Open Science Monitoring

Impact Case Study – Structural Genomics Consortium
Author information and acknowledgements

This study was delivered by RAND Europe, with the support of Deloitte and Digital Science & Research Solutions, including Altmetric.com and figshare. The project was commissioned by the European Commission Directorate-General for Research and Innovation and received funding under specific contract 2015/RTD/A6/SC/PP-02741-2015.\(^1\)

The team at RAND Europe included: Elta Smith (project leader), Salil Gunashekar (project manager), Sarah Parks, Catherine Lichten, Louise Lepeitit, Molly Morgan Jones, Catriona Manville and Calum MacLure. The team at Deloitte included: Valentina Cilli, Maruca De Ramón, Vilma Zotou, Benoit Vandresse and Sebastiaan van der Peijl.

Data sources

We would like to thank the following organisations for generously contributing their data for use on the monitor: 101innovations, Altmetric, arXiv, bioRxiv, Clarivate Analytics, F1000Research, figshare, GenBank, Helmholtz-Centre for Environmental Research – UFZ and German Centre for Integrative Biodiversity Research (iDiv), Nature Publishing Group and Palgrave Macmillan, OpenAIRE, PeerJ preprints, Publons, re3data.org, RePEc, SHERPA/RoMEO, SHERPA/JULIET, Taylor & Francis, and Wiley.

Expert panel

We are grateful to our expert panel members who provided helpful advice and guidance: Aletta Bonn (German Centre for Integrative Biodiversity Research (iDiv) and Helmholtz Center for Environmental Research (UFZ), Friedrich-Schiller-Universität Jena, Germany); Thomas Crouzier (School of Biotechnology, Royal Institute of Technology (KTH), Sweden); Ivo Grigorov (National Institute of Aquatic Resources, Technical University of Denmark); Peter Kraker (Postdoctoral Fellow, Know-Center Social Computing Group, Graz University of Technology, Austria); Eleni Malliou (National Documentation Center, Greece); and Paul Wouters (CWTS-Centre for Science and Technology Studies, Leiden University, The Netherlands).

Acknowledgements:

We would like to thank the following individuals for their contributions to the monitor: Susan Guthrie (RAND Europe), Gavin Cochrane (RAND Europe), Alexandra Pollitt (King’s College London), and Marion Bywater who acted as our quality assurance reviewers and provided thoughtful and timely comments on our analyses; Jessica Plumridge (RAND Europe) who designed the graphics presented in the online monitor; Monica Hertzman (RAND Corporation) who provided technical assistance in relation to the design of the website; Steven Wooding (RAND Europe) who provided advice and support throughout the study; Vincent Larivière (Observatoire des Sciences et des Technologies) for his insights and advice; Joanna Chataway (University of Sussex) for her advice and support; Rebecca Ioppolo (RAND Europe) who provided research assistant support; Jack Melling (RAND Europe) who helped in the dissemination activities related to the launch of the monitor; and the many individuals and organisations who participated in an online consultation we conducted as part of the development of the monitor.

For more information about this document or the Open Science Monitor, please contact:

Elta Smith
RAND Europe | Westbrook Centre, Milton Road, Cambridge CB4 1YG, United Kingdom
Telephone: +44 (1223) 353 329 | E-mail: eltas@rand.org

**Structural Genomics Consortium**

**Summary**

The Structural Genomics Consortium (SGC) was established in response to concerns about decreased productivity in drug discovery, which its founders attributed to aggressive patenting policies and the increasing complexity of the underpinning science at a time when ‘quick wins’ are exhausted. The SGC aims to foster a larger and more coordinated field of researchers in order to share the risk of exploring new areas of genomics research and to increase the speed of progress.

The SGC opted to forego patent claims and to provide open access to its outputs in order to enable their uptake by as many scientists as possible across as many fields as possible. By placing all newly solved protein structures in the Protein Data Bank and by sending samples to researchers on request, the SGC aims to remove the legal and financial barriers that prevent a more diverse range of actors from engaging in drug discovery work.

The SGC has solved the structures of more than 1,500 proteins that it has identified as having potential relevance to human medicine, and it is the fourth-largest contributor to the Protein Data Bank, providing around 10 per cent of new structures deposited each year. It has put 75 novel human kinase structures in the public domain, which is more than the entirety of academia and industry put together. In addition, between 2007 and 2014, the SGC distributed more than 1,200 expression clones to academic organisations and more than 200 to industry. Building on the characterisation of protein structures, the SGC has discovered and published more than 40 chemical compounds (probes) that enable modulation of proteins, as well recombinant antibodies and antibody-like reagents targeting more than 100 human proteins involved in epigenetic signalling. In terms of academic outputs, the SGC produced 452 peer-reviewed journal publications and 8 books in its first 10 years. These include 19 papers in *Nature* and its subsidiary journals.

A number of SGC projects have produced findings with the potential to catalyse advances in drug discovery. The generation of this type of new knowledge has been the principal impact of the SGC’s work. In a number of cases, the knowledge generated has also catalysed the development of new products. The SGC’s work has also had commercial and economic impacts. Although no monetary gain accrues to the SGC itself, values can be calculated for the products it has developed. A previous study valued the cost per protein structure solved by the SGC at $219,000, and the SGC values each chemical probe developed at over $2 million. The SGC’s research has also led to companies being established or increasing in value. Three spin-out companies have come out of SGC projects. Furthermore, the SGC produces substantial cost savings for funders of research. In 2010, Biohub estimated that the SGC’s work to establish the effectiveness of available antibodies will save over $750 m in research spending. Moreover, the SGC’s model avoids the cost of establishing material transfer agreements when sharing samples of chemical probes or establishing collaborations. For comparison, it is estimated that pharmaceutical companies spend hundreds of thousands of dollars on this type of expense for a single collaboration.

**Background**

The SGC uses an open science model to catalyse drug discovery by conducting pre-competitive research in less studied areas of structural biology and placing the results of that research in the public domain. It was established in 2004 in response to diminishing returns on pharmaceutical R&D investment in the preceding decades. This productivity crisis in drug discovery has been attributed to a combination of strategic patenting of drug discovery targets and the ever-increasing complexity of the scientific problems to be solved, both of which increase the cost of R&D and the risk to investors (Hu et al. 2007; Pammolli et al. 2011). The SGC seeks to address these problems by determining 3D protein structures of
biomedical importance, and in some cases using those as a basis to develop chemical and biological probes, through a collaborative approach based on open access to research data. It aims to de-risk this early-stage R&D work by allowing investors to share its cost through pooled funding, and to foster a larger and more coordinated field of researchers to build on this work by foregoing patent claims and providing open access to findings.

The work of the SGC is carried out primarily by researchers based at the University of Oxford in the UK, the University of Toronto in Canada, the Universidade Estadual de Campinas (Unicamp) in Brazil, the University of North Carolina at Chapel Hill in the USA, the Karolinska Institutet in Sweden and Johann Wolfgang Goethe Universität in Germany. It is funded by a combination of pharmaceutical companies, public bodies and non-profit organisations in the UK, Canada and Brazil. These partners, together with project-based financing from conventional research funders, have provided the SGC with around $400 m across three phases of funding. The role of open science in leveraging that funding is explored in detail below.

The SGC’s key achievements have been the rapid determination of previously unsolved or unpublished protein structures of medical relevance, and the development of chemical probes capable of modifying those proteins. In some cases these findings have enabled follow-on drug discovery work, including first-in-human trials of drugs targeting diseases for which no treatment previously existed.

**Methods and the role of open science**

The SGC’s goal is to foster a critical mass of scientific effort in previously neglected fields by allowing its findings to be accessed and taken forward by the largest possible number of scientists with the broadest possible range of expertise. Moreover, the SGC seeks not only to attract pharmaceutical companies to novel areas of structural biology by de-risking early-stage research, but also to enable the participation of more diverse organisations by removing legal and financial barriers to entry (Perkmann & Schildt 2015). This is to be achieved by depositing all newly solved protein structures in the Protein Data Bank (2016) without any intellectual property restrictions.

This openness has also enabled the SGC to secure investment. The policy of foregoing patents and publishing all findings helped to attract the public and philanthropic supporters who provided the first phase of funding for the SGC, which enabled it to demonstrate the value of its approach to pharmaceutical companies. In keeping with its collaborative working model, the SGC allows partner organisations to help shape its scientific agenda by specifying up to 200 target proteins in exchange for their financial backing. As a result, the SGC has been able to secure funding from nine of the world’s largest pharmaceutical companies (Euroscientist 2014).

The collaborative approach enabled by open science principles shapes the way the SGC sets its priorities and conducts research. By allowing partner organisations to contribute to its research agenda and share risk through pooled public-private funding, the SGC creates an incentive for partners to target novel proteins (Morgan Jones et al. 2014). Conventional approaches to research, influenced by the peer review process for academic grants and the commercial constraints of industry, have been shown to make both sectors reluctant to pursue novel targets (Institute of Medicine 2011). The impact of the SGC’s model in countering this risk aversion is illustrated by the example of protein kinases. There are more than 500 protein kinases in the human genome, but over 90 per cent of papers published in this area of genomics focus on just 10 per cent of kinases (or just over 50 kinases) (ibid). In
comparison, the SGC alone has solved the catalytic domain structure of 75 kinases (SGC 2016a).

Once priorities have been established, the SGC relies on its open approach to generate collaborations to carry out the research efficiently. Conventional genomics research is conducted by pharmaceutical companies working separately, behind closed doors. Among 18 leading pharmaceutical companies, this approach has been found to result in the duplication of 86 per cent of drug discovery targets across companies (Leeson & St. Gallay 2011). Moreover, research has identified a disconnect between the different types of expertise required to advance genomics research. For example, pharmaceutical companies interviewed for a previous study identified a lack of in-house epigenetic skills, as certain types of biology expertise are found mainly in academia (Institute of Medicine 2011; Morgan Jones et al. 2014). By removing proprietary barriers, the SGC facilitates collaborations that connect researchers across fields, sectors and countries to increase the breadth and depth of expertise concentrated on novel research problems (Morgan Jones et al. 2014).

**Outputs and findings**

The SGC has solved the structures of more than 1,500 proteins that they have identified as having potential relevance to human medicine (SGC 2016b). The 75 protein kinases characterised by the SGC represent around one third of known kinase structures (SGC 2016c). They include kinases of medical importance, such as LIMK1 (the overexpression of which is found in malignant melanoma, breast cancer and prostate cancer) (SGC 2011a) and ACVR2A (which has been found to be mutated in certain forms of colon cancer) (SGC 2011b). Building on the characterisation of protein structures, the SGC has discovered and published more than 40 chemical compounds (probes) that enable modulation of proteins and may therefore provide targets for drug discovery (SGC 2016b). The SGC has also produced recombinant antibodies and antibody-like reagents, which help to pinpoint the expression of proteins, targeting more than 100 human proteins involved in epigenetic signalling (ibid).

A high-profile example of the SGC’s findings comes from its work on the JQ1 probe. The SGC collaborated with the Dana-Farber Cancer Institute at the Harvard Medical School to develop a probe, JQ1, which was shown to inhibit the BRD4 protein (Filippakopoulos et al. 2010). BRD4, the structure of which had previously been solved by the SGC, was found to be associated with NUT midline carcinoma – a rare, incurable form of cancer. BRD4 was thus identified as a potential target for drug discovery.

The SGC has also produced findings which have advanced the understanding of fibrodysplasia ossificans progressiva (FOP) – another rare, incurable disease that causes fibrous tissue to turn to bone when damaged and affects one person in every two million. This work followed the identification of a mutation of ACVR1 – a kinase whose structure had been characterised by SGC - in FOP (Mohedas et al. 2013; SGC 2013). In partnership with charities working on FOP, the SGC collaborated with geneticists with the aim of developing new compounds targeting ACVR1. The characterisation of ACVR1 has led to the development of animal models of FOP, which have the potential to play a vital role in drug discovery work (Chakkalakal et al. 2012). Studies have also found that ACVR1 mutations are present in a rare type of childhood brain tumour (Taylor et al. 2014).

As well as facilitating collaboration in the conduct of research, the SGC seeks to engage a broad range of stakeholders in the uptake of its findings. With regard to conventional stakeholders in drug discovery, the SGC’s abovementioned efforts to de-risk novel areas of science have generally been welcomed by the pharmaceutical industry (ibid.). Representatives of the industry interviewed for a previous study emphasised the attraction of being able to nominate targets for the SGC to investigate and then build on the SGC’s publicly available findings in their company’s internal R&D work (Perkmann & Schildt 2015). However, a more mixed picture has been reported by representatives of small biotechnology
firms, representatives of which have stated that their business models are reliant on their ability to claim intellectual property rights, and that this presents a barrier to engagement with the SGC (Morgan Jones et al. 2014).

In academia, the absence of proprietary barriers has enabled researchers to access SGC outputs, thereby avoiding duplication of effort. An SGC representative interviewed for this case study gave the example of the Consortium sharing its findings with an academic group seeking to apply the AVCR1 work to brain tumours (Lee, pers. comm. 2016). The representative noted that, had the academic group not seen the SGC’s publications and been able to access its findings, the group would have had to duplicate the lengthy and costly process of developing chemical compounds and animal models. However, previous research on attitudes in academia has found that there is still some reluctance to engage with the SGC due to a trend among academic institutions to prioritise the protection of their intellectual property (ibid.).

The SGC’s approach has also enabled engagement with stakeholders outside of academia, including patient-focused non-profit organisations that have an interest in supporting work towards a treatment for a specific disease but may have been priced out of involvement in R&D under a proprietary model. For example, following the development of the JQ1 probe, the Leukaemia and Lymphoma Society raised $7.5 m – a relatively small amount in the context of drug discovery expenditure - to engage a contract research organisation, which produced a clinical candidate based on JQ1 (Lee 2015). The work on FOP had a similar effect, resulting in collaboration between the SGC and The Brain Tumour Charity to investigate the AVCR1 kinase’s association with a certain type of brain tumour (SGC 2016d). According to the SGC, the engagement of patient-focused organisations such as these has the added advantage of facilitating access to suitable participants for clinical trials. This engagement has also extended to the general public through such activities as public lectures and museum exhibits, according to a survey of SGC researchers (Morgan Jones et al. 2014).

In seeking to remove barriers to organisations engaging with its findings as outlined above, the SGC has increased the number and diversity of projects building on those findings. According to an SGC representative, within a month of the publication of the paper on JQ1, SGC had provided more than 100 samples to more than 100 different research groups. As of 2016, samples of JQ1 have been sent to more than 400 laboratories. An earlier probe was distributed to more than 250 groups worldwide (ibid.). The case of JQ1 also illustrates the tendency for SGC findings to be applied to multiple disease areas, as work building on JQ1 has produced results in leukaemia and multiple myeloma, as well as diseases outside of cancer, such as sepsis (Lee, pers. comm. 2016).³

The SGC’s open model has enabled the translation of its research findings into a high number of outputs. In terms of the number of protein structures placed in the public domain, the SGC is the fourth-largest contributor to the Protein Data Bank, providing around 10 per cent of new structures deposited each year (Morgan Jones et al. 2014). It has put 75 novel human kinase structures in the public domain (SGC 2016b), which is more than the entirety of the contribution of academia and industry together (Knapp & Sundström 2014). The SGC also provides researchers with expression clones on request, and between 2007 and 2014 it distributed more than 1,200 clones to academic organisations and more than 200 to industry (Morgan Jones et al. 2014).

In terms of academic outputs, the SGC produced 452 peer-reviewed journal publications and 8 books in its first 10 years (ibid.). These included 19 papers in *Nature* and its subsidiary journals. These publications were supported by other dissemination activities, with SGC scientists attending or presenting at more than 250 conferences between 2007 and 2011,

³ For examples of BET domain inhibitors in clinical development, see von Schaper (2016).
including 87 invited talks resulting directly from the researchers’ work with the SGC. Public-facing outputs are also part of the SGC’s dissemination work: in 2012, its researchers contributed 12 press articles and gave 7 interviews for TV and other media.

**Impacts**

The principal impact of the SGC has been in the generation of new knowledge. In a number of cases, the knowledge generated has catalysed the development of new products. Studies building on the SGC’s work on JQ1 are a notable example of this. GlaxoSmithKline started an internal proprietary programme based on JQ1, and in 2012, less than two years after the publication of the initial JQ1 paper, it began the first clinical trial using BET bromodomain chemical probes. In total, there are now 12 registered clinical trials targeting this type of bromodomain (Lee 2015). This increased product development activity is reflected in an increase in the number of patents relating to the bromodomain proteins targeted by JQ1, which has passed 100 (Arshad et al. 2016; Scott 2016). A survey of SGC researchers found that 47 per cent of respondents expected their research to lead to the development or trialling of a therapeutic pharmaceutical product in the future, while 18 per cent stated that their research had already led to products entering phase 1 or 2 clinical trials, and 24 per cent stated that their research led to pharmaceutical innovations/processes that are cited by patents or other intellectual property arising (Morgan Jones et al. 2014).

The SGC’s work has also led to economic impacts by enabling cost- and time-efficient research. In the survey cited above, 81 per cent of SGC researchers reported economic or commercial returns resulting from their work (ibid.). Although no monetary gain accrues to the SGC itself, values can be calculated for the products it has developed. A previous study valued the cost per protein structure solved by the SGC at $219,000 (ibid.),4 and the SGC values each chemical probe developed at over $2 million.5

The SGC’s research has also led to companies being established or increasing in value. Three spin-out companies have come out of SGC projects, the most notable being Tensha Therapeutics, which was set up by the SGC’s academic partner on the JQ1 work. The company, which aimed to build on the work with SGC on bromodomain inhibitors, attracted $15 million in initial investment and was sold within a year for $535 million (Johnston 2016; Tensha Therapeutics 2011). The other spin-out companies resulting from SGC work are Harbinger Biotech (started by SGC researchers) and 1DegreeBio (started by a former SGC employee). In addition to direct spin-outs, the SGC has had an impact on the ability of start-ups and small firms to carry out drug discovery work based on SGC findings. An SGC representative we interviewed asserted that the SGC’s non-proprietary model reduces the financial barriers to participation in drug discovery work (Lee, pers. comm. 2016).

Furthermore, the SGC produces substantial cost savings for funders of research. In 2010, Biohub estimated that the SGC’s work to establish the effectiveness of available antibodies will save over $750 million in research spending (Morgan Jones et al. 2014).6 The SGC’s model avoids the cost of establishing material transfer agreements when sharing samples of chemical probes or establishing collaborations (ibid.). For comparison, it is estimated that pharmaceutical companies spend hundreds of thousands of dollars on this type of expense for a single collaboration (Lee 2015). According to the SGC, the fact that the SGC has

---

4 Value converted from CAD to USD at the 23 September 2016 rate of 1 CAD:0.758836 USD. Source: http://www.xe.com

5 Value provided through correspondence between the research team and the SGC.

6 Value converted from CAD to USD at the 23 September 2016 rate of 1 CAD:0.758836 USD. Source: http://www.xe.com
collaborated with hundreds of laboratories and distributed thousands of samples of chemical probes means that considerable cost savings can be attributed to its open approach (ibid.).

At the sectoral level, the SGC’s work appears to have had an impact on perceptions of open access to research tools and results, particularly in the pharmaceutical industry. A representative of SGC partner Pfizer interviewed in Nature stated that JQ1 has resulted in a change of mindset at Pfizer, which has started to publish more information on related tools it has developed itself, in recognition of the potential for information sharing to aid scientific progress (Scott 2016). Despite this, representatives of the SGC concede that attitudes towards information sharing in the pharmaceutical industry are deeply ingrained, and that the SGC has so far only been able to reshape companies’ thinking to a limited extent (Morgan Jones et al. 2014). Moreover, impacts on policy have also been limited, according to SGC researchers interviewed for a previous study. Forty-seven per cent of respondents to the study stated that they were not aware of any policy impact, while 18 per cent stated that there had been no policy impact from their work. Although 24 per cent of respondents stated that their research had an impact on policy within industry and 18 per cent identified impacts on public policy, the vast majority (81 per cent) of respondents did not know whether the policy impact of their research would have been any different under a traditional, non-open science working model (ibid.). Finally, the majority of respondents to the same survey believed that, through the public engagement activities enabled by the model, the SGC’s work had improved public engagement with science.

**Sources**

**Interviews**

We gratefully acknowledge the following interviewee, who provided information for this case study: Dr Wen Hwa Lee of the Structural Genomics Consortium, September 2016.

**Bibliography**


---

7 A previous study by RAND Europe (Morgan Jones et al. 2014) found that a much larger study would be required to validate the monetary values assigned to SGC outputs.


———. 2011b. 'Human Activin Receptor Type-IIA (ACVR2A) Kinase Domain in Complex with Quinazolin.' As of 16 January 2017: http://www.thesgc.org/structures/3SOC.


———. 2016b. 'Key Achievements.' As of 16 January 2017: http://www.thesgc.org/about/key_achievements

———. 2016c. 'Progress in Protein Kinase Structural Biology.' As of 16 January 2017: http://www.thesgc.org/scientists/resources/kinases

