bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Complex polymorphisms in endocytosis genes suggest alpha-cyclodextrin against metastases in breast cancer

Alpha-cyclodextrin as a treatment for breast cancer

KNUT M. WITTKOWSKI[1*], CHRISTINA DADURIAN[1], HAN SANG KIM[2], AYUKO HOSHINO[2], DAVID LYDEN[2]

[1] Center for Clinical and Translational Science, The Rockefeller University, New York, New York, United States of America.

[2] Department of Pediatrics, and Cell and Developmental Biology Weill Medical College of Cornell University, New York, New York, United States of America

* Corresponding author

E-mail: kmw@rockefeller.edu (KMW)

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract (300 Words)

Most breast cancer (BC) deaths are caused by metastasis and treatment options beyond radiation and cytotoxic drugs are urgently needed.

This study reanalyzed existing data from genome-wide association studies (GWAS) using a novel computational biostatistics approach (muGWAS), which had been validated in epilepsy and identified a potential drug against disruption of language development in autism. Compared to many traditional bioinformatics approaches, muGWAS jointly analyzes several neighboring single nucleotide polymorphisms while incorporating knowledge about genetics of heritable diseases into the statistical method and about GWAS into the rules for determining adaptive genome-wide significance in samples of 600–2000 subjects only.

Results from three independent breast cancer GWAS not only confirmed cell-cycle control and receptor/AKT signaling, but, for the first time in GWAS, also consistently identified many endo-/exocytosis (EEC) pathway genes, most of which had already been observed in functional and expression studies of breast cancer, including genes that translocate (*ATP8A1*, *ATP8B1, ANO4, ABCA1*) and metabolize (*AGPAT3*, *AGPAT4*, *DGKQ*, *LPPR1*) phospholipids entering the phosphatidylinositol cycle, which controls EEC. These novel findings suggest scavenging phospholipids via alpha-cyclodextrin (αCD) as a potential intervention to control packaging of exosomes (which prepare distant microenvironment for organ-specific metastases) and endocytosis of β1 integrins (which are required for mesenchymal migration of tumor cells).

Beta-cyclodextrin (βCD) had already been shown to be effective *in vitro* and animal studies of breast cancer, but exhibits cholesterol-related ototoxicity. The smaller αCD also scavenges phospholipids, but cannot fit cholesterol. An *in-vitro* study presented here confirmed hydroxypropyl (HP)-αCD to be at least twice as effective as HPβCD against migration of human cells of both receptor negative and estrogen-receptor positive breast cancer.

If the previous successful animal studies with βCDs are replicated with the safer and more effective αCDs, clinical trials of αCDs are warranted in women with triple-negative breast cancer, who have few treatment options and poor prognosis.

Author Summary (200 Words)

Although a family history of breast cancer (BC) substantially increases breast cancer risk, the known mutations in *BRCA1/2* explain only 5–10% of all cases. Patients with hormone-receptor positive cancer initially respond well to anti-hormone therapy, but later have as limited treatment options as women with triple-negative breast cancer. Taking advantage of increases in computer memory since the advent of 32-bit operating system in 2001, we use a novel computational biostatistics approach, which does not screen for individual genetic “letters” (SNPs, which would typically be selected against if they increased risk), but for genetic “words” (combinations of neighboring SNPs, for which selection is less effective). This “nonparametric” approach also avoids false positive results due to unrealistic assumptions. Finally, we adjust decision criteria to the specifics of the study. In combination, these methodological advances confirm hormone-receptor signaling and transcriptional control consistently across three independent studies, but also point to upregulation of tumor cell migration as a novel common risk factor. In an *in vitro* study, alpha-cyclodextrin (αCD) was twice as effective as βCD in reducing migration of breast cancer cells. Hence preventing metastases with αCD might reduce breast cancer deaths with fewer side effects (including hair loss) than cytotoxic drugs and radiation.

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table of Content

Introduction **4**

Materials and Methods **4**

Ethics Statement 4
Subjects 4
Statistical Methods 5
In vitro Assay 7

Results **8**

Additional ssGWAS CGEM results complement known breast cancer risk factors 8
Novel ssGWAS aGWS results in EPIC and PBCS complement CGEM results. 8
muGWAS aGWS results are cross-validated across CGEM, EPIC, and PBCS. 10
muGWAS results confirm known signaling and nuclear disease pathways in BC 10
muGWAS results highlight Endo-/Exocytosis (EEC) as a disease pathway in BC. 12
muGWAS identified genes causing dysfunction of EEC, a known BC risk factor. 12
muGWAS identifies PI cycle dysregulation as novel breast cancer risk factor 14
Results for EEC regulation and function are consistent across populations 15
PI supply into the PI cycle as a drug target in BC 15
HPaCD is more effective than HPbCD against migration of breast cancer cells 15

Discussion **17**

Replication and complementation of previously identified genes 17
Computational biostatistics approach to genetic data 17
Replication of findings across populations 18
Dysregulated EEC in breast cancer metastasis, angiogenesis, and progression 18
The PI cycle in Breast Cancer 19
βCDs are effective in cancer models of migration, invasion, and angiogenesis 20
Migration and invasion in BC involves processes unrelated to cholesterol 20
αCD specifically scavenge PLs 21
HPαCD as a novel treatment in breast cancer 21

Figure Captions **29**

References **23**

S1 Supplementary Material **1**

S1 Figures 1
S1 Tables 5
S1 References 8
S1 Regional Plots 8

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

x

Introduction

Breast cancer (BC) is the most common cancer in women worldwide.[Rojas 2016] In 2016, 246,660 new U.S. cases were estimated.[Siegel 2016] The highly penetrant, but rare mutations in *BRCA1* and *BRCA2* point to DNA repair deficiencies as a factor in breast cancer etiology, but explain only 5 to 10 percent of all breast cancer cases. Patients with breast cancer positive for estrogen receptor (ER) or human epidermal growth factor (GF) receptor type 2 (*HER2*) initially respond well to anti-estrogen or anti-HER2 therapy, respectively, but inevitably become refractory.[Hayashi 2015] Moreover, triple-negative breast cancer, which also lacks expression of progesterone receptor (PR), harbors an aggressive clinical phenotype with limited treatment options.

In women of European ancestry, 127 single nucleotide polymorphisms (SNPs) have been associated with breast cancer [Burdett] at the conventional fixed $s = -\log(p) = 7.3$ level for genome-wide statistical significance (GWS)[Barsh 2012] (s is used throughout for significance). These SNPs map to 51 genes with known function; all but 16 involved in three known pathways: 27 genes are associated with nuclear function (DNA repair, transcription, cell-cycle control). Six genes (*FGFR2*, *PTHLH*, *ESR1*, *ITPR1*, *KCNN4, TGFBR2*) are involved in receptor signaling, ion channels, and mammary gland development (KEGG *Homo sapience* pathway hsa04915) and two genes (*MAP3K1, TRAPPC2*) are associated with AKT signaling (hsa04064).[Kendellen 2014]

The present evaluation is based on separate analyses of three independent populations of women of European ancestry (about 50% cases in all populations, see Subjects). Two of the populations (EPIC, PBCS) had never been analyzed individually, because their sample size was deemed insufficient for conventional statistical approaches.

Most breast cancer death are not due to the primary tumor, but to metastases, often in the bone, lung, liver, and brain. Our analysis generates a testable biological hypotheses and suggests a novel treatment strategy, which is confirmed in an *in-vitro* analysis. Our results implicate dysregulation and dysfunction of endo-/exocytosis (EEC), which is involved in cell migration and invasion, as well as organ targeting, and, thus, suggest overall downregulation of phosphoinositides (PI) via (derivatives of) alpha-cyclodextrin (αCD) as a potential avenue to prevent metastases without the side effects of radiation and cytotoxic therapies.

Materials and Methods

Ethics Statement

The study was approved by The Rockefeller University IRB on Aug 24, 2015 (ref# 330390, exempt).

Subjects

This reanalysis is based on data from three GWAS in women of European ancestry:

(a) the NHS cases from the Nurses' Health Study as part of the Cancer Genetic Markers project (CGEM, phs000147/39389-2/GRU, 1145 cases / 1142 controls), [Hunter 2007; Haiman 2011]

(b) estrogen-negative (ER-neg) cases from the nested case-control study of the European Prospective Investigation into Cancer (EPIC, phs000812/39395-2/HMB-PU, 511 cases / 500 controls),[Siddiq 2012]

(c) ER-neg cases from the Polish Breast Cancer Case-Control Study (PBCS, phs000812/39397-2, 543 cases / 511 controls),[Siddiq 2012]

The study was approved by The Rockefeller University IRB on Aug 24, 2015 (ref# 330390, exempt).

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Statistical Methods

In this analysis, conventional single-SNP GWAS (ssGWAS) are complemented with a computational biostatistics approach (muGWAS, GWAS using muStat [Wittkowski 2012]) that incorporates knowledge about genetics into the method [Wittkowski 2010, Sections 4.3.4 and 4.4.2; Wittkowski 2013] and knowledge about the nature of GWAS into the decision strategy.[Wittkowski 2014]

Statistical methods tend to have higher power if they are based on more realistic assumptions, which, in biology, tend to be weak. In contrast, methods based on stronger assumptions, such as additivity of allelic effects and independence of SNPs within an LD block (LDB), may generate more significant results when errors happen to fulfill these assumptions than for true effects. With millions of test statistics calculated, even a rare false positive result due to model-misspecification (1/10,000 tests, say), may result in the 100 most significant results all being false positives. U-statistics for multivariate data in GWAS (muGWAS) rely only on weak, realistic assumptions, but require large amounts of memory and GPU enabled cloud instances, which became available only after 2001 and 2009, respectively.

After excluding non-informative or low-quality SNPs and SNPs in high LD with an immediate neighbor [Ioannidis 2009] (20–25%) to avoid "loss of power [when] including irrelevant SNPs".[Li 2012], an initial traditional ssGWAS was performed, using the u-test for univariate data. [Mann 1947; Wilcoxon 1954; Kruskal 1957] The same data was then analyzed using a u-test for genetically structured multivariate data.[Wittkowski 2013] U-statistics avoid model-misspecification biases by replacing linear/logistic[Wu 2010b] with non-parametric kernels.[Li 2012]

Below, we describe the assumptions about genetics and GWAS that are implemented in the statistical method and decision strategy and refer to published empirical validation of this approach.

1.1 Heterodominance: The degree of dominance at a particular SNP is not assumed to be either recessive (aA = aa), additive (aA = (aa+AA)/2), or dominant (aA = AA), but merely monotonic (aa < aA < AA). Accordingly, the information contributed by a particular SNP is represented as a matrix containing for each of the $n \times n$ pairs of n subjects whether the genetic risk carried by the row subject is lower, the same, or higher than the risk of column subject, or unknown ("?") in case of missing data in one or both of the subjects. Below, the possible genetic risk constellations (left) are compared to models with different degrees of dominance (right). While the left matrix is similar to the matrix for dominant effects (all non-zero elements are ±2), the inequalities are not (numerically) equivalent. In effect, the single-SNP results based on the adaptive u-scores for aa, aA, and AA are similar to results from the Cochran-Armitage test for additive codominance, [Cochran 1954; Armitage 1955] which uses fixed scores 0, 1, and 2.

X~Y	aa	aA	AA	??
aa	=	<	<	?
aA	>	=	<	?
AA	>	>	=	?
??	?	?	?	?

X–Y	aa	aA	AA	??
aa	±0	+2/+1/±0	+2	?
aA	±0/–1/–2	±0	+2/+1/±0	?
AA	–2	±0/–1/–2	±0	?
??	?	?	?	?

1.2 LD-structure: A basic assumption underlying GWAS, in general, is that a disease locus should be in LD with both neighboring SNPs (unless they are separated by a recombination hotspot). Hence, the information from two neighboring SNPs is AND-ed (but not ADD-ed) using the function Λ, where Λ(x,y) is "?" if x and y are conflicting ("<", ">") or both "?", else "=" for a combination of "=", and "?", else "<"/">" for a combination of "<"/">", "=", or "?":

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

S_k	<	<	<	<	=	=	=	=	>	>	>	>	?	?	?	?
S_{k+1}	<	=	>	?	<	=	>	?	<	=	>	?	<	=	>	?
$I_{k,\,k+1} = \Lambda(S_k,\, S_{k+1})$	<	<	?	<	<	=	>	=	?	>	>	>	<	=	>	?

Non-informative SNPs are added between LDBs to prevent intervals from spanning LDBs.

1.3 Cis-epistasis, including compound-heterozygosity: To account for interactions between functional polymorphisms,[Aslibekyan 2013a] a natural extension of Λ is then used to combine information from corresponding elements of the $n \times n$ matrices containing information about neighboring pairs. Assuming, for the sake of simplicity, the case of only four SNPs within in the same LD block, the aggregated diplotype information for one pair of subjects is

$$\Lambda(I_{k,k+1}, \ldots, I_{k+2,k+3}) = \Lambda(\Lambda(S_k, S_{k+1}), \Lambda(S_{k+1}, S_{k+2}), \Lambda(S_{k+2}, S_{k+3})) \neq \Lambda(S_k, \ldots, S_{k+3})\ ),$$

which can be one of the following:

| | | | | | | | | | | | | | | | |

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

tial interaction terms (Wittkowski 2013)), where many dependent tests are performed at overlapping window positions, the expected QR curve (see S1 Fig 3) may be even more convex. The expected distribution curve is estimated from the 50% of chromosomes with the fewest outliers rising above a convex fit.(Wittkowski 2014) The empirical adaptive (study-specific) aGWS cut-off is the median apex (highest point) of a convex curve fitted against these chromosomes' QR plot. For the purpose of validation, results are considered "replicated" if they are significant at the aGWS/2 level.

2.2. Replication: Complex diseases may involve different SNPs in high LD with causal variants across populations,(Pickrell 2016) epistasis between several SNPs per locus, several loci per gene, and several genes per function, with risk factors differing across populations (see above). Hence, we will consider SNPs within a locus, loci within a gene, and genes with a direct mechanistic relationship (paralogs, binding partners, ...) for replication. (Peng 2010; Aslibekyan 2013a)

Validation: The above approaches have been validated in two published analyses, where previous analyses using ssGWAS and fixed GWS also had identified not more than a few apparently unrelated SNPs.

- In epilepsy,(Wittkowski 2013) muGWAS confirmed the Ras pathway and known drug targets (ion channels, *IL1B*). In that analysis, muGWAS was also compared with a parametric analogue, logistic regression with interaction terms for neighboring SNPs (lrGWAS). muGWAS produced fewer apparent false positives (isolated highly significant results far away from coding regions) (Wittkowski 2013, Suppl. Fig 2) and higher sensitivity for genes downstream of Ras, which are involved in more complex cis-epistatic interactions,(Wittkowski 2013, Fig 3, blue) than ion channels, which were also implicated by lrGWAS.(Wittkowski 2013, Fig 3, red)
- In autism,(Wittkowski 2014) muGWAS identified sets of mechanistically related genetic risk factors for mutism in autism (independently confirmed in functional studies (Guglielmi 2015) and a pathway network analysis (Wen 2016)). In this analysis, adaptive GWS was validated against three analyses with randomly permutated phenotypes. Only one gene (DMD, not aGWS) appeared in one of the other analyses (also not aGWS). Moreover, there is no noticeable overlap between aGWS genes between breast cancer and either mutism (Wittkowski 2014) or epilepsy (Wittkowski 2014, Suppl. Fig 7), while there is considerable functional overlap between mutism in autism and epilepsy.

In vitro Assay

A 24-well plate (CBA-120, Cell BioLabs Inc., San Diego, CA) with CytoSelect Wound Healing Inserts was warmed up at room temperature for 10 min. A cell suspension used contained 0.5–1.0 × 10^6 cells/ml in media containing 10% fetal bovine serum (FBS) was prepared and 1 mL of this suspension was added to each well. Cells were then incubated for 12 h, after which time the insert was removed and cells were washed with new media to remove dead cells and debris. FBS with/without CDs (Sigma-Aldridge, St. Louis, MO) was added to start the wound healing process. Cells were incubated for 2 h, washed with PBS, fresh control media was added, and cells were incubated for another 12 h. After removing the fixation solution, 400 µL of Cell Stain Solution were added to each well and incubated for 15 min at room temperature, after which stained wells were washed thrice with deionized water and left to dry at room temperature. Cells that migrated into the wounded area or protruded from the border of the wound were visualized and photographed under an inverted microscope to determine migrated cell surface area.

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Additional ssGWAS CGEM results complement known breast cancer risk factors

The original CGEM analysis had identified the fibroblast GF receptor *FGFR2* [Entrez Gene 2263] (rs1219648: $s = 5.49$, rs2420946: 5.46),[Hunter 2007] which affects mammary epithelial cell growth and migration.[Czaplinska 2014] The second SNP identified in that analysis was subsequently located to a long variant of the mitotic checkpoint protein *BUB3* [9184] (rs10510126: 6.25, >1 MB apart from *FGFR2*). These two genes are also the only genes in the present analysis with one and two SNPs, respectively, slightly above the diagonal in the summary ssGWAS quantile-rank (QR, often: QQ) plot (Fig 1 left), although the QR plots of several individual chromosomes show evidence for association in chromosomes 4 (the *SNCA–MMRN1* [22915] region), 5 (breast cancer associated transcript *BRCAT54* [100506674], non-coding), 6 (*PARK2* [5071], the Parkinson's disease [PD] ubiquitin ligase Parkin), and 9 (*LPPR1* [54886], phospholipid [PL] phosphatase-related 1) (S1 Fig 2).

In the present analysis, a total of 22 genes and *BRCAT49*[Iyer 2015] reached aGWS in CGEM (Fig 1, left, blue). A total of 21, 11, and 24 genes with known function or relation to breast cancer exceeded muGWAS aGWS in CGEM, EPIC, and PBCS, respectively.

Novel ssGWAS aGWS results in EPIC and PBCS complement CGEM results.

In EPIC, the two most (statistically) significant SNPs (rs4791889, 5.66 and rs9596958, 5.42) are located 4.5 kB upstream of the chromodomain helicase DNA binding protein *CHD3* [1107] and the transcription factor (TF) *SOHLH2* [54937], respectively (see S1 Table 2 and S1 Fig 5).

In PBCS, the two most significant SNPs (rs2297075, 5.83, rs943628, 5.55, 100 kB apart) are located in *DOCK8* [81704], a guanine nucleotide exchange factor for Rac1, which drives mesenchymal cell movement.[Wang 2015b] Significance of *FGFR2* relies on the above two previously reported and a third SNP (rs11200014) within intron 2.[Cui 2016] Significance in *BUB3* is driven by three SNPs in high LD (rs10510126, rs17663978, rs7916600, spanning 30 kB). These findings are consistent with significance of the top five genes in ssGWAS depending on a single polymorphism each. Lack of evidence in the other populations (S1 Table 2) is consistent with different variations developing in different parts of Europe (Iceland vs. Sicily, Portugal vs Azerbaijan).

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

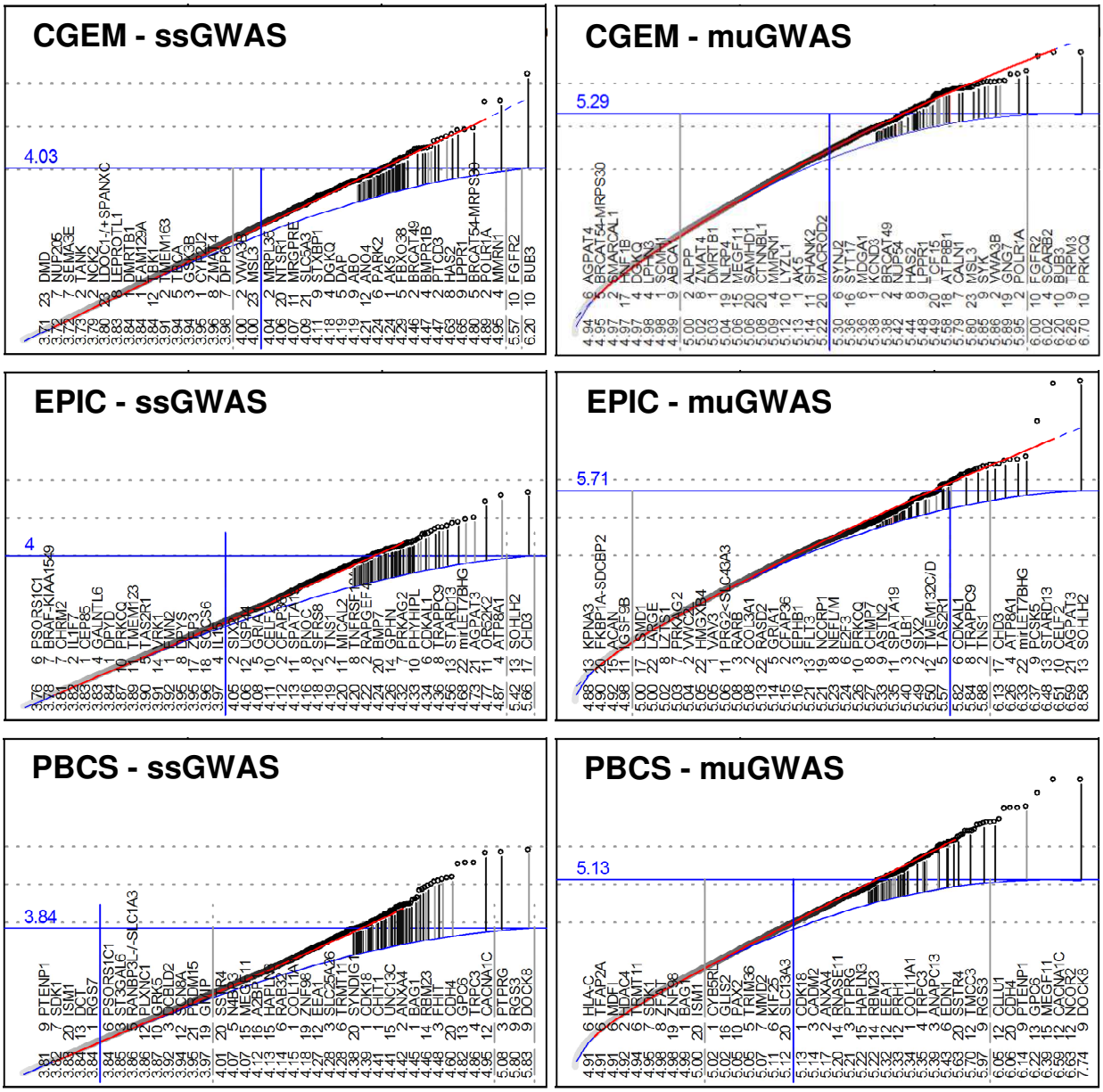

Fig 1: GWAS Quantile-Rank (QR) plots.
Left: ssGWAS, right: muGWAS (each point represents the most significant result among all diplotypes centered at the same SNP) Results are ranked by significance (bottom). For the most significant results and other results of interest, the location of SNPs to genes is shown in S1 Fig 5. Upper curve (red): convex fit against points; dashed extension: projection; lower curve (blue): population-specific expectation. Vertical lines between curves connect the highest *s*-values ($-\log_{10} p$) of a gene (dot) with its expected value for genes with known function. Light gray vertical lines indicate genes omitted from the list due to low reliability (either low µIC or reliance on a single SNP), respectively. Genes to the right of the vertical dark line are above the aGWS cut-off. See S1 Fig 1 for Manhattan plots. The horizontal solid line at highest point at the end of the expected curve indicates the estimate for adjusted GWS (aGWS). All points above the horizontal line (and genes to the right of the vertical blue line) are "significant" at the aGWS level.

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

muGWAS aGWS results are cross-validated across CGEM, EPIC, and PBCS.

In CGEM, the top gene was the PL/diacylglycerol (DAG)-dependent protein kinase *PRKCQ* [5588] (chr10: 6,540,724-6,573,883), which induces cell migration and invasion.[(Belguise 2012; Zafar 2014)] The same SNP (rs661891) was also implicated in EPIC. The three most significant SNPs and the most significant regions in muGWAS were all located within the same 34 kB LD block. The second most significant gene was a long EST of the transient receptor potential cation channel *TRPM3* [80036], which controls oncogenic autophagy in renal cell carcinoma,[(Hall 2014)] supported by a part of the promoter region of the shorter main form in PBCS. *BUB3* was also significant in -muGWAS, followed by the endo-/lysosomal receptor *SCARB2* [950] and the nuclear RNA polymerase subunit *POLR1A* [25885] (rs10779967).

In EPIC, the top gene in muGWAS (and in ssGWAS, see above), was the transcription factor (TF) *SOHLH2*, followed by *AGPAT3* [56894] (rs8132053 in CGEM and EPIC), whose paralog *AGPAT4* [56895] is included in Fig 1 (4.94, right panel). *CELF2* [10659], a RNA binding protein, and *STARD13* [10948], a breast cancer tumor suppressor that regulates cell migration and invasion[(Hanna 2014)] also reached aGWS. *CHD* [364663] depends entirely on SNP rs4791889 (see Statistical Methods, 2.2. Replication, for replication criteria).

In PBCS, the top gene in muGWAS, as in ssGWAS, was *DOCK8* [81704], followed by the nuclear receptor coreceptor *NCOR2* [9612], which has been implicated in tamoxifen resistance in breast cancer.[(van Agthoven 2009; Zhang 2013b)] *CACNA1C* [775] (3rd) is highly up-regulated in breast cancer.[(Wang 2015a)] The multiple epidermal GF-like domains protein 11 (*MEGF11* [84465], 4th), like *MEGF10* [84466] an ortholog of *C. elegans* Ced-1 and the *Drosophila* draper, had been implicated in colorectal cancer. [(Cicek 2012)]

Both CGEM and EPIC have a significant P-type ATPase, which import phosphatidylserine (PS, *ATP8B1* [5205]) and phosphatidylcholine (PC, *ATP8A1* [10396]), respectively, the substrates for phospholipase D (PLD) to produce phosphatidic acid (PA) for the synthesis of phosphatidylinositol (PI).[(Daleke 2007)] *BMP7* [655] (ss: 4.24) and its receptor *BMPR1B* [658] (ss: 4.47) are significant in EPIC and CGEM, respectively, and BMP signaling is known to regulates mitotic checkpoint protein levels in human breast cancer cells, including levels of *BUB3* (see above).[(Yan 2012)] *DGKQ* [1609] (rs2290405) which converts DAG into PA, was replicated in CGEM and PBCS, while *LPPR1* [54886], which is involved in the conversion of PA into PI was replicated in CGEM and EPIC.

As expected in samples from the general population, the known risk factors for rare early-onset breast cancer (*BRCA1/2* [672/675], *HER2* [2064], *RB1* [5925]) do not show association and many receptor-related genes are absent in ER neg. populations. Except for the genes with highest significance in ssGWAS (*BUB3* in CGEM, *SOHL2* in EPIC, and *DOCK8* in PBCS), all of the aGWS genes in muGWAS have support in least one of the other two populations (2nd block of S1 Table 2). This observation is consistent with muGWAS identifying primarily old cis-epistatic variations, rather than *de novo* mutations favored by ssGWAS. S1 Table 2 gives an overview about the significance and replication of the genes identified and supportive evidence in the literature.

muGWAS results confirm known disease pathways in breast cancer

Consistent with the published results in the NHGRI-EBI catalog, a total of 16, 15, and 18 genes among the aGWS results in CGEM, EPIC, and PBCS, respectively, are involved in the three known disease pathways, such as membrane-associated receptor signaling (G protein–coupled receptors [GPCR], Fc receptors [FcR], hemagglutinin [HA], receptor tyrosine kinases [RTK], or ion channels), MAP kinases, and in nuclear proteins involved in cell cycle control, transcription, or splicing in breast cancer (Table 1).

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1: Breast cancer genes associated with pathways by Study.
Each block of columns shows the genes identified sorted by s-values in muGWAS (s6) down and ssGWAS (s1) up; top: genes above aGWS in muGWAS (CGEM: 5.29, EPIC: 5.71, PBCS: 5.13); bottom: genes with SNPs above aGWS in ssGWAS only (CGEM: 4.03, EPIC: 4.00, PBCS: 3.84, ssGWAS results for genes also implicated in muGWAS are shown next to the muGWAS results); center: other genes shown in Fig 1; Mbrn (green): membrane-associated genes (GPCR, FcR, HA, RTK, Ion channels), PI/EC (pink): PI cycle/EEC, MPK (violet): MAP kinases, Ncls (blue): nucleus (cell cycle control, transcription, splicing). See S1 Table 1 for Entrez Gene identifiers and S1 Table 2 for replication across populations, which is indicated in bold.

CGEM							EPIC							PBCS						
s6	s1	Othr	Mbrn	PI/EC	MPK	Ncls	s6	s1	Othr	Mbrn	PI/EC	MPK	Ncls	s6	s1	Othr	Mbrn	PI/EC	MPK	Ncls
6.70					**PRKCQ**		**8.58**	**5.42**					SOHLH2	**7.74**	**5.83**		DOCK8			
6.26			**TRPM3**				**6.59**	**4.73**			**AGPAT3**			**6.63**						**NCOR2**
6.20	**6.20**					BUB3	**6.51**	**4.11**					**CELF2**	**6.59**	**4.95**		CACNA1C			
6.02				**SCARB2**			**6.48**	**4.56**					**STARD13**	**6.39**	**4.07**			**MEGF11**		
6.00	**5.57**		**FGFR2**				**6.37**			PCSK5				**6.22**	**4.82**		GPC6			
5.95	**4.89**					POLR1A	**6.33**	**4.68**		mirLET7B				**6.14**				**PTENP1**		
5.89			GNG7				**6.26**	**4.87**			**ATP8A1**			**6.06**	**4.60**		CDH4			
5.86			VWA3B				**6.13**	**5.66**					CHD3	**6.05**			CLLU1			
5.85			SYK				**5.88**	**4.19**			**TNS1**			**5.97**	**5.80**		RGS3			
5.81						MSL3	**5.85**	**4.36**					**TRAPPC9**	**5.70**			TMCC3			
5.79			CALN1				**5.82**	**4.34**					CDKAL1	**5.63**			SSTR4			
5.58				**ATP8B1**										**5.43**			EDN1			
5.49						TCF15	**5.6**			TAS2R1				**5.39**						ANAPC13
5.48	**4.65**			**LPPR1**			5.50			TMEM132C/D				**5.35**	**4.86**		**TRPC3**			
5.44	**4.63**		HAS2				5.49	**4.05**					SIX2	**5.34**	**4.15**		COL11A1			
5.42						NUP54	5.40				GLB1			**5.33**			PRH1			
5.38	**4.46**		BRCAT49				5.35	**4.13**		SPATA19				**5.32**	**4.27**			**EEA1**		
5.38			KCND3				5.33				ASTN2			**5.22**	**4.46**					RBM23
5.37			MDGA1				5.27				CHMP7			**5.22**	**4.13**		HAPLN3			
5.36				SYT17			5.26						**PRKCQ**	**5.21**	**5.08**		PTPRG			
5.30				**SYNJ2**			5.24						E2F3	**5.21**			RNASE11			
							5.23			NEFL/M				**5.17**	**4.42**			ANXA4		
5.22						MACROD2	5.21						NCCRP1	**5.14**			CADM2			
5.14			SHANK2				5.21			FLT3				**5.13**	**4.39**					CDK18
5.13	**4.24**		AK5				5.16			EPHB1										
5.12			LYZL1				5.15	**4.12**		CFAP36				5.12			SLC13A3			
5.09	**4.95**		MMRN1				5.14	**4.08**		GRIA1				5.11			KIF25			
5.08						CTNNBL1	5.13			RASD2				5.07			MMD2			
5.06						SAMHD1	5.08			COL3A1				5.05			TRIM36			
5.06				**MEGF11**			5.08						RARB	5.05						PAX2
5.04				**NLRP4**			5.06			SLC43A3				5.02						GLIS2
5.03						DMRTB1	5.05				**VAV3**			5.02			CYB5RL			
5.02						ZMAT4	5.05						HMGXB4	5.00	**3.83**					ISM1
5.00			ALPP				5.04			VWC2L				4.99	**4.45**		BCL10			
4.99				**ABCA1**			5.03	**4.32**					**PRKAG2**	4.98	**4.19**		ZNF98			
4.98						SCMH1	5.02						LZTS1	4.98						ZFAT
4.98			LPHN3				5.00			LARGE				4.95			SDK1			
4.97						HNF1B	5.00			LSMD1				4.95	**4.29**		TRMT11			
4.97						SMARCAL1	4.98			IGSF9B				4.92						HDAC4
4.95	**4.81**		BRCAT54				4.93			ACAN				4.91						MDFI
4.94				**AGPAT4**			4.90			FKBP1A	SDCBP2			4.91						TFAP2A
4.94						ZBTB20	4.88			KPNA3				4.91			HLA-C			
4.52	**4.47**		PTCD3				4.77	**4.77**		OR52K2				4.61	**4.48**					FHIT
4.89	**4.47**		BMPR1B				4.64	**4.26**		GPHN				4.91	**4.41**			**UNC13C**		
4.29	**4.29**					FBXO38	4.83	**4.24**		BMP7				4.42	**4.41**		CHIT1			
4.86	**4.24**			**PARK2**			4.22	**4.22**			**RAPGEF4**			4.38	**4.38**		SYNDIG1			
4.21	**4.21**			ANO4			3.68	**4.20**		NFRSF10A				4.74	**4.28**		SLC25A26			
4.24	**4.19**		ABO				4.20	**4.20**		MICAL2				4.84	**4.14**			RAB32		
4.97	**4.18**			**DGKQ**			3.90	**4.18**					SFRS8	4.12	**4.12**					A2BP1
4.28	**4.19**		DAP				3.57	**4.16**		PNOC				4.59	**4.07**			N4BP3		
4.59	**4.11**			STXBP1			4.81	**4.06**		USP44										
4.49	**4.07**		MRGPRE																	
4.82	**4.06**		NTSR1																	
4.04	**4.04**		MRPL35																	

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

muGWAS results highlight Endo-/Exocytosis (EEC) as a pathway in breast cancer.

The cell's major fibronectin-binding integrin (α5β1) is key to the survival and migration of tumor cells.(Dozynkiewicz 2012) Results of various expression and functional studies have pointed to EEC of β1 integrins as a functional component of "derailed endocytosis" in cancers, including breast cancer (Fig 2).(Mosesson 2008; Morgan 2009; De Franceschi 2015).

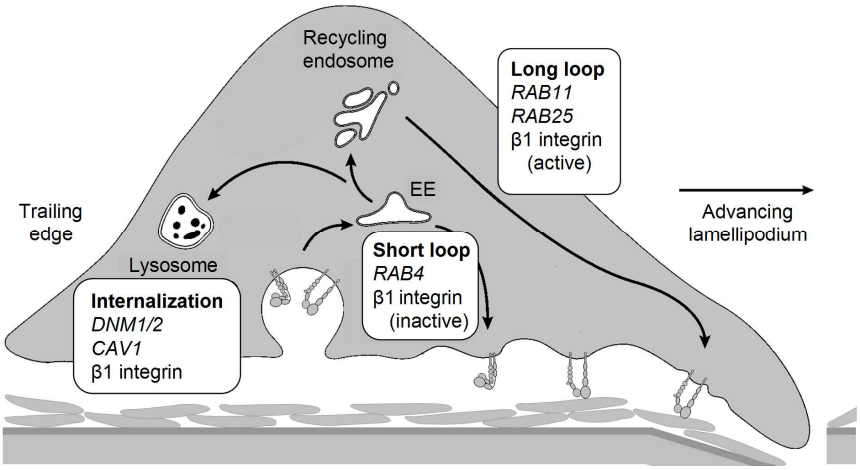

Fig 2: EEC of β1 Integrin underlying mesenchymal tumor cell migration and invasion.
Cell migration necessitates trafficking of b1 integrin, whose internalization is controlled by dynamin. Both clathrin- and caveolin 1 (CAV1)-coated domains of the plasma membrane are involved. Once in early endosomes (EE), integrins may be sorted for degradation in lysosomes, recycled to the plasma membrane through a RAB4-dependent route, or transported to the recycling endosome (RE). Recycling from the RE requires Rab11 family members, such as RAB25 which is often aberrantly expressed in human tumors, including luminal B breast cancer.(Mitra 2016) (adopted from (Mosesson 2008; Morgan 2009; De Franceschi 2015))

Among the 15 genes not associated with known pathways in the NHGRI-EBI catalog (excluding the ambiguous locus between *MDM4* [4194] and *PIK3C2B* [5287]), only four GWS genes are involved in EEC (*PDE4D*, *SNX32*, *STXBP4*, *DNAJC1*, marked with "*" in S1 Table 1), all from ssGWAS of a combined analysis of 9 studies,.(Michailidou 2013) including the three studies analyzed separately here. A String(http://string-db.org/) analysis of the subset of aGWS genes that are not part of the above three pathways identified two separate clusters related to EEC:

- **Function of EEC:** *PARK2*, *PTEN (from PTENP1)*, *SYNJ2*, *STXBP1*, *UNC13* (consistent with previous functional studies, see S1 Table 3))
- **Regulation of EEC:** *AGPAT3* and *DGKQ* (S1 Fig 4).

muGWAS identified genes causing dysfunction of EEC, a known BC risk factor.

Further subset analyses and a literature review by the authors identified additional aGWS genes related to EEC function (genes in parenthesis replaced in the String search by binding partner with known function in String). They include endocytosis (hsa04144): *DNM1* (from *MEGF11)*, *EEA1*, *PDE4D*, *SNX32*, *NEDD4* (from *N4BP3*) (FDR = .018) and synaptic vesicle cycle (hsa04721): *STXBP1*, *UNC13C, VAMP2*; (FDR = .0001).

Fig 3 integrates the genes identified in the present GWAS analysis (pink, see S1 Table 1 for details) with results from expression and functional studies of β1 integrin EEC in breast cancer (see S1 Table 3 for details).

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

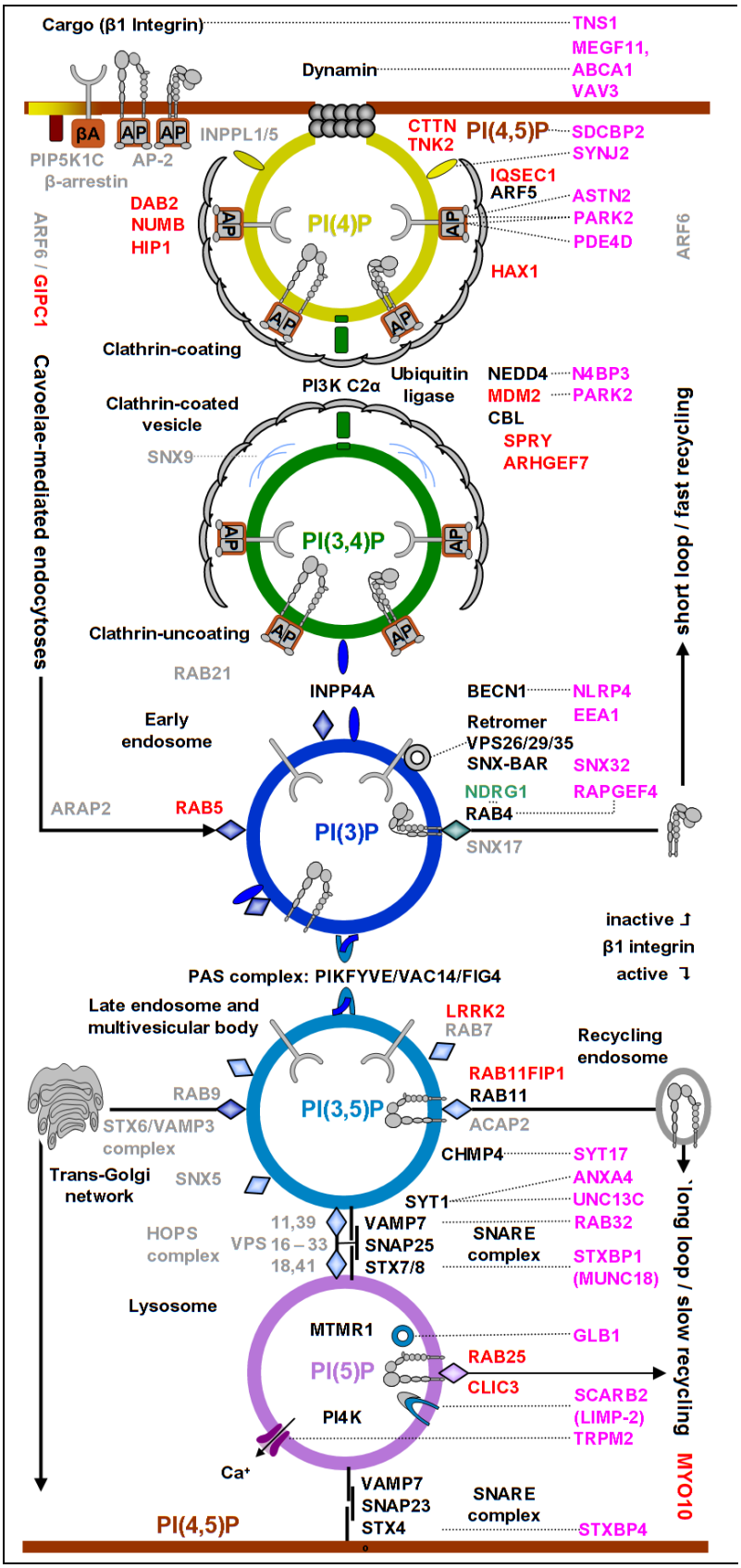

Fig 3: Endo-/exocytosis pathway.

Pink: genes identified in this analysis, most of which have been implicated in breast cancer (S1 Table 1), by stage of EEC: Formation of clathrin-coated vesicles, E3 ubiquitination, separation of inactive integrin (fast recycling) from active integrins (slow recycling), sorting between secretory, lysosomal, and (slow) recycling pathway, and lysosomal degradation. Red and green genes are known breast cancer promoters and suppressors, respectively (S1 Table 3).

Clathrin-mediated endocytosis (CME) begins with co-assembly of the hetero-tetrameric clathrin adaptor complex AP-2 with clathrin at PI(4,5)P_2-rich plasma membrane sites. AP-2 in its open conformation recruits clathrin and a number of additional endocytic proteins, many of which also bind to PI(4,5)P_2.

CCP maturation may be accompanied by SHIP-2-mediated dephosphorylation of PI(4,5)P_2 to PI(4)P. Synthesis of PI(3,4)P_2 is required for assembly of the PX-BAR domain protein SNX9 at constricting CCPs and may occur in parallel with PI(4,5)P_2 hydrolysis to PI(4)P via synaptojanin, thereby facilitating auxilin-dependent vesicle uncoating by the clathrin-dependent recruitment and activation of PI3KC2α, a class II PI3-kinase. PI(3,4)P_2 may finally be converted to PI(3)P en route to endosomes by the 4-phosphatases *INPP4A/B*, effectors of the endosomal GTPase Rab5. Adapted from (Posor 2015)

In the early endosome, β1 integrins are sorted. Inactive integrins undergo fast "short loop" recycling; active integrins go to the late endosome / multivesicular body for slow "long group" recycling (*RAB11*), lysosomal degeneration (unless rescued by *RAB25*/*CLIC3*), or secretion via the trans-Golgi-network (TGN) mediated by *RAB9*.

Fast recycling of epidermal GF receptor drives proliferation,(Tomas 2014) so one would expect gain-of-function mutations in the upper part of the Figure.

Lysosomal and synaptic vesicle exocytosis share many similarities. Endo-lysosome-localized PIs may regulate lysosomal trafficking.(Samie 2014)

(derived, in part from Kegg pathways hsa04144 and hsa04721).

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

muGWAS identifies PI cycle dysregulation as novel breast cancer risk factor

In relation to EEC regulation, both CGEM and EPIC identified a PL-translocating ATPase, *ATP8B1* (PE) and *ATP8A1* (PS), respectively. *AGPAT3* is the second most significant gene in EPIC (mu: 6.59, ss: 4.73); *AGPAT4* is among the supportive genes in CGEM (Fig 1, mu: 4.94). Both acyltransferases transform LPA into PA. CGEM identified the scramblase *ANO4* [121601] (ss: 4.21), a PS exporter, and the plasma membrane PC/PS efflux pump *ABCA1* [19] (mu: 4.99). After including the latter two genes, String identified functional enrichment in

ATP8A1, *ATP8B1*, *ANO4*, *ABCA1*

GO:0097035 (biol. process) Regulation of membrane lipid distribution: FDR = 0.012
GO:0015914 (biol. process) phospholipid transport: 0.0407
GO:0005548 (mol. function) phospholipid transporter activity: 0.00968

As shown in Fig 4 (upper left corner), 8 (including 6 aGWS) genes are involved in providing the PI cycle with its substrate, PI (and the MAPK signaling pathway with PA.[(hsa04072)]).

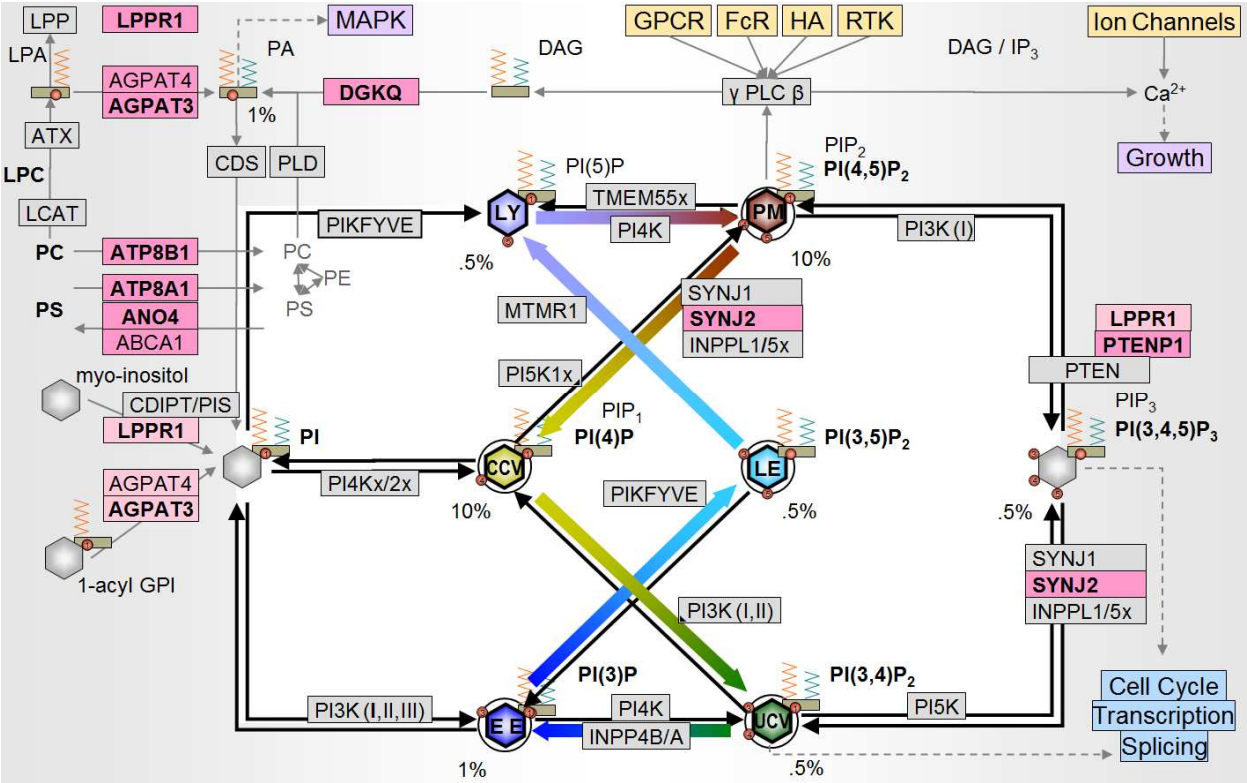

Fig 4: Functional relation of the PI/EC genes. PI is synthesized from myo-inositol (imported by HMIT) and PA (via CDP-DAG) which can be synthesized from lysophosphatic acid (LPA), PC, or PS, or salvaged from IP_3 and DAG. It can also be synthesized from 1-acyl GPI. Arrows: PIs are phosphorylated at a 3-, 4-, or 5- position by PI-kinases (left to right) and hydrolyzed by phosphatases (right-to-left). Genes associated with breast cancer in this GWAS are highlighted in pink (bold: aGWS). See Table 1 for other box colors. Colored arrows in the center indicate the sequence of PIs involved in EEC (Fig 3). Percent values indicate approximate proportion of PLs.[(Viaud 2016)].

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results for EEC regulation and function are consistent across populations

All three populations show aGWS association with EEC genes (CGEM: 4 in ssGWAS only / 4 in muGWAS only / one in both ; EPIC: 1/0/3; PBCS: 3/1/3). Most are validated in at least one of the other two populations, either by the same SNP involved (*AGPAT3*, *DGKQ*), the same region (*SYNJ2*, *PDE4D*), the same gene (see S1 Fig 5), or a functionally related gene (*AGPAT3*/*AGPAT4*, *LPPR1*/*DGKQ*, *ATP8A1*/*ATP8B1*, *STXBP1*/*UNC13C*, *TNS1*/*PTENP1*, see S1 Table 2 for details).

PI supply into the PI cycle as a drug target in breast cancer

After loss-of-function in *PTEN* and gain-of-function in *PI3K* suggested a mechanism for upregulation of PI(3,4,5)P_3 in cancer, blocking *PI3K* with Wortmannin [Powis 1995] or drugs with similar function [McNamara 2011] were considered for treatment of cancers, including breast cancer. Upregulation in PI(3,4)P_2 (gain-of-function in *SYNJ1/2* or *INPPL1* [Bunney 2010]) and PI(3)P (gain-of-function in *INPP4B*), [Woolley 2015] have also been associated with breast cancer. Recently, components to lower PI(3,4)P_2 by inhibiting *SYNJ2* have been identified. [Ben-Chetrit 2015]

Targeting individual phosphotransferases is unlikely to succeed due to the robustness of the PI cycle. [Powis 1995] All PIs regulating EEC, except for the evolutionarily recent *MTMR1* link (Fig 4), are regulated by both three kinases and three groups of phosphatases. Given the plethora of genes involved in EEC (Fig 3) identifying the appropriate set of phosphotransferase to interfere with endocytosis or to correct for functional deficits in exocytosis may be impractical.

Regulating EEC through the availability of PLs, however, while leaving functional interactions within the PI cycle intact, may be feasible. In fact, adding of either exogenous PS or PE led to an enhancement of endocytosis. [Farge 1999] As EEC is an essential and highly conserved mechanism for tissue morphogenesis [Emery 2006; Bokel 2014] and neuronal migration, [Wilson 2010; Cosker 2014; Kawauchi 2015] loss-of-function mutations would likely terminate embryonal development. Accordingly, the overall effect of the variations identified (S1 Table 3) is expected to be gain-of-function.

HPaCD is more effective than HPbCD against migration of breast cancer cells

In 2014, it was reported the benefit attributed to the neurosteroid allopregnanolone in the treatment of Niemann-Pick type C (NPC) disease was, in fact, due to the expedient, 2-hydroxypropyl-beta-cyclodextrin (HPβCD). Cyclodextrins are hydrophilic rings of ≥6 starch molecules (Fig 5). The lipophilic cavity can transport lipid drugs, such as allopregnanolone. Empty CDs, at therapeutic doses, form a pool in the aqueous phase into which, in the case of βCDs cellular cholesterol is extracted, [Ohtani 1989] the mechanism of action in NPC. [Vance 2014]

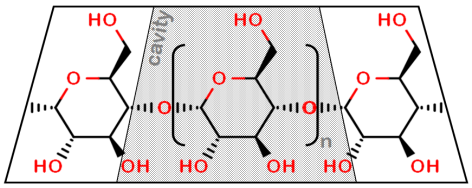

Fig 5: Structure of cyclodextrins.
Cyclodextrins are toroids formed of six (n=4, αCD), seven (n=5, βCD), or eight (n=6, γCD) starch molecules. The cavity is lipophilic, while the surface is hydrophilic.

Given the focus on cholesterol in NPC, it has often been overlooked that βCDs also scavenge PLs. The above GWAS results (Table 1), however, suggested defects in PL, rather than cholesterol function so that the efficacy of HPβCD in breast cancer might rather be due to its ability to scavenge PLs.

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

HPβCD is known to inhibit migration of human MDA-MB 231 breast cancer cells.[Liu 2007; Donatello 2012; Guerra 2016, Figure 3B] To determine whether inhibition of migration is caused by HPβCD depleting cholesterol, as assumed by the authors, or by it depleting PLs, as implicated by the above genetics results, the published activity from wound healing experiments comparing HPβCD against control was replicated, and complemented with novel activity results comparing HPαCD against control. Both comparisons were done in both MDA-MB 231 (ER^-) and MCF-7 (ER^+) human breast epithelial cell lines.

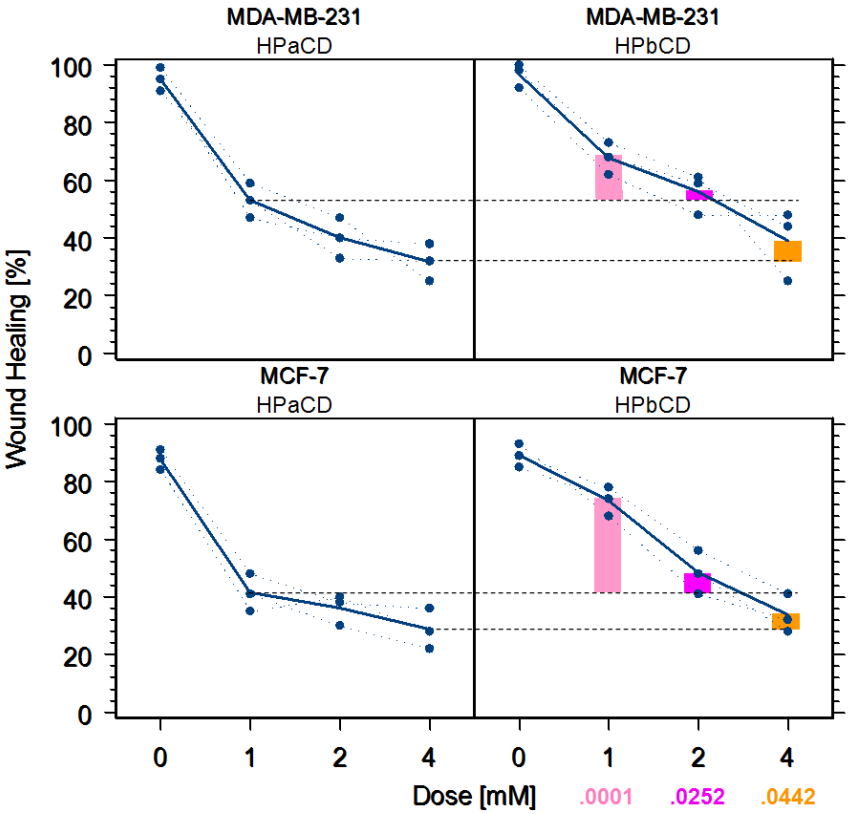

Fig 6: Wound healing by cyclodextrins in breast cancer **cell lines.**

Cells were grown in triplicates for 12 h and incubated with either of the CDs for 2 h at the concentration indicated (0–4 mM), before a 0.9 mm wide gap was opened and cells were allowed to migrate into the "wound" for 12 h.

HPβCD is more than 10× as toxic as HPαCD, which at <100 mM does not affect epithelial cell viability.[Leroy-Lechat 1994; Roka 2015]

Dashed horizontal line indicates inhibition of wound healing in HPaCD at 1 and 4 mM respectively..

ANOVA results:

indep: HPαCD vs HPβCD (fixed)

block: MCF-7/MDA-MB-231 (fixed)

dep: %change in wound healing

1 mM α vs 1 mM β, p = .0001 ***
1 mM α vs 2 mM β, p = .0252 *
4 mM α vs 4 mM β, p = .0442 *

Fig 6 shows that 1 mM HPαCD is more effective than 2 mM HPβCD against migration of both ER^- and ER^+ tumor cells (p = .0252) while more than 10× less toxic, irrespective of cell type,[Leroy-Lechat 1994] and, thus, that the effect previously seen with HPβCD is, in fact, likely the effect of it scavenging PLs, rather than cholesterol.

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

Our analysis confirmed previous GWAS, which pointed to receptor/AKT signaling and nuclear functions as critical components in breast cancer etiology. The present results from a reanalysis of data found previously inconclusive provides the first GWAS evidence for the contribution of EEC dysfunction, in general, and novel evidence for overstimulation of EEC in mesenchymal tumor cell migration and invasion. These findings, confirmed by an in vitro study on the activity of HPαCD vs HPβCD against breast cancer cell migration, suggest the novel hypothesis that reducing the influx of PLs, including PS, PC, and LPC, into the PI cycle via HPaCD may impair integrin translocation and thus provide an urgently needed treatment option against metastases in TNBC.

Replication and complementation of previously identified genes

A previous analysis of the CGEM data reported only two genes, *FGFR2* and *BUB3,* as risk factors for breast cancer. The EPIC and PBCS data have been published only as part of two meta-analysis, which also included CGEM. Among ER-negative cases, the first meta-analysis [Siddiq 2012] confirmed two SNPs each in *BABAM1* (7.31) (a nuclear *BRCA1* complex component), *PTHLH* (12.8) (which regulates epithelial-mesenchymal interactions during the formation of mammary glands), and the estrogen receptor *ESR1* (9.6). Our findings of *BMP7* (EPIC) and *BMPRT1B* (CGEM) are consistent with the previous finding of *PTHLH,* which forms a nuclear complex with *BMP4*. The second meta-analysis,[Garcia-Closas 2013] pointed to the *PIK3C2B-MDM4* region (11.68), *LGR6* (7.85) (a GPCR), and *FTO* (7.40) (a regulator of nuclear mRNA splicing). Hence, ssGWAS in all three populations point to receptor/AKT signaling and nuclear processes, although the individual genes differ.

Three of the four EEC genes identified in previous ssGWAS;[Michailidou 2013] were confirmed in muGWAS at aGWS/2 (CGEM: 2.56 / EPIC: 2.86 / PBCS: 2.57, S1 Table 1) in regions in LD (r^2):[Machiela 2015]

PDE4D	rs1353747	(4.56/4.46/2.84),	$r^2 \leq .213$
SNX32	rs3903072	(2.92/ ---- / ----),	$r^2 \leq .482$ rs7114014
STXBP4	rs6504950	(2.85/ ---- / ----),	$r^2 \leq .238$
DNAJC1	rs11814448	(---- / ---- / ----).	

The other EEC genes identified in this GWAS analysis (with the exception of *AGPAT3/4*, *ASTN2*, and *EEA1*), have previously been shown to be associated with breast cancer in gene expression and functional studies (S1 Table 1).

Computational biostatistics approach to genetic data

The analysis approach,[Wittkowski 2013] used here integrates genetics concepts into the statistical method, rather than considering them during visual inspection of p-values calculated one SNP at a time and correlations among SNPs within an LD block. In particular, this approach avoids assumptions about the degree of dominance, reflects that both SNPs next to a disease locus should be in LD (unless they are separated by a recombination hotspot), increases resolution within LD blocks (by distinguishing between members of the same tag sets being in a different order), integrates information from different disease loci within the same region (similar effects, compound heterozygosity), and draws on a measure of “information content” to prioritize results.

Screening for cis-epistatic regions (which may plausibly have evaded selective pressure) prioritizes biologically plausible results while de-emphasizing individual SNPs, which may be significant because of population selection biases, unless they cause exclusively late-onset pheno-

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

types, such as age-related macular degeneration.[Klein 2005] Avoiding strong model assumptions (additivity, independence) reduces model misspecification biases. Increasing the sample size, instead, does not guard against these biases, so that imposing a higher fixed GWS level in ssGWAS may, somewhat counterintuitively, favor "false positives" over biologically plausible cis-epistatic effects. The main limitation of u-statistics for multivariate data (conceived in the 1940s [Hoeffding 1948]) is that the amount of memory required became available only with 32-bit operating systems, in 2001, and computations became feasible only with the advent of GPU-enabled cloud computing.

To improve upon the conventional "overly conservative correction"[Pearson 2008] of 7.3, a systematic analysis of GWA studies suggested lowering the GWS level to 7.0 (fixed),[Panagiotou 2012] and then further by using study-specific empirical approaches.[Aslibekyan 2013a] The empirical aGWS decision rule used here accounts for GWAS not being randomized, the absence of a traditional 'null hypothesis' in a heritable disease, differences in MAF causing the expected distributions in a QR plot to be convex, and tests in overlapping diplotypes being related.[Wittkowski 2014]

The combination of a method with higher specificity and a decision strategy with higher sensitivity increased the number of genes above the cut-off (for aGWS) while ensuring that the vast majority of genes implicated was related to known pathways in breast cancer etiology, including dysregulation of EEC.

Replication of findings across populations

Conventionally, the level required for replication is lower. At the aGWS/2 level, none of most significant ssGWAS results (CGEM: *FGFR2*, *BUB3*, *MMRN1*; EPIC: *CHD3*, *SOHLH2*; PBCS: *DOCK8*) was replicated in another population (S1 Table 2). Only three genes (*AGPAT3*, *MEGF11*, and *TRAPPC9*) were replicated in both of the other populations, but none for the same SNP. These results are consistent with common lack of replication in ssGWAS.[Ioannidis 2013] With muGWAS, in contrast, many genes were replicated in at least one population and seven genes were replicated in both of the other populations (*FGFR2*, *TRPM3*, *AGPAT3*, *NCOR2*, *MEGF11*, *GPC6*, and *RGS3*), although not necessarily in the same region. Hence, analyses combining the data from several studies (often called "meta-analyses", even though subject-level data may be used) carries the risk of some of the populations diluting the risk factors present in others, resulting in overall effect sizes close to the null.[Ioannidis 2013]

Our results are consistent with ssGWAS finding rare, highly penetrant mutations, which may differ across populations, while muGWAS has higher power for common cis-epistatic variations, which are more likely to be shared across populations. Even more likely to be shared are genes that carry different variations and different genes with similar contribution to the etiology,[Aslibekyan 2013b] consistent with previous findings that gene expression signatures for breast cancer have little overlap across populations.[Haibe-Kains 2008]

Dysregulated EEC in breast cancer metastasis, angiogenesis, and progression

Genes involved in EEC (e.g., Rab GTPases) are aberrantly expressed in human cancers.[Mosesson 2008] Dysregulation of endocytosis-mediated recycling of oncoproteins (e.g., GF receptors and adhesion molecules, including integrins and annexins), can promote progression, migration, and invasion [Mosesson 2008; Maji 2016]. Cell migration and invasion, which are promoted by EEC of integrins, are also "essential features" of angiogenesis.[Demircioglu 2016] In addition, endocytic uptake of lipoproteins is critical for adaptation of cancer to its microenvironment.[Menard 2016]

Tumor-derived exosomes, 30–150 nm sized extracellular vesicles formed by dysregulated endocytosis, are critical mediators of intercellular communication between tumor cells and the recipient stromal cells in both local and distant microenvironments.[Costa-Silva 2015; Zhang 2015a] Several

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Rab proteins (Rab2b, Rab9a, Rab5a, Rab27a, and Rab27b) have been known to function in the selective packaging and production of exosomes in tumor cells.(Ostrowski 2010) Rab27a knockdown in highly metastatic melanoma cells led to a significant decrease in exosome production, primary tumor growth, and metastasis.(Peinado 2012)

Dysregulated EEC alters not only exosome biogenesis (vesicular packaging and trafficking), but also the composition of exosomal cargos. Tumor-specific proteins, such as integrins, were enriched in exosomes and correlated with migration and invasion.(Keerthikumar 2015) A recent study revealed that exosomal integrin expression patterns enriched in cancer-derived exosomes, involving specific αβ combinations, can determine organ targeting. Proteomic analysis revealed that the exosomal integrin αvβ5 binds to Kupffer cells that mediate liver metastasis, integrins α6β1 and α6β4 are associated with lung metastasis in breast cancer, while integrin β1 was non-specific. (Hoshino 2015)

Additionally, other tumor-specific exosomal proteins, such as annexins (calcium-dependent PL-binding proteins known to regulate membrane trafficking and EEC), which is known to correlate with migration and invasion, are also packaged in cancer exosomes (Leca 2016; Maji 2016). Annexins are frequently overexpressed in breast, liver, prostate, and pancreatic cancers and participate in multiple functions in cancer, including angiogenesis, tumor migration and invasion.(Mussunoor 2008) For instance, exosomal annexin A2 promotes angiogenesis and vascularization via tissue plasminogen activator (tPA) in breast cancer.(Maji 2016) Exosomal annexin A6 from cancer-associated fibroblasts contributes to tumor cell survival and invasion through annexin A6/LDL receptor-related protein 1/thrombospondin 1 complex formation in pancreatic cancer.(Leca 2016)

The PI cycle in Breast Cancer

Our findings of *PTENP1* (PBCS), *TNS1* (EPIC), and *SYNJ2* (CGEM) are consistent with known breast cancer mutations in *PI3K/PTEN*(Varticovski 1991; Li 1997) and *SYNJ2*. That both PI(3,4,5)P_3 and PI(3,4)P_2 are required to achieve and sustain a malignancy, has been formulated as the "two PI hypothesis"(Kerr 2011) Except for the known *PRCKQ*, which is regulated by PLs via the PI(4,5)P_2–PLC–DAG route, however, our analysis identified few genes along the *AKT/TSC/mTOR* pathway, which is controlled by the "two PIs". Instead, our results point to EEC, in which virtually all PIs are involved. The closely related set of genes involved in recycling of DAG (*DGKQ*), influx of PC and PS (*ATP8B1*, *ATP8A1*), and influx of LPA and 1-acyl GPI (*AGPAT3*, *AGPAT4*) suggests the downregulation of circulating PLs as a novel strategy to reduce EEC.

LPA, a known promoter of cell migration and invasion in breast cancer,(Mills 2003; Wang 2016) is produced from lysophosphatidylcholine (LPC) by autotaxin (ATX).(Benesch 2016) While *ATX* mouse knockouts are embryonically lethal, mice that overexpress LPA or *ATX* develop spontaneous metastatic mammary tumors. A mechanism mediated by G-coupled LPA receptors may cause mesenchymal tumors via endocytosis upregulation involving β-arrestin2 (Alemayehu 2013) and Arf6.(Hashimoto 2016)

LPA and LPC in physiologic concentrations have been shown to strongly induce migration of rhabdomyosarcoma (RMS) cells and to be increased by irradiation and chemotherapy in bone marrow.(Schneider 2014) The authors suggested the development of inhibitors of LPA/LPC signaling or "molecules that bind these bioactive lipids" after radio/chemotherapy. However, targeting a single among several redundant receptor/ligand complex may not be sufficiently effective to prevent metastases.(Ratajczak 2016)

Alkyl-LPCs, which compete with LPC, are in clinical use for treatment of cutaneous metastases in breast cancer, but have shown little activity (and substantial GI side effects) in advanced metastatic breast cancer.(Ríos-Marco 2017) From the results in this study, this is consistent with reducing LPC is most effective while cells are still migrating.

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

As our results suggest, overall EEC upregulation may be caused by various variations affecting the PI cycle. Thus, reducing EEC by diminishing the overall PL pool might be a more effective breast cancer treatment than blocking one or even two phosphotransferases, a strategy for which the highly robust PI cycle is designed to compensate. Given the ability of biologic systems to prioritize scarce resources, one would expect this effect to be stronger for tumor cells than for host cells whose functions are routinely prioritized when supplies are scarce. A related approach, substituted myo-inositol (MI) analogues, had already been considered, but was found unlikely to be effective *in vivo*, because even physiological concentration of MI antagonized the growth inhibitory activity of such analogues.(Powis 1995)

βCDs are effective in cancer models of migration, invasion, and angiogenesis

A plethora of studies have investigated the effect of methyl-β-cyclodextrin (MβCD) *in vitro*. For instance, MβCD suppressed invasion activity in three H7 Lewis lung cancer cell lines where highly metastatic cell lines had more b1 integrin.(Zhang 2006). Breast and prostate cancer cell lines were more sensitive to MβCD-induced cell death than their normal counterparts.(Li 2006) In particular, MβCD treatment induced a substantial decrease (40%) in activity of breast cancer resistance protein (*BCRP/ABCG2*),(Storch 2007) which transports PS and PC analogues.(Daleke 2007) In subsequent functional studies, MβCD inhibited spheroid migration and invasion of MDA-MB-241 and ZR751 breast cancer cells (Raghu 2010) and also endocytosis (Palaniyandi 2012) and migration (Guerra 2016) of MCF7 breast cancer cells. MβCD was more toxic for invasive than for non-invasive urothelial cancer cells,(Resnik 2015) and interfered with RTK-[PI2]-PI3K-[PI3]-AKT signaling in HeLa cells.(Yamaguchi 2015) Finally, MβCD reduced breast cancer-induced osteoclast activity in RAW264.7 cells and osteoclastogenic gene expression in MCF-7 cells.(Chowdhury 2017)

Sulfated SβCD also inhibits epithelial cell migration and invasion, but not proliferation (Watson 2013) and prevents angiogenesis *ex vivo* in an rat aortic ring assay and an chick embryo collagen on-plant assay.(Watson 2013)

The relevance of the above *in vitro* findings was confirmed by several *in vivo* studies. MβCD had higher concentration in tumor than in other cells (except kidney and liver) and reduced tumor volume in mice xenografted with MCF-7 breast cancer or A2780 ovarian carcinoma cells at least as effectively and with less toxicity than doxycyclin,(Grosse 1998) reduced the number of lung metastases in mice implanted with H7-O Lewis lung cancer cells,(Zhang 2006) reduced invasiveness of melanoma,(Fedida-Metula 2008) and inhibited growth of primary effusion lymphoma (PEL) in mice.(Gotoh 2014) HPβCD was necessary in triple combination treatment for tumor regression in mice implanted with renal cancer cells.(Yamaguchi 2015) and prolonged survival in leukemia mouse models.(Yokoo 2015)

However, while HPβCD was effective against tumors in animal models and well tolerated in most peripheral and central organ systems,(Cronin 2015) it was shown to carry the risk of causing permanent hearing loss in mice,(Crumling 2012) cats,(Ward 2010; Vite 2015) and at least one human.(Maarup 2015) This ototoxicity is believed to be due to depriving prestin *(SLC26A5*) in outer hair cells of cholesterol. (Kamar 2012; Yamashita 2015; Takahashi 2016)

Migration and invasion in breast cancer involve cholesterol-unrelated processes

The role of PLs emerging from our results, however, suggests a different mechanism than scavenging of cholesterol. This mechanism is consistent with previously reported *in vivo* results: *CAV1* expression in breast cancer stroma increases tumor migration and invasion (Goetz 2011) and *CAV1* is required for invadopodia formation specifically by breast cancer cells, where *CAV1* knockdown cannot be rescued by cholesterol.(Yamaguchi 2009) Growing MDA-MB-231 breast cancer cells in lipoprotein depleted medium resulted in an 85% decrease in cell migration.(Antalis 2011) LPA

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

activates the Arf6-based mesenchymal pathway for migration and invasion of renal cancer cells, which originate, like breast cancer cells, from cells located within epithelial ductal structures. (Kamar 2012; Yamashita 2015; Hashimoto 2016; Takahashi 2016)

Limiting the availability PIs would be particularly effective for PI(4)P and PI(4,5)P_2 (each at <10%, see Fig 4) and, thus, would likely reduce endocytosis more than lysosomal degradation. In addition, cyclodextrins have been shown to exert their role in NPC treatment by activating rather than downregulating, Ca-dependent lysosomal exocytosis.(Chen 2010)

From the mechanism of βCD in NPC and elevated cholesterol levels seen in several cancers, including breast cancer,(Yokoo 2015) βCDs were thought to reduce cancer growth by lowering cholesterol levels. Early evidence that this might not be the case emerged from the study of exosomes, which play a key role in development of breast cancer.(Peinado 2011; Lowry 2015) Treatment of MDA-MB-231 breast cancer cells with MβCD inhibited the internalization of exosomes containing integrins,(Hoshino 2015) but did so independently of cholesterol.(Koumangoye 2011)

αCD specifically scavenge PLs, reducing AEs and increasing effectiveness

The effect of βCDs on tumor (and other cells) is widely believed to be due to "cholesterol depletion",(Gotoh 2014; Badana 2016) yet βCDs also scavenges PLs.(Ohtani 1989) The genetics results presented here (Fig 4) suggest that, at least for breast cancer, the effect of βCDs is related primarily to scavenging of PLs. The cavity of αCDs is too small for cholesterol, but large enough for PLs. (Rajnavolgyi 2014; Shityakov 2016) From the above *in vitro* results (Fig 6) αCDs may achieve the same anti-tumor effect as βCDs, at half the dose and without the risk of cholesterol-related ototoxicity.

Two types of "controls" have been used: repletion of cholesterol via βCDs "loaded" with cholesterol, and reduction of cholesterol production via statins. Repletion of cholesterol, however, also increases production of PLs by freeing acetyl-CoA, the precursor of both PLs and cholesterol (Shiratori 1994; Ridgway 1999; Lagace 2015) and statins also lower phospholipids.(Snowden 2014) Hence, neither of these two strategies can "controls" against βCDs scavenging phospholipids, rather than cholesterols.

Using αCD as a control, however, can answer this question and the above *in vitro* results suggest that equi-molar αCDs are, in fact, at least twice as effective as βCDs, as one would expect if the effect of either CD is caused by its ability to scavenge phospholipids. Hence, our results suggest that many of the previous experiments with βCDs should be redone, this time using αCDs as a control.

αCD is generally recognized as safe (GRAS)(FDA, GRN000155) and approved as an expedient for i.v. alprostadil.(Loftsson 2010) Due to higher watersolubility, αCD has lower nephrotoxicity than βCD.(Frank 1976) HP derivatives of αCD and βCD increase water solubility from 145 and 18.5, respectively to ≥500 g/L. In mice, the observed ototoxicity order of HPβCD $>_{[p<.002]}$ HPγCD $>_{[p<.02]}$ HPαCD [$\approx_{[NS]}$ vehicle] matches the reported order for hemolysis and toxicities in various cell types.(Leroy-Lechat 1994; Davidson 2016) In humans, a single dose of up to 3 g/kg/d HPβCD and seven daily doses of 1 g/kg/d were reported to have no adverse effects.(Gould 2005) In 5-yr old children treated for NPC, 800 mg/kg/d HPβCD i.v. for 12 months was well tolerated.(Hastings 2009)

HPαCD as a novel treatment in breast cancer

Given significant redundancy pro-metastatic ligand-receptor complexes, the paradigm of targeting a single receptor-ligand complex has recently been challenged.(Ratajczak 2016) Although targeting EEC is a promising therapeutic strategy to prevent and treat metastasis,(Chew 2016) a therapeutic agent is yet to be determined. Our results suggest that metastases in breast cancer rely on upregulation of the highly robust PI cycle and various types of dysregulation along the complex

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

EEC pathway, rather than a simple linear PI pathway. Hence targeting the PI cycle in its entirety may be more efficacious than targeting individual phosphatases or kinases, or specific genes along the EEC pathway. The cyclodextrins are attractive candidates for a polyvalent approach to treat breast cancer. The ability of CDs to modulate several pathways involved in breast cancer, such as altering exosome production and packaging, and impede metastatic colonization is likely to confer greater protective effects than molecules that have single targets, thus allowing them to be effective at lower doses. The selectivity of the smaller αCDs to PLs would minimize side effects (e.g., ototoxicity) from βCDs also capturing cholesterol. Given that some CDs are already routinely used clinically, and their pharmacokinetic and toxicity profiles are well established, encouraging results for HPαCD in animal studies could lead rapidly to clinical efficacy trials.

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 # References

Alemayehu M, Dragan M, Pape C, et al. (2013). beta-Arrestin2 regulates lysophosphatidic acid-induced human breast tumor cell migration and invasion via Rap1 and IQGAP1. *PLoS One* **8**(2): e56174.

Antalis CJ, Uchida A, Buhman KK, et al. (2011). Migration of MDA-MB-231 breast cancer cells depends on the availability of exogenous lipids and cholesterol esterification. *Clin Exp Metastasis* **28**(8): 733-41.

Armitage P (1955). Tests for linear trends in proportions and frequencies. *Biometrics* **11**(3): 375-86.

Aslibekyan S, Claas SA, Arnett DK (2013a). To replicate or not to replicate: the case of pharmacogenetic studies: Establishing validity of pharmacogenomic findings: from replication to triangulation. *Circ Cardiovasc Genet* **6**(4): 409-12; discussion 12.

Aslibekyan S, Claas SA, Arnett DK (2013b). To replicate or not to replicate: the case of pharmacogenetic studies: Have pharmacogenomics failed, or do they just need larger-scale evidence and more replication? - Response to John P.A. Ioannidis, MD, DSc. *Circ Cardiovasc Genet* **6**(4): 418.

Badana A, Chintala M, Varikuti G, et al. (2016). Lipid Raft Integrity Is Required for Survival of Triple Negative Breast Cancer Cells. *J Breast Cancer* **19**(4): 372-84.

Barsh GS, Copenhaver GP, Gibson G, et al. (2012). Guidelines for genome-wide association studies. *PLoS Genet* **8**(7): e1002812.

Belguise K, Milord S, Galtier F, et al. (2012). The PKCtheta pathway participates in the aberrant accumulation of Fra-1 protein in invasive ER-negative breast cancer cells. *Oncogene* **31**(47): 4889-97.

Ben-Chetrit N, Chetrit D, Russell R, et al. (2015). Synaptojanin 2 is a druggable mediator of metastasis and the gene is overexpressed and amplified in breast cancer. *Sci Signal* **8**(360): ra7.

Benesch MG, Tang X, Venkatraman G, et al. (2016). Recent advances in targeting the autotaxin-lysophosphatidate-lipid phosphate phosphatase axis in vivo. *J Biomed Res* **30**(4): 272-84.

Bokel C, Brand M (2014). Endocytosis and signaling during development. *Cold Spring Harb Perspect Biol* **6**(3): a017020.

Bunney TD, Katan M (2010). Phosphoinositide signalling in cancer: beyond PI3K and PTEN. *Nat Rev Cancer* **10**(5): 342-52.

Burdett T, Hall PN, Hastings E, et al. The NHGRI-EBI Catalog of published genome-wide association studies. Retrieved 2016-05-16, from http://www.ebi.ac.uk/gwas.

Chen FW, Li C, Ioannou YA (2010). Cyclodextrin induces calcium-dependent lysosomal exocytosis. *PLoS One* **5**(11): e15054.

Chew CL, Chen M, Pandolfi PP (2016). Endosome and INPP4B. *Oncotarget* **7**(1): 5-6.

Chowdhury K, Sharma A, Sharma T, et al. (2017). Simvastatin and MBCD Inhibit Breast Cancer-Induced Osteoclast Activity by Targeting Osteoclastogenic Factors. *Cancer Invest*: 1-11.

Cicek MS, Cunningham JM, Fridley BL, et al. (2012). Colorectal cancer linkage on chromosomes 4q21, 8q13, 12q24, and 15q22. *PLoS One* **7**(5): e38175.

Cochran (1954). Some methods of strengthening the commen chi-square tests. *Biometrics* **10**: 417-51.

Cosker KE, Segal RA (2014). Neuronal signaling through endocytosis. *Cold Spring Harb Perspect Biol* **6**(2): a020669.

Costa-Silva B, Aiello NM, Ocean AJ, et al. (2015). Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat Cell Biol* **17**(6): 816-26.

Cronin S, Lin A, Thompson K, et al. (2015). Hearing Loss and Otopathology Following Systemic and Intracerebroventricular Delivery of 2-Hydroxypropyl-Beta-Cyclodextrin. *J Assoc Res Otolaryngol* **16**(5): 599-611.

Crumling MA, Liu L, Thomas PV, et al. (2012). Hearing loss and hair cell death in mice given the cholesterol-chelating agent hydroxypropyl-beta-cyclodextrin. *PLoS One* **7**(12): e53280.

Cui F, Wu D, Wang W, et al. (2016). Variants of FGFR2 and their associations with breast cancer risk: a HUGE systematic review and meta-analysis. *Breast Cancer Res Treat* **155**(2): 313-35.

Czaplinska D, Turczyk L, Grudowska A, et al. (2014). Phosphorylation of RSK2 at Tyr529 by FGFR2-p38 enhances human mammary epithelial cells migration. *Biochim Biophys Acta* **1843**(11): 2461-70.

Daleke DL (2007). Phospholipid flippases. *J Biol Chem* **282**(2): 821-5.

Davidson CD, Fishman YI, Puskas I, et al. (2016). Efficacy and ototoxicity of different cyclodextrins in Niemann-Pick C disease. *Ann Clin Transl Neurol* **3**(5): 366-80.

De Franceschi N, Hamidi H, Alanko J, et al. (2015). Integrin traffic - the update. *J Cell Sci* **128**(5): 839-52.

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Demircioglu F, Hodivala-Dilke K (2016). alphavbeta3 Integrin and tumour blood vessels-learning from the past to shape the future. *Curr Opin Cell Biol* **42**: 121-7.

Donatello S, Babina IS, Hazelwood LD, et al. (2012). Lipid raft association restricts CD44-ezrin interaction and promotion of breast cancer cell migration. *Am J Pathol* **181**(6): 2172-87.

Dozynkiewicz MA, Jamieson NB, Macpherson I, et al. (2012). Rab25 and CLIC3 collaborate to promote integrin recycling from late endosomes/lysosomes and drive cancer progression. *Dev Cell* **22**(1): 131-45.

Emery G, Knoblich JA (2006). Endosome dynamics during development. *Curr Opin Cell Biol* **18**(4): 407-15.

Farge E, Ojcius DM, Subtil A, et al. (1999). Enhancement of endocytosis due to aminophospholipid transport across the plasma membrane of living cells. *Am J Physiol* **276**(3 Pt 1): C725-33.

Fedida-Metula S, Elhyany S, Tsory S, et al. (2008). Targeting lipid rafts inhibits protein kinase B by disrupting calcium homeostasis and attenuates malignant properties of melanoma cells. *Carcinogenesis* **29**(8): 1546-54.

Frank DW, Gray JE, Weaver RN (1976). Cyclodextrin nephrosis in the rat. *Am J Pathol* **83**(2): 367-82.

Frommlet F, Nuel G (2016). An Adaptive Ridge Procedure for L0 Regularization. *PLoS One* **11**(2): e0148620.

Garcia-Closas M, Couch FJ, Lindstrom S, et al. (2013). Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat Genet* **45**(4): 392-8, 8e1-2.

Goetz JG, Minguet S, Navarro-Lerida I, et al. (2011). Biomechanical remodeling of the microenvironment by stromal caveolin-1 favors tumor invasion and metastasis. *Cell* **146**(1): 148-63.

Gotoh K, Kariya R, Alam MM, et al. (2014). The antitumor effects of methyl-beta-cyclodextrin against primary effusion lymphoma via the depletion of cholesterol from lipid rafts. *Biochem Biophys Res Commun* **455**(3-4): 285-9.

Gould S, Scott RC (2005). 2-Hydroxypropyl-beta-cyclodextrin (HP-beta-CD): a toxicology review. *Food Chem Toxicol* **43**(10): 1451-9.

Grosse PY, Bressolle F, Pinguet F (1998). Antiproliferative effect of methyl-beta-cyclodextrin in vitro and in human tumour xenografted athymic nude mice. *Br J Cancer* **78**(9): 1165-9.

Guerra FS, Sampaio LdS, Konig S, et al. (2016). Membrane cholesterol depletion reduces breast tumor cell migration by a mechanism that involves non-canonical Wnt signaling and IL-10 secretion. *Translational Medicine Communications* **1**(1): 3.

Guglielmi L, Servettini I, Caramia M, et al. (2015). Update on the implication of potassium channels in autism: K(+) channelautism spectrum disorder. *Front Cell Neurosci* **9**: 34.

Haibe-Kains B, Desmedt C, Piette F, et al. (2008). Comparison of prognostic gene expression signatures for breast cancer. *BMC Genomics* **9**: 394.

Haiman CA, Chen GK, Vachon CM, et al. (2011). A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat Genet* **43**(12): 1210-4.

Hajek J, Sidak Z (1967). *Theory of rank tests*. New York, NY, Academic.

Hall DP, Cost NG, Hegde S, et al. (2014). TRPM3 and miR-204 establish a regulatory circuit that controls oncogenic autophagy in clear cell renal cell carcinoma. *Cancer cell* **26**(5): 738-53.

Hanna S, Khalil B, Nasrallah A, et al. (2014). StarD13 is a tumor suppressor in breast cancer that regulates cell motility and invasion. *Int J Oncol* **44**(5): 1499-511.

Hashimoto S, Mikami S, Sugino H, et al. (2016). Lysophosphatidic acid activates Arf6 to promote the mesenchymal malignancy of renal cancer. *Nat Commun* **7**: 10656.

Hastings C. (2009). Addi and Cassi Hydroxy-Propyl-Beta-Cyclodextrin Plan. Compassionate Use Clinical Study. Treatment Plan Version #2. Retrieved 2016-11-13, from http://addiandcassi.com/wordpress/wp-content/uploads/2009/09/FDA-Submission-for-Addi-and-Cassi-Cyclodextrin-Treatment-Plan.pdf.

Hayashi S, Kimura M (2015). Mechanisms of hormonal therapy resistance in breast cancer. *Int J Clin Oncol* **20**(2): 262-7.

Hoeffding W (1948). A Class of Statistics with Asymptotically Normal Distribution. *Ann Math Stat* **19**(3): 293-325.

Hoshino A, Costa-Silva B, Shen T-L, et al. (2015). Tumour exosome integrins determine organotropic metastasis. *Nature* **527**(7578): 329-35.

Hunter DJ, Kraft P, Jacobs KB, et al. (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nature Genet* **39**(7): 870-4.

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ioannidis JP (2013). To replicate or not to replicate: the case of pharmacogenetic studies: Have pharmacogenomics failed, or do they just need larger-scale evidence and more replication? *Circ Cardiovasc Genet* **6**(4): 413-8; discussion 8.

Ioannidis JP, Thomas G, Daly MJ (2009). Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet* **10**(5): 318-29.

Iyer MK, Niknafs YS, Malik R, et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* **47**(3): 199-208.

Kamar RI, Organ-Darling LE, Raphael RM (2012). Membrane Cholesterol Strongly Influences Confined Diffusion of Prestin. *Biophysical Journal* **103**(8): 1627-36.

Kawauchi T (2015). Cellullar insights into cerebral cortical development: focusing on the locomotion mode of neuronal migration. *Front Cell Neurosci* **9**: 394.

Keerthikumar S, Gangoda L, Liem M, et al. (2015). Proteogenomic analysis reveals exosomes are more oncogenic than ectosomes. *Oncotarget* **6**(17): 15375-96.

Kendellen MF, Bradford JW, Lawrence CL, et al. (2014). Canonical and non-canonical NF-kappaB signaling promotes breast cancer tumor-initiating cells. *Oncogene* **33**(10): 1297-305.

Kerr WG (2011). Inhibitor and activator: dual functions for SHIP in immunity and cancer. *Ann N Y Acad Sci* **1217**: 1-17.

Klein RJ, Zeiss C, Chew EY, et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science* **308**(5720): 385-9.

Koumangoye RB, Sakwe AM, Goodwin JS, et al. (2011). Detachment of Breast Tumor Cells Induces Rapid Secretion of Exosomes Which Subsequently Mediate Cellular Adhesion and Spreading. *PLoS One* **6**(9): e24234.

Kruskal WH (1957). Historical notes on the Wilcoxon unpaired two-sample test. *J Am Statist Assoc* **52**: 356-60.

Lagace TA (2015). Phosphatidylcholine: Greasing the Cholesterol Transport Machinery. *Lipid insights* **8**(Suppl 1): 65-73.

Leca J, Martinez S, Lac S, et al. (2016). Cancer-associated fibroblast-derived annexin A6+ extracellular vesicles support pancreatic cancer aggressiveness. *J Clin Invest* **126**(11): 4140-56.

Leroy-Lechat F, Wouessidjewe D, Andreux JP, et al. (1994). Evaluation of the cytotoxicity of cyclodextrins and hydroxypropylated derivatives. *International Journal of Pharmaceutics* **101**(1-2): 97-103.

Li H (2012). U-statistics in genetic association studies. *Hum Genet* **131**(9): 1395-401.

Li J, Yen C, Liaw D, et al. (1997). PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* **275**(5308): 1943-7.

Li YC, Park MJ, Ye SK, et al. (2006). Elevated levels of cholesterol-rich lipid rafts in cancer cells are correlated with apoptosis sensitivity induced by cholesterol-depleting agents. *Am J Pathol* **168**(4): 1107-18; quiz 404-5.

Liu Y, Sun R, Wan W, et al. (2007). The involvement of lipid rafts in epidermal growth factor-induced chemotaxis of breast cancer cells. *Molecular membrane biology* **24**(2): 91-101.

Loftsson T, Brewster ME (2010). Pharmaceutical applications of cyclodextrins: basic science and product development. *J Pharm Pharmacol* **62**(11): 1607-21.

Lowry MC, Gallagher WM, O'Driscoll L (2015). The Role of Exosomes in Breast Cancer. *Clin Chem* **61**(12): 1457-65.

Maarup TJ, Chen AH, Porter FD, et al. (2015). Intrathecal 2-hydroxypropyl-beta-cyclodextrin in a single patient with Niemann-Pick C1. *Mol Genet Metab* **116**(1-2): 75-9.

Machiela MJ, Chanock SJ (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**(21): 3555-7.

Maji S, Chaudhary P, Akopova I, et al. (2016). Exosomal Annexin A2 Promotes Angiogenesis and Breast Cancer Metastasis. *Mol Cancer Res*.

Mann HB, Whitney DR (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* **18**(1): 50-60.

McNamara CR, Degterev A (2011). Small-molecule inhibitors of the PI3K signaling network. *Future medicinal chemistry* **3**(5): 549-65.

Menard JA, Christianson HC, Kucharzewska P, et al. (2016). Metastasis Stimulation by Hypoxia and Acidosis-Induced Extracellular Lipid Uptake Is Mediated by Proteoglycan-Dependent Endocytosis. *Cancer Res* **76**(16): 4828-40.

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Michailidou K, Hall P, Gonzalez-Neira A, et al. (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* **45**(4): 353-61, 61e1-2.

Mills GB, Moolenaar WH (2003). The emerging role of lysophosphatidic acid in cancer. *Nat Rev Cancer* **3**(8): 582-91.

Mitra S, Federico L, Zhao W, et al. (2016). Rab25 acts as an oncogene in luminal B breast cancer and is causally associated with Snail driven EMT. *Oncotarget* **7**(26): 40252-65.

Morgan MR, Byron A, Humphries MJ, et al. (2009). Giving off mixed signals--distinct functions of alpha5beta1 and alphavbeta3 integrins in regulating cell behaviour. *IUBMB Life* **61**(7): 731-8.

Mosesson Y, Mills GB, Yarden Y (2008). Derailed endocytosis: an emerging feature of cancer. *Nat Rev Cancer* **8**(11): 835-50.

Mussunoor S, Murray GI (2008). The role of annexins in tumour development and progression. *J Pathol* **216**(2): 131-40.

Ohtani Y, Irie T, Uekama K, et al. (1989). Differential effects of α-, β- and γ-cyclodextrins on human erythrocytes. *European Journal of Biochemistry* **186**(1-2): 17-22.

Ostrowski M, Carmo NB, Krumeich S, et al. (2010). Rab27a and Rab27b control different steps of the exosome secretion pathway. *Nat Cell Biol* **12**(1): 19-30; sup pp 1-13.

Palaniyandi K, Pockaj BA, Gendler SJ, et al. (2012). Human Breast Cancer Stem Cells Have Significantly Higher Rate of Clathrin-Independent and Caveolin-Independent Endocytosis than the Differentiated Breast Cancer Cells. *Journal of cancer science & therapy* **4**(7): 214-22.

Panagiotou OA, Ioannidis JPA, Project ftG-WS (2012). What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *International Journal of Epidemiology* **41**(1): 273-86.

Pearson TA, Manolio TA (2008). How to interpret a genome-wide association study. *JAMA* **299**(11): 1335-44.

Peinado H, Aleckovic M, Lavotshkin S, et al. (2012). Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nat Med* **18**(6): 883-91.

Peinado H, Lavotshkin S, Lyden D (2011). The secreted factors responsible for pre-metastatic niche formation: old sayings and new thoughts. *Semin Cancer Biol* **21**(2): 139-46.

Peng G, Luo L, Siu H, et al. (2010). Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur J Hum Genet* **18**(1): 111-7.

Pickrell JK, Berisa T, Liu JZ, et al. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48**(7): 709-17.

Posor Y, Eichhorn-Grunig M, Haucke V (2015). Phosphoinositides in endocytosis. *Biochim Biophys Acta* **1851**(6): 794-804.

Powis G, Berggren M, Gallegos A, et al. (1995). Advances with phospholipid signalling as a target for anticancer drug development. *Acta Biochim Pol* **42**(4): 395-403.

Raghu H, Sodadasu PK, Malla RR, et al. (2010). Localization of uPAR and MMP-9 in lipid rafts is critical for migration, invasion and angiogenesis in human breast cancer cells. *BMC Cancer* **10**: 647.

Rajnavolgyi E, Laczik R, Kun V, et al. (2014). Effects of RAMEA-complexed polyunsaturated fatty acids on the response of human dendritic cells to inflammatory signals. *Beilstein J Org Chem* **10**: 3152-60.

Ratajczak MZ, Suszynska M, Kucia M (2016). Does it make sense to target one tumor cell chemotactic factor or its receptor when several chemotactic axes are involved in metastasis of the same cancer? *Clinical and translational medicine* **5**(1): 28.

Resnik N, Repnik U, Kreft ME, et al. (2015). Highly Selective Anti-Cancer Activity of Cholesterol-Interacting Agents Methyl-beta-Cyclodextrin and Ostreolysin A/Pleurotolysin B Protein Complex on Urothelial Cancer Cells. *PLoS One* **10**(9): e0137878.

Ridgway ND, Byers DM, Cook HW, et al. (1999). Integration of phospholipid and sterol metabolism in mammalian cells. *Progress in Lipid Research* **38**(4): 337-60.

Ríos-Marco P, Marco C, Gálvez X, et al. (2017). Alkylphospholipids: An update on molecular mechanisms and clinical relevance. *Biochimica et Biophysica Acta (BBA) - Biomembranes*.

Rojas K, Stuckey A (2016). Breast Cancer Epidemiology and Risk Factors. *Clin Obstet Gynecol* **59**(4): 651-72.

Roka E, Ujhelyi Z, Deli M, et al. (2015). Evaluation of the Cytotoxicity of alpha-Cyclodextrin Derivatives on the Caco-2 Cell Line and Human Erythrocytes. *Molecules* **20**(11): 20269-85.

Samie MA, Xu H (2014). Lysosomal exocytosis and lipid storage disorders. *J Lipid Res* **55**(6): 995-1009.

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Schneider G, Sellers ZP, Abdel-Latif A, et al. (2014). Bioactive lipids, LPC and LPA, are novel prometastatic factors and their tissue levels increase in response to radio/chemotherapy. *Mol Cancer Res* **12**(11): 1560-73.

Shiratori Y, Okwu AK, Tabas I (1994). Free cholesterol loading of macrophages stimulates phosphatidylcholine biosynthesis and up-regulation of CTP: phosphocholine cytidylyltransferase. *J Biol Chem* **269**(15): 11337-48.

Shityakov S, Salmas RE, Salvador E, et al. (2016). Evaluation of the potential toxicity of unmodified and modified cyclodextrins on murine blood-brain barrier endothelial cells. *J Toxicol Sci* **41**(2): 175-84.

Siddiq A, Couch FJ, Chen GK, et al. (2012). A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum Mol Genet* **21**(24): 5373-84.

Siegel RL, Miller KD, Jemal A (2016). Cancer statistics, 2016. *CA: a cancer journal for clinicians* **66**(1): 7-30.

Snowden SG, Grapov D, Settergren M, et al. (2014). High-dose simvastatin exhibits enhanced lipid-lowering effects relative to simvastatin/ezetimibe combination therapy. *Circ Cardiovasc Genet* **7**(6): 955-64.

Storch CH, Ehehalt R, Haefeli WE, et al. (2007). Localization of the human breast cancer resistance protein (BCRP/ABCG2) in lipid rafts/caveolae and modulation of its activity by cholesterol in vitro. *J Pharmacol Exp Ther* **323**(1): 257-64.

Takahashi S, Homma K, Zhou Y, et al. (2016). Susceptibility of outer hair cells to cholesterol chelator 2-hydroxypropyl-beta-cyclodextrine is prestin-dependent. *Sci Rep* **6**: 21973.

Tomas A, Futter CE, Eden ER (2014). EGF receptor trafficking: consequences for signaling and cancer. *Trends Cell Biol* **24**(1): 26-34.

van Agthoven T, Sieuwerts AM, Veldscholte J, et al. (2009). CITED2 and NCOR2 in anti-oestrogen resistance and progression of breast cancer. *Br J Cancer* **101**(11): 1824-32.

Vance JE, Karten B (2014). Niemann-Pick C disease and mobilization of lysosomal cholesterol by cyclodextrin. *Journal of Lipid Research* **55**(8): 1609-21.

Varticovski L, Daley GQ, Jackson P, et al. (1991). Activation of phosphatidylinositol 3-kinase in cells expressing abl oncogene variants. *Mol Cell Biol* **11**(2): 1107-13.

Viaud J, Mansour R, Antkowiak A, et al. (2016). Phosphoinositides: Important lipids in the coordination of cell dynamics. *Biochimie* **125**: 250-8.

Vite CH, Bagel JH, Swain GP, et al. (2015). Intracisternal cyclodextrin prevents cerebellar dysfunction and Purkinje cell death in feline Niemann-Pick type C1 disease. *Sci Transl Med* **7**(276): 276ra26.

Wang CY, Lai MD, Phan NN, et al. (2015a). Meta-Analysis of Public Microarray Datasets Reveals Voltage-Gated Calcium Gene Signatures in Clinical Cancer Patients. *PLoS One* **10**(7): e0125766.

Wang J, Sun Y, Qu J, et al. (2016). Roles of LPA receptor signaling in breast cancer. *Expert review of molecular diagnostics* **16**(10): 1103-11.

Wang SJ, Cui HY, Liu YM, et al. (2015b). CD147 promotes Src-dependent activation of Rac1 signaling through STAT3/DOCK8 during the motility of hepatocellular carcinoma cells. *Oncotarget* **6**(1): 243-57.

Ward S, O'Donnell P, Fernandez S, et al. (2010). 2-hydroxypropyl-beta-cyclodextrin raises hearing threshold in normal cats and in cats with Niemann-Pick type C disease. *Pediatr Res* **68**(1): 52-6.

Watson CA, Vine KL, Locke JM, et al. (2013). The antiangiogenic properties of sulfated beta-cyclodextrins in anticancer formulations incorporating 5-fluorouracil. *Anti-cancer drugs* **24**(7): 704-14.

Wen Y, Alshikho MJ, Herbert MR (2016). Pathway Network Analyses for Autism Reveal Multisystem Involvement, Major Overlaps with Other Diseases and Convergence upon MAPK and Calcium Signaling. *PLoS One* **11**(4): e0153329.

Wilcoxon F (1954). Individual comparisons by ranking methods. *Biometrics* **1**: 80-3.

Wilson PM, Fryer RH, Fang Y, et al. (2010). Astn2, A Novel Member of the Astrotactin Gene Family, Regulates the Trafficking of ASTN1 during Glial-Guided Neuronal Migration. *The Journal of Neuroscience* **30**(25): 8529-40.

Wittkowski KM, Sonakya V, Bigio B, et al. (2014). A novel computational biostatistics approach implies impaired dephosphorylation of growth factor receptors as associated with severity of autism. *Transl Psychiatry* **4**: e354.

Wittkowski KM, Sonakya V, Song T, et al. (2013). From single-SNP to wide-locus: genome-wide association studies identifying functionally related genes and intragenic regions in small sample studies. *Pharmacogenomics* **14**(4): 391-401.

bioRxiv preprint doi: https://doi.org/10.1101/152405; this version posted June 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Wittkowski KM, Song T (2010). Nonparametric methods for molecular biology. *Methods Mol Biol* **620**: 105-53.

Wittkowski KM, Song T. (2012). muStat. from https://CRAN.R-project.org/package=muStat.

Woolley JF, Dzneladze I, Salmena L (2015). Phosphoinositide signaling in cancer: INPP4B Akt(s) out. *Trends Mol Med* **21**(9): 530-2.

Wu MC, Kraft P, Epstein MP, et al. (2010b). Powerful SNP-set analysis for case-control genome-wide association studies. *Am J Hum Genet* **86**(6): 929-42.

Yamaguchi H, Takeo Y, Yoshida S, et al. (2009). Lipid rafts and caveolin-1 are required for invadopodia formation and extracellular matrix degradation by human breast cancer cells. *Cancer Res* **69**(22): 8594-602.

Yamaguchi R, Perkins G, Hirota K (2015). Targeting cholesterol with beta-cyclodextrin sensitizes cancer cells for apoptosis. *FEBS Lett* **589**(24 Pt B): 4097-105.

Yamashita T, Hakizimana P, Wu S, et al. (2015). Outer Hair Cell Lateral Wall Structure Constrains the Mobility of Plasma Membrane Proteins. *PLoS Genet* **11**(9): e1005500.

Yan H, Zhu S, Song C, et al. (2012). Bone morphogenetic protein (BMP) signaling regulates mitotic checkpoint protein levels in human breast cancer cells. *Cell Signal* **24**(4): 961-8.

Yokoo M, Kubota Y, Motoyama K, et al. (2015). 2-Hydroxypropyl-beta-Cyclodextrin Acts as a Novel Anticancer Agent. *PLoS One* **10**(11): e0141946.

Zafar A, Wu F, Hardy K, et al. (2014). Chromatinized protein kinase C-theta directly regulates inducible genes in epithelial to mesenchymal transition and breast cancer stem cells. *Mol Cell Biol* **34**(16): 2961-80.

Zhang L, Gong C, Lau SL, et al. (2013b). SpliceArray profiling of breast cancer reveals a novel variant of NCOR2/SMRT that is associated with tamoxifen resistance and control of ERalpha transcriptional activity. *Cancer Res* **73**(1): 246-55.

Zhang L, Zhang S, Yao J, et al. (2015a). Microenvironment-induced PTEN loss by exosomal microRNA primes brain metastasis outgrowth. *Nature* **527**(7576): 100-4.

Zhang Q, Furukawa K, Chen HH, et al. (2006). Metastatic potential of mouse Lewis lung cancer cells is regulated via ganglioside GM1 by modulating the matrix metalloprotease-9 localization in lipid rafts. *J Biol Chem* **281**(26): 18145-55.