Family Study Designs Informed by Tumor Heterogeneity and Multi-Cancer Pleiotropies: The Power of the Utah Population Database

Heidi A. Hanson[1,2,3], Claire L. Leiser[1,4], Michael J. Madsen[1], John Gardner[1], Stacey Knight[5], Melissa Cessna[6,7], Carol Sweeney[1,8,9], Jennifer A. Doherty[1,8,10], Ken R. Smith[1,2,11], Philip S. Bernard[1,12], and Nicola J. Camp[1,2,9]

ABSTRACT

Background: Previously, family-based designs and high-risk pedigrees have illustrated value for the discovery of high- and intermediate-risk germline breast cancer susceptibility genes. However, genetic heterogeneity is a major obstacle hindering progress. New strategies and analytic approaches will be necessary to make further advances. One opportunity with the potential to address heterogeneity via improved characterization of disease is the growing availability of multisource databases. Specific to advances involving family-based designs are resources that include family structure, such as the Utah Population Database (UPDB). To illustrate the broad utility and potential power of multisource databases, we describe two different novel family-based approaches to reduce heterogeneity in the UPDB.

Methods: Our first approach focuses on using pedigree-informed breast tumor phenotypes in gene mapping. Our second approach focuses on the identification of families with similar pleiotropies. We use a novel network-inspired clustering technique to explore multi-cancer signatures for high-risk breast cancer families.

Results: Our first approach identifies a genome-wide significant breast cancer locus at 2q13 [$P = 1.6 \times 10^{-8}$, logarithm of the odds (LOD) equivalent 6.64]. In the region, *IL1A* and *IL1B* are of particular interest, key cytokine genes involved in inflammation. Our second approach identifies five multi-cancer risk patterns. These clusters include expected coaggregations (such as breast cancer with prostate cancer, ovarian cancer, and melanoma), and also identify novel patterns, including coaggregation with uterine, thyroid, and bladder cancers.

Conclusions: Our results suggest pedigree-informed tumor phenotypes can map genes for breast cancer, and that various different cancer pleiotropies exist for high-risk breast cancer pedigrees.

Impact: Both methods illustrate the potential for decreasing etiologic heterogeneity that large, population-based multisource databases can provide.

See all articles in this *CEBP Focus* section, "**Modernizing Population Science.**"

Introduction

The use of the family study design, and high-risk pedigrees in particular, was instrumental in the discovery of germline breast cancer susceptibility genes and our understanding of their pleiotropies (1, 2). However, breast cancers, like other complex diseases, have many sources of heterogeneity that can hinder gene discovery. Efforts to identify additional etiologic risk factors are hampered by these complexities and new methods to identify and reduce sources of heterogeneities are needed to identify novel disease loci. Deconstructing within-site heterogeneity and identification of across-site pleiotropies will require large multisource data resources and computational techniques to mine them. Many large multisource data resources are currently under development throughout the United States and the world (3–9), providing potential opportunities for a new wave of discoveries. In Utah, an established statewide multisource database (the Utah Population Database, UPDB) with linked biobank resources exists. Here, we will describe two different novel family-based approaches using the UPDB, designed to address heterogeneity and identify pleiotropies, to illustrate the broad utility of multisource databases.

Fundamentally necessary to family studies are data for relationship structure and disease, as well as knowledge of population expectations of disease. The former is critical for defining phenotypes that cluster in families and therefore has potential power for genetic discovery. The UPDB is currently the only statewide resource in the United States that links statewide genealogies (5 million records that span 3–18 generations) with a statewide Surveillance, Epidemiology, and End Results (SEER) Program cancer registry [Utah Cancer Registry (UCR), since 1966]. Hence, it allows for both family construction and designation of significant clustering of disease. Other data sources are also linked to the UPDB (https://uofuhealth.utah.edu/huntsman/utah-population-database/data/), including: electronic medical records (1996–present); historical census data (1880; 1900–1940); vital statistics (1905–present); residential histories (back to 1900); linkages to environmental

[1]Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah. [2]Utah Population Database, University of Utah, Salt Lake City, Utah. [3]Department of Surgery, University of Utah, Salt Lake City, Utah. [4]Department of Epidemiology, University of Washington, Seattle, Washington. [5]Intermountain Healthcare, Salt Lake City, Utah. [6]Intermountain Biorepository, Intermountain Healthcare, Salt Lake City, Utah. [7]Department of Pathology, Intermountain Medical Center, Intermountain Healthcare, Salt Lake City, Utah. [8]Utah Cancer Registry, University of Utah, Salt Lake City, Utah. [9]Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah. [10]Department of Population Sciences, University of Utah School of Medicine, Salt Lake City, Utah. [11]Department of Family and Consumer Studies, University of Utah, Salt Lake City, Utah. [12]Department of Pathology, University of Utah, Salt Lake City, Utah.

measures (geographic based); and biobanks. This multisource database is unique and can be harnessed for many designs to study cancer risk and survivorship across the lifespan and across generations (10–14).

Breast cancer is a prime example of a common, complex disease. Substantial etiologic heterogeneity exists both within and across breast cancers in high-risk pedigrees. Reducing heterogeneity is an important design issue in family-based genetic research. For example, even within high-risk pedigrees, the discovery of *BRCA1* and *BRCA2* (*BRCA1/2*) required restriction to early-onset disease to clarify segregation (15, 16). It is well-established that gene expression varies across tumors, and hence tumor expression phenotypes may hold promise for deconstructing heterogeneity. In breast cancer, tumor gene expression has been shown to differentiate tumors into intrinsic subtypes (Luminal A, Luminal B, HER2-enriched, and Basal-like; refs. 17, 18), of which Basal-like has increased *BRCA1* susceptibility (19). The first approach we describe integrates tumor expression phenotypes with gene mapping in high-risk pedigrees. This approach was made possible by record linkages between genealogy, cancer diagnoses, hospital medical records, and biobanks, all available via the UPDB. We previously defined quantitative tumor expression phenotypes associated with high-risk pedigrees not attributed to *BRCA1/2*, and illustrated power for mapping breast cancer loci in one large pedigree (20). Here we apply the same approach to a second large, high-risk breast cancer pedigree.

Cancer pleiotropies are a well-accepted phenomenon, and crucial to genetic counselling for accurate risk predictions. In breast cancer, pleiotropies are known to vary by the risk gene involved (**Fig. 1**). Hence, characterizing families by their patterns of familial cancer risk could provide new opportunities to identify families with similar genetic risk factors. Gene mapping focusing on multi-cancer patterns could also elucidate molecular factors that underlie pleiotropies. For example, Basal-like breast tumors show more gene expression similarities to high-grade serous ovarian cancer than other breast tumor types (21, 22). The multiple linked data sources in the UPDB provide a platform to describe multi-cancer patterns of familial risk. Furthermore, links to biorepositories could support investigations into the molecular factors underlying pleiotropies, and links to environmental data investigations to shared exposures. In the second approach, we illustrated how data-driven methods make it possible to uncover familial multi-cancer signatures. We recently introduced this novel multi-cancer clustering technique and defined four familial multi-

cancer signatures in high-risk bladder cancer families (23). Here, we focus on multi-cancer signatures for high-risk breast cancer families.

Materials and Methods

The UPDB

The vast majority of individuals residing in Utah are represented in the UPDB (24–27). Core to the UPDB is an immense genealogy that is record-linked to many other statewide datasets (including the UCR), with annual updates. The full genealogic dataset contains nearly 5 million people with 28 million records and the linking of multiple distinct records for a specific person allows the UPDB to depict the life history of an individual based on medical and administrative data. There are currently 336,000 cancer records from the UCR with diagnoses beginning in 1966 that are linked to the UPDB. The UPDB is linked to the pathology records of two healthcare systems (University of Utah, Salt Lake City, UT and Intermountain Healthcare, Salt Lake City, UT) that together serve over 85% of the state, and facilitates access to over 4 million formalin-fixed, paraffin-embedded (FFPE) tissue blocks linked to clinical data. It is also linked to external data repositories using a statewide federated ID, including approximately 85% of outpatient claims in the state of Utah (1996–present).

The data contained in the UPDB may be used for biomedical and health-related research. It is a rich and unique resource for cancer research that can support genetic, epidemiologic, public health, and healthcare delivery studies. Overseeing ethical approvals for use of the UPDB data for research is the Resource for Genetic and Epidemiological Research (RGE) body, which was established by Executive Order of the Governor of Utah in 1982. RGE administers access to UPDB through a formal review process to ensure the protection of privacy and confidentiality of the persons and data held in UPDB, and protects the interests of the data contributors (28). A summary list of data contributors can be found in Supplementary Table S1.

Approach 1: Reducing heterogeneity: Breast cancer gene mapping using a tumor expression phenotype

Breast cancer pedigrees were identified in the UPDB using record linkage between the 18-generation genealogy and statewide cancer records from the UCR. High-risk status was defined as a statistical excess of breast cancer–compared UPDB internal rates ($P < 0.05$). Pedigrees known to be attributable to *BRCA1/2* from previous Utah studies were removed (i.e., screen positive or linked to chromosomes
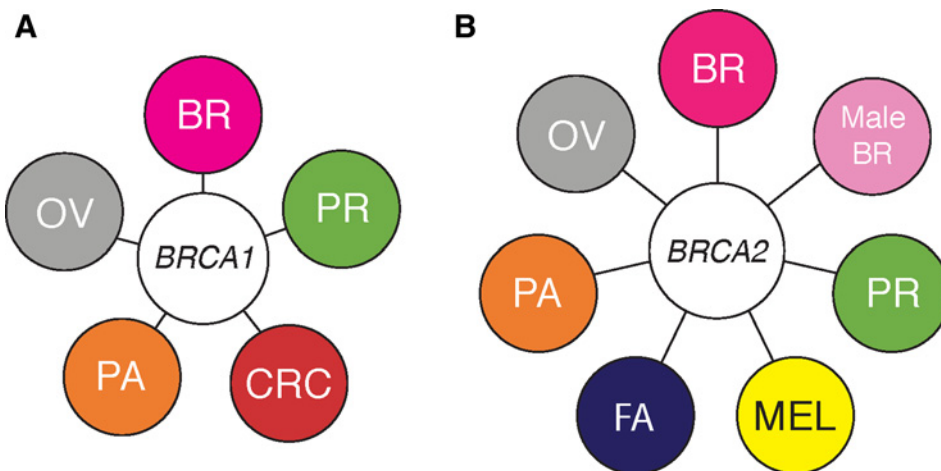

Figure 1.
Familial multiphenograms for *BRCA1* (**A**) and *BRCA2* (**B**). BR, breast cancer; CRC, colorectal cancer; FA, Fanconi anemia; MEL, melanoma; OV, ovarian cancer; PA, pancreatic cancer; PR, prostate cancer. Source: https://ghr.nlm.nih.gov.

17q21 or 13q13). Record linkage between the UPDB and pathology records in the University of Utah (Salt Lake City, UT) and Intermountain Healthcare Systems (Salt Lake City, UT) allowed identification of pathology records and archived tissue blocks. We pursued matched tumor and GU FFPE tissues for 25 high-risk pedigrees. GU refers to tissue that is histologically determined to contain 0% tumor. In the absence of peripheral blood, DNA extracted from GU tissue can be used for germline (inherited) DNA (see Supplementary Materials and Methods for more detail). Eleven of the 25 pedigrees contained at least 15 cases for whom tumor blocks were available. These 11 pedigrees were selected for tumor and germline experiments. Tumor RNA was used for gene expression and GU DNA for germline genotyping. Tumor gene expression was measured using the PAM50 RT-qPCR research assay (29). We used the OmniExpress high-density SNP array for germline genotyping. Quality control included: duplicate check, sex check, SNP call-rate (95%), sample call rate (90%), and failure of Hardy–Weinberg equilibrium ($P \leq 1 \times 10^{-5}$). All women were of European ancestry. Ethical approvals for the study were governed by RGE and Institutional Review Boards (IRB) at the University of Utah (IRB_00096990; Salt Lake City, UT) and Intermountain Healthcare (IRB_1015580; Salt Lake City, UT).

We previously used a set of population-based breast tumors (30) and identified five principal components from the 50 PAM50 classifier genes, referred to as dimensions PC1—PC5 (31). PC3 and PC5 were shown to be significantly different between the population and the pedigree tumors and hence potentially powerful phenotypes for gene mapping in pedigrees. Here we concentrate on high-risk pedigree 1822 (**Fig. 2**) and dimension PC3 as the phenotype of interest. Tumors in

pedigree 1822 were identified as the most significantly different of all 11 pedigrees to population tumors for PC3 ($P = 4.0 \times 10^{-5}$; ref. 20). Germline DNA was available for 46 breast cancer cases and tumor RNA for 31. As described previously (20), we considered breast cancer cases with tumors in the top decile of PC3 in the population as "extreme," resulting in 10 PC3-extreme breast cancer cases for gene mapping in pedigree 1822.

We used Shared Genomic Segment (SGS) analysis (32), a single-pedigree method which identifies chromosomal identity-by-state (IBS) sharing at consecutive SNPs. Segregation from a common ancestor is implied if the observed IBS sharing is significantly longer than expected by chance (33, 34). To address any residual heterogeneity, sharing evidence is assessed over all possible subsets. Statistical significance was determined empirically using a gene-drop approach. Briefly, a gene-drop assigns haplotypes randomly to pedigree founders under the null hypothesis [i.e., according to a population distribution, we used 1000Genomes Project (ref. 35) data for our linkage disequilibrium model; ref. 36]. Mendelian segregation and recombination are simulated through the pedigree structure (37) to generate genotypes for all pedigree members. We used the established Rutgers genetic map (38) for simulating recombination events. For each simulated configuration of genotypes in the pedigree, shared segments are assessed and result in one genome-wide expectation of sharing under the null hypothesis. The gene-drop procedure was repeated to generate a null distribution of sharing from which an empirical estimate of significance for the observed sharing was made. For accurate interpretation, a genome-wide significance threshold was established, which corrects for the subsets within the pedigree and the whole-genome framework. After 1
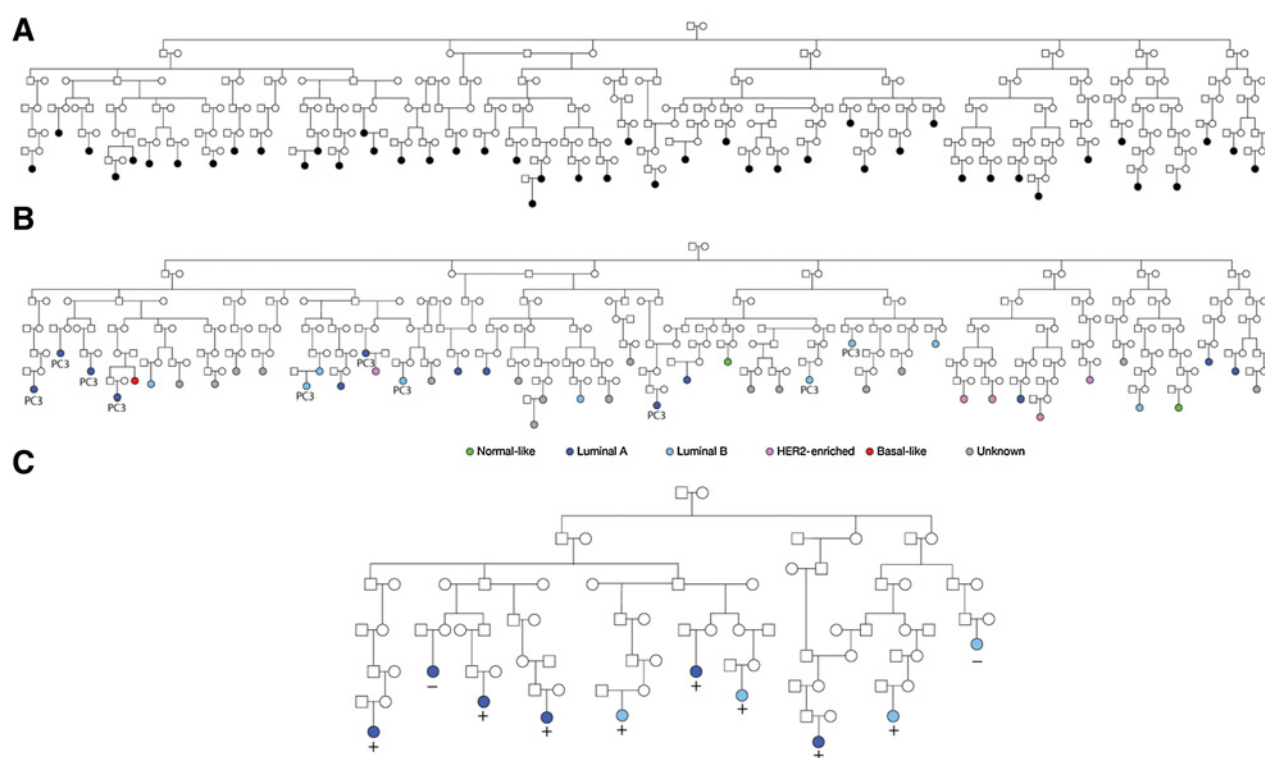
Figure 2.
Pedigree 1822. **A,** Forty-six breast cancer cases with germline DNA available (colored black). **B,** Thirty-one breast cancer cases with tumor expression data (colored green, light/dark blue, purple, or red). Gray color indicates no tumor data. "PC3" indicates cases whose tumors were established to be extreme for PC3. **C,** The reduced pedigree structure for 1822 based on the 10 PC3-extreme breast cancer cases. These 10 were the focus of the SGS gene mapping

million simulations, a gamma distribution was fit to the observed *P* values across the genome. The genome-wide significance threshold was derived from this distribution using the theory of large deviations (39).

Approach 2: Identifying pleiotropic patterns—multi-cancer signatures for familial breast cancer

High-risk breast cancer families were the focus of the clustering to identify multi-cancer pleiotropies. Linked genealogic, demographic, and cancer data from the UPDB were used. First, all individuals with breast cancer ("probands") and their first- (FDR), second- (SDR), and third-degree relatives (TDR) were identified using the UPDB. Only family members known to reside in Utah for at least 1 year from 1966–2017 were included. We identified 27,635 probands with at least one TDR and 1,696,913 family members. Second, this set was reduced to only families with at least 10 relatives to allow for family risk assessment. Familial risk for a cancer type was measured using standardized incidence risk (SIR) ratios accounting for the sex, age, birth-cohort, and person-years of the pedigree members (for a detailed description of SIR calculations, see Supplementary Materials and Methods). Person-years were calculated using the minimum of the first year residing in Utah or 1966 to the year of first cancer diagnosis, last year of residence in Utah (due to death or migration), or 2017. Finally, a total of 5,045 families (including 326,024 family members) were determined as high risk for breast cancer, defined as a statistical excess of cases compared with the age- and sex-adjusted internal rates of the UPDB ($P < 0.05$). These were the basis of our study. This study was approved by IRBs at the University of Utah (IRB_00088870 and IRB_00079328).

Each of the 5,045 high-risk breast cancer families were further characterized by risk for 25 additional cancer types (26 total, including breast cancer). Other cancers were selected on the basis of SEER site codes and frequency (see Supplementary Table S2 for detailed information; ref. 40).

Two risk metrics were used to capture a family's multi-cancer signature. First, *wSIR*, the SIR weighted by the *P* value. This incorporated both the magnitude and significance of the familial risk, and was calculated using the following equation. This metric allowed us to include, but down-weight, SIR values that were not significantly different than the overall population.

$$wSIR_{ij} = SIR_{ij} * \left(1 - \log(p_{ij})\right)$$

Where *p* is the *P value*, *i* is the family, and *j* is the cancer type.

For robustness, and to avoid bias due to large SIRs (especially for rare cancers), we imposed a maximum value such that any *wSIR* values larger than the 90th percentile were set to the 90th percentile value across all families for the cancer type.

$$wSIR = \begin{cases} wSIR_{ij}; & \text{if } wSIR_{ij} \leq wSIR_{j90} \\ wSIR_{j90}; & \text{otherwise} \end{cases}$$

where 90 indicates the 90th percentile for cancer *j*.

Second, we included a dichotomous indicator of risk (I_{SIR}). Families were considered to have "high risk" status for a cancer type ($I_{SIR} = 1$) if the SIR was statistically significant ($P < 0.05$) and "population risk" ($I_{SIR} = 0$) otherwise. As all families were selected to be high risk for breast cancer by design, we substituted the I_{SIR} for breast with an indicator variable for male breast cancer. Our final matrix included 52 risk metrics per family (26 *wSIR* and 26 I_{SIR}).

Clustering was performed on the 5,045 × 52 data matrix (families × risk metrics). A Gower general coefficient (ade4 R package) was used as

the distance metric for clustering as it allows for the simultaneous use of our two risk metric types (*wSIR* continuous and I_{SIR} categorical; detailed information can be found in the Supplementary Data). We used partitioning around medoids (PAM or K-medoids clustering package in R; ref. 41) to measure similarities between the multi-cancer risk signatures of families. K was selected by running a series of iterative models from k = 2 to k = 20 and using Silhouette (Supplementary Fig. S1) and elbow plots to identify the point of diminishing improvement in average Silhouette width.

Bootstrapping was used to evaluate the reproducibility of the clustering (*clustboot* function in R) with 200 random draws. Results from each draw were transformed into a consensus matrix using the ward linkage algorithm and the (*consensusmatrix* function in R) and then plotted in a heatmap used for visualization. The results for k = 5 were stable (Supplementary Fig. S2).

Each cluster in the matrix represents a familial multi-cancer configuration (FMC) signature for high-risk breast cancer families. To describe and compare these clusters (FMCs), we used Cox proportional hazard models to estimate cluster-specific differences in cancer incidence and their 95% confidence intervals (CI) using the R package *survival*. All models controlled for birth year and sex.

Results

Approach 1: Reducing heterogeneity: Breast cancer gene mapping using a tumor expression phenotype

Figure 2 illustrates pedigree 1822, showing the 46 breast cancer cases with germline DNA available (**Fig. 2A**) and the subset of 31 with tumor expression data (**Fig. 2B**). Their intrinsic subtype (the usual purpose of the PAM50) is also indicated for comparison. The 10 PC3-extreme breast cancer cases used in the SGS analyses are shown in **Fig. 2C**. The SGS genome-wide significance threshold for 1822 was determined to be $\alpha = 2.0 \times 10^{-8}$, and one 0.6 Mb region at chromosome 2q13 surpassed this ($P = 1.6 \times 10^{-8}$, from 113.2 to 113.8 Mb). This segment was shared by eight of the 10 extreme PC3 breast cancer cases and was inherited through 38 meioses (**Fig. 2C**). Ten genes are contained in the 2q13 locus: *TTL*; *POLR1B*; *CHCHD5*; *SLC20A1*; *NT5DC4*; *CKAP2L*; *IL1A*; *IL1B*; *IL37*; and *IL36G*.

We explored fine-mapping of the 2q13 locus within the pedigree by assessing the possibility that the shared haplotype inherited to others. We defined the eight SGS sharers as "core sharers" and ranked all other breast cancer cases with genotype data based on their IBS sharing with them at this locus. We sequentially added these breast cancer cases to core sharers based on their ranking, and reassessed SGS sharing across the full set after each addition. **Figure 3** shows how the possible sharing narrows as cases are added. As a *post hoc* analysis, this cannot be formally tested for significance, but it indicates there may be an additional 15 cases who inherit the same 120,567 bp region. This reduced region contains only *NT5DC4, CKAP2L, IL1A*, and *IL1B*.

Approach 2: Identifying pleiotropic patterns—multi-cancer signatures for familial breast cancer

The 5,045 high-risk breast cancer families in the UPDB ranged in size from 10 to 284 relatives (FDR, SDR, and TDR). **Figure 4** shows the hazard rate ratios (HRR) for all 5,045 familial breast cancer families relative to the Utah population and for each familial multi-cancer configuration (FMC$_{1-5}$). The clustering algorithm identified five family types based on their multi-cancer risks: FMC$_1$ (2,159 families, 42.8%), FMC$_2$ (657, 13.0%), FMC$_3$ (625, 12.4%), FMC$_4$ (1,004, 19.9%), and FMC$_5$ (600, 11.9%). While, by definition, all clusters contained a statistical excess of breast cancer, the magnitude of breast cancer risk
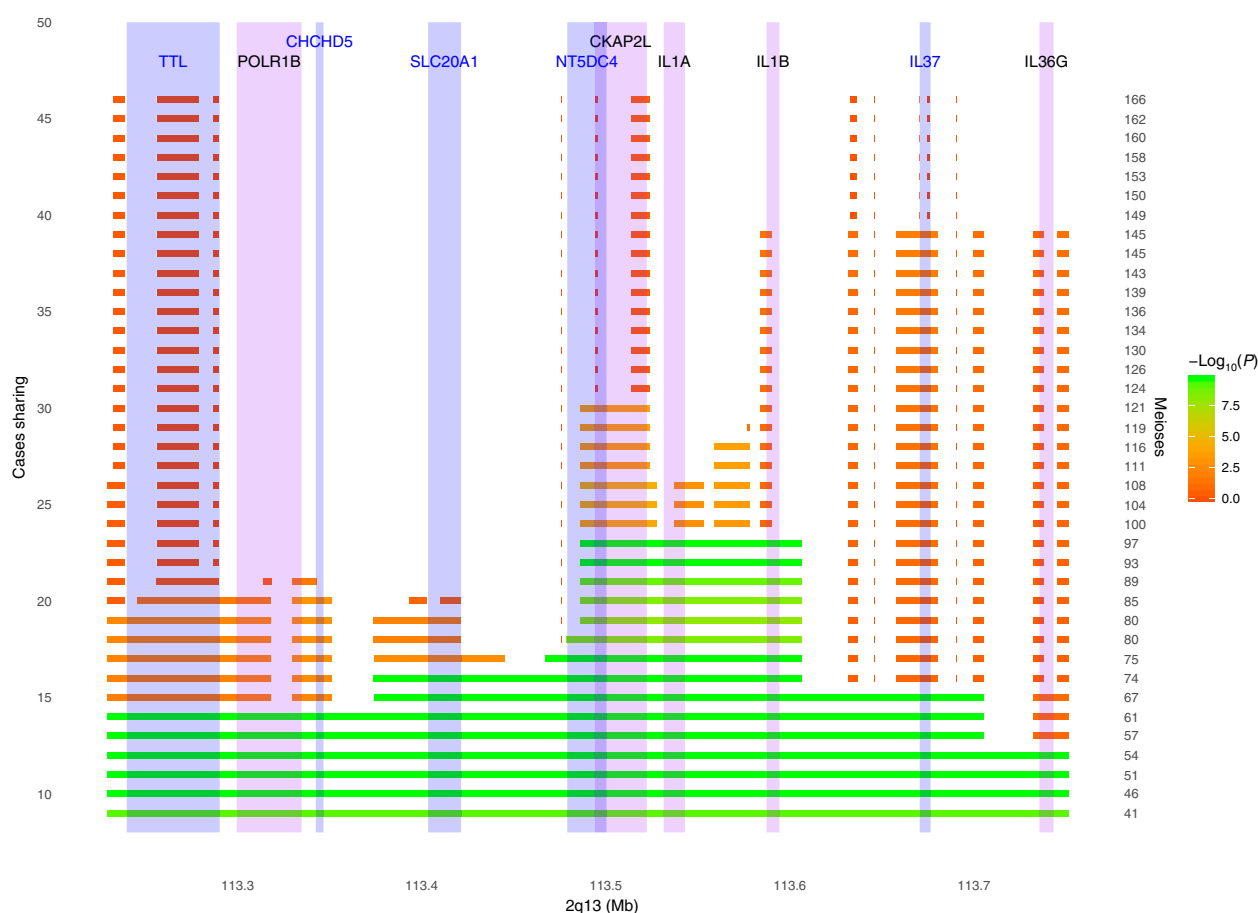
Figure 3.
The left-hand side *y*-axis indicates the number of individuals sharing. From the established 8 "core sharers," individuals are added, thus the range is 9 to 46 individuals (total cases with DNA). The right-hand side *y*-axis indicates the number of meioses separating the set of sharers on each row. The *x*-axis indicates the SNP markers across 2q13. Each horizontal rectangle is a shared segment, with the color indicating the *P* value for the segment (green, highly significant, to red, not significant). The number in the white box on the green rectangles indicates the number of simulations more extreme than the observed segment. At 23 sharers (15 additional to the core 8), significance disappears and returns to that expected.

varied across clusters (see **Table 1**): FMC$_1$ HRR = 3.05 (95% CI, 2.98–3.12), FMC$_2$ HRR = 4.32 (4.14–4.50), FMC$_3$ HRR = 3.79 (3.64–3.94), FMC$_4$ HRR = 6.16 (5.96–6.37), and FMC$_5$ HRR = 3.24 (3.12–3.37).

Separating high-risk breast cancer families into clusters with similar patterns of multi-cancer risk uncovered many differences in effect sizes of cancer risks (including opposing directions), and identified previously undiscovered pleiotropic associations (**Table 1**; **Fig. 4**; Supplementary Fig. S3). We found that the risk of ovarian cancer, an established coaggregation with breast cancer for known risk genes, varied widely by cluster. Ovarian cancer risk for each of the five FMCs was significantly different than the risk estimated from all families together (overall HRR = 1.17; 95% CI, 1.09–1.26; **Table 1**). FMC$_5$ captured extreme increased risk (HRR = 6.10; 95% CI, 5.64–6.61, while the remaining four FMCs showed negative associations (significant decreased risk; **Table 1**; **Fig. 4**). Melanoma, another established cancer associated with breast cancer, was found to vary widely across clusters (**Table 1**; **Fig. 4**). Novel coaggregations were also evident. There was neither established association for larynx cancer, nor a signal for risk to larynx cancer when all high-risk breast cancer families were considered together. However, significant risks (increased and decreased) were seen for larynx cancer in all five FMCs [e.g., FMC$_3$

HRR = 4.93 (95% CI, 4.58–5.31) and FMC$_4$ HRR = 0.19 (95% CI, 0.14–0.27); **Table 1**].

Prostate cancer risk was consistent and modest (1.05–1.20) across all clusters, significantly elevated in four of the FMCs, and borderline in the fifth. Some cancers were consistently absent: bladder, brain, cranial nerves and other nervous system (central nervous system), myeloma, and small intestine. The remaining cancers provided patterns that differentiated FMCs. Families in FMC$_1$ were at moderately increased risk for prostate cancer and acute lymphocytic leukemia (ALL) and had decreased risk for 11 cancers (**Fig. 4**; **Table 1**), with notable decreases in ovarian (HRR = 0.19; 95% CI, 0.15–0.24) and cancer of the larynx (HRR = 0.41; 95% CI, 0.37–0.46). The FMC$_2$ cluster alone showed strong coaggregation of melanoma (HRR = 4.17; 95% CI, 3.95–4.40) and moderate increases in risk for cancers that are usually seen in adolescents, such as testicular, thyroid, non-Hodgkin lymphoma, acute lymphocytic leukemia, and acute myeloid leukemia (**Fig. 4**; **Table 1**). This cluster had increased risk for eight cancer sites, the highest of the FMCs, and decreased risk for two sites, the lowest of the FMCs. FMC$_3$ was the only cluster to exhibit substantial and significant risk for cancer of the larynx (HRR = 4.93; 95% CI, 4.58–5.31) and Hodgkin lymphoma (HRR = 1.53; 95% CI, 1.15–
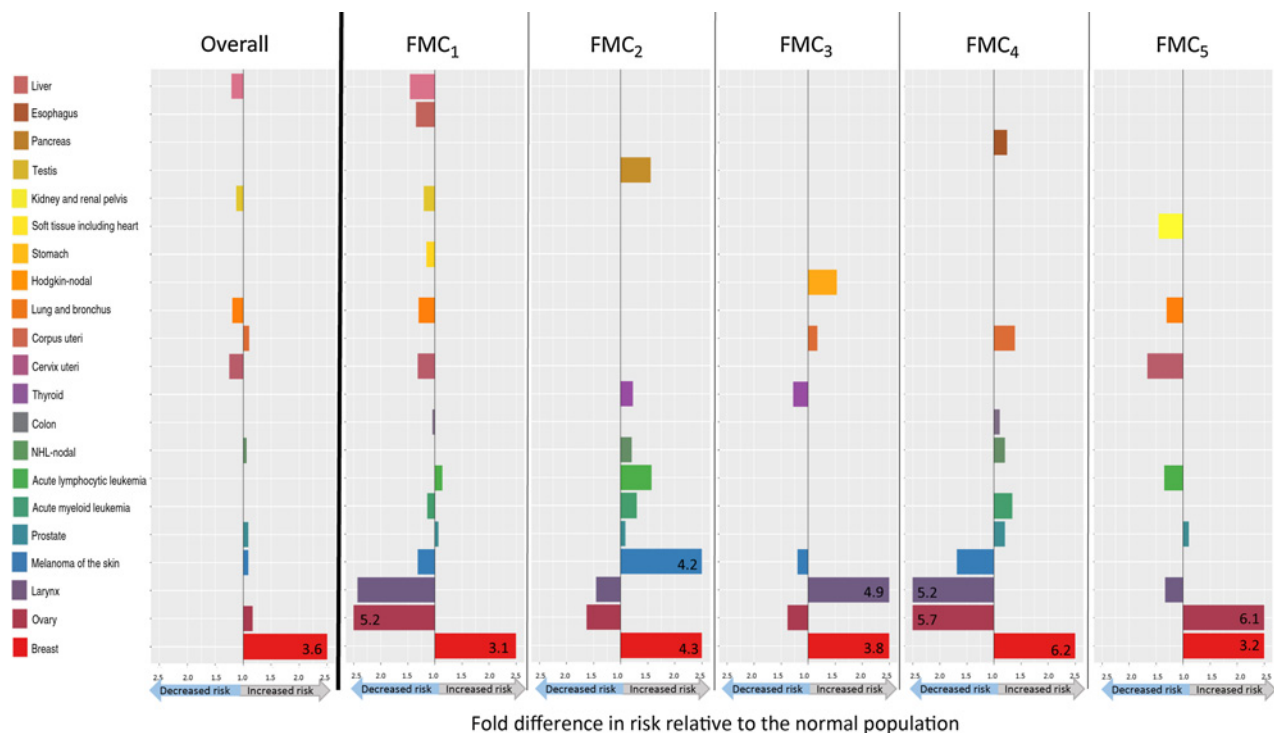
Figure 4.
Familial multiphenograms illustrating the patterns of familial cancer risk across the five high-risk FMCs. The "Overall" column shows the fold difference in risk for all familial breast cancer families relative to the control population. The x-axis is truncated at 2.5, and values larger than 2.5 are noted within the horizontal bars on the chart. Columns FMC$_1$–FMC$_5$ show the unique familial cancer patterns by FMC. These patterns differ significantly from one another and the overall pattern of cancer clustering in familial breast cancer families. NHL, non-Hodgkin lymphoma.

2.04). Families in FMC$_4$ had an increased risk of uterine cancer (HRR = 1.39; 95% CI, 1.22–1.57), and the lowest risk of cancer of the larynx (HRR = 0.19; 95% CI, 0.14–0.27) and ovary (HRR = 0.17; 95% CI, 0.11–0.28). Finally, the FMC$_5$ cluster was the only to capture strong coaggregation with ovarian cancer (HRR = 6.10; 95% CI, 5.64–6.61).

Discussion

Large multisource database resources are being developed in several healthcare systems across the United States and country-wide initiatives are becoming more common across the world (42–44). Each of these immense resources has its particular strength and together these resources hold the potential for paradigm-shifting opportunities in Population Science research. However, these will only be realized with consummate advances in computational approaches to interrogate the data. In Utah, a strength of the UPDB is an immense genealogy linked to a statewide health data. Here, we have described two different novel approaches that focus on high-risk pedigrees to understand and address etiologic heterogeneity and define pleiotropic patterns. Both rely on the UPDB to provide the necessary linked databases of genealogy, cancer data, demographic, and medical/clinical information. These data are available on nearly the entire population of Utah starting with the original European settlers of Utah in the 1800s (the earliest records) and extending to current residents of the state (where all sources of records are represented). The UPDB is a dynamic resource that continues to expand as the population grows and as linked data sources develop. For example, a recent SEER-funded pilot project by the UCR illustrated a 73.6% success rate for identifying FFPE

tumor blocks for breast cancers diagnosed from 2000 to 2015 across the state. Such streamlining of tumor acquisition by the UCR would further benefit UPDB studies.

The techniques and findings here rely on a large multisource population database and cannot easily be replicated. However, the Statistics Sweden Multigeneration Register, which has been used extensively to identify familial associations between concordant and discordant cancers (45, 46), is one of the potential data source that can be used to test the reproducibility of our findings. Notably, previous genetic discoveries using UPDB have proven generalizable, such as for breast cancer (*BRCA1/BRCA2*), neurofibromatosis type I (*NF1*), familial adenomatous polyposis coli (*APC*), and melanoma (*CDKN2A*). Once other large databases become ready, the methods described here may enable and accelerate the path to discovery elsewhere. Conversely, our methods also have the potential to be broadened, for example, to explore genetic pleiotropy through multiple primaries (22, 47).

In Approach 1, we highlighted a strategy for reducing heterogeneity, and utilized a novel tumor expression phenotype, PC3, previously shown to be increased in high-risk pedigrees in the UPDB (20). We performed gene mapping in a large high-risk pedigree that contained an unusual number of breast cancer cases whose tumors were extreme for PC3. Using SGS, a method specifically designed for identifying segregating haplotypes in very large families (32, 34), we identified a 0.6 Mb genome-wide significant segment in pedigree 1822 at 2q13 ($P = 1.6 \times 10^{-8}$, LOD equivalent 6.64). A *post hoc* search for additional carriers (not restricted to those with tumor data) indicates the region may only be 120 kb. Only 4 genes are contained in the smaller region, and of particular interest are *IL1A* and *IL1B*. ILs are key regulators of

Table 1. The HRR of cancer diagnosis by cancer site. Results are displayed for all familial breast cancer families ($N = 5,045$) combined and by FMC.

	Overall HR (95% CI)	FMC_1 HR (95% CI)	FMC_2 HR (95% CI)	FMC_3 HR (95% CI)	FMC_4 HR (95% CI)	FMC_5 HR (95% CI)
Breast	3.64 (3.57–3.70)	3.05 (2.98–3.12)	4.32 (4.14–4.50)	3.79 (3.64–3.94)	6.16 (5.96–6.37)	3.24 (3.12–3.37)
Ovary	1.17 (1.09–1.26)	0.19 (0.15–0.24)	0.61 (0.46–0.82)	0.72 (0.57–0.92)	0.17 (0.11–0.28)	6.10 (5.64–6.61)
Larynx	0.99 (0.93–1.05)	0.41 (0.37–0.46)	0.69 (0.56–0.85)	4.93 (4.58–5.31)	0.19 (0.14–0.27)	0.75 (0.64–0.88)
Melanoma of the skin	1.09 (1.05–1.13)	0.76 (0.72–0.80)	4.17 (3.95–4.40)	0.83 (0.75–0.92)	0.59 (0.53–0.67)	0.94 (0.86–1.02)
Prostate	1.09 (1.06–1.11)	1.07 (1.03–1.10)	1.08 (1.01–1.16)	1.05 (0.99–1.12)	1.20 (1.13–1.27)	1.11 (1.05–1.17)
Acute myeloid leukemia	1.00 (0.91–1.10)	0.87 (0.77–0.99)	1.30 (1.02–1.65)	0.97 (0.75–1.24)	1.34 (1.08–1.65)	1.00 (0.80–1.24)
Acute lymphocytic leukemia	1.06 (0.97–1.15)	1.14 (1.02–1.27)	1.57 (1.28–1.93)	0.85 (0.66–1.09)	0.92 (0.73–1.17)	0.74 (0.58–0.94)
Hodgkin—nodal	1.04 (0.91–1.19)	0.95 (0.79–1.14)	1.20 (0.83–1.72)	1.53 (1.15–2.04)	1.14 (0.83–1.57)	0.75 (0.51–1.08)
NHL—nodal	1.06 (1.00–1.11)	1.05 (0.98–1.12)	1.20 (1.05–1.39)	0.91 (0.79–1.05)	1.20 (1.06–1.36)	0.98 (0.87–1.11)
Colon	1.00 (0.97–1.03)	0.96 (0.91–1.00)	1.06 (0.96–1.16)	1.02 (0.93–1.11)	1.10 (1.01–1.20)	1.03 (0.95–1.11)
Thyroid	1.01 (0.95–1.07)	1.03 (0.95–1.12)	1.23 (1.04–1.45)	0.78 (0.65–0.94)	0.93 (0.79–1.10)	1.04 (0.89–1.20)
Cervical	0.80 (0.74–0.86)	0.76 (0.69–0.84)	0.98 (0.80–1.19)	1.02 (0.86–1.21)	0.84 (0.70–1.01)	0.60 (0.49–0.74)
Uterine	1.11 (1.05–1.17)	1.05 (0.97–1.12)	1.08 (0.93–1.27)	1.17 (1.02–1.34)	1.39 (1.22–1.57)	1.06 (0.93–1.20)
Lung and bronchus	0.84 (0.80–0.88)	0.77 (0.72–0.82)	0.92 (0.81–1.05)	1.06 (0.95–1.18)	0.94 (0.84–1.06)	0.77 (0.68–0.86)
Stomach	0.92 (0.84–1.01)	0.87 (0.76–0.98)	1.25 (0.99–1.58)	0.92 (0.73–1.17)	0.98 (0.78–1.24)	0.87 (0.70–1.08)
Soft tissue including heart	1.02 (0.90–1.15)	1.03 (0.87–1.21)	1.31 (0.94–1.81)	1.03 (0.75–1.43)	1.14 (0.84–1.55)	0.69 (0.48–0.98)
Kidney and renal pelvis	0.89 (0.83–0.96)	0.83 (0.75–0.91)	0.99 (0.80–1.22)	1.00 (0.83–1.20)	0.87 (0.72–1.06)	0.97 (0.82–1.15)
Testis	1.05 (0.92–1.19)	1.00 (0.84–1.20)	1.55 (1.13–2.12)	1.04 (0.74–1.47)	1.09 (0.79–1.50)	0.80 (0.56–1.15)
Pancreas	1.04 (0.97–1.11)	0.98 (0.89–1.07)	1.12 (0.93–1.36)	1.06 (0.89–1.26)	1.24 (1.05–1.46)	1.03 (0.88–1.21)
Esophagus	0.88 (0.77–1.01)	0.74 (0.61–0.90)	0.79 (0.51–1.21)	1.16 (0.85–1.60)	1.31 (0.98–1.76)	0.84 (0.60–1.16)
Liver	0.83 (0.71–0.96)	0.68 (0.55–0.85)	0.58 (0.34–1.01)	0.89 (0.60–1.32)	1.14 (0.81–1.61)	1.17 (0.85–1.59)
Brain	0.98 (0.90–1.06)	0.90 (0.80–1.02)	1.13 (0.89–1.43)	1.01 (0.81–1.25)	1.03 (0.83–1.27)	1.05 (0.86–1.27)
CNS	0.94 (0.86–1.03)	0.89 (0.79–1.01)	0.92 (0.70–1.21)	1.05 (0.84–1.32)	0.99 (0.79–1.25)	1.00 (0.81–1.24)
Myeloma	1.03 (0.95–1.13)	0.98 (0.87–1.01)	1.26 (0.99–1.60)	1.08 (0.86–1.36)	1.07 (0.85–1.34)	1.02 (0.83–1.26)
Small intestine	1.01 (0.87–1.17)	0.97 (0.79–1.19)	1.22 (0.82–1.83)	0.97 (0.65–1.45)	0.95 (0.63–1.42)	1.07 (0.75–1.51)
Urinary bladder	0.99 (0.94–1.04)	0.96 (0.90–1.03)	1.04 (0.90–1.22)	1.02 (0.89–1.17)	1.05 (0.92–1.20)	0.94 (0.83–1.07)

Note: The overall estimates and 95% CIs are displayed in column 2. The FMC configuration (FMC_{1-5})–specific HRRs are reported in columns 3–7.
Abbreviations: CNS, cranial nerves, other nervous system; NHL, non-Hodgkin lymphoma.

inflammation and immune response with roles in cell growth, angiogenesis, and regulation of inflammatory process, and therefore strong candidate genes for breast cancer risk and mortality. In case–control studies, *IL1B* SNPs have been associated with breast cancer risk (48, 49). *IL1B* has also been studied as a candidate for metastatic progression, particularly with respect to invasiveness and the epithelial–mesenchymal transition (50–56), as well as resistance to therapy (57). *IL1A* has been shown to play a role in chronic inflammation driving tumorigenesis and chemotherapy resistance (58). With these compelling candidates, the natural next step will be to sequence the shared haplotype for functional variants.

In Approach 2, we highlighted the ability to identify pleiotropies and described five FMCs for high-risk breast cancer families. This novel, network-inspired approach simultaneously considered risk of multiple cancer types to classify families into clusters with similar patterns of familial cancer risk. Several cancer types that have previously been shown to coaggregate with breast cancer were identified in the signatures of our agnostic clustering approach (prostate, ovary, uterine, and melanoma; refs. 59–61). However, we show that these risks may vary widely across clusters (ovarian and melanoma, in particular). New coaggregations were also identified. Notably, risk for larynx cancer (FMC_3 HRR = 4.93) and lymphomas (FMC_3 Hodgkin HRR = 1.53 and FMC_2 ALL HRR = 1.57). These findings improve resolution and our understanding of cancer family risks and have potential implications for screening and prevention. Also, while it is common for familial studies to focus only on increased risk, we also considered cancers with decreased risk.

Isolating patterns of extreme decrease in risk, such as the multiple cancers at decreased risk in FMC_1, could aid in the discovery of etiologic factors that have opposing pleiotropic effects (i.e., a genetic mutation that increases risk for one cancer but is protective for others) or are single cause–single phenotype relationships. Another interesting pattern that may provide avenues to better understand etiology was identified in FMC_2, which showed increased risk for several cancers often seen in adolescent and young adults. Other studies have shown similar clustering patterns: Hodgkin lymphoma and other lymphoid neoplasms; (10, 62–64); testicular and non-Hodgkin lymphoma (65); and testicular, breast, and melanoma (66). Our multi-cancer signatures of risk have the potential to improve characterization of different subtypes of breast cancer and provide new avenues to explore common etiologic pathways including gene–environment factors. Subtypes provide the potential to reduce heterogeneity and increase power. The method could also be extended to noncancer phenotypes that may have an underlying genetic link to cancer, such as Parkinson disease (60). Cancer is a complex phenotype and by embracing large multisource databases and computational tools, such as machine learning, it will be possible to seek out important combinations, beyond individual factors, to further our knowledge of the disease.

The goal of both approaches was to increase homogeneity to improve genetic studies, the first by defining cases within a pedigree that are similar and second by selecting groups of pedigrees that are similar (and indicative of genetics, rather than environment). It is important to note that findings from both approaches are sensitive

to parameters of the methods. In Approach 1, the phenotype used to select cases is critical to power (extreme-PC3, previously shown to cluster in pedigrees). Without restriction, there is no signal at 2q13, or elsewhere in the genome. We note that sharing in the eight cases in 1822 ($P = 1.6 \times 10^{-8}$) compares in significance with the best single-*BRCA1* pedigree published (equivalent $P = 6.2 \times 10^{-8}$; ref. 67) or best *BRCA2* pedigree ($P = 1.8 \times 10^{-5}$; ref. 2). In Approach 2, as with all clustering techniques, the clusters are sensitive to the distance metrics and weighing scheme used. This is important to consider when interpreting findings. To improve authenticity and generalizability and reduce spurious patterns, these parameters can be grounded with domain-specific knowledge or logical theories.

Large, population-based, multi-faceted databases, such as the UPDB, represent a new era for Population Sciences. Together with novel approaches, such as we have described here, these will play a critical role in advancing knowledge of cancer risk, elucidating the interplay between factors at the molecular level to individual interactions with the environment, and determine how these factors vary between people. Datasets that link family structure will also allow for important questions about the transgenerational nature of disease. We have illustrated that tumor phenotypes identified using high-risk status can map genes for breast cancer, and that various different cancer pleiotropies exist in high-risk breast cancer pedigrees. These types of discoveries will offer new avenues for defining germline susceptibilities, cancer prevention, and multi-cancer risk management.

Disclosure of Potential Conflicts of Interest

P.S. Bernard has ownership interest (including patents) in Bioclassifier LLC. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: H.A. Hanson, C.L. Leiser, N.J. Camp
Development of methodology: H.A. Hanson, C.L. Leiser, M.J. Madsen, J. Gardner, S. Knight, N.J. Camp
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): H.A. Hanson, S. Knight, M. Cessna, C. Sweeney, K.R. Smith, P.S. Bernard
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): H.A. Hanson, C.L. Leiser, M.J. Madsen, J. Gardner, K.R. Smith, P.S. Bernard, N.J. Camp
Writing, review, and/or revision of the manuscript: H.A. Hanson, C.L. Leiser, S. Knight, M. Cessna, C. Sweeney, J.A. Doherty, K.R. Smith, P.S. Bernard, N.J. Camp
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): H.A. Hanson, P.S. Bernard
Study supervision: H.A. Hanson, K.R. Smith, N.J. Camp

References

1. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science 1990;250: 1684–9.
2. Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science 1994;265:2088–90.
3. Wood GC, Chu X, Manney C, Strodel W, Petrick A, Gabrielsen J, et al. An electronic health record-enabled obesity database. BMC Med Inform Decis Mak 2012;12:45.
4. Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE–An integrated standards-based translational research informatics platform. AMIA Annu Symp Proc 2009;2009:391–5.
5. Mullins IM, Siadaty MS, Lyman J, Scully K, Garrett CT, Miller WG, et al. Data mining and clinical data repositories: Insights from a 667,000 patient data set. Comput Biol Med 2006;36:1351–77.
6. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Am J Hum Genet 2010;86:560–72.
7. Brüggenjürgen B, Burkowitz J, Willich SN. Utilisation of medical resources of patients with pain undergoing an outpatient opioid therapy. Gesundheitswesen 2007;69:353–8.
8. Webster PC. Sweden's health data goldmine. CMAJ 2014;186:E310.
9. Collins R. What makes UK Biobank special? Lancet 2012;379:1173–4.
10. Kerber RA, O'Brien E. A cohort study of cancer risk in relation to family histories of cancer in the Utah Population Database. Cancer 2005;103:1906–15.
11. Martin C, Leiser CL, O'Neil B, Gupta S, Lowrance WT, Kohlmann W, et al. Familial cancer clustering in urothelial cancer: a population-based case–control study. J Nat Cancer Inst 2017;110:527–33.
12. Hanson HA, Horn KP, Rasmussen KM, Hoffman JM, Smith KR. Is cancer protective for subsequent Alzheimer's disease risk? Evidence from the Utah Population Database. J Gerontol B Psychol Sci Soc Sci 2016;72:1032–43.
13. Soisson S, Ganz PA, Gaffney D, Rowe K, Snyder J, Wan Y, et al. Long-term, adverse genitourinary outcomes among endometrial cancer survivors in a large, population-based cohort study. Gynecol Oncol 2018;148:499–506.
14. Ou JY, Hanson HA, Ramsay JM, Leiser CL, Zhang Y, VanDerslice JA, et al. Fine particulate matter and respiratory healthcare encounters among survivors of childhood cancers. Int J Environ Res Public Health 2019;16. pii: E1081.
15. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 1994;266:66–71.
16. Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, et al. The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. Nat Genet 1996;12:333–7.
17. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.
18. Chia SK, Bramwell VH, Tu D, Shepherd LE, Jiang S, Vickery T, et al. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. Clin Cancer Res 2012;18:4465–72.
19. Gorski JJ, James CR, Quinn JE, Stewart GE, Staunton KC, Buckley NE, et al. BRCA1 transcriptionally regulates genes associated with the basal-like phenotype in breast cancer. Breast Cancer Res Treat 2010;122:721–31.
20. Madsen MJ, Knight S, Sweeney C, Factor R, Salama M, Stijleman IJ, et al. Reparameterization of PAM50 expression identifies novel breast tumor dimensions and leads to discovery of a genomewide significant breast cancer locus at 12q15. Cancer Epidemiol Biomarkers Prev 2018;27:644–52.
21. Cancer Genome Atlas Research Network, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. Integrated genomic characterization of endometrial carcinoma. Nature 2013;497:67–73.
22. Begg CB, Rice MS, Zabor EC, Tworoger SS. Examining the common aetiology of serous ovarian cancers and basal-like breast cancers using double primaries. Br J Cancer 2017;116:1088–91.
23. Hanson HA, Leiser CL, Martin C, Gupta S, Smith KR, Dechet C, et al. Redefining the bladder cancer phenotype using patterns of familial risk. medRxiv 19003681 [Preprint]. 2019. Available from: https://www.medrxiv.org/content/10.1101/19003681v1.
24. Bean LL, May DL, Skolnick M. The Mormon historical demography project. Hist Methods 1978;11:45–53.

25. Bishop DT, Skolnick MH. Genetic epidemiology of cancer in Utah genealogies: a prelude to the molecular genetics of common cancers. J Cell Physiol Suppl 1984; 3:63–77.

26. Skolnick M BL, Dintelman S, Mineau GP. A computerized family history database system. Sociol Social Res 1979;63:506–23.

27. O'Brien E, Rogers AR, Beesley J, Jorde LB. Genetic structure of the Utah Mormons: comparison of results based on RFLPs, blood groups, migration matrices, isonymy, and pedigrees. Hum Biol 1994;66:743–59.

28. Wylie JE, Mineau GP. Biomedical databases: protecting privacy and promoting research. Trends Biotechnol 2003;21:113–6.

29. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27: 1160–7.

30. Sweeney C, Bernard PS, Factor RE, Kwan ML, Habel LA, Quesenberry CP Jr, et al. Intrinsic subtypes from PAM50 gene expression assay in a population-based breast cancer cohort: differences by age, race, and tumor characteristics. Cancer Epidemiol Biomarkers Prev 2014;23:714–24.

31. Camp NJ, Madsen MJ, Herranz J, Rodríguez-Lescure Á, Ruiz A, Martín M, et al. Re-interpretation of PAM50 gene expression as quantitative tumor dimensions shows utility for clinical trials: application to prognosis and response to paclitaxel in breast cancer. Breast Cancer Res Treat 2019;175:129–39.

32. Waller RG, Darlington TM, Wei X, Madsen MJ, Thomas A, Curtin K, et al. Novel pedigree analysis implicates DNA repair and chromatin remodeling in multiple myeloma risk. PLoS Genet 2018;14:e1007111.

33. Thomas A, Camp NJ, Farnham JM, Allen-Brady K, Cannon-Albright LA. Shared genomic segment analysis. Mapping disease predisposition genes in extended pedigrees using SNP genotype assays. Ann Hum Genet 2008;72:279–87.

34. Knight S, Abo RP, Abel HJ, Neklason DW, Tuohy TM, Burt RW, et al. Shared genomic segment analysis: the power to find rare disease variants. Ann Hum Genet 2012;76:500–9.

35. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature 2015; 526:68–74.

36. Abel HJ, Thomas A. Accuracy and computational efficiency of a graphical modeling approach to linkage disequilibrium estimation. Stat Appl Genet Mol Biol 2011;10:5.

37. Thomas A. Assessment of SNP streak statistics using gene drop simulation with linkage disequilibrium. Genet Epidemiol 2010;34:119–24.

38. Matise TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, et al. A second-generation combined linkage physical map of the human genome. Genome Res 2007;17:1783–6.

39. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995;11:241–7.

40. Norris D, Stone J. WHO classification of tumours of haematopoietic and lymphoid tissues. Geneva (Switzerland): World Health Organization; 2008.

41. Kaufman L, Rousseeuw PJ. Partitioning around medoids (program PAM). In: Finding groups in data: an introduction to cluster analysis. Hoboken (NJ): John Wiley & Sons; 1990. p. 68–125.

42. Polubriaginof FCG, Vanguri R, Quinnies K, Belbin GM, Yahi A, Salmasian H, et al. Disease heritability inferred from familial relationships reported in medical records. Cell 2018;173:1692–704.

43. Machluf Y, Tal O, Navon A, Chaiter Y. From population databases to research and informed health decisions and policy. Front Public Health 2017;5:230.

44. Meyer A-M, Olshan AF, Green L, Meyer A, Wheeler SB, Basch E, et al. Big data for population-based cancer research: the integrated cancer information and surveillance system. N C Med J 2014;75:265–9.

45. Weires M, Bermejo JL, Sundquist J, Hemminki K. Clustering of concordant and discordant cancer types in Swedish couples is rare. Eur J Cancer 2011;47:98–106.

46. Frank C, Sundquist J, Yu H, Hemminki A, Hemminki K. Concordant and discordant familial cancer: familial risks, proportions and population impact. Int J Cancer 2017;140:1510–6.

47. Mauguen A, Zabor EC, Thomas NE, Berwick M, Seshan VE, Begg CB. Defining cancer subtypes with distinctive etiologic profiles: an application to the epidemiology of melanoma. J Am Stat Assoc 2017;112:54–63.

48. Peng S, Lu B, Ruan W, Zhu Y, Sheng H, Lai M. Genetic polymorphisms and breast cancer risk: evidence from meta-analyses, pooled analyses, and genome-wide association studies. Breast Cancer Res Treat 2011;127:309–324.

49. He B, Zhang Y, Pan Y, Xu Y, Gu L, Chen L, et al. Interleukin 1 beta (IL1B) promoter polymorphism and cancer risk: evidence from 47 published studies. Mutagenesis 2011;26:637–42.

50. Perez-Yepez EA, Ayala-Sumuano JT, Lezama R, Meza I. A novel beta-catenin signaling pathway activated by IL-1beta leads to the onset of epithelial-mesenchymal transition in breast cancer cells. Cancer Lett 2014;354:164–71.

51. Escobar P, Bouclier C, Serret J, Bièche I, Brigitte M, Caicedo A, et al. IL-1beta produced by aggressive breast cancer cells is one of the factors that dictate their interactions with mesenchymal stem cells through chemokine production. Oncotarget 2015;6:29034–47.

52. Oh K, Lee O-Y, Park Y, Seo MW, Lee D-S. IL-1β induces IL-6 production and increases invasiveness and estrogen-independent growth in a TG2-dependent manner in human breast cancer cells. BMC Cancer 2016;16:724.

53. Voigt C, May P, Gottschlich A, Markota A, Wenk D, Gerlach I, et al. Cancer cells induce interleukin-22 production from memory CD4(+) T cells via interleukin-1 to promote tumor growth. Proc Natl Acad Sci U S A 2017;114: 12994–9.

54. Castano Z, San Juan BP, Spiegel A, Pant A, DeCristo MJ, Laszewski T, et al. IL-1beta inflammatory response driven by primary breast cancer prevents metastasis-initiating cell colonization. Nature Cell Biol 2018;20:1084–97.

55. Tulotta C, Lefley DV, Freeman K, Gregory WM, Hanby AM, Heath PR, et al. Endogenous production of IL1B by breast cancer cells drives metastasis and colonization of the bone microenvironment. Clin Cancer Res 2019;25: 2769–82.

56. Martinez-Reza I, Diaz L, Barrera D, Segovia-Mendoza M, Pedraza-Sánchez S, Soca-Chafre G, et al. Calcitriol inhibits the proliferation of triple-negative breast cancer cells through a mechanism involving the proinflammatory cytokines IL-1beta and TNF-alpha. J Immunol Res 2019;2019:6384278.

57. Mendoza-Rodriguez MG, Ayala-Sumuano JT, Garcia-Morales L, Zamudio-Meza H, Perez-Yepez EA, Meza I. IL-1β inflammatory cytokine-induced TP63 isoform NP63α signaling cascade contributes to cisplatin resistance in human breast cancer cells. Int J Mol Sci 2019;20:270.

58. Liu S, Lee JS, Jie C, Park MH, Iwakura Y, Patel Y, et al. HER2 overexpression triggers an IL1α proinflammatory circuit to drive tumorigenesis and promote chemotherapy resistance. Cancer Res 2018;78:2040–51.

59. Goggins W, Gao W, Tsao H. Association between female breast cancer and cutaneous melanoma. Int J Cancer 2004;111:792–4.

60. Olsen JH, Friis S, Frederiksen K. Malignant melanoma and other types of cancer preceding Parkinson disease. Epidemiology 2006;17:582–7.

61. Kar SP, Beesley J, Amin Al Olama A, Michailidou K, Tyrer J, Kote-Jarai Z, et al. Genome-wide meta-analyses of breast, ovarian, and prostate cancer association studies identify multiple new susceptibility loci shared by at least two cancer types. Cancer Discov 2016;6:1052–67.

62. Linabery AM, Erhardt EB, Richardson MR, Ambinder RF, Friedman DL, Glaser SL, et al. Family history of cancer and risk of pediatric and adolescent Hodgkin lymphoma: a Children's Oncology Group study. Int J Cancer 2015; 137:2163–74.

63. Crump C, Sundquist K, Sieh W, Winkleby MA, Sundquist J. Perinatal and family risk factors for Hodgkin lymphoma in childhood through young adulthood. Am J Epidemiol 2012;176:1147–58.

64. Pang D, Alston RD, Eden TO, Birch JM. Cancer risks among relatives of children with Hodgkin and non-Hodgkin lymphoma. Int J Cancer 2008;123: 1407–10.

65. Nordsborg RB, Meliker JR, Wohlfahrt J, Melbye M, Raaschou-Nielsen O. Cancer in first-degree relatives and risk of testicular cancer in Denmark. Int J Cancer 2011;129:2485–91.

66. Zhang L, Yu H, Hemminki O, Försti A, Sundquist K, Hemminki K. Familial associations in testicular cancer with other cancers. Sci Rep 2018;8:10880.

67. Goldgar DE, Cannon-Albright LA, Oliphant A, Ward JH, Linker G, Swensen J, et al. Chromosome 17q linkage studies of 18 Utah breast cancer kindreds. Am J Hum Genet 1993;52:743–8.