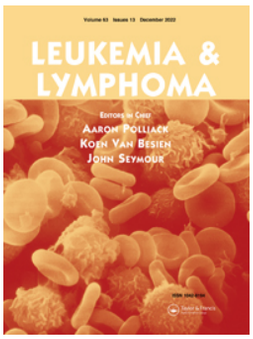


Leukemia & Lymphoma

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ilal20

Development of a distributed international patient data registry for hairy cell leukemia

Leslie A. Andritsos, Mirela Anghelina, Jasmine Neal, James S. Blachly, Puneet Mathur, Omkar Lele, Claire Dearden, Sunil Iyengar, Matthew Cross, Clive S. Zent, Kerry A. Rogers, Narendranath Epperla, Gerard Lozanski, Christopher C. Oakes, Eric Kraut, Amy S. Ruppert, Qiuhong Zhao, Seema A. Bhat, Francesco Forconi, Versha Banerji, Sasanka Handunnetti, Constantine S. Tam, John F. Seymour, Monica Else, Robert J. Kreitman, Alan Saven, Timothy Call, Sameer A. Parikh, Farhad Ravandi, James B. Johnston, Enrico Tiacci, Xavier Troussard, Martin S. Tallman, Sascha Dietrich, Tamar Tadmor, Alessandro Gozzetti, Pier Luigi Zinzani, Tadeusz Robak, Graeme Quest, Judit Demeter, Kanti Rai, Soledad A. Fernandez & Michael Grever

To cite this article: Leslie A. Andritsos, Mirela Anghelina, Jasmine Neal, James S. Blachly, Puneet Mathur, Omkar Lele, Claire Dearden, Sunil Iyengar, Matthew Cross, Clive S. Zent, Kerry A. Rogers, Narendranath Epperla, Gerard Lozanski, Christopher C. Oakes, Eric Kraut, Amy S. Ruppert, Qiuhong Zhao, Seema A. Bhat, Francesco Forconi, Versha Banerji, Sasanka Handunnetti, Constantine S. Tam, John F. Seymour, Monica Else, Robert J. Kreitman, Alan Saven, Timothy Call, Sameer A. Parikh, Farhad Ravandi, James B. Johnston, Enrico Tiacci, Xavier Troussard, Martin S. Tallman, Sascha Dietrich, Tamar Tadmor, Alessandro Gozzetti, Pier Luigi Zinzani, Tadeusz Robak, Graeme Quest, Judit Demeter, Kanti Rai, Soledad A. Fernandez & Michael Grever (2022) Development of a distributed international patient data registry for hairy cell leukemia, Leukemia & Lymphoma, 63:13, 3021-3031, DOI: 10.1080/10428194.2022.2109157

To link to this article: https://doi.org/10.1080/10428194.2022.2109157

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 07 Sep 2022.

Submit your article to this journal

Article views: 2367

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ilal20

LEUKEMIA & LYMPHOMA
2022, VOL. 63, NO. 13, 3021–3031
https://doi.org/10.1080/10428194.2022.2109157

Taylor & Francis
Taylor & Francis Group

REVIEW

OPEN ACCESS Check for updates

Development of a distributed international patient data registry for hairy cell leukemia

Leslie A. Andritsos[a], Mirela Anghelina[b], Jasmine Neal[c], James S. Blachly[d], Puneet Mathur[e], Omkar Lele[b], Claire Dearden[f], Sunil Iyengar[f], Matthew Cross[f], Clive S. Zent[g], Kerry A. Rogers[d], Narendranath Epperla[d], Gerard Lozanski[h], Christopher C. Oakes[d], Eric Kraut[d], Amy S. Ruppert[d], Qiuhong Zhao[d], Seema A. Bhat[d], Francesco Forconi[i], Versha Banerji[j,k], Sasanka Handunnetti[l], Constantine S. Tam[l], John F. Seymour[l], Monica Else[m], Robert J. Kreitman[n], Alan Saven[o], Timothy Call[p], Sameer A. Parikh[p], Farhad Ravandi[q], James B. Johnston[j,k], Enrico Tiacci[r], Xavier Troussard[s], Martin S. Tallman[t], Sascha Dietrich[u], Tamar Tadmor[v], Alessandro Gozzetti[w], Pier Luigi Zinzani[x], Tadeusz Robak[y], Graeme Quest[z], Judit Demeter[aa], Kanti Rai[ab], Soledad A. Fernandez[b] and Michael Grever[d]

[a]Division of Hematology Oncology, University of New Mexico Comprehensive Cancer Center, Albuquerque, NM, USA; [b]Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA; [c]Department of Internal Medicine, Division of Hematology, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA; [d]Department of Internal Medicine, Division of Hematology, The Ohio State University, Columbus, OH, USA; [e]Department of Research Information Technology, College of Medicine, The Ohio State University, Columbus, OH, USA; [f]The Royal Marsden NHS Foundation Trust, London, UK; [g]James P. Wilmot Cancer Institute, University of Rochester Medical Center, Rochester, NY, USA; [h]Department of Pathology, The Ohio State University Medical Center, Columbus, OH, USA; [i]Cancer Sciences and Haematology Department, University of Southampton Hospital Trust, Southampton, UK; [j]Department of Internal Medicine & Biochemistry and Medical Genetics, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada; [k]Research Institute in Oncology and Hematology, Winnipeg, Canada; [l]Haematology Department, Peter MacCallum Cancer Centre & Royal Melbourne Hospital, University of Melbourne, Melbourne, Australia; [m]The Institute of Cancer Research, London, UK; [n]Laboratory of Molecular Biology, National Cancer Institute, NIH, Bethesda, MD, USA; [o]Division of Hematology and Oncology, Scripps Clinic, La Jolla, CA, USA; [p]Division of Hematology, Mayo Clinic, Rochester, MN, USA; [q]Department of Leukemia, University of Texas MD Anderson Cancer Center, Houston, TX, USA; [r]Department of Medicine and Surgery, Institute of Hematology, University and Hospital of Perugia, Perugia, Italy; [s]Department of Hematology, Centre Hospitalier Universitaire Cote de Nacre, Caen, France; [t]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA; [u]Department of Hematology, University Hospital of Heidelberg, Heidelberg, Germany; [v]Hematology Division, Bnai Zion Medical Center and The Ruth and Bruce Rappaport Faculty of Medicine, Technion, Haifa, Israel; [w]Department of Medicine, University of Siena Policlinico S Maria alle Scotte, Siena, Italy; [x]Dipartimento di Medicina Specialistica, Diagnostica e Sperimentale Università degli Studi, IRCCS Azienda Ospedaliero-Universitaria di Bologna and Istituto di Ematologia "Seràgnoli", Bologna, Italy; [y]Department of Hematology, Medical University of Lodz, Lodz, Poland; [z]Pathology and Molecular Medicine, Queen's University and Kingston Health Sciences Centre, Kingston, Canada; [aa]Department of Internal Medicine and Oncology, Semmelweis University, Budapest, Hungary; [bb]Zucker School of Medicine at Hofstra/Northwell, New Hyde Park, NY, USA

ABSTRACT

Hairy cell leukemia (HCL) is a rare lymphoproliferative disorder, comprising only 2% of all leukemias. The Hairy Cell Leukemia Foundation (HCLF) has developed a patient data registry to enable investigators to better study the clinical features, treatment outcomes, and complications of patients with HCL. This system utilizes a centralized registry architecture. Patients are enrolled at HCL Centers of Excellence (COE) or via a web-based portal. All data are de-identified, which reduces regulatory burden and increases opportunities for data access and re-use. To date, 579 patients have been enrolled in the registry. Efforts are underway to engage additional COE's to expand access to patients across the globe. This international PDR will enable researchers to study outcomes in HCL in ways not previously possible due to the rarity of the disease and will serve as a platform for future prospective research.

ARTICLE HISTORY
Received 13 June 2022
Revised 25 July 2022
Accepted 26 July 2022

KEYWORDS
HCL; registry; outcomes

CONTACT Leslie Andritsos LAndritsos@salud.unm.edu University of New Mexico Comprehensive Cancer Center, 1201 Camino de Salud NE, Albuquerque, NM 87131, USA

Supplemental data for this article is available online at https://doi.org/10.1080/10428194.2022.2109157

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Hairy cell leukemia

Hairy cell leukemia (HCL) is a rare, chronic B-cell leukemia that was initially described as a unique clinical entity by Dr. Bertha Bouroncle et al. [1]. Originally called leukemic reticuloendotheliosis, HCL was associated with short overall survival (OS) and had inadequate therapeutic options [2]. Initially, HCL was thought to be a single homogenous disease. However, in 2016, the World Health Organization differentiated two distinct entities. Classic HCL (cHCL) demonstrates a unique immunophenotype and almost always carries the BRAFV600E mutation [3], whereas variant HCL (vHCL) demonstrates several distinctive morphologic and immunophenotypic features, and the BRAFV600E mutation is absent though activating mutations in the downstream kinase MAP2K1 may be seen [4]. Patients with cHCL typically present with fatigue, pancytopenia, splenomegaly, and a greatly increased risk for infection, which is the leading cause of serious morbidity [5]. Patients with vHCL frequently present with a malignant lymphocytosis in addition to the pancytopenia, and exhibit a more aggressive clinical course with inferior responses to chemotherapy. Since their initial discovery, enormous progress has been made in the understanding of the treatment and biology of these diseases [6]. Prior to the 1980s, treatment options were limited primarily to splenectomy and the projected median OS was only approximately 5 years. With the discovery of purine nucleoside analogs, the overall response rates (ORRs) and complete response (CR) rates in cHCL are now approximately 100% and 80–90%, respectively, and most patients can achieve durable remissions. In fact, cHCL patients may now experience a life expectancy approaching that of unaffected age matched controls [6]. vHCL remains difficult to treat, with shorter progression-free and OS rates in comparison to cHCL [7].

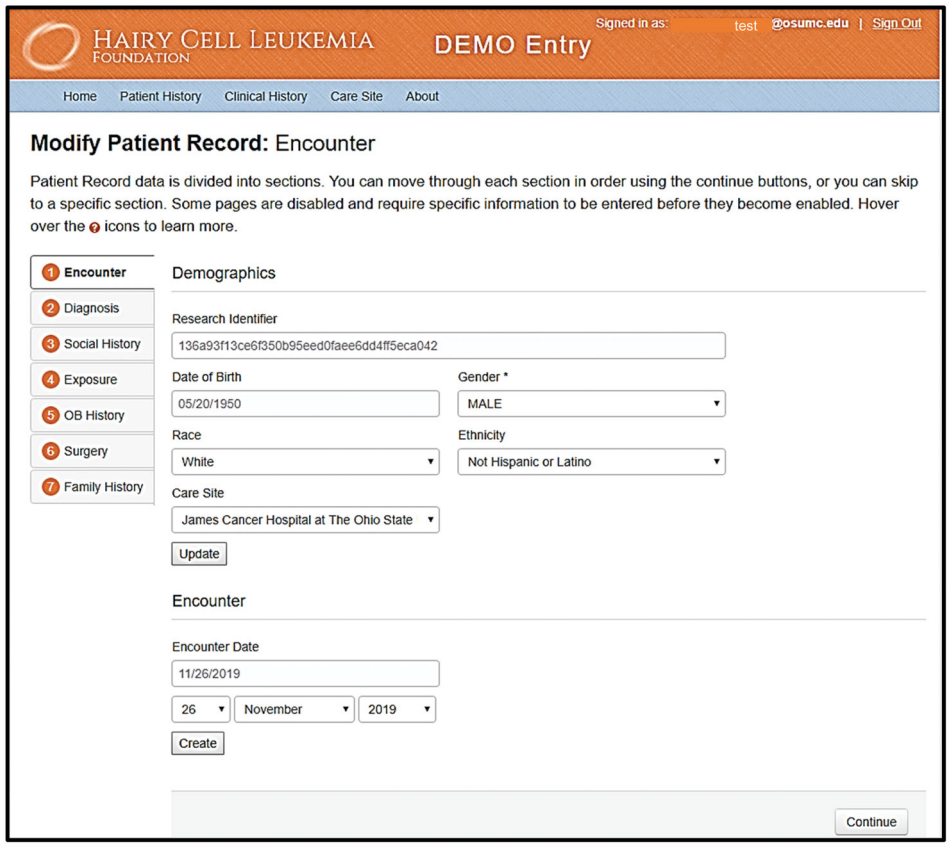
Many important questions remain unanswered in HCL. As patients are now living longer due to improved therapeutic options, there is evidence that at least 40% of patients with cHCL will eventually suffer disease relapse and require additional treatment. Patients with vHCL have a higher relapse rate, with shorter duration of remissions. More research is needed to elucidate outcomes in specific disease subtypes, complications of the diseases and their therapies, and other issues pertinent to patients such as quality of life. Furthermore, risk stratification and disease staging have not been updated since 1982 [8], prior to the introduction of highly effective therapies, leaving patients and clinicians without meaningful prognostic information. In order to derive meaningful clinical conclusions regarding integral aspects of disease staging and management of these rare adult lymphoproliferative disorders, cooperation, and collaboration are essential in defining important questions and collecting sufficient information in a large cohort of patients.

The conduct of clinical and translational studies requires access to high quality longitudinal data, often derived from a significant number of representative patients. In most instances, such data and resources are collected, formalized, stored, and retrieved using study-specific and localized data repositories. Unfortunately, in rare diseases such as HCL, the availability of these types of resources is greatly limited due to the low numbers of patients seen at any single institution [9]. It is difficult to conduct impactful clinical research for a rare disease at a single center. Instead, it is essential to cooperate and collaborate across multiple institutions to develop and execute successful clinical studies. In 2008, following a meeting to discuss potential advances in the management of HCL, an international consortium of recognized experts was formed to drive advances in the field. The collective effort of these experts ultimately merged into an existing patient-based organization, the Hairy Cell Leukemia Foundation (HCLF) [10]. As a result of inquiries submitted via the Foundation's web portal regarding requests for information related to highly relevant yet unanswered questions, the Foundation developed a patient data registry that could provide access to high quality, disease-specific data to advance clinical and translational research in HCL. The database was carefully constructed with input from experts in the field to ensure inclusion of all relevant disease-specific elements. The final database captures patient demographics, including family history, chemical and radiation exposure, clinical data such as comorbidities, pathology reports, immunophenotyping, genetic testing and hematology, as well as information on disease treatment and outcome, including complications (Figure 1).

Constructing the international HCL patient data registry

Rare disease research requires the collaboration of multiple organizations which combine their research grade data. However, such data sharing presents numerous regulatory, technical, operational, and legal obstacles. These obstacles are particularly challenging when collaborators may not have control over data access and use. Federated data registries, in which a

A



B

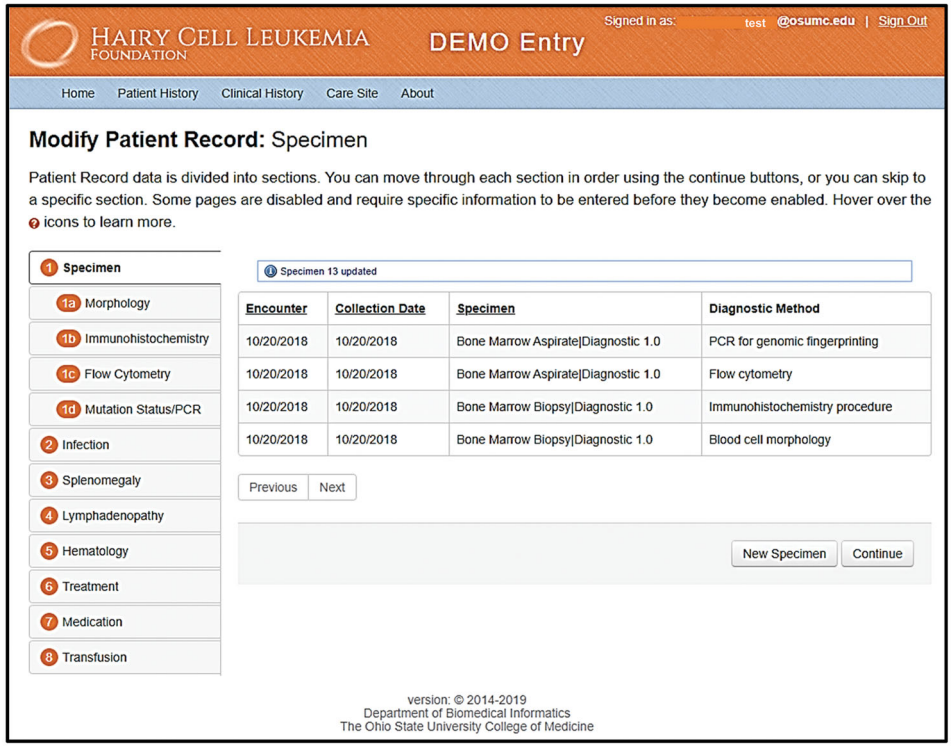


Figure 1. The web-based manual data entry portal (MDE) facilitates entry of information from source documents including pathology reports. The MDE portal can be navigated based on two major tabs divided into patient history (A) and clinical history (B) comprising the entire HCL-PDR data set.

network of collaborators facilitates the exchange of data among participating sites in an on-demand manner while each institution stores and controls its own local data, have proven to be a successful means of enabling data sharing for clinical and translational research [11–13]. Therefore, we initially chose a federated design for the HCL-PDR to allow for institutional autonomy in IRB oversight and local management of patient data, features which also facilitate international collaboration. Under this system, participating institutions uploaded patient data into local 'staging' databases via both manual and automated extract-transform-load (ETL) processes. ETL updates occurred on a scheduled basis, allowing for prospective, longitudinal data capture from each site's electronic health record (EHR). Part of the data were entered manually at the time of study entry and thereafter annually or when a study participant reported an event, such as a new infection or signs of relapse. Data existing in progress notes and external primary source data were entered manually in order to supplement the ETL process (Figure 1). Thereafter, aggregate and de-identified data were able to be queried and exchanged via the use of a shared federated query processor (FQP) communicating with those 'staging databases' in a secure and scalable manner, enabled via secure grid architecture, all made accessible by a web-based query portal. The federated model gave each institution the autonomy to obtain approval from its own Institutional Review Board (IRB) to create and maintain the local 'staging' database (see Supplemental Data for a detailed description of the original HCL-PDR architecture).

However, despite the many theoretical advantages of this system, it quickly became apparent that the ETL component of the registry created enormous difficulties with obtaining security agreements and data sharing agreements, which significantly delayed the on-boarding of new institutions to the registry protocol. These delays added to the expense of the project and in addition the structure of the registry required a third-party vendor to maintain the query portal. The federated system also suffered many disadvantages on the technical side. The heterogeneity in EHR models across sites and lack of standardization in the mapping of clinical data elements to the registry database made it difficult to query the registry data. To address the lack of common data elements across sites, the federated model required the development and deployment of a manual data entry (MDE) portal at each site. The burden for local IT resources made the federated model infrastructure difficult to deploy and maintain. Lastly, we identified that some sites interested in participating in the registry that did not have a functional EHRs system and therefore were unable to join. These issues prompted our group to reconsider the registry structure, and ultimately decided to reconfigure the system to a centralized registry with the inclusion of a MDE portal in order to enable any institution to participate independent of an EHR system.

Development of the revised HCL-PDR centralized software platform

In collaboration with The Ohio State University Wexner Medical Center (OSU) Department of Biomedical Informatics (BMI) and Research Information Technology, the existing federated system was revised to a centralized platform consisting of a MDE portal, an ETL process for automation of data collection (in use only at OSU), and a query portal termed 'Scarlet' [14]. Scarlet is a secure data query portal developed by the OSU BMI, software engineering team used for disease-specific research registries (Figure 2). The MDE portal only allows MDE; data are collected via web forms, de-identified and stored in a structured query language (SQL) server relational database (Figure 3). Each participating site has a dedicated server hosting the MDE application and its own SQL server dataset to store de-identified data. Only authorized users can authenticate and enter data at each site. Separate user accounts are created for each site and authorized by an administrator before data can be entered via the MDE portal. Centralized identity management is performed using OAuth2, an authentication and authorization framework that enables secure access to the MDE application. OAuth2, which is an industry standard process, allows for seamless management of users and applications. The data schema is a common data model (CDM) based on OMOP (Observational Medical Outcomes Partnership) definitions. Adhering to OMOP CDM4 ensures that data from all sites are stored using a common format. All infrastructure components are physically hosted internally and protected by the OSU's enterprise firewall. OSU is the hub (centralized model) for all software elements and thus streamlines and facilitates the upgrades and maintenance of the system. Also, OSU has the capacity to enter data via both the manual and ETL mechanisms. The weekly ETL process pulls data from OSU's EHR and loads it into the registry's OMOP schema. The *merge* ETL process extracts data from each of the site's staging

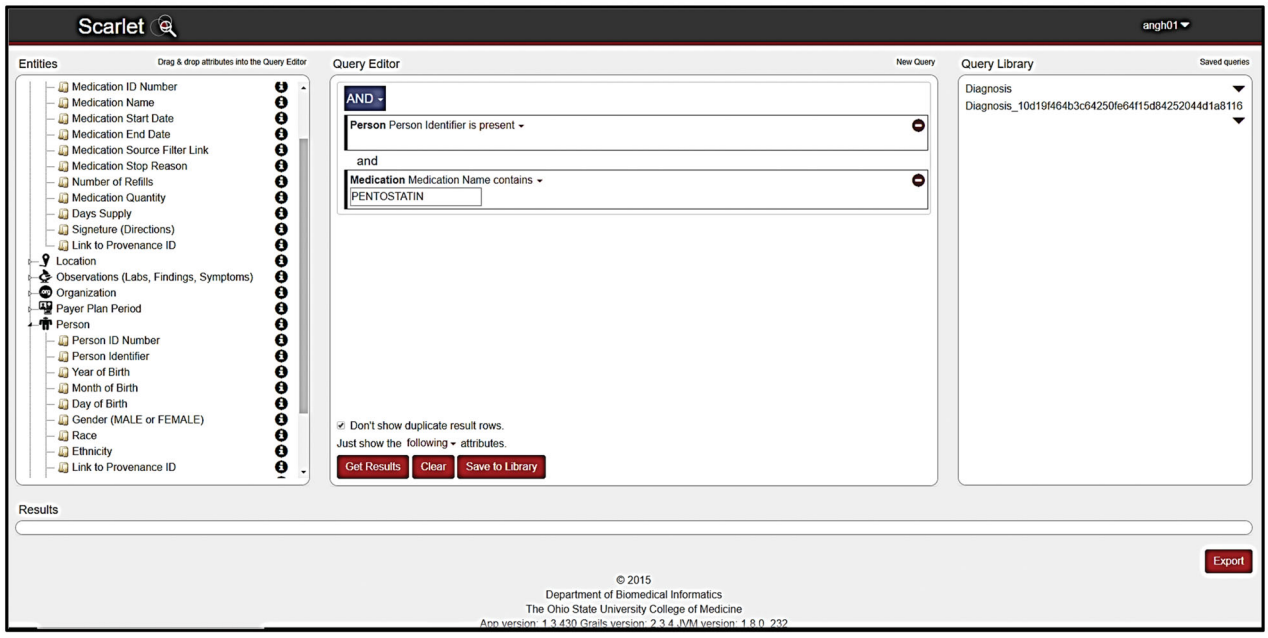


Figure 2. Scarlet query portal showing the fields that are available for query.

databases and consolidates into a single de-identified data registry (Figure 4(A)).

Scarlet, the query portal, allows secure access to the consolidated registry's de-identified data. Query operations are carried out using a web-portal interface and query-construction wizard. The Scarlet query portal is restricted to authorized users, and queries are only performed once investigators have received steering committee approval for data acquisition.

Institutional registry participation and data capture

Following registry software access approval and institutional attainment of regulatory approval, participating institutions can upload patient data into their centralized password-protected staging databases. The MDE data fields have been configured so that a relevant data set can be captured as structured data including diagnostic information such as pathology, laboratory, and radiology results, clinical outcomes, treatment, complications of treatment, comorbidities, family and social history, occupational and exposure history, and obstetric history can be entered in this manner. In addition, narrative data may be more easily entered vial the MDE portal rather than captured electronically (Figure 1). Thereafter, aggregate de-identified data may be queried. This model allows each institution the option of either obtaining approval from its own IRB or to cede reliance to a lead reviewing IRB. This design additionally facilitates international research as it allows each institution to follow its own country's regulatory guidelines.

Patient enrollment

Enrollment of patients is conducted in two ways. Patients treated at a participating COE are consented to the protocol during routine medical visits. Patients in the U.S. who do not have access to a participating site may enter the registry via the HCLF website (www.hairycellleukemia.org), which includes the contact information of a member of the research team who will facilitate informed consent (Figure 4(A)). Source data are then acquired by the research team from the patient's health care provider for upload into the registry. The principal investigator at each participating site is responsible for maintaining the confidentially coded information and the access to that information. To date, 579 patients with HCL have been consented and enrolled across three sites (Figure 4(B)).

Patient data de-identification

Given the sensitive nature of the data stored within the registry, de-identification of patient data is essential for the secure storage of this information. Patient data that enter the HCL-PDR at each site's 'staging database' is pre-processed using both automated and human-mediated processes in order to remove any HIPAA-defined unique identifiers (i.e. name, date of birth, MRN, etc.) (Figure 5(A)). In addition, all dates are subjected to a date-shifting process to obfuscate the

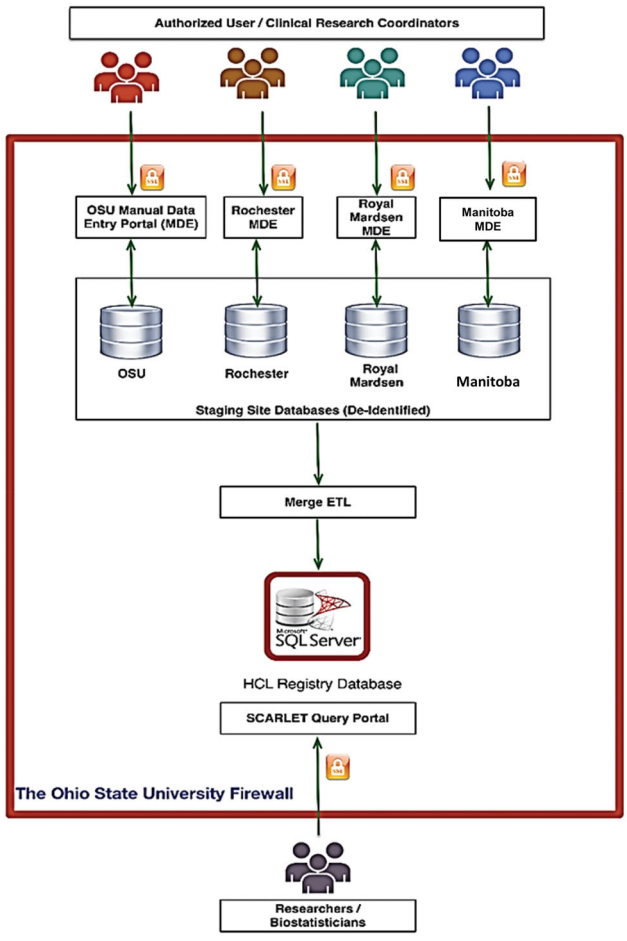


Figure 3. HCL PDR architecture. In this architecture, each institution has a staging database hosted at OSU and has full control of its patients' data. MDE: manual data entry; ETL: extraction, transfer, and load process.

original date but maintain the temporal intervals between dates corresponding to the course of a patient's condition or care. During this process, each patient is assigned a unique research identifier via a secure, nonreversible mathematical transformation of demographic data including initials and birthdate. As the mathematical transformation is one-way, the resulting unique ID cannot be reversed back to any HIPAA-defined unique identifiers (Figure 5(B)). However, each principal investigator maintains a key in order to re-identify patients at their own site should the need arise. This de-identification process was partially adapted from Jacobson RS et al. [13].

Hairy cell leukemia patient data registry network

The HCL-PDR project was initially launched at three institutions: The Ohio State University (Columbus, OH), The University of Rochester (Rochester, NY), and The Royal Marsden Hospital (London, UK). Following implementation at each of these institutions, the protocol was submitted for regulatory approval and activation at The Peter MacCallum Cancer Center (Melbourne, Australia), The University of New Mexico (Albuquerque, NM), The University of Manitoba (Winnipeg, Canada), and The University of Southampton (Southampton, UK). The long-term goal is to include all 29 Centers of Excellence (COE) for HCL around the world, greatly enhancing the demographic diversity of the stored data.

Data sharing and governance

All participating institutions are required to sign the Participation Agreement with the HCLF and The Ohio State University as the hosting site. The Participation Agreement covers details of membership, the HCL registry platform and procedures and responsibility of storing, controlling and sharing the data, and includes a Data Use Agreement which governs policies and procedures for using the system.

The HCL-PDR Scientific Review Committee (SRC), consisting of one representative from each participating COE, a designated representative from the HCLF board of directors, as well as a patient representative, oversees the policies and processes for operating the registry. The SRC considers new data requests and monitors the access and use data. Proposals are evaluated for scientific merit, responsible use of registry resources, and alignment with HCLF goals and strategies. Upon SRC approval of proposals and institutional IRB approval, researchers may access de-identified data from the HCL-PDR through the query portal. This system ensures that all participating institutions are informed about requests to the HCL-PDR and input into how their data are used.

Discussion and future goals of HCL-PDR

The HCL-PDR collaboration consists of institutions that enroll patients on site, in addition to the option for patients to enroll via the web portal from any US location. There are five national and international sites currently undergoing the regulatory approval process. The ultimate goal of this collaboration is for all 29 HCL Foundation COE to achieve full participation. The HCL-PDR contains 429 complete patient records, presenting the opportunity to query a complete data set and launch important studies in HCL that were not previously feasible. Of significant interest to the HCL community are projects examining the demographics and

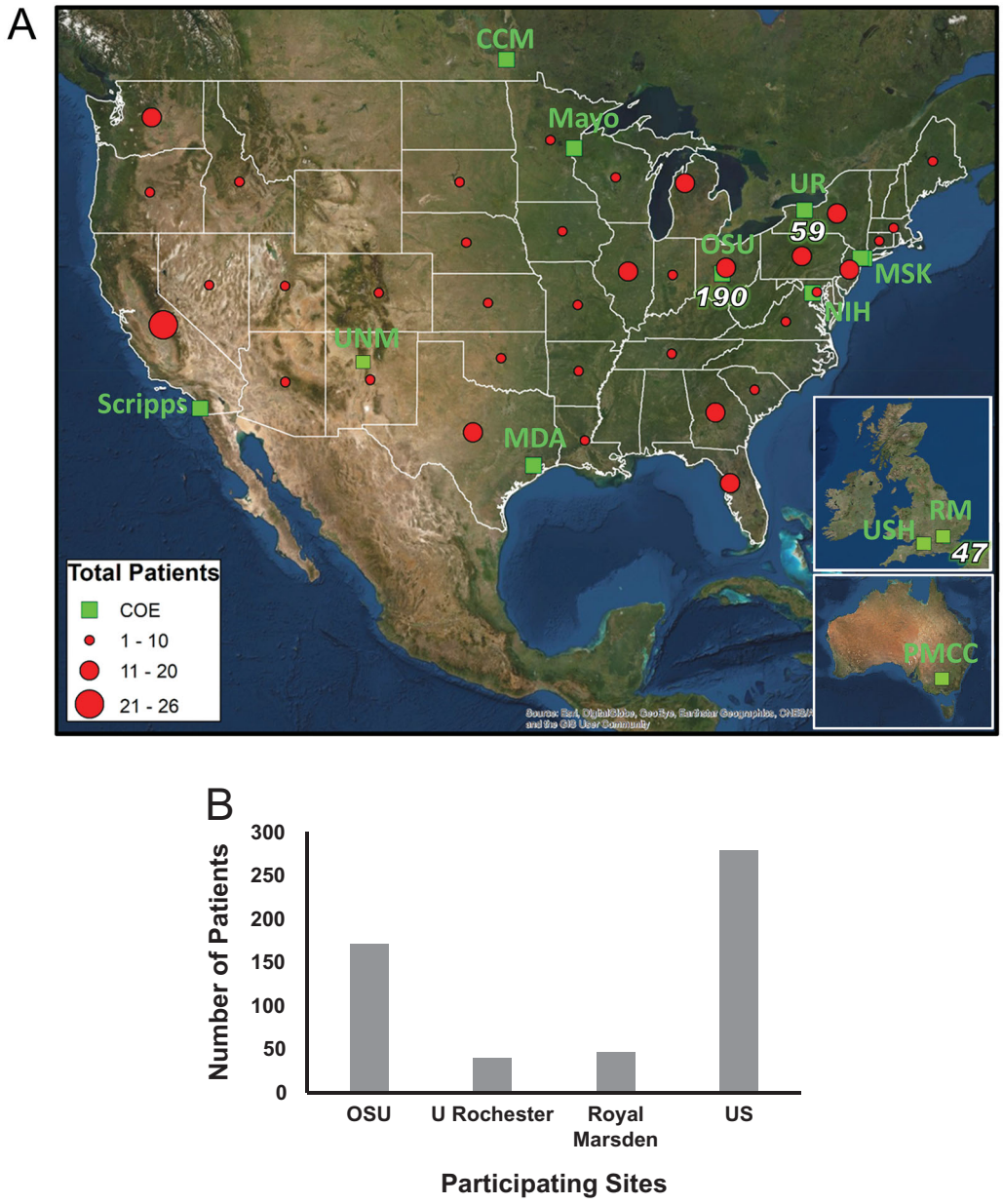
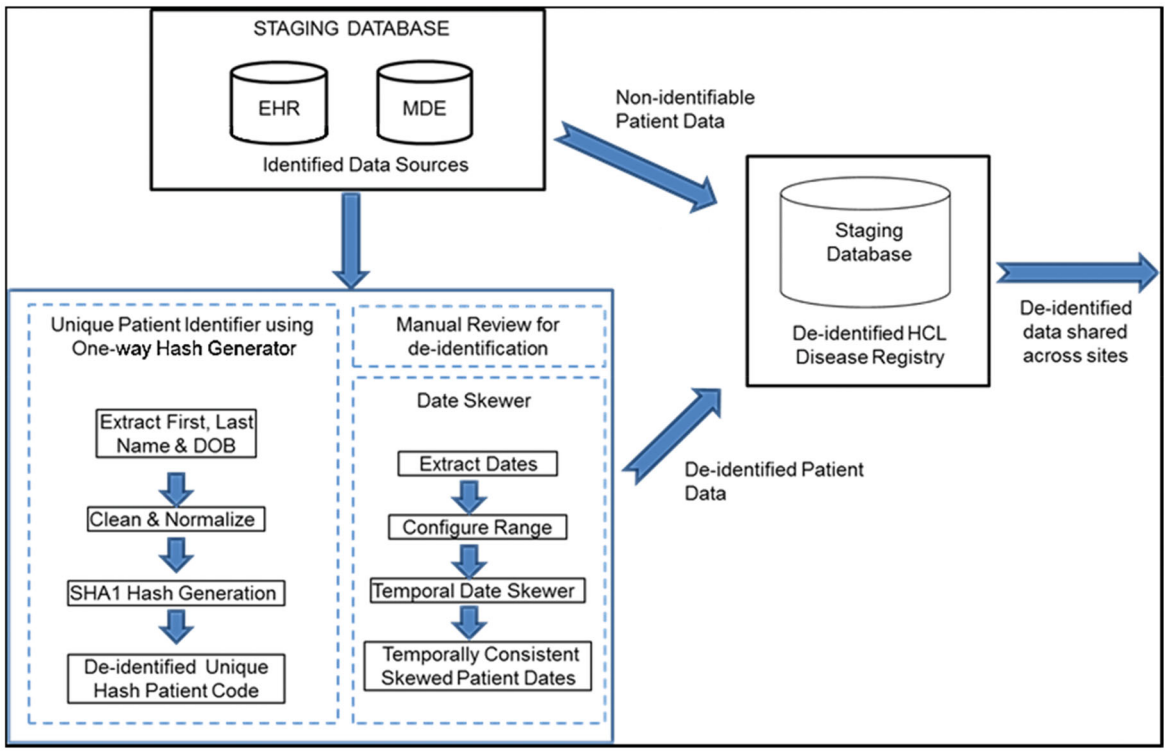


Figure 4. Registry participation. (A) Web-based patients' participation by location (circles) along with the Centers of Excellence (squares). Centers of Excellence that are already part of the registry project have the number of patients enrolled to date indicated next to the representative square. COE: Centers of Excellence; MDA: MD Anderson, Texas, US; MSK: Memorial Sloan Kettering, New York, US; OSU: The Ohio State University, Columbus, US; UR: University of Rochester, Rochester, US; UNM: University of New Mexico, Albuquerque, New Mexico, US; RM: Royal Marsden Hospital, London, UK; USH: University of Southampton, Southampton, UK; PMCC: Peter McCallum Cancer Center, Melbourne, Australia; CCM: Cancer Care Manitoba, Manitoba, Canada. (B) Total number of consented patients at each site. US: patients from United States and outside Center of Excellence, consented via HCLF website.

characteristics of this rare disease in a global patient population, and identifying important patterns and trends in clinical practice. This prospectively collected, longitudinal data may enable the detection of novel predictors of outcome and survival, the definition of disease subtypes and phenotypes, and characterization of complications in the disease course of HCL that are currently poorly understood.

There have been previous efforts to create databases with medical information from patients living with rare diseases. Examples include NORD's Patient registries, Orphanet, and ASCO's Cancer-Linq [15–17]. The Registries for Rare Diseases project launched by NORD is a US effort to collect data from all patients living with rare diseases around the world, while Orphanet has the same mission with an initial focus

A



B

Example of Unique Patient Identifier using One-way Hash Generator

Step 1: Extract Elements

First name: **Joe** Last name: **Smith**

DOB: **20 May 1950**

Step 2: Preprocessing Text

Format: <First 2 characters of First Name : First 2 characters of Last Name : DOB(YYYY-MM-DD)

New text**: <jo:sm:1950-05-20>**

Step 3: SHA1 Hash Code (one-way)

Unique patient code: **136a93f13ce6f350b95eed0faee6dd4ff5eca042**

Example of Date Skewing

Step 1: Configure Range:

Range: maxDaysForward = **0**; maxDaysReverse = **360**

Step 2: Temporal Date Skewer:

De-identified patient hash code based Duration : **300 days**

Skewed Dates :

12 March, 2015 --> 4 Jan, 2016

Figure 5. HCL-PDR data de-identification process. A schematic of the de-identification workflow is shown (A). The de-identification process is demonstrated using a sample patient name and date of birth (B). Example of unique patient identifier generator process. EHR: electronic health records; MDE: manual data entry; HCL: hairy cell leukemia; ETL: extraction, transfer, and load process.

on European countries. Our registry is unique in that it was developed by HCL patients in collaboration with HCL experts from around the world in order to design a database that would contain large number of variables known or suspected to influence HCL outcomes. In addition, the registry architecture was designed to accommodate structural alterations to capture of new discoveries. To our knowledge, this is the first multi-institutional HCL specific registry ever created.

This multi-institutional data registry can be used to develop and implement studies with appropriate sample sizes, and a more diverse patient population for these rare diseases, whereas single center registries may be confounded by a lack of diversity in the patient population, as well as a homogeneous approach to treatment. Furthermore, our network design allows patients who are not being seen at a COE to enroll in the HCL-PDR via the HCLF website, allowing us to evaluate possible differentiating factors in the care and outcomes of patients who are being treated in community care sites. The data from these patients is housed at only one site, The Ohio State University, constituting a centralized data capture mechanism.

While the federated model presented many innovative achievements in registry development and utilization, the drawbacks we encountered such as delays due to institutional concerns regarding security considerations around registry software deployment connecting the query portal to the EHR could not be overcome despite concentrated efforts by all parties involved. Therefore, we transitioned our project to a centralized model, ensuring that all institutions are ultimately able to participate and thereby enhancing the reach of the registry project. Only de-identified data from individual databases are accessed by the Scarlet query portal. Although the data from participating institutions are stored in a centralized location, each institution has its own password-protected staging databases and has direct control over identified data.

We believe that the architectural approach of the HCL-PDR provides a novel and highly scalable model for enabling data sharing in the context of a rare disease-focused consortium. With the infrastructure for the HCL-PDR now in place, we are planning for the incorporation of a multi-institutional biobank for HCL patient samples in order to link novel molecular and genetic findings to clinical data. In addition, a wide range of parallel and future studies will be supported by the HCL-PDR, including longitudinal quality of life assessments, updates to diagnostic and risk categorizations, and longitudinal assessments of treatment outcomes and long-term complications. The existence of such a resource vastly improves access to research-grade data, helps to organize and inform the community of clinicians and scientists, and serves as a platform for the development of new clinical trials, with the ultimate goal of improved patient outcomes.

Acknowledgements

We are grateful to Joshua W. Sadvari from OSU University Libraries, for his help with mapping the registry patients, as well as to the following developers whose contribution has been instrumental in making the registry software platform: Wei-Chao Chuang; John Strickler; John Thompson. We also acknowledge Philip Payne for his contribution to the initial design of the federated registry platform.

Author contributions

All authors contributed to the initial design of the registry. LA, MA, JN, PM, SAF, and MRG performed the study and/or treated study participants. PM, OL, and SAF designed and implemented the registry platform. LAA, JSB, CD, SI, MC, CZ, KAR, NE, EK, SB, and MRG enrolled patients in the registry. LAA, MA, and MRG wrote the first draft of the manuscript. ASR, QZ, and MA contributed to MDE design. All authors reviewed the manuscript and agreed with submission for publication.

Disclosure statement

The Ohio State University (OSU) and Signet Accel., LLC (SA) had a financial conflict of interest in that SA provided technical support and maintained the database infrastructure for the Hairy Cell Leukemia Patient Data Registry, a project initiated and developed at OSU. OSU had a financial interest in SA as part of a licensing agreement. Dr. Phillip Payne had a financial interest in SA. SA is no longer providing database support, and these functions are now accessed through the Department of Research Information Technology at The Ohio State University. LA previously provided consultation services for Innate Pharma and AstraZeneca. The other authors declare no competing interests. MA has consulted for AstraZeneca in 2019; received travel funding from Hairy Cell Leukemia Foundation. JSB was on Advisory Board and has received consulting fees from: AbbVie, AstraZeneca, INNATE Pharma, KITE Pharma; has received research funding from MingSight Pharmaceuticals; has patents and intellectual property on a leukemia diagnostic device (patent pending). OL is currently employed by Bristol-Myers Squibb Company. His contribution to the work described in this manuscript was during his employment with The Ohio State University which ended as of June 2018. CD has been an advisor for Medimmune/Innate Pharma (moxetumomab); consulting/Advisory Board for Abbie and Jansen. SI: Advisory and speaker fees: Gilead and Takeda; Advisory: Beigene; Speaker fees: Janssen, Takeda and Gilead. CSZ: funding through the University of Rochester for laboratory research, from Acerta/

AstraZeneca and TG Therapeutics, and from the Hairy Cell Leukemia Foundation. KAR receives research funding from Genentech, AbbVie, Novartis, and Janssen, has consulted for Acerta Pharma, Genentech, AbbVie, Pharmacyclics, Innate Pharma, and AstraZeneca, and received travel funding from AstraZeneca; NE is on the Speakers Bureau: Incyte; Ad Board: TG therapeutics, Beigene, Pharmacyclics; Consultancy: Novartis; GL has received travel funding from Hairy Cell Leukemia Foundation. ASR is currently employed by Eli Lily and Company, though contributions to the work described in this manuscript occurred during employment with The Ohio State University which ended as of January 2022; served on an independent DSMB for Telios Pharma; SAB served on Advisory board for Pharmacyclics and Janssen, Beigene and AstraZeneca; received honorarium from OncLive; received travel grant from Arqule; VB serves on the Advisory Boards of Janssen, AstraZeneca, and Abbvie and has had research funding from CIHR, LLSC, CCMF, Roche, Janssen, Abbvie, and Lundbeck. VB also received fees from BIOGEN for patented compounds unrelated to this study; SH: Abbvie (Travel Grant), Novartis (Travel Grant), Roche (Honoraria), Gilead (Honoraria); CST: Honoraria from Janssen, AbbVie, and Beigene, and his hospital receives research funding from Janssen, AbbVie, and Beigene; JFS: AbbVie, Advisory board, speakers' bureau, research funding; Astra Zeneca, Advisory board; Celgene, Advisory board, speakers' bureau, research funding, expert testimony; Genentech, Advisory board; Gilead, Advisory Board; Janssen, Advisory board, research funding; Mei Pharma, Advisory board; Morphosys, Advisory board; Roche, Advisory board, speakers' bureau, research funding, expert testimony; Sunesis, Advisory board; Takeda, Advisory boardME; RJK: Employment: National Institutes of Health and Regional Cancer Care Associates; Honoraria: PlatformQ, OncLive, Cure; Research Funding: Innate, AstraZeneca, Novartis, Genetech, Pfizer, Teva, Hairy Cell Leukemia Foundation; Patents: Coinventor for NIH patent for Moxetumomab Pasudotox; AS: Consultant and Advisory Boards: AstraZeneca and Innate Pharmaceuticals; Speakers Bureau: AbbVie and Pharmacyclics; SAP: Research funding has been provided to the institution from Pharmacyclics, Janssen, AstraZeneca, TG Therapeutics, Merck, AbbVie, and Ascentage Pharma for clinical studies in which Sameer A. Parikh is a principal investigator. SAP has also participated in Advisory Board meetings of Pharmacyclics, AstraZeneca, Genentech, GlaxoSmithKline, Innate Pharma, Adaptive Biotechnologies, and AbbVie (he was not personally compensated for his participation); FR: Consultancy and Honoraria – Celgene, BMS, Amgen, Astellas, Xencor, Agios, AstraZeneca, Orsenix, Innate Pharma, Syros, Taiho, Novartis; Research Funding – BMS, Amgen, Xencor, Macrogenics, Orsenix, Abbvie, Taiho, Prelude, Astex; ET: Consultant for Innate Pharma. Research funding: Roche. Travel cost: Shire. Holder of a patent on the use of mutant BRAF as HCL biomarker. Our research work in HCL is funded by the Hairy Cell Leukemia Foundation, the Leukemia and Lymphoma Society and the Associazione Italiana Ricerca sul Cancro (AIRC); XT: Consultant for Innate Pharma, AstraZeneca; Advisor: Abbvie; MST: Research funding: AbbVie, Amgen, Biosight, Glycomimetics, Orsenix, Rafael; Advisory boards: Amgen, Daiichi-Sankyo, Delta Fly Pharma, Innate pharmaceuticals, Jazz, Kahr, Kura, Novartis, Orsenix, Roche, Syros; Royalties: UpToDate; AG: Advisory board: Amgen, Takeda. Research funding: Jansenn and Cilag SPA. Honoraria: Abbvie, Jansenn Cilag, Celgene, Amgen, Takeda; PLZ: Speakers Bureau: Verastem, Celltrion, Gilead, Janssen-Cilag, BMS, Servier, MSD, TG Therap., Takeda, Roche, Eusapharma, Kyowa Kirin; Advisory Board: Verastem, Celltrion, Gilead, Janssen-Cilag, BMS, Servier, Sandoz, MSD, TG Therap., Takeda, Roche, Eusapharma, Kyowa Kirin, ADC Therap.; Consultant: Verastem, MSD, Eusapharma, Sanofi; TR: Research funding: AstraZeneca, Medimmune, Roche, Janssen, Abbvie, Advisory board; AstraZeneca, Jassen, Abbvie; Travel grant: Roche, Janssen; Honoraria: Abbvie, AstraZeneca, Janssen; TT: Consultant Abbvie, BMS, Roche, Janssen, Novartis, Pfizer, Takeda, Astra Zeneca, Gilead; JD: Has participated in advisory committees in Hungary for Novartis, Bristol Myers Squibb, Amicus, Angelini, Pfizer, Amgen, and Roche; MG has been consultant for AstraZeneca, Pharmacyclics, Ascerta, Axio, Inc.; EMD Serono (Merck) Research and Development Institute; has received research funding and travel expenses from Hairy Cell Leukemia Foundation; Scientific Board Chair: Hairy Cell Leukemia Foundation Scientific Board (no reimbursement); Scientific Honorarium: University of Pittsburgh. The other authors declare no competing interests.

Funding

Research reported in this publication was supported by Hairy Cell Leukemia Foundation. OSU services utilized include the College of Medicine Research Information Technology (National Center For Advancing Translational Sciences Grant: UL1TR002733, OSU Comprehensive Cancer Center Support Grant: P30CA016058, and National Center for Advancing Translational Sciences, Grant UL1TR002733).

ORCID

Leslie A. Andritsos http://orcid.org/0000-0003-0222-1766
Clive S. Zent http://orcid.org/0000-0001-6099-3313
Narendranath Epperla http://orcid.org/0000-0002-8216-3457
Robert J. Kreitman http://orcid.org/0000-0001-5141-1844
Sameer A. Parikh http://orcid.org/0000-0002-3221-7314
Farhad Ravandi http://orcid.org/0000-0002-7621-377X
Alessandro Gozzetti http://orcid.org/0000-0003-0769-6891
Tadeusz Robak http://orcid.org/0000-0002-3411-6357

References

[1] Bouroncle BA, Wiseman BK, Doan CA. Leukemic reticuloendotheliosis. Blood. 1958;13(7):609–630.
[2] Bouroncle BA. Leukemic reticuloendotheliosis (hairy cell leukemia). Blood. 1979;53(3):412–436.
[3] Pettirossi V, Santi A, Imperi E, et al. BRAF inhibitors reverse the unique molecular signature and phenotype of hairy cell leukemia and exert potent antileukemic activity. Blood. 2015;125(8):1207–1216.
[4] Shao H, Calvo KR, Gronborg M, et al. Distinguishing hairy cell leukemia variant from hairy cell leukemia: development and validation of diagnostic criteria. Leuk Res. 2013;37(4):401–409.

[5] Kraut E. Infectious complications in hairy cell leukemia. Leuk Lymphoma. 2011;52(Suppl. 2):50–52.
[6] Grever MR, Abdel-Wahab O, Andritsos LA, et al. Consensus guidelines for the diagnosis and management of patients with classic hairy cell leukemia. Blood. 2017;129(5):553–560.
[7] Matutes E. Diagnostic and therapeutic challenges in hairy cell leukemia-variant: where are we in 2021? Expert Rev Hematol. 2021;14(4):355–363.
[8] Jansen J, Hermans J. Clinical staging system for hairy-cell leukemia. Blood. 1982;60(3):571–577.
[9] Teras LR, DeSantis CE, Cerhan JR, et al. 2016 US lymphoid malignancy statistics by world health organization subtypes. CA Cancer J Clin. 2016;66(6):443–459.
[10] HCL. Hairy Cell Leukemia Foundation; 2021, www.haircellleukemia.org.
[11] Ohno-Machado L, Agha Z, Bell DS, et al. pSCANNER: patient-centered scalable national network for effectiveness research. J Am Med Inform Assoc. 2014;21(4):621–626.
[12] Weber GM, Murphy SN, McMurry AJ, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. J Am Med Inform Assoc. 2009;16(5):624–630.
[13] Jacobson RS, Becich MJ, Bollag RJ, et al. A federated network for translational cancer research using clinical data and biospecimens. Cancer Res. 2015;75(24):5194–5201.
[14] Mathur P. Design of a scalable and affordable local research registry pipeline [abstract]. AMIA Joint Summits Proceedings; 2016.
[15] Orphanet, The Portal for Rare Diseases and Orphan Drugs, 2021, https://www.orpha.net/consor/cgi-bin/index.php
[16] NORD – National Organization for Rare Disorders. Hairy cell leukemia; 2013, https://rarediseases.org.
[17] Schilsky RL, Michels DL, Kearbey AH, et al. Building a rapid learning health care system for oncology: the regulatory framework of CancerLinQ. J Clin Oncol. 2014;32(22):2373–2379.