Somatic mutational landscapes of ductal and lobular breast carcinomas in the GENIE Cohort v8.1: real world actionability assessment in 8,756 patients

Alessandro Leal[1], Patricia Taranto[1], Poliana Bergamaschine Giovani Blasi[1], Bianca Baron Geronimo[1], Carlos Tadeu Garrote[2], Fernando Moura[1]

[1]Centro de Oncologia e Hematologia Einstein Dayan-Daycoval, Hospital Israelita Albert Einstein, Sao Paulo, Brazil; [2]Instituto do Cancer do Estado de Sao Paulo, Sao Paulo, Brazil

Background

Invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) of the breast typically present distinct clinicopathological characteristics and responsiveness to systemic therapy[1]. In addition, breast cancer data from The Cancer Genome Atlas (TCGA) have shown these two pathological subtypes also present distinct genomic features when analyzed using DNA copy number arrays and whole exome sequencing platforms[2]. More recently, the AACR Project GENIE Consortium, which is a publicly accessible international cancer registry of real-world data assembled through data sharing among leading cancer centers in the world, have allowed in-depth analyses of clinical actionability using patient-level data from clinical next-generation sequencing (NGS) assays[3]. In this study, we assessed the somatic mutational landscapes of a large cohort (n = 8,756) of invasive breast carcinomas from 19 institutions participating in the GENIE Consortium Cohort (v8.1) and examine clinical actionability of unique mutations identified in each breast cancer subtype.

Methods

We assessed the eighth data release of the GENIE Consortium Cohort v8.1 (**Fig. 1**) encompassing targeted sequencing data from 7,647 IDC and 1,109 ILC cases. Clinical features and somatic mutations including single-nucleotide variants, small indels, fusions, and copy number alterations (CNAs) were retrieved from cBioportal and SAGE Bionetworks. All patient samples were de-identified and encoded with GENIE sample codes. Gene actionability was examined using CiVIC, OncoKB, and ESCAT publicly available knowledgebases.

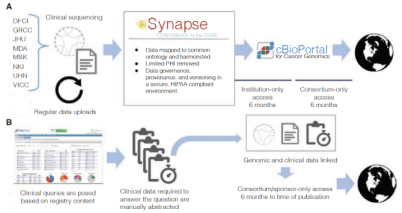

Fig. 1. AACR Project GENIE. A) Variant calls and a limited clinical dataset from patients at each of the participating centers are sent to the Synapse platform, developed by Sage Bionetworks, where the data are harmonized and protected health information (PHI) removed in a secure Health Insurance Portability and Accountability Act (HIPAA) – compliant environment that provides data governance. Once harmonized, these data are viewed and analyzed in the cBioPortal for Cancer Genomics. Value is provided to both the data generators and the consortium by establishing 6-month periods of exclusivity to each prior to the data becoming available to the broader research community. **B)** Once data are available in the cBioPortal, clinical research projects are proposed and vetted by the project steering committee. Clinical teams are then assembled to defi ne the clinical attributes required to answer the approved research question; these data are then manually curated from the relevant medical records and deposited in an electronic data capture system. The detailed clinical data are then transferred to Synapse where they are linked with the appropriate genomic and limited clinical data and are viewable and analyzable in the cBioPortal platform (extracted from AACR Project GENIE Consortium. *Cancer Discov* 8, 818-831, 2017).

Results

Patients with IDC tumors were 5 years younger than patients with ILC tumors at the time sequencing data was reported (median 55 *versus* median 60, *Kruskal-Wallis*, p < 10e-10) (**Fig. 2A**). Both IDC and ILC had on average 2 mutations per tested sample. Overall, IDC and ILC tumors had median fractions of 22% and 14% of their genomes altered, respectively (*Kruskal-Wallis*, p < 10e-10) (**Fig. 2B**).

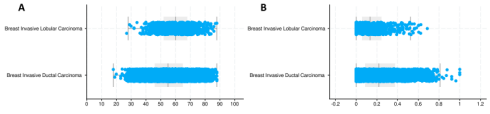

Fig. 2. Overview of reported breast cancer samples at Genie Consortium Cohort v8.1. A) Age distribution at which sequencing was report for invasive lobular carcinoma and invasive ductal carcinoma. **B)** Fraction of genome altered for invasive lobular carcinoma and invasive ductal carcinoma.

A gene enrichment analysis including 938 genes with point mutations and indels identified CDH1 (LR 4.66, p<1e-10), RHOA (LR 2.81, p=1.3e-10), PTK2B (LR 2.68, p=5.2e-4), ERBB2 (LR 1.80, p<1e-10), TBX3 (LR 1.72, p<1e-10), FOXA1 (LR 1.49, p=2.5e-10) and RUNX1 (LR 1.25, p=3.1e-9) as genes significantly enriched in ILC tumors. On the other hand, mutations in GATA3 (LR = 1.67, p<1e-10) and TP53 (LR = 1.55, p<1e-10) were significantly enriched in IDC tumors (**Fig. 3A, Table 1**). A further gene enrichment analysis for copy-number alterations in 1139 genes showed amplification in PARP1 (LR 1.55 p=2.5e-3) and deep deletions in IKZF1 (LR 2.8, p=2.2e-3) and CDH1 (LR = 1.88, p=1.7e-4) as the most enriched genes with CNAs in ILC. In parallel, amplifications in ERBB2 (LR 1.65, p=1e-10), MYC (LR 1.64, p=1e-10), COL22A1 (LR 1.19, p=1.6e-5), BRIP1 (LR 2.66, p<1e-10), CDK12 (LR 1.55, p=2.6e-9), PPM1D (LR 3.1, p<1e-10), RAD51C (LR 2.85, p=3.8e-8), AURKA (LR 3.2, p=1e-8) and deep deletion in CDKN2A (LR 2.1, p=1.9e-6) were enriched in IDC tumors (**Fig. 3B, Table 2**).

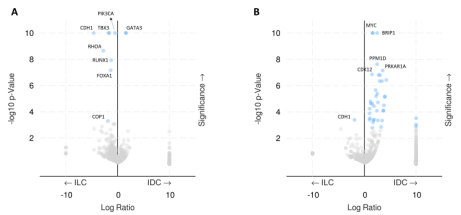

Fig. 3. Differential point mutations and copy number alterations (CNAs) at Genie Consortium Cohort v8.1. A) Volcano plot showing the log ratios for point mutations across invasive lobular carcinoma (ILC) and invasive ductal carcinoma (IDC). **B)** Volcano plot showing the log rations for CNAs across ILC and IDC.

Table 1. Genes harboring missense and nonsense mutations which are significantly enriched in invasive ductal carcinoma (blue) or invasive lobular carcinoma (red).

Gene	Cytoband	Invasive Ductal Carcinoma	Invasive Lobular Carcinoma	Log Ratio	p-Value	q-Value	Enriched in
CDH1	16q22.1	207 (3.06%)	631 (66.49%)	-4.44	< 10E-10	< 10E-10	Lobular
TP53	17p13.1	2967 (43.70%)	148 (15.53%)	1.49	< 10E-10	< 10E-10	Ductal
PIK3CA	3q26.32	2308 (34.00%)	466 (48.90%)	-0.52	< 10E-10	< 10E-10	Lobular
TBX3	12q24.21	197 (4.06%)	89 (12.90%)	-1.67	< 10E-10	< 10E-10	Lobular
ERBB2	17q12	199 (2.93%)	86 (9.02%)	-1.62	< 10E-10	< 10E-10	Lobular
GATA3	10p14	803 (13.67%)	44 (5.16%)	1.4	< 10E-10	< 10E-10	Ductal
RHOA	3p21.31	19 (0.61%)	22 (4.07%)	-2.74	5.86E-09	1.30E-06	Lobular
RUNX1	21q22.12	183 (3.07%)	62 (7.14%)	-1.22	3.82E-08	7.42E-06	Lobular
FOXA1	14q21.1	119 (3.78%)	47 (8.70%)	-1.2	2.40E-06	4.14E-04	Lobular
CBFB	16q22.1	214 (4.07%)	55 (7.30%)	-0.84	1.17E-04	0.0182	Lobular
BRIP1	17q23.2	104 (1.79%)	3 (0.36%)	2.33	3.46E-04	0.0489	Ductal

Table 2. Genes harboring amplifications or deep deletions which are significantly enriched in invasive ductal carcinoma (blue) or invasive lobular carcinoma (red).

Gene	Cytoband	Type	Invasive Ductal Carcinoma	Invasive Lobular Carcinoma	Log Ratio	p-Value	q-Value	Enriched in
ERBB2	17q12	Amp	826 (14.34%)	41 (4.86%)	1.56	< 10-10	< 10-10	Ductal
BRIP1	17q23.2	Amp	345 (6.01%)	9 (1.07%)	2.49	< 10-10	1.83E-09	Ductal
MYC	8q24.21	Amp	429 (10.90%)	23 (3.51%)	1.64	< 10-10	1.76E-08	Ductal
PPM1D	17q23.2	Amp	192 (6.84%)	6 (1.23%)	2.47	2.35E-08	6.57E-06	Ductal
PRKAR1A	17q24.2	Amp	140 (3.58%)	2 (0.31%)	3.54	7.09E-08	1.59E-05	Ductal
CDK12	17q12	Amp	316 (9.38%)	19 (3.36%)	1.48	1.41E-07	2.28E-05	Ductal
CD79B	17q23.3	Amp	162 (4.14%)	4 (0.61%)	2.75	1.52E-07	2.28E-05	Ductal
SPOP	17q21.33	Amp	147 (4.36%)	3 (0.53%)	3.04	1.63E-07	2.28E-05	Ductal
AKIN2	17q24.1	Amp	108 (3.51%)	1 (0.19%)	4.24	3.74E-07	4.66E-05	Ductal
AURKA	20q13.2	Amp	141 (3.60%)	3 (0.46%)	2.97	4.53E-07	4.66E-05	Ductal
RNF43	17q22	Amp	126 (3.74%)	2 (0.35%)	3.4	4.57E-07	4.66E-05	Ductal
GNAS	20q13.32	Amp	165 (4.19%)	6 (0.91%)	2.2	2.22E-06	2.07E-04	Ductal
RAD51C	17q22	Amp	136 (4.04%)	4 (0.71%)	2.51	4.15E-06	3.57E-04	Ductal
GATA3	10p14	Amp	165 (2.87%)	5 (0.60%)	2.27	6.38E-06	4.95E-04	Ductal
EGFR	7p11.2	Amp	106 (1.84%)	1 (0.12%)	3.96	6.63E-06	4.95E-04	Ductal
CCNE1	19q12	Amp	93 (2.36%)	1 (0.15%)	3.95	7.34E-06	5.14E-04	Ductal
COL22A1	8q24.23-q24.3	Amp	400 (21.89%)	18 (9.57%)	1.19	1.57E-05	1.04E-03	Ductal
NCOA3	20q13.12	Amp	131 (3.05%)	3 (0.49%)	2.64	1.90E-05	1.18E-03	Ductal
RAD21	8q24.11	Amp	252 (7.48%)	18 (3.11%)	1.26	2.30E-05	1.33E-03	Ductal
RECQL4	8q24.3	Amp	199 (5.59%)	12 (1.96%)	1.51	2.37E-05	1.33E-03	Ductal
IGF1R	15q26.3	Amp	83 (2.11%)	1 (0.15%)	3.79	3.18E-05	1.70E-03	Ductal
CDKN2A	9p21.3	DeepDel	173 (3.00%)	7 (0.83%)	1.86	3.63E-05	1.85E-03	Ductal
TG	8q24.22	Amp	421 (22.62%)	21 (10.99%)	1.04	5.60E-05	2.73E-03	Ductal
PRDM1	6q21	Amp	77 (1.97%)	1 (0.15%)	3.68	7.74E-05	3.48E-03	Ductal
RTEL1	20q13.33	Amp	75 (4.21%)	1 (0.33%)	3.66	8.06E-05	3.48E-03	Ductal
GH1	17q23.3	Amp	158 (8.65%)	3 (1.60%)	2.44	8.08E-05	3.48E-03	Ductal
UBR5	8q22.3	Amp	392 (21.36%)	20 (10.53%)	1.02	1.28E-04	5.32E-03	Ductal
ROS1	6q22.1	Amp	59 (1.02%)	0 (0.00%)	>10	3.02E-04	0.0117	Ductal
STMN2	8q21.13	Amp	330 (18.06%)	16 (8.51%)	1.09	3.02E-04	0.0117	Ductal
AGO2	8q24.3	Amp	109 (6.12%)	5 (1.66%)	1.88	3.83E-04	0.0141	Ductal
DCAF4L2	8q21.3	Amp	351 (19.21%)	18 (9.57%)	1	3.97E-04	0.0141	Ductal
CDH1	16q22.1	DeepDel	24 (0.42%)	13 (1.55%)	-1.89	4.04E-04	0.0141	Lobular
CDKN2B	9p21.3	DeepDel	121 (3.08%)	6 (0.91%)	1.75	4.18E-04	0.0142	Ductal
IKZF3	17q12-q21.1	Amp	76 (13.24%)	2 (2.15%)	2.62	4.41E-04	0.0144	Ductal
SOX9	17q24.3	Amp	64 (1.67%)	1 (0.15%)	3.43	4.51E-04	0.0144	Ductal
PIK3CA	3q26.32	Amp	133 (2.31%)	6 (0.71%)	1.7	6.38E-04	0.0196	Ductal
NF1	17q11.2	Amp	51 (0.89%)	0 (0.00%)	>10	9.11E-04	0.0276	Ductal
ZNF217	20q13.2	Amp	47 (4.00%)	0 (0.00%)	>10	1.30E-03	0.0382	Ductal
PTPRD	9p24.1-p23	Amp	65 (1.24%)	1 (0.13%)	3.23	1.39E-03	0.0398	Ductal
BCAS3	17q23.2	Amp	169 (9.25%)	6 (3.19%)	1.54	1.62E-03	0.0454	Ductal

We identified 981 genes with point mutations across all 8,756 samples. From these, there are OncoKB curated information for 539 (54.9%) genes. Regarding variants and genes actionability for breast cancer, OncoKB and ESCAT present data for 16 (1.6%) and 11 (1.1%) genes, respectively (**Table 3**). Among enriched alterations for each histological subtype, the knowledgebase CiVIC does not present curated data available for genes TBX3, FOXA1, GATA3, COL22A1, BRIP1, PPM1D, and RAD51C. OncoKB only missed genes PTK2B and COL22A1.

Table 3. Genes harboring clinically actionable alterations in breast cancer at OncoKB and ESCAT.

Hugo Symbol	OncoKB Breast Cancer Data	OncoKB Breast Cancer Drugs List	OncoKB Breast Cancer Level	OncoKB Breast Cancer Number of Citations	ESCAT Breast Cancer Data	ESCAT Tier
PIK3CA	Yes	Alpelisib + Fulvestrant	1	3	Yes	IA
AKT1	Yes	AZD5363	3A	4	Yes	IIB
BRCA1	Yes	Talazoparib, Olaparib	2	5	Yes	IA
BRCA2	Yes	Talazoparib, Olaparib	2	4	Yes	IA
ERBB2	Yes	Lapatinib + Trastuzumab; Pertuzumab + Trastuzumab; Tucatinib + Capecitabine; Trastuzumab; Ado- Trastuzumab; Emtansine; Lapatinib; Neratinib; Deruxtecan	1	20	Yes	IA ou IIB
ESR1	Yes	AZD9496 e Fulvestrant	3A	3	Yes	IIA
NTRK2	Yes	Larotrectinib e Entrectinib	1	7	Yes	IC
NTRK3	Yes	Larotrectinib e Entrectinib	1	7	Yes	IC
PTEN	Yes	GSK2636771 e AZD8186	4	2	Yes	IIA
FGFR1	Yes	AZD4547; BGJ398; Erdafitinib; Debio 1347	4	8	No	NA
FGFR2	Yes	AZD4547; BGJ398; Erdafitinib; Debio 1347	4	8	No	NA
FGFR3	Yes	AZD4547; BGJ398; Erdafitinib; Debio 1347	4	12	No	NA
KRAS	Yes	Cobimetinib; Binimetinib; Trametinib	4	8	No	NA
MET	Yes	Crizotinib	4	4	No	NA
MTOR	Yes	Temsirolimus; Everolimus	4	5	No	NA
NF1	Yes	Cobimetinib; Trametinib	4	5	No	NA
MDM2	No	NA	NA	NA	Yes	IIA
NTRK1	No	NA	NA	NA	Yes	IC

Conclusions

Real-world genomic data from the GENIE Consortium Cohort support that breast cancer presents distinct mutational landscapes for IDC and ILC tumors. For each histological subtype, we confirmed there are different levels of enrichments for shared mutations in actionable genes. Even though publicly available knowledgebases present curated information about commonly mutated genes in cancer, we noticed that actionability data for important cancer genes are still scarce.

References:

1. Luveta, J., Parks, R.M., Heery, D.M. *et al.* Invasive Lobular Breast Cancer as a Distinct Disease: Implications for Therapeutic Strategy. *Oncol Ther* 8, 1–11 (2020).
2. The Cancer Genome Atlas Network., Genome sequencing centres: Washington University in St Louis., Koboldt, D. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70 (2012).
3. AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov* 8, 818-831 (2017).

Contact: aleal@einstein.br