Landscape Analysis of Breast Cancer and Acute Myeloid Leukemia Trials Using the My Cancer Genome Clinical Trial Data Model

Neha M. Jain, PhD; Marilyn Holt, PhD; Christine Micheel, PhD; and Mia Levy, PhD

PURPOSE The field of oncology is expanding rapidly. New trials are opening as an increasing number of therapeutic agents are being investigated before they can become approved therapies. Aggregate views of these data, particularly data associated with diseases, biomarkers, and drugs, can be helpful in understanding the trends in current research as well as existing gaps in cancer care.

METHODS In this paper, we performed a landscape analysis for breast cancer and acute myeloid leukemia related trials with structured, curated data from clinical trials using the My Cancer Genome clinical trial knowledgebase.

RESULTS We have performed detailed analytics on breast cancer (N = 1,128) and acute myeloid leukemia trial sets (N = 483) to highlight the top biomarkers, drug classes, and drugs—thereby supporting a full view of biomarkers, biomarker groups, and drugs that are currently being explored in these respective diseases.

CONCLUSION Analysis and data visualization of the cancer clinical trial landscape can inform strategic planning for new trial designs and trial activation at a particular site.

INTRODUCTION The term cancer is used to describe many conditions, and cancer’s complexity and diversity is reflected in cancer treatments as well. It is not surprising, therefore, that a quarter of all drug approvals in the United States are for oncology. Before these drugs are approved, they go through a multistep process of approval—preclinical studies, clinical trials, regulatory review, etc—to generate data demonstrating safety and efficacy. The ability to view these data for all oncology trials in an aggregated form can be a powerful tool for providing unique insights into the clinical trial enterprise.

Currently, publicly available sources such as cancer.gov and ClinicalTrials.gov have information about clinical trials. Most of this information, excluding some clinical trial metadata, is unstructured and trapped in free-text documents. This makes it hard to computationally analyze these documents and extract trial elements for quick visualization. Developing informatics tools that support structuring these data elements and make these data machine-readable will open up avenues for downstream applications.

Several efforts have been made to develop tools to structure clinical trial data, including My Cancer Genome, Matchminer, IBM Watson, MolecularMatch, JAX-CKB, and TrialProspector, to name a few. Some of these models are fully commercial and therefore details about the data elements are not readily available. Additionally, the aggregated data are not made publicly available. Some groups have published aggregate data analyses, but the scope of curation is limited to individual genomic biomarkers and does not take into consideration cytogenetic and protein biomarkers, which are becoming increasingly therapeutically relevant. Certain biomarkers are typically reported together in clinical trial documents, or are related via a single pathway, or may be altered in relation to specific cancer diagnosis, such as human epidermal growth factor receptor 2 (HER2)-positive, estrogen receptor (ER)-positive, EGFR-activating mutations, and poor-risk mutations in acute myeloid leukemia (AML) as defined by National Comprehensive Cancer Network (NCCN) guidelines. Curating these together as an entity rather than individual alterations can add value to the curated trial content making it more relevant for patient populations. We have discussed this in detail elsewhere.

Most groups do not curate information about clinical trial interventional arms, the respective drugs in each of the arms related to the disease group and biomarker eligibility criteria, nor the treatment context for the trial (ie, neoadjuvant or adjuvant metastatic). This information when curated accurately can provide additional context to data and can provide meaningful insights.
In collaboration with Genomoncology LLC,12 we developed our own clinical trial data model that has been used to manually curate > 9,100 oncology trials. The individual and aggregate data from all these trials are publicly available on the My Cancer Genome website.5 Although these data from the My Cancer Genome website offer an insight into the design of individual trials and the landscape of the oncology trial spectrum, assessing therapy trends across all cancer trials and understanding biomarkers being studied for drug development in selected diseases of interest warrants a deeper analysis. We chose breast cancer and AML for our analysis because these diseases especially highlight the importance of curating protein and cytogenetic biomarkers, which is a unique strength of our curation model. The model has been described in detail elsewhere.11 Furthermore, breast cancer is now the most common cancer in the world13 and therefore is a disease of wide interest. AML is one of the most common hematologic malignancies in adults and also most often associated with clinical trials. Therefore, we performed a comprehensive analysis of all clinical trials related to breast cancer and AML to assess treatment trends and gather a snapshot view of the clinical trial spectrum across the globe.

**METHODS**

Using the clinical trial data model described earlier,11 we curated all therapeutic interventional breast cancer and AML trials regardless of the presence of a biomarker-driven eligibility criterion or trial recruiting status. For these trials, the clinical trial document was reviewed, and the information related to the diagnosis, biomarker eligibility criteria (wherever applicable), treatment context as well as drugs was captured in a structured format. All trials curated till December 3, 2019, were included in our analysis to allow a full view of trials for this disease, and not just the trials that were open and recruiting at the time of analysis.

All curated data for breast cancer and AML trials were accessed from the backend of the model and extracted out in a structured form as a CSV file (Data Supplement). As described in our previous publication,11 a trial arm is a combination of disease (with or without a biomarker) and a specific treatment. Therefore, a clinical trial with breast cancer exploring two different drug combinations would be organized into two trial arms. Drug classes were manually added and associated with all drugs associated with trial arms. Here, a trial arm refers to Drug class information was obtained from Drugbank,14 NCIt,15 and PubMed.16 Some previous studies have classified vaccine therapy, viral therapy, and cell therapy as immunotherapy,17 but for the purposes of this paper, we defined them as individual categories. Immunotherapy was used to define cytokotoxic T-cell lymphocyte-4 inhibitors, programmed cell death protein 1 (PD-1), or programmed death ligand-1 (PD-L1) inhibitors, whereas chimeric antigen receptor T-cells (CAR-T) cell therapies and other engineered cell therapies that directly or indirectly modulate the immune system were classified as cell-therapy based immunotherapies. Some drugs could not be clearly classified into one category, whereas others were relatively new and did not have much information available, while yet others were extremely specialized and could not be added in the previously defined categories. These drugs were grouped in the other category. A full list of all drugs and drug classes can be found in the Data Supplement. All data were sorted and visualized using Microsoft Excel.

For the triple-negative breast cancer (TNBC)-focused analysis, we extracted trial phase, drug names, and drug classes for all TNBC trials that were using pembrolizumab as one of the treatment drugs. These trials were manually reviewed and classified further by metastatic line of therapy into first line, second line or greater, first to third line, and any line. The data were visualized in multiple ways with sunburst plots, first oriented by line of metastatic therapy, and then by drug class. These analyses are not directly available from the website and need downloadable access to the full data set as well significant manual data sorting and enhancements. The aggregate data can be found in the Data Supplement.

**RESULTS**

**Landscape Analysis of Breast Cancer Trials**

Figure 1 shows the landscape of biomarkers investigated in breast cancer trials curated in the My Cancer Genome knowledgebase. One thousand eight hundred eighty-one clinical trial arms from 1,128 clinical trials related to breast...
cancer were reviewed for this analysis between June 2012 and December 2019. Unsurprisingly, the top biomarker alterations across breast trials (Fig 1A) are related to protein expression of the ER and progesterone receptor (PR), and amplification or protein expression of HER2. *BRCA1* and *BRCA2* mutations and *PIK3CA* mutations also had a significantly high number of associated trials owing to the multiple studies investigating poly (ADP-ribose) polymerase inhibitors and *PIK3CA* inhibitors, respectively. These trials ultimately led to the approval of several poly (ADP-ribose) polymerase inhibitors and *PIK3CA* inhibitors, respectively. These trials were used for this analysis.

Figure 1A shows the breakdown of breast cancer trials by the drug classes being investigated in the different cohorts. Cytotoxic therapy was most widely used in clinical trial arms (n = 1,023), closely followed by tyrosine kinase inhibitors (TKIs; n = 571) and hormonal therapy (n = 458). Figure 2B shows the number of unique drugs in each of the studied drug classes. TKIs with a total of 105 agents being investigated was the top category, followed by cytotoxic therapy (n = 82) and immunotherapy or cell therapy (n = 78). Cytotoxic agents comprised 32% of the drugs in breast cancer trials.

Since tumor ER, PR, and HER2 status is routinely assessed for all invasive breast cancer diagnoses, we wanted to classify all breast cancer trials according to this criterion to highlight the distribution of drug discovery across this trio of biomarkers. The highest number of trial arms were found associated with HER2-negative breast cancer (n = 638; 34%). TNBC was the second largest category with 553 trials (29%). Interestingly, there were
only six trials (0.3%) that were recruiting patients who had both a positive HER2 status and a negative hormone receptor status (Fig 3A). This highlights the importance of protein biomarkers in breast cancers and makes a strong case for developing models that support curation for protein biomarkers.

In an attempt to more deeply interrogate the TNBC trials, Figure 3B shows TNBC trials that were in the metastatic setting and used pembrolizumab as one of the treatment drugs. Most trials explored cytotoxic therapies in combination with pembrolizumab. Kinase inhibitors with pembrolizumab was the second most common combination. Most trials were in the phase II setting (n = 19), some in phase III (n = 4), and a few more in phase I (n = 7). Majority of the trials were in metastatic: any line, closely followed by metastatic first line. Only two trials were exploring single-agent pembrolizumab.

Landscape Analysis of AML Clinical Trials
The landscape of global AML trials has been shown in Figure 4. Figure 4A shows the top biomarker alterations featured across all AML-related clinical trial arms. Seven hundred seventy-nine clinical trial arms from 483 clinical trials related to AML were reviewed for this analysis between June 2012 and December 2019. KMT2A fusion, FLT3 mutation, and DEK-NUP214 fusion were the top biomarker alterations across these trials. Seventy-five percent of the biomarkers in Figure 4A were cytogenetic markers, highlighting the importance and prevalence of these markers in AML trials. Figure 4B displays the top biomarker alteration groups encountered while curating AML-related trial arms. NCCN® and European Leukemia Net (ELN®) guideline-driven alteration groups feature in this list along with other biomarker groups that were created by curators to facilitate efficient curation of biomarkers that are frequently featured.
of treatment (on line of treatment (summarized in Table 1. The AML trials were divided based immunotherapy, n = 29). cytotoxic therapy (n = 39) and engineered cell therapy (cell-agents being investigated was the top category, followed by each of the studied drug classes. TKIs with a total of 41 therapies. Figure 5B shows the number of unique drugs in breakdown of AML trials by the drug classes being investigated in the different cohorts. Cytotoxic therapy was together in multiple clinical trials. Figure 5A shows the breakdown of AML trials by the drug classes being investigated in the different cohorts. Cytotoxic therapy was still most widely used in clinical trial arms (n = 888), closely followed by transplant therapy (n = 160), TKIs (n = 138), and a variety of other targeted therapies and immunotherapies. Figure 5B shows the number of unique drugs in each of the studied drug classes. TKIs with a total of 41 agents being investigated was the top category, followed by cytotoxic therapy (n = 39) and engineered cell therapy (cell-based immunotherapy, n = 29).

The treatment context associated with AML trials has been summarized in Table 1. The AML trials were divided based on line of treatment (first-line v relapse or refractory), stage of treatment (induction, consolidation, or maintenance), as well as treatment modality (cellular or transplant therapies, noncellular therapy, or post-transplant therapy). Therefore, each trial could be a part of one or all three of these categories. The highest number of trials was found to be associated with relapse or refractory, followed closely by noncellular therapy.

DISCUSSION

In this manuscript, we describe a landscape analysis of breast cancer and AML clinical trials. This curation effort supports several unique analytics that required significant manual input and are currently not available on the My Cancer Genome website.5

The analysis presented here highlights the importance of biomarker and biomarker groups for curating oncology trials especially for breast cancer and AML. Breast cancer and AML were used for this analysis for their relative abundance in the population. Furthermore, these cancers also highlight the importance of model expressivity of the My Cancer Genome clinical trial curation and data model. It is not surprising to see the large number of trials associated with the ER, PR, and HER2-related biomarkers for breast cancer, or the numerous AML trials associated with the risk groups. But the unique strength in our model is the inclusion of protein biomarkers for breast cancer trials (which are usually measured through immunohistochemistry tests, not genomic tests) as well as the cytogenetic markers (often evaluated using fluorescent in situ hybridization tests or karyotyping) for AML trials. Other efforts at creating a landscape analysis have only focused on genomic biomarkers, which does not offer a complete view into the complexity of oncology biomarkers. The curation of biomarker groups and cytogenetic markers for AML and protein biomarkers for breast cancer trials helps appreciate the significance and prevalence of these biomarkers in the full view of other curated biomarkers.

Another unique contribution of this manuscript is the detailed analysis related to investigational drugs, drug classes, and treatment context. It is not surprising that cytotoxic chemotherapy regimens still remain the backbone of most cancer treatments and are seen in abundance in clinical trials—for breast cancer as well as AML. Cytotoxic therapy–related regimens are widely used as treatment controls and their abundance can therefore be accounted. Although transplant therapies have always been an important cornerstone of AML treatments, emergence of cell-based immunotherapies or engineered cell therapies has been an exciting development. There were a total of 242 drugs evaluated in AML trials, compared with the 3124 drugs approved for AML over the period explored in this analysis. Similarly, for the 479 agents being investigated in breast cancer, 6825 agents are US Food and Drug Administration–approved for treating breast cancer. This is not surprising, given the recent data that < 3.5% of oncology drugs that are investigated in clinical trials actually get approved.26

Despite the relative rarity of triple-negative breast cancer (TNBC),27 a significantly high number of trials were found to be investigating it. This is because of the lethality of the disease as well as the lack of therapies in this area. TNBC is an aggressive form of breast cancer, which currently does not have targeted therapy standard of care options. It is noteworthy that the frequent inclusion of patients with TNBC in multidisease trials and phase I trials also contributed to this high percentage.
This analysis also allowed us to highlight the complexity of treatment context related to AML (Table 1). These could be overlapping and are therefore more difficult to interpret. For example, a particular treatment could be classified as first-line therapy, induction therapy, and noncellular therapy. The information reported on the publicly available clinical trial document was relied upon to perform this classification. Since trial document can sometimes be hard to interpret, the classification was not always straightforward. Most trials could be distinguished based on line of treatment and treatment modality but stage of treatment was not always readily available in trial document. In future, simplifying or unifying treatment contexts for AML may help describe the treatment context and clinical appropriateness for these trials more efficiently.

One of the important observations was the enormous number of clinical trials that were open for both of the diseases studied. Low trial enrollment has been a stumbling block, preventing many trials from reaching completion.
In the light of this knowledge, it seems counterproductive to open multiple trials exploring either the same or similar drugs resulting in a division of efforts rather than a consolidated unified effort in one direction. Tang et al. found that only 5% of the 1,105 combination trials exploring PD-1 and PD-L1 agents are exploring new immunotherapy agents, whereas the remaining 95% are evaluating the previously approved five agents (avelumab, atezolizumab, durvalumab, nivolumab, and pembrolizumab) in combination with other immunotherapies, targeted therapies, or cytotoxic therapies.

In an attempt to investigate the existing overlap of trials in our data set, we performed a deeper analysis for a smaller subset: metastatic TNBC trials exploring pembrolizumab. At first instance, these trials look considerably different with minimal overlap in terms of the treatment setting and the actual drug being explored. But a closer look revealed overlap in terms of the drug classes being explored. For example, a clinician looking for a first-line metastatic trial will have to choose from nine different available trial options, four of which are exploring pembrolizumab in combination with cytotoxic therapy (Fig 3B). Such fragmentation of efforts can result in significant duplication of work, making the whole enrollment effort counterproductive. Investing in tools that support visualization of drug and drug

FIG 4. Landscape analysis of all AML trial arms. (A) Top biomarker alterations and (B) top biomarker groups. A total of 779 clinical trial arms from 483 clinical trials were used for this analysis. AML, acute myeloid leukemia; ELN, European Leukemia Net; FAB, French-American-British; NCCN, National Comprehensive Cancer Network.
FIG 5. Landscape of drugs and drug classes in AML trial arms. (A) The top drug classes being studied in AML trial arms and (B) the number of unique drugs in each of the drug classes being studied for AML trial arms. A total of 779 clinical trial arms from 483 clinical trials were used for this analysis. ADC, antibody-drug conjugate; AML, acute myeloid leukemia; MAb, monoclonal antibody; TKI, tyrosine kinase inhibitor.
In conclusion, in this paper, we display the landscape of AML and breast cancer trials. Our clinical trial data model allows us to record several data elements in a structured exportable format. In this paper, we presented aggregate data analysis for all breast cancer and AML trials in an effort to illuminate the oncology landscape for these two diseases. At an institutional level, such aggregated data can help visualize the trial portfolio of an institution, thereby identifying strengths and areas of gap in cancer care. Globally, this can help understand trends in oncology both in research as well as standard of care treatments. Such visualizations can help assess the landscape of oncology and guide the decision to open a new trial after careful assessment of existing trials, with the ultimate goal of preventing duplication of efforts.

**AFFILIATIONS**
1. Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN
2. Division of Hematology/Oncology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN
3. Division of Hematology/Oncology, Department of Internal Medicine, Rush University Medical Center, Chicago, IL
4. Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN

**CORRESPONDING AUTHOR**
Neha M. Jain, PhD, Vanderbilt University Medical Center, 2525 West End Ave, Suite 1582, Nashville, TN 37203; e-mail: neha.jain@vumc.org.

**SUPPORT**
Supported by the Susan G. Komen Foundation (SAC160070), the National Human Genome Research (U01 HG007253), and the GE Healthymagination award.

**AUTHOR CONTRIBUTIONS**
Conception and design: Neha M. Jain, Christine Micheel, Mia Levy
Financial support: Mia Levy
Administrative support: Neha M. Jain
Provision of study materials or patients: Neha M. Jain, Marilyn Holt
Collection and assembly of data: Neha M. Jain, Marilyn Holt, Mia Levy
Data analysis and interpretation: Neha M. Jain, Christine Micheel, Mia Levy
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

**REFERENCES**
1. Cancer Drugs Account for Over a Quarter of All New Drug Approvals in the US—The ASCO Post. https://www.ascopost.com/news/september-2019/cancer-drugs-account-for-over-a-quarter-of-all-new-drug_approvals-in-the-us/
2. Find NCI-Supported Clinical Trials—National Cancer Institute, 2016. https://www.cancer.gov/about-cancer/treatment/clinical-trials/search
3. Home—ClinicalTrials.gov. https://clinicaltrials.gov/
4. Jain NM, Culley A, Knoop T, et al: Conceptual framework to support clinical trial optimization and end-to-end enrollment workflow. JCO Clin Cancer Inform 3:1-10, 2019

**TABLE 1. Curated Treatment Contexts in Acute Myeloid Leukemia**

| Category               | No. of Trial Arms Associated With Each Category |
|------------------------|-----------------------------------------------|
| First-line therapy     | 300                                           |
| Relapse or refractory  | 337                                           |
| Induction therapy      | 77                                            |
| Consolidation therapy  | 63                                            |
| Maintenance therapy    | 39                                            |
| Cellular therapy       | 127                                           |
| Noncellular therapy    | 332                                           |
| Post-transplant therapy| 49                                            |

class data can elucidate trial data and potentially help conserve resources and prevent duplication of efforts in the future.
5. Home—My Cancer Genome. https://www.mycancergenome.org/
6. MatchMiner—An Open Source Computational Platform for Matching Patient-Specific Genomic Profiles to Precision Cancer Medicine Clinical Trials. https://matchminer.org/
7. IBM Watson Oncology Clinical Trial Matching-Overview-United States. https://www.ibm.com/us-en/marketplace/clinical-trial-matching-oncology
8. Clinical Intelligence I Precision Oncology I Clinical Decision Support. https://www.molecularmatch.com/
9. Patterson SE, Statz CM, Yin T, et al: Utility of the JAX Clinical Knowledgebase in capture and assessment of complex genomic cancer data. Npj Precis Oncol 3:1, 2019
10. Sahoo SS, Tao S, Parchman A, et al: Trial prospector: Matching patients with cancer research studies using an automated and scalable approach. Cancer Inform 13:157-166, 2014
11. Jain N, Mittendorf KF, Holt M, et al: The My Cancer Genome clinical trial data model and trial curation workflow. J Am Med Inform Assoc 27:1057-1066, 2020
12. Precision Oncology Solutions: GenomOncology. https://www.genomoncology.com
13. Breast Cancer Now Most Common Form of Cancer: WHO Taking Action. https://www.who.int/news/item/03-02-2021-breast-cancer-now-most-common-form-of-cancer-who-taking-action
14. Wishart DS, Feunang YD, Guo AC, et al: DrugBank 5.0: A major update to the DrugBank database for 2018. Nucleic Acids Res 46:D1074-D1082, 2018
15. NCI Thesaurus. https://ncit.nci.nih.gov/ncitbrowser/pages/home.jsf?version=20.03e
16. PubMed: PubMed. https://pubmed.ncbi.nlm.nih.gov/
17. Tang J, Shalabi A, Hubbard-Lucey VM: Comprehensive analysis of the clinical immuno-oncology landscape. Ann Oncol 29:84-91, 2018
18. Robson M, Im SA, Senkus E, et al: Olaparib for metastatic breast cancer in patients with a germline BRCA mutation. N Engl J Med 377:523-533, 2017
19. Litton JK, Rugo HS, Ettl J, et al: Talazoparib in patients with advanced breast cancer and a germline BRCA mutation. N Engl J Med 379:753-763, 2018
20. Novartis Pharmaceuticals: A Phase III Randomized Double-Blind Placebo Controlled Study of Alpelisib in Combination With Fulvestrant for Men and Postmenopausal Women With Hormone Receptor Positive, HER2-Negative Advanced Breast Cancer Which Progressed on or After Aromatase Inhibitor Treatment. Clinical trial registration NCT02437318, 2020. https://clinicaltrials.gov/ct2/show/NCT02437318
21. André F, Ciruelos E, Rubovszky G, et al: Alpelisib for PIK3CA-mutated, hormone receptor–positive advanced breast cancer. N Engl J Med 380:1929-1940, 2019
22. NCCN—Evidence-Based Cancer Guidelines, Oncology Drug Compendium, Oncology Continuing Medical Education. https://www.nccn.org/default.aspx
23. Drugs Approved for Leukemia—National Cancer Institute, 2011. https://www.cancer.gov/about-cancer/treatment/drugs/leukemia
24. Drugs Approved for Breast Cancer—National Cancer Institute, 2011. https://www.cancer.gov/about-cancer/treatment/drugs/breast
25. Wong CH, Siah KW, Lo AW: Estimation of clinical trial success rates and related parameters. Biostatistics 20:273-286, 2019
26. Scott LC, Mobley LR, Kuo T-M, et al: Update on triple-negative breast cancer disparities for the United States: A population-based study from the United States Cancer Statistics database, 2010 through 2014. Cancer 125:3412-3417, 2019