January 2017

Genetics Of Chemotherapy Response In Triple Negative Breast Cancer

Charles Li
Yale University

GENETICS OF CHEMOTHERAPY RESPONSE IN TRIPLE NEGATIVE BREAST CANCER

A Thesis Submitted to the Yale University School of Medicine in Partial Fulfillment of the Requirements for the Degree of Doctor of Medicine

by

Charles Li

2017

GENETICS OF CHEMOTHERAPY RESPONSE IN TRIPLE NEGATIVE BREAST CANCER

Li and Lifton. Department of Genetics, Yale School of Medicine, New Haven, CT

Abstract

Triple Negative Breast Cancer (TNBC) encompasses a wide range of treatment responses, however there are no predictive biomarkers approved for clinical use to target therapy. With a novel exome analysis method, we discovered that the overall proportion of the homologous recombination repair (HRR) genes affected by structural variation can accurately predict both positive and negative chemotherapy response prior to initiation of therapy in the large majority of patients in our cohort.

We analyzed exome sequences of unpaired tumor samples, collected prior to ACT chemotherapy, in 17 TNBC patients who exhibit complete pathologic response to neoadjuvant chemotherapy (pCR) and 15 patients who had extensive residual disease (RD).

In the process, we created one of the first analytical pipelines capable of performing comprehensive integrated analysis of somatic point mutations and structural variation in unpaired tumor exome samples. Validation on tumor-

normal matched samples demonstrated >95% specificity for point mutations, LOH, and CNV calling compared to standard tumor-normal somatic analysis.

When applied to the TNBC cohort, our somatic mutation caller identified multiple damaging somatic mutations in genes linked to EMT. LOH analysis showed significantly greater LOH in pCR (Complete Response) than RD patients (Residual Disease) (p=6.5E-12). The five regions with greatest LOH difference between pCR and RD subgroups each contained a HRR gene locus. Overall high LOH burden was associated with the presence of TP53 point mutations (p=0.002).

By integrating data from all three methods, we found significantly more pCR patients with high mutation burden (including CNV and LOH) in homologous recombination repair genes than RD patients (83% vs 20%). With this metric, we can predict 83% of positive response and 80% of negative response based on our patients' genomic profiles prior to chemotherapy initiation (OR=18.7, 95% CI= 3.2 to 110.3, p=0.0012).

This result offers a potential significant improvement in our ability to personalize therapy in TNBC and may facilitate development of targeted PARP inhibitor therapeutics.

Acknowledgements

I would like to express my deepest gratitude to my advisor, Dr. Richard Lifton, who has guided and supported my work over every hill and across every valley. In the face of seemingly insurmountable challenges, I was comforted by the fact that Dr. Lifton always knew the way ahead. He has been the mentor and the life-long scholar I aspire to be.

I'd like to extend a special thanks to the members of Lifton Lab, in particular Bixiao Zhao, Jungmin Choi, Samir Zaidi, and Durga Thakral for teaching me nearly all I know about coding and running genomic analysis.

Finally, I am particularly grateful to my family, loved ones, and my best friend Tanner, who have provided unwavering support through difficult stretches and long nights.

Chapter 1: Introduction

Breast cancer is one of the most prevalent types of cancers and a leading cause of morbidity and mortality for women both young and old. In 2016, there were an estimated 246,660 new cases of breast cancer, comprising 14.6% of total cancer cases, and an estimated 40,450 deaths from breast cancer, comprising 6.8% of total cancer deaths. Over a lifetime, a woman has a 12.4% chance of being diagnosed with breast cancer. [1]

In the past decade, breast cancer treatment has been revolutionized by new therapies designed to target specific molecular profiles allowing for treatment to be personalized for most eligible patients.

The subsequent improvement in outcomes has led to a beacon of hope in a devastating disease. While new cases of breast cancer have remained relatively constant between 1992 and 2013, the death rate has decreased significantly from 31/100,000 women to 20.7/100,000.[1]

In an early example of personalized medicine in oncology, clinicians found the first targetable subgroup within breast cancer with the discovery that endocrine therapy appeared effective in treating breast cancer in patients who overexpress ER/PR. Tamoxifen, a high effective breast cancer treatment that specifically targets ER, was shown to decrease breast cancer recurrence by 47% in patients

with ER+ breast cancer and reduce mortality by 26%, changing the prognosis for thousands of patients to come. [2]

However, endocrine therapy is rarely effective in patients without estrogen receptor and progesterone receptor. Additionally, tumors which lose expression of estrogen receptor and progesterone receptor during metastases were found to be similarly resistant to endocrine therapy.[3]

A third marker, HER2+, was first discovered as a proto-oncogene amplified in 30% of breast tumors. Subsequent studies found that amplification of HER2 predicted poor survival in breast cancer. Based on this, trastuzumab, a monoclonal antibody targeting HER2, was developed and then found to reduce risk of death by 33% in patients who overexpress HER2.[4]

All targeted therapies share the trait of working exceedingly well in patients who express the targeted molecular profile while showing minimal effects in patients lacking expression. When tested in breast cancer as a whole, without regard to patients' molecular profiles, the average benefit of any of these therapies is greatly reduced.

As such, breast cancer patients routinely undergo molecular profiling as part of their diagnostic work-up. The 85% of cases who express at least one marker

are directed towards first-line therapeutics targeted towards their individual tumor characteristics.

Section 1A – Introduction to TNBC

Triple Negative Breast Cancer by definition lack all 3 markers used today to personalize treatment. Without a targeted therapy available, TNBC patients have a worse prognosis than other subtypes with reduced overall survival and greater recurrence.[5]

However, as a diagnosis of exclusion, tumors in TNBC are not necessarily a single disease, and in fact demonstrate great diversity in histopathologic features [6] and genetic profiles. A Cancer Genome Atlas (TCGA) study of 510 exome-sequenced breast tumors found TP53 to be the only gene in which there were point mutations in the majority of TNBC tumors, with 80% of patients harboring a mutation. PIK3CA was found to be the next most commonly mutated gene with 8% of patients harboring mutations. Most other somatic mutations were scattered among a multitude of genes at low frequency.[7]

Clinically, TNBC patients have a wide range of outcomes and chemotherapy responses. While TNBC has a worse prognosis than other types of breast cancer, patients who achieve pathologic complete response (pCR) after chemotherapy, as defined by complete absence of residual tumor tissue on pathologic examination, have a similar progression free survival to patients of

other breast cancer types. Conversely, patients with Residual Disease (RD) burden have significantly worse survival. [8]

Section 1B-Introduction to Next Generation Sequencing:

The emergence of next generation sequencing as a key tool the lab and clinic has transformed the way we test and diagnose patients. Gene sequencing has been a key component of genetics since the discovery of Sanger Sequencing in 1977[9] which sequences genes one base at a time in series. While exceedingly accurate, Sanger Sequencing becomes impractical for more than a small set of targets. As such, it has been most useful in studying single gene diseases such as Cystic Fibrosis and Huntington's Disease. [10]

This approach becomes particularly problematic in understanding complex phenotypes involving multiple genes, such as hypertension, autism, and cancer. In such diseases, it can be difficult to narrow down a list of likely causal genes prior to sequencing. Diseases which have been thoroughly sequenced with other methods have demonstrated that there is often a broad range of genetic variation contributing to pathogenesis. [11] [12]

Next Generation Sequencing utilizes a parallel sequencing approach to exponentially increase sequencing output. Rather than sequence single bases, next generation sequencing analyzes millions of bases at once.[13] This has led to

a dramatic decrease in cost and has made whole exome sequencing of large cohorts feasible.

In practice, the accessibility of next generation sequencing has allowed clinicians and scientists alike rapidly obtain data on the nearly complete genomic landscape of patients, and has enabled the discovery of novel and often times unexpected mechanisms of disease in both Mendelian disease and cancer.

Section 1C- Impact of Genomics in Oncology

In oncology, next generation sequencing (NGS) has had a particularly profound impact. Despite its recent discovery, the vast trove of data produced by next generation sequencing has already impacted the way we understand, diagnose, and treat cancer.

Prior to the advent of next generation sequencing, imatinib proved the concept of targeting a specific mutation to treat cancer by rationally designing an inhibitor for the BCR-ABL fusion protein found in most CML patients. [14]

Next generation sequencing has greatly accelerated this process of discovering new targetable mechanisms and driver mutations.

Section 1D – Genomics in Oncology: Discovery of New Mechanisms

One of the key advantages of next generation sequencing lies in its ability to sequence an entire exome or genome. For rare and poorly understood diseases in which we do not have known mechanisms, whole exome sequencing can reveal entirely novel and unexpected disease mechanisms.

As an example, in fibrolamellar hepatocellular carcinoma (FHC), a rare liver tumor, sequencing has led to the rapid development of a new targeted therapy. The first study of its kind sequenced 15 patients with FHC, and found that all patients harbored a novel chimeric transcript: DNAJB1-PRKACA.[15] Based on this research, a compound targeting one of the components of the chimeric protein is currently in clinical trials less than 2 years after discovery.

Similarly, small cell carcinoma of the ovary, hypercalcemic type, had long been a poorly understood disease that often went undiagnosed due to a lack of reliable specific markers for diagnosis. Exome sequencing revealed to be driven by SMARCA4, previously not implicated in this disease, as 30/32 cases in the study contained SMARCA4 mutations. This discovery has not only led to new insights on the pathogenesis of this disease, but also a possible diagnostic marker and direction for therapeutic research.[16]

Section 1E – Genomics in Oncology: Personalizing Therapy

For better studied diseases in which we already have an understanding of the pathophysiology, gene sequencing can be used to characterize new molecular

subtypes that can serve as a basis for personalizing current therapies to fit a patient's molecular profile and as a basis for discovery of new treatment strategies.

In non-small cell lung carcinoma, the discovery of key driver mutations has led to a new treatment paradigm guided by a patient's somatic mutation landscape. Patients with somatic mutations in EGFR were found to respond well to Gefitinib, a drug targeting EGFR. (Lynch, 2004)[17] Based up on the discovery of a second recurrent somatic mutation, the EML4-ALK fusion protein, a new treatment was developed specifically to target this mutation.[18]

In the near future, it may in fact be feasible to target the mutation profile of a patient rather than cancer type. The National Cancer Institute identified 24 gene mutations with targeted treatments, and is currently undertaking a clinical trial to evaluate the efficacy of treating patients with at least one of these 24 mutations based on their molecular profiles rather than cancer type. [19]

Section 1F - Genomics in Oncology: Improving upon existing therapeutics

For patients who do not respond to therapy or who develop resistance, gene sequencing has been used to discovery mechanisms behind drug resistance.

In melanoma, next generation sequencing has been used to discover multiple mechanisms of resistance in patients treated vemurafenib, a BRAF targeted

therapy, including novel amplifications, [20] downstream mutations [21], and activation of alternate pathways. [22]

For melanoma patients treated with anti PD-1 immunotherapy, exome sequencing was used to discovery mechanisms of resistance even prior to FDA approval. 3/4 patients who had developed resistance in clinical trials were sequenced and found to have acquired mutations in genes involved in interferon-receptor signaling and antigen presentation upon relapse. [23]

These discoveries not only demonstrate the potential of next generation sequencing to rapidly discover mechanism of resistance, they also set the foundation for development of second line therapy and next generation targeted therapies.

Chapter 2: Development of Methods

All methods described in this section were coded and validated by the author, under the close guidance of Dr. Lifton.

Section 2A – The Challenge of Analyzing Unpaired Tumors:

In order to perform comprehensive analysis on this cohort of unpaired TNBC tumor samples and enable analysis of other unpaired tumor exome samples, we've developed a novel method which conducts specific analysis of point mutations, loss of heterozygosity, and copy number variations in unpaired tumor exome data.

Current gold standard algorithms rely on differences between tumor and normal samples to distinguish somatic mutations and germline mutations. In the absence of paired normal samples, we've developed and validated novel algorithms that rely on prior distributions to specifically call somatic point mutations and structural variation.

Tumor biopsies have been routinely obtained for diagnosis and cancer grading for decades. [24] Thousands of samples are currently stored in tissue banks and may contain many new insights in their somatic mutation landscapes. [25] [26]

However, among the many routine biopsies obtained annually, few have matched blood controls. [27] Most samples contain only tumor tissue, from which we can obtain limited data with today's genomic tools. As a result, TNBC and many other tumors have fallen behind the full potential of next generation sequencing to understand their biology and potential therapeutic strategies.

A reliable method analyzing unpaired tumor samples can enable exome sequencing to unlock genomic insights in thousands of unmatched tumor samples which are unanalyzable with today's methods.

Section 2B – Challenge of Calling Somatic Point Mutations

Somatic mutations and germline mutations play important yet distinct roles in pathogenesis of cancer. Whereas germline mutations affect cancer predisposition, somatic mutations drive the complex evolution and biology of a tumor. The Hallmarks of Cancer, as famously outlined Hanahan and Weinberg, define essential capabilities for tumor cells to survive and proliferate. While certain germline mutations may contribute to this process, these capabilities are typically acquired during the evolution of tumor cells.[28] Thus, sensitive and specific identification of somatic mutations is essential for understanding the biology of malignant transformation, treatment resistance, and molecularly profiling of tumors.

However, there are significant challenges in reliably detecting and calling somatic mutations. Germline variants within most tumors vastly outnumber somatic variants as somatic point mutations tend to occur at low frequency.[29] Additionally, minor allele frequencies of somatic mutations are highly variable due to tumor heterogeneity and presence of sub-clones within a tumor. [30]

Even in highly conserved tumor suppressor genes such as TP53, variants within a tumor sample can include somatic driver mutations acquired in tumorigenesis as well as benign germline variants present since birth. [31] [32]

Current methods overcome these challenges by comparing the sequence of a tumor sample to that of a non-tumor sample from the same patient to call

somatic mutations. Since somatic mutations by definition arise in the tumor, mutations detected only in the tumor sample are most likely somatic. Nearly all somatic mutation callers today use variations of this approach[33] which precludes analysis of non-paired cohorts, such as this TNBC cohort.

In order to reliably analyze this cohort and other unpaired tumor cohorts, we developed and validated a novel somatic mutation caller that only requires a tumor sample.

Section 2C – Development of Bayesian Classification Algorithm

Tumor samples typically contain a mix of tumor cells and contaminating normal cells, including stromal cells and immune cells[34]. As such, we expected and observed that the distribution of germline mutation minor allele frequencies centers around 0.5, as both alleles are present in equal proportions in both normal cells and tumor cells in the absence of LOH, whereas somatic MAFs should have a maxima lower than 0.5 as these are only present in the tumor cells as shown in Figure 1.

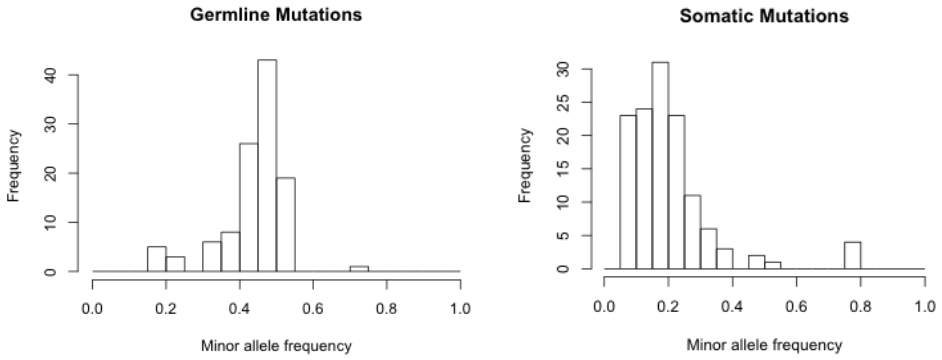

Figure 1: **Germline vs Somatic Minor Allele Frequency –** Minor allele frequencies of Germline and Somatic mutations from a tumor-normal paired sample exhibit distinct distributions.

Based upon this concept, we applied a Bayesian inference based algorithm (Figure 2) to accurately calculate the probability of mutation to be germline, with probability of being somatic calculated as the complement. (1-p(germline))

$$p(g \mid maf) = \frac{p(maf \mid g)\,p(g)}{p(maf)}$$

Figure 2: **Bayesian Equation to determine probability of germline mutation -** Probability of a germline mutation given a minor allele frequency *p(g|maf)* is calculated with the distribution of minor allele frequency of rare variants in normal samples represented in *p(maf|g)*, the estimated proportion of germline variants in the sample *p(g)*, and the overall distribution of minor allele in the tested sample, *p(maf).*

The prior distribution of germline variants, p(g), was collected from novel germline variants in a panel of normal controls from other cohorts. We estimated the prior probability of a germline variant within the cohort by clustering minor allele frequencies with a parameterized Gaussian mixture model. [35]

The use of the prior probability is important for accurate calling as somatic mutation counts vary significantly between tumors. We expect that the probability of a mutation to be a somatic is lower in a tumor with few somatic mutations than a tumor with a hyper mutation phenotype at any given minor allele frequency.

Once all the parameters are inputted, the algorithm calculates posterior probability of germline vs somatic origin for each minor allele frequency and calls somatic mutations based off of a probability cutoff.

One key challenge in calling somatic mutations with minor allele frequency (MAF) is loss of heterozygosity. MAFs of germline variants within regions of LOH can be altered to similar values as somatic variants MAFs. (Figure 3)

Therefore, we developed an automated algorithm to filter calls in LOH regions. To strike a balance between maintaining high specificity while not losing potential driver mutations, we first analyzed non-LOH calls with the Bayesian Algorithm to generate a list of high confidence somatic mutations. We then used novel variants found in LOH regions to inform our rankings and analysis of these genes.

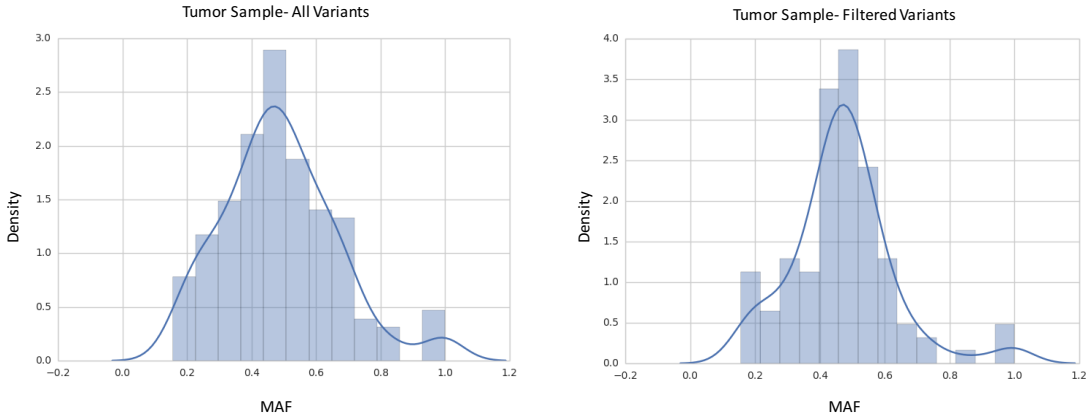

Figure 3: **Minor Allele Frequency in High LOH Sample –** LOH in this sample alters minor allele frequencies of germline variants, which can mask somatic calls and reduce specificity, particularly in a sample with few somatic point mutations such as this sample. By filtering out variants in LOH regions, we can maintain high specificity in tumors with high LOH.

Section 2D – Validation of Bayesian Classification Algorithm

We validated the Bayesian Algorithm against GATK and Mutect2, the two most commonly used germline and somatic mutation calling methods today. These samples are paired samples were randomly selected from an ongoing ovarian cancer exome sequencing project. Samples without significant somatic mutation burden, and with extremes of tumor purity were excluded to maintain comparability between the validation cohort and the TNBC cohort.

In each pair, the normal sample was removed and the remaining tumor-only sample was analyzed with the Bayesian Algorithm. As a gold-standard comparison, we called known somatic variants with the Mutect2 somatic variant caller and called known germline mutations by analyzing only normal samples with GATK.

Compared to Mutect2 paired calling, tumor-only analysis with the Bayesian Algorithm achieved 87% sensitivity and 98% specificity.

Sample #	True Positive	False Positve	True Negative	False Negative
1	18	4	158	0
2	28	1	93	1
3	12	0	125	1
4	18	2	57	8
5	21	1	130	2
6	7	2	46	0
7	3	2	156	0
8	5	7	67	5

Bayesian Algorithm Validation Cohort – A summary of the sequencing quality as output by GATK. On average, tumors were sequenced to high depth of coverage with a mean coverage of 163.9, a necessity for accurate somatic mutation calling.

Section 2E – Role of Loss of Heterozygosity (LOH) in Cancer

LOH and aneuploidy are key sources of tumor genetic variation. Mechanisms behind LOH include chromosomal recombination, both homologous and nonhomologous, as well as mitotic non-disjunction for chromosome-wide events. [37]

In cancer, these mechanisms can play a key role in deleting tumor suppressors and creating copy number changes. As proposed in the Knudson Two-Hit Hypothesis, certain key transformations in cancer require two separate mutation events, particularly if a single healthy allele can maintain tumor suppressor function as seen in retinoblastoma. [38] LOH can serve as the second event by removing the healthy allele and over expressing the disease allele leading to

cancer. [39] As such, in Li-Fraumeni Syndrome, where patients have a deleterious germline mutation in TP53, patients with tumors often demonstrate LOH at TP53 alongside germline mutations in TP53. [40]

Other studies have shown that LOH plays a key role in the oncogenesis of multiple types of cancer, including loss of VHL in Renal cell carcinoma[41] and LOH in PTEN in multiple cancers including breast cancer, endometrial cancer, and thyroid cancer.[42] Additionally, LOH in specific loci have been proposed as markers for cancer risk prediction based on correlation with clinical characteristics. [43]

Section 2F - Current challenges in LOH Calling

Current algorithms for LOH calling, including ExomeCNV[44] and ExomeAI[45], call LOH by comparing the difference of B allele frequency deviation from 0.5 between case and controls.

Without LOH, most SNVs have B-Allele frequencies that center around 0.5. With LOH, we expect heterozygous mutations to have B allele frequencies to form a bimodal peak at <0.5 and >0.5. This difference is a key metric used for calling LOH and is quantified as the absolute deviation from expected B allele frequencies: |BAF – 0.5| Current methods call LOH by filtering for regions where the average deviation in the tumor is greater than the average deviation in the control.

Manual calling, or visual calling, can be used to call LOH as well. However, this presents challenges in consistent calling and may be prone to experimenter bias without proper blinding.

Section 2G – Development of Novel LOH Calling Method

With unpaired samples, we found that relying on average deviation alone risks overcalling LOH, particularly in tumor samples which have high variability in B allele frequencies due to tumor heterogeneity and tumor impurity. In recurrent regions of LOH, it can be difficult to determine whether the changes in BAF deviation (|BAF-0.5|) are due to a recurrent artifact or a true significant mutation (Figure 4).

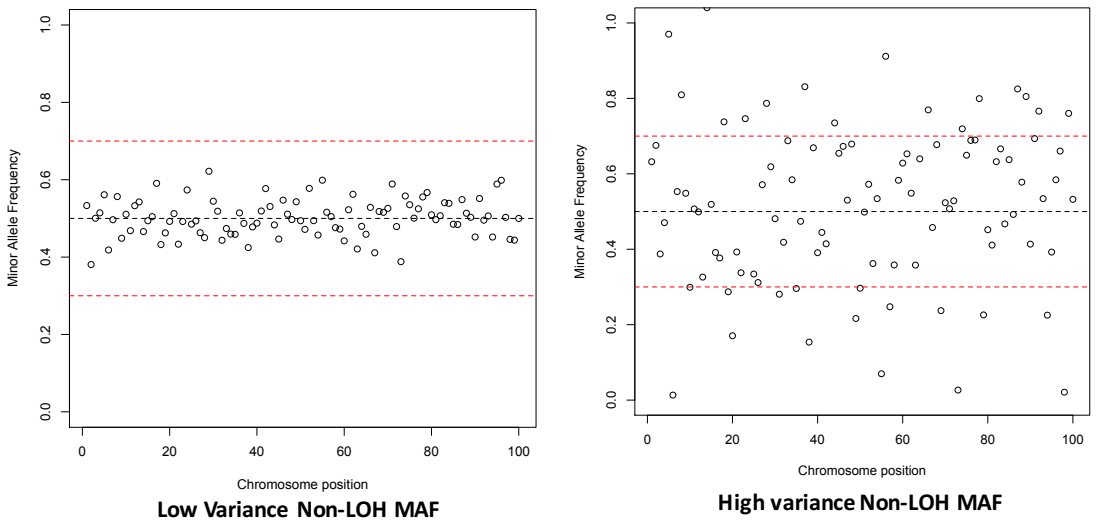

Figure 4: **Effect of Variance on LOH Calling -** We created a simulated a dataset to demonstrate the effect of variance in false-positive LOH calls. Neither of these two regions have LOH, given the presence of expected heterozygous alleles. However, the average deviation from expected B allele frequency, as measured by distance from 0.5, is markedly different between these samples. Based on this metric alone, as most methods use, the high variance sample may be indistinguishable from actual LOH, particularly in impure tumor samples.

Current algorithms avoid false positive calls by using a paired control to filter effects of local variance. To accurately call LOH without paired normal samples, we developed a novel method that accounts for the observed evidence as well as the prior distribution of expected B allele frequencies.

Similar to current methods, we use the presence of deviated B allele frequencies. However, we also incorporate the absence of expected heterozygous B-allele frequencies near 0.5 to create a new metric for LOH calling. By using both metrics rather than BAF deviation alone, we can increase sensitivity and specificity of LOH calling without a paired tumor sample.

Section 2H – Validation of LOH Calling Method

We validated our approach in 3 steps. We first confirmed that regions of LOH called with our algorithm aligned with visual calling in a set of 10 tumor samples. To evaluate sensitivity and specificity, we analyzed a set of 3 tumor-normal breast cancer paired samples. We found that our novel algorithm was >98% sensitive and >98% specific when compared to paired calling.

Finally, to further test the frequency of false positive calls, we tested our algorithm on a panel of 25 normal samples collected from multiple tumor projects including ovarian cancer and uterine serious carcinoma with the

assumption that normal samples should not contained significant LOH. In this test, unpaired calling was shown to be >98% specific compared to paired calling.

Section 2I – The Role of Copy Number Variation in Cancer

CNVs form a key component of genomic diversity in disease. [46] Compared to SNVs, which have a relatively constant mutation rate of $1.8–2.5 \times 10^{-8}$ per base pair, CNVs have a highly variable mutation rate across diseases ranging from 1.7×10^{-6} to 1.0×10^{-4}. [47] In breast cancer, this rate may be even higher given the key role of homologous recombination repair deficiency.

As such, CNVs have been found to be a key driver of several Mendelian diseases, including Spinal Muscular Atrophy[48] and Williams syndrome [49]. In cancer, recurrent CNVs have been found in multiple tumor types, often in proteins required for cancer cell survival. [50] Recurrent somatic CNVs have also been linked to prognosis and shown to be a necessary component for tumor proliferation in some cancers[51].

Section 2J – Summary of Current CNV Calling Methods

Reliable calling of copy number variation in exome data faces unique challenges due to the nature of exome sequencing. Compared to whole genome sequencing and SNP Chip sequencing, exome data has far more variation of read depth due to differences in specificity of probes. Also, by sequencing only

the exome, the regions that are captured are not continuous and often have significant gaps.[52] Therefore, reliable copy number variation analysis in exome data has required the development of exome specific algorithms, such as EXCAVATOR, ExomeCNV [44] and CoNIFER [53].

Given this level of variability and the challenges in calling CNV in exome data, CNV calling algorithms specifically optimized for tumor CNVs, such as ExomeCNV and EXCAVATOR, rely on a tumor-normal pair or a case-control. Other pipelines such as CoNIFER use statistical methods optimized for rare germline CNVs to remove artifacts for unpaired calling and thus may not fully account for tumor specific phenomena such as aneuploidy and tumor impurity.

Section 2K – Development of Novel Unpaired CNV Method

In the absence of paired normal controls, we have developed a novel method that creates a simulated normal sample composed of the chromosomes from the same cohort of tumor-only samples with the least likelihood of containing CNV. We have found that this method can achieve comparable results to standard tumor-normal CNV calling.

We observed that in CNV data called with tumor-normal read depth methods, chromosomes without any detectable LOH or allelic imbalance rarely contained copy number variation. Biologically, this is consistent with the shared structural variation mechanisms behind LOH and CNV. [54]

Given the variability of read counts in exome capture, we hypothesized that in the absence of paired normal controls, the next best control should be a control sequenced in the same cohort. The samples would have nearly the same collection techniques, storage conditions, and would have been sequenced with the same machine, the same probes, and possibly even the same flow-cell.

Within an entirely tumor-only cohort, we hypothesized that tumor chromosomes without structural variation (i.e. LOH or CNV) may be able to serve as a surrogate for a normal control.

Therefore, we utilized the specificity of our LOH calling algorithm to estimate the probability of structural rearrangements in all chromosomes. For each chromosome in the tumor-only cohort, our algorithm picks at least three tumor chromosomes with the lowest likelihood of containing structural variation based on our LOH algorithm. We use these “clean” chromosomes to simulate a normal control for the cohort as a whole.

To call CNV in tumor samples, we compared read counts within each exome capture probe between the tumor sample and each of the control samples, all normalized to overall median coverage to account for differences in average sample coverage.

At each region, we selected the median coverage ratio between all control samples to remove possible outliers and undetected structural variation under the observation that read counts in most controls are closely aligned and that any undetected structural variation should be excluded as an outlier.

The combined relative read-count data for each sample, containing estimated log2 coverage ratios for each of the >80,000 regions, was segmented with a CBS algorithm.[55]

All results within each group were analyzed with GISTIC to determine significantly mutated recurrent regions of CNV across the cohort.[56] While we used an in house script based on methods previously described[57], this method can be used to adapt any tumor-normal calling pipeline for tumor-only CNV calling.

Section 2L – Validation of Novel CNV Method

We compared the performance of our tumor-only CNV calling pipeline to tumor-normal paired calling on a cohort of 29 tumor-normal pairs from an ongoing ovarian cancer exome sequencing project. Our tumor-only pipeline demonstrated high specificity as 95% of CNVs called by the unpaired pipeline were true positive calls found in the tumor-normal paired pipeline results. Additionally, we observed an improved signal-to-noise ratio in tumor only calling,

with the preservation of significant CNV calls and a reduction of nonsignificant CNV calls.

In validating our method, we first confirmed that coverage ratios at individual probes and segmentation break points were comparable between tumor-only calling and standard tumor-normal paired calling. We found that the log2 ratios calculated in tumor-only calling and standard paired calling typically had an average difference lower than the standard deviation.

We analyzed the results of both pipelines with GISTIC to test whether tumor-only CNV calling can reliably call significant regions of recurrent CNV within a cohort.

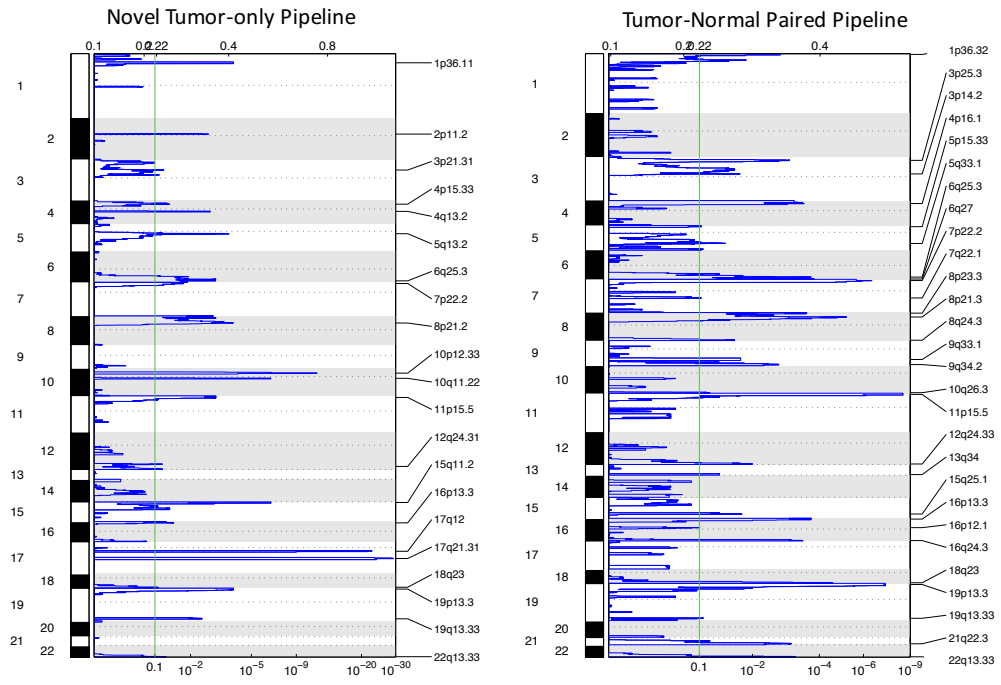

Figure 5: **Comparison of Deletions called by Tumor-Only Pipeline and Tumor-Normal Paired Pipeline** – Recurrent CNV analysis was conducted with GISTIC with segmented CNV output from tumor-only analysis and standard tumor-normal analysis. The X-axis represents q-value and y-axis represents chromosome location.

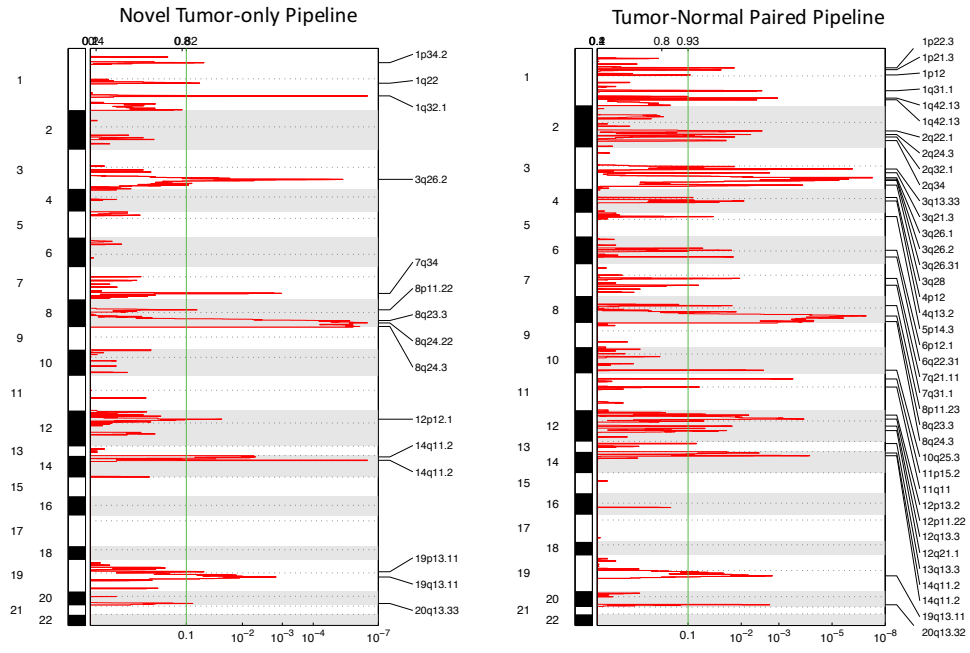

Figure 6: **Comparison of Amplifications called by Tumor-Only Pipeline and Tumor-Normal Paired Pipeline -** Recurrent CNV analysis was conducted with GISTIC with segmented CNV output from tumor-only analysis and standard tumor-normal analysis. The X-axis represents q-value and y-axis represents chromosome location.

Overall, the tumor-only pipeline demonstrated a 95% specificity for recurrent CNVs. False positive calls in tumor-only calling are usually also present in tumor-normal paired results, albeit at a below significant level. Therefore, these false calls appear to arise from differences in q-value rather than true artifacts.

We prioritized high specificity in our analysis given that we are utilizing a novel method. By optimizing specificity, we increase the likelihood that CNV calls in the TNBC cohort are most likely true positive CNVs.

Sensitivity can be difficult to characterize given that that small differences can affect the q-value. We estimate a 68% sensitivity compared to tumor-normal calling for recurrent segments. However, this is likely an underestimate as many of the missed calls were “borderline” calls. Overall, most significant recurrent regions of CNV called in Tumor-Normal paired calling, most were captured with the tumor only pipeline. The specificity of our method for highly significant regions is further supported by our confirmation of all previously reported recurrent CNVs in the most comprehensive exome analysis of TNBC as shown in Section 3F.

The differences in sensitivity may arise from the methods used to call CNVs with paired exome data. Current pipelines call tumor CNVs relative to coverage in the normal samples, therefore accuracy of CNV calling is dependent on variability in tumor samples as well as normal samples. High variation in normal control coverage as well as exome sequencing artifacts can create a false call even in the absence of true tumor CNV.

In our novel simulated control approach, by comparing tumor coverage to a panel of controls samples within the same cohort, we reduce variability in the controls as reflected in the lower number of non-significant calls in the novel unpaired pipeline results compared to standard paired calling.

One significant weakness of this approach is the inability to distinguish germline from somatic CNVs. Germline CNVs are purported to contribute to cancer predisposition and may be present alongside somatic variants.[58] Tumor-paired calling by design will filter out germline variation leaving only somatic variation, unlike our tumor-only calling method.

Chapter 3 – Comprehensive Analysis of Triple Negative Breast Cancer

All analysis was performed by the author as directed by Dr. Lifton

Section 3A - Study Overview

We analyzed a cohort of TNBC tumor-only samples with our novel methods to investigate genomic drivers of chemotherapy response and resistance in TNBC that can provide new insights into the biology of cancer treatment and novel markers to personalize therapy.

Our cohort consists of 32 unmatched TNBC FNA biopsies collected from patients prior to neoadjuvant ACT chemotherapy, a regimen consisting of Adriamycin, cyclophosphamide, and paclitaxel. These patients were selected as representatives of two opposite responses to chemotherapy in TNBC: 17 patients had complete pathologic response and 15 had extensive residual disease.

This study is part of a larger effort to examine chemotherapy response in breast cancer. Samples were obtained from a cohort of patients treated at the MD Anderson Cancer Center and used with permission from Dr. Pusztai. These samples originally part of a larger study analyzing predictors of chemotherapy response in breast cancer overall, including both TNBC and ER+ breast cancer, led by Dr. Lajos Pusztai. (109)

All samples were exome sequenced at high depth to a mean coverage of 163.9, median coverage of 131. (Table 1)

Cohort Sequencing Report	Average
Read Length:	74
Num reads (M):	192.8
Num bases (G):	14.3
Mean coverage:	163.9
Median coverage:	131
PCR duplicates:	8.30%
Multiply mapped:	6.00%
Unmapped:	0.14%
Reads on-target:	65.02%
Bases on-target:	53.91%
Mean error rate:	0.37%
1x target base coverage:	93.65%
2x target base coverage:	92.17%
4x target base coverage:	90.93%
8x target base coverage:	89.60%
10x target base coverage:	89.00%
15x target base coverage:	87.56%
20x target base coverage:	86.10%
30x target base coverage:	82.97%
40x target base coverage:	79.63%
50x target base coverage:	76.17%
100x target base coverage:	58.61%

Table 1**: Exome Sequencing Quality Metrics –** These metrics measure the quality, accuracy, and depth of coverage of sequencing. Most notably, these samples were sequenced to a high average depth of coverage with a mean coverage of 163.9, which is necessary for sensitive somatic mutation detection.

Section 3B - Mutation Count Results

In our initial analysis, we first calculated the overall mutation counts in the tumor samples, including both germline and somatic variants, and compared pCR (pathologic complete response) and RD (Residual Disease) patients. We filtered for rare and novel variants using the ExAC database which contains germline variants from 60,706 individuals.[59] Given the size and diversity of the ExAC database, these criteria should filter out most common germline variants.

Overall, mutation counts were similar between both cohorts. Filtering for novel variants leads to an increase of Nonsilent:Silent (NS/S) ratio, a measure of selection pressure. As driver mutations are expected to be non-silent, this increase in NS/S ratio in novel mutations compared to rare mutations suggests an enrichment of driver mutations (Table 2). [60]

Based on data from the TCGA Breast Cancer study, we expected to detect approximately 60 somatic mutations per sample. [61] Given the average novel mutation count of 217 (Table 3), it appears that while filtering for novel mutations

may enrich somatic mutations within the results, most the filtered mutations are still likely germline.

Rare Variant (ExAC Freq < 0.001)	pCR (n=17)	RD (n=15)	Total
Mean Mutation Count	723	640	684
Mean Non-Silent Mutation Count	471	424	449
Overall Nonsilent:Silent Ratio	2.12	2.08	2.10

Table 2**: Mutation Counts and NS/S for Rare Variants** (ExAC Freq <0.001)

Novel Variants	pCR (n=17)	RD (n=15)	Total
Mean Mutation Count	217	210	214
Mean Non-Silent Mutation Count	163	157	160
Overall Nonsilent:Silent Ratio	3.03	2.90	2.97

Table 3**: Mutation Counts and NS/S for Novel Variants**

Section 3C - Somatic Point Mutation Results

We combined two strategies to analyze somatic point mutations. We first analyzed novel mutations in a list of 187 known cancer driver genes as published in the TCGA Pan-Cancer analysis and Vogelstein et al. [62] [63] This approach focuses our analysis from the 20,000 protein coding genes to 187 genes with the highest probability of containing somatic driver mutations. We restricted our analysis to ExAC novel mutations to exclude known germline variants.

Gene	Alleles	Damaging	Gene Size
TP53	12	12	1222
NOTCH1	4	3	7804
PIK3CA	3	3	3287
ATRX	3	2	7619
ARID1B	2	1	6830
RB1	2	2	2895

Table 4: **Cancer Gene Mutation Analysis (n=32)** - Novel mutations were counted in each cancer gene from the Vogelstein and TCGA Pan-Cancer lists. Damaging is defined as CADD score > 10, which indicates that the mutation is within the most deleterious 10% of possible SNVs. (Kircher, 2013) Gene Size measures the size of the largest transcript of each gene.

Within the top genes found in this analysis, TP53 (12/32), PIK3CA (3/32), ARID1A (2/32), and RB1 (2/32) have been reported as significantly mutated genes in breast cancer[61]. We also found multiple mutations in two genes which have not been previously reported as significantly mutated in breast cancer: NOTCH1 (4/32), APRX (3/32).

In addition to analyzing novel mutations in cancer genes, we conducted a comprehensive analysis of point mutations across all genes with the Bayesian Algorithm. We also analyzed rare mutations occurring in regions of LOH, which were not included in the Bayesian analysis, as well as gene size to create a relative ranking of top candidates (Table 5).

Gene	Size	Somatic	Incl. LOH
TP53	1222	3	21
HELZ2	7,827	2	4
FBXL14	1265	2	3
WDR90	5411	2	5
COL18A1	4719	3	4
SEC14L3	1251	2	2
IDS	1689	2	2
VEZT	2388	2	2
DLGAP2	2972	2	2
CCDC141	4689	2	2
ADAMTS20	5889	2	3
SCAF4	3524	2	2

Table 5**: Bayesian Algorithm Somatic Mutations (n=32) -** Genes with multiple somatic mutations as identified with the Bayesian Algorithm. Size measures the length of the transcript. The "Incl. LOH" column indicates number of somatic mutations in addition to novel variants found in regions of LOH.

Consistent with previous studies, TP53 had the highest somatic burden with most other somatic mutations spread amongst many genes at relatively low abundance. Our TP53 mutation count in the Bayesian Algorithm results was lower than expected, likely due to the fact that TP53 occurs in a region of LOH in most samples and was therefore excluded from our point mutation analysis. Once these LOH variants were included in our integrated analysis, we managed to retain most TP53 mutations.

Section 3D - Enrichment of EMT Related Genes

Within the top genes identified with our Bayesian Algorithm, we found four candidate driver genes with known functions that may play key roles in Endothelial to Mesenchymal Transition (EMT). EMT, a fundamental process in tumor progression and a hallmark of cancer, describes the loss of adhesion in

epithelial cells to their neighboring cells, which enables invasion and metastasis[64].

ADAMTS20, a secreted matrix metalloproteinase, contained 2 novel damaging somatic mutations in TSP domains, which control substrate specificity. While this family of proteins have known roles in breaking down the extracellular matrix [65], ADAMTS20 has not been previously implicated in cancer. COL18A1, which had two damaging somatic mutations, forms a precursor of endothelin, a promotor of EMT and angiogenesis.[66] VEZT, in which we found 2 damaging mutations, is an adherens junction, one of the 3 key cell-cell anchoring protein types that become disassembled in EMT.[67] FBXL14, which had two somatic mutations a single patient, functions as an ubiquitin ligase for SNAIL1, a central regulator of EMT in breast cancer.[68] NOTCH1, a protein shown to trigger EMT in breast cancer, contains three mutations identified in cancer gene analysis[69]. Of note, the NOTCH1 variants are of uncertain somatic status as they were found in regions of LOH. However, given that it's a known cancer gene, we've included it in this analysis.

Gene	Alleles	Non-Silent	Damaging	Gene Size
NOTCH1	3	3	3	7804
FBXL14	2	2	2	1265
COL18A1	4	4	2	4719
VEZT	2	2	2	2388
ADAMTS20	2	2	2	5889

Table 6: **EMT Related Genes -** Analysis of NS/S and Damaging Alleles of possible EMT somatic mutations show that these mutations are nearly universally non-silent and damaging. (Damaging is defined as CADD > 10)

In total, 31% of our cohort contained at least one somatic mutation in genes possibly involved in EMT. Based on enrichment of nonsilent mutations and damaging mutations, there is a high likelihood that these genes may contain driver mutations that have undergone positive selection[70]. Within this group of genes, we found a 100% NS rate and a 92% damaging rate which is significantly higher than expected based on chance alone. For novel mutations, we expect a NS/S ratio of 3 based on our mutation count data, and 10% of these to meet our baseline criteria of damaging (CADD > 10) without selection.

Section 3E - LOH Results

Overall, pCR patients had a higher prevalence of LOH compared to RD patients. Since LOH occurs at the chromosome level, we quantified LOH burden by comparing the number of patients with any amount of LOH in a chromosome between pCR and RD subgroups. (Table 7)

Chromosome	pCR (n=17)	RD (n=15)	p-value
1	71%	60%	0.71
2	71%	40%	0.15
3	88%	80%	0.65
4	82%	73%	0.68
5	76%	60%	0.45
6	82%	53%	0.13
7	82%	40%	0.03
8	76%	80%	1.00
9	88%	80%	0.65
10	65%	53%	0.72
11	82%	33%	0.01
12	65%	60%	1.00
13	76%	47%	0.14
14	71%	53%	0.47
15	76%	53%	0.27
16	71%	47%	0.28
17	82%	67%	0.42
18	53%	47%	1.00
19	41%	33%	0.73
20	59%	33%	0.18
21	47%	20%	0.15
22	47%	33%	0.49
24	71%	40%	0.15
Overall	**68%**	**43%**	**6.45E-12**

Table 7: **Prevalence of LOH -** For each chromosome, we calculated the percent of patients in clinical subgroups group with LOH of any length in the chromosome. P-value was calculated with a two tailed Fisher Exact Test.

In 22/23 chromosomes (including the X chromosome), pCR patients had a greater prevalence of LOH than RD patients. Chromosomes 11 and 7 showed the highest difference in LOH burden between subgroups, with a relative LOH prevalence of 82% in pCR patients vs 33% in RD patients in chromosome 11(p=0.01) and 82% vs 40% in chromosome 7 (p=0.03).

To quantify overall differences in LOH prevalence, we compared the odds for a chromosome in a clinical subgroup to contain any length of LOH. We found that

68% of chromosomes in pCR patients had detectable LOH, compared to 43% of chromosomes in RD patients. ($p < 0.001$) We also calculated average proportion of the exome exhibiting LOH between pCR and RD patients and found a similarly significantly difference.

To detect recurrent LOH regions, we compared LOH levels across both subgroups and between subgroups. As expected, both 17p and 17q had LOH across most samples as BRCA1 and TP53 are contained in these segments. Patterns of LOH appear to be similar between pCR and RD patients, with the exception of a few regions where pCR demonstrates a markedly higher LOH prevalence than RD.

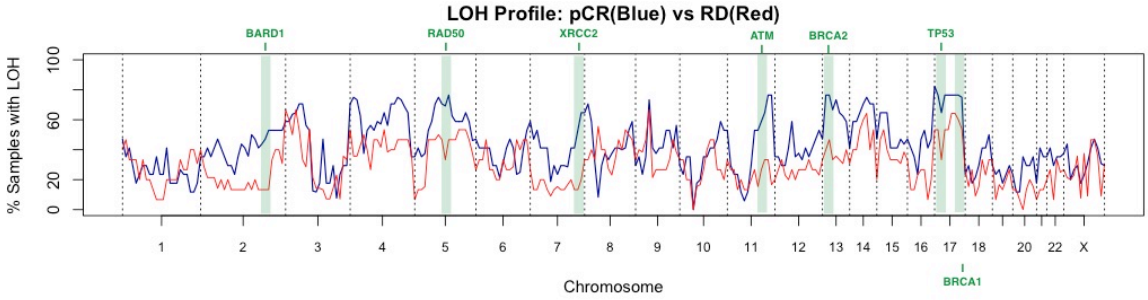

Figure 7: **LOH Profile** Relative abundance of LOH is plotted for both pCR (Responders) and RD (Non-responders) subgroups across all chromosomes. Areas of highest LOH prevalence difference correspond to homologous recombination repair gene loci. Key mediators of homologous recombination repair, as well as TP53, are highlighted in green.

We cross-referenced these regions with lists of significantly mutated cancer genes and found that each of these 5 regions (11q, 7q, 13p, 5q, and 2q) contains a key mediator of homologous recombination repair. (Figure 7)

The recurrent region in 11q contains ATM and MLL1, both of which have been linked to breast cancer risk. In the presence of double stranded DNA breaks, ATM delays the cell cycle and activates proteins responsible for repairing double-stranded DNA breaks.

The region within 2q with the greatest difference between pCR and RD patients contains BARD1 which acts as a partner for BRCA1, and is believed to be essential for BRCA1 activation of DNA damage response mediators, including ATM, MSH6, and MSH2. Additionally, similar to BRCA1, loss of BARD1 leads to genomic instability in cells. [71]

Region 7q corresponds to XRCC2 and MLL3, both of which are significantly mutated genes in breast cancer. [72] [73]. Similar to BRCA1 and ATM, XRCC2 is a key mediator of homologous recombination repair, [74]and has been shown to repair DNA damage from crosslinking. [75]

In chromosome 5, the region with greatest difference contains RAD50, a key mediator of homologous recombination repair that functions to sense double stranded breaks in DNA.[76]

Section 3F - CNV Result Overview

Our unpaired CNV analysis reconfirmed key regions of CNV that have been reported in multiple previous studies of TNBC and also found novel regions of

recurrent CNV across both cohorts. Additionally, we detected significant recurrent regions of CNV unique to pCR patients, suggesting the presence of possible markers and drivers of chemotherapy response in these regions. (Tables 8 and 9)

We first compared our data to CNV data from the TCGA Breast Cancer study and found that all 5 characteristic CNVs of TNBC reported in the TCGA study, 1q gain, 10p gain, 5q loss, 8p loss, and focal Myc amplification at 8q24, were also called as significant CNVs in our cohort as well. [61]

Our CNV calls also included four additional TNBC specific CNVs reported in a second comprehensive study of TNBC including 6p gain, 4q loss, 15q22 loss, and 17q25 focal gain. [77]

Detecting all TCGA reported CNVs for TNBC as significantly recurrent in our cohort, despite having a smaller sample size, further validates the sensitivity of our novel tumor-only CNV calling method.

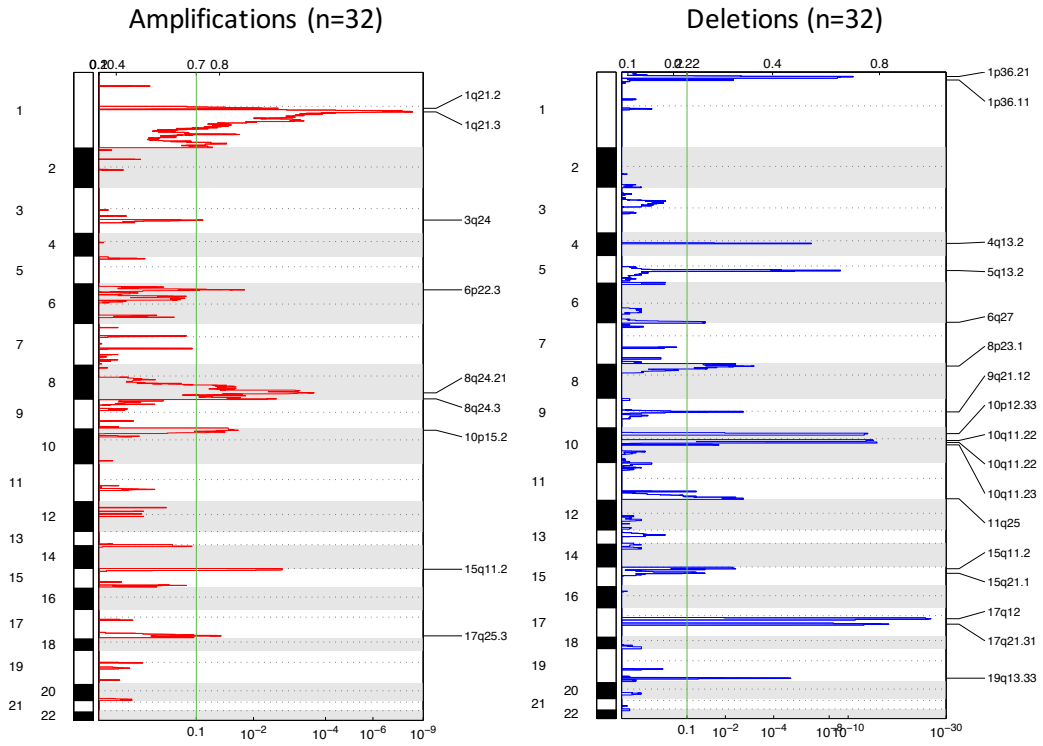

Figure 8: **GISTIC Analysis of Deletions and Amplifications Across all TNBC Samples –** Recurrent Amplifications and Deletions from cohort-wide CNV analysis as identified by GISTIC based on our Novel Unpaired CNV calling method. The X-axis represents q-value and y-axis represents chromosome location.

Significant Amplifications (n=32)		
Region	**q-value**	**Genes in Region**
1q21	5.20E-09	MCL1,PI4KB,CTSS, CTSK, MUC1
8q24	2.47E-04	**Myc**
15q11	2.10E-03	**NDN**
6p22	1.50E-02	E2FB, CDKAL1
10p15	2.90E-02	**GATA3**
17q25	4.10E-02	RECQL5

Table 8: **Cohort-wide Significant Amplifications –** Purported genes were identified with GISTIC annotation, cross-reference with known breast cancer genes in the region. Bolded genes are known breast cancer genes.

Significant Deletions (n=32)		
Region	**q-value**	**Genes in Region**
17q21	1.19E-16	**BRCA1**
10q11	1.18E-14	BMP9, BMP2
1p36	2.48E-11	CHD5
5q13	2.90E-02	GTF2H2
4q13	3.52E-07	UGT2B15, UGT2B17
19q13	1.02E-05	
8p23	8.56E-04	DLGAP1,CSMD1
9q21	2.38E-03	
11q24.2	2.40E-03	**CHEK1**
15q11	5.40E-03	**NDN**
6q25	3.81E-02	GTF2H5
15q21	3.81E-02	**RAD51**

Table 9**: Cohort-wide Significant Deletions -** Purported genes were identified with GISTIC annotation, cross-reference with known breast cancer genes in the region. Bolded genes are known breast cancer genes.

Section 3G - Novel TNBC Associated CNVs

In addition to the regions reported in TCGA and Begamaschi, we have found several novel CNVs that have not been previously linked to TNBC or pCR/RD status. Many of these regions have been associated with breast cancer in various studies, which supports their role in breast cancer pathogenesis.

LOH in 1p36 and 6q25 have been reported as correlated with poorer prognosis in breast cancer [78] [79] and LOH in region 11q24 and amplifications in 15q have been associated with general early onset breast cancer, both of which are clinical features of TNBC. [80] [81]

19q13 has been described as a Luminal B associated CNV in Begamaschi et al and linked to poor response to endocrine therapy in Luminal B cancer, a characteristic expected in TNBC patients. However, it has not been previously described as a TNBC specific CNV. [77] [82]

Section 3H - 10q11 Correlation with EMT and Metastasis

10q11 deletion, a highly significant CNV in our cohort(q=1.18E-14), has been previously described as a CNV preferentially expressed in brain metastasess. We've found that key genes in this locus suggest that it may be a marker for EMT.

In a study comparing breast tumors and distal metastases to the brain, 10q11 was found to be the most differentially expressed CNV with a 60% deletion rate in metastases and 3% deletion rate in primary tumors. [83]

In our analysis, which has a higher resolution than previous studies using microsatellite AI analysis, the peak of this region appears to tightly correspond to GDF2 (BMP-9) and GDF10(BMP-3B), two members of the TGF-beta super family of proteins which are the central mediators of EMT[84]. Multiple *in vitro* and *in vivo* functional studies have demonstrated that GDF2 inhibits proliferation and invasion of breast cancer, consistent with the loci's previously discovered correlation with metastasis and our finding of EMT-related point mutation enrichment.[85] [86]

Section 3I - pCR/RD Specific CNVs

To find potential markers and drivers of chemotherapy response in TNBC, we analyzed the pCR and RD groups separately and found that 5q (peak at 5q13) deletions, 8p23 deletions and a broad 6p amplification reached significance uniquely in the pCR cohort. (Figures 9 and 10)

While 5q loss is reported as a general TNBC characteristic CNV in the TCGA study, we found that this CNV was exclusively enriched in the pCR group. This region contains RAD17, RAD50, RAP80, key mediators of BRCA1 dependent DNA repair, consistent with our finding of strong LOH in loci corresponding to homologous recombination repair genes[87].

Our finding of recurrent 6p22 amplification only in pCR patients correlates with a key clinical study of cisplatin response in TNBC which found that the overexpression of E2F3, which is found in the boundaries of this region, is as predictive of cisplatin response as BRCA1 expression[88].

8p23 deletion (q=0.00086) has shown to be a marker for high tumor stage and grade[89] and invasiveness[90] in general studies of breast tumors, which is consistent with our finding of its enrichment in TNBC overall. However, there are no well characterized tumor suppressors ascribed to 8p23 deletion.

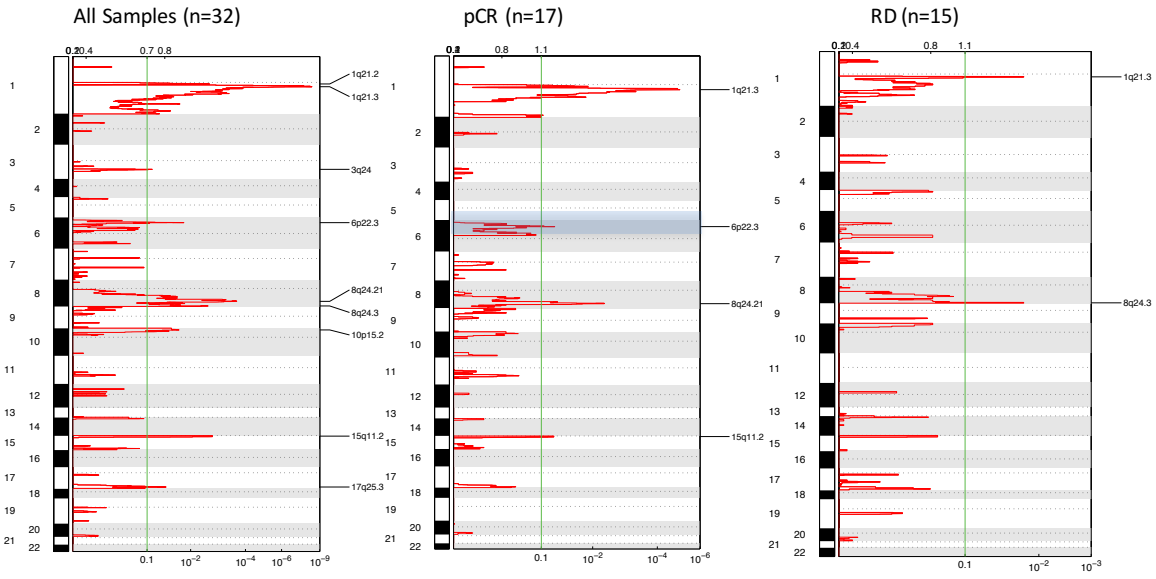

Figure 9: **Comparison of Amplifications:** 6p22 (Highlighted in Blue) is significantly recurrent in pCR patients and nearly absent in RD patients. Analysis was conducted with GISTIC. Both groups and the combined group were analyzed separately. The X-axis represents q-value and y-axis represents chromosome location.

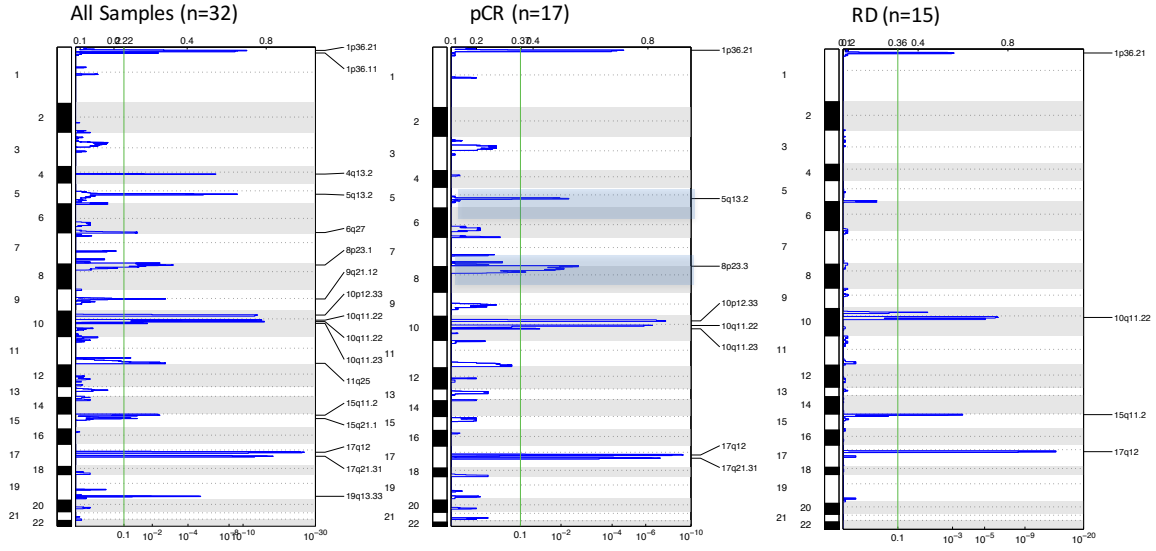

Figure 10: **Comparison of Deletions:** 5q13 and 8p23 regions (Highlighted in Blue) were only found in significance in pCR Patients. Analysis was conducted with GISTIC. Both groups and the combined group were analyzed separately. The X-axis represents q-value and y-axis represents chromosome location.

Section 3J - Integrated Analysis of Breast and Ovarian Cancer Genes

We integrated copy number variation, LOH, and point mutation data to analyze how the full spectrum of somatic variation affects key genes and pathways in breast cancer[91].

To get an initial overview of the mutation landscape of TNBC, we first analyzed a list of 65 genes most strongly associated with breast and ovarian cancer found in the BROCA list, as this list contained both predisposition genes and possible somatic driver genes as well for breast and ovarian cancer [92].

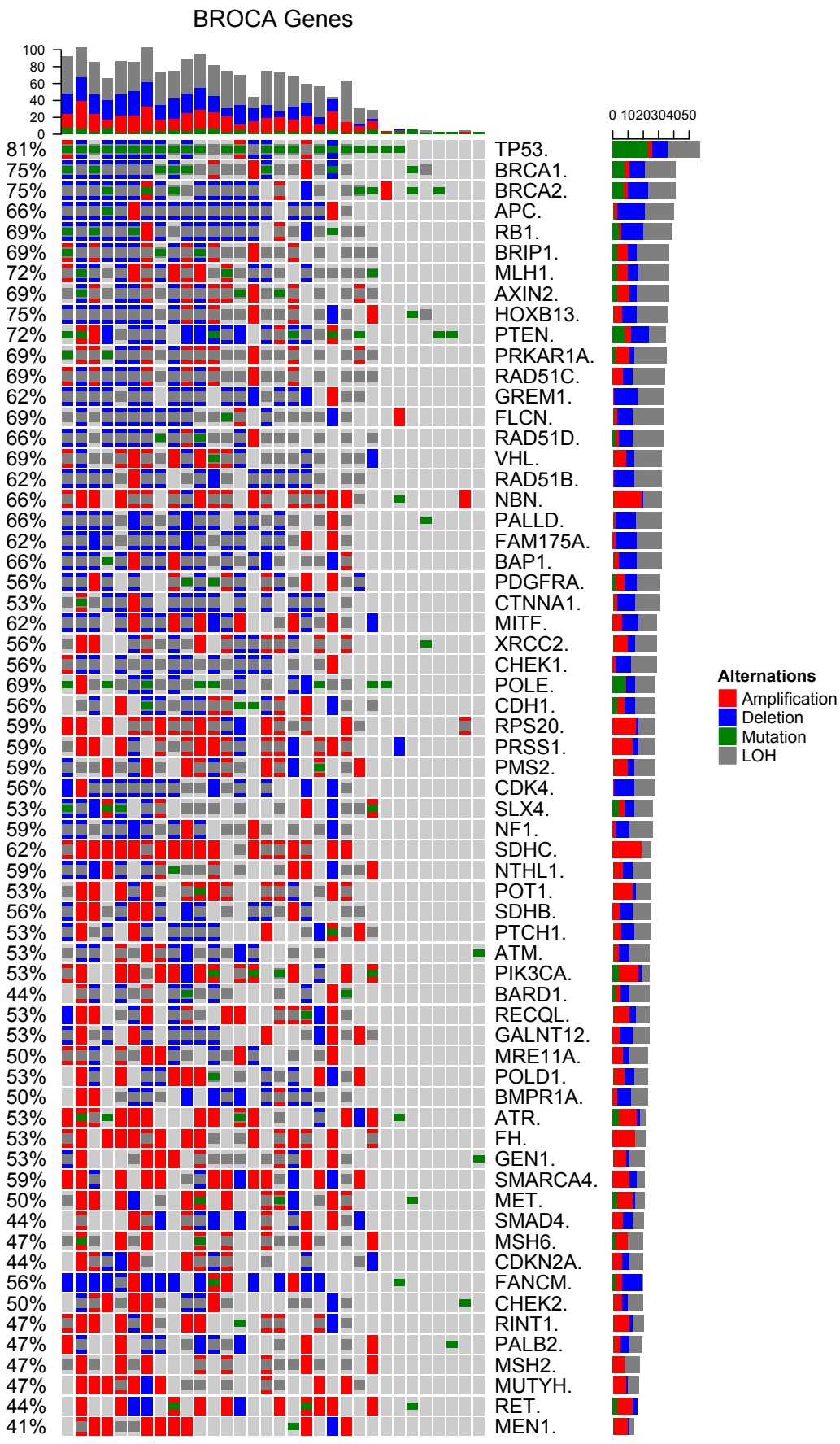
BROCA Genes
100
80
60
40
20
0
0 1020304050
81% TP53.
75% BRCA1.
75% BRCA2.
66% APC.
69% RB1.
69% BRIP1.
72% MLH1.
69% AXIN2.
75% HOXB13.
72% PTEN.
69% PRKAR1A.
69% RAD51C.
62% GREM1.
69% FLCN.
66% RAD51D.
69% VHL.
62% RAD51B.
66% NBN.
66% PALLD.
62% FAM175A.
66% BAP1.
56% PDGFRA.
53% CTNNA1.
62% MITF.
56% XRCC2.
56% CHEK1.
69% POLE.
56% CDH1.
59% RPS20.
59% PRSS1.
59% PMS2.
56% CDK4.
53% SLX4.
59% NF1.
62% SDHC.
59% NTHL1.
53% POT1.
56% SDHB.
53% PTCH1.
53% ATM.
53% PIK3CA.
44% BARD1.
53% RECQL.
53% GALNT12.
50% MRE11A.
53% POLD1.
50% BMPR1A.
53% ATR.
53% FH.
53% GEN1.
59% SMARCA4.
50% MET.
44% SMAD4.
47% MSH6.
44% CDKN2A.
56% FANCM.
50% CHEK2.
47% RINT1.
47% PALB2.
47% MSH2.
47% MUTYH.
44% RET.
41% MEN1.
Alternations
Amplification
Deletion
Mutation
LOH

Figure 11: **BROCA Cancer Predisposition –** Integrated analysis of the entire cohort with BROCA breast/ovarian cancer predisposition genes demonstrates consistent patterns of copy number variation across samples. These patterns appear to match closely with previously published CNV patterns in breast cancer. Each column in this analysis represents a single sample within the cohort, columns are ordered based on mutational burden. Each line represents a known breast and ovarian cancer predisposition gene and is ordered based on prevalence within the cohort. Figure was generated with Oncoprint.

As expected, TP53, BRCA1, and BRCA2 have the highest mutation burden, including CNVs and LOH, confirming that structural variations in this cohort follow expected patterns of mutation. (Figure 11)

As a further validation step for this analysis, we confirmed that our results are in line with previously reported data. Our CNV patterns appear to align well with the reported CNVs in the TCGA Breast Cancer study, with 6/6 reported CNV genes (TP53, PTEN, RB1, MLL3, PIK3CA, and MAP2K4) showing similar patterns to our data. In addition to the above 6 genes, BRCA1 was significantly deleted in the TCGA study (p=3.14E-09), consistent with our data.

We found a high prevalence and consistency of CNV and LOH across known breast cancer genes. (Figure 11) Kinases such as PIK3CA and ATR are consistently amplified among samples with significant CNV while tumor suppressors such as BRCA1, BRCA2, and RB1 are consistently deleted suggesting that CNV and LOH in these genes may play a role in tumor progression by affecting these genes.

Interestingly, there appears a distinct divide between samples with high CNV and LOH and samples with very little. This phenotype correlates strongly with

the presence of a TP53 mutation: 88% of high CNV samples contain a TP53 mutation compared to 25% of low CNV samples. (p=0.0021)

Section 3K - Test of TCGA Breast Cancer Somatic Mutations

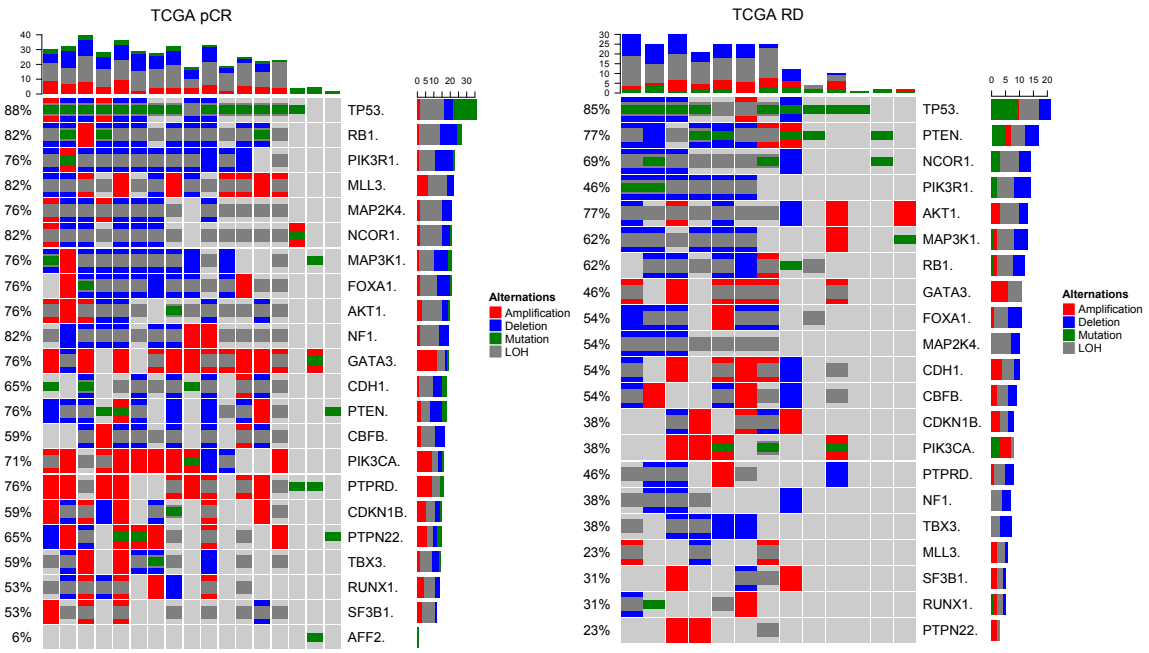

Figure 13: **Integrated Analysis of Known Breast Cancer Genes** – Known significantly mutated genes in breast cancer were analyzed with integrated analysis. CNVs followed consistent and expected patterns based on function. Additionally, pCR patients have a overall higher burden of CNV and LOH.

In order to find differences between pCR and RD, we applied our integrated analysis to significantly mutated genes in the TCGA breast cancer study as we expect that much of the pCR/RD phenotype is driven by a tumor's unique evolution rather than a patient's genetic predisposition.

Among 23 genes identified as significantly mutated in breast cancer, TP53 and RB1, and PTEN showed consistently high mutation rates and deletion status between cohorts consistent with their known driver status in breast cancer.

MLL3 showed the highest difference in mutation rates between pCR and RD, with 82% of pCR patients exhibiting CNV or LOH at the locus vs 23% of RD patients, however this may be related to its proximity to XRCC2, a homologous recombination repair gene. Other differentially mutated loci include GATA3 (76% vs 46%), which has been previously reported as a significant gene Luminal A/B subtypes, and NF1 (82% vs 38%).

Interestingly, there appears to be a clear correlation between overall CNV and LOH burden and pCR/RD status. We further investigated this by focusing our analysis on homologous recombination repair genes, the most likely mechanism responsible for structural mutation burden.

Section 3L - Test of Homologous Recombination Repair Mutation Burden

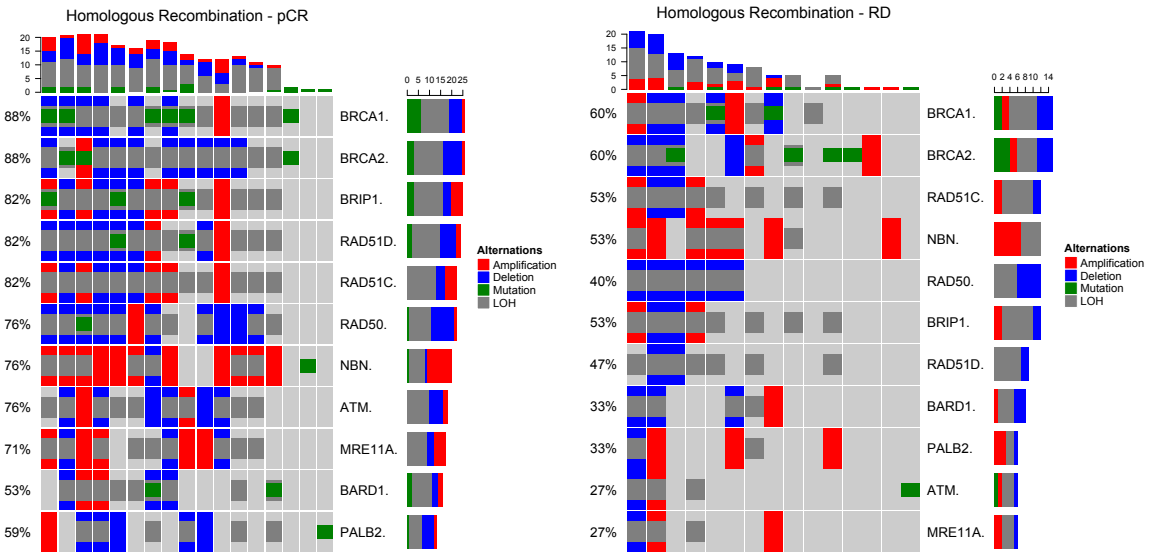

Figure 14: **Integrated Analysis of Homologous Recombination Repair Genes**– Known significantly mutated genes in breast cancer were analyzed with integrated analysis. CNVs followed consistent and expected patterns based on function. Additionally, pCR patients have a overall higher burden of CNV and LOH. (pCR – Pathologic Complete Response to Chemotherapy, RD – Residual Disease)

Based on the key role that Homologous Recombination Repair plays in breast cancer pathogenesis and in creating genomic instability, we hypothesized that genomic aberrations may be enriched in HRR genes.

Compared to our more general analysis of likely somatic genes identified by TCGA (Figure 13), genes involved in HRR demonstrate the clearest difference between pCR and RD (Figure 14). Not only is mutation burden in these genes correlated to pCR and RD status, there appears to be two distinct groups of patients that segregate to each clinical group.

Section 3M - Discriminant Analysis

Based on our findings in integrated analysis, we retrospectively evaluated whether this can be used to predict pCR and RD status prior to therapy.

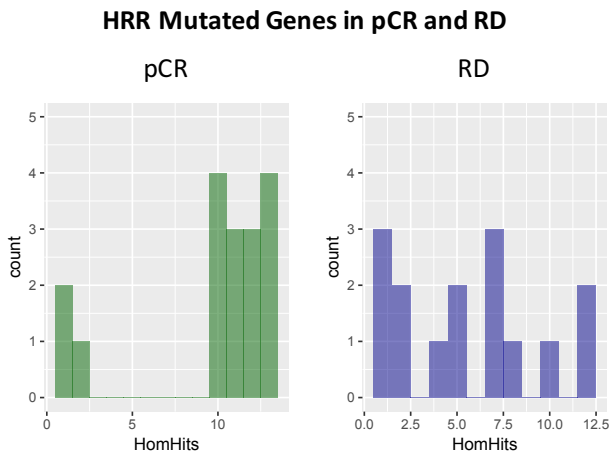

Figure 15**: Distribution of HRR Mutation Counts** In nearly all pCR patients (Responders), the majority of homologous recombination repair genes are in regions of structural variations, whereas this is the case in only a minority of RD patients (Non-responders).

We found that 82% of pCR patients have mutations across most HRR genes, as defined by >75% of HRR genes mutated, compared to 20% of RD patients. In addition, pCR patients demonstrate a clear cutoff at 75% of genes affected, with a few outliers possibly explained by lack of TP53 point mutations. With this metric, we can predict chemotherapy response in 82% of pCR patients and 80% of RD patients. (OR=18.7, 95% CI= 3.2 to 110.3, p=0.0012).

Currently, there are few available markers for chemotherapy response in TNBC and for breast cancer in general. BRCA1/2 point mutations have been proposed and tested as markers for chemotherapy response [93] [94]and homologous recombination repair deficiency. However, as shown in our analysis below,

BRCA point mutations alone do not correlate strongly with pCR and RD status in our cohort.

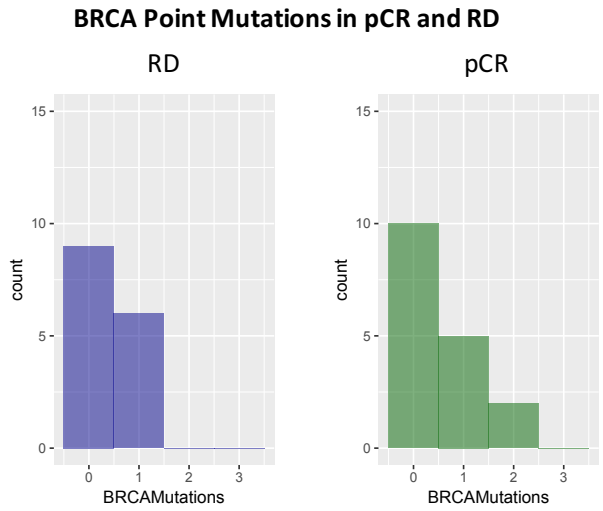

Figure 16: **Distribution of BRCA Point Mutation Counts** – BRCA1/2 point mutations, which have been previously described as predictors of chemotherapy response, show a weak correlation to chemotherapy response in this cohort.

Chapter 4 – Discussion

Patients and clinicians are faced with an ever growing array of treatment options. Similar to HER2 status and ER/PR status, a reliable marker for response to specific therapies in TNBC can help direct patients to the most effective treatment option. With better prediction of response to ACT chemotherapy, patients likely to respond can prioritize ACT chemotherapy, while other patients can consider possibly more effective alternatives.

In our study, we found that extensive CNVs and LOH significantly correlates with chemotherapy response, particularly in regions containing homologous recombination repair (HRR) genes. This suggests HRR may play a key role in

chemotherapy response in this cohort and supports the utility of CNV and LOH as markers for this process.

While there is a general consensus about the central role of HRR in breast cancer[95], and chemotherapy response [96] [97], no clear marker has been found beyond BRCA1/2 deficiency despite many efforts [98] [99]. A clinical test using three broad measures of genomic instability, including general LOH burden, has shown promise in clinical trials[100]. However, sensitivity and specificity is not as strong nor does this test utilize CNV data and specific cancer gene loci, which we've found to be more predictive than LOH alone.

Based on our results, we propose that structural variation in regions containing Homologous Recombination Repair (HRR) genes may serve as a strong predictor of response to chemotherapy and targeted PARP inhibitors. Additionally, this may be evidence that structural variation in the form of CNV and LOH may contribute to HRR deficiency.

Section 4A - LOH and CNVs as Driver Mutations

There are six original Hallmarks of Cancer, key functions tumors must obtain to proliferate, with several more proposed[28]. However, the somatic point mutations found in this cohort do not come close to accounting for these necessary changes. Given the high mutation rate of CNV and LOH in our study, we

theorized that structural variation may be the missing link driving these required transformations.

As demonstrated in our integrated analysis, not only do CNVs and LOH follow expected patterns of mutation (i.e. amplification of kinases, deletion of tumor suppressors), they provide the strongest correlation to response out of any detected genetic feature suggesting their driving role in tumor evolution.

Section 4B - LOH and CNV as Predictors of Chemotherapy Response

However, this begs the question of why tumors with this powerful ability to amplify kinases and delete suppressors paradoxically respond better to chemotherapy.

The answer likely lies in the mechanism of ACT chemotherapy. High LOH and CNV burden has been shown to reflect higher genomic instability[101]. When already fragile chromosomes are treated with Adriamycin and cyclophosphamide, which function by creating damage in DNA to hindering DNA replication, the cells may be more likely to succumb to apoptosis hence the correlation LOH to chemo-sensitivity. This is consistent with evidence that BRCA1 mutated tumors are more sensitive to DNA damaging agents, as found in ACT chemotherapy. [102]

Section 4C - Homologous Recombination Repair Deficiency in Breast Cancer

Previous studies have found LOH to correlate with BRCA1/2 mutations, which are central genes in HRR[103 104]. LOH has also been shown to reflect homologous recombination repair deficiency in ovarian cancer[105]. Given the association between HRR deficiency and genomic instability, HRR is a likely driver of the abundant CNVs and LOH found in our cohort.

With the significant role HRR plays in breast cancer tumor evolution and chemo-sensitivity, there has been recent significant interest in understanding HHR deficiency beyond BRCA1/2 mutations[95]. Somatic and germline point mutations in HRR genes have been extensively studied as possible markers, including in a recent study of 560 breast cancer whole genome sequences, however no predictive patterns in point mutations have been found. [106]

Our data suggests that structural variation in HHR loci may be markers of genomic instability in these genes driving HHR deficiency. We hypothesize that defective HRR function may confer a survival advantage to cells under the intense selection pressure in clonal expansion by allowing cells to amplify kinases and delete suppressors as needed[107]. Therefore, the advantage of reduced HRR function for tumor mutagenesis may lead to selection for structural instability and deletions in HHR genes, as seen in our data.

Section 4D - Development of PARP Inhibitors

Recognition of homologous recombination repair deficiency is particularly important in the context of PARP inhibitors in breast cancer, a new class of targeted therapies in development.

PARP inhibitors have shown significant activity within BRCA1/2 deficient patients, however, <5% of breast cancer patients fall under this category. Across all breast cancer patients, this effect disappears. As such, clinical development has been delayed in favor of applications in ovarian cancer which has a 10% prevalence of BRCA1/2 deficiency[108].

A more sensitive and specific marker for homologous recombination repair deficiency will allow for more patients to become eligible for treatment and thus facilitate the continuation of PARP inhibitor clinical development in breast cancer.

Our data shows that LOH and CNV provide stronger correlations with chemotherapy response than BRCA1/2 mutations alone, the current standard biomarker for PARP inhibitor response. As these variations are clustered around HRR genes, our metric of overall mutation burden at HRR gene loci may potentially serve as an improved biomarker for HRR deficiency and PARP inhibitor response in clinical development and therapy.

Section 4E - Conclusion

In conclusion, our data provides evidence that LOH and CNV burden, particularly in genes of the homologous recombination repair pathway, predicts chemotherapy response and may contribute to the development of homologous recombination repair deficiency.

Additionally, as one of the first comprehensive analyses of unpaired tumor exome samples, these data demonstrate the potential insights obtainable in previously unanalyzable unpaired tumor exome data with our novel analytical pipeline.

Bibliography

1. National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch. Surveillance, Epidemiology, and End Results (SEER) Program [Internet]. Available from: https://seer.cancer.gov/statfacts/html/breast.html
2. Osborne CK. Tamoxifen in the treatment of breast cancer. N Engl J Med 1998;339(22):1609–18.
3. Osborne CK, Yochmowitz MG, Knight WA 3rd, McGuire WL. The value of estrogen and progesterone receptors in the treatment of breast cancer. Cancer 1980;46(12 Suppl):2884–8.
4. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. N Engl J Med 2005;353(16):1659–72.
5. Dent R, Trudeau M, Pritchard KI, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res Off J Am Assoc Cancer Res 2007;13(15 Pt 1):4429–34.

6. Turner NC, Reis-Filho JS. Tackling the diversity of triple-negative breast cancer. Clin Cancer Res Off J Am Assoc Cancer Res 2013;19(23):6380–8.

7. Shah SP, Roth A, Goya R, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012;486(7403):395–9.

8. Liedtke C, Mazouni C, Hess KR, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J Clin Oncol Off J Am Soc Clin Oncol 2008;26(8):1275–81.

9. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977;74(12):5463–7.

10. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. Cell 2013;155(1):27–38.

11. Sanders SJ, Murtha MT, Gupta AR, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 2012;485(7397):237–41.

12. Brennan CW, Verhaak RGW, McKenna A, et al. The somatic genomic landscape of glioblastoma. Cell 2013;155(2):462–77.

13. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010;11(1):31–46.

14. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J. Structural Mechanism for STI-571 Inhibition of Abelson Tyrosine Kinase. Science 2000;289(5486):1938–42.

15. Honeyman JN, Simon EP, Robine N, et al. Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. Science 2014;343(6174):1010–4.

16. Witkowski L, Carrot-Zhang J, Albrecht S, et al. Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. Nat Genet 2014;46(5):438–43.

17. Lynch TJ, Bell DW, Sordella R, et al. Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. N Engl J Med 2004;350(21):2129–39.

18. Sasaki T, Rodig SJ, Chirieac LR, Jänne PA. The Biology and Treatment of EML4-ALK Non-Small Cell Lung Cancer. Eur J Cancer Oxf Engl 1990 2010;46(10):1773–80.

19. McNeil C. NCI-MATCH launch highlights new trial design in precision-medicine era. J Natl Cancer Inst 2015;107(7).

20. Nazarian R, Shi H, Wang Q, et al. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. Nature 2010;468(7326):973–7.

21. Wagle N, Emery C, Berger MF, et al. Dissecting Therapeutic Resistance to RAF Inhibition in Melanoma by Tumor Genomic Profiling. J Clin Oncol 2011;29(22):3085–96.

22. Shi H, Hugo W, Kong X, et al. Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. Cancer Discov 2014;4(1):80–93.

23. Zaretsky JM, Garcia-Diaz A, Shin DS, et al. Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. N Engl J Med 2016;375(9):819–29.

24. Trojani M, Contesso G, Coindre JM, et al. Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system. Int J Cancer 1984;33(1):37–42.

25. Watson PH, Snell L, Parisien M. The NCIC-Manitoba Breast Tumor Bank: a resource for applied cancer research. CMAJ Can Med Assoc J 1996;155(3):281–3.

26. Chu T-Y, Hwang K-S, Yu M-H, Lee H-S, Lai H-C, Liu J-Y. A research-based tumor tissue bank of gynecologic oncology: characteristics of nucleic acids extracted from normal and tumor tissues from different sites. Int J Gynecol Cancer Off J Int Gynecol Cancer Soc 2002;12(2):171–6.

27. Meric-Bernstam F, Johnson A, Holla V, et al. A decision support framework for genomically informed investigational cancer therapy. J Natl Cancer Inst 2015;107(7).

28. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144(5):646–74.

29. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31(3):213–9.

30. Nik-Zainal S, Van Loo P, Wedge DC, et al. The life history of 21 breast cancers. Cell 2012;149(5):994–1007.

31. Whibley C, Pharoah PDP, Hollstein M. p53 polymorphisms: cancer implications. Nat Rev Cancer 2009;9(2):95–107.

32. Marsh A, Spurdle AB, Turner BC, et al. The intronic G13964C variant in p53 is not a high-risk mutation in familial breast cancer in Australia. Breast Cancer Res BCR 2001;3(5):346–9.

33. Xu H, DiCarlo J, Satya RV, Peng Q, Wang Y. Comparison of somatic mutation calling methods in amplicon and whole exome sequence data. BMC Genomics 2014;15:244.

34. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nat Commun 2015;6:8971.

35. Fraley C, E. Raftery A. mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. J Stat Softw 2007;18(6).

36. Choi M, Scholl UI, Yue P, et al. K+ channel mutations in adrenal aldosterone-producing adenomas and hereditary hypertension. Science 2011;331(6018):768–72.

37. Thiagalingam S, Laken S, Willson JKV, et al. Mechanisms underlying losses of heterozygosity in human colorectal cancers. Proc Natl Acad Sci U S A 2001;98(5):2698–702.

38. Knudson AG. Mutation and Cancer: Statistical Study of Retinoblastoma. Proc Natl Acad Sci U S A 1971;68(4):820–3.

39. Zhu X, Dunn JM, Goddard AD, et al. Mechanisms of loss of heterozygosity in retinoblastoma. Cytogenet Cell Genet 1992;59(4):248–52.

40. Varley JM, McGown G, Thorncroft M, et al. Germ-line mutations of TP53 in Li-Fraumeni families: an extended study of 39 families. Cancer Res 1997;57(15):3245–52.

41. Lott ST, Chandler DS, Curley SA, et al. High frequency loss of heterozygosity in von Hippel-Lindau (VHL)-associated and sporadic pancreatic islet cell tumors: evidence for a stepwise mechanism for malignant conversion in VHL tumorigenesis. Cancer Res 2002;62(7):1952–5.

42. Hollander MC, Blumenthal GM, Dennis PA. PTEN loss in the continuum of common cancers, rare syndromes and mouse models. Nat Rev Cancer 2011;11(4):289–301.

43. Zhang L, Poh CF, Williams M, et al. Loss of heterozygosity (LOH) profiles--validated risk predictors for progression to oral cancer. Cancer Prev Res Phila Pa 2012;5(9):1081–9.

44. Sathirapongsasuti JF, Lee H, Horst BAJ, et al. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. Bioinforma Oxf Engl 2011;27(19):2648–54.

45. Nadaf J, Majewski J, Fahiminiya S. ExomeAI: detection of recurrent allelic imbalance in tumors using whole-exome sequencing data. Bioinformatics 2015;31(3):429–31.

46. Freeman JL, Perry GH, Feuk L, et al. Copy number variation: new insights in genome diversity. Genome Res 2006;16(8):949–61.

47. Zhang F, Gu W, Hurles ME, Lupski JR. Copy Number Variation in Human Health, Disease, and Evolution. Annu Rev Genomics Hum Genet 2009;10:451–81.

48. Prior TW, Nagan N, Sugarman EA, Batish SD, Braastad C. Technical standards and guidelines for spinal muscular atrophy testing. Genet Med Off J Am Coll Med Genet 2011;13(7):686–94.

49. Jabbi M, Kippenhan JS, Kohn P, et al. The Williams syndrome chromosome 7q11.23 hemideletion confers hypersocial, anxious personality coupled with altered insula structure and function. Proc Natl Acad Sci U S A 2012;109(14):E860-866.

50. Beroukhim R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. Nature 2010;463(7283):899–905.

51. Tseng Y-Y, Moriarity BS, Gong W, et al. PVT1 dependence in cancer with MYC copy-number increase. Nature 2014;512(7512):82–6.

52. Magi A, Tattini L, Cifola I, et al. EXCAVATOR: detecting copy number variants from whole-exome sequencing data. Genome Biol 2013;14(10):R120.

53. Krumm N, Sudmant PH, Ko A, et al. Copy number variation detection and genotyping from exome sequence data. Genome Res 2012;22(8):1525–32.

54. Hastings P, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nat Rev Genet 2009;10(8):551–64.

55. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinforma Oxf Engl 2007;23(6):657–63.

56. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 2011;12(4):R41.

57. Zhao S, Bellone S, Lopez S, et al. Mutational landscape of uterine and ovarian carcinosarcomas implicates histone genes in epithelial–mesenchymal transition. Proc Natl Acad Sci 2016;201614120.

58. Shlien A, Malkin D. Copy number variations and cancer. Genome Med 2009;1(6):62.

59. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536(7616):285–291.

60. Hodis E, Watson IR, Kryukov GV, et al. A Landscape of Driver Mutations in Melanoma. Cell 2012;150(2):251–63.

61. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490(7418):61–70.

62. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. Science 2013;339(6127):1546–58.

63. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 2013;45(10):1113–20.

64. Wang Y, Zhou BP. Epithelial-mesenchymal Transition---A Hallmark of Breast Cancer Metastasis. Cancer Hallm 2013;1(1):38–49.

65. Murphy G. The ADAMs: signalling scissors in the tumour microenvironment. Nat Rev Cancer 2008;8(12):929–41.

66. Rosanò L, Spinella F, Di Castro V, et al. Endothelin-1 promotes epithelial-to-mesenchymal transition in human ovarian cancer cells. Cancer Res 2005;65(24):11649–57.

67. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol 2014;15(3):178–96.

68. Viñas-Castells R, Beltran M, Valls G, et al. The hypoxia-controlled FBXL14 ubiquitin ligase targets SNAIL1 for proteasome degradation. J Biol Chem 2010;285(6):3794–805.

69. Bolós V, Mira E, Martínez-Poveda B, et al. Notch activation stimulates migration of breast cancer cells and promotes tumor growth. Breast Cancer Res BCR 2013;15(4):R54.

70. Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. Nature 2007;446(7132):153–8.

71. Irminger-Finger I, Jefford CE. Is there more to BARD1 than BRCA1? Nat Rev Cancer 2006;6(5):382–91.

72. Park DJ, Lesueur F, Nguyen-Dumont T, et al. Rare mutations in XRCC2 increase the risk of breast cancer. Am J Hum Genet 2012;90(4):734–9.

73. Stephens PJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature 2012;486(7403):400–4.

74. Johnson RD, Liu N, Jasin M. Mammalian XRCC2 promotes the repair of DNA double-strand breaks by homologous recombination. Nature 1999;401(6751):397–9.

75. Liu N, Lamerdin JE, Tebbs RS, et al. XRCC2 and XRCC3, new human Rad51-family members, promote chromosome stability and protect against DNA cross-links and other damages. Mol Cell 1998;1(6):783–93.

76. Lee J-H, Paull TT. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. Science 2005;308(5721):551–4.

77. Bergamaschi A, Kim YH, Wang P, et al. Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. Genes Chromosomes Cancer 2006;45(11):1033–40.

78. Henrich K-O, Schwab M, Westermann F. 1p36 tumor suppression--a matter of dosage? Cancer Res 2012;72(23):6079–88.

79. Noviello C, Courjal F, Theillet C. Loss of heterozygosity on the long arm of chromosome 6 in breast cancer: possibly four regions of deletion. Clin Cancer Res Off J Am Assoc Cancer Res 1996;2(9):1601–6.

80. Gentile M, Olsen K, Dufmats M, Wingren S. Frequent allelic losses at 11q24.1-q25 in young women with breast cancer: association with poor survival. Br J Cancer 1999;80(5–6):843–9.

81. Weber-Mangal S, Sinn H-P, Popp S, et al. Breast cancer in young women (≤ 35 years): genomic aberrations detected by comparative genomic hybridization. Int J Cancer 2003;107(4):583–592.

82. Khan S, Fagerholm R, Rafiq S, et al. Polymorphism at 19q13.41 Predicts Breast Cancer Survival Specifically after Endocrine Therapy. Clin Cancer Res Off J Am Assoc Cancer Res 2015;21(18):4086–96.

83. Wikman H, Lamszus K, Detels N, et al. Relevance of PTEN loss in brain metastasis formation in breast cancer patients. Breast Cancer Res BCR 2012;14(2):R49.

84. Li Q, Gu X, Weng H, et al. Bone morphogenetic protein-9 induces epithelial to mesenchymal transition in hepatocellular carcinoma cells. Cancer Sci 2013;104(3):398–408.

85. Wang K, Feng H, Ren W, et al. BMP9 inhibits the proliferation and invasiveness of breast cancer cells MDA-MB-231. J Cancer Res Clin Oncol 2011;137(11):1687–96.

86. Ren W, Liu Y, Wan S, et al. BMP9 inhibits proliferation and metastasis of HER2-positive SK-BR-3 breast cancer cells through ERK1/2 and PI3K/AKT pathways. PloS One 2014;9(5):e96816.

87. Weigman VJ, Chao H-H, Shabalin AA, et al. Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival. Breast Cancer Res Treat 2012;133(3):865–80.

88. Silver DP, Richardson AL, Eklund AC, et al. Efficacy of neoadjuvant Cisplatin in triple-negative breast cancer. J Clin Oncol Off J Am Soc Clin Oncol 2010;28(7):1145–53.

89. Choschzick M, Lassen P, Lebeau A, et al. Amplification of 8q21 in breast cancer is independent of MYC and associated with poor patient outcome. Mod Pathol Off J U S Can Acad Pathol Inc 2010;23(4):603–10.

90. Yaremko ML, Kutza C, Lyzak J, Mick R, Recant WM, Westbrook CA. Loss of heterozygosity from the short arm of chromosome 8 is associated with invasive behavior in breast cancer. Genes Chromosomes Cancer 1996;16(3):189–95.

91. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci Signal 2013;6(269):pl1.

92. Walsh CS. Two decades beyond BRCA1/2: Homologous recombination, hereditary cancer risk and a target for ovarian cancer therapy. Gynecol Oncol 2015;137(2):343–50.

93. James CR, Quinn JE, Mullan PB, Johnston PG, Harkin DP. BRCA1, a potential predictive biomarker in the treatment of breast cancer. The oncologist 2007;12(2):142–150.

94. Faneyte IF, Schrama JG, Peterse JL, Remijnse PL, Rodenhuis S, van de Vijver MJ. Breast cancer response to neoadjuvant chemotherapy: predictive markers and relation with outcome. Br J Cancer 2003;88(3):406–12.

95. Lord CJ, Ashworth A. BRCAness revisited. Nat Rev Cancer 2016;16(2):110–20.

96. Lips EH, Mulder L, Hannemann J, et al. Indicators of homologous recombination deficiency in breast cancer and association with response to neoadjuvant chemotherapy. Ann Oncol Off J Eur Soc Med Oncol 2011;22(4):870–6.

97. Afghahi A, Timms KM, Vinayak S, et al. Tumor BRCA1 Reversion Mutation Arising During Neoadjuvant Platinum-Based Chemotherapy in Triple-Negative Breast Cancer Is Associated with Therapy Resistance. Clin Cancer Res Off J Am Assoc Cancer Res 2017;

98. Graeser M, McCarthy A, Lord CJ, et al. A marker of homologous recombination predicts pathological complete response to neoadjuvant chemotherapy in primary breast cancer. Clin Cancer Res Off J Am Assoc Cancer Res 2010;16(24):6159–68.

99. R K, E S, E P, et al. Polymorphism of the homologous recombination repair genes RAD51 and XRCC3 in breast cancer. Exp Mol Pathol 2009;87(1):32–5.

100. Telli ML, Timms KM, Reid J, et al. Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. Clin Cancer Res Off J Am Assoc Cancer Res 2016;22(15):3764–73.

101. Geigl JB, Obenauf AC, Schwarzbraun T, Speicher MR. Defining "chromosomal instability." Trends Genet TIG 2008;24(2):64–9.

102. Kennedy RD, Quinn JE, Mullan PB, Johnston PG, Harkin DP. The role of BRCA1 in the cellular response to chemotherapy. J Natl Cancer Inst 2004;96(22):1659–68.

103. Holstege H, Horlings HM, Velds A, et al. BRCA1-mutated and basal-like breast cancers have similar aCGH profiles and a high incidence of protein truncating TP53 mutations. BMC Cancer 2010;10:654.

104. Stefansson OA, Jonasson JG, Johannsson OT, et al. Genomic profiling of breast tumours in relation to BRCA abnormalities and phenotypes. Breast Cancer Res 2009;11(4):R47.

105. Abkevich V, Timms KM, Hennessy BT, et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. Br J Cancer 2012;107(10):1776–82.

106. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016;534(7605):47–54.

107. Greaves M, Maley CC. CLONAL EVOLUTION IN CANCER. Nature 2012;481(7381):306–13.

108. Guha M. PARP inhibitors stumble in breast cancer. Nat Biotechnol 2011;29(5):373–4.

109. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A, et al. A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. JAMA. 2011;305:1873–81. doi: 10.1001/jama.2011.593. pmid:21558518