Contents lists available at ScienceDirect

Cancer Treatment and Research Communications

journal homepage: www.sciencedirect.com/journal/cancer-treatment-and-research-communications

Breast cancer biobank from a single institutional cohort in an urban setting in india: Tumor characteristics and survival outcomes

Laleh Busheri[a], Santosh Dixit[a,b], Smeeta Nare[a], Rashmi Alhat[a], George Thomas[a], Mangal Jagtap[a], Ruth Navgire[a], Priya Shinde[a], Rituja Banale[a], Rohini Unde[a], Ruhi Reddy[a], Shahin Shaikh[a], Aishwarya Konnur[a], Namrata Namewar[a], Ashwini Bapat[a], Ankita Patil[a,b], Rateeka Johari[a], Roli Kushwaha[a,b], Wimpy Kumari[a], Beenu Varghese[a], Pooja Deshpande[a,#], Chetan Deshmukh[a], Devaki A. Kelkar[a,b], L S Shashidhara[b,c], Chaitanyanand B Koppiker[a,b], Madhura Kulkarni[a,b,*]

[a] *Prashanti Cancer Care Mission, Pune*
[b] *Center for Translational Cancer Research, a Joint venture between Prashanti Cancer Care Mission and IISER Pune*
[c] *Ashoka University, Sonipat, Delhi*

ARTICLE INFO

Keywords:
Breast Cancer Biobank
Single Surgeon – Cohort
Indian Cohort

ABSTRACT: 250 WORDS

Background: A breast cancer biobank with retrospectively collected patient data and FFPE tissue samples was established in 2018 at Prashanti Cancer Care Mission, Pune, India. It runs a cancer care clinic with support from a single surgeon's breast cancer practice. The clinical data and tissue sample collection is undertaken with appropriate patient consent following ethical approval and guidelines.

Methods: The biobank holds clinical history, diagnostic reports, treatment and follow-up information along with FFPE tumor tissue specimens, adjacent normal and, in few cases, contralateral normal breast tissue. Detailed family history and germline mutational profiles of eligible and consenting patients and their relatives are also deposited in the biobank.

Results: Here, we report the first audit of the biobank. A total number of 994 patients with breast disease have deposited consented clinical records in the biobank. The majority of the records (80%, $n = 799$) are of patients with infiltrating ductal carcinoma (IDC). Of 799 IDC patients, 434 (55%) have deposited tumor tissue in the biobank with consent. In addition, germline mutation profiles of 84 patients and their family members are deposited. Follow-up information is available for 85% of the 434 IDC patients with an average follow-up of 3 years.

Conclusion: The biobank has aided the initiation of translational research at our center in collaboration with eminent institutes like IISER Pune and SJRI Bangalore to evaluate profiles of breast cancer in an Indian cohort. The biobank will be a valuable resource to the breast cancer research community, especially to understand South Asian profiles of breast cancer.

Background

Breast cancer is one of the cancers with the highest lifetime risks and a common cause of cancer death amongst women worldwide [1]. Breast cancers are characterized based on histopathological parameters such as tumor size, tumor grade, lymph node involvement [2] and classified by the expression of molecular hormone receptors such as estrogen receptor (ER), progesterone receptor (PR) and ERBB2 (HER2) [3]. Both the clinical stage and the molecular receptor expression affect the prognosis of the disease and also guide treatment decisions. Most breast cancers (70%−85%) express hormone receptors ER/PR and/or HER2. Such receptor-positive tumors show a better response to targeted therapies

[*] Madhura Kulkarni, PhD, Senior Scientist and DBT-Ramalingaswami Fellow. Center for Transltional Cancer Research, a joint inititative of Prashanti Cancer Care Mission and IISER Pune. And Prashanti Cancer Care Mission Pune India. 1-2 Kapil Vastu, Senapati Bapat Road Pune 411016.

E-mail address: madhura.kulkarni@me.com (M. Kulkarni).

[#] deceased and contribution is honoured with the authorship

https://doi.org/10.1016/j.ctarc.2021.100409

Available online 1 June 2021

and have limited relapse [4, 5]. The remaining 15–30% that do not express hormone receptors or HER2 are referred to as triple-negative breast cancer (TNBC), lack targeted therapy and require systemic chemotherapy [6].

In India, 14% of all newly diagnosed cancers are breast cancers, and 27% of all cancer deaths are attributed to breast cancer with high proportions below age 50 [7], unlike that of western countries [1, 8]. In a span of 10 years, from 2016 to 2026, it is estimated that the number of breast cancer cases will rise by 32%, i.e. with 240,000 new incidence cases estimated in 2026 [9]. There is a growing need to understand and investigate histopathological, molecular, and genetic profiles of breast cancer in the Indian context to better treat and manage the disease in the coming years. Establishing dedicated research centers with tumor tissue repositories is essential to understand the trajectory of breast cancers in India in terms of their response to therapy, recurrence, and survival rates, along with their molecular and genetic profiles.

With the aim to profile breast cancer in an Indian cohort, we established a breast cancer patient database and tissue repository at our center in 2018. The retrospective cohort consists of consented and annotated patient data and tissue samples collected and curated from breast cancer patients who visited the onco-surgeon and the clinic from 2010 till 2018. Since the center is a single surgeon unit, the patients have received uniform treatment and are followed up by one clinician through diagnosis, treatment and yearly follow-ups. Our center is one of the few centres in India that performs oncoplastic surgery. Hence, along with adjacent normal, the biobank holds contralateral breast tissue samples as a valuable source of true normal breast tissue.

In this report, we audit the breast cancer patient cohort for benign and malignant tumor cases with their clinical, follow-up data and tissue repository for the first time since the establishment of the biobank.

Methods

Oncological management at the unit

The general patient flow through diagnosis and treatment is described previously in a detailed performance audit of the clinic [10, 11]. Patients visit the clinic either for a routine breast screening or present with specific (symptoms) complaints such as lump, pain, nipple discharge, nipple retraction, skin changes or axillary lump. These patients undergo clinical breast examination followed by screening or diagnostic 2D digital mammograms with 3D tomosynthesis. Ultrasonography is performed in cases with abnormal or inconclusive mammograms. Suspicious findings at radiology are recommended for a biopsy.

In most cases, a core needle biopsy of the breast lesion is performed on the same day. A fine-needle aspiration biopsy is performed in specific situations, e.g. abnormal axillary lymph nodes. All procedures are performed under local anesthesia with proper aseptic precautions and with written informed consent. After biopsy diagnosis, patients with non-malignant conditions are recommended for follow-up or surgery as required. Some of them undergo vacuum-assisted biopsy procedures in the clinic under local anesthesia where indicated. Patients with malignant disease are counselled for an appropriate line of treatment. Patients eligible for genetic testing are recommended for in-house genetic counselling and HBOC testing as per NCCN guidelines.

Ethics approval

To initiate the biobank with breast cancer patient clinical data and FFPE tissue, a proposal was developed for the retrospective collection of de-identified data and tissue samples for submission to the institute's ethics committee. The committee approved the consent forms, data collection forms, SOPs and tissue collection and storage SOPs on 21st July 2018. The ethics committee's approval letter is provided as supplementary document 1.

With this approval, the patients diagnosed with breast-related diseases from 2010 to 2018 treated with the oncosurgeon on board were contacted and requested consent for clinical data and tissue deposition.

Patient consenting

The patients listed in the clinic's yearly roster were contacted by a trained clinical staff either via telephone or in-person during their annual check-up. The trained clinical staff requested an in-person visit to explain the purpose of the biobank and consenting. During the in-person visit, a consent form either in English (supplementary document 2) or the vernacular language (Marathi) (supplementary document 3) was shared and explained to the patient. The patient's signed consent form was filed in the clinic records.

Patient data collection

For all the patients who consented, a copy of the clinical records and notes, test reports, discharge reports, treatment regimens were filed in the clinic record files with a unique patient ID. Designated and qualified staff curated individual files within twelve modules. Each module was tagged with three unique identifiers, 1. birth date, 2. first visit date at the clinic and 3. unique patient ID designated at the clinic. The clinical information tagged with each module is listed below.

1 Patient information/history: biographical information, patient habits, medical history, family medical history, reproductive history, symptoms and mode of breast cancer detection, metastatic workup.

2 Radiology: mammography, ultrasonography, automated breast volume scanning (ABVS), MRI data reports are curated for lesion location and features, node features at diagnosis. This module is based on the ACR BI-RADS Atlas (Fifth Edition). Radiology images are stored in the Picture Archiving and Communication System (PACS) in Digital Imaging and Communications in Medicine (DICOM) format.

3 PET scans: details of any PET scans performed at diagnosis, during NACT or follow up. Results for the brain, thorax, abdomen are divided into abnormal and normal sections entered as free text. Breast related observations are entered as a structured text for lesion size, SUV status, node observations and SUV status. The presence of metastatic features and location is curated separately.

4 Biopsy Reports Histopathology reports for a type of tumor, grade, lymphovascular invasion, tumor infiltrating lymphocytes, etc. Immunohistochemistry assessments for Estrogen Receptor (ER), Progesterone Receptor (PR), HER2 and ki67 expression and HER2 FISH, wherever applicable, are logged in for biopsy tissue.

5 Neo-Adjuvant Therapy: details of neoadjuvant treatment – both hormone and chemotherapy. Details include regimen prescribed, drugs, dose, patient weight, toxicity and treatments, and residual tumor size assessment. Details of residual tumor localization by clip/wire insertion.

6 Surgery: Surgery type of conventional surgery or oncoplastic breast surgery is mentioned. Surgical details such as the type of incision used and the weight of tissue excised are captured. Most surgeries have unique elements of oncoplasty, such as immediate breast reconstruction after a mastectomy using implants or autologous tissue such as *Latisimmus Dorsi* (LD) flaps, breast conservation using techniques such as simple oncoplastic closure, therapeutic mammoplasty, perforator flaps etc. Details specific to these techniques such as pedicles, nipple-areola complex graft are captured. Details of sentinel node biopsy and frozen tissue are also collected in this module. Finally, post-surgical complications and their treatment are collected in this module.

7 Surgery tissue: Frozen tissue histopathology report, tumor size, margins, histopathology report of FFPE tissue blocks to include

tumor, adjacent normal tissue and contralateral breast tissue. IHC report, if any. Follow-up surgery report, if any.

8 Adjuvant Chemotherapy: details of post-surgery chemotherapy – drugs, dose, toxicity and treatments.

9 Radiotherapy: type of radiotherapy given, reasons for discontinuation, if any, and toxicity. This data is taken from a radiation therapy discharge summary created by the radiotherapy department at an allied Hospital.

10 Hormone Therapy and Survival: details of long-term hormone therapy if given and response, ovarian suppression details if given, recurrence and survival status at last follow-up.

11 Follow-up Data: detail of every follow up visit and tests done; investigation and treatment for recurrence; status at every follow up. For some patient's follow-up is taken from a telephonic conversation with a qualified nurse.

12 Germline genetic testing and counselling report: After genetic counselling, collecting family pedigree and obtaining written informed consent, germline genetic testing is performed using genomic DNA isolated from blood/saliva of the proband (diagnostic testing) and/or family members (preventive testing) in allied CAP-certified genetic laboratories. Exome NGS is performed using an ACMG-consensus multi-gene panel comprising key genes involved in hereditary cancers. The pedigree and the test results for the mutations and their pathogenic or VUS (variant of unknown significance) status are noted in the database.

Data storage and handling

A data entry program PCCM_DB is created as a command-line based entry system in python 3.6 [12] to enter patient data across the Breast Cancer diagnosis, treatment and follow-up workflow. The program is divided into modules based on the patient workflow as described above. Each workflow module described above represents a unit of the relational database. Surgical details are captured in a surgery datasheet by the assistant surgeon and filed at the clinic in patient files as a file_number, patient name, and surgery date labelled form until the database module is created. Radiology images are stored manually tagged by patient file number. These will be incorporated into the database and marked with appropriate identifiers in the future. For all other modules, each patient record is tagged with three unique identifiers, as mentioned earlier. Data is entered as a yes/no or multiple-choice question to maintain a defined vocabulary. For a limited number of descriptive questions (such as PET reports, NACT regime, treating doctor name), data is entered as free text. Information is stored in a separate SQLite3 database [13] for each data entry person and transferred to a common SQLite3 database at regular intervals. A tool has been created to extract data in a Microsoft Excel format with separate sheets that represent each module. Separate output files with de-identified data are exported out as excel files for researchers.

Tissue storage and handling

As part of the biobank, the tissue repository is built with FFPE tissue donated by the consenting patients. Biopsy and surgery tissue blocks and clinicopathological reports are received in the biobank facility and the signed informed consent. The biobank staff identifies the unique patient ID and histopathology report from the PCCM file system. Once the block ID written on the physical block is verified with the report and patient ID, the tissue blocks are placed in the designated tissue cabinet. Typically, one column in the tissue cabinet drawer is allotted to one patient. The blocks are arranged in alphabetical and/or numerical order. The placeholder and column header are allocated with three identifiers: Serial No - Patient ID - Block location ID: cabinet No-drawer No-column No. The placeholder or location ID, as it is referred to, is logged in the database as the storage coordinates. The tissue repository contains FFPE blocks with biopsy tissue, surgery tissue and adjacent normal tissue. For some of the cases, normal contralateral breast tissue derived from an oncoplastic surgery is deposited.

Genetic counselling and testing

The genetic clinic at the center was established in 2017 as per the National Comprehensive Cancer Network (NCCN) and American Society for Medical Genetics (ACMG) guidelines. Accordingly, we apply eligibility criteria for genetic counselling and testing of breast cancer patients and their family members for determining the risk of Hereditary Breast and Ovarian Cancer (HBOC). At the genetic clinic, a multidisciplinary team of onco-clinicians (i.e., radiologist, onco-surgeon or medical oncologist), geneticist and genetic counsellors collaborate to integrate genetic information with clinical decision-making algorithms.

On referral by onco-clinicians, eligible patients (diagnostic testing) and their relatives (preventive testing) undergo pre-test counselling to discuss the importance of genetic testing on cancer management and risk assessment. Detailed lifestyle, medical and 3-generational family history is collected along with informed consent for genetic testing. Using genomic DNA from blood samples, next-generation exome sequencing is performed using a multi-gene germline mutation consensus panel recommended by ACMG. Whenever required for cascade testing, Sanger sequencing is used. If clinically actionable germline mutations are found, a post-test counselling session is planned to explain the implications to the patient and family. With inputs from geneticists/counsellors, the onco-clinicians plan tailored surveillance and clinical management for such patients.

Demographic analysis of pccm and ffpe cohorts

Demographic and clinicopathological data of IDC patients were statistically analyzed using R Ver. 3.6.3 [14]. For clinical tumor size (cT), the longest tumor dimension reported in radiology reports was considered to define the T1-T3 category. The description of axillary nodes (loss of fatty hilum, node dimensions, number of enlarged nodes) in ultrasound and PET reports was assessed to derive cN status. The 7th edition AJCC guidelines were used to derive cT and cN [15]. In a follow-up analysis, cases with less than 30 days of follow up post-biopsy or surgery were reported as lost to follow-up.

The mean age of patients within three molecular subtypes based on the receptor expression is compared with One-way ANOVA. For all other parameters where data is available, the distribution of the subcategories within molecular subtypes is assessed by Chi-Square analysis. For menopausal status comparison, distribution within pre-and post-menopausal status is compared. P-value of <0.05 is taken as significant. Median follow up was calculated using the reverse KM method of Schemper and Smith [16].

Kaplan-Meir plots of patient survival at ten years (overall survival and disease-free survival) were plotted using SPSS (IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp). Significance at 95% confidence was estimated with Mantel Cox's Log Rank test. For overall survival, cases that were noted as Deceased at follow-up were taken as events. The time interval was estimated in months from biopsy to last follow-up or date of death when known. For disease-free survival, localized disease recurrence or systemic metastasis were taken as events. The time interval was taken as the time in months from surgery to last follow up or disease recurrence. Survival was further analyzed using Cox Proportional Hazard Analysis in R using the survival [17] and survminer [18] packages.

Results

Prashanti Cancer Care Mission (PCCM) is a public charitable trust established to provide affordable holistic cancer care. PCCM established a dedicated breast clinic managed by a single onco-surgeon that offers comprehensive breast cancer care. It is one of the few centers in India to provide oncoplastic surgery routinely. To understand the incidence, progression and response to therapy in an Indian cohort of breast cancer patients, a translational cancer research center (center for Translational Cancer Research – CTCR) was established in 2017 in collaboration with an academic institute, the Indian Institute of Science Education and Research (IISER Pune). The center established a biobank with breast tissue and patient clinical data in 2018 with appropriate guidelines and ethical approvals in place to aid clinical and translational research. This report is the first audit of the biobank to assess clinical and follow-up data along with the deposited tissue parameters.

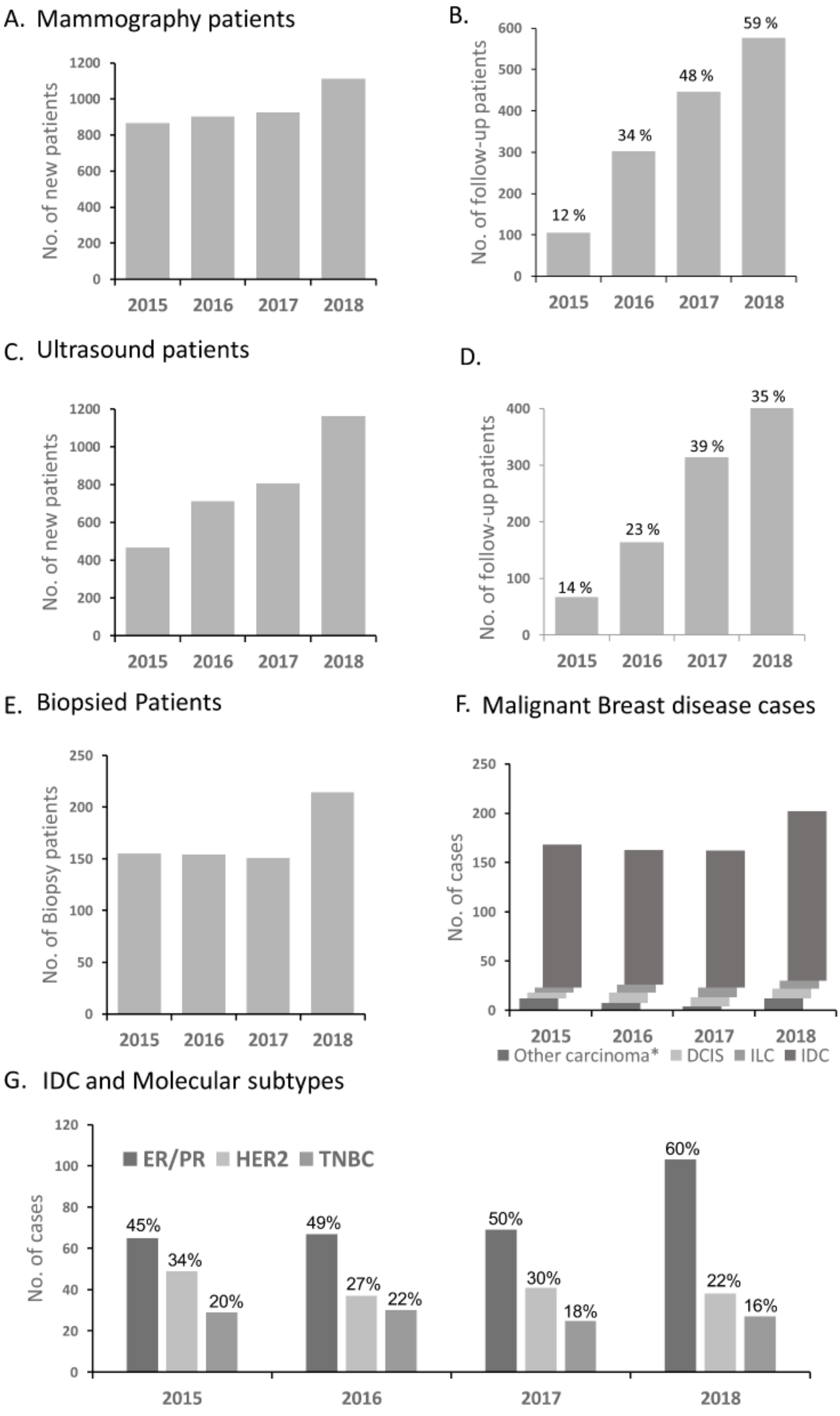


Fig. 1. Patient flow in numbers from testing to diagnosis of malignant breast disease in years 2015 to 2018 at the clinic. Number of unique female patients that were screened per year for breast abnormalities by mammography (A) or ultrasound (C). The number of unique entries for female patients with breast abnormalities that were tested for follow-up per year for mammography (B) and for ultrasound (D). The percentage of unique number of follow-up patients per year is reflected above each bar (B and D). The number of female patients per year that underwent biopsy to confirm the indication of a breast disease with mammography and/or ultrasound test (E). The number of female patients presented and treated at the clinic per year with malignant breast disease, diagnosed at the center or elsewhere (F). The distribution of IDC cases into molecular subtypes that were presented and treated at the clinic per year (G).

Diagnostic screening at the clinic

Diagnosis of breast abnormalities is performed at the clinic with in-house mammography (Siemens Revelation digital mammography with 3D tomosynthesis) and ultrasound equipment (Siemens Acuson S 2000) that have been operational since 2015. The data logs from mammography and ultrasound machines were extracted to identify the number of unique patients, and the number of follow-up patients screened every year. The number of female patients screened with mammography was on average 950 per year with a modest increment per year (Fig. 1A). The percent number of patients screened for follow-up showed a remarkable upward trend from 12% in 2015 to 59% in 2018 (Fig. 1B). The number of patients screened with ultrasound for breast disease increased from 467 in 2015 to 1161 in 2018 (Fig. 1C). The number of patients who underwent follow-up screening/test using ultrasound also went up from 14% to 35% from 2015 to 2018 (Fig. 1D). The number of patients biopsied at the clinic to confirm the breast disease indicated in screening was steady for 2015 to 2017, with a 25% increase in 2018 to 214 (Fig. 1E).

Breast disease composition of the diagnosed cases

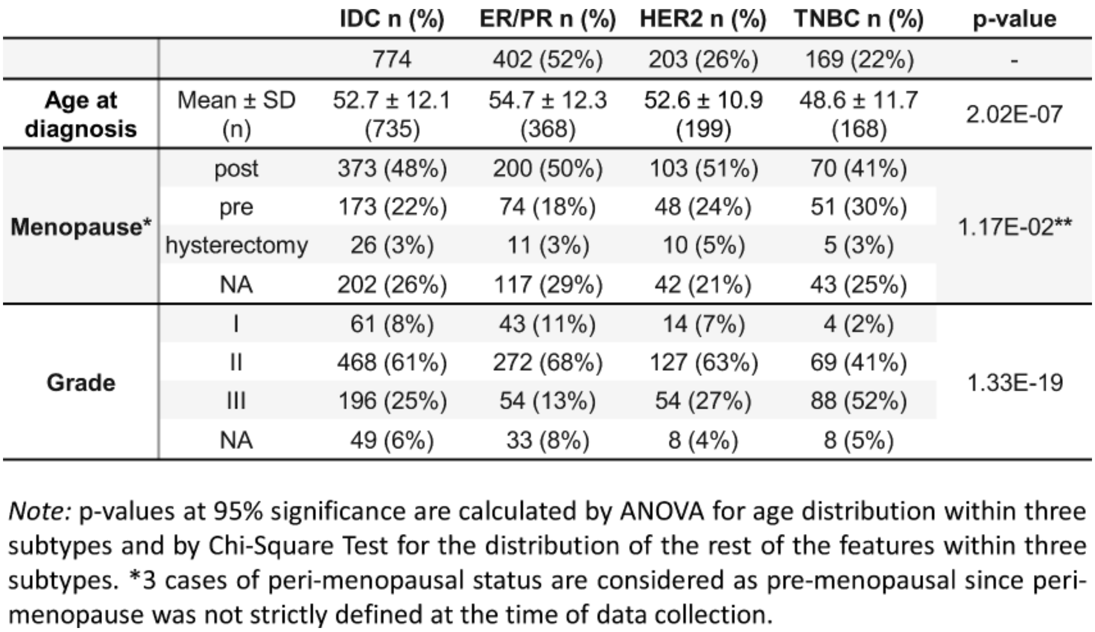
A total of 994 breast cancer cases were diagnosed and treated at the clinic, including benign and malignant disease. Benign breast tumours included fibroadenomas, phyllodes, etc. Malignant breast tumors included DCIS, Invasive lobular carcinoma, invasive papillary carcinoma and invasive ductal carcinomas. Very few cases presented with DCIS, as in India, only symptomatic patients visit for screening due to lack of any official screening program. The majority, i.e. 85% of the total diagnosed malignant tumors, are invasive ductal carcinoma types. The proportion of invasive/infiltrating ductal carcinoma (IDC) cases diagnosed with biopsy in 2015 was biased with 90% representation, which gradually dropped to 78% by 2018 (Fig. 1F). Molecular subtype identification of IDC cases was by immunohistochemistry, and HER2 FISH was performed at an allied pathology laboratory. Over the years from 2015 to 2018, the proportion of ER-positive IDC increased from 45.5% to 61.3%, while the proportion of HER2 positive and TNBC IDC cases reduced from 34.4% to 20.3% and 22.6% to 16.1%; respectively (Fig. 1G).

Demographic and clinicopathological details of idc cohort

The number of IDC cases diagnosed and with data deposited in the clinic from 2015 to 2018 is 580. A few patients who visited for follow-up at the clinic during these years but were diagnosed in the previous years also consented to the deposition of clinical data and breast tissue blocks to the biobank. Hence, clinical data for a total of 774 IDC patients is deposited in the biobank with appropriate consent from patients diagnosed between 2010 and 2018. The clinical data at this stage consisted of demographic details, histopathological diagnosis, and immunohistochemistry reports for molecular subtyping. The demographic information of 774 IDC cases is described in Fig. 2. Overall, the cohort contains 52% ER-positive, 26% HER2 positive and 22% TNBC cases. Mean age at incidence is observed to be significantly younger for TNBC patients (48.6 ± 11.7 yrs) as compared to HER2+ve (52.6 ± 10.9 yrs) and ER/PR+ve (54.7 ± 12.3yrs) patients. Similarly, TNBC patients comprised a higher proportion of premenopausal patients (50%), significantly higher than HER2+ve (24%) and ER/PR+ve (18%) patients. TNBC presented with grade 3 tumors 50% of the times, significantly higher than HER2+ve (27%) and ER/PR+ve (13%) tumors, which presented with higher proportions of grade 2 tumors. The TNBC subtype presents with a significantly higher proportion of aggressive features such as younger and premenopausal age at the incidence and high-grade tumors (Fig. 2). High proportion and aggressive features of TNBC at presentation in India have been reported often [19].

Breast tissue biobank

Patients diagnosed and treated at the clinic within 2010–2018 were actively approached for consenting and were requested to deposit their



		IDC n (%)	ER/PR n (%)	HER2 n (%)	TNBC n (%)	p-value
		774	402 (52%)	203 (26%)	169 (22%)	-
Age at diagnosis	Mean ± SD (n)	52.7 ± 12.1 (735)	54.7 ± 12.3 (368)	52.6 ± 10.9 (199)	48.6 ± 11.7 (168)	2.02E-07
Menopause*	post	373 (48%)	200 (50%)	103 (51%)	70 (41%)	1.17E-02**
	pre	173 (22%)	74 (18%)	48 (24%)	51 (30%)	
	hysterectomy	26 (3%)	11 (3%)	10 (5%)	5 (3%)	
	NA	202 (26%)	117 (29%)	42 (21%)	43 (25%)	
Grade	I	61 (8%)	43 (11%)	14 (7%)	4 (2%)	1.33E-19
	II	468 (61%)	272 (68%)	127 (63%)	69 (41%)	
	III	196 (25%)	54 (13%)	54 (27%)	88 (52%)	
	NA	49 (6%)	33 (8%)	8 (4%)	8 (5%)	

Note: p-values at 95% significance are calculated by ANOVA for age distribution within three subtypes and by Chi-Square Test for the distribution of the rest of the features within three subtypes. *3 cases of peri-menopausal status are considered as pre-menopausal since peri-menopause was not strictly defined at the time of data collection.

Fig. 2. Demographic table of IDC cases (2010–2018) with clinical data The number of IDC cases that were presented at and were treated at the clinic from 2010 till 2018 is compiled, and the demographic distribution of the clinical and pathological features is presented in the table. The total number of IDC cases are 799; out of those, 10 bilateral cases of IDC and 12 cases without IHC reports are not considered for this analysis. Out of 774, 593 cases were diagnosed in the years 2015 to 2018; the rest 181 cases were diagnosed prior to 2015. *p*-values at 95% significance are calculated by ANOVA for age distribution within the three subtypes and by Chi-Square Test for the distribution of the rest of the features within the three subtypes. Three cases with menopause recorded as peri-menopausal status were considered post-menopausal since perimenopause was not strictly defined at the time of data collection. Chi-square test for menopause status has been carried out for post vs premenopausal data.

biopsy and/or tumor tissue to the biobank post-treatment completion. Overall, 111 patients with benign tumors and 883 patients with malignant breast tumors (diagnosed between 2010 and 2018) have deposited clinical data to the repository. Of those, 72 patients with benign disease (Fig. 3A) and 482 patients with malignant disease (Fig. 3B) consented and deposited the tissue blocks. Deposited FFPE blocks include both biopsy and surgery tissue.

Benign tumour cases

Benign tumours are defined with histopathology report that included one of the following terms as part of the diagnosis: Benign, Fibroadenoma, Hyperplasia, Phyllodes, Cyst, Intraductal Papilloma, Mastitis, Inflammation Cancer (Fig. 3A). Out of 72 patients with the benign disease who consented to deposit the tissue, biopsy and surgery tissue blocks are deposited for 37 and 36 cases, respectively. In comparison, for six patients, both biopsy and surgery blocks are deposited (Fig. 4A). With the follow-up data, 18 patients with benign tumours as a primary disease were observed to return with the recurrence of benign disease post-surgery.

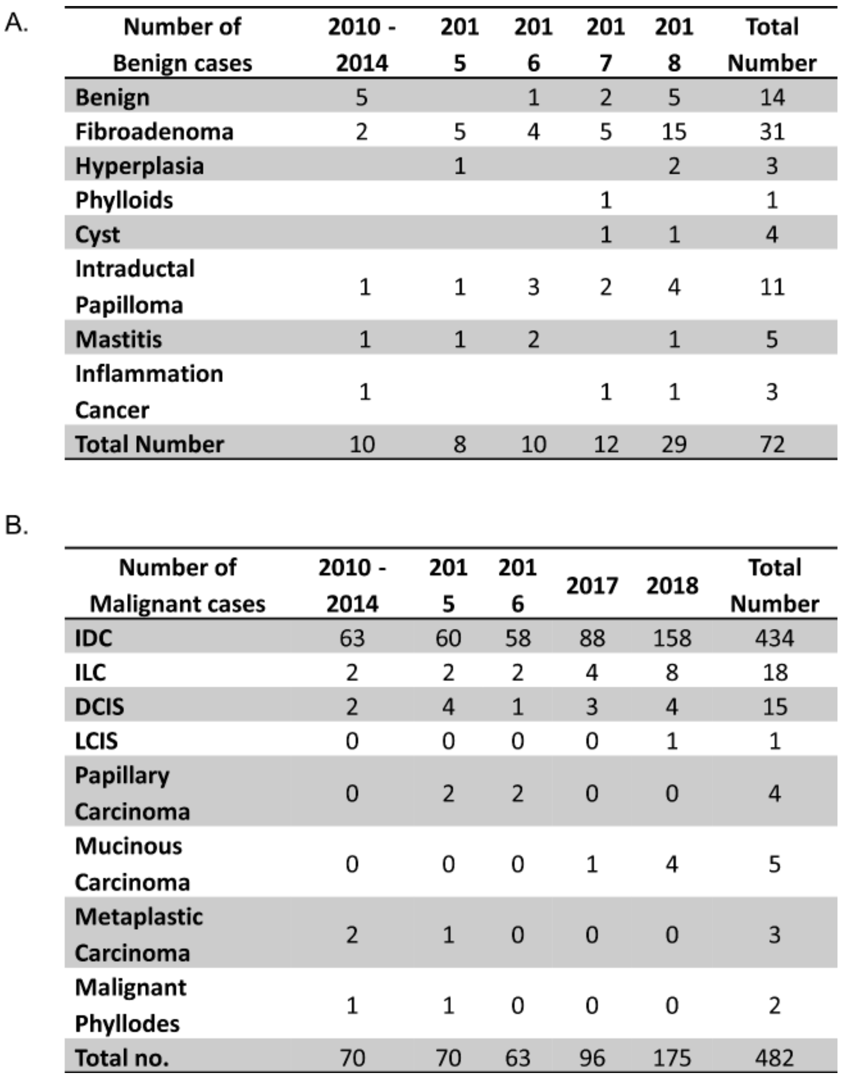
Malignant tumour cases

For malignant tumors, 482 out of 883 patients had access to and consented to deposit their FFPE tissue blocks. Of these 482 cases, 192 deposited biopsy tissue and 374 deposited surgery tissue, and 112 patients deposited both – biopsy and surgery tissue (Fig. 4A). A majority of the malignant tissue deposits comprise IDC tumors ($n = 434$), while only 48 are non-IDC diseases, including DCIS and ILC.

IDC cases

Of 434 IDC cases deposited with tumor tissue, 74.8% ($n = 325$) are primary tumors (Fig. 4B). The primary tumor tissue includes biopsy ($n = 170$) and treatment naïve surgery ($n = 155$). 142 patients deposited tumor blocks as post-treatment surgery, where 45 have deposited both primary tumor tissue and the post-treatment surgery tumor tissue. Molecular subtypes wise numbers of primary and post-treatment tumor deposits are tabulated in Fig. 4B.

One of the advantages of establishing a biobank at Prashanti Cancer Care Mission is access to contralateral normal breast tissue since the onco-surgeon on board is one of the pioneering oncoplastic surgeons in India. Within the first year of ethical clearance, the biobank has been



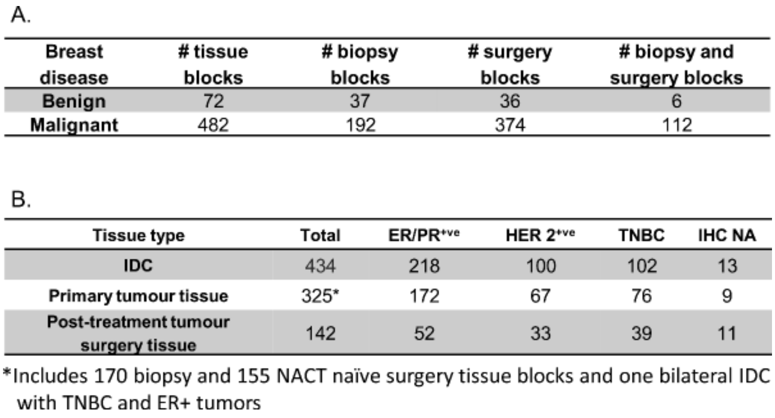
A.

Number of Benign cases	2010 - 2014	2015	2016	2017	2018	Total Number
Benign	5		1	2	5	14
Fibroadenoma	2	5	4	5	15	31
Hyperplasia		1			2	3
Phylloids				1		1
Cyst				1	1	4
Intraductal Papilloma	1	1	3	2	4	11
Mastitis	1	1	2		1	5
Inflammation Cancer	1			1	1	3
Total Number	10	8	10	12	29	72

B.

Number of Malignant cases	2010 - 2014	2015	2016	2017	2018	Total Number
IDC	63	60	58	88	158	434
ILC	2	2	2	4	8	18
DCIS	2	4	1	3	4	15
LCIS	0	0	0	0	1	1
Papillary Carcinoma	0	2	2	0	0	4
Mucinous Carcinoma	0	0	0	1	4	5
Metaplastic Carcinoma	2	1	0	0	0	3
Malignant Phyllodes	1	1	0	0	0	2
Total no.	70	70	63	96	175	482

Fig. 3. Clinical records deposited in the biobank with consent for patients with the Breast disease The number of patients per year with benign breast disease (A), malignant breast disease (B) that consented to deposit clinicopathological data in the biobank. The types of benign breast diseases and malignant breast diseases are listed in table A and table B, respectively. The mode of diagnosis of malignant breast diseases is tabulated.



A.

Breast disease	# tissue blocks	# biopsy blocks	# surgery blocks	# biopsy and surgery blocks
Benign	72	37	36	6
Malignant	482	192	374	112

B.

Tissue type	Total	ER/PR^{+ve}	HER 2^{+ve}	TNBC	IHC NA
IDC	434	218	100	102	13
Primary tumour tissue	325*	172	67	76	9
Post-treatment tumour surgery tissue	142	52	33	39	11

*Includes 170 biopsy and 155 NACT naïve surgery tissue blocks and one bilateral IDC with TNBC and ER+ tumors

Fig. 4. Number of breast tumor tissue deposited in the biobank with consent The number of patients with breast disease that deposited FFPE blocks with the breast tissue along with the clinicopathological data is tabulated. A. Number of cases with tissue blocks for benign disease and malignant disease. B. Number of IDC cases that deposited tumor tissue, pre and post-surgery. The numbers are tabulated for breast cancer and according to the subtype.

deposited with contralateral normal breast tissue from 40 breast cancer patients who underwent oncoplastic surgery as a part of the cancer treatment. Other than these, 85 of all IDC tissues deposited contain adjacent normal breast tissue.

Clinicopathological features of idc patients with tumor tissue

The demographic details and clinicopathological features of IDC patients with tumor tissue deposited in the biobank are summarized in Fig. 5. Out of 434 IDC cases, eight patients had bilateral IDCs with different clinical features of the disease and hence are not included in the summary. Within IDC tumor deposits, 52% are ER-positive, 23% HER2 positive, and 24% are TNBC tumors. TNBC tumor tissue may have been over-represented as a result of collection bias due to subtype targeted studies. Similar to the clinical cohort, TNBC patients present with significantly younger mean age and a higher proportion of high-grade disease as compared to patients with HER2+ve and ER+ve disease. Clinical tumor size and nodal involvement data could be retrieved for only 50% of patients. Within these, a majority (50%) of the patients presented with cT2 tumors, with equal distribution within N0-N2 nodal involvement. Amongst 413 IDC patients, 251 received surgery as a primary treatment, while 149 received NACT followed by surgery. Amongst the subtypes, 149 who received NACT comprised 29% ER+ve, 41% of HER2+ve and 46% of TNBC patients. Interestingly, 57% of the tumors where primary treatment was surgery are pathological tumor size T3 with N0 or N1 node status. A majority of the tumors that were removed post-treatment are T0-T2, with either N0 or N1 node involvement status (Fig. 5).

Germline mutations profile of hboc qualified idc patients

Genetic testing for index cancer patients is termed as diagnostic testing, while preventive/ cascade testing is used when healthy, unaffected individuals with a strong family history are tested. Between 2017–2018, 84 IDC patients underwent diagnostic testing, and 65 healthy unaffected individuals underwent preventing testing at the genetic clinic. The majority of IDC patients were diagnosed with the TNBC subtype.

Out of the 84 IDC patients tested for germline mutations, 40 were found to be carriers of *BRCA* pathogenic mutations (47.6%). Of these 40, 31 (77.5%) were *BRCA1* carriers, and 9 (22.5%) were *BRCA2* carriers (Fig. 6). Twelve cases were found to carry *BRCA* variants of unknown significance (VUS) (5 for *BRCA1*; 7 for *BRCA2*). Since we use a multigene panel for HBOC risk profiling, pathogenic mutations were also observed in non-BRCA genes (6%, 9/149) that include *ATM, APC, BRIP1, CHEK2, CDH1, MLH1, MSH2, MSH6, NBN, NF1, PALB2, PTEN, RAD51D, STK11, TP53* (Fig. 6). A large number of VUS (38%, 32/84) were observed in the non-BRCA genes.

Amongst the unaffected individuals, 4 out of 65 individuals were positive for *BRCA1 and BRCA2* pathogenic mutations each. 2 VUS were found in *BRCA2* and 3 in non-*BRCA* genes. These BRCA1/2 carriers were offered appropriate advice for medical management (for index patients) and risk mitigation (for unaffected family members) as per NCCN guidelines. No further medical interventions were offered for BRCA 1/2 mutation carriers with variants of unknown significance. As both BRCA and non-BRCA genes are involved in cancer risk to other organs, appropriate counselling and medical advice were provided for risk mitigation via medical surveillance. Oncological follow-up data for these index patients are maintained in the database. The clinic has a quarterly follow-up schedule for the cascades (unaffected, healthy individuals) to assess the surveillance and oncological follow-up.

Follow-up data of breast cancer patients

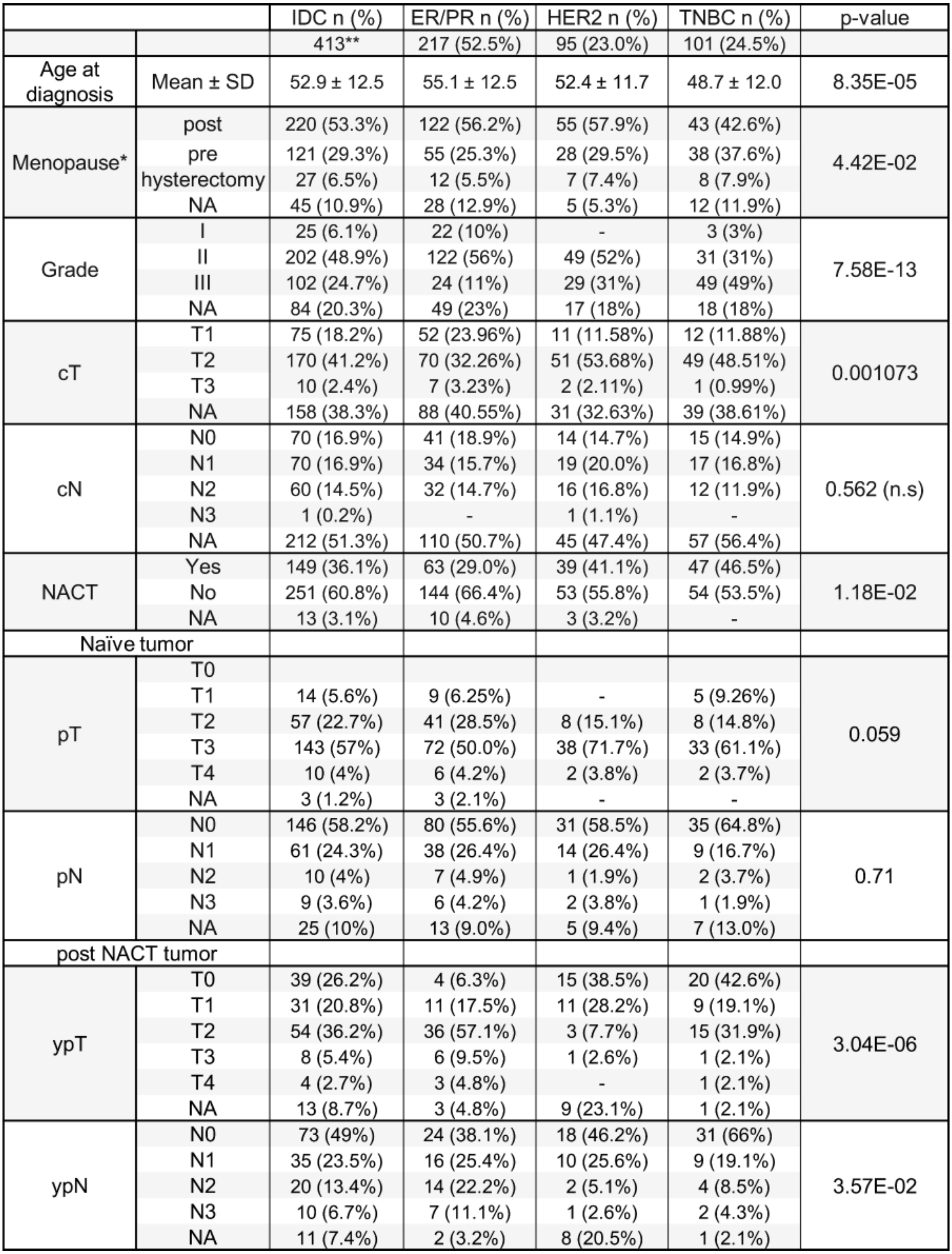
Detailed file keeping and data curation has revealed three of the initial DCIS cases presented with IDC later in their yearly follow-up check-ups. Two of the three cases presented with IDC in 3rd year, while one case presented with IDC in the 4th year. Both DCIS and IDC diagnosed tissues are consented to and are preserved in our tissue repository.

The clinic follows a quarterly follow-up regime for the first year post-surgery and an annual check-up thereafter for malignant cases. Amongst the patients with IDC who consented to deposit clinical data ($n = 799$), 620 patients (77.5%) have follow-up information deposited. The average follow-up in the database is 35 months since diagnosis and 36 months since surgery, with a median follow-up of 28 months, respectively (Fig. 7A). The number of patients lost to follow-up post-diagnosis is 82, and post-surgery is 44, i.e. less than 10%.

For 434 patients with tumor tissue deposited, 369 cases have follow-up information available with an average of 30 months post-diagnosis as well as post-surgery and a median follow-up of 26 and 25 months, respectively (Fig. 7B). Of the 369 cases with follow-up data deposited, 11 returned with metastatic disease. For the relapsing cases, both primary and recurrent tumor tissue is deposited in the biobank.

Survival outcomes of breast cancer patients

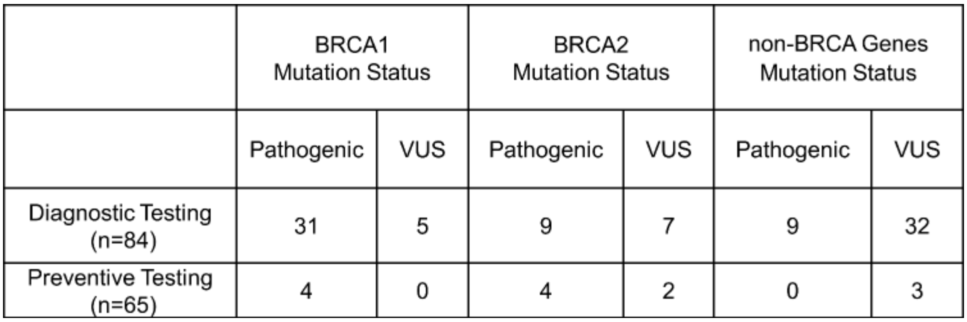
To assess survival outcomes of the cohort, Overall survival and disease-free survival are computed for IDC patients from the clinical database cohort and for the tissue biobank cohort. As reported in Table 7, the clinical database with 799 IDC patients has longer follow-up information than the tissue biobank cohort. Three molecular subtypes



		IDC n (%)	ER/PR n (%)	HER2 n (%)	TNBC n (%)	p-value
		413**	217 (52.5%)	95 (23.0%)	101 (24.5%)	
Age at diagnosis	Mean ± SD	52.9 ± 12.5	55.1 ± 12.5	52.4 ± 11.7	48.7 ± 12.0	8.35E-05
Menopause*	post	220 (53.3%)	122 (56.2%)	55 (57.9%)	43 (42.6%)	4.42E-02
	pre	121 (29.3%)	55 (25.3%)	28 (29.5%)	38 (37.6%)	
	hysterectomy	27 (6.5%)	12 (5.5%)	7 (7.4%)	8 (7.9%)	
	NA	45 (10.9%)	28 (12.9%)	5 (5.3%)	12 (11.9%)	
Grade	I	25 (6.1%)	22 (10%)	-	3 (3%)	7.58E-13
	II	202 (48.9%)	122 (56%)	49 (52%)	31 (31%)	
	III	102 (24.7%)	24 (11%)	29 (31%)	49 (49%)	
	NA	84 (20.3%)	49 (23%)	17 (18%)	18 (18%)	
cT	T1	75 (18.2%)	52 (23.96%)	11 (11.58%)	12 (11.88%)	0.001073
	T2	170 (41.2%)	70 (32.26%)	51 (53.68%)	49 (48.51%)	
	T3	10 (2.4%)	7 (3.23%)	2 (2.11%)	1 (0.99%)	
	NA	158 (38.3%)	88 (40.55%)	31 (32.63%)	39 (38.61%)	
cN	N0	70 (16.9%)	41 (18.9%)	14 (14.7%)	15 (14.9%)	0.562 (n.s)
	N1	70 (16.9%)	34 (15.7%)	19 (20.0%)	17 (16.8%)	
	N2	60 (14.5%)	32 (14.7%)	16 (16.8%)	12 (11.9%)	
	N3	1 (0.2%)	-	1 (1.1%)	-	
	NA	212 (51.3%)	110 (50.7%)	45 (47.4%)	57 (56.4%)	
NACT	Yes	149 (36.1%)	63 (29.0%)	39 (41.1%)	47 (46.5%)	1.18E-02
	No	251 (60.8%)	144 (66.4%)	53 (55.8%)	54 (53.5%)	
	NA	13 (3.1%)	10 (4.6%)	3 (3.2%)	-	
Naïve tumor						
pT	T0					0.059
	T1	14 (5.6%)	9 (6.25%)	-	5 (9.26%)	
	T2	57 (22.7%)	41 (28.5%)	8 (15.1%)	8 (14.8%)	
	T3	143 (57%)	72 (50.0%)	38 (71.7%)	33 (61.1%)	
	T4	10 (4%)	6 (4.2%)	2 (3.8%)	2 (3.7%)	
	NA	3 (1.2%)	3 (2.1%)	-	-	
pN	N0	146 (58.2%)	80 (55.6%)	31 (58.5%)	35 (64.8%)	0.71
	N1	61 (24.3%)	38 (26.4%)	14 (26.4%)	9 (16.7%)	
	N2	10 (4%)	7 (4.9%)	1 (1.9%)	2 (3.7%)	
	N3	9 (3.6%)	6 (4.2%)	2 (3.8%)	1 (1.9%)	
	NA	25 (10%)	13 (9.0%)	5 (9.4%)	7 (13.0%)	
post NACT tumor						
ypT	T0	39 (26.2%)	4 (6.3%)	15 (38.5%)	20 (42.6%)	3.04E-06
	T1	31 (20.8%)	11 (17.5%)	11 (28.2%)	9 (19.1%)	
	T2	54 (36.2%)	36 (57.1%)	3 (7.7%)	15 (31.9%)	
	T3	8 (5.4%)	6 (9.5%)	1 (2.6%)	1 (2.1%)	
	T4	4 (2.7%)	3 (4.8%)	-	1 (2.1%)	
	NA	13 (8.7%)	3 (4.8%)	9 (23.1%)	1 (2.1%)	
ypN	N0	73 (49%)	24 (38.1%)	18 (46.2%)	31 (66%)	3.57E-02
	N1	35 (23.5%)	16 (25.4%)	10 (25.6%)	9 (19.1%)	
	N2	20 (13.4%)	14 (22.2%)	2 (5.1%)	4 (8.5%)	
	N3	10 (6.7%)	7 (11.1%)	1 (2.6%)	2 (4.3%)	
	NA	11 (7.4%)	2 (3.2%)	8 (20.5%)	1 (2.1%)	

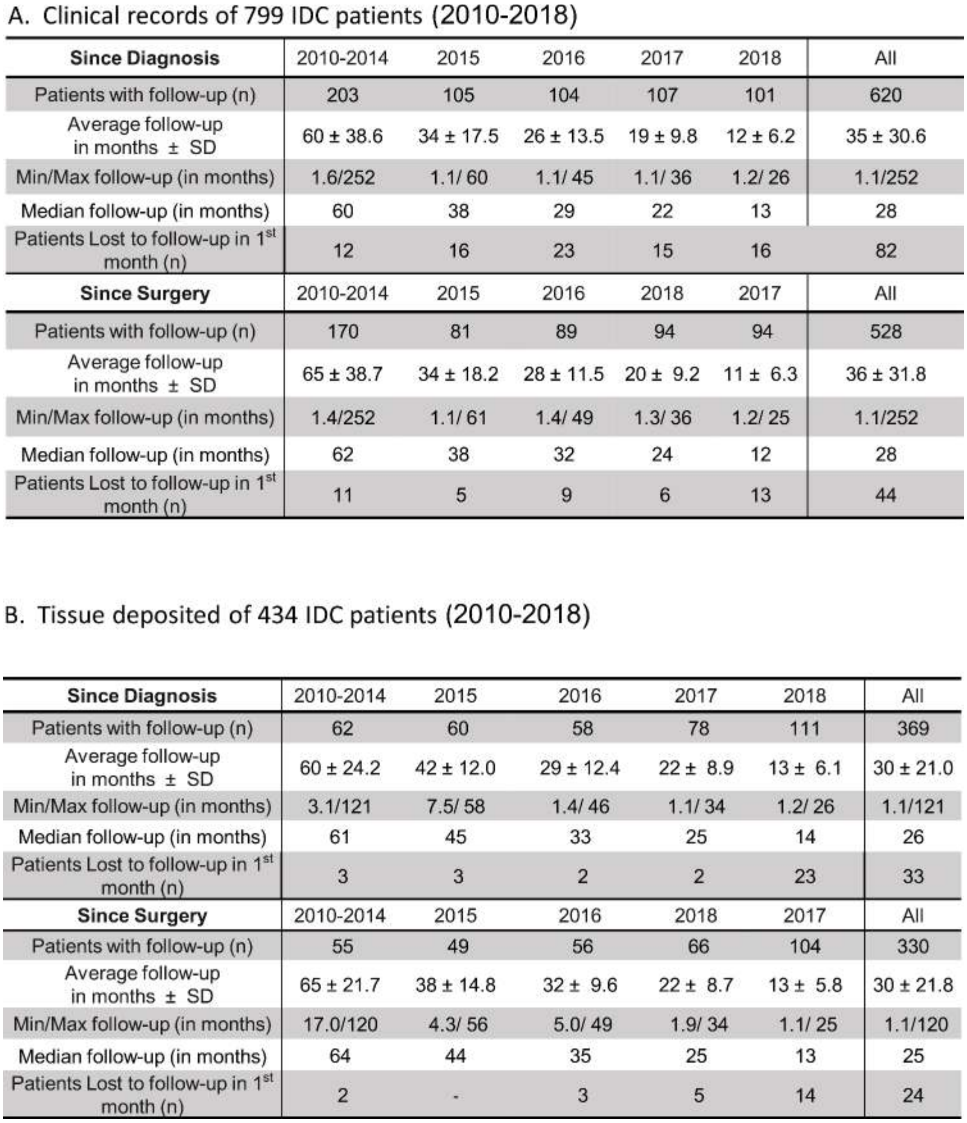
Fig. 5. Demographic table of IDC cases (2010–2018) with tumor tissue Clinical tumor and node status were derived from sizes and observations in diagnostic mammography (9%), ultrasound (87%) or PET (4%) reports and based on AJCC guidelines (8th AJCC guideline). ypT and ypN refer to the post-treatment pathological tumor and node status. All *p*-values except for age represent 95% confidence levels for Chi-square tests for distribution of the parameter amongst the three subtypes. 13 Peri menopausal cases have been considered as post-menopausal. *Chi-square test for menopause status has been carried out for post vs premenopausal data. One way ANOVA was computed for age distribution across subtypes. **8 bilateral samples have not been included in this analysis.

varied significantly in overall outcomes, with worse outcomes over 10 years for TNBC (Fig. 8A), followed by HER2 patients. ER-positive subtype showed better overall survival but not disease-free survival (Fig. 8A). Hazard Ratios for both HER2 and TNBC indicate worse prognosis as compared to ER-positive subtype in overall (HER2 4.42 [CI95%: 1.33–14.81]; TNBC: 6.20 [CI95%: 1.99 −19.23]) and disease-free survival HER2: 1.31 [CI95%: 0.78–2.26; TNBC: 1.81 [CI95%: 1.08–3.01]).



	BRCA1 Mutation Status		BRCA2 Mutation Status		non-BRCA Genes Mutation Status	
	Pathogenic	VUS	Pathogenic	VUS	Pathogenic	VUS
Diagnostic Testing (n=84)	31	5	9	7	9	32
Preventive Testing (n=65)	4	0	4	2	0	3

Fig. 6. HBOC screening summary Data is representative of diagnostic and preventive testing activities at PCCM genetic clinic (2017–18) as per NCCN guidelines on HBOC risk assessment. After pre-test counselling and patient consent, multi-gene germline mutation panels are used for exome NGS or Sanger sequencing for target-specific mutations. Mutations are deemed pathogenic or variants of unknown significance (VUS) after comprehensive bioinformatics analysis as per ACMG guidelines. The non-BRCA genes include *ATM, APC, BRIP1, CHEK2, CDH1, MLH1, MSH2, MSH6, NBN, NF1, PALB2, PTEN, RAD51D, STK11, TP53*. While post-test counselling is undertaken for assessment of risk in family members, onco-clinicians advise appropriate medical management protocols for patients that are carriers of pathogenic mutations.



A. Clinical records of 799 IDC patients (2010-2018)

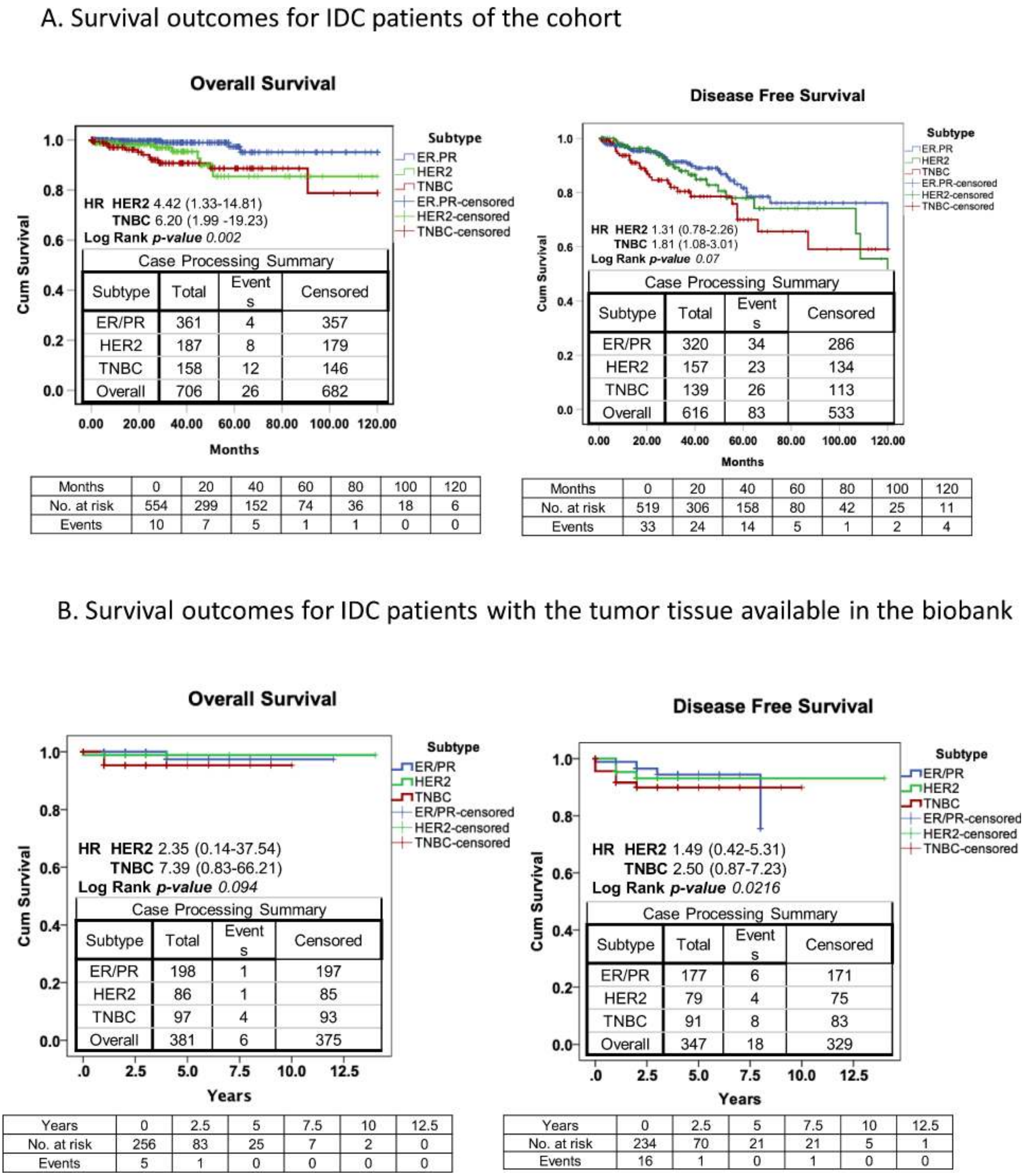
Since Diagnosis	2010-2014	2015	2016	2017	2018	All
Patients with follow-up (n)	203	105	104	107	101	620
Average follow-up in months ± SD	60 ± 38.6	34 ± 17.5	26 ± 13.5	19 ± 9.8	12 ± 6.2	35 ± 30.6
Min/Max follow-up (in months)	1.6/252	1.1/ 60	1.1/ 45	1.1/ 36	1.2/ 26	1.1/252
Median follow-up (in months)	60	38	29	22	13	28
Patients Lost to follow-up in 1st month (n)	12	16	23	15	16	82
Since Surgery	2010-2014	2015	2016	2018	2017	All
Patients with follow-up (n)	170	81	89	94	94	528
Average follow-up in months ± SD	65 ± 38.7	34 ± 18.2	28 ± 11.5	20 ± 9.2	11 ± 6.3	36 ± 31.8
Min/Max follow-up (in months)	1.4/252	1.1/ 61	1.4/ 49	1.3/ 36	1.2/ 25	1.1/252
Median follow-up (in months)	62	38	32	24	12	28
Patients Lost to follow-up in 1st month (n)	11	5	9	6	13	44

B. Tissue deposited of 434 IDC patients (2010-2018)

Since Diagnosis	2010-2014	2015	2016	2017	2018	All
Patients with follow-up (n)	62	60	58	78	111	369
Average follow-up in months ± SD	60 ± 24.2	42 ± 12.0	29 ± 12.4	22 ± 8.9	13 ± 6.1	30 ± 21.0
Min/Max follow-up (in months)	3.1/121	7.5/ 58	1.4/ 46	1.1/ 34	1.2/ 26	1.1/121
Median follow-up (in months)	61	45	33	25	14	26
Patients Lost to follow-up in 1st month (n)	3	3	2	2	23	33
Since Surgery	2010-2014	2015	2016	2018	2017	All
Patients with follow-up (n)	55	49	56	66	104	330
Average follow-up in months ± SD	65 ± 21.7	38 ± 14.8	32 ± 9.6	22 ± 8.7	13 ± 5.8	30 ± 21.8
Min/Max follow-up (in months)	17.0/120	4.3/ 56	5.0/ 49	1.9/ 34	1.1/ 25	1.1/120
Median follow-up (in months)	64	44	35	25	13	25
Patients Lost to follow-up in 1st month (n)	2	-	3	5	14	24

Fig. 7. Follow-up data of IDC patients The table summarises the number of IDC patients for whom follow-up information is available in the biobank. The follow-up summary is presented in two parts, A. for the clinical data cohort ($n = 799$) and B. for tissue and clinical data cohort ($n = 434$). Time to follow-up is derived from the time of diagnosis as well as from the time of treatment, which is surgery in cases of IDC patients. The columns depict yearly and overall summary numbers.

The tissue biobank cohort is a recent subset of the clinical database cohort, and although with comprehensive follow-up, the duration of follow-up for most of the number of patients is still within two years. Overall survival for all three subtypes is within 90%, while disease-free survival worse for TNBC as compared to HER2 positive and ER-positive patients (Fig. 8B). Hazard ratios indicate worse prognosis for HER2 and



A. Overall Survival — Case Processing Summary

Subtype	Total	Events	Censored
ER/PR	361	4	357
HER2	187	8	179
TNBC	158	12	146
Overall	706	26	682

Months	0	20	40	60	80	100	120
No. at risk	554	299	152	74	36	18	6
Events	10	7	5	1	1	0	0

A. Disease Free Survival — Case Processing Summary

Subtype	Total	Events	Censored
ER/PR	320	34	286
HER2	157	23	134
TNBC	139	26	113
Overall	616	83	533

Months	0	20	40	60	80	100	120
No. at risk	519	306	158	80	42	25	11
Events	33	24	14	5	1	2	4

B. Overall Survival — Case Processing Summary

Subtype	Total	Events	Censored
ER/PR	198	1	197
HER2	86	1	85
TNBC	97	4	93
Overall	381	6	375

Years	0	2.5	5	7.5	10	12.5
No. at risk	256	83	25	7	2	0
Events	5	1	0	0	0	0

B. Disease Free Survival — Case Processing Summary

Subtype	Total	Events	Censored
ER/PR	177	6	171
HER2	79	4	75
TNBC	91	8	83
Overall	347	18	329

Years	0	2.5	5	7.5	10	12.5
No. at risk	234	70	21	21	5	1
Events	16	1	0	1	0	0

Fig. 8. Survival Outcomes of the IDC patients Kaplan Meier (KM) curves for overall and disease-free survival are plotted for IDC patients according to their molecular subtypes. Significance is estimated with Mantel Cox's Log Rank test. A risk table is provided for each KM plot. Hazard ratios (HR) with 95% confidence limits are derived from Cox proportional hazards analysis. The ER-positive subtype is taken as the reference value to calculate the hazard ratio for HER2 and TNBC subtypes. A. IDC cohort with consented clinical data deposited in the biobank. B. IDC cohort with consented data and tumor tissue deposited in the biobank.

TNBC patients compared to ER-positive patients for overall (HER2: 2.35 [CI95%: 0.14–37.54]; TNBC 7.39 [CI95%: 0.83–66.21]) and disease-free (HER2: 1.49 [CI95%: 0.42–5.31]; TNBC: 2.50 [CI95%: 0.87–7.23]) survival.

Comparison of idc cohorts from the biobank with TCGA

The clinical data deposited in our biobank cab be a representative subset of the breast cancer cohort in India. We, therefore, compared the demographic and subtype distribution of our cohort with that from the TCGA breast cancer dataset [20, 21] primarily (71%), a western cohort (Supplementary Figure S1 and S2). Overall clinical features (Fig. 2) of our IDC cohort were compared to that of TCGA IDC cases (Supplementary Figure S1A). The IDCs in our biobank had significantly different subtype distribution ($p = 1.54 \times 10^{-9}$) with a greater proportion of HER2 (26% vs 13%, $p = 1.65 \times 10^{-14}$) and TNBC cases (22% vs 18%, $p = 0.028$). IDC patients were also found to be collectively younger at diagnosis than TCGA IDC cases (52.7 ± 12.1 vs 57.7 ± 13.3, $p =$

4.3810^{-11}). There was no significant difference ($p = 0.098$) in the distribution of menopause status (post-menopausal and premenopausal) between the two datasets.

To compare survival characteristics of TCGA, we used the PANCAN clinical data resource dataset [22]. Supplementary Figure S1B shows the comparison of follow-up data for TCGA to the biobank data in (Fig. 7A). Our efforts reflect far better follow-up numbers, where the number of patients that are lost to follow-up within 1st month of the surgery is considerably less in the biobank (only 10.5%) compared to that of TCGA (44%). Data collection in the TCGA IDC cohort spans 30 years (from 1989 to 2003) across 164 sources, while the PCCM data is more recent (2010–2018) and from one source. This may have contributed to better follow-up rates in our cohort.

Kaplan-Meir plots for overall survival and disease-free survival (Supplementary data S2A and SB) also show a quite different distribution of survival outcomes compared to that of our cohort (Fig. 8A). Our cohorts show better overall survival, especially for ER/PR^{+}subtype, possibly due to the recent advancements in ER targeted therapies in the past decade. The disease-free outcomes, though, are worse in all the subtypes in our cohort compared to that of TCGA, something that needs to be followed up.

Summary of the biobank

To summarize, a total number of 994 clinical records are available in the biobank with breast tumors. 11% ($n = 111$) of those are benign tumors and 89% ($n = 883$) are malignant breast tumors. The biobank is biased, with 88% ($n = 799$) of the malignant cases with IDC disease, while only 11% ($n = 84$) represent non-IDC disease. Within IDC deposits, 50% of the clinical data, as well as tissue deposits, are of ER-positive subtype, 26–23% are HER2 positive, and 22–24% are of TNBC subtype (Fig. 9). The biobank has follow-up information for 77% ($n = 620$) of IDC patients with clinical data alone, while 85% ($n = 369$) of IDC patients who have deposited tumor tissue along with the clinical data.

Ongoing utilization of the bio-banked data

Since its inception, five independent research projects have been initiated with ethical approvals to profile breast cancer tumors at the molecular and genomic level and explore novel molecular markers.

The tissue repository has been a valuable resource to assess the distribution of infiltrating tumor lymphocytes (TILs) in molecular subtypes of IDC and their association with response to chemotherapy [23]. Knowing higher prevalence and aggressive presentation of TNBC, the center has undertaken the detailed characterization of TNBC tumor tissue for known and novel markers by IHC.

Bio-banked data is also proving invaluable to audit surgical techniques employed in the clinic and their evolution through the last few years. As one of the few centers to offer oncoplastic surgery in India, it is essential that we examine and audit various aspects of the surgical practice and its oncological and cosmetic outcomes. At the present time, an in-depth study of 131 NACT cases operated from 2015 to 2019 has been undertaken to analyze the effects of oncoplastic surgery on breast conservation post-NACT. Given the high rate of late-stage cancer presentation in India, mastectomy is most often the surgery offered to patients. However, mastectomies have been shown to have an adverse effect on psychosocial outcomes for the patient. To improve this, increased breast conservation using oncoplastic techniques or immediate breast reconstruction after a conservative skin or nipple-sparing mastectomy are practiced at PCCM. An audit of breast reconstructions

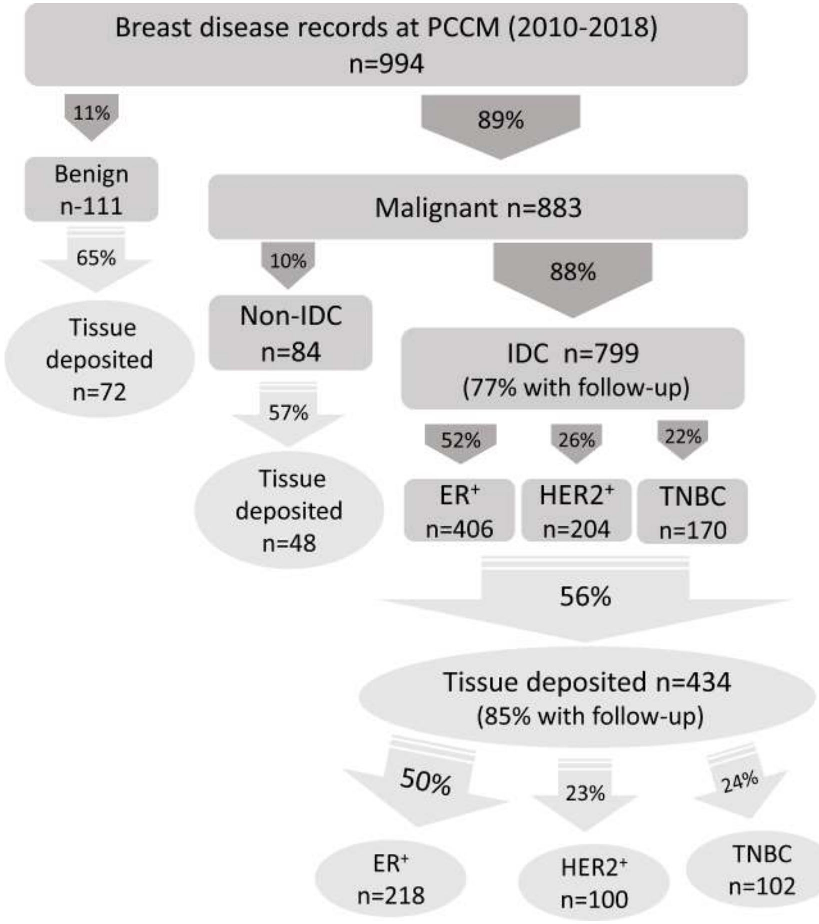


Fig. 9. Summary Flow Chart of clinical and tissue records in the biobank flowchart to summarize the clinical and tissue records in the biobank for the years 2010 to 2018. The data distribution for benign disease and malignant disease is represented in absolute numbers and percentages. The malignant disease records for non-IDC and IDC cases, which are reported according to their molecular subtypes. Finally, the percent no. of records with follow-up information is mentioned with each category.

carried out from 2010 – 2019 is currently underway, making use of the bio-banked data and the various data collection methods reported here. The clinical databases generated through the PCCM biobank have been valuable to report the Breast Onco-plasty efforts in India for analysis of clinical outcomes [11, 24, 25]. Further to this, a detailed audit of the oncoplastic surgery will be prepared with the biobank data, which is will 1st of a kind audit from the country.

Given the high prevalence of TNBCs in India [19] and strong links between TNBC phenotype and BRCA1/2 mutational status, PCCM has established a TNBC germline mutation cohort linked to the PCCM biobank. As a result, spin-off research projects for understanding the influence of biological phenomenon (i.e. BRCAness and homologous recombination deficiency) on clinical outcomes of TNBC patients are underway.

The PCCM radiology database is richly populated with high-resolution images from 2D digital mammograms with 3D tomosynthesis, ultrasonography and automated breast volume scanner from a large number of breast cancer patients and unaffected individuals. This image repository has led to research projects aimed at understanding (a) differentiating radiological features of IDC for development of AI and ML model (b) radio-genomics studies to investigate unique mammographic and ultrasound features of TNBCs with BRCA1/2 mutations (c) technical capability of USG and mammography for applications in planning and monitoring of NAST outcomes.

In future, a Graphical User Interface (GUI) customized portal will be developed that will host the de-identified clinical data and link to the availability and location of tissue resources. Clinical data will be associated with the relevant clinical and pathological reports to facilitate data quality checks. Radiological images, where available, will be directly linked to each case. Clinical observations, reports and follow-up will also be rapidly linked to the database. Surgical procedural images and cosmetic scores (PROMS) will also be entered into the database with appropriate patient approvals. The database will therefore have the most up-to-date information on the patient. This will facilitate clinical research as described above since the creation of clinical feature specific cohorts (e.g. the case of central quadrant tumors, surgeries using specific techniques Perforator flaps) will be based on queries to a single database. Currently, such information is scattered across databases (albeit often linked by common identifiers). Finally, this close linkage of clinical data and tissue resources at our center will surely facilitate and give impetus to translational projects to characterize breast cancer in Indian patients.

Conclusion

Tissue repositories with annotated patient data have been proven to

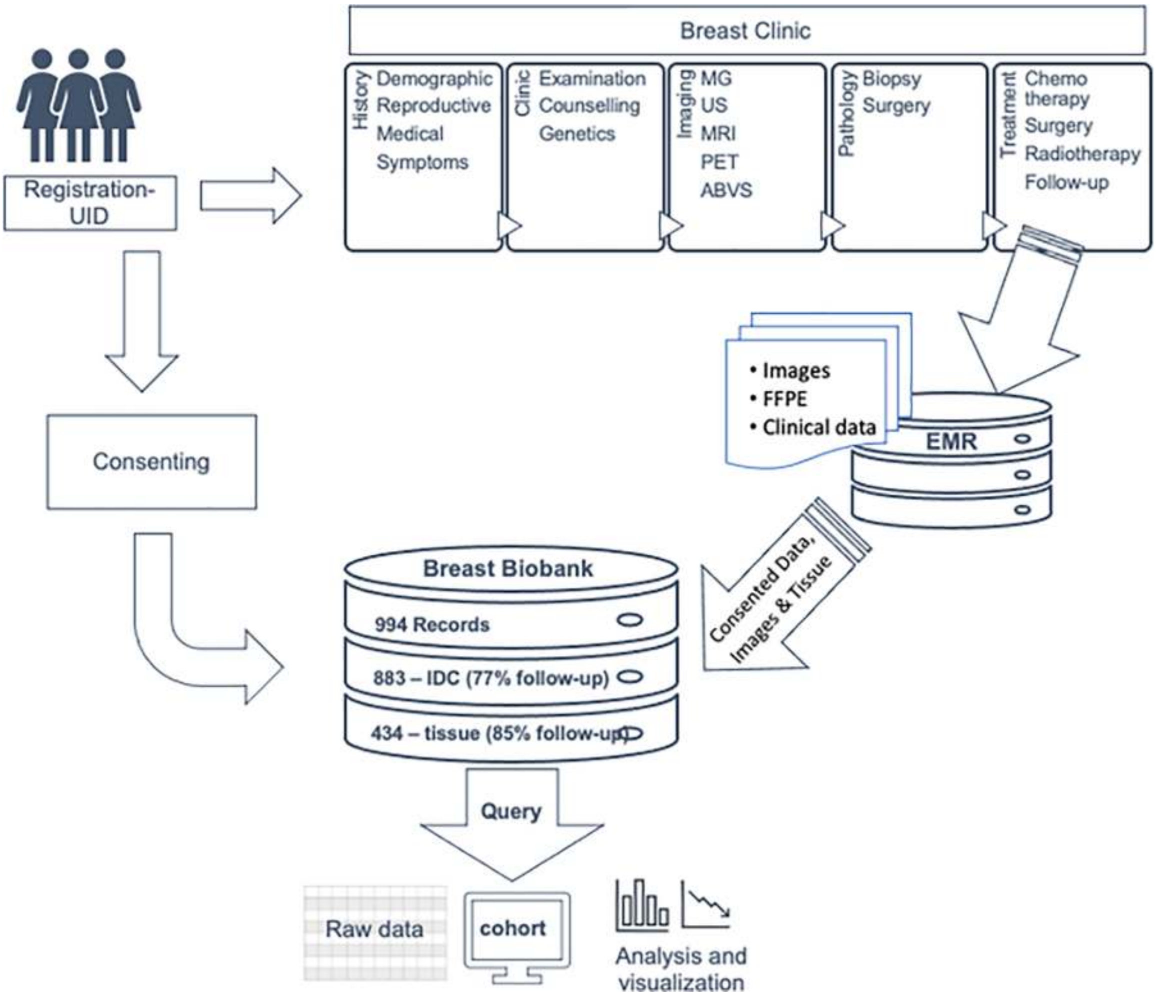


Fig. 10. Structured Summary of the biobanking process At registration, after the generation of a UID, patients are requested for consent to use clinical data and tissues for research. At the clinic, data for every patient is collected at presentation, examination, counselling, diagnosis, imaging investigations, histopathology, treatment, and follow up. At recurrence, follow-up and treatment regime is recorded. Data, images and tissue for patients who have consented are transferred at regular intervals by trained curators to the database and the biobank. Currently, the biobank holds 994 patient records, of which 883 are Invasive Ductal Carcinoma cases (77% have follow-up). Of these, 434 have corresponding tissue (biopsy and/or surgery) with 85% follow-up. The biobank is subjected to specific queries from researchers to create cohorts that are then analyzed and reported.

be of immense importance to understand cancer profiles and identify targeted and personalized therapies (e.g. TCGA). Ours will prove one such valuable resource within the country as well as globally to aid understanding of breast cancer in the Indian context (Fig. 10). This report may serve useful for upcoming and future breast cancer research projects to identify a specific breast cancer cohort to explore a scientific/clinical question.

Abbreviations

Not Applicable.

Declarations

Ethics approval and consent to participate

Ethics approval (dated 21st July 2018) is obtained from the ethics committee for the consent forms, both in English and vernacular language (Marathi), and to collect and store consented and de-identified clinical data and tissue material.

The 'Institutional Ethics Committee' for the institute 'Prashanti Cancer Care Mission Pune' is approved by DCGI – Drug Controller General of India under the registration number of 'ECR/298/Indt/MH/2018′.

Consent for publication

Not applicable, as no personal information is deposited in the database.

Availability of data and materials

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Author contributions

LB initiated the biobank set-up and coordinated patient follow-up, SD coordinated ethical approval, SN, RA, MJ, RN, PS are clinical nurse staff who are instrumental in collecting informed consent with patient information and follow-up, GT recorded chemotherapy information, RB, RU, RR, SS curated the clinical data and diagnostic reports, AK, NN, AB curated genetic testing data and family tree, AP, RJ, RK, WK plotted the data, BV and PD (deceased) collected radiology data, CD oncologist on board, DK generated SQL database for data curation and trained the staff to curate the data, LSS conceptualized the biobank, CBK is an onco-surgeon on board, MK generated a system for collection and organised storage of consented tissue, organised data, supervised data analysis and wrote the manuscript. All authors have read and approved the manuscript.

Funding

MK would like to acknowledge the DBT- Ramalingaswami' re-entry' fellowship awarded by DBT-India. LSS would like to acknowledge DST-JC Bose Research Fellowship. Research grant to Dr Koppiker is supported by Bajaj Auto Ltd.

Funding from Bajaj Auto Ltd is instrumental in building the infrastructure and manpower needed for storage of FFPE tissue material as well as building digital infrastructure to store consented clinical data. DBT-Ramalingaswami Fellowship and DST-JC Bose fellowship supports peripheral activities associated with the biobank and research projects.

CRediT authorship contribution statement

Laleh Busheri: Conceptualization, Project administration. **Santosh Dixit:** Methodology, Supervision, Writing - review & editing. **Smeeta Nare:** Methodology, Resources. **Rashmi Alhat:** Methodology, Resources. **George Thomas:** Methodology, Resources. **Mangal Jagtap:** Methodology, Resources. **Ruth Navgire:** Methodology, Resources. **Priya Shinde:** Methodology, Resources. **Rituja Banale:** Data curation. **Rohini Unde:** Data curation. **Ruhi Reddy:** Data curation. **Shahin Shaikh:** Data curation. **Aishwarya Konnur:** Data curation. **Namrata Namewar:** Data curation. **Ashwini Bapat:** Data curation. **Ankita Patil:** Formal analysis. **Rateeka Johari:** Formal analysis. **Roli Kushwaha:** Formal analysis. **Wimpy Kumari:** Formal analysis. **Beenu Varghese:** Investigation. **Pooja Deshpande:** Investigation. **Chetan Deshmukh:** Investigation. **Devaki A. Kelkar:** Software, Data curation, Formal analysis, Supervision, Writing - review & editing. **L S Shashidhara:** Project administration, Funding acquisition. **Chaitanyanand B Koppiker:** Project administration, Funding acquisition. **Madhura Kulkarni:** Supervision, Formal analysis, Funding acquisition, Visualization, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Not applicable.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ctarc.2021.100409.

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA. Canc. J. Clin. 68 (2018) 394–424.

[2] C.W. Elston, I.O. Ellis, S.E. Pinder, Pathological prognostic factors in breast cancer, Crit. Rev. Oncol. Hematol. 31 (1999) 209–223.

[3] M.E.H. Hammond, D.F. Hayes, M. Dowsett, D.C. Allred, K.L. Hagerty, S. Badve, American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version), Arch. Pathol. Lab. Med. 134 (2010). https://pubmed.ncbi.nlm.nih.gov/20586616/. Accessed 10th May 2021.

[4] N. Bentzon, M. Düring, B.B. Rasmussen, H. Mouridsen, N. Kroman, Prognostic effect of estrogen receptor status across age in primary breast cancer, Int. J. Canc. 122 (2007) 1089–1094.

[5] Cooke T., Reeves J., Lanigan A., Stanton P. HER2 As a Prognostic and Predictive Marker For Breast Cancer. 2001.

[6] F. Pareja, J.S. Reis-Filho, Triple-negative breast cancers-a panoply of cancer types, Nat. Rev. Clin. Oncol. 15 (2018) 347–348.

[7] S. Sundar, P. Khetrapal-Singh, J. Frampton, E. Trimble, P. Rajaraman, R. Mehrotra, Harnessing genomics to improve outcomes for women with cancer in india: key priorities for research, Lancet. Oncol. 19 (2018) e102–e112.

[8] M.H. Forouzanfar, K.J. Foreman, A.M. Delossantos, R. Lozano, A.D. Lopez, C.J. L. Murray, Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis, Lancet. 378 (2011) 1461–1484.

[9] Dsouza N., Murthy N.S., Aras R.Y. Projection of cancer incident cases for india -till 2026. 2013;14:4379–86.

[10] J. Baptist, L. Busheri, L. Krishnan, R. Alhatz, Evaluating the performance of an advanced breast cancer diagnosis unit in india, Ind. J. Pub. Heat. Res. Dev. 8 (2017) 598–604.

[11] S. Nare, T. Patil, S. Dixit, P. Jere, B. Verghese, L. Krishnan, Establishment of a breast cancer biobank for translational research: a single institutional pilot study, Ind. J. Pub. Heal. Res. Dev. 9 (2018) 469–477.

[12] G. van Rossum, J. de Boer, Interactively testing remote servers using the python programming language, CWI. Q. 4 (1991) 283–303.

[13] G. Allen, M. Owens, The Definitive Guide to SQLite, Apress, 2010.

[14] R. Core Team. R: A Language and Environment for Statistical Computing. 2020. https://www.r-project.org/.

[15] G. Hortobagyi, J.L. Connolly, C.J. D'Orsi, S.B. Edge, E.A. Mittendorf, H.S. Rugo, Breast - AJCC Eight Edition, The American College of Surgeons (ACS), Chicago, Illinois, 2017.

[16] M. Schemper, T.L. Smith, A note on quantifying follow-up in studies of failure time, Contr. Clin. Trial. 17 (1996) 343–346.

[17] Therneau T. A Package for Survival Analysis in R R package Version 3.1-12. 2020. https://cran.r-project.org/package=survival.

[18] Kassambara A., Kosinski M., Przemyslaw B. survminer: Drawing Survival Curves using "ggplot2". R package version 0.4.7. 2020. https://cran.r-project.org/package=survminer.

[19] A. Kulkarni, D. Kelkar, N. Parikh, L.S. Shashidhara, C.B. Koppiker, M. Kulkarni, Meta-analysis of prevalence of triple negative breast cancer and its clinical features at incidence in indian breast cancer patients, JCO. Glob. Oncol. 6 (2020) 1052–1060. Jul.

[20] K.A. Hoadley, C. Yau, T. Hinoue, D.M. Wolf, A.J. Lazar, E. Drill, Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of, Canc. Cell. 173 (2018) 291–304, https://doi.org/10.1016/j.cell.2018.03.022, e6.

[21] D.C. Koboldt, R.S. Fulton, M.D. McLellan, H. Schmidt, J. Kalicki-Veizer, J. F. McMichael, Comprehensive molecular portraits of human breast tumours, Nat. 490 (2012) 61–70.

[22] J. Liu, T. Lichtenberg, K.A. Hoadley, L.M. Poisson, A.J. Lazar, A.D. Cherniack, An integrated tcga pan-cancer clinical data resource to drive high-quality survival outcome analytics, Cell 173 (2018) 400–416, e11.

[23] P.M. Vaid, A.K. Puntambekar, R.A. Banale, R.R. Reddy, R.R. Unde, N.P. Namewar, Stromal tumor infiltrating lymphocytes (stils) as a putative prognostic marker to identify a responsive subset of tnbc in an, Ind. Bre. Canc. Cohort. medRxiv. (2020), https://doi.org/10.1101/2020.08.19.20177865, 2020.08.19.20177865.

[24] C.B. Koppiker, A.U. Noor, S. Dixit, R. Mahajan, G. Sharan, U. Dhar, Implant-Based breast reconstruction with autologous lower dermal sling and radiation therapy outcomes, Ind. J. Surg. 81 (2019) 543–551.

[25] C.B. Koppiker, A.U. Noor, S. Dixit, L. Busheri, G. Sharan, U. Dhar, Extreme oncoplastic surgery for multifocal/multicentric and locally advanced breast cancer, Int. J. Bre. Canc. 2019 (2019) 1–8.