The Breast 41 (2018) 14–18

Contents lists available at ScienceDirect

The Breast

journal homepage: www.elsevier.com/brst

Original article

Transmission of breast cancer polygenic risk based on single nucleotide polymorphisms

David G. Cox [a], Pierre-Etienne Heudel [b], Julie Henry [c], Xavier Pivot [d, *]

[a] INSERM U1052, 28 rue Laënnec, 69008, Lyon, France
[b] Oncology Department, Centre Léon Bérard, 28 rue Laënnec, 69008, Lyon, France
[c] Ecole Normale Superieur de Lyon, UMR5037, 15 parvis René-Descartes BP 7000, 69342, Lyon Cedex 07, France
[d] Centre Paul Strauss, Porte de l'Hopital Strasbourg, France

ARTICLE INFO

Article history:
Received 4 January 2018
Received in revised form
5 May 2018
Accepted 11 June 2018
Available online 14 June 2018

ABSTRACT

Aim: The goal of the present study was to further refine how polygenic risk scores may be used in a large population and to quantify the transmission of risk score through generations.
Methods: Allele frequencies from the 1000 Genomes data for 159 single nucleotide polymorphisms associated with breast cancer risk were used. A breast cancer risk score was calculated among 100,000 people. Choosing two "parents" and the alleles they transmit at random, 100,000 "daughters" were simulated. The population was divided by deciles of risk score. Comparing mean risk score in the mother and daughter populations provided information regarding the general relationship at a population level. By examining the distribution of daughter's risk score within each decile of maternal risk score, the transmission was evaluated at the subject level.
Results: Mean values of risk score were 85.1 (St Dev = 7.5) and 85.0 (St Dev = 7.5) for the populations of mothers and daughters, respectively (mean absolute difference = 0.02, p = 0.48). When examining the transmission of risk score from mothers to daughters in specific deciles of risk, statistically significant differences were observed in all deciles (ranged between 0.001 and $< 2.2 \times 10^{-16}$).
Conclusion: The relationship at the subject level will not provide information regarding prevention and screening of offspring based on the knowledge of parents' risk score alone. The present results show that risk estimation by polygenic risk scores is personal, and evaluation of risk score is required for each individual.

1. Introduction

Breast cancer has long been known to have a hereditary component [1]. Linkage studies in the 1990s led to the identification of two major breast cancer susceptibility genes, *BRCA1* [2] and *BRCA2* [3]. The increased risk associated with hereditary mutations in these genes often appears in family trees through their history of breast and/or ovarian cancer occurrence [4]. The inheritance model of familial breast cancer is generally considered as autosomal dominant transmission. In other words, a single mutation conferring a very high risk is transmitted directly through families. However, a large proportion of women referred for sequencing of *BRCA1/2* based on their family history are not found to be carriers of predisposing mutations in either gene.

The completion of the human genome project in the early 2000s provided the capacity to search for correlation between common variants and human phenotype. The most abundant type of variants is the single nucleotide polymorphism (SNP). The architecture of SNP distribution allows the identification of polymorphisms that capture a large proportion of variability in the genome. Genotyping arrays were developed to measure this variability and used for Genome Wide Association Studies (GWAS). Enthusiasm for the first breast cancer GWAS were limited due to the poor relationship between few variants and the risk. Over the past decade, the accumulation of evidence suggests the emergence of a relationship between breast cancer risk [5,6] and the constellation of numerous common germline genetic variants. Recently, analysis of more than 200,000 women definitively established a relationship between 159 SNPs and breast cancer risk [7]. Each of these variants

* Corresponding author. Centre Paul Strauss 3 rue de la porte de l'Hopital BP 30042 67065, Strasbourg Cedex, France.
E-mail address: xpivot@strasbourg.unicancer.fr (X. Pivot).

https://doi.org/10.1016/j.breast.2018.06.006
0960-9776/

independently has a limited effect on breast cancer risk [8]. However, when combined with non-genetic risk factors in a composite risk score, the identification of high- and low-risk groups of women are established [9,10]. The search for additional variants continues, and these variants will likely further increase the ability to provide a risk score that stratifies the population by their risk of developing breast cancer.

In the present manuscript, we focus on the context of risk estimates based on a constellation of very low-penetrance alleles. Intuitively, there should be some correlation in risk score from one generation to the next. However, given that each variant is independent, and the choice of which allele is transmitted at each locus is apparently random, direct estimation of the risk score among children based on their parents' score is not straightforward. Fisher had shown mathematically the transmission of polygenic traits from generation to generation, and emphasized the uncertain estimation for a risk transmission in offspring by polygenic risk models [11]. In order to quantify Fisher's calculations, and make them more readily interpretable for non-statistical geneticists, we have modeled risk transmission of polygenic risk scores from "mothers" to "daughters". The results of this study provide information essential to the debate regarding the use of risk scores that include SNP information in precision estimates of disease risk for personalized breast cancer prevention and screening.

2. Materials and methods

2.1. Study design

We used allele frequencies from the 1000 Genomes data for 159 single nucleotide polymorphisms presently associated with breast cancer risk, to randomly generate genotypes for a hypothetical population of 100,000 people. As the human genome is diploid, a subject may carry between zero and two copies of the same allele at each locus. At each locus, there is an allele at high risk (coded 1) and one at low risk (coded 0). The possible result for a subject, at any given locus, is therefore 0, 1 or 2 based on the number of risk alleles carried. A breast cancer risk score can then be calculated for each "person" by summing the number of risk alleles carried at each locus. This population was divided by deciles of risk score, defining risk profiles.

Choosing two "parents" and the alleles they transmit at random, 100,000 "daughters" were simulated from the parent population. Only one daughter per couple was generated. The same variant-based risk score was calculated among the daughter population. The "mother" was identified at random between the two parents and the breast cancer risk score was the point of reference for her daughter's risk score. The mother-daughter relationship was retained, to allow evaluation of risk transmission.

The relationship between the risk score of mothers and their daughters were assessed using paired t-tests. Comparing mean risk score in the overall mother and daughter populations provides information regarding the general relationship at a population level. By using a paired t-test, as well as the distribution of daughter's risk score within each decile of maternal risk score, a more fine-grain evaluation of the transmission of risk can be evaluated at the subject level. All simulations and analyses were carried out in the R environment, and code is available from the authors upon request.

3. Results

Resulting from the hypothetical population of 100,000 people, the decile cutoffs of polygenic risk score subgroups were as follows: <76, 79, 81, 83, 85, 87, 89, 91, >95. The mean values of risk score were 85.1 (StDev = 7.5) and 85.0 (StDev = 7.5) for the populations of mothers and daughters, respectively. There was no difference in mean polygenic risk score between the parental and daughter populations (mean absolute difference = 0.02, $p = 0.48$).

Table 1
Paired t-test of mean polygenic risk scores between mothers and daughters in each decile of maternal risk.

Maternal Risk Score deciles	Mean difference (95% CI) between maternal deciles and daughter risk score	Paired t-test p-value
Overall	0.02 (−0.03–0.06)	0.48
>95	6.45 (6.32–6.58)	$<2.2 \times 10^{-16}$
91–95	3.69 (3.57–3.80)	$<2.2 \times 10^{-16}$
89–91	2.17 (2.04–2.31)	$<2.2 \times 10^{-16}$
87–89	1.28 (1.16–1.41)	$<2.2 \times 10^{-16}$
85–87	0.20 (0.08–0.32)	0.001
83–85	−0.77 (−0.89–−0.64)	$<2.2 \times 10^{-16}$
81–83	−1.71 (−1.84–−1.58)	$<2.2 \times 10^{-16}$
79–81	−2.80 (−2.94–−2.66)	$<2.2 \times 10^{-16}$
76–79	−3.87 (−4.00–3.74)	$<2.2 \times 10^{-16}$
<76	−6.49 (−6.62–−6.37)	$<2.2 \times 10^{-16}$

However, when examining the transmission of risk score from mothers to daughters in specific deciles of risk score, statistically significant differences in mean ranged between $p = 0.001$ and $p < 2.2 \times 10^{-16}$ (Table 1). The box-plots with lines connecting mothers and daughters for each decile are presented in Fig. 1, showing the large distribution of daughter risk scores, despite restraining maternal risk scores. This is further highlighted by Fig. 2 and Table 2, which show the distribution of the daughter's risk score, based on maternal decile risk group.

4. Discussion

As our capacity to explore and exploit the information contained in the human genome has grown, the potential to use the correlation between genetic variability and disease risk has become a reality. To properly exploit this potential, understanding transmission of risk is of paramount importance. Health care professionals have become accustomed to risk transmission in the context of monogenic familial syndromes. The inheritance through a family of predisposing *BRCA* mutations, taking into account their autosomal dominant transmission, is well known. Other genes related to moderate breast cancer risk, such as *PALB2*, *ATM*, *CHEK2*, have subsequently been identified, and the transmission of these isolated genes is also easy to follow through offspring.

However, little is published regarding how to appreciate the risk that may be estimated using multi-variant risk scores. Accurately modeling the transmission of this risk is a major step for using the information contained in our genomes to provide more personalized and precise health care, including prevention and screening programs.

In the context of monogenic highly penetrant mutations, the probability that relatives carry the same mutations is easily calculated, and the implications of this probability on the health of relatives that carry the mutation are straightforward. In the present study, a multifactorial risk score based on many independent variants showed correlation at the population level between risk scores in mothers and daughters. Nevertheless, at the subject level there was a high variability of risk score resulting from the transmission from parents to offsprings. Therefore, we demonstrated that the constellation of risk alleles carried by any given individual is truly personal, and their risk needs to be evaluated individually, regardless of the risk levels of their parents, siblings, or offspring.

Even if estimation of risk score categories of daughters from their mother can be estimated, daughters with all possible deciles