Published in final edited form as:

Cancer Causes Control. 2018 March ; 29(3): 305–314. doi:10.1007/s10552-018-1006-3.

Feasibility of analyzing DNA copy number variation in breast cancer tumor specimens from 1950–2010: how old is too old?

Nancy Krieger, PhD,

Dept of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, 677
Huntington Avenue, Kresge 717, Boston, MA 02130

Sheida Nabavi, PhD, MMSc,

Dept of Computer Science and Engineering, University of Connecticut, 371 Fairfield Way, Storrs,
CT 06269

Pamela D. Waterman, MPH,

Dept of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, 677
Huntington Avenue, Boston, MA 02130

Ninah S. Achacoso, MS,

Division of Research, Kaiser Permanente Northern California, 2000 Broadway, Oakland, CA
94612

Luana Acton, BS,

retired: *affiliation at time of participation in the study*: Division of Research, Kaiser Permanente
Northern California, 2000 Broadway, Oakland, CA 94612

Stuart J. Schnitt, MD, and

Chief of Breast Oncologic Pathology, Dana-Farber/Brigham and Women's Cancer Center;
Associate Director, Dana-Farber Cancer Institute/Brigham and Women's Hospital Breast
Oncology Program, Department of Pathology, Brigham and Women's Hospital, 75 Francis Street,
Boston, MA, 02115; *affiliation at the time of conducting this study*: Department of Pathology, Beth
Israel Deaconess Medical Center and Harvard Medical School, 330 Brookline Avenue, Boston,
MA 02215

CORRESPONDING AUTHOR: Nancy Krieger, PhD, Dept of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Kresge 717, Boston, MA 02130, ORCID ID: 0000-0002-4815-5947, office: 617-432-1571; fax: 617-432-3123; nkrieger@hsph.harvard.edu.

AUTHOR CONTRIBUTIONS

NK led design of the study, arranged for the assays, oversaw the analyses, and led manuscript preparation. SN interpreted the assay results and contributed to text on the methods and results. LH oversaw the initial selection of cases and retrieval of the tumor specimens, and assisted in all logistics pertaining to handling of the specimens at KPNC, as carried out by LA and NA. PDW assisted with all logistics pertaining to the handling of the specimens and with arranging the assays, once the specimens were at Harvard. SJS oversaw the histological characterization of the tumor specimens and the extraction of tumor DNA for the study assay. All authors contributed to manuscript preparation and reviewed and approved the final manuscript prior to submission.

ETHICAL STANDARDS

CONFLICT OF INTEREST: The authors declare they have no conflicts of interest to declare.

HUMAN SUBJECTS AND INFORMED CONSENT: This study was approved by the Institutional Review Boards (IRBs) of the Harvard T.H. Chan School of Public Health (#CR-20929-02) and Kaiser Permanente Northern California (#CN-13Labe-O3-H). All procedures involving human subjects were in accordance with the ethical standards of these IRBs and the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For the type of retrospective study conducted, formal consent was not required for inclusion in this study.

Laurel A. Habel, PhD

Division of Research, Kaiser Permanente Northern California, 2000 Broadway, Oakland, CA 94612

Abstract

Purpose—To assess the feasibility of quantifying long-term trends in breast tumor DNA copy number variation (CNV) profiles.

Methods—We evaluated CNV profiles in formalin-fixed paraffin-embedded (FFPE) tumor specimens from 30 randomly selected Kaiser Permanente Northern California health plan women members diagnosed with breast cancer from 1950–2010. Assays were conducted for five cases per decade who had available tumor blocks and pathology reports.

Results—As compared to the tumors from the 1970s to 2000s, the older tumors dating back to the 1950s and 1960s were much more likely to: (1) fail quality control, and (2) have fewer CNV events (average: 23 and 31 vs. 58 to 69), fewer CNV genes (average: 5.1k and 3.7k vs 8.1k to 10.3k), shorter CNV length (average: 2,440k and 3,300k vs. 5,740k to 9,280k), fewer high frequency Del genes (37% and 25% vs. 54% to 76%) and fewer high frequency high_Amp genes (20% vs. 56% to 73%). On average, assay interpretation took an extra 60 minutes/specimen for cases from the 1960s vs. 20 minutes/specimen for the most recent tumors.

Conclusions—Assays conducted in the mid-2010s for CNVs may be feasible for FFPE tumor specimens dating back to the 1980s, but less feasible for older specimens.

Keywords

archival specimen; breast cancer; biomarker; DNA copy number variation; epidemiology; historical trend

During the past decade, new assays for DNA copy number variations (CNVs) have yielded novel information critical both for guiding disease treatment and understanding etiology, given that evidence indicates some CNVs may be caused by environmental exposures [1–3]. In the case of breast cancer, research indicates both histologic type and molecular phenotype are associated with genetic alterations – variously involving amplification, translocation, and deletion – of several genes, including *MYC*, *CCND1*, *HER2*, *TOP2A*, *PIK3CA*, *PRNC1*, *DBC1*, *DEC1*, *TSPAN1*, *EGFR*, *ESR1*, and *EMSY1* [4–9]. These breast cancer CNVs have been shown to be prognostic for survival, recurrence, and metastasis, and predictive of response to treatment [4–6], independent of estrogen receptor status [4,5] and BRCA1 and BRCA2 status [6].

From an epidemiologic perspective, it is intriguing to consider evaluation of CNVs in breast cancer tissue over time, as they may result from important time-period dependent etiologic exogenous exposures. If current population distributions of CNVs differ from those observed in prior decades this may help inform investigations into specific etiologic exogenous exposures. [10–12]. Attesting to the value of re-analyzing older specimens, a study of Norwegian women born between 1886–1977 compared breast cancer incidence rates by molecular subtypes, using re-analyzed specimens, for women born between 1886–1928 and

1929–1977, and observed evidence of a secular increase, among women ages 50 to 69 at diagnosis, solely for the Luminal A and Luminal B (HER2-) subtypes [13].

Limited data on time-trends in tumor CNVs, however, exist. One reason is that assays suitable for use in large-scale population based studies have only recently been developed [1–3]; a second is the paucity of data about whether such assays can be used on archival formalin-fixed paraffin-embedded (FFPE) tumor specimens [11,12]. Only one published investigation has, to our knowledge, investigated the use of CNV assays on FFPE specimens older than 15 years [14]; this study used the leading CNV assay (OncoScan® CNV FFPE Assay [15]) and reported successfully conducting assays on FFPE specimens up to 28 years old, equivalent to the mid-1980s [14]. One other study, performed with specimens dating 12 years back, found equivalent DNA expression based on fresh frozen vs. FFPE tumor specimens [16]. To the best of our knowledge, no studies have assessed use of a CNV assay on specimens that pre-date the mid-1980s.

In this investigation, we accordingly built on our prior study to locate decades-old population-based archival FFPE tumor specimens, spanning from 1950–2010 [11]. This prior work was designed to assess the feasibility of retrieving old FFPE specimens and analyzing their tumor profiles using immunostains for clinically relevant biomarkers; key findings were the high rate of successful retrieval of linked pathology reports and specimens (83% of 60 randomly selected breast cancer cases) and the high quality and reliability of the assays [11]. In the current study, we sought to assess both the quality of results and time required to use a CNV assay on these previously retrieved breast cancer tumor specimens. Our *a priori* hypothesis was that the CNV assay would be equally valid to use, given the well-known stability of DNA preserved even in harsh environments, e.g., forensic and paleoarcheological specimens [17].

METHODS

As previously described [11], we analyzed FFPE specimens obtained from randomly selected women diagnosed with invasive breast cancer who were members of Kaiser Permanente, Northern California (KPNC) (IRB approval: Harvard School of Public Health, # CR-20929-02; KPNC, # CN-13LHabe-03-H). Established as an integrated health care delivery system in the 1940s for workers employed in World War II shipyards [18], KPNC's membership has, since inception, mirrored the economic and racial/ethnic diversity of San Francisco and California's Central Valley [19]. Its cancer registry extends back to cases diagnosed in 1947 [20].

Between 1947 and 2009, 60,904 breast cancer cases were diagnosed and included in the KPNC cancer registry, of which 7,150 met our feasibility study's eligibility criteria: women who were 50–64 years old at diagnosis and had an invasive tumor ≥ 1 cm. We randomly selected 10 eligible cases per each of 6 time periods (hereafter referred to as decades: 1947–1959; 1960–1969; ...; 2000–2009), and used information available as of 1987 to restrict sampling to cases with lymph node positive tumors. Among the 60 randomly selected cases (10 per decade), 50 cases had locatable eligible blocks containing tumor tissue. From these 50 cases we selected a random sample of five cases per decade (total $n = 30$) for biomarker

immunohistochemical analysis [11], which comprise the same cases used for the CNV assays.

For this study, DNA was extracted from the tumor specimens at the Molecular Epidemiology Research Laboratory at the Beth Israel Deaconess Medical Center. Using a randomly assigned ID number that could be linked back to the tumor data only by the study team, we shipped the tumor DNA to Affymetrix, which employed its OncoScan® CNV FFPE Assay [15]. The Affymetrix lab was blinded to all information about the tumor, including the date of each specimen.

Briefly stated, the OncoScan® CNV FFPE Assay [15] uses molecular inversion probes (MIPs) to provide integrated whole genome copy number alteration and copy-neutral loss of heterozygosity (LOH) data for more than 335,000 markers across the genome, including approximately 900 cancer-associated genes. The assay provides copy number data in both log2 scale as well as linear copy number calls, and for quality assessment, it also provides the sample 2-point relative standard error (2p-RSE), which “takes the median of the relative standard error of a whole genome pairwise comparison and uses the median value to avoid counting abnormal breakpoints that are frequently detected in tumor specimens” [14]. The quality of results and how long it took to interpret the assay data were assessed blind to the date of each tumor specimen. Once the data analysis was completed, the data were incorporated into a table that grouped the tumors by diagnostic date, allowing for calculation of average values for the five tumors per decade.

For data analysis and visualization, we employed the OncoScan™ Nexus Express Software [21]. To investigate and compare CNV profiles of samples in different decades, we used CNV and B-allele frequency (BAF) results of all samples presented by this software [22]. In general, this software uses information from the OncoScan Consol software that obtains data from scanned arrays (stored in CEL files), normalizes the data, performs quality control (QC) checks, determines CNV segments, calculates BAF, and detects LOH.

To assess the feasibility of capturing CNV profiles for specimens from different decades, we focused on general characteristics of the assays’ outcome and CNV profiles. We considered quality control (QC) metrics as the main characteristics of the assay. We also investigated the number and length of CNV events captured by the assay, number of captured CNV genes, and the percentage of published high frequency CNV genes in cancer captured by the assay for all samples. In addition, we measured extra time required to validate reported CNV results by eye inspection of captured CNV events for each assay.

To validate a CNV segment and its copy number, especially for assays with low QC metrics, it is necessary to study together the allele frequency (BAF) data and log2 ratio data of each CNV event. For QC checks we used Median Absolute Pairwise Difference of log2 ratios (MAPD) and SNP Quality Control of Normal Diploid Markers (ndSNPQC) metrics. The MAPD metric compares the log2 ratios of each adjacent pair of probes and reflects noisiness of log2 ratios. The lower this value the greater the quality of CNV detection. The ndSNPQC metric estimates distances between each genotype (AA, AB, and BB genotype calls), and reflects the separation of the homozygous and heterozygous alleles. The higher the

ndSNPQC, the better identification of each genotype and better BAF results. We used MAPD 0.3 and ndSNPQC 26 to call high quality assays, as recommended by Affymetrix [15, 21].

To investigate the ability of capturing CNV events from specimens belonging to different decades, we compared the number and length of CNV events on average across the time blocks (decades). First, we performed eye inspection to validate the CNV segments; then we used a threshold of 0.2 to filter out non-amplified CNV segments and to call amplified segments and a threshold of -0.2 to filter out non-deleted CNV segments and to call deleted segments. To investigate the biological and functional relevance and meaningfulness of the captured CNV segments, we annotated the CNV segments to obtain CNV genes, and calculated how many of the identified CNV genes were among published high frequency CNV genes. We used the data published by Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) for obtaining high frequency CNV genes. The METABRIC dataset includes CNV profiles derived from more than 2000 breast tumors. These tumors have been collected from participants of the METABRIC trial [23]. The METABRIC data are also available through cBioPortal website (http://www.cbioportal.org/study?id=brca_metabric#summary). This dataset provides the frequencies of deletion, amplification and high-amplification of all genes across all the samples in the METABRIC study. We considered genes with deletion, amplification and high-amplification frequency greater than 5% as high frequency CNV genes.

RESULTS

As previously reported [11], we were able to locate pathology reports for 55 cases (92%) among the original random sample of 60 selected cases, and 50 of these cases (83% of the 60) had blocks that contained tumor tissue. From these 50 cases, we selected a random sample of five cases per decade, yielding 30 cases, all of which displayed excellent morphology [11].

Table 1 provides key data on the tumor characteristics and the assay outcomes. All 30 cases yielded sufficient DNA (>75 ng) for analysis using the OncoScan™ assay. Among these cases, all five specimens from the 1950s failed both the MAPD QC and ndSNPQC QC tests, as did four of the five specimens from the 1960s, and two of the specimens from the 1970s. By contrast, only one out of the five specimens respectively from the most recent three decades (1980s–2000s) failed the MAPD QC test, and only one out of the five specimens from the most recent two decades (1990s–2000s) failed the ndSNPQC QC test.

With regard to the number of detected CNV segments, notably fewer CNV events after filtering were detected for the specimens from the 1950s–1960s (average number: 30.8 and 23.4, respectively); by contrast, for the specimens from the 1970s–2000s, this average number ranged between 51 and 69. The average CNV length after filtering was likewise shorter for the earlier specimens (1950s and 1960s: average CNV length of 3,296,353 and 2,438,713, respectively, versus 1970s–2000s average length ranging from 5,740,525 to 9,276,281), as was the average number of CNV genes (1950s and 1960s: average of 5,085.6 and 3,674.4, respectively, versus 1970s–2000s average number ranging from 8,058.6 to

10,283). The shorter length of CNV segments and fewer number of detected CNV segments for the earlier decades (1950s–1960s) resulted in fewer CNV genes.

The earlier specimens also were much more likely to have a lower percentage of high frequency CNV genes (Table 1). Among at least one out of the five cases per decade, the percentage of high frequency deleted genes in the 1950s and 1960s equaled 37% and 25%, respectively, versus the 1970s–2000s range of 54% to 76%. The earlier specimens likewise had a lower percentage of high frequency high_Amp genes in at least one sample out of the five samples per decade (1950s and 1960s: both 20%, versus the 1970s–2000s range of 56% to 73%). For the most recent specimens, from the 2000s, 99% of high frequency amplified CNV genes were called as amplified genes in at least one sample out of five samples.

Lastly, the average extra time for validation and evaluation of detected CNVs was greatest for the tumors from the 1950s and 1970s (61.6 and 59.6 minutes per specimen) and least for those from the 2000s (19.2 minutes per specimen). For the tumors from the 1960s, 1980s, and 1990s, the average extra time required was both intermediate and similarly long (range: 41.2 to 46.4 minutes per specimen).

DISCUSSION

Our results indicate assays conducted in the mid-2010s for CNVs may be feasible for FFPE tumor specimens dating back to the 1980s, but less feasible for older specimens. By contrast, specimens pre-dating 1980 (i.e., older than 30 years) would require more assays and more resources to yield interpretable results.

Supporting our interpretation of the study results are two key findings. First, the data presented provide novel empirical evidence that many relevant CNV genes were not being captured from the older specimens (1950s–1960s). This phenomenon may potentially be due to degradation of DNA from the older samples.

Second, the results demonstrate that the extra evaluation time required for the older specimens was directly related to the assay quality. Stated briefly, the relationship between evaluation time and assay time is as follows. For assays with high QC metric where the data are less noisy and the separation between the genotypes is clear, there is no need for inspection of the detected CNV segments, and the CNV data can be used as reported. However, for assays with marginal or low QC metric, the detected CNV segments need to be validated by eye inspection and adjustment to the software parameters needs to be applied to avoid false positives. Although this extra validation time could potentially be feasible (in terms of time and resources required) for studies with a manageable number of samples, it would be less feasible (and costlier) for large studies with thousands of samples.

The main limitation of our study is small numbers, but this limitation is offset by our ability to analyze a population-based random sample of tumors extending back five decades (1950s–2000s). Another limitation is a focus solely on breast cancer tumor specimens. However, a recent validation study of the OncoScan® CNV FFPE Assay found comparable performance for “76 FFPE tumor specimens of diverse tissue origin, including breast ($n = 28$), brain ($n = 12$), colon ($n = 12$), lung ($n = 11$), skin ($n = 4$), endometrium ($n = 2$), gastric

($n = 2$), and others ($n = 5$)"; no data were provided regarding the years in which the biopsy specimens were obtained [24]. It is also unlikely that problems with the conduct of the assay affected our study results, since the assays were performed by Affymetrix, the company that developed the OncoScan™ [15], and Affymetrix conducted the assays blinded to any information about the tumor specimens, including their date. Factors affecting specimen quality, as related to the age of the specimens, rather than the conduct of the assays, are thus the most plausible explanation for the difference in results for the older vs. more recent specimens. Whether our results can be generalized to other older specimens is also unknown, since all specimens were from a single institution, and pre-analytic variables (e.g., ischemia time, type of fixative, time of fixation, etc.) that could affect assay results potentially could vary across institutions.

Additional evidence supporting analysis of tumor specimens up to three decades old is provided by a study of breast cancer tumor FFPE specimens obtained from 1212 women diagnosed between 1985 and 2000. Among these cases, 56 specimens (4.6%) had insufficient DNA for tumor extraction, 153 (12.6%) had DNA extraction failure, and 32 (2.6%) had MIP assay failure, and among the 81.2% of specimens successfully analyzed, 95% passed the 2p-RSE threshold [5]. These findings, in conjunction with our results, support the use of a DNA CNV assay on tumor specimens dating back to the mid-1980s. Our findings, however, did not provide support for our *a priori* hypothesis that a DNA-based assay would be robust to specimen age (over a period of 50 years) and instead raise cautions for any study that seeks to analyze specimens from earlier time periods, i.e., dating back to before 1980, and hence more than 30 years old at the time of the assay.

Our results suggest two new avenues for research. First, future studies should explore whether the difficulties we encountered with analyzing tumors from cases diagnosed before the 1980s was a reflection of the time period during which the tumors were preserved, versus the actual age of the specimen at the time the assay was conducted, or both. Second, future studies could include a similar evaluation of samples from other tumor sites to determine consistency of findings across tumor types as another validation of the methods used in the work presented here. In light of the importance of tumor CNVs for treatment modality choice [4–7] and etiology [1–3,8,9], more research is warranted on methods for analyzing the unique resource of archival FFPE specimens [11,12].

Acknowledgments

This study was funded by NIH grant 5R03CA193078 (PI: Krieger), which was supported by the National Cancer Institute (NCI) and the NIH Office of Behavioral and Social Sciences Research, in the Office of the Director, National Institutes of Health (OD).

We also would like to thank Marvella A. Villaseñor, BA (Division of Research, Kaiser Permanente Northern California, 2000 Broadway, Oakland, CA 94612; email: Marvella. A. Villaseñor@kp.org) for her administrative assistance with study logistics.

FUNDING: This study was funded by NIH grant 5R03CA193078 (PI: Krieger), which was supported by the National Cancer Institute (NCI) and the NIH Office of Behavioral and Social Sciences Research, in the Office of the Director, National Institutes of Health (OD). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the article; and decision to submit the article for publication.

References

1. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet.* 2009; doi: 10.1146/annurev.genom.9.081307.164217
2. Tang Y-C, Amon A. Gene copy-number alterations: a cost-benefit analysis. *Cell.* 2013; 152:394–405. [PubMed: 23374337]
3. Hart L, O'Driscoll M. Causes and consequences of structural genomic alterations in the human genome. *eLS.* 2013; doi: 10.1002/9780470015902.a0024976
4. Thompson PA, Brewster AM, Kim-Anh D, Baladandayuthapani V, Broom BM, Edgerton ME, Hahn KM, Murray JL, Sahin A, Tsavachidis S, Wang Y, Zhang L, Hortobagyi GN, Mills GB, Bondy ML. Selective genomic copy number imbalances and probability of recurrence in early-stage breast cancer. *PLoS One.* 2011; 6:e23543. [PubMed: 21858162]
5. Brewster AM, Thompson P, Sahin AA, Do K, Edgerton M, Murray JL, Tsavachidis S, Zhou R, Liu Y, Zhang L, Mills G, Bondy M. Copy number imbalances between screen- and symptom-detected breast cancers and impact on disease-free survival. *Cancer Prev Res (Phila).* 2011; 4:1609e1616. [PubMed: 21795423]
6. Liu Y, Zhou R, Baumbusch LO, Tsavachidis S, Brewster AM, Do K-A, Sahin A, Horobagyi GN, Taube JH, Mani SA, Aarøe J, Wärnberg F, Børresen-Dale A-L, Mills GB, Thompson PA, Bondy ML. Genomic copy number imbalances associated with bone and non-bone metastasis of early-stage breast cancer. *Breast Cancer Res.* 2014; 143:189–201.
7. Cava C, Bertoli G, Castiglioni I. Integrating genetics and epigenetics in breast cancer: biological insights, computational methods and therapeutic potential. *BMC Syst Biol.* 2015; 9:62.doi: 10.1186/s12918-015-0211-x [PubMed: 26391647]
8. Santarpia L, Bottai G, Kelly CM, Gyoffry B, Székely B, Puszlai L. Deciphering and targeting oncogenic mutations and pathways in breast cancer. *Oncologist.* 2016; 21:1063–1078. [PubMed: 27384237]
9. Rennhack J, To B, Wermuth H, Andrecheck ER. Mouse models of breast cancer share amplification and deletion events with human breast cancer. *J Mammary Gland Biol Neoplasia.* 2017; 22:71–84. [PubMed: 28124185]
10. Krieger N. History, biology, and health inequities: emergent embodied phenotypes & the illustrative case of the breast cancer estrogen receptor. *Am J Public Health.* 2013; 103:22–27. [PubMed: 23153126]
11. Krieger N, Habel LA, Waterman PD, Shabani M, Ellison-Loschmann L, Achacoso NS, Acton L, Schnitt SJ. Analyzing historical trends in breast cancer biomarker expression: a feasibility study (1947–2009). *npj Breast Cancer.* 2015; 1:15016.doi: 10.1038/npjbcancer.2015.16 [PubMed: 26807440]
12. Kokkat TJ, Patel MS, McGarvey D, LiVolsi VA, Baloch ZW. Archived formalin-fixed paraffin-embedded (FFPE) blocks: a valuable underexploited resource for extraction of DNA, RNA, and protein. *Biopreservation Biobanking.* 2013; 11:101–106. [PubMed: 24845430]
13. Valla M, Vatten LJ, Engstrøm MJ, Haugen OA, Akslen LA, Bjørngaard JH, Hagen AI, Ytterhus B, Bofin AM, Ordahl S. Molecular subtypes of breast cancer: long-term incidence trends and prognostic differences. *Cancer Epidemiol Biomarkers Prev.* 2016; 25:1625–1634. [PubMed: 27672056]
14. Wang Y, Cottman ME, Schiffman JD. Molecular inversion probes: a novel microarray technology and its application in cancer research. *Cancer Genetics.* 2012; 205:341–355. [PubMed: 22867995]
15. Affymetrix. [Accessed 8 January 2018] OncoScan® CNV FFPE Assay Kit. http://www.affymetrix.com/support/technical/byproduct.affx?product=oncoscan_cnv_ffpe_assay_kit
16. Mittempergher L, de Ronde JJ, Nieuwland M, Kerkhoveen RM, Simon I, Rutgers EEJT, Wessels LFA, Van't Veer LJ. Gene expression profiles from formalin fixed paraffin embedded breast cancer tissue are largely comparable to fresh frozen matched tissue. *PLoS ONE.* 2011; 6(2):e17163. [PubMed: 21347257]
17. Herrmann, B., Hummel, S., editors. *Ancient DNA: Recovery and Analysis of Genetic Material from Paleontological, Archeological, Museum, Medical, and Forensic Specimens.* New York: Springer-Verlag; 1994.

18. Hendricks, RL. A Model for National Health Care: The History of Kaiser Permanente. Rutgers University Press; New Brunswick, NJ: 1993.
19. Gordon, N. What the Member Health Surveys Project Tells Us about the Kaiser Permanent Northern California Adult Membership: Demographics, IT Access, Behavioral/Lifestyle Risks, and Health: Trends, Race-Ethnic Differences, and Variation across Service Populations. Oakland, CA: Kaiser Permanente Division of Research; 2008 Feb. 2008. https://divisionofresearch.kaiserpermanente.org/projects/memberhealthsurvey/SiteCollectionDocuments/What_Member_Health_Surveys_Project_Tells_about_KPNC_Population_Feb_2008.pdf [Accessed 8 January 2018]
20. Kaiser Permanente Division of Research. [Accessed 8 January 2018] Cancer Genomics. 2017. <https://divisionofresearch.kaiserpermanente.org/Pages/cancergenomics.aspx>
21. [Accessed 8 January 2018] OncoScan™ Nexus Express Software. https://www.affymetrix.com/products_services/software/specific/oncoscan_nexus.affx#1_1https://tools.thermofisher.com/content/sfs/manuals/oncoscan_nexus_express_software_qrc.pdf
22. Jung H, Lefferts JA, Tsongalis GJ. Utilization of the Oncoscan microarray assay in cancer diagnostics. Applied Cancer Research. 2017; 37:1.
23. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S. METABRIC Group. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012; 486:346–352. [PubMed: 22522925]
24. Singh RR, Mehrotra M, Chen H, Almohammedsalim A, Sahin A, Bosamra A, Patel KP, Routbort MJ, Lu X, Ronald A. Comprehensive screening of gene copy number aberrations in formalin-fixed, paraffin-embedded solid tumors using molecular inversion probe-based single-nucleotide polymorphism array. J Molec Diagnostics. 2016; 18:676–687.

DNA copy number variation (CNV) assay results for 30 primary invasive breast cancer tumor formalin-fixed paraffin-embedded (FFPE) specimens, 1950s–2000s: 5 randomly selected cases per decade, based on tumor specimens obtained from women members of the Kaiser Permanente Northern California Health Plan.

Year Range	ID	Year of diagnosis	Histologic Diagnosis	Histologic Grade	Total DNA (ng)	MA PD	MA PD QC test	MA PD QC test : pas vs. fail	ndSN PQC test	ndSN PQC test: pass vs. fail	Need verification for ploidy	#CNV after filtering 0.2 and -0.2	average CNV length (after filtering)	# of CNV genes	extra time for evaluation (minutes)	max CNV length	min CNV length	Average LOH length	min LOH length	max LOH length	High frequency Amp genes	High frequency high Amp genes	High frequency Deleted genes
1950–1959	1-1	1955	Invasive carcinoma with ductal and lobular features	2	1,095.62	0.36	Fail	Fail	18.73	Fail	No	24	1,682,422	2,749	24*2	13,411,570	25,680	5,279,027	2,510,712	28,926,345	61%	20%	37%
	1-2	1957			1,182.51	0.36	Fail	Fail	18.67	Fail	No	18	338,577	2,113	18*2	1,300,793	15,750	3,700,365	2,529,793	6,703,841			
	1-3	1959	invasive ductal carcinoma (IDC)	3	660.03	0.45	Fail	Fail	13.31	Fail	No	37	4,033,237	7,856	37*2	36,061,869	10,818	8,267,304	2,527,521	40,024,461			
	1-4	1956	invasive lobular carcinoma	2	467.96	0.41	Fail	Fail	15.35	Fail	No	36	5,535,344	4,666	36*2	1.04E+08	84,896	9,109,480	2,612,904	1.09E+08			
	1-5	1956	IDC	3	873.77	0.38	Fail	Fail	15.09	Fail	No	39	4,892,183	8,044	39*2	61,981,422	45,594	13,844,743	2,635,882	60,226,159			
average												30.8	3,296,353	5085.6	61.6								
percentage in at least one sample																							
1960–1969	2-6	1968	solid papillary carcinoma	2	883.56	0.38	Fail	Fail	15.86	Fail	No	17	1,028,303	942	17*2	11,514,689	81,272	3,822,958	2,561,422	10,052,453			
	2-7	1969			135.07	0.36	Fail	Fail	20.18	Fail	No	28	2,476,968	2,327	28*2	27,384,234	1	5,873,802	2,500,211	27,266,659			
	2-8	1969	IDC	2	448.59	0.24	Pass	Pass	38.73	Pass	No	1	243,707	4	0	243,707	243,707	10,494,780	2,673,029	45,636,027			
	2-9	1968	invasive mucinous carcinoma	2	1,019.97	0.37	Fail	Fail	16.20	Fail	No	44	1,884,659	3,796	44*2	7,095,473	25,695	3,438,664	2,529,227	7,852,817			
	2-10	1965	IDC with medullary features	3	736.67	0.33	Fail	Fail	17.09	Fail	No	27	6,559,930	11,303	27*2	1.23E+08	86,147	14,246,463	2,626,244	1.1E+08			
average												23.4	2,438,713	3674.4	46.4								
percentage in at least one sample																							
1970–1979	3-11	1973	IDC	1	269.99	0.43	Fail	Fail	12.87	Fail	No	100	5,490,496	7,073	100*2	35,062,281	21,070	6,281,059	2,504,259	34,883,122			
	3-12	1974	IDC	3	437.25	0.31	Fail	Fail	20.48	Fail	No	49	15,574,496	13,053	49*2	1.43E+08	47,710	29,675,247	2,670,444	1.1E+08			
	3-13	1977	IDC	3	682.76	0.22	Pass	Pass	37.82	Pass	No	25	1,541,987	5,087	0	7,042,968	43,667	3,280,221	2,502,596	5,578,689			
average																							
percentage in at least one sample																							
																					72%	20%	25%

Year Range	ID	Year of diagnosis	Histologic Diagnosis	Histologic Grade	Total DNA (ng)	MA PD	MA PD QC test	MA PD QC test : pass vs. fail	ndSN PQC	ndSN PQC test: pass vs. fail	Need verification for ploidy	#CNV after filtering and -0.2	average CNV length (after filtering)	# of CNV genes	extra time for evaluation (minutes)	max CNV length	min CNV length	Average LOH length	min LOH length	max LOH length	High frequency Amp genes	High frequency high_Amp genes	High frequency Del genes
1980–1989	3-14	1979	solid papillary carcinoma	2	1,173.16	0.24	Pass	Pass	33.52	Pass	No	112	11,308,458	14,983	0	1.03E+08	93,769	36,864,056	2,502,872	1.09E+08	80%	61%	54%
	3-15	1974		3	992.45	0.24	Pass	Pass	40.73	Pass	No	6	488,407	97	0	1,307,382	198,539	3,554,236	2,523,683	6,042,236			
	average										58.4	6,880,769	8,058.6	59.6									
	percentage in at least one sample																						
	4-16	1988	IDC	3	510.44	0.26	Pass	Pass	33.43	Pass	No	27	6,788,831	9,936	0	33,309,995	90,738	19,613,682	2,850,366	96,803,567			
	4-17	1980	IDC	3	150.92	0.37	Fail	27.03	27.03	Pass	No	103	2,351,078	12,288	103*2	32,300,086	15,854	4,075,081	2,507,018	10,435,818			73%
4-18	1985	IDC	3	145.78	0.27	Pass	39.89	39.89	Pass	No	51	3,135,476	4,215	0	93,174,821	24,138	8,710,679	2,609,203	43,696,697				
4-19	1986	tubular carcinoma	1	83.39	0.25	Pass	41.46	41.46	Pass	No	73	2,726,320	6744	0	26,342,526	38,963	10,695,057	2,588,535	47,254,294				
4-20	1985	IDC	1	242.03	0.26	Pass	32.40	32.40	Pass	Yes, Fish	86	13,700,922	14,173	0	1.52E+08	18,696	44,596,680	2,529,793	1.48E+08				
average											68	5,740,525	9471.2	41.2									
percentage in at least one sample																							
1990–1999	5-21	1999	IDC	3	595.00	0.28	Pass	Pass	33.52	Pass	No	43	23,364,608	12,354	0	1.73E+08	10,973	3,6911,557	2,509,179	1.21E+08	75%	73%	76%
	5-22	1994	mixed IDC and mucinous carcinoma	3	98.32	0.27	Pass	Pass	39.96	Pass	No	45	1,749,286	3,775	0	32,109,295	25,496	10,189,556	2,508,123	56,514,458			
	5-23	1996	invasive carcinoma with ductal and lobular features	2	588.56	0.27	Pass	Pass	35.10	Pass	No	21	7,883,124	2,124	0	74,140,774	92,479	7,492,691	2,552,335	43,696,697			
	5-24	1995	invasive carcinoma with ductal and lobular features	2	178.35	0.27	Pass	Pass	40.82	Pass	No	75	7,747,462	7,074	0	64,273,327	13,751	16,970,838	2,638,266	90,167,746			
	5-25	1998	invasive carcinoma with ductal and lobular features	2	1,275.15	0.32	Fail	14.96	14.96	Fail	No	110	5,636,925	15,560	110*2	62,423,174	13,524	19,823,897	2,634,336	87,828,838			
	average											58.8	9,276,281	8,177.4	44								
percentage in at least one sample																							

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Year Range	ID	Year of diagnosis	Histologic Diagnosis	His- to- logi- c Grade	Total DNA (ng)	MA PD	MA PD QC test : pas s vs. fail	ndSN PQC test: pass vs. fail	ndSN PQC test: pass vs. fail	Need verification for ploidy	#CN V after filter ing for 0.2 and -0.2	average CNV length (after filterin g)	# of CNV genes	extra time for evalua- tion (minu- tes)	max CNV length	min CNV length	Average LOH length	min LOH length	max LOH length	High frequency Amp genes	High frequency high_ Del genes
2000–2010	6-26	2006	IDC	3	1,072.92	0.24	Pass	39.43	Pass	Yes, Fish	49	13,926,103	8,165	0	1.03E+08	37,093	22,196,487	2,556,130	1.03E+08	99%	56%
	6-27	2002	IDC	3	47.37	0.29	Pass	31.62	Pass	No	131	10,001,618	14,657	0	1.39E+08	12,927	14,713,013	2,563,197	87,062,942		
	6-28	2000	invasive carcinoma with ductal and lobular features	2	1,257.89	0.32	Fail	18.65	Fail	No	48	7,725,056	9,529	48*2	1.31E+08	44,481	12,335,936	2,589,466	62,266,230		
	6-29	2002	IDC with medullary features	3	80.88	0.27	Pass	39.90	Pass	No	6	431,092	1,209	0	1,128,185	41,592	3,669,066	2,622,587	4,451,715		
	6-30	2000	IDC	3	1,469.39	0.26	Pass	29.67	Pass	No	111	10,761,343	17,855	0	98,630,207	47,661	28,773,328	2,568,877	1.13E+08		
	average percentage in at least one sample										69	8,569,042	10,283	19.2						99%	74%