analysis of 2658 whole-cancer genomes derived from 38 tumor types and their matching normal tissues identified sets of genes that suffered somatic driver mutations across cancers, including *TP53*, *CDKN2A*, *ARID1A*, *KRAS*, and *PTEN*, among others [327]. Therefore, many shared genes, including those mentioned above, were repeatedly found across the different lists of cancer-associated genes included in our analysis. Besides, significant genetic correlations were also observed between some cancer pairs, which included BRCA and OVCA or BRCA and PRCA, among others. However, the genetic correlations between the explored cancer types tended to be more modest than those observed between CNS

disorders, indicating that even if somatic alterations are frequently shared among cancers, genetic susceptibility mechanisms could differ to a greater extent between them.

Finally, we analyzed the interactome-based overlaps and genetic correlations between CNS disorders and cancers. One common trait of all the interactome-based disease-disease overlap analyses carried out under different settings was that no significant CNS-cancer overlaps were observed. Despite this fact, a number of negative interactome distance measures between CNS and cancers were observed. Some authors have recognized that the current state of knowledge limits network-based approaches' predictive power and that the high-throughput methods used to determine protein-protein interactions cover about 20% of all potential pairs in humans [32]. In addition, the information regarding disease-gene associations still is incomplete. Future studies employing more complete interactomes and better defined disease-associated gene lists could shed more light on this matter.

A number of significant genetic correlations between CNS disorders and cancers were identified in our analyses, which included the negative correlations observed between ASD and BRCA and PRCA, as well as several instances of positive genetic correlations. BD, MD, and SCZ were found to be positively correlated with BRCA. However, the correlation values were low and ranged between Rg = 0.09 in the case of MD and Rg = 0.14 in the case of SCZ. MD was found to be positively correlated with LGCA, and PD presented positive correlations with PRCA and SCKM. SCZ was also found to be present significant genetic correlations with OVCA. The assessment of the potential impact of such correlations should be the object of further research.

Previous studies have reported significant genetic correlations between AD and BRCA [38]. However, we could not reproduce these results. These differences could be explaining by the use of different datasets. The same authors reported significant genetic correlations between AD and LGCA, which we could not reproduce either. Probably due to the low statistical power of the LGCA dataset included in our analyses, which included a reduced number of cases.

Although efforts were made to obtain the LGCA and BRCA datasets employed in the previous reference, data access was not feasible due to time constraints.

In summary, we provided some specific instances of genetic correlations between CNS disorders and cancers, which implies that shared genetic variability could be involved in comorbidities observed between them.

Chapter 5. Role of pharmacological treatment in the CNS and cancer associations

5.1 Introduction

Pharmacological treatment could play a pivotal role in CNS and cancer comorbidities. Disulfiram (DSF), a drug that has been used for decades to treat alcohol dependence, provides a paradigmatic instance of the previous statement. Epidemiological research has shown that patients diagnosed with cancer who were under DSF treatment presented a reduced all-cancer and site-specific cancer mortality compared to those who had been previous users of DSF but who had stopped its use at least one year before cancer diagnosis [13].

The mechanistic description of the anti-cancer effect of DSF, which relies on NPL4 protein targeting and aggregation, has also been reported [13, 328]. Other instances of CNS medications acting as potential cancer modulators are available in the literature. For example, antipsychotics have been linked to an increase in breast cancer risk in SCZ patients [85], and the use of levodopa has been implicated in the increased melanoma risk observed in PD patients [86].

In the previous chapter, we showed that genes associated with a particular disorder tend to cluster in the same network neighborhood, called the disease module, which represents a connected sub-network of the human interactome enriched in disease proteins.

It has been hypothesized that effective causative treatments have molecular targets placed significantly closer to the disease module of the disorder they are indicated to treat, whereas palliative medications do not. This point has been tested using Interactome-based drug-disease proximity measures [329]. In this context, research has shown that most traditionally used drugs lack sensitivity towards the genetic causes of disease, targeting their symptoms instead.

In this chapter, we used two alternative but complementary methods to identify medications that could influence CNS and cancer comorbidities. The first method relies on the computation of network-based proximity measures between drug targets and disease modules. With this aim, we first compiled a list of drug indications frequently used to treat CNS disorders and cancers and obtained information about their molecular targets. Network distances were computed between pairs of drug-associated proteins and disease modules. Using this approach, we identified medications which targets were significantly closer to specific disease modules than expected by chance. Then, we looked for CNS indications placed in the interactome vicinity of cancer-related modules and for cancer indications targeting proteins placed in the same interactome neighborhood as CNS disease modules.

The second method was based on the computation of correlations between the differential expression profiles derived from our disease-specific meta-analyses (computed in **Chapter 3**) and the differential gene expression profiles of cell lines treated with the indicated drugs retrieved from the LINCS L1000 project [330]. LINCS L1000 contains an extensive collection of gene expression profiles generated using thousands of perturbagens (i.e., small molecules, ligands, micro-environments, CRISPR gene over-expression, and knockdown perturbations) and different cell lines, doses, and exposure times. Positive correlations suggest that a specific medication has the potential to produce transcriptomic changes that tend to mimic those observed in a specific disease, whereas negative correlations indicate that a particular drug could revert the transcriptomic changes in a specific disorder.

Using this approach, we identified medications-indicated for the treatment of CNS disorders that induce transcriptomic changes that significantly mimic or reverse those observed in cancer and drugs used for cancer treatment with the potential to mimic or reverse the transcriptomic changes observed in CNS gene expression signatures.

5.2 Material and methods

5.2.1 Data sources

- **Drug indications**

Drug indications for each of the included disorders were obtained from MEDI-an [331]. MEDI-an is a medication indication repository that gathers information from multiple resources including, RxNorm, MedlinePlus, and SIDER2) constructed using combining ontology-based and natural language processing (NLP) techniques. We selected the MEDI high precision subset (MEDI-HPS), which contains 13400 unique indications regarding 2136 medications. The file (MEDI_HPS.csv) was downloaded from the site, and indications for all the studied CNS disorders and cancer types were selected by inspecting disease names and ICD codes. MEDI-HPS uses concept ids CUI IDs derived from the UMLS thesaurus to identify each available drug. Rxcui IDs were translated to Drugbank IDs using the restful web API from RxNav [332] through the use of the R packages httr [333]and rjson [334].

- **Drug targets**

The molecular targets of the indicated drugs were retrieved from DrugBank [335]. In short, the XML file containing the full database was downloaded from DrugBank's website and parsed using custom R scripts. Drugs and their targets were retrieved alongside information regarding their mechanism of action and the mapping to external identifiers such as PubChem substance and compound IDs. The drugs that did not have any target in the employed interactomes were excluded. The details of the used interactomes can be found in **Chapter's 4** material and methods section.

- **Differential expression profiles of cell lines treated with the indicated drugs**

The differential gene expression profiles of those cell lines treated with the indicated drugs were retrieved from LINCS L1000. LINCS L1000 contains an extensive collection of gene expression profiles generated using thousands of perturbagens (i.e., small molecules, ligands, micro-environments, CRISPR gene over-expression, and knockdown perturbations) and different cell lines, doses, and exposure times. Datasets from phases I (GSE92742) and II (GSE70138) of the project were downloaded from GEO. LINCS's perturbation IDs were mapped to Drugbank IDs through Inchi keys using the UnChem RESTful service [336]. All analyses were carried out using LINCS L1000 Level 5 data, which includes differential gene expression signatures computed by comparing three technical replicates of the same perturbation to appropriate controls. In Level 5 datasets, each perturbation (drug treatment) is represented by several differential gene expression signatures generated by treating different cell lines at different times and concentrations. We constructed an

individual signature for each drug by combining all the available signatures generated using it, applying Stouffer's method, as explained in (**See section 5.2.3**).

5.2.2 Network-based proximity measures between drugs and diseases

We measured drug-disease distances in the human interactome through a previously reported metric, which computes the average shortest path length between a specific drug's targets and the nearest disease protein. This metric has been found to better capture the associations in cases in which a drug does not directly target all disease-associated proteins and has shown superior performance in discriminating between the known and unknown drug-disease associations compared to similar metrics [329].

Given sets S (the set of disease-associated proteins) and T (the set of protein targets for a specific drug) , let $d(s,t)$, be the shortest path length between nodes s and t in the interactome. The distance function, called the closest distance, is defined as follows:

$$d_c(S,T) = \frac{1}{\|T\|}\sum_{t \in T} min_{s \in S}\, d(s,t)$$

were $\|T\|$ is the number of genes in set T. Basically, for each protein in the set of drug-associated proteins T , the minimum distance to any element of the set of disease-associated proteins S is selected. Then the average of the resulting values for all drug-associated proteins is computed.

The significance of the obtained d_c values was assessed by constructing null distributions by randomly sampling 1000 sets of genes of the same dimension as those found in both the disease- and drug-associated sets. Z-values were computed as follows using the actual closest distance and the null distribution.

$$z(S,T) = \frac{d_c(S,T) - \langle d_{c\,rand}(S,T)\rangle}{\sigma d_{c\,rand}(S,T)}$$

Where $\langle d_{c\,rand}(S,T)\rangle$ is the average of the random distribution and $\sigma d_{c\,rand}(S,T)$ its standard deviation. P-values were computed by comparing the obtained Z-values to the standard normal distribution. The same degree-preserving sampling procedure described in **Chapter 4** was applied to construct the null distributions accounting for the impact of the node degree distributions of the disease- and drug-associated sets.

5.2.3 Identification of drugs producing transcriptomic effects that mimic or reverse the differential gene expression signatures of the included disorders

To determine if a particular drug could induce transcriptomic changes that could resemble or revert the differential gene expression profiles of the disorders studied in this Thesis, we computed correlations between the drug signatures and the Z-values derived from the differential gene expression meta-analyses of each disorder obtained in **Chapter 3**.

Significantly positively correlated drugs exhibiting correlation values higher than 0.15 were selected as candidates to generate differential expression changes that would partially mimic those found in a particular disorder, whereas drugs showing significant correlation values lower than -0.15 were selected as potential candidates to revert the differential gene expression profile generated by a particular disorder.

In LINCS L1000, each perturbation is represented by many differential gene expression signatures, which have been generated by treating different cell lines at different times and concentrations and by comparing them to appropriate controls. A consensus signature was computed for each perturbation using Stouffer's method.

For each signature, weights were computed as the average correlation to all other signatures. Then a consensus signature was created using the following formula applied to each gene.

$$Z \sim \frac{\sum_{i=1}^{k} w_i Z_i}{\sqrt{\sum_{i=1}^{k} w_i^2}}$$

Where w_i is the weight of the i-th signature and Z_i is the differential expression value of a particular gene in the i-th signature.

Finally, Pearson's correlations between each disease signature and all drug consensus signatures were computed.

5.3 Results

5.3.1 Identified indications

Three-hundred and fifty drug indications encompassing two hundred and twenty-five drugs were found after merging MEDI-HPS and DrugBank data. Two-hundred and eight of them had known protein targets reported in DrugBank. Several medications were indicated for the treatment of more than one disorder. In the case of CNS diseases, selegiline, an antidepressive and antiparkinsonian which selectively inhibits monoamine oxidase B, was found to be indicated for AD, MD, PD, and SCZ, whereas risperidone, an atypical antipsychotic targeting several 5-HT serotonin receptor subtypes, was found to be indicated for the treatment of ASD, BD, MD, and SCZ. Several other drugs were indicated to treat more than one CNS disorder, including quetiapine, aripiprazole, valproic acid, and olanzapine, among others. Cyclophosphamide, a chemotherapeutic agent and immune system suppressor, was found to be indicated for the treatment of six cancer types, including AML, BRCA, BRNCA, CLL, CML, and LGCA, whereas cisplatin did it for BLCA, BRNA, CERV, HANC, LGCA, and OVCA treatment. The complete list of drug indications can be found in **Appendix 8**.

5.3.2 Interactome-based distances between drug targets and disease modules

The interactome-based distances between drug targets and disease modules were computed using the interactome one described in **Chapter 4**. and both sets of disease-associated genes and variants (stringent and relaxed). We will refer to the analyses carried out using the stringent and relaxed lists of disease-associated genes as setting one and setting two, respectively. Two hundred and two drugs which targets were included in our interactome were selected for downstream analysis. Overall, 5656 distances between drugs and disease modules were computed under settings one and two.

5.3.2.1 Closest distances d_c between disease modules and their drug indications.

First, we investigated if drugs tended to present lower closest distances (d_c) with the disorders for which they were indicated (target disorders) compared to those for which they were not (off-target disorders).

Under setting one, the mean interactome distance between drugs and their target disorders (n = 312) was 1.72 (sd = 0.57) whereas in the case of drugs and off-target disorders pairs (n = 5344) was 1.96 (sd = 0.44). The differences in means between both groups were found to be significant (t = 7.35, p-val = 1.5e-12) (**Figure 20 A**). More pronounced differences between groups were observed under setting two, where the mean d_c value between drugs and their target disorders was found to be 0.97 (sd = 0.71) compared to the mean d_c value of 1.56 observed in off-target drug-disease pairs (t = 14.38, p-val < 2.2e-16) (**Figure 20 B**). A Pearson's correlation value of

0.51 was found between the Z-scores computed under settings one and two (**Figure 20 C**), and the overlap coefficient of the significant associations (FDR < 0.05) identified under the two settings was 0.72 (**Figure 20 D**).

Next, we wondered if medications and the disease modules of their target disorders presented shorter interactome-based distances than expected by chance with their target disorders.

Under setting one, we were able to compute 312 closest distances between disease modules and their indicated drugs' targets. One hundred and twenty of them (38.46%) were found to present d_c values lower than expected by chance (p-val < 0.05), of those, 61 (19.6%) remained significant after adjustment by multiple comparisons (FDR < 0.05). The most significantly associated indications in the case of CNS disorders were two PD drugs, carbidopa (p-adj = 1.39e-05), and selegiline (p-adj = 4.16e-04), pimozide, which is indicated for the treatment of SCZ (6.00e-04), and clomipramine (p-adj = 6.15e-04) and olanzapine (p-adj = 2.28e-03), which were significantly associated with the disease modules of MD and BD, respectively. In the case of cancer drugs, the most significant associations were found between PRCA and cyproterone (p-adj = 3.75e-11), BRCA and methyltestosterone (p-adj = 6.61e-10), ALL and dasatinib (p-adj = 1.54e-05), and BRCA and toremifene (p-adj = 5.45e-05). **Supplementary Appendix 1 Table 42** shows all drug indications displaying significant adjusted p-values under setting one.

One hundred and seventy-four drug-disease indications (55.8%) presented d_c values lower than expected by chance (nominal p-values < 0.05) in setting two analyses. After adjustment by multiple comparisons, 144 (46.15%) remained significant. For CNS disorders, the top associations were those found between the following drugs and disease modules, amoxapine and MD (p-adj = 8.02e-23), loxapine and SCZ (p-adj = 5.85e-22), and trimipramine and MD (p-adj = 5.15e-21), whereas in the case of cancer the most significant associations were those found between capecitabine and CRCA (p-adj = 1.48e-07), mitoxantrone and PRCA (p-adj = 1.17e-06), and fluorouracil and CRCA (p-adj = 2.52e-06). **Supplementary Appendix 1 Table 43** includes all drug indications displaying significant FDR-adjusted p-values under setting two.

5.3.2.2 Significant off-target drug-disease interactome-based closest distances

In this section, we focus on off-target drug-disease associations. Those were defined as significant interactome-based associations between drugs and the disease modules of the disorders for which they were not indicated.

Under setting one, eighty-six drugs indicated for the treatment of one or more CNS disorders presented significant off-target associations with the disease modules of other CNS disorders. Those included the associations found between phenelzine, a monoamine oxidase inhibitor indicated for the treatment of MD, which was found to be associated with the PD module (p-adj = 2.47e-07), mesoridazine, an SCZ medication, which was associated with PD (p-adj = 3.17e-05), and tranylcypromine, another monoamine oxidase inhibitor indicated for the treatment of MD, which was found to be significantly associated with the PD module (p-adj = 1.72e-04).

Under setting two, 243 drugs indicated for the treatment of one or more CNS disorders were found to present interactome-based associations with the disease modules of other CNS disorders for which they were not indicated. The most significant were those observed between loxapine, an antipsychotic drug indicated for the treatment of SCZ, which was significantly associated with the MD (p-adj = 1.12e-26) BD (p-adj = 2.10e-22) disease modules. The tetracyclic antidepressant amoxapine, which was indicated for the treatment of MD and was found to be heavily associated with the disease module of BD (p-adj = 7.08e-24) and SCZ (p-adj = 8.02e-23), and aripiprazole, an atypical antipsychotic medication indicated for the treatment of ASD, SCZ, and BD which presented significant associations with the disease module of MD (p-adj = 5.91e-21).

Seventy-three off-target interactions were observed between drugs indicated for the treatment of one or more cancers and other cancer types under setting 1. Lapatinib, a tyrosine kinase inhibitor indicated for the treatment of BRCA, was significantly associated with the BRNCA module (p-adj = 1.56e-07). Imatinib, a chemotherapy medication used for CML treatment, was associated with the ALL module (p-adj = 1.30e-05). Dasatinib, another CML drug that acts as a tyrosine kinase inhibitor, was also found to be among the top off-target associations observed in cancer drugs and showed a significant closest distance with the ALL module (p-adj = 1.54e-05).

Two hundred and eighty-seven off-target significant closest distances were observed between drugs indicated to treat one or more cancer types and other cancers. Those included the association found between idarubicin, an anthracycline indicated for the treatment of AML and CML, and the PRCA disease module (p-adj = 7.82e-10) and pemetrexed, a chemotherapy medication indicated for the treatment of LGCA, which was found to be significantly linked to the FLYMPH disease module (p-adj = 4.78e-09), among others.

Under setting one, thirty-nine drugs indicated for treating one or more CNS disorders were significantly associated with cancer modules. The most significant included CNS drugs that were

found to target the disease module of ALL, such as imipramine (p-adj = 3.88e-05), aripiprazole (p-adj = 3.43e-04), and clozapine (p-adj = 6.15e-04). Selegiline, a medication indicated for the treatment of SCZ, AD, MD, and PD, was associated with the disease module of HANC (p-adj = 6.15e-04).

Seventy-nine drugs indicated for the treatment of one or more CNS disorders presented significantly shorter closest distances than expected by chance with cancer disease modules under setting two. Rivastigmine, a cholinesterase inhibitor indicated for AD and PD treatment, was found to be associated with the KDNCA disease module (p-adj = 1.97e-12). Tacrine, an acetylcholinesterase inhibitor used in AD with was also associated with the KDNCA disease module (p-adj = 6.66e-06). Fluoxetine, a selective serotonin reuptake inhibitor used in MD and BD, was found to be associated with the LGCA module (p-adj = 1.40e-04), and donepezil, a reversible acetylcholinesterase inhibitor indicated for the treatment of AD, was found to be linked to the BRCA disease module (p-adj = 1.76e-04).

Under setting 1, eighteen drugs indicated for the treatment of one or more cancers were found to be associated with CNS modules, including crizotinib, which is indicated for LGCA and was found to be associated with the ASD module (p-adj = 2.28e-04), docetaxel, a chemotherapy medication indicated for the treatment of several cancers which was associated with the AD module (p-adj = 7.13e-04), and gemtuzumab, a CD33 antibody indicated for the treatment of AML, which was found to be associated with the disease module of AD (p-adj = 2.01e-03),

Seventy drugs indicated for cancer treatment were found to present significant associations with CNS disease modules under setting two. Different drugs indicated for the treatment of breast cancer were found to present significant associations with the disease modules of CNS disorders that were placed among the most significant associations, including toremifene, a selective estrogen receptor modulator which was associated with AD's disease module (p-adj = 6.30e-11), testosterone and testosterone enanthate, which were found to be significantly associated with MD (p-adj = 4.85e-09, 4.48e-07), and fulvestrant, a selective estrogen receptor degrader, which was found to be significantly associated with the AD module (p-adj = 2.11e-08). Other medications, such as arsenic trioxide, a compound used to treat AML, was found to be associated with the SCZ module (p-adj = 1.04e-07), whereas cisplatin, a chemotherapy medication which is indicated for BRNCA, CERV, BLCA, LGCA, HANC, OVCA, was found to be significantly associated with AD (p-adj = 4.41e-05). **Table 24** shows the most significant off-target associations between drugs indicated for the treatment of CNS disorders and cancer modules and the top significant drugs indicated for cancer treatment associated with CNS modules for settings 1 and 2. **Supplementary Appendix 1 Tables 44**

and 45 provide the complete list of significant off-target associations for both setting one and two, respectively.

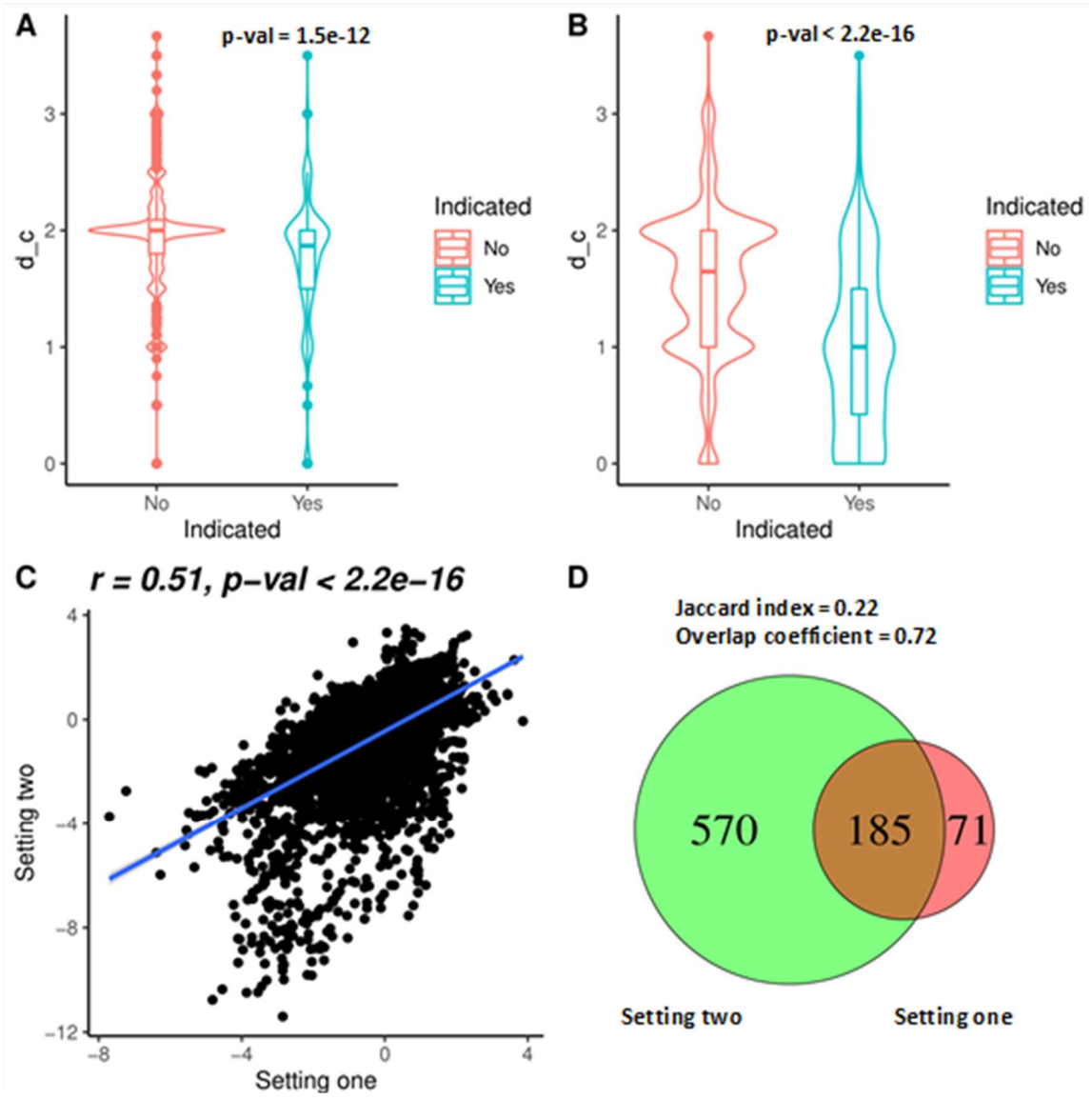

Figure 20: Closest distance analysis results. A) Differences between disease-drug target and off-target associations under setting one. B) Differences between disease-drug target and off-target associations under setting two. C) Scatter plot depicting the correlations between the Z-scores derived from the disease-drug associations computed under settings one and two. D) Venn diagram showing the overlap between the significant drug-disease associations in settings one and two.

A) CNS indications presenting significant associations with cancer modules under setting 1							
Indicated disorder	**Associated disorder**	**DB ID**	**Drug name**	d_c	**Z**	**p-val**	**p-adj**
MD	ALL	DB00458	Imipramine	1.56	-5.25	7.54e-08	3.88e-05
MD, ASD, SCZ, BD	ALL	DB01238	Aripiprazole	1.73	-4.7	1.33e-06	3.43e-04
BD, SCZ	ALL	DB00363	Clozapine	1.7	-4.51	3.25e-06	6.15e-04
BD, SCZ, MD	ALL	DB00334	Olanzapine	1.67	-4.36	6.62e-06	1.07e-03
AD	ALL	DB01043	Memantine	1.43	-4.29	9.07e-06	1.35e-03
SCZ	ALL	DB00408	Loxapine	1.79	-4.2	1.36e-05	1.79e-03
SCZ, AD, MD, PD	HANC	DB01037	Selegiline	1	-4.13	1.79e-05	2.11e-03
SCZ, MD	ALL	DB00477	Chlorpromazine	1.65	-4.12	1.92e-05	2.22e-03
MD	HANC	DB01247	Isocarboxazid	1	-4.07	2.38e-05	2.49e-03
MD	ALL	DB00726	Trimipramine	1.69	-4.03	2.76e-05	2.73e-03
MD	ALL	DB01142	Doxepin	1.64	-4.02	2.91e-05	2.77e-03
MD	ALL	DB00321	Amitriptyline	1.7	-3.97	3.65e-05	3.22e-03
MD	HANC	DB00752	Tranylcypromine	1	-3.93	4.26e-05	3.55e-03
SCZ, BD	ALL	DB00246	Ziprasidone	1.76	-3.86	5.62e-05	4.24e-03
MD	ALL	DB00543	Amoxapine	1.74	-3.77	8.26e-05	5.52e-03
HD, SCZ	ALL	DB01100	Pimozide	1.25	-3.71	1.03e-04	6.20e-03
SCZ	ALL	DB00679	Thioridazine	1.4	-3.66	1.28e-04	6.82e-03
PD	FLYMPH	DB01367	Rasagiline	1	-3.62	1.45e-04	7.60e-03
SCZ	ALL	DB00434	Cyproheptadine	1.5	-3.59	1.66e-04	8.29e-03
B) Cancer idications presenting significant associations with CNS modules under setting 1							
Indicated disorder	**Associated disorder**	**DB ID**	**Drug name**	d_c	**Z**	**p-val**	**p-adj**
LGCA	ASD	DB08865	Crizotinib	1	-4.79	8.47e-07	2.28e-04
LGCA, SKCM, PRCA	AD	DB01248	Docetaxel	1.33	-4.46	4.04e-06	7.13e-04
AML	AD	DB00056	Gemtuzumab ozogamicin	1.58	-4.16	1.60e-05	2.01e-03
LGCA, BRCA, HANC	AD	DB01229	Paclitaxel	1.33	-4.08	2.23e-05	2.38e-03
CLL	AD	DB00087	Alemtuzumab	1.64	-3.67	1.21e-04	6.71e-03
CML, ALL	AD	DB01254	Dasatinib	1.64	-3.65	1.29e-04	6.82e-03
BRCA	AD	DB08871	Eribulin	1	-3.49	2.37e-04	1.07e-02
LGCA, BRCA, CRCA, KDNCA	AD	DB00112	Bevacizumab	1.64	-3.49	2.45e-04	1.08e-02
CRCA	AD	DB00002	Cetuximab	1.67	-3.43	3.04e-04	1.27e-02
BRNCA, CERV, BLCA, LGCA, HANC, OVCA	AD	DB00515	Cisplatin	1.5	-3.31	4.64e-04	1.65e-02

PRCA, BRCA	MD	DB00351	Megestrol	0.5	-3.12	8.89e-04	2.65e-02
AML	MD	DB00755	Tretinoin	1.59	-3.13	8.89e-04	2.65e-02
KDNCA	ASD	DB06287	Temsirolimus	1	-2.97	1.49e-03	3.84e-02
BRCA	AD	DB00072	Trastuzumab	1.69	-2.97	1.51e-03	3.84e-02
AML	SCZ	DB01169	Arsenic trioxide	1.2	-2.92	1.73e-03	4.22e-02
LGCA	ASD	DB00317	Gefitinib	1	-2.9	1.87e-03	4.40e-02
SKCM	SCZ	DB06186	Ipilimumab	1	-2.88	1.99e-03	4.60e-02
ALL, CML	AD	DB00619	Imatinib	1.67	-2.85	2.20e-03	4.88e-02

A) CNS indications presenting significant associations with cancers modules under setting 2

Indicated disorder	**Associated disorder**	**DB ID**	**Drug name**	d_c	**Z**	**p-val**	**p-adj**
AD, PD	KDNCA	DB00989	Rivastigmine	0	-7.54	2.33e-14	1.97e-12
AD	KDNCA	DB00382	Tacrine	0.67	-5.07	2.03e-07	6.66e-06
MD, BD	LGCA	DB00472	Fluoxetine	0.67	-4.4	5.53e-06	1.40e-04
AD	BRCA	DB00843	Donepezil	0.5	-4.34	7.18e-06	1.76e-04
PD	STCA	DB00248	Cabergoline	1.21	-4.21	1.26e-05	2.82e-04
SCZ	STCA	DB00408	Loxapine	1.39	-4.06	2.49e-05	5.12e-04
AD, PD	BRCA	DB00989	Rivastigmine	0	-4.03	2.84e-05	5.77e-04
PD	FLYMPH	DB01367	Rasagiline	1	-3.94	4.12e-05	8.09e-04
MD, ASD, SCZ, BD	STCA	DB01238	Aripiprazole	1.36	-3.85	5.86e-05	1.10e-03
BD, SCZ	STCA	DB06216	Asenapine	1.24	-3.83	6.45e-05	1.20e-03
AD	KDNCA	DB00843	Donepezil	1.17	-3.82	6.78e-05	1.24e-03
SCZ, MD	STCA	DB00477	Chlorpromazine	1.25	-3.76	8.60e-05	1.51e-03
AD, BD, MD	SKCM	DB00313	Valproic Acid	0.89	-3.75	8.73e-05	1.52e-03
AD	BRCA	DB00382	Tacrine	0.33	-3.71	1.02e-04	1.74e-03
BD, SCZ	STCA	DB00363	Clozapine	1.35	-3.69	1.14e-04	1.92e-03
MD	ALL	DB00458	Imipramine	1.5	-3.66	1.27e-04	2.11e-03
MD	STCA	DB01151	Desipramine	1.27	-3.48	2.51e-04	3.66e-03
PD	LGCA	DB00810	Biperiden	0.5	-3.45	2.77e-04	4.00e-03
PD	HANC	DB01367	Rasagiline	0.5	-3.43	3.03e-04	4.33e-03
SCZ	STCA	DB04946	Iloperidone	1.11	-3.41	3.23e-04	4.54e-03

B) Cancer indications presenting significant associations with CNS modules under setting 2

Indicated disorder	**Associated disorder**	**DB ID**	**Drug name**	d_c	**Z**	**p-val**	**p-adj**
BRCA	AD	DB00539	Toremifene	0	-7.04	9.47e-13	6.30e-11
BRCA	MD	DB00624	Testosterone	0	-6.37	9.18e-11	4.85e-09
BRCA	AD	DB00947	Fulvestrant	0	-6.13	4.37e-10	2.11e-08

AML	SCZ	DB01169	Arsenic trioxide	0.4	-5.85	2.45e-09	1.04e-07
BRCA	MD	DB13944	Testosterone enanthate	0	-5.59	1.14e-08	4.48e-07
BRCA	MD	DB06710	Methyltestosterone	0	-4.89	5.01e-07	1.53e-05
BRNCA, CERV, BLCA, LGCA, HANC, OVCA	AD	DB00515	Cisplatin	1	-4.66	1.56e-06	4.41e-05
BRCA	SCZ	DB00675	Tamoxifen	0.56	-4.49	3.58e-06	9.55e-05
BRCA	BD	DB00539	Toremifene	0	-4.45	4.27e-06	1.11e-04
KDNCA	SCZ	DB00603	Medroxyprogesterone	0.25	-4.35	6.75e-06	1.67e-04
BRCA	MD	DB00675	Tamoxifen	0.67	-4.25	1.06e-05	2.42e-04
BRCA	BD	DB00947	Fulvestrant	0	-4.18	1.44e-05	3.14e-04
BRCA	AD	DB06710	Methyltestosterone	0.5	-4.17	1.49e-05	3.25e-04
PRCA	MD	DB01196	Estramustine	0.25	-4.15	1.70e-05	3.65e-04
BRCA	SCZ	DB00947	Fulvestrant	0	-4.1	2.03e-05	4.26e-04
KDNCA	MD	DB06287	Temsirolimus	0	-4.01	3.09e-05	6.22e-04
KDNCA, PRCA	SCZ	DB01041	Thalidomide	0.4	-3.99	3.35e-05	6.67e-04
KDNCA	MD	DB01590	Everolimus	0	-3.89	4.97e-05	9.59e-04
KDNCA, PRCA	MD	DB01041	Thalidomide	0.4	-3.82	6.54e-05	1.21e-03
PRCA	SCZ	DB00513	6-Aminocaproic Acid	0	-3.81	7.08e-05	1.29e-03

Table 24: Off-target associations between CNS indicated drugs and cancer modules and cancer indicated drugs and CNS modules.. Section A presents the top CNS indications displaying significant associations with cancer modules under setting 1. Section B shows the top cancer indications presenting significant associations with CNS modules under setting 1 whereas sections C and D present the same results for setting 2.

5.3.3 LINCS L1000 analyses results

Transcriptomic signatures were available for one hundred and forty-nine out of the two hundred and twenty-five drugs indicated for the treatment of the studied disorders, which were represented by a total of twelve-thousand and sixty-one signatures. On average, each drug was used to produce 81 signatures employing different cell lines, treatment times, and concentrations. We applied Stouffer's method to combine the signature profiles generated using the same perturbations and obtained 149 consensus signatures. Drugs producing transcriptomic signatures presenting FDR-adjusted correlation p-values lower than 0.05 and absolute values of correlation higher than 0.15 were selected as medications with the potential of mimicking or reversing the transcriptomic changes produced by a specific disorder.

5.3.3.1 Correlations between the transcriptomic signatures of drugs and the differential gene expression profiles of their target disorders

Forty-seven disease-drug correlations presented absolute values larger than 0.15 and adjusted p-values lower than 0.05. Twenty-one of them were negative, whereas twenty-six were positive. Among the top negatively correlated indications for CNS disorder, we found two anti-epileptics, valproic acid (r = -0.33) and oxcarbazepine (r = -0.33) indicated for the treatment of BD, as well as isocarboxazid, a non-selective irreversible monoamine oxygenase inhibitor (MOI) used for the treatment of MD (r = -0.32). In the case of drugs indicated for cancer treatment, the highest negative correlations were found for topotecan, a topoisomerase inhibitor indicated for the treatment of cervical cancer (r = -0.32), doxorubicin, which is indicated for pancreatic cancer (r = -0.29), and epirubicin used for the treatment of STCA (r = -0.25). In addition, the transcriptomic signatures of several drugs were found to be positively correlated with the gene expression profiles of the disorders for which they are indicated. For instance, BD's profile was found to be positively correlated with carbamazepine (r = 0.37), phenelzine, a non-selective monoamine oxidase inhibitor used as an antidepressant, presented positive correlations with MD (r = 0.33), and capecitabine, a chemotherapy medication presented positive correlations with one of the cancer types for which it is indicated, CRCA (r = 0.23). **Supplementary Appendix 1 Table 46** shows the LINCS L1000 analysis results for the available drug indications and their target disorders.

5.3.3.2 Correlations between CNS disorders indicated drugs and cancer profiles and cancer indicated drugs and CNS disorders profiles

Several drugs indicated to treat CNS disorders were found to present significant negative correlations lower than -0.15 with diverse cancer types.

First, two sodium channel blockers used as anticonvulsants, lamotrigine, and carbamazepine presented negative correlations with different cancers' gene expression signatures. Lamotrigine is used to treat epilepsy, and it is also indicated to prevent depressive episodes in BD and MD. The gene expression signature of lamotrigine was found to be negatively correlated with the differential expression profiles of 4 cancer types. Carbamazepine, primarily used in the treatment of epilepsy but also in schizophrenia and as a second-line agent in bipolar disorder, was negatively correlated with the differential gene expression profiles of nine cancer types with correlations ranging from r = -0.15 in the case of DLBCL and r = -0.30 in the case of PACA.

Second, many instances of antipsychotic drugs, including five dopamine receptors antagonists, such as pimozide, prochlorperazine, haloperidol, trifluoperazine, chlorpromazine, and one typical antipsychotic (risperidone), were found to be negatively correlated to 8, 7, 10, 7, 3, and 6 different cancer types respectively. The cancer types negatively correlate with more antipsychotic

drugs were CERV (n = 6), STCA (n = 6), LIVCA (n = 5), HANC (n = 5), PACA (n = 4), CRCA (n = 4), CHLCA (n = 5), and BRNCA (n = 4). Correlations ranged from r = -0.3 in the case of pimozide and CERV and r = -0.15 in the case of trifluoperazine and BLCA.

Third, antidepressant drugs (tricyclic and tetracyclic), selective serotonin reuptake inhibitors, and monoamine oxidase inhibitors were also found to present negative correlations with the gene expression signatures of different cancer types. The transcriptomic signatures produced by three tricyclic antidepressants, amitriptyline, doxepin, and desipramine, were negatively correlated with the differential gene expression profiles of 11 cancer types, including BRNCA, CERV, CHLCA, CRCA, DLBCL, FLYMPH, HANK, KDNCA, LIVCA, PACA, and STCA. Correlations ranged from r = -0.33 in the case of PACA and desipramine and r = -0.16 in the case of desipramine and FLYMPH. The tetracyclic antidepressant maprotiline was negatively correlated with the signatures of eight cancer types. Three selective serotonin reuptake inhibitors, sertraline, fluoxetine, and paroxetine, were anticorrelated with the transcriptomic signatures of 11, 5, and 11 cancer types, respectively. Furthermore, one monoamine oxidase inhibitor used in the treatment of PD, AD, and SCZ, selegiline, was negatively correlated with 11 cancer types and displayed correlations that ranged from r = -0.32 in the case of PACA and r = -0.15 in the case of KDNCA.

One drug of the benzodiazepine family, alprazolam, and carbidopa, a medication that inhibits the peripheral metabolism of levodopa used for PD treatment, were negatively correlated with the transcriptomic signatures of 10 and 3 cancer types, respectively.

Finally, in addition to selegiline, two acetyl-cholinesterase inhibitors, donepezil, and galantamine, presented negative correlations with the differential gene expression profiles of 13 and 7 cancer types. Correlations ranged from r = -0.3 in the case of galantamine and LIVCA and r = -0.15 in galantamine and BRCA.

Positive correlations were also found between the transcriptomic signatures of drugs used to treat CNS disorders and several cancer types. Oxcarbazepine, an anticonvulsant drug also used for the treatment of BD, was positively correlated with the expression profiles of 10 cancers. Mirtazapine, a tetracyclic antidepressant, was positively correlated with the expression profiles of three cancer types. Buspirone, an antianxiety serotonin receptor inhibitor indicated for the treatment of some patients with MD, presented positive correlations with the differential expression profiles of eight cancer types, isocarboxazid which acts as a monoamine oxidase inhibitor the treatment of MD, was found to be positively associated with the differential gene expression profiles of 11 cancers. Valproic acid, which is indicated for the treatment of MD, BD, and AD, was positively correlated with the differential gene expression profiles of 11 cancer types. Finally, Three Parkinson's treatments, trihexyphenidyl hydrochloride, entacapone, and pergolide, were linked to 4, 10, and 13 cancer types, respectively. In addition, three antipsychotic drugs used

for the treatment of schizophrenia treatments, loxapine, iloperidone, fluphenazine, were linked to 8, 9, and 7 cancer types. **Supplementary Appendix 1 Table 47** includes the complete information of the CNS drugs correlated with cancer gene expression signatures.

Thirty-three unique drugs indicated for the treatment of one or more cancers were significantly correlated with CNS disorders' transcriptomic signatures. The central nervous system disorders involved in the associations were AD, BD, and MD.

Cancer-related drugs showing negative correlations with AD, BD, and MD included hormones and drugs targeting hormone receptors, such as raloxifene, an agonist/antagonist of the estrogen receptor, exemestane, an aromatase inhibitor, which acts as an anti-estrogen. Lapatinib, an inhibitor of the Her2/Neu receptor, nilutamide, a non-steroidal antiandrogen, and estradiol. Some chemotherapy agents were also found to be negatively correlated with AD, BD, and MD and included the alkylating agent tiptepa, capecitabine, and 5-fluorouracil, which act as thymidylate synthase inhibitors, and vincristine. Two additional drugs, pazopanib, a tyrosine kinase inhibitor, and thalidomide, presented negative correlations with AD, BD, and MD.

Finally, a number of anti-cancer medications also presented positive correlations with AD, BD, and MD transcriptomic signatures. Those included instances of hormones and hormone receptors (i.e., methyltestosterone, medroxyprogesterone, a hormonal medication of the progestin type, and flutamide, a non-steroidal anti-estrogen), chemotherapy agents, such as alkylating agents (i.e., temozolomide, Chlorambucil, and cyclophosphamide), topoisomerase inhibitors (i.e., idarubicin, doxorubicin, epirubicin, and topotecan), a purine analog (fludarabine), and an RNA polymerase inhibitor (dactinomycin). In addition, everolimus, an MTORC1 inhibitor, was also found to be positively correlated with the differential gene expression profiles of AD, BD, and MD. **Supplementary Appendix 1 Table 48** includes all the information related to cancer drugs correlated with CNS disorder signatures.

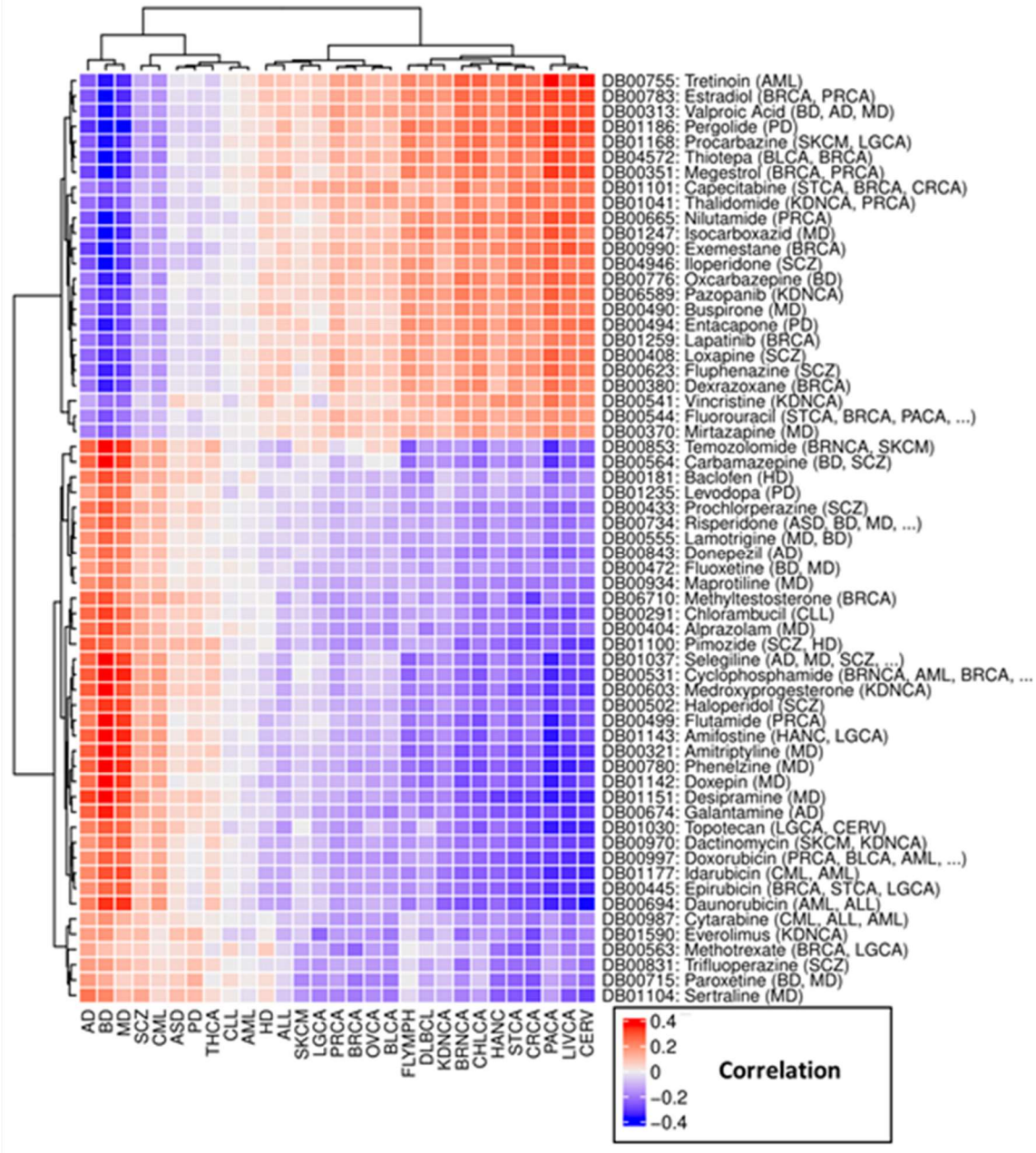

Figure 21: Heatmap depicting the strongest correlations observed between drugs and disorders gene expression profiles. The disease(s) for which each drug is indicated can be found in brackets.

5.4 Discussion

This chapter served to identify several drugs and drug-families that could play a role in modulating CNS disorders and cancer associations. The interactome-based analysis results were found to clearly depend upon the use of relaxed or stringent lists of disease-associated genes and produce a low yield in terms of off-target drug disease-module associations between drugs indicated for the treatment of CNS disorder and cancer modules and vice versa. However, some interesting candidates were extracted from this analysis. Three acetylcholinesterase inhibitor drugs used to treat AD (rivastigmine, tacrine, and donepezil) were associated with the disease module of kidney and breast cancers. In the nervous system, acetylcholinesterase (AChE) rapidly hydrolyzes the neurotransmitter acetylcholine and participates in the impulse transmission termination. In non-neural contexts, AChE is thought to participate in the control of cell proliferation, differentiation, and apoptosis and has been suggested to play a tumor suppressor role [337, 338]. Some cancer types such as leukemia, non-small cell lung cancers, and breast cancers have been reported to present increased activities of AChE, whereas hepatocellular carcinomas, gastric and prostate cancers seem to present reduced activities [338]. It has been shown that there is a clear link between AChE expression and cell cycle progression in hepatoma cell lines, in which inhibition of AChE activity increased cell proliferation, which was associated with a downregulation of p27 and cyclins [337]. Some authors have defined AChE as an apoptotic marker and a promising tumor suppressor [339]. Therefore, the use of acetylcholinesterase inhibitors, such as rivastigmine, could have a cancer-promoting effect in some tumors. However, this seems to be a more complex issue than previously anticipated. Three different isoforms of AChE exist. They differ in their carboxy-terminal sequences and include synaptic AChE (AChE-S), erythrocytic AChE (AChE-E), and read-through AChE (AChE-R). Whereas AChE-S promotes cell death and plays a vital role in the apoptosome's formations, AChE-R positively regulates cell proliferation [340]. Given the dual role of AChE in cell survival and proliferation regulation, some authors have suggested the use of AChE inhibitors as anti-cancer agents for cancers displaying elevated AChE activities [340]. In addition, our analyses showed that cell lines treated using three acetylcholinesterase inhibitors (selegiline, donepezil, and galantamine) presented negative correlations with the differential expression profiles of several cancer types, suggesting that they could have anti-cancer effects. Therefore, acetylcholinesterase inhibitors' role in the reduction of cancer risk observed in AD patients cannot be excluded.

The interactome-based analysis results also pointed towards the effect of breast cancer drugs and their association with the disease-modules of AD, MD, BD, and SCZ. Those drugs included selective estrogen receptor modulators (SERM) as toremifene and tamoxifen, selective estrogen

receptor degraders (SERD) like fulvestrant, and testosterone and derivatives, such as testosterone enanthate, methyltestosterone. In our analyses, toremifene was associated with AD and BD's disease modules and tamoxifen with the disease module of SCZ. Toremifene and the closely related molecule, tamoxifen, can impair mice's learning and memory [341]. Estrogens are known to have a neuroprotective and antioxidant effect and inhibit neurotoxic glutamate and amyloid-beta accumulation, and therefore, to protect against cognitive dysfunction [342]. In fact, it has been observed that postmenopausal women present a higher incidence of AD than controls [343]. SERMs can act both as agonists or antagonists of estrogen receptors in a context-dependent fashion. Some authors have proposed that SERMs could present the same benefits as estrogen treatment but reduce the risks associated with it, like the development of estrogen-dependent tumors [344, 345]. In this respect, SERMs have been shown to display neuroprotective effects by reducing local brain inflammation [346].

In our analyses, tamoxifen was found to be linked to the SCZ disease module. A series of studies have demonstrated that the use of estrogen is useful in schizophrenia patients and to treat the manic phase of bipolar disorder [345, 347, 348]. Estrogen treatment has also been found to reduce psychiatric symptoms in postmenopausal women with schizophrenia and bipolar mania. In addition, tamoxifen has been proven useful in preclinical models of schizophrenia-like symptoms and women with bipolar disorder experiencing episodes of mania [349]. Toremifene was also significantly associated with the disease module of bipolar disorder. In the past, testosterone was the most common form of hormonal therapy in breast cancer. We found that testosterone was associated with the disease module of major depression. Testosterone was a common form of hormonal therapy used for the treatment of breast cancer. Despite being progressively abandoned during the last four decades, some reports have tested its efficacy in refractory metastatic breast cancer, finding significant therapeutic activity [350]. A meta-analysis suggests that testosterone treatment could be effective in reducing depressive symptoms in men [351]. However, a recent randomized control trial examining the effects of low-dose testosterone in women with treatment-resistant depression did not show efficacy compared with placebo [352]. Altogether, this data indicates that medications linked to the hormonal axis could play a role in the modulation of CNS and cancer associations.

Cisplatin is a chemotherapy medication used to treat an extensive collection of cancer types, including BRNCA, BLCA, LGCA, HANC, LGCA, and OVCA. In our analyses, cisplatin was found to present interactome associations with Alzheimer’s disease. There is evidence relating cisplatin treatment with the emergence of dementia-like symptoms. For instance, cisplatin is known to induce cell death in mice's hippocampus through changes in the expression of apoptotic-related genes [353] and accelerated the development of tauopathy signs and the loss of synaptic integrity

[354]. Besides, cognitive deficits are present in up to 75 percent of cancer patients treated with chemotherapy [355].

LINCS L1000 analysis allowed us to identify a set of drugs indicated for the treatment of CNS disorders, which were found to produce opposite patterns of gene expression than those observed in the differential gene expression analysis profiles of several cancer types. The identified sets included sodium channel blockers, antipsychotics, antidepressants, benzodiazepines, and carbidopa, as well as acetylcholinesterase inhibitors.

Different degrees of evidence have been gathered regarding the anti-cancer properties of these compounds. In the case of sodium channel blockers, some studies have reported that patients treated with them present a reduced risk of several tumor types, including colorectal, lung, gastric, and hematological malignancies [356]. In addition, it has been observed that lamotrigine inhibits breast cancer growth in cell lines and rodent models [357]. Furthermore, carbamazepine has been suggested to act as a histone deacetylase inhibitor, reduce the number of cells in S and G2-M phases, and decrease the viability of colon cancer cell lines. However, one study linked the use of carbamazepine and lamotrigine to an increase in the risk of skin cancer melanoma [358].

Antipsychotic drugs have also been linked to anti-cancer properties. Pimozide presents a potent antitumor effect in breast cancer cell lines and xenograft models and reduces the number of lung metastases by blocking the vascular endothelial growth factor [359]. Pimozide also reduces the cell migration and invasion capacity of various cancer cells and inhibits the proliferation of liver and prostate cancer cell lines [360-362]. Prochlorperazine has been found to inhibit cell viability in a concentration-dependent fashion in melanoma cells [363]. Haloperidol presents growth-inhibiting properties in prostate cancer cell line cultures [364].

Moreover, trifluoperazine suppresses tumor growth and brain metastases in triple-negative breast cancer cell lines and mouse xenografts by inducing G0/G1 arrest and apoptosis [365] and also reduces the angiogenic and invasive potential of aggressive cancer cells by modulating the beta-catenin pathway [366]. Finally, chlorpromazine has been shown to present antiproliferative activity in brain tumors and colorectal cancer cell-lines [367]. In addition to the previous antipsychotics, which belong to the dopamine receptor antagonists' family, risperidone, an atypical antipsychotic was also found to be negatively correlated to cancer profiles. Second-generation antipsychotics in general and risperidone, in particular, have been associated with a lower risk of colorectal cancer through the induction of ROS-mediated apoptosis [368].

Different families of antidepressants, including tricyclic (TCA), tetracyclic, selective serotonin reuptake inhibitors (SSRI), and monoamine oxidases (MAO), were the next family of drugs that were found to be negatively correlated with cancer profiles. Among the tricyclic antidepressants, amitriptyline has been proposed as a potential medication for oxidative cancer

therapy. It induces cell toxicity due to an increase of oxidative stress by augmenting ROS levels and causing mitochondrial dysfunction [369] through the inhibition of mitochondrial complex III [370]. Desipramine, another tricyclic antidepressant, induces apoptosis in hepatocellular carcinoma [371]. However, a phase II study for the repositioning of desipramine in small cell lung cancer was discontinued, and no clinal benefits were observed in desipramine treated patients [372]. Maprotiline, a tetracyclic antidepressant, has demonstrated anti-tumor activity in previous drug repositioning studies [373] and has been observed to induce autophagic cell death in Burkitt's lymphoma [374]. Three members of the family of selective serotonin reuptake inhibitors (SSRI) were present among the antidepressant drugs negatively correlated with cancer profiles. SSRIs have been linked to anti-proliferative effects in experimental models. In particular, sertraline has been found to have anti-cancer activity in colon cancer cell lines and cancer xenograft models [375] and has been proposed to target prostate cancer stem cells [376]. Sertraline has also been found to have beneficial effects on tumor growth and cell dissemination in a mouse model of lymphoma. Furthermore, fluoxetine (Prozac) also presents anti-tumor progression effects on hepatocellular carcinoma and non-small cell lung cancer animal models [377]. Selegiline, a member of the monoamine oxidase family, also indicated for the treatment of SCZ, MD, PD, and AD was negatively correlated to several cancer transcriptomic signatures. Selegiline has been shown to induce apoptosis in a melanoma cell line and in acute myelogenous leukemia cells by inhibiting mitochondrial respiration [378, 379]. Downregulation of mitochondrial respiration genes was a shared feature of the transcriptomic alterations found in all the studied CNS disorders. This fact raises questions about the potential pharmacological origin of the mitochondrial dysfunction observed in CNS disorder transcriptomes. The overexpression MAO has been reported in different cancer types [380, 381]. In addition, phenelzine, a non-selective MAO inhibitor, is in a phase II clinical trial for the treatment of recurrent prostate cancer. Despite the fact that experimental evidence seems to point towards an anti-cancer effect of antidepressants, a number of epidemiological studies have not found changes in the risk of cancer according to antidepressant administration. One study found that overall cancer risk was altered in antidepressant users neither for TCA nor for SRRI [382]. Another study found no association between the use of antidepressants (TCA and SRRI) and epithelial ovarian cancer risk [383]. Furthermore, a nation-wide case-control study found no association between any class of antidepressant prescription and the risk of lung cancer [384], and meta-analysis did not find evidence of an association between exposure to SSRIs and colorectal cancer [385]. Exposure to SRRI was not associated with an increase in prostate cancer risk [386] neither in testicular cancer [387]. Finally, a meta-analysis found no association between the use of antidepressants en colorectal cancer morbidity or mortality. In particular, a lack of association was found for SSRIs or SNRIs, but a risk reduction was observed for TCA [388], and a

single study linked SSRI inhibitors to an increased risk of breast cancer. In particular sertraline, and paroxetine presented (RR = 1.58, 95% CI: 1.03-2.41) and (RR = 1.55, 95% CI: 1.00-2.40) (PMID: 14681256). This contrasting evidence suggests that maybe antidepressant treatment has no effect on cancer initiation but could help to reduce the rates of tumor growth in already formed cancers.

Two Acetylcholinesterase inhibitors donepezil, galantamine, were found to generate opposite patterns of transcriptomic changes compared to those observed in several cancer types. Donepezil induces apoptosis in human promyelocytic leukemia cell-lines though the mitochondria-mediated caspase-dependent pathways [389].

Alprazolam, a benzodiazepine used to treat anxiety and indicated for depression, also presented negative correlations with ten cancer types. We found negative correlations between the differential gene expression profile generated by alprazolam and the transcriptomic signatures of 10 cancer types. Surprisingly a 2016 meta-analysis showed that benzodiazepine treatment was associated with an increase in cancer risk in a dose-response manner. Alprazolam has been found to increase the levels of inflammation mediators [390]. A more recent meta-analysis also showed an increase of cancer risk in benzodiazepine treated patients for all-cancer and several site-specific cancers, including breast, ovarian, renal, brain, esophagus, prostate, liver, stomach, pancreatic, and lung cancers, as well as, for skin cancer melanoma [391].

Finally, carbidopa also showed a negative correlation with three cancers. Carbidopa has been found to inhibit pancreatic cancer cell proliferation in vitro and in vivo [392] and prostate cancer xenografted into immunocompromised mice [393].

To a lesser extent, some instances of drugs used for the treatment of CNS disorders belonging to the previously reported families also presented positive correlations with the differential gene expression profiles of several cancer types, which included instances of tetracyclic antidepressants (mirtazapine), anticonvulsant drugs (oxcarbazepine), the serotonin receptor buspirone or the MAO Isocarboxazid, as well as the antipsychotic drugs, loxapine, iloperidone, and fluphenazine. Surprisingly valproic acid, an antiseizure medication known to act as HDAC inhibitor that induces HDAC2 proteasomal degradation leading to cellular differentiation, cell growth arrest, inhibition of angiogenesis, and cell death [394], was found to be positively correlated with the differential gene expression profiles of 11 cancer types.

Several drugs indicated for cancer treatment were found to be correlated with the differential gene expression profiles of AD, BD, or MD. Interestingly, hormonal drugs or drugs acting on hormonal receptors were positively and negatively correlated with those CNS disorders. Among the negatively correlated drugs, we found antiestrogens, agonists/antagonists of the estrogen receptor, estradiol, and antiandrogens, whereas testosterone derivatives and progesterone were found to be positively correlated with AD, BD, and MD profiles. This strengthens the involvement

of this axis in AD, BD, and MD and complements the results found in the interactome-based analysis section. Furthermore, it provides a potential link to explain the comorbidities between those disorders and BRCA and PRCA.

6. General discussion, limitations, future directions and conclussions

6.1 Final discussion, study limitations, and future directions

This manuscript provides a picture of the current knowledge about the epidemiological and molecular associations between CNS disorders and cancer. However, due to data availability limitations, both quantitative and qualitative, this landscape is still far from being completed. Here we discuss the limitations of our work and propose future directions to overcome them.

Despite several of the epidemiological associations reported in this Thesis were found to be well supported by epidemiological evidence, there was an evident lack of adjustment for some cancer-related covariates (e.g., smoking status and BMI) in most of the included observational studies. Therefore, the possibility that the reported associations are driven by uneven distributions of those covariates in the exposed and unexposed cohorts cannot be excluded. Further epidemiological research conducted through careful design will help to elucidate if those factors are responsible for the observed associations. In addition, the number of available studies dedicated to the analysis of each CNS-cancer pair was highly heterogeneous. It is fair to say that most of the potential CNS and site-specific cancer pairs are still poorly studied. This is especially evident in the case of mortality studies. Besides, for some CNS and overall cancer associations (HD and ASD), the number of available studies were found to be also low. We intend to keep conducting epidemiological research following two strategies, first, through the design and implementation of observational studies using the Spanish population's medical records. Second, by conducting updated systematic reviews and meta-analyses as new data is released.

In the context of the transcriptomic associations, we acknowledge some important limitations. The gene expression data included in our analyses provide a picture of a specific time-point of the disease physiopathology. In the case of CNS disorders, working with post-mortem brain tissues gives us a view that is likely biased towards the disease's latest stages. This fact prevents us from characterizing the complex changes that take place in the disorder along the temporal axis. An interesting example of this is the role of the cell cycle in neurodegeneration. Despite there is previous evidence of cell cycle activation in post-mitotic cells preceding neurodegeneration, we observed the downregulation of cell cycle-related processes in both AD and PD. Besides, the transcriptomic datasets employed were derived from tissues that are composed of a heterogeneous set of cell types. Although this approach allowed us to identify the joint involvement of some biological processes in CNS disorders and cancer, such as the immune system, the cell cycle, and the oxidative phosphorylation, it precludes the possibility of clearly understand the changes that operate in specific cell types, as well as to determine if the alterations observed are due to changes in the patterns of gene expression of specific cell types or to differences in tissue

composition. This is a fundamental question in neurodegenerative disorders and cancer where the respective death and proliferation of specific cell types occur and is followed by a reconfiguration of the tissue's cellular architecture.

Furthermore, our analyses compared the differential gene expression profiles of disorders, which, in most instances, take place in different tissues. Supposing that the joint alterations in genes and pathways are the drivers of the reported comorbid associations, we should be able to find them in tissues other than those directly linked to the disease physiopathology. For instance, if the opposite patterns of transcriptomic deregulation observed in oxidative phosphorylation genes in AD and LGCA were to be, to some extent, the drivers of the inverse comorbidity observed at a population level between both disorders, we should expect to find alterations in oxidative phosphorylation processes in lung tissues derived from AD patients. However, little is known regarding the presence of those changes in tissues other than those directly linked to the disease physiopathology, with the exception of transcriptomic studies derived from peripheral blood. Some steps have already been taken in this direction. For instance, it has been argued that PD's pathological manifestations seem to extend beyond brain tissues. In particular, Lewy bodies have been found in both the peripheral and the enteric nervous system of PD patients, as well as changes in dopamine expression and mitochondrial and ubiquitin-proteasome dysfunctions [254]. Increased cell cycle activity has also been observed in immortalized lymphocytes derived from PD patients [254].

The lack of control for potential confounding covariates in differential gene expression and gene co-expression analyses due to data availability and methodological matters constitutes another limitation of our study.

Further research is necessary to shed length on these topics. The analysis of transcriptomic deregulation patterns along the temporal axis would allow us to characterize the characteristic alterations of specific disorders better. The emergence of single-cell RNA-seq studies will allow the characterizations of the specific transcriptomic changes occurring in cell-types, and the analysis of transcriptomic changes in tissues other than those directly linked to disease could strengthen the evidence of the implication of specific molecular alterations as modulators in the comorbid associations. We aim to conduct new transcriptomic-based studies as new data suitable to explore these unanswered questions are available.

A number of inverse comorbid associations detected at a population level between neurodegenerative disorders and site-specific cancers were found to translate into significant opposite patterns of transcriptomic deregulation. Instances of those were the inverse associations reported between AD and LGCA, AD and LIVCA, AD and SKCM, PD and BRCA, PD and LGCA, HD and BRCA, and HD and PRCA in both epidemiological and transcriptomic analyses. On the other hand,

PD patients were found to be at an increased risk of brain cancer, and same-direction significant patterns of transcriptomic deregulation were observed between PD and BRNCA. Instances of significant epidemiological associations that did not translate into altered joint patterns of gene expression were also observed. PD and SKCM constitute an interesting example of the latter. We observed a well-supported risk increase in SKCM incidence in PD patients, which did not translate into significant direct transcriptomic patterns of association.

In general, the lack of complete reliable epidemiological and transcriptomic data covering all CNS and site-specific cancer associations did not allow us to test the predictive power that the presence of significant patterns of transcriptomic deregulation have on epidemiological associations. This should be the object of future research.

The study of genetic correlations between disorders is a promising tool to establish comorbidity's potential biological bases. Two big steps forward are enabling researchers to conduct large scale studies based on the computation of genetic correlations. The first is the availability of specific data repositories that facilitate the searches and direct access to GWAS data, such as the GWAS catalog. The second is linked to the development of methods to compute genetic correlations between disorders using GWAS summary statistics data alone, such as cross-trait LD score regression, which provides a computationally affordable approach compared to competing methods. Using this method, we were able to identify positive genetic correlations between CNS disorders that presented increased risks of different cancer types, including SCZ and BRCA, MD and LGCA, MD and BRCA, and PD and SKCM.

Despite this, we could not obtain GWAS summary statistics for most of the studied site-specific cancers, which are in some cases under restricted access. In those instances, the data access application procedures have proven more challenging and time-consuming than anticipated precluding their use in our analyses. We think that the ongoing transitions towards open access data models will improve the future's access conditions. Therefore, we intend to complete the landscape of the genetic correlations between CNS disorders and cancers as more datasets are released. Furthermore, our genetic correlation analyses are limited to populations of European ancestry. Therefore, these studies should be extended to datasets derived from different populations. The study of the potential role of medications in disease-disease associations can be used to deepen our understanding of comorbidity but also opens the door to the identification of drug repurposing candidates. Drug development is a slow process. The complete journey from drug design until the completion of phase 3 clinical trials can take more than 12 years [395, 396]. In this context, we identified drugs indicated for the treatment of CNS disorders that targeted the disease module of several cancer types and produced transcriptomic alterations that could reverse

those found in some site-specific cancers. A number of these drugs have shown anticancer activity in previous *in vitro* and animal model studies.

Finally, although our study could be used as a tool to generate hypotheses, it needs to be completed through experimental research. The use of animal models of CNS disorders could constitute a valuable approach to establish the molecular bases of comorbidity. This raises questions about which models are best fitted to this end since most of the studied disorders are complex and heterogeneous in nature.

6.2 Conclusions

1) CNS disorders present diverse epidemiological patterns of comorbidity with cancer. Neurodegenerative disorders (AD, PD, and HD) display a reduced risk of subsequent all-cancer incidence and mortality. BD, SCZ, and ASD do not present altered all-cancer incidence patterns compared to controls, whereas MD patients are at increased risk of subsequent all-cancer incidence. In contrast, BD, MD, SCZ, and ASD are associated with an increased risk of all-cancer mortality.
2) The best-supported associations between CNS disorders and site-specific cancers were the following: AD presented a reduced incidence risk of liver and lung cancers. HD patients were found to be at a reduced risk of colorectal and prostate cancer incidence. PD presented reduced incidence risks of bladder, colorectal, larynx, lip and oral cavity, and lung cancers, whereas the incidence risks of brain and melanoma were found to be increased in PD patients. Colorectal, lung, pancreatic, and stomach cancer mortality risks were reduced in PD patients compared to controls. MD patients are at a higher risk of brain, lung, and pancreatic cancer than controls, and depressed women at a higher risk of breast cancer incidence and mortality. BD patients were found at an increased risk of stomach cancer. Women with SZC presented an increased risk of uterine and breast cancer incidence and were also found to be at an increased risk of breast cancer death, whereas SCZ men present a reduced risk of prostate cancer. The risks of mortality due to colorectal, liver, lung, and pancreatic cancer are also increased in SCZ patients compared to controls. Melanoma, nasopharynx, and thyroid cancer incidences were found to be reduced in SCZ patients.
3) The risk of smoking-related cancers was found to be reduced in PD patients, but not the risk of cancers not linked to smoking, whereas MD presented an increased risk in smoking-related cancers but not in the smoking-unrelated cancers. This observation suggests the role of smoking as a potential modulator in some of the reported CNS and cancer associations. However, this is an indirect way of measuring smoking's impact, therefore these results should be confirmed by further epidemiological research which should take into account smoking status information by design or as an adjustment factor. In addition, gender-based differences in all-cancer incidence were observed in the case of SCZ. Men presented a significant reduction, whereas women presented a non-significant trend towards an increase in cancer risk. This could be due to the significant increase in the risk of breast cancer incidence and mortality observed in SCZ women and the significant reduction in prostate cancer observed in PD patients.

4) The reported meta-analysis results should be taken with caution since some of them are based on meta-analyses carried out using a small number of observational studies. In addition, important confounding variables, such as smoking status, are not accounted for in the vast majority of cases. Other potential sources of bias, such as ascertainment bias or diagnostic overshadowing, cannot be excluded either. There is considerable heterogeneity in the number of published studies reporting cancer associations for the different CNS disorders. ASD and HD are the poorest studied disorders in the context of their associations with cancer. In the case of mortality studies, site-specific cancer information was very scarce.
5) Neurodegenerative disorders (AD, PD, and HD) and ASD presented a substantial amount of differentially expressed genes, whereas BD, MD, and SCZ did not. Cancers presented large amounts of differentially expressed genes when compared to healthy control tissues in both the array and RNA-seq-based analyses.
6) Several significant patterns of joint transcriptomic deregulation were found between CNS disorders and cancers using array-based studies. Some of the associations were further validated using alternative RNA-seq cancer datasets. AD presented direct transcriptomic associations with BRNCA and inverse with BLCA, BRCA, LGCA, LIVCA, and PRCA. PD was found to present direct transcriptomic associations with BRNCA and KDNCA and inverse with BRCA, LGCA, and PRCA. HD was directly associated with BRNCA, KDNCA, and inversely with BRCA, LGCA. Finally, ASD was found to present significant direct patterns of transcriptomic deregulation with BRNCA and KDNCA.
7) Overrepresentation, GSEA, and consensus co-expression module overlap analyses suggest the joint involvement of alterations in biological processes in CNS and cancer, which included diverse instances of immune system-related pathways, as well as others, such as P53 signaling, unfolded protein response, mTORC1 signaling, DNA repair, cell cycle, apoptosis, and the oxidative phosphorylation, among others.
8) CNS disorders and cancers did not present significant interactome-based overlaps. In contrast, significant genetic correlations were observed. ASD was found to present significant negative genetic correlations with PRCA and BRCA. BD was positively correlated with BRCA, whereas MD did it with BRCA and LGCA. PD presented positive genetic correlations with PRCA and SKCM. Finally, SCZ was also found to present significant positive genetic correlations with BRCA and OVCA. These observations suggest that shared variants could jointly influence the susceptibility to CNS and cancers.
9) Several drugs indicated for the treatment of CNS disorders were placed close to the disease module of cancers and produced gene expression changes that could mimic or reverse

those found in different cancer types, suggesting their potential role in modulating CNS disorders and cancer associations. Instances span drugs from several families, including anticonvulsants, antipsychotics, antidepressants, and benzodiazepines, among others. Drugs indicated for cancer also have the capacity to produce transcriptomic changes that mimic or reverse the transcriptomic signatures of CNS disorders.

7. Bibliography

1. Catala-Lopez, F., et al., *Cancer and central nervous system disorders: protocol for an umbrella review of systematic reviews and updated meta-analyses of observational studies.* Syst Rev, 2017. **6**(1): p. 69.
2. Feinstein, A.R., *The Pre-Therapeutic Classification of Co-Morbidity in Chronic Disease.* J Chronic Dis, 1970. **23**(7): p. 455-68.
3. Valderas, J.M., et al., *Defining comorbidity: implications for understanding health and health services.* Ann Fam Med, 2009. **7**(4): p. 357-63.
4. Charlson, M.E., et al., *A new method of classifying prognostic comorbidity in longitudinal studies: development and validation.* J Chronic Dis, 1987. **40**(5): p. 373-83.
5. Linn, B.S., M.W. Linn, and L. Gurel, *Cumulative illness rating scale.* J Am Geriatr Soc, 1968. **16**(5): p. 622-6.
6. Greenfield, S., et al., *The importance of co-existent disease in the occurrence of postoperative complications and one-year recovery in patients undergoing total hip replacement. Comorbidity and outcomes after hip replacement.* Med Care, 1993. **31**(2): p. 141-54.
7. Safford, M.M., J.J. Allison, and C.I. Kiefe, *Patient complexity: more than comorbidity. the vector model of complexity.* J Gen Intern Med, 2007. **22 Suppl 3**: p. 382-90.
8. Tabares-Seisdedos, R. and J.L. Rubenstein, *Inverse cancer comorbidity: a serendipitous opportunity to gain insight into CNS disorders.* Nat Rev Neurosci, 2013. **14**(4): p. 293-304.
9. Tabares-Seisdedos, R., et al., *No paradox, no progress: inverse cancer comorbidity in people with other complex diseases.* Lancet Oncol, 2011. **12**(6): p. 604-8.
10. Ording, A.G. and H.T. Sorensen, *Concepts of comorbidities, multiple morbidities, complications, and their clinical epidemiologic analogs.* Clin Epidemiol, 2013. **5**: p. 199-203.
11. Cohn, A., A.M. Sofia, and S.S. Kupfer, *Type 1 diabetes and celiac disease: clinical overlap and new insights into disease pathogenesis.* Curr Diab Rep, 2014. **14**(8): p. 517.
12. Xu, W., et al., *Use of antidepressants and risk of colorectal cancer: a nested case-control study.* Lancet Oncol, 2006. **7**(4): p. 301-8.
13. Skrott, Z., et al., *Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4.* Nature, 2017. **552**(7684): p. 194-199.
14. Gulbinat, W., et al., *Cancer incidence of schizophrenic patients. Results of record linkage studies in three countries.* Br J Psychiatry Suppl, 1992(18): p. 75-83.
15. Dupont, A., et al., *Incidence of cancer in patients diagnosed as schizophrenic in Denmark.* Elsevier Science, 1986(Psychiatric Case Registries in Public Health Amsterdam): p. 229–239.
16. Shekelle, R.B., et al., *Psychological depression and 17-year risk of death from cancer.* Psychosom Med, 1981. **43**(2): p. 117-25.
17. Faraoni, D. and S.T. Schaefer, *Randomized controlled trials vs. observational studies: why not just live together?* BMC Anesthesiol, 2016. **16**(1): p. 102.
18. Elwood, P., *The first randomized trial of aspirin for heart attack and the advent of systematic overviews of trials.* J R Soc Med, 2006. **99**(11): p. 586-8.
19. Egger, M., G.D. Smith, and D.G. Altman, *Systematic Reviews in Health Care*. 2001.
20. McGee, R., S. Williams, and M. Elwood, *Depression and the development of cancer: a meta-analysis.* Soc Sci Med, 1994. **38**(1): p. 187-92.
21. Catala-Lopez, F., et al., *Inverse and direct cancer comorbidity in people with central nervous system disorders: a meta-analysis of cancer incidence in 577,013 participants of 50 observational studies.* Psychother Psychosom, 2014. **83**(2): p. 89-105.
22. Gormley, M., *The first 'molecular disease': a story of Linus Pauling, the intellectual patron.* Endeavour, 2007. **31**(2): p. 71-7.
23. Sansonetti, P.J., *Moving molecular medicine.* EMBO Mol Med, 2017. **9**(4): p. 395.

24. Gusella, J.F., et al., *A polymorphic DNA marker genetically linked to Huntington's disease.* Nature, 1983. **306**(5940): p. 234-8.
25. Koenig, M., et al., *Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals.* Cell, 1987. **50**(3): p. 509-17.
26. Macdonald, M., *A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes.* Cell, 1993. **72**(6): p. 971-983.
27. Ng, S.B., et al., *Exome sequencing identifies the cause of a mendelian disorder.* Nat Genet, 2010. **42**(1): p. 30-5.
28. International HapMap, C., *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
29. Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.
30. Zuk, O., et al., *The mystery of missing heritability: Genetic interactions create phantom heritability.* Proc Natl Acad Sci U S A, 2012. **109**(4): p. 1193-8.
31. Goh, K.I., et al., *The human disease network.* Proc Natl Acad Sci U S A, 2007. **104**(21): p. 8685-90.
32. Menche, J., et al., *Disease networks. Uncovering disease-disease relationships through the incomplete interactome.* Science, 2015. **347**(6224): p. 1257601.
33. Yildirim, M.A., et al., *Drug-target network.* Nat Biotechnol, 2007. **25**(10): p. 1119-26.
34. Ibanez, K., et al., *Molecular evidence for the inverse comorbidity between central nervous system disorders and cancers detected by transcriptomic meta-analyses.* PLoS Genet, 2014. **10**(2): p. e1004173.
35. Sanchez-Valle, J., et al., *A molecular hypothesis to explain direct and inverse co-morbidities between Alzheimer's Disease, Glioblastoma and Lung cancer.* Sci Rep, 2017. **7**(1): p. 4474.
36. Choi, J.K., et al., *Combining multiple microarray studies and modeling interstudy variation.* Bioinformatics, 2003. **19 Suppl 1**: p. i84-90.
37. Frain, L., et al., *Association of cancer and Alzheimer's disease risk in a national cohort of veterans.* Alzheimers Dement, 2017. **13**(12): p. 1364-1370.
38. Feng, Y.A., et al., *Investigating the genetic relationship between Alzheimer's disease and cancer using GWAS summary statistics.* Hum Genet, 2017. **136**(10): p. 1341-1351.
39. Burns, A. and S. Iliffe, *Alzheimer's disease.* BMJ, 2009. **338**: p. b158.
40. Selkoe, D.J., *Translating cell biology into therapeutic advances in Alzheimer's disease.* Nature, 1999. **399**(6738 Suppl): p. A23-31.
41. Yankner, B.A., L.K. Duffy, and D.A. Kirschner, *Neurotrophic and neurotoxic effects of amyloid beta protein: reversal by tachykinin neuropeptides.* Science, 1990. **250**(4978): p. 279-82.
42. Tuchman, R., S.L. Moshe, and I. Rapin, *Convulsing toward the pathophysiology of autism.* Brain Dev, 2009. **31**(2): p. 95-103.
43. Disease, G.B.D., I. Injury, and C. Prevalence, *Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015.* Lancet, 2016. **388**(10053): p. 1545-1602.
44. Geschwind, D.H., *Advances in autism.* Annu Rev Med, 2009. **60**: p. 367-80.
45. Fernandez, B.A. and S.W. Scherer, *Syndromic autism spectrum disorders: moving from a clinically defined to a molecularly defined approach.* Dialogues Clin Neurosci, 2017. **19**(4): p. 353-371.
46. Geschwind, D.H. and M.W. State, *Gene hunting in autism spectrum disorder: on the path to precision medicine.* Lancet Neurol, 2015. **14**(11): p. 1109-20.
47. Yoo, H., *Genetics of Autism Spectrum Disorder: Current Status and Possible Clinical Applications.* Exp Neurobiol, 2015. **24**(4): p. 257-72.
48. Hsiao, E.Y., *Immune dysregulation in autism spectrum disorder.* Int Rev Neurobiol, 2013. **113**: p. 269-302.

49. Onore, C., M. Careaga, and P. Ashwood, *The role of immune dysfunction in the pathophysiology of autism.* Brain Behav Immun, 2012. **26**(3): p. 383-92.
50. Patterson, P.H., *Maternal infection and immune involvement in autism.* Trends Mol Med, 2011. **17**(7): p. 389-94.
51. Conus, P., et al., *The proximal prodrome to first episode mania--a new target for early intervention.* Bipolar Disord, 2008. **10**(5): p. 555-65.
52. Duffy, A., et al., *Early stages in the development of bipolar disorder.* J Affect Disord, 2010. **121**(1-2): p. 127-35.
53. Merikangas, K.R., et al., *Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative.* Arch Gen Psychiatry, 2011. **68**(3): p. 241-51.
54. Edvardsen, J., et al., *Heritability of bipolar spectrum disorders. Unity or heterogeneity?* J Affect Disord, 2008. **106**(3): p. 229-40.
55. Salvadore, G., et al., *The neurobiology of the switch process in bipolar disorder: a review.* J Clin Psychiatry, 2010. **71**(11): p. 1488-501.
56. Lahera, G., N. Freund, and J. Saiz-Ruiz, *Salience and dysregulation of the dopaminergic system.* Rev Psiquiatr Salud Ment, 2013. **6**(1): p. 45-51.
57. Michael, N., et al., *Acute mania is accompanied by elevated glutamate/glutamine levels within the left dorsolateral prefrontal cortex.* Psychopharmacology (Berl), 2003. **168**(3): p. 344-6.
58. Schizophrenia Working Group of the Psychiatric Genomics, C., *Biological insights from 108 schizophrenia-associated genetic loci.* Nature, 2014. **511**(7510): p. 421-7.
59. consortium, C., *Sparse whole-genome sequencing identifies two loci for major depressive disorder.* Nature, 2015. **523**(7562): p. 588-91.
60. Hyde, C.L., et al., *Identification of 15 genetic loci associated with risk of major depression in individuals of European descent.* Nat Genet, 2016. **48**(9): p. 1031-6.
61. Segurado, R., et al., *Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder.* Am J Hum Genet, 2003. **73**(1): p. 49-62.
62. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. 2013, Washington, DC.
63. Kessler, R.C., et al., *The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R).* JAMA, 2003. **289**(23): p. 3095-105.
64. *World Health Organization*. Available from: http://www.who.int/mediacentre/factsheets/fs369/en/.
65. Keller, M.B., et al., *Optimizing outcomes in depression: focus on antidepressant compliance.* Int Clin Psychopharmacol, 2002. **17**(6): p. 265-71.
66. Jeon, S.W. and Y.K. Kim, *Molecular Neurobiology and Promising New Treatment in Depression.* Int J Mol Sci, 2016. **17**(3): p. 381.
67. Walker, F.O., *Huntington's disease.* Lancet, 2007. **369**(9557): p. 218-28.
68. Frank, S., *Treatment of Huntington's disease.* Neurotherapeutics, 2014. **11**(1): p. 153-60.
69. Pringsheim, T., et al., *The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis.* Mov Disord, 2012. **27**(9): p. 1083-91.
70. Dayalu, P. and R.L. Albin, *Huntington disease: pathogenesis and treatment.* Neurol Clin, 2015. **33**(1): p. 101-14.
71. Lobsiger, C.S. and D.W. Cleveland, *Glial cells as intrinsic components of non-cell-autonomous neurodegenerative disease.* Nat Neurosci, 2007. **10**(11): p. 1355-60.
72. Liu, Z., et al., *Oxidative Stress in Neurodegenerative Diseases: From Molecular Mechanisms to Clinical Applications.* Oxid Med Cell Longev, 2017. **2017**: p. 2525967.
73. Cuenca, L., et al., *Parkinson's disease: a short story of 200 years.* Histol Histopathol, 2019. **34**(6): p. 573-591.
74. Goetz, C.G., et al., *Gender-, age-, and race/ethnicity-based differential item functioning analysis of the movement disorder society-sponsored revision of the Unified Parkinson's disease rating scale.* Mov Disord, 2016. **31**(12): p. 1865-1873.

75. Linazasoro, G., *Classical Parkinson disease versus Parkinson complex--reflections against staging and in favour of heterogeneity.* Eur J Neurol, 2007. **14**(7): p. 721-8.
76. Klein, C. and A. Westenberger, *Genetics of Parkinson's disease.* Cold Spring Harb Perspect Med, 2012. **2**(1): p. a008888.
77. Kalia, L.V. and S.K. Kalia, *alpha-Synuclein and Lewy pathology in Parkinson's disease.* Curr Opin Neurol, 2015. **28**(4): p. 375-81.
78. Graybiel, A.M., *The basal ganglia: learning new tricks and loving it.* Curr Opin Neurobiol, 2005. **15**(6): p. 638-44.
79. Sullivan, P.F., K.S. Kendler, and M.C. Neale, *Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies.* Arch Gen Psychiatry, 2003. **60**(12): p. 1187-92.
80. Psychiatric, G.C.S.C., *A framework for interpreting genome-wide association studies of psychiatric disorders.* Mol Psychiatry, 2009. **14**(1): p. 10-7.
81. Rees, E., M.C. O'Donovan, and M.J. Owen, *Genetics of schizophrenia.* Current Opinion in Behavioral Sciences, 2015. **2**: p. 8-14.
82. Global Burden of Disease Cancer, C., et al., *Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study.* JAMA Oncol, 2019. **5**(12): p. 1749-1768.
83. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
84. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
85. Ahn, H.K., et al., *Risk of cancer among patients with depressive disorder: a meta-analysis and implications.* Psychooncology, 2016. **25**(12): p. 1393-1399.
86. Catts, V.S., et al., *Cancer incidence in patients with schizophrenia and their first-degree relatives - a meta-analysis.* Acta Psychiatr Scand, 2008. **117**(5): p. 323-36.
87. Kilian, R., et al., *Health behavior in psychiatric in-patients compared with a German general population sample.* Acta Psychiatr Scand, 2006. **114**(4): p. 242-8.
88. McIntyre, R.S., et al., *Obesity in bipolar disorder and major depressive disorder: results from a national community health survey on mental health and well-being.* Can J Psychiatry, 2006. **51**(5): p. 274-80.
89. Strine, T.W., et al., *The association of depression and anxiety with obesity and unhealthy behaviors among community-dwelling US adults.* Gen Hosp Psychiatry, 2008. **30**(2): p. 127-37.
90. Beydoun, M.A., et al., *Role of depressive symptoms in explaining socioeconomic status disparities in dietary quality and central adiposity among US adults: a structural equation modeling approach.* Am J Clin Nutr, 2009. **90**(4): p. 1084-95.
91. Lord, O., D. Malone, and A.J. Mitchell, *Receipt of preventive medical care and medical screening for patients with mental illness: a comparative analysis.* Gen Hosp Psychiatry, 2010. **32**(5): p. 519-43.
92. Viron, M.J. and T.A. Stern, *The impact of serious mental illness on health and healthcare.* Psychosomatics, 2010. **51**(6): p. 458-65.
93. Boyd, C.A., et al., *The effect of depression on stage at diagnosis, treatment, and survival in pancreatic adenocarcinoma.* Surgery, 2012. **152**(3): p. 403-13.
94. Kisely, S., E. Crowe, and D. Lawrence, *Cancer-related mortality in people with mental illness.* JAMA Psychiatry, 2013. **70**(2): p. 209-17.
95. Wang, P.S., et al., *Dopamine antagonists and the development of breast cancer.* Arch Gen Psychiatry, 2002. **59**(12): p. 1147-54.
96. Fiala, K.H., J. Whetteckey, and B.V. Manyam, *Malignant melanoma and levodopa in Parkinson's disease: causality or coincidence?* Parkinsonism Relat Disord, 2003. **9**(6): p. 321-7.

97. Collaborators, G.B.D.C.o.D., *Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017.* Lancet, 2018. **392**(10159): p. 1736-1788.
98. Catala-Lopez, F., et al., *The increasing burden of mental and neurological disorders.* Eur Neuropsychopharmacol, 2013. **23**(11): p. 1337-9.
99. DALYs, G.B.D. and H. Collaborators, *Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015.* Lancet, 2016. **388**(10053): p. 1603-1658.
100. Atun, R., *Transitioning health systems for multimorbidity.* Lancet, 2015. **386**(9995): p. 721-2.
101. Barnett, K., et al., *Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study.* Lancet, 2012. **380**(9836): p. 37-43.
102. Olsen, J.H., et al., *Atypical cancer pattern in patients with Parkinson's disease.* Br J Cancer, 2005. **92**(1): p. 201-5.
103. Rugbjerg, K., et al., *Malignant melanoma, breast cancer and other cancers in patients with Parkinson's disease.* Int J Cancer, 2012. **131**(8): p. 1904-11.
104. Fois, A.F., et al., *Cancer in patients with motor neuron disease, multiple sclerosis and Parkinson's disease: record linkage studies.* J Neurol Neurosurg Psychiatry, 2010. **81**(2): p. 215-21.
105. Bushe, C.J., et al., *Schizophrenia and breast cancer incidence: a systematic review of clinical studies.* Schizophr Res, 2009. **114**(1-3): p. 6-16.
106. Liu, R., et al., *Meta-analysis of the relationship between Parkinson disease and melanoma.* Neurology, 2011. **76**(23): p. 2002-9.
107. Roe, C.M., et al., *Alzheimer disease and cancer.* Neurology, 2005. **64**(5): p. 895-8.
108. Roe, C.M., et al., *Cancer linked to Alzheimer disease but not vascular dementia.* Neurology, 2010. **74**(2): p. 106-12.
109. Driver, J.A., et al., *Inverse association between cancer and Alzheimer's disease: results from the Framingham Heart Study.* BMJ, 2012. **344**: p. e1442.
110. Musicco, M., et al., *Inverse occurrence of cancer and Alzheimer disease: a population-based incidence study.* Neurology, 2013. **81**(4): p. 322-8.
111. Zhang, Q., et al., *Inverse relationship between cancer and Alzheimer's disease: a systemic review meta-analysis.* Neurol Sci, 2015. **36**(11): p. 1987-94.
112. Driver, J.A., et al., *A prospective cohort study of cancer incidence following the diagnosis of Parkinson's disease.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(6): p. 1260-5.
113. Elbaz, A., et al., *Risk of cancer after the diagnosis of Parkinson's disease: a historical cohort study.* Mov Disord, 2005. **20**(6): p. 719-25.
114. Bajaj, A., J.A. Driver, and E.S. Schernhammer, *Parkinson's disease and cancer risk: a systematic review and meta-analysis.* Cancer Causes Control, 2010. **21**(5): p. 697-707.
115. Shamseer, L., et al., *Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation.* BMJ, 2015. **350**: p. g7647.
116. Moher, D., et al., *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.* PLoS Med, 2009. **6**(7): p. e1000097.
117. Stroup, D.F., et al., *Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group.* JAMA, 2000. **283**(15): p. 2008-12.
118. Sampson, M., et al., *Surveillance search techniques identified the need to update systematic reviews.* J Clin Epidemiol, 2008. **61**(8): p. 755-62.
119. Wells, G., *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses.* 2013.
120. DerSimonian, R. and N. Laird, *Meta-analysis in clinical trials.* Control Clin Trials, 1986. **7**(3): p. 177-88.

121. Cochran, W.G., *The Combination of Estimates from Different Experiments.* Biometrics, 1954. **10**(1).
122. Viechtbauer, W., *Conducting Meta-Analyses inRwith themetaforPackage.* Journal of Statistical Software, 2010. **36**(3).
123. Viechtbauer, W., *Publication bias in meta-analysis: Prevention, assessment and adjustments.* Psychometrika, 2007. **72**(2): p. 269-271.
124. Egger, M., et al., *Bias in meta-analysis detected by a simple, graphical test.* BMJ, 1997. **315**(7109): p. 629-34.
125. Begg, C.B. and M. Mazumdar, *Operating characteristics of a rank correlation test for publication bias.* Biometrics, 1994. **50**(4): p. 1088-101.
126. Saku, M., et al., *Mortality in psychiatric patients, with a specific focus on cancer mortality associated with schizophrenia.* Int J Epidemiol, 1995. **24**(2): p. 366-72.
127. Laursen, T.M., et al., *Increased mortality among patients admitted with major psychiatric disorders: a register-based study comparing mortality in unipolar depressive disorder, bipolar affective disorder, schizoaffective disorder, and schizophrenia.* J Clin Psychiatry, 2007. **68**(6): p. 899-907.
128. Batty, G.D., et al., *Impact of mental health problems on case fatality in male cancer patients.* Br J Cancer, 2012. **106**(11): p. 1842-5.
129. Ajetunmobi, O., et al., *Early death in those previously hospitalised for mental healthcare in Scotland: a nationwide cohort study, 1986-2010.* BMJ Open, 2013. **3**(7).
130. Guan, N.C., et al., *Cancer mortality in patients with psychiatric diagnoses: a higher hazard of cancer death does not lead to a higher cumulative risk of dying from cancer.* Soc Psychiatry Psychiatr Epidemiol, 2013. **48**(8): p. 1289-95.
131. Almeida, O.P., et al., *Mortality among people with severe mental disorders who reach old age: a longitudinal study of a community-representative sample of 37,892 men.* PLoS One, 2014. **9**(10): p. e111882.
132. Angst, F., et al., *Mortality of patients with mood disorders: follow-up over 34-38 years.* J Affect Disord, 2002. **68**(2-3): p. 167-81.
133. Hung, Y.N., et al., *Cancer incidence in people with affective disorder: nationwide cohort study in Taiwan, 1997-2010.* Br J Psychiatry, 2014. **205**(3): p. 183-8.
134. Hippisley-Cox, J., et al., *Risk of malignancy in patients with schizophrenia or bipolar disorder: nested case-control study.* Arch Gen Psychiatry, 2007. **64**(12): p. 1368-76.
135. Hiroeh, U., et al., *Deaths from natural causes in people with mental illness: a cohort study.* J Psychosom Res, 2008. **64**(3): p. 275-83.
136. Daumit, G.L., et al., *Pattern of mortality in a sample of Maryland residents with severe mental illness.* Psychiatry Res, 2010. **176**(2-3): p. 242-5.
137. McGinty, E.E., et al., *Cancer incidence in a sample of Maryland residents with serious mental illness.* Psychiatr Serv, 2012. **63**(7): p. 714-7.
138. Castagnini, A., L. Foldager, and A. Bertelsen, *Excess mortality of acute and transient psychotic disorders: comparison with bipolar affective disorder and schizophrenia.* Acta Psychiatr Scand, 2013. **128**(5): p. 370-5.
139. Crump, C., et al., *Comorbidities and mortality in persons with schizophrenia: a Swedish national cohort study.* Am J Psychiatry, 2013. **170**(3): p. 324-33.
140. Osborn, D.P., et al., *Relative incidence of common cancers in people with severe mental illness. Cohort study in the United Kingdom THIN primary care database.* Schizophr Res, 2013. **143**(1): p. 44-9.
141. Lin, G.M., et al., *Cancer incidence in patients with schizophrenia or bipolar disorder: a nationwide population-based study in Taiwan, 1997-2009.* Schizophr Bull, 2013. **39**(2): p. 407-16.
142. Knekt, P., et al., *Elevated lung cancer risk among persons with depressed mood.* Am J Epidemiol, 1996. **144**(12): p. 1096-103.

143. Lawrence, D., et al., *Excess cancer mortality in Western Australian psychiatric patients due to higher case fatality rates.* Acta Psychiatr Scand, 2000. **101**(5): p. 382-8.
144. Kisely, S., S. Forsyth, and D. Lawrence, *Why do psychiatric patients have higher cancer mortality rates when cancer incidence is the same or lower?* Aust N Z J Psychiatry, 2016. **50**(3): p. 254-63.
145. Freedman, D.M., et al., *Associations between cancer and Parkinson's disease in U.S. elderly adults.* Int J Epidemiol, 2016. **45**(3): p. 741-51.
146. Chen, Y.J., et al., *Does Parkinsons Disease Dementia Reduce Cancer Risk more than Alzheimers Disease Alone?* Neuropsychiatry, 2018. **07**(04).
147. Freedman, D.M., et al., *Associations between cancer and Alzheimer's disease in a U.S. Medicare population.* Cancer Med, 2016. **5**(10): p. 2965-2976.
148. Yarchoan, M., et al., *Association of Cancer History with Alzheimer's Disease Dementia and Neuropathology.* J Alzheimers Dis, 2017. **56**(2): p. 699-706.
149. Schmidt, S.A., et al., *Non-melanoma skin cancer and risk of Alzheimer's disease and all-cause dementia.* PLoS One, 2017. **12**(2): p. e0171527.
150. Bowles, E.J.A., et al., *Risk of Alzheimer's disease or dementia following a cancer diagnosis.* PLoS One, 2017. **12**(6): p. e0179857.
151. Hodgson, K.D., et al., *A meta-analysis of the effects of chemotherapy on cognition in patients with cancer.* Cancer Treat Rev, 2013. **39**(3): p. 297-304.
152. Surma-aho, O., et al., *Adverse long-term effects of brain radiotherapy in adult low-grade glioma patients.* Neurology, 2001. **56**(10): p. 1285-90.
153. Olsen, J.H., S. Friis, and K. Frederiksen, *Malignant melanoma and other types of cancer preceding Parkinson disease.* Epidemiology, 2006. **17**(5): p. 582-7.
154. Levine, R.L., J.C. Jones, and N. Bee, *Stroke and Parkinson's disease.* Stroke, 1992. **23**(6): p. 839-42.
155. D'Amelio, M., et al., *Tumor diagnosis preceding Parkinson's disease: a case-control study.* Mov Disord, 2004. **19**(7): p. 807-811.
156. Driver, J.A., et al., *Prospective case-control study of nonfatal cancer preceding the diagnosis of Parkinson's disease.* Cancer Causes Control, 2007. **18**(7): p. 705-11.
157. Elbaz, A., et al., *Nonfatal cancer preceding Parkinson's disease: a case-control study.* Epidemiology, 2002. **13**(2): p. 157-64.
158. Powers, K.M., et al., *Diabetes, smoking, and other medical conditions in relation to Parkinson's disease risk.* Parkinsonism Relat Disord, 2006. **12**(3): p. 185-9.
159. Rajput, A.H., et al., *A case-control study of smoking habits, dementia, and other illnesses in idiopathic Parkinson's disease.* Neurology, 1987. **37**(2): p. 226-32.
160. Coarelli, G., et al., *Low cancer prevalence in polyglutamine expansion diseases.* Neurology, 2017. **88**(12): p. 1114-1119.
161. Takahashi, K., et al., *Prolyl isomerase, Pin1: new findings of post-translational modifications and physiological substrates in cancer, asthma and Alzheimer's disease.* Cell Mol Life Sci, 2008. **65**(3): p. 359-75.
162. Liou, Y.C., et al., *Role of the prolyl isomerase Pin1 in protecting against age-dependent neurodegeneration.* Nature, 2003. **424**(6948): p. 556-61.
163. Driver, J.A. and K.P. Lu, *Pin1: a new genetic link between Alzheimer's disease, cancer and aging.* Curr Aging Sci, 2010. **3**(3): p. 158-65.
164. Ganguli, M., *Cancer and Dementia: It's Complicated.* Alzheimer Dis Assoc Disord, 2015. **29**(2): p. 177-82.
165. Majd, S., J. Power, and Z. Majd, *Alzheimer's Disease and Cancer: When Two Monsters Cannot Be Together.* Front Neurosci, 2019. **13**: p. 155.
166. Shepherd, J.E., *Effects of estrogen on congnition mood, and degenerative brain diseases.* J Am Pharm Assoc (Wash), 2001. **41**(2): p. 221-8.

167. Muti, P., et al., *Estrogen metabolism and risk of breast cancer: a prospective study of the 2:16alpha-hydroxyestrone ratio in premenopausal and postmenopausal women.* Epidemiology, 2000. **11**(6): p. 635-40.
168. Francis, P.T., et al., *The cholinergic hypothesis of Alzheimer's disease: a review of progress.* J Neurol Neurosurg Psychiatry, 1999. **66**(2): p. 137-47.
169. Lau, J.K., et al., *Inhibition of cholinergic signaling causes apoptosis in human bronchioalveolar carcinoma.* Cancer Res, 2013. **73**(4): p. 1328-39.
170. Behrens, M.I., C. Lendon, and C.M. Roe, *A common biological mechanism in cancer and Alzheimer's disease?* Curr Alzheimer Res, 2009. **6**(3): p. 196-204.
171. Schernhammer, E., H. Chen, and B. Ritz, *Circulating melatonin levels: possible link between Parkinson's disease and cancer risk?* Cancer Causes Control, 2006. **17**(4): p. 577-82.
172. Tan, L.C., et al., *Dietary cholesterol, fats and risk of Parkinson's disease in the Singapore Chinese Health Study.* J Neurol Neurosurg Psychiatry, 2016. **87**(1): p. 86-92.
173. Sorensen, S.A., K. Fenger, and J.H. Olsen, *Significantly lower incidence of cancer among patients with Huntington disease: An apoptotic effect of an expanded polyglutamine tract?* Cancer, 1999. **86**(7): p. 1342-6.
174. Murmann, A.E., et al., *Small interfering RNAs based on huntingtin trinucleotide repeats are highly toxic to cancer cells.* EMBO Rep, 2018. **19**(3).
175. White, M.C., et al., *Age and cancer risk: a potentially modifiable relationship.* Am J Prev Med, 2014. **46**(3 Suppl 1): p. S7-15.
176. Guerreiro, R. and J. Bras, *The age factor in Alzheimer's disease.* Genome Med, 2015. **7**: p. 106.
177. Ganguli, M., et al., *Alzheimer disease and mortality: a 15-year epidemiological study.* Arch Neurol, 2005. **62**(5): p. 779-84.
178. Jagger, C., M. Clarke, and A. Stone, *Predictors of survival with Alzheimer's disease: a community-based study.* Psychol Med, 1995. **25**(1): p. 171-7.
179. Paradise, M., et al., *Prediction of survival in Alzheimer's disease--the LASER-AD longitudinal study.* Int J Geriatr Psychiatry, 2009. **24**(7): p. 739-47.
180. Tsai, P.H., et al., *Survival of ethnic Chinese with Alzheimer's disease: a 5-year longitudinal study in Taiwan.* J Geriatr Psychiatry Neurol, 2007. **20**(3): p. 172-7.
181. van Dijk, P.T., D.W. Dippel, and J.D. Habbema, *Survival of patients with dementia.* J Am Geriatr Soc, 1991. **39**(6): p. 603-10.
182. Zanetti, O., S.B. Solerte, and F. Cantoni, *Life expectancy in Alzheimer's disease (AD).* Arch Gerontol Geriatr, 2009. **49 Suppl 1**: p. 237-43.
183. Hernan, M.A., et al., *A meta-analysis of coffee drinking, cigarette smoking, and the risk of Parkinson's disease.* Ann Neurol, 2002. **52**(3): p. 276-84.
184. Ou, S.M., et al., *Does Alzheimer's disease protect against cancers? A nationwide population-based study.* Neuroepidemiology, 2013. **40**(1): p. 42-9.
185. Akushevich, I., et al., *Morbidity risks among older adults with pre-existing age-related diseases.* Exp Gerontol, 2013. **48**(12): p. 1395-401.
186. Burke, W.J., et al., *Occurrence of cancer in Alzheimer and elderly control patients: an epidemiologic necropsy study.* Alzheimer Dis Assoc Disord, 1994. **8**(1): p. 22-8.
187. Attner, B., et al., *Low cancer rates among patients with dementia in a population-based register study in Sweden.* Dement Geriatr Cogn Disord, 2010. **30**(1): p. 39-42.
188. Raji, M.A., et al., *Effect of a dementia diagnosis on survival of older patients after a diagnosis of breast, colon, or prostate cancer: implications for cancer care.* Arch Intern Med, 2008. **168**(18): p. 2033-40.
189. Heflin, M.T., et al., *The effect of comorbid illness on receipt of cancer screening by older people.* J Am Geriatr Soc, 2002. **50**(10): p. 1651-8.
190. Gorin, S.S., et al., *Treatment for breast cancer in patients with Alzheimer's disease.* J Am Geriatr Soc, 2005. **53**(11): p. 1897-904.

191. Turner, M.R., R. Goldacre, and M.J. Goldacre, *Reduced cancer incidence in Huntington's disease: record linkage study clue to an evolutionary trade-off?* Clin Genet, 2013. **83**(6): p. 588-90.
192. Ferreira, J.J., et al., *Summary of the recommendations of the EFNS/MDS-ES review on therapeutic management of Parkinson's disease.* Eur J Neurol, 2013. **20**(1): p. 5-15.
193. Pariante, C.M. and S.L. Lightman, *The HPA axis in major depression: classical theories and new developments.* Trends Neurosci, 2008. **31**(9): p. 464-8.
194. Scott, D. and B. Happell, *The high prevalence of poor physical health and unhealthy lifestyle behaviours in individuals with severe mental illness.* Issues Ment Health Nurs, 2011. **32**(9): p. 589-97.
195. DiMatteo, M.R., H.S. Lepper, and T.W. Croghan, *Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence.* Arch Intern Med, 2000. **160**(14): p. 2101-7.
196. Dhar, A.K. and D.A. Barton, *Depression and the Link with Cardiovascular Disease.* Front Psychiatry, 2016. **7**: p. 33.
197. Nemeroff, C.B. and P.J. Goldschmidt-Clermont, *Heartache and heartbreak--the link between depression and cardiovascular disease.* Nat Rev Cardiol, 2012. **9**(9): p. 526-39.
198. Yang, L., et al., *Pre-existing depression predicts survival in cardiovascular disease and cancer.* J Epidemiol Community Health, 2018. **72**(7): p. 617-622.
199. Howard, L.M., et al., *Cancer diagnosis in people with severe mental illness: practical and ethical issues.* Lancet Oncol, 2010. **11**(8): p. 797-804.
200. Mortensen, P.B. and K. Juel, *Mortality and causes of death in schizophrenic patients in Denmark.* Acta Psychiatr Scand, 1990. **81**(4): p. 372-7.
201. Saha, S., D. Chant, and J. McGrath, *A systematic review of mortality in schizophrenia: is the differential mortality gap worsening over time?* Arch Gen Psychiatry, 2007. **64**(10): p. 1123-31.
202. Inagaki, T., et al., *Factors disturbing treatment for cancer in patients with schizophrenia.* Psychiatry Clin Neurosci, 2006. **60**(3): p. 327-31.
203. Jones, S., L. Howard, and G. Thornicroft, *'Diagnostic overshadowing': worse physical health care for people with mental illness.* Acta Psychiatr Scand, 2008. **118**(3): p. 169-71.
204. Kempster, P.A., et al., *Asymmetry of substantia nigra neuronal loss in Parkinson's disease and its relevance to the mechanism of levodopa related motor fluctuations.* J Neurol Neurosurg Psychiatry, 1989. **52**(1): p. 72-6.
205. Agid, Y., et al., *Are dopaminergic neurons selectively vulnerable to Parkinson's disease?* Adv Neurol, 1993. **60**: p. 148-64.
206. Dwyer, P.K., et al., *Plantar malignant melanoma in a white Caucasian population.* Br J Dermatol, 1993. **128**(2): p. 115-20.
207. Harbour, J.W., et al., *Association between choroidal pigmentation and posterior uveal melanoma in a white population.* Br J Ophthalmol, 2004. **88**(1): p. 39-43.
208. De Marco, F., et al., *Ectopic deposition of melanin pigments as detoxifying mechanism: a paradigm for basal nuclei pigmentation.* Biochem Biophys Res Commun, 2004. **314**(2): p. 631-7.
209. Quik, M., et al., *Nicotine and Nicotinic Receptor Drugs: Potential for Parkinson's Disease and Drug-Induced Movement Disorders.* Int Rev Neurobiol, 2015. **124**: p. 247-71.
210. Skibba, J.L., et al., *Multiple primary melanoma following administration of levodopa.* Arch Pathol, 1972. **93**(6): p. 556-61.
211. Sandyk, R., *Accelerated growth of malignant melanoma by levodopa in Parkinson's disease and role of the pineal gland.* Int J Neurosci, 1992. **63**(1-2): p. 137-40.
212. Constantinescu, R., et al., *Malignant melanoma in early Parkinson's disease: the DATATOP trial.* Mov Disord, 2007. **22**(5): p. 720-2.
213. Spiegel, D. and J. Giese-Davis, *Depression and cancer: mechanisms and disease progression.* Biol Psychiatry, 2003. **54**(3): p. 269-82.

214. Giese-Davis, J., et al., *Decrease in depression symptoms is associated with longer survival in patients with metastatic breast cancer: a secondary analysis.* J Clin Oncol, 2011. **29**(4): p. 413-20.
215. Wernli, K.J., et al., *Antidepressant medication use and breast cancer risk.* Pharmacoepidemiol Drug Saf, 2009. **18**(4): p. 284-90.
216. Sharpe, C.R., et al., *The effects of tricyclic antidepressants on breast cancer risk.* Br J Cancer, 2002. **86**(1): p. 92-7.
217. Howard, L.M., et al., *The general fertility rate in women with psychotic disorders.* Am J Psychiatry, 2002. **159**(6): p. 991-7.
218. Oksbjerg Dalton, S., et al., *Schizophrenia and the risk for breast cancer.* Schizophr Res, 2003. **62**(1-2): p. 89-92.
219. Harvey, P.W., D.J. Everett, and C.J. Springall, *Adverse effects of prolactin in rodents and humans: breast and prostate cancer.* J Psychopharmacol, 2008. **22**(2 Suppl): p. 20-7.
220. McCreadie, R., et al., *Dietary intake of schizophrenic patients in Nithsdale, Scotland: case-control study.* BMJ, 1998. **317**(7161): p. 784-5.
221. Coodin, S., *Body mass index in persons with schizophrenia.* Can J Psychiatry, 2001. **46**(6): p. 549-55.
222. Bai, Y.M., et al., *Risk of developing diabetes mellitus and hyperlipidemia among patients with bipolar disorder, major depressive disorder, and schizophrenia: a 10-year nationwide population-based prospective cohort study.* J Affect Disord, 2013. **150**(1): p. 57-62.
223. Sweeney, T.E., et al., *Methods to increase reproducibility in differential gene expression via meta-analysis.* Nucleic Acids Res, 2017. **45**(1): p. e1.
224. Song, C. and G.C. Tseng, *HYPOTHESIS SETTING AND ORDER STATISTIC FOR ROBUST GENOMIC META-ANALYSIS.* Ann Appl Stat, 2014. **8**(2): p. 777-800.
225. Carvalho, B.S. and R.A. Irizarry, *A framework for oligonucleotide microarray preprocessing.* Bioinformatics, 2010. **26**(19): p. 2363-7.
226. Gautier, L., et al., *affy--analysis of Affymetrix GeneChip data at the probe level.* Bioinformatics, 2004. **20**(3): p. 307-15.
227. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
228. Du, P., W.A. Kibbe, and S.M. Lin, *lumi: a pipeline for processing Illumina microarray.* Bioinformatics, 2008. **24**(13): p. 1547-8.
229. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
230. Forés-Martos, J., et al., *Transcriptomic metaanalyses of autistic brains reveals shared gene expression and biological pathway abnormalities with cancer.* Mol Autism, 2019. **10**: p. 17.
231. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC Bioinformatics, 2008. **9**: p. 559.
232. Langfelder, P. and S. Horvath, *Fast R Functions for Robust Correlations and Hierarchical Clustering.* J Stat Softw, 2012. **46**(11).
233. Colaprico, A., et al., *TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data.* Nucleic Acids Res, 2016. **44**(8): p. e71.
234. Kauffmann, A., R. Gentleman, and W. Huber, *arrayQualityMetrics--a bioconductor package for quality assessment of microarray data.* Bioinformatics, 2009. **25**(3): p. 415-6.
235. Wang, X., et al., *An R package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection.* Bioinformatics, 2012. **28**(19): p. 2534-6.
236. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-3.
237. Nakaya, C.P.D.L.H., *MetaVolcanoR: Gene Expression Meta-analysis Visualization Tool. R package version 1.4.0.* 2020.

238. Langfelder, P., P.S. Mischel, and S. Horvath, *When is hub gene selection better than standard meta-analysis?* PLoS One, 2013. **8**(4): p. e61505.
239. Langfelder, P. and S. Horvath, *Eigengene networks for studying the relationships between co-expression modules.* BMC Syst Biol, 2007. **1**: p. 54.
240. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* OMICS, 2012. **16**(5): p. 284-7.
241. Langfelder, P., *anRichment: Collections and annotation data for use with anRichmentMethods*. 2020.
242. Sergushichev, A., *An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation*. 2016.
243. *MSigDB*. Available from: https://www.gsea-msigdb.org/gsea/msigd.
244. Dubois, B., et al., *Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria.* Lancet Neurol, 2014. **13**(6): p. 614-29.
245. Trist, B.G., D.J. Hare, and K.L. Double, *Oxidative stress in the aging substantia nigra and the etiology of Parkinson's disease.* Aging Cell, 2019. **18**(6): p. e13031.
246. Agus, F., et al., *The caudate nucleus undergoes dramatic and unique transcriptional changes in human prodromal Huntington's disease brain.* BMC Med Genomics, 2019. **12**(1): p. 137.
247. Maletic, V. and C. Raison, *Integrated neurobiology of bipolar disorder.* Front Psychiatry, 2014. **5**: p. 98.
248. Petinou, K. and D. Minaidou, *Neurobiological Bases of Autism Spectrum Disorders and Implications for Early Intervention: A Brief Overview.* Folia Phoniatr Logop, 2017. **69**(1-2): p. 38-42.
249. Courchesne, E., et al., *Neuron number and size in prefrontal cortex of children with autism.* JAMA, 2011. **306**(18): p. 2001-10.
250. Ibáñez, K., et al., *Molecular evidence for the inverse comorbidity between central nervous system disorders and cancers detected by transcriptomic meta-analyses.* PLoS Genet, 2014. **10**(2): p. e1004173.
251. McCall, M.N., B.M. Bolstad, and R.A. Irizarry, *Frozen robust multiarray analysis (fRMA).* Biostatistics, 2010. **11**(2): p. 242-53.
252. Moh, C., et al., *Cell cycle deregulation in the neurons of Alzheimer's disease.* Results Probl Cell Differ, 2011. **53**: p. 565-76.
253. Zhu, X., et al., *Alzheimer's disease: the two-hit hypothesis.* Lancet Neurol, 2004. **3**(4): p. 219-26.
254. Esteras, N., et al., *G1/S Cell Cycle Checkpoint Dysfunction in Lymphoblasts from Sporadic Parkinson's Disease Patients.* Mol Neurobiol, 2015. **52**(1): p. 386-98.
255. Manickam, N., et al., *Cell cycle re-entry of neurons and reactive neuroblastosis in Huntington's disease: Possibilities for neural-glial transition in the brain.* Life Sci, 2020. **263**: p. 118569.
256. Pramparo, T., et al., *Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers.* Mol Syst Biol, 2015. **11**(12): p. 841.
257. Gabay, M., Y. Li, and D.W. Felsher, *MYC activation is a hallmark of cancer initiation and maintenance.* Cold Spring Harb Perspect Med, 2014. **4**(6).
258. Lee, H.G., et al., *The neuronal expression of MYC causes a neurodegenerative phenotype in a novel transgenic mouse.* Am J Pathol, 2009. **174**(3): p. 891-7.
259. Ignacio, Z.M., et al., *New perspectives on the involvement of mTOR in depression as well as in the action of antidepressant drugs.* Br J Clin Pharmacol, 2016. **82**(5): p. 1280-1290.
260. Xu, F., et al., *Roles of the PI3K/AKT/mTOR signalling pathways in neurodegenerative diseases and tumours.* Cell Biosci, 2020. **10**: p. 54.
261. Bryan, M.R. and A.B. Bowman, *Manganese and the Insulin-IGF Signaling Network in Huntington's Disease and Other Neurodegenerative Disorders.* Adv Neurobiol, 2017. **18**: p. 113-142.

262. Cui, W., et al., *Inhibition of PTEN Attenuates Endoplasmic Reticulum Stress and Apoptosis via Activation of PI3K/AKT Pathway in Alzheimer's Disease.* Neurochem Res, 2017. **42**(11): p. 3052-3060.
263. Wang, G., J. Pan, and S.D. Chen, *Kinases and kinase signaling pathways: potential therapeutic targets in Parkinson's disease.* Prog Neurobiol, 2012. **98**(2): p. 207-21.
264. Khwanraj, K., et al., *Comparative mRNA Expression of eEF1A Isoforms and a PI3K/Akt/mTOR Pathway in a Cellular Model of Parkinson's Disease.* Parkinsons Dis, 2016. **2016**: p. 8716016.
265. Machado-Vieira, R., et al., *Decreased AKT1/mTOR pathway mRNA expression in short-term bipolar disorder.* Eur Neuropsychopharmacol, 2015. **25**(4): p. 468-73.
266. Leibrock, C., et al., *Akt2 deficiency is associated with anxiety and depressive behavior in mice.* Cell Physiol Biochem, 2013. **32**(3): p. 766-77.
267. Matsuda, S., et al., *Roles of PI3K/AKT/GSK3 Pathway Involved in Psychiatric Illnesses.* Diseases, 2019. **7**(1).
268. Kelleher, R.J., 3rd and M.F. Bear, *The autistic neuron: troubled translation?* Cell, 2008. **135**(3): p. 401-6.
269. Troca-Marin, J.A., A. Alves-Sampaio, and M.L. Montesinos, *Deregulated mTOR-mediated translation in intellectual disability.* Prog Neurobiol, 2012. **96**(2): p. 268-82.
270. Levitt, P. and D.B. Campbell, *The genetic and neurobiologic compass points toward common signaling dysfunctions in autism spectrum disorders.* J Clin Invest, 2009. **119**(4): p. 747-54.
271. Ehninger, D. and A.J. Silva, *Rapamycin for treating Tuberous sclerosis and Autism spectrum disorders.* Trends Mol Med, 2011. **17**(2): p. 78-87.
272. Yeung, K.S., et al., *Identification of mutations in the PI3K-AKT-mTOR signalling pathway in patients with macrocephaly and developmental delay and/or autism.* Mol Autism, 2017. **8**: p. 66.
273. Aoki, M. and T. Fujishita, *Oncogenic Roles of the PI3K/AKT/mTOR Axis.* Curr Top Microbiol Immunol, 2017. **407**: p. 153-189.
274. Perez Ortiz, J.M. and R.H. Swerdlow, *Mitochondrial dysfunction in Alzheimer's disease: Role in pathogenesis and novel therapeutic opportunities.* Br J Pharmacol, 2019. **176**(18): p. 3489-3507.
275. Cardoso, S., et al., *Alzheimer's Disease: From Mitochondrial Perturbations to Mitochondrial Medicine.* Brain Pathol, 2016. **26**(5): p. 632-47.
276. Terada, T., et al., *In vivo mitochondrial and glycolytic impairments in patients with Alzheimer disease.* Neurology, 2020. **94**(15): p. e1592-e1604.
277. Intihar, T.A., E.A. Martinez, and R. Gomez-Pastor, *Mitochondrial Dysfunction in Huntington's Disease; Interplay Between HSF1, p53 and PGC-1alpha Transcription Factors.* Front Cell Neurosci, 2019. **13**: p. 103.
278. Clay, H.B., S. Sillivan, and C. Konradi, *Mitochondrial dysfunction and pathology in bipolar disorder and schizophrenia.* Int J Dev Neurosci, 2011. **29**(3): p. 311-24.
279. Bansal, Y. and A. Kuhad, *Mitochondrial Dysfunction in Depression.* Curr Neuropharmacol, 2016. **14**(6): p. 610-8.
280. Griffiths, K.K. and R.J. Levy, *Evidence of Mitochondrial Dysfunction in Autism: Biochemical Links, Genetic-Based Associations, and Non-Energy-Related Mechanisms.* Oxid Med Cell Longev, 2017. **2017**: p. 4314025.
281. Seo, J. and M. Park, *Molecular crosstalk between cancer and neurodegenerative diseases.* Cell Mol Life Sci, 2020. **77**(14): p. 2659-2680.
282. Szybinska, A. and W. Lesniak, *P53 Dysfunction in Neurodegenerative Diseases - The Cause or Effect of Pathological Changes?* Aging Dis, 2017. **8**(4): p. 506-518.
283. Enocksson, E. and B. Jonsson, *[Does neonatal care of very low birth weight infants mean an increased number of children with multiple handicaps?].* Lakartidningen, 1989. **86**(50): p. 4420-1.
284. Jodeiri Farshbaf, M. and K. Ghaedi, *Huntington's Disease and Mitochondria.* Neurotox Res, 2017. **32**(3): p. 518-529.

285. Houck, A.L., S. Seddighi, and J.A. Driver, *At the Crossroads Between Neurodegeneration and Cancer: A Review of Overlapping Biology and Its Implications.* Curr Aging Sci, 2018. **11**(2): p. 77-89.
286. Doty, K.R., M.V. Guillot-Sestier, and T. Town, *The role of the immune system in neurodegenerative disorders: Adaptive or maladaptive?* Brain Res, 2015. **1617**: p. 155-73.
287. Ellrichmann, G., et al., *The role of the immune system in Huntington's disease.* Clin Dev Immunol, 2013. **2013**: p. 541259.
288. Meltzer, A. and J. Van de Water, *The Role of the Immune System in Autism Spectrum Disorder.* Neuropsychopharmacology, 2017. **42**(1): p. 284-298.
289. Bennett, F.C. and A.V. Molofsky, *The immune system and psychiatric disease: a basic science perspective.* Clin Exp Immunol, 2019. **197**(3): p. 294-307.
290. Pandya, P.H., et al., *The Immune System in Cancer Pathogenesis: Potential Therapeutic Approaches.* J Immunol Res, 2016. **2016**: p. 4273943.
291. Stark, A.K., et al., *PI3K inhibitors in inflammation, autoimmunity and cancer.* Curr Opin Pharmacol, 2015. **23**: p. 82-91.
292. Song, G., G. Ouyang, and S. Bao, *The activation of Akt/PKB signaling pathway and cell survival.* J Cell Mol Med, 2005. **9**(1): p. 59-71.
293. Wang, D.B., et al., *p53 and mitochondrial function in neurons.* Biochim Biophys Acta, 2014. **1842**(8): p. 1186-97.
294. Bulik-Sullivan, B., et al., *An atlas of genetic correlations across human diseases and traits.* Nat Genet, 2015. **47**(11): p. 1236-41.
295. *The Human Reference Interactome (HuRI)*. Available from: http://www.interactome-atlas.org/.
296. Luck, K., et al., *A reference map of the human binary protein interactome.* Nature, 2020. **580**(7803): p. 402-408.
297. *Genotype-Tissue Expression (GTEx)*. Available from: https://www.gtexportal.org/home/datasets.
298. *The comprehensive resource of mammalian protein complexes (CORUM)*. Available from: http://mips.helmholtz-muenchen.de/corum/.
299. Matys, V., et al., *TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes.* Nucleic Acids Res, 2006. **34**(Database issue): p. D108-10.
300. Matys, V., et al., *TRANSFAC: transcriptional regulation, from patterns to profiles.* Nucleic Acids Res, 2003. **31**(1): p. 374-8.
301. *Harmonizome*. Available from: http://amp.pharm.mssm.edu/Harmonizome/dataset/TRANSFAC+Curated+Transcription+Factor+Targets.
302. *PhosphoSitePlus* Available from: https://www.phosphosite.org/homeAction.action.
303. *Kyoto Encyclopedia of Genes and Genomes (KEGG)*. Available from: https://www.genome.jp/kegg/.
304. Zhang, J.D. and S. Wiemann, *KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor.* Bioinformatics, 2009. **25**(11): p. 1470-1.
305. *STRING*. Available from: https://string-db.org/.
306. *BioGriD*. Available from: https://thebiogrid.org/.
307. *DisGeNet*. Available from: https://www.disgenet.org/.
308. *The Phenotype Genotype Integrator (PheGenI)*. Available from: https://www.ncbi.nlm.nih.gov/gap/phegeni.
309. *eDGAR*. Available from: http://edgar.biocomp.unibo.it/gene_disease_db/.
310. *GWAS catalog*. Available from: https://www.ebi.ac.uk/gwas/downloads/summary-statistics.
311. *UK Biobank*. Available from: http://www.nealelab.is/uk-biobank.
312. Lambert, J.C., et al., *Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.* Nat Genet, 2013. **45**(12): p. 1452-8.

313. Moreno-Grau, S., et al., *Genome-wide association analysis of dementia and its clinical endophenotypes reveal novel loci associated with Alzheimer's disease and three causality networks: The GR@ACE project.* Alzheimers Dement, 2019. **15**(10): p. 1333-1347.
314. Jansen, I.E., et al., *Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk.* Nat Genet, 2019. **51**(3): p. 404-413.
315. Grove, J., et al., *Identification of common genetic risk variants for autism spectrum disorder.* Nat Genet, 2019. **51**(3): p. 431-444.
316. Stahl, E.A., et al., *Genome-wide association study identifies 30 loci associated with bipolar disorder.* Nat Genet, 2019. **51**(5): p. 793-803.
317. Wray, N.R., et al., *Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression.* Nat Genet, 2018. **50**(5): p. 668-681.
318. Nalls, M.A., et al., *Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies.* Lancet Neurol, 2019. **18**(12): p. 1091-1102.
319. Chang, D., et al., *A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci.* Nat Genet, 2017. **49**(10): p. 1511-1516.
320. Schumacher, F.R., et al., *Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci.* Nat Genet, 2018. **50**(7): p. 928-936.
321. Zhang, H., et al., *Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses.* Nat Genet, 2020. **52**(6): p. 572-581.
322. Michailidou, K., et al., *Association analysis identifies 65 new breast cancer risk loci.* Nature, 2017. **551**(7678): p. 92-94.
323. Phelan, C.M., et al., *Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer.* Nat Genet, 2017. **49**(5): p. 680-691.
324. Bishop, D.T., et al., *Genome-wide association study identifies three loci associated with melanoma risk.* Nat Genet, 2009. **41**(8): p. 920-5.
325. Almeida, J.R., et al., *Pattern recognition analysis of anterior cingulate cortex blood flow to classify depression polarity.* Br J Psychiatry, 2013. **203**(3): p. 310-1.
326. Fusar-Poli, P., et al., *Transdiagnostic psychiatry: a systematic review.* World Psychiatry, 2019. **18**(2): p. 192-207.
327. Consortium, I.T.P.-C.A.o.W.G., *Pan-cancer analysis of whole genomes.* Nature, 2020. **578**(7793): p. 82-93.
328. Skrott, Z., et al., *Disulfiram's anti-cancer activity reflects targeting NPL4, not inhibition of aldehyde dehydrogenase.* Oncogene, 2019. **38**(40): p. 6711-6722.
329. Guney, E., et al., *Network-based in silico drug efficacy screening.* Nat Commun, 2016. **7**: p. 10331.
330. Duan, Q., et al., *LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W449-60.
331. *MEDI-an*. Available from: https://www.vumc.org/cpm/cpm-blog/medi-ensemble-medication-indication-resource-0.
332. *RxNev RESTful*. Available from: https://rxnav.nlm.nih.gov/RxNormAPIREST.html.
333. Wickham, H., *httr Tools for Working with URLs and HTTP.* 2019.
334. Couture-Beil, A., *rjson: JSON for R.* 2018.
335. *DrugBank*. Available from: https://www.drugbank.ca/.
336. *UniChem RESTful*. Available from: https://www.ebi.ac.uk/unichem/rest.
337. Perez-Aguilar, B., et al., *Acetylcholinesterase is associated with a decrease in cell proliferation of hepatocellular carcinoma cells.* Biochim Biophys Acta, 2015. **1852**(7): p. 1380-7.
338. Lazarevic-Pasti, T., et al., *Modulators of Acetylcholinesterase Activity: From Alzheimer's Disease to Anti-Cancer Drugs.* Curr Med Chem, 2017. **24**(30): p. 3283-3309.

339. Xi, H.J., et al., *Role of acetylcholinesterase in lung cancer.* Thorac Cancer, 2015. **6**(4): p. 390-8.
340. Jiang, H. and X.J. Zhang, *Acetylcholinesterase and apoptosis. A novel perspective for an old enzyme.* FEBS J, 2008. **275**(4): p. 612-7.
341. Chen, D., et al., *Tamoxifen and toremifene cause impairment of learning and memory function in mice.* Pharmacol Biochem Behav, 2002. **71**(1-2): p. 269-76.
342. Zabihi, H., et al., *The effects of tamoxifen on learning, memory and brain tissues oxidative damage in ovariectomized and naive female rats.* Adv Biomed Res, 2014. **3**: p. 219.
343. Pike, C.J., et al., *Protective actions of sex steroid hormones in Alzheimer's disease.* Front Neuroendocrinol, 2009. **30**(2): p. 239-58.
344. Coman, L., et al., *Alzheimer's Disease - Estrogens and Selective Estrogen Receptor Modulators, Friends or Foes?* Acta Endocrinol (Buchar), 2017. **13**(1): p. 77-83.
345. Khan, M.M., *Translational Significance of Selective Estrogen Receptor Modulators in Psychiatric Disorders.* Int J Endocrinol, 2018. **2018**: p. 9516592.
346. Manthey, D., et al., *Estrogen induces a rapid secretion of amyloid beta precursor protein via the mitogen-activated protein kinase pathway.* Eur J Biochem, 2001. **268**(15): p. 4285-91.
347. Gogos, A., et al., *A Role for Estrogen in Schizophrenia: Clinical and Preclinical Findings.* Int J Endocrinol, 2015. **2015**: p. 615356.
348. Kulkarni, J., et al., *Estradiol for treatment-resistant schizophrenia: a large-scale randomized-controlled trial in women of child-bearing age.* Mol Psychiatry, 2015. **20**(6): p. 695-702.
349. Kulkarni, J., et al., *A pilot study of hormone modulation as a new treatment for mania in women with bipolar affective disorder.* Psychoneuroendocrinology, 2006. **31**(4): p. 543-7.
350. Boni, C., et al., *Therapeutic activity of testosterone in metastatic breast cancer.* Anticancer Res, 2014. **34**(3): p. 1287-90.
351. Walther, A., J. Breidenstein, and R. Miller, *Association of Testosterone Treatment With Alleviation of Depressive Symptoms in Men: A Systematic Review and Meta-analysis.* JAMA Psychiatry, 2019. **76**(1): p. 31-40.
352. Dichtel, L.E., et al., *Low-Dose Testosterone Augmentation for Antidepressant-Resistant Major Depressive Disorder in Women: An 8-Week Randomized Placebo-Controlled Study.* Am J Psychiatry, 2020. **177**(10): p. 965-973.
353. Manohar, S., S. Jamesdaniel, and R. Salvi, *Cisplatin inhibits hippocampal cell proliferation and alters the expression of apoptotic genes.* Neurotox Res, 2014. **25**(4): p. 369-80.
354. Chiang, A.C.A., et al., *Chemotherapy accelerates age-related development of tauopathy and results in loss of synaptic integrity and cognitive impairment.* Brain Behav Immun, 2019. **79**: p. 319-325.
355. Huo, X., et al., *Cisplatin treatment induces attention deficits and impairs synaptic integrity in the prefrontal cortex in mice.* Sci Rep, 2018. **8**(1): p. 17400.
356. Takada, M., et al., *Inverse Association between Sodium Channel-Blocking Antiepileptic Drug Use and Cancer: Data Mining of Spontaneous Reporting and Claims Databases.* Int J Med Sci, 2016. **13**(1): p. 48-59.
357. Pellegrino, M., et al., *FoxO3a Mediates the Inhibitory Effects of the Antiepileptic Drug Lamotrigine on Breast Cancer Growth.* Mol Cancer Res, 2018. **16**(6): p. 923-934.
358. Akbarzadeh, L., T. Moini Zanjani, and M. Sabetkasaei, *Comparison of Anticancer Effects of Carbamazepine and Valproic Acid.* Iran Red Crescent Med J, 2016. **18**(10): p. e37230.
359. Dakir, E.H., et al., *The anti-psychotic drug pimozide is a novel chemotherapeutic for breast cancer.* Oncotarget, 2018. **9**(79): p. 34889-34910.
360. Choi, J., et al., *Pimozide suppresses cancer cell migration and tumor metastasis through binding to ARPC2, a subunit of the Arp2/3 complex.* Cancer Sci, 2019. **110**(12): p. 3788-3801.
361. Chen, J.J., et al., *Antipsychotic agent pimozide promotes reversible proliferative suppression by inducing cellular quiescence in liver cancer.* Oncol Rep, 2019. **42**(3): p. 1101-1109.
362. Zhou, W., et al., *The antipsychotic drug pimozide inhibits cell growth in prostate cancer through suppression of STAT3 activation.* Int J Oncol, 2016. **48**(1): p. 322-8.

363. Otreba, M., M. Pajor, and J.D. Warncke, *Antimelanoma activity of perphenazine and prochlorperazine in human COLO829 and C32 cell lines.* Naunyn Schmiedebergs Arch Pharmacol, 2019. **392**(10): p. 1257-1264.
364. Friedman, G.D., et al., *Haloperidol and Prostate Cancer Prevention: More Epidemiologic Research Needed.* Perm J, 2020. **24**.
365. Feng, Z., et al., *The antipsychotic agent trifluoperazine hydrochloride suppresses triple-negative breast cancer tumor growth and brain metastasis by inducing G0/G1 arrest and apoptosis.* Cell Death Dis, 2018. **9**(10): p. 1006.
366. Pulkoski-Gross, A., et al., *Repurposing the antipsychotic trifluoperazine as an antimetastasis agent.* Mol Pharmacol, 2015. **87**(3): p. 501-12.
367. Lee, W.Y., et al., *Repositioning antipsychotic chlorpromazine for treating colorectal cancer by inhibiting sirtuin 1.* Oncotarget, 2015. **6**(29): p. 27580-95.
368. Chen, V.C., et al., *New Use for Old Drugs: The Protective Effect of Risperidone on Colorectal Cancer.* Cancers (Basel), 2020. **12**(6).
369. Cordero, M.D., et al., *Acute oxidant damage promoted on cancer cells by amitriptyline in comparison with some common chemotherapeutic drugs.* Anticancer Drugs, 2010. **21**(10): p. 932-44.
370. Villanueva-Paz, M., et al., *Amitriptyline induces mitophagy that precedes apoptosis in human HepG2 cells.* Genes Cancer, 2016. **7**(7-8): p. 260-277.
371. Yang, D.K. and S.J. Kim, *Desipramine induces apoptosis in hepatocellular carcinoma cells.* Oncol Rep, 2017. **38**(2): p. 1029-1034.
372. Riess, J.W., et al., *A Phase Iia Study Repositioning Desipramine in Small Cell Lung Cancer and Other High-Grade Neuroendocrine Tumors.* Cancer Treat Res Commun, 2020. **23**: p. 100174.
373. Lee, H., S. Kang, and W. Kim, *Drug Repositioning for Cancer Therapy Based on Large-Scale Drug-Induced Transcriptional Signatures.* PLoS One, 2016. **11**(3): p. e0150460.
374. Cloonan, S.M. and D.C. Williams, *The antidepressants maprotiline and fluoxetine induce Type II autophagic cell death in drug-resistant Burkitt's lymphoma.* Int J Cancer, 2011. **128**(7): p. 1712-23.
375. Gil-Ad, I., et al., *Evaluation of the potential anti-cancer activity of the antidepressant sertraline in human colon cancer cell lines and in colorectal cancer-xenografted mice.* Int J Oncol, 2008. **33**(2): p. 277-86.
376. Chinnapaka, S., V. Bakthavachalam, and G. Munirathinam, *Repurposing antidepressant sertraline as a pharmacological drug to target prostate cancer stem cells: dual activation of apoptosis and autophagy signaling by deregulating redox balance.* Am J Cancer Res, 2020. **10**(7): p. 2043-2065.
377. Hsu, L.C., et al., *Beneficial effect of fluoxetine on anti-tumor progression on hepatocellular carcinoma and non-small cell lung cancer bearing animal model.* Biomed Pharmacother, 2020. **126**: p. 110054.
378. Szende, B., K. Magyar, and Z. Szegedi, *Apoptotic and antiapoptotic effect of (-)deprenyl and (-)-desmethyl-deprenyl on human cell lines.* Neurobiology (Bp), 2000. **8**(3-4): p. 249-55.
379. Ryu, I., et al., *LDeprenyl exerts cytotoxicity towards acute myeloid leukemia through inhibition of mitochondrial respiration.* Oncol Rep, 2018. **40**(6): p. 3869-3878.
380. Shih, J.C., *Monoamine oxidase isoenzymes: genes, functions and targets for behavior and cancer therapy.* J Neural Transm (Vienna), 2018. **125**(11): p. 1553-1566.
381. Gaur, S., et al., *Effect of Monoamine oxidase A (MAOA) inhibitors on androgen-sensitive and castration-resistant prostate cancer cells.* Prostate, 2019. **79**(6): p. 667-677.
382. Eom, C.S., S.M. Park, and K.H. Cho, *Use of antidepressants and the risk of breast cancer: a meta-analysis.* Breast Cancer Res Treat, 2012. **136**(3): p. 635-45.
383. Huo, Y.L., J.M. Qiao, and S. Gao, *Association between antidepressant medication use and epithelial ovarian cancer risk: a systematic review and meta-analysis of observational studies.* Br J Clin Pharmacol, 2018. **84**(4): p. 649-658.

384. Tsai, C.J., et al., *Antidepressant Prescription and Risk of Lung Cancer: A Nationwide Case-Control Study.* Pharmacopsychiatry, 2019. **52**(3): p. 134-141.
385. Wang, Y.C., et al., *Increased Risk of Dementia in Patients with Antidepressants: A Meta-Analysis of Observational Studies.* Behav Neurol, 2018. **2018**: p. 5315098.
386. Tamim, H.M., et al., *Antidepressants and risk of prostate cancer: a nested case-control study.* Prostate Cancer Prostatic Dis, 2008. **11**(1): p. 53-60.
387. Friedman, G.D., et al., *Antidepressants and testicular cancer.* Cancer Causes Control, 2014. **25**(2): p. 251-8.
388. Chen, L., et al., *Antidepressant use and colorectal cancer morbidity and mortality: A dose-response meta analysis.* Medicine (Baltimore), 2020. **99**(22): p. e20185.
389. Ki, Y.S., et al., *Donepezil, a potent acetylcholinesterase inhibitor, induces caspase-dependent apoptosis in human promyelocytic leukemia HL-60 cells.* Biol Pharm Bull, 2010. **33**(6): p. 1054-9.
390. Kim, H.B., et al., *Use of benzodiazepine and risk of cancer: A meta-analysis of observational studies.* Int J Cancer, 2017. **140**(3): p. 513-525.
391. Zhang, T., et al., *Benzodiazepine drug use and cancer risk: a dose-response meta analysis of prospective cohort studies.* Oncotarget, 2017. **8**(60): p. 102381-102391.
392. Ogura, J., et al., *Carbidopa is an activator of aryl hydrocarbon receptor with potential for cancer therapy.* Biochem J, 2017. **474**(20): p. 3391-3402.
393. Chen, Z., et al., *Carbidopa suppresses prostate cancer via aryl hydrocarbon receptor-mediated ubiquitination and degradation of androgen receptor.* Oncogenesis, 2020. **9**(5): p. 49.
394. Brodie, S.A. and J.C. Brandes, *Could valproic acid be an effective anticancer agent? The evidence so far.* Expert Rev Anticancer Ther, 2014. **14**(10): p. 1097-100.
395. Mohs, R.C. and N.H. Greig, *Drug discovery and development: Role of basic biological research.* Alzheimers Dement (N Y), 2017. **3**(4): p. 651-657.
396. DiMasi, J.A., et al., *Trends in risks associated with new drug development: success rates for investigational drugs.* Clin Pharmacol Ther, 2010. **87**(3): p. 272-7.
397. Ma, L.L., et al., *Association between cancer and Alzheimer's disease: systematic review and meta-analysis.* J Alzheimers Dis, 2014. **42**(2): p. 565-73.
398. Shi, H.B., et al., *Alzheimer disease and cancer risk: a meta-analysis.* J Cancer Res Clin Oncol, 2015. **141**(3): p. 485-94.
399. Papageorgakopoulos, T.N., et al., *The association between Alzheimer's disease and cancer: Systematic review - Meta-analysis.* Hell J Nucl Med, 2017. **20 Suppl**: p. 45-57.
400. Hayes, J.F., et al., *A systematic review and meta-analysis of premature mortality in bipolar affective disorder.* Acta Psychiatr Scand, 2015. **131**(6): p. 417-25.
401. Wulsin, L.R., G.E. Vaillant, and V.E. Wells, *A systematic review of the mortality of depression.* Psychosom Med, 1999. **61**(1): p. 6-17.
402. Oerlemans, M.E., et al., *A meta-analysis on depression and subsequent cancer risk.* Clin Pract Epidemiol Ment Health, 2007. **3**: p. 29.
403. Pinquart, M. and P.R. Duberstein, *Depression and cancer mortality: a meta-analysis.* Psychol Med, 2010. **40**(11): p. 1797-810.
404. Cuijpers, P., et al., *Comprehensive meta-analysis of excess mortality in depression in the general community versus patients with specific illnesses.* Am J Psychiatry, 2014. **171**(4): p. 453-62.
405. Sun, H.L., et al., *Depression and the risk of breast cancer: a meta-analysis of cohort studies.* Asian Pac J Cancer Prev, 2015. **16**(8): p. 3233-9.
406. Jia, Y., et al., *Depression and cancer risk: a systematic review and meta-analysis.* Public Health, 2017. **149**: p. 138-148.
407. Huang, P., et al., *The association between Parkinson's disease and melanoma: a systematic review and meta-analysis.* Transl Neurodegener, 2015. **4**: p. 21.

408. Wang, T., *The link between Parkinson's disease and breast and prostate cancers: A meta-analysis.* Int J Neurosci, 2015. **125**(12): p. 895-903.
409. Ye, R., et al., *The Relationship between Parkinson Disease and Brain Tumor: A Meta-Analysis.* PLoS One, 2016. **11**(10): p. e0164388.
410. Xie, X., et al., *Risk of lung cancer in Parkinson's disease.* Oncotarget, 2016. **7**(47): p. 77319-77325.
411. Xie, X., X. Luo, and M. Xie, *Association between Parkinson's disease and risk of colorectal cancer.* Parkinsonism Relat Disord, 2017. **35**: p. 42-47.
412. Gangaraju, R., et al., *Thrombotic Thrombocytopenic Purpura Associated With Pegylated Interferon Alfa-2a Use in a Patient With Polycythemia Vera.* J Natl Compr Canc Netw, 2017. **15**(6): p. 757-760.
413. Xu, D., et al., *Lower risk of liver cancer in patients with schizophrenia: a systematic review and meta-analysis of cohort studies.* Oncotarget, 2017. **8**(60): p. 102328-102335.
414. Li, H., et al., *The incidence rate of cancer in patients with schizophrenia: A meta-analysis of cohort studies.* Schizophr Res, 2018. **195**: p. 519-528.
415. Beard, C.M., et al., *Cause of death in Alzheimer's disease.* Ann Epidemiol, 1996. **6**(3): p. 195-200.
416. Attems, J., et al., *Cause of death in demented and non-demented elderly inpatients; an autopsy study of 308 cases.* J Alzheimers Dis, 2005. **8**(1): p. 57-62.
417. Chamandy, N. and C. Wolfson, *Underlying cause of death in demented and non-demented elderly Canadians.* Neuroepidemiology, 2005. **25**(2): p. 75-84.
418. Lai, S.W., et al., *No correlation between Alzheimer's disease and risk of hepatocellular carcinoma in older people: an observation in Taiwan.* Geriatr Gerontol Int, 2014. **14**(1): p. 231-2.
419. Todd, S., S. Barr, and A.P. Passmore, *Cause of death in Alzheimer's disease: a cohort study.* QJM, 2013. **106**(8): p. 747-53.
420. Romero, J.P., et al., *Alzheimer's disease is associated with decreased risk of cancer-specific mortality: a prospective study (NEDICES).* J Alzheimers Dis, 2014. **40**(2): p. 465-73.
421. Shavelle, R.M., D.J. Strauss, and J. Pickett, *Causes of death in autism.* J Autism Dev Disord, 2001. **31**(6): p. 569-76.
422. Lauritsen, M.B., et al., *Medical disorders among inpatients with autism in Denmark according to ICD-8: a nationwide register-based study.* J Autism Dev Disord, 2002. **32**(2): p. 115-9.
423. Atladottir, H.O., et al., *Patterns of contact with hospital for children with an autism spectrum disorder: a Danish register-based study.* J Autism Dev Disord, 2012. **42**(8): p. 1717-28.
424. Chiang, H.L., et al., *Risk of cancer in children, adolescents, and young adults with autistic disorder.* J Pediatr, 2015. **166**(2): p. 418-23 e1.
425. Darbro, B.W., et al., *Autism Linked to Increased Oncogene Mutations but Decreased Cancer Rate.* PLoS One, 2016. **11**(3): p. e0149041.
426. Hirvikoski, T., et al., *Premature mortality in autism spectrum disorder.* Br J Psychiatry, 2016. **208**(3): p. 232-8.
427. Mouridsen, S.E., B. Rich, and T. Isager, *Risk of cancer in adult people diagnosed with infantile autism in childhood: A longitudinal case control study based on hospital discharge diagnoses.* Research in Autism Spectrum Disorders, 2016. **23**: p. 203-209.
428. Norton, B. and L.J. Whalley, *Mortality of a lithium-treated population.* Br J Psychiatry, 1984. **145**: p. 277-82.
429. Weeke, A. and M. Vaeth, *Excess mortality of bipolar and unipolar manic-depressive patients.* J Affect Disord, 1986. **11**(3): p. 227-34.
430. Hoyer, E.H., P.B. Mortensen, and A.V. Olesen, *Mortality and causes of death in a total national sample of patients with affective disorders admitted for the first time between 1973 and 1993.* Br J Psychiatry, 2000. **176**: p. 76-82.

431. Osby, U., et al., *Excess mortality in bipolar and unipolar disorder in Sweden.* Arch Gen Psychiatry, 2001. **58**(9): p. 844-50.
432. Dalton, S.O., et al., *Depression and cancer risk: a register-based study of patients hospitalized with affective disorders, Denmark, 1969-1993.* Am J Epidemiol, 2002. **155**(12): p. 1088-95.
433. Dutta, R., et al., *Suicide and other causes of mortality in bipolar disorder: a longitudinal study.* Psychol Med, 2007. **37**(6): p. 839-47.
434. BarChana, M., et al., *Enhanced cancer risk among patients with bipolar disorder.* J Affect Disord, 2008. **108**(1-2): p. 43-8.
435. Rosenshfir, E., et al., *Reduced Risk of Cancer in Patients with Bipolar Affective Disorder.* Neuroscience and Medicine, 2011. **02**(01): p. 57-60.
436. Crump, C., et al., *Comorbidities and mortality in bipolar disorder: a Swedish national cohort study.* JAMA Psychiatry, 2013. **70**(9): p. 931-9.
437. Kanani, R., et al., *The association of mood disorders with breast cancer survival: an investigation of linked cancer registration and hospital admission data for South East England.* Psychooncology, 2016. **25**(1): p. 19-27.
438. Martinsson, L., et al., *Lithium treatment and cancer incidence in bipolar disorder.* Bipolar Disord, 2016. **18**(1): p. 33-40.
439. Kahan, N.R., et al., *No apparent association between bipolar disorder and cancer in a large epidemiological study of outpatients in a managed care population.* Int Clin Psychopharmacol, 2018. **33**(2): p. 73-78.
440. Sorensen, S.A. and K. Fenger, *Causes of death in patients with Huntington's disease and in unaffected first degree relatives.* J Med Genet, 1992. **29**(12): p. 911-4.
441. Ji, J., K. Sundquist, and J. Sundquist, *Cancer incidence in patients with polyglutamine diseases: a population-based study in Sweden.* Lancet Oncol, 2012. **13**(6): p. 642-8.
442. McNulty, P., et al., *Reduced Cancer Incidence in Huntington's Disease: Analysis in the Registry Study.* J Huntingtons Dis, 2018. **7**(3): p. 209-222.
443. Solberg, O.K., et al., *Age at Death and Causes of Death in Patients with Huntington Disease in Norway in 1986-2015.* J Huntingtons Dis, 2018. **7**(1): p. 77-86.
444. Whitlock, F.A. and M. Sisking, *Depression and cancer: a follow-up study.* Psychol Med, 1979. **9**(4): p. 747-52.
445. Thomas, C.B. and O.L. McCabe, *Precursors of premature disease and death: habits of nervous tension.* Johns Hopkins Med J, 1980. **147**(4): p. 137-45.
446. Persky, V.W., J. Kempthorne-Rawson, and R.B. Shekelle, *Personality and risk of cancer: 20-year follow-up of the Western Electric Study.* Psychosom Med, 1987. **49**(5): p. 435-49.
447. Hahn, R.C. and D.B. Petitti, *Minnesota Multiphasic Personality Inventory-rated depression and the incidence of breast cancer.* Cancer, 1988. **61**(4): p. 845-8.
448. Kaplan, G.A. and P. Reynolds, *Depression and cancer mortality and morbidity: prospective evidence from the Alameda County study.* J Behav Med, 1988. **11**(1): p. 1-13.
449. Zonderman, A.B., P.T. Costa, Jr., and R.R. McCrae, *Depression as a risk for cancer morbidity and mortality in a nationally representative sample.* JAMA, 1989. **262**(9): p. 1191-5.
450. Linkins, R.W. and G.W. Comstock, *Depressed mood and development of cancer.* Am J Epidemiol, 1990. **132**(5): p. 962-72.
451. Friedman, G.D., *Psychiatrically-diagnosed depression and subsequent cancer.* Cancer Epidemiol Biomarkers Prev, 1994. **3**(1): p. 11-3.
452. Vogt, T., et al., *Mental health status as a predictor of morbidity and mortality: a 15-year follow-up of members of a health maintenance organization.* Am J Public Health, 1994. **84**(2): p. 227-31.
453. Penninx, B.W., et al., *Chronically depressed mood and cancer risk in older persons.* J Natl Cancer Inst, 1998. **90**(24): p. 1888-93.

454. Whooley, M.A. and W.S. Browner, *Association between depressive symptoms and mortality in older women. Study of Osteoporotic Fractures Research Group.* Arch Intern Med, 1998. **158**(19): p. 2129-35.
455. Bradvik, L. and M. Berglund, *Late mortality in severe depression.* Acta Psychiatr Scand, 2001. **103**(2): p. 111-6.
456. Berglund, M. and K. Nilsson, *Mortality in severe depression. A prospective study including 103 suicides.* Acta Psychiatr Scand, 1987. **76**(4): p. 372-80.
457. Yasuda, N., et al., *The differential influence of distinct clusters of psychiatric symptoms, as assessed by the general health questionnaire, on cause of death in older persons living in a rural community of Japan.* J Am Geriatr Soc, 2002. **50**(2): p. 313-20.
458. Nyklicek, I., et al., *Depression and the lower risk for breast cancer development in middle-aged women: a prospective study.* Psychol Med, 2003. **33**(6): p. 1111-7.
459. Goodwin, J.S., D.D. Zhang, and G.V. Ostir, *Effect of depression on diagnosis, treatment, and survival of older women with breast cancer.* J Am Geriatr Soc, 2004. **52**(1): p. 106-11.
460. Kisely, S., et al., *Excess cancer mortality in psychiatric patients.* Can J Psychiatry, 2008. **53**(11): p. 753-61.
461. Kroenke, C.H., et al., *Depressive symptoms and prospective incidence of colorectal cancer in women.* Am J Epidemiol, 2005. **162**(9): p. 839-48.
462. Goldacre, M.J., et al., *Cancer in people with depression or anxiety: record-linkage study.* Soc Psychiatry Psychiatr Epidemiol, 2007. **42**(9): p. 683-9.
463. Mykletun, A., et al., *Anxiety, depression, and cause-specific mortality: the HUNT study.* Psychosom Med, 2007. **69**(4): p. 323-31.
464. Kawamura, T., et al., *Survival rate and causes of mortality in the elderly with depression: a 15-year prospective study of a Japanese community sample, the Matsunoyama-Niigata suicide prevention project.* J Investig Med, 2007. **55**(3): p. 106-14.
465. Gross, A.L., J.J. Gallo, and W.W. Eaton, *Depression and cancer risk: 24 years of follow-up of the Baltimore Epidemiologic Catchment Area sample.* Cancer Causes Control, 2010. **21**(2): p. 191-9.
466. Gallo, J.J., et al., *Major depression and cancer: the 13-year follow-up of the Baltimore epidemiologic catchment area sample (United States).* Cancer Causes Control, 2000. **11**(8): p. 751-8.
467. Chen, Y.H. and H.C. Lin, *Increased risk of cancer subsequent to severe depression--a nationwide population-based study.* J Affect Disord, 2011. **131**(1-3): p. 200-6.
468. Hamer, M., C.J. Bates, and G.D. Mishra, *Depression, physical function, and risk of mortality: National Diet and Nutrition Survey in adults older than 65 years.* Am J Geriatr Psychiatry, 2011. **19**(1): p. 72-8.
469. Liang, J.A., et al., *The analysis of depression and subsequent cancer risk in Taiwan.* Cancer Epidemiol Biomarkers Prev, 2011. **20**(3): p. 473-5.
470. Wyman, L., R.M. Crum, and D. Celentano, *Depressed mood and cause-specific mortality: a 40-year general community assessment.* Ann Epidemiol, 2012. **22**(9): p. 638-43.
471. Gallo, J.J., et al., *Long term effect of depression care management on mortality in older adults: follow-up of cluster randomized clinical trial in primary care.* BMJ, 2013. **346**: p. f2570.
472. Lai, S.W., et al., *No Association between Depression and Risk of Hepatocellular Carcinoma in Older People in Taiwan.* ISRN Psychiatry, 2013. **2013**: p. 901987.
473. Lemogne, C., et al., *Depression and the risk of cancer: a 15-year follow-up study of the GAZEL cohort.* Am J Epidemiol, 2013. **178**(12): p. 1712-20.
474. Almeida, O.P., et al., *Depression, antidepressant use and mortality in later life: the Health In Men Study.* PLoS One, 2010. **5**(6): p. e11266.
475. Archer, G., H. Pikhart, and J. Head, *Do depressive symptoms predict cancer incidence?: 17-year follow-up of the Whitehall II study.* J Psychosom Res, 2015. **79**(6): p. 595-603.

476. Chang, H.Y., et al., *Depression as a risk factor for overall and hormone-related cancer: the Korean cancer prevention study.* J Affect Disord, 2015. **173**: p. 1-8.
477. Huang, T., et al., *Depression and risk of epithelial ovarian cancer: Results from two large prospective cohort studies.* Gynecol Oncol, 2015. **139**(3): p. 481-6.
478. O'Neill, S., et al., *Associations between DSM-IV mental disorders and subsequent self-reported diagnosis of cancer.* J Psychosom Res, 2014. **76**(3): p. 207-12.
479. Liang, X., et al., *Effect of depression before breast cancer diagnosis on mortality among postmenopausal women.* Cancer, 2017. **123**(16): p. 3107-3115.
480. Westlund, K., *Cancer as a Cause of Death among Patients with Other Chronic Diseases.* Journal of the American Medical Association, 1956. **162**(10).
481. Barbeau, A. and J.G. Joly, *[Parkinsonism and cancer].* Union Med Can, 1963. **92**: p. 169-74.
482. Hoehn, M.M. and M.D. Yahr, *Parkinsonism: onset, progression and mortality.* Neurology, 1967. **17**(5): p. 427-42.
483. Kessler, II, *Epidemiologic studies of Parkinson's disease. II. A hospital-based survey.* Am J Epidemiol, 1972. **95**(4): p. 308-18.
484. Harada, H., S. Nishikawa, and K. Takahashi, *Epidemiology of Parkinson's disease in a Japanese city.* Arch Neurol, 1983. **40**(3): p. 151-4.
485. Jansson, B. and J. Jankovic, *Low cancer rates among patients with Parkinson's disease.* Ann Neurol, 1985. **17**(5): p. 505-9.
486. Gorell, J.M., C.C. Johnson, and B.A. Rybicki, *Parkinson's disease and its comorbid disorders: an analysis of Michigan mortality data, 1970 to 1990.* Neurology, 1994. **44**(10): p. 1865-8.
487. Ben-Shlomo, Y. and M.G. Marmot, *Survival and cause of death in a cohort of patients with parkinsonism: possible clues to aetiology?* J Neurol Neurosurg Psychiatry, 1995. **58**(3): p. 293-9.
488. Wermuth, L., et al., *Mortality in patients with Parkinson's disease.* Acta Neurol Scand, 1995. **92**(1): p. 55-8.
489. Louis, E.D., et al., *Mortality from Parkinson disease.* Arch Neurol, 1997. **54**(3): p. 260-4.
490. Hely, M.A., et al., *The sydney multicentre study of Parkinson's disease: progression and mortality at 10 years.* J Neurol Neurosurg Psychiatry, 1999. **67**(3): p. 300-7.
491. Vanacore, N., et al., *Mortality cancer risk in parkinsonian patients: a population-based study.* Neurology, 1999. **52**(2): p. 395-8.
492. Minami, Y., et al., *Mortality and cancer incidence in patients with Parkinson's disease.* J Neurol, 2000. **247**(6): p. 429-34.
493. Beyer, M.K., et al., *Causes of death in a community-based study of Parkinson's disease.* Acta Neurol Scand, 2001. **103**(1): p. 7-11.
494. Fall, P.A., et al., *Survival time, mortality, and cause of death in elderly patients with Parkinson's disease: a 9-year follow-up.* Mov Disord, 2003. **18**(11): p. 1312-6.
495. Guttman, M., et al., *Parkinsonism in Ontario: comorbidity associated with hospitalization in a large cohort.* Mov Disord, 2004. **19**(1): p. 49-53.
496. D'Amelio, M., et al., *Long-term survival of Parkinson's disease: a population-based study.* J Neurol, 2006. **253**(1): p. 33-7.
497. Morgante, L., et al., *Parkinson disease survival: a population-based study.* Arch Neurol, 2000. **57**(4): p. 507-12.
498. Leibson, C.L., et al., *Comorbid conditions associated with Parkinson's disease: a population-based study.* Mov Disord, 2006. **21**(4): p. 446-55.
499. Driver, J.A., et al., *Parkinson disease and risk of mortality: a prospective comorbidity-matched cohort study.* Neurology, 2008. **70**(16 Pt 2): p. 1423-30.
500. Becker, C., et al., *Cancer risk in association with Parkinson disease: a population-based study.* Parkinsonism Relat Disord, 2010. **16**(3): p. 186-90.
501. Lo, R.Y., et al., *Comorbid cancer in Parkinson's disease.* Mov Disord, 2010. **25**(12): p. 1809-17.

502. Posada, I.J., et al., *Mortality from Parkinson's disease: a population-based prospective study (NEDICES).* Mov Disord, 2011. **26**(14): p. 2522-9.
503. Sun, L.M., et al., *Analysis of Parkinson's disease and subsequent cancer risk in Taiwan: a nationwide population-based cohort study.* Neuroepidemiology, 2011. **37**(2): p. 114-9.
504. Kareus, S.A., et al., *Shared predispositions of parkinsonism and cancer: a population-based pedigree-linked study.* Arch Neurol, 2012. **69**(12): p. 1572-7.
505. Moller, H., et al., *Occurrence of different cancers in patients with Parkinson's disease.* BMJ, 1995. **310**(6993): p. 1500-1.
506. Lai, S.W., et al., *Parkinson's disease and lung cancer: a population-based case-control study in Taiwan.* Geriatr Gerontol Int, 2013. **13**(1): p. 238-40.
507. Constantinescu, R., et al., *Malignant melanoma in early-treated Parkinson's disease: the NET-PD trial.* Mov Disord, 2014. **29**(2): p. 263-5.
508. Ong, E.L., R. Goldacre, and M. Goldacre, *Differential risks of cancer types in people with Parkinson's disease: a national record-linkage study.* Eur J Cancer, 2014. **50**(14): p. 2456-62.
509. Pinter, B., et al., *Mortality in Parkinson's disease: a 38-year follow-up study.* Mov Disord, 2015. **30**(2): p. 266-9.
510. Wirdefeldt, K., et al., *Parkinson's disease and cancer: A register-based family study.* Am J Epidemiol, 2014. **179**(1): p. 85-94.
511. Lin, P.Y., et al., *Association Between Parkinson Disease and Risk of Cancer in Taiwan.* JAMA Oncol, 2015. **1**(5): p. 633-40.
512. Boursi, B., et al., *Parkinson's disease and colorectal cancer risk-A nested case control study.* Cancer Epidemiol, 2016. **43**: p. 9-14.
513. Jespersen, C.G., M. Norgaard, and M. Borre, *Parkinson's disease and risk of prostate cancer: A Danish population-based case-control study, 1995-2010.* Cancer Epidemiol, 2016. **45**: p. 157-161.
514. Peretz, C., et al., *Cancer incidence among Parkinson's disease patients in a 10-yrs time-window around disease onset: A large-scale cohort study.* Parkinsonism Relat Disord, 2016. **28**: p. 68-72.
515. Tang, C.F., et al., *Increased risk of brain tumor in patients with Parkinson's disease: a nationwide cohort study in Taiwan.* Acta Neurol Scand, 2016. **134**(2): p. 148-53.
516. Tsuang, M.T., R.F. Woolson, and J.A. Fleming, *Causes of death in schizophrenia and manic-depression.* Br J Psychiatry, 1980. **136**: p. 239-42.
517. Herrman, H.E., J.A. Baldwin, and D. Christie, *A record-linkage study of mortality and general hospital discharge in patients diagnosed as schizophrenic.* Psychol Med, 1983. **13**(3): p. 581-93.
518. Brook, O.H., *Mortality in the long-stay population of Dutch mental hospitals.* Acta Psychiatr Scand, 1985. **71**(6): p. 626-35.
519. Kendler, K.S., *A twin study of mortality in schizophrenia and neurosis.* Arch Gen Psychiatry, 1986. **43**(7): p. 643-9.
520. Buda, M., M.T. Tsuang, and J.A. Fleming, *Causes of death in DSM-III schizophrenics and other psychotics (atypical group). A comparison with the general population.* Arch Gen Psychiatry, 1988. **45**(3): p. 283-5.
521. Allebeck, P. and B. Wistedt, *Mortality in schizophrenia. A ten-year follow-up based on the Stockholm County inpatient register.* Arch Gen Psychiatry, 1986. **43**(7): p. 650-3.
522. Mortensen, P.B., *The incidence of cancer in schizophrenic patients.* J Epidemiol Community Health, 1989. **43**(1): p. 43-7.
523. Zilber, N., N. Schufman, and Y. Lerner, *Mortality among psychiatric patients--the groups at risk.* Acta Psychiatr Scand, 1989. **79**(3): p. 248-56.
524. Mortensen, P.B. and K. Juel, *Mortality and causes of death in first admitted schizophrenic patients.* Br J Psychiatry, 1993. **163**: p. 183-9.
525. Mortensen, P.B., *The occurrence of cancer in first admitted schizophrenic patients.* Schizophr Res, 1994. **12**(3): p. 185-94.

526. Salazar-Fraile, J., et al., *Mortality among psychiatric patients referred to the mental health services in Valencia.* Soc Psychiatry Psychiatr Epidemiol, 1998. **33**(5): p. 224-9.
527. Osby, U., et al., *Mortality and causes of death in schizophrenia in Stockholm county, Sweden.* Schizophr Res, 2000. **45**(1-2): p. 21-8.
528. Joukamaa, M., et al., *Mental disorders and cause-specific mortality.* Br J Psychiatry, 2001. **179**: p. 498-502.
529. Lichtermann, D., et al., *Incidence of cancer among persons with schizophrenia and their relatives.* Arch Gen Psychiatry, 2001. **58**(6): p. 573-8.
530. Barak, Y., et al., *Reduced cancer incidence among patients with schizophrenia.* Cancer, 2005. **104**(12): p. 2817-21.
531. Barak, Y., et al., *Breast cancer in women suffering from serious mental illness.* Schizophr Res, 2008. **102**(1-3): p. 249-53.
532. Raviv, G., et al., *Risk of prostate cancer in patients with schizophrenia.* Compr Psychiatry, 2014. **55**(7): p. 1639-42.
533. Dalton, S.O., et al., *Risk for cancer in a cohort of patients hospitalized for schizophrenia in Denmark, 1969-1993.* Schizophr Res, 2005. **75**(2-3): p. 315-24.
534. Goldacre, M.J., et al., *Schizophrenia and cancer: an epidemiological study.* Br J Psychiatry, 2005. **187**: p. 334-8.
535. Grinshpoon, A., et al., *Cancer in schizophrenia: is the risk higher or lower?* Schizophr Res, 2005. **73**(2-3): p. 333-41.
536. Heila, H., et al., *Mortality among patients with schizophrenia and reduced psychiatric hospital care.* Psychol Med, 2005. **35**(5): p. 725-32.
537. Kisely, S., et al., *Mortality in individuals who have had psychiatric treatment: population-based study in Nova Scotia.* Br J Psychiatry, 2005. **187**: p. 552-8.
538. Amaddeo, F., *Risk of death from natural causes increased in people with psychiatric admissions.* Evid Based Ment Health, 2008. **11**(4): p. 106.
539. Tran, E., et al., *Cancer mortality in patients with schizophrenia: an 11-year prospective cohort study.* Cancer, 2009. **115**(15): p. 3555-62.
540. Brown, S., et al., *Twenty-five year mortality of a community cohort with schizophrenia.* Br J Psychiatry, 2010. **196**(2): p. 116-21.
541. Brown, S., H. Inskip, and B. Barraclough, *Causes of the excess mortality of schizophrenia.* Br J Psychiatry, 2000. **177**: p. 212-7.
542. Chou, F.H., et al., *The incidence and relative risk factors for developing cancer among patients with schizophrenia: a nine-year follow-up study.* Schizophr Res, 2011. **129**(2-3): p. 97-103.
543. Truyers, C., et al., *Incident somatic comorbidity after psychosis: results from a retrospective cohort study based on Flemish general practice data.* BMC Fam Pract, 2011. **12**: p. 132.
544. Talaslahti, T., et al., *Mortality and causes of death in older patients with schizophrenia.* Int J Geriatr Psychiatry, 2012. **27**(11): p. 1131-7.
545. Whitley, E., et al., *Psychiatric disorder as a risk factor for cancer: different analytic strategies produce different findings.* Epidemiology, 2012. **23**(4): p. 543-50.
546. Ji, J., et al., *Incidence of cancer in patients with schizophrenia and their first-degree relatives: a population-based study in Sweden.* Schizophr Bull, 2013. **39**(3): p. 527-36.
547. Kredentser, M.S., et al., *Cause and rate of death in people with schizophrenia across the lifespan: a population-based study in Manitoba, Canada.* J Clin Psychiatry, 2014. **75**(2): p. 154-61.
548. Perini, G., et al., *Cancer mortality among psychiatric patients treated in a community-based system of care: a 25-year case register study.* Soc Psychiatry Psychiatr Epidemiol, 2014. **49**(5): p. 693-701.
549. Grigoletti, L., et al., *Mortality and cause of death among psychiatric patients: a 20-year case-register study in an area with a community-based system of care.* Psychol Med, 2009. **39**(11): p. 1875-84.

550. Lesage, A., et al., *A Surveillance System to Monitor Excess Mortality of People With Mental Illness in Canada.* Can J Psychiatry, 2015. **60**(12): p. 571-9.
551. Olfson, M., et al., *Premature Mortality Among Adults With Schizophrenia in the United States.* JAMA Psychiatry, 2015. **72**(12): p. 1172-81.
552. Park, S., S.Y. Kim, and J.P. Hong, *Cause-specific mortality of psychiatric inpatients and outpatients in a general hospital in Korea.* Asia Pac J Public Health, 2015. **27**(2): p. 164-75.
553. Wu Chou, A.I., et al., *Female schizophrenia patients and risk of breast cancer: A population-based cohort study.* Schizophr Res, 2017. **188**: p. 165-171.
554. Chen, L.Y., et al., *Cancer incidence in young and middle-aged people with schizophrenia: nationwide cohort study in Taiwan, 2000-2010.* Epidemiol Psychiatr Sci, 2018. **27**(2): p. 146-156.
555. Dalton, S.O., et al., *Impact of schizophrenia and related disorders on mortality from breast cancer: A population-based cohort study in Denmark, 1995-2011.* Breast, 2018. **40**: p. 170-176.
556. Tanskanen, A., J. Tiihonen, and H. Taipale, *Mortality in schizophrenia: 30-year nationwide follow-up study.* Acta Psychiatr Scand, 2018. **138**(6): p. 492-499.
557. Blalock, E.M., et al., *Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses.* Proc Natl Acad Sci U S A, 2004. **101**(7): p. 2173-8.
558. Readhead, B., et al., *Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus.* Neuron, 2018. **99**(1): p. 64-82 e7.
559. Bronner, I.F., et al., *Comprehensive mRNA expression profiling distinguishes tauopathies and identifies shared molecular pathways.* PLoS One, 2009. **4**(8): p. e6826.
560. Hokama, M., et al., *Altered expression of diabetes-related genes in Alzheimer's disease brains: the Hisayama study.* Cereb Cortex, 2014. **24**(9): p. 2476-88.
561. Miller, J.A., et al., *Genes and pathways underlying regional and cell type changes in Alzheimer's disease.* Genome Med, 2013. **5**(5): p. 48.
562. Sarvari, M., et al., *Menopause leads to elevated expression of macrophage-associated genes in the aging frontal cortex: rat and human studies identify strikingly similar changes.* J Neuroinflammation, 2012. **9**: p. 264.
563. Wang, M., et al., *Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease.* Genome Med, 2016. **8**(1): p. 104.
564. Voineagu, I., et al., *Transcriptomic analysis of autistic brain reveals convergent molecular pathology.* Nature, 2011. **474**(7351): p. 380-4.
565. Chow, M.L., et al., *Genome-wide expression assay comparison across frozen and fixed postmortem brain tissue samples.* BMC Genomics, 2011. **12**: p. 449.
566. Gupta, S., et al., *Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism.* Nat Commun, 2014. **5**: p. 5748.
567. Hernandez, D.G., et al., *Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain.* Neurobiol Dis, 2012. **47**(1): p. 20-8.
568. Lanz, T.A., et al., *Postmortem transcriptional profiling reveals widespread increase in inflammation in schizophrenia: a comparison of prefrontal cortex, striatum, and hippocampus among matched tetrads of controls with subjects diagnosed with schizophrenia, bipolar or major depressive disorder.* Transl Psychiatry, 2019. **9**(1): p. 151.
569. Hagenauer, M.H., et al., *Inference of cell type content from human brain transcriptomic datasets illuminates the effects of age, manner of death, dissection, and psychiatric diagnosis.* PLoS One, 2018. **13**(7): p. e0200003.

570. Jurata, L.W., et al., *Comparison of microarray-based mRNA profiling technologies for identification of psychiatric disease and drug signatures.* J Neurosci Methods, 2004. **138**(1-2): p. 173-88.
571. Struyf, J., S. Dobrin, and D. Page, *Combining gene expression, demographic and clinical data in modeling disease: a case study of bipolar disorder and schizophrenia.* BMC Genomics, 2008. **9**: p. 531.
572. Hodges, A., et al., *Regional and cellular gene expression changes in human Huntington's disease brain.* Hum Mol Genet, 2006. **15**(6): p. 965-77.
573. Durrenberger, P.F., et al., *Common mechanisms in neurodegeneration and neuroinflammation: a BrainNet Europe gene expression microarray study.* J Neural Transm (Vienna), 2015. **122**(7): p. 1055-68.
574. Chang, L.C., et al., *A conserved BDNF, glutamate- and GABA-enriched gene module related to human depression identified by coexpression meta-analysis and DNA variant genome-wide association studies.* PLoS One, 2014. **9**(3): p. e90980.
575. Lesnick, T.G., et al., *A genomic pathway approach to a complex disease: axon guidance and Parkinson disease.* PLoS Genet, 2007. **3**(6): p. e98.
576. Duke, D.C., et al., *The medial and lateral substantia nigra in Parkinson's disease: mRNA profiles associated with higher brain tissue vulnerability.* Neurogenetics, 2007. **8**(2): p. 83-94.
577. Cantuti-Castelvetri, I., et al., *Effects of gender on nigral gene expression and parkinson disease.* Neurobiol Dis, 2007. **26**(3): p. 606-14.
578. Zheng, B., et al., *PGC-1alpha, a potential therapeutic target for early intervention in Parkinson's disease.* Sci Transl Med, 2010. **2**(52): p. 52ra73.
579. Dijkstra, A.A., et al., *Evidence for Immune Response, Axonal Dysfunction and Reduced Endocytosis in the Substantia Nigra in Early Stage Parkinson's Disease.* PLoS One, 2015. **10**(6): p. e0128651.
580. Riley, B.E., et al., *Systems-based analyses of brain regions functionally impacted in Parkinson's disease reveals underlying causal mechanisms.* PLoS One, 2014. **9**(8): p. e102909.
581. Mirnics, K., et al., *Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex.* Neuron, 2000. **28**(1): p. 53-67.
582. McCullumsmith, R.E., et al., *Expression of transcripts for myelination-related genes in the anterior cingulate cortex in schizophrenia.* Schizophr Res, 2007. **90**(1-3): p. 15-27.
583. Maycox, P.R., et al., *Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function.* Mol Psychiatry, 2009. **14**(12): p. 1083-94.
584. Narayan, S., et al., *Molecular profiles of schizophrenia in the CNS at different stages of illness.* Brain Res, 2008. **1239**: p. 235-48.
585. Kuhnl, A., et al., *High BAALC expression predicts chemoresistance in adult B-precursor acute lymphoblastic leukemia.* Blood, 2010. **115**(18): p. 3737-44.
586. Homminga, I., et al., *Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia.* Cancer Cell, 2011. **19**(4): p. 484-97.
587. Coustan-Smith, E., et al., *New markers for minimal residual disease detection in acute lymphoblastic leukemia.* Blood, 2011. **117**(23): p. 6267-76.
588. Hatirnaz Ng, O., et al., *A Possible Role for WNT5A Hypermethylation in Pediatric Acute Lymphoblastic Leukemia.* Turk J Haematol, 2015. **32**(2): p. 127-35.
589. Hirabayashi, S., et al., *ZNF384-related fusion genes define a subgroup of childhood B-cell precursor acute lymphoblastic leukemia with a characteristic immunotype.* Haematologica, 2017. **102**(1): p. 118-129.
590. Payton, J.E., et al., *High throughput digital quantification of mRNA abundance in primary human acute myeloid leukemia samples.* J Clin Invest, 2009. **119**(6): p. 1714-26.

591. Bosman, M.C., et al., *The TAK1-NF-kappaB axis as therapeutic target for AML.* Blood, 2014. **124**(20): p. 3130-40.
592. de la Bletiere, D.R., et al., *Routine use of microarray-based gene expression profiling to identify patients with low cytogenetic risk acute myeloid leukemia: accurate results can be obtained even with suboptimal samples.* BMC Med Genomics, 2012. **5**: p. 6.
593. Cramer-Morales, K., et al., *Personalized synthetic lethality induced by targeting RAD52 in leukemias identified by gene mutation and expression profile.* Blood, 2013. **122**(7): p. 1293-304.
594. Jung, N., et al., *An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis.* Nat Commun, 2015. **6**: p. 8489.
595. Pont, M.J., et al., *Microarray Gene Expression Analysis to Evaluate Cell Type Specific Expression of Targets Relevant for Immunotherapy of Hematological Malignancies.* PLoS One, 2016. **11**(5): p. e0155165.
596. Mengual, L., et al., *DNA microarray expression profiling of bladder cancer allows identification of noninvasive diagnostic markers.* J Urol, 2009. **182**(2): p. 741-8.
597. Kim, S.K., et al., *Identification of S100A8-correlated genes for prediction of disease progression in non-muscle invasive bladder cancer.* BMC Cancer, 2010. **10**: p. 21.
598. Borisov, N., et al., *A method of gene expression data transfer from cell lines to cancer patients for machine-learning prediction of drug efficiency.* Cell Cycle, 2018. **17**(4): p. 486-491.
599. Maraver, A., et al., *NOTCH pathway inactivation promotes bladder cancer progression.* J Clin Invest, 2015. **125**(2): p. 824-30.
600. El Behi, M., et al., *An essential role for decorin in bladder cancer invasiveness.* EMBO Mol Med, 2013. **5**(12): p. 1835-51.
601. Richardson, A.L., et al., *X chromosomal abnormalities in basal-like human breast cancer.* Cancer Cell, 2006. **9**(2): p. 121-32.
602. Chen, D.T., et al., *Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue.* Breast Cancer Res Treat, 2010. **119**(2): p. 335-46.
603. Pedraza, V., et al., *Gene expression signatures in breast cancer distinguish phenotype characteristics, histologic subtypes, and tumor invasiveness.* Cancer, 2010. **116**(2): p. 486-96.
604. Sabatier, R., et al., *Down-regulation of ECRG4, a candidate tumor suppressor gene, in human breast cancer.* PLoS One, 2011. **6**(11): p. e27656.
605. Clarke, C., et al., *Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis.* Carcinogenesis, 2013. **34**(10): p. 2300-8.
606. Tan, T.Z., et al., *Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients.* EMBO Mol Med, 2014. **6**(10): p. 1279-93.
607. Gruosso, T., et al., *Chronic oxidative stress promotes H2AX protein degradation and enhances chemosensitivity in breast cancer patients.* EMBO Mol Med, 2016. **8**(5): p. 527-49.
608. Sun, L., et al., *Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain.* Cancer Cell, 2006. **9**(4): p. 287-300.
609. French, P.J., et al., *Identification of differentially regulated splice variants and novel exons in glial brain tumors using exon expression arrays.* Cancer Res, 2007. **67**(12): p. 5635-42.
610. Gravendeel, L.A., et al., *Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology.* Cancer Res, 2009. **69**(23): p. 9065-72.
611. Liu, Z., et al., *Gene expression profiling in human high-grade astrocytomas.* Comp Funct Genomics, 2011. **2011**: p. 245137.
612. Grzmil, M., et al., *MAP kinase-interacting kinase 1 regulates SMAD2-dependent TGF-beta signaling pathway in human glioblastoma.* Cancer Res, 2011. **71**(6): p. 2392-402.

613. Henriquez, N.V., et al., *Comparative expression analysis reveals lineage relationships between human and murine gliomas and a dominance of glial signatures during tumor propagation in vitro.* Cancer Res, 2013. **73**(18): p. 5834-44.
614. Lambert, S.R., et al., *Differential expression and methylation of brain developmental genes define location-specific subsets of pilocytic astrocytoma.* Acta Neuropathol, 2013. **126**(2): p. 291-301.
615. Griesinger, A.M., et al., *Characterization of distinct immunophenotypes across pediatric brain tumor types.* J Immunol, 2013. **191**(9): p. 4880-8.
616. Madhavan, S., et al., *Rembrandt: helping personalized medicine become a reality through integrative translational research.* Mol Cancer Res, 2009. **7**(2): p. 157-67.
617. de Bont, J.M., et al., *Differential expression and prognostic significance of SOX genes in pediatric medulloblastoma and ependymoma identified by microarray analysis.* Neuro Oncol, 2008. **10**(5): p. 648-60.
618. Pyeon, D., et al., *Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers.* Cancer Res, 2007. **67**(10): p. 4605-19.
619. Espinosa, A.M., et al., *Mitosis is a source of potential markers for screening and survival and therapeutic targets in cervical cancer.* PLoS One, 2013. **8**(2): p. e55975.
620. den Boon, J.A., et al., *Molecular transitions from papillomavirus infection to cervical precancer and cancer: Role of stromal estrogen receptor signaling.* Proc Natl Acad Sci U S A, 2015. **112**(25): p. E3255-64.
621. Saha, S.S., et al., *Expression signatures of HOX cluster genes in cervical cancer pathogenesis: Impact of human papillomavirus type 16 oncoprotein E7.* Oncotarget, 2017. **8**(22): p. 36591-36602.
622. Andersen, J.B., et al., *Genomic and genetic characterization of cholangiocarcinoma identifies therapeutic targets for tyrosine kinase inhibitors.* Gastroenterology, 2012. **142**(4): p. 1021-1031 e15.
623. Oishi, N., et al., *Transcriptomic profiling reveals hepatic stem-like gene signatures and interplay of miR-200c and epithelial-mesenchymal transition in intrahepatic cholangiocarcinoma.* Hepatology, 2012. **56**(5): p. 1792-803.
624. Sia, D., et al., *Integrative molecular analysis of intrahepatic cholangiocarcinoma reveals 2 classes that have different outcomes.* Gastroenterology, 2013. **144**(4): p. 829-40.
625. Seol, M.A., et al., *Genome-wide expression patterns associated with oncogenesis and sarcomatous transdifferentation of cholangiocarcinoma.* BMC Cancer, 2011. **11**: p. 78.
626. Meyers, J.A., D.W. Su, and A. Lerner, *Chronic lymphocytic leukemia and B and T cells differ in their response to cyclic nucleotide phosphodiesterase inhibitors.* J Immunol, 2009. **182**(9): p. 5400-11.
627. Vargova, K., et al., *MYB transcriptionally regulates the miR-155 host gene in chronic lymphocytic leukemia.* Blood, 2011. **117**(14): p. 3816-25.
628. Wang, L., et al., *Somatic mutation as a mechanism of Wnt/beta-catenin pathway activation in CLL.* Blood, 2014. **124**(7): p. 1089-98.
629. Maura, F., et al., *Insulin growth factor 1 receptor expression is associated with NOTCH1 mutation, trisomy 12 and aggressive clinical course in chronic lymphocytic leukaemia.* PLoS One, 2015. **10**(3): p. e0118801.
630. Filarsky, K., et al., *Kruppel-like factor 4 (KLF4) inactivation in chronic lymphocytic leukemia correlates with promoter DNA-methylation and can be reversed by inhibition of NOTCH signaling.* Haematologica, 2016. **101**(6): p. e249-53.
631. Gerber, J.M., et al., *Genome-wide comparison of the transcriptomes of highly enriched normal and chronic myeloid leukemia stem and progenitor cell populations.* Oncotarget, 2013. **4**(5): p. 715-28.
632. Abraham, S.A., et al., *Dual targeting of p53 and c-MYC selectively eliminates leukaemic stem cells.* Nature, 2016. **534**(7607): p. 341-6.

633. Hong, Y., et al., *A 'metastasis-prone' signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics.* Clin Exp Metastasis, 2010. **27**(2): p. 83-90.
634. Matsuyama, T., et al., *MUC12 mRNA expression is an independent marker of prognosis in stage II and stage III colorectal cancer.* Int J Cancer, 2010. **127**(10): p. 2292-9.
635. Skrzypczak, M., et al., *Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability.* PLoS One, 2010. **5**(10).
636. Danielsen, S.A., et al., *Phospholipase C isozymes are deregulated in colorectal cancer--insights gained from gene set enrichment analysis of the transcriptome.* PLoS One, 2011. **6**(9): p. e24419.
637. Abba, M., et al., *Look who's talking: deregulated signaling in colorectal cancer.* Cancer Genomics Proteomics, 2012. **9**(1): p. 15-25.
638. Kemper, K., et al., *Mutations in the Ras-Raf Axis underlie the prognostic value of CD133 in colorectal cancer.* Clin Cancer Res, 2012. **18**(11): p. 3132-41.
639. Musella, V., et al., *Effects of warm ischemic time on gene expression profiling in colorectal cancer tissues and normal mucosa.* PLoS One, 2013. **8**(1): p. e53406.
640. Galamb, O., et al., *Dysplasia-carcinoma transition specific transcripts in colonic biopsy samples.* PLoS One, 2012. **7**(11): p. e48547.
641. Marisa, L., et al., *Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value.* PLoS Med, 2013. **10**(5): p. e1001453.
642. Perilli, L., et al., *Human miRNome profiling in colorectal cancer and liver metastasis development.* Genom Data, 2014. **2**: p. 184-8.
643. Moreno, V., et al., *Colon-specific eQTL analysis to inform on functional SNPs.* Br J Cancer, 2018. **119**(8): p. 971-977.
644. Chen, X., et al., *Comparison of Nanostring nCounter(R) Data on FFPE Colon Cancer Samples and Affymetrix Microarray Data on Matched Frozen Tissues.* PLoS One, 2016. **11**(5): p. e0153784.
645. An, N., et al., *Discovery of a Novel Immune Gene Signature with Profound Prognostic Value in Colorectal Cancer: A Model of Cooperativity Disorientation Created in the Process from Development to Cancer.* PLoS One, 2015. **10**(9): p. e0137171.
646. Weniger, M.A., et al., *Human CD30+ B cells represent a unique subset related to Hodgkin lymphoma cells.* J Clin Invest, 2018. **128**(7): p. 2996-3007.
647. Garcia-Ramirez, I., et al., *Crebbp loss cooperates with Bcl2 overexpression to promote lymphoma in mice.* Blood, 2017. **129**(19): p. 2645-2656.
648. Norgaard, C.H., et al., *Subtype assignment of CLL based on B-cell subset associated gene signatures from normal bone marrow - A proof of concept study.* PLoS One, 2018. **13**(3): p. e0193249.
649. Killian, J.K., et al., *Large-scale profiling of archival lymph nodes reveals pervasive remodeling of the follicular lymphoma methylome.* Cancer Res, 2009. **69**(3): p. 758-64.
650. Guo, S., et al., *EZH2 mutations in follicular lymphoma from different ethnic groups and associated gene expression alterations.* Clin Cancer Res, 2014. **20**(12): p. 3078-86.
651. Newman, A.M., et al., *Robust enumeration of cell subsets from tissue expression profiles.* Nat Methods, 2015. **12**(5): p. 453-7.
652. Ye, H., et al., *Transcriptomic dissection of tongue squamous cell carcinoma.* BMC Genomics, 2008. **9**: p. 69.
653. Bhosale, P.G., et al., *Chromosomal Alterations and Gene Expression Changes Associated with the Progression of Leukoplakia to Advanced Gingivobuccal Cancer.* Transl Oncol, 2017. **10**(3): p. 396-409.
654. Peng, C.H., et al., *A novel molecular signature identified by systems genetics approach predicts prognosis in oral squamous cell carcinoma.* PLoS One, 2011. **6**(8): p. e23452.

655. Chen, C., et al., *Gene expression profiling identifies genes predictive of oral squamous cell carcinoma.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(8): p. 2152-62.
656. Rentoft, M., et al., *Transcriptional profiling of formalin fixed paraffin embedded tissue: pitfalls and recommendations for identifying biologically relevant changes.* PLoS One, 2012. **7**(4): p. e35276.
657. Demokan, S., et al., *Identification of guanine nucleotide-binding protein gamma-7 as an epigenetically silenced gene in head and neck cancer by gene expression profiling.* Int J Oncol, 2013. **42**(4): p. 1427-36.
658. Tomar, S., et al., *Human papillomavirus status and gene expression profiles of oropharyngeal and oral cancers from European American and African American patients.* Head Neck, 2016. **38 Suppl 1**: p. E694-704.
659. Krishnan, N.M., et al., *A Minimal DNA Methylation Signature in Oral Tongue Squamous Cell Carcinoma Links Altered Methylation with Tumor Attributes.* Mol Cancer Res, 2016. **14**(9): p. 805-19.
660. Kort, E.J., et al., *The E2F3-Oncomir-1 axis is activated in Wilms' tumor.* Cancer Res, 2008. **68**(11): p. 4034-8.
661. Ding, Y., et al., *Combined gene expression profiling and RNAi screening in clear cell renal cell carcinoma identify PLK1 and other therapeutic kinase targets.* Cancer Res, 2011. **71**(15): p. 5225-34.
662. Pena-Llopis, S. and J. Brugarolas, *Simultaneous isolation of high-quality DNA, RNA, miRNA and proteins from tissues for genomic applications.* Nat Protoc, 2013. **8**(11): p. 2240-55.
663. Wozniak, M.B., et al., *Integrative genome-wide gene expression profiling of clear cell renal cell carcinoma in Czech Republic and in the United States.* PLoS One, 2013. **8**(3): p. e57886.
664. Valletti, A., et al., *Genome-wide analysis of differentially expressed genes and splicing isoforms in clear cell renal cell carcinoma.* PLoS One, 2013. **8**(10): p. e78452.
665. Eckel-Passow, J.E., et al., *Somatic expression of ENRAGE is associated with obesity status among patients with clear cell renal cell carcinoma.* Carcinogenesis, 2014. **35**(4): p. 822-7.
666. Liep, J., et al., *Cooperative Effect of miR-141-3p and miR-145-5p in the Regulation of Targets in Clear Cell Renal Cell Carcinoma.* PLoS One, 2016. **11**(6): p. e0157801.
667. Thibodeau, B.J., et al., *Characterization of clear cell renal cell carcinoma by gene expression profiling.* Urol Oncol, 2016. **34**(4): p. 168 e1-9.
668. Takahashi, M., et al., *Downregulation of WDR20 due to loss of 14q is involved in the malignant transformation of clear cell renal cell carcinoma.* Cancer Sci, 2016. **107**(4): p. 417-23.
669. Sanchez-Palencia, A., et al., *Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer.* Int J Cancer, 2011. **129**(2): p. 355-64.
670. Hou, J., et al., *Gene expression-based classification of non-small cell lung carcinomas and survival prediction.* PLoS One, 2010. **5**(4): p. e10312.
671. Lu, T.P., et al., *Identification of regulatory SNPs associated with genetic modifications in lung adenocarcinoma.* BMC Res Notes, 2015. **8**: p. 92.
672. Yamauchi, M., et al., *Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma.* PLoS One, 2012. **7**(9): p. e43923.
673. Lin, J., et al., *Lung cancer transcriptomes refined with laser capture microdissection.* Am J Pathol, 2014. **184**(11): p. 2868-84.
674. Selamat, S.A., et al., *Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression.* Genome Res, 2012. **22**(7): p. 1197-211.
675. Rousseaux, S., et al., *Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers.* Sci Transl Med, 2013. **5**(186): p. 186ra66.
676. Zhang, Y., et al., *USP44 regulates centrosome positioning to prevent aneuploidy and suppress tumorigenesis.* J Clin Invest, 2012. **122**(12): p. 4362-74.
677. Kim, I.J., et al., *Rewiring of human lung cell lineage and mitotic networks in lung adenocarcinomas.* Nat Commun, 2013. **4**: p. 1701.

678. Kabbout, M., et al., *ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer.* Clin Cancer Res, 2013. **19**(13): p. 3383-95.
679. Girard, L., et al., *An Expression Signature as an Aid to the Histologic Classification of Non-Small Cell Lung Cancer.* Clin Cancer Res, 2016. **22**(19): p. 4880-4889.
680. Wurmbach, E., et al., *Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma.* Hepatology, 2007. **45**(4): p. 938-47.
681. Archer, K.J., et al., *Identifying genes for establishing a multigenic test for hepatocellular carcinoma surveillance in hepatitis C virus-positive cirrhotic patients.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(11): p. 2929-32.
682. Satow, R., et al., *Combined functional genome survey of therapeutic targets for hepatocellular carcinoma.* Clin Cancer Res, 2010. **16**(9): p. 2518-28.
683. Lim, H.Y., et al., *Prediction of disease-free survival in hepatocellular carcinoma by gene expression profiling.* Ann Surg Oncol, 2013. **20**(12): p. 3747-53.
684. Neumann, O., et al., *Methylome analysis and integrative profiling of human HCCs identify novel protumorigenic factors.* Hepatology, 2012. **56**(5): p. 1817-27.
685. Boormans, J.L., et al., *Identification of TDRD1 as a direct target gene of ERG in primary prostate cancer.* Int J Cancer, 2013. **133**(2): p. 335-45.
686. Yildiz, G., et al., *Genome-wide transcriptional reorganization associated with senescence-to-immortality switch during human hepatocellular carcinogenesis.* PLoS One, 2013. **8**(5): p. e64016.
687. Zubiete-Franco, I., et al., *SUMOylation regulates LKB1 localization and its oncogenic activity in liver cancer.* EBioMedicine, 2019. **40**: p. 406-421.
688. Melis, M., et al., *Viral expression and molecular profiling in liver tissue versus microdissected hepatocytes in hepatitis B virus-associated hepatocellular carcinoma.* J Transl Med, 2014. **12**: p. 230.
689. Schulze, K., et al., *Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets.* Nat Genet, 2015. **47**(5): p. 505-511.
690. Grinchuk, O.V., et al., *Tumor-adjacent tissue co-expression profile analysis reveals pro-oncogenic ribosomal gene signature for prognosis of resectable hepatocellular carcinoma.* Mol Oncol, 2018. **12**(1): p. 89-113.
691. Makowska, Z., et al., *Gene expression analysis of biopsy samples reveals critical limitations of transcriptome-based molecular classifications of hepatocellular carcinoma.* J Pathol Clin Res, 2016. **2**(2): p. 80-92.
692. Tone, A.A., et al., *Decreased progesterone receptor isoform expression in luteal phase fallopian tube epithelium and high-grade serous carcinoma.* Endocr Relat Cancer, 2011. **18**(2): p. 221-34.
693. Bowen, N.J., et al., *Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells.* BMC Med Genomics, 2009. **2**: p. 71.
694. Elgaaen, B.V., et al., *ZNF385B and VEGFA are strongly differentially expressed in serous ovarian carcinomas and correlate with survival.* PLoS One, 2012. **7**(9): p. e46317.
695. Yeung, T.L., et al., *TGF-beta modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment.* Cancer Res, 2013. **73**(16): p. 5016-28.
696. Yamamoto, Y., et al., *In vitro and in vivo correlates of physiological and neoplastic human Fallopian tube stem cells.* J Pathol, 2016. **238**(4): p. 519-530.
697. Idichi, T., et al., *Regulation of actin-binding protein ANLN by antitumor miR-217 inhibits cancer cell aggressiveness in pancreatic ductal adenocarcinoma.* Oncotarget, 2017. **8**(32): p. 53180-53193.
698. Hiraoka, N., et al., *CXCL17 and ICAM2 are associated with a potential anti-tumor immune response in early intraepithelial stages of human pancreatic carcinogenesis.* Gastroenterology, 2011. **140**(1): p. 310-21.

699. Donahue, T.R., et al., *Integrative survival-based molecular profiling of human pancreatic cancer.* Clin Cancer Res, 2012. **18**(5): p. 1352-63.
700. Frampton, A.E., et al., *MicroRNAs cooperatively inhibit a network of tumor suppressor genes to promote pancreatic tumor growth and progression.* Gastroenterology, 2014. **146**(1): p. 268-77 e18.
701. Lunardi, S., et al., *IP-10/CXCL10 induction in human pancreatic cancer stroma influences lymphocytes recruitment and correlates with poor survival.* Oncotarget, 2014. **5**(22): p. 11064-80.
702. Janky, R., et al., *Prognostic relevance of molecular subtypes and master regulators in pancreatic ductal adenocarcinoma.* BMC Cancer, 2016. **16**: p. 632.
703. Yang, S., et al., *A Novel MIF Signaling Pathway Drives the Malignant Character of Pancreatic Cancer by Targeting NR3C2.* Cancer Res, 2016. **76**(13): p. 3838-50.
704. Wang, J., et al., *Splice variants as novel targets in pancreatic ductal adenocarcinoma.* Sci Rep, 2017. **7**(1): p. 2980.
705. Taylor, B.S., et al., *Integrative genomic profiling of human prostate cancer.* Cancer Cell, 2010. **18**(1): p. 11-22.
706. Borno, S.T., et al., *Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation.* Cancer Discov, 2012. **2**(11): p. 1024-35.
707. Labbe, D.P., et al., *High-fat diet fuels prostate cancer progression by rewiring the metabolome and amplifying the MYC program.* Nat Commun, 2019. **10**(1): p. 4358.
708. Mortensen, M.M., et al., *Expression profiling of prostate cancer tissue delineates genes associated with recurrence after prostatectomy.* Sci Rep, 2015. **5**: p. 16018.
709. Ross-Adams, H., et al., *Integration of copy number and transcriptomics provides risk stratification in prostate cancer: A discovery and validation cohort study.* EBioMedicine, 2015. **2**(9): p. 1133-44.
710. Zhang, L., et al., *MNX1 Is Oncogenically Upregulated in African-American Prostate Cancer.* Cancer Res, 2016. **76**(21): p. 6290-6298.
711. Talantov, D., et al., *Novel genes associated with malignant melanoma but not benign melanocytic lesions.* Clin Cancer Res, 2005. **11**(20): p. 7234-42.
712. Riker, A.I., et al., *The gene expression profiles of primary and metastatic melanoma yields a transition point of tumor progression and metastasis.* BMC Med Genomics, 2008. **1**: p. 13.
713. Raskin, L., et al., *Transcriptome profiling identifies HMGA2 as a biomarker of melanoma progression and prognosis.* J Invest Dermatol, 2013. **133**(11): p. 2585-2592.
714. Kabbarah, O., et al., *Integrative genome comparison of primary and metastatic melanomas.* PLoS One, 2010. **5**(5): p. e10770.
715. Staaf, J., et al., *Primary melanoma tumors from CDKN2A mutation carriers do not belong to a distinct molecular subclass.* J Invest Dermatol, 2014. **134**(12): p. 3000-3003.
716. D'Errico, M., et al., *Genome-wide expression profile of sporadic gastric cancers with microsatellite instability.* Eur J Cancer, 2009. **45**(3): p. 461-9.
717. Cui, J., et al., *Gene-expression signatures can distinguish gastric cancer grades and stages.* PLoS One, 2011. **6**(3): p. e17819.
718. Holbrook, J.D., et al., *Deep sequencing of gastric carcinoma reveals somatic mutations relevant to personalized medicine.* J Transl Med, 2011. **9**: p. 119.
719. Oh, S.C., et al., *Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype.* Nat Commun, 2018. **9**(1): p. 1777.
720. Jin, Y., et al., *Overexpression of HS6ST2 is associated with poor prognosis in patients with gastric cancer.* Oncol Lett, 2017. **14**(5): p. 6191-6197.
721. He, H., et al., *The role of microRNA genes in papillary thyroid carcinoma.* Proc Natl Acad Sci U S A, 2005. **102**(52): p. 19075-80.
722. Pita, J.M., et al., *Gene expression profiling associated with the progression to poorly differentiated thyroid carcinomas.* Br J Cancer, 2009. **101**(10): p. 1782-91.

723. Handkiewicz-Junak, D., et al., *Gene signature of the post-Chernobyl papillary thyroid cancer.* Eur J Nucl Med Mol Imaging, 2016. **43**(7): p. 1267-77.
724. Tarabichi, M., et al., *Revisiting the transcriptional analysis of primary tumours and associated nodal metastases with enhanced biological and statistical controls: application to thyroid cancer.* Br J Cancer, 2015. **112**(10): p. 1665-74.
725. von Roemeling, C.A., et al., *Aberrant lipid metabolism in anaplastic thyroid carcinoma reveals stearoyl CoA desaturase 1 as a novel therapeutic target.* J Clin Endocrinol Metab, 2015. **100**(5): p. E697-709.

8. Appendices

Appendix 1: ICD-9 and ICD-10 codes for chapter's two primary and secondary outcomes

CNS disorders	**ICD codes 9**	**ICD codes 10**	**Smoking- related**
Alzheimer's disease	331.0	G30.9	NA
Autism spectrum disorders	299.0	F84.0	NA
Bipolar disorder	296.80	F31	NA
Huntington's disease	333.4	G10	NA
Major depression	296.3	F33	NA
Parkinson's disease	332	G20	NA
Schizophrenia	295	F20.9	NA
Cancers	**ICD codes 9**	**ICD codes 10**	**Smoking-related**
Overall cancer	140-239	C00-D48	NA
Acute lymphoblastic leukemia	204-208.92 (leukemia)	C91-C95.92 (leukemia)	Yes
Acute myeloid leukemia	204-208.92 (leukemia)	C91-C95.92 (leukemia)	Yes
Bladder cancer	188-188.9	C67-C67.9	Yes
Brain and nervous system cancer	191-192.9	C70-C72.9	No
Breast cancer	174-175.9	C50-C50.929	No
Cervical cancer	180-180.9	C53-C53.9	Yes
Cholangiocarcinoma	156-156.9, 209.25-209.27 (Gallbladder and biliary tract cancer)	C20.8, C23-C24.9 (Gallbladder and biliary tract cancer)	No

Chronic lymphocytic leukemia	204-208.92 (leukemia)	C91-C95.92 (leukemia)	Yes
Chronic myeloid leukemia	204-208.92 (leukemia)	C91-C95.92 (leukemia)	Yes
Colorectal cancer	153-154.9, 155.5, 155.9, 209.1-209.17	C18-C20.0, C20.9-C21.8	Yes
Diffuse large B-cell lymphoma	200-200.9, 202-202.98 (non-Hodgkin lymphoma)	C82-C86.6, C96-C97.9 (non-Hodgkin lymphoma)	No
Esophageal cancer	150-150.9	C15-C15.9	Yes
Follicular Lymphoma	200-200.9, 202-202.98 (non-Hodgkin lymphoma)	C82-C86.6, C96-C97.9 (non-Hodgkin lymphoma)	No
Gallbladder and biliary tract cancer	156-156.9, 209.25-209.27	C20.8,C23-C24.9	No
Stomach cancer	151-151.9, 209.23	C16-C16.9	Yes
Head and neck carcinoma	195.0	C76.0	Yes
Hodgkin lymphoma	201-201.98	C81-C81.99	No
Kidney cancer	189.0, 189.1, 209.24	C64-C65.9	Yes
Lip and oral cavity cancer	140-145.9	C00-C08.9	Yes
Larynx cancer	161-161.9, 162.1	C32-C32.9	Yes
Leukemia	204-208.92	C91-C95.2	Yes
Liver cancer	155-155.3	C22-C22.9	Yes
Tracheal, bronchus, and lung cancer	162, 162.0, 162.2-162.9, 209.21	C33-C34.9	Yes
Melanoma	172-172.9	C43-C43.9, C4A	No
Mesothelioma	58-158.9, 163-163.3, 163.8, 163.9	C45-C45.9	No
Nasopharynx cancer	147-147.9	C11-C11-9	Yes

Non-Hodgkin lymphoma	200-200.9, 202-202.98	C82-C86.6, C96-C97-9	No
Multiple myeloma	203-203.9	C88-C90-32	No
Ovarian cancer	183, 183.0	C56-C56.9	Yes
Oral and lip	140-145.9	C00-C08.9	Yes
Pancreatic cancer	157-157.9	C25-C25.9	Yes
Prostate cancer	185-185.9	C61-C61.9	No
Testicular cancer	186-186.9	C62-C62.92	No
Thyroid cancer	193-192.9	C73-C73.9	No
Uterine cancer	182-182.8	C54-C54.9	No

Appendix 1: Table showing the ICD codes associates with each disorder and the classification of tumors based on their association with smoking.

Appendix 2: Templates for NOS observational studies quality appraisal

that the evaluation of some specific items of the scale involves the choice of some parameters by the researcher. For instance, item two of the outcome section of the Newcastle-Ottawa scale of cohort studies asks the researcher if the follow-up time was long enough, but it does not specify a particular threshold. We set this threshold to three years. In item four of the selection section of cohort studies quality appraisal, we only assign a star for cancer mortality studies if the particular study reported that the CNS patients did not present incident cancer before the CNS disorder diagnosis.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation ✸
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases ✸
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls ✸
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) ✸
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview where blind to case/control status ✸
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes ✸
   b) no

3) Non-Response rate
   a) same rate for both groups ✸
   b) non respondents described
   c) rate different and no designation

Appendix 2 Figure 1: Newcastle-Ottawa quality assessment scale for case-controls studies.

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
COHORT STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability

Selection

1) Representativeness of the exposed cohort
   a) truly representative of the average _____ (describe) in the community ✸
   b) somewhat representative of the average _____ in the community ✸
   c) selected group of users eg nurses, volunteers
   d) no description of the derivation of the cohort

2) Selection of the non exposed cohort
   a) drawn from the same community as the exposed cohort ✸
   b) drawn from a different source
   c) no description of the derivation of the non exposed cohort

3) Ascertainment of exposure
   a) secure record (eg surgical records) ✸
   b) structured interview ✸
   c) written self report
   d) no description

4) Demonstration that outcome of interest was not present at start of study
   a) yes ✸
   b) no

Comparability

1) Comparability of cohorts on the basis of the design or analysis
   a) study controls for _____ (select the most important factor) ✸
   b) study controls for any additional factor ✸ (This criteria could be modified to indicate specific control for a second important factor.)

Outcome

1) Assessment of outcome
   a) independent blind assessment ✸
   b) record linkage ✸
   c) self report
   d) no description

2) Was follow-up long enough for outcomes to occur
   a) yes (select an adequate follow up period for outcome of interest) ✸
   b) no

3) Adequacy of follow up of cohorts
   a) complete follow up - all subjects accounted for ✸
   b) subjects lost to follow up unlikely to introduce bias - small number lost - > ____ % (select an adequate %) follow up, or description provided of those lost) ✸
   c) follow up rate < ____% (select an adequate %) and no description of those lost
   d) no statement

Appendix 2 Figure 2: Newcastle-Ottawa quality assessment scale for cohort studies.

Appendix 3: Previous systematic reviews identified

Disorders	Author, year	Cancer Incidence or Mortality
AD, ALS, DS, HD,MS, PD, SCZ	Catalá, 2014 [21]	Incidence
AD	Ma, 2014 [397]	Incidence
AD	Shi, 2015 [398]	Both
AD	Zhang, 2015 [111]	Incidence
AD	Papageorgakopoulos, 2017 [399]	Incidence
BD	Hayes, 2015 [400]	Mortality
MD	McGee, 1994 [20]	Incidence
MD	Wulsin, 1999 [401]	Mortality
MD	Oerlemans, 2007 [402]	Incidence
MD	Pinquart, 2010 [403]	Mortality
MD	Cuijpers, 2014 [404]	Mortality
MD	Sun, 2015 [405]	Incidence (Breast Cancer)
MD	Ahn, 2016 [85]	Incidence
MD	Jia, 2017 [406]	Incidence
PD	Bajaj, 2010 [114]	Incidence
PD	Liu, 2011 [106]	Incidence (Melanoma)
PD	Huang, 2015 [407]	Incidence (Melanoma)
PD	Wang, 2015 [408]	Incidence (Breast and prostate cancer)
PD	Ye, 2016 [409]	Incidence (Brain cancer)
PD	Xie, 2016 [410]	Incidence (Lung cancer)
PD	Xie, 2017 [411]	Incidence (Colorectal cancer)
SCZ	Catts, 2008 [86]	Incidence
SCZ	Bushe, 2009 [105]	Incidence (Breast cancer)
SCZ	Zhuo, 2017 [412]	Mortality
SCZ	Xu, 2018 [413]	Incidence
SCZ	Li, 2018 [414]	Incidence

Appendix 3: Previous systematic reviews and meta-analysis identified including studies analyzing the associations between CNS disorders and subsequent cancer risk or mortality due to cancer.

Appendix 4: Characteristics of the observational studies included in meta-analyses

Source	Study Design (Country)	Setting; Coverage	Study Years (Follow-up, y)	No. of participants with CNS	Characteristics of Participants (Sex; Age)	No. of Cancer Cases	Main Cancer Outcome	Site Specific Cancer	Exposure	Outcome Definitions	End Point Measure	Adjustment for Confounding factors	NOS quality and risk of bias
AD													
Beard, 1996 [415] (Rochester Epidemiology Project)	Retrospective case-control (USA)	Community; Population-based	1960-1984 (NA)	917	**Female:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer mortality	No	HICDA-8 codes (ICD-8 Modification)	HICDA-8 codes (ICD-8 Modification) (Death certificates)	OR	Age, sex, and interaction term	6 Moderate risk
Attems, 2005 [416]	Retrospective cohort (Austria)	Inpatient; Single-center	1998-2004 (NA)	135	**Female:** 51.8% **Age:** 83.5 **Ethnicity:** NA **Smoking history:** NA	4	Overall cancer mortality	No	Neuropathological analyses CERAD; Braak scores,; NIA-ReaganInstitute (NIA-RI) Criteria	Death certificates and autopsy.	RR	None	2 High risk
Chamandy. 2005 [417] Sample of the Canadian Study of Health and Aging (CSHA)	Prospective cohort	Community; Population-based (It says community dwelling and institutionaliziedCanadians	1991-1996 (NA)	547	**Female %:** 72.9% **Age:** 86.6 **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer mortality	No	MMSE; DSM-III-R; ICD-10 codes; NINCDS-ADRDA	Death certificates	OR	Age, sex, region, and community/ institution residence	6 Moderate risk
Ganguli, 2005 [177] Monongahela Valley Independent Elders Survey	Prospective cohort (USA)	Community; Population-based	1987-2002 (10.3 y	273	**Female %:** 57.8% **Age:** 74 y **Ethnicity:**	29 in AD	Overall cancer mortality	No	DSM-III; CDR; NINCDS-ADRDA	Death certificates	RR	None, But the cognitive intact group was matched on Age, sex, and education.	6 Moderate risk
Roe, 2005 [107]	Prospective cohort (USA)	Outpatient; Single-center	1992-2003 (3.2 y in AD group and 4.3 years in the control group)	395	**Female %:** 63.8% in AD group **Age:** 75.5 y in AD group (mean) **Ethnicity:** 83% White, 16.7% Black, 0.3% Asian	45 (All participants AD + controls)	Overall cancer	No	CDR; Semi structured interview; Histopathologic AD in 93% of individuals at autopsy	Interview/Self-reports	HR	Sex, age, education	7 Low risk
Roe, 2010 [108] Cardiovascular Health Study-Cognition Substudy	Prospective cohort (USA)	Community; Population-based	1992-2003 (8.3 y)	71	**Female %:** 59.8% in the prevalent dementia group. **Age:** 77.9 y in incident dementia group. (mean)	376 in the whole study group	Overall cancer	No	NINCDS-ADRDA; ADDTC MRI (Modified Mini-Mental State Examination.	ICD-9	HR	Sex, race, age, education, income, CHS clinic, smoking status, obesity, physical activity	8 (Low-risk)

					Ethnicity:								
Driver, 2012 [109]	Prospective case-control (USA)	Community; Population-based	1986-2008 (10 y)	327	**Female:** 61.2% **Age:** 76 **Ethnicity:**	159	Overall cancer	No	MMSE; NINCDS-ADRDA	ICD-O	HR	Age, sex, smoking, and body mass index	8 (Low-risk)
Musicco, 2013 [110]	Prospective cohort (Italy)	Mixed; Population-based	2004-2009 (5.3y)	2832	**Female %:** 66.6% **Age:** 78.1 y in AD group (mean) **Ethnicity:**	68	Overall cancer	Yes	Payment exemptions, hospitalizations, drug prescriptions	ICD-10	RR	Age, sex, calendar year	8 (Low-risk)
Ou, 2013 [184]	Retrospective cohort (Taiwan)	Mixed; Population-based	1997-2009 (4.2y, median)	6960	**Female %:** 60.3 % **Age:** 76 y in AD group (median) **Ethnicity:**	405	Overall cancer	Yes	MME; CDR: Diagnosis ;anti-dementia drug prescription	Clinical diagnosis from Taiwan National Cancer Registry.	SIR	Age, sex, calendar year	9 (Low risk)
Lai, 2014 [418]	Retrospective case-control (Taiwan)	Mixed; Population-based	2000-2010 (NA)	60	**Female:** NA **Age:** 74 y **Ethnicity:** 100% Asian	3281	Specific cancer	NA	ICD-9	ICD-9	OR	Diabetes mellitus, cirrhosis, alcoholic liver damage, other chronic hepatitis, hepatitis B infection and hepatitis C infection	2 (High risk)
Todd, 2013 [419]	Prospective cohort (Ireland, UK)	Community; Population-based	2002-2010 (4.27y) mean	85	**Female %** 68.3 **Age:** 78.6 (mean) **Ethnicity:**	10	Overall cancer mortality	No	NINCDS-ADRDA; ADL index; MMSE	Death certificates; ICD	SMR	Age, sex	7 (Low risk)
Romero, 2014 [420] Neurological Diseases In Central Spain. NEDICES prospective population-based study.	Prospective cohort (Spain)	Community; Population-based	1993-2007 (10.1y) mean	306	**Female %:** 68.6% **Age:** 82.7 y **Ethnicity:**	16	Overall cancer mortality	No	37-MMSE; FAQ; Neurological evaluation, DSM-IV; NINCDS-ADRDA	ICD-9; ICD-10	HR	Age, gender, educational level, current smoker, current drinker, depressive symptoms or antidepressant use, hypertension, and diabetes mellitus	8 (Low risk)
Freedman, 2016 [147] SEER Medicare dataset.	Retrospective case-control (USA)	Mixed; Population-based	1992-2005 (NA)	7321	**Female %:** 45.5% **Age:** 74 y (median) **Ethnicity:**	5961	Overall cancer	Yes	One hospital or two physician/outpatient AD claims at least 30 days apart ICD-9	ICD-O-3	OR	Sex, age category, and calendar year	6 (Moderate risk)
Chen, 2017 [146] Taiwan national health insurance database (NHIRD). Registry for Catastrophic Illness Patients (HV)	Retrospective cohort (Taiwan)	Mixed; Population-based	1997-2010 (4.4 y)	25557	**Female %:** 61.8% **Age:** 78.2 **Ethnicity:** Asian (100%)	1341	Overall cancer	Yes	ICD-9-CM	ICD-9-CM	SIR	Age, sex	9 (Low risk)
ASD													

Shavelle, 2001 [421]	Retrospective cohort (USA)	Outpatient; Population-based	1983-1997 (15 y)	13111	**Female %:** 20.6% **Age:** NA **Ethnicity:** 50% White, 13% Black, 6% Asian	21	Overall cancer mortality	No	ICD-9 codes	ICD-9 codes	SMR	Age	7 (Low-risk)
Lauritsen, 2002 [422]	Retrospective case-control (Denmark)	Mixed; Population-based	1978-1993 (NA)	244	**Female %:** NA **Age:** NA **Ethnicity:** NA	1	Specific cancer	No	ICD-8 codes	ICD-8 codes	SIR	Age, sex, and calendar year.	6 (Moderate-risk)
Atladóttir 2012 [423]	Retrospective cohort (Denmark)	Mixed; Population-based	1994-2002 (10.5 y)	4995	**Female %:** 47.0% **Age:** 7.1 y **Ethnicity:** NA	44	Overall cancer	No	ICD-10 codes	ICD-10 codes	HR	Age, sex, birthweight, parity, congenital malformation, and parental psychiatric history	8 (Low risk)
Chiang, 2015 [424]	Retrospective cohort (Taiwan)	Mixed; Population-based	1997-2011 (9.13 y)	8438	**Female %:** 17.9% **Age:** 5.3 y (mean) **Ethnicity:** 100% Asian	20	Overall cancer	Yes	ICD-9 codes	ICD-9 codes	SIR	Age, sex, and calendar year	7 (Low risk)
Darbro, 2016 [425]	Retrospective case-control (USA)	Mixed, Single-center	2009-2015 (NA)	1837	**Female %:** 19.3% **Age:** most 0-14 **Ethnicity:** NA	10	Overall cancer	Yes	ICD-9 codes	ICD-9 codes	OR	Age and sex	4 (Moderate risk)
Hirvikoski, 2016 [426]	Retrospective case-control (Sweden)	Mixed, Population-based	1987-2009 (NA)	27122	**Female %:** 31.1% **Age:** 19.8 y (mean) **Ethnicity:** NA	88	Overall cancer mortality	No	ICD-9 or ICD-10 codes; DSM-V	ICD-9 or ICD-10 codes	OR	Sex	7 (Low risk)
Mouridsen, 2016 [427]	Prospective case-control (Denmark)	Mixed, Population-based	1977-2014 (37.2 y)	118	**Female %:** 30% **Age:** 4.9 y (mean) **Ethnicity:** NA	8	Overall cancer	Yes	IDC-9, ICD-10	ICD-8, ICD-10	OR	Age, Sex, place of birth, and social group	9 (Low risk)
BD													
Norton, 1984 [428]	Retrospective, cohort (UK)	Mixed, Population-based	1967-1976 (NA)	791 (Subjects who received lthium for at least two months.)	**Female %:** NA **Age:** 38.5 y (mean) **Ethnicity:** NA	4	Overall cancer mortality	No	Feighner criteria	Death certificates	SMR	Age and sex	3 (High risk)
Weeke, 1986 [429]	Retrospective, cohort (Denmark)	Inpatient, Multi-center	1970-1977 (5 y)	2168 manic-depressive	**Female %:** NA **Age:** NA **Ethnicity:** NA	53	Overall cancer mortality	No	ICD-8 codes	ICD-8 codes	SMR	Age, sex, and time at risk	7 (Low risk)

Saku, 1993 [126]	Retrospective, cohort (Japan)	Inpatient, Single-center	1948-1985 NA	187 manic participants	**Female %:** 36.4% **Age:** NA **Ethnicity:** 100% Asian	6	Overall cancer mortality	No	DSM-III-R codes	ICD-9 codes	SMR	Age, sex, and time at risk	5 (Moderate risk)
Høyer, 2000 [430]	Retrospective, cohort (Denmark)	Inpatient, Population-based	1973-1993 NA	NA	**Female %:** NA **Age:** NA **Ethnicity:** NA	2602	Overall cancer death	No	ICD-8 codes	ICD-8 codes	SMR	Age, sex, and illness duration	5 (Moderate risk)
Ösby, 2001 [431]	Retrospective, cohort (Sweden)	Inpatient, Population-based	1973-1995 (10 y)	15386 BD patients	**Female %:** 57.2% **Age:** 43 y (mean) **Ethnicity:** NA	475	Overall cancer death	No	ICD-8 and ICD-9 codes	Death certificates	SMR	Age, sex, and calendar period	7 (Low risk)
Angst, 2002 [132]	Prospective, cohort (Switzerland)	Inpatient, Single-center	1959-1997 (22 y)	220 BD	**Female %:** 71.7% **Age:** NA **Ethnicity:** NA	22	Overall cancer death	No	ICD-8 codes	ICD-8 codes	SMR	Age, sex, and calendar period	7 (Low risk)
Dalton, 2002 [432]	Retrospective, cohort (Denmark)	Inpatient, Population-based	1969-1993 (11.3 y)	9876 BD patients	**Female %:** 58.0% **Age:** 44.6 y **Ethnicity:** NA	1217	Overall cancer	Yes	ICD-7 and ICD-8	ICD-7 and ICD-8	SIR	Age, sex, and calendar period	7 (Low risk)
Dutta, 2007 [433]	Prospective, cohort (England, United Kingdom)	Mixed, Single-center	1965-1999 (19 y)	235 BD	**Female %:** 56.6% **Age:** 32.7 y (mean) **Ethnicity:** NA	11	Overall cancer mortality	No	DSM-IV codes	Death certificates	SMR	Age and sex	8 (Low risk)
Hippisley-Cox, 2007 [134]	Retrospective, case-control (England, United Kingdom)	Outpatient, Population-based	1995-2005 NA	614 BD	**Female %:** 57.3% **Age:** 68 y **Ethnicity:** NA	123	Specific cacner	Yes	Clinical diagnosis from clinical records	Clinical diagnosis from clinical records	OR	Age, sex, smoking, obesity, socio-economic status, comorbidities, and use of medication	8 (Low risk)
Laursen, 2007 [127]	Retrospective, cohort (Denmark)	Inpatient, Population-based	1973-2000 (>10 y)	11648 BD	**Female %:** NA **Age:** NA **Ethnicity:** NA	NA	Overall cancer death	No	ICD-10 codes	Death certificates	SMR	Age, sex, and calendar year	7 (Low risk)
BarChana, 2008 [434]	Retrospective, cohort (Israel)	Inpatient, Multi-center	1980-2005 NA	2121 BD	**Female %:** 52.5% **Age:** NA **Ethnicity:** NA	90	Overall cancer	No	ICD-10 codes	ICD-10 codes	SIR	Age, sex, and calendar year	5 (Moderate risk)
Hiroeh, 2008 [135]	Prospective, cohort (Denmark)	Mixed, Population-based	1973-1993 (21 y)	NA BD	**Female %:** NA **Age:**	2740	Overall cancer death	No	ICD-8 codes	ICD-8 codes	SMR	Age, sex, and calendar year	8 (Low risk)

					NA **Ethnicity:** NA								
Daumit, 2010 [136]	Retrospective, cohort (USA)	Community, Population-based	1992-2001 >5 y (Up to 8 years)	NA	**Female %:** 51% **Age:** 41.6 **Ethnicity:** 51% Black	NA	Overall cancer death	No	Clinical diagnosis	Medical records	SMR	Age, sex, and ethnicity	8 (Low risk)
Rosenshfir, 2011 [435]	Retrospective, cohort (Israel)	Inpatient, Single-center	1990-2006 NA	1638 BD	**Female %:** 40% **Age:** 47.2 y **Ethnicity:** NA	72	Overall caner	Yes	DSM criteria	ICD-9 codes	SIR	Sex	3 (High risk)
Batty, 2012 [128]	Retrospective, cohort (Sweden)	Inpatient, Multi-center	1950-1976 (10 y)	36 BD	**Female %:** 0% **Age:** >18y **Ethnicity:** NA	9	Overall cancer mortality	No	ICD-8, ICD-9, and ICD-10 codes	ICD-8, ICD-9, and ICD-10 codes	HR	Socioeconomic status in childhood, highest educational attainment, body mass index, comorbidities, and age at cancer diagnosis	7 (Low risk)
McGinty, 2012 [137]	Retrospective, cohort (United States)	Community, Population-based	1994-2004 >5 y	1002	**Female %:** 63% **Age:** 42.9 **Ethnicity:** 54% White, 45% Black	75	Overall cancer	Yes	ICD-9	ICD-9	SIR	Race, sex, age	6 (Moderate risk)
Ajetunmobi, 2013 [129]	Retrospective, cohort (Scotland, United Kingdom)	Inpatient, Population-based	1986-2010 >5 y	3876	**Female %:** 56.6% **Age:** NA **Ethnicity:** NA	15	Overall cancer mortality	No	ICD-9 and ICD-10 codes	ICD-9 and ICD-10 codes	SMR	Age, sex, calendar year, and deprivation	6 (Moderate risk)
Castagnini, 2013 [138]	Retrospective, cohort (Denmark)	Mixed, Population-based	1995-2008 6.6 y	3200 BD	**Female %:** NA **Age:** 41.8 y (mean) **Ethnicity:** NA	21	Overall cancer death	No	ICD-10 codes	ICD-10 codes	SMR	Age and sex	8 (Low risk)
Crump, 2013 [436]	Prospective, cohort (Sweden)	Mixed, Population-based	2003-2009 7 y	6618 BD	**Female %:** 59.2% **Age:** Most 40-69 **Ethnicity:** NA	178	Overall cancer mortality	Yes	ICD-10 codes	ICD-10 codes	HR	Age, sex, sociodemographic status (marital, educational level, employment income) and substance abuse disorder	9 (Low risk)
Guan, 2013 [130]	Retrospective, cohort (Netherlands)	Mixed, Population-based	1999-2008 6.4 y	2077 BD	**Female %:** 55.2% **Age:** 44.8 y (mean) **Ethnicity:** 81.1% White	31	Overall cancer mortality	No	DSM-IV codes	ICD-10 codes	HR	Age, sex, ethnicity, and income	9 (Low risk)
Lin, 2013 [141]	Retrospective, cohort	Mixed Population-based	1999-2009 5.6 y	20567 BD	**Female %:** 54.3% **Age:**	367	Overall cancer	Yes	ICD-9 codes	ICD-9 codes	SIR	Age and sex	8 (Low risk)

	(Taiwan, Republic of China)				38.5 y (mean) **Ethnicity:** 100% Asian								
Osborn, 2013 [140]	Retrospective, cohort (England, United Kingdom)	Outpatient, Population-based	1990-2008 6.4 y	4848 BD	**Female %:** 49.1% **Age:** 43.7 y **Ethnicity:** NA	NA	Overall cancer	Yes	Clinical diagnosis in medical records.	Clinical diagnosis in medical records	IRR	Age, Sex, General practices, age, sex, social deprivation, smoking status, obesity	9 (Low risk)
Almeida, 2014 [131]	Retrospective, cohort (Australia)	Community, Population-based	1996-2010 14 y	101 BD	**Female %:** 0% **Age:** 65-85 y **Ethnicity:** NA	11	Overall cancer death	No	ICD-9 or ICD-10 codes	ICD-9 or ICD-10 codes	HR	Age, time at risk (competing risks)	8 (Low risk)
Hung, 2014 [133]	Prospective, cohort (Taiwan, Republic of China)	Inpatient, Population-based	1997-2010 9.5 y	10207 BD	**Female %:** 50.8% **Age:** Most < 40 y **Ethnicity:** 100% Asian	409	Overall cancer	Yes	ICD-9 codes	ICD-9 codes	SIR	Age and sex	7 (Low risk)
Kanani, 2016 [437]	Retrospective cohort (England, United Kingdom)	Mixed, Multi-center	2000-2009 10 y	131 BD	**Female %:** 100% Women **Age:** >30 (mean age) **Ethnicity:** 87.8% White, 3.0% Black, 2.3% Asian	NA	Specific cancer death	Yes	ICD-10 codes	ICD-10 codes	HR	Age, sex, ethnicity, comorbidity stage at diagnosis and treatment	
Martinsson, 2016 [438]	Retrospective, cohort (Sweden)	Inpatient, Population-based	1987-2009 4.5 y	5442 BD	**Female %:** 58.8% **Age:** 63.1 y (mean) **Ethnicity:** NA	327	Overall cancer	Yes	ICD-9 and ICD-10 codes	ICD-9 and ICD-10 codes	IRR	Age and sex	7 (Low risk)
Kahan, 2017 [439]	Retrospective, cohort (Israel)	Outpatient, Population-based	2000-2012 5.9 y	3304 BD	**Female %:** 51.15% **Age:** NA **Ethnicity:** NA	110	Overall cancer	No	ICD-9 codes	ICD-9 codes	SIR	Age and sex	7 (Low risk)
HD													
Sørensen, 1992 [440]	Retrospective, cohort (Denmark)	Mixed, population-based	1943-1981 NA	395 HD	**Female:** 47% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	5	Overall cancer death	No	Clinical diagnosis	Death certificates	RR	Age	3 (High risk)
Sørensen, 1999 [173]	Retrospective, case-control (Denmark)	Mixed, population-based	1943-1993 16 y	694 HD	**Female:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	55	Overall cancer	Yes	ICD-7 codes	ICD-7 codes	SIR	Age, sex, and calendar year	8 (Low risk)
Ji, 2012 [441]	Prospective, cohort (Sweden)	Mixed, population-based	1969-2008 >5 y	1510 HD	**Female:** 52% **Age:**	91	Overall cancer	Yes	ICD-7-ICD-10 codes	ICD-7-ICD-10 codes	SIR	Age, calendar year, socio-economic status,	9 (Low risk)

					54 y (median) **Ethnicity:** NA **Smoking history:** NA							region of residence	
Turner, 2013 [191]	Retrospective, cohort (England, United Kingdom)	Inpatient, multicenter	1999-2010 >5 y	4865 HD	**Female:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	160	Overall cancer	Yes	ICD-7, ICD-8, ICD-9, and ICD-10 codes	ICD-7, ICD-8, ICD-9, and ICD-10 codes	RR	Age, sex, calendar year, and district of residence	6 (Moderate risk)
McNulty, 2018 [442]	Prospective, cohort (Europe)	NA, NA	NA NA	6540 HD	**Female:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer	Yes	Huntington's disease network REGISTRY	ICD-10	SIR	Age and sex	6 (Moderate risk)
Solberg, 2018 [443]	Retrospective, cohort (Sweden)	Mixed, population-based	1986-2015 NA	559 HD	**Female:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer death	No	ICD-10	ICD-10	OR	No	6 (Moderate risk)
MD													
Whitlock, 1979 [444]	Prospective, cohort (Australia)	Inpatient, Single-center	1973-1978 3.5 y	129 MD	**Female:** 69.8% **Age:** <40y **Ethnicity:** NA	6	Overall cancer death	No	Medical records	Medical records	SMR	Age and sex	6 (Moderate risk)
Thomas, 1980 [445]	Retrospective, case-control, (United States)	Community, Single-center	1947-1979 NA	36 medical graduates with depression MD	**Female;** NA **Age:** NA **Ethnicity:** NA	12	Overall cancer	No	HNTQ	NA	OR	NA	2 (High risk)
WEHS, 1987 (Shekelle 1981 [16] and Persky 1987 [446])	Prospective, cohort (United States)	Community, Single-center	1957-1978 20 y	379 MD	**Female:** 0% **Age:** 40-55 y **Ethnicity:** White	15	Overall cancer; Overall cancer death	No	MMPI	Death certificates, ICD-8	RR	Age and sex	6 (Moderate risk)
Hahn, 1988 [447]	Prospective, cohort (United States)	Inpatient, Single-center	1968-1982 14 y	836 MD	**Female:** 100% **Age:** 25-44 most **Ethnicity:** 99% White	15	Specific cancer	Yes	MMPI	Death certificates (biopsy) and healthy records	RR	Age and sex	6 (Moderate risk)
Kaplan, 1988 [448]	Prospective, cohort (United States)	Mixed, population-based	1965-1982 17 y	6928 participants (depressed and non)	**Female:** NA **Age:** NA **Ethnicity:**	443	Overall cancer; Overall cancer mortality	Yes	HPL Depression Index	Health records and death certificates	HR	Age and sex	6 (Moderate risk)

					NA								
Zonderman, 1989 [449]	Prospective, cohort (United States)	Community, population-based	1971-1986 10 y	1002 MD	**Female:** NA **Age:** NA **Ethnicity:** NA	110	Overall cancer; Overall cancer death	No	CES-D GWB-D	Death and hospital records	RR	Age, sex, marital status, smoking, family history of cancer, hypertension and cholesterol level	9 (Low risk)
Linkins, 1990 [450]	Prospective, cohort (United States)	Community, population-based	1971-1987 12 y	368 MD	**Female:** 69.8% **Age:** 25-64 most **Ethnicity:** NA **Smoking:** 39% never, 11% past	25	Overall cancer	No	CES-D	Death records, clinical confirmation	RR	Age and sex	8 (Low risk)
Friedman, 1994 [451]	Retrospective, cohort (United States)	Outpatient, single-center	1969-1988 19 y	923 MD	**Female:** 68.8% **Age:** 15-54 Most **Ethnicity:** 83% White, 10% Black, 7% Other	70	Overall cancer; Overall cancer death	Yes	Clinical diagnosis (hospital admission, medical records)	Clinical diagnosis (hospital admission, medical records)	SIR; SMR	Age and sex	8 (Low risk)
Vogt, 1994 [452]	Prospective, cohort (United States)	Community, population-based	1971-1985 15 y	NA	**Female %:** NA **Age:** NA **Ethnicity:** NA	NA	Overall cancer	No	DSM-III-R	ICD-7 codes	HR	Age, sex, socioeconomic status, length of health plan, membership, subjective, health status, smoking	8 (Low risk)
Saku, 1995 [126]	Retrospective, cohort (Japan)	Inpatient, single-center	1948-1982 NA	180 MD	**Female %:** 50.5% **Age:** NA **Ethnicity:** 100% Asian	12	Overall caner mortality	No	DSM-III-R	ICD-9 codes	SMR	Age, sex and time at risk	5 (Moderate risk)
Knekt, 1996 [142]	Prospective, cohort (Finland)	Community, population-based	1978-1991 14 y	324 MD	**Female %:** NA **Age:** NA **Ethnicity:** NA	29	Overall cancer	Yes	PSE and standardized psychiatric interview	Cancer records registry	RR	Age and sex	8 (Low risk)
Penninx, 1998 [453]	Prospective, cohort (USA)	Community, population-based	1982-1992 3.8 y	146 MD	**Female %:** 85.6% **Age:** 80.7 y **Ethnicity:** 0.9% Black **Smoking history:** 74% never, 17% past, 9% current	16	Overalll cancer;Overall cancer mortality	Yes	CES-D and DSM-III	ICD-9 codes	HR	Age, sex, ethnicity, physical disability, hospital admissions, smoking, and alcohol abuse	8 (Low risk)
Whooley, 1998 [454]	Prospective, cohort (USA)	Community, population-based	1988-1994 6 y	473 MD	**Female %:** 100% **Age:** 72.4 y **Ethnicity:** NA **Smoking history:** 15% Current	NA	Overall cancer death	Yes	Geriatric Depression Scale	ICD-9-R codes	HR	Age, sex, and history of morbidity (stroke, myocardial infarction, hypertension, smoking,…)	8 (Low risk)

Van den Heuvel, 1999 (NA INFORMATION EXTRACTED FROM Oerlemans meta-analysis)	Prospective, cohort (Netherlands)	Community, population-based	1084-1994 10 y	216 MD	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	9	Overall cancer	No	GP diagnosis	ICHPPC-2 codes	HR	NA	5 (Moderate risk)
Lawrence, 2000 (+ Lawrence 2001) [143]	Retrospective, cohort (Australia)	Mixed, population-based	1966-1995 13 y	8793 MD	**Female %:** 68.9% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	331	Overall cancer, Overeall cancer death	Yes	ICD-9	ICD-9	RR	Age and sex	8 (Low risk)
Brådvik, 2001 (+Berglund, 1987) [455, 456]	Prospective, cohort (Sweden)	Inpatient, single-center	1949-2000 >10 y	1206 MD	**Female %:** 58.0% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	121	Overall cancer death	No	DSM-III	ICD-8	SMR	Sex	3 (High risk)
Schuurman, 2001 (NA INFORMATION EXTRACTED FROM Oerlemans mata-analysis)	Prospective, cohort (Netherlands)	Community, population-based	1975-2000 25 y	1341 MD	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	95	Overall cancer	Yes	GP diagnosis	ICHPPC-2 codes	RR	Age, sex, and socioeconomic status	7 (Low risk)
Angst, 2002 [132]	Prospective, cohort (Switzerland)	Inpatient, single center	1959-1997 22 y	186MD	**Female %:** 71.7% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	23	Overall cancer death	No	ICD-8 codes	ICD-8 codes	SMR	Age, sex, and calendar period	7 (Low risk)
Yasuda, 2002 [457]	Prospective, cohort (Japan)	Community, population-based	1991-1998 7 y	NA	**Female %:** 61% **Age:** 72 y **Ethnicity:** 100% Asian **Smoking history:** NA	46	Overall cancer death	No	GHQ-Depression subscale	ICD-9 codes	HR	Age and sex	6 (Moderate risk)
Nyklícek, 2003 [458]	Prospective, cohort (Netherlands)	Comminity, population-based	1994-1999	840 MD	**Female %:** 100% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	3	Specific cancer	Yes	EDS and clinical diagnosis	clinical diagnosis	OR	Age, family history of cancer, body mass index, menopause, education, breastfeeding, estrogens use, exercise, alcohol intake, and other	5 (Moderate risk)

Goodwin, 2004 [459]	Retrospective, cohort (USA)	Inpatient, multi-center	1993-1996 3y	1841 MD	**Female %:** 100% **Age:** 76.5 **Ethnicity:** 90.8% White, 4% Black, 3% Hispanic, 2% Other **Smoking history:** NA	NA	Specific cancer death	Yes	ICD-9-CM codes	ICD-9-CM codes	HR	Age, sex, ethnicity, comorbidity, cancer stage	6 (Moderate risk)
Kisely, 2005 it is 2008?? [460]	Retrospective, cohort (Canada)	Couuminity, multi-center	1992-2000 >5y	NA	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer death	No	DSM-IV	ICD-9 codes	RR	Age and sex	7 (Low risk)
Kroenke, 2005 [461] (Nurses' Health Study I)	Prospective, cohort (USA)	Community, multi-center	1992-2000 >5y	5402 MD	**Female %:** 100% **Age:** 57 **Ethnicity:** NA **Smoking history:** 18.6% current	33	Specific cancer	Yes	Clinical diagnosis (e.g. 5-item Mental Health Index - MHI-5 score >53)	medical records, pathology reports to confirm cancer diagnosis	HR	Age, family history, diabetes, smoking, BMI, energy intake, physical activity, aspirin, menopausal status, hormone use, contraceptive use	8 (Low risk)
Goldacre, 2007 [462]	Retrospective, cohort (England, UK)	Inpatient, population-based	1963-1999 NA	27816 MD	**Female %:** 63.6% **Age:** most 20-59 **Ethnicity:** NA **Smoking history:** NA	1319	Overall cancer	Yes	ICD-7, ICD-8, ICD-9, ICD-10 codes	ICD-7, ICD-8, ICD-9, ICD-10 codes	RR	Age, sex, calendar year, and residence	6 (Moderate risk)
Mykletun, 2007 [463]	Prospective cohort, (Norway)	Community, population-based	1995-2001 4.4 y	3032 MD	**Female %:** NA **Age:** 48 y **Ethnicity:** non-White < 3% **Smoking history:** NA	NA	Overall cancer death	No	HADS scale, DSM-III/IV	ICD-8 and ICD-10 codes	OR	Age, sex, somatic symptoms, and physical diagnoses	6 (Moderate risk)
Kawamura, 2007 [464]	Prospective, cohort (Japan)	Community, population-based	1985-2000 6 y	111 MD	**Female %:** 75% **Age:** 77.4 **Ethnicity:** 100% Asian **Smoking history:** NA	NA	Overall cancer death	No	Geriatric depression screening test: Zung's SDS scale, NSDS, Research Diagnostic Criteria, and clinical interview	ICD-10 codes	RR	Age and sex	8 (Low risk)
Laursen, 2007 [127]	Retrospective, cohort (Denmark)	Inpatient. population-based	1973-2000 >10 y	72164 MD	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:**	NA	Overall cancer death	No	ICD-10-codes	Death certificates	SMR	Age, sex, and calendar year	7 (Low risk)

					NA								
Gross, 2010 (Gallo 2000) [465, 466]	Prospective, cohort (United states)	Community, population-based	1981-2005 24 y	163 MD	**Female %:** 62% **Age:** most 45-64 **Ethnicity:** 61% White, 35% Black, 3% Other **Smoking history:** 34% never, 38% past, 27% current	17	Overall cancer	Yes	DSM-III and DSM-IV (Diagnostic Interview Survey)	Self-reports and death certificates	HR	Age, sex, smoking, and parity (only for breast cancer)	9 (Low risk)
Chen, 2011 [467]	Prospective, cohort (Taiwan, Republic of China)	Community, population-based	1998-2008 5 y	778 MD	**Female %:** 60.7% **Age:** most 30-54 **Ethnicity:** 100% Asian **Smoking history:** NA	61	Overall cancer	Yes	ICD-9 codes	ICD-9 codes	HR	Age, sex, calendar year, income, geographic location, and urbanization level	7 (Low risk)
Hamer, 2011 [468]	Prospective, cohort (United Kingdom)	Community, population-based	1994-2008 9.2 y	210 MD	**Female %:** 56.2 **Age:** 79 y **Ethnicity:** NA **Smoking history:** 19% current smoking	NA	Overall cancer death	No	(GDS-15)	ICD-9 and ICD-10 codes	HR	Age and sex	5 (Moderate risk)
Liang, 2011 [469]	Retrospective, cohort (Taiwan, Republic of China)	Outpatient, population-based	2000-2008 NA	8419 MD	**Female %:** 60.5 % **Age:** NA **Ethnicity:** 100% Asian **Smoking history:** NA	321	Overall cancer	Yes	ICD-9-CM codes	ICD-9-CM codes	RR	Age, sex, urbanization and comorbidity	7 (Low risk)
Batty, 2012 [128]	Retrospective, cohort (Sweden)	Inpatient, multi-center	1950-1976 10 y	230MD	**Female %:** 0% **Age:** -< 18 y **Ethnicity:** NA **Smoking history:** NA	76	Overall cancer death	No	ICD-8 and ICD-9 codes	ICD-8 and ICD-9 codes	HR	Age, sex (only men), socioeconomic status. education, body mass index, and comrbidities	7 (Low risk)
Wyman, 2012 [470]	Retrospective, cohort (United States)	Community, population-based	1971-2011 40 y	476 MD	**Female %:** 70.5% **Age:** 47 y **Ethnicity:** 97.8% White/Netherlands **Smoking history:** NA	50	Overall cancer mortality	No	CES-D	ICD-8,ICD-9, and ICD-10	HR	Age, sex, calendar year, deprivation	7 (Low risk)
Ajetunmobi, 2013 [129]	Retrospective, cohort (Scotland,, United Kingdom)	Inpatient, population-based	1986-2010 >5 y	33161 MD	**Female %:** 61.5% **Age:** 47 y **Ethnicity:** NA **Smoking history:**	15	Overall cancer death	No	ICD-9 and ICD-10 codes	ICD-9 and ICD-10 codes	SMR	Age, sex, calendar year, deprivation	7 (Low risk)

					NA								
Gallo, 2013 [471]	Prospective, cohort (United States)	Community, population-based	1980-1994 13 y	90 MD	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	8	Overall cancer	Yes	DSM-III (Diagnostic Interview Schedule)	Surveillance, Epidemiology, and End Results (SEER)	RR	Age, sex, smoke, socioeconomic status	7 (Low risk)
Guan, 2013 [130]	Retrospective, cohort (Netherlands)	Mixed, population-based	1999-2008 5.7 y	15130 MD	**Female %:** 64.6% **Age:** 44.1 **Ethnicity:** 70.7% White/Netherlands **Smoking history:** NA	274	Overall cancer death	No	DSM-IV codes	ICD-10 codes	HR	Age, sex, ethnicity, and income	9 (Low risk)
Kisely, 2013 [94]	Retrospective, cohort (Western Australia, Australia)	Mixed, population-based	1988-2007 >5 y	NA	**Female %:** 52.2% **Age:** 67 y **Ethnicity:** NA **Smoking history:** NA	961 cancers and 401 cancer deaths	Overall cancer; Overall cancer death	No	ICD-9 and ICD-10 codes	ICD-O codes	RR	Age and sex	8 (Low risk)
Lai, 2013 [472]	Retrospective, case-control (Taiwan, Republic of China)	Mixed, population-based	2000-2010 NA	499 MD	**Female %:** 37.6% **Age:** > 65 y **Ethnicity:** 100% Asian **Smoking history:** NA	114	Specific cancer	No	ICD-9 codes	ICD-9 codes	OR	Age, diabetes mellitus, cirrhosis, alcohol, chronic hepatitis	5 (Moderate risk)
Lemogne, 2013 [473]	Prospective, cohort (France)	Community (Occupational/Work), multi-center	1994-2009 5.3 y	1183 Chronic or recurrend depression	**Female %:** 28.3% **Age:** > 40 y **Ethnicity:** NA **Smoking history:** NA	NA	Specific cancer	Yes	ICD-9 codes and CES-D-ICD-10 codes	ICD-9 codes and CES-D-ICD-10 codes	HR	Age, sex, occupational grade, alcohol, smoking, diet physical activity, height, weight, perceived health status	7 (Low risk)
Almeida, 2014 (+ Almeida, 2010) [131, 474]	Retrospective, cohort (Australia)	Community, population-based	1996-2010 14 y	958 MD	**Female %:** 0.0% **Age:** 65-85 y **Ethnicity:** NA **Smoking history:** 29% Never, 57%Past, 14% current	185	Overall cancer death	No	ICD-9 or ICD-10 codes	ICD-9 or ICD-10 codes	HR	Age, sex (only men), time at risk (competing risks)	8 (Low risk)
Hung, 2014 [133]	Prospective cohort (Taiwan, Republic of China)	Inpatient, population-based	1997-2010 8.6 y	9826 MD	**Female %:** 52.4% **Age:** most < 40 y **Ethnicity:** 100% Asian **Smoking history:** NA	507	Overall cancer	Yes	ICD-9 codes	ICD-9 codes	SIR	Age and sex	7 (Low risk)

Archer, 2015 [475] (Whitehall II study)	Prospective, cohort (England, United Kingdom)	Community (occupational/work), multi-center	1988-2008 17.4 y	1816 depressed (14% with chronic depression)	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer	No	GHQ-30 depression subscale	ICD-9 or ICD-10	HR	Age, sex, socio-economic status, health behaviors, health status, medication, social support	8 (Low risk)
Chang, 2015 [476]	Retrospective, cohort (South Korea)	Community (occupational/work), multi-center	1992-2012 19 y	47317 MD	**Female %:** 21.4% **Age:** most > 35 y **Ethnicity:** 100% Asian **Smoking history:** 5-10% never smoking, 8-15% current, 7-14% past	4412 cancers	Overall cancer	Yes	DSM-IV	ICD-10 codes	HR	Age, sex, smoking status, alcohol consumption, exercise, BMI, cholesterol, blood sugar, hypertension, cancer family history	NA TO DO
Huang, 2015 [477] (Nurses' Health Study I)	Prospective, cohort (United States)	Community, multi-center	1992-2010 >5 y	7509 MD	**Female %:** 100% **Age:** 66.7 y **Ethnicity:** NA **Smoking history:** 11% current	61	Specific cancer	Yes	MHI-5 score <= 52, antidepressant use, or physician-diagnosed depression.	Pathology reports to confirm cancer diagnosis.	HR	Age, parity, hormone, history of cancer, BMI, physical activity, smoking, caffeine, lactose	8 (Low risk)
Huang, 2015 [477] (Nurses' Health Study II)	Prospective, cohort (United States)	Community, multi-center	1993-2011 >5 y	19936 MD	**Female %:** 100% **Age:** 48.5 y **Ethnicity:** NA **Smoking history:** 11% current	48	Specific cancer	Yes	MHI-5 score <= 52, antidepressant use, or physician-diagnosed depression.	Pathology reports to confirm cancer diagnosis.	HR	Age, parity, hormone, history of cancer, BMI, physical activity, smoking, caffeine, lactose	8 (Low risk)
O'Neil, 2015 [478]	Retrospective, case-control (Population surveys in nineteen countries)	Community, population-based	NA	NA	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer	No	DSM-IV	Self-reports of diagnosis	OR	Age, Gender, Parson-year, Country and dummy variables for all mental disorders	6 (Moderate risk)
Kanani, 2016 [437]	Retrospective, cohort (England, United Kingdom)	Mixed, multi-center	2000-2009 10 y	955 MD	**Female %:** 100% **Age:** >30 y **Ethnicity:** 86.7% White, 3.1% Black, 2.8% Asian **Smoking history:** NA	NA	Specific cancer	Yes	ICD-10 codes	ICD-10 codes	HR	Age, sex, ethnicity, deprivation, comorbidity, stage at diagnosis and treatment	8 (Low risk)
Kisely, 2016 [144]	Retrospective, cohort (Australia)	Mixed, population-based	1993-2003 3 y	NA	**Female %:** 44% **Age:** 67 y **Ethnicity:** NA **Smoking history:** NA	292 cancers 111 cancer deaths	Overall cancer, Overall cancer death	No	ICD-9 codes	ICD-O codes	RR	Age and sex	7 (Low risk)

Liang, 2017 [479] – Women's Health Initiative	Prospective, cohort (United States)	Outpatient, population-based	1993-2003 3 y	386 MD	**Female %:** 100% **Age:** 61.2 y **Ethnicity:** 90% White **Smoking history:** 41% Never smoker	29	Specific cancer death	Yes	CES-D	National death registry	HR	Age, race, body mass index, smoking, history of postmenopausal hormone therapy, comorbidity, mammography, tumor state, tumor grade, estrogen receptor and progesterone receptor status.	6 (Moderate risk)
Yang, 2018 [198]	Retrospective, cohort (Finland)	Mixed, poputaltion-based	1987-2012 NA	113754 MD	**Female %:** 50.4% **Age:** 67 y **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer mortality	No	Purchase of antidepressants (ATC codes)	ICD-8, ICD-9, and ICD-10	HR	Age, calendar, education, occupational social class, economic activity, marital status, and sex	5 (Moderate risk)
PD													
Westlund, 1956 [480]	Prospective, cohort (Norway)	Inpatient, single-center	1917-1950 >5 y	146 PD	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	7	Overall cancer death	No	Clinical diagnosis (hospital discharge)	Clinical diagnosis (hospital discharge)	SMR	Age, calendar year	4 (Moderate risk)
Barbeu and Joly, 1963 [481]	Retrospective, cohort (Canada)	Inpatient, single-center	1950-1961	502 PD	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	9	Overall cancer	No	Clinical diagnosis (medical charts)	Clinical diagnosis (medical charts)	SIR	Age	5 (Moderate risk)
Hoehn and Yahr, 1967 [482]	Retrospective, cohort (Unitied States)	Inpatient, single-center	1949-1964 9 y	672 PD	**Female %:** 39.9% **Age:** 55.3 **Ethnicity:** NA **Smoking history:** NA	24 cancer deaths	Overall cancer death	No	Clinical diagnosis (medical charts)	Clinical diagnosis (medical charts)	RR	Sex	4 (Moderate risk)
Kessler, 1972 [483]	Prospective, cohort (United States)	Inpatient, single-center	1965-1967 3 y	468 PD	**Female %:** 48.1% **Age:** 67 **Ethnicity:** most white **Smoking history:** 72% non-smokers	17	Overall cancer death	Yes	Clinical diagnosis (medical charts and interviews)	ICD-8	RR	Sex	4 (Moderate risk)
Harada, 1983 [484]	Retrospective, cohort (Japan)	Mixed, population-based	1975-1981 5 y	101 PD	**Female %:** 62.4% **Age:** 68 **Ethnicity:**	5	Overall cancer death	No	Clinical diagnosis (Survey and medical charts)	Clinical diagnosis (Survey and medical charts)	RR	None	6 (Moderate risk)

					100% Assian **Smoking history:** NA								
Jansson and Jankovik, 1985 [485]	Prospective, cohort (United States)	Inpatient, single-center	1978.1984 8.6 y	406 PD	**Female %:** 40.4% **Age:** 40 **Ethnicity:** NA **Smoking history:** 24.9% smokers, 75.1% smokers	19	Overall cancer	Yes	Clinical diagnosis (medical charts)	Clinical diagnosis (medical charts)	RR	Age, sex, smoking	6 (Moderate risk)
Grorell, 1994 [486]	Retrospective, cohort (United States)	Community, population-based	1970-1989 >5 y	8629 PD	**Female %:** NA **Age:** >40 y **Ethnicity:** NA **Smoking history:** NA	548	Overall cancer death	No	ICD-8 and ICD-9 (medical records, death registry)	ICD-8 and ICD-9 (medical records, death registry)	PMR	Age	6 (Moderate risk)
Ben-Shlomo and Marmot, 1995 [487]	Retrospective, cohort (England and Wales, United Kingdom)	Outpatient, population-based	1970-1991 20 y	220PD	**Female %:** NA **Age:** >40 y **Ethnicity:** NA **Smoking history:** NA	9	Overall cancer death	Yes	ICD-9 (medical records)	ICD-9 (medical records)	HR	Age, sex, region	5 (Moderate risk)
Wermuth, 1995 [488]	Retrospective, cohort (Denmark)	Mixed, population-based	1973-1991 NA	458 PD	**Female %:** 50.6% **Age:** >60 y **Ethnicity:** NA **Smoking history:** NA	26	Overall cancer death	No	Clinical diagnosis (medical reords)	Death certificates	SMR	Age and sex	6 (Moderate risk)
Louis, 1997 [489]	Retrospective, cohort (United States)	Community, population-based	1988-NA 2.3 y	288 PD	**Female %:** 49.6% **Age:** 74 y **Ethnicity:** NA **Smoking history:** NA	2	Overall cancer death	No	Clinical diagnosis (medical records)	Death certificates	RR	None	3 (High risk)
Hely, 1999 [490]	Prospective, cohort (Australia)	Mixed, multi-center	1984-1997 10 y	149 PD	**Female %:** 44.4% **Age:** 62 y **Ethnicity:** NA **Smoking history:** NA	12	Overall cancer death	No	Clinical diagnosis	ICD-9	SMR	Age, calendar year	6 (Moderate risk)
Vancore, 1999 [491]	Retrospective, cohort (Italy)	Outpatient, population-based	1987-1994 5.7 y	10322 PD	**Female %:** 54.9% **Age:** 74 y **Ethnicity:** NA **Smoking history:**	448	Overall cancer death	Yes	Antiparkinsonian prescriptions and outpatient records	ICD-9	SMR	Age, sex, and calendar period	7 (Low risk)

					NA								
Minami, 2000 [492]	Retrospective, cohort (Japan)	Community, population-based	1984-1992 5 y	228 PD	**Female %:** 57.7% **Age:** 66.2 y **Ethnicity:** 100% Asian **Smoking history:** NA	15	Overall cancer	Yes	Clinical diagnosis (Survey)	Cancer registry	SIR	Age and sex	7 (Low risk)
Beyer, 2001 [493]	Prospective, cohort (Norway)	Community, population-based	1993-1996 4 y	245 PD	**Female %:** 51% **Age:** 70 y **Ethnicity:** NA **Smoking history:** NA	14	Overall cancer death	No	Clinical diagnosis	ICD-9 or ICPC	RR	Age and sex	8 (Low risk)
Fall, 2003 [494]	Prospective, cohort (Sweden)	Community, population-based	1989-1998 9.4 y	170 PD	**Female %:** 40% **Age:** 65.6 y **Ethnicity:** NA **Smoking history:** NA	10	Overall cancer death	Yes	Clinical diagnosis	ICD-9	HR	Age and sex	7 (Low risk)
Gutman, 2004 [495]	Prospective, cohort (Canada)	Mixed, population-based	1993-1999 6 y	15306 PD	**Female %:** NA **Age:** most > 60 **Ethnicity:** NA **Smoking history:** NA	NA	Overall caner	Yes	ICD-9 (medical records, hospitalizations)	ICD-9 (medical records, hospitalizations)	RR	Age and sex	6 (Moderate risk)
Elbaz, 2005 [113]	Retrospective, cohort (United States)	Community, population-based	1976-2002 >5 y	196 PD	**Female %:** 38.3% **Age:** 71 y **Ethnicity:** NA **Smoking history:** NA	50 incident cancers	Overall cancer	Yes	Clinical diagnosis (medical records)	ICD Adapted Code for Hospitals (H-ICDA)	RR	Age, sex, and smoking	9 (Low risk)
D'Amelio, 2006 (+ Morgante, 2000) [496, 497]	Prospective, cohort (Italy)	Community, population-based	1987-2001 >5 y	59 PD	**Female %:** 55.9% **Age:** 74 y **Ethnicity:** NA **Smoking history:** NA	3	Overall cancer death	No	Clinical diagnosis	Clinical diagnosis	RR	Age, sex	7 (Low risk)
Leibson, 2006 [498] All Olmsted County Minnesota	Prospective, cohort (USA)	Mixed, population-based	1976-NA	197 PD	**Female %:** 39% **Age:** 70 y **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer	No	REP diagnostic index + complete medical records review	ICD-9 and ICD-9-CM	RR	Sex and birth year	8 (Low risk)

Constantinescu, 2007 [212]	Prospective, cohort (Canada and USA)	Inpatient, multicenter	1987-1994 NA	800 PD	**Female %:** 44% **Age:** 61.1 y **Ethnicity:** 97.6% White **Smoking history:** NA	2	Specific cancer	Yes	Clinical diagnosis (records from observation period of a randomizer trial)	Clinical diagnosis (records from observation period of a randomizer trial)	SIR	Age and sex.	4 (Moderate risk)
Driver, 2007 [112]	Cohort, prospective (USA)	Community (Physician's Health Study), multi-center	1981-2000 5.2 y	487 PD	**Female %:** 0% **Age:** 59.7 y **Ethnicity:** 92.2% White **Smoking history:** 51.3% never, 42.3% past, 6.4% current	53	Overall cancer	Yes	Self-report by US male physician (validation study by medical records)	Self-report by US male physician (validation study by medical records)	RR	Smoking, alcohor use, BMI category, and exercise	7 (Low risk)
Driver, 2008 [499]	Prospective, cohort (USA)	Community (Physician's Health Study), multi-center	1982-2006 5.8 y	560 PD	**Female %:** 0% **Age:** 59.8 y **Ethnicity:** NA **Smoking history:** 51.3% never, 42.3% past, 6.5% current	22	Overall cancer death	Yes	Self-report by US male physician (validation study by medical records)	Self-report by US male physician (validation study by medical records)	HR	Age and smoking	7 (Low risk)
Becker, 2010 [500]	Retrospective, cohort (England, United Kingdom)	Outpatient, population-based	1994-2005 NA	2993 PD	**Female %:** 36.7% Cancers in women **Age:** >40 y **Ethnicity:** NA **Smoking history:** NA	188	Overall caner	Yes	Clinical diagnosis (Oxform Medical Information System)	Clinical diagnosis (Oxform Medical Information System)	IRR	None (Crude IRR in the cohort study)	5 (Moderate risk)
Lo, 2010 [501]	Retrospective, cohort (United States)	Mixed, population-based	1994-2008 >5 y	692 PD	**Female %:** 37.4% **Age:** 66 Y **Ethnicity:** 80.3% White, 8.2% Hispanic, 7.7% Asian, 3.2% Black **Smoking history:** 50.4% never, 49.4% ever, 0.1% missing	90	Overall cancer	Yes	Clinical diagnosis (medical records and cancer registry)	Clinical diagnosis (medical records and cancer registry)	RR	Age, sex, smoking, alcohol, body mass index	9 (Low risk)
Fois, 2010 [104]	Retrospective, cohort (England, United Kingdom)	Inpatient, population-based	1963-1999 3.2 y	4355 PD	**Female %:** 50.9% **Age:** >70 y **Ethnicity:** NA **Smoking history:** NA	219	Overall cancer	Yes	Clinical diagnosis (hospital admissions)	Clinical diagnosis (hospital admissions)	RR	Age, sex, calendar year, and district of residence	7 (Low risk)
Posada, 2011 [502]	Prospective, cohort (Spain)	Community, population-based	1994-2007 12 y	81 PD	**Female %:** 59.3% **Age:** 77y **Ethnicity:**	9	Overall cancer death	No	Clinical diagnosis/Unified Parkinson's Disease Rating Scale (UPDRS)	ICD-9 codes	RR	None	7 (Low risk)

					NA **Smoking history:** 34.8% ever smoker								
Sun, 2011 [503]	Retrospective, cohort (Taiwan, Republic of China)	Mixed, population-based	2000-2008 >5 y	4957 PD	**Female %:** 48.3% **Age:** 63.5 **Ethnicity:** 100% Asian **Smoking history:** NA	NA	Overall cancer	Yes	ICD-9-CM codes	ICD-9-CM codes	HR	Age, sex, occupation and comorbidities (diabetes, hypertension, heart disease)	9 (Low risk)
Kareus, 2012 [504]	Retrospective, cohort (United States)	Mixed, population-based	1904-2008 NA	2998 PD	**Female %:** NA **Age:** >70 u **Ethnicity:** NA **Smoking history:** NA	NA	Specific cancer death	Yes	Clinical diagnosis (death records)	ICD-10	RR	None	4 (Moderate risk)
Rugbjerg, 2012 (+ Olsen, 2005, + Moller, 1995) [102, 103, 505]	Retrospective, cohort (Denmark)	Mixed, population-based	1977-2008 5.7 y	20343 PD	**Female %:** 47.3% **Age:** 72.7 y **Ethnicity:** NA **Smoking history:** NA	2218	Overall cancer	Yes	ICD-8 and ICD-10 codes	ICD-8 and ICD-10 codes	SIR	Age and sex	8 (Low risk)
Lai, 2013 [506]	Retrospective, case-control (Taiwan, Republic of China)	Mixed, population-based	2000-2009 NA	397 PD	**Female %:** 32.3% women with lung cancer **Age:** 68.3 y in lung cancer cohort **Ethnicity:** NA **Smoking history:** NA	NA	Specific cancer	Yes	ICD-9	ICD-9	OR	sex, age and index date Adjusted for age, sex, Parkinson's disease, pulmonary tuberculosis, chronic obstructive pulmonary disease, pneumoconiosis, asbestosis, alcoholism and tobacco use	8 (Low risk)
Constantinescu, 2014 [507]	Prospective, cohort (USA)	Inpatient, multicenter	2007-2013	1737 PD	**Female %:** 35.6% **Age:** 60 y **Ethnicity:** 90% White **Smoking history:** NA	19	Specific cancer	Yes	Clinical diagnosis (records from observation period of a randomizer trial)	Clinical diagnosis (records from observation period of a randomizer trial)	SIR	Age and sex	3 (High risk)
Ong, 2014 [508]	Retrospective, cohort (England, United Kingdom)	Mixed, population-based	1999-2011 >5 y	219194 PD	**Female %:** 43.0% **Age:** 92% > 65 y **Ethnicity:** NA **Smoking history:** NA	17524	Overall cancer	Yes	Clinical diagnosis (hospital admissions)	ICD-10	RR	Age, sex, calendar year and district of residence	8 (Low risk)

Pinter, 2014 [509]	Retrospective, cohort (Austria)	Outpatient, single-center	1974-2012 32.5 y	237 PD	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	25	Overall cancer death	No	ICD-10	ICD-10	SMR	Age, sex	7 (Low risk)
Wirdefeldt, 2014 [510]	Prospective, cohort (Sweden)	Mixed, population-based	1958-2009 >5 y	11786 PD	**Female %:** 39,5% **Age:** 62.5 y **Ethnicity:** NA **Smoking history:** NA	562	Overall cancer	Yes	ICD-7, ICD-8, ICD-9, and ICD-10 codes	ICD-7, ICD-8, ICD-9, and ICD-10 codes	HR	Age, education	7 (Low risk)
Liao, 2015	Retrospective, case-control (Taiwan, Republic of China)	Mixed, population-based	2005-2011 NA	170 PD	**Female %:** 65.7% in pancreatic cancer group **Age:** 64.9 y in the pancreatic cancer group **Ethnicity:** 100% Asian **Smoking history:** NA	NA	Specific cancer	Yes	ICD-9	ICD-9	OR	sex, age, comorbidities, and index year of diagnosing pancreatic cancer	8 (Low risk)
Lin, 2015 [511]	Retrospective, cohort (Taiwan, Republic of China)	Mixed, population-based	2004-2012 3.4 y	62023 PD	**Female %:** 50.8% **Age:** most > 60 y **Ethnicity:** NA **Smoking history:** NA	2925	Overall cancer	Yes	ICD-9-M codes	ICD-9-M codes	HR	Age, sex	7 (Low risk)
Boursi, 2016 [512]	Retrospective, case-control (United Kingdom)	Outpatient, population-based	1995-2013 6 y	657 PD	**Female %:** 44.9% **Age:** 72.3 y **Ethnicity:** NA **Smoking history:** 46.8 ever	117	Specific cancer	Yes	Clinical diagnosis (THIN database codes)	Clinical diagnosis (THIN database codes)	OR	Age, sex, obesity, alcohol, diabetes mellitus, aspirin use, hormone replacement therapy, antibiotics.	9 (Low risk)
Freedman, 2016 [145]	Retrospective, case-control (United States)	Mixed, population-based	1992-2005 NA	6994 PD	**Female %:** 45% **Age:** 74 y **Ethnicity:** 86.1% White, 13.9% Non-white, 7.8% African American, 2.5% Asian, 1.4% Hispanic, and 0.2 Native-American Indian **Smoking history:** NA	5829	Overall cancer	Yes	Clinical diagnosis (ICD-9, with full claim information in the registry system)	ICD-O-3 codes	OR	Age, sex, race, number of doctor's visits, cancer registry area	7 (Low risk)

Jespersen, 2016 [513]	Retrospective, case-control (Denmark)	Mixed, population-based	1995-2010 NA	1901 PD	**Female %:** 0% **Age:** 72 y **Ethnicity:** NA **Smoking history:** NA	245	Specific cancer	Yes	Clinical diagnosis (ICD-8, ICD-9 codes)	Clinical diagnosis (ICD-8, ICD-9 codes)	OR	Age, sex, and comorbidity	8 (Low risk)
Peretz, 2016 [514]	Retrospective, cohort (Israel)	Outpatient, population-based	2000-2012 105 y	7125 PD	**Female %:** 46.3% **Age:** 71.1 y **Ethnicity:** NA **Smoking history:** NA	1302	Overall cancer	Yes	Clinical diagnosis	Clinical diagnosis	SIR	Age, sex	7 (Low risk)
Tang, 2016 [515]	Retrospective, cohort (Taiwan, Republic of China)	Mixed, population-based	1999-2010 NA	2998 PD	**Female %:** 47.6% **Age:** 68 y **Ethnicity:** 100% Asian **Smoking history:** NA	19	Specific cancer	Yes	ICD-9-CM codes	ICD-9-CM codes	HR	Age and sex	6 (Moderate risk)
Chen, 2017 [146]	Retrospective, cohort (Taiwan, Republic of China)	Mixed, population-based	1997-2010 4.4 y	2527 PD	**Female %:** 57% **Age:** 77.1 y **Ethnicity:** 100% Asian **Smoking history:** NA	110	Overall cancer	Yes	ICD-9-CM codes	ICD-9-CM codes	SIR	Age and sex	7 (Low risk)
Liao, 2017	Retrospective, case-control (Taiwan, Republic of China)	Mixed, population-based	2005-2011 NA	641 PD	**Female %:** 42.7% in the colorectal cancer group **Age:** 64.7 y in the colorectal cancer group **Ethnicity:** 100% Asian **Smoking history:** NA	255	Specific cancer	Yes	ICD-9	ICD-9	OR	Sex, age, comorbidities, and index year of diagnosis of colorectal cancer	8 (Low risk)
SCZ													
Tsuang, 1980 [516]	Prospective, cohort (United Sates)	Inpatient, single-center	1934-1974 35 y	195 SCZ	**Female %:** 48.7% **Age:** 29 y **Ethnicity:** NA **Smoking history:** NA	8	Overall cancer death	No	Clinical diagnosis and research criteria	Clinical diagnosis and research criteria	SMR	Age, sex, and admission pay status (public or private)	7 (Low risk)
Herrman, 1983 [517]	Retrospective, cohort (England, United Kingdom)	Mixed, population-based	1970-NA 10 y	592 SCZ	**Female %:** 49.8% **Age:** NA **Ethnicity:** NA	23 cancers and 6 cancer deaths	Overall cancer, Overall cancer death	No	ICD-8 codes (medical records, hospitalizations)	ICD-8 codes (medical records, hospitalizations)	SMR	Age and sex	7 (Low risk)

					Smoking history: NA								
Brook, 1985 [518]	Retrospective, cohort (Netherlands)	Inpatient, population-based	1979-1981 2 y	4743 SCZ	**Female %:** 49.8% **Age:** >40 **Ethnicity:** NA **Smoking history:** NA	20	Overall cancer death	No	Clinical diagnosis	Clinical diagnosis	SMR	Age and sex	3 (High risk)
Kendler, 1986 [519]	Retrospective, cohort (United States)	Community (military/work)	1946-1981 35 y	590 SCZ	**Female %:** 0% **Age:** NA **Ethnicity:** 100% White **Smoking history:** NA	20	Overall cancer death	No	ICD-8 codes	ICD-8 codes	SMR	Age	5 (Moderate risk)
Nakane, 1986	Retrospective, cohort (Japan)	Mixed, population-based	1960-1978 >5 y	3107 SCZ	**Female %:** 55.3% **Age:** NA **Ethnicity:** 100% Asian **Smoking history:** NA	44	Overall cancer	Yes	Clinical diagnosis	Clinical diagnosis	RR	Age and sex	8 (Low risk)
Buda, 1988 [520]	Prospective, cohort (United States)	Inpatient, single-center	1934-1974 40 y	322 SCZ	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	14	Overall cancer death	No	DSM-III	ICD-8 codes	SMR	Age and sex	6 (Moderate risk)
Allebeck, 1989 [521]	Retrospective, cohort (Sweden)	Inpatient, population-based	1971-NA 10 y	1190 SCZ	**Female %:** 51.9 **Age:** Most > 30 **Ethnicity:** NA **Smoking history:** NA	34	Overall cancer death	No	ICD-8 codes	ICD-8 codes	SMR	Age and sex	6 (Moderate risk)
Mortensen, 1989 (+ Dupont 1986; Mortensen, 1990) [200, 522]	Retrospective, cohort (Denmark)	Inpatient, population-based	1957-1984 >5 y	6152 SCZ	**Female %:** 51.9% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	1028	Overall cancer, Overall cancer death	Yes	Clinical diagnosis/Medical records	Clinical diagnosis/Medical records	IRR	Age and sex	7 (Low risk)
Zilber, 1989 [523]	Retrospective, cohort (Israel)	Inpatient, multicenter	1978-1983 5 y	45893 patient-years	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	74	Overall cancer death	No	ICD-9 codes	ICD-9 codes	SMR	Age and sex	5 (Moderate risk)

Gulbinat, 1992 [14]	Retrospective, cohort (Denmark, United States, and Japan)	Mixed, population-based	1957-1980 >5 y	16236 SCZ	**Female %:** 45.3% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	895	Overall cancer	Yes	Clinical diagnosis and DMS-III-R codes	ICD-8 codes	RR	Age and sex	8 (Low risk)
Mortensen, 1993 + Mortensen, 1994 [524, 525]	Retrospective, cohort (Denmark)	Inpatient, population-based	1970-1987 >5 y	9156 SCZ	**Female %:** 38.2% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	133	Overall cancer; Overall cancer death	Yes	ICD-8/DSM-III-R	medical records (cancer registry)	SIR; SMR	Age, sex, and calendar period	7 (Low risk)
Saku, 1995 [126]	Retrospective, cohort (Japan)	Inpatient, single-center	1948-1982 NA	2268 SCZ	**Female %:** 36.8% **Age:** NA **Ethnicity:** 100% **Smoking history:** NA	41	Overall cancer death	No	DSM-III-R codes	ICD-9 codes	SMR	Age, sex, and time at risk	5 (Moderate risk)
Knekt, 1996 [142]	Prospective, cohort (Finland)	Community, population-based	1978-1991 14 y	164 SCZ	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	17	Overall cancer	Yes	PSE and standardized psychiatric interview schedule	Cancer records (registry)	RR	Age and sex	8 (Low risk)
Slazar-Fraile, 1998 [526]	Prospective, cohort (Spain)	Mixed, population-based	1986-1993 NA	NA	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer deaths	No	ICD-9 codes	ICD-9 codes	HR	Age and sex	7 (Low risk)
Lawrence, 2000 (+ Lawrence, 2001) [143]	Retrospective, cohort (Australia)	Mixed, population-based	1966-1995 13 y	9721 SCZ	**Female %:** 38.3% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	496	Overall cancer; Overall cancer death	Yes	ICD-9 codes	ICD-9 codes	RR	Age and sex	8 (Low risk)
Ösby, 2000 [527]	Retrospective, cohort (Sweden)	Inpatient, population-based	1973-1995 >5 y	7784 SCZ	**Female %:** 49.5% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	225	Overall cancer death	No	Clinical diagnosis	ICD-8 codes	SMR	Age and sex	6 (Moderate risk)
Joukamaa, 2001 [528]	Retrospective, cohort (Finland)	Community, population-based	1978-1994 17 y	93 SCZ	**Female %:** 57.0% **Age:**	7	Overall cancer death	No	Clinical diagnosis (coputer catego-ID)	ICD-8-codes	RR	Age, sex	7 (Low risk)

					NA **Ethnicity:** NA **Smoking history:** NA								
Lichtermann, 2001 [529]	Retrospective, cohort (Finland)	Mixed, population-based	1971-1996 >5 y	26996 SCZ	**Female %:** 42.3% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	724	Overall cancer	Yes	DSM-III-R, ICD-8 , and ICD-9 codes	ICD-8 , and ICD-9 codes	SIR	Age, sex, and follow-up	8 (Low risk)
Barak, 2005 (+ Barak, 2008, + Raviv, 2014) [530-532]	Retrospective, cohort (Israel)	Inpatient, single-center	1990-2011 >5 y	6337 SCZ	**Female %:** 31.7% **Age:** 49 y **Ethnicity:** NA **Smoking history:** NA	320	Overall cancer	Yes	DSM-IV	ICD-C and ICD-9 codes	SIR	Age, sex	7 (Low risk)
Dalton, 2005 (+ Dalton 2003 for breast) [218, 533]	Retrospective, cohort (Denmark)	Inpatient, population-based	1969-1993 12.8 y	22766 SCZ	**Female %:** 42.8% **Age:** 38 y **Ethnicity:** NA **Smoking history:** NA	1299	Overall cancer	Yes	ICD-7 and ICD-8 codes	ICD-7 and ICD-8 codes	SIR	Age, sex, and calendar year	6 (Moderate risk)
Goldacre, 2005 [534] (+ Baldwin 1980)	Retrospective, cohort (England, United Kingdom)	Inpatient, population-based	1963-1999 12.6 y	9649 SCZ	**Female %:** NA **Age:** 40 y **Ethnicity:** NA **Smoking history:** NA	486	Overall cancer	Yes	ICD-9 codes (medical reocrds, hospitalizations)	ICD-9 codes (medical reocrds, hospitalizations)	RR	Age, sex, and calendar year	6 (Moderate risk)
Grinshpoon, 2005 [535]	Retrospective, cohort (Israel)	Inpatient, population-based	1962-2001 >5 y	33372 SCZ	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	1504	Overall cancer	Yes	ICD-10 codes	ICD-03 codes	SIR	Age, sex, place of birth	7 (Low risk)
Heilä, 2005 [536]	Retrospective, cohort (Finland)	Inpatient, population-based	1980-1996 >5 y	58761 SCZ	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	2320	Overall cancer death	No	ICD-8, ICD-9, and ICD-10 codes	ICD-8, ICD-9, and ICD-10 codes	RR	Age, sex, calendar year	6 (Moderate risk)
Kisely, 2005 [537]	Retrospective, cohort (Canada)	Mixed, population-based	1995-2000 5 y	NA	**Female %:** NA **Age:** NA **Ethnicity:** NA	NA	Overall cancer death	No	DSM-IV	ICD-9 codes	RR	Age and sex	7 (Low risk)

					Smoking history: NA								
Hippisley-Cox, 2007 [134]	Retrospective, case-control (England, United Kingdom)	Outpatient, population-based	1995-2005 NA	710 SCZ	**Female %:** 51.4% **Age:** NA **Ethnicity:** NA **Smoking history:** 34% Current	139	Specific cancer	Yes	Clinical diagnosis (medical)	Clinical diagnosis (medical)	OR	Age, sex, smoking, obesity, soci-economic status, comorbidities and use of medication	8 (Low risk)
Laursen, 2007 [127]	Retrospective, cohort (Denmark)	Inpatient, population-based	1973-2000 >10 y	17660 SCZ	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer death	No	ICD-10 codes	Death certificates	SMR	Age, sex, and calendar year	7 (Low risk)
Hiroeh, 2008 [538]	Prospective, cohort (Denmark)	Mixed, population-based	1973-1993 21 y	NA	**Female %:** NA **Age:** NA **Ethnicity:** NA **Smoking history:** NA	467	Overall cancer death	No	ICD-8 codes	ICD-8 codes	SMR	Age, sex, and calendar year	8 (Low risk)
Tran, 2009 [539]	Prospective, cohort (France)	Mixed, population-based	1993-2003 11 y	3470 SCZ	**Female %:** 37.8% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	74	Overall cancer death	Yes	ICD-10 codes	ICD-10 codes	SMR	Age, sex, BMI, smoking, alcohol use, duration of illness, hospitalization, and antipsychotic treatment	8 (Low risk)
Brown, 2010 (+ Brown 2000) [540, 541]	Prospective, cohort (England, United Kingdom)	Outpatient, population-based	1981-2006 25 y	370 SCZ	**Female %:** 42.4% **Age:** 39-43 y mean age **Ethnicity:** NA **Smoking history:** NA	30	Overall cancer death	Yes	ICD-9 and ICD-10 codes	ICD-9 and ICD-10 codes	SMR	Age and sex	7 (Low risk)
Daumit, 2010 [136]	Retrospective, cohort (United States)	Community, population-based	1992-2001 >5 y	NA	**Female %:** 51% **Age:** 41.6 y mean age **Ethnicity:** 51% Black **Smoking history:** NA	NA	Overall cancer death	No	Clinical diagnosis (medical records)	Clinical diagnosis (medical records)	SMR	Age, sex, and ethnicity	7 (Low risk)
Chou, 2011 [542]	Retrospective, cohort (Taiwan, Republic of China)	Mixed, population-based	1999-2008 9 y	59257 SCZ	**Female %:** 44.2% **Age:** 40.5 y mean age **Ethnicity:** 100% Asian **Smoking history:** NA	1145	Overall cancer	Yes	ICD-9 codes	ICD-9 codes	HR	Age, sex, comorbidity, socio-economic status	9 (Low risk)

Truyers, 2011 [543]	Retrospective, cohort (Belgium)	Outpatient, population-based	1994-2007 >5 y	894 SCZ	**Female %:** 53% **Age:** 48.8 y mean age **Ethnicity:** 100% Asian **Smoking history:** NA	NA	Overall cancer	No	ICD-10 codes, ICPC-2 codes	ICD-10 codes, ICPC-2 codes	HR	Age, sex, other (unspecified)	5 (Moderate risk)
Batty, 2012 [128]	Retrospective, cohort (United States)	Community, population-based	1950-1976 10 y	111 SCZ	**Female %:** 0.0% **Age:** >18 **Ethnicity:** NA **Smoking history:** NA	40	Overall cancer death	No	ICD-8 and ICD-9 codes	ICD-8 and ICD-9 codes	HR	Age, sex (only men), socio-economic status, education, body mass index, and comorbidities	7 (Low risk)
McGinty, 2012 [137]	Retrospective, cohort (United States)	Community, population-based	1994-2004 >5 y	2315 SCZ	**Female %:** 48% **Age:** 41.5 y **Ethnicity:** 43% White, 56% Black **Smoking history:** NA	155	Overall cancer	Yes	ICD-9 codes (administrative claims data)	ICD-9 codes (administrative claims data)	SIR	Age, sex, ethnicity (race)	7 (Low risk)
Talaslahti, 2012 [544]	Retrospective, cohort (Finland)	Mixed, population-based	1999-2008 9 y	9461 SCZ	**Female %:** 68.7% **Age:** 35 y **Ethnicity:** NA **Smoking history:** NA	827	Overall cancer death	No	ICD-8, ICD-9, and ICD-10	ICD-8, ICD-9, and ICD-10	SMR	Age, sex	7 (Low risk)
Whitley, 2012 [545]	Prospective, cohort (Sweden)	Inpatient, population-based	1969-2004 22 y	6669 SCZ	**Female %:** 0% **Age:** 31 y **Ethnicity:** NA **Smoking history:** NA	22	Overall cancer	No	Clinical diagnosis (medical records)	Clinical diagnosis (medical records)	SIR	Age, sex	7 (Low risk)
Ajetunmobi, 2013 [129]	Retrospective, cohort (Scotland, United Kingdom)	Inpatient, population-based	1986-2010 >5 y	15965 SCZ	**Female %:** 43.8% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	12	Overall cancer death	No	ICD-9 and ICD-10 codes	ICD-9 and ICD-10 codes	SMR	Age, sex, calendar year, deprivation	7 (Low risk)
Castagnini 2013 [138]	Retrospective, cohort (Denmark)	Mixed, population-based	1995-2008 6.6 y	4576 SCZ	**Female %:** NA **Age:** 32.6 y **Ethnicity:** NA **Smoking history:** NA	29	Overall cancer death	No	ICD-10 codes	ICD-10 codes	SMR	Age and sex	8 (Low risk)
Crump, 2013 [139]	Prospective, cohort (Sweden)	Mixed, population-based	2003-2009 7 y	8377 SCZ	**Female %:** 42.2% **Age:**	487	Overall cancer; Overall cancer death	Yes	ICD-10 codes	ICD-10 codes	HR	Age, sex, income, socio-demographic	9 (Low risk)

					NA **Ethnicity:** NA **Smoking history:** NA							status, substance use disorders	
Guan, 2013 [130]	Retrospective, cohort (Netherlands)	Mixed, population-based	1999-2008 6.3 y	4590 SCZ	**Female %:** 43.4% **Age:** 41.8 y **Ethnicity:** 69.3% White/Netherlands **Smoking history:** NA	87	Overall cancer death	No	DSM-IV codes	ICD-10 codes	HR	Age, sex, ethnicity, and income	9 (Low risk)
Ji, 2013 [546]	Retrospective, cohort (Sweden)	Mixed, population-based	1965-2008 >5 y	59233 SCZ	**Female %:** 45.6% **Age:** 38-47 y **Ethnicity:** NA **Smoking history:** NA	5101	Overall cancer	Yes	ICD-7, ICD-8, ICD-9, and ICD-10 codes	ICD-7, ICD-8, ICD-9, and ICD-10 codes	SIR	Age, sex, calendar year, socio-economic status, residential area and comorbidity	9 (Low risk)
Kisely, 2013 [94]	Retrospective, cohort (Western Australia, Australia)	Mixed, population-based	1988-2007 >5 y	NA	**Female %:** 52.2% **Age:** 67 y **Ethnicity:** NA **Smoking history:** NA	275 cancers 147 cancer deaths	Overall cancer; Overall cancer death	No	ICD-9 or ICD-10 codes	ICD-O codes	RR	Age and sex	8 (Low risk)
Lin, 2013 [141]	Retrospective, cohort (Taiwan, Republic of China)	Mixed, population-based	1997-2009 6.6 y	71317 SCZ	**Female %:** 46.7 % **Age:** 37 y **Ethnicity:** 100% Asian **Smoking history:** NA	1129	Overall cancer	Yes	ICD-9 codes	ICD-9 codes	SIR	Age and sex	8 (Low risk)
Osborn, 2013 [140]	Retrospective, cohort (England, United Kingdom)	Outpatient, population-based	1990-2008 6.4 y	6845	**Female %:** 49.1 % **Age:** 43.7 y **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer	Yes	Clinical diagnosis (medical records)	Read codes	IRR	Age, sex, calendar year, social deprivation, smoking and obesity	9 (Low risk)
Almeida, 2014 [131]	Retrospective, cohort (Australia)	Community, population-based	1996-2010 14 y	444	**Female %:** 0 % **Age:** 65-85 mean **Ethnicity:** NA **Smoking history:** 16% never, 67% past, 16% current	61	Overall cancer death	No	ICD-9 or ICD-10 codes	ICD-9 or ICD-10 codes	HR	Age, sex (only men), time at risk (competing risks)	8 (Low risk)
Hendrie, 2014 (NA)	Retrospective, cohort (United States)	Outpatient, population-based	1999-2008 5.7 y	757 SCZ	**Female %:** 69.2% **Age:** 69.9 y	30	Overall cancer death	No	ICD-9 codes	ICD-9 codes	RR	None	7 (Low risk)

					Ethnicity: 40.8% Black **Smoking history:** 56.6% Ever								
Kredentser 2014 [547]	Retrospective, cohort (United States)	Outpatient, population-based	1999-2008 10 y	9038 SCZ	**Female %:** 50.7% **Age:** 48.7 y **Ethnicity:** NA **Smoking history:** NA	NA	Overall cancer death	Yes	ICD-9 and ICD-10 codes	ICD-9 and ICD-10 codes	RR	Age and Sex	7 (Low risk)
Perini, 2014 (+ Grigoletti 2009) [548, 549]	Retrospective, cohort (Italy)	Community, population-based	1982-2006 25 y	7875 SCZ	**Female %:** 56.9% **Age:** 50.3 y **Ethnicity:** NA **Smoking history:** NA	17	Overall cancer death	Yes	ICD-9 or ICD-10 codes	ICD-9 or ICD-10 codes	SMR	Age and sex	7 (Low risk)
Lesage, 2015 [550]	Retrospective, cohort (Canada)	Mixed, population-based	2000-2009 >5 y	328340 SCZ	**Female %:** 43.4% **Age:** NA **Ethnicity:** NA **Smoking history:** NA	966	Overall cancer	No	ICD-9 or ICD-10 codes	ICD-9 or ICD-10 codes	RR	Age and sex	7 (Low risk)
Olfson, 2015 [551]	Retrospective, cohort (United States)	Mixed, population-based	2001-2007 >5 y	1138853 SCZ	**Female %:** 46.4% **Age:** most > 35 **Ethnicity:** 52% White, 29% Black, 6% Hispanic, 4% Other **Smoking history:** NA	9638	Overall cancer death	Yes	ICD-10 codes	ICD-10 codes	SMR	Age, sex, ethnicity and geographic region	7 (Low risk)
Park, 2015 [552]	Retrospective, cohort (South Korea)	Mixed, population-based	1995-2009 NA	NA	**Female %:** NA **Age:** NA **Ethnicity:** 100% Asian **Smoking history:** NA	NA	Overall cancer deaths	No	ICD-10	ICD-10	SMR	NA	7 (Low risk)
Kisely, 2016 [144]	Retrospective, cohort (Queensland, Australia)	Mixed, population-based	2002-2007 5 y	NA	**Female %:** 44% **Age:** 64.3 y **Ethnicity:** NA **Smoking history:** NA	236 cancers; 87 cancer-deaths	Overall cancer; Overall cancer death	No	ICD-9	ICD-O	RR	Age and sex	7 (Low risk)
Wu Chou, 2017 [553]	Retrospective, cohort (Taiwan, Republic of China)	Mixed, population-based	1998-2011 >5 y	10727 SCZ	**Female %:** 100% **Age:** 44.1 y **Ethnicity:**	119	Specific cancer	Yes	ICD-9 codes	ICD-9 codes	HR	Age, occupation, income, comorbidities, and medications	9 (Low risk)

					100% Asian **Smoking history:** NA								
Chen, 2018 [554]	Retrospective, cohort (Taiwan, Republic of China)	Inpatient, population-based	2000-2010 7.29 y	32731 SCZ	**Female %:** 45.1% **Age:** NA **Ethnicity:** 100% Asian **Smoking history:** NA	514	Overall cancer	Yes	ICD-9 codes	ICD-9 codes	SIR	Age and sex	7 (Low risk)
Dalton, 2018 [555]	Retrospective, cohort (Denmark)	Mixed, population-based	1995-2011 5.5 y	499	**Female %:** 100% **Age:** 60.5 y **Ethnicity:** NA **Smoking history:** NA	86	Specific cancer deaths	No	ICD-8 and ICD-10 codes	ICD-8 and ICD-10 codes	HR	Age at diagnosis, calendar year, education, Charlson comorbidity, cancer related factors, treatment	8 (Low risk)
Tanskanen, 2018 [556]	Retrospective, cohort (Finland)	Mixed, population-based	1984-2014	79877 SCZ	**Female %:** NA **Age:** NA **Ethnicity:** 100% White **Smoking history:** NA	NA	Overall cancer deaths	No	ICD-8, ICD-9. and ICD-10 codes	ICD-8, ICD-9. and ICD-10 codes	SMR	Age and sex	7 (Low risk)

Appendix 4: observational studies included in the meta-analyses of observational studies included in Chapter 2

.

Appendix 5: Transcriptomic datasets included in the differential gene expression meta-analyses and the consensus co-expression module analyses.

Disease	Study.ID	Year	Platform	Initial samples	Total samples	Cases	Controls	QC SMR	Included
AD	GSE1297 [557]	2004	1	31	31	22	9	5.12	Yes
AD	GSE5281 [558]	2006	2	161	23	10	13	4.75	Yes
AD	E_MEXP_2280 [559]	2010	2	31	12	7	5	2.88	Yes
AD	GSE36980 [560]	2013	3	79	17	7	10	3.75	Yes
AD	GSE29378 [561]	2013	4	63	27	15	12	5.38	Yes
AD	GSE48350 [562]	2014	2	253	61	19	42	1.38	Yes
AD	GSE84422 [563]	2016	2	2004	55	44	11	4.75	Yes
ASD	GSE28521 [564]	2011	5	79	72	16	56	1.25	Yes
ASD	GSE28475 [565]	2012	5	143	65	8	57	2.5	Yes
ASD	Gupta [566]	2014	NA	120	27	10	17	2.25	Yes
ASD contols	GSE36192 [567]	2012	3	911	NA	NA	NA	NA	NA
BD	GSE53987 [568]	2014	2	36	36	17	19	2.75	Yes
BD	GSE92538 [569]	2016	2	386	65	12	53	2.12	Yes
BD	AltarC [570]	NA	1	22	22	11	11	3	Yes
BD	Dobrin [571]	NA	2	86	50	26	24	2.12	Yes
HD	GSE3790 [572]	2005	1	404	70	38	32	NA	Yes
HD	GSE26927 [573]	2011	5	118	19	9	10	NA	Yes
MD	GSE54570 [574]	2014	1	26	26	13	13	4.25	Yes
MD	GSE54567 [574]	2014	2	28	28	14	14	3.38	Yes
MD	GSE54568 [574]	2014	2	30	29	14	15	2.88	Yes
MD	GSE53987 [568]	2014	2	205	36	17	19	2.88	Yes
MD	GSE92538 [569]	2016	2	363	82	29	53	3.25	Yes
MD	AltarC [570]	NA	1	72	22	11	11	4.38	Yes
PD	GSE7621 [575]	2007	2	25	25	16	9	3	Yes
PD	GSE8397 [576]	2008	1	94	38	24	14	1.88	Yes
PD	E_MEXP_1416 [577]	2008	6	16	16	8	8	7.5	No
PD	GSE20141 [578]	2010	2	18	17	9	8	6.75	Yes
PD	GSE20163 [578]	2010	1	17	17	8	9	5.88	Yes
PD	GSE20164 [578]	2010	1	11	11	6	5	8	No
PD	GSE20295 [578]	2010	1	93	29	11	18	4.5	Yes
PD	GSE20159 [578]	2011	4	33	33	16	17	7.62	No
PD	GSE49036 [579]	2013	2	28	23	15	8	5.25	Yes
PD	GSE54282 [580]	2014	3	33	6	3	3	8.38	No
PD	GSE24378 [578]	NA	6	17	17	8	9	7.25	No
SCZ	Mirnics [581]	2000	1	18	18	12	6	5.12	Yes
SCZ	Haroutunian [582]	2007	7	38	38	20	18	6.5	Yes

SCZ	GSE17612 [583]	2009	2	51	51	28	23	2.62	Yes
SCZ	GSE21138 [584]	2010	2	59	57	29	28	3.38	Yes
SCZ	GSE53987 [568]	2014	2	205	34	15	19	3.5	Yes
SCZ	GSE92538 [569]	2016	2	363	79	27	52	4.12	Yes
SCZ	Dobrin [571]	NA	2	86	48	24	24	5.12	Yes
SCZ	AltarC [570]	NA	1	72	22	11	11	5.62	Yes
ALL	GSE13204 [585]	2009	2	3248	824	750	74	1.88	Yes
ALL	GSE26713 [586]	2011	2	124	124	117	7	2.38	Yes
ALL	GSE28497 [587]	2011	1	288	288	284	4	3.38	Yes
ALL	GSE46170 [588]	2013	2	38	38	32	6	3.75	Yes
ALL	GSE79533 [589]	2017	2	226	226	223	3	3.62	Yes
AML	GSE12662 [590]	2008	2	106	81	76	5	5.75	Yes
AML	GSE13204 [585]	2009	2	3248	574	501	73	2.88	Yes
AML	GSE30029 [591]	2011	4	121	121	90	31	3.38	Yes
AML	GSE34577 [592]	2011	4	89	75	57	18	3.88	Yes
AML	GSE48558 [593]	2013	3	170	21	18	3	5.88	Yes
AML	GSE67936	2015	7	168	168	150	18	4.88	Yes
AML	GSE68172	2015	2	77	15	10	5	8.25	No
AML	GSE63270 [594]	2016	2	104	104	42	62	3.62	Yes
AML	GSE76340 [595]	2016	4	166	18	15	3	6.5	Yes
BLCA	GSE7476 [596]	2007	2	12	12	9	3	1.88	Yes
BLCA	GSE13507 [597]	2008	8	256	84	75	9	3.5	Yes
BLCA	GSE52519 [598]	2013	9	12	8	5	3	3.38	Yes
BLCA	GSE38264 [599]	2014	3	51	38	28	10	2.25	Yes
BLCA	E_MTAB_1940 [600]	2015	2	86	86	82	4	4	Yes
BRCA	GSE7904 [601]	2007	2	62	47	40	7	6.12	Yes
BRCA	GSE10780 [602]	2009	2	185	185	42	143	3.62	Yes
BRCA	GSE10810 [603]	2009	2	58	58	31	27	2.38	Yes
BRCA	GSE29431	2011	2	66	54	44	10	4.38	Yes
BRCA	GSE31448 [604]	2011	2	357	356	352	4	3.88	Yes
BRCA	GSE42568 [605]	2013	2	121	114	98	16	3.12	Yes
BRCA	GSE54002 [606]	2014	2	433	433	417	16	6.12	Yes
BRCA	GSE45827 [607]	2016	2	155	149	140	9	6.38	Yes
BRNCA	GSE4290 [608]	2006	2	180	176	153	23	3.25	Yes
BRNCA	GSE9385 [609]	2008	10	55	55	49	6	4.5	Yes
BRNCA	GSE16011 [610]	2010	2	284	278	270	8	2.25	Yes
BRNCA	GSE21354 [611]	2010	2	18	18	14	4	9	No
BRNCA	GSE15824 [612]	2011	2	45	32	27	5	7.12	No
BRNCA	GSE42656 [613]	2013	4	73	55	44	11	6.5	Yes
BRNCA	GSE44971 [614]	2013	2	58	58	49	9	7.5	No
BRNCA	GSE50161 [615]	2013	2	130	128	115	13	4.75	Yes
BRNCA	GSE68848 [616]	2015	2	580	482	454	28	3.5	Yes
BRNCA	GSE74195 [617]	2015	2	51	44	39	5	6.62	Yes
CERV	GSE6791 [618]	2007	2	84	28	20	8	3.75	Yes

CERV	GSE39001 [619]	2013	3	79	24	19	5	3.08	Yes
CERV	GSE63514 [620]	2015	2	128	52	28	24	1.67	Yes
CERV	GSE67522 [621]	2015	7	42	35	15	20	2.5	Yes
CHLCA	GSE26566 [622]	2012	5	169	17	11	6	1.75	Yes
CHLCA	GSE32879 [623]	2012	3	37	23	16	7	1.75	Yes
CHLCA	GSE32225 [624]	2013	1	155	88	83	5	3.38	Yes
CHLCA	GSE22633 [625]	2014	8	63	24	20	4	3.12	Yes
CLL	GSE13204 [585]	2009	2	3248	522	448	74	2	Yes
CLL	GSE13987 [626]	2009	2	24	8	4	4	3.62	Yes
CLL	GSE26725 [627]	2011	2	17	17	12	5	5	Yes
CLL	GSE31048 [628]	2013	2	221	221	188	33	4.25	Yes
CLL	GSE51528 [629]	2015	3	2299	229	217	12	3.12	Yes
CLL	GSE67640 [630]	2017	7	24	24	15	9	3	Yes
CML	GSE13204 [585]	2009	2	3248	139	66	73	1.88	Yes
CML	GSE43754 [631]	2013	10	20	20	10	10	2	Yes
CML	GSE47927 [632]	2013	3	67	67	52	15	2.12	Yes
CRCA	GSE9348 [633]	2007	2	82	82	70	12	4.88	Yes
CRCA	GSE18105 [634]	2010	2	111	94	77	17	11.38	No
CRCA	GSE20916 [635]	2010	2	145	105	81	24	5.12	Yes
CRCA	GSE24550 [636]	2011	10	167	90	77	13	13.25	No
CRCA	GSE31279 [637]	2011	5	110	53	30	23	8.38	No
CRCA	GSE33113 [638]	2011	2	96	96	90	6	9.38	No
CRCA	GSE31737	2012	10	80	80	40	40	6.62	Yes
CRCA	GSE37182 [639]	2013	4	172	168	84	84	8.62	No
CRCA	GSE37364 [640]	2013	2	94	52	14	38	4.62	Yes
CRCA	GSE39582 [641]	2013	2	585	585	566	19	5.62	Yes
CRCA	GSE35834 [642]	2014	10	158	53	30	23	9	No
CRCA	GSE44076 [643]	2014	11	246	246	98	148	5.38	Yes
CRCA	GSE41657	2015	12	88	37	25	12	10.88	No
CRCA	GSE62932 [644]	2015	2	68	68	64	4	10.25	No
CRCA	GSE71187 [645]	2017	12	189	111	99	12	6.62	Yes
DLBCL	GSE12453 [646]	2008	2	67	36	11	25	1.88	Yes
DLBCL	GSE12195 [647]	2009	2	136	88	73	15	2	Yes
DLBCL	GSE56315 [648]	2015	2	122	122	89	33	2.12	Yes
FLYMPH	GSE12453 [646]	2008	2	67	30	5	25	4.62	Yes
FLYMPH	GSE12195 [647]	2009	2	136	53	38	15	2.25	Yes
FLYMPH	GSE14214 [649]	2011	13	13	12	9	3	4.62	Yes
FLYMPH	GSE48047	2014	14	55	26	18	8	4.12	Yes
FLYMPH	GSE55267 [650]	2014	2	69	69	63	6	2.62	Yes
FLYMPH	GSE65135 [651]	2015	2	28	24	14	10	2.75	Yes
HANC	GSE9844 [652]	2008	2	38	38	26	12	3.12	Yes
HANC	GSE23558 [653]	2011	12	32	28	23	5	4.62	Yes
HANC	GSE25099 [654]	2011	10	79	79	57	22	5	Yes
HANC	GSE30784 [655]	2011	2	229	212	167	45	2.12	Yes
HANC	GSE34105 [656]	2012	15	78	57	49	8	9	Yes

HANC	GSE29330 [657]	2014	2	18	18	13	5	4.25	Yes
HANC	GSE55550 [658]	2014	16	155	86	73	13	6.12	Yes
HANC	GSE59102	2014	12	42	32	23	9	4	Yes
HANC	GSE75538 [659]	2016	15	28	10	5	5	6.75	Yes
KDNCA	GSE11024 [660]	2008	2	79	22	10	12	6.38	Yes
KDNCA	GSE17895 [661]	2010	2	160	157	135	22	6.88	Yes
KDNCA	GSE36895 [662]	2012	2	76	52	29	23	2.62	Yes
KDNCA	GSE40435 [663]	2013	7	202	202	101	101	4.88	Yes
KDNCA	GSE47032 [664]	2013	10	40	20	10	10	5.12	Yes
KDNCA	GSE46699 [665]	2014	2	130	94	49	45	4.62	Yes
KDNCA	GSE76351	2015	17	24	24	12	12	7.62	No
KDNCA	GSE66272 [666]	2016	2	54	54	27	27	3.38	Yes
KDNCA	GSE68417 [667]	2016	3	49	49	35	14	6.38	Yes
KDNCA	GSE71963 [668]	2016	12	48	48	32	16	7.12	No
LGCA	GSE12236	2008	10	40	40	20	20	8.25	No
LGCA	GSE18842 [669]	2010	2	91	91	46	45	7	No
LGCA	GSE19188 [670]	2010	2	156	156	91	65	3.38	Yes
LGCA	GSE19804 [671]	2011	2	120	120	60	60	5.12	Yes
LGCA	GSE31210 [672]	2011	2	246	246	226	20	6.88	Yes
LGCA	GSE31552 [673]	2011	3	131	75	43	32	9.88	No
LGCA	GSE32863 [674]	2012	17	116	87	33	54	7.12	No
LGCA	GSE40275	2012	10	10	72	29	43	11.38	No
LGCA	GSE30219 [675]	2013	2	307	307	293	14	8.12	No
LGCA	GSE40791 [676]	2013	2	194	194	94	100	3.38	Yes
LGCA	GSE32665 [677]	2013	8	179	170	81	89	10.88	No
LGCA	GSE43458 [678]	2013	3	110	110	80	30	9.62	No
LGCA	GSE33532	2014	2	100	34	14	20	8.38	No
LGCA	GSE75037 [679]	2016	9	134	134	51	83	5.62	Yes
LIVCA	GSE6764 [680]	2007	2	75	75	65	10	7.12	No
LIVCA	GSE17967 [681]	2009	1	63	63	16	47	10.25	No
LIVCA	GSE12941 [682]	2010	10	20	20	10	10	8	No
LIVCA	GSE36376 [683]	2012	7	433	273	80	193	5.88	Yes
LIVCA	GSE50579 [684]	2013	14	80	34	26	8	9.12	No
LIVCA	GSE41804 [685]	2013	2	40	40	20	20	5.12	Yes
LIVCA	GSE17548 [686]	2013	2	37	37	17	20	6.88	Yes
LIVCA	GSE54236 [687]	2014	12	161	133	55	78	6.38	Yes
LIVCA	GSE55092 [688]	2014	2	140	44	20	24	4.38	Yes
LIVCA	GSE62232 [689]	2014	2	91	86	81	5	6.12	Yes
LIVCA	GSE76427 [690]	2015	7	167	101	50	51	5.12	Yes
LIVCA	GSE64041 [691]	2016	3	44	125	60	65	3.62	Yes
OVCA	GSE10971 [692]	2008	2	37	37	13	24	3.62	Yes
OVCA	GSE14407 [693]	2009	2	24	19	7	12	4.12	Yes
OVCA	GSE36668 [694]	2012	2	12	12	8	4	3.38	Yes
OVCA	GSE40595 [695]	2014	2	77	75	61	14	2.38	Yes
OVCA	GSE69428 [696]	2015	2	29	17	8	9	3.5	Yes

OVCA	GSE66957	2015	18	69	62	57	5	4	Yes
PACA	GSE15471 [697]	2009	2	78	72	36	36	4.38	Yes
PACA	GSE19650 [698]	2010	2	22	19	12	7	6.62	Yes
PACA	GSE32676 [699]	2011	2	32	32	25	7	5.62	Yes
PACA	GSE41368 [700]	2013	3	12	12	6	6	3.75	Yes
PACA	GSE43795	2013	7	31	11	6	5	6.12	Yes
PACA	GSE55643 [701]	2014	12	53	21	18	3	6.75	Yes
PACA	GSE62165 [702]	2016	11	131	131	118	13	2.38	Yes
PACA	GSE62452 [703]	2016	3	130	130	69	61	2.88	Yes
PACA	GSE63111 [704]	2017	10	35	35	28	7	6.5	Yes
PRCA	GSE21034 [705]	2010	10	370	160	131	29	3.5	Yes
PRCA	GSE29079 [706]	2011	10	95	95	47	48	5.12	Yes
PRCA	GSE62872 [707]	2014	3	424	424	264	160	2.38	Yes
PRCA	GSE46602 [708]	2015	2	50	50	36	14	4.12	Yes
PRCA	GSE70768 [709]	2015	7	199	199	125	74	3.12	Yes
PRCA	GSE71016 [710]	2016	16	95	91	45	46	2.75	Yes
SKCM	GSE3189 [711]	2005	1	70	69	44	25	3.75	Yes
SKCM	GSE7553 [712]	2008	2	87	18	14	4	3.12	Yes
SKCM	GSE15605 [713]	2012	2	74	62	46	16	2.38	Yes
SKCM	GSE46517 [714]	2013	1	121	45	30	15	2	Yes
SKCM	GSE57715 [715]	2014	19	297	292	275	17	3.75	Yes
STCA	GSE13911 [716]	2008	2	69	69	38	31	3.88	Yes
STCA	GSE13195	2009	10	100	50	25	25	4.5	Yes
STCA	GSE27342 [717]	2011	10	160	146	73	73	5	Yes
STCA	GSE29998 [718]	2012	4	99	99	50	49	4.12	Yes
STCA	GSE30727	2014	10	60	60	30	30	4.5	Yes
STCA	GSE51575	2014	14	52	36	15	21	6.25	Yes
STCA	GSE26899 [719]	2016	4	108	52	41	11	5.88	Yes
STCA	GSE79973 [720]	2016	2	20	20	10	10	4.88	Yes
STCA	GSE54129	2017	2	132	132	111	21	6	Yes
THCA	GSE3467 [721]	2005	2	18	18	9	9	2.75	Yes
THCA	GSE53157 [722]	2013	2	27	27	24	3	4.5	Yes
THCA	GSE35570 [723]	2015	2	116	116	65	51	1.75	Yes
THCA	GSE60542 [724]	2015	2	92	59	29	30	2.62	Yes
THCA	GSE65144 [725]	2015	2	25	25	12	13	3.38	Yes

Appendix 5: Study characteristics for CNS disorders. Platform codes: 1- hgu133a, 2- hgu133plus2, 3- Human Gene 1.0 ST, 4- Illumina HumanHT-12 V3.0, 5- Illumina humanRef-8 v2.0, 6- Affymetrix Human X3P Array, 7- Illumina HumanHT-12 V4.0, 8-Illumina human-6 v2.0 expression beadchip, 9- Illumina HumanWG-6 v3.0, 10- Affymetrix Human Exon 1.0 ST Array, 11- Affymetrix Human Genome U219 Array, 12- Agilent-014850 Whole Human Genome Microarray 4x44K G4112F, 13- Sentrix HumanRef-8 Expression BeadChip, 14- Agilent-028004 SurePrint G3 Human GE 8x60K Microarray, 15- Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip, 16- Agilent-039494

SurePrint G3 Human GE v2 8x60K Microarray 039381, 17- Affymetrix Human Gene 1.1 ST Array, 18- Rosetta/Merck Human RSTA Custom Affymetrix 2.0, 19- Illumina HumanRef-8 WG-DASL v3.0.

Appendix 6: Site-specific cancers available in the TCGA validation cohort

TCGA Dataset	TCGA dataset description	Correspondance to array analysis abbreviations	N	Nº Cases	Nº Controls	Nº Tested Genes	Nº DEGs	Nº Up	Nº Down
BLCA	Bladder Urothelial Carcinoma	BLCA	433	414	19	16251	9043	6045	2998
BRCA	Breast Invasive Carcinoma	BRCA	1215	1102	113	16251	13665	9026	4639
GBM	Glioblastoma Multiforme	BRNCA	161	156	5	16251	9187	5639	3548
CESC	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma	CERV	307	304	3	16251	4418	2407	2011
CHOL	Cholangiocarcinoma	CHOL	45	36	9	16251	8989	6055	2934
COAD	Colon Adenocarcinoma	CRCA	519	478	41	16250	12705	7938	4767
READ	Rectum Adenocarcinoma	CRCA	176	166	10	16251	9481	5934	3547
HNSC	Head and Neck Squamous Cell Carcinoma	HANC	544	500	44	16251	11334	6668	4666
KICH	Kidney Chromophobe	KDNCA	89	65	24	16251	11665	7257	4408
KIRC	Kidney Renal Clear Cell Carcinoma	KDNCA	610	538	72	16251	13291	7999	5292
LUAD	Lung Adenocarcinoma	LGCA	592	533	59	16251	12834	8590	4244
LUSC	Lung Squamous Cell Carcinoma	LGCA	551	502	49	16251	13367	8417	4950
LIHC	Liver Hepatocellular Carcinoma	LIVCA	421	371	50	16251	11597	8490	3107
PAAD	Pancreatic Adenocarcinoma	PACA	181	177	4	16251	842	94	748
PRAD	Prostate Adenocarcinoma	PRCA	550	498	52	16229	11644	7023	4621
STAD	Stomach Adenocarcinoma	STCA	407	375	32	16251	10306	5714	4592
THCA	Thyroid Carcinoma	THCA	560	502	58	16251	11638	6040	5598

Appendix 6: Table summarizing the types of cancers derived from TCGA and the results of the differential gene expression analyses.

Appendix 7: Disease identifiers used to query each databse. Disease associated genes and variant-genes for each disorder included in interactome 1 stringent analysis and genes included in the largest connected component.

Disorder	IDs used in the different DBs.	Disease associated genes and variant-genes (stringent)	Genes included in the largest connected component in the interactome 1 stringent analysis
AD	**-DisGeNet MESH:** D000544 **-EDGAR:** 104300 **-PhenGeID trait:** Alzheimer Disease	SORL1, NOS3, ACE, PLAU, MPO, APP, PSEN1, APBB2, PAXIP1, BLMH, PSEN2, HFE, A2M, ADAM10, GSK3B, APOE, MAPT, TREM2, BACE1, IDE, IL1B, INSR, LEP, NPY, BCL2, BDNF, CASP3, IL6R, CR1, MRPL50P4, SPON1, MS4A2, MS4A6A, MIR6503, PICALM, MMP3, BCAS3, CYB561, ABCA7, PVRL2, TOMM40, APOC1, BIN1, ACKR2, CCRL2, LOC102724297, SUCLG2, SNAR-I, RANP6, LOC100289673, HLA-DRB1, CD2AP, GAPDHP15, BZW2, EPHA1-AS1, PTK2B, CLU, MIR6843, HNF4G, SLC16A9, MS4A4A, FNTAP1, MMP12, CYP27C1, OSTN, FBXO8, HLA-DQA1, RBBP4P4, LINC01111	PSEN1, APP, TOMM40, GSK3B, PSEN2, MAPT
ASD	**-DisGeNet MESH:** D000067877 **-EDGAR:** PS209850, 605309 **-PhenGeID** trait: Autism Spectrum Disorder|Autistic Disorder|Child Development Disorders, Pervasive|Asperger Syndrome	NLGN3, CHD8, MECP2, NLGN4X, SLC9A9, MET, EN2, CNTNAP2, RPL10, PTCHD1, EIF4E, SHANK2, TMLHE, PTEN, NRXN2, SHANK3, NRXN1, ITIH3, TRIM26	EIF4E
BD	**-DisGeNet MESH:** D001714 **-EDGAR:** NA **-PhenGeID trait:** Bipolar Disorder	S100B, COMT, ANK3, CACNA1C, NCAN, SP4, ADCY2, POLG, LMAN2L, FADS2, DRD1, GAD1, NR3C1, GSK3B, ITIH1, MTHFR, BDNF, SLC6A4, CLOCK, ITIH3, TRIM26, MAD1L1	ANK3, NR3C1, GSK3B, CLOCK, BDNF
HD	**-DisGeNet MESH:** D006816 **-EDGAR:** 143100 **-PhenGeID trait:** NA	HTT	HTT
MD	**-DisGeNet MESH:** D003866 **-EDGAR:** 608516 **-PhenGeID trait:** Depression|Depressive Disorder|Depressive Disorder, Major	FKBP5, TPH2, APOE, DISC1, SLC6A4, FGFR1, SOD1, CRH, NR3C1, IL6, KCNK2, MTHFR, NPY, PTGS2, BDNF, S100A10, TOMM40, SEPT3, WBP2NL, CYP2D6, ITIH3, TRIM26, CYP2C9	NR3C1, BDNF
PD	**-DisGeNet MESH:** D010300 **-EDGAR:** PS168600 **-PhenGeID trait:** Parkinson Disease	GIGYF2, GBA, PODXL, PINK1, HTRA2, CHCHD2, LRRK2, PARK7, ATP13A2, PARK2, PLA2G6, SYNJ1, DNAJC13, VPS13C, NR4A2, UCHL1, VPS35, SNCAIP, TBP, EIF4G1, GLUD2, ADH1C, MAPT, SNCA, ATXN2, ATXN3, FGF20, DDC, DRD2, MAOB, SLC18A2, TH, RAB25, NUCKS1, RAB29, TIAL1, INPP5F, SLC2A13, CNTN1, GCH1, TMEM229B, TPM1, BCKDK, CRHR1,	CHCHD2, PARK7, PARK2, SNCA, PINK1

		SPPL2C, NSF, WNT3, RIT2, DCUN1D1, MCCC1, GAK, TMEM175, DGKQ, BST1, FAM47E-STBD1, LHFPL2, HLA-DRB1, HLA-DQB1, FAM126A, GPNMB, KRTCAP2, SLC41A1, SYT10, ACMSD, CERS6, DDRGK1, MMRN1, HLA-DRA, HLA-DQA1	
SCZ	**-DisGeNet MESH:** D012559 **-EDGAR:** 181500|604906|603013|600850|613950|615232 **-PhenGeID trait:** Schizophrenia	DRD3, DISC1, MTHFR, DAO, COMT, SETD1A, AKT1, RTN4R, CHI3L1, DTNBP1, APOL4, SYN2, APOL2, MC4R, SLC1A1, SHANK3, PRODH, NRG1, MAGI2, CHRNA7, GRIN2B, NOS1, RELN, SRR, TCF4, NRXN1, SP4, PPP3R1, SYNGAP1, MDK, GRM5, GSK3B, ZDHHC8, APOE, NR4A2, SLC6A3, PPP1R1B, KMO	AKT1, GSK3B
ALL	**-DisGeNet MESH:** D054198 **-EDGAR:** 613065 **-PhenGeID trait:** Precursor B-Cell Lymphoblastic Leukemia-Lymphoma|Precursor Cell Lymphoblastic Leukemia-Lymphoma	TAL1, TCF3, TAL2, BAX, PAX5, NBN, NUP214, FLT3, BCR, IKZF1, CDKN2A, ABL1, GATA3, LHPP, CEBPE, C14orf119	TCF3, ABL1, GATA3, TAL1, FLT3, BCR, CEBPE
AML	**-DisGeNet MESH:** D015470 **-EDGAR:** 601626 **-PhenGeID trait:** Myeloproliferative Disorders	*TGM6, SETBP1, SH3GL1, FLT3, CHIC2, CEBPA, NPM1, WHSC1L1, CBFB, JAK2, NUP214, TERT, MLF1, MLLT10, LPP, GATA2, KRAS, ETV6, DDX41, KIT, RUNX1, NSD1, PICALM, DNMT3A, PTPN11, IDH1, IDH2, NRAS, TP53, WT1, SBDS, CREBBP, KMT2A, SPI1*	*CBFB, PICALM, MLLT10, CHIC2, DNMT3A, RUNX1, NSD1, KMT2A, CREBBP, NRAS, SPI1, TP53, CEBPA, ETV6, LPP, WHSC1L1*
BLCA	**-DisGeNet MESH:** D001749 **-EDGAR:** 109800 **-PhenGeID trait:** Urinary Bladder Neoplasms	*HRAS, FGFR3, RB1, KRAS, ATM, CDH1, NQO1, ERCC2, GSTP1, TP53, CDKN2A, TSC1, SLC14A1, CBX6, TACC3, CLPTM1L, CWC27, NAT2, PSCA, CCNE1, APOBEC3A, PSD3*	*ATM, TP53*
BRCA	**-DisGeNet MESH:** D001943 **-EDGAR:** 114480 **-PhenGeID trait:** Breast Neoplasms|Triple Negative Breast Neoplasms	*TSG101, HMMR, ATM, NQO2, AKT1, BRIP1, XRCC3, RB1CC1, PPM1D, RAD54L, FAM175A, NBN, CDH1, CHEK2, BRCA1, BRCA2, BARD1, KRAS, TP53, SLC22A18, ESR1, PHB, PALB2, CASP8, RAD51, PIK3CA, ERBB2, PTEN, CAV1, EP300, FGFR2, NOTCH2, CDKN1B, PARP1, NQO1, AKT2, ESR2, FGF3, FGFR1, FLT1, FN1, GATA3, FOXA1, HRAS, IGF1, AR, MDM2, MMP1, NOS2, NOTCH1, ROR1, FBXW7, PTHLH, RB1, STAT1, TBX3, NCOA3, BAP1, FGF4, TRIM33, MDM4, MLLT10, DNAJC1, ZNF365, ZMIZ1, LSP1, KRT8, USP44, PAX9, PELI2, RAD51B, NTRK3, TOX3, FTO, CDYL2, HNF1B, STXBP4, BABAM1, ELL, ERBB4,*	*IGF1, SLC4A7, HMMR, MLLT10, AKT2, BABAM1, GATA3, CDKN1B, NQO2, BARD1, ESR2, AKT1, ATM, TNXB, PAX9, BAG6, CDSN, MDM4, EP300, PRRC2A, RAD54L, TRIM33, PIK3CA, RB1CC1, NBN, RAD51, NOTCH4, RB1, PHB, NCOA3, BAP1, CAV1, PPM1D, STXBP4, ELL, STAT1, TP53, MDM2, BRCA1, CCND1, ESR1, BRCA2, AR, CHEK2, ERBB2, FGFR2, NOTCH1, FAM175A, SEMA3A, FOXA1, ERBB4, ENPP2, HNF1B, BRIP1, HNF4G, PTEN, PTHLH, NTRK3, USP44, GLI2, TERT*

		TNP1, NRIP1, CYYR1, MKL1, ITPR1, NEK10, SLC4A7, TERT, MAP3K1, MIER3, EBF1, CDSN, PSORS1C1, DDX39B, MCCD1, AIF1, BAG6, EHMT2, C2, TNXB, NOTCH4, HLA-DRA, HLA-DRB1, HLA-DQA1, HLA-DQB1, TAP2, ADGRB3, UST, CCDC170, SEMA3A, ARHGEF5, NOV, SNX32, MYEOV, CCND1, CCDC91, GLI2, MRPS30, GPBP1, HLA-C, PRRC2A, HLA-DRB5, HLA-DOB, TAB2, HNF4G, ENPP2	
BRNCA	**-DisGeNet MESH:** D001932 **-EDGAR:** PS137800 **-PhenGeID trait:** Glioblastoma|Glioma	*ERBB2, BRCA2, TP53, PTEN, IDH2, IDH1, POT1, PPARG, ALK, RTEL1, TERT, SEC61G, ZBTB16, PHLDB1, POLR2A, EGFR, PHLDA1*	*TP53, EGFR, ERBB2, PTEN*
CERV	**-DisGeNet MESH:** D002583 **-EDGAR:** 603956 **-PhenGeID trait:** Uterine Cervical Neoplasms	*FGFR3, HLA-DRB1, HLA-DQA1*	*HLA-DRB1, HLA-DQA1*
CHLCA	**-DisGeNet MESH:** D018281 **-EDGAR:** NA **-PhenGeID trait:** NA	*IDH1, TP53, KRAS*	*TP53*
CLL	**-DisGeNet MESH:** D015451 **-EDGAR:** NA **-PhenGeID trait:** Leukemia, Lymphocytic	*TP53, ATM, PLCG2, POT1, ID3, ACTA2, FAS, TSPAN32, C11orf21, BMF, MNS1, RPLP1, PHLPP1, BCL2, ACOXL, BCL2L11, CFLAR, SP110, SP140, FARP2, EOMES, ULK4, IRF4, HLA-DRB1, BAK1, TMPRSS5, BUB1B, ZNF280D, IRF8, CMC1, EXOC2, HLA-DQA1*	*BMF, ATM, ZNF280D, TP53, BCL2*
CML	**-DisGeNet MESH:** D054438 **-EDGAR:** NA **-PhenGeID trait:** Leukemia, Myeloid, Chronic-Phase	*SETBP1, BCR*	*SETBP1*
CRCA	**-DisGeNet MESH:** D015179 **-EDGAR:** 114500 **-PhenGeID trait:** Colorectal Neoplasms	*MT-CO1, BUB1B, MLH3, BAX, PDGFRL, CTNNB1, AKT1, ODC1, AXIN2, RAD54B, PIK3CA, NRAS, AURKA, DLC1, TRIM28, PLA2G2A, EP300, FGFR3, TP53, CCND1, PTPN12, MCC, KAT5, TLR2, BRAF, TLR4, APC, PTPRJ, MLH1, MSH2, CHEK2, SMAD3, SMAD4, MMP2, TCF7L2, BUB1, KLF5, MSH6, IGFBP3, KDR, KRAS, ABCB1, POLD1, POLE, RET, STK11, DPYD, TYMS, NKX2-3, CYP17A1, FGFR2, MYRF, POLD3, SPSB2, ETV6, KRT8, BRCA2, GREM1, HNF1B, SMAD7, UTP23, SLC25A28, MYEOV, RPS21, MAP3K1, EIF3H, RAD21*	*MLH1, SMAD4, AKT1, TCF7L2, SMAD3, BUB1, TRIM28, EP300, BRAF, BUB1B, PIK3CA, RAD21, MSH6, TYMS, AURKA, APC, STK11, TP53, MSH2, CTNNB1, RET, CHEK2, FGFR2, KDR, FGFR3, NKX2-3*
DLBCL	**-DisGeNet MESH:** D016403 **-EDGAR:** NA **-PhenGeID trait:** Lymphoma, Large B-Cell, Diffuse	*EZH2, MYD88, PIK3CD, CARD11, CD79B, IRF4, EXOC2*	*CD79B*

FLYMPH	**-DisGeNet MESH:** D008224 **-EDGAR:** NA **-PhenGeID trait:** Lymphoma, Follicular	*EZH2, BCL2, HLA-DQB1, CXCR5, FLI1, C6orf15, HLA-DRB5*	*EZH2*
HANC	**-DisGeNet MESH:** D006258 **-EDGAR:** 275355 **-PhenGeID trait:** Carcinoma, squamous cell of head and neck|Head and Neck Neoplasms	*PTEN, TNFRSF10B, EGFR, ING1, ING3, ADH1B, ADH7*	*EGFR, PTEN*
KDNCA	**-DisGeNet MESH:** D002292 **-EDGAR:** 144700|605074 **-PhenGeID trait:** Carcinoma, Renal Cell	*DIRC2, HNF1B, VHL, HNF1A, RNF139, OGG1, PRCC, MET, EPAS1, PTEN, SETD2, PBRM1, PTGS2, MTOR, TSC1, KDM5C, BAP1*	*SETD2, PBRM1, HNF1A, PTGS2, HNF1B*
LGCA	**-DisGeNet MESH:** D008175 **-EDGAR:** 211980 **-PhenGeID trait:** Adenocarcinoma of lung|Carcinoma, Non-Small-Cell Lung|Small Cell Lung Carcinoma	*ERBB2, IRF1, SLC22A18, ERCC6, RASSF1, PIK3CA, EGFR, MAP3K8, CASP8, PARK2, FASLG, BRAF, PPP2R1B, KRAS, CYP2A6, CDKN2A, TP53, BRCA2, PTEN, GSTP1, ERCC1, STK11, VTI1A, FGFR2, KRT8, HNF1B, BPTF, TP63, TERT, BAG6, APOM, MYEOV, MAP3K1*	*GSTP1, BRAF, STK11, TP53, EGFR, ERBB2, FGFR2, PARK2, IRF1, CYP2A6, CDKN2A, PTEN*
LIVCA	**-DisGeNet MESH:** D006528 **-EDGAR:** 114550 **-PhenGeID trait:** Carcinoma, Hepatocellular	*TP53, AXIN1, MTUS1, CDKN3, PIK3CA, CASP8, PDGFRL, CTNNB1, MET, APC, IGF2R, HNF1A, CDKN2A, IGF2, KRAS, HTATIP2, ARID2, FOXM1, GPC3, GNMT, MYC, ABCB1, PTEN, PTGS2, PTK2, HAMP, SKP2, TERT, TGFA, CCNE1, CDK1, KIF1B, HLA-DRB1, HLA-DQB1, HLA-DQA1*	*CASP8, MTUS1, PTK2, FOXM1, CDK1, AXIN1, APC, TP53, CTNNB1, MYC, MET, IGF2R, HNF1A, PTGS2, GPC3*
OVCA	**-DisGeNet MESH:** D010051 **-EDGAR:** 167000 **-PhenGeID trait:** Ovarian Neoplasms|Ovarian epithelial cancer	*BRCA1, PIK3CA, RRAS2, PARK2, CTNNB1, AKT1, SEPT9, CDH1, OPCML, BRCA2, ERBB2, KRAS, MLH1, MSH2, PTEN, TP53, RAD51C, BRIP1, PMS2, ESR1, CCNE1, RAD51D, BRAF, RSPO1, FGFR2, KRT8, HNF1B, NSF, SKAP1, BABAM1, HOXD3, MYEOV, MAP3K1, IFNL3*	*MLH1, AKT1, BRAF, PMS2, TP53, BRCA1, MSH2, ESR1, CDH1, CTNNB1, BRCA2, ERBB2, FGFR2, BRIP1, PTEN*
PACA	**-DisGeNet MESH:** D010190 **-EDGAR:** 260350|613348|606856|614320|613347 **-PhenGeID trait:** Pancreatic Carcinoma|Pancreatic Neoplasms	*RBBP8, TP53, ACVR1B, SMAD4, STK11, KRAS, PALLD, BRCA2, PALB2, BRCA1, CDKN2A, BACH1, TFF2, TERT, CLPTM1L, ABO, NR5A2, PRLHR, TFF1*	*RBBP8, BRCA1, BACH1, BRCA2*
PRCA	**-DisGeNet MESH:** D011471 **-EDGAR:** 176807|614731|601518|611928|611868 **-PhenGeID trait:** Prostatic Neoplasms	*EPHB2, BRCA2, PTEN, MAD1L1, HIP1, CD82, ZFHX3, KLF6, AR, MXI1, CDH1, FGFR4, MSR1, CHEK2, HOXB13, RNASEL, MSMB, ELAC2, EHBP1, APC, SPOP, CTNNB1, IGF1, NKX3-1, KRAS, CCND1, TGFBR2, KCND3, GOLPH3L, MDM4, FGFR2, MMP7, TUBA1C, KRT8, FERMT2, VPS53, HNF1B, DPF1, PCAT19, KLK15, KLK3, GGCX,*	*TBX5, FERMT2, APC, CDH1, CTNNB1, FGFR2, PTEN*

		TANC1, ITGA6, ADNP, ZGPAT, XAGE3, TEX11, SIDT1, PRKCI, SKIL, AFM, PDLIM5, TERT, RFX6, RGS17, SLC22A1, JAZF1, EBF2, FAM84B, RAD23B, ASCL2, MYEOV, PRPH, TBX5, PPP1R14A, KLK2, LILRA5, VAMP8, BIK, NUDT11, SLC7A3, CLDN11, MAP3K1, SLC22A2	
SKCM	**-DisGeNet MESH:** D008545 **-EDGAR:** PS155600 **-PhenGeID trait:** Melanoma	*CDKN2A, STK11, POT1, PTEN, CDK4, BRAF, MITF, XRCC3, TERT, TYR, HRAS, ERBB4, GNA11, GNAQ, NRAS, MAP2K1, TP53, BAP1, MAP2K2, NF1, RAC1, FTO, CDK10, AFG3L1P, RALY, PIGU, MYH7B, SLC45A2, IRF4, MTAP, CCND1, EXOC2*	*MTAP, RALY, MAP2K2, FTO, MAP2K1, NF1, BRAF, GNAQ, STK11, TP53, GNA11, TYR, PTEN, MITF*
STCA	**-DisGeNet MESH:** D013274 **-EDGAR:** 137215|613659 **-PhenGeID trait:** Stomach Neoplasms	*ERBB2, MUTYH, IRF1, PIK3CA, CASP10, KLF6, APC, FGFR2, KRAS, CDH1, IL1RN, IL1B, MET, ATM, MUC1, ASH1L, PRKAA1, DNAH11, PSCA*	*ERBB2, IRF1, KLF6*
THCA	**-DisGeNet MESH:** D013964 **-EDGAR:** PS188550 **-PhenGeID trait:** Thyroid Neoplasms	*NKX2-1, FOXE1, BRAF, HABP2, RET, KRAS, TSHR, TP53, PCNXL2, OBFC1, NRG1, SLK, MBIP*	*NRG1, BRAF, TP53, TSHR*

Appendix 7: Table including the disease identifiers employed to query the gene-disease association databases. Genes and variant-genes identified for each disorder using the stringent selection setting, which are included in the interactome 1. Genes included in the largest connected component.

Appendix 8: Drug indications identified for all the included disorders

Disorder	ICD9	Drug name	Drugbank_ID	RXCUI_IN
AD	331.0	Galantamine	DB00674	4637
AD	331.0	Memantine	DB01043	6719
AD	331.0	Selegiline	DB01037	9639
AD	331.0	Tacrine	DB00382	10318
AD	331.0	Valproic Acid	DB00313	11118
AD	331.0	Vitamin E	DB00163	11256
AD	331.0	Tocopherol Acetate	DB14003	39625
AD	331.0	donepezil	DB00843	135447
AD	331.0	rivastigmine	DB00989	183379
AD	331.0	alpha-Tocopherol Acetate	DB14002	1046243
ASD	299.0;299.00	Fenfluramine	DB00574	4328
ASD	299.0;299.00	Risperidone	DB00734	35636
ASD	299.0;299.00	venlafaxine	DB00285	39786
ASD	299.0;299.00	quetiapine	DB01224	51272
ASD	299.0;299.00	aripiprazole	DB01238	89013
BD	296.7;296.80	Carbamazepine	DB00564	2002
BD	296.80	Clonazepam	DB01068	2598
BD	296.80	Clozapine	DB00363	2626
BD	296.80	Fluoxetine	DB00472	4493
BD	296.80	Valproic Acid	DB00313	11118
BD	296.80	gabapentin	DB00996	25480
BD	296.7;296.80	lamotrigine	DB00555	28439
BD	296.80	tiagabine	DB00906	31914
BD	296.80	oxcarbazepine	DB00776	32624
BD	296.80	Paroxetine	DB00715	32937
BD	296.80	Risperidone	DB00734	35636
BD	296.80	topiramate	DB00273	38404
BD	296.80	Lithium Carbonate	DB14509	42351
BD	296.7;296.80	quetiapine	DB01224	51272
BD	296.80	lithium citrate	DB14507	52105
BD	296.7;296.80	olanzapine	DB00334	61381
BD	296.80	aripiprazole	DB01238	89013
BD	296.80	ziprasidone	DB00246	115698
BD	296.7;296.80	Asenapine	DB06216	784649
BD	296.80	gabapentin enacarbil	DB08872	1101333
HD	333.4	Baclofen	DB00181	1292
HD	333.4	Perphenazine	DB00850	8076
HD	333.4	Pimozide	DB01100	8331
HD	333.4	coenzyme Q10	DB09270	21406
MD	311	Alprazolam	DB00404	596
MD	311	Amitriptyline	DB00321	704
MD	311	Amoxapine	DB00543	722

MD	311	Buspirone	DB00490	1827
MD	311	Chlorpromazine	DB00477	2403
MD	296.20;311	Citalopram	DB00215	2556
MD	311	Clomipramine	DB01242	2597
MD	311	Desipramine	DB01151	3247
MD	311	Dexamethasone	DB01234	3264
MD	311	Doxepin	DB01142	3638
MD	296.20;311	Fluoxetine	DB00472	4493
MD	296.20;311	Imipramine	DB00458	5691
MD	311	Isocarboxazid	DB01247	6011
MD	311	Isoflurane	DB00753	6026
MD	311	Lithium	DB01356	6448
MD	311	Lorazepam	DB00186	6470
MD	311	Maprotiline	DB00934	6646
MD	311	Methylphenidate	DB00422	6901
MD	296.20;311	Nortriptyline	DB00540	7531
MD	311	Phenelzine	DB00780	8123
MD	311	Protriptyline	DB00344	8886
MD	296.20;311	Selegiline	DB01037	9639
MD	311	Temazepam	DB00231	10355
MD	311	Tranylcypromine	DB00752	10734
MD	296.20;311	Trazodone	DB00656	10737
MD	296.20;311	Trimipramine	DB00726	10834
MD	311	Tryptophan	DB00150	10898
MD	296.20	Valproic Acid	DB00313	11118
MD	296.20;311	Mirtazapine	DB00370	15996
MD	311	lamotrigine	DB00555	28439
MD	311	nefazodone	DB01149	31565
MD	296.20;311	Paroxetine	DB00715	32937
MD	311	Risperidone	DB00734	35636
MD	296.20;311	Sertraline	DB01104	36437
MD	296.20;311	venlafaxine	DB00285	39786
MD	296.20;311	Bupropion	DB01156	42347
MD	311	Lithium Carbonate	DB14509	42351
MD	311	Fluvoxamine	DB00176	42355
MD	311	quetiapine	DB01224	51272
MD	311	lithium citrate	DB14507	52105
MD	311	olanzapine	DB00334	61381
MD	296.20	duloxetine	DB00476	72625
MD	296.20;311	aripiprazole	DB01238	89013
MD	311	Pramipexole dihydrochloride	DB00413	236747
MD	311	Kava preparation	DB01322	285228
MD	296.20;311	Escitalopram	DB01175	321988
MD	296.20	Desvenlafaxine	DB06700	734064
MD	296.20	vilazodone	DB06684	1086769

PD	332	Amantadine	DB00915	620
PD	332	Apomorphine	DB00714	1043
PD	332	Trihexyphenidyl Hydrochloride	DB00376	1115
PD	332	Hyoscyamine Sulfate	DB00424	1225
PD	332	Benztropine	DB00245	1424
PD	332	Biperiden	DB00810	1589
PD	332	Bromocriptine	DB01200	1760
PD	332	Carbidopa	DB00190	2019
PD	332	Levodopa	DB01235	6375
PD	332	Pergolide	DB01186	8047
PD	332	Selegiline	DB01037	9639
PD	332	cabergoline	DB00248	47579
PD	332	entacapone	DB00494	60307
PD	332	ropinirole	DB00268	72302
PD	332	tolcapone	DB00323	72937
PD	332	rasagiline	DB01367	134748
PD	332	rivastigmine	DB00989	183379
PD	332	Pramipexole dihydrochloride	DB00413	236747
PD	332	Rotigotine	DB05271	616739
SCZ	295;295.9;295.90	Carbamazepine	DB00564	2002
SCZ	295;295.9;295.90	Chlorpromazine	DB00477	2403
SCZ	295.3	Chlorprothixene	DB01239	2406
SCZ	295;295.9;295.90	Clonazepam	DB01068	2598
SCZ	295;295.9;295.90	Clozapine	DB00363	2626
SCZ	295;295.9;295.90	Cyproheptadine	DB00434	3013
SCZ	295;295.9;295.90	Fluphenazine	DB00623	4496
SCZ	295;295.9;295.90	Haloperidol	DB00502	5093
SCZ	295;295.9;295.90	Loxapine	DB00408	6475
SCZ	295;295.9;295.90	Mesoridazine	DB00933	6779
SCZ	295;295.9;295.90	Molindone	DB01618	7019
SCZ	295;295.9;295.90	Perphenazine	DB00850	8076
SCZ	295;295.9;295.90	Pimozide	DB01100	8331
SCZ	295;295.9;295.90	Prochlorperazine	DB00433	8704
SCZ	295;295.9;295.90	Reserpine	DB00206	9260
SCZ	295;295.9;295.90	Selegiline	DB01037	9639
SCZ	295;295.9;295.90	Thioridazine	DB00679	10502
SCZ	295;295.9;295.90	Thiothixene	DB01623	10510
SCZ	295;295.9;295.90	Trazodone	DB00656	10737
SCZ	295;295.9;295.90	Trifluoperazine	DB00831	10800
SCZ	295;295.9;295.90	acetophenazine	DB01063	16735
SCZ	295;295.9;295.90	Risperidone	DB00734	35636
SCZ	295;295.9;295.90	quetiapine	DB01224	51272
SCZ	295;295.9;295.90	olanzapine	DB00334	61381
SCZ	295;295.9;295.90	deserpidine	DB01089	62174

SCZ	295;295.9;295.90	iloperidone	DB04946	73178
SCZ	295;295.9;295.90	aripiprazole	DB01238	89013
SCZ	295;295.9;295.90	ziprasidone	DB00246	115698
SCZ	295;295.9;295.90	paliperidone	DB01267	679314
SCZ	295;295.9;295.90	Asenapine	DB06216	784649
SCZ	295;295.9;295.90	Lurasidone Hydrochloride	DB08815	1040027
ALL	204.0;204.00	mercaptopurine	DB01033	103
ALL	204.0;204.00	ASPARAGINASE	DB00023	1156
ALL	204.0;204.00	Cytarabine	DB00987	3041
ALL	204.0;204.00	Daunorubicin	DB00694	3109
ALL	204.0;204.00	Doxorubicin	DB00997	3639
ALL	204.0;204.00	Teniposide	DB00444	10362
ALL	204.0;204.00	pegaspargase	DB00059	34132
ALL	204.0;204.00	clofarabine	DB00631	44151
ALL	204.0;204.00	imatinib	DB00619	282388
ALL	204.0;204.00	dasatinib	DB01254	475342
AML	205.00	ASPARAGINASE	DB00023	1156
AML	205.0;205.00	Cyclophosphamide	DB00531	3002
AML	205.0;205.00	Cytarabine	DB00987	3041
AML	205.0;205.00	Daunorubicin	DB00694	3109
AML	205.0;205.00	Doxorubicin	DB00997	3639
AML	205.0;205.00	Etoposide	DB00773	4179
AML	205.0;205.00	Idarubicin	DB01177	5650
AML	205.0;205.00	Mitoxantrone	DB01204	7005
AML	205.0;205.00	Thioguanine	DB00352	10485
AML	205.00	Tretinoin	DB00755	10753
AML	205.0;205.00	arsenic trioxide	DB01169	18330
AML	205.0;205.00	sargramostim	DB00020	69634
AML	205.0;205.00	Gemtuzumab ozogamicin	DB00056	1294580
BLCA	188;188.9;239.4	Mitomycin	DB00305	632
BLCA	188;188.9;239.4	Cisplatin	DB00515	2555
BLCA	188.9;239.4	Doxorubicin	DB00997	3639
BLCA	239.4	Etoposide	DB00773	4179
BLCA	188;188.9	Fluorouracil	DB00544	4492
BLCA	188;188.9;239.4	Thiotepa	DB04572	10473
BLCA	188;188.9	gemcitabine	DB00441	12574
BLCA	188;188.9;239.4	valrubicin	DB00385	31435
BLCA	188;188.9;239.4	Carboplatin	DB00958	40048
BRCA	174.9	Cyclophosphamide	DB00531	3002
BRCA	174.9	Epirubicin	DB00445	3995
BRCA	174.9	Estradiol	DB00783	4083
BRCA	174.9	Estrogens Conjugated (USP)	DB00286	4099
BRCA	174.9	Fluorouracil	DB00544	4492
BRCA	174.9	Fluoxymesterone	DB01185	4494
BRCA	174.9	Ifosfamide	DB01181	5657

BRCA	174.9	Megestrol	DB00351	6703
BRCA	174.9	Methotrexate	DB00563	6851
BRCA	174.9	Methyltestosterone	DB06710	6904
BRCA	174.9;233.0	Tamoxifen	DB00675	10324
BRCA	174.9	Testolactone	DB00894	10378
BRCA	174.9	Testosterone	DB00624	10379
BRCA	174.9	Thiotepa	DB04572	10473
BRCA	174.9	pamidronate	DB00282	11473
BRCA	174.9	gemcitabine	DB00441	12574
BRCA	174.9	testosterone enanthate	DB13944	37859
BRCA	174.9	Toremifene	DB00539	38409
BRCA	174.9	vinorelbine	DB00361	39541
BRCA	174.9	Dexrazoxane	DB00380	42736
BRCA	174.9	Goserelin	DB00014	50610
BRCA	174.9	Paclitaxel	DB01229	56946
BRCA	174.9;233.0	Raloxifene	DB00481	72143
BRCA	174.9	letrozole	DB01006	72965
BRCA	174.9	zoledronic acid	DB00399	77655
BRCA	174.9	anastrozole	DB01217	84857
BRCA	174.9	capecitabine	DB01101	194000
BRCA	174.9	Estrogens Esterified (USP)	DB09381	214549
BRCA	174.9	trastuzumab	DB00072	224905
BRCA	174.9	bevacizumab	DB00112	253337
BRCA	174.9	exemestane	DB00990	258494
BRCA	174.9	fulvestrant	DB00947	282357
BRCA	174.9	ixabepilone	DB04845	337523
BRCA	174.9	lapatinib	DB01259	480167
BRCA	174.9	eribulin	DB08871	1045453
BRNCA	191.9	Carmustine	DB00262	2105
BRNCA	191.9	Cisplatin	DB00515	2555
BRNCA	191;191.9	Cyclophosphamide	DB00531	3002
BRNCA	191.9	temozolomide	DB00853	37776
CERV	180.9	Cisplatin	DB00515	2555
CERV	180.9	Ifosfamide	DB01181	5657
CERV	180.9	Topotecan	DB01030	57308
CLL	204.1	mercaptopurine	DB01033	103
CLL	204.1	Chlorambucil	DB00291	2346
CLL	204.1	Cyclophosphamide	DB00531	3002
CLL	204.1	Etoposide	DB00773	4179
CLL	204.1	Ifosfamide	DB01181	5657
CLL	204.1	Mechlorethamine	DB00888	6674
CLL	204.1	Pentostatin	DB00552	8011
CLL	204.1	fludarabine	DB01073	24698
CLL	204.1	pegaspargase	DB00059	34132

CLL	204.1	Immunoglobulins Intravenous	DB00028	42386
CLL	204.1	Cladribine	DB00242	44157
CLL	204.1	alemtuzumab	DB00087	117055
CLL	204.1	bendamustine	DB06769	134547
CLL	204.1	ofatumumab	DB06650	712566
CML	205.1	Busulfan	DB01008	1828
CML	205.1	Cyclophosphamide	DB00531	3002
CML	205.1	Cytarabine	DB00987	3041
CML	205.1	hydroxyurea	DB01005	5552
CML	205.1	Idarubicin	DB01177	5650
CML	205.1	Interferon Alfa-2b	DB00105	5880
CML	205.1	Plicamycin	DB06810	6995
CML	205.1	Thioguanine	DB00352	10485
CML	205.1	imatinib	DB00619	282388
CML	205.1	dasatinib	DB01254	475342
CRCA	153.9	Fluorouracil	DB00544	4492
CRCA	153.9	oxaliplatin	DB00526	32592
CRCA	153.9	irinotecan	DB00762	51499
CRCA	153.9	capecitabine	DB01101	194000
CRCA	153.9	bevacizumab	DB00112	253337
CRCA	153.9	cetuximab	DB00002	318341
HANC	195.0	Bleomycin	DB00290	1622
HANC	195.0	Cisplatin	DB00515	2555
HANC	195.0	Amifostine	DB01143	4126
HANC	195.0	Vinblastine	DB00570	11198
HANC	195.0	Carboplatin	DB00958	40048
HANC	195.0	Paclitaxel	DB01229	56946
KDNCA	189.0	Dactinomycin	DB00970	3100
KDNCA	189.0	hydroxyurea	DB01005	5552
KDNCA	189.0	Interferon Alfa-2a	DB00034	5879
KDNCA	189.0	Interferon Alfa-2b	DB00105	5880
KDNCA	189.0	Interferon gamma-1b	DB00033	5882
KDNCA	189.0	Medroxyprogesterone	DB00603	6691
KDNCA	189.0	Thalidomide	DB01041	10432
KDNCA	189.0	Vincristine	DB00541	11202
KDNCA	189.0	interferon alfacon-1	DB00069	59744
KDNCA	189.0	Aldesleukin	DB00041	70223
KDNCA	189.0	peginterferon alfa-2a	DB00008	120608
KDNCA	189.0	everolimus	DB01590	141704
KDNCA	189.0	bevacizumab	DB00112	253337
KDNCA	189.0	peginterferon alfa-2b	DB00022	253453
KDNCA	189.0	sunitinib	DB01268	357977
KDNCA	189.0	sorafenib	DB00398	495881
KDNCA	189.0	temsirolimus	DB06287	657797

KDNCA	189.0	pazopanib	DB06589	714438
LGCA	162.9	Cisplatin	DB00515	2555
LGCA	162.9	Cyclophosphamide	DB00531	3002
LGCA	162.9	Epirubicin	DB00445	3995
LGCA	162.9	Amifostine	DB01143	4126
LGCA	162.9	Etoposide	DB00773	4179
LGCA	162.9	hydroxyurea	DB01005	5552
LGCA	162.9	Ifosfamide	DB01181	5657
LGCA	162.9	Methotrexate	DB00563	6851
LGCA	162.9	Procarbazine	DB01168	8702
LGCA	162.9	Vinblastine	DB00570	11198
LGCA	162.9	gemcitabine	DB00441	12574
LGCA	162.9	vinorelbine	DB00361	39541
LGCA	162.9	Carboplatin	DB00958	40048
LGCA	162.9	irinotecan	DB00762	51499
LGCA	162.9	Paclitaxel	DB01229	56946
LGCA	162.9	Topotecan	DB01030	57308
LGCA	162.9	pemetrexed	DB00642	68446
LGCA	162.9	docetaxel	DB01248	72962
LGCA	162.9	bevacizumab	DB00112	253337
LGCA	162.9	gefitinib	DB00317	328134
LGCA	162.9	erlotinib	DB00530	337525
LGCA	162.9	crizotinib	DB08865	1148495
LIVCA	155.0	Interferon Alfa-2a	DB00034	5879
LIVCA	155.0	Interferon Alfa-2b	DB00105	5880
LIVCA	155.0	Interferon gamma-1b	DB00033	5882
LIVCA	155.0	interferon alfacon-1	DB00069	59744
LIVCA	155.0	peginterferon alfa-2a	DB00008	120608
LIVCA	155.0	peginterferon alfa-2b	DB00022	253453
LIVCA	155.0;155.2	sorafenib	DB00398	495881
OVCA	183.0	Cisplatin	DB00515	2555
OVCA	183.0	hydroxyurea	DB01005	5552
PACA	157.9	Doxorubicin	DB00997	3639
PACA	157.9	Fluorouracil	DB00544	4492
PACA	157.9	gemcitabine	DB00441	12574
PACA	157.9	erlotinib	DB00530	337525
PRCA	185	6-Aminocaproic Acid	DB00513	99
PRCA	185	Cyproterone	DB04839	3014
PRCA	185	Doxorubicin	DB00997	3639
PRCA	185	Estradiol	DB00783	4083
PRCA	185	Estramustine	DB01196	4089
PRCA	185	Estrogens Conjugated (USP)	DB00286	4099
PRCA	185	Flutamide	DB00499	4508
PRCA	185	Ketoconazole	DB01026	6135
PRCA	185	Megestrol	DB00351	6703

PRCA	185	Mitoxantrone	DB01204	7005
PRCA	185	Thalidomide	DB01041	10432
PRCA	185	nilutamide	DB00665	31805
PRCA	185	Triptorelin	DB06825	38782
PRCA	185	Leuprolide	DB00007	42375
PRCA	185	Goserelin	DB00014	50610
PRCA	185	histrelin	DB06788	50975
PRCA	185	docetaxel	DB01248	72962
PRCA	185	zoledronic acid	DB00399	77655
PRCA	185	bicalutamide	DB01128	83008
PRCA	185	Estrogens Esterified (USP)	DB09381	214549
PRCA	185	abarelix	DB00106	301739
PRCA	185	degarelix	DB06699	475230
PRCA	185	denosumab	DB06643	993449
PRCA	185	cabazitaxel	DB06772	996051
PRCA	185	sipuleucel-T	DB06688	997261
PRCA	185	abiraterone acetate	DB05812	1100071
SKCM	172.9	Bleomycin	DB00290	1622
SKCM	172.9	Carmustine	DB00262	2105
SKCM	172.9	Dacarbazine	DB00851	3098
SKCM	172.9	Dactinomycin	DB00970	3100
SKCM	172.9	hydroxyurea	DB01005	5552
SKCM	172.9	Interferon Alfa-2a	DB00034	5879
SKCM	172.9	Interferon Alfa-2b	DB00105	5880
SKCM	172.9	Interferon gamma-1b	DB00033	5882
SKCM	172.9	Lomustine	DB01206	6466
SKCM	172.9	Melphalan	DB01042	6718
SKCM	172.9	Procarbazine	DB01168	8702
SKCM	172.9	temozolomide	DB00853	37776
SKCM	172.9	interferon alfacon-1	DB00069	59744
SKCM	172.9	Aldesleukin	DB00041	70223
SKCM	172.9	docetaxel	DB01248	72962
SKCM	172.9	peginterferon alfa-2a	DB00008	120608
SKCM	172.9	peginterferon alfa-2b	DB00022	253453
SKCM	172.9	ipilimumab	DB06186	1094833
SKCM	172.9	Vemurafenib	DB08881	1147220
STCA	151.9	Mitomycin	DB00305	632
STCA	151.9	Epirubicin	DB00445	3995
STCA	151.9	Fluorouracil	DB00544	4492
STCA	151.9	capecitabine	DB01101	194000
THCA	193	Vandetanib	DB05294	1098413

Appendix 8: Table showing the drug indications identified for all the included disorders.