The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.

Mary-Claire King

The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.

“The Race” to Clone \textit{BRCA1}

Mary-Claire King

The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.

Mary-Claire King

The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.

Mary-Claire King

The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.

Mary-Claire King

The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.

Mary-Claire King

The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.

Mary-Claire King

The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.

Mary-Claire King

The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.

Mary-Claire King

The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.

Mary-Claire King

The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.

Mary-Claire King

The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.

Mary-Claire King

The existence of \textit{BRCA1} was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lay triggered efforts by public and private groups to clone and sequence it. The press baptized the competition “the race” and reported on it in detail for the next 4 years. \textit{BRCA1} was positionally cloned in September 1994. Twenty years later, I reflect on “the race” and its consequences for breast cancer prevention and treatment.
The 173rd marker that we tested was D17S74, a highly polymorphic VNTR on chromosome 17q21. Linkage of breast cancer to this marker, using our model-based linkage parameters, yielded odds of $10^6$ to 1 in favor of linkage for the seven families in our series with an average age of breast cancer onset $\leq 45$ years and evidence against linkage for the 16 families with average age of breast cancer onset $>45$ years. We published the results in 1990 (3).

Through the 1980s, with the help of oncologists, their patients, and the patients’ families, we enrolled 23 extended kindreds severely affected with breast cancer. No detailed human genetic map yet existed, so gene hunters worldwide developed the genetic map collaboratively, in parallel with linkage analysis for each group’s own projects. The first linkage markers were protein polymorphisms (10), then restriction fragment length polymorphisms (11), then variable number of tandem repeat (VNTR) markers (12), and—after the discovery of the polymerase chain reaction (13)—short sequence repeat markers (14).

The development of the human genetic map had organizational homes at the Centre d’Etude du Polymorphisme Humain in Paris (15), led by Jean Dausset, Jean Weissenbach, Jean Marc Lalouel, and Mark Leppert; and with the beginning of the Human Genome Project, at the National Institutes of Health (NIH), led by Jim Watson. The human gene mapping period was characterized by open collaboration, in parallel: continue genetic mapping with new markers as quickly as we learned of them or could develop them ourselves. Each marker was genotyped individually, the vast majority by Southern blot, in all informative persons.

To keep the activities in our Berkeley, Ann Arbor, Dallas, and Research Triangle Park labs minimal surgical treatment for breast cancer. “You’re looking for the cause of invasive breast cancer,” he said. “Don’t get distracted.” In taking his advice, we sought families including multiple relatives with invasive ductal carcinoma of the breast. In many of these families, breast cancer was early onset and often bilateral, and occasionally also appeared in men. We did not broaden the phenotype to include the far more common atypical hyperplasia, despite its being an established risk factor for breast cancer. Focusing on a narrow disease definition limited the number of persons in each family defined as “affected” and thus reduced the sample size for linkage, but far more importantly, it eliminated false positives. Keeping our eyes on the prize proved critical to successful mapping.

A rigorous definition of the phenotype did not solve other complexities, including causal heterogeneity and dependence of breast cancer expression on gender and age. To address these complexities, my colleague Ming Lee adapted the elegant linkage methods developed by Newton Morton, Robert Elston, and Jurg Ott (9) to incorporate the parameters of our previous population-based model into our calculations of the likelihoods of linkage of breast cancer to each of our genetic markers. To take an extreme example, an unaffected male or an unaffected female in her 20s did not contribute any information against linkage, whereas an affected male or an affected female in her 20s was highly informative.

In our 23 extended families, Jeff Hall and I genotyped new markers as quickly as we learned of them or could develop them ourselves. Each marker was genotyped individually, the vast majority by Southern blot, in all informative persons. The mainstream estimate of the number of human genes was 100,000, based on inaccurate estimates of gene size and density. Very few genes were characterized and even fewer mapped. Sequencing was done by hand. There was no e-mail or Internet. The revolutionary advance in data sharing was the fax, with curly paper that found comfort hiding beneath your desk.

Following successful mapping, the process to gene discovery was positional cloning, which was experimentally challenging but fun. For BRCA1 (which in 1991 I was allowed to name, because its existence was now widely accepted), we knew the chromosomal locale, defined by linkage and bounded by recombination at genetic markers in informative members of our families. The corresponding physical region was completely unknown. My group at Berkeley developed a collaboration with Francis Collins, then at the University of Michigan; with Anne Bowcock, then at the University of Texas Southwestern Medical Center; and later with Jeff Boyd, then at the National Institute of Environmental Health Sciences. Our joint strategy was to carry out four activities in parallel: continue genetic mapping with new markers to narrow the linkage region; generate a complete physical map of the linked region via a path of overlapping DNA-containing clones; hybridize the clones representing the region to a cDNA library of genes expressed in our tissue of interest; and isolate and sequence these revealed genes from the cDNA library, then sequence each gene in DNA from affected members of our families to discover mutations disabling the gene.

Twenty years later, with the complete human genome sequence, it is now clear that the physical size of the region linked to BRCA1 in our families was 22 Mb in 1990, 4.5 Mb by late 1992, and 1.0 Mb by early 1994, when all groups had resolved the critical recombination events in all families (16–18). Given the difficulty of physically mapping and cloning, the most productive period, by far, was after the region was reduced to a manageable size by linkage. Constructing the physical map was a tremendous challenge. At the time, DNA had been cloned into 40-kb cosmids by NIH and U.S. Department of Energy labs, and into much larger yeast artificial chromosomes (YACs) by Maynard Olson’s lab (19). From their previous gene hunts, Francis Collins’ lab was adept with chromosome walking with cosmids, but it was soon clear that the 17q21 region had many messy features that made these walks more difficult than previous ones. A path that began nicely enough with two or three adjacent cosmids would soon turn back on itself, yielding more a meander through a swamp than a walk from one signpost to another.

"A path that began nicely enough with two or three adjacent cosmids would soon turn back on itself, yielding more a meander through a swamp than a walk from one signpost to another."

**Positional Cloning**

The Human Genome project was also born in 1990, so “the race” to find “the breast cancer gene,” began with no genome sequence, no integrated physical maps, no awareness of genomic architecture, and certainly no genome browser. In many bizarre architectural features. Odd genomic architecture predisposes to errors at mitosis and therefore to the somatic mutations that are the second hits of tumor suppressor genes.

For physical cloning, YACs were a cause for celebration because they captured much more DNA than did cosmids, so each tile (clone) was larger and the number of tiles needed to span a region was smaller. The corresponding challenges of YACs were that the longer the cloned insert, the more likely it was to be chimeric; that is, to include pieces of multiple chromosomes and to have internal deletion of elements of the chromosome of interest. The secret was not to be greedy, to work with YACs that were 100 to 200 kb—so longer than cosmids, but not grandiose. Bacterial artificial chromosomes (BACs) (20) had been developed at about the same time as YACs, but were not reduced to practice for human gene mapping for several years. BACs were intermediate in size between cosmids and YACs, so they are more stable than YACs and more informative than cosmids. BACs became a critical backbone of the Human Genome Project and, indeed, the Myriad group identified BRCA1 from a BAC (1).
straight, every day I wrote an “Order of the Day” (OOTD) that included progress with every cosmid, YAC, probe, library, genetic marker, and family. The OOTD was sent around by fax, then various small groups of people would chat by phone to plan the next experiments. This continued, if not 24/7, at least 14/6. The OOTD and the organization connected to it became even more important after Francis Collins moved from Michigan to the NIH in 1993 to begin the Human Genome project from Jim Watson. With Francis’ lab now in two places, we were four groups in five towns in three time zones, keeping in contact without Internet or cell phones. It seemed normal at the time.

The collaboration painstakingly identified hundreds of cosmids and YACs that mapped to our linked region. To maximize our success in capturing complete and critical genes, each group probed our shared clones with different cDNA libraries. In my lab, Eric Lynch created a beautiful cDNA library from ovarian epithelium like a delicate blanket, so epithelial cells could be gently removed, the RNA isolated after visiting our lab, we were looking for an expressed genes. When the BRCA1 sequence was published, how close were we to finding it? In retrospect, we had a marker inside BRCA1 and did not know it. Our most informative linkage marker was AFM248yg9, in intron 20 of BRCA1, within a few kilobases of mutations in several of our families. Of course, we did not know this marker was inside BRCA1; we knew only that we could not identify a cosmid or YAC that carried it. In retrospect, cosmids carrying BRCA1 were particularly difficult to identify because most of the intronic sequence of BRCA1 is Alu. By masking Alu sequences to hybridize to the cosmid library, we masked the marker in BRCA1 as well. Our physical map of the 1-Mb region had only one gap, exactly at BRCA1 (Fig. 1). Of course, at the time, we had no idea of the size of the gap. Part of the plan was to create probes with single-copy DNA flanking each genetic marker and use them to find more clones. This would have revealed a cosmid containing BRCA1, then BRCA1 itself. All groups in “the race” used essentially the same positional cloning strategy. The limiting factors were time and resources. As Maynard Olson remarked after visiting our lab, we were looking for a pot of gold at the end of a rainbow in the backyard with a hoe, when what we needed was a backhoe.

The Gene

Even after it was cloned, BRCA1 continued to confound its pursuers. For example, there is no exon 4 in BRCA1, because in the cDNA sequence published by the Myriad group, exon 4 was an Alu sequence with stops in all reading frames (1). The faux exon was soon gone from the sequence, but exon numbering did not change. More substantially, the genomic structure of BRCA1 is distinct, with a long central exon that encodes ~60% of the protein and is remarkably tolerant of amino acid substitutions. The 21 small flanking coding exons encode virtually all the functionally important regions of the protein. Likely for reasons of their parallel evolutionary histories, the genomic structure of BRCA2 is almost identical to that of BRCA1, despite the two genes sharing no similarity in primary sequence (23).

At the time of its discovery, the biological function of BRCA1 was completely unknown. There were no homologous genes and no recognizable motifs other than the RING domain (24), which was an acronym for “really interesting new gene,” not a description of function. BRCA1 could only have been found by a genetic approach. Because its function was unknown, it would not have been selected as a candidate gene based on a biological role. A genetic approach, however, is blissfully tolerant of total functional ignorance.

**Fig. 1. Linkage and physical mapping of the BRCA1 region as of early September 1994.** Drawings of chromosome 17 at the top of the figure indicate refinement of the BRCA1 region (black boxes) by linkage in 1990, 1992, and 1994. The final region defined by recombination in families was bounded by genetic markers (purple arrows), in retrospect known to lie at 40.705 and 41.710 Mb on 17q21.31, so yielding a linked region 1 Mb in length. This region was captured in YACs, P1 clones, and cosmid pools (CPs) and assembled in physical contigs (green bars). In this 1-Mb region, we identified 15 genes, 5 previously known but not mapped (blue bars) and 10 previously not known (red bars). BRCA1 was not captured because it lay in a 100-kb gap in the physical contig that was only fully sequenced 2 years later (37). In retrospect, the very dense packing of Alus in BRCA1 led to it being refractory to clone capture. Thanks to the completion of the Human Genome Project, no gene hunter ever need face this problem again.
Twenty years after its discovery, the biological roles and evolutionary origins of BRCA1 are still being elucidated. Genetics revealed that BRCA1 is a tumor suppressor gene following the two-hit model (25). Cancer develops as the result of one inherited loss-of-function mutation followed by a somatic mutation causing loss of the remaining wild-type allele in a vulnerable cell type. The central puzzle is why complete loss of function of BRCA1 leads to cancer. Solving this puzzle has been especially challenging, because the BRCA1 protein is involved in multiple essential biological functions (26).

As part of a multiprotein complex, BRCA1 repairs double-strand DNA breaks via the homologous recombination repair pathway. The C-terminal BRCT domain interacts with histone deacetylases and is involved in transcriptional regulation. The N-terminal RING domain heterodimerizes with a sister domain of BARD1 and acts as a ubiquitin ligase of the estrogen receptor (27). Missense mutations that abrogate the function of the RING domain lead to breast cancer. Virtually all other cancer-causing mutations of BRCA1 are truncations: nonsense mutations, frameshifts, or large genomic deletions or duplications leading to stops and loss of the C-terminal domain.

BRCA1 is ubiquitously expressed, so it has been a mystery why BRCA1 mutations lead specifically to breast and ovarian cancer and, to a lesser degree, to pancreatic and prostate cancer. The estrogen receptor is a substrate of the ubiquitin ligase activity of the BRCA1 RING domain, and missense mutations in critical residues of this domain lead to breast cancer predisposition (23). Very recent work indicates that estrogen controls the survival of BRCA1-deficient cells via a PI3K/NRF2-regulated pathway (28). BRCA1 has revealed other breast cancer genes by virtue of the functional relationships of their encoded proteins. In particular, other genes critical for DNA repair—including TP53, PALB2, CHEK2, BARD1, BRIP1, ATM, RAD51C, and RAD51D—harbor mutations leading to inherited breast and ovarian cancer (29). Thousands of different disease-causing mutations have been detected in BRCA1 and BRCA2. Each loss-of-function mutation is individually rare, and each independently confers very high risk for breast and ovarian cancer. The other breast and ovarian cancer genes also harbor many different rare, recent damaging mutations with effect sizes ranging from twofold increased risk for CHEK2 to 10-fold for TP53.

Of the seven families in our 1990 linkage analysis with young-onset breast cancer (3), six families harbor mutations in BRCA1, and one harbors a mutation in BRCA2. Of the 16 families in that analysis that we predicted would not carry mutations in BRCA1, six are explained by BRCA2; one each is explained by PALB2, CDH1, and SLX4; and seven remain unsolved. There are more breast cancer genes to be found.

Genetic heterogeneity of inherited predisposition to breast cancer serves as a model for other complex illnesses. The disorder results from any one of thousands of different mutations in any one of multiple genes. The critical genes known thus far encode proteins in the same and related pathways. The discovery of BRCA1 and its sister genes illustrates that the degree of biological complexity underlying a phenotype is an excellent predictor of its genetic heterogeneity (30).

Our Genomes, Ourselves

In June 2013, the U.S. Supreme Court ruled unanimously that genes are products of nature and therefore cannot be patented (31), nullifying the Myriad patents on BRCA1 and BRCA2. The ruling was a victory for science and for patients and led immediately to broader availability of clinical genetic testing.

For nearly 20 years, while Myriad was the only commercial source in the United States for genetic testing of BRCA1 and BRCA2, cost was a major deterrent to widespread screening. The cost to women of BRCA1 and BRCA2 testing is now dropping, due both to the end of the monop-oly and to two scientific developments that have changed the landscape. First, there are now enough genes identified with mutations predisposing to breast and ovarian cancer that multigene screening panels can be developed and effectively implemented. Second, genomic technology now offers the opportunity to sequence at costs orders of magnitude lower than the cost of Sanger sequencing (32). Previously, clinical genetic testing was carried out gene by gene, based on specific clinical indications and family histories, with each test costing thousands of dollars. With the advent of massively parallel sequencing, large panels of genes are now screened simultaneously at far lower cost (33).

There was another barrier to genetic testing for inherited breast and ovarian cancer. Some patients and physicians worried that a positive finding would lead to loss of health care coverage. In consequence, mutations were not identified in some women who could have been saved by risk-reducing surgery. Clinical guidelines have been established for women harboring damaging mutations in BRCA1 and BRCA2, including increased surveillance, surgical removal of ovaries and fallopian tubes (salpingo-oophorectomy) by age 40 years or younger, and the possibility of risk-reducing mastectomy (34, 35). The Genetic Information Nondiscrimination Act of 2008 (Pub-lish Law 110–233), which protects mutation carriers against loss of health care coverage, should have removed fear as a barrier to testing, so that women with mutations in BRCA1 and BRCA2 can be identified without economic reprisal.

So what next? Given that 50% of BRCA1 and BRCA2 mutations are inherited from unaffected fathers, and given the small size of modern families, almost 50% of women with BRCA1 and BRCA2 mutations have little or no family history of breast or ovarian cancer. Yet, cancer risks to mutation carriers with no cancer family history are as high as risks to mutation carriers from severely affected families (36). Identification of cancer-causing mutations in BRCA1 and BRCA2 has clear and actionable implications for prevention. BRCA1 and BRCA2 screening as part of routine health care for young adult women is sensible and feasible. As in any population-screening program, genetic or otherwise, few participants will prove positive, but for women who learn that they carry mutations in BRCA1 or BRCA2, the consequences are enormous, addressable, and life-saving.

Until there are no more breast or ovarian cancers among women with BRCA1 or BRCA2 mutations, the real race is not over.