Automated eligibility screening and monitoring for genotype-driven precision oncology trials

RECEIVED 29 August 2015

REVISED 23 December 2015

ACCEPTED 31 January 2016

PUBLISHED ONLINE FIRST 25 March 2016

OXFORD
UNIVERSITY PRESS

Michael H Eubank,¹ David M Hyman,² Amritha D Kanakamedala,¹ Stuart M Gardos,¹ Jonathan M Wills,¹ Peter D Stetson^{2,3}

ABSTRACT

The Information Systems Department at Memorial Sloan Kettering Cancer Center developed the DARWIN Cohort Management System (DCMS). The DCMS identifies and tracks cohorts of patients based on genotypic and clinical data. It assists researchers and treating physicians in enrolling patients to genotype-matched IRB-approved clinical trials. The DCMS sends automated, actionable, and secure email notifications to users with information about eligible or enrolled patients before their upcoming appointments. The system also captures investigators' input via annotations on patient eligibility and preferences on future status updates. As of August 2015, the DCMS is tracking 159,893 patients on both clinical operations and research cohorts. 134 research cohorts have been established and track 64,473 patients. 51,192 of these have had one or more genomic tests including MSK-IMPACT, comprising the pool eligible for genotype-matched studies. This paper describes the design and evolution of this Informatics solution.

Keywords: clinical research informatics, genomics, precision oncology, individualized medicine, informatics

BACKGROUND AND SIGNIFICANCE

Clinical trials are critical to advancing oncology care and their success is dependent on enrollment. In the United States, only 2–4% of newly diagnosed adult cancer patients participate in clinical trials.¹ Manual screening for patient eligibility is expensive and inefficient.² Automated systems to increase accrual to clinical trials are needed.

Clinical research informatics tools may improve the efficiency of identifying patients for clinical trials.^{3–5} However, few studies have reported on tools that enable genotype-driven enrollment. The Clinical and Translational Science Award's (CTSA) Integrated Data Repository (IDR) Group evaluated IDR implementation across 46 national CTSA centers and found only 20% of IDRs included genomic and proteomic data.⁶ Multiplex mutational profiling has been shown to have the potential to facilitate genotype-driven trial enrollment, but only in small pilot studies typically evaluating a single disease.⁷ A recent evaluation of the utility of tumor genomic screening at a major cancer center found that only 4% of patients subsequently enrolled in genotype matched studies and less than 50% of patients with actionable alterations were notified of available ongoing studies.⁸

These considerations establish the need for informatics solutions to improve genomics-driven clinical trial enrollment. Challenges for precision oncology research^{9,10} include linking clinical and genotypic data, filtering trial inclusion/exclusion criteria, ensuring optimal detection sensitivity and specificity, and notifying investigators with actionable information in a timely manner. We believe real-time tools to notify physicians of potential genotype-matched trials are needed to increase protocol enrollment.¹¹ We report here on our informatics solutions for genotype-driven enrollment in precision oncology trials.

INTRODUCTION

Description of Memorial Sloan Kettering Cancer Center and Initiatives in Precision Oncology

Memorial Sloan Kettering Cancer Center (MSKCC), a National Cancer Institute–designated Comprehensive Cancer Center, has approximately

840 active therapeutic clinical trials. The center has reported previously on the development of the Memorial Sloan Kettering–Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT),¹² a hybridization capture-based next-generation sequencing assay for targeted deep sequencing of all exons and select introns of cancer genes in formalin-fixed, paraffin-embedded tumors. Since the study's publication, the number of cancer genes MSK-IMPACT targets has increased from 341 to 410. Calendar year 2014 produced 2461 somatic MSK-IMPACT studies, and 2015 is on track for over 6000. The majority of this testing is performed under the auspices of a center-wide genomic testing protocol (ClinicalTrials.gov number NCT01775072).

Prior Approaches to Patient Accrual at MSKCC and the Introduction of DARWIN

Historically, patients were manually screened and accrued to clinical trials based on reports from MSKCC's enterprise data warehouse, the Institutional Database (IDB). In 2006 a Web-based self-service analytics platform, DARWIN, was introduced for searching the IDB. DARWIN includes demographics, diagnoses, procedures, medications, radiology, radiation therapy, labs, admissions, appointments, protocol accruals, tissue bank, and Diagnostic Molecular Pathology (DMP) results. DARWIN can present de-identified data stripped of Protected Health Information, including unstructured notes. Fully identified data is available to authorized users. As Figure 1 depicts, DARWIN has evolved significantly to support the search for eligible patients based on combined genotypic/phenotypic criteria. This case report focuses on the addition of discrete genomic aberration data and the DARWIN Cohort Management System in support of precision oncology trials.

MATERIALS AND METHODS

We use the following definitions in this report:

- Cohort: a population of patients identified in our clinical data warehouse using automated inclusion/exclusion criteria defined

Correspondence to Michael H Eubank, 633 3rd Ave, 2nd Floor, New York, NY 10016, USA; eubankm@mskcc.org For numbered affiliations see end of article.

© The Author 2016. Published by Oxford University Press on behalf of the American Medical Informatics Association. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Figure 1: Timeline of the Darwin Cohort Management System Development

by users (not all research protocols at MSK have been modeled with a cohort definition, so not all research patients are in cohorts).

- Eligible: patients in cohorts meeting criteria for inclusion in a study, but not yet enrolled.
- Enrolled: patients in cohorts who have consented and are part of a clinical trial.

Design Goals and Evolution of the DARWIN Cohort Management System (DCMS)

Prior to 2012, DMP results were available exclusively as unstructured text at MSKCC. Researchers used free-text searches and chart review to find potentially eligible patients. To provide a more efficient and accurate method of identifying patients, the Information Systems department developed a daily Structured Query Language (SQL)-driven regular expression (RegEx) process to parse unstructured genomic aberration data (including gene, locus, amino acid substitutions, and cDNA changes) from DMP reports into results warehoused like any other structured data.

By 2012, however, analysis of researcher queries identified 4 opportunities for enhancements: (1) structured genomic aberration data (without the need for RegEx parsing), (2) automated identification of potentially eligible patients, (3) automated data abstraction for eligibility confirmation, and (4) incorporation of structured investigator feedback into the system with eligibility confirmations/rejections. In response, a user-centered design approach was employed to develop these enhancements. This new tool set became known as the DARWIN Cohort Management System (DCMS).

The DCMS was developed to identify, track, and manage cohorts. By early 2013, cohorts based on complex genetic and clinical criteria

were being created. In 2014, it was enhanced to push out secure email notifications of newly identified patients and patient activity. Researchers could customize notifications by specifying aberrations and events of interest. Notifications were actionable and automated, as described below in the Workflow section. The system also pushes data to other portals, such as the MSKCC cBioPortal, a tool designed by MSKCC to simplify access to complex genomic data for cancer researchers so they can analyze and visualize de-identified genetic information.¹³ Present-day DCMS provides additional metadata, greater detail for targeted patient identification, and latest-generation genetic testing data directly from the Pathology department. A timeline of DCMS development is displayed in Figure 1. Figure 2 illustrates application components and data flows.

Workflow Support by DCMS and Notification System

DCMS workflow follows these steps:

- Step 1. Following an IRB-approved expedited protocol and Privacy Office approval, logic-defining study inclusion criteria are developed, typically SQL against the data warehouse. This logic is captured as an attribute of the new cohort in the DCMS Entity Attribute Value architecture. Researchers dictate the cohort refresh schedule and whether patients can be automatically removed.
- Step 2. A onetime historical notification is sent to the Principal Investigator (PI) and team, containing identified patients and upcoming appointments and scans. Investigators can set a patient's status for a cohort via a provided link to DARWIN. A screenshot of email notification and status update screens is found in Figure 3.
- Step 3. Users select the frequency and scope of subsequent notifications containing newly identified patients and upcoming activity for the entire cohort. Notifications provide a link to specific tumor

Figure 2: Application Components and Data Flows

samples in the cBioPortal when applicable. Notification emails can assist with generating a patient-specific email, notifying the treating physician that a patient may be eligible for a study. The content is customized to the patient, tumor aberration(s) of interest, and appointment schedule. All wording in these emails is customized by the researcher.

Technical Specifications of DCMS

The DCMS leverages an existing IBM DB2 data warehouse and was designed using open-source software where possible. The rules engine stores and executes logic from database tables and stored procedures. Notifications are generated from scheduled queries using an open-source Java email package implemented as a database user-defined function to provide messaging from the SQL statement. A Web application using jQuery open-source JavaScript libraries allows system administrators to capture cohort logic and settings. The same user interface provides end users the ability to annotate and set patient status as it relates to their study. Notification queries produce XML transformed into HTML for delivery as content-rich emails.

Logic for cohorts can be flexible and updated as requested by the PI. Numerous cohorts have proven either too restrictive or too permissive and have had their inclusion criteria refined. The system then refreshes the cohort and will add or drop patients based on the evolving criteria.

RESULTS

Out of 1 459 348 unique patients in the data warehouse, 159 893 (11%) are associated with at least 1 DCMS research or clinical operations cohort. As of August 2015, 134 research cohorts have been established and track 64 473 patients. Of these, 51 192 have had 1 or more molecular diagnostic tests, including MSK-IMPACT, and represent the pool of potential patients for genotype-matched studies. Table 1 contains additional details regarding cohort types and the associated patient counts.

Since the last major update of DCMS at the end of 2014, it meets all 4 of the specified design goals. The system has pushed out over 1000 email notifications and is designed to generate at most 1 notification per study per day, including any new patients and upcoming clinical events; see Figure 3 for sample email. The latest version also captures end-user feedback on patient eligibility to influence future notification and annotate cohorts at the patient level. Finally, the

system tracks key scheduled clinical events pertinent to the study protocol, such as an upcoming radiology study. Currently there are 33 users registered across 20 cohorts for these advanced management features. See Figure 3 for status and annotation interfaces.

DISCUSSION

The DCMS is an informatics platform to screen for and manage large cohorts of patients on genotype-matched clinical trials. The evolution of IDB, DARWIN, and MSK-IMPACT supports patient accrual to genotype-matched trials. The DCMS has enabled results to be shared earlier and in a targeted manner via notifications with actionable information.

It should be noted that MyCancerGenome.org, Personalized CancerTherapy.org, and similar references, which are curated clinical annotation repositories of genomic variants, provide guidance to treating physicians about the clinical significance of specific genomic variants. The DCMS provides the ability to identify patients with specific genomic alterations in order to approach them regarding trial enrollment. In this way, variant curation databases and our DCMS are complementary and non-overlapping resources.

Studying tumor molecular profiles to guide treatment decisions is increasingly common.¹⁴ Specifically, there is growing interest in using genomic data to enroll patients in genotype-matched trials.^{15,16} To increase patient accrual rates, not only is it important to screen patients broadly and maintain a large number of clinical trial options, it is also important for study PIs to have efficient tools to manage enrolled patients and to notify treating physicians in a timely manner about potential genotype-matched studies for their patients.¹¹

The DCMS extends prior work on oncology clinical trial alert systems in several ways. Our system tracks genotype-specific cohorts on a larger scale than did prior studies.^{7,8} It provides workflow support by integrating data on upcoming scheduled patient events with the clinical and genomic information for targeted push notifications to investigators. Notifications delivered to the PI can generate subsequent secure email messages to the treating physician with context and tailored content, thus making the notifications actionable. DCMS is set up to capture investigators' input via annotations on patient eligibility and preferences on future status updates. Finally, the notification email includes a link to the cBioPortal, allowing for a de-identified visual analysis of a patient's genomic information in context with other patients with similar variations, for discovery of new associations.^{13,17}

Figure 3: Email Notification and Status Update

Sample Email Notification

darwin cohort notification for:
CD274 9p24 amplification patients
27 active patients

2 Newly identified patients

Days on Cohort	Patient	DMP Results of interest	Next Scan	Next Dr Visit
0	[MRN] [Last Name] [First Name]	[Path Report ID] 15 aberrations DMP Provided Dx: cBio site: Brain tumor: Glioblastoma Multiforme mets: Not Applicable CD274 (NM_014143 - 9p24.1) Amplification (Fold Change: 2.2) JAK2 (NM_004972 - 9p24.1) Amplification (Fold Change: 4.6) PTPRD (NM_002839 - 9p24.1) Amplification (Fold Change: 3.1)		
0	[MRN] [Last Name] [First Name]	[Path Report ID] 11 aberrations DMP Provided Dx: cBio site: Urethra tumor: Mucosal Melanoma mets: Not Applicable CD274 (NM_014143 - 9p24.1) Amplification (Fold Change: 3.6) JAK2 (NM_004972 - 9p24.1) Amplification (Fold Change: 3.6)	Jan 12 - 12:00 CT CHEST	Jan 12 - 14:45 [Treating Physician]

1 upcoming scan

Days on Cohort	Patient	DMP Results of interest	Next Scan	Next Dr Visit
68	[MRN] [Last Name] [First Name]	[Path Report ID] 6 aberrations DMP Provided Dx: cBio site: Soft Tissue tumor: Myxofibrosarcoma mets: Not Applicable CD274 (NM_014143 - 9p24.1) Amplification (Fold Change: 2.6) JAK2 (NM_004972 - 9p24.1) Amplification (Fold Change: 3.6)	Aug 30 - 09:40 CT CHEST	Jul 27 - 15:55 [Treating Physician]

Link to DCMS to manage patient status and chart review

Link to specimen in cBioPortal.mskcc.org

Email to Treating Physician

To: [Treating Physician Email Address]
Cc: [PI and team email addresses]
Bcc: [PI and team email addresses]
Subject: Regarding patient [Patient Name] - [MRN]

[Treating Physician],

A patient of yours [LAST NAME], [FIRST NAME] – [MRN] with an [PI dictated mutation of interest] mutation potentially qualifies them for our study.

[Additional text as dictated by Principal Investigator]

Cohort Status Update

darwin Patient Viewer
data through Aug 26

MRN [Medical Record Number] Age/Gender 62 year old Female
Name [Last Name], [First Name] Deceased Date

Logoff

PT Dx Path Rad Labs Rx Surg IR RT Protocol EMR Docs Cohorts

Status Title Status Details

● Active CD274 9p24 amplification patients [insert comment regarding patient status here]
○ Inactive

Limitations of this approach include dependency on the high-throughput molecular profiling pipeline in place with MSK-IMPACT and incomplete coverage to date of all research patients in DCMS cohorts because of the work involved in developing the inclusion criteria and implementing them as structured rules in the DCMS.

Future steps include exploring point-of-care notifications to the treating physician and further integration of data visualization and summaries. Although this paper describes use of the DCMS for clinical research, MSKCC also utilizes the system for population health and operational purposes. Examples include cohorts based on scheduled

appointments/scans, pediatric patients, and patients with surgical procedures in combination with radiology reports. These allow for operational users to easily interact with dynamic groups of patients that relate to their work.

CONCLUSION

The DARWIN Cohort Management System facilitates the integration of genotypic and clinical data to assist with accruing large cohorts of patients to genotype-matched precision oncology trials. The DCMS provides workflow support in the form of timely, actionable information for

Table 1: Number of cohorts with unique patient by earliest ID date and cohort type

First Cohort ID Date	Total Count of Cohorts	Unique Patient Count by Cohort Type		
		Research	Clinical Operations	Overall
2013-Q1	2	78		78
2013-Q2	3	103		103
2013-Q3	5	125		125
2013-Q4	5	153		153
2014-Q1	5	715		715
2014-Q2	22	2339	375	2708
2014-Q3	105	53 253	13 009	53 652
2014-Q4	108	55 565	19 070	56 035
2015-Q1	118	59 241	121 875	149 645
2015-Q2	136	62 929	125 287	153 774
2015-08-21	147	64 473	131 701	159 943

researchers looking to accrue patients to targeted studies. It also integrates key event data to support prospective management of enrolled patients. Finally, it provides a means for researchers to annotate and set patient statuses, thereby allowing them to manage future notifications.

CONTRIBUTORS

M.E.: Architect and developer of the Darwin Cohort Management System. Lead author who oversaw all aspects of drafting, revision, and final approval of the manuscript. Agrees to be accountable for all aspects of the work. D.M.H.: Principal champion and user of the Darwin Cohort Management System, responsible for providing overall project direction and conceptual requirements. Participated in the drafting, revision, and final approval of the manuscript. Agrees to be accountable for all aspects of the work. A.D.K.: Provided significant contributions toward the acquisition, analysis, and interpretation of the data for this manuscript. Participated in the drafting, revision, and final approval of the manuscript. Agrees to be accountable for all aspects of the work. S.G.: Provided significant contributions toward the acquisition, analysis, and interpretation of the data for this manuscript. Participated in the drafting, revision, and final approval of the manuscript. Agrees to be accountable for all aspects of the work. J.W.: Provided significant contributions toward the acquisition, analysis, and interpretation of the data for this manuscript. Participated in the drafting, revision, and final approval of the manuscript. Agrees to be accountable for all aspects of the work. P.D.S.: Provided significant contributions to formulation of the manuscript's core content, theme, and direction and toward the acquisition, analysis, and interpretation of the data for this manuscript. Senior author, and participated in all aspects of drafting, revision, and final approval of the manuscript. Agrees to be accountable for all aspects of the work.

AUTHOR AFFILIATIONS

¹Information Systems, Memorial Sloan Kettering Cancer Center, New York, NY

²Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sector.

COMPETING INTERESTS

The authors have no competing interests to declare.

REFERENCES

- Lara PN Jr, Higdon R, Lim N, et al. Prospective evaluation of cancer clinical trial accrual patterns: identifying potential barriers to enrollment. *J Clin Oncol*. 2001;19(6):1728–1733.
- Penberthy LT, Dahman BA, Petkov VI, DeShazo JP. Effort required in eligibility screening for clinical trials. *J Oncol Practice*. 2012;8(6):365–370.
- Miotto R, Weng C. Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials. *J Am Med Inform Assoc*. 2015;22(e1):e141–e150.
- Ni Y, Wright J, Perentesis J, et al. Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients. *BMC Med Inform Dec Mak*. 2015;15:28.
- Seroussi B, Bouaud J. Using OncoDoc as a computer-based eligibility screening system to improve accrual onto breast cancer clinical trials. *Artificial Intel Med*. 2003;29(1-2):153–167.
- MacKenzie SL, Wyatt MC, Schuff R, Tenenbaum JD, Anderson N. Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey. *J Am Med Inform Assoc*. 2012;19(e1):e119–e124.
- Lovly CM, Dahlman KB, Fohn LE, et al. Routine multiplex mutational profiling of melanomas enables enrollment in genotype-driven therapeutic trials. *PLoS One*. 2012;7(4):e35309.
- Meric-Bernstam F, Brusco L, Shaw K, et al. Feasibility of large-scale genomic testing to facilitate enrollment onto genomically matched clinical trials. *J Clin Oncol*. 2015;33(25):2753–2762.
- Committee on a Framework for Development a New Taxonomy of Disease (National Research Council). Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington, DC, National Academy Press; 2011.
- Garraway LA, Verweij J, Ballman KV. Precision oncology: an overview. *J Clin Oncol*. 2013;31(15):1803–1805.
- Hyman DM, Solit DB. Tumor genetic screening programs: a call to action. *J Clin Oncol*. 2015;33(25):2725–2726.
- Cheng DT, Mitchell TN, Zehir A, et al. Memorial Sloan Kettering–Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J Mol Diagnostics*. 2015;17(3):251–264.
- Sirintrapun SJ, Zehir A, Syed A, Gao J, Schultz N, Cheng DT. Translational Bioinformatics and Clinical Research (Biomedical) Informatics. *Surg Pathol Clinics*. 2015;8(2):269–288.
- Kummar S, Williams PM, Lih CJ, et al. Application of molecular profiling in clinical trials for advanced metastatic cancers. *J Natl Cancer Inst*. 2015;107(4). pii: djv003. doi: 10.1093/jnci/djv003.
- Hyman DM, Puzanov I, Subbiah V, et al. Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 mutations. *New Engl J Med*. 2015;373(8):726–736.
- Hunter DJ, D'Agostino RB, Sr. Let's Not Put All Our Eggs in One Basket. *New Engl J Med*. 2015;373(8):691–693.
- Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.