DOCTORAL THESIS

DECIPHERING BREAST CANCER INTERTUMOR HETEROGENEITY : TUMOR SUBTYPING AND ANALYSIS OF GERMLINE GENETIC VARIANTS

ÁNGELA SANTONJA CLIMENT

Programa de Doctorado en Biología Celular y Molecular.

Facultad de Ciencias. Universidad de Málaga

DIRECTOR	DIRECTOR	TUTOR
Emilio Alba Conejo	Aurelio A. Moya García	M. Gonzalo Claros Díaz

Málaga, Spain 2018

To my brother,
who awoke in me the
passion for research

To my parents,
who always encouraged
me to go beyond

ACKNOWLEDGMENTS

I would like to offer my deepest appreciation to my Supervisor **Emilio Alba**, for his tremendous support and constant encouragement during these last years. Thanks for the great opportunity you gave me when you decided to bet for me that first time, and all the times that followed. I would also like to express my gratitude to my Supervisor **Aurelio A. Moya**, for his insightful comments and maddening attention to details, that helped me feel backed up during my writing. I appreciate the feedback offered by my Tutor **M. Gonzalo Claros**, that improved my scientific writing skills in both languages used in this document.

I would like to express my gratitude to the **Medical Oncology Service** and the **Pathology Department** of the Hospital Virgen de la Victoria and the **patients' and relatives** that willingly donated their samples for these and many other research projects.

I would also like to thank all members from the **Human Genotyping Unit** at CNIO, for their training in genome-wide genotyping and variant association analysis as well as the members from the **Microarray Analysis Service** at IMIM, who trained me in whole-transcriptome analysis.

My grateful thanks are extended to all the members of the **Caldas Lab** (CRUK Cambridge Institute) for their warm welcoming in their group and their constructive discussions that, even though are not directly related with this dissertation, greatly contributed to my career. Particularly, I would like to thank **Suet-Feung Chin**, an outstanding supervisor, for all the skills she taught me, the illuminating discussions and her huge help.

Thanks are also due to Instituto de Salud Carlos III (ISCIII, under the grants PI13/00730 and PI11/022217, ISCIII FI12/00489 and M-AESMV16-00044), Mutua Madrileña and Sociedad Española de Oncología Médica (SEOM) for their financial support.

My deepest gratitude to **Patricia Cháves**, who encouraged me to face the challenge of a doctoral research and who got me in touch with **Luis G. Pérez**, who guided me in the first steps of my Doctoral Thesis and has always been an invaluable source of knowledge.

This Doctoral Thesis would have been impossible without the help of my lab mates, who deserve my most heartfelt gratitude. **Cristina E. Fernández, Vanessa de Luque, M. Rosario (Rous) Chica, Rocío Lavado, Cynthia Robles, Mª José Lozano, Alicia Garrido** and **Aurora Laborda**, thanks for being my teachers in lab skills, for your technical support, your encouragement, all the invaluable suggestions, constructive discussions and your moral support in and out the lab, that not only built me as a researcher but also contributed to my personal growth. I would like to single out **Martina Álvarez**, for always fighting for all of us and for her constructive comments and warm encouragements. Special thanks also to my colleagues at the Biomedical Research Lab at Hospital Universitario Virgen de la Victoria, in particular to **Alicia Céspedes, Isabel Moreno** and **Mª del Carmen Rueda** for their caring, the stimulating discussions and their support in and out of the lab.

I am particularly grateful to **Jose M. Roldán**, who edited the artwork of this Thesis as well as from the derived published papers (and yet to publish). Thanks for your invaluable help, that made this dissertation visually appealing, and thanks for your endless support

Last but not least, I lack of words to express my gratitude to my **family and friends**. The accomplishment of this milestone so far from home would not have been possible without your warm love, continued patience, and endless support.

RESUMEN (SUMMARY)

INTRODUCCIÓN

El cáncer de mama no es una única enfermedad, sino un conjunto de enfermedades que, a pesar de localizarse en el mismo órgano, presentan una gran variabilidad en todos sus componentes: características histopatológicas, respuesta al tratamiento y supervivencia [1]. Esta variabilidad se observa tanto dentro de un único tumor—heterogeneidad intratumoral—como entre tumores de distintos pacientes—heterogeneidad intertumoral—a diferentes niveles: morfológico, molecular y genómico [2]. Desde el punto de vista de la heterogeneidad intertumoral, la identificación de subgrupos de pacientes con características homogéneas nos ayudaría a mejorar el conocimiento de esta enfermedad y, por lo tanto, la atención a los pacientes. Además de la heterogeneidad observada entre tumores de distintos pacientes, la alta variabilidad existente en el microambiente del tumor (estroma o matriz extracelular) así como en las células que puedan participar en la metástasis (células del sistema inmunitario, células sanguíneas o células de tejidos diana) contribuye a aumentar la heterogeneidad observada entre pacientes con cáncer de mama [3]. Por tanto, la heterogeneidad tumoral se configura como una de las áreas con más relevancia en la investigación en el cáncer de mama, indispensable para mejorar el diagnóstico de los pacientes, identificar biomarcadores predictivos y pronósticos, y diseñar las estrategias de tratamiento.

Ante todo lo mencionado anteriormente, nuestra hipótesis es que la disección del cáncer de mama en subgrupos clínicamente relevantes y el estudio del contexto genético del paciente (conjunto de variantes genéticas heredables del paciente, también denominadas variantes genéticas germinales) pueden ser útiles en la selección del tratamiento y la supervivencia de pacientes con cáncer de mama. En este Proyecto de Tesis, hemos abordado la heterogeneidad intertumoral tanto en el propio tumor como en el contexto genético de los pacientes. Para estudiar el tumor, hemos elegido dos escenarios: el cáncer de mama triple negativo y el cáncer de mama en los

varones. En cuanto al contexto genético, hemos estudiado si existen variantes genéticas germinales en las pacientes con cáncer de mama que modulen la diseminación metastásica de la enfermedad. En las siguientes secciones describiremos cada una de estas estrategias. Las tablas (*tables)* y figuras (*figures*) a las que se hace referencia en este resumen están incluidas en el texto principal de la Tesis Doctoral y se encuentran indexadas en las páginas xxv-xxvii.

HETEROGENEIDAD TUMORAL EN EL CÁNCER DE MAMA TRIPLE NEGATIVO

En la práctica clínica, el cáncer de mama se clasifica en función de la determinación inmunohistoquímica de los receptores de estrógenos (RE) y de progesterona (PR), y de la sobreexpresión o amplificación del factor de crecimiento epidérmico humano 2 (HER2) por inmunohistoquímica o hibridación *in situ* [4]. En función de estos marcadores, el cáncer de mama se clasifica como receptor hormonal positivo o luminal (tumores que expresan el ER y/o el PR; ER+ y/o PR+), HER2-positivo (tumores con sobreexpresión o amplificación de HER2; HER2+) o triple negativo (tumores que no expresan ninguno de estos tres marcadores; ER -, PR- y HER2-). El estado de estos marcadores es importante para predecir la respuesta al tratamiento hormonal (tumores con receptores hormonales positivos) y dirigido a HER2. Debido a que aún no se han identificado dianas moleculares apropiadas, no existen tratamientos dirigidos para pacientes triple negativo; el tratamiento recomendado es la quimioterapia basada en antraciclinas y taxanos, aunque su respuesta es moderada y muy heterogénea [5, 6].

El cáncer de mama triple negativo representa del 15 al 20 % de los carcinomas de mama, y presenta una historia natural más agresiva y peor supervivencia que otros subtipos [5, 6]. Aunque la quimioterapia basada en sales de platino se ha incorporado al tratamiento de pacientes en neoadyuvancia (tratamiento antes de la intervención quirúrgica) y de pacientes metastásicas (tratamiento tras la diseminación del tumor a otros órganos), la respuesta es variable, lo que sugiere que no todas las pacientes triple negativo son sensibles a esta terapia [7].

Tras el tratamiento neoadyuvante con quimioterapia, alrededor del 30 % de las pacientes triple negativo presentan una respuesta patológica completa contra la enfermedad (ausencia de células tumorales invasivas en el tejido obtenido durante la intervención quirúrgica de la mama y ganglios axilares). Las pacientes con respuesta patológica completa tienen un pronóstico y una supervivencia similares al del resto de subtipos, pero las que presentan un tumor residual tras la quimioterapia neoadyuvante tienen peor pronóstico y supervivencia [8–10]. Por lo tanto, la respuesta patológica completa a la quimioterapia neoadyuvante es un marcador subrogado de la supervivencia en pacientes con cáncer de mama triple negativo.

Se han publicado varias clasificaciones del cáncer de mama triple negativo con el objetivo de descifrar la heterogeneidad observada entre pacientes, aunque todavía no se ha demostrado que estas clasificaciones sean predictivas de la eficacia del tratamiento [11–13]. Es por tanto necesario validar la relevancia de los subtipos del cáncer de mama triple negativo en cuanto a sus características moleculares y respuesta al tratamiento. En este estudio, hemos evaluado la relevancia clínica de los subtipos de Lehmann mediante el análisis de la respuesta patológica completa al tratamiento neoadyuvante como un subrogado de la supervivencia. También hemos evaluado si alguno de estos subtipos es sensible a las sales de platino combinadas con quimioterapia, ya que este tipo de agentes han demostrado ser eficaces contra los tumores con deficiencias en el sistema de reparación del DNA, como sucede en el caso de los subtipos basales de Lehmann.

Materiales y métodos

Realizamos un análisis retrospectivo de 125 carcinomas de mama invasivo triple negativo de distinta procedencia: 45 (36 %) muestras de una recolección prospectiva del ensayo clínico aleatorizado de fase II GEICAM/2006-03 [14] y 80 (64 %) muestras recogidas de forma retrospectiva en cuatro hospitales colaboradores. En este estudio se incluyeron pacientes con confirmación histológica de carcinoma invasivo de mama tipo triple negativo (ER-, PR- y HER2-) que hubieran recibido antraciclinas y/o taxanos con o sin carboplatino como quimioterapia neoadyuvante.

Todos los análisis se realizaron en bloques de tejido tumoral incluidos en parafina de biopsias diagnósticas obtenidas antes del tratamiento neoadyuvante.

Al analizar el cáncer de mama mediante sus perfiles de expresión génica se ha observado que el cáncer de mama triple negativo está compuesto por todos los subtipos intrínsecos (luminal A, luminal B, HER2-enriquecido o de tipo basal), si bien el subtipo basal es el más común (~70 %) (Figura 2) [15]. Para explorar la distribución de las muestras en subtipos intrínsecos, determinamos su patrón de expresión génica con el equipo nCounter Analysis System (Nanostring): las muestras del ensayo GEICAM/2006-03 se clasificaron con un algoritmo de predicción de subtipos basado en la firma PAM50 [16] y las de la colección retrospectiva con el ensayo de Prosigna [17], que incluye un algoritmo basado en la firma PAM50.

De forma paralela, clasificamos las muestras en los subtipos de Lehmann (basal-like 1, BL1; basal-like 2, BL2; immunomodulatory, IM; mesenchymal, M; mesenchymal stem-like, MSL; luminal androgen receptor, LAR; y un subtipo inestable o unstable, UNS) mediante un análisis del transcriptoma completo con la plataforma Affymetrix GeneChip® (Affymetrix) y la herramienta online TNBCtype [18]. Tras su clasificación, realizamos un análisis estadístico univariante y multivariante para explorar las características clinicopatológicas de los subtipos de Lehmann y su asociación con la respuesta patológica completa a los distintos tratamientos recibidos. Debido a que Prat y colaboradores han sugerido que el subtipo MSL suele contener muestras ricas en tejido normal [19], realizamos todos los análisis estadísticos con y sin el subtipo MSL, en caso de que su inclusión nos llevara a perder resultados relevantes en los estudios de asociación.

Resultados

En el estudio de asociación de las características clinicopatológicas con los subtipos de Lehmann observamos que la tasa de expresión de Ki-67 fue la única variable asociada a los subtipos: BL1 presentaba la tasa más alta de proliferación (88.2 % de BL1 frente al 63.7 % del resto de pacientes con Ki-67 > 50 %; p = 0,02) y LAR la más baja (71 % de LAR frente al 27 % del resto de pacientes con Ki-67 ≤ 50 %; p = 0,002). También observamos una asociación estadísticamente

significativa entre la clasificación en subtipos de Lehmann y en subtipos intrínsecos, principalmente porque LAR es el único subtipo (exceptuando el subtipo inestable o UNS) que incluye muestras de tipo no basal (luminal A, luminal B o HER2-enriquecido; véase la Figura 9). Comprobamos que las muestras LAR sobreexpresaban el receptor de andrógenos (AR) y que eran histopatológicamente compatibles con carcinoma endocrino.

Al estudiar la asociación entre variables clinicopatológicas y respuesta patológica completa, sólo la tasa de expresión de Ki-67 y el tamaño tumoral se asociaron a respuesta en el análisis bivariante, y sólo el tamaño tumoral en el análisis multivariante. Considerando todos los tratamientos, observamos un amplio margen de respuestas en los distintos subtipos de Lehmann (desde 47,1 % de BL1 a 14,3 % de LAR con respuesta patológica completa; véase la Tabla 3). Las pacientes LAR fueron las más quimiorresistentes (14,3 % de LAR tuvieron respuesta completa frente a 41,9 % del resto de subtipos combinados; $p = 0,077$), siendo más acusada esta diferencia en la respuesta al excluir las pacientes MSL (14,3 % de LAR tuvieron respuesta completa frente a 42,7 % del resto de subtipos combinados excepto MSL; $p = 0,046$).

Finalmente, estudiamos la respuesta a los distintos tratamientos recibidos (Figura 10). En nuestra cohorte, no hubo diferencias en la respuesta patológica completa entre el grupo de pacientes tratadas con y sin carboplatino. En el grupo de pacientes tratadas con antraciclinas y taxanos, tampoco observamos diferencias en respuesta en los distintos subtipos de Lehmann. Sin embargo, dentro del grupo de pacientes tratadas con quimioterapia y carboplatino, las pacientes clasificadas como BL1 fueron las que más se beneficiaron del tratamiento (80 % BL1 con respuesta patológica completa frente a 23 % del resto de subtipos combinados; $p = 0,027$).

Discusión

Nuestros resultados sugieren que, entre los subtipos de Lehmann del cáncer de mama triple negativo, el subtipo LAR es el menos proliferativo y el más quimiorresistente. A pesar de su menor respuesta a la quimioterapia neoadyuvante, el subtipo LAR se asocia generalmente a buen pronóstico [20, 21]. Esto puede ser debido, al menos en parte, a que el subtipo LAR es el único que incluye tumores de

tipo no basal según PAM50 [22, 23], lo que explicaría la baja proliferación y la baja tasa de respuesta patológica completa observadas en nuestro estudio. La mayoría de estas muestras sobreexpresaban el receptor de andrógenos y eran consistentes histológicamente con carcinoma apocrino [24]. Las publicaciones recientes de ensayos clínicos de fase II sugieren que estas pacientes (con AR+) pueden beneficiarse de tratamientos con antiandrógenos [25–27].

En nuestra cohorte, el subtipo BL1 es el más proliferativo y parece ser particularmente sensible a regímenes de quimioterapia que incluyen sales de platino. Recientemente ha habido mucho interés en el uso sales de platino para tratar pacientes con cáncer de mama triple negativo, ya que los tumores con deficiencias en la recombinación homóloga son sensibles a estos agentes que inducen muerte celular. A pesar de que los estudios de fase II que incluyen pacientes triple negativo no seleccionadas en el contexto neoadyuvante han sido contradictorios [28, 29], los tumores triple negativo con alta tasa de deficiencia de recombinación homóloga parecen beneficiarse de los tratamientos con sales de platino [30].

Nuestros resultados confirman la gran diversidad molecular y de respuesta al tratamiento de los tumores triple negativo. Sin embargo, estos tumores no parecen pertenecer a categorías aisladas, sino que deberían considerarse como un espectro de tumores con variaciones en sus características clínicas y moleculares. En un extremo de este espectro estarían los tumores BL1, muy proliferativos y posiblemente sensibles a las sales de platino. En el otro extremo estarían los LAR, poco proliferativos y resistentes a la quimioterapia neoadyuvante. Entre estos dos extremos encontraríamos una gran diversidad de tumores, que no podrían clasificarse en un grupo con características únicas. La identificación de estas pacientes extremas podría llevar a una mejora en su evaluación clínica y en la elección del tratamiento respecto a cuando se encuentran diluidas en un grupo tan amplio y heterogéneo como el triple negativo.

HETEROGENEIDAD TUMORAL EN EL CÁNCER DE MAMA EN LOS VARONES

El cáncer de mama en los varones es una enfermedad rara que no se conoce en profundidad: representa menos del 1 % de los tumores de mama y menos del 1 % de los cánceres diagnosticados en los hombres [31]. Por ello, su conocimiento biológico y tratamiento médico se basa principalmente en hallazgos del cáncer de mama en las mujeres o de estudios con un número pequeño de pacientes.

La clasificación inmunohistoquímica se usa de forma sistemática en las mujeres con cáncer de mama para determinar su tratamiento. Sin embargo, la mayoría de los tumores de mama en los varones son ER+, por lo que la clasificación inmunohistoquímica no es suficiente para identificar subtipos clínicamente relevantes: no refleja la heterogeneidad del cáncer de mama en los varones.

A pesar de las diferencias biológicas entre el cáncer de mama de hombres y mujeres, actualmente no hay datos sobre los subtipos de cáncer de mama en los varones basados en firmas pronósticas, incluida la firma PAM50 [16]. Por lo tanto, es necesario identificar subtipos moleculares en el cáncer de mama en el varón que reflejen su biología, predigan su evolución clínica y guíen las decisiones de tratamiento. El objetivo de este estudio es comparar las características clinicopatológicas y la supervivencia de los hombres con cáncer de mama clasificados según los subtipos intrínsecos (basados en la firma PAM50) frente a un panel de subrogados inmunohistoquímicos.

Materiales y métodos

En este estudio se incluyeron 67 varones con carcinoma de mama; todos los análisis se realizaron en bloques de tejido tumoral incluido en parafina de muestras obtenidas en la intervención quirúrgica. En base a un panel inmunohistoquímico con seis marcadores, las muestras se clasificaron como luminal A (ER+ y/o PR+, HER2-, Ki-67 < 14 %), luminal B (ER+ y/o PR+, HER2-, Ki-67 ≥ 14 %), HER2-positivo (HER2+, independientemente del estado de los ER y PR), basal (ER-, PR-, HER2-, EGFR+ y/o CK5/6+) y triple negativo no basal (ER-, PR-,

HER2-, EGFR-, CK5/6-). Las tinciones y la valoración inmunohistoquímica se realizaron por triplicado en micromatrices de tejido.

Clasificamos las muestras en subtipos intrínsecos (luminal A, luminal B, HER2-enriquecido o de tipo basal) mediante un análisis de expresión génica con un equipo nCounter Analysis System usando el kit PAM50 y el algoritmo de Prosigna. Una vez clasificadas por ambos métodos, analizamos las características clínicopatológicas de cada subtipo, así como la supervivencia libre de progresión y la supervivencia global, comparando ambos métodos de clasificación.

Resultados

Las características clínicopatológicas de los pacientes se presentan en la Tabla 4. La mayoría de los pacientes tenían tumores ER+ (96 %) y/o PR+ (84 %). Sólo un paciente tenía un tumor HER2+ y recibió tratamiento anti-HER2 (trastuzumab).

Tanto por inmunohistoquímica como por la firma PAM50, la mayoría de las muestras se clasificaron como luminal B (51 % y 60 %, respectivamente) seguidas por luminal A (44 % y 30 %, respectivamente). Ninguna muestra se clasificó como de tipo basal por PAM50, y sólo tres fueron triple negativo (basal y no basal) por inmunohistoquímica.

Encontramos una fuerte correlación al comparar los subtipos determinados por inmunohistoquímica y por PAM50 ($p = 0{,}018$), a pesar de que más de la mitad de las muestras clasificadas como luminal A por inmunohistoquímica y alrededor del 20 % de las luminal B se clasificaron en otro grupo por PAM50 (Figura 12). Tan sólo uno de los pacientes HER2-enriquecido era HER2-positivo por inmunohistoquímica y CISH (hibridación cromogénica *in situ*).

Al analizar las características clinicopatológicas entre los distintos subtipos, no encontramos diferencias significativas entre las características de los pacientes luminal A y luminal B por PAM50 (Tabla 5). Tampoco encontramos diferencias significativas en la supervivencia libre de progresión ni en la supervivencia global entre los pacientes luminal A y luminal B clasificados por inmunohistoquímica ni por PAM50. Observamos, sin embargo, una peor supervivencia global en los pacientes con

tumores HER2-enriquecido (Figura 13) al compararlos con el resto de tumores (mediana de 71 meses frente a 128 meses, respectivamente; OR = 2,59 [0,47-14,29]; p = 0,046).

Discusión

Las características clinicopatológicas observadas en esta cohorte de varones con cáncer de mama difieren de las que se esperarían en una población de mujeres: tienen una mayor tasa de positividad de ER y PR, menor positividad de HER2, edad de presentación más tardía y mayor proporción de ganglios afectados. Estos resultados son comparables a los de estudios previos [32]. Varios grupos de investigación han intentado clasificar el cáncer de mama en varones en base a subrogados inmunohistoquímicos. En línea con nuestros resultados, todos obtuvieron un menor número de tumores HER2-positivo y triple negativo que el esperado en las mujeres, que se asociaron con mal pronóstico. En cuanto a los tumores luminales, es complejo realizar una comparativa entre estudios debido a que no hay un consenso en la diferenciación entre luminal A y luminal B. Esta diferenciación puede estar basada en la sobreexpresión o amplificación de HER2 (luminal A cuando HER2 es negativo y luminal B cuando HER2 es positivo) o en la tasa de expresión de Ki-67 (luminal A cuando la tasa de expresión de Ki-67 es baja y luminal B cuando es alta). La clasificación basada en Ki-67 es la recomendada en el último panel de expertos de St. Gallen [33], aunque debido a la falta de reproducibilidad de la determinación de Ki-67 y a que no se ha definido un punto de corte óptimo, esta clasificación está en entredicho. Por lo tanto, el uso de marcadores inmunohistoquímicos puede llevar a una clasificación errónea de los tumores comparada con la información generada por métodos basados en los patrones de expresión génica, como es el caso del clasificador PAM50.

Este es el primer estudio que clasifica el cáncer de mama en el varón en subtipos intrínsecos basados en la firma PAM50. Nuestros resultados confirman que, también a nivel genómico, es un cáncer principalmente de tipo luminal. La mayoría de las muestras se clasificaron como luminal B, seguidas por luminal A. Los tumores luminal B son más agresivos, más proliferativos y con peor respuesta a los tratamientos

endocrinos que los luminal A. Más de la mitad de los pacientes luminal A definidos por inmunohistoquímica se clasificaron en otro grupo por PAM50, pero no sabemos si esto puedo tener un impacto en su tratamiento. En cualquier caso, no encontramos diferencias significativas en las características clinicopatológicas ni en la supervivencia entre los tumores luminal A y luminal B. Muchos de estos tumores se diagnosticaron en pacientes de edad avanzada, algunos de los cuales fallecieron por causas no relacionadas con el tumor, y las recaídas pueden ocurrir incluso mucho tiempo después el diagnóstico, lo que podría explicar al menos en parte por qué no encontramos diferencias de supervivencia entre los luminal A y los luminal B.

Nuestros resultados sugieren que el subtipo HER2-enriquecido es el de peor pronóstico. La mayoría de estos pacientes eran HER2- por inmunohistoquímica, por lo que no recibieron ningún tratamiento anti-HER2. La incorporación de los tratamientos anti-HER2 ha cambiado la historia natural de los tumores de mama HER2 positivo en mujeres, pasando de un subtipo históricamente agresivo a un subtipo con supervivencia similar a los subtipos negativos para HER2 [34]. Hay estudios que sugieren que los tumores HER2-enriquecido en mujeres tienen mayor porcentaje de respuestas patológicas completas que otros subtipos y que pueden beneficiarse de terapias anti-HER2 [35, 36].

HETEROGENEIDAD EN EL CONTEXTO GENÉTICO DE PACIENTES CON CÁNCER DE MAMA

La mayoría de las muertes asociadas al cáncer están causadas por la diseminación metastásica de la enfermedad, un proceso extraordinariamente complejo que aún no se conoce en profundidad [3]. Diversos estudios sugieren que la metástasis es un proceso en el que intervienen no sólo las células del tumor, sino también del microambiente del tumor y de los tejidos diana de la metástasis. De este modo, el contexto genético del paciente podría influir en el desarrollo de metástasis, sólo o en combinación con las características del tumor [3, 37].

Sin embargo, no se ha conseguido identificar con claridad las variantes genéticas germinales asociadas a la metástasis, principalmente por la falta de potencia estadística

de los estudios tradicionales de asociación de genoma completo (GWAS) para detectar variantes relevantes para los fenotipos complejos como la enfermedad metastásica [38, 39]. Para superar estos problemas, hemos reducido la heterogeneidad fenotípica mediante la selección de pacientes en los extremos de la distribución del riesgo a metástasis, pero con un comportamiento inverso al esperado; es decir, con fenotipos extremos discordantes (Figura 14) [40, 41]. Estos fenotipos extremos discordantes son más útiles que los individuos de la población general a la hora de buscar los rasgos genéticos asociados a las metástasis, ya que deben estar enriquecidos en dichos rasgos debido a sus características únicas, lo que aumenta la potencia estadística y requiere menos muestras que en un estudio tradicional de asociación.

El objetivo de este estudio es identificar las variantes genéticas germinales asociadas a la metástasis, ya sea variantes con un fuerte efecto individual (suponiendo que el efecto de cada variante en el fenotipo es independiente de otras variantes) [42] o conjuntos de variantes con efecto moderado e interacciones sinérgicas.

Materiales y métodos

En este estudio incluimos 97 pacientes con cáncer de mama, con seguimiento superior a cinco años y fenotipos extremos discordantes procedentes de ocho hospitales. Estas pacientes se clasificaron en dos grupos: casos de buen pronóstico o casos de mal pronóstico. Los casos de buen pronóstico incluyen pacientes con tumores de bajo riesgo (tumores $\leq$ 2 cm y sin ganglios afectados) que, sin embargo, recayeron en los primeros cinco años tras la intervención quirúrgica. Los casos de mal pronóstico incluyen pacientes con alto riesgo de desarrollar metástasis (más de 10 ganglios afectados independientemente del tamaño del tumor) que no recayeron en ese periodo.

Genotipamos el genoma completo de las pacientes con muestras de sangre periférica con el HumanOmin5-Quad Beadchip de Illumina y preprocesamos los datos con PLINK [43]. Después, buscamos variantes con un efecto individual fuerte en metástasis mediante un análisis de asociación entre variantes genéticas en los casos de buen y mal pronóstico también con PLINK. Realizamos un análisis de redes de epistasia con Encore [44] para detectar variantes y genes con efecto sinérgico en

metástasis, es decir, cuya interacción tiene un efecto acumulativo. Consideramos que estos genes pueden ser relevantes para la metástasis por su interacción con muchos otros genes relevantes y por lo tanto los hemos denominado "genes que influyen en la metástasis" o GIM. Finalmente, identificamos cuáles de nuestros GIM eran reguladores y el conjunto de genes bajo su regulación (sus regulones) y realizamos un análisis de supervivencia en una base de datos pública de cáncer de mama [45]. En este análisis evaluamos la asociación del patrón de expresión génica de cada factor de transcripción y sus regulones con la supervivencia libre de metástasis a distancia como una forma de determinar su impacto en la metástasis

Resultados

No encontramos ninguna variante genética germinal con un efecto individual fuerte en la metástasis que alcanzara la significación estadística requerida en estudios de GWAS ($p < 5 \times 10^{-8}$). Sin embargo, 13 variantes alcanzaron valores de p en el orden de 10^{-6} (Tabla 7). La mayoría de estas variantes están en o cerca de genes involucrados en procesos relacionados con la metástasis o el desarrollo tumoral [46–48], tales como *COPZ1* [49], *NGEF* [50] y *MUC16* [51, 52]. A pesar de no alcanzar la significación estadística, el análisis funcional de estos genes combinado con los altos valores de la razón de probabilidades (OR, *odds ratio*) obtenidos en el análisis de asociación indican que no podemos descartar que estas variantes puedan tener algún efecto individual en la susceptibilidad a la metástasis.

Para identificar el conjunto de variantes con efectos sinérgicos en la metástasis, modelamos una red de epistasia genética que codifica la susceptibilidad a metástasis en nuestra cohorte e identificamos los genes que reflejan la mayoría de estas interacciones—están mejor asociados a la susceptibilidad a la metástasis—mediante su centralidad en la red (*community centrality measure*) [53]. Nuestra hipótesis es que estos genes (GIM) afectan directamente a la metástasis mediante la influencia en muchos otros genes—que también pueden estar alterados por variantes genéticas germinales—codificados en la red de epistasia [54]. Esta red contiene 1 431 genes y aproximadamente 5 600 conexiones entre ellos. Es una red grande y densa (hay muchas conexiones entre los genes), lo que sugiere que la contribución germinal a un

rasgo complejo como es la metástasis es poligénica. Además, en esta red la mayoría de genes pueden alcanzarse desde cualquier otro gen por unos pocos pasos y los genes están fuertemente interconectados formando comunidades. Los genes más relevantes de esta red son los GIM (Tabla 8), ya que interconectan diferentes comunidades, integrándolas en la topología de la red.

El 10 % de los GIM con mayor centralidad están sobrerrepresentados en rutas de KEGG, tales como la interacción con receptores de la matriz extracelular (KEGG: map04512) y el establecimiento de puntos de contacto entre las células y la matriz extracelular (KEGG: map04510), lo que implica que algunas variantes genéticas germinales de pacientes con cáncer de mama afectan a genes involucrados en el mecanismo de la metástasis.

Para analizar el impacto en la metástasis de los GIM hemos utilizado la supervivencia libre de metástasis a distancia (definida como tiempo desde el diagnóstico hasta la aparición de un foco metastásico o muerte asociada al tumor), basada en datos de expresión génica y supervivencia de una base de datos pública de cáncer de mama [45] (Tabla 8). Observamos que los perfiles de expresión de la mayoría de nuestros GIM están alterados en células tumorales y que su expresión está significativamente asociada a la supervivencia libre de metástasis a distancia. Como los análisis de supervivencia se realizaron en células tumorales, podemos inferir que aquellos GIM asociados a la supervivencia están alterados en el tumor, mientras que el resto están alterados en el microambiente o en los tejidos diana (son relevantes para la metástasis debido a su centralidad en la red, pero no se encuentran en las células tumorales). Como resultado, hemos identificado variantes genéticas germinales que afectan a genes expresados en el tumor y que promueven la migración de las células tumorales (*TNP1* y *MAP2K4*) o que están expresados en el microambiente del tumor o de los tejidos diana y favorecen el crecimiento de los focos metastásicos (*FN1*, *COMP*).

Las variantes genéticas germinales suelen asociarse con fenotipos complejos como la metástasis mediante la regulación de la expresión génica [55, 56]. Para evaluar si los GIM regulan la expresión de genes tumorales, obtuvimos las interacciones de

regulación entre los GIM y los genes bajo su regulación de datos de regulación específica de cáncer de mama [57]. Como esta base de datos contiene información de líneas celulares de cáncer de mama, nos permite identificar sólo los genes que se expresan en el tumor, pero no los que se expresan en el microambiente tumoral o en los tejidos diana. Entre nuestros GIM, 20 son reguladores, un número más alto que el esperado por azar (test binomial; $p < 0{,}001$), y regulan la expresión de alrededor de 800 genes. Encontramos una asociación estadísticamente significativa entre la supervivencia libre de metástasis a distancia y la expresión del conjunto de genes bajo la regulación de varios GIM. También observamos que varios de los reguladores identificados entre nuestros GIM eran parálogos a factores de transcripción descritos en el cáncer de mama [58, 59], lo que apoya la teoría de que las variantes genéticas germinales en estos genes producen una desregulación en los patrones de expresión en el tejido tumoral que dan lugar a un aumento de la tasa de metástasis.

Discusión

El genotipado de genoma completo realizado en nuestra cohorte de fenotipos extremos discordantes no identifica ninguna variante genética germinal con un efecto individual estadísticamente significativo en la metástasis. Estos resultados están en línea con los de estudios previos [38, 60, 61]. Sin embargo, observamos que algunas de las variantes con los valores de p más bajos tenían altos valores de OR, y que además se encontraban en genes (o cerca de genes) con funciones relacionadas con la metástasis o el desarrollo tumoral, lo que sugiere que no podemos descartar que exista una relación entre dichas variantes y la metástasis, enmascaradas por el bajo número de muestras analizadas.

Mediante el análisis de redes identificamos diversos genes alterados por variantes genéticas germinales que influyen en la metástasis a través de su interacción sinérgica con múltiples genes de la red de epistasia, cada uno con un pequeño efecto acumulativo en la metástasis. Estos genes, denominados "genes que influyen en la metástasis" o GIM, se encuentran tanto en el tumor como en genes que se expresan en su microambiente o en los tejidos diana: las alteraciones genéticas germinales en

estos genes favorecen la diseminación de células tumorales o bien facilitan que el tejido diana sea el adecuado para que se desarrollen los focos metastásicos.

También observamos que algunos GIM localizados en el tumor afectan a la metástasis mediante su expresión y patrones de regulación: su expresión desregulada en el tumor promueve la metástasis y/o altera la expresión de un conjunto de genes tumorales que conducen al desarrollo de metástasis. Debido a que la función de los GIM está alterada por variantes del contexto genético de la paciente (que existe antes del desarrollo del carcinoma de mama) nuestros resultados sugieren que el potencial metastásico de una célula tumoral de mama no depende exclusivamente de la acumulación de mutaciones en la misma, sino también de la desregulación preprogramada de la expresión génica en el contexto genético de la paciente. De este modo, las mujeres que tengan ciertas variantes germinales en su contexto genético desplegarán patrones de expresión que favorecerán (o protegerán de) la metástasis en el caso de que desarrollen un carcinoma de mama.

Los resultados de este estudio sugieren que la gran heterogeneidad del cáncer de mama no se debe exclusivamente a las características propias del tumor, sino también al contexto genético de las pacientes. De este modo, las pacientes con cáncer de mama se podrían beneficiar de una mejor atención médica y un tratamiento más personalizado cuando se diagnostican mediante una aproximación en la que se consideren tanto las características del tumor como las del contexto genético.

CONCLUSIONES

En este trabajo de investigación hemos observado que la disección del cáncer de mama puede desvelar subtipos biológica y clínicamente relevantes, y que la combinación del análisis de rasgos del tumor combinado con el estudio de las características del contexto genético de los pacientes puede mejor el pronóstico del cáncer de mama y guiar las decisiones de tratamiento. En el estudio de la heterogeneidad tumoral hemos explorado dos escenarios: el cáncer de mama triple negativo y el cáncer de mama en el varón.

Las pacientes con cáncer de mama triple negativo parecen distribuirse en un espectro de tumores en cuyos extremos encontraríamos pacientes altamente proliferativos y sensibles a las sales de platino y pacientes con poca proliferación y que parecen resistentes a la quimioterapia estándar. Nuestros resultados apuntan a que el subtipaje de Lehmann puede ser útil en la identificación de estos grupos de pacientes extremos, que podrían beneficiarse de un tratamiento más personalizado. En el caso del cáncer de mama en los varones, hemos observado que la firma PAM50 identifica un grupo de pacientes HER2-enriquecido con mal pronóstico, que no se detecta mediante los métodos tradicionales basados en subrogados inmunohistoquímicos, que podrían beneficiarse de terapias anti-HER2. De este modo, tanto en el cáncer de mama triple negativo como en el cáncer de mama en los varones, el subtipaje molecular confirma su gran diversidad molecular y de respuesta al tratamiento. En ambos casos, este subtipaje podría llevar a la elección de un tratamiento más personalizado y a una mejora en la evolución clínica de un subconjunto de pacientes.

La alta heterogeneidad que presentan los pacientes con cáncer de mama no puede explicarse completamente mediante las características del tumor ni la elección del tratamiento: la variabilidad en las características del contexto genético de las pacientes podría explicar esta heterogeneidad. Nuestros resultados sugieren que existen variantes en el contexto genético de las pacientes con cáncer de mama que modulan la aparición de metástasis. Los genes afectados por estas variantes parecen promover la dispersión de las células tumorales y facilitar que el tejido diana de la metástasis sea el apropiado para que se desarrollen nuevos focos tumorales. Por tanto, la atención a los pacientes con cáncer de mama y la selección del tratamiento deberían basarse no sólo en la información derivada de las características clínicas y moleculares del tumor, sino también información sobre el contexto genético del paciente. Estos métodos convergen a la hora de promover la dispersión de las células tumorales, así como facilitar que tejido diana sea apropiado para que se desarrollen los focos tumorales.

En resumen, la identificación y el uso de subtipos clínicamente relevantes de cáncer de mama, así como la determinación de las características del contexto genético de los pacientes, podría facilitar la selección del tratamiento y mejorar su evolución clínica.

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
BL	Basal-like
CK	Cytokeratin
CISH	Chromogenic *In Situ* Hybridization
CNIO	Centro Nacional de Investigaciones Oncológicas *(National Centre of Oncological Research)*
EGFR	Epidermal Growth Factor Receptor
ER	Estrogen Receptor
FFPE	Formalin-Fixed Paraffin-Embedded
FISH	Fluorescence *In Situ* Hybridization
GWAS	Genome-Wide Association Study
GEICAM	Grupo Español de Investigación en Cáncer de Mama *(Breast Cancer Research Spanish Group)*
GIM	Genes que Influyen en la Metástasis *(Metastasis Influence Genes)*
HER2	Human Epidermal Growth Factor Receptor 2
IACR	International Agency of Research on Cancer
IHC	Immunohistochemical
IM	Immunomodulated
IMIM	Institut Hospital del Mar d' Investigacions Mèdiques *(Hospital del Mar Medical Research Institute)*
IntClust 1-10	Integrative Clusters
LAR	Luminal androgen receptor
M	Mesenchymal
MAF	Minimum Allele Frequency

METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
MSL	Mesenchymal stem-like
OR	Odds Ratio
pCR	Pathological Complete Response
PBS	Phosphate-buffered saline
PR	Progesterone Receptor
REDECAN	Spanish Network of Cancer Registries *(Red Española de Registros de Cáncer)*
RMA	Robust Multichip Analysis
SAM	Servei d'Anàlisi de Microarrays *(Microarray Analysis Service)*
SNP	Single Nucleotide Polymorphism
TCGA	The Cancer Genome Atlas
TNBC	Triple negative breast cancer
UNS	Unstable
WHO	World Health Organization

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. GENERAL BACKGROUND

Science is simply the word we use to describe
a method of organizing our curiosity

Tim Minchin (1975 -)

1.1 EPIDEMIOLOGY OF FEMALE BREAST CANCER

Breast cancer is the second most commonly diagnosed cancer overall and by far the most common cancer in women worldwide. The most recent data about incidence and mortality worldwide were published in 2012 by GLOBOCAN [62], a project to provide data on estimates of incidence, mortality and prevalence from 184 countries by cancer site (excluding skin non-melanoma) and sex. In 2012, breast cancer represented a 25 % (1.7 million new cases) of all cancers diagnosed in women in the world and a 30 % in Europe (361 608 new cases). The incidence of breast cancer is expected to raise in the next years, estimating an increase of 17 % (422 005 breast cancers) of new cases diagnosed in Europe by the year 2035. Breast cancer has the highest mortality of all cancers in women and is the 8th cause of death overall [63]. In Spain, 27 747 new breast cancers were diagnosed in 2015 (28 % of all cancers diagnosed in women) [64]. Breast cancer is a major health burden also in Spain; it is the cancer that harvests more lives (6 212 deaths in 2015) and the second cause of death overall in women.

1.2 BREAST CANCER HETEROGENEITY

Breast cancer is not a single disease but a collection of neoplasms that originate in the same organ and yet, widely differ in almost all its components: histopathological and molecular features, clinical outcome, and treatment response [1]. Heterogeneity exists within a given tumor—intratumor heterogeneity—and between different patients—intertumor heterogeneity—in different layers: morphologic, protein expression, and genomic [2]. The high interpatient variability of microenvironmental components of the tumor such as stromal cells or the extracellular matrix, together with the variability in other cells of the patients that may participate in the metastatic dissemination of the disease such as cells of the immune system, blood cells or cells from target tissues, add an extra level of complexity to breast cancer heterogeneity [3]. This way, **heterogeneity is one of the most relevant areas of breast cancer research**, paramount to improve diagnosis, to identify prognostic and predictive biomarkers, and to design therapeutic strategies.

As this Thesis is focused on intertumor heterogeneity, the most relevant breast cancer classifications to date are summarized below, from the more traditional histopathological classification to the recent integrative analyses. The dissection of breast cancer into a number of more homogeneous entities allows a better prediction of short and long-term outcome as well as the selection of targeted therapies [65]. Thus, **subtyping can improve the routine management of individual patients**, getting us a step closer to personalized breast cancer treatment.

1.2.1 Morphologic heterogeneity: histopathological classification

The histopathological classification of breast cancer includes the histological grade and the histological type. The **histological grade** evaluates in a scale from 1 to 3 how similar is a tumor to normal epithelium: a low-grade tumor is more similar to normal epithelium than a high-grade tumor [66]. Based on their morphological and cytological pattern, breast cancers can be classified in 17 **histological types** according the World Health Organization (WHO) [67]. The vast majority of breast cancers are classified as invasive carcinoma of no special type (ductal not otherwise specified, > 70 %) or lobular invasive carcinoma (~ 10 %), grouping together tumors that have diverse biological and clinical profiles [1]. Therefore, the **histological classification is unable to reflect the wide heterogeneity observed in breast cancer**, even though it remains an essential component of pathological reports.

1.2.2 Protein expression heterogeneity: immunohistochemical subtypes

Breast cancer subtyping to determine prognosis and treatment options is routinely determined by the **immunohistochemical expression** of estrogen receptor (ER) and progesterone receptor (PR), and overexpression or amplification of human epidermal growth factor 2 (HER2) [4]. These biomarkers led to the classification of breast cancer in three subtypes: 1) hormone-receptor positive or luminal tumors express the ER and/or the PR (ER+ and/or PR+), 2) HER2-positive tumors have overexpression or amplification of HER2 (HER2+) and 3) triple negative tumors lack the expression of these three biomarkers (ER-, PR-, HER2-). An example of immunostained sections of these subtypes is shown in Figure 1.

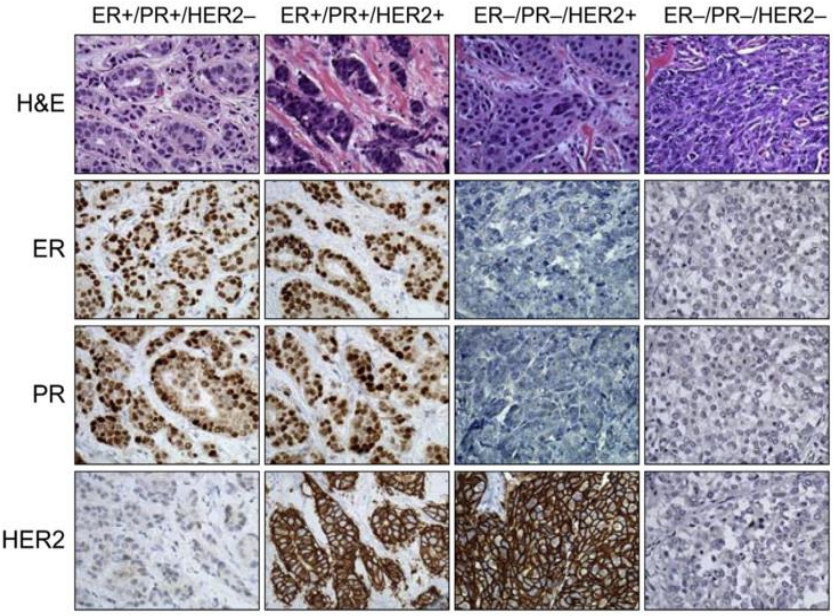

Figure 1. Immunohistochemical breast cancer subtypes. Cancer histology is depicted using hematoxilin and eosin staining (magnification 40x). *Image from Rivenbark et al. 2013 [1]*

These subtypes have improved breast cancer outcome by the administration of targeted therapies in the hormone-receptor positive (hormonal therapy) and HER2-positive (anti-HER2 therapy) subtypes. **Although the clinical use of these biomarkers has proven useful, they cannot predict responses to emerging targeted therapies neither provide sufficient prognostic accuracy to aid in chemotherapy treatment decisions**. Therefore, more predictive biomarker panels are needed.

1.2.3 Gene-expression heterogeneity: intrinsic subtypes

Recent advances in gene-expression analysis have led to the development of several breast cancer classifications based on comprehensive gene-expression profiling. The pioneer studies were conducted by Perou, Sørlie and colleagues [68, 69], who classified breast cancer into five categories with prognostic value that were predictive of treatment response: luminal A, luminal B, HER2-enriched, normal-like and basal-like. These subtypes are known as "intrinsic subtypes" as they are defined by a set of "intrinsic" genes with greater variation in expression between subtypes than between paired samples form the same tumor (their expression patterns are characteristic of each tumor and not due to tissue sampling).

Luminal A (50 to 60 % of all breast tumors in females) is usually positive for hormone receptors (ER+ and/or PR+), has low proliferation and the lowest relapse rate (the best prognosis) [69]. **Luminal B** (15 to 20 % of all breast cancers) is positive for hormone receptors and highly proliferative, with higher grade, worse prognosis and lower response to endocrine treatment than luminal A. **HER2-enriched** breast cancer (15 to 20 % of all breast tumors) is defined by the overexpression of HER2 (or other genes in amplicon 17q12) and morphologically is characterized by high proliferation, high histological grade and high nuclear grade. **Normal-like** samples express many genes characteristic of the adipose tissue and other nonepithelial cell types as well as basal epithelial genes, and have low expression of luminal epithelial cells. This group resembles normal breast tissue and is thought to be highly contaminated with stroma (non-tumor tissue) so is usually not included in predictive or prognosis studies [16, 69]. **Basal-like** (15 to 20 % of all cancers) is the most heterogeneous subtype and is mainly composed of tumors that lack the expression of hormone receptors and HER2 overexpression/amplification (i.e. triple negative breast tumors in the immunohistochemical-based classification). Basal-like tumors express high levels of basal cytokeratins, and present a more aggressive natural history and worse disease-specific survival than other breast cancer subtypes [65, 70]. Subsequent studies that focused on basal-like and/or triple-negative breast cancer revealed additional subtypes within his group [11–13]. Due to its poor prognosis, high intertumor heterogeneity—a wide clinical outcome and response to treatment is observed between different patients—and lack of targeted therapies, we consider that basal-like is of paramount interest in deciphering breast cancer heterogeneity for improving the disease management and its treatment. A more detailed description of this subtype can be found below (section 1.3 "Triple negative and basal-like breast cancer").

Based on these intrinsic subtypes, a 50-gene signature (PAM50) with significant prognostic and predictive value was developed by Parker and colleagues [16] and led to the development of Prosigna™, an FDA-cleared and CE marked test that can identify a tumor's intrinsic subtype and estimate the risk of distant recurrence of hormone-receptor positive tumors. Other gene-expression based tests such as

Oncotype DX™ [71], MammaPrint® [45] and Endopredict® [72] predict recurrence; still, do not provide information about breast cancer intrinsic subtypes.

Several studies have assessed the relevance of the intrinsic subtypes in prognosis and response to therapy, establishing that they represent unique biological entities that may guide patients' management and treatment selection [36, 69, 73]. Nevertheless, this classification has not yet reached clinical implementation. In the last St. Gallen Conference (St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017), the Experts Panel exposed that a classification based on **immunohistochemical surrogates** to reproduce intrinsic subtypes (Table 1) was clinically valuable and should be used for therapy selection [7]. They highlighted that patients with clear indications on the use or not of chemotherapy do not benefit of routine gene-expression testing. However, in patients in the zone in between that are classified as equivocal cases by clinical surrogates (e.g. women with ER+, HER2-, with tumors between 1-3 cm, with 0-3 positive lymph nodes, and intermediate Ki-67 proliferation rate) the question is whether they require chemotherapy in addition to hormonotherapy. In this set of patients, the Experts Panel recommend the use of gene-expression tests to decide for or against chemotherapy, as they provide a more realistic assessment of prognosis than immunohistochemical surrogates.

Table 1. Immunohistochemical surrogates to determine breast cancer intrinsic subtypes

Intrinsic subtype	Biomarker profile
Luminal A	ER+ and/or PR+, HER2-, low Ki-67 (< 14 %)[a]
Luminal B	ER+ and/or PR+, HER2-, high Ki-67 (≥ 14 %)[a]
HER2-enriched	ER-, PR-, HER2+
Basal-like	ER-, PR-, HER2-, CK5/6+ and/or EGFR+[b]

[a] *Classification of luminal A and luminal B depending on Ki-67 is based on Cheang et al. 2009 [74]. Luminal A is differentiated from luminal B in some studies based on HER2 status [33, 74], although in the last St. Gallen Conference, the method recommended was the one based on Ki-67 status.*

[b] *Definition of basal-like based on Cheang et al. 2008 [75]*

1.2.4 Integrative analysis

The most recent classifications have gone a step further by analyzing **integrative data** at multiple levels to better understand the heterogeneity of breast cancer. The Cancer Genome Atlas Network (TCGA) integrated information of multiple platforms (copy number profiling, DNA methylation, exome sequencing, gene expression analysis, miRNA sequencing and reverse-phase protein arrays) and found a convergent classification of breast tumors in four major subtypes (luminal A, luminal B, HER2-positive, triple-negative) [76]. The study of Curtis et al. [77] integrated gene expression, copy number profiling and single nucleotide polymorphisms data revealing a refined breast cancer taxonomy with 10 different subclasses (Integrative Clusters, IntClust 1-10) associated with distinct clinical features and outcomes.

I have so far explained the current state of the art regarding the diverse layers of breast cancer intertumor heterogeneity. In this Thesis Project I explored this subject by focusing on the **tumor tissue heterogeneity** in 1) triple negative breast cancer as the most heterogeneous subtype of breast cancer—patients present a wide range of clinical outcome and treatment response, so there is a need to identify subgroups that share biological features and response to treatment—and 2) male breast cancer as an example of an atypical type of breast cancer that has been rarely studied, the molecular/genomic heterogeneity of which is not well defined; and also focusing on **non-tumor heterogeneity** by 3) exploring the host genetic factors that determine the different aggressiveness of breast cancer independently of the tumor's clinicopathological characteristics. In the next sections, I will address these three items in more depth.

1.3 TRIPLE NEGATIVE AND BASAL-LIKE BREAST CANCER

1.3.1 Definitions and clinicopathological characteristics

Triple negative breast cancer is the most challenging subtype from a molecular and clinical perspective because its poor survival, wide heterogeneity in treatment response—i.e. intertumor heterogeneity—and its lack of targeted therapies [5, 6].

Gene-expression analysis studies have shown that triple negative breast cancer is composed of all the intrinsic subtypes, being basal-like the most common (~ 70 %) [15]. Although triple negative and basal-like are used sometimes as equivalent terms, they are not synonymous. Triple negative is a clinically defined term based on immunohistochemical expression: tumors are triple negative when they lack ER and PR expression, and HER2 amplification/overexpression. Basal-like is a term based on gene-expression analysis: tumors are basal-like when their gene-expression pattern resembles that of basal myoepithelial cells. Thus, not all triple negative tumors defined by immunohistochemistry are basal-like by gene expression, not all basal-like are triple negative (Figure 2). There are however immunohistochemical surrogates that can help to distinguish basal-like (CK5/6+ and/or EGFR+) from non-basal-like (CK5/6- and EGFR-) tumors within triple negative breast cancer [75].

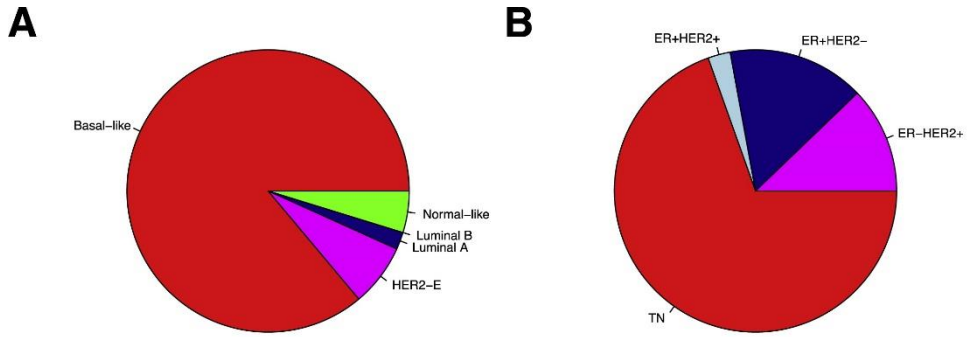

Figure 2. Overlapping of triple negative and basal-like breast cancer in the METABRIC study [77]**.** A) Distribution of intrinsic subtypes within triple negative breast cancer. B) Distribution of immunohistochemical subtypes within basal-like breast cancer. *Abbreviations: HER2-E, HER2-enriched; ER, estrogen receptor; TN, triple negative. Image from Engebraaten et al. 2013 [132]*

Basal-like tumors express high levels of basal myoepithelial markers (CK5, 6, 14, 17), laminin, P-cadherin, fascin, caveolins 1 and 2, alpha-beta crystallin and EGFR. They have high genomic instability, with frequent mutations in TP53, deregulated integrin expression and inactivation of the retinoblastoma pathway [70]. They tend to occur in younger patients, are usually high grade, highly proliferative and poorly differentiated.

1.3.2 Clinical implications: lack of targeted therapies

Due to the lack of appropriate biomarkers and therefore of targeted therapies, anthracycline and taxane-based chemotherapy has been traditionally the mainstay of therapy for triple negative breast cancer patients. However, their response is moderate and widely heterogeneous. After receiving neoadjuvant chemotherapy (a chemotherapy administered before breast surgery), about 30 % of triple negative breast cancer patients present a pathological complete response—complete absence of invasive tumor. Achieving a pathological complete response improves patients' prognoses to the point that their disease-free survival and overall survival are similar to patients with less aggressive tumors. However, triple negative breast cancer patients with residual disease after chemotherapy have worse survival and prognosis than those non-triple negative [8–10]. Therefore, pathological complete response to neoadjuvant chemotherapy is a surrogate for survival in triple negative breast cancer patients.

The response of triple negative breast cancer patients to platinum-based chemotherapy varies widely, such that clinical guides are against its routine use and only recommend it in patients with mutations in BRCA1/2 genes in addition to standard chemotherapy [7]. This should not be surprising given the heterogeneity of triple negative breast cancer: several studies have further subdivided triple negative in subgroups with unique molecular characteristics and response to treatment [5, 6]. To improve patients' management, different therapeutic agents could be selected according the triple negative subtype. Some therapies that are nowadays being tested beyond platinum-based therapies are anti-EGFR, anti-VEGF (vascular endothelial growth factor receptor), inhibitors of PARP (poly ADP ribose polymerase), and inhibitors of the androgen receptor, among others.

1.3.3 Triple negative breast cancer subtyping

Several triple negative breast cancer classifications have been published to date. Lehmann, Bauer et al. classified triple negative breast cancer in seven subtypes based on gene expression profiling. This classification included 6 stable subtypes consisting in two basal-like (BL1 and BL2), an immunomodulatory (IM), a mesenchymal (M), a

mesenchymal stem-like (MSL), and a luminal androgen receptor (LAR) subtype; as well as an unstable (UNS) subtype. These subtypes were reproduced and pharmacologically targeted in breast cancer cell lines suggesting that they can inform therapy selection [11]. However, Prat and colleagues highlighted that Lehmann subtyping ignores samples that are highly contaminated with normal breast tissue, which are mostly classified as MSL [19]. Another classification is the one published by Burstein and colleagues, that classifies triple negative breast cancer in four subgroups with different prognoses: luminal androgen receptor, mesenchymal, basal-like immunosuppressed and basal-like immune-activated [12]. Similarly, Liu et al. used human transcriptome microarrays to classify 165 triple negative breast cancers into four distinct clusters, including an immunomodulatory (IM), a luminal androgen receptor (LAR), a mesenchymal-like (MES) and a basal-like and immune suppressed (BLIS) subtype [13]. **Although these subtypes may have unique biological and molecular characteristics, we do not know yet if they are predictive of treatment selection and efficacy.**

1.4. MALE BREAST CANCER

1.4.1 Epidemiology of male breast cancer

Male breast cancer is a rare disease, accounting for less than 1 % of all breast cancers diagnosed and less than 1 % of all cancers in men. Its incidence is very low, only 1 in 1 000 men will ever be diagnosed with breast cancer, in contrast to the much higher rates in female breast cancer [31]. The reason of the lower incidence in men is the lesser amount of breast tissue along with the difference in their hormonal environment. Countries with high or low female breast cancer incidence tend to have also high or low male breast cancer incidence rates. This correlation may indicate common risk factors in both men and women [78].

Breast cancer has a higher mortality in men, primarily because less awareness about the disease. Men are less likely to assume that a lump could be due to breast cancer, leading to a delay in diagnosis, higher levels of comorbidity and more advanced stages at presentation [79].

1.4.2 Male breast cancer vs. female breast cancer characteristics

Breast tissue is basically the same in men and women until puberty, when female breasts undergo complex changes while male breasts remains underdeveloped [80]. Female breast contains ducts, glandular epithelium and non-adipose stroma, while the male breast is mostly adipose tissue with few ducts and periductal stroma. A representation of male and female's normal breast tissue is depicted in Figure 3.

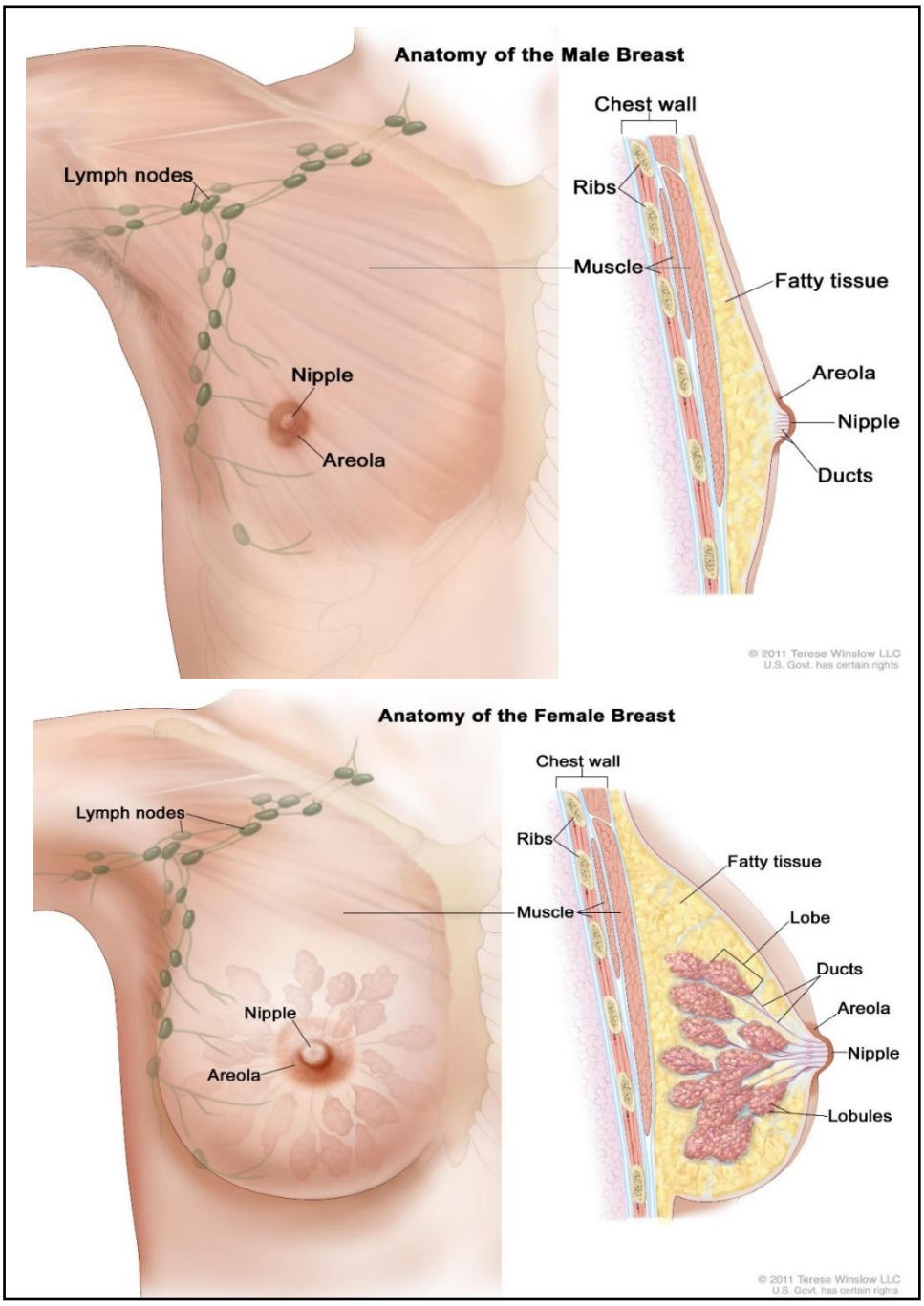

Figure 3. Anatomy of male and female's normal breast. *Image from https://www.cancer.gov [accessed on 10/10/2017]*

Breast cancer is diagnosed at more advanced stages in men than in women and also later in life (67 years vs. 62 years in females), mainly do the lack of awareness in the male population [81, 82]. Male breast cancers are mainly hormone receptor positive (ER+ in 80 to 81 % of males vs. 76 to 78 % in females and PR+ in 80 to 81 % vs. 67 %, respectively) and very rarely HER2-positive (in contrast with 15-20 % of females). Due to the simpler anatomy of male breast compared with female breast, its histology is more homogeneous, henceforth, male breast cancer is mainly classified as invasive carcinoma of no special type (90 to 97 % of male tumors *vs.* 83 to 86 % of female tumors). Male breast cancer almost never has an invasive lobular morphology since lobular differentiation is unusual in males. The rare cases with lobular morphology may be due to Klinefelter's syndrome (chromosomal XXY disorder), the intake of estrogens as the case of transsexuals or to the development of prostate cancer, but in general have no known endocrine risk factor [80]. Male breast cancer also differs from female breast cancer in its mutational status repertoire and the mutational frequency of the most commonly mutated genes; males have less PIK3CA and TP53 mutations than females of the same immunohistochemical profile, but display more frequently mutations in genes associated with DNA repair genes [83] and have higher incidence of mutations in BRCA2, CHECK2 and PALB2 than females [84].

Although traditionally male breast cancer has been considered similar to postmenopausal female breast cancer, a closer examination of the emerging data suggests that molecular and clinical characteristics of males do not overlap females. Nevertheless, due to its low incidence and the lack of data from prospective randomized trials, male breast cancer recommendations for treatment are extrapolated from studies with small number of patients and knowledge derived from female breast cancer studies. This way, **the deepening in the molecular knowledge of male breast cancer is essential to increase the understanding of its etiology and the development of an appropriate therapeutic approach.**

1.4.3 Male breast cancer subtyping

Several research groups have attempted to classify male breast cancer into molecular subtypes using the same surrogate immunohistochemical classification of

female breast cancer (summarized in table 3). These molecular profiles indicate that female and male breast cancer are different diseases: male breast cancer is mainly a luminal disease, with fewer triple-negative and HER2-positive tumors than females.

Table 3: Distribution of immunohistochemical subtypes in male and female breast cancer

Gender	Study	Luminal A	Luminal B	HER2-positive	Basal-like
Male	[a]Ge *et al.* 2009 [85] (N = 42)	35 (83 %)	7 (17 %)	0	0
	[b]Sánchez-Muñoz *et al.* 2012 [86] (N = 43)	19 (44 %)	22 (51 %)	0	2 (5 %)
	[b]Kornegoor *et al.* 2012 [87] (N = 129)	98 (76 %)	27 (21 %)	0	4 (3 %)
	[a]Shaaban *et al.* 2012 [88] (N = 203)	199 (98 %)	0	0	4 (2 %)
	[a]Yu *et al.* 2013[89] (N = 68)	41 (60 %)	17 (25 %)	6 (9 %)	4 (6 %)
	[a]Leone *et al.* 2015 [90] (N = 960)	815 (85 %)	111 (11 %)	6 (1 %)	28 (3 %)
	[b]Piscuoglio *et al.* 2015 [83] (N = 59)	17 (29 %)	42 (71 %)	0	0
Female	[a]Blows *et al.* 2010 [91] (N = 10 159)	7 243 (72 %)	639 (6 %)	632 (6 %)	1 645 (16 %)
	[a]Shabaan *et al.* 2012 [88] (N = 220)	197 (90 %)	14 (6 %)	4 (2 %)	5 (2 %)

[a]*Luminal A: ER+ and/or PR+, HER2-; luminal B: ER+ and/or PR+, HER2+*

[b]*Luminal A: ER+ and/or PR+, HER2-, low Ki-67; luminal B: ER+ and/or PR+, HER2+ and/or high Ki-67*

Regarding luminal tumors, there has been no consensus in the definition of immunohistochemical luminal A and luminal B, so a comparison of results is difficult. When this differentiation is based on HER2 positivity [85, 88–90] (luminal A defined

as ER+ and/or PR+, HER2-; and luminal B as ER+ and/or PR+, HER2+), the great majority of male breast tumors are classified as luminal A (60 to 98 %). This predominance of luminal A over luminal B is not so clear in those studies that differentiated between luminal A and B according to differences in the expression of the proliferation marker Ki-67 [83, 86, 87] (luminal A defined as ER+ and/or PR+, low Ki-67; and luminal B as ER+ and/or PR+, high Ki-67). This last definition was adopted by the St. Gallen Expert Consensus Panel recommendation guidelines for the systemic treatment of early breast cancer and improved the distinction between luminal A and luminal B tumors [7, 33]. However, there are still many concerns about this classification method due to the considerable lack of reproducibility between laboratories—as methods are not yet standardized—and the lack of an optimal Ki-67 cut-off point. Therefore, the use of immunohistochemical markers may result in the misclassification of tumors compared with the information generated by gene expression profiling methods.

As described above (section 1.2.3 "Gene-expression heterogeneity: intrinsic subtypes"), gene-expression based signatures provide useful prognostic data in female breast cancer patients. In contrast, the genetic landscape of male breast cancer remains fundamentally uncharacterized. Johansson et al. classified male breast cancer into two molecular subgroups based on copy number alterations and gene-expression profiling, luminal M1 and luminal M2 [92, 93]. Luminal M1 included more aggressive tumors with high levels of chromosomal changes and had upregulated genes in cell proliferation, migration, tumor invasion and metastasis. Luminal M2 tumors show upregulated immune response genes and ER signaling associated genes. Still, they did not resemble any of the subgroups reported for females. Even though most male breast tumors are ER+, they behave differently to ER+ female breast cancer because of a gender associated landscape in hormone receptor pathways.

This study highlights the intertumor heterogeneity that exists in male breast cancer patients and their biological differences with female breast cancer patients. However, prognostic gene signatures (like PAM50 in females) have yet to be evaluated or developed to improve the prediction of clinical outcome and guide treatment decisions in male breast cancer patients.

1.5. INFLUENCE OF THE GENETIC BACKGROUND IN BREAST CANCER METASTASIS

1.5.1 Metastasis and breast cancer survival

We have so far exposed evidences supporting that breast tumors are highly heterogeneous at various levels and have described the histopathological and molecular features that can be used to characterize the different breast cancer subtypes defined to date. These intrinsic characteristics are also predictive of a patient's prognosis and can help in treatment decisions or in the search of targeted treatments. However, the development of a tumor is not the direct reason of cancer's high mortality, since up to 90 % of cancer deaths are due to complications arising from the metastatic dissemination of the disease [94]. Even if metastatic breast cancer can be treated, it remains an incurable disease with a median overall survival of 2 to 3 years and almost universally fatal within 5 to 10 years [95, 96]. Hence, unravelling the molecular mechanisms that drive tumor cells to metastasize could lead to the prevention of metastasis in early diagnosed breast cancer patients, with the result of a better patient management and a more personalized treatment.

1.5.2 Metastasis and the seed and soil hypothesis

Metastasis is an extraordinarily complex process in which tumor cells acquire a set of abilities that allow them to develop new tumors in secondary sites (Figure 4). The mechanisms that trigger the metastatic cascade and the factors that regulate these processes are not yet fully understood. Even when several models have been proposed to explain the phenomena underlying tumor dissemination, none can fully explain the biological and clinical observations associated to metastasis [3].

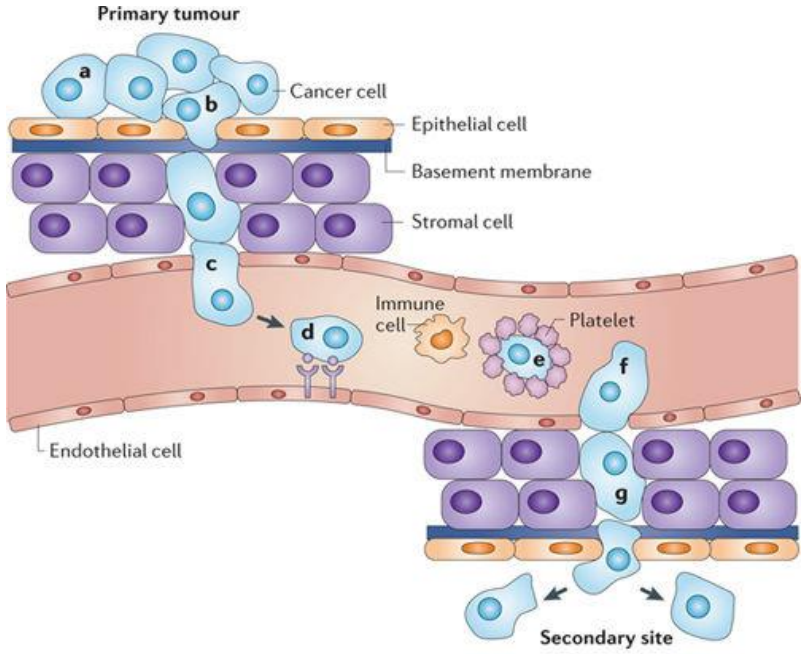

Figure 4. The steps of the metastatic cascade. Cancer cells detach from the primary tumor by a) reducing adhesion to neighbor cells and b) clearing a path for migration into the stroma. Then, they enter the bloodstream by c) intravasation. Once in the bloodstream, cancer cells can express receptors that bind to d) metastasis-supporting sites or to e) platelets which protect them from the immune system. After reaching the secondary site, cancer cells must f) exit the bloodstream and g) proliferate to form a secondary tumor. *Image from Schroeder et al. 2011 [133]*

The metastatic cascade requires multiple steps, which encompass not only tumor cells but also the cooperation of different cells of the host: cells from the tumor stroma, bloodstream and secondary sites [97, 98]. There are evidences from epidemiological studies and animal models suggesting that the risk of developing metastasis after breast cancer diagnosis depends not only on the tumor characteristics but also on the germline genetic variants of the host [99–101]. Therefore, despite the tumor acquisition of further mutations that allow its cells to disseminate and proliferate in a secondary site (**metastatic seeds**), certain characteristics of the host's genetic background could provide either a congenial **soil** promoting the development of metastasis or a set of features that may be protective against metastasis, independently or in conjunction with certain tumor characteristics [3]. This way, the genetic background of a patient would modulate the metastatic dissemination of a tumor likewise constitutive polymorphisms are responsible of other traits or characteristics of an individual.

1.5.3 Germline genetic variants associated to metastasis

Several candidate gene and genome-wide association studies (GWAS) have attempted to identify human germline genetic variants with breast cancer survival at nominal (p-value < 0.05) or genome-wide (p-value < 5 x 10^{-8}) significance levels. However, the single nucleotide polymorphisms (SNP) identified to date are not consistent between different studies or have not been validated in further studies [3, 38]. This limitation is not a particular feature of breast cancer, as GWAS have had a limited success in detecting genetic variants associated to heritability of complex traits [39].

In complex traits such as metastasis, multiple genetic features that tend to be spread across the genome and to interact with one another are responsible of small differences in patient's survival. That is, variants with individual, strong effects are rare and the detection of the set of variants with cumulative weak effect on metastasis requires a greater research effort than the achieved with the individual variant analysis performed through GWAS [102]. In these cases, the study of all the spectrum of samples could difficult the identification of the allelic variants that truly determine the overall expression of a metastatic phenotype. A useful and intuitive approximation to overcome this limitation would be the minimization of the phenotype heterogeneity by the selection of samples that most likely harbor relevant genetic information. These approaches increase the ability to identify allelic variants that predispose to or prevent from the particular phenotype under study [40, 41, 103].

2. HYPOTHESES AND OBJECTIVES

If we knew what it was we were doing,
it would not be called research, would it?

Albert Einstein (1879 - 1955)

2.1 HYPOTHESES

Breast cancer is a wide term that encompasses an heterogeneous set of diseases with unique biological, molecular and clinical characteristics. Heterogeneity is present in all features of breast cancer both within a given patient—i.e. intratumor heterogeneity—and among different patients—i.e. intertumor heterogeneity. Focusing on intertumor heterogeneity, a rational way to handle breast cancer would be the identification of a set of manageable and homogenous subgroups, that help us improve the knowledge of the disease and patients' management. Going a step further, there are evidences supporting that non-tumor tissues—normal tissue in the tumor microenvironment or metastasis target tissue, among others—could also be an important element to unravel the complexity of the heterogeneity observed among patients. We believe that tumor subtyping and the analysis of germline genetic features could contribute to decipher breast cancer intertumor heterogeneity, improving breast cancer treatment selection and patients' outcome.

Given the aforementioned, this Thesis is focused on exploring intertumor heterogeneity (between different patients) both in the tumor tissue and in the host (non-tumor tissue) through the following approaches:

1.- Triple negative breast cancer is the most heterogeneous breast cancer subtype, with the worst prognosis and the widest range of response to treatment. There is a need to identify clinically relevant subtypes that can improve therapy selection and patients' outcome. Our hypothesis is that the Lehmann classification of triple negative breast cancer subtypes can be useful in predicting the response to neoadjuvant chemotherapy.

2.- Male breast cancer is a mostly unexplored type of breast cancer, the management and treatment of which is largely based on female breast cancer studies. We believe that it is paramount to classify male breast cancer in clinically validated subtypes (as it has been done in female breast cancer) that reflect its unique biology, predict clinical outcome and guide treatment decision. Our hypothesis is that the molecular classification of breast cancer in intrinsic subtypes with the PAM50 signature could

be more useful than the traditional clinical classification based on immunohistochemical surrogates to reveal the molecular characteristics of male breast cancer, and to predict its outcome and treatment response.

3.- The wide heterogeneity observed in breast cancer survival cannot be explained only by the tumor characteristics or treatment selection. It has been suggested that normal tissue around the tumor (tumor microenvironment) as other tissues of the organism (blood stream, metastasis target tissues) may play a role in breast cancer survival, mostly due to their influence in the metastatic dissemination of the disease. Therefore, our hypothesis is that the genetic background of the host modulates the metastatic dissemination of the disease and this knowledge, combined with the analysis of the tumor characteristics, may improve the prognosis of breast cancer patients and unveil new biomarkers for targeted therapies.

2.2 OBJECTIVES

The three hypotheses above lead to the objectives of this Doctoral Thesis, as outlined below.

1.- To validate the clinical relevance of Lehman triple negative breast cancer subtypes by assessing their correlation with pathological complete response to standard neoadjuvant chemotherapy with or without platinum salts.

2.- To evaluate the relevance of male breast cancer PAM50-based subtypes on clinicopathological features and clinical outcome compared with a validated panel of immunohistochemical surrogates.

3.- To investigate the role of the host's genetic background in the development of breast cancer metastasis through the search of germline genetic variants associated to metastasis susceptibility.

3. GENERAL METHODS

An experiment is a question which science poses to Nature,
and a measurement is the recording of Nature's answer.

Max Planck, (1858 – 1947)

3.1 SAMPLES PROCESSING

All patients analyzed were informed of the purpose of the study they were recruited for and gave their written informed consent to participate. The study protocols were approved by the corresponding ethical committee (Comité Coordinador de Ética de la Investigación Biomédica de Andalucía).

3.1.1 Tumor tissue samples

The Málaga Hospital-IBIMA Biobank managed and/or provided most of the tumor tissue blocks analyzed in this project; tumor blocks of patients that had participated in the clinical trial GEICAM/2006-03 were provided by GEICAM Biobank. Samples were retrieved from a biopsy (obtained for diagnosis) or during surgery and included in paraffin blocks (FFPE, formalin-fixed paraffin-embedded) to maintain cellular morphology and preserve the tissue over time. To determine the tumor area of each sample or if it was tumor tissue left after treatment we performed a hematoxylin and eosin staining of a slide-mounted tumor section, that was reviewed and marked by a pathologist.

We performed immunohistochemical staining of 5 μm sections in an automatic immunostainer (Autostainer Plus, Dako) using the EnVision FLEX System (Dako). As chromogen, we used diaminobenzidine counterstained with hematoxylin. We used a nonimmune serum to replace the primary antibody as a negative control and specific positive controls to each antibody. The antibodies used were: estrogen receptor, ER (rabbit monoclonal antibody, clone SP1, Master Diagnostica); progesterone receptor, PR (rabbit monoclonal antibody, clone Y85, Master Diagnostica); cytokeratin 5/6, CK5/6 (mouse monoclonal, clone D5/16B4, Master Diagnostica); epidermal growth factor receptor, EGFR (rabbit monoclonal antibody, clone EP38Y, Master Diagnostica); Ki-67 (rabbit monoclonal antibody, clone SP6, Master Diagnostica); androgen receptor, AR (mouse monoclonal antibody, Dako) and human epidermal growth factor receptor 2, HER2 (Kit Herceptest, Dako). The stains were evaluated and scored by a pathologist. Positivity was defined as any nuclear staining (> 1 %) [104] for ER, PR and AR, and as any membranous or cytoplasmic staining for EGFR

and CK5/6. Ki-67 index was determined quantifying the percentage of nuclear-positive neoplastic cells. HER2 status was evaluated according international practice guidelines [105] and categorized as negative if immunohistochemical results were 0 to +1; when results were +2 we investigated the gene amplification by fluorescence (FISH) or chromogenic (CISH) *in situ* hybridization.

Prior RNA purification, we manually macrodissected the tumor area to get rid of normal tissue and obtained 3-6 sections of 10 μm thick. Purification was performed with the RNeasy FFPE Kit (Qiagen) following manufacturer's guidelines. We determined the concentration and quality of the purified RNA with a Nanodrop 1 000 spectrophotometer (Thermo Scientific).

3.1.2 Blood samples

In each collaborating hospital, peripheral blood from the patients was collected in 3 mL tubes with sodium citrate as anticoagulant. Whole blood was frozen at -20 ºC and shipped to our lab (Laboratorio de Investigación Biomédica, Hospital Universitario Virgen de la Victoria, Málaga) in order to centralize sample's processing. We isolated the buffy coat (blood fraction that contains most of the white blood cells and, therefore, the patient's DNA) by centrifugation before DNA purification to improve its yield. To isolate and purify the genomic DNA we used the QIAamp DNA Blood Mini Kit (Qiagen) according manufacturer's guidelines. In all cases we started from the buffy coat pellet resuspended in 200 μL of PBS (Phosphate-buffered saline). After purification, we determined DNA concentration and quality with a Nanodrop 1 000 spectrophotometer (Thermo Scientific).

3.2 INTRINSIC SUBTYPES

We classified triple negative and male breast tumor samples into 1 of the 4 intrinsic subtypes (luminal A, luminal B, HER2-enriched, basal-like) using an nCounter Dx Analysis System (Nanostring). This system measures directly the gene expression of multiple targets attached to specific color-coded barcodes. These targets can be individually counted, offering high levels of precision and sensitivity (< 1 copy per cell) and avoiding the biases of amplification-based methods as PCR or

microarrays. The nCounter System consists in a PrepStation module for sample processing and a Digital Analyzer for direct digital counting.

Prior RNA processing, we selected RNA samples passing Nanostring quality criteria (concentration ≥ 12.5 ng/μL and A260/280 ratio of 1.7-2.3, measured with Nanodrop). RNA was processed according to the manufacturer's guidelines. A detailed description of the workflow is depicted in Figure 5.

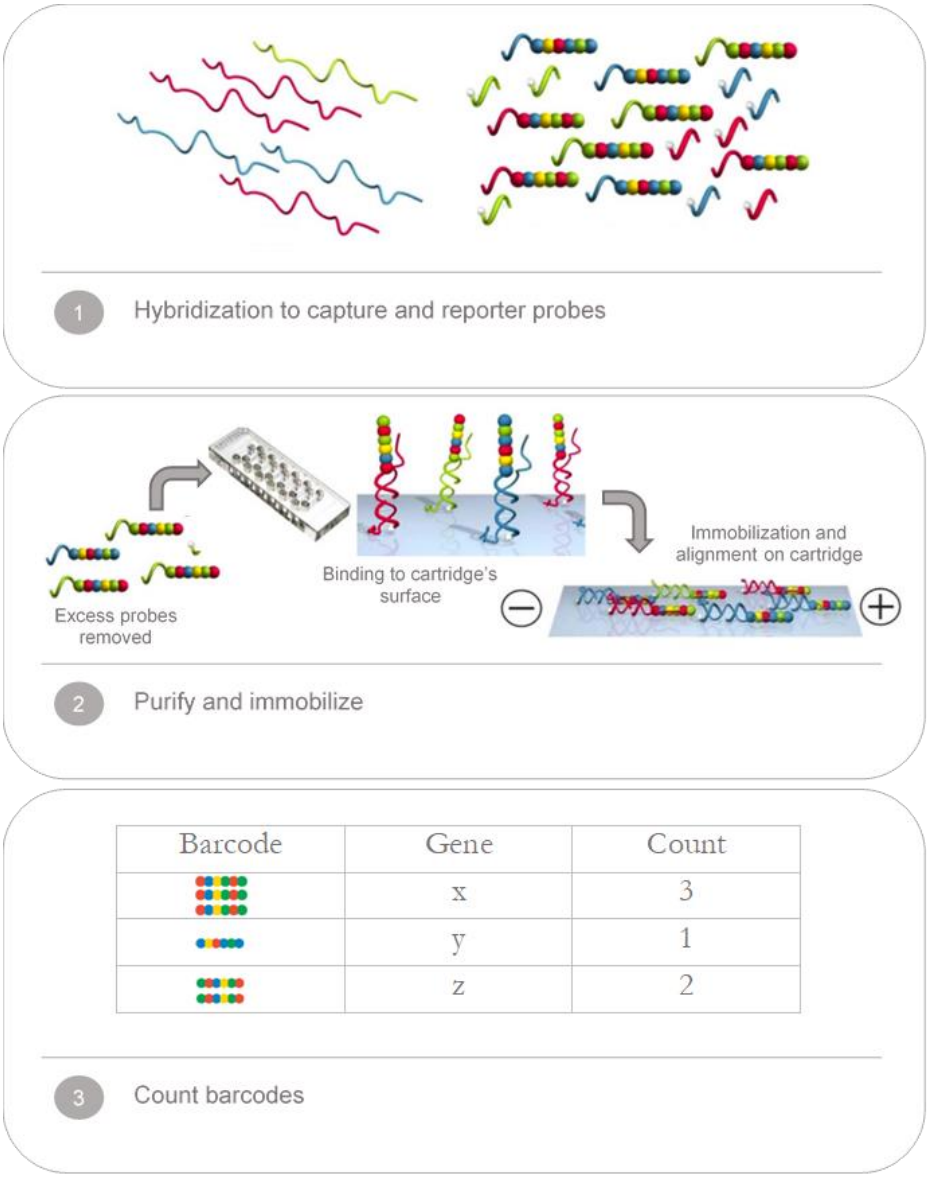

Figure 5. nCounter System workflow. 1) Each mRNA is hybridized to a reporter probe (carries the signal) and to a capture probe (allows immobilization in the cartridge). 2) Hybridized samples are loaded into the PrepStation, where the excess of probes are removed and the probe/target complexes are immobilized and aligned in the cartridge surface. 3) Cartridges are placed in the Digital Analyzer for data collection; color codes on the surface of the cartridge are counted and assigned to each target. *The images used to build this workflow were from Nanostring's website (https://www.nanostring.com/scientific-content/technology-overview/ncounter-technology)*

More information about the kits and data analysis used for the different cohorts can be found in the sections 4.2.3 "Triple negative breast cancer subtyping: intrinsic and Lehmann subtypes" and 5.2.3 "Intrinsic subtypes: PAM50 signature".

3.3 LEHMANN SUBTYPES

3.3.1 Whole-transcriptome analysis

Gene expression analysis and Lehmann subtyping classification were performed at the Microarray Analysis Service (SAM) core facility from Hospital del Mar Medical Research Institute (IMIM) using the Affymetrix GeneChip® Platform.

For whole-transcriptome analysis we used the SensationPlus™ FFPE Amplification and WT Labeling kit combined with the GeneChip® Human Transcriptome Array 2.0. This array contains > 6 million different probes that cover coding and non-coding transcripts. It was designed with ten probes per exon and four per exon-exon splice junction, ensuring accurate and reproducible data even in highly degraded RNA samples. The SensationPlus™ FFPE Amplification kit provides a one round of whole-transcriptome amplification using T7 RNA polymerase and a linear amplification method. It generates high quality amplified RNA from degraded RNA overcoming the challenges associated with FFPE samples (RNA degradation, chemical modification and low yield). A set of poly-A RNA controls supplied by Affymetrix are added to the RNA before starting the protocol in order to monitor the whole process independently from the quality of the starting RNA samples. After amplification, the SensationPlus™ WT Labeling Kit converts the senseRNA into fragmented and biotin-labeled double-stranded cDNA, which is ready for hybridization.

Prior RNA processing, we evaluated its integrity with a Bioanalyzer 2 100 (Agilent Technologies). Gene expression analysis was performed according the manufacturer's instructions as depicted in Figure 6. All details about the protocol and more technical information about the GeneChip® Human Transcriptome Array 2.0. can be found at ThermoFisher's website (https://www.thermofisher.com).

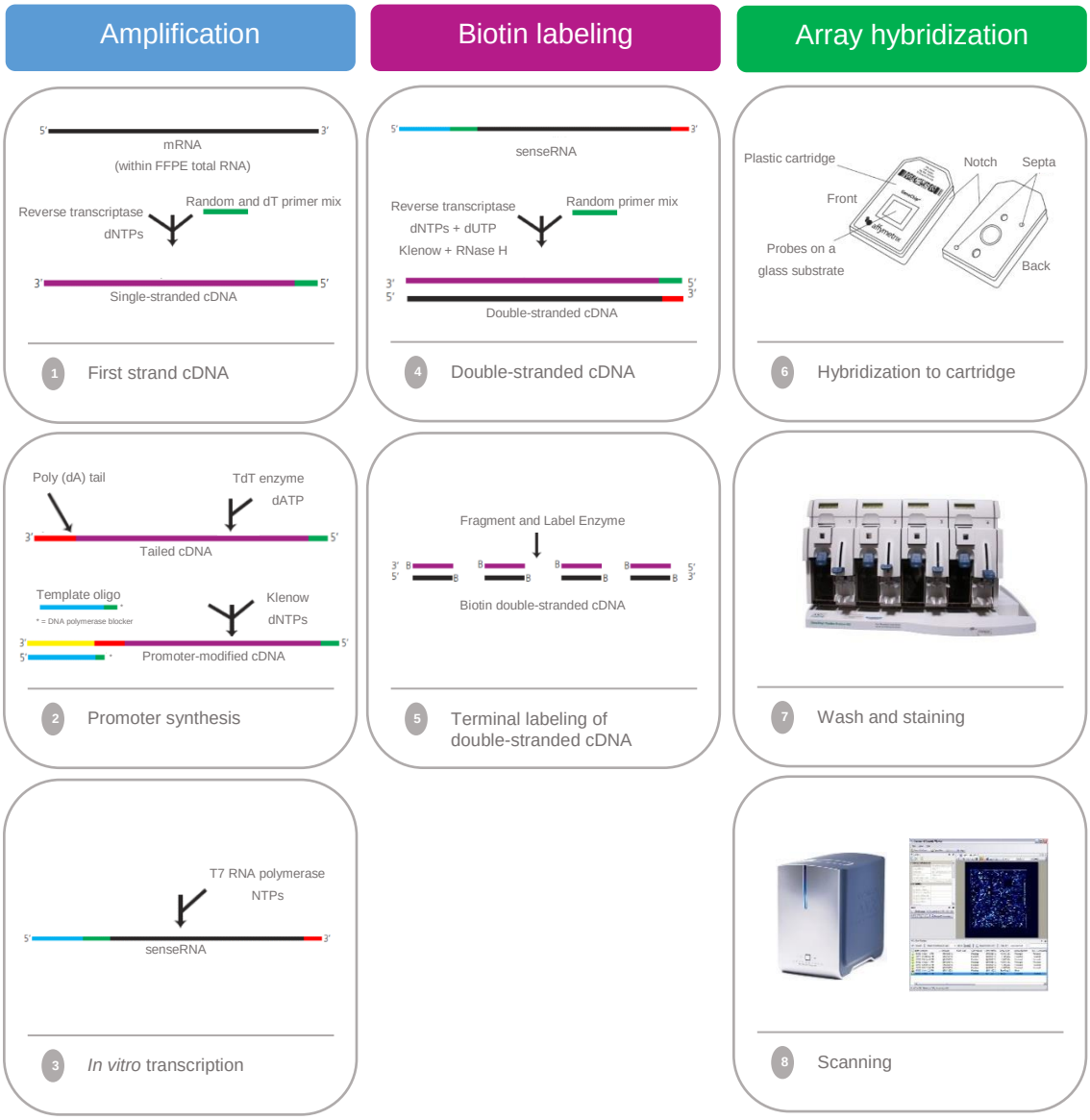

Figure 6. Whole-transcriptome analysis workflow**.** 1) Total RNA (200 ng) is reverse transcribed into a single-stranded cDNA using random primer and dT primers. 2) cDNA is poly (dA) tailed on the 3' end using a terminal deoxynucleotidyl transferase. A T7 dT oligo with a 3' blocking group is hybridized to the 3' tail of the cDNA. Klenow and dNTP mix generate a double stranded T7 promoter region on the single stranded cDNA. 3) The promoter-modified cDNA is *in vitro* transcribed using T7 RNA polymerase, resulting in senseRNA. 4) After this round of whole transcriptome amplification, the senseRNA is reverse transcribed and a second strand cDNA is made with Klenow and RNase H. 5) Double-stranded cDNA is fragmented and end-labeled with biotin. 6) The biotin-double-stranded cDNA is hybridized to the array in the GeneChip Hybridization Oven 645. 7) The arrays are washed and stained in the Fluidics Station 450. 8) Finally, the intensity of fluorescence of each probe is quantified in a GeneChip 3000 scanner. Image processing and initial quality control analysis were performed using Affymetrix GeneChip Command Console (AGCC) v. 4.0 and Affymetrix Expression Console (EC, v. 1.4.1) software, respectively. All reagents, equipment and software used are from Affymetrix. *Most images used to build this workflow were from the Sensation Plus™ Amplification and WT Labeling kit User Guide available at http://tools.thermofisher.com/content/sfs/manuals/sensationplus_wt_ffpe_user_guide.pdf*

3.3.2 Data analysis and subtyping

Data were normalized with the Robust Multichip Analysis (RMA) algorithm in the Affymetrix Expression Console (EC, v. 1.4.1). RMA creates an expression matrix from Affymetrix data. The raw intensity values are background corrected, transformed and normalized to obtain an expression measure for each probe set on each array. Data were annotated in the statistical computing environment R (v. 3.2.3) using hg19 human genome built and duplicated genes were mean summarized. Subtypes were identified with the web-based tool TNBCtype [18]. When uploading a normalized gene expression data matrix, this tool displays the predicted Lehmann subtype with a corresponding correlation coefficient and the permutation *p*-value. It includes a filter that removes samples expressing estrogen receptor—i.e. erroneously classified as triple negative—as they can influence normalization and prediction results. Gene expression data generated in this study are available in the GEO repository at: www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE106977 (accession GSE106977).

3.4. GENOME-WIDE ASSOCIATION STUDY

3.4.1 Genome-wide genotyping

Genome-wide genotyping of patient's genetic background was performed in collaboration with the Human Genotyping Unit from the National Centre of Oncological Research (CNIO). This Unit, headed by Dr. Anna González Neira, is part of the National Genotyping Centre (CeGen, www.cegen.org) and is the first genetics laboratory providing high throughput technology accredited in Europe, ensuring expertise and reliability of the results obtained.

For genotyping we used the Illumina Infinium LCG Quad Assay protocol with the HumanOmni5-Quad Beadchip. This chip contains ca. 4.3 million SNPs (4 301 331 SNPs) selectively distributed and with an average distance of 0.68 kb. This elevated number of polymorphic sequences offers the required coverage of the genome to detect frequent (in at least 5 % of the population or with a minimum allele frequency (MAF) of 5 %) and rare (MAF < 1 %) variants. Prior to genotyping, we

quantified again the DNA by spectrofluorometry using the PicoGreen dsDNA quantitation assay (ThermoFisher). This measure is more precise that the obtained with Nanodrop, which overestimates the DNA concentration, although it does not provide any quality measure. Once quantified, DNA was diluted to a final concentration of 50 ng/μL. In order to determine if the results were reproducible, we included a technical replicate in the setup (one same sample was included twice).

Genotyping was performed according the manufacturer's instructions, which are depicted in Figure 7. All details about the protocol and more technical information about the HumanOmin5-Quad Beadchip can be found at Illumina's website (http://www.illumina.com).

3.4.2 Individual effect variant detection

We looked for germline genetic variants with **strong individual effects** in the metastatic cascade by performing an association analysis (GWAS) between allelic variants in the good and poor prognosis cases with the PLINK [43] library within the Encore pipeline [44]. We performed a quality control before the analysis, applying filters of missing rate (per person and per SNP) and Hardy-Weinberg equilibrium. Missingness filter excludes both individuals with many missing genotype data and SNPs that are not genotyped in a minimum percentage of the samples; we set the missingness filter to 0.1, meaning that samples with more than 10 % missing genotypes or SNPs not genotyped in more than 10 % of the samples were excluded. The Hardy-Weinberg equilibrium filter is used to estimate if the homozygous and heterozygous genotypes significantly differ from the prediction of equilibrium, which would indicate genotyping errors or batch effects; we set this filter to 10^{-6} as recommended by the software.

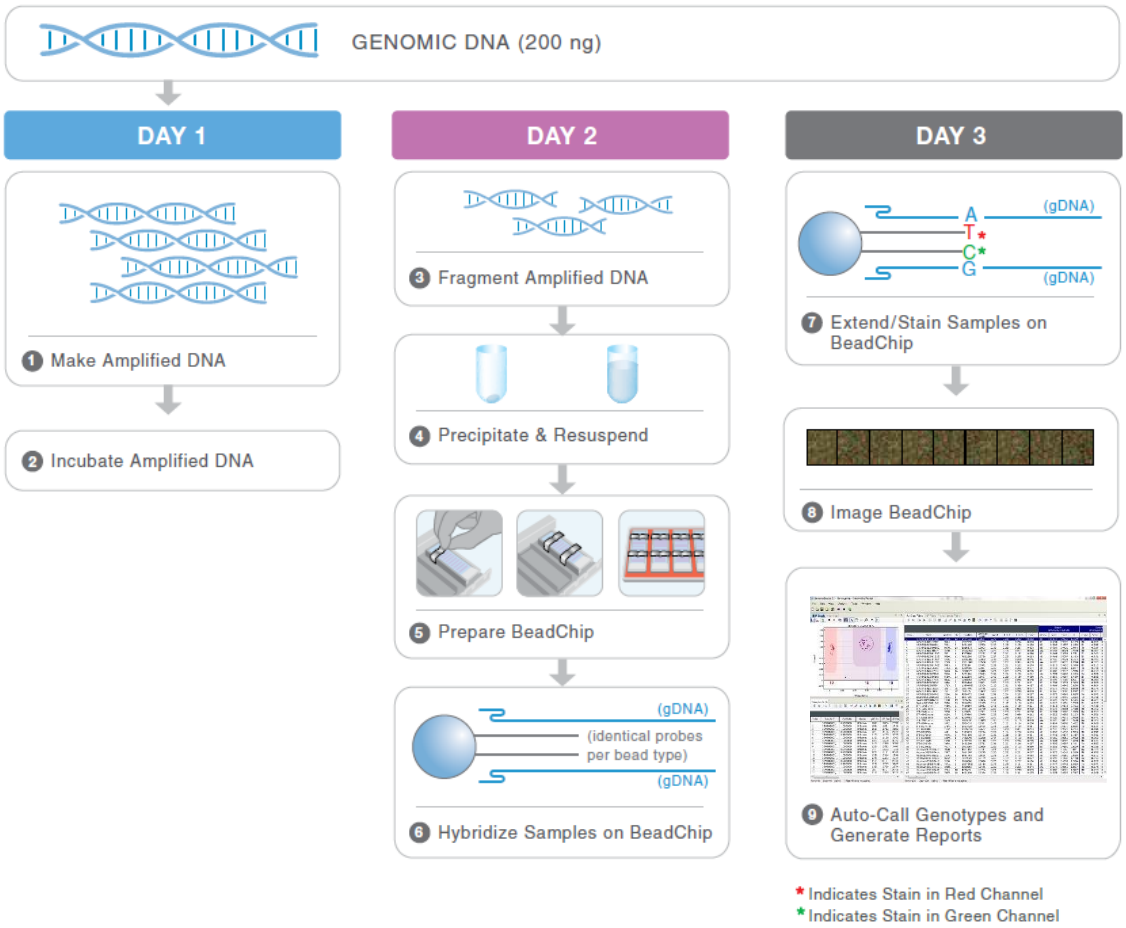

Figure 7. Infinium Assay Workflow. Genomic DNA is amplified overnight (1-2) and fragmented by a controlled enzymatic process (3). After alcohol precipitation and DNA resuspension (4), the BeadChip is assembled for hybridization (5). Denaturalized DNA is then loaded to the BeadChip and hybridized overnight to specific probes bond to the beads—one bead type corresponds to each allele per SNP per locus (6). After washing the chips, the primers hybridized to DNA on the BeadChip are extended and fluorescently stained (7). The fluorescence intensities are detected with the HiScan System, which excites the fluorophore-labelled nucleotides with two lasers and records the wavelength as high-resolution images (8), and analyzed with GenomeStudio software for genotype calling (9). All reagents and software used were from Illumina.
Image from: https://www.illumina.com/Documents/products/workflows/workflow_infinium_ii.p

We used a confidence interval of 95 % and selected all the SNPs with a *p*-value ≤ 10^{-4} for further analysis, although only those with a *p*-value ≤ 10^{-8} are considered statistically significant in GWAS. Due to the high density of SNPs in the HumanOmni5-Quad Beadchip, it is common to genotype several SNPs so close that are always inherited together—i.e. are in linkage disequilibrium—and provide redundant information; we evaluated the linkage disequilibrium between SNPs using Haploreg v. 3 [106]. We queried the top SNPs in the NCBI [47], ENSEMBL [46]

and WikiGene [48] databases to evaluate their chromosomic location and the function of the nearest gene.

3.4.3 Epistasis network analysis

We looked for germline genetic variants which affect metastasis through their genetic interactions by performing an epistasis network analysis with the Encore pipeline [44]. Encore is an open-source command-line tool for analysis of GWAS and other types of biological data with an increased power to detect important variants using epistasis network centrality; it discovers the importance of variants without a strong individual effect but whose relevance to the phenotype is derived from **gene-gene interactions** [107]. Encore includes data pre-processing with PLINK [43], machine learning-based modelling of the statistical epistasis network and network centrality ranking of variants based on gene-gene interactions and main effects.

Our gene epistasis network encodes the susceptibility to metastasis in our cohort. Thus, the links between genes are conditional to the development of metastasis and central genes contain most of the metastasis information through their interactions. To identify these central genes—which we termed "metastasis influence genes"—we used the community centrality, which measures the importance of a gene based on the number of network communities or groups of interconnected genes to which that gene belongs [53].

As our dataset contains millions of variants (ca. 4.3 million variants per sample), it was computationally prohibitive to calculate the interactions between all variants in a personal computer. For that reason we modelled the epistasis network in the Supercomputing and Bioinformatics Centre of the University of Malaga. All subsequent network analysis were performed with the iGraph package [108] and the R platform for statistical computing (v. 3.2.3). The workflow we used in this analysis, which implements the Encore pipeline, is depicted in Figure 8. We modelled the epistasis network for the susceptibility to metastasis and determined the most relevant genes—i.e. the metastasis influence genes—as those participating in more communities.

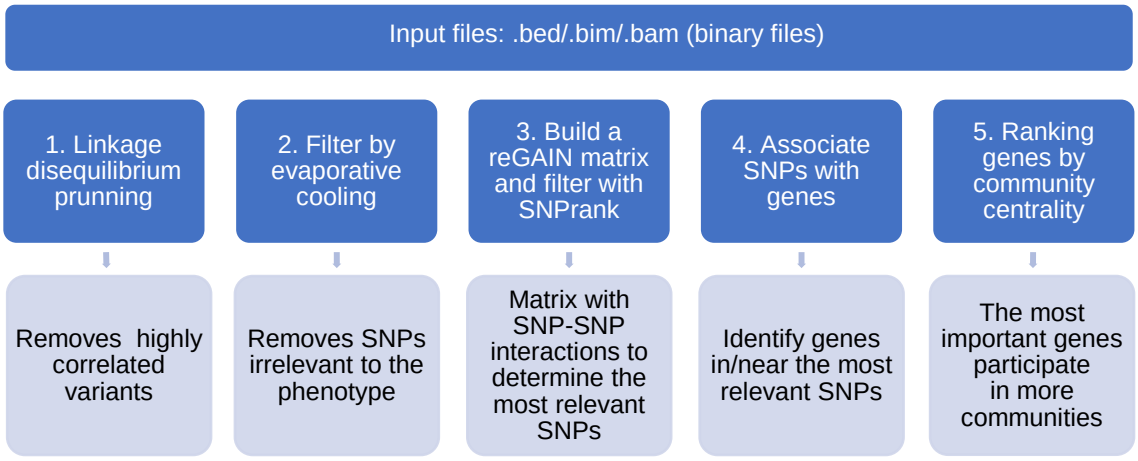

Figure 8. Encore pipeline. The Encore pipeline includes a variety of filtering methods to minimize the computational time. 1) The linkage disequilibrium pruning step removes highly correlated—i.e. low informative—SNPs. 2) Evaporative cooling is a machine learning method that integrates multiple importance scores while removing irrelevant genetic variants. In this step, we kept the 10 000 most relevant SNPs, which constitutes a significant reduction from the initial ~ 4.3 million. 3) After filtering, Encore calculates the pairwise interaction for the 10 000 SNPs with a generalized linear model and generates a symmetric matrix (reGAIN matrix) with the main effect of each variant along the diagonal. From that matrix, SNPs are ranked and filtered (with SNPrank) based on their importance to the phenotype (metastasis), which is measured based on their network centrality (i.e. interactions with relevant SNPs); we kept ~2 000 SNPs. 4) We obtained the genes names in or near which the most relevant SNPs lie with Biomart [134]. 5) Finally, we ranked the most relevant genes by their community centrality (using link communities); genes are important if they participate in more communities.

3.4.4 Regulation of gene expression and survival analysis

Germline genetic variants tend to associate with complex phenotypes (as the development of metastasis) through the regulation of gene expression [109, 110]; our metastasis influence genes could influence metastasis directly or by regulating the expression of other genes that influence metastasis. Since the development of metastasis is what determines survival of breast cancer patients, we used the distant metastasis-free survival (time from breast cancer diagnosis to the detection of distant metastasis or breast cancer death) to analyze the impact in metastasis of both metastasis influence genes and the genes under their regulation—i.e. their regulons.

The first step was to identify which of our metastasis influence genes regulate other genes—i.e. are transcription factors—and which are their regulons. For that, we obtained the regulatory circuit modelled after two breast carcinoma cell lines from the FANTOM5 project [57] and developed by Marbach and collaborators [111]

through the integration of transcription factor sequence motifs with promoter and enhancer activity data. This gene regulatory circuit reflects the cell-type specificity of the mechanisms underlying the effects of variants associated with breast cancer traits.

In parallel, we extracted data about gene expression and its association with distant metastasis-free survival from a public breast cancer dataset [45]. With this information, we divided the public dataset into two groups based on the gene expression pattern of our metastasis influence genes and their regulons and performed a Kaplan-Meier survival test for each of them with the SigCheck R package [112]. The resulting *p*-values revealed if the difference in survival between the two groups were significant, indicating whether the metastasis influence genes or regulons were associated with distant metastasis-free survival.

4. TRIPLE NEGATIVE BREAST CANCER HETEROGENEITY

Our goal should be to understand our differences.

James D. Watson (1928 -)

4.1 BACKGROUND AND OBJECTIVE OF THE CHAPTER

Triple negative breast cancers are homogeneous in their clinical definition (ER-, PR-, HER2-), which is widely used to manage the disease. However, they encompass a wide range of tumors with diverse biology and response to treatment, leading to diverse clinical outcomes [15]. They are also more aggressive and present worse disease-specific survival than other breast cancer subtypes [65, 70].

Chemotherapy remains the only systemic treatment option for triple negative breast cancer patients; unlike other breast cancer subtypes they do not benefit from targeted therapies since no appropriate molecular targets have been identified yet [5, 6]. Even though platinum-based chemotherapy has been incorporated into the neoadjuvant and metastatic settings, many triple negative patients do not benefit from this therapy as evidenced by the broad variation in response rate [7, 14].

Several classifications have been published to date in an attempt to dissect triple negative breast cancer heterogeneity [11, 12] although it is unclear whether they are predictive of treatment efficacy; there is a **need for proper validation of the value of triple negative breast cancer subtyping** regarding its molecular characteristics and response to different treatments. In this chapter we will evaluate the clinical relevance of Lehmann subtypes [11] by analyzing their pathological complete response to neoadjuvant chemotherapy as a surrogate of survival. We will also evaluate if any Lehmann subtype benefits from the incorporation of platinum salts to the standard chemotherapy, given that drugs with DNA-damaging mechanisms of action have proven to be effective in tumors with DNA repair defects, as is the case of Lehmann's basal-like tumors.

4.2 POPULATION OF STUDY AND METHODS' SUMMARY

4.2.1 Patients and samples

We performed a retrospective analysis on 125 patients with invasive triple negative breast cancer from two sources: 45 (36 %) patients from a randomized phase II trial (GEICAM/2006-03, ClinicalTrials.gov: NCT00432172) with a prospective collection of formalin-fixed paraffin-embedded (FFPE) tumor samples and associated clinical data and 80 (64 %) from four Spanish hospitals (Hospitales Universitarios Regional y Virgen de la Victoria, Málaga; Complejo Hospitalario de Jaén, Jaén; Hospital Universitario Gregorio Marañón, Madrid; and Hospital Costa del Sol, Marbella) for whom we had retrospectively collected FFPE tumor samples and clinical data. GEICAM/2006-03 clinical trial was the first to investigate whether adding carboplatin to a standard neoadjuvant chemotherapy combination (epirubicin/cyclophosphamide followed by docetaxel) increased the pathological complete response rate in basal-like breast cancer patients. It offers a unique source of patients treated with platinum salts and associated clinical data. Further information about this trial can be found elsewhere [14].

Patients included in this study met the following criteria: to be adult females (over 18 years old); to have histologically confirmed invasive triple negative breast cancer (ER-, PR-, HER2-); to have received neoadjuvant chemotherapy consisting of anthracyclines and/or taxanes with or without carboplatin; and to have already undergone surgery during which data on their pathological response was assessed and collected. According to the Miller & Payne criteria, we defined pathological complete response as the absence of invasive carcinoma in the breast and lymph nodes [113]. All analyses were performed in FFPE tumor tissue blocks from a diagnosis biopsy obtained prior neoadjuvant treatment.

4.2.2 Immunohistochemistry

FFPE tumor tissue blocks were sectioned for immunohistochemical subtyping and RNA purification. All samples were defined as triple negative by immunohistochemistry as ER-, PR- and HER2-. All GEICAM samples were also CK5/6+ and/or EGFR+ and therefore defined as *core basal* [114].

4.2.3 Triple negative breast cancer subtyping: intrinsic and Lehmann subtypes

We classified the samples in intrinsic subtypes by gene expression analysis with the nCounter Analysis System (Nanostring) and grouped them as basal-like or non-basal-like (luminal A, luminal B and HER2-enriched samples), as we expected to obtain a low percentage of non-basal-like samples. Samples from the GEICAM/2006-03 clinical trial were profiled using the PAM50 classifier and analyzed by means of a clinical algorithm for subtype prediction [16], discarding samples classified as normal-like. The retrospective collection of patients was classified according the Prosigna assay [17], which includes a proprietary algorithm based on the PAM50 gene signature [68].

We classified the samples in Lehmann subtypes (basal-like 1, BL1; basal-like 2, BL2; immunomodulatory, IM; mesenchymal, M; mesenchymal stem-like, MSL; luminal androgen receptor, LAR; and unstable, UNS) [11] by whole-transcriptome analysis using the Affymetrix GeneChip Platform® and the online classification tool TNBCtype [18].

4.2.4 Statistical analysis

We used Pearson's chi-square test to perform contingency table and goodness-of-fit tests, and Fisher's exact test when any of expected values in cells were less than 5. Student's t-test was used to test the null hypothesis under the assumption that the two populations have equal means. We performed a logistic regression multivariate analysis using a stepwise forward and backward selection procedure to select the most important variables of the model based on the Akaike information criterion (AIC). All statistical analyses were conducted in the statistical computing environment
R (v. 3.3.1).

As Prat and colleagues underscored that Lehmann subtyping often ignores samples contaminated with normal breast tissue (which are mostly classified as MSL) we performed all analyses with and without the MSL group [19] to avoid missing relevant associations between clinicopathological variables and Lehmann subtyping.

4.3 RESULTS

4.3.1 Characteristics of the population

Patients' clinicopathological characteristics in the global population as well as in the clinical trial and the retrospective cohort are shown in Table 2. Overall, patients were mainly pre-menopausal (56 %), with high histological grade (63 % of the patients had grade 3) and high proliferation (66.4 % of patients with Ki-67 > 50 %). When comparing the cohorts, almost all patients from the clinical trial were lymph node positive (96 % vs. 39 % in the retrospective cohort) and had a higher percentage of patients treated with carboplatin (56 % vs. 11 %, respectively). The pathological complete response rate in the global population was 37 %, although it was unevenly distributed across Lehmann subtypes (Table 3).

4.3.2 Clinicopathological characteristic of triple negative breast cancer subtypes

We collected RNA with sufficient quantity and quality for whole-transcriptome analysis—and therefore Lehmann subtyping—from 119 (95 %) out of the 125 patients with evaluable pathological complete response after neoadjuvant chemotherapy: 102 (86 %) were classified in a stable subtype and 17 (14 %) were unstable (UNS) (Table 2). A detailed description of the clinicopathological characteristics of every Lehmann subtype is shown in Table 3. Ki-67 index (dichotomized using a cut-off > 50 %) was significantly associated with Lehmann subtyping. **We observed that BL1 samples had the highest proliferation rates** (88.2 % of BL1 patients vs. 63.7 % of patients with other subtypes had Ki-67 > 50 %, *p*-value = 0.02) and **LAR the lowest** (71 % of LAR patients vs. 27 % of patients with other subtypes had Ki-67 ≤ 50 %, *p*-value = 0.002).

Table 2. Patients' characteristics, N (%) and *p*-value of the comparison of the two cohorts (clinical trial GEICAM/2006-03 and retrospective cohort)

Characteristics	Global (N = 125)	Clinical trial (N = 45)	Retrospective (N = 80)	*p*-value
Age at diagnosis (years)				
Median	48	49	47	
Range	29 to 76	29 to 76	34 to 75	0.7954
Tumor size (cm)				
< 2	9 (7.2 %)	1 (2 %)	8 (10 %)	
2-5	90 (72 %)	34 (76 %)	56 (70 %)	
> 5	24 (19.2 %)	10 (22 %)	14 (18 %)	
Unknown	2 (1.6 %)	0	2 (3 %)	0.2774
Lymph node status				
Positive	74 (59.2 %)	43 (96 %)	31 (39 %)	
Negative	38 (30.4 %)	2 (4 %)	36 (45 %)	
Unknown	13 (10.4 %)	0	13 (16 %)	< 0.001*
Histological grade				
1	3 (2.4 %)	2 (4 %)	1 (1 %)	
2	28 (22.4 %)	14 (31 %)	14 (18 %)	
3	79 (63.2 %)	29 (64 %)	50 (63 %)	
NA	15 (12 %)	0	15 (19 %)	0.2984
Ki-67 (%)				
≤ 50	39 (31.2 %)	10 (22 %)	29 (36 %)	
> 50	83 (66.4 %)	32 (71 %)	51 (64 %)	
Unknown	3 (2.4 %)	3 (7 %)	0	0.2318
Intrinsic subtypes				
Basal-like	104 (83.2 %)	33 (73 %)	71 (89 %)	
Non-basal-like	6 (4.8 %)	0	6 (8 %)	
Unknown	15 (12 %)	12 (27 %)	3 (4 %)	0.1756
Lehmann subtypes				
BL1	17 (13.6 %)	6 (13 %)	11 (14 %)	
BL2	15 (12 %)	3 (7 %)	12 (15 %)	
M	22 (17.6 %)	5 (11 %)	17 (21 %)	
MSL	9 (7.2 %)	2 (4 %)	7 (9 %)	
IM	25 (20 %)	10 (22 %)	15 (19 %)	
LAR	14 (11.2 %)	5 (11 %)	9 (11 %)	
UNS	17 (13.6 %)	8 (18 %)	9 (11 %)	
Unknown	6 (4.8 %)	6 (13 %)	0	0.5869
Treatment				
A+T	91 (72.8 %)	20 (44 %)	71 (89 %)	
A and/or T + Cb	34 (27.2 %)	25 (56 %)	9 (11 %)	< 0.001*
pCR				
Yes	46 (36.8 %)	14 (31 %)	32 (40 %)	
No	79 (63.2 %)	31 (69 %)	48 (60 %)	0.4261

*Abbreviations: BL1, basal-like 1; BL2, basal-like 2; M, mesenchymal; MSL, mesenchymal stem-like; IM, immunomodulatory; LAR, luminal androgen receptor; UNS, unstable; A, anthracyclines; T, taxanes; Cb, carboplatin; pCR, pathological complete response. * p-value ≤ 0.05*

Table 3. Patients' characteristics by Lehmann subtype, N (%) and *p*-values of the comparison between all subtypes and all subtypes except MSL

Characteristics	BL1 (N = 17)	BL2 (N = 15)	M (N = 22)	MSL (N = 9)	IM (N = 25)	LAR (N = 14)	UNS (N = 17)	*p*-value all subtypes	*p*-value excluding MSL
Age (years)									
< 50	10 (58.8 %)	9 (60 %)	12 (54.5 %)	3 (33.3 %)	20 (80 %)	6 (42.9 %)	10 (58.8 %)	0.1791	0.2686
≥ 50	7 (41.2 %)	6 (40 %)	10 (45.5 %)	6 (66.7 %)	5 (20 %)	8 (57.1 %)	7 (41.2 %)		
Tumor size (cm)									
< 2	3 (17.6 %)	0	3 (13.6 %)	0	2 (8 %)	0	1 (5.9 %)	0.5033	0.3979
2-5	13 (76.5 %)	13 (86.7 %)	16 (72.7 %)	7 (77.8 %)	17 (68 %)	9 (64.3 %)	11 (64.7 %)		
> 5	1 (5.9 %)	2 (13.3 %)	2 (9.1 %)	2 (22.2 %)	6 (24 %)	4 (28.6 %)	5 (29.4 %)		
Unknown	0	0	1 (4.6 %)	0	0	1 (7.1 %)	0		
Lymph node status									
Positive	8 (47.1 %)	8 (53.3 %)	12 (54.6 %)	5 (55.6 %)	16 (64 %)	10 (71.4 %)	10 (58.8 %)	0.7903	0.7449
Negative	8 (47.1 %)	4 (26.7 %)	5 (22.7 %)	4 (44.4 %)	7 (28 %)	3 (21.4 %)	6 (35.3 %)		
Unknown	1 (5.8 %)	3 (20 %)	5 (22.7 %)	0	2 (8 %)	1 (7.2 %)	1 (5.9 %)		
Histological grade									
1	0	1 (6.7 %)	0	0	0	1 (7.1 %)	1 (5.9 %)	0.1438	0.0912
2	5 (29.4 %)	0	7 (31.8 %)	2 (22.2 %)	5 (20 %)	5 (35.7 %)	3 (17.6 %)		
3	10 (58.8 %)	13 (86.6 %)	11 (50 %)	4 (44.5 %)	18 (72 %)	7 (50.1 %)	11 (64.7 %)		
Unknown	2 (11.8 %)	1 (6.7 %)	4 (18.2 %)	3 (33.3 %)	2 (8 %)	1 (7.1 %)	2 (11.8 %)		
Ki-67 (%)									
≤ 50	1 (5.9 %)	4 (26.7 %)	9 (40.9 %)	5 (55.6 %)	4 (16 %)	10 (71.4 %)	5 (29.4 %)	0.0015*	< 0.001*
> 50	15 (88.2 %)	11 (73.3 %)	13 (59.1 %)	4 (44.4 %)	21 (84 %)	4 (28.6 %)	12 (70.6 %)		
Unknown	1 (5.9 %)	0	0	0	0	0	0		
Intrinsic subtypes									
Basal-like	16 (94.1 %)	15 (100 %)	20 (90.9 %)	8 (88.9 %)	25 (100 %)	5 (35.7 %)	12 (70.6 %)	< 0.001*	< 0.001*
Non-basal-like	0	0	0	0	0	5 (35.7 %)	1 (5.9 %)		
Unknown	1 (5.9 %)	0	2 (9.1 %)	1 (11.1 %)	0	4 (28.6 %)	4 (23.5 %)		
Treatment									
A+T	12 (70.6 %)	12 (80 %)	18 (81.8 %)	7 (77.8 %)	17 (68 %)	10 (71.4 %)	12 (70.6 %)	0.9447	0.8916
A and/or T+Cb	5 (29.4 %)	3 (20 %)	4 (18.2 %)	2 (22.2 %)	8 (32 %)	4 (28.6 %)	5 (29.4 %)		
pCR									
Yes	8 (47.1 %)	7 (46.7 %)	9 (40.9 %)	3 (33.3 %)	10 (40 %)	2 (14.3 %)	7 (41.2 %)	0.5714	0.4539
No	9 (52.9 %)	8 (53.3 %)	13 (59.1 %)	6 (66.7 %)	15 (60 %)	12 (85.7 %)	10 (58.8 %)		

*Abbreviations: BL1, basal-like 1; BL2, basal-like 2; M, mesenchymal; MSL, mesenchymal stem-like; IM, immunomodulatory; LAR, luminal androgen receptor; UNS, unstable; NA, not available A, anthracyclines; T, taxanes; Cb, carboplatin; pCR, pathological complete response. * p-value ≤ 0.05*

We had enough material to further classify 110 (88 %) of the samples into intrinsic subtypes: 104 (94.5 %) were basal-like and 6 (5.4 %) non-basal-like (Table 2). This low percentage of non-basal-like samples prevented us from extracting conclusions of the clinicopathological characteristics and the pathological complete response based on the intrinsic subtypes. As 36 % of the tumors were from the GEICAM/2006-03 clinical trial which eligibility criteria included the *core basal* definition (triple negative definition plus CK5/6+ and/or EGFR+), we investigated if this could have caused an overrepresentation of basal-like tumors compared to the triple negative-only definition. It was not the case as we observed a comparable proportion of basal-like samples when excluding this subset of *core basal* samples determined by immunohistochemistry (94.5 % of basal-like samples in the global cohort vs. 92.2 % in the triple negative-only cohort, *p*-value = 0.56).

We found a strong concordance between Lehmann and intrinsic subtyping (Table 3) mainly **because the only stable group with non-basal-like samples was the LAR subtype**; from the 6 non-basal-like samples 5 were classified as LAR and 1 as unstable. The distribution of Lehmann subtypes into basal-like and non-basal-like intrinsic subtypes is shown in Figure 9.

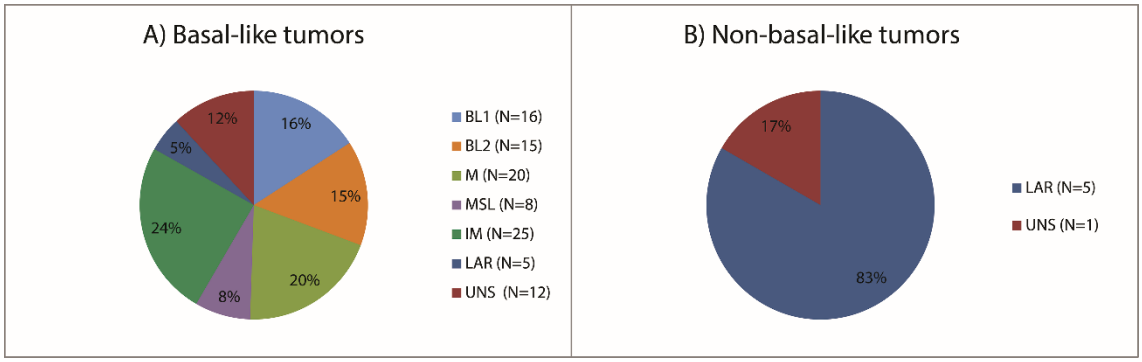

Figure 9. Distribution of Lehmann subtypes within intrinsic subtypes. A) Distribution of Lehmann subtypes in molecular basal-like tumors. B) Distribution of Lehmann subtypes in molecular non-basal-like tumors. *Abbreviations: BL1, basal-like 1; BL2, basal-like 2; M, mesenchymal; MSL, mesenchymal stem-like; IM, immunomodulatory; LAR, luminal androgen receptor; UNS, unstable*

The LAR samples classified as non-basal-like included 4 HER2-enriched and 1 luminal A. We had enough tumor tissue to successfully determine the expression of the androgen receptor in 4 out of these 5 LAR non-basal-like samples; all

overexpressed the androgen receptor and were histopathologically consistent with apocrine carcinomas.

4.3.3 Correlation with pathological response to neoadjuvant treatment

We analyzed the association between clinicopathological variables and the pathological complete response to the different neoadjuvant treatments. In a bivariate analysis, we found that the expression rate of Ki-67 (≤ 50 % vs. > 50 %, *p*-value = 0.037) and clinical tumor size (≤ 2 cm vs. > 2 cm, *p*-value = 0.024) were the only variables associated with pathological complete response. In the multivariate analysis, only clinical tumor size remained associated with pathological complete response (*p*-value = 0.002).

As aforementioned in the methodology of this chapter (section 4.2 "Population of study and methods' summary"), we analyzed the differences among Lehmann subtypes with and without the MSL subtype, as this subtype is supposed to contain normal breast tissue. We found no statistically significant association between Lehmann subtypes and pathological complete response to overall treatment (*p*-value = 0.571) in spite of the wide range of complete responses observed (from 47.1 % in BL1 to 14.3 % in LAR). **LAR patients were the most chemoresistant** (14.3 % of LAR had pathological complete response vs. 41.9 % of the remaining subtypes combined, *p*-value = 0.077) and when we excluded MSL samples from the analysis, this difference in response appeared to be more pronounced (14.3 % of LAR had pathological complete response vs. 42.7 % of the remaining groups except MSL, *p*-value = 0.046).

Figure 10 shows the pathological complete response rates to the different treatments received by Lehmann subtypes. We found no differences in the global population when treated with and without carboplatin (40.9 % vs. 32.3 % of patients achieved a pathological complete response, respectively; *p*-value = 0.521). We neither observed a difference in pathological complete response rates to standard chemotherapy (sequential anthracyclines plus taxanes) by Lehmann subtype (*p*-value = 0.556). When comparing the rates in patients treated with carboplatin in

each of the Lehmann subtypes, we observed that **BL1 patients were the most benefited of the addition of carboplatin** (80 % of BL1 vs. 23 % of the remaining groups had a pathological complete response, *p*-value = 0.027), while no differences were found in the remaining subtypes.

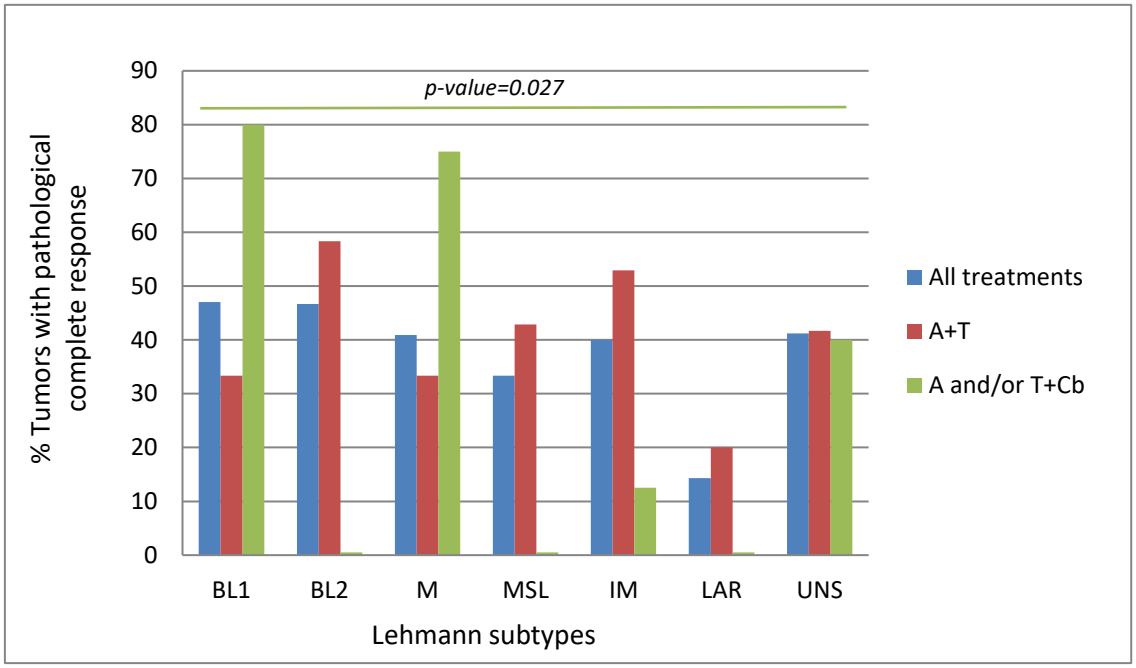

Figure 10. Percentage of tumors from each Lehmann subtype with pathological complete response to the different treatments. The green horizontal line represents the comparison of the percentage of pathological complete response to anthracyclines and/or taxanes plus carboplatin (A and/or T+Cb) of BL1 versus the rest of patients and its associated *p*-value. The number of patients receiving every treatment. within each Lehmann subtype can be found at Table 3. *Abbreviations: A, anthracyclines; T, taxanes; Cb, carboplatin; BL1, basal-like 1; BL2, basal-like 2; M, mesenchymal; MSL, mesenchymal stem-like, IM, immunomodulatory; LAR, luminal androgen receptor; UNS, unstable*

4.4 Discussion

Triple negative breast cancer is a commonly used umbrella term for a histologic group of tumors that are vastly heterogeneous from a molecular perspective, including a wide range of entities differing in biology and response to chemotherapy and targeted therapies. The different classifications of triple negative breast cancer result in inconsistent definitions of disease subgroups and their corresponding clinical outcomes. Only the subtypes termed as LAR appear to be consistent across all the studies, though it is unclear whether these classifications are predictive of treatment efficacy.

In this Thesis Project, we analyzed a combined dataset of triple negative breast cancer patients treated with anthracyclines and/or taxanes +/- carboplatin in the neoadjuvant setting. First, we classified them into the Lehmann subtypes and evaluated their clinicopathological characteristics. Then, we explored the chemosensitivity of these subtypes to the different neoadjuvant treatments administrated. To our knowledge, this is the first study evaluating the prognostic role of Lehmann triple negative breast cancer subtypes in the neoadjuvant setting of patients treated with and without platinum salts.

Based on our results, **LAR was the least proliferative tumor subtype and the most chemoresistant** one. Despite its significantly lower response to neoadjuvant chemotherapy in comparison to the other subtypes, LAR has been associated with a favorable prognosis [20, 21]. This may be, in part, because LAR is the only subtype including non-basal-like tumors [22, 23] which could also explain the low proliferation and pathological complete response rates observed in our study. As expected, most of these samples expressed androgen receptor and were histologically consistent with apocrine carcinomas [24]. Recently published early phase II clinical trials results suggest that antiandrogen therapy may target the androgen receptor-positive subset of triple negative breast cancers [25–27].

In our study, **BL1 was the most proliferative subtype and appeared to be particularly sensitive to chemotherapy regimens including a platinum agent**. This is of major significance because in the past few decades there has been considerable interest in platinum salts as treatment for triple negative breast cancer given that homologous recombination deficiency sensitizes tumor cells to these agents inducing cell death. Although results from phase II studies involving unselected triple negative breast cancer patients in the neoadjuvant setting have been conflicting [28, 29], triple negative tumors harboring a high homologous recombination deficit score seem to benefit from platinum-based therapies [30].

The results of this study should be interpreted in the context of its limitations. First, the actual number of samples analyzed under each Lehmann subtype is limiting; second, we used paraffined samples for gene expression analysis, which could present

differences when compared to analysis performed in fresh/frozen tissue; and third, patients did not receive homogeneous neoadjuvant treatments, although all received anthracyclines and/or taxanes +/- carboplatin regimes.

Our results confirm the high genetic diversity within triple negative breast tumors, although they should be considered a spectrum of tumors with varying clinical and molecular characteristics rather than falling into discreet categories. On one extreme of this spectrum we would have BL1 tumors, a highly proliferative subtype with its likely deficiencies in homologous recombination that lead to a high pathological complete response to platinum-based therapies. On the other extreme lies LAR, a tumor subtype characterized by low proliferation and low response to standard chemotherapy. In between, we would find tumors that cannot be classified into any subgroup further than their standard definition of triple negative by immunohistochemistry. These extreme patients may benefit from a better clinical management and a more personalized treatment than when buried in the wide term of "triple negative".

5. MALE BREAST CANCER HETEROGENEITY

Knowledge and wonder are the dyad of our worthy lives as intellectual beings.

Stephen Jay Gould (1941 – 2002)

5.1 BACKGROUND AND OBJECTIVE OF THE CHAPTER

Male breast cancer is a rare disease that is still poorly understood. Consequently, its biological knowledge and medical treatment is largely based either on female breast cancer studies or in studies with small number of patients, representing a special challenge within breast cancer management.

Recent discoveries in female breast cancer have allowed us to move from a purely anatomical and pathological classification to a new classification based on molecular criteria. However, as there are no data on male breast cancer molecular subtyping based on prognostic gene signatures, it still relies on the traditional clinicopathological features for its management and treatment.

There is therefore a need for the classification of male breast cancer in validated subtypes that reflect male breast cancer biology, predict clinical outcome and guide treatment decision. The aim of this chapter is to **classify male breast cancer in intrinsic subtypes based on the PAM50 signature and investigate their correlation with immunohistochemical surrogates and clinical outcome.**

5.2 POPULATION OF STUDY AND METHODS' SUMMARY

5.2.1 Patients and samples

We performed a retrospective analysis on 67 male breast cancer patients from four different Spanish pathology laboratories (Hospitales Universitarios Regional y Virgen de la Victoria, Málaga; Complejo Hospitalario de Jaén, Jaén; Hospital de la Serranía, Ronda; and Hospital Costa del Sol, Marbella). Medical record's data were reviewed to collect clinicopathological data and follow-up. Analyses were performed on FFPE tumor samples obtained at the time of surgery.

5.2.2 Immunohistochemical subtypes

We performed immunohistochemical staining and assessment in tissue microarrays for triplicate for each antibody. We obtained cores sized 0.6 mm from each tumor to build the tissue microarrays in a manual tissue arrayer (Beecher Instruments) and performed immunohistochemical staining on 5 μm sections from the tissue microarray blocks (Figure 11). The immunohistochemical scoring was conducted by two experienced pathologists independently and blinded to other features.

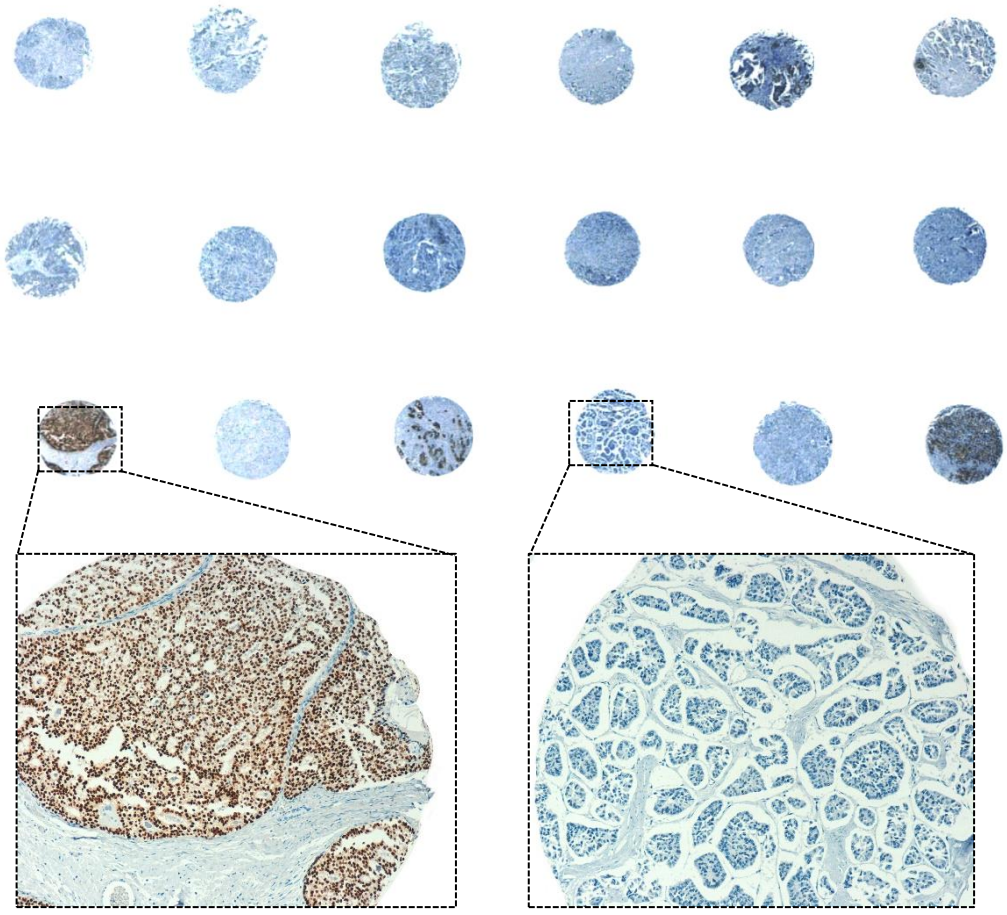

Figure 11. Tissue microarray with estrogen receptor immunostaining. Above, overall tissue microarray (magnification 2x). Below, an example of positive (left) and negative (right) estrogen receptor immunostaining (magnification 10x). *Images taken in a Leica DMD108 microscope.*

Based on a validated 6-marker immunohistochemical panel, samples were classified as luminal A (ER+ and/or PR+, HER2-, Ki-67 < 14 %), luminal B (ER+ and/or PR+, HER2-, Ki-67 ≥ 14 %), HER2-positive (HER2+, independently of ER and PR status), basal-like (ER-, PR-, HER2-, EGFR+ and/or CK5/6+) and non-basal triple negative (ER-, PR-, HER-2, EGFR-, CK5/6-).

5.2.3 Intrinsic subtypes: PAM50 signature

We classified the samples in intrinsic subtypes (luminal A, luminal B, HER2-enriched or basal-like) by gene expression analysis with an nCounter Analysis System (Nanostring) using the PAM50 assay according to the manufacturer's guidelines. Nanostring Technologies team analyzed the data with the Prosigna algorithm and provided us the classification into intrinsic subtypes.

5.2.4 Statistical analysis

We performed the data analysis in the statistical computing environment R (v. 3.3.0) and the packages Survival (v. 2.38) and gmodels (v. 2.16.2), with a general descriptive analysis of the variables included in the study. Qualitative variables were described according to absolute and relative frequency distributions. Quantitative variables were evaluated using central trend measures (mean and median) and scatter measures (standard deviation). Both Pearson's chi-square and Fisher's exact tests were performed for testing the null hypothesis of independence of variables in contingency tables. Disease-free survival was defined as the time from diagnosis until disease progression or death by any cause on the date of the last follow-up. Overall survival was defined as the time from diagnosis until death by any cause. We used the Kaplan–Meier method to estimate disease-free survival and overall survival curves. The survival distributions for the different values of the patients' characteristics were compared using the log-rank test. Cox proportional hazards models for disease-free survival and overall survival were also fitted, using the Efron approximation for tie handling.

5.3 RESULTS

5.3.1 Characteristics of the population

All male breast cancer patients were successfully classified by both the validated 6-marker immunohistochemical panel and PAM50 signature. Patient's clinicopathological characteristics are shown in Table 4.

Table 4. Characteristics of male breast cancer patients cohort

Characteristic	**Number (%)**
Age at diagnosis (years)	
Median	64
Range	23 to 92
Tumor size (cm)	
≤ 2	31 (46 %)
> 2	36 (54 %)
Lymph node status	
Positive	28 (42 %)
Negative	39 (58 %)
Histology	
No special type (ductal NOS)	60 (90 %)
Other	7 (10 %)
Histological grade	
1	15 (23 %)
2	30 (45 %)
3	19 (28 %)
Unknown	3 (4 %)
Immunohistochemical subtype	
Luminal A	29 (44 %)
Luminal B	34 (51 %)
Non-basal triple negative	2 (3 %)
Basal-like	1 (1 %)
HER2-positive	1 (1 %)
PAM50 subtype	
Luminal A	20 (30 %)
Luminal B	40 (60 %)
HER2-enriched	7 (10 %)
Basal-like	0

Most cases were ER+ (96%) and/or PR+ (84%) and three patients (5%) had a tumor that was both ER- and PR-. The only patient displaying a HER2+ tumor by immunohistochemistry was confirmed to be positive by chromogenic *in situ* hybridization (CISH) and was treated with anti-HER2 therapy (trastuzumab) in the adjuvant setting (after surgery).

All patients were treated surgically with either mastectomy—total removal of the breast (85 % of the patients)—or lumpectomy—breast-conserving surgery (15 % of the patients). Patients were treated with radiotherapy (37 %) hormonal therapy (85 %) and/or chemotherapy (55 %). Only three patients (5 %) presented advanced disease at diagnosis. Twelve (18 %) patients had a family history of breast cancer and only three (5 %) a personal history of gynecomastia. Five patients (7 %) developed a second cancer after breast cancer: 2 prostate cancers, 1 bladder cancer and 2 melanomas.

5.3.2 Male breast cancer subtyping: immunohistochemical panel and PAM50 signature

None of the samples were classified as basal-like according PAM50 (Table 4). Those triple negative samples (basal and non-basal) by immunohistochemistry were 1 luminal A, 1 luminal B and 1 HER2-enriched by PAM50.

We found a strong correlation between immunohistochemistry and PAM50 subtyping in our cohort (p-value = 0.018) when comparing patients grouped as luminal A, luminal B, HER2-positive/HER2-enriched or triple negative/basal-like by both classifications. The distribution of PAM50 subtypes within luminal A and luminal B defined by immunohistochemistry is displayed in Figure 12. We observe that, despite the strong statistical association between immunohistochemical and PAM50 subtyping, more than half of luminal A patients by immunohistochemistry and around 20 % of luminal B are classified in a different group by PAM50. Only one of the patients classified as HER2-enriched by PAM50 was HER2+ by immunohistochemistry and chromogenic *in situ* hybridization (CISH). Of note, this HER2-enriched/HER2+ patient was also ER+ and PR+, and could be considered as Luminal B-like (HER2-positive) depending on the classification criteria <used [33,

74]. In order to verify if there were any HER2 subclones missed by the placement of small-sized cores into the tissue microarrays, we assessed the HER2 status of the 6 HER2-enriched/ HER2- tumor samples in their original tissue blocks and confirmed that all were negative.

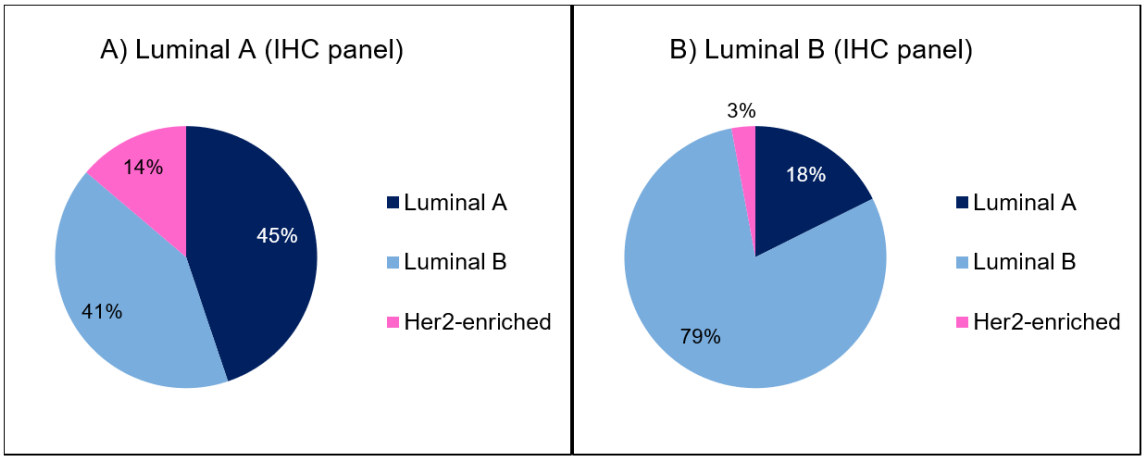

Figure 12. Distribution of PAM50 subtypes within immunohistochemical (IHC) subtypes. A) PAM50 subtypes within immunohistochemical luminal A tumors (defined as ER+ and/or PR+, HER2-, Ki-67 < 14 %). B) PAM50 subtypes within immunohistochemical luminal B tumors (defined as ER+ and/or PR+, HER2-, Ki-67 ≥ 14 %

We found no significant differences between the clinicopathological features of our cohort of PAM50 luminal A and luminal B tumors (Table 5), only a trend toward a higher percentage of luminal B subtype tumors that were poorly differentiated (grade 3) (38 % vs. 10 %, *p*-value = 0.08) compared with luminal A tumors by PAM50.

Table 5. Male breast cancer patients' characteristics by PAM50 Luminal A or Luminal B subtype, N (%) and *p*-values of the comparison between subtypes

Characteristics	**Luminal A (N = 20)**	**Luminal B (N = 40)**	***p*-value**
Age at diagnosis (years)			
< 65	11 (55 %)	22 (55 %)	1
≥ 65	9 (45 %)	18 (45 %)	
Tumor size (cm)			
≤ 2	10 (50 %)	17 (42 %)	0.58
> 2	10 (50 %)	23 (58 %)	
Lymph node status			
Positive	17 (85 %)	37 (93 %)	0.36
Negative	13 (65 %)	21 (53 %)	
Histology	7 (35 %)	19 (47 %)	
No special type			0.36
Other	3 (15 %)	3 (7 %)	
Histological grade			
1-2	17 (85 %)	24 (60 %)	
3	2 (10 %)	15 (38 %)	0.08
Unknown	1 (5 %)	1 (2 %)	

5.3.3 Survival analysis

Patients had a median follow up of 128 months (range 93 to 162 months). Seventeen of them (25 %) had recurrence: 2 had local relapse and 15 distant metastases. Twenty patients (30 %) died during follow up, 12 (18 %) from progression disease, and 8 (12 %) from other causes unrelated to the tumor. We found no significant differences between luminal A and luminal B by PAM50 in disease-free survival (OR = 0.29 [0.07 to 3.11], *p*-value = 0.291) or overall survival (OR = 0.46 [0.15 to 4.91], *p*-value = 0.8). Similarly, we found no differences between luminal A and luminal B based on the immunohistochemical panel in disease-free survival or overall survival (data not shown). However, when we analyzed the differences in overall survival between PAM50 HER2-enriched and non-HER2-enriched patients (Figure 13) **we found a significant worse overall survival in HER2-enriched tumors** (median of 71 vs. 128 months, respectively; OR 2.59 = [0.47 to 14.29], *p*-value = 0.046).

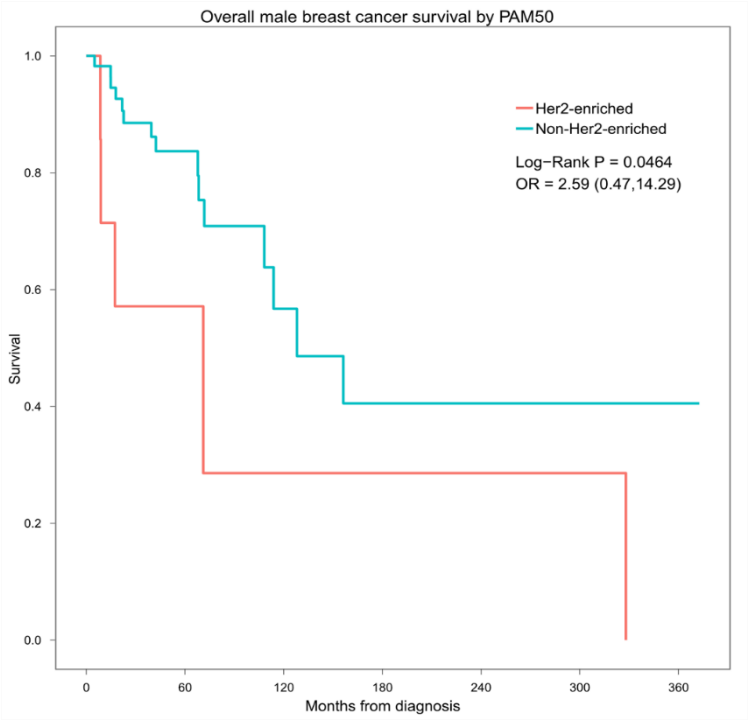

Figure 13. Overall survival of PAM50 HER2-enriched vs. non-HER2-enriched tumors

5.4 Discussion

Immunohistochemical markers and molecular signatures identify distinct subtypes of female breast cancer with different clinical outcomes and responses to systemic treatment. However, there are currently no data on validated molecular subtypes in male breast cancer that reflect their unique characteristics, leading to a less personalized treatment than in females. In this chapter, we classified a set of male breast tumors based on a validated 6-marker immunohistochemical panel and into intrinsic subtypes with the PAM50 signature, and analyzed their clinicopathological features and survival.

As expected, we observed several **unique characteristics of male breast cancer compared with female breast cancer**, including higher rates of ER positivity (96 %) and PR positivity (84 %), lower HER2 positivity (1 %), older age at presentation (median age 64 years), and higher proportion of nodal disease (43 %), that were in line with results from previous male breast cancer studies [32], although we lack a direct comparison between both types of tumors. Like in postmenopausal

women, there are low levels of circulating estrogens in males and most of the estrogen is synthesized in the peripheral tissue. This locally synthesized estrogen is believed to play an important role in the pathogenesis and development of hormone-dependent breast cancers, which could explain the overwhelming predominance of hormone-receptor positive cancers in males [115]. However, the true cause is still unknown. Although most men with breast cancer have no identifiable risk factors, several have been associated with an excess of estrogen or lack of androgen (chronic liver diseases, cryptorchidism, Klinefelter´s syndrome). Despite most male breast cancers are hormone receptor-positive, not all of them are similar and they may behave differently to hormone receptor-positive female breast cancers due to a gender associated landscape in hormone receptor pathways. Callari et al. [116] reported a different pattern of steroid receptors' gene expression between ER+ male and female breast patients with similar clinicopathological features, indicating that there may be many differences between the biology of male and female breast cancer.

This is the first study that classifies male breast tumors in PAM50 molecular subtypes and correlates them to immunohistochemical surrogates. Our results confirm that **male breast cancer is mainly a genomic luminal disease** with a strong correlation between PAM50 and a validated 6-marker immunohistochemical panel. Most of the samples were luminal B followed by luminal A. Luminal B tumors are characterized by a more aggressive biology, higher proliferation and less endocrine-responsiveness than the luminal A subtype in female breast cancer. This predominance of luminal B tumors was reproduced by the panel of immunohistochemical surrogates. We found that the definition of luminal B tumors based on Ki-67 status (applied in this study) overcomes the definition based on HER2 positivity: with this last definition our cohort would be mainly luminal A (94 %), not reproducing the subtyping based on gene-expression profiling. We observed some discordances between immunohistochemical and PAM50 subtyping despite they were significantly associated. More than 50 % of luminal A patients by immunohistochemistry were classified in a different group by PAM50 although we do not know if this may have an impact in the treatment of male breast cancer patients. Nevertheless, **we did not find significant differences in the**

clinicopathological features or in the outcome between luminal A and luminal B subtypes. Most of the tumors in this cohort were diagnosed in elderly patients, some of whom died from causes unrelated to the tumor, and relapses can occur even a long time after the initial diagnosis, which may explain at least partially why we did not observe differences in disease-free survival and overall survival between luminal A and luminal B subtypes.

We found a set of **HER2-enriched patients by PAM50 that were HER2- by immunohistochemistry; they did not receive any targeted therapy and had worse outcome than luminal tumors.** The incorporation of anti-HER2 therapies have changed the natural history of HER2+ female breast cancer from a historically aggressive disease to a subtype with similar survival than HER2- tumors [34]. Several studies suggest that females with HER2-enriched tumors have the highest pathological complete response compared to HER2+ or any other subtype and might benefit the most from anti-HER2 therapies [35, 36]. Prat et al. reported a benefit from an anti-HER2 agent (lapatinib) in a group of females with HER2-/HER2-enriched metastatic breast cancer in an unplanned retrospective analysis from the EGF30008 phase 3 clinical trial [117].

Although the results from our study are limited by the small number of patients analyzed and the study's retrospective nature, we consider that it provides useful information on the biology of male breast cancer patients and their outcome. In contrast with female breast cancer, in which a third of the cases are non-luminal, our findings suggest that male breast cancer is mainly a genomic luminal disease based on the PAM50 signature. More research is needed to identify the reasons for the scarcity of non-luminal breast tumors in males, the predominance of the luminal B over luminal A subtypes and the clinical significance of these genomic subtypes. In addition, we found a proportion HER2-/HER2-enriched patients untreated with any anti-HER2 therapy that had a worse outcome than luminal patients. Identifying these patients and the subsequent treatment with anti-HER2 therapy could change the natural history of their disease.

6. GENETIC BACKGROUND HETEROGENEITY

Somewhere, something incredible is waiting to be known

Carl Sagan (1934 – 1996)

6.1 BACKGROUND AND OBJECTIVE OF THE CHAPTER

The overwhelming majority of cancer related deaths are caused by the metastatic dissemination of the disease, an extraordinarily complex process that is still poorly understood [3]. Many studies suggest that metastasis is a multistep process that requires not only cells from the tumor but also from the tumor microenvironment and metastasis target tissues; the genetic makeup of the host may influence metastasis alone or in conjunction with tumor characteristics [3, 37].

However, there has been limited success in determining which are the germline genetic variants that influence metastasis, mainly because of the lack of power of traditional genome-wide association studies (GWAS) to detect relevant variants underlying complex phenotypes as the metastatic disease [38, 39]. To overcome this limitation, we reduced the phenotype heterogeneity by selecting patients at the extreme distribution of metastasis risk but with an atypical behavior—i.e. with an **extreme discordant phenotype** [40, 41]. In one extreme we would have patients with good prognosis (small tumors and no lymph node involvement) that nevertheless develop metastasis: they have a genetic background that turns them more sensitive to metastasis than expected. On the other extreme, there would be poor prognosis patients (high lymph node involvement) that do not develop a disseminated disease: their genetic background protects them again metastasis (Figure 14). These extreme discordant patients are more useful in the search of genetic traits associated to metastasis than overall individuals as they should be enriched in germline variants associated to metastasis, increasing the statistical power and requiring less patients than in a traditional gene association study.

We used two complementary tools to identify genetic variants associated to metastasis: GWAS, which focuses on genetic variants with a strong individual effect on metastasis (assuming that the effect of each variant on the phenotype is independent of other variants) [42]; and an epistasis network analysis, that can detect synergistic interactions among many genetic variants, each with a moderate or weak individual effect.

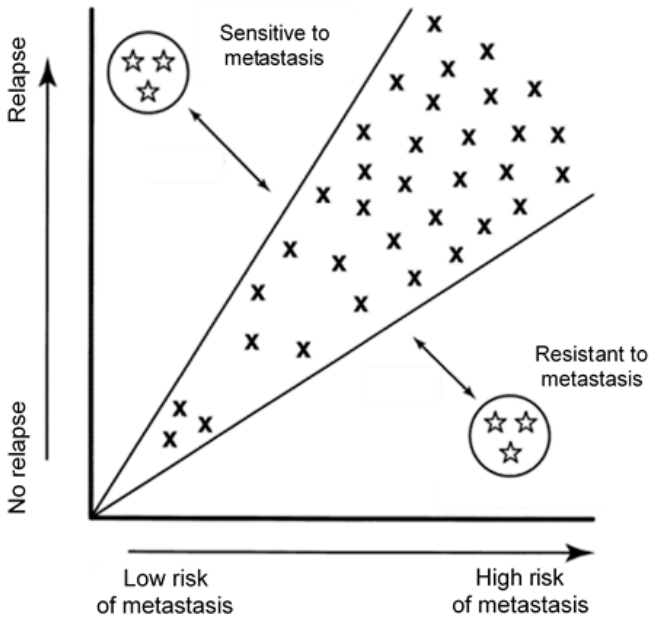

Figure 14. Extreme discordant phenotype framework. In general, the risk of metastasis increases due to certain clinicopathological characteristics (higher tumor size or higher number of lymph nodes affected). However, there are patients with an atypical behavior regarding risk to metastasis: low-risk patients who relapse and can be considered as more "sensitive" to metastasis than expected by their clinicopathological characteristics and high-risk patients who do not relapse as they are more "resistant" to metastasis. *Modified from Nebert 2000 [40]*

6.2 POPULATION OF STUDY AND METHOD'S SUMMARY

6.2.1 Patients

We recruited patients from eight Spanish hospitals (Hospitales Universitarios Regional y Virgen de la Victoria, Málaga; Complejo Hospitalario de Jaén, Jaén; Hospital Universitario Gregorio Marañón, Madrid; Hospital Costa del Sol, Marbella; Hospital Clínico Universitario, Valencia; Hospital Provincial de Córdoba, Córdoba; Hospital Virgen del Rocío, Sevilla; and Hospital San Carlos, Madrid). Patients included in this study were adult females with histologically confirmed invasive breast cancer that had undergone surgery and with a follow-up of more than five years. Patients with bilateral breast cancer and second primary tumors were excluded.

Using an **extreme discordant phenotype framework**, we selected patients at the extremes of the risk to develop metastasis but with an inverse behavior than expected and classified them as good or poor prognosis cases. Good prognosis cases

included patients with low risk of developing metastasis (with tumors ≤ 2 cm and no lymph nodes affected) that nevertheless relapsed within five years after surgery. Poor prognosis cases included patients with high risk of developing metastasis (≥ 10 lymph nodes affected regardless of tumor size) who did not relapse in that period.

6.2.2 Genotyping and identification of variants associated to metastasis

We performed genome-wide genotyping of 97 peripheral blood samples using the HumanOmni5-Quad Beadchip (Illumina), including a technical replicate to determine reproducibility (which was of 0.99). We analyzed the effect of **strong individual variants** by performing an association analysis between SNPs in the good and poor prognosis cases with PLINK [43]. We performed an epistasis network analysis with the Encore pipeline [44] to find **variants and genes with synergistic effect in metastasis**; these are genes relevant to metastasis through their interaction with many other relevant genes and therefore we termed them "metastasis influence genes". Since germline genetic variants tend to associate with complex phenotypes through the regulation of gene expression [55, 56], we hypothesized that our metastasis influence genes might regulate the expression of tumor genes that influence metastasis. Therefore, we studied the impact of our metastasis influence genes on metastasis by testing the role of the sets of genes under their regulation in distant metastasis-free survival in public database.

6.3 RESULTS

6.3.1 Characteristics of the population

Patients' characteristics are shown in Table 6. The median age at diagnosis was 52 years (range 29 to 89). Patients were mainly postmenopausal (55 %), with luminal tumors (76 %) and histological grade 2 (53 %). We observed similar characteristics between good and poor prognosis cases except that good prognosis cases included a higher proportion of postmenopausal patients than poor prognosis cases, although they were similar in age. The intrinsic characteristics that define both groups (tumor

size and lymph node involvement) imply that more patients in the poor prognosis cases received adjuvant treatment.

Table 6. Patients' characteristics, N (%) and *p*-values of the comparison between good and poor prognosis cases

Characteristic	**Good prognosis cases (N = 34)**	**Poor prognosis cases (N = 63)**	***p*-value**
Age at diagnosis (years)			
Median	56	49	0.3209
Range	34 to 89	29 to 85	
Menopausal status			
Premenopausal	9 (26 %)	31 (49 %)	0.0474*
Postmenopausal	22 (65 %)	27 (43 %)	
Unknown	3 (9 %)	5 (8 %)	
Histological grade			
1	1 (3 %)	5 (8 %)	0.2839
2	19 (56 %)	26 (41 %)	
3	9 (26 %)	25 (40 %)	
Unknown	5 (15 %)	7 (11 %)	
Hormone receptor status			
Positive	24 (71 %)	44 (70 %)	0.9043
Negative	8 (23 %)	12 (19 %)	
Unknown	2 (6 %)	7 (11 %)	
HER2 status			
Positive	6 (18 %)	9 (14 %)	0.8556
Negative	20 (59 %)	40 (64 %)	
Unknown	8 (23 %)	14 (22 %)	
Adjuvant chemotherapy			
Yes	14 (41 %)	54 (86 %)	0.0009*
No	20 (59 %)	7 (11 %)	
Unknown	0	2 (3 %)	
Adjuvant hormonotherapy			
Yes	18 (23 %)	48 (76 %)	0.0248*
No	16 (47 %)	14 (22 %)	
Unknown	0	1 (2 %)	
Adjuvant radiotherapy			
Yes	19 (56 %)	52 (83 %)	0.0133*
No	11 (32 %)	7 (11 %)	
Unknown	4 (12 %)	4 (6 %)	

*$p\text{-value} \leq 0.05$

6.3.2 Germline genetic variants with single effect in breast cancer metastasis

We found that **none of the germline variants analyzed had a strong single-effect in metastasis** that reached the required GWAS statistical significance (p-value $< 5 \times 10^{-8}$) in our extreme discordant phenotype framework. However, 19 SNPs achieved p-values in the order of 10^{-6}, 13 of which were not in linkage disequilibrium ($r^2 < 0.8$) [106] (Table 7).

Table 7. Top SNPs, p-values and odds ratios from the single-variant association analysis in our metastasis extreme discordant phenotypes (good prognosis vs. poor prognosis cases)

Chr	SNP ID	p-value	OR	Location	Type of variant
16	rs8043586	2.96 X 10^{-6}	5.47	122 kb 3' of XYLT1	intergenic
12	rs6580981	2.96 X 10^{-6}	0.21	COPZ1	intronic
11	rs643711	3.72 X 10^{-6}	4.71	131 kb 5' of RP11-379J13.2	intergenic
16	rs28452734	4.50 X 10^{-6}	4.42	146 kb 3' of XYLT1	upstream gene
14	rs146483988	5.80 X 10^{-6}	7.13	TDRD9	synonymous
1	rs16831607	6.26 X 10^{-6}	27.46	RP5-1198O20.4	intronic
2	rs62193961	6.98 X 10^{-6}	4.29	NGEF	intronic
23	rs5907175	8.38 X 10^{-6}	0.22	39 kb 5' of SPANXD	intergenic
23	rs138833797	8.48 X 10^{-6}	NA	33 kb 3' of IDS	intergenic
16	rs55655642	9.26 X 10^{-6}	4.85	106 kb 3' of XYLT1	intergenic
22	rs9611014	9.68 X 10^{-6}	6.81	IGLV8-61	missense
18	rs2848788	9.79 X 10^{-6}	6.36	507 kb 5' of PIK3C3	intergenic
19	rs17000770	9.95 X 10^{-6}	4.13	MUC16	missense

Abbreviations: Chr, chromosome; SNP ID, reference SNP identifier or "rs" ID number; OR, odds ratio; NA, not available.

Data of location and type of variant were obtained from Haploreg v. 3 [106] and Ensembl databases [46]

Most of these SNPs are in/close to genes involved in processes related to metastasis or tumor development [46–48]. To give some examples, COPZ1 (coatomer protein complex subunit zeta 1) is a protein that induces apoptosis in tumor cells when knocked down [49]; NGEF (neuronal guanine nucleotide exchange factor) is a gene associated with cell differentiation and positive regulation of apoptotic processes that has transforming potential in cell culture and can induce tumors in nude mice [50]; and MUC16 (mucin 16, also known as ovarian carcinoma antigen CA125) is a membrane-associated mucin involved in cell adhesion and migration that increases proliferation and decreases apoptosis in breast cancer cells [51, 52]. The functional analysis combined with the elevated odds ratios observed in the association analysis (Table 7) suggest that, although not reaching statistical significance, we cannot dismiss that these variants might have some individual effect on the susceptibility to metastasis.

6.3.3 Genetic variants with synergistic interactions that modulate metastasis in breast cancer

We next looked for germline variants that affect the susceptibility to metastasis through synergistic interactions, each with a moderate or weak individual effect on genes that associate on regulatory networks and pathways [118–120]. To identify these genes, we modelled a gene epistasis network that encodes the susceptibly to metastasis in our cohort and identified the genes that reflect most of these interactions—are best associated with the susceptibility to metastasis—by the community centrality measure [53]. These central genes or "metastasis influence genes" are expected to play a direct role on metastasis by their influence on many other genes—which may also be perturbed by germline variants—encoded by the epistasis network [54]. A graphical representation of our epistasis network is shown in Figure 15. It illustrates the community centrality using as examples FOXQ1 (forkhead box Q1) and RXRA (retinoid X receptor alpha), two genes that influence the susceptibility to develop metastasis by interacting with several communities in the epistasis network. Our epistasis network contains 1 431 genes and ca. 5 600 links among them. It is a large and dense network (there are many links among genes), which reveals the polygenic nature of the germline contribution to a complex trait such as metastasis. It is also a

small world network: most genes can be reached from every other gene by a few steps and genes are tightly interconnected forming communities. The most relevant genes of the network are the metastasis influence genes (Table 8), that interconnect different communities integrating them in the network topology; they are relevant because they modulate the contribution of many other genes to metastasis (which is the phenotype encoded in the network).

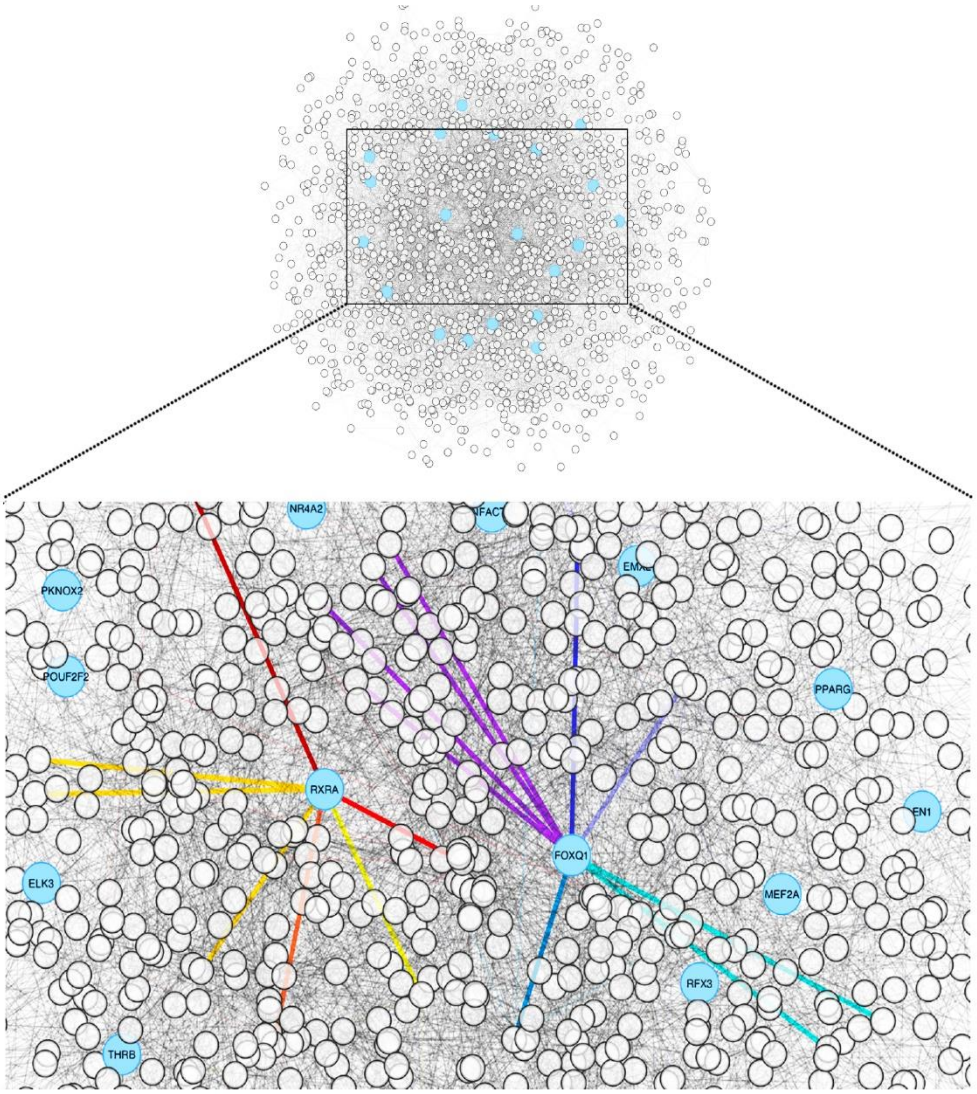

Figure 15. Epistasis network encoding the susceptibility to metastasis in our cohort. The metastasis influence genes, those with high community centrality, are represented in blue. The panel below highlights the participation of two metastasis influence genes (FOXQ1 and RXRA) in several communities by the color of their links (each community is represented with a different color)

We found that the top 10 % of the metastasis influence genes (ranked by community centrality) are overrepresented in KEGG pathways (overrepresentation test; multiple comparison adjusted and false discovery rate adjusted *p*-value < 0.05) such as the interaction with extracellular matrix receptors (KEGG: map04512) and the establishment of cell-extracellular matrix contact points (KEGG: map04510), which implies that **some germline variants of breast cancer patients affect genes mechanistically involved in metastasis**.

Table 8. Top metastasis influence gene ranked by community centrality and *p*-value of the survival test of each metastasis influence gene

Gene	Community centrality	Gene KM *p*-value[a]
FN1	9.101	0.369
TNP1	8.248	0.042*
COMP	7.370	0.331
MAP2K4	6.789	< 0.001*
ZAP70	6.507	0.798
FOXQ1	6.489	0.007*
RBMS3	6.448	0.007*
RAB11FIP2	5.886	0.003*
DENND1A	5.678	0.019*
COL4A1	5.570	0.029*
PER2	5.558	0.010*
TNN	5.373	0.008*
RASGRF1	5.274	0.096
FRS2	5.262	0.012*
MET	5.187	0.660
SPSB1	5.155	0.001*
GABRA5	5.129	0.001*
GREM2	4.718	0.035*
SMARCA4	4.716	0.014*

[a] *p-value of the Kaplan-Meier survival analysis: association of the gene expression profile of each regulator with distant metastasis-free survival in a breast cancer public dataset [45].*
** p-value ≤ 0.05*

6.3.4 Metastasis influence genes and breast cancer survival

We used distant metastasis-free survival (based on gene-expression and survival data from a public breast cancer dataset [45]) to analyze the impact of our metastasis influence genes on metastasis (Table 8). We observed that the expression profile of most of our metastasis influence genes was significantly altered in breast cancer tumor cells (as we found these genes among those differentially expressed in the breast cancer public dataset) and that their expression was significantly associated with distant metastasis-free survival. Previous data support that the dysregulated expression of these genes in breast cancer cells and tissue causes metastasis, for example somatic alteration in TNP1 (transition protein 1) and in RBMS3 (RNA binding motif single stranded interacting protein 3) bring forth metastasis in breast cancer [121, 122].

As the survival analysis was performed on tumor cells we can infer that those metastasis influence genes significantly associated with survival (gene KM p-value ≤ 0.05) are altered in the tumor tissue, while the rest are altered in the tumor microenvironment or metastasis target tissues—they are relevant to metastasis due to their community centrality but are not found in tumor cells (Table 8). In the tumor tissue, we found genes that are known to contribute to breast cancer: TNP1 (transition protein 1) is a DNA-binding protein that participates in chromatin remodeling that is associated with overall breast cancer risk [123]; and MAP2K4 (mitogen-activated protein kinase 4) is a member of a protein family that acts as an integration point for multiple biochemical signals and is involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development [121]. Our research suggests that these (and other metastasis influence genes) might be altered in the host cells and that their expression in the tumor tissue may influence the development of metastases. We also found non-tumor genes that affect breast cancer metastasis: FN1 (fibronectin 1) contributes to cell migration through the continual formation of the extracellular matrix and, when silenced in the lungs, favors the metastatic seeding of mouse mammary tumor cells [124]; and COMP (cartilage oligomeric matrix protein) is a non-collagenous extracellular matrix protein which increases invasiveness and tumor cell viability.

As a result of the network and survival analysis, **we found germline genetic variants that affect genes that either are expressed in the tumor and promote the migration of tumor cells—dissemination of tumor seeds—or are expressed in tumor microenvironment or target tissues and favor the growth of the tumor in metastatic sites—provide a congenial soil for the seeding of the primary tumor cells.**

6.3.5 Effect of gene expression regulation in breast cancer metastasis

Germline genetic variants tend to associate with complex phenotypes as metastasis through the regulation of gene expression [55, 56]. Therefore, our metastasis influence genes might also be relevant to metastasis by the regulation of the expression of a set of tumor genes that influence metastasis.

To evaluate if any of our metastasis influence genes regulate the expression of tumor genes we obtained the regulatory interactions between them and their regulons—genes under their regulation—from a breast cancer-specific regulatory network [57]. We found that 353 of our metastasis influence genes were present in this breast cancer regulatory gene network. The proportion of regulators among our metastasis influence genes was significantly higher than expected by chance (binomial test p-value < 0.001), with 20 regulators out of the 353 genes. They regulate the expression of ca. 8 000 target genes in breast cancer cell lines, with an average of 404 regulated genes per regulator. For each of the regulators, we investigated whether their regulons were associated with distant metastasis-free survival to find if the metastasis influence genes from the tumor tissue regulate genes involved in breast cancer metastasis. **We observed a statistically significant association between distant metastasis-free survival and the regulon expression patterns of many metastasis susceptibility genes** (Table 9).

We found that several of these regulators are paralogous of known transcriptional master regulators in breast cancer: FOXQ1 (forkhead box Q1); AR (androgen receptor); RXRA (retinoid X receptor alpha); and PPARG (peroxisome proliferator activated receptor gamma) among others [58, 59]. This further supports

that the germline variants in such genes produce a dysregulated expression pattern in the tumor tissue that results in increased metastasis.

Table 9. Regulators among the top metastasis influence genes ranked by community centrality, number of genes under their regulation (regulon size) and *p*-value of the survival test for each regulon

Gene	Community centrality	Regulon size	Regulon KM *p*-value[a]
FOXQ1	6.489	131	0.002*
EN2	4.633	195	0.008*
RFX3	4.264	1187	< 0.001*
THRB	3.848	199	0.200
POU6F2	3.624	182	0.527
POU2F2	3.483	531	< 0.001*
NR4A2	2.846	478	< 0.001*
MEF2A	2.833	474	0.001*
IRX4	2.778	35	0.001*
PKNOX2	2.750	330	< 0.001*
AR	2.742	346	< 0.001*
RXRA	2.714	963	< 0.001*
EN1	2.667	274	0.961
PPARG	2.600	806	< 0.001*
SHOX2	2.556	126	0.970
EMX2	2.500	142	0.257
NFATC1	2.500	193	< 0.001*
ID4	2.429	247	0.001*
BACH2	2.429	371	0.002*
ELK3	2.429	871	< 0.001*

[a] *p-value of the Kaplan-Meier survival analysis: association of the gene expression pattern of each regulon with distant-metastasis survival in a breast cancer public dataset. * p-value ≤ 0.05*

6.4 DISCUSSION

The results of this study move onward the seed and soil hypothesis: the genetic background of the host contributes to the development of metastasis. To identify the germline genetic variants that modulate metastasis in breast cancer we designed an extreme discordant phenotype GWAS framework with enhanced statistical power to find variants with strong single-effect combined with an epistasis network analysis to

detect synergistic interactions among genetic variants that collectively influence the susceptibility to develop metastasis in breast cancer.

In the same line than previous studies that aimed to identify germline genetic variants associated with breast cancer survival [38, 60, 61], our genome-wide genotyping of extreme discordant phenotypes did not reveal any variant with a statistically significant individual effect on metastasis. Nevertheless, we found that some of the top SNPs identified held high odds ratios, suggesting that there may be a correlation between some of these SNPs and metastasis masked by the low sample size. We also found several genes such as COPZ1 and MUC16 expressed in the tumor tissue that, when holding specific variants, might favor the ability of tumor cells to disseminate—and hence might influence the susceptibility to develop metastasis—suggesting a biological effect of these variants in metastasis.

Through the analysis of epistasis networks, we identified several metastasis influence genes. These are **genes altered by germline genetic variants that influence metastasis through their synergistic interaction with multiple genes** from our epistasis network, each with accumulative weak effects in metastasis. These metastasis influence genes affect the susceptibility to develop metastasis either through molecular mechanisms that favor tumor spread (TNP1, MAP2K4) or they influence genes expressed in the tumor microenvironment or the metastasis target tissues (FN1, COMP): **the germline alterations in these genes either favor the dissemination of metastatic seeds or provide a congenial soil for them to grow.**

We also observed that some of the metastasis influence genes from the tumor tissue (FOXQ1, RXRA, PPARG) affect metastasis through their expression and regulatory patterns: either their dysregulated expression in the tumor promotes metastasis and/or these genes alter the expression of a set of other tumor genes that result in the development of metastasis. Since the function of the metastasis influence genes is altered by germline genetic variants—that therefore exist before the development of breast tumors—**our results confirm that the metastatic potential of a breast tumor cell depends not only on the cumulative mutations on the primary tumor but also on the pre-programmed dysregulation of gene**

expression in the host and point to the keyplayers of this dysregulated gene expression. This way, women who harbor certain variants in these genes will deploy gene expression patterns that favor (or protect against) metastasis in case they develop breast cancer.

More research should be done to complement our results. As breast cancer prognosis is significantly different in each breast cancer subtype [125], it would be useful to analyze if the effect of the host genetic variants in metastasis vary among breast cancer subtypes. Besides, the influence of the host's background in breast cancer metastasis may not be due exclusively to the effect of a set of germline genetic variants, but could also be modulated by other factors such as hormone factors, methylation or regulation through miRNA.

Our work provides further support of the seed and soil hypothesis. We identified a set of genes that, when altered by germline mutations, modulate metastasis by facilitating the dispersion of the primary tumor seeds or providing a congenial soil for them to grow. We also found that some of them operate through the regulation of the expression of other tumor genes: the host genetic makeup can modulate the metastatic dissemination independently or in conjunction with the primary tumor characteristics and the genetic susceptibility to metastasis in breast cancer patients operates through the genetic regulation of the tumor. These results suggest that **breast cancer vast heterogeneity is not only due to the tumor characteristics but also to the patient's genetic background**: patients would benefit from a better management and more personalized treatment when diagnosed in a more systemic approach that considers both the tumor characteristics and the host genetic background.

7. GENERAL DISCUSSION

Your assumptions are your windows on the world.
Scrub them off every once in a while, or the light won't come in

Isaac Asimov

In these research studies we have explored the intertumor heterogeneity of breast cancer—i.e. differences observed between patients—by the use of triple negative breast cancer and male breast cancer as a proxy to explore intertumor heterogeneity in tumor tissues and germline genetic variation as a proxy of intertumor heterogeneity in non-tumor tissues. We found that the subtyping of breast cancer can reveal biologically and clinically relevant subsets, and that the combined analysis of tumor features with host genetic features could improve breast cancer prognosis and guide treatment decisions.

Triple negative breast cancer is the subtype with the poorest survival and wider range of treatment response. Due to the lack of targeted therapies, chemotherapy is its only systemic treatment option [5, 6]; platinum-based therapies are also used in the neoadjuvant and metastatic settings, although its use in unselected triple negative patients is controversial [7, 14]. There is a need to identify the subset of patients that actually benefit from each treatment and to identify potential biomarkers to guide targeted therapies in resistant patients.

Although male breast cancer is a unique disease, its treatment is largely extrapolated from female breast cancer studies. In fact, there are few male breast cancer specific recommendations in clinical practice guidelines [7, 95, 126, 127]. These guidelines have in common that tamoxifen is the standard recommended therapy as the great majority of male breast cancers are ER+. Although some alternative options are offered for ER+ male patients (aromatase inhibitors or gonadotropic hormone-releasing hormone analogues), none of the guides offer specific recommendations on the few ER- or HER2+ cases. In fact, it has been suggested that the poorer survival of males compared with females with breast cancer may be due to its undertreatment, to which the inconsistent application of treatments contributes [128]. The immunohistochemical classification of breast cancer appears to be insufficient to identify male breast cancer subtypes that guide treatment decisions and predict clinical outcome—i.e. it cannot capture male breast cancer heterogeneity [7].

Several molecular classifications have been published both for triple negative and male breast cancer in order to unveil their intertumor heterogeneity. Although useful in assessing the molecular diversity within these tumors, it is unclear if these classifications are predictive of treatment efficacy. Therefore, there is a need to identify clinically relevant subtypes that reveal sets of patients with different prognosis and response to treatment: the identification of these patients can lead to an improvement in their management, treatment selection and clinical outcome.

We found in both triple negative and male breast cancer that subtyping may identify sets of patients with different prognosis and guide treatment selection. In triple negative breast cancer, we found a set of patients that may benefit from platinum-based neoadjuvant chemotherapy. We also observed a set of triple negative tumors that appear to be luminal instead, as they had the lowest proliferation rates and none was basal-like according to PAM50 signature. In male breast cancers, we found that the PAM50-based classification explains the interpatient heterogeneity better than the immunohistochemical classification. The PAM50 signature identifies a set of tumors associated with poor survival that could improve their outcome through targeted therapies.

On the other hand, we studied intertumor heterogeneity in non-tumor tissues by the analysis of germline genetic variants associated with metastasis. The survival of breast cancer patients cannot be fully explained by tumor characteristics and treatment decisions. The tumor microenvironment plays a key role in breast cancer progression, is associated with clinical outcome and can modulate the response to commonly applied therapies [129, 130]. These data suggest that germline genetic variants, combined with the intrinsic characteristics of the tumor tissue, could help explaining the heterogeneity observed in breast cancer survival. The genetic variants studied may be implicated in cellular processes involved in tumor progression, and their identification may lead to targets for future therapies.

We found a set of genes that, when altered by germline genetic variants, modulate metastasis through their synergistic interactions. These genes are located in the tumor tissue, tumor microenvironment and target tissues: they promote the dispersion of tumor cells or facilitate that a target tissue is suitable for them to grow. The aggressiveness of a tumor is therefore determined not only by tumor characteristics but also from the pre-programmed dysregulation of gene expression in the host. The combination of these variants adds more layers to the heterogeneity observed between patients in breast cancer survival and suggests that they should be considered for developing treatment decisions. The identification of phenotypes more sensitive or resistant to develop metastasis could lead to improved prognosis and to novel anti-metastatic therapies.

Further studies are needed to validate the clinical relevance of these subtyping methods and to validate the role of these germline genetic variants in breast cancer metastasis. In triple negative breast cancer, randomized clinical trials could be performed to evaluate the benefit of standard chemotherapy vs. platinum-based therapies in BL1 and the benefit of standard chemotherapy vs. antiandrogen therapies in LAR. These clinical trials may lead to the inclusion of new treatment recommendations in clinical guidelines. In male breast cancer, an improvement in survival could come from the assessment of the benefit on anti-HER2 therapies in either HER2-enriched tumors or in HER2+ by immunohistochemistry; although some studies suggest their use following the same recommendations for female cancer [131], the treatment of male breast cancer with anti-HER2 therapies is not included in current clinical guidelines. Finally, the confirmation of the host's genetic background role in breast cancer metastasis opens a new field of knowledge. More research is needed to validate the implication of the germline genetic variants identified in breast cancer metastasis as well as to discover the mechanisms through which they affect metastasis. All these studies could have a great impact in improving breast cancer patients' survival and in modifying clinical guide recommendations for specific subsets of female and male breast cancer.

To sum up, we have confirmed that breast cancer subtyping can help overcoming the great heterogeneity observed between breast cancer patients, improving prognosis and treatment decisions. The patient's genetic background seem to also play a role in the differences observed in breast cancer survival, suggesting that a more comprehensive management of breast cancer patients would include not only information about breast cancer subtyping but also about certain germline genetic variants.

8. Conclusions

Science is not only a disciple of reason but,
also, one of romance and passion.

Stephen Hawking (1942 -)

The results of this research suggest that the identification of clinically relevant breast cancer subtypes and the determination of germline genetic variants are useful to reveal the wide heterogeneity that exists between breast cancer patients and could improve treatment selection and patient's outcome, getting us a step closer to personalized medicine. The main conclusions of this Thesis are the following:

1. Triple negative cancer, rather than falling into discreet categories, seems to be a continuous spectrum of tumors ranging from highly proliferative tumors sensitive to platinum-based chemotherapy (basal-like 1 tumors) to lowly proliferative tumors chemoresistant to standard chemotherapy (luminal androgen receptor tumors). Lehmann subtyping has proven useful in identifying the patients at the extremes of this spectrum, who could benefit of a better clinical management and a more personalized treatment.

2. Male breast cancer classification in intrinsic subtypes based on the PAM50 signature confirms that male breast cancer is mainly a genomic luminal disease (contrary to female breast cancer) and identifies a poor prognosis subgroup missed by traditional methods based on immunohistochemical surrogates. The identification of these poor prognosis patients and their treatment with targeted therapies could improve the natural history of their disease.

3. Host genetic factors modulate metastasis in breast cancer patients—independently or in conjunction with the tumor characteristics—by diverse methods that include altered gene expression in tumor and non-tumor tissues and regulation of the expression of sets of tumor genes. These methods converge in promoting the dispersion of metastatic seeds and providing a congenial soil that allows them to root and grow. The management and care of breast cancer patients should move beyond the dissection of tumor characteristics to a more systemic approach that also incorporates data about their genetic background.

9. Conclusiones

The most beautiful experience we can have is the mysterious.
It is the fundamental emotion that stands at the cradle of true art and true science

Albert Einstein (1879 - 1955)

Los resultados de este trabajo de investigación sugieren que la identificación de subtipos tumorales clínicamente relevantes y la identificación de variantes en el contexto genético de pacientes con cáncer de mama son útiles para explorar la heterogeneidad existente entre pacientes con cáncer de mama. De este modo, podrían facilitar la elección del tratamiento y mejorar la evolución clínica de los pacientes, acercándonos un poco más a la medicina personalizada. Las conclusiones principales que se derivan de esta Tesis Doctoral son las siguientes:

1. El cáncer de mama triple negativo, en vez de subdividirse en categorías discretas, parece ser un espectro continuo de tumores. En un extremo de este espectro encontraríamos tumores muy proliferativos que son sensibles al tratamiento con sales de platino (tumores *basal-like 1*) y en el otro, tumores poco proliferativos aparentemente a la quimioterapia estándar (tumores *luminal androgen receptor*). Los subtipos de Lehmann resultan útiles en la identificación de estos pacientes extremos, que podrían beneficiarse de una mejor atención clínica y de un tratamiento más personalizado.

2. La clasificación en subtipos intrínsecos basados en la firma PAM50 confirma que el cáncer de mama en los varones es principalmente una enfermedad luminal genómica (al contrario de lo que sucede en el cáncer de mama en las mujeres) e identifica un subgrupo de pacientes de mal pronóstico, que no se detecta mediante los métodos tradicionales basados en subrogados inmunohistoquímicos. La identificación de estos pacientes con mal pronóstico y su tratamiento con terapias dirigidas podría mejorar la historia natural de su enfermedad.

3. Existen factores en el contexto genético de las pacientes con cáncer de mama que modulan la aparición de metástasis—independientemente o junto a las características del tumor—mediante diversos métodos que incluyen la alteración de la expresión en tejidos tumorales y no tumorales, y la regulación de la expresión de un conjunto de genes tumorales. Estos métodos convergen a la hora de promover la dispersión de las células tumorales, así como de facilitar que el tejido diana sea apropiado para que se

desarrollen focos tumorales. La atención a estas pacientes y la selección del tratamiento debería ir más allá de la disección de las características del tumor hacia una aproximación más global que incorpore también información sobre su contexto genético.

10. REFERENCES

Nothing in life is to be feared, it is only to be understood.
Now is the time to understand more, so that we may fear less

Marie Curie

1. Rivenbark, A. G., O'Connor, S. M., and Coleman, W. B. (2013). Molecular and cellular heterogeneity in breast cancer: challenges for personalized medicine. Am J Pathol *183*, 1113–1124.

2. Badve, S., and Gökmen-polar, Y. (2015). Tumor Heterogeneity in Breast Cancer. Advances In Anatomic Pathology.

3. Ribelles, N., Santonja, A., Pajares, B., Llácer, C., and Alba, E. (2014). The seed and soil hypothesis revisited: current state of knowledge of inherited genes on prognosis in breast cancer. Cancer Treat Rev *40*, 293–299.

4. Duffy, M. J., Harbeck, N., Nap, M., Molina, R., Nicolini, A., Senkus, E., and Cardoso, F. (2017). Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). Eur J Cancer *75*, 284–298.

5. Dent, R., Trudeau, M., Pritchard, K. I., Hanna, W. M., Kahn, H. K., Sawka, C. A., Lickley, L. A., Rawlinson, E., Sun, P., and Narod, S. A. (2007). Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res *13*, 4429–4434.

6. Foulkes, W. D., Smith, I. E., and Reis-Filho, J. S. (2010). Triple-negative breast cancer. N Engl J Med *363*, 1938–1948.

7. Curigliano, G., Burstein, H. J., P Winer, E., Gnant, M., Dubsky, P., Loibl, S., Colleoni, M., Regan, M. M., Piccart-Gebhart, M., Senn, H. J., et al. (2017). De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. Ann Oncol *28*, 1700–1712.

8. Liedtke, C., Mazouni, C., Hess, K. R., André, F., Tordai, A., Mejia, J. A., Symmans, W. F., Gonzalez-Angulo, A. M., Hennessy, B., Green, M., et al. (2008). Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J Clin Oncol *26*, 1275–1281.

9. Carey, L. A., Dees, E. C., Sawyer, L., Gatti, L., Moore, D. T., Collichio, F., Ollila, D. W., Sartor, C. I., Graham, M. L., and Perou, C. M. (2007). The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res *13*, 2329–2334.

10. Cortazar, P., Zhang, L., Untch, M., Mehta, K., Costantino, J. P., Wolmark, N., Bonnefoi, H., Cameron, D., Gianni, L., Valagussa, P., et al. (2014). Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. The Lancet *384*, 164–172.

11. Lehmann, B. D., Bauer, J. A., Chen, X., Sanders, M. E., Chakravarthy, A. B., Shyr, Y., and Pietenpol, J. A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest *121*, 2750–2767.

12. Burstein, M. D., Tsimelzon, A., Poage, G. M., Covington, K. R., Contreras, A., Fuqua, S. A. W., Savage, M. I., Osborne, C. K., Hilsenbeck, S. G., Chang, J. C., et al. (2015). Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin Cancer Res *21*, 1688–1698.

13. Liu, Y.-R., Jiang, Y.-Z., Xu, X.-E., Yu, K.-D., Jin, X., Hu, X., Zuo, W.-J., Hao, S., Wu, J., Liu, G.-Y., et al. (2016). Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer. Breast Cancer Res *18*, 33.

14. Alba, E., Chacon, J. I., Lluch, A., Anton, A., Estevez, L., Cirauqui, B., Carrasco, E., Calvo, L., Segui, M. A., Ribelles, N., et al. (2012). A randomized phase II trial of platinum salts in basal-like breast cancer patients in the neoadjuvant setting. Results from the GEICAM/2006-03, multicenter study. Breast Cancer Res Treat *136*, 487–493.

15. Prat, A., and Perou, C. M. (2011). Deconstructing the molecular portraits of breast cancer. Mol Oncol *5*, 5–23.

16. Parker, J. S., Mullins, M., Cheang, M. C. U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol *27*, 1160–1167.

17. Prosigna [Package Insert]. (2015). Seattle, WA: Nanostring Technologies, Inc. Available at: http://prosigna.com/docs/Prosigna_Packet_Insert_US.pdf [Accessed January 10, 2016].

18. Chen, X., Li, J., Gray, W. H., Lehmann, B. D., Bauer, J. A., Shyr, Y., and Pietenpol, J. A. (2012). TNBCtype: A Subtyping Tool for Triple-Negative Breast Cancer. Cancer Inform *11*, 147–156.

19. Prat, A., Fan, C., Fernández, A., Hoadley, K. A., Martinello, R., Vidal, M., Viladot, M., Pineda, E., Arance, A., Muñoz, M., et al. (2015). Response and survival of breast cancer intrinsic subtypes following multi-agent neoadjuvant chemotherapy. BMC Med *13*, 303.

20. Masuda, H., Baggerly, K. A., Wang, Y., Zhang, Y., Gonzalez-Angulo, A. M., Meric-Bernstam, F., Valero, V., Lehmann, B. D., Pietenpol, J. A., Hortobagyi, G. N., et al. (2013). Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. Clin Cancer Res *19*, 5533–5540.

21. Loibl, S., Müller, B. M., von Minckwitz, G., Schwabe, M., Roller, M., Darb-Esfahani, S., Ataseven, B., du Bois, A., Fissler-Eckhoff, A., Gerber, B., et al. (2011). Androgen receptor expression in primary breast cancer and its predictive and prognostic value in patients treated with neoadjuvant chemotherapy. Breast Cancer Res Treat *130*, 477–487.

22. Prat, A., Lluch, A., Albanell, J., Barry, W. T., Fan, C., Chacón, J. I., Parker, J. S., Calvo, L., Plazaola, A., Arcusa, A., et al. (2014). Predicting response and survival in chemotherapy-treated triple-negative breast cancer. Br J Cancer *111*, 1532–1541.

23. Lehmann, B. D., and Pietenpol, J. A. (2014). Identification and use of biomarkers in treatment strategies for triple-negative breast cancer subtypes. J Pathol *232*, 142–150.

24. Pareja, F., Geyer, F. C., Marchiò, C., Burke, K. A., Weigelt, B., and Reis-Filho, J. S. (2016). Triple-negative breast cancer: the importance of molecular and histologic subtyping, and recognition of low-grade variants. NPJ Breast Cancer *2*, 16036.

25. Gucalp, A., Tolaney, S., Isakoff, S. J., Ingle, J. N., Liu, M. C., Carey, L. A., Blackwell, K., Rugo, H., Nabell, L., Forero, A., et al. (2013). Phase II trial of bicalutamide in patients with androgen receptor-positive, estrogen receptor-negative metastatic Breast Cancer. Clin Cancer Res *19*, 5505–5512.

26. Traina, T. A., Miller, K., Yardley, D. A., O'Shaughnessy, Y., Cortes, J., Awada, A., Kelly, C. M., Trudeau, M. E., Peter Shmid, Gianni, L., et al. Results from a phase 2 study of enzalutamide (ENZA), an androgen receptor (AR) inhibitor, in advanced AR+ triple-negative breast cancer (TNBC). J Clin Oncol *33,* Suppl 15

27. Bonnefoi, H., Grellety, T., Tredan, O., Saghatchian, M., Dalenc, F., Mailliez, A., L'Haridon, T., Cottu, P., Abadie-Lacourtoisie, S., You, B., et al. (2016). A phase II trial of abiraterone acetate plus prednisone in patients with triple-negative androgen receptor positive locally advanced or metastatic breast cancer (UCBG 12-1). Ann Oncol *27*, 812–818.

28. Von Minckwitz, G., Schneeweiss, A., Loibl, S., Salat, C., Denkert, C., Rezai, M., Blohmer, J. U., Jackisch, C., Paepke, S., Gerber, B., et al. (2014). Neoadjuvant carboplatin in patients with triple-negative and HER2-positive early breast cancer (GeparSixto; GBG 66): a randomised phase 2 trial. Lancet Oncol *15*, 747–756.

29. Sikov, W. M., Berry, D. A., Perou, C. M., Singh, B., Cirrincione, C. T., Tolaney, S. M., Kuzma, C. S., Pluard, T. J., Somlo, G., Port, E. R., et al. (2015). Impact

of the addition of carboplatin and/or bevacizumab to neoadjuvant once-per-week paclitaxel followed by dose-dense doxorubicin and cyclophosphamide on pathologic complete response rates in stage II to III triple-negative breast cancer: CALGB 40603 (Alliance). J Clin Oncol *33*, 13–21.

30. Telli, M. L., Timms, K. M., Reid, J., Hennessy, B., Mills, G. B., Jensen, K. C., Szallasi, Z., Barry, W. T., Winer, E. P., Tung, N. M., et al. (2016). Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. Clin Cancer Res *22*, 3764–3773.

31. Rizzolo, P., Silvestri, V., Tommasi, S., Pinto, R., Danza, K., Falchetti, M., Gulino, M., Frati, P., and Ottini, L. (2013). Male breast cancer: genetics, epigenetics, and ethical aspects. Ann Oncol *24 Suppl 8*, viii75–viii82.

32. Vermeulen, M. A., Slaets, L., Cardoso, F., Giordano, S. H., Tryfonidis, K., van Diest, P. J., Dijkstra, N. H., Schröder, C. P., van Asperen, C. J., Linderholm, B., et al. (2017). Pathological characterisation of male breast cancer: Results of the EORTC 10085/TBCRC/BIG/NABCG International Male Breast Cancer Program. Eur J Cancer *82*, 219–227.

33. Goldhirsch, A., Winer, E. P., Coates, A. S., Gelber, R. D., Piccart-Gebhart, M., Thürlimann, B., Senn, H. J., and Panel members (2013). Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol *24*, 2206–2223.

34. Dawood, S., Broglio, K., Buzdar, A. U., Hortobagyi, G. N., and Giordano, S. H. (2010). Prognosis of women with metastatic breast cancer by HER2 status and trastuzumab treatment: an institutional-based review. J Clin Oncol *28*, 92–98.

35. Carey, L. A., Berry, D. A., Cirrincione, C. T., Barry, W. T., Pitcher, B. N., Harris, L. N., Ollila, D. W., Krop, I. E., Henry, N. L., Weckstein, D. J., et al. (2016). Molecular heterogeneity and response to neoadjuvant human epidermal growth

factor receptor 2 targeting in CALGB 40601, a randomized phase III trial of paclitaxel plus trastuzumab with or without lapatinib. J Clin Oncol *34*, 542–549.

36. Prat, A., Bianchini, G., Thomas, M., Belousov, A., Cheang, M. C. U., Koehler, A., Gómez, P., Semiglazov, V., Eiermann, W., Tjulandin, S., et al. (2014). Research-based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. Clin Cancer Res *20*, 511–521.

37. De Groot, A. E., Roy, S., Brown, J. S., Pienta, K. J., and Amend, S. R. (2017). Revisiting seed and soil: examining the primary tumor and cancer cell foraging in metastasis. Mol Cancer Res *15*, 361–370.

38. Pirie, A., Guo, Q., Kraft, P., Canisius, S., Eccles, D. M., Rahman, N., Nevanlinna, H., Chen, C., Khan, S., Tyrer, J., et al. (2015). Common germline polymorphisms associated with breast cancer-specific survival. Breast Cancer Res *17*, 58.

39. De, R., Bush, W. S., and Moore, J. H. (2014). Bioinformatics challenges in genome-wide association studies (GWAS). Methods Mol Biol *1168*, 63–81.

40. Nebert, D. W. (2000). Extreme discordant phenotype methodology: an intuitive approach to clinical pharmacogenetics. Eur J Pharmacol *410*, 107–120.

41. Pérez-Gracia, J. L., Gúrpide, A., Ruiz-Ilundain, M. G., Alfaro Alegría, C., Colomer, R., García-Foncillas, J., and Melero Bermejo, I. (2010). Selection of extreme phenotypes: the role of clinical observation in translational research. Clin Transl Oncol *12*, 174–180.

42. Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., and Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet *11*, 446–450.

43. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., et al. (2007). PLINK: a

tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559–575.

44. Davis, N. A., Lareau, C. A., White, B. C., Pandey, A., Wiley, G., Montgomery, C. G., Gaffney, P. M., and McKinney, B. A. (2013). Encore: Genetic Association Interaction Network centrality pipeline and application to SLE exome data. Genet Epidemiol *37*, 614–621.

45. Van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A. M., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med *347*, 1999–2009.

46. Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C. G., et al. (2017). Ensembl 2018. Nucleic Acids Res.

47. Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2015). GenBank. Nucleic Acids Res *43*, D30–5.

48. Hoffmann, R. (2008). A wiki for the life sciences where authorship matters. Nat Genet *40*, 1047–1051.

49. Shtutman, M., Baig, M., Levina, E., Hurteau, G., Lim, C.-U., Broude, E., Nikiforov, M., Harkins, T. T., Carmack, C. S., Ding, Y., et al. (2011). Tumor-specific silencing of COPZ2 gene encoding coatomer protein complex subunit ζ 2 renders tumor cells dependent on its paralogous gene COPZ1. Proc Natl Acad Sci U S A *108*, 12449–12454.

50. Rodrigues, N. R., Theodosiou, A. M., Nesbit, M. A., Campbell, L., Tandle, A. T., Saranath, D., and Davies, K. E. (2000). Characterization of Ngef, a novel member of the Dbl family of genes expressed predominantly in the caudate nucleus. Genomics *65*, 53–61.

51. Gubbels, J. A. A., Belisle, J., Onda, M., Rancourt, C., Migneault, M., Ho, M., Bera, T. K., Connor, J., Sathyanarayana, B. K., Lee, B., et al. (2006). Mesothelin-MUC16 binding is a high affinity, N-glycan dependent interaction that facilitates peritoneal metastasis of ovarian tumors. Mol Cancer *5*, 50.

52. Lakshmanan, I., Ponnusamy, M. P., Das, S., Chakraborty, S., Haridas, D., Mukhopadhyay, P., Lele, S. M., and Batra, S. K. (2012). MUC16 induced rapid G2/M transition via interactions with JAK2 for increased proliferation and anti-apoptosis in breast cancer cells. Oncogene *31*, 805–817.

53. Ahn, Y.-Y., Bagrow, J. P., and Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. Nature *466*, 761–764.

54. McKinney, B. A., and Pajewski, N. M. (2011). Six degrees of epistasis: statistical network models for GWAS. Front Genet *2*, 109.

55. Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., and Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A *106*, 9362–9367.

56. Bai, L., Yang, H. H., Hu, Y., Shukla, A., Ha, N.-H., Doran, A., Faraji, F., Goldberger, N., Lee, M. P., Keane, T., et al. (2016). An Integrated Genome-Wide Systems Genetics Screen for Breast Cancer Metastasis Susceptibility Genes. PLoS Genet *12*, e1005989.

57. FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A. R. R., Kawaji, H., Rehli, M., Baillie, J. K., de Hoon, M. J. L., Haberle, V., Lassmann, T., Kulakovskiy, I. V., Lizio, M., et al. (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470.

58. Tovar, H., García-Herrera, R., Espinal-Enríquez, J., and Hernández-Lemus, E. (2015). Transcriptional master regulator analysis in breast cancer genetic networks. Comput Biol Chem *59 Pt B*, 67–77.

59. Lim, W. K., Lyashenko, E., and Califano, A. (2009). Master regulators used as breast cancer metastasis classifier. Pac Symp Biocomput, 504–515.

60. Azzato, E. M., Pharoah, P. D. P., Harrington, P., Easton, D. F., Greenberg, D., Caporaso, N. E., Chanock, S. J., Hoover, R. N., Thomas, G., Hunter, D. J., et al. (2010). A genome-wide association study of prognosis in breast cancer. Cancer Epidemiol Biomarkers Prev *19*, 1140–1143.

61. Scannell Bryan, M., Argos, M., Andrulis, I. L., Hopper, J. L., Chang-Claude, J., Malone, K., John, E. M., Gammon, M. D., Daly, M., Terry, M. B., et al. (2017). Limited influence of germline genetic variation on all-cause mortality in women with early onset breast cancer: evidence from gene-based tests, single-marker regression, and whole-genome prediction. Breast Cancer Res Treat *164*, 1–11.

62. Ervik, M., Lam, F., Ferlay, J., Mery, L., Soerjomataram, I., and Bray, F. (2016). Cancer Today. Lyon, France: International Agency for Research on Cancer. Cancer Today. Available at: http://gco.iarc.fr/today [Accessed August 28, 2017].

63. WHO | Global Health Observatory (GHO) data Available at: http://www.who.int/gho/en/ [Accessed November 6, 2017].

64. Galceran, J., Ameijide, A., Carulla, M., Mateos, A., Quirós, J. R., Rojas, D., Alemán, A., Torrella, A., Chico, M., Vicente, M., et al. (2017). Cancer incidence in Spain, 2015. Clin Transl Oncol *19*, 799–825.

65. Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., and Shi, B. (2015). Breast cancer intrinsic subtype classification, clinical use and future trends. Am J Cancer Res *5*, 2929–2943.

66. Elston, C. W., and Ellis, I. O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology *19*, 403–410.

67. International Agency for Research on Cancer (2012). WHO Classification of Tumours of the Breast 4th ed. S. R. Lakhani and I. O. Ellis, eds. (Lyon: International Agency for Research in Cancer).

68. Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., et al. (2000). Molecular portraits of human breast tumours. Nature *406*, 747–752.

69. Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A *98*, 10869–10874.

70. Yersal, O., and Barutca, S. (2014). Biological subtypes of breast cancer: Prognostic and therapeutic implications. World J. Clin. Oncol. *5*, 412–424.

71. Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F. L., Walker, M. G., Watson, D., Park, T., et al. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med *351*, 2817–2826.

72. Filipits, M., Rudas, M., Jakesz, R., Dubsky, P., Fitzal, F., Singer, C. F., Dietze, O., Greil, R., Jelen, A., Sevelda, P., et al. (2011). A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res *17*, 6012–6020.

73. Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., et al. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A *100*, 8418–8423.

74. Cheang, M. C. U., Chia, S. K., Voduc, D., Gao, D., Leung, S., Snider, J., Watson, M., Davies, S., Bernard, P. S., Parker, J. S., et al. (2009). Ki67 index, HER2 status,

and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst *101*, 736–750.

75. Cheang, M. C. U., Voduc, D., Bajdik, C., Leung, S., McKinney, S., Chia, S. K., Perou, C. M., and Nielsen, T. O. (2008). Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res *14*, 1368–1376.

76. Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

77. Curtis, C., Shah, S. P., Chin, S.-F., Turashvili, G., Rueda, O. M., Dunning, M. J., Speed, D., Lynch, A. G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

78. Ly, D., Forman, D., Ferlay, J., Brinton, L. A., and Cook, M. B. (2013). An international comparison of male and female breast cancer incidence rates. Int J Cancer *132*, 1918–1926.

79. Giordano, S. H. (2005). A review of the diagnosis and management of male breast cancer. Oncologist *10*, 471–479.

80. Fentiman, I. S. (2016). Male breast cancer is not congruent with the female disease. Crit Rev Oncol Hematol *101*, 119–124.

81. Yalaza, M., İnan, A., and Bozer, M. (2016). Male Breast Cancer. J. Breast Health *12*, 1–8.

82. Johansson, I., Killander, F., Linderholm, B., and Hedenfalk, I. (2014). Molecular profiling of male breast cancer - lost in translation? Int J Biochem Cell Biol *53*, 526–535.

83. Piscuoglio, S., Ng, C. K. Y., Murray, M. P., Guerini-Rocco, E., Martelotto, L. G., Geyer, F. C., Bidard, F.-C., Berman, S., Fusco, N., Sakr, R. A., et al. (2016). The genomic landscape of male breast cancers. Clin Cancer Res *22*, 4045–4056.

84. Anderson, W. F., Althuis, M. D., Brinton, L. A., and Devesa, S. S. (2004). Is male breast cancer similar or different than female breast cancer? Breast Cancer Res Treat *83*, 77–86.

85. Ge, Y., Sneige, N., Eltorky, M. A., Wang, Z., Lin, E., Gong, Y., and Guo, M. (2009). Immunohistochemical characterization of subtypes of male breast carcinoma. Breast Cancer Res *11*, R28.

86. Sánchez-Muñoz, A., Román-Jobacho, A., Pérez-Villa, L., Sánchez-Rovira, P., Miramón, J., Pérez, D., Sáez, M.-I., de Luque, V., Medina, L., Ramírez-Tortosa, C. L., et al. (2012). Male breast cancer: immunohistochemical subtypes and clinical outcome characterization. Oncology *83*, 228–233.

87. Kornegoor, R., Verschuur-Maes, A. H. J., Buerger, H., Hogenes, M. C. H., de Bruin, P. C., Oudejans, J. J., van der Groep, P., Hinrichs, B., and van Diest, P. J. (2012). Molecular subtyping of male breast cancer by immunohistochemistry. Mod Pathol *25*, 398–404.

88. Shaaban, A. M., Ball, G. R., Brannan, R. A., Cserni, G., Di Benedetto, A., Dent, J., Fulford, L., Honarpisheh, H., Jordan, L., Jones, J. L., et al. (2012). A comparative biomarker study of 514 matched cases of male and female breast cancer reveals gender-specific biological differences. Breast Cancer Res Treat *133*, 949–958.

89. Yu, X.-F., Feng, W.-L., Miao, L.-L., Chen, B., and Yang, H.-J. (2013). The prognostic significance of molecular subtype for male breast cancer: a 10-year retrospective study. Breast *22*, 824–827.

90. Leone, J. P., Leone, J., Zwenger, A. O., Iturbe, J., Vallejo, C. T., and Leone, B. A. (2015). Prognostic significance of tumor subtypes in male breast cancer: a population-based study. Breast Cancer Res Treat *152*, 601–609.

91. Blows, F. M., Driver, K. E., Schmidt, M. K., Broeks, A., van Leeuwen, F. E., Wesseling, J., Cheang, M. C., Gelmon, K., Nielsen, T. O., Blomqvist, C., et al. (2010). Subtyping of breast cancer by immunohistochemistry to investigate a

relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS Med *7*, e1000279.

92. Johansson, I., Nilsson, C., Berglund, P., Strand, C., Jönsson, G., Staaf, J., Ringnér, M., Nevanlinna, H., Barkardottir, R. B., Borg, A., et al. (2011). High-resolution genomic profiling of male breast cancer reveals differences hidden behind the similarities with female breast cancer. Breast Cancer Res Treat *129*, 747–760.

93. Johansson, I., Nilsson, C., Berglund, P., Lauss, M., Ringnér, M., Olsson, H., Luts, L., Sim, E., Thorstensson, S., Fjällskog, M.-L., et al. (2012). Gene expression profiling of primary male breast cancers reveals two unique subgroups and identifies N-acetyltransferase-1 (NAT1) as a novel prognostic biomarker. Breast Cancer Res *14*, R31.

94. Chaffer, C. L., and Weinberg, R. A. (2011). A perspective on cancer cell metastasis. Science *331*, 1559–1564.

95. Cardoso, F., Costa, A., Senkus, E., Aapro, M., André, F., Barrios, C. H., Bergh, J., Bhattacharyya, G., Biganzoli, L., Cardoso, M. J., et al. (2017). 3rd ESO–ESMO international consensus guidelines for advanced breast cancer (ABC 3). Ann Oncol *28*, 3111–3111.

96. Yates, L. R., Knappskog, S., Wedge, D., Farmery, J. H. R., Gonzalez, S., Martincorena, I., Alexandrov, L. B., Van Loo, P., Haugland, H. K., Lilleng, P. K., et al. (2017). Genomic evolution of breast cancer metastasis and relapse. Cancer Cell *32*, 169–184.e7.

97. Paget, S. (1989). The distribution of secondary growths in cancer of the breast. 1889. Cancer Metastasis Rev *8*, 98–101.

98. Hunter, K. W., Crawford, N. P. S., and Alsarraj, J. (2008). Mechanisms of metastasis. Breast Cancer Res *10 Suppl 1*, S2.

99. Lifsted, T., Le Voyer, T., Williams, M., Muller, W., Klein-Szanto, A., Buetow, K. H., and Hunter, K. W. (1998). Identification of inbred mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression. Int J Cancer *77*, 640–644.

100. Hartman, M., Lindström, L., Dickman, P. W., Adami, H.-O., Hall, P., and Czene, K. (2007). Is breast cancer prognosis inherited? Breast Cancer Res *9*, R39.

101. Hemminki, K., Ji, J., Försti, A., Sundquist, J., and Lenner, P. (2008). Survival in breast cancer is familial. Breast Cancer Res Treat *110*, 177–182.

102. Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.

103. Reichardt, J. K. V., and Mehrian-Shai, R. (2009). The future of the human SNP identification: which individuals to sequence? Proc Natl Acad Sci U S A *106*, E50.

104. Hammond, M. E. H., Hayes, D. F., Dowsett, M., Allred, D. C., Hagerty, K. L., Badve, S., Fitzgibbons, P. L., Francis, G., Goldstein, N. S., Hayes, M., et al. (2010). American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). Arch Pathol Lab Med *134*, e48–72.

105. Wolff, A. C., Hammond, M. E. H., Hicks, D. G., Dowsett, M., McShane, L. M., Allison, K. H., Allred, D. C., Bartlett, J. M. S., Bilous, M., Fitzgibbons, P., et al. (2013). Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. J Clin Oncol *31*, 3997–4013.

106. Ward, L. D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res *40*, D930–4.

107. Davis, N. A., Crowe, J. E., Pajewski, N. M., and McKinney, B. A. (2010). Surfing a genetic association interaction network to identify modulators of antibody response to smallpox vaccine. Genes Immun *11*, 630–636.

108. Csardi, G., and Nepusz, T. (2006). The Igraph Software Package for Complex Network Research. InterJournal *Complex Systems*, 1695.

109. Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S., and Cheung, V. G. (2004). Genetic analysis of genome-wide variation in human gene expression. Nature *430*, 743–747.

110. Lukes, L., Crawford, N. P. S., Walker, R., and Hunter, K. W. (2009). The origins of breast cancer prognostic gene expression profiles. Cancer Res *69*, 310–318.

111. Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z., and Bergmann, S. (2016). Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nat Methods *13*, 366–370.

112. Stark, R., and Norden, J. (2016). SigCheck: Check a gene signature's prognostic performance against random signatures, known signatures, and permuted data/metadata.

113. Ogston, K. N., Miller, I. D., Payne, S., Hutcheon, A. W., Sarkar, T. K., Smith, I., Schofield, A., and Heys, S. D. (2003). A new histological grading system to assess response of breast cancers to primary chemotherapy: prognostic significance and survival. Breast *12*, 320–327.

114. Nielsen, T. O., Hsu, F. D., Jensen, K., Cheang, M., Karaca, G., Hu, Z., Hernandez-Boussard, T., Livasy, C., Cowan, D., Dressler, L., et al. (2004). Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res *10*, 5367–5374.

115. Subramanian, A., Salhab, M., and Mokbel, K. (2008). Oestrogen producing enzymes and mammary carcinogenesis: a review. Breast Cancer Res Treat *111*, 191–202.

116. Callari, M., Cappelletti, V., De Cecco, L., Musella, V., Miodini, P., Veneroni, S., Gariboldi, M., Pierotti, M. A., and Daidone, M. G. (2011). Gene expression analysis reveals a different transcriptomic landscape in female and male breast cancer. Breast Cancer Res Treat *127*, 601–610.

117. Prat, A., Cheang, M. C. U., Galván, P., Nuciforo, P., Paré, L., Adamo, B., Muñoz, M., Viladot, M., Press, M. F., Gagnon, R., et al. (2016). Prognostic Value of Intrinsic Subtypes in Hormone Receptor-Positive Metastatic Breast Cancer Treated With Letrozole With or Without Lapatinib. JAMA Oncol. *2*, 1287–1294.

118. Boyle, E. A., Li, Y. I., and Pritchard, J. K. (2017). An expanded view of complex traits: from polygenic to omnigenic. Cell *169*, 1177–1186.

119. Hu, T., Sinnott-Armstrong, N. A., Kiralis, J. W., Andrew, A. S., Karagas, M. R., and Moore, J. H. (2011). Characterizing genetic interactions in human disease association studies using statistical epistasis networks. BMC Bioinformatics *12*, 364.

120. Lareau, C. A., and McKinney, B. A. (2015). Network theory for data-driven epistasis networks. Methods Mol Biol *1253*, 285–300.

121. Lefebvre, C., Bachelot, T., Filleron, T., Pedrero, M., Campone, M., Soria, J.-C., Massard, C., Lévy, C., Arnedos, M., Lacroix-Triki, M., et al. (2016). Mutational profile of metastatic breast cancers: A retrospective analysis. PLoS Med *13*, e1002201.

122. Yang, Y., Quan, L., and Ling, Y. (2017). RBMS3 inhibits the proliferation and metastasis of breast cancer cells. Oncol Res.

123. Han, M.-R., Deming-Halverson, S., Cai, Q., Wen, W., Shrubsole, M. J., Shu, X.-O., Zheng, W., and Long, J. (2015). Evaluating 17 breast cancer susceptibility loci in the Nashville breast health study. Breast Cancer *22*, 544–551.

124. Zhang, C., Wu, M., Zhang, L., Shang, L.-R., Fang, J.-H., and Zhuang, S.-M. (2016). Fibrotic microenvironment promotes the metastatic seeding of tumor cells via activating the fibronectin 1/secreted phosphoprotein 1-integrin signaling. Oncotarget *7*, 45702–45714.

125. Ribelles, N., Perez-Villa, L., Jerez, J. M., Pajares, B., Vicioso, L., Jimenez, B., de Luque, V., Franco, L., Gallego, E., Marquez, A., et al. (2013). Pattern of recurrence of early breast cancer is different according to intrinsic subtype and proliferation index. Breast Cancer Res *15*, R98.

126. Senkus, E., Kyriakides, S., Ohno, S., Penault-Llorca, F., Poortmans, P., Rutgers, E., Zackrisson, S., Cardoso, F., and ESMO Guidelines Committee (2015). Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol *26 Suppl 5*, v8–30.

127. Gradishar, W. J., Anderson, B. O., Balassanian, R., Blair, S. L., Burstein, H. J., Cyr, A., Elias, A. D., Farrar, W. B., Forero, A., Giordano, S. H., et al. (2016). Invasive breast cancer version 1.2016, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw *14*, 324–354.

128. Da Silva, T. L. (2016). Male breast cancer: Medical and psychological management in comparison to female breast cancer. A review. Cancer Treat. Commun. *7*, 23–34.

129. Finak, G., Bertos, N., Pepin, F., Sadekova, S., Souleimanova, M., Zhao, H., Chen, H., Omeroglu, G., Meterissian, S., Omeroglu, A., et al. (2008). Stromal gene expression predicts clinical outcome in breast cancer. Nat Med *14*, 518–527.

130. Bergamaschi, A., Tagliabue, E., Sørlie, T., Naume, B., Triulzi, T., Orlandi, R., Russnes, H. G., Nesland, J. M., Tammi, R., Auvinen, P., et al. (2008).

Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. J Pathol *214*, 357–367.

131. Sousa, B., Moser, E., and Cardoso, F. (2013). An update on male breast cancer and future directions for research and treatment. Eur J Pharmacol *717*, 71–83.

132. Engebraaten, O., Vollan, H. K. M., and Børresen-Dale, A.-L. (2013). Triple-negative breast cancer and the need for new therapeutic targets. Am J Pathol *183*, 1064–1074.

133. Schroeder, A., Heller, D. A., Winslow, M. M., Dahlman, J. E., Pratt, G. W., Langer, R., Jacks, T., and Anderson, D. G. (2011). Treating metastatic cancer with nanotechnology. Nat Rev Cancer *12*, 39–50.

134. Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc *4*, 1184–1191.