Effy Vayena[1] *& Barbara Prainsack*[2]

[1]*University of Zurich, Institute of Biomedical Ethics, Zurich, Switzerland.* [2]*King's College London, Department of Social Science, Health and Medicine, London, UK.*
e-mail: vayena@ethik.uzh.ch

1. Wright, C.F. & Gregory-Jones, S. *Genet. Med.* **12**, 594 (2010).
2. Wright, C. & MacArthur, D.G. Direct to consumer genetic testing. in *Molecular Genetics and Personalized Medicine. Molecular and Translational Medicine* (eds. Best, D.H. & Swensen, J.J.) 215 (Springer, 2012).
3. European Science Foundation (ESF). *Personalised Medicine for the European Citizen.* A Forward Look Report (ESF, 2012).
4. Topol, E. The Creative Destruction of Medicine: How the Digital Revolution Will Create Better Health Care (Basic Books, 2012).
5. Kaphingst, K.A. *et al. Genet. Med.* **14**, 681–687 (2012).
6. Bloss, C. *et al. N. Engl. J. Med.* **364**, 524–534 (2011).
7. Anonymous. *Lancet* 379, 386 (2012).
8. Borry, P. *et al. Eur. J. Hum. Genet.* **20**, 715–721 (2012).
9. European Academies Science Advisory Council & the Federation of European Academies of Medicine. Direct-to-Consumer Genetic Testing for Health-Related Purposes in the European Union: The View from EASAC and FEAM (EASAC policy report 18) (EASAC, 2012). <http://www.easac.eu/home/reports-and-statements/detail-view/article/direct-to-co.html>

The International Serious Adverse Events Consortium's data sharing model

To the Editor:

Although the exchange of data among scientists has long been considered a norm of scientific practice[1], it was not until the mid-1990s that formalized structures for sharing scientific data before publication began to emerge. Most notably, in 1996 the leaders of the Human Genome Project adopted the 'Bermuda Principles'[2], which established for the first time that genomic sequence data be released to publicly accessible databases almost immediately after their generation and that intellectual property (IP) protection of such data be discouraged. The Bermuda Principles continue to shape data release practices today, both in the genomics research community and other fields[3,4]. Among the many genomics research projects adopting rapid prepublication data release requirements have been the International HapMap Project[5], the International Cancer Genome Consortium[6], the Malaria Genomic Epidemiology Network (MalariaGEN)[7], the Genetic Association Information Network (GAIN)[8] and the eMERGE Consortium[9]. Each of these projects was funded, to a large degree, by governmental and charitable sources. Though the data-sharing requirements and practices of such 'public' projects have been well-documented, data sharing in the private sector has received less attention. Nevertheless, the existence and study of private sector data-sharing initiatives is important, as the motivations leading private enterprises to share data openly are likely to be different from the motivations of public agencies and charitable foundations, and the data generated by private sector efforts may be of greater interest to other private sector researchers than publicly generated data. Two early private sector initiatives, the Merck Gene Index[10] and the SNP Consortium[11], have been widely cited as exemplars of private sector data sharing, but it has been a decade since these projects completed their work.

Here, we describe the data release and IP policies of the International Serious Adverse Events Consortium (iSAEC), a biomedical research consortium led and funded primarily by private pharmaceutical firms together with the Wellcome Trust (London). The data-sharing model adopted by iSAEC embodies many of the characteristics of the Bermuda Principles and later public genomics projects. We envision it will provide a useful reference for other public-private consortiums seeking to facilitate precompetitive research.

Originally founded in 2007 as a US tax-exempt organization, iSAEC today includes nine major US, European and Japanese pharmaceutical manufacturers, two information technology providers, a major US hospital network and the Wellcome Trust. Members are either voting or non-voting, depending on different membership criteria (that is, financial or nonfinancial contribution). The consortium's board of directors consists of one representative of each voting member, plus one of us (A.L.H.) as its independent chairman and CEO.

The mission of the iSAEC is to coordinate and fund research on the identification of DNA markers that confer a risk on individuals for drug-induced serious adverse events. The iSAEC engages academic collaborators and networks around the world to collect DNA samples and associated phenotypic data, and then to conduct genome-wide association studies (GWAS), targeted sequencing and statistical analyses to identify potential markers and associations of interest, including drug-induced liver injury[12,13] and serious skin rash[14].

Because drug-induced serious adverse events for marketed substances are rare and occurrences are geographically scattered and symptoms are often underdiagnosed or misdiagnosed, the iSAEC faces numerous challenges in assembling properly sized sample collections to conduct statistically meaningful analyses. Fostering collaboration with academic and clinical networks around the world has thus become a critical element of iSAEC's research program (**Supplementary Table 1**). To date, the iSAEC has established relationships with more than 50 academic and clinical groups and has formed two dedicated clinical networks (the International DILI (drug-induced liver injury) Consortium and the International Consortium on Drug Hypersensitivity) to accelerate the collection of suitable cases and related DNA samples. It has also supported phenotype standardization efforts[15–17] to promote the rational and consistent identification of case subjects across multiple collection sites. Given the importance of iSAEC's research to overall drug safety, the US Food and Drug Administration (FDA) has been an active participant in the consortium's programs, though neither FDA nor any other governmental entity has provided financial support for iSAEC's activities.

Although the members of the iSAEC are primarily commercial entities that individually hold important patent assets, the express policy of the iSAEC is to make its research results available to the public free of any patent encumbrances. By removing IP protection constraints, this policy aims to provide both members and nonmembers with unfettered access to valuable data and to promote biomedical research in drug-induced serious adverse events. This public commitment of IP serves as a cornerstone of iSAEC's charitable tax-exempt status and also alleviates concerns regarding potential

© 2013 Nature America, Inc. All rights reserved.

antitrust challenges to this coordinated research activity.

To achieve the desired end, iSAEC's members, as well as its academic and commercial collaborators, are contractually prohibited from seeking patent protection for associations or other discoveries made with iSAEC support. This prohibition does not, however, limit a member's or a collaborator's ability to seek to patent downstream discoveries enabled by the results developed through iSAEC.

To limit the ability of third parties to patent the genetic associations and other results identified by the iSAEC, the consortium has adopted a 'protective' patent strategy modeled on that of the SNP Consortium. Under this approach, the iSAEC files US patent applications claiming various DNA markers that have been identified through its studies, with the intention that these applications be abandoned after publication. This approach ensures that the relevant genetic markers are included in the US Patent and Trademark Office's 'prior art' database, which is used by patent examiners to assess the originality of claimed inventions, but that no patents will ever issue based on these applications. The patent-defeating effect of such patent filings extends to any other country that is a treaty partner of the United States.

GWAS, sequence and other data generated by iSAEC studies are managed by the iSAEC Data Analysis and Coordination Center (DACC) at Columbia University in New York. Data analysis is coordinated through iSAEC's Scientific Management Committee (SMC), which consists of investigators representing iSAEC members, consultants and selected academic collaborators. The SMC is responsible for creating the data analysis plans for all iSAEC studies, whereas the DACC is responsible for executing the plans, often in conjunction with non-DACC academic collaborators.

Like the publicly funded genomic science projects discussed above, the iSAEC releases a large amount of data to the public. Only data that have been de-identified and are not linked to any individual patient are made publicly accessible. Data release is accomplished through the iSAEC data portal (https://dataportal.saeconsortium.org/), a controlled-access database developed and maintained by the DACC. Data are made available for public access no later than 12 months after they are generated, and often sooner. Before data are publically

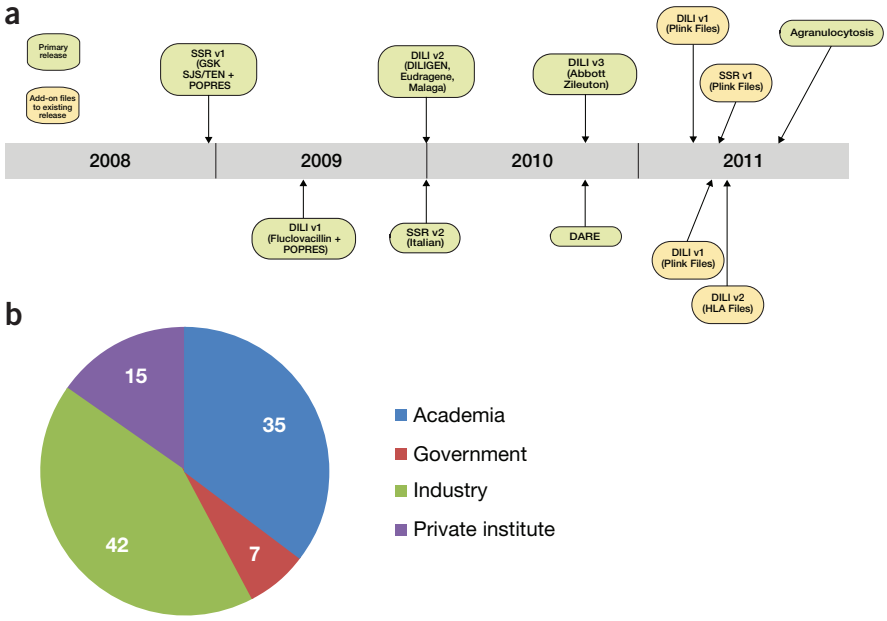

Figure 1 iSAEC data release program. (**a**) Profile and timeline of iSAEC public data releases (showing data sets based on studies of drug-induced liver injury (DILI), serious skin rash (SSR), Stevens-Johnson Syndrome/toxic epidermal necrolysis (SJS/TEN) and agranulocytosis. (**b**) iSAEC public data requests by sector (percentage).

released, no investigator accessing iSAEC data may make use of such data for any purpose other than the advancement of iSAEC projects (**Fig. 1a**).

The iSAEC's sixth and most recent data release (not shown) occurred in March 2012 and includes anonymous clinical and genotyping data on 360 individuals exposed to the anti-HIV drug nevirapine (Viramune), which is known to cause skin hypersensitivity reactions including Stevens-Johnson Syndrome and toxic epidermal necrolysis.

Any researcher may apply for access to the iSAEC's data through the DACC, which requires the applicant to disclose its institutional affiliation and specific research purpose. As of October 2011, the DACC had received ~100 separate requests for access to iSAEC data. All requests that stated a research purpose and identified a legitimate institutional affiliation were granted.

Approximately 42% of iSAEC data access requests are received from private industry (more than any other sector; **Fig. 1b**). This figure contrasts with the much lower rate of industry access to public databases, such as dbGaP, which has been reported at ~8% (as of August 2010)[18]. This finding demonstrates that private sector users are interested in, and make efforts to access, publicly accessible data on drug safety biomarkers, which is clearly useful and relevant to their internal research programs.

All recipients of iSAEC data must enter into a data use agreement pursuant to which they agree to use these data (i) solely for the purposes stated in the data access application, (ii) in accordance with any specific restrictions identified in the data file, (iii) in a manner that does not seek to identify any individual data subject, (iv) in a manner that is reasonably secure, and (v) in compliance with all applicable laws, rules, regulations and orders.

Data users must also agree not to publish any results based on the analysis of the iSAEC data until the publication embargo date associated with the relevant data set on the iSAEC data portal. This embargo date is typically set in the range of 9–12 months from the initial public release of the data and is intended to afford the consortium time to conclude its own data analyses and make protective patent filings, as described above. In addition, data users must agree not to seek patents claiming any DNA markers or associations disclosed in, or derived from, the iSAEC data, or that would otherwise block access to, or use of, such data. To our knowledge, there have been no violations of these policies by users of iSAEC data.

Although the iSAEC currently releases data through its own publicly accessible data portal, it is also in the process of working with the US National Institutes of Health (NIH) to make these data available by means of a dedicated drug safety genomic database after the expiration of applicable restriction periods.

© 2013 Nature America, Inc. All rights reserved.

Rapid, prepublication data sharing is viewed by many as an essential component of contemporary genomic research[19]. Accordingly, substantial attention has been given in the literature to data-sharing models employed by NIH-administered databases, such as GenBank and dbGaP, and in connection with publicly funded programs, such as GAIN and eMERGE. Although these resources have contributed greatly to public knowledge, they do not reveal the entire story. As described above, an increasing amount of data are being released to the public on a voluntary basis by privately funded collaborations, such as iSAEC.

There are numerous possible motivations that could lead for-profit enterprises to adopt this public-spirited approach. These range from corporate philanthropy and public relations to facilitating collaboration with academic, clinical and governmental research groups. Whatever the motivations are, however, the fact that industry-based collaborations such as iSAEC continue to contribute to public knowledge in critical research areas should be recognized and encouraged by policy makers and the broader scientific community.

The open data sharing and release policies adopted by iSAEC, as well as the consortium's adherence to a strict nonpatenting approach, have facilitated collaborations with academic and clinical networks across the world. Indeed, iSAEC's adherence to such policies has eliminated the need for the protracted and often difficult negotiations over patent rights and data access that often stymie collaboration between the private sector and nonprofit research organizations.

Thus, iSAEC's open data model is pivotal to maximizing the scientific value of its research into the genetic basis of drug-induced serious adverse events. Timely public release of the consortium's data offers biomedical researchers and informatics scientists the opportunity to pursue additional scientific findings by applying novel statistical methods and/or by integrating additional data sets. It is hoped that this approach will accelerate the development of scientific understanding and validation far beyond the prevalent traditional biomedical paradigm of publication without release of supporting primary data sets. This approach is in the best interests of science and society at large.

Note: Supplementary information is available at http://www.nature.com/doifinder/10.1038/nbt.2470.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available at http://www.nature.com/doifinder/10.1038/nbt.2470.

Jorge L Contreras[1], Aris Floratos[2] & Arthur L Holden[3]

[1]American University-Washington College of Law, Washington, DC, USA. [2]Columbia University, New York, New York, USA. [3]International SAE Consortium Ltd., Chicago, Illinois, USA.
e-mail: contreras@wcl.american.edu

1. Merton, R.K.The normative structure of science. in *The Sociology of Science* (ed. Storer, N.W.) 267–278 (Univ. Chicago Press, Chicago, 1973)
2. Human Genome Organisation (HUGO). Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing, 25–28 February 1996; Bermuda (HUGO, Geneva, 1996). <http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml#1>
3. Contreras, J.L. *Science* **329**, 393–394 (2010).
4. Kaye, J. *et al. Nat. Rev. Genet.* **10**, 331–335 (2009).
5. The International HapMap Consortium. *Nature* **426**, 789–796 (2003).
6. Zhang, J. *et al. Database (Oxford)* published online, doi: 10.1093/database/bar026 (16 September 2011)
7. Chokshi, D.A., Parker, M. & Kwiatkowski, D.P. Data Sharing and Intellectual Property in a Genomic Epidemiology Network: Policies for Large-Scale Research Collaboration. Bulletin of the World Health Organisation (WHO, Geneva, May 2006).
8. The GAIN Collaborative Research Group. *Nat. Genet.* **39**, 1045–1051 (2007).
9. McGuire, A.L. *et al. Genome Res.* **21**, 1001–1007 (2011).
10. Contreras, J.L. *Minn. J. Law Sci. & Technol.* **12**, 61–125 (2011).
11. Holden, A.L. *Biotechniques* **32**, 22–26 (2002).
12. Pirmohamed, M. *et al. Clin. Pharmacol. Ther.* **89**, 784–785 (2011).
13. Pirmohamed, M. *et al. Clin. Pharmacol. Ther.* **89**, 896–901 (2011).
14. Aithal, G.P. *et al. Clin. Pharmacol. Ther.* **89**, 806–815 (2011).
15. Daly, A.K. *et al. Nat. Genet.* **41**, 816 (2009).
16. Lucena, M.I. *et al. Gastroenterology* **141**, 338–347 (2011).
17. Shen, Y. *et al. Pharmacogenomics J.* **12**, 96–104 (2012).
18. Walker, L., Starks, H., West, K.M. & Fullerton, S.M. *Sci. Transl. Med.* **3**, 113cm34 (2011).
19. Green, E., Guyer, M. & NHGRI *Nature* **470**, 204–213 (2011).

Lack of evidence for existence of noncanonical RNA editing

To the Editor:

RNA editing, the post-transcriptional modification of genomic information, results in variation that can be observed in the transcriptome. In mammals, only two types of canonical RNA editing are known to exist: the pervasive adenosine-to-inosine (A-to-I; I is recognized as G) editing and the rare cytosine-to-uracil (C-to-U) editing. The recent advent of next-generation sequencing technologies allows comparative analysis of RNA sequencing (RNA-seq) and genome sequencing data from the same individual. A large number of putative noncanonical RNA editing sites have been reported[1]. However, the detection of RNA editing sites is often riddled with problems of a technical nature[2–5]. Our groups have recently developed refined methods and found that the vast majority of mismatches between genomic DNA and RNA are of the A-to-G type, indicative of A-to-I RNA editing[6,7]. Whereas one of our groups (at Stanford University) did not find evidence of noncanonical RNA editing events[7], the other (at BGI) reported successful validation for a number of them[6]. Therefore, the question remains: do the noncanonical RNA editing events exist? Here, our analysis unambiguously suggests that the previously identified noncanonical sites were derived mostly from mismapping reads to a highly similar genomic duplicate region.

Previously, eleven genomic loci (defined by PCR amplicons), each containing at least one non-A-to-G site, were successfully validated by PCR and Sanger sequencing (ref. 6, Supplementary Table 8). Often multiple sites of different types were present within each locus. We queried each of the 11 PCR amplicon sequences against the human reference genome using the global alignment tool BLAT[8], which has the ability to detect the correct location of spliced reads. To our surprise, each one of them had a second-best hit in the human genome. Mismatches between the original amplicons and the second-best hits were consistent with the previously suggested editing types at these positions (**Fig. 1a**). This suggests that for all 11 genomic loci, the observed non-A-to-G mismatches were a result of wrongly mapped RNA-seq reads that originated from highly similar regions, which were undistinguishable even by the primers used in the original validation.

Therefore, we designed novel primer pairs that exclusively amplify the second-best hit region in both genomic DNA (gDNA) and cDNA for four genomic loci with BLAT double hits (**Supplementary Table 1**). Using the previously designed and the new primers, we successfully obtained gDNA and cDNA products of the first- and second-best BLAT hit regions (**Supplementary Notes** and **Supplementary Table 2**). In all cases, the

© 2013 Nature America, Inc. All rights reserved.