

# Journal Pre-proof

Dissecting the heritable risk of breast cancer: from statistical methods to susceptibility genes

Viola Fanfani, Martina Zatopkova, Adrian L. Harris, Francesco Pezzella, Giovanni Stracquadanio



PII: S1044-579X(20)30134-6  
DOI: <https://doi.org/10.1016/j.semcancer.2020.06.001>  
Reference: YSCBI 1828

To appear in: *Seminars in Cancer Biology*

Received Date: 15 November 2019  
Revised Date: 1 June 2020  
Accepted Date: 2 June 2020

Please cite this article as: Viola Fanfani, Martina Zatopkova, Adrian L. Harris, Francesco Pezzella, Giovanni Stracquadanio, Dissecting the heritable risk of breast cancer: from statistical methods to susceptibility genes, *Seminars in Cancer Biology* (2020), doi: <https://doi.org/10.1016/j.semcancer.2020.06.001>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

# Dissecting the heritable risk of breast cancer: from statistical methods to susceptibility genes

Viola Fanfani <sup>\*</sup>, Martina Zatopkova <sup>†</sup>, Adrian L. Harris, <sup>‡</sup>Francesco Pezzella, <sup>§</sup>  
Giovanni Stracquadanio <sup>¶</sup>

June 1, 2020

## Abstract

Decades of research have shown that rare highly penetrant mutations can promote tumorigenesis, but it is still unclear whether variants observed at high-frequency in the broader population could modulate the risk of developing cancer. Genome-wide Association Studies (GWAS) have generated a wealth of data linking single nucleotide polymorphisms (SNPs) to increased cancer risk, but the effect of these mutations are usually subtle, leaving most of cancer heritability unexplained. Understanding the role of high-frequency mutations in cancer can provide new intervention points for early diagnostics, patient stratification and treatment in malignancies with high prevalence, such as breast cancer.

Here we review state-of-the-art methods to study cancer heritability using GWAS data and provide an updated map of breast cancer susceptibility loci at the SNP and gene level.

## 1 Introduction

Breast cancer is the most frequent cancer among women worldwide, representing approximately one third of all diagnosed malignancies. Breast cancer has a cumulative risk of 5%, that is 5 in 100 newborns are expected to develop this malignancy during their lifetime. While the survival in first-world countries is usually very high, about 70% of all cases, breast cancer was still responsible for more than 600,000 deaths in 2018 [1, 2].

The mechanisms affecting cancer predisposition, tumorigenesis and progression are still unclear; in the majority of cases, tumors are triggered by the accumulation of somatic mutations, which impair critical cellular functions, like those controlled by the p53 tumor suppressor pathway [3]. While the causal role of somatic mutations has been confirmed by in-vitro and in-vivo models, there is limited understanding of whether inherited mutations mediate the risk of developing cancer. Familial and cancer syndrome studies have shown a causal role of inherited variants; usually, low frequency highly penetrant variants

<sup>\*</sup>Institute of Quantitative Biology, Biochemistry, and Biotechnology, SynthSys, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3BF, UK

<sup>†</sup>Department of Clinical Studies, Faculty of Medicine, University of Ostrava, Ostrava, Czech Republic

<sup>‡</sup>Molecular Oncology Laboratories, Department of Oncology, The Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK.

<sup>§</sup>Nuffield Department of Clinical Laboratory Sciences, University of Oxford, John Radcliffe Hospital, Oxford, UK

<sup>¶</sup>Institute of Quantitative Biology, Biochemistry, and Biotechnology, SynthSys, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3BF, UK. Phone: +44 (0) 131 6507193, Email: giovanni.stracquadanio@ed.ac.uk. Corresponding author.

in cancer susceptibility genes (CSG), directly increase the risk of cancer in first-degree relatives. In particular, breast cancer has been one of the first malignancies for which evidence of inheritance has been found and whose CSGs have been identified [4], including the well known BRCA1/2 genes [5, 6].

However, rare mutations explain only a small fraction of the risk of cancer in the broader population, suggesting that cancer risk could be somehow mediated by high-frequency low-penetrance mutations, such as single nucleotide polymorphisms (SNPs). Recent advances in high-density genotyping arrays and DNA sequencing technologies allow genotyping SNPs in large cohorts, paving the way to population-scale Genome Wide Association Studies (GWAS). Currently, more than 100,000 SNP alleles have been associated with various traits and diseases, of those around 5000 variants are associated with various tumor types, including breast cancer [7].

The contribution of germline mutations to the inherited risk of cancer is estimated through heritability analysis. Heritability estimates for cancer have been usually obtained through familial studies; however, these estimates have not been replicated when analysing inherited mutations in the broader population, thus leading to the concept of missing cancer heritability [8]. Missing heritability could be apportioned to a number of factors, including structural variants, gene-gene and gene-environment interactions, as well as rare highly penetrant variants [9, 10]. Ultra rare variants, which are difficult to detect with current technologies, have also shown to have a significant role in complex diseases [11]. However, even accounting for rare highly penetrant variants and genome-wide significant SNPs, the difference in risk between individuals is not completely explained [12].

There is strong evidence suggesting that the risk of complex diseases, such as cancer, can be explained by the co-inheritance of a large number of frequent variants with subtle effects [13]. In this case, we consider a disease to be polygenic [14], thus we are interested in quantifying the contribution of low-penetrance inherited mutations to cancer risk. Here, we focus on narrow sense heritability,  $h^2$ , that is the cumulative effect of all loci on the phenotype variance [15]. Interestingly, using GWAS data, we can estimate the heritability explained by SNPs regardless of their statistical significance. Heritability analysis is becoming a crucial step in recent cancer GWAS analyses, providing insights on the inherited risk of many malignancies, including prostate [16, 17], cervical [18], testicular germ cell tumor [19], and breast cancer [20].

Here we aim at providing an overview of state-of-the-art methods to estimate the amount of heritability explained by SNPs and an updated reference of the genetic architecture of breast cancer at the SNP and gene level. We organised this review as follows; in section 2, we introduce common notation and standard statistical analyses performed in GWAS, and we then present state-of-the-art methods for the estimation of heritability. Finally, in section 3, we systematically characterize current GWAS data available for breast cancer, and propose a curated resource of SNPs and genes that can be used for further investigations.

## 2 Estimating the risk of cancer explained by high-frequency inherited mutations

DNA sequencing technologies have enabled the discovery of thousands of rare and common variants that are associated with complex traits and diseases. While high-throughput whole genome sequencing is now routinely used to detect both common and low-frequency mutations across relatively small cohorts (< 10,000 individuals), cost-

effective genotyping arrays allow to carry out genetic studies at a population scale, albeit limited to only known loci.

Population scale genotyping is pivotal to understand the role of high-frequency low penetrant inherited mutations as genomic modifiers controlling quantitative traits and disease risk in the broader population. While highly penetrant mutations are often identified in relatively small cohorts [21], quantifying the contribution of high-frequency but low penetrance mutations requires genotyping large number of individuals.

In the last 30 years, genome-wide association studies (GWAS) have identified thousands of SNPs associated with increased risk of many diseases. In this context, cancer is not an exception; GWAS have been carried out on a broad spectrum of malignancies leading to the identification of a plethora of SNPs associated with increased risk of cancer [22]. However, experimental and analytical challenges have limited GWAS contribution in understanding the mechanisms underpinning cancer heritability.

Since the focus of this review is on computational methods for cancer GWAS analysis, we will focus on the methodological limits of SNP association tests, rather than issues arising from different experimental designs. GWAS have also complex interpretability limits; in particular, since variants often reside in non coding genomic regions, associations between SNP genotype and a trait provides limited mechanistic insights.

Here, we will introduce methods for heritability analysis as a framework to dissect the contribution of SNPs to the heritable risk of a disease, focusing on how to use these methods to study cancer heritability.

## 2.1 Tests of association

We refer to a single nucleotide polymorphism (SNP), as a locus where two or more distinct nucleotides are observed in a given population. Hereby, we assume SNPs to be bi-allelic, that is only 2 nucleotides are observed or considered at a given locus; this is a reasonable assumption for the vast majority of loci in the human genome.

We denote the most frequent nucleotide, as the major allele  $B$ , and the other as the minor allele,  $b$ . Since human cells are diploid, there are three possible genotypes, namely homozygous major ( $BB$ ), heterozygous ( $Bb$ ), and homozygous minor ( $bb$ ).

For a binary phenotype, such as case-control studies, the association between the genotype and the disease status (e.g. 0: normal, 1: affected) can then be tested using a  $\chi^2$  test with 2 degrees of freedom. For each SNP, the test is carried out by comparing genotype counts in cases and controls,  $g_{ij}$ , with their expected value,  $\hat{g}_{ij}$ , as follows:

$$\chi^2 = \sum_i \sum_j \left[ \frac{(g_{ij} - \hat{g}_{ij})^2}{\hat{g}_{ij}} \right] \quad (1)$$

where  $i$  is the disease status,  $j$  is one of the three possible genotypes and  $\hat{g}_{ij} = f_j N_i$ , with  $f_j$  being the genotype frequency. While the above is the general formulation, the  $\chi^2$  association test can be adapted to different hypotheses and data [23].

Logistic regression can instead be used to account for confounders, like age or sex. For a GWAS with  $N$  individuals and  $M$  SNPs, a logistic regression model can be defined as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2)$$

where  $\mathbf{Y} : N \times 1$  is a binary vector encoding the disease status,  $\mathbf{X} : N \times M$  is the genotype matrix, with  $x_{ij}$  being the number of minor alleles for the  $i$ -th locus of  $j$ -th individual.

Under the model in Eq. 2,  $\beta$  represents the effect-size of all SNPs and  $\epsilon$  are the error introduced by confounders. In presence of other covariates,  $C_i$ , the regression is extended to include those terms, such that  $Y = X\beta + C_1\beta_1 + C_2\beta_2 + \dots + \epsilon$ . Under the null hypothesis of no association between the SNP and the disease,  $\beta_j \sim N(0, \sigma^2)$ ; thus, the statistical significance of each effect-size can be tested using a Wald-test or a likelihood-ratio test. While the above formulations are useful to understand the idea behind association testing, in practice, these analyses require more complex models, which account for population biases, such as structure and relatedness, and genotype uncertainty.

While association analysis provides a mathematically tractable framework for testing whether a SNP genotype is associated with a trait, it is prone to false discoveries. This is largely due to SNP co-inheritance, a phenomenon usually referred to as linkage disequilibrium (LD); during meiotic crossing-over, proximal SNPs are more likely to be inherited together, resulting in a non-independence of their occurrence. From a statistical point of view, LD inflates the test statistic of the variants co-inherited with true causal SNPs, ultimately hindering the discovery of causal variants. LD for all genotyped SNPs in a GWAS can be represented as a lower-triangular matrix,  $R : M \times M$ , where  $r_{ij}^2$  is the LD between the alleles in SNP  $i$  and  $j$ . However, it is important to note that LD estimates are population-dependent and are biased by non-genotyped variants. Ultimately, finding causal variants usually requires integration of functional data to prioritize alleles within a given set of SNPs in LD [24].

To limit the number of false positives, GWAS studies usually apply a stringent family-wise error correction; in this context, empirical studies have concluded that  $5 \times 10^{-8}$  is a reasonable threshold to filter false positives out [13]. While for large populations and easily measurable phenotypes, such as height or blood pressure, it is possible to identify robust associations for a large number of loci, in cancer studies only a handful of SNPs pass correction for multiple hypotheses testing, resulting in the contribution of other loci with subtle effects to be neglected.

Thus, it is becoming apparent that methods able to estimate the cumulative contribution of multiple SNPs will be pivotal to maximize the information gained from GWAS. The rationale behind grouping SNPs together is based on the hypothesis that multiple variants in the same gene or pathway are more likely to have a stronger association with the phenotype regardless of their individual statistical significance. This is particularly true for cancer, whose inherited risk is thought to be mediated by a polygenic genetic architecture.

## 2.2 Estimating the heritable risk

Whenever referring to heritability, clarity is paramount; in GWAS analysis, inheritance does not refer to the amount of familial resemblance, rather to the effects of all inherited genomic loci to the phenotypic variance [25]. While broad sense heritability encompasses the effects of all genetic factors, narrow sense heritability accounts only for additive genetic effects. Thus, narrow sense heritability can be estimated from GWAS data as the cumulative contribution of all SNPs to the inherited risk.

Heritability in the narrow sense is defined as the portion of variance explained by the variance of the additive genetic effects,  $h^2 = \frac{\sigma_{Add}^2}{\sigma_P^2}$  [15]. The phenotype  $Y$  can be partitioned into two terms: a genotype term  $G$  and an environmental term  $E$ . The genotype contribution can be further partitioned into an additive genetic effect (*add*), a dominant genetic effect (*dom*) and an epistatic genetic effect (*epi*). Thus, a phenotype,  $Y$ , can be expressed as  $Y = G + E = (add + dom + epi) + E$ .

By estimating heritability from GWAS data, we are assuming the variance of the phenotype  $P$  to be  $\text{Var}(P) = \text{Var}(G + E)$ . Assuming independence between the terms, the overall phenotype variance is explained by narrow sense heritability, other genetic factors, and environmental effects as follows:

$$\frac{\sigma_P^2}{\sigma_P^2} = h^2 + \frac{\sigma_{epi}^2 + \sigma_{dom}^2}{\sigma_P^2} + \frac{\sigma_E^2}{\sigma_P^2} \quad (3)$$

GWAS can be used to estimate narrow sense heritability, since germline variants are accounting for additive genetic effects. However, the estimate obtained from the genotyped SNPs,  $h_{SNP}^2$ , is a lower bound of the narrow sense heritability,  $h_{SNP}^2 \leq h^2$ , since the genotyped loci are usually a subset of all the variants in the genome. Hereby, we will refer to the term heritability as a synonym of narrow-sense heritability, which we will denote as  $h^2$ .

Advances in statistical genetics are leading to an increasing number of methods to estimate the heritability explained by all genotyped SNPs, a quantity we will refer to as genome-wide heritability. However, these methods provide limited insights into the genetic architecture of a disease. While the reasons for this stall are probably multifaceted, there are many challenges that affect the accuracy of heritability estimation methods. In general, we would like to measure the contribution of the SNPs to a binary trait, that is the disease status. However, many popular methods to estimate  $h^2$  are working under the assumption of continuous traits. This problem is overcome by introducing the concept of liability [26]. Since most continuous traits can be approximated by a normal distribution, binary traits have been modelled by a liability threshold model; thus, there is the underlying assumption that disease risk follows a normal distribution, which represents the sum of many independent and normally distributed genetic and environmental effects. Thus, the binary phenotype represents whether the liability score exceeds a certain threshold  $t$ . Hence, in a normally distribute population, the quantile distribution function at  $t$  is the probability of the disease and is usually set from the observed prevalence in the population. In this framework, the observed value of heritability,  $h_{observed}^2$ , can be easily translated on the liability scale,  $h_{liability}^2$ , as follows:

$$h_{observed}^2 = \frac{z(t)^2 h_{liability}^2}{K(1 - K)} \quad (4)$$

where  $K$  is the incidence,  $z$  is the standard gaussian density.

Although mapping  $h^2$  from the observed to the liability scale is straightforward, it is important to check whether the assumptions made by a method hold for the study under consideration. In particular, for many cancer types, the incidence can be extremely low and so are the values of  $h_{observed}^2$ ; in both cases, the case-control ratio of the GWAS is incremented by design. While this procedure increments  $h_{observed}^2$ , thus making heritability detectable, it introduces a bias due to the difference between the real prevalence of the disease and the one in the cohort.

We now move forward describing methods to estimate heritability from GWAS data, highlighting their strength and weaknesses in the context of cancer GWAS analysis.

### 2.3 Methods for the estimation of genome-wide heritability

Estimates of genome-wide heritability can be obtained using a plethora of methods, each working under specific hypotheses, using different estimators, and requiring different input data. However, these methods estimate the heritability explained by genotyped SNPs



and it is common to refer to this quantity as array-heritability,  $h_{array}^2$ , or SNP heritability,  $h_{SNP}^2$ .

Here we present state-of-the-art methods classified based on the required input, that is either genotype data or SNP summary statistics. Methods using raw data require genotype and covariates for each patient. Conversely, methods using summary statistics require only SNP test statistics and standard errors, along with population-level parameters that can be estimated from reference panels.

Here we describe methods using genotype data first as they are regarded as the gold-standard in the field; we then introduce those using summary statistics highlighting differences and advantages between the other class.

### 2.3.1 Estimating heritability from genotype data

Heritability is obtained by regressing the variance of the phenotype against the variance of the genotype as defined in Eq. 2.

To do that, the vast majority of methods regress  $h^2$  using linear mixed models (LMM) [27, 28, 29, 30, 31]. The genomic-relatedness-based restricted maximum-likelihood approach (GREML, [27]), was the first to be introduced and it is routinely used for heritability studies. GREML uses genotype data with allele frequency as input and regress  $h^2$  using restricted maximum-likelihood. GREML assumes that effect sizes  $\beta$  and errors  $\epsilon$  in Eq. 2 are normally distributed with variance  $\sigma_g^2$  and  $\sigma_e^2$ , respectively. The variance of the phenotype then becomes:

$$\text{var}[\mathbf{Y}] = \mathbf{G}\sigma_g^2 + \mathbf{I}\sigma_e^2 \quad (5)$$

where  $\mathbf{G} = \mathbf{X}\mathbf{X}^T/M$  is the genetic relationship matrix (GRM) between pairs of individuals at  $M$  loci.

This method has been extended to account for differences in allele frequencies and relatedness. GREML has also been applied to binary traits [29], transforming the observed heritability estimates on the liability scale  $h_l^2$ , following the procedure outlined in Eq. 4. However, this procedure should be used with caution when analysing cancer data, since GREML works under the assumption that the phenotype is normally distributed. While the liability model is a good approximation for diseases with high prevalence, REML assumptions do not hold when study prevalence does not match the true population prevalence; this leads to consistently biased estimates [32], thus suggesting that GREML-like approaches are not appropriate to analyse cancer data [33].

A second class of methods adapts the Haseman-Elston regression [34] to GWAS analysis, specifically focusing on case-control studies [35, 32, 36]. The Phenotype Correlation - Genotype Correlation method (PCGC) does not rely on normality assumptions, but instead obtains heritability estimates by considering the relationship between phenotypic and genotypic correlations between individual  $i$  and individual  $j$ . The phenotypic correlation,  $E(y_i y_j)$  can be written as a generic function of the heritability and the genotypic correlation :

$$E(y_i y_j) = f(h^2, G_{ij}) \quad (6)$$

In its simplest formulation, considering only additive quantitative phenotypes and no specific study design confounders,  $f(h^2, G_{ij}) = h^2 G_{ij}$  and  $h^2$  can be estimated by least squares as follows:

$$h^2 = \text{argmin} \sum_{i,j,i \neq j} [y_i y_j - h^2 G_{ij}] \quad (7)$$

Case-control studies, extreme phenotypes, studies with related individuals are all modelled by using an appropriate  $f(h^2, G_{ij})$ .

For binary phenotypes, the phenotypic correlation,  $E(y_i y_j)$ , is accounted for to obtain estimates of heritability on the liability scale. The general consensus is that PCGC is better suited for binary phenotypes, being more robust to different covariates and cohort size.

Methods using genotype data are considered the gold-standard for heritability analysis and are readily available as part of many bioinformatics packages [37, 38]. However, these methods require access to high-performance computing (HPC) infrastructures and genotype data; while HPC facilities are routinely found in academic and industrial environments, access to cancer patients' genotype is usually difficult, due to privacy concerns, thus limiting their use in practice.

### 2.3.2 Estimating heritability from summary statistics data

There has been an increasing interest in estimating heritability using GWAS summary statistics to overcome the limitations imposed by methods requiring genotype data [39, 40, 41, 42, 43]. Summary statistics are usually publicly available, since genotype information cannot be traced back from regression weights, and the analysis is not computationally taxing. Here we review how genome-wide heritability can be estimated from GWAS summary statistics.

The most widely used approach to estimate heritability from summary statistics is the LD score (LDSC) regression method [39, 44]. LDSC computes heritability estimates by regressing  $h^2$  as follows:

$$E[\chi_j^2] = \frac{N}{M} h^2 l_j + Na + 1 \quad (8)$$

where  $\chi_j^2$  is the summary statistic of the  $j$ -th SNP for a GWAS with  $N$  individuals and  $M$  variants. Here  $l_j$  is a quantity called LD score, computed as  $l_j = \sum_{i=0}^K r_{ij}^2$ , that is by summing up the correlation coefficients of all the SNPs in a window of prefixed size from the  $j$ -th variant. Here,  $Na + 1$  is a term introduced to account for confounding bias, which can be estimated as the intercept of the linear regression between the LD score of each variant and its test statistic. The heritability is regressed using reweighted least squares, where the weights are adjusted to account for heteroscedasticity of the test statistic. LDSC is also implemented as part of the SUMHER software, which improves the original LDSC model by taking into account allele frequency [42].

Recently, PCGC has also been extended to take summary statistics in input (s-PCGC, [41]). It has been shown that LDSC and s-PCGC are almost equivalent in absence of covariates with strong effects [45], although s-PCGC is recommended in presence of effects that could severely skew the liability distribution.

The methods discussed so far have been shown to be sensitive to the input data and trait properties, e.g. low or high heritability, low or high disease prevalence. This is due to the assumptions made by each model, which must be carefully considered as part of the analysis [46, 47, 45].

Although estimating heritability from genotype data usually leads to more accurate estimates, summary statistics proved to be sufficient to obtain accurate heritability estimates across a number of phenotypes [48]. Moreover, the negligible computational burden of summary statistics methods has made them the preferred approach for population scale studies [49] and the steppingstone to estimate the heritability of SNP groups.

Nonetheless, genome-wide heritability analyses have major interpretability limits. The estimate of  $h^2$  gives a measure of the contribution of all genotyped SNPs to the heritable



risk, which is usually an underestimation for cancer, in part due to the low prevalence of the disease. Importantly, current heritability studies do not provide insights into the mechanisms underpinning disease risk; thus, the focus has shifted on estimating the heritability explained by SNPs in functional genomic regions to provide a mechanistic interpretation of GWAS associations.

## 2.4 Methods for partitioning heritability

The vast majority of methods providing genome-wide heritability estimates usually assume that all genotyped SNPs have the same contribution to heritability. This assumption has already been questioned in literature [28], since it is more reasonable to assume that the amount heritability explained by a group of SNPs depends on the genomic region where they are located, e.g. promoter or coding regions. Thus, it is becoming apparent that estimating the heritability explained by SNPs residing in functional loci could give further insights in the genetic architecture of a disease.

Finucane et al. proposed a stratified LD score regression method (s-LDSC, [50]), which has been used to study the UK Biobank cohort. The method computes the heritability explained by SNPs belonging to a list of 53 functional binary classes, such as coding regions or histone marks. To do that, s-LDSC estimates the heritability explained by  $C$  functional categories, as follows:

$$E[\chi_j^2] = N \sum_{c \in C} \tau_c l(j, c) + Na + 1 \quad (9)$$

where the LD-score is computed only over the SNPs within the  $c$ -th class and  $\tau_c$  is the per-SNP heritability contribution of the  $c$ -th class. Thus, the portion of heritability of one class with  $L_c^{SNP}$  variants is:  $h_i^2 = L_c^{SNP} \tau_c$ . Recently, the model has been extended to account for continuous annotations, such as GC content or recombination rate [51].

An alternative approach uses the heritability estimator from summary statistics, HESS [52], to partition the genome in 1703 independent loci [53] and to then estimate the explained heritability as follows:

$$h_{local}^2 = \frac{N\beta R^{-1}\beta - M}{N - M}$$

where  $\beta$  are the summary statistics for a GWAS with  $N$  individuals and  $M$  SNPs.  $R^{-1}$  is the inverse of the LD matrix approximated by a singular value decomposition, since the inverse usually does not exist due to linkage disequilibrium between SNPs. While for each category, s-LDSC partitions the whole genome in just two classes, HESS divides the genome in multiple regions (see supplementary figure S1). The scope of partitioning is to test whether a category has an heritability enrichment, that is the SNPs in the category explains a larger amount of  $h^2$  compared to the genome-wide estimate. If  $h_k^2$  is the heritability explained by the  $M_k$  SNPs belonging to annotation  $k$ , the quantity  $(h_k^2/M_k)M$  is on the same scale of the genome-wide estimate; thus, in absence of any enrichment, the heritability for the single SNP  $h_k^2/M_k$  should be approximately equal to the genome-wide estimate  $h^2/M$ .

While partitioning methods could provide insights into genomic regions explaining a large proportion of heritability, there are still limits to use partitioned heritability to study cancer GWAS. Both HESS and LDSC are not robust for small sample sizes and low heritability diseases; this usually has the effect of providing erroneous negative local heritability estimates, suggesting that new robust estimators are needed to maximize the utility of these analyses.

Table 1: **Breast cancer heritability estimates in European populations.** For each study, we report the heritability estimate on the liability scale ( $h_l^2$ ), the reported standard error or the 95% confidence intervals (CI) and the disease prevalence.

Cancer (subtype)	$h_l^2$	Cases/controls
Breast (ER negative) [57]	0.096 (CI=[0, 0.199])	1,998/3,263
Breast (Self-reported) [49]	0.1104 (s.e.=0.0221)	7,480/329,679
Breast [58]	0.13 (s.e.=0.011)	122,977/105,974

### 3 The genetic landscape of breast cancer

The genetics of breast cancer has been extensively studied due to its relatively high prevalence and incidence in the broader population. The first three GWAS on breast cancer were published in 2007 and new targeted studies have been conducted in different populations. To date, the Breast Cancer Association Consortium (BCAC) is the largest breast cancer GWAS in Europeans, including more than 120,000 cases [54]; moreover, new genome-wide significant SNPs have been recently found in the same cohort using imputation [55]. Conversely, the UK Biobank (UKBB, [56]) represents the study with the largest total number of individuals ( $N > 300,000$ ) and unbiased disease prevalence.

In this section we review the main results on breast cancer heritability, and then summarise and characterise susceptibility loci and genes for this malignancy.

#### 3.1 Heritability estimates

The estimation of heritability from high-frequency variants for cancer presents multiple challenges and the results are highly dependent on the cohort and downstream processing. However, as novel studies with large cohorts are released and targeted GWAS are carried out, it is reasonable to expect that understanding cancer risk in the broader population will be possible.

While the exact heritability estimate varies across GWAS studies, there is a consensus estimate of breast cancer heritability being  $h^2 \sim 0.1$  on the liability scale (see Tab. 1). This value is significantly smaller than previous familial estimates,  $h^2 \sim 0.3$ , although there is mounting evidence that this value could be an overestimation [31, 33]. Sampson et al. report values of heritability, estimated via GREML, between 0.092 and 0.25, after adjusting for age, minor allele frequency and gender [57]. While the authors analysed GWAS data calibrated for cancer studies, the cohort is considerably smaller than the UKBB and BCAC cohorts. Jiang et al. analysed the BCAC cohort using LDSC regression [58], finding an heritability estimate  $h^2 \sim 0.13$ ; interestingly, when excluding genome-wide significant SNPs and their linked loci, the heritability estimate is significantly smaller, suggesting that up to 45% of the total heritability is explained by genome-wide significant variants. Estimates obtained by LDSC on the UKBB cohort show remarkably coherent estimates, despite the prevalence of the malignancy being significantly smaller than other studies[49].

Recently, there has been increasing interest in identifying functional elements, such as histone mark or DNA I hypersensitive regions, explaining breast cancer heritability. However, analyses performed using stratified LDSC regression on the UKBB and BCAC cohorts were inconclusive [49]. Nonetheless, there is evidence suggesting that taking into account SNP location and functional effects in the analysis could provide useful insights on the role of inherited variants for cancer [59, 60, 61]. On this point, using local co-

heritability between breast, lung, and prostate cancer [62, 58], a pattern of local risk inheritance has been found. This result provides preliminary evidence that improvements in the analysis of partitioned heritability could be useful to discover loci across the human genome mediating the risk of multiple cancers.

### 3.2 Breast cancer risk loci across the human genome

Heritability studies have shown that high-frequency inherited mutations explain a significant proportion of breast cancer risk. We then move forward to identify SNPs and genes that are associated with increased risk of breast cancer in the broader population; ultimately, we aim at providing an updated map of breast cancer susceptibility genes across the human genome.

We obtained SNPs data from the GWAS Catalog [7], which reports more than 143,000 SNPs across 3,522 studies. We then retrieved SNPs associated with breast cancer in European populations and mapped SNPs to genes, after applying quality control filters (see Supplementary Methods and Supplementary Figure S2). We also discarded SNPs, approximately half of the total reported, that did not reach genome-wide statistical significance set at  $p < 5 \times 10^{-8}$ ; usually, p-values above this threshold are indicative of a small population size or old genotyping arrays, thus we preferred to filter those out as a conservative approach for our downstream analysis.

We found 719 significant variants (see Fig. 1A) reported by 26 different studies, which are within 50kb from 311 genes (see Supplementary Table 1, we consider a 50kb window to include regulatory regions in the analysis). Interestingly, of those 719 reported variants, 108 are reported in more than one study, while 311 are reported only once; while this provides preliminary evidence to support the robustness of a reported association, differences in tag SNP selection and reporting criteria across studies will likely result in different SNPs being reported for the same susceptibility haplotype (see Supplementary Methods and Supplementary Figure S2).

We observed that most variants account for limited increase in risk, with average odds ratio  $OR : 1.11$ , and ranging from 1.02 for rs17529111 to 1.59 for rs62235635 (See Supplementary Figure S3,S4); moreover, the odds ratio for rs62235635 is still well below the strongest reported cancer association, that is for SNP rs995030-G in testicular germ cell tumors ( $OR: 2.26$ ) [63].

The risk allele frequency for breast cancer is 0.37 on average, ranging from 0.005 to 0.98 (Supplementary Figure S5). Unsurprisingly, the data suggests a negative-correlation between cancer risk and allele frequency (see Fig. 1B). In particular, SNP rs62235635 in PITPNB, which is the variant with the lowest frequency, is also the one with the highest odds ratio  $OR : 1.589$ . This is consistent with other studies, which have shown that SNPs with detrimental impact are less frequently observed in the broader population because are likely to be subject to negative selection [51, 64].

We then analyzed the functional impact of each SNP associated with breast cancer (see Fig. 1C) and found that the vast majority of SNPs reside in introns or intergenic regions, with only a negligible fraction located in coding regions and possibly causing detrimental changes, such as missense variations or stop codon gain. While functional genomics techniques are continuously improving, testing functional effects of cancer SNPs will likely remain challenging, since phenotypic changes are going to be subtle and difficult to detect (Supplementary Figure S6). Nonetheless, we found that 89% of breast cancer SNPs are in or around a coding region, suggesting that most of them could act as cis-regulator of an upstream or downstream gene. We then used this information to compile a draft panel of genes associated with breast cancer heritability.

**Table 2: Breast cancer susceptibility loci in European populations.** We report SNPs associated with increased risk of breast cancer, whose odds-ratios (OR) are in the 95<sup>th</sup> percentile among all those reported in the GWAS catalog for this malignancy. For each SNP, we report the rsid, the cytogenic region, the reported odds ratio (OR), the functional consequence as sequence ontology term, the nearest gene, the reported risk allele frequency and the PUBMED id of the study.

SNPS	REGION	OR	CONTEXT	GENES	RISK ALLELE FREQUENCY	PUBMEDID
rs62235635	22q12.1	1.59	intron variant	PITPNB	0.0065	29059683
rs11571833	13q13.1	1.58	stop gained	BRCA2	0.01	29058716
rs62235681	22q12.1	1.58	intergenic variant	CHEK2	0.0085	29059683
rs1314913	14q24.1	1.57	intron variant	RAD51B		23001122
rs62237615	22q12.1	1.55	intron variant	TTC28	0.0082	29059683
rs62237573	22q12.1	1.53	intron variant	TTC28	0.0092	29059683
rs3803662	16q12.1	1.5	non coding transcript exon variant	CASC16		23001122
rs2229882	5q11.2	1.45	synonymous variant	MAP3K1	0.06	24493630
rs2981579	10q26.13	1.43	intron variant	FGFR2	0.42	20453838
rs10771399	12p11.22	1.39	intergenic variant	PTHLH		24325915
rs16886448	5q11.2	1.37	intron variant	MAP3K1	0.07	24493630
rs7726354	5q11.2	1.37	intron variant	MIER3	0.06	24493630
rs16886034	5q11.2	1.36	intergenic variant		0.08	24493630
rs16886364	5q11.2	1.36	intron variant	MAP3K1	0.07	24493630
rs3822625	5q11.2	1.36	synonymous variant	MAP3K1	0.07	24493630
rs16886397	5q11.2	1.36	intron variant	MAP3K1	0.07	24493630
rs16886113	5q11.2	1.35	regulatory region variant		0.08	24493630
rs614367	11q13.3	1.34	intergenic variant	LINC01488	0.16	24493630
rs78540526	11q13.3	1.34	intergenic variant	LINC01488	0.08	25751625
rs1017226	5q11.2	1.33	intron variant	AC008937.2;MAP3K1	0.08	24493630
rs9397437	6q25.1	1.32	intergenic variant	CCDC170	0.07	29058716
rs1219648	10q26.13	1.32	intron variant	FGFR2	0.42	20872241
rs75915166	11q13.3	1.31	regulatory region variant		0.06	25751625

### 3.3 Genes associated with breast cancer susceptibility

We analysed 104 genes, out of the 311 in total, reported in at least 2 studies and associated with a Hugo symbol (Supplementary figure S7). It is worth noting that a gene can be reported multiple times because the same variant might have been reported in multiple studies or because different variants are mapped to the same genes.

We assigned the highest reported odds ratio,  $OR_{max}$ , and focused on those with the highest effect-size (see Fig. 2A). There are 20 genes with an  $OR_{max} > 1.2$ , with the top 10 genes having  $OR_{max} > 1.28$ ; we hereby refers to these genes as breast cancer susceptibility genes (BCSGs, see Fig. 2B,C).

We then analyzed the functional role of BCSGs to identify possible mechanisms mediating breast cancer heritability. After performing literature curation, we found that 4 BCSGs control cell cycle, whereas 5 others are involved in DNA repair and invasion (see Fig. 3), which are fundamental processes underpinning all cancers [65, 66]. It is important to note that CASC16 has been reported as a cancer susceptibility gene, but its functional role remains unclear.

We identified 4 BCSGs, namely CHEK2, FGFR2, MAP3K1 and TTC28, which control critical steps of the cell cycle. CHEK2 is a tumour suppressor gene activated upon DNA damage, which activates genes controlling basic cellular activities, such as apoptosis, DNA repair, and cell cycle arrest. The mechanism is triggered via activation of TP53, BRCA1 or BRCA2 proteins [67]. Mutations in this gene are known to lead to the dysregulation of cell cycle and thus facilitate malignant transformation of the cell, and development of various types cancer, including breast cancer [68]. Mutations in CHEK2 gene mediate response to anthracycline based chemotherapy in breast cancer patients [69]. FGFR2 (Fibroblast growth factor receptor 2) negatively modulates activity of ESR1 and can inhibit estrogen signalling [70]. It has been clearly shown that FGFR2 mediates cancer susceptibility and mutations at this locus can account for an increase in the risk of breast cancer of up to 16% [71]. FGFR2 is also a member of the fibroblast growth factor receptor (FGFR) family, which controls upregulation of MAPK, PI3K/AKT, STAT and PLC $\gamma$  signaling pathways. These pathways are involved in cancer mediating processes, such as cancer cell proliferation, differentiation, invasion, survival and carcinogenesis [72, 73, 74]. The mitogen-activated protein kinase kinase kinase 1 (MAP3K1) is a serine/threonine kinase having a role in signal transduction cascades, like MAPK, ERK, NF- $\kappa$ B, JNK or JUN pathway, which control critical cellular processes, including apoptosis, proliferation and differentiation [75]. Mutations in this gene affect kinase activity and are identified as oncogenic drivers [76]. TTC28 is a gene with oncogenic activity required during the cell cycle for condensation of spindle midzone microtubules, formation of the midbody, and completion of cytokinesis [77]. The gene resides in the proximity of the CHEK2 gene, thus suggesting a possible pattern of co-inheritance.

A second group includes 3 genes, namely BRCA2, RAD51B and LINC01488, which mediates repair mechanisms upon double-strand DNA breaks. BRCA2 is a well known cancer susceptibility gene, whose mutations are associated with 69% increase in risk of breast cancer and 17% increase in risk of ovarian cancer[78]. Mutations in this gene are also linked to other malignancies, including stomach, pancreatic and prostate cancer [79]. BRCA2 is also a therapeutic target of the FDA approved PARP inhibitors Rucaparib [80] and Niraparib [81]. For RAD51B there is evidence of association with familial breast cancer due to common variations [82]. In detail, RAD51B (RAD51 paralog B) encodes a protein which creates a complex with other RAD51 paralogs promoting binding of RAD51 upon DNA damage [83, 84]. Damaged DNA prevents successful replication and cause a cell cycle arrest and apoptosis. Overexpression of RAD51 is usually found in tumors and mediates drug resistance [85]. Haploinsufficiency of RAD51B causes mild hypersensitiv-



ity to DNA-damaging agents favoring chromosome aberrations and aneuploidy in human cells by impairing RAD51 function [86]. LINC01488, also known as CUPID1, is a long non coding RNA regulated by estrogen and located in the 11q13 cytogenic band, which is associated with increased risk of breast cancer[87]. CUPID1, and the neighboring lncRNA CUPID2, have been shown to affect homologous-repair (HR) and non-homologous end joining (NHEJ) DNA repair mechanisms by impairing RAD51 recruitment.

We finally report 2 BCSGs, namely MIER3 and PTHLH, that are known to control invasion. MIER3 (MIER family member 3) together with MIER1/2 and BAHD1 (vertebrate protein that promotes heterochromatin formation and gene repression) repress expression of the steroid hormone receptor gene ESR1 [88]. MIER3 is reported to act as tumor suppressor [89] and is a known cancer susceptibility gene [90]. The Parathyroid Hormone Like Hormone (PTHLH), which encodes the Parathyroid hormone-related protein (PTHrP), is a gene responsible for the humoral hypercalcemia of the malignancy, mammary development and lactation[91, 92]. During lactation it facilitates delivery of maternal calcium to milk and thus play a role in regulation of bone and mineral metabolism. By action through PTH1 receptors, PTHrP contributes to formation of bone metastasis through promotion of osteoclast formation and bone resorption [93]. It is important to note that FGFR2, MIER3 and LINC01488 are also involved in estrogen signaling, which regulates mammary gland development and is one of the main risk factors for breast cancer.

Taken together, the BCSGs identified in our analysis directly mediates cancer phenotypes and co-morbidities related to breast cancer. Upon further investigation, we also found these genes to be reported in many cancer panels (see Supplementary Figure 8) [94, 95], thus suggesting also a possible link between somatic and inherited mutations.

## 4 Future directions

Decades of familial cancer studies provide evidence for a causal role of inherited genomic mutations, but these results have not been replicated by GWAS, when analyzing high-frequency mutations in the broader population. However, recent advances in sequencing and genotyping technologies, combined with accurate statistical methods, are enabling the identification of variants and quantify the heritable risk of many common malignancies, including breast cancer.

Here we provided an updated overview of SNPs and genes associated with breast cancer susceptibility, showing how variants in genes controlling cell cycle, DNA repair and invasion could modulate the risk of developing this disease. Since breast cancer susceptibility genes are often mutated in breast tumors, we speculate that a possible link between inherited and somatic mutations might exist and could provide new targets for clinical applications, including treatment and patients stratification. In particular, it is still difficult to dissect the functional role of the polymorphisms and how they may interact on a common mechanism, such as RAD51 regulation.

It will be of interest in long term follow up studies e.g. 'Generations' study, to see whether the type of breast cancer that develops is related to these polymorphisms, and to understand prevention studies e.g. hormone suppression in those with estrogen regulated polymorphic genes.

However, current experimental and analytical limitations lead us to believe that identifying the biological components modulating the risk of breast cancer and other oncological diseases will require substantial advances in statistical genetics. Moreover, experimental systems should be put in place to systematically validate the findings, and update and improve models. Taken together, heritability analysis is emerging as a powerful tool



to quantify the effect of variants with subtle effects, but new robust methods able to identify biological units, such as genes or pathways, are needed to translate analytical results into biological and clinical findings.

## **5 Conflict of interest**

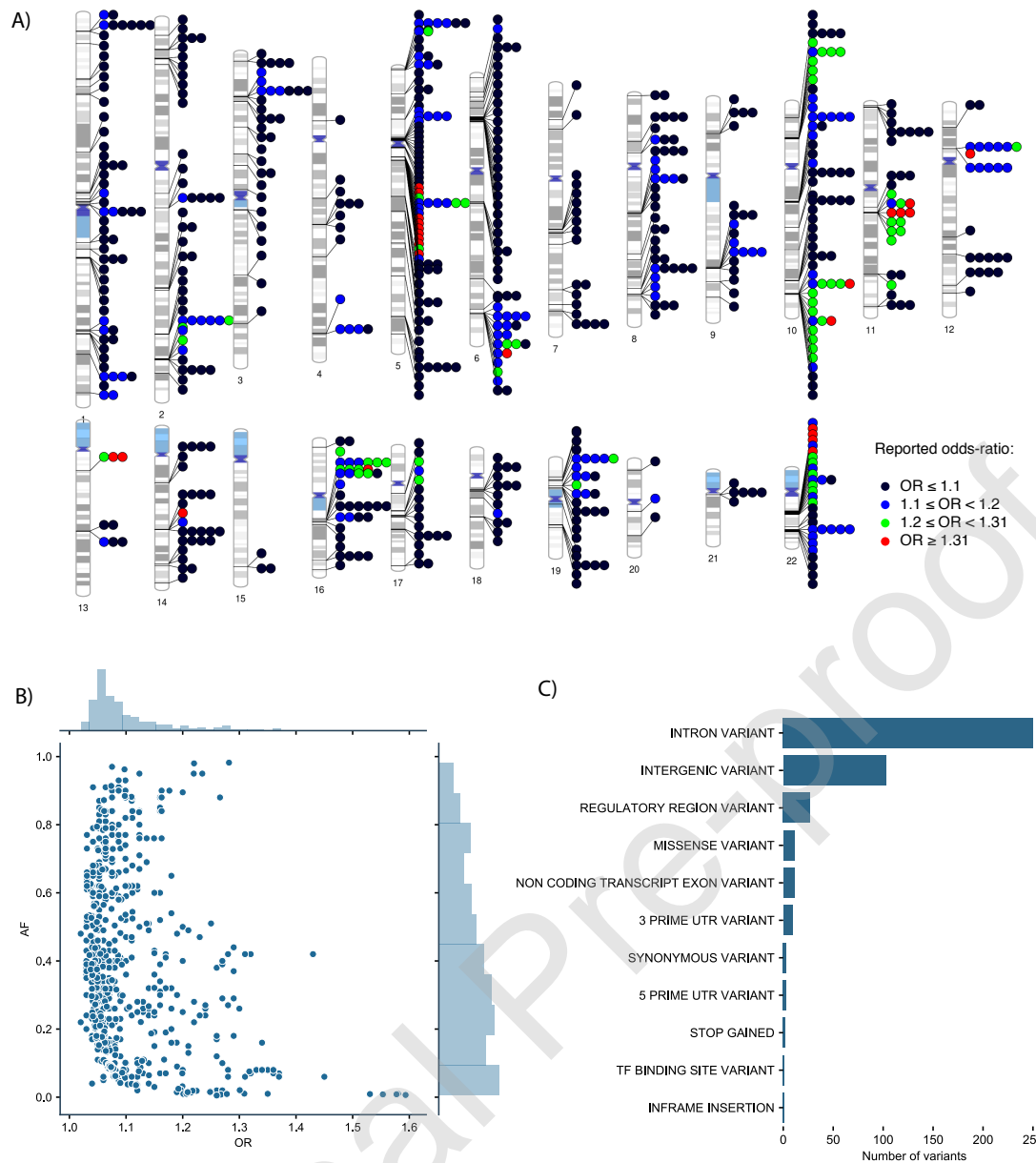
The authors declare there are no conflicts of interest.

## **6 Acknowledgments**

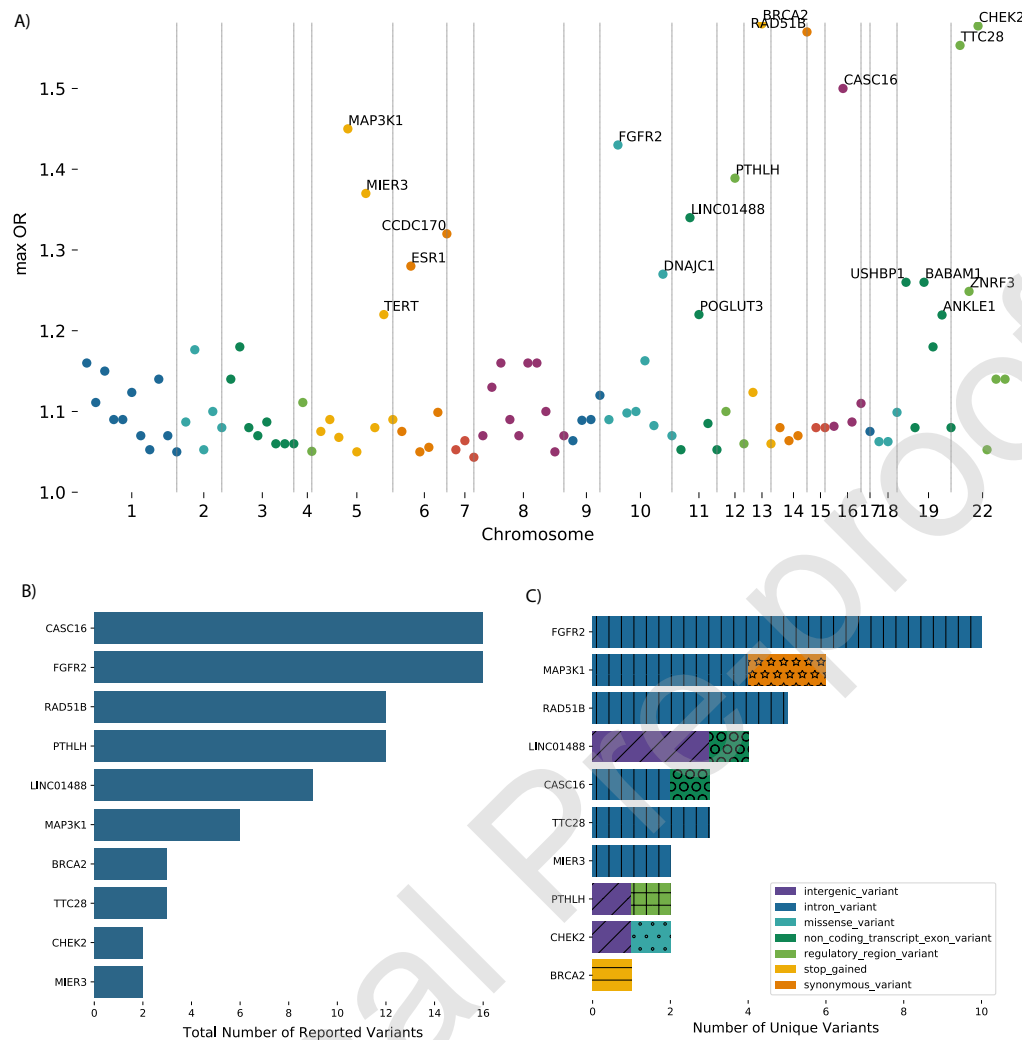
This work has been supported by the Wellcome Trust Seed Award in Science (207769/A/17/Z) to G.S. M.Z was supported by the European Regional Development Fund, Project ENOCH (No. CZ.02.1.01\0.0\0.0\16\_019\0000868).

## 7 Figures

Journal Pre-proof



**Figure 1: Breast cancer susceptibility loci across the human genome.** A) Phenogram [96] of the 719 reported SNPs associated with breast cancer. Each SNP is represented by circles, and stacked symbols represent a locus for which multiple studies have reported an association. The color codes distinguish the reported odds-ratio (OR). Red circles denote those with stronger effect,  $\text{OR} \geq 1.31$ , that are only 5% of the total. B) Distribution of the odds-ratios (OR) and risk allele frequencies (AF). The central scatter plot shows the ORs and AFs for each SNP, where the top and right side are the corresponding histograms of OR and AF, respectively. For SNPs reporting only regression coefficients,  $\beta$ , we transformed these values in odds-ratios as follows  $\text{OR} = \exp(\beta)$ . ORs are characterized by a long-tail distribution, whereas AF seems uniformly distributed. It is important to note the correlation between OR and AF, with rare variants have consistently stronger effects. C) Functional classification of the variants reported by the GWAS catalog.



**Figure 2: Breast cancer susceptibility genes.** A) For each gene, we report the variants that are mapped within 50Kb of the gene body and the corresponding odds-ratios (ORs); variants reporting only regression coefficients were transformed into ORs by computing  $OR = \exp(\beta)$ . The 10 genes with highest OR were further characterized below. B) Number of reported variants for each gene. It is important to note that the same gene could harbor different variants, or the same variant could have been reported in multiple studies. C) Number of unique variants grouped by gene and mutation effect. Only BRCA2, CHEK2 and MAP3K1 harbor exon variants.

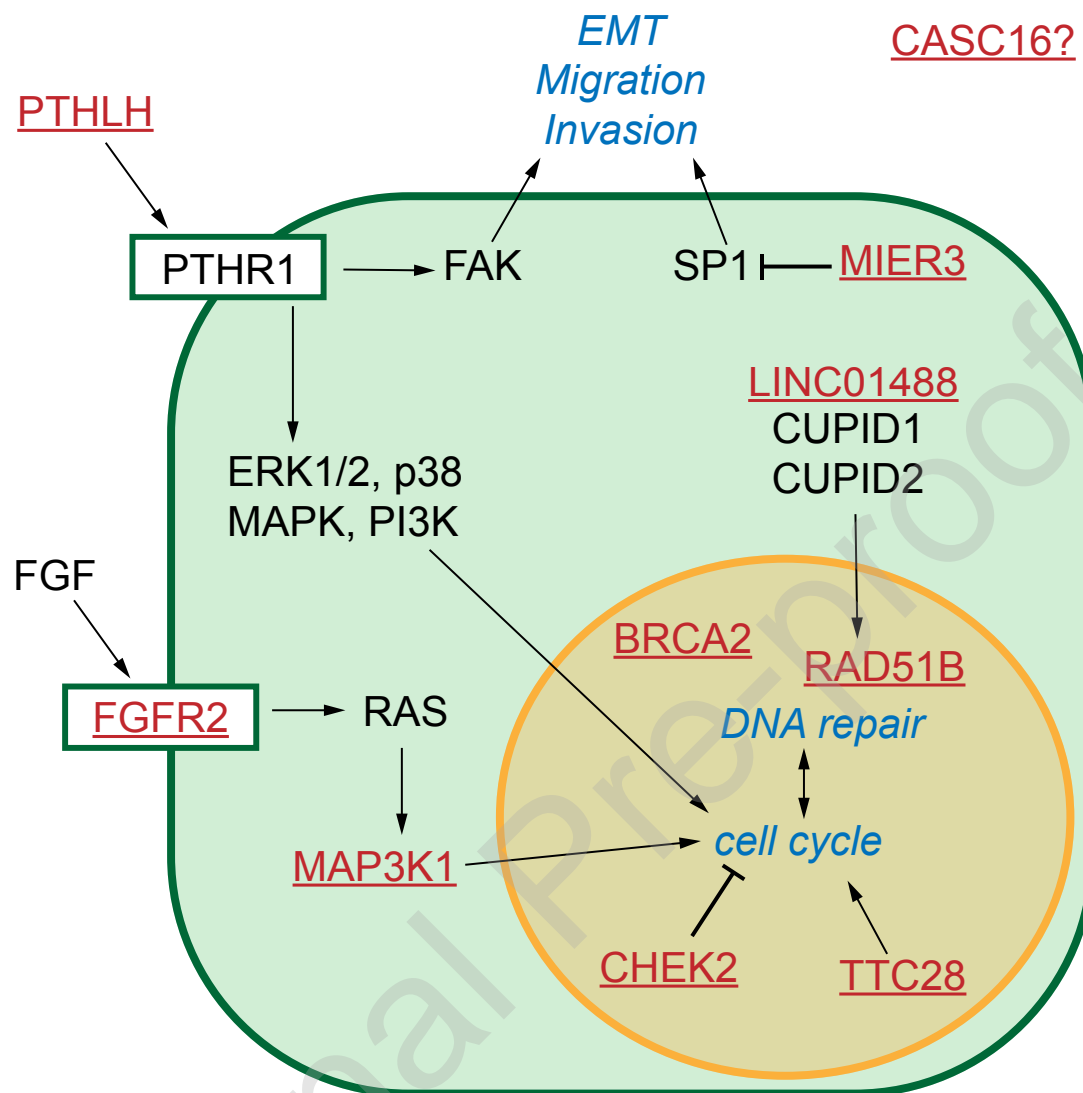


Figure 3: **Function and location of the breast cancer susceptibility genes.** Breast cancer susceptibility genes (BCSGs, in red) are linked to three main biological processes (italic blue), namely cell cycle, DNA repair and invasion. When coupled to its ligand, FGFR2 triggers the RAS pathway, which activates downstream MAP3K1, thus promoting cell cycle. TTC28 and the hormone PTHLH also promote cell cycle while CHEK2 inhibits it. PTHLH induces FAK phosphorylation, leading to increased invasion, which is in turn inhibited by MIER3. Finally, both RAD51B and BRCA2 are active in DNA repair, whereas LINC01488 (CUPID2) mediates this process by impairing RAD51 recruitment.

## References

- [1] Freddie Bray et al. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." In: *CA: a cancer journal for clinicians* (Sept. 2018). ISSN: 1542-4863. DOI: [10.3322/caac.21492](https://doi.org/10.3322/caac.21492). URL: <http://www.ncbi.nlm.nih.gov/pubmed/30207593>.
- [2] Ferlay J et al. *Global Cancer Observatory: Cancer Today*. Lyon, France: International Agency for Research on Cancer. 2018. URL: <https://gco.iarc.fr/today>.
- [3] Bert Vogelstein et al. "Cancer genome landscapes". In: *Science* 340.6127 (Mar. 2013), pp. 1546–1558. ISSN: 10959203. DOI: [10.1126/science.1235122](https://doi.org/10.1126/science.1235122). URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3749880/>.
- [4] David E. Anderson. "Genetic study of breast cancer: Identification of a high risk group". In: *Cancer* 34.4 (Oct. 1974), pp. 1090–1097. ISSN: 0008-543X. DOI: [10.1002/1097-0142\(197410\)34:4<1090::AID-CNCR2820340419>3.0.CO;2-J](https://doi.org/10.1002/1097-0142(197410)34:4<1090::AID-CNCR2820340419>3.0.CO;2-J). URL: <http://doi.wiley.com/10.1002/1097-0142%28197410%2934%3A4%3C1090%3A%3AAID-CNCR2820340419%3E3.0.CO%3B2-J>.
- [5] Y Miki et al. "A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1". In: *Science* 266.5182 (Oct. 1994), pp. 66–71. ISSN: 0036-8075. DOI: [10.1126/science.7545954](https://doi.org/10.1126/science.7545954). URL: <http://www.sciencemag.org/cgi/doi/10.1126/science.7545954>.
- [6] Richard Wooster et al. "Identification of the breast cancer susceptibility gene BRCA2". In: *Nature* 378.6559 (Dec. 1995), pp. 789–792. ISSN: 00280836. DOI: [10.1038/378789a0](https://doi.org/10.1038/378789a0).
- [7] Jacqueline MacArthur et al. "The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog)". en. In: *Nucleic Acids Research* 45.D1 (Jan. 2017), pp. D896–D901. ISSN: 13624962. DOI: [10.1093/nar/gkw1133](https://doi.org/10.1093/nar/gkw1133). URL: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw1133>.
- [8] Antonella Galvan, John P.A. Ioannidis, and Tommaso A. Dragani. "Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer". In: *Trends in Genetics* 26.3 (Mar. 2010), pp. 132–141. ISSN: 0168-9525. DOI: [10.1016/J.TIG.2009.12.008](https://doi.org/10.1016/J.TIG.2009.12.008). URL: <https://www.sciencedirect.com/science/article/pii/S0168952509002662>.
- [9] Or Zuk et al. "The mystery of missing heritability: Genetic interactions create phantom heritability." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.4 (Jan. 2012), pp. 1193–8. ISSN: 1091-6490. DOI: [10.1073/pnas.1119675109](https://doi.org/10.1073/pnas.1119675109). URL: <http://www.ncbi.nlm.nih.gov/pubmed/22223662> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3268279>.
- [10] Naomi R. Wray and Robert Maier. "Genetic Basis of Complex Genetic Disease: The Contribution of Disease Heterogeneity to Missing Heritability". In: *Current Epidemiology Reports* 1.4 (Dec. 2014), pp. 220–227. ISSN: 2196-2995. DOI: [10.1007/s40471-014-0023-3](https://doi.org/10.1007/s40471-014-0023-3). URL: <http://link.springer.com/10.1007/s40471-014-0023-3>.
- [11] Ryan D. Hernandez et al. "Ultrarare variants drive substantial cis heritability of human gene expression". In: *Nature Genetics* 51.9 (Sept. 2019), pp. 1349–1355. ISSN: 1061-4036. DOI: [10.1038/s41588-019-0487-7](https://doi.org/10.1038/s41588-019-0487-7). URL: <http://www.nature.com/articles/s41588-019-0487-7>.



- [12] Teri A. Manolio et al. "Finding the missing heritability of complex diseases". In: *Nature* 461.7265 (2009), pp. 747–753. ISSN: 00280836. DOI: [10.1038/nature08494](https://doi.org/10.1038/nature08494). URL: <http://dx.doi.org/10.1038/nature08494>.
- [13] Peter M. Visscher et al. "10 Years of GWAS Discovery: Biology, Function, and Translation". In: *The American Journal of Human Genetics* 101.1 (July 2017), pp. 5–22. ISSN: 00029297. DOI: [10.1016/j.ajhg.2017.06.005](https://doi.org/10.1016/j.ajhg.2017.06.005). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0002929717302409>.
- [14] Evan A. Boyle, Yang I. Li, and Jonathan K. Pritchard. "An Expanded View of Complex Traits: From Polygenic to Omnigenic". In: *Cell* 169.7 (2017), pp. 1177–1186. ISSN: 10974172. DOI: [10.1016/j.cell.2017.05.038](https://doi.org/10.1016/j.cell.2017.05.038). URL: <http://linkinghub.elsevier.com/retrieve/pii/S0092867417306293>.
- [15] Peter M. Visscher, William G. Hill, and Naomi R. Wray. "Heritability in the genomics era - Concepts and misconceptions". In: *Nature Reviews Genetics* 9.4 (Apr. 2008), pp. 255–266. ISSN: 14710056. DOI: [10.1038/nrg2322](https://doi.org/10.1038/nrg2322). URL: <http://www.nature.com/doifinder/10.1038/nrg2322>.
- [16] Nicholas Mancuso et al. "The contribution of rare variation to prostate cancer heritability". In: *Nature Genetics* 48.1 (Jan. 2016), pp. 30–35. ISSN: 1061-4036. DOI: [10.1038/ng.3446](https://doi.org/10.1038/ng.3446). URL: <http://www.nature.com/articles/ng.3446>.
- [17] Edward J. Saunders et al. "Fine-Mapping the HOXB Region Detects Common Variants Tagging a Rare Coding Allele: Evidence for Synthetic Association in Prostate Cancer". In: *PLoS Genetics* 10.2 (Feb. 2014). Ed. by Greg Gibson, e1004129. ISSN: 1553-7404. DOI: [10.1371/journal.pgen.1004129](https://doi.org/10.1371/journal.pgen.1004129). URL: <http://dx.plos.org/10.1371/journal.pgen.1004129>.
- [18] Dan Chen et al. "Analysis of the genetic architecture of susceptibility to cervical cancer indicates that common SNPs explain a large proportion of the heritability". In: *Carcinogenesis* 36.9 (Sept. 2015), pp. 992–998. ISSN: 0143-3334. DOI: [10.1093/carcin/bgv083](https://doi.org/10.1093/carcin/bgv083). URL: <https://academic.oup.com/carcin/article-lookup/doi/10.1093/carcin/bgv083>.
- [19] Kevin Litchfield et al. "Quantifying the heritability of testicular germ cell tumour using both population-based and genomic approaches". In: *Scientific Reports* 5.1 (Nov. 2015), p. 13889. ISSN: 2045-2322. DOI: [10.1038/srep13889](https://doi.org/10.1038/srep13889). URL: <http://www.nature.com/articles/srep13889>.
- [20] Yadav Sapkota. "Germline DNA variations in breast cancer predisposition and prognosis: a systematic review of the literature." In: *Cytogenetic and genome research* 144.2 (2014), pp. 77–91. ISSN: 1424-859X. DOI: [10.1159/000369045](https://doi.org/10.1159/000369045). URL: <http://www.ncbi.nlm.nih.gov/pubmed/25401968>.
- [21] Pak C. Sham and Shaun M. Purcell. "Statistical power and significance testing in large-scale genetic studies". In: *Nature Reviews Genetics* 15.5 (Apr. 2014), pp. 335–346. ISSN: 14710064. DOI: [10.1038/nrg3706](https://doi.org/10.1038/nrg3706). URL: <http://www.nature.com/doifinder/10.1038/nrg3706>.
- [22] Amit Sud, Ben Kinnersley, and Richard S. Houlston. "Genome-wide association studies of cancer: current insights and future perspectives". In: *Nature reviews. Cancer* 17.11 (2017), pp. 692–704. ISSN: 14741768. DOI: [10.1038/nrc.2017.82](https://doi.org/10.1038/nrc.2017.82). URL: <http://dx.doi.org/10.1038/nrc.2017.82>.

- [23] Geraldine M. Clarke et al. "Basic statistical analysis in genetic case-control studies". In: *Nature Protocols* 6.2 (Feb. 2011), pp. 121–133. ISSN: 17542189. DOI: [10.1038/nprot.2010.182](https://doi.org/10.1038/nprot.2010.182). URL: <http://www.nature.com/doifinder/10.1038/nprot.2010.182>.
- [24] Michael D. Gallagher and Alice S. Chen-Plotkin. "The Post-GWAS Era: From Association to Function". In: *American Journal of Human Genetics* 102.5 (May 2018), pp. 717–730. ISSN: 15376605. DOI: [10.1016/j.ajhg.2018.04.002](https://doi.org/10.1016/j.ajhg.2018.04.002).
- [25] A. Jacquard. "Heritability: One Word, Three Concepts". In: *Biometrics* 39.2 (June 1983), p. 465. ISSN: 0006341X. DOI: [10.2307/2531017](https://doi.org/10.2307/2531017). URL: <https://www.jstor.org/stable/2531017?origin=crossref>.
- [26] Albert Tenesa and Chris S Haley. "The heritability of human disease: estimation, uses and abuses". In: (2013). DOI: [10.1038/nrg3377](https://doi.org/10.1038/nrg3377). URL: <https://www.nature.com/articles/nrg3377.pdf>.
- [27] Jian Yang et al. "Common SNPs explain a large proportion of the heritability for human height". In: *Nature Genetics* 42.7 (July 2010), pp. 565–569. ISSN: 1061-4036. DOI: [10.1038/ng.608](https://doi.org/10.1038/ng.608). URL: <http://www.nature.com/articles/ng.608>.
- [28] Jian Yang et al. "Genome partitioning of genetic variation for complex traits using common SNPs". In: *Nature Genetics* 43.6 (June 2011), pp. 519–525. ISSN: 10614036. DOI: [10.1038/ng.823](https://doi.org/10.1038/ng.823). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4295936/>.
- [29] Sang Hong Lee et al. "Estimating missing heritability for disease from genome-wide association studies". In: *American Journal of Human Genetics* 88.3 (2011), pp. 294–305. ISSN: 00029297. DOI: [10.1016/j.ajhg.2011.02.002](https://doi.org/10.1016/j.ajhg.2011.02.002). URL: [https://ac.els-cdn.com/S0002929711000206/1-s2.0-S0002929711000206-main.pdf?\\_tid=b134eef1-4774-41f6-b049-b67b4c6d1e3c&acdnat=1527155214\\_16ca4da6f407d1615fdb88c5b7fe7ba2](https://ac.els-cdn.com/S0002929711000206/1-s2.0-S0002929711000206-main.pdf?_tid=b134eef1-4774-41f6-b049-b67b4c6d1e3c&acdnat=1527155214_16ca4da6f407d1615fdb88c5b7fe7ba2).
- [30] Doug Speed et al. "Improved Heritability Estimation from Genome-wide SNPs". In: *The American Journal of Human Genetics* 91 (2012), pp. 1011–1021. DOI: [10.1016/j.ajhg.2012.10.010](https://doi.org/10.1016/j.ajhg.2012.10.010). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3516604/pdf/main.pdf>.
- [31] Noah Zaitlen et al. "Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits". In: *PLoS Genetics* 9.5 (2013). ISSN: 15537390. DOI: [10.1371/journal.pgen.1003520](https://doi.org/10.1371/journal.pgen.1003520). URL: <http://journals.plos.org/plosgenetics/article/file?id=10.1371/journal.pgen.1003520%7B%5C%7Dtype=printable>.
- [32] David Golan, Eric S. Lander, and Saharon Rosset. "Measuring missing heritability: Inferring the contribution of common variants". In: *Proceedings of the National Academy of Sciences* 111.49 (2014), E5272–E5281. ISSN: 0027-8424. DOI: [10.1073/pnas.1419064111](https://doi.org/10.1073/pnas.1419064111). URL: <http://www.pnas.org/lookup/doi/10.1073/pnas.1419064111>.
- [33] Jian Yang et al. "Concepts, estimation and interpretation of SNP-based heritability". In: *Nature Genetics* (2017). ISSN: 15461718. DOI: [10.1038/ng.3941](https://doi.org/10.1038/ng.3941). arXiv: [arXiv: 1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- [34] J. K. Haseman and R. C. Elston. "The investigation of linkage between a quantitative trait and a marker locus". In: *Behavior Genetics* 2.1 (Mar. 1972), pp. 3–19. ISSN: 0001-8244. DOI: [10.1007/BF01066731](https://doi.org/10.1007/BF01066731). URL: <http://link.springer.com/10.1007/BF01066731>.

- [35] Guo Bo Chen. “Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman-Elston regression”. In: *Frontiers in Genetics* 5.APR (2014). ISSN: 16648021. DOI: [10.3389/fgene.2014.00107](https://doi.org/10.3389/fgene.2014.00107). URL: [www.frontiersin.org](http://www.frontiersin.org).
- [36] Yue Wu and Sriram Sankararaman. “A scalable estimator of SNP heritability for biobank-scale data”. In: *Bioinformatics* 34.13 (2018), pp. i187–i194. ISSN: 14602059. DOI: [10.1093/bioinformatics/bty253](https://doi.org/10.1093/bioinformatics/bty253). URL: <https://github.com/sriramlab/RHE-reg>.
- [37] Jian Yang et al. “GCTA: A Tool for Genome-wide Complex Trait Analysis”. In: *The American Journal of Human Genetics* 88 (2011), pp. 76–82. DOI: [10.1016/j.ajhg.2010.11.011](https://doi.org/10.1016/j.ajhg.2010.11.011). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3014363/pdf/main.pdf>.
- [38] Doug Speed et al. “Reevaluation of SNP heritability in complex human traits”. In: *Nature Genetics* 49.7 (May 2017), pp. 986–992. ISSN: 1061-4036. DOI: [10.1038/ng.3865](https://doi.org/10.1038/ng.3865). URL: <http://www.nature.com/doifinder/10.1038/ng.3865>.
- [39] Brendan Bulik-Sullivan et al. “LD score regression distinguishes confounding from polygenicity in genome-wide association studies”. In: *Nature Genetics* 47.3 (2015), pp. 291–295. ISSN: 15461718. DOI: [10.1038/ng.3211](https://doi.org/10.1038/ng.3211). URL: <http://dx.doi.org/10.1038/ng.3211>.
- [40] Xiang Zhou, Peter Carbonetto, and Matthew Stephens. “Polygenic Modeling with Bayesian Sparse Linear Mixed Models”. en. In: *PLoS Genetics* 9.2 (Feb. 2013), e1003264. ISSN: 15537390. DOI: [10.1371/journal.pgen.1003264](https://doi.org/10.1371/journal.pgen.1003264). URL: <http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003264>.
- [41] Omer Weissbrod, Jonathan Flint, and Saharon Rosset. “Estimating SNP-Based Heritability and Genetic Correlation in Case-Control Studies Directly and with Summary Statistics.” In: *American journal of human genetics* 103.1 (July 2018), pp. 89–99. ISSN: 1537-6605. DOI: [10.1016/j.ajhg.2018.06.002](https://doi.org/10.1016/j.ajhg.2018.06.002). URL: <http://www.ncbi.nlm.nih.gov/pubmed/29979983>  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6035374>.
- [42] Doug Speed and David J. Balding. “SumHer better estimates the SNP heritability of complex traits from summary statistics”. In: *Nature Genetics* 51.2 (Feb. 2019), pp. 277–284. ISSN: 1061-4036. DOI: [10.1038/s41588-018-0279-5](https://doi.org/10.1038/s41588-018-0279-5). URL: <http://www.nature.com/articles/s41588-018-0279-5>.
- [43] Kangcheng Hou et al. “Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture”. In: *Nature Genetics* 51.8 (Aug. 2019), pp. 1244–1251. ISSN: 1061-4036. DOI: [10.1038/s41588-019-0465-0](https://doi.org/10.1038/s41588-019-0465-0). URL: <http://www.nature.com/articles/s41588-019-0465-0>.
- [44] Jie Zheng et al. “LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis”. In: *Bioinformatics* 33.2 (Jan. 2017), pp. 272–279. ISSN: 1367-4803. DOI: [10.1093/bioinformatics/btw613](https://doi.org/10.1093/bioinformatics/btw613). URL: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw613>.
- [45] Brendan Bulik-Sullivan. “Relationship between LD Score and Haseman-Elston Regression”. In: *bioRxiv* (Apr. 2015), p. 018283. DOI: [10.1101/018283](https://doi.org/10.1101/018283). URL: <https://www.biorxiv.org/content/10.1101/018283v1.full>.

- [46] Luke M. Evans et al. "Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits". In: *Nature Genetics* 50.5 (2018), pp. 737–745. ISSN: 15461718. DOI: [10.1038/s41588-018-0108-x](https://doi.org/10.1038/s41588-018-0108-x). URL: <https://www.nature.com/articles/s41588-018-0108-x.pdf%20https://www.biorxiv.org/content/early/2017/03/10/115527>.
- [47] Steven Gazal et al. "Reconciling S-LDSC and LDK functional enrichment estimates". In: *Nature Genetics* 51.8 (Aug. 2019), pp. 1202–1204. ISSN: 1061-4036. DOI: [10.1038/s41588-019-0464-1](https://doi.org/10.1038/s41588-019-0464-1). URL: <http://www.nature.com/articles/s41588-019-0464-1>.
- [48] Bogdan Pasaniuc and Alkes L. Price. "Dissecting the genetics of complex traits using summary association statistics". In: *Nature Reviews Genetics* 18.2 (2017), pp. 117–127. ISSN: 14710064. DOI: [10.1038/nrg.2016.142](https://doi.org/10.1038/nrg.2016.142).
- [49] Tian Ge et al. "Phenome-wide heritability analysis of the UK Biobank". In: *PLoS Genetics* 13.4 (2017), pp. 1–21. ISSN: 15537404. DOI: [10.1371/journal.pgen.1006711](https://doi.org/10.1371/journal.pgen.1006711).
- [50] Hilary K. Finucane et al. "Partitioning heritability by functional annotation using genome-wide association summary statistics". In: *Nature Genetics* 47.11 (2015), pp. 1228–1235. ISSN: 15461718. DOI: [10.1038/ng.3404](https://doi.org/10.1038/ng.3404).
- [51] Steven Gazal et al. "Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection". In: *Nature Genetics* 49.10 (Oct. 2017), pp. 1421–1427. ISSN: 1061-4036. DOI: [10.1038/ng.3954](https://doi.org/10.1038/ng.3954). URL: <http://www.nature.com/articles/ng.3954>.
- [52] Huwenbo Shi, Gleb Kichaev, and Bogdan Pasaniuc. "Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data". In: *American Journal of Human Genetics* 99.1 (2016), pp. 139–153. ISSN: 15376605. DOI: [10.1016/j.ajhg.2016.05.013](https://doi.org/10.1016/j.ajhg.2016.05.013). URL: <http://dx.doi.org/10.1016/j.ajhg.2016.05.013>.
- [53] Tomaz Berisa and Joseph K. Pickrell. "Approximately independent linkage disequilibrium blocks in human populations". In: *Bioinformatics* 32.2 (Sept. 2015), btv546. ISSN: 1367-4803. DOI: [10.1093/bioinformatics/btv546](https://doi.org/10.1093/bioinformatics/btv546). URL: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv546>.
- [54] *The Breast Association Consortium*. 2019. URL: <http://bcac.ccge.medschl.cam.ac.uk/>.
- [55] Kyriaki Michailidou et al. "Association analysis identifies 65 new breast cancer risk loci". In: *Nature* 551.7678 (Oct. 2017), pp. 92–94. ISSN: 0028-0836. DOI: [10.1038/nature24284](https://doi.org/10.1038/nature24284). URL: <http://www.nature.com/doifinder/10.1038/nature24284>.
- [56] Cathie Sudlow et al. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age". In: *PLOS Medicine* 12.3 (Mar. 2015), e1001779. ISSN: 1549-1676. DOI: [10.1371/journal.pmed.1001779](https://doi.org/10.1371/journal.pmed.1001779). URL: <https://dx.plos.org/10.1371/journal.pmed.1001779>.
- [57] Joshua N Sampson et al. "Analysis of Heritability and Shared Heritability Based on Genome-Wide Association Studies for Thirteen Cancer Types." In: *Journal of the National Cancer Institute* 107.12 (Dec. 2015), djv279. ISSN: 1460-2105. DOI: [10.1093/jnci/djv279](https://doi.org/10.1093/jnci/djv279). URL: <http://www.ncbi.nlm.nih.gov/pubmed/26464424%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4806328>.

- [58] Xia Jiang et al. "Shared heritability and functional enrichment across six solid cancers". In: *Nature Communications* 10.1 (Dec. 2019), p. 431. ISSN: 2041-1723. DOI: [10.1038/s41467-018-08054-4](https://doi.org/10.1038/s41467-018-08054-4). URL: <http://www.nature.com/articles/s41467-018-08054-4>.
- [59] Wouter van Rheenen et al. "Genetic correlations of polygenic disease traits: from theory to practice". In: *Nature Reviews Genetics* 20.10 (2019), pp. 567–581. ISSN: 14710064. DOI: [10.1038/s41576-019-0137-z](https://doi.org/10.1038/s41576-019-0137-z). URL: [www.nature.com/nrg](http://www.nature.com/nrg).
- [60] Manuel A. Ferreira et al. "Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer". In: *Nature Communications* 10.1 (Dec. 2019). ISSN: 20411723. DOI: [10.1038/s41467-018-08053-5](https://doi.org/10.1038/s41467-018-08053-5).
- [61] Xingyi Guo et al. "A Comprehensive cis-eQTL Analysis Revealed Target Genes in Breast Cancer Susceptibility Loci Identified in Genome-wide Association Studies". In: *American Journal of Human Genetics* 102.5 (May 2018), pp. 890–903. ISSN: 15376605. DOI: [10.1016/j.ajhg.2018.03.016](https://doi.org/10.1016/j.ajhg.2018.03.016).
- [62] Huwenbo Shi et al. "Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits". In: *The American Journal of Human Genetics* 101.5 (Nov. 2017), pp. 737–751. ISSN: 00029297. DOI: [10.1016/j.ajhg.2017.09.022](https://doi.org/10.1016/j.ajhg.2017.09.022). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0002929717303919>.
- [63] Elise Ruark et al. "Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14". In: *Nature Genetics* 45.6 (June 2013), pp. 686–689. ISSN: 10614036. DOI: [10.1038/ng.2635](https://doi.org/10.1038/ng.2635).
- [64] Armin P. Schoech et al. "Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection". In: *Nature Communications* 10.1 (Dec. 2019), p. 790. ISSN: 2041-1723. DOI: [10.1038/s41467-019-08424-6](https://doi.org/10.1038/s41467-019-08424-6). URL: <http://www.nature.com/articles/s41467-019-08424-6>.
- [65] Douglas Hanahan and Robert A. Weinberg. *The hallmarks of cancer*. Jan. 2000. DOI: [10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9).
- [66] Douglas Hanahan and Robert A. Weinberg. "Hallmarks of cancer: The next generation". In: *Cell* 144.5 (Mar. 2011), pp. 646–674. ISSN: 00928674. DOI: [10.1016/j.cell.2011.02.013](https://doi.org/10.1016/j.cell.2011.02.013).
- [67] Panagiotis Apostolou and Ioannis Papanotiriou. "Current perspectives on CHEK2 mutations in breast cancer". In: *Breast Cancer: Targets and Therapy* 9 (May 2017), pp. 331–335. ISSN: 11791314. DOI: [10.2147/BCTT.S111394](https://doi.org/10.2147/BCTT.S111394).
- [68] C Cybulski et al. "CHEK2 is a multiorgan cancer susceptibility gene." In: *American journal of human genetics* 75.6 (Dec. 2004), pp. 1131–5. ISSN: 0002-9297. DOI: [10.1086/426403](https://doi.org/10.1086/426403). URL: <http://www.ncbi.nlm.nih.gov/pubmed/15492928>  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1182149>.
- [69] Stian Knappskog et al. "Low expression levels of ATM may substitute for CHEK2/TP53 mutations predicting resistance towards anthracycline and mitomycin chemotherapy in breast cancer." In: *Breast cancer research : BCR* 14.2 (Mar. 2012), R47. ISSN: 1465-542X. DOI: [10.1186/bcr3147](https://doi.org/10.1186/bcr3147). URL: <http://www.ncbi.nlm.nih.gov/pubmed/22420423>  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3446381>.
- [70] Thomas M. Campbell et al. "Era binding by transcription factors NFIB and YBX1 enables FGFR2 signaling to modulate estrogen responsiveness in breast cancer". In: *Cancer Research* 78.2 (2018), pp. 410–421. ISSN: 15387445. DOI: [10.1158/0008-5472.CAN-17-1153](https://doi.org/10.1158/0008-5472.CAN-17-1153).



- [71] David J. Hunter et al. "A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer". In: *Nature Genetics* 39.7 (July 2007), pp. 870–874. ISSN: 10614036. DOI: [10.1038/ng2075](https://doi.org/10.1038/ng2075).
- [72] Ellen Margrethe Haugsten et al. "Roles of fibroblast growth factor receptors in carcinogenesis". In: *Molecular Cancer Research* 8.11 (2010), pp. 1439–1452. ISSN: 15417786. DOI: [10.1158/1541-7786.MCR-10-0168](https://doi.org/10.1158/1541-7786.MCR-10-0168).
- [73] Shuwei Wang and Zhongyang Ding. "Fibroblast growth factor receptors in breast cancer". In: *Tumor Biology* 39.5 (2017). ISSN: 14230380. DOI: [10.1177/1010428317698370](https://doi.org/10.1177/1010428317698370).
- [74] Young Kwang Chae et al. "Inhibition of the fibroblast growth factor receptor (FGFR) pathway: The current landscape and barriers to clinical application". In: *Oncotarget* 8.9 (2017), pp. 16052–16074. ISSN: 19492553. DOI: [10.18632/oncotarget.14109](https://doi.org/10.18632/oncotarget.14109).
- [75] Vasudha Sehgal and Prahlad T. Ram. "Network Motifs in JNK Signaling". In: *Genes and Cancer* 4.9-10 (2013), pp. 409–413. ISSN: 19476019. DOI: [10.1177/1947601913507577](https://doi.org/10.1177/1947601913507577).
- [76] Magali Michaut et al. "Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer". In: *Scientific Reports* 6.July 2015 (2016), pp. 1–13. ISSN: 20452322. DOI: [10.1038/srep18517](https://doi.org/10.1038/srep18517).
- [77] Tomohiro Izumiyama et al. "A novel big protein TPRBK possessing 25 units of TPR motif is essential for the progress of mitosis and cytokinesis". In: *Gene* 511.2 (Dec. 2012), pp. 202–217. ISSN: 03781119. DOI: [10.1016/j.gene.2012.09.061](https://doi.org/10.1016/j.gene.2012.09.061).
- [78] Karoline B. Kuchenbaecker et al. "Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers". In: *JAMA - Journal of the American Medical Association* 317.23 (2017), pp. 2402–2416. ISSN: 15383598. DOI: [10.1001/jama.2017.7112](https://doi.org/10.1001/jama.2017.7112).
- [79] Helen Cavanagh and Katherine M A Rogers. "The role of BRCA1 and BRCA2 mutations in prostate, pancreatic and stomach cancers". In: *Hereditary Cancer in Clinical Practice* 13.1 (2015), pp. 1–7. ISSN: 18974287. DOI: [10.1186/s13053-015-0038-x](https://doi.org/10.1186/s13053-015-0038-x).
- [80] Amit M. Oza et al. "Antitumor activity and safety of the PARP inhibitor rucaparib in patients with high-grade ovarian carcinoma and a germline or somatic BRCA1 or BRCA2 mutation: Integrated analysis of data from Study 10 and ARIEL2". In: *Gynecologic Oncology* 147.2 (2017), pp. 267–275. ISSN: 10956859. DOI: [10.1016/j.ygyno.2017.08.022](https://doi.org/10.1016/j.ygyno.2017.08.022).
- [81] Kathleen N. Moore et al. "Niraparib monotherapy for late-line treatment of ovarian cancer (QUADRA): a multicentre, open-label, single-arm, phase 2 trial". In: *The Lancet Oncology* 20.5 (2019), pp. 636–648. ISSN: 14745488. DOI: [10.1016/S1470-2045\(19\)30029-4](https://doi.org/10.1016/S1470-2045(19)30029-4).
- [82] Liisa M. Pelttari et al. "RAD51B in familial breast cancer". In: *PLoS ONE* 11.5 (2016), pp. 1–18. ISSN: 19326203. DOI: [10.1371/journal.pone.0153788](https://doi.org/10.1371/journal.pone.0153788).
- [83] Natsuko Suwaki, Kerstin Klare, and Madalena Tarsounas. "RAD51 paralogs: Roles in DNA damage signalling, recombinational repair and tumorigenesis". In: *Seminars in Cell and Developmental Biology* 22.8 (Oct. 2011), pp. 898–905. ISSN: 10963634. DOI: [10.1016/j.semcd.2011.07.019](https://doi.org/10.1016/j.semcd.2011.07.019).
- [84] Rohit Prakash et al. "Homologous recombination and human health: The roles of BRCA1, BRCA2, and associated proteins". In: *Cold Spring Harbor Perspectives in Biology* 7.4 (2015), pp. 1–27. ISSN: 19430264. DOI: [10.1101/cshperspect.a016600](https://doi.org/10.1101/cshperspect.a016600).



- [85] Hannah L. Klein. "The consequences of Rad51 overexpression for normal and tumor cells". In: *DNA Repair* 7.5 (May 2008), pp. 686–693. ISSN: 15687864. DOI: [10.1016/j.dnarep.2007.12.008](https://doi.org/10.1016/j.dnarep.2007.12.008).
- [86] Osamu Date et al. "Haploinsufficiency of RAD51B causes centrosome fragmentation and aneuploidy in human cells". In: *Cancer Research* 66.12 (2006), pp. 6018–6024. ISSN: 00085472. DOI: [10.1158/0008-5472.CAN-05-2803](https://doi.org/10.1158/0008-5472.CAN-05-2803).
- [87] Joshua A. Betts et al. "Long Noncoding RNAs CUPID1 and CUPID2 Mediate Breast Cancer Risk at 11q13 by Modulating the Response to DNA Damage". In: *American Journal of Human Genetics* 101.2 (Aug. 2017), pp. 255–266. ISSN: 15376605. DOI: [10.1016/j.ajhg.2017.07.007](https://doi.org/10.1016/j.ajhg.2017.07.007).
- [88] Goran Lakisic et al. "Role of the BAHD1 Chromatin-Repressive Complex in Placental Development and Regulation of Steroid Metabolism". In: *PLoS Genetics* 12.3 (2016), pp. 1–26. ISSN: 15537404. DOI: [10.1371/journal.pgen.1005898](https://doi.org/10.1371/journal.pgen.1005898).
- [89] Man Peng et al. "MIER3 suppresses colorectal cancer progression by down-regulating Sp1, inhibiting epithelial-mesenchymal transition". In: *Scientific Reports* 7.1 (Dec. 2017). ISSN: 20452322. DOI: [10.1038/s41598-017-11374-y](https://doi.org/10.1038/s41598-017-11374-y).
- [90] Aaron D. DenDekker et al. "Rat Mcs1b is concordant to the genome-wide association-identified breast cancer risk locus at human 5q11.2 and MIER3 is a candidate cancer susceptibility gene". In: *Cancer Research* 72.22 (2012), pp. 6002–6012. ISSN: 00085472. DOI: [10.1158/0008-5472.CAN-12-0748](https://doi.org/10.1158/0008-5472.CAN-12-0748).
- [91] Gordon J. Strewler. "The physiology of parathyroid hormone-related protein". In: *New England Journal of Medicine* 342.3 (Jan. 2000), pp. 177–185. ISSN: 00284793. DOI: [10.1056/NEJM200001203420306](https://doi.org/10.1056/NEJM200001203420306).
- [92] Wonnam Kim and John J. Wysolmerski. "Calcium-sensing receptor in breast physiology and cancer". In: *Frontiers in Physiology* 7.SEP (2016), pp. 1–11. ISSN: 1664042X. DOI: [10.3389/fphys.2016.00440](https://doi.org/10.3389/fphys.2016.00440).
- [93] T. John Martin and Rachelle W. Johnson. "Multiple actions of parathyroid hormone-related protein in breast cancer bone metastasis". In: *British Journal of Pharmacology* (2019). ISSN: 14765381. DOI: [10.1111/bph.14709](https://doi.org/10.1111/bph.14709).
- [94] Debyani Chakravarty et al. "OncoKB: A Precision Oncology Knowledge Base." In: *JCO precision oncology* 2017 (July 2017). ISSN: 2473-4284. DOI: [10.1200/P0.17.00011](https://doi.org/10.1200/P0.17.00011). URL: <http://www.ncbi.nlm.nih.gov/pubmed/28890946><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5586540>.
- [95] Zbyslaw Sondka et al. "The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers." In: *Nature reviews. Cancer* 18.11 (Nov. 2018), pp. 696–705. ISSN: 1474-1768. DOI: [10.1038/s41568-018-0060-1](https://doi.org/10.1038/s41568-018-0060-1). URL: <http://www.ncbi.nlm.nih.gov/pubmed/30293088>.
- [96] *Ritchie Lab visualisation tools*. URL: <http://visualization.ritchielab.org/home/index>.