Society for Immunotherapy of Cancer clinical and biomarkers data sharing resource document: Volume I—conceptual challenges

Sergio Rutella,[1,2] Michael A Cannarile,[3] Sacha Gnjatic,[4] Bruno Gomes,[5] Justin Guinney,[6] Vaios Karanikas,[7] Mohan Karkada,[8] John M Kirkwood,[9] Beatrix Kotlan,[10] Giuseppe V Masucci,[11] Els Meeusen,[12] Anne Monette,[13] Aung Naing ![ORCID],[14] Vésteinn Thorsson,[15] Nicholas Tschernia,[16] Ena Wang,[17] Daniel K Wells,[18] Timothy L Wyant,[19] Alessandra Cesano[20]

ABSTRACT

The sharing of clinical trial data and biomarker data sets among the scientific community, whether the data originates from pharmaceutical companies or academic institutions, is of critical importance to enable the development of new and improved cancer immunotherapy modalities. Through data sharing, a better understanding of current therapies in terms of their efficacy, safety and biomarker data profiles can be achieved. However, the sharing of these data sets involves a number of stakeholder groups including patients, researchers, private industry, scientific journals and professional societies. Each of these stakeholder groups has differing interests in the use and sharing of clinical trial and biomarker data, and the conflicts caused by these differing interests represent significant obstacles to effective, widespread sharing of data. Thus, the Society for Immunotherapy of Cancer (SITC) Biomarkers Committee convened to identify the current barriers to biomarker data sharing in immuno-oncology (IO) and to help in establishing professional standards for the responsible sharing of clinical trial data. The conclusions of the committee are described in two position papers: Volume I—conceptual challenges and Volume II—practical challenges, the first of which is presented in this manuscript. Additionally, the committee suggests actions by key stakeholders in the field (including organizations and professional societies) as the best path forward, encouraging the cultural shift needed to ensure responsible data sharing in the IO research setting.

THE UNMET NEED FOR DATA SHARING IN IMMUNO-ONCOLOGY

Advances in cancer immunotherapy have significantly improved life expectancy, quality of life and overall survival for some patients with cancer. However, only a relatively small percentage of patients receiving immuno-oncology (IO) treatments currently achieve durable responses and lasting disease-free survival, and the mechanisms and patient attributes underlying immune responsiveness and resistance have yet to be fully understood.[1] Given the high costs and potential side effects associated with these therapies, it is crucial to identify the subsets of patients who will clinically benefit from immunotherapy. In addition, understanding the biological basis of mechanisms of resistance to IO treatments is essential for identifying new therapeutic approaches, including the design of rational therapeutic combinations.[2]

It is now widely accepted that in the field of IO, the conventional single biomarker approach has failed to have a substantial impact in improving treatment efficacy or in informing the next generation of IO therapeutics due to the broad heterogeneity of malignant diseases and limited sizes of patient subgroups providing data.[3 4] In addition, the complex and dynamic nature of the tumor microenvironment, together with variability in host genetic background and environmental factors, makes the identification and subsequent validation of biomarkers from single trials a nearly impossible endeavor. Since the determinants of cancer immune responsiveness are multifactorial (including genetic makeup of the patient, tumor genomic instability, epigenetic adaptation, and external modifiers such as the microbiome, concomitant medications, comorbidities, etc), collection and analysis of 'big data' sets for biomarker discovery are becoming the norm in the IO clinical research field.[5–7] Therefore, converging efforts in biomarker identification and validation using existing datasets and combined patient populations with properly collected clinical and biomarker data (ie, meta-analyses) are greatly needed to guide patient selection, treatment

decisions, and advances in new therapy developments in IO.

To date, the development and validation of new biomarkers for IO has been limited by the fragmentation of clinical research efforts, often conducted in a drug-centric way, statistically designed to determine efficacy or safety end points instead of biomarker validation endpoints, and have used investigator/sponsor preferred non-standardized technology platforms, with different levels of analytic validation applied to relatively small data sets (usually with low numbers of responders). The combination of all of these factors results in a lack of intrinsic robustness and underpowered statistical outputs, making most observations hypothesis-generating, at best.

In this context, harmonized clinical and biomarker data sharing can leverage, unify, and maximize sample sizes; enhance biomarker identification, validation, and impact on clinical decision making; and provide a launching pad to accelerate new discoveries. This can be achieved by using existing and completed studies, potentially avoiding future duplicative trials (unless there is a clear mechanistic reason that a drug within an existing class would have better efficacy in a different study population and/or indication). The success and benefit of data sharing will include facilitating the identification and validation of standardized biomarkers through meta-analysis and enabling maximal scientific knowledge and benefits to be gained from the efforts of clinical trial participants and investigators. However, the enabling of this paradigm shift in the field of clinical data sharing and IO biomarker discovery and validation will require a concerted effort from all involved stakeholders.

Current challenges to data sharing can be divided into two major categories:
► Conceptual Challenges: Key stakeholders may have conflicting interests (both financial and academic) and ethical and regulatory restrictions, which need to be taken into consideration and balanced for data sharing to become the norm in clinical research.[8–12]
► Practical Challenges: Effective data sharing requires the improvement and/or standardization of the systems and protocols governing clinical research, including minimal information required for data alignment, standardized statistical approaches, data management infrastructure, technology for data acquisition and analysis, workforce training, and equitable distribution of the costs of research.[12–14]

The Society for Immunotherapy of Cancer (SITC) Biomarkers Committee formed the Clinical and Biomarkers Data Sharing Subcommittee to discuss challenges from both of these categories. Conceptual challenges and related possible mitigation steps to data sharing in IO are described in detail in this manuscript, while the practical challenges and recommended mitigation strategies are discussed in the companion manuscript 'The SITC clinical and biomarkers data sharing resource document: Volume II—practical challenges.' Importantly, these challenges need to be addressed in parallel and in a holistic way in order to make data sharing in IO the standard in clinical research.

An unprecedented example of the power of open data sharing has come from the COVID-19 pandemic. In January 2020, just days after isolation of the SARS-CoV-2 virus, its genome sequence was published and shared with the scientific community.[15] Since then, characterization of the virus and associated disease has continued, with additional isolates sequenced and data openly shared, allowing researchers to monitor and track the progression of the pandemic. For example, NextStrain has been tracking and updating information on the progression and changes observed in the viral genome.[16] Sequences can be uploaded and shared through the Global Initiative on Sharing all Influenza Data.[17] Epidemiology data are being made available by the Johns Hopkins University of Medicine,[18] Worldometers,[19] the nCoV-2019 data working group,[20] Our World in Data,[21] and the World Health Organization (WHO)[22] dashboards on the coronavirus disease pandemic, which keep ongoing statistics on the spread of the virus, and appraise health officials and the world of the situation, therefore, allowing for critical movement of resources to impacted regions. The WHO has acted to ensure open access to scientific publications, creating a database of 77103 citations on the COVID-19 pandemic and SARS-CoV-2 as of October 7 2020.[23] Importantly for the concerns of oncologists, the COVID-19 and Cancer Consortium was rapidly formed to share data on cancer patients with COVID-19 among many institutions.[24 25] Additionally, many journals and professional societies, including SITC, have encouraged and enabled open access to articles related to COVID-19 research.[26] As a consequence of efforts like those discussed above, the rapid availability of data such as epitope sequences including the binding site of SARS-CoV-2 spike protein to its entry receptor, angiotensin converting enzyme 2 (ACE2), is propelling the generation of diagnostic tests, vaccines, and antibody therapeutics.[27 28] Additionally, data sharing during the COVID-19 pandemic has been facilitated through the availability of pre-print services such as medRxiv and bioRxiv; however, the sheer number of preprint articles in COVID-19-related research has prompted calls to parse this non-peer-reviewed literature and grade articles by quality.[29 30]

CONCEPTUAL CHALLENGES TO DATA SHARING IN IO
Ethical considerations: patient privacy and 'big data' challenges
Even though most informed consent processes approved by ethics committees have embedded clauses allowing the future nonprofit use of deidentified data, patients' willingness to share their de-identified data does come with specific requirements in order to protect their privacy, choices, and needs for information regarding the use of their data. Efforts to balance patients' right to data protection/confidentiality against the need to promote medical science advances through clinical research has been a

topic of significant debate in recent years.[31] To ensure ethical and regulatory compliance and the sustainability and success of data sharing initiatives in cancer research, appropriate legislation and guidelines need to be implemented. Joint multi-stakeholder agreements and efforts need to be pursued to foster cultural (ie, conceptual) and technological (ie, practical) adaptation or discovery-led changes that will enable patients' preferences regarding sharing of their own health data to be respected.

In the US, the Health Insurance Portability and Accountability Act (HIPAA) is a federal law that governs the use and disclosure of personal health information by covered entities, defined as health plans, healthcare clearinghouses, and healthcare providers. The general rule is that personal health information cannot be disclosed without the patient's authorization. Three primary measures protect patient privacy and confidentiality: (1) informed consent, (2) study review and approval by Institutional Review Boards (IRBs) and Research Ethics Committees (RECs), and (3) data use limited to deidentified information.

The intent behind informed consent regulations is to ensure that, before providing their consent, participants are fully aware of the study they are volunteering for, and that they have been given a clear and accurate account of the potential risks and expected benefits of the study. In relation to data sharing, it is important that the informed consent form includes appropriate language explaining the information collected including data generated from integrative genetic/genomic assays on the patients' specimens and how this data will be stored and used, including the potential for being shared outside the boundaries of the specific study in which they were collected, such as for integrative meta-analyses.

IRBs and RECs are tasked with reviewing, revising, and approving clinical investigation protocols involving humans with the goal of protecting research participants and ensuring that they are treated ethically in the course of their participation in studies.[32–34] These committees play close attention to the informed consent process and should provide guidance for investigators and informed consent templates for participants to enable responsible data sharing. In addition, IRBs should influence the responsible sharing of individual participant data by adding considerations for data sharing plans when assessing the benefits and risks of clinical trials in IO. Moreover, consistent policies within and between institutions are needed so that investigators, IRBs, and other institutional officials know what level of protection they can promise to participants and participants can make decisions based on accurate and truthful information.

Data deidentification by removing individually identifiable health information from data sets is another measure aimed to protect patient privacy. However, the increase of research involving 'big data,' particularly in IO, may impair efforts to maintain patient privacy. Because big data can contain family history and genetic information, absolute data deidentification can be challenging.[35 36] Indeed, one of the main ethical challenges and risks posed to clinical trial participants that has been identified regarding data sharing is the possible reidentification of previously deidentified data. The standard data protections used to anonymize or deidentify patient-representative data were devised prior to current deep sequencing, multiplex technologies, and multiomics platforms that produce higher resolution, more detailed information on patient samples.[37]

The National Science Foundation funded an initiative to establish the Council for Big Data, Ethics, and Society, which discusses the importance of maintaining ethical standards in the relatively new field of big data analysis. The council has also noted that big data often resides outside of existing systems designed to ensure that research is ethical, and noted several unique issues associated with big data research. These unique issues include: non-physical harm resulting from research (harm by enabling surveillance or discrimination), unpredictable uses of data in the future (beyond the scope of the original study), and the possibility that individually 'safe' data sets may be combined to generate data capable of causing harm to participants.[38] The council also made a series of recommendations regarding policy, pedagogy, and networking.

Disease advocacy organizations share many of the same concerns, roles, and responsibilities as those of other non-profit funders and clinical trial sponsors with regard to data sharing. Whether through direct financial support (either alone or as part of a funding syndicate) or other forms of assistance (eg, participant recruitment or clinical research networks), disease advocacy organizations make significant contributions to the development and execution of clinical trials. These efforts give these organizations an opportunity to influence policies and strategies to encourage responsible sharing of clinical trial data.[39]

Data ownership: protection of intellectual property
Historically, the culture surrounding clinical research in industry has not encouraged the proactive sharing of clinical trial data, including biomarker data. There are several reasons for this reluctance. Investors and industry are mostly driven by business models that require the generation of revenue. Therefore, data gathered with industry sponsorship or in support of marketing applications is often considered commercially confidential and competitively advantageous, representing a significant resource investment by the sponsoring organization. As a consequence, new findings that could be competitively advantageous are commonly considered protected intellectual property (IP), which impedes data sharing. Unwillingness to relinquish control of data for fear of misuse and the potential for incidental findings that require regulatory application resubmission have been additional concerns among the pharmaceutical industry.

Over the last 5 years, data sharing within the pharmaceutical industry has evolved from being virtually non-existent to being a landscape in which most companies

have participated in one form or another, both for the public good and for the opportunity to uncover new insights and add value to existing data. This has been particularly true during the current COVID-19 pandemic response, and for the field of IO, where access to 'big' datasets is crucial. In general, access to, and collaboration with, top scientists and a perceived trustworthiness of academic involvement in data analysis is an important factor in the sharing of clinical trial data by pharmaceutical companies.[40]

The Foundation for the National Institutes of Health (FNIH) has launched and coordinated the Partnership for Accelerating Cancer Therapies (PACT) program, aimed at building an oncology clinical study database with state-of-the art study design, technology implementation, and pipelines for databases, bioinformatics, and biostatistics, as well as planning for future public open access to the database. This program is truly a partnership effort contributed to by the US government, industry, biotech, biopharma, and academia. It lays the foundation for clinical data sharing in principle and in practice in the future and promotes the sharing of existing data for retrospective exploratory biomarker discovery. However, for the pharmaceutical sector to truly benefit, data collaboration needs to be incorporated into business as usual, rather than remaining the domain of special projects.

Participant motivations: formal recognition and career incentives

In most academic research communities, publications are the primary currency. Promotions, grants, and recruitment policies are often based on publication track-records. The demands and offers of publication outlets, therefore, have an impact on the individual researcher's data sharing disposition. Because academic success depends on published outputs, and because only a small fraction of the accrued data is published in the primary report of a clinical trial, many academics have been reticent to share data. A 2018 European survey showed that more than 30% of researchers do not routinely share scientific data.[41] Fear that competitors may publish novel findings and/or misuse the original data represents an additional hurdle to broad data sharing. This understandable response to a lack of incentives and protection is in line with other situations in the biomedical field, such as the publication of negative data.[42] Thus, there is a cultural gap and a lack of governing framework that needs to be addressed by educating the scientific community and by policy/guideline development. Data sharing in academia is a multidimensional effort that includes a diverse set of stakeholders, entities, and individual interests. In the fast moving and fast publishing modern era of biomedical sciences, unless sufficient recognition is paid to the intellectual and physical efforts involved in designing, accruing, and curating comprehensive and usable clinical trial data sets slated for sharing with the greater medical community, individual investigators will continue to work in silos, and forgo the data sharing process that is essential to continued advancement.

Integral steps toward remedying this issue are for governing frameworks to be built with data protection and data sharing in mind, and for clinical investigators to recognize the value of data sharing and the accompanying infrastructure. This could be promoted by inviting collaboration and co-authorship on publications resulting from the use of existing data in new research. In addition, funders can accelerate this process by providing tangible rewards to clinical investigators for data sharing activities. Research institutions and universities should also make sharing of clinical trial data a consideration in promotion of faculty members and assessment of programs. Finally, training for data science and collaboration with quantitative scientists to facilitate sharing and analysis of clinical trial data should also be emphasized.

Journals: impetus and platform for data sharing

Biomedical journals can play a central role in championing and providing recognition for data sharing as well. Currently, the academic h-index of success is calculated based on the set of the scientist's most cited papers and the number of citations that they have received in other publications.[43] Authors actively participating in data sharing should also be scored and recognized accordingly for advancing the field. With fewer restrictions in publication lengths and upper limits on numbers of references resulting from electronic/online journals (in contrast to past hard copy issues), journals should be responsible for ascertaining that all primary data set contributors be properly cited to increase their individual professional scores, enhancing academic or industrial recognition, and thereby sustaining the motivation to share data. More importantly, journals can take the lead in promoting data sharing norms.

While a statement recommending data sharing has been accepted as a step in the right direction by the International Committee of Medical Journal Editors initiative, response was starkly divided, and prominent research funders including the Wellcome Trust, National Institutes of Health (NIH), Medical Research Council (MRC), Cancer Research UK, and Bill & Melinda Gates Foundation declared that the mandates are still vague and open for interpretation.[44]

Nevertheless, many higher impact journals already mandate data set deposition to repositories (eg, the Gene Expression Omnibus or the database of Genotypes and Phenotypes) at the time of manuscript submission, in addition to mandates on explicit clarity on algorithms used for data analyses reported. Though repositories provide templates and guidelines for correct summary reporting for gene expression and DNA sequencing data set submissions, there is a lack of common clinical data elements from patients. Such a set of common elements would enhance the usability and harmonization of data sets for meta-analyses investigating predictive biomarkers for immunotherapies, and could be established by

journal-linked repositories. A notable set of meta-studies demonstrated that journals that require data sharing by authors for peer review and publication typically have significantly higher impact factors, suggesting that journals may also benefit from mandating and streamlining standards (eg, providing guidelines and templates) that ensure sustainable data sharing.[45 46]

Additionally, preprint servers, including bioRxiv and medRxiv, operated by Cold Spring Harbor Laboratory, could accelerate the pace of science and expedite data sharing with a large audience long before research data are published in peer-reviewed journals, while also ensuring that ideas are not 'stolen' and that a record of priority is established. This avenue of data sharing has been used frequently, more recently during the COVID-19 pandemic, motivated by the need to share emerging relevant information in real time with the medical and scientific community. Major research funders, including the MRC, the NIH, and the Howard Hughes Medical Institute, have recently begun to encourage the citation of pre-prints and urge grant applicants to cite pre-prints in their funding proposals. However, it should be made clear that pre-prints reporting the results of clinical trials have not been peer-reviewed, and that they should therefore not be used to guide clinical decision making by other researchers.

Further, to create a responsible medium for biomarker data sharing, societies such as SITC and journals (including the *Journal for Immunotherapy of Cancer* (*JITC*)) can initiate and support an independent body that promotes good practices and standardized data stewardship to ensure that biomarker data are fairly and consistently presented; valid based on sample size and methodology; and accessible, interpretable, reusable, and interoperable as per the 'Findable, Accessible, Interpretable and Reusable' guiding principles for scientific data management and stewardship.[47]

Professional societies: hubs for data sharing and biomarker development

At the intersection of improvements in patient outcomes and the advancement of IO research are professional societies. Organizations including SITC, the American Association for Cancer Research (AACR), the American Association of Immunologists (AAI), the American Society of Clinical Oncology (ASCO), the American Society of Hematology (ASH), and the American Society for Transplantation and Cellular Therapy (ASTCT), among several others, represent the collective interests of thousands of physicians, research scientists, advanced practitioners, nurses, biostatisticians, and patient advocates.

With the historical strength of professional societies as a guide, these organizations have great leverage in the pursuit of large-scale collaboration and data sharing. As an example, ASH is creating infrastructure and is designing a process for sharing genomic and clinical data on multiple myeloma called the ASH Research Collaborative (ASH RC), an initiative that could serve as a model for other

diseases. The ASH RC Data Hub will also host industry or government datasets and will be able to accommodate data in different formats, including patient-reported instruments and manual chart abstraction.[48]

Another example, the AACR Project Genomics Evidence Neoplasia Information Exchange (GENIE) is a publicly accessible international cancer registry assembled through data sharing.[49] The registry links high-quality next-generation cancer genomic sequencing data with clinical outcomes obtained during routine medical practice from nearly every patient with cancer treated at 19 of the leading oncology institutions in the world, including individuals with rare cancers. The AACR Project GENIE therefore potentially provides the statistical power required to inform clinical decision making and to foster novel clinical and translational research.

The leverage possessed by professional societies could also be used to aid in setting standards or in establishing commonplace norms for data sharing within the immunotherapy space. For example, societies can allocate resources to convene workshops and establish committees of experts that can develop standard operating procedures, set standards for research and clinical practice within the field, and publish white papers advocating for particular research priorities or policy positions. For example, the SITC Biomarkers Committee has held workshops and generated outputs aimed at providing a platform for representatives to discuss the advantages and challenges of data sharing from federal, academic, and industrial perspectives.[50] The Biomarkers Committee has also drawn attention to the importance of standardized and validated assays, a crucial component of successful data sharing.[51–53] Furthermore, the society's open-access journal, *JITC*, has also been on the leading edge of important developments in the field since its inception through the publication of original research within its dedicated Immunotherapy Biomarkers section.

Societies may also provide funding for grants or programs that enable or enhance data sharing, both in the clinical and basic and translational research settings. These funding initiatives can play a vital role in unifying ongoing projects and preventing duplication of efforts. One such example was the establishment of a concept called TimIOs that would function as an honest broker for data set amalgamation between biotech companies that was established through SITC's Sparkathon Program.[50] The eventual goal of the project is to identify biomarkers that differentiate between patients with tumors that are likely to respond strongly or weakly to anti-PD-1 therapies.

A more ambitious, but potentially more rewarding proposal would be for societies to initiate or promote the development of an immunotherapy outcomes registry. The proposed registry network of IO data, contributed on a volunteer basis, could be initiated in principle by leveraging the political capital of medical professional societies through policy and advocacy efforts toward a larger national or international initiative. An example that could serve as a model registry program for society-led efforts is

the FNIH PACT, which is engaged in efforts on behalf of several academic institutions, biotech and pharmaceutical companies, and the federal government and is involved in standardizing immunotherapy assays with the Cancer Immune Monitoring and Analysis Centers and in housing this data from privately or NIH-funded studies through a cloud based, central data repository as part of the Cancer Immunologic Data Commons Network. However, the FNIH PACT effort does not include a larger mechanism to globally collect biomarker and outcomes data for those oncology patients treated with immuno-therapeutics in 'real-world' settings,[50] which is an area of need that societies could potentially address with their own registry programs.

Professional society initiatives aimed at clinical data sharing have proven successful. In one of the most prominent examples, the world of hematopoietic stem cell transplantation offers guidance. In 1970, the International Bone Marrow Transplant Registry (IBMTR) was founded as a division of the American College of Surgeons/National Institutes of Health Organ Transplant Registry.[54] By 1976, with the dissolution of the American College of Surgeons Organ Transplant Registry, the IBMTR became a standalone agency based in Milwaukee, Wisconsin and supported by the US Department of Health, Education and Welfare. In the years since that time the IBMTR merged with the National Bone Marrow Donor Registry in 2004 and formed one comprehensive program, the Center for International Blood & Marrow Transplant Research, which has continued to accrue invaluable data related to hematopoietic stem cell transplantation.[55] These data have played a pivotal role in shaping the development of hematopoietic stem cell transplants.

Another example of a federal-level program that could serve as an effective model for future society programs is The Cancer Genome Atlas (TCGA), a successful data sharing initiative that has generated and compiled comprehensive genomic, transcriptomic, epigenomic, and proteomic data from bulk tumor samples across 33 human malignancies.[56 57] Through this approach and collaborative effort, six immune subtypes of solid tumors of prognostic significance have been identified.[56] All data and results are intended to serve as a resource for future studies and can be freely accessed at the National Cancer Institute Genomic Data Commons, which houses data not only from TCGA but from several other cancer genomic projects, and through the Cancer Research Institute iAtlas portal, which provides interactive visualization of immune response in tumors. An in-depth atlas of the immune microenvironment of human tumors has been developed in clear cell renal cell carcinoma and early lung cancer, representing an invaluable tool for the rational design of targeted therapies as well as combination immunotherapy clinical trials.[58 59] Additionally, the Human Protein Atlas is another example of successful data sharing. The program was initiated in 2003 in Sweden and aims to map all the human proteins in cells, tissues and organs using integrative-omics technologies. The online platform provides access to the distribution of the proteins across tissues, organs, the subcellular localization of proteins, and potential prognostic impacts of protein expression on cancer patient survival.

CONCLUSIONS

In summary, IO data sharing is an idea that is not only beneficial in concept, but one that promises to be a powerful tool in combating cancer once made routine. Joint efforts from diverse stakeholders described in this paper are required to champion, adopt, and foster a culture of data sharing with realistic goals. Data sharing also requires entities that can provide management and oversight to ensure the protection of patient privacy, IP for funding organizations, and primary publication rights. These entities should also work to ensure the quality, integrity, and completeness of shared information, and that protocols governing data access and reuse are ethical, fair, and transparent. Data sharing will help break down silos that can inadvertently lead to the duplication of work and wasted research dollars. It is not only a possibility, but only a matter of time until the cycle of data silo-ing will be broken. In the future, clinical study data will not be boxed away after the initial study-focused analysis, but reused to accelerate the pace of discovery in the field. As we take stock of those stakeholders who hold the potential to shift our culture of competition into one of collaboration, we see the role that professional societies such as SITC have played, and can continue to act as champions for such change.

Worldwide, research policy makers support the accessibility of research data. This can be seen in the US with efforts by the NIH and in Europe, with the European Union's Horizon 2020 Program.[60–63] In order to develop consequential policies for data sharing, policy-makers need to understand and address the perspectives of all involved parties. We hope that the framework that we present in this paper will encourage a better understanding of the prevailing issues and provide insights into the underlying dynamics of academic and Pharma data sharing.

Author affiliations
[1]John van Geest Cancer Research Centre, Nottingham Trent University, Nottingham, UK
[2]Centre for Health, Ageing and Understanding Disease (CHAUD), Nottingham Trent University, Nottingham, UK
[3]Roche Pharmaceutical Research and Early Development Oncology, Roche Innovation Center, Penzberg, Germany
[4]Department of Medicine, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA
[5]Roche Pharmaceutical Research and Early Development Oncology, Roche Innovation Center, Basel, Switzerland
[6]Sage Bionetworks, Seattle, Washington, USA
[7]Roche Pharmaceutical Research and Early Development Oncology, Roche Innovation Center, Zurich, Switzerland
[8]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts, USA
[9]Department of Medicine, Division of Hematology/ Oncology, University of Pittsburgh School of Medicine and Melanoma Center at UPMC Hillman Cancer Center, Pittsburgh, Pennsylvania, USA
[10]National Institute of Oncology, Budapest, Hungary

[11]Department Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden
[12]CancerProbe Pty Ltd, Prahran, Victoria, Australia
[13]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada
[14]Department of Investigational Cancer Therapeutics, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA
[15]Institute for Systems Biology, Seattle, Washington, USA
[16]Department of Medicine, Division of Hematology/Oncology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[17]Allogene Therapeutics, South San Francisco, California, USA
[18]Parker Institute for Cancer Immunotherapy, San Francisco, California, USA
[19]Biolojic Design Inc, Cambridge, Massachusetts, USA
[20]ESSA Pharma Inc, South San Francisco, California, USA

ORCID iD
Aung Naing http://orcid.org/0000-0002-4803-8513

REFERENCES
1 Haslam A, Prasad V. Estimation of the percentage of US patients with cancer who are eligible for and respond to checkpoint inhibitor immunotherapy drugs. *JAMA Netw Open* 2019;2:e192535–e.
2 Naing A. Being realistic and optimistic in curing cancer. *J Immunother Precis Oncol* 2018;1:53.
3 Nakamura Y. Biomarkers for immune checkpoint inhibitor-mediated tumor response and adverse events. *Front Med* 2019;6:119.
4 Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer* 2019;19:133–50.
5 Auslander N, Zhang G, Lee JS, *et al*. Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *Nat Med* 2018;24:1545–9.
6 Heo YJ, Kang SY, Kim ST, *et al*. Combined biomarker for prediction of response to an immune checkpoint inhibitor in metastatic gastric cancer. *Precis Future Med* 2019;3:165–75.
7 Nassar AH, Mouw KW, Jegede O, *et al*. A model combining clinical and genomic factors to predict response to PD-1/PD-L1 blockade in advanced urothelial carcinoma. *Br J Cancer* 2020;122:555–63.
8 Antman E. Data sharing in research: benefits and risks for clinicians. *BMJ* 2014;348:g237.
9 Bull S, Cheah PY, Denny S, *et al*. Best practices for ethical sharing of individual-level health research data from low- and middle-income settings. *J Empir Res Hum Res Ethics* 2015;10:302–13.
10 Bull S, Roberts N, Parker M. Views of ethical best practices in sharing individual-level data from medical and public health research: a systematic scoping review. *J Empir Res Hum Res Ethics* 2015;10:225–38.
11 Mazor KM, Richards A, Gallagher M, *et al*. Stakeholders' views on data sharing in multicenter studies. *J Comp Eff Res* 2017;6:537–47.
12 van Panhuis WG, Paul P, Emerson C, *et al*. A systematic review of barriers to data sharing in public health. *BMC Public Health* 2014;14:1144.
13 Kuchinke W, Krauth C, Bergmann R, *et al*. Legal assessment tool (LAT): an interactive tool to address privacy and data protection issues for data sharing. *BMC Med Inform Decis Mak* 2016;16:81.
14 Hajduk GK, Jamieson NE, Baker BL, *et al*. It is not enough that we require data to be shared; we have to make sharing easy, feasible and accessible too! *BMJ Glob Health* 2019;4:e001550-e:e001550.
15 Lu R, Zhao X, Li J, *et al*. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020;395:565–74.
16 Nextstrain. All SARS-CoV-2 situation reports. Available: https://nextstrain.org/ncov-sit-reps
17 Global Initiative on Sharing all Influenza Data. Genomic epidemiology of hCoV-19. Available: https://www.gisaid.org/epiflu-applications/next-hcov-19-app/
18 Johns Hopkins University. Coronavirus resource center. Available: https://coronavirus.jhu.edu/
19 Worldometer. COVID-19 coronavirus pandemic. Available: https://www.worldometers.info/coronavirus/
20 Healthmap. COVID-19. Available: https://healthmap.org/covid-19/
21 Our World in Data. Coronavirus pandemic (COVID-19). Available: https://ourworldindata.org/coronavirus
22 World Health Organization. Coronavirus disease (COVID-19) pandemic. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019
23 World Health Organization. COVID-19 global literature on coronavirus disease. Available: https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/
24 Kuderer NM, Choueiri TK, Shah DP, *et al*. Clinical impact of COVID-19 on patients with cancer (CCC19): a cohort study. *Lancet* 2020;395:1907–18.
25 Warner JL, Rubinstein S, Grivas P, *et al*. Clinical impact of COVID-19 on patients with cancer: data from the COVID-19 and cancer Consortium (CCC19). *JCO* 2020;38:LBA110.
26 The Society for Immunotherapy of Cancer. SITC Advocates for scientific journals to grant open access to COVID-19 publications. Available: https://www.sitcancer.org/research/covid-19-resources/open-access
27 Hoffmann M, Kleine-Weber H, Schroeder S, *et al*. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 2020;181:271–80.
28 Wrapp D, Wang N, Corbett KS, *et al*. Cryo-Em structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020;367:1260–3.
29 Vabret N, Samstein R, Fernandez N, *et al*. Advancing scientific knowledge in times of pandemics. *Nat Rev Immunol* 2020;20:338.
30 Human immune monitoring center at Mount Sinai, Fernandez N. reviewing COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv. Available: https://observablehq.com/@ismms-himc/covid-19-sars-cov-2-preprints-from-medrxiv-and-biorxiv
31 Mascalzoni D, Bentzen HB, Budin-Ljøsne I, *et al*. Are requirements to deposit data in research Repositories compatible with the European

Union's General data protection regulation? *Ann Intern Med* 2019;170:332–4.

32 U.S. Food and Drug Administration. Institutional review boards frequently asked questions, 1998. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/institutional-review-boards-frequently-asked-questions

33 European Patients' Academy. Research ethics committees (rec). Available: https://www.eupati.eu/clinical-development-and-trials/research-ethics-committees-rec/

34 Gelling L. Role of the research ethics Committee. *Nurse Educ Today* 1999;19:564–9.

35 Tene O, Polonetsky J. Big data for all: privacy and user control in the age of analytics. *Northwest J Technol Intellect Prop* 2012;11.

36 Politou E, Alepis E, Patsakis C. Forgetting personal data and revoking consent under the GDPR: challenges and proposed solutions. *J Cybersecur* 2018;4.

37 Kaye J. The tension between data sharing and the protection of privacy in genomics research. *Annu Rev Genomics Hum Genet* 2012;13:415–31.

38 Metcalf J, Keller E, Boyd D. Perspectives on big data, ethics, and society, 2016. Available: https://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/

39 Lo B. Sharing clinical trial data: maximizing benefits, minimizing risk. *JAMA* 2015;313:793–4.

40 Harris L. Understanding corporate data sharing decisions: practices, challenges, and opportunities for sharing corporate data with researchers, 2017. Available: https://fpf.org/2017/11/14/understanding-corporate-data-sharing-decisions-practices-challenges-and-opportunities-for-sharing-corporate-data-with-researchers/

41 Wouters P, Haak W E, *et al*. Open Data- the researcher perspective 2018.

42 Weintraub PG. The importance of publishing negative results. *J Insect Sci* 2016;16:109.

43 Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A* 2005;102:16569–72.

44 Kiley R, Peatfield T, Hansen J, *et al*. Data Sharing from Clinical Trials - A Research Funder's Perspective. *N Engl J Med* 2017;377:1990–2.

45 Resnik DB, Morales M, Landrum R, *et al*. Effect of impact factor and discipline on Journal data sharing policies. *Account Res* 2019;26:139–56.

46 Vasilevsky NA, Minnier J, Haendel MA, *et al*. Reproducible and reusable research: are Journal data sharing policies meeting the mark? *PeerJ* 2017;5:e3208-e.

47 Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al*. The fair guiding principles for scientific data management and stewardship. *Sci Data* 2016;3:160018.

48 ASH Research Collaborative. Ash rC data hub. Available: https://www.ashresearchcollaborative.org/s/data-hub

49 AACR. Aacr project genie: Powering precision medicine. Available: https://www.aacr.org/professionals/research/aacr-project-genie/

50 Butterfield LH, Disis ML, Fox BA, *et al*. SITC 2018 workshop report: Immuno-Oncology biomarkers: state of the art. *J Immunother Cancer* 2018;6:138.

51 Gulley JL, Berzofsky JA, Butler MO, *et al*. Immunotherapy biomarkers 2016: overcoming the barriers. *J Immunother Cancer* 2017;5:29.

52 Masucci GV, Cesano A, Hawtin R, *et al*. Validation of biomarkers to predict response to immunotherapy in cancer: Volume I - pre-analytical and analytical validation. *J Immunother Cancer* 2016;4:76.

53 Dobbin KK, Cesano A, Alvarez J, *et al*. Validation of biomarkers to predict response to immunotherapy in cancer: Volume II - clinical validation and regulatory considerations. *J Immunother Cancer* 2016;4:77.

54 Goldman JM, Horowitz MM. The International bone marrow transplant registry. *Int J Hematol* 2002;76 Suppl 1:393–7.

55 Bortin MM, Rimm AA. *Factors influencing success and failure of human marrow transplantation: a review from the International bone marrow transplant registry*. New York, NY: Springer New York, 1979.

56 Thorsson V, Gibbs DL, Brown SD, *et al*. The immune landscape of cancer. *Immunity* 2018;48:e14:812–30.

57 National Cancer Institute. Tcga research network publications. Available: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/publications

58 Chevrier S, Levine JH, Zanotelli VRT, *et al*. An immune atlas of clear cell renal cell carcinoma. *Cell* 2017;169:e18:736–49.

59 Lavin Y, Kobayashi S, Leader A, *et al*. Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell* 2017;169:e17:750–65.

60 National Institutes of Health. Nih data sharing policy and implementation guidance, 2003. Available: https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

61 National Institutes of Health. Nih data sharing policies, 2019. Available: https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html

62 European Commission. Horizon 2020, 2020. Available: https://ec.europa.eu/programmes/horizon2020/en

63 European Commission, Horizon. H2020 Online Manual: Open access & Data management. Available: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm