Phase II Competition - Finalist Application Form

Created: 10/24/2016
Last updated: 11/22/2016

In this form, the six finalist teams for the Open Science Prize are asked to describe the work they have done to develop their prototypes, and make the case for why they should be considered for the phase II Prize.

The information you submit on this form will be considered alongside the prototype you have developed in deciding which team will receive the Phase II Prize.

Please note that, unless otherwise stated, the information you submit on this form will be made available publicly via the Open Science Prize website (under a CC BY 4.0 license), so that it can be assessed as part of the public voting process. All fields, except the final box for additional information are mandatory.

Your application must be completed by 11:59pm Pacific Standard Time on 21 November 2016 . You may edit this form as many times as you like before the deadline.

If you have any questions about the Prize or the review process or if you would like to provide any further information that you would not wish to be made public, please contact David Carr (d.carr@wellcome.ac.uk) or Elizabeth Kittrie (elizabeth.kittrie@nih.gov). Any technical questions regarding this form or the web platform should be directed to (openscience@wellcome.ac.uk).

Page 1

Executive Summary

Please provide a brief Executive Summary of no more than 150 words for the public voting page on the Open Science Prize website. This should be suitable for an informed lay audience, and should briefly describe your prototype and why it should be considered for the phase II Prize. [150 words]

Approximately 350 million people worldwide have a rare disease, most of which are Mendelian conditions caused by mutations of a single gene. However, of ~8,000 known Mendelian conditions, the causal gene is known for only half. This is a major gap in knowledge that limits diagnosis, prognosis, and treatment of rare diseases.

A major obstacle to understanding the genetic basis of all rare diseases is insufficient sharing of genetic and health data. MyGene2 makes it easy for families with rare conditions, doctors, and

researchers to share data publicly and equitably. MyGene2 analyzes these data to spot discoveries, and summarizes discoveries and the health problems caused by each gene in a public report. Users can explore their own data or data from other users, be automatically notified of discoveries, and network with each other. Radically-open data sharing using MyGene2 will transform the study, diagnosis, and care of families with rare diseases.

Weblink for prototype

Please provide the public URL for your prototype tool or service (this will be viewed by the public for purposes of public voting):

https://www.mygene2.org

Your Prototype

Purpose and need

Please provide a brief summary description of the purpose of the prototype you have developed and the key challenges or needs it is seeking to address [200 words]

The purpose of MyGene2 is to:

1. enable families, doctors and researchers to publicly share health and genetic data effectively
2. help families with a rare disease find a diagnosis and doctors understand the health problems caused by each gene
3. accelerate the pace of gene discovery and publicly report discoveries in real-time
4. empower families to act as citizen-scientists and engage doctors as shared decision-makers

MyGene2 overcomes the following key challenges:

1. health information and genetic data needed to discover genes for rare diseases are conventionally not openly shared between doctors and researchers, let alone with families
2. gene discoveries and information on gene-disease relationships are made public slowly, or not at all
3. it is hard for families to effectively share health and genetic data that would accelerate the pace of diagnosis and discovery

4. tools for families to explore their own data are not widely available and typically complex to use
5. most families cannot effectively participate in gene discovery or research gene-disease relationships as citizen-scientists because they lack access to both the necessary genetic and health data as well the tools for analysis

Please summarise the work you have taken forward to develop your prototype since you were awarded the Phase I Prize in April 2016.

i. Progress

The key milestones achieved and the extent to which the goals and challenges you set out to address in your original application were delivered [400 words]

We delivered on the three key milestones to which we committed and six additional features that overcome the key challenges identified in our original application:

1. Join the MatchMaker Exchange (MME), a network of restricted access databases designed to share health and genetic data among doctors and researchers to promote discovery

MyGene2 joined the MME. All novel candidate genes shared by doctors and researchers in MyGene2 are now automatically shared with MME. To date, families have been forbidden to participate in MME. MyGene2 successfully lobbied MME to allow data submission from families and development to do so is underway.

2. Enable families to browse their own exome or genome sequence data

MyGene2 integrated gene.iobio, a tool that enables families to easily browse their sequence data. The gene.iobio instance was customized for MyGene2 to ensure a family-friendly experience.

3. Develop an automated pipeline for identifying candidate gene matches and publicly share all candidate match reports

MyGene2 developed an automated pipeline for identifying candidate gene matches and created a mechanism to publish public “match reports” in real-time. These reports highlight candidate gene discoveries and summarize the health problems found in each family with a mutation in a

particular gene. The latter allows for doctors to better understand the relationship between genes and health problems. Match reports are available to both families and doctors at point of care, which facilitates shared decision-making about diagnosis and treatment.

We delivered six additional features including:

1. families can assign a doctor or genetic counselor as a proxy to assist with entry of health information and genetic data, provide guidance about matches, and moderate communication with researchers and other doctors
2. families can apply directly for research exome sequencing by the University of Washington Center for Mendelian Genomics or for self-pay CLIA-certified research exome sequencing with return of a primary result report, “raw” sequence data, and candidate genes
3. improved search capability that enables profiles in MyGene2 to be searched by a clinical finding or name of a gene
4. families can indicate their membership(s) in different advocacy organizations and MyGene2 can be searched by membership
5. case-classification of profiles submitted by doctors and researchers ensures only novel information is shared with MME--this reduces the likelihood of making a “false match” with other MME nodes
6. MyGene2 became a member of the GA4GH Beacon Network

Finally, we made numerous modifications to improve the user experience.

ii. Team Contributions

The contributions of the team members to the development of the prototype [200 words]

Drs. Chong, Bamshad, and Groza have both individual and shared responsibilities for the development, management, marketing and governance of MyGene2. All three investigators meet weekly via video conference and ad hoc as needed to discuss overall project management, monitor progress, and troubleshoot bugs. Primary responsibilities for Drs. Chong and Bamshad include design of the interface, development of workflows, interaction with users and stakeholders (i.e. patient advocacy groups, genetic testing companies, doctors, genetic counselors, and researchers), and management of advisory groups. Dr. Groza is responsible for both the front- and back-end development, design and implementation of data matching algorithms and search

functions, search optimization, and integration of services and tools (e.g., gene.iobio, MME, Beacon Network).

MyGene2 has also added both a Scientific Advisory Committee (SAC) and a Family Advisory Committee (FAC) to provide guidance on strategic planning, development, and safety. The SAC consists of experts in genetics, genetic counseling, pediatric bioethics, consenting, data sharing, return of results, and regulatory issues. The FAC consists of leaders of various rare disease advocacy groups, parents of children with rare conditions, and active citizen scientists.

iii. Significant Achievements

Any significant achievements or key success metrics you wish to highlight - this might include, for example, numbers of users, key endorsements or engagements with users, new partnerships, external funding, and so forth [200 words]

Stakeholder engagement:

MyGene2 has engendered widespread support, both nationally and internationally, from a broad variety of stakeholder communities (families, genetic testing laboratories, doctors, genetic counselors, researchers).

Partnerships:

MyGene2 has established or has pending partnerships with multiple stakeholder groups. Among rare disease advocacy groups, established partners include Syndromes Without a Name (SWAN-USA), Rare and Undiagnosed Network (RUN), and RunMyDNA. Partnerships are pending with Citizens United for Research in Epilepsy (CURE) and Global Genes. MyGene2 established a partnership with the University of Washington Center for Precision Diagnostics to provide self-pay CLIA-certified research exome sequencing. In conjunction with the Monarch Initiative, partnerships with several medical journals to integrate “phenopackets” as means to directly push family and genetic data from case reports and case series into MyGene2 are pending. Finally, discussions are underway with several genetic testing laboratories to directly submit de-identified health information and genetic data from families to MyGene2.

Memberships:

MyGene2 has become a member of the Global Alliance for Genomics and Health (GA4GH), MME, and the Beacon Network.

Learning Points

Please briefly highlight any key learning points you took from the work that you undertook to develop your prototype [200 words max]

Key learning points we identified during Phase 1 include:

Families with a rare disease need encouragement and education to understand the extent to which sharing health information and genetic data can empower them to: find a diagnosis, facilitate gene discovery, and help doctors and other families understand the health problems caused by each gene.

Matching on a combination of variant type, inheritance pattern, phenotype, and gene is more precise than matching on candidate gene alone, but requires development and integration of additional tools to facilitate collection of this data as well as empirical testing of how to weight these data types in the matching algorithm itself.

Integration of MyGene2 into academic and industry workflows for evaluating families with rare disease, returning results of genetic testing to providers and families, and publishing health information and genetic data from families with rare disease using conventional approaches (e.g., publication in an academic journal) will need to take place in a series of transitional steps beginning with passive notification of families/customers about MyGene2, to moderated assistance for families using MyGene2, to developing APIs that enable families to initiate direct submission.

Case for Phase II Prize

Please make the case for why your prototype should be considered for the Phase II Open Science Prize against the following key criteria [100 words each]:

i. Impact

The current and potential future impact of the tool or service in terms of advancing research and generating health and societal benefit

Widespread adoption of MyGene2 will transform gene discovery and the study of genotype-phenotype relationships for rare diseases by accelerating the pace of discovery, increasing the total numbers of genes discovered for rare conditions, and making information on discoveries equitably available to all. This will fundamentally alter interactions between families and doctors as well as the relationships between families and researchers. It will enable rapid, precise genetic diagnosis of most families with rare diseases, accurate prediction of natural history and outcome, and development of more effective treatments, if not cures. Accordingly, the impact of MyGene2 will be deep and sustained.

ii. Innovation

The degree of innovation associated with the tool or service

MyGene2 is highly innovative as both a tool and a service. Public sharing at scale of health information and genetic data among families, doctors, and researchers is unprecedented. Indeed, MyGene2 is unique among a crowded field of data sharing tools. Other platforms (e.g., GeneMatcher, PatientKind/RareConnect) are moving to adapt elements of our novel approach to data sharing and matching into their services. We applaud and support these efforts as our overarching aim is to empower discovery of genes underlying 90% of rare diseases over the next decade. This will require a paradigm shift that MyGene2 hopes to inspire.

iii. Utility

The level of demand and utility associated with the proposed service or tool

Demand for MyGene2 has far exceeded our expectations despite the fact we have not yet marketed MyGene2 broadly or aggressively. Our current usage reflects the high demand of families for a web-based tool that enables them to effectively share health information and genetic data in order to facilitate diagnosis and gene discovery, engage with researchers as citizen scientists, and network with other families with similar conditions.

Usage statistics:

- ~1,000 unique visitors/month
- >500 profiles created
- ~50% of profiles submitted by families; 50% submitted by doctors or researchers (i.e., submissions balanced by stakeholder type)
- >470 candidate genes submitted

iv. Feasibility & Technical Merit

The feasibility and technical merit of the prototype

Feasibility:
MyGene2 is not only feasible, it is in active use.

Technical merit:

We leveraged multiple open resources and services to validate/annotate user-entered data in order to accelerate the release of MyGene2 and concentrate on developing features for which no service exists. These software resources include: Monarch Initiative, Human Phenotype Ontology, Mutalyzer, mygene.info, and myvariant.info. MyGene2 uses these tools to make data more accessible and useful to families. Families can upload sequence data to MyGene2 and browse it via the custom Gene.iobio instance in the absence of access to high performance computing resources.

Development & sustainability plan

Please briefly describe your vision, and any tangible steps you have taken, to develop your prototype into a sustainable tool or service that advances the goals of open science [400 words]

Our vision:

We envision MyGene2 as a central hub for collecting, organizing, and streamlining sharing and matching of health and genetic data submitted by families, doctors, researchers in order to facilitate gene discovery, understanding of gene-disease relationships, and ultimately improve the care and treatment of families with rare diseases.

Sustainability:

MyGene2 is poised to become a sustainable tool and service that will transform gene discovery for rare disease and dissemination / accessibility of information on the health problems associated with each mutation in each gene. Sustainability depends on many factors, perhaps most importantly, adoption by stakeholders, financial support, and ongoing innovation.

To inform the public and specific stakeholder groups about MyGene2 and encourage its use as a platform for open data sharing, we have undertaken the following:
● developed a multimedia marketing plan to publicize the advantages of using MyGene2
● partnered with several national science writers to develop stories and podcasts about MyGene2
● identified national and international meetings of stakeholder groups at which MyGene2 will be promoted
● partnered with genetic testing laboratories to advertise MyGene2 to their users and to develop direct data submission from each lab
● partnered with genetic testing laboratories to offer CLIA-certified research sequencing services directly to families
● partnered with the genetic counseling community to educate counselors in training about MyGene2 and integrate use of MyGene2 directly into clinical workflows
● partnered with rare disease advocacy groups to serve as their health and genetic data repository
● partnered with rare disease-focused social networking sites to enable families to directly “push” health and genetic data to MyGene2
● partnered with large-scale gene discovery programs to directly submit candidate gene data to MyGene2
● partnered with scientific journals to enable direct submission of candidate / known gene data linked to health information directly to MyGene2

To secure financial support for the continued development and maintenance of MyGene2, we have undertaken the following:
● partnered with commercial services offering genetic testing, social networking, etc. and initiated negotiations to support elements of MyGene2 that improve their existing workflows and market appeal
● created a donor case statement and planned gift account enabling donations directly to MyGene2
● formulated a development plan for hosting research studies on MyGene2—an additional benefit is that families in MyGene2 would have early access to new tools for diagnosis and case-matching

Final comments

Please use the box below to provide any further information you would like to add, that has not

been addressed in the questions above [200 words]

Further development of MyGene2 will concentrate on four key areas:

1. integration into clinical, research, and diagnostic testing workflows
a. enable doctors, researchers, and diagnostic labs to digitally transfer ownership of de-identified profiles to families
b. enable doctors to upload photos for families and families to moderate submissions
c. enable direct submission of data from diagnostic labs and medical journals

2. improved matching and data accessibility
a. develop matching algorithms that also utilize phenotype and variant type data
b. enable users to re-analyze sequence data using graphical browser-based tools (e.g., GxBrowse)
c. enable crowd-sourced evaluation of match reports.
d. automate posting of match reports to social media (e.g., Twitter, Facebook, Reddit)

3. platform enhancements
a. mobile compatibility to increase accessibility
b. infrastructure to enable translation of site content
c. professional content review and revisions to improve comprehension
d. incorporation of multimedia content to broaden audience appeal

4. regulatory and bioethics
a. revise consenting process and develop modular elements supported by multimedia that better convey potential benefits and risks of data sharing
b. develop an automated backend to support reconsenting minors who reach age of majority
c. create platform to identify issues important to the stakeholders, study issues, and report findings