Genomic complexity predicts resistance to endocrine therapy and CDK4/6 inhibition in hormone receptor-positive (HR+)/HER2-negative metastatic breast cancer

A. A. Davis[1], J. Luo[2], T. Zheng[3], C. Dai[3], X. Dong[3], L. Tan[3], R. Suresh[1], F. O. Ademuyiwa[1], C. Rigden[1], T. P. Rearden[1], K. Clifton[1], K. Weilbaecher[1], A. Frith[1], P. K. Tandra[4], T. Summa[1], B. Haas[1], S. Thomas[1], L. Hernandez-Aya[1], L. L. Peterson[1], X. Wang[3], S. J. Luo[3], K. Zhou[3], P. Du[3], S. Jia[3], B. L. King[3], J. Krishnamurthy[4], C. X. Ma[1]

[1]Department of Medicine, Division of Oncology, Washington University School of Medicine in St. Louis, MO
[2]Division of Public Health Science, Department of Surgery, Biostatistics Shared Resource, Washington University in St. Louis, MO
[3]Predicine, Inc., Hayward, CA
[4]Division of Oncology/Hematology, University of Nebraska Medical Center, Omaha, NE.

Corresponding authors:

*__Corresponding author,__ Dr. Andrew A. Davis, Section of Medical Oncology, Division of Oncology, Department of Medicine, Siteman Cancer Center, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, MO 63110, USA. Tel: 314-273-3581; Email: aadavis@wustl.edu

*__Corresponding author,__ Dr. Cynthia X. Ma, Section of Medical Oncology, Division of Oncology, Department of Medicine, Siteman Cancer Center, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, MO 63110, USA. Tel: 314-362-9383, Fax: 314-362-7086. Email: cynthiaxma@wustl.edu; Orcid ID: https://orcid.org/0000-0002-8156-7492

SUPPLEMENTARY METHODS

Blood collection and cfDNA/gDNA extraction

Each single draw of 8 mL of whole blood was collected into a Streck tube before undergoing a two-step centrifugation to separate plasma and buffy coat compartments. Aliquoted samples were stored at -80°C for batch processing. Circulating cell-free DNA (cfDNA) was extracted from plasma samples using the QIAamp circulating nucleic acid kit (Qiagen, Hilden, Germany). Quantity and quality of the purified cfDNA were checked using a Qubit fluorimeter (ThermoFisher Scientific, Waltham, Massachusetts, USA) and Bioanalyzer 2100 (Agilent Technologies, California, USA. For cfDNA samples with severe genomic contamination from peripheral blood cells, a bead-based size selection was performed to remove large genomic fragments (AMPure XP beads, Beckman Coulter, California, USA). Genomic DNA (gDNA) was extracted from matched peripheral blood mononuclear cells (PBMCs) using the QIAamp DNA Blood Mini Kit (Qiagen), then enzymatically fragmented and purified.

Library preparation, hybrid capture and sequencing

Five to 30 ng of extracted cfDNA or 30-50 ng of fragmented PBMC gDNA were then processed for library construction including end-repair dA-tailing and adapter ligation. Ligated library fragments with appropriate adapters were amplified via PCR. The amplified DNA libraries were then further checked using a Bioanalyzer 2100 and samples with sufficient yield were advanced to hybrid capture.

Hybrid capture was conducted using Biotin labelled DNA probes. In brief, each library was hybridized overnight with a Predicine NGS panel and paramagnetic beads. The unbound fragments were washed away, and the enriched fragments were amplified via PCR amplification. The purified product was checked on a Bioanalyzer 2100 and then loaded into an Illumina NovaSeq 6000 (San Diego, CA, USA) for NGS sequencing with paired-end 2x150bp sequencing kits.

Analyses of NGS data from cfDNA

NGS data from cfDNA were analyzed using the Predicine DeepSea NGS analysis pipeline, which starts from the raw sequencing data (BCL files) and outputs the final mutation calls. Briefly, the pipeline first performed adapter trimming, barcode checking, and correction. Cleaned paired FASTQ files were aligned to human reference genome build hg19 using the BWA alignment tool (BWA, RRID:SCR_010910). Consensus bam files were then derived by merging paired-end reads originated from the same molecules (based on mapping location and unique molecular identifiers) as single strand fragments. Single strand fragments from the same double strand DNA molecules were further merged as double stranded. By using the error suppression method described in [Newman, 2016], both sequencing and PCR errors were mostly corrected during this process.

Somatic mutation identification

Candidate variants were called by comparing with local variant background (defined based on plasma samples from healthy donors and historical data). Variants were further filtered by log-odds (LOD) threshold [Cibulskis 2013], base and mapping quality thresholds, repeat regions and other quality metrics.

Candidate somatic mutations were further filtered on the basis of gene annotation to identify those occurring in protein-coding regions. Intronic and silent changes were excluded, while mutations resulting in missense mutations, nonsense mutations, frameshifts, or splice site alterations were retained. Mutations annotated as benign or likely benign were also filtered out based on the ClinVar, database (ClinVar, RRID:SCR_006169), or as common germline variants in databases including 1000 genomes, ExAC, gnomAD and KAVIAR with population allele frequency >0.5%. Finally, hematopoietic expansion-related variants that have been previously described, including those in DNMT3A, ASXL1, TET2, and specific alterations within ATM (residue 3008), GNAS (residue 201, 202), or JAK2 (residue 617) were marked as CHIP-related mutations .

Germline DNA analysis

Germline variants were determined by concurrent sequencing of buffy coat PBMCs. Candidate variants with low base quality, mapping scores, and other poor quality metrics were filtered. Candidate variants with an allelic frequency <5% or with less than 8 distinct reads containing the mutation were excluded. Unknown variants in repeat regions were also excluded. Details of the analytical workflow are provided above in "Analyses of NGS data generated from cfDNA".

Copy number analysis by targeted panel

Copy number variations were estimated at the gene level. The pipeline calculates the on-target unique fragment coverage based on consensus bam files, which is first corrected for GC bias, and is then adjusted for the probe level bias (estimated from a pooled reference). Each adjusted coverage profile is self-normalized (assuming a diploid status of each sample) first, and then compared against correspondingly adjusted coverages from a group of normal reference samples to estimate the significance of the copy number variant. To call an amplification or deletion of a gene requires the absolute z-score and copy number change pass minimum thresholds.

DNA re-arrangements

DNA re-arrangement was detected by identifying the alignment breaking points based on the bam files before the consensus step. Suspicious alignments were filtered based on repeat regions, local entropy calculation and similarity between reference and alternative alignments. Larger than 3 unique alignments (at least one of them should be double stranded) are required to report a DNA fusion.

ctDNA fraction

ctDNA fractions are estimated based on the allele fractions of autosomal somatic mutations as described previously [Vandekerkhove, 2017]. Briefly, the mutant allele fraction (MAF) and ctDNA fraction are related as MAF = (ctDNA * 1) / [(1 - ctDNA) * 2 + ctDNA *1], and thus ctDNA = 2 / ((1

/ MAF) + 1). Somatic mutations in genes with a detectable copy number change are omitted from ctDNA fraction estimation.

Tumor Fraction

ichorCNA, adapted from the ABSOLUTE method [Carter, 2012], was applied to estimate tumor fraction using a hidden Markov model (HMM) based on genome-wide copy number variation at a segment level.

bTMB score estimation

Blood-based tumor mutational burden (bTMB) was defined as the number of somatic coding SNVs including synonymous and nonsynonymous variants within panel target regions. The bTMB score was then normalized by the total effective targeted panel size within the coding region [Gandara, 2018].

Copy number burden analysis using low-pass whole genome sequencing

Low-pass whole genome sequencing (LP-WGS) with an overall average coverage of 5x was performed on patient samples. The ichorCNA algorithm [Adalsteinsson, 2017] was applied to GC and mappability-normalized reads to estimate plasma copy number variations using the hidden Markov model (HMM). The pipeline first estimated the segment level (1Mb genomic regions) copy number deviation as log2 ratio of the normalized reads between the test sample and a pool of normal plasma samples, then the sample level copy number burden (CNB score) was calculated as the logarithm of the sum of absolute CNV z-scores, where higher/lower CNB score indicates higher/lower CNV abnormality compared with normal background. The CNB score cutoff of 5.6 was defined as three standard deviations away from the population mean of normal plasma CNB scores.

Mutational signature analysis

Mutational signature analysis was performed to compare patterns of single base substitutions (SBS) with previously reported SBS signatures [Alexandrov, 2013, 2020] now available in the

COSMIC data base using the maftools package (version 2.4.15) in R (version 3.6.3). Briefly, each of the 96 possible mutation substitution types is defined by one of six substitution types (T>A, T>C, T>G, C>A, C>G, C>T) and the bases immediately 5' and 3' to the mutated base. For each sample, the number of mutation substitutions was counted. Non-negative matrix factorization was used to decompose the count matrix into n signatures. The number of signatures (n) best fitting the data was estimated using Cophenetic correlation. Signatures were compared to the 78 available SBS COSMIC signatures (v3.2 - March 2021, https://cancer.sanger.ac.uk/signatures/sbs/) using cosine similarity. A dominant signature was defined as the signature with the maximum signature score in each sample.

Oncogenic signaling pathway analysis

To compare the relative proportion of mutations within key oncogenic signaling pathways in high vs. low bTMB patients, we filtered the list of genes included in a previous publication describing oncogenic signaling pathways [Sanchez-Vega, 2018] to include only those identified as breast cancer driver genes [Dietlein, 2020; Martinez-Jiminez, 2020]. The list of resulting genes is shown in the table below. The frequency of SNVs across these genes was compared between high vs. low bTMB patients and statistical significance was evaluated using the Fisher's Exact Test.

Pathway	Gene List
HRD	BRCA1,BRCA2
DDR	BRCA1,TP53,BAP1,BRCA2,PTEN,ATM,ARID1A,POLD1
PI3K	PIK3CA,AKT1,PTEN,PIK3R1,MTOR
Cell_Cycle	RB1,CDKN2A,CDKN1B
TP53	TP53,ATM
RTK-RAS	EGFR,FGFR2,BRAF,NF1,KRAS,ERBB3,ERBB4,JAK2,ALK,PDGFRB,HRAS,ERBB2
Hippo	FAT1,FAT4,FAT3
NOTCH	NOTCH2,CREBBP,NCOR2,FBXW7,SPEN
WNT	
MYC	
NRF2	
TGF-Beta	SMAD2,SMAD4

References.

1. Patel, P.G., et al., *Preparation of Formalin-fixed Paraffin-embedded Tissue Cores for both RNA and DNA Extraction.* J Vis Exp, 2016(114).
2. Newman, A.M., et al., *Integrated digital error suppression for improved detection of circulating tumor DNA.* Nat Biotechnol, 2016. **34**(5): p. 547-555.
3. Cibulskis, K., et al., *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.* Nat Biotechnol, 2013. **31**(3): p. 213-9.
4. Vandekerkhove, G., et al., *Circulating Tumor DNA Reveals Clinically Actionable Somatic Genome of Metastatic Bladder Cancer.* Clin Cancer Res, 2017. **23**(21): p. 6487-6497.
5. Carter, S.L., et al. *Absolute quantification of somatic DNA alterations in human cancer. Nature Biotechnology, 2012. **30**(5): p. 413.*
6. Gandara, D.R., et al., *Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab.* Nat Med, 2018. **24**(9): p. 1441-1448.
7. Adalsteinsson, V.A., et al., *Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors.* Nat Commun, 2017. **8**(1): p. 1324.
8. Alexandrov, L.B., et al., *The repertoire of mutational signatures in human cancer.* Nature, 2020. **578**(7793): p. 94-101.
9. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415-21.
10. Sanchez-Vega, F., et al., *Oncogenic Signaling Pathways in The Cancer Genome Atlas.* Cell, 2018. **173**(2): p. 321-337 e10.
11. Dietlein, F., et al., *Identification of cancer driver genes based on nucleotide context.* Nat Genet, 2020. **52**(2): p. 208-218.
12. Martinez-Jimenez, F., et al., *A compendium of mutational cancer driver genes.* Nat Rev Cancer, 2020. **20**(10): p. 555-572.