

ALFRED P. SLOAN FOUNDATION

sloan.org | [proposal guidelines](#)

PROPOSAL COVER SHEET

Project Information

Principal Investigator

Daniel Himmelstein

Postdoctoral Fellow

daniel.himmelstein@gmail.com

Grantee Organization: University of Pennsylvania

Amount Requested: \$48,784

Requested Start Date: July 1, 2018

Requested End Date: June 30, 2019

Project URL (if any): <https://github.com/greenelab/manubot-rootstock>

Project Goal

The grant would support the next year of Manubot development. During this time Manubot would like to become a stable and full-featured system for writing scholarly documents on GitHub with substantial adoption.

Objectives

Manubot focuses on automating scholarly writing and publishing to the greatest extent possible. It combines existing computational tools like markdown, pandoc, git, GitHub, and continuous integration to achieve reproducible authoring. The process is modeled after the workflow of open source software development.

Proposed Activities

The proposed activities are continued Manubot development, documentation, and outreach. Outreach consists of holding workshops to onboard new users.

Expected Products

Manubot is an open source software tool. The products of this grant will be improvements to the Manubot software including improved functionality and documentation. Furthermore, this grant would support dissemination and outreach efforts, such as the PIs time to publish descriptions of Manubot and to perform workshops to grow the userbase.

Expected Outcomes

Manubot aims to make publishing more transparent and reproducible. At scale, Manubot will make the entire history of a scientific study transparent including where each result originated and who authored which parts. In addition, Manubot shows that high quality scholarly publishing is achievable without any costs, hopefully paving the way to decreased price barriers to disseminating knowledge.

Manubot Proposal to the Sloan Foundation

1. What is the main issue, problem, or subject and why is it important? The proposal should start with a discussion of the main issues, problems, or subjects that the proposer expects to address. It should go on to discuss why these are important and how they are related to the particular Foundation program that would provide the grant support.

Manubot is a tool for writing scholarly documents, using a workflow adapted from open source software. It addresses several problems with the current scholarly writing stack and publishing landscape:

1. **Opaque manuscripts.** Scientific stakeholders are beginning to realize the importance of open science. Currently, the pace of discovery is limited by the closed nature of most stages of the research life cycle. Manubot aims to make scholarly manuscripts (and documents more generally) transparent and auditable. Currently, it is difficult to know who wrote which parts of manuscripts and where the data and figures came from. Manubot adopts the concept of end-to-end transparency and reproducibility whereby every aspect of a manuscript is traceable. This should help keep scholars honest and provide greater public access into the research incorporated by a manuscript.
2. **Proprietary software.** For maximum reproducible and reusability, studies should not depend on proprietary software which requires permission to use. Currently, most scholars prepare and publish manuscripts using proprietary tools such as Microsoft Word, Endnote, or proprietary web applications. Manubot aims to establish a scholarly writing and publishing stack that is entirely open source. Open source infrastructure massively reduces centralization risk, whereby the scientific enterprise becomes dependent on specific entities and individuals.
3. **Expensive publishing.** While the shift to open access models of publishing has been a huge achievement in the past decade and is set to continue, many open access journals

charge high article processing charges. Much of these fees goes not to intellectually valuable contributions (such as peer review or copyediting), but instead towards converting documents between different formats. Manubot adopts the latest publishing standards and protocols, such that most aspects of publishing become automated and cost-free. In addition, Manubot allows scholars to produce documents that are technically superior to those published by many journals. Combined with preprinting, authors will now have a cost-free route to high-quality publishing, placing healthy competition on the scholarly publishing market.

2. What is the major related work in this field? The proposal should contain a summary of similar projects or other work addressing the same subjects, issues, or problems. This summary should include references to important publications or results of other significant creative efforts in the field. The proposer should describe how his or her work differs from, contributes to, or complements this work.

There have been several successful projects that have used git and GitHub to write scholarly documents. In 2013, two dozen mathematicians [created](#) the 600-page “[HoTT Book](#)” on “Homotopy Type Theory”, writing collaboratively in LaTeX on [GitHub](#). Two technical books on cryptocurrency — [Mastering Bitcoin](#) and [Mastering Ethereum](#) — written on GitHub in asciidoc format have engaged hundreds of contributors. Several books have been written using git and [bookdown](#), which extends markdown with several scholarly writing features. Manubot draws from many of the technical successes of these projects and offers a more comprehensive and featurefull solution for authoring scholarly documents. The goal of the Manubot is to create the most advanced and complete system for writing manuscripts on GitHub. Primarily, Manubot integrates existing software tools in a new way and thus shares many pieces with previous approaches. While previous projects have combined many of the same components, they were often custom one-time solutions. Manubot, on the otherhand, provides a template for creating manuscripts on GitHub.

There are several popular webapps for writing papers online that are not based on git repositories. [ShareLaTeX](#) and [Overleaf](#) allow writing LaTeX documents in your browser. [Authorea](#) is another webapp that allows mixing markdown and LaTeX sections in the same document. Online word processors, such as [Google Docs](#) and [Microsoft Word Online](#), offer a purely WYSIWYG authoring experience and can be extended with add-ons for scholarly features, such as [Paperpile](#) for reference management. The issue with these services is that they are generally proprietary and make their users dependent on a single centralized provider. In addition, they are not customizable. Finally, most of them do not achieve the degree of traceability and workflow customizability enabled by git. While services like Authorea can export manuscripts to git repositories, git is not used as the primary workflow tool. Hence, contributors cannot propose changes via “pull requests” and they have difficulty scaling to highly collaborative manuscripts. Nonetheless, these tools have been used for some collaborative manuscripts. For example, a 2017 [perspective](#) on the future of peer review was written by 32 authors using Overleaf. In addition, a Google Doc discussion, which involved over 150 individuals, was later [condensed](#) into a traditional [journal commentary](#).

3. Why is the proposer(s) qualified to address the issue or subject for which funds are being sought? The proposer’s education, previous accomplishments, etc., related to the proposed project should be discussed briefly.

Daniel Himmelstein received his PhD in Biological & Medical Informatics from the University of California, San Francisco in 2016. His past research has focused on integrating biomedical data to [predict](#) new therapeutic uses for drugs. Coming from an academic background, Daniel has himself authored several manuscripts and knows the pain points scholars currently face. In addition, Daniel’s computation expertise makes him well suited to lead development of the Manubot. For example, Daniel led Project Cognoma from 2016–2018, which was a volunteer meetup where over 35 individuals contributed source code. Daniel is an avid open source

developer (see his [GitHub activity](#)) and has contributed to major projects such as Pandas, Python, and NetworkX. Furthermore, Daniel takes a forward-thinking approach towards scientific publishing, having been one of the first to perform data-driven investigation of [publishing delays](#), [citation styles](#), and the [growth of Sci-Hub](#). Daniel currently is a member of the ASAPbio licensing task force.

Casey Greene received his PhD from Dartmouth College and currently is a Assistant Professor Of Pharmacology at the University of Pennsylvania. [His lab](#) applies cutting-edge data science and machine learning techniques to large public datasets in biology. In addition, the lab is interested in the future of scholarly communications. Casey [helped lead](#) the Deep Review, the project from which the Manubot was born. This project received over 40 contributions and illustrated massively collaborative scholarly writing was feasible using the open source workflow. Casey recently founded [biOverlay](#), a site that releases high-quality public reviews of studies, (primarily preprints). Casey has [been active](#) in bringing [Continuous Integration](#) practices from software development to scientific workflows.

4. What is the approach being taken? The proposal should contain a discussion of the overall approach to the subject, issue, or problem, and a detailed description of the proposed activities. The proposal should describe steps to be taken, by whom and when. If the proposal involves collaboration with other individuals or organizations, those collaborators and their expected contributions to the project should be listed. A management and staffing plan for current staff or staff to be hired should be provided, clearly indicating who will do what jobs or have what responsibilities. If applicable, the proposal should also describe the status of the work to date, and plans for dissemination and/or sustainability.

The following paragraphs, until the horizontal rule, were adapted from our [Manubot 2018 development proposal](#) — which we previously prepared for this funding request.

Manubot is a system for writing scholarly documents on GitHub. Manubot aims to **transform publishing** to be transparent & reproducible, immediate & permissionless, versioned &

automated, collaborative & open, linked & provenanced, decentralized & hackable, interactive & annotated, and free of charge.

Manubot transfers the lessons from the **open source software movement** to academic writing. Documents are written in the open, creating an easy path for readers to become contributors. We recommend following a workflow where anyone can propose changes, which are then reviewed by project members and revised as necessary before being incorporated (for example, [see how](#) this proposal was created).

Manubot is based on a collection of open source **tools and standards**. Manuscripts are written in markdown, using git for version control and GitHub for collaboration and review. Users can cite standard persistent identifiers and Manubot retrieves bibliographic details to automatically create a reference list. Conversion between formats (e.g. from markdown to HTML, PDF, and DOCX) is done using [Pandoc](#). When the manuscript changes, continuous integration automatically rebuilds and deploys it.

Currently, the Manubot consists of two repositories: [greenelab/manubot-rootstock](#) (a blank manuscript with the required files) and [greenelab/manubot](#) (a backend Python package). As of December 2017, [10](#) individuals have contributed to the **Manubot codebase**, which has [55](#) stars on GitHub. For more information on the Manubot, see [these slides](#) as well as the in-progress [Meta Review](#).

We initially created the Manubot to enable the **Deep Review**, a review paper on deep learning that we collaboratively drafted [on GitHub](#). Over 30 authors contributed to the Deep Review, using [issues](#) to discuss source material and [pull requests](#) to propose manuscript changes. The community appreciated the breadth of perspectives on a quickly-developing topic and the Deep

Review was the [most-viewed bioRxiv preprint](#) from April–June 2017. We’ve also used the Manubot to write our [Sci-Hub Coverage Study](#) — the second [most viewed PeerJ Preprint](#) ever — and to reproduce the [Bitcoin Whitepaper](#). In the Manubot’s first year, its collaborative, open, and review-driven *authoring process* yielded two of the most popular preprints of 2017.

Our **long-term vision** is an improved ecosystem of scholarly communication. Imagine you could click an author to highlight only the text they wrote. Or highlight all results and figures that descend from source code contributions by the selected author. Or highlight all results that depend on a selected upstream experiment or database. Or, if a new version of software used in the manuscript or an updated dataset is released, rebuild the manuscript with the updated version and see the resulting changes... regardless of whether you’re the original authors or a curious third-party! Manubot forms the foundation for the traceability of all aspects of a scholarly manuscript, which is the key principle that underlies this potential future.

Management and Staffing Plan

- [Daniel Himmelstein](#) (Principal Investigator, 3.6 Calendar Months, estimated at \$25,000).
Dr. Himmelstein will be responsible for setting priorities, managing budgets, implementing certain features, and managing contractors engaged in manubot development.
- [Casey Greene](#) (Co-Investigator, Scientist, 0.12 CM, estimated cost of \$1,700). Dr. Greene will assist Dr. Himmelstein at identifying priorities and engaging with contractors performing development.
- [Anthony Gitter](#) (Independent Contractor): Dr. Gitter will provide guidance on priority setting, continued assistance as a lead project maintenance on GitHub, and continued testing and prototyping. In addition, he will help with Manubot outreach at University of

Wisconsin-Madison, where he is an assistant professor, and will lead a workshop for local researchers.

- [Ariel Rodriguez Romero](#) (Independent Contractor). Mr. Rodriguez is the frontend developer who produced the [Sci-Hub Stats Browser](#) for our recent [study](#). He is interested in contributing more to academic projects and is a talented developer.
- **Additional Independent Contractors:** We plan to contract with additional developers based on specific needs and skill-sets. Fortunately, we have a strong network with the local developer community, since [we led](#) the volunteer Cognoma project, which was associated with the DataPhilly and Code for Philly meetups. Specific developers whom we've worked with before include [Derek Goss](#), [Andrew Madonna](#), and [Patrick Miller](#). We estimate the cost of contractor hours at \$125.00.
- **Workshops:** Dr. Himmelstein intends to lead several hands on workshops to demonstrate Manubot and onboard new users. The plan would be to do three workshop series: an east coast series, a west coast series, and a special session at a computational research conference. Tentatively, the east coast series would include major research centers close to Philadelphia, such as New York City, Baltimore, and Washington DC (estimated cost for 3 destinations of \$1,000). The west coast series could include Seattle, San Francisco, and Los Angeles (estimated cost of \$3,500 for a week trip with 3-5 workshops). The conference special session could be at the [Pacific Symposium for Biocomputing](#) in January, 2019 (estimated cost of \$4,000).

This proposal seeks funding for further development and outreach. Specifically, we envision four feature enhancements in addition to continued maintenance and technical support (**proposed technical activities**):

1. JATS is the predominant archival format for scholarly manuscripts that is optimized for machine-readability and interoperability. We will implement JATS export and incorporate an existing **JATS viewer**, such as [Lens](#), into Manubot webpages. (estimated time of 35 developer hours, target completion date of September 2018)
2. Currently all changes to a manuscript's source are tracked using git. We will implement **rendered diff functionality** to show how a manuscript changed from the reader's perspective and make creating "tracked changes" documents for journal resubmission easy. (estimated time of 50 developer hours, target completion date of September 2018)
3. **End-to-end reproducibility** is when the provenance of every manuscript result can be traced to its source. We will continue to evaluate and improve the Manubot template variables functionality, which allows users to embed results and method parameters directly from the analyses that created them. (estimated time of 15 hours, target completion date of December 2018)
4. The Manubot's **citation metadata** system is a powerful and versatile approach for bibliographic information. We will continue to expand Manubot's citation-by-persistent-ID capabilities. (estimated time of 10 hours, target completion date of June 2019).

We receive a considerable amount of comments from the community via [GitHub issues](#), which include questions, bug reports, and feature requests. This project would help support addressing **user-posted issues**, especially ones that require code changes.

5. What will be the output from the project? The proposal should include a discussion of anticipated products such as publications, programs, conferences, briefings, training of students, etc., that the project is expected to yield.

The first primary output will be **improvements to the Manubot**. Since Manubot is a software product, improvements fall into several categories, such as feature enhancements, better documentation, bug fixes, expanded test coverage, expanded compatibility, and easier setup.

Specifically, we are excited about the feature aims listed above as well as continuing maintenance of the project.

The second primary output will be **creating an community** of users and contributors. Manubot is an open source project and long-term success will require building a broad userbase from which contributors will arise. Fortunately, we have experience building open source communities. For example, we led *Project Cognoma* which [engaged](#) the local Philadelphia Meetup scene to create a [webapp](#) for cancer research. In addition, we [orchestrated](#) the [Deep Review](#) — the collaborative review paper from which Manubot was born — that received contributions from over 40 researchers.

Manubot already follows the best practices for open source workflows. Therefore, greater usage and engagement will come mainly from **improved documentation, researching user stories, and performing outreach activities**. Outreach activities include demonstrations and getting the word out at conferences. In addition, we'd like to help with onboarding by providing one-on-one support users that're just getting started. Since writing manuscripts with Manubot requires basic knowledge of git, this project also contains a **training mission** to teach scholars the open source workflow so they can most effectively use Manubot.

We are in the process of writing a [manuscript](#) describing the Manubot, which we intend to publish in a peer reviewed journal. However, the best way to judge the Manubot's success is by how many *other publications* were created with it!

6. What is the justification for the amount of money requested? The proposer should justify the budget request and why it is consistent with the associated output that the funds will support. This should be a general discussion, with additional detail provided in the budget section. (See “Budget”, below.)

The amount of money requested is relatively modest. Fortunately, Manubot relies on free open source tools and has few expenses that are not directly related to new development or outreach.

7. What other sources of support does the proposer have in hand or has he/she applied for to support the project? If funds are being provided or requested from other sources, the proposal should list those sources, amounts and the current status of funding from each. This narrative information can be integrated into the previous question, with detail provided in the budget section. (See “Budget, below.)

Currently, the project is supported in part by the Gordon and Betty Moore Foundation’s Data-Driven Discovery Initiative (Grant Number [GBMF4552](#) to Casey Greene). This grant currently pays a small portion of Daniel’s salary and is not specifically for Manubot development. Hence, most work on the Manubot up to this point has been provided on a voluntary basis by its contributors. Presently, we are not requesting Manubot support from other funders than the Sloan Foundation.

Sample Budget for a Proposal to the Alfred P. Sloan Foundation

Name of Organization: Manubot / UPenn
 Timeframe: 7/1/18-6/30/19

- * Not all the entries below may apply to your organization. Leave those items blank that do not apply.
- * For grants \$50,000 and under, the Foundation does not allow indirect/overhead charges
- * For multi-year grant proposals, please add columns for each year.

Project Expenses	Sloan Foundation Request, yr 1	Other non-Sloan Funds, if any	Total
Salaries and wages:			
Daniel Himmelstein: salary, 3.6 calendar months	20,703		20,703
Casey Greene: salary, 0.12 calendar months	1,700		1,700
			-
Administrative salaries and wages:			
List job titles and nature of support			-
			-
Student Support:			
Identify whether graduate or undergraduate and nature of support			-
			-
			-
Consultant/Contract services:			
110 contractor hours at \$125 per hour	13,750		13,750
			-
Employee fringe benefits:	2,421		2,421
Other Direct Costs:			
Travel (provide assumptions separately)	8,500		8,500
Supplies and materials			-
Printing, publications, copying			-
Postage, shipping			-
Computer support, telecommunications			-
Equipment (specify)			-
Other (specify):			
Post Doc Health Insurance (Daniel)	1,710		1,710
Total Direct Expenses	\$ 48,784	\$ -	\$ 48,784
Overhead/indirect (0%)	0		-
Total Expenses	\$ 48,784	\$ -	\$ 48,784

Daniel S. Himmelstein

Curriculum vitae

RESEARCH

Digital craftsman of the biodata revolution focused on human disease and open science. I study and apply het-nets, machine learning, data science, blockchains, graph databases, and open science.

2016 – 2017

University of Pennsylvania
Casey Greene Lab

Postdoctoral Fellow. Develop search algorithms for het-nets. Advocate for open licensing through data-driven analyses. Create Manubot to bring the open source workflow to scholarly manuscripts. Led the Cognoma meetup.

2012 – 2016

Univ of California, San Francisco
Sergio Baranzini Lab

Introduced *metapath*-based heterogeneous network analysis to bioinformatics. Improved upon edge prediction techniques to prioritize disease-associated genes and uncover multiple sclerosis susceptibility-genes. Predicted new uses for existing drugs and identified mechanisms of efficacy.

2012 – 2014

Independent Research
Cancer Epidemiology

Found that elevation is negatively associated with lung cancer, suggesting that atmospheric oxygen increases lung cancer risk.

2012

Univ of California, San Francisco
Graduate Rotations

Integrated high-throughput bioinformatics databases into a heterogeneous network in the Baranzini Lab. Compared the prevalence of disease-associated SNPs under additive and epistatic models in the John Witte Lab. Simulated case-control sequencing data capturing complex human demographic history in the Ryan Hernandez Lab.

SUMMERS OF 2010 & 2011

Princeton University
Olga Troyanskaya Lab

Developed a network-based information propagation technique to improve the prediction of tissue-specific gene function.

SUMMERS OF 2007 – 2011

Dartmouth Medical School
Jason Moore Lab

Developed methods for studying epistasis including a technique for evolving complex datasets and a modified permutation test.

📍 Philadelphia, PA
☎ +1 (603) 369-8394
✉ daniel.himmelstein@gmail.com
🌐 dhimmel.com

EDUCATION

2011 – 16 **Biological & Medical Informatics**
PHD
Univ of California, San Francisco

2007 – 11 **Biometry & Statistics**
BACHELOR OF SCIENCE
Cornell University

2004 – 07 **Hanover High School, NH**

GRANTS

2012 **NSF Graduate Research Fellowship**
\$132,000 for graduate education

2015 **Thinklab Pilot Funding**
\$5,000 for community participation

2014 **Graduate Student Research Award**
\$2,000 for HPC-workstation memory

ACHIEVEMENTS

2016 Philly Geek Awards
Nominated for Scientist of the Year

National Debate Tournament
only 3-time qualifying team in Cornell history

TEACHING & SERVICE

Peer Reviewer
80 merit points on Publons

2017 **ASAPbio Licensing Taskforce**

20— **Editor**
'15 ESR Journal
'14 Let's Talk Bitcoin






20— **Teaching Assistant**
'15 Systems Pharmacology
'13 Computational Evolutionary Genomics

2013 **Invited Participant**
Innovation in Medical Education Forum
Univ Francisco Marroquín, Guatemala





2007 **Swimming Captain**
Hanover High School

PUBLICATIONS

Systematic integration of biomedical knowledge prioritizes drugs for repurposing.

Himmelstein DS, Lizée A, Hessler C, Brueggeman L, Chen SL, Hadley D, Green A, Khankhanian P, Baranzini SE
2017 – *eLife* –  open notebook on *Thinklab*  40 GitHub repositories,  Hetionet Neo4j Browser,  data licensing issues in *Nature News*,  Russ Altman's 2018 Year in Review of 25 Translational Bioinformatics papers









Sci-Hub provides access to nearly all scholarly literature.

Himmelstein DS, Romero AR, Levernier JG, Munro TA, McLaughlin SR, Greshake Tzovaras B, Greene CS
2018 – *eLife* –  interview with *Science*,  article in the *International Business Times*,  *eLife* podcast,  greenelab/sci-hub

Heterogeneous network edge prediction: a data integration approach to prioritize disease-associated genes.

Himmelstein DS, Baranzini SE
2015 – *PLOS Computational Biology* –  featured in the *Biomedical Computation Review*,  PrePrint

Lung cancer incidence decreases with elevation: evidence for oxygen as an inhaled carcinogen.

Simeonov KP , **Himmelstein DS** 
2015 – *PeerJ* –  covered by 100+ news articles including the New York Times,  *PeerJ* featured article,  *PeerJ* top cancer biology paper,  Abramson Cancer Center 2015 basic research prize,  *SciShow* episode,  author interview




Opportunities and obstacles for deep learning in biology and medicine.

Ching T, **Himmelstein DS**, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, ...Gitter A, Greene CS
2018 – *Journal of the Royal Society Interface* –  greenelab/deep-review,  Nature Jobs TechBlog

Association-guided analysis of gene networks to discover the genetic basis of complex traits.

Greene CS, **Himmelstein DS**
2016 – *Circulation: Cardiovascular Genetics* –  dhimmel/snpIentiful

Understanding multicellular function and disease with human tissue-specific networks.

Greene CS , Krishnan A , Wong AK , Ricciotti E, Zelaya RA, **Himmelstein DS**, ...Troyanskaya OG
2015 – *Nature Genetics*

Meta analysis of genome-wide association studies reveals genetic overlap between Hodgkin lymphoma and multiple sclerosis.

Khankhanian P, Cozen W, **Himmelstein DS**, Madireddy L ...Hjalgrim H
2016 – *International Journal of Epidemiology* –  IJE blog,  dhimmel/hodgkins

iCTNet2: integrating heterogeneous biological interactions to understand complex traits.

Wang L, **Himmelstein DS**, Santaniello A, Parvin M, Baranzini SE
2015 – *F1000Research*

Evolving hard problems: generating human genetics datasets with a complex etiology.

Himmelstein DS, Greene CS, Moore JH
2011 – *BioData Mining* –  greenelab/model-free-data

Enabling personal genomics with an explicit test of epistasis.

Greene CS, **Himmelstein DS**, Nelson HH, Kelsey KT, Williams SM, Andrew AS, Karagas MR, Moore JH
2010 – *The Pacific Symposium on Biocomputing*

Multifactor dimensionality reduction for graphics processing units enables genome-wide testing of epistasis in sporadic ALS.

Greene CS, Sinnott-Armstrong NA, **Himmelstein DS**, Park PJ, Moore JH, Harris BT
2010 – *Bioinformatics*

A model free method to generate human genetics datasets with complex gene-disease relationships.

Greene CS, **Himmelstein DS**, Moore JH
2010 – *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*

The informative extremes: using both nearest and farthest individuals can improve Relief algorithms in the domain of human genetics.

Greene CS, **Himmelstein DS**, Kiralis J, Moore JH
2010 – *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*

SELECT BLOG POSTS

The most interesting case of scientific irreproducibility?
2017 – *Satoshi Village* – 📄 resulted in study retraction

On author versus numeric citation styles
2018 – *Satoshi Village* – 📄 dhimmel/pmc-citation-styles

The licensing of *bioRxiv* preprints
2016 – *Satoshi Village* – 📄 covered by *Nature News*

The history of publishing delays
2016 – *Satoshi Village* – 📄 covered by *Nature News*

Publication delays at PLOS and 3,475 other journals
2015 – *Satoshi Village* – 📄 covered by *Nature News* & *The Publication Plan*

RESOURCES

Manubot

Manubot is tool for writing scholarly manuscripts on GitHub, which aims to transform scientific publishing to be permissionless, transparent, automated, and free of charge – 📄 the most viewed preprints of 2017 on both *bioRxiv* and *PeerJ Preprints* were authored with Manubot. 📄 2018 Development Proposal

Open source software

GitHub – 📄 79 public repositories; 3,128 public contributions in the last year; have contributed to **pandas**, **python**, **networkx**, **tidyr** & **seaborn**

Leading Project Cognoma

Cognoma is a webapp that puts machine learning in the hands of cancer biologists. 📄 This ongoing project is organized through the DataPhilly Meetup group with 100s of individuals having attended and over 40 having contributed source code.

PRESENTATIONS

- 2017-02-22 **Models, Inference & Algorithms, Broad Institute**
CAMBRIDGE, MA
Integrating biomedical knowledge to predict new uses for existing drugs
📄 recording, 📄 slides
- 2016-10-13 **GraphConnect**
SAN FRANCISCO, CA
Hetionet Awakens: Integrating all of Biology into a Public Neo4j Database
📄 recording, 📄 slides, 📄 travel fellowship
- 2014-07-13 **ISMB: Intelligent Systems for Molecular Biology**
BOSTON, MA
Heterogeneous Network Link Prediction Prioritizes Disease-Associated Genes
📄 travel fellowship
- 2014-01-30 **BCATS: Biomedical Computation at Stanford Annual Symposium**
PALO ALTO, CA
Heterogeneous Network Link Prediction Prioritizes GWAS
- 2010-04-08 **EvoBio: European Conference on Evolutionary Computation in Biology**
ISTANBUL, TURKEY
A Model Free Method to Generate Human Genetics Datasets

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Casey S. Greene

eRA COMMONS USER NAME (credential, e.g., agency login): csgreene

POSITION TITLE: Assistant Professor of Systems Pharmacology and Translational Therapeutics

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Berry College	B.S.	05/05	Chemistry
Dartmouth College	Ph.D.	11/09	Computational Genetics
Princeton University	Postdoctoral	07/12	Bioinformatics

A. Personal Statement

The goal of the proposed work is to continue the development of Manubot. Daniel Himmelstein, a postdoc in the group, developed Manubot to construct a GitHub authored review of deep learning. Our review, written in the open on GitHub, cites more than 500 papers. A PDF of the manuscript was downloaded more than 20,000 times in 2017 making it the most downloaded preprint posted in all of that year leading to it being listed in the article, "2017 in news: The science events that shaped the year" from the journal *Nature*.

Since that time, Manubot has taken on a life of its own. Daniel has shepherded its development. During the course of this award I will help Daniel with strategic planning around Manubot's future and help him engage with the broader community. I look forward to working with Daniel as he manages this project.

Four peer reviewed publications that highlight experience and qualifications for this project:

1. **Greene, C.S.***, Krishnan, A.*, Wong, A.K.*, Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C., Chasman, D.I., FitzGerald, G.A., Dolinski, K., Grosser, T., Troyanskaya, O.G. Understanding multicellular function and disease with human tissue-specific networks. *Nat Genetics*. 2015. PMID: PMC4828725
2. Tan, J., Hammond, J.H, Hogan, D.A., **Greene, C.S.** ADAGE-based integration of publicly available pseudomonas aeruginosa gene expression data with denoising autoencoders illuminates microbe-host interactions. *mSystems*. 1(1):e00025-15. PMID: 5069748
3. Tan J, Doing G, Lewis KA, Price CE, Chen KM, Cady KC, Perchuk B, Laub MT, Hogan DA, **Greene CS.** Unsupervised extraction of stable expression signatures from public compendia with an ensemble of neural networks. *Cell Systems*. 2017. 5:63-71. PMID: 5532071
4. Beaulieu-Jones, B.K., **Greene, C.S.** Reproducibility of computational workflows is automated using continuous analysis. *Nat Biotech*. 2017. 35:342-346. PMID: 28288103

B. Positions and Honors**Employment**

2009-2012 Postdoctoral Research Associate, Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ.

2010 Lecturer, Department of Computer Science, Princeton University, Princeton, NJ.

2012-2015 Assistant Professor, Department of Genetics, Geisel School of Medicine, Hanover, NH.

2015- Assistant Professor, Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Philadelphia, PA

Honors

2014 Moore Investigator in Data-Driven Discovery, Gordon and Betty Moore Foundation.

Other Experience and Professional Memberships

- 2011- Member International Society for Computational Biology
- 2013- Program Committee for Intelligent Systems for Molecular Biology (ISMB)
- 2013-2014 Co-chair of the "Text and Data Mining for Biomedical Discovery" session at PSB.
- 2015 Co-chair of the "Computational Approaches to Study Microbes and Microbiomes" workshop at PSB.
- 2016- Co-chair. Posters. Intelligent Systems for Molecular Biology (ISMB) Meeting

C. Contributions to Science

* indicates co-first author.

Neural Network Methods for Unsupervised Analysis of Gene Expression

Gene expression data provide a broad lens for generating hypotheses about biological systems. Because these data are generated in a genome-wide manner, algorithms that mine these data are less biased toward well studied features of biology. Analyses that apply unsupervised machine learning algorithms to large compendia that include experiments covering diverse processes can be particularly well suited to discovering general properties of the measured biological systems. This work was performed in my own lab. During this work I developed and evaluated new algorithms for mining gene expression data, I drafted and reviewed manuscripts, and I supervised a PhD student, Jie Tan, and undergraduate student, Kathleen Chen, in my own lab. Programming support was carried out in part by Matt Huyck, Dongbo Hu, and Rene Zelaya who are programmers in my research group. Molecular experiments to validate computational predictions were performed in Dr. Deb Hogan's lab.

1. Tan, J., Ung, M., Cheng, C., and **Greene, C.S.** Unsupervised feature construction and knowledge extraction from genome-wide assays of breast cancer with denoising autoencoders. *Pac Symp Biocomput.* 2015; 20:132-43. PMID: PMC4299935
2. Tan, J., Hammond, J.H., Hogan, D.A., **Greene, C.S.** ADAGE-based integration of publicly available pseudomonas aeruginosa gene expression data with denoising autoencoders illuminates microbe-host interactions. *mSystems.* 1(1):e00025-15. PMID: PMC5069748
3. Tan J, Doing G, Lewis KA, Price CE, Chen KM, Cady KC, Perchuk B, Laub MT, Hogan DA, **Greene CS.** Unsupervised extraction of stable expression signatures from public compendia with an ensemble of neural networks. *Cell Systems.* 2017. 5:63-71. PMID: 5532071
4. Tan, J., Huyck, M., Hu, D., Zelaya, R.A., Hogan, D.A., **Greene, CS.** ADAGE signature analysis: differential expression analysis with data-defined gene sets. *BMC Bioinformatics.* 2017. 18(1):512. PMID: 5700673

Context-specific Functional Relationship Networks Extracted from Large-Scale Data

Proteins act in concert to carry out the biological processes required to develop and sustain living organisms. Because of the limits of annotation specificity and experimental coverage, pathway databases and protein-protein networks frequently treat an interaction at any measured time as a universal interaction. We showed that these context-specific networks were useful for generating hypotheses related to the development of asymmetry in zebrafish [1,2], phase-specific interactions in the human cell cycle [3], and the tissue-specific response to pro-inflammatory cytokines [4]. In our work, we have experimentally validated such discoveries. For these contributions I developed new algorithms, created the software infrastructure to enable the integration of large-scale databases, evaluated the resulting predictions, and wrote the manuscripts.

1. Wong, A.K.*, Park, C.Y.*, **Greene, C.S.***, Bongo, L.A., Guan, Y., and Troyanskaya, O.G. IMP: A multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks. *Nucleic Acids Res.* 2012 Jul;40(Web Server issue):W484-490. PMID: PMC3394282
2. Park, C.Y.*, Wong, A.K.*, **Greene, C.S.***, Rowland, J., Guan, Y., Bongo, L.A., Burdine, R.D., and Troyanskaya, O.G. Functional knowledge transfer for high-accuracy prediction of under-studied biological processes. *PLoS Comput Biol.* 2013;9(3):e1002957. PMID: PMC3597527
3. Tan, J., Grant, G.D., Whitfield, M.L., and **Greene, C.S.** 2013. Time-Point Specific Weighting Improves Coexpression Networks from Time-Course Experiments. *Evolutionary Computation, Machine Learning, and Data Mining in Bioinformatics.* 7833:11-22.
4. **Greene, C.S.***, Krishnan, A.*, Wong, A.K.*, Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C., Chasman, D.I., FitzGerald, G.A., Dolinski, K., Grosser, T., Troyanskaya, O.G. Understanding multicellular function and disease with human tissue-specific networks. *Nat Genetics.* 2015. PMID: PMC4828725

New Computing Approaches to Identify gene-gene Interactions in Genome-wide Association Studies

Gene-gene interactions represent a potential source of the missing heritability associated with common human diseases. As a graduate student, I developed new computational approaches to identify gene-gene interactions in genetic association data. These algorithms were focused on the particularly challenging problem of identifying gene-gene associations in which no SNP had a main effect. I developed methods based on evolutionary computation [1], heuristic approaches [3], and statistical approaches [4] to test for interactions in the context of main effects. I also developed a simulation approach to evaluate the implications of gene-gene interactions on standard genome-wide association study designs [2]. For these papers I developed and implemented new algorithms, evaluated these algorithms in the context of both simulation studies and real data analysis, and wrote the manuscripts.

1. **Greene, C.S.**, White, B.C., and Moore, J.H. Ant colony optimization for genome-wide genetic analysis. *Ant Colony Optimization and Swarm Intelligence*. 2008. 5217:37-47.
2. **Greene, C.S.**, Penrod, N.M., Williams, S.M., and Moore, J.H. Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS ONE*. 2009 June 2;4(6):e5639. PMID: PMC2685469
3. **Greene, C.S.**, Penrod, N.M., Kiralis, J., and Moore, J.H. Spatially Uniform ReliefF (SURF) for computationally-efficient filtering of gene-gene interactions. *BioData Min*. 2009 Sep 22;2(1):5. PMID: PMC2761303
4. **Greene, C.S.**, Himmelstein, D.S., Nelson, H.H., Kelsey, K.T., Williams, S.M., Andrew, A.S., Karagas, M.R., and Moore, J.H. Enabling personal genomics with an explicit test of epistasis. *Pac Symp Biocomput*. 2010:327-36. PMID: PMC2916690

Discovery of Gene-Phenotype and Drug-Phenotype Associations from Genetic Association Data

Algorithms that we developed and discussed above have been used by others to identify gene-gene interactions associated with human phenotypes. I have also participated in studies to identify gene-gene interactions associated with disease using our methods and those developed by others. In part, this required the development of new computing approaches. In Beretta et al. [1], I performed an analysis of gene-gene interactions from a candidate SNP dataset. In Greene et al. [2], we developed a new computing framework, graphics cards, to perform a genome-wide analysis of epistasis because the problem was not tractable on traditional computing platforms. In Mahoney et al. [3], I guided the analysis of students and postdocs using functional networks to analyze and interpret genetic association results. In Cordell et al. [4], I used an integrative bioinformatics approach to prioritize small molecules based on GWAS-associated pathways. For these projects, I participated in the development and evaluation of methods, the analysis of genetic association data, the guidance of undergraduate, graduate students, and postdocs, and the drafting of manuscripts.

1. Beretta, L., Cappiello, F., Moore, J.H., Barili, M., **Greene, C.S.**, Scorza, R. Ability of epistatic interactions of cytokine single-nucleotide polymorphisms to predict susceptibility to disease subsets in systemic sclerosis patients. *Arthritis Rheum*. 2008 Jul 15;59(7):974-83. PMID:18576303
2. **Greene, C.S.***, Sinnott-Armstrong N.A.*, Himmelstein, D.S., Park, P.J., Moore, J.H., and Harris, B.T. Multifactor dimensionality reduction for graphics processing units enables genome-wide testing of epistasis in sporadic ALS. *Bioinformatics*. 2010 Mar 1;26(5):694-5. PMID: PMC2828117
3. Mahoney, J.M., Taroni, J., Martyanov, V., Wood, T.A., **Greene, C.S.**, Pioli, P.A., Hinchcliff, M.E., Whitfield, M.L. Systems level analysis of systemic sclerosis shows a network of immune and profibrotic pathways connected with genetic polymorphisms. *PLoS Comput Biol*. 2015. Jan 8;11(1):e1004005. PMID: PMC4288710
4. Cordell, H.J., Han, Y., Li, Mells, G.F., Hirshfield, G.M., **Greene, C.S.**, Xie, G., Juran, B.D., Zhu, D., Qian, D.C., Floyd, J.A.B., Morley, K.I., Prati, D., Lleo, A., Cusi, D., Canadian-US PBC Consortium, Italian PBC Genetics Study Group, UK-PBC Consortium, Gershwin, M.E., Anderson, C.A., Lazaridis, K.N., Invernizzi, P., Seldin, M.F., Sandford, R.N., Amos, C.I., and Siminovitch, K.A. An international genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and highlights pathogenic pathways for drug targeting. *Nat Commun*. PMID: PMC4580981

Webservers for Search, Mining, and Exploration of Large-scale Compendia of Experimental Results

In addition to developing and applying new computational approaches to address biological challenges, I have also developed user-friendly webserver to make these methods available to the broader research community. I solely developed or co-developed these and other servers. PILGRM (sole developer) and IMP (co-developed with Aaron Wong) were developed while I was a postdoc. A programmer in my lab (Rene Zelaya) and I co-developed GIANT with a member of my postdoctoral lab (Aaron Wong). Tribe is a project of my lab (primary developer: Rene Zelaya with collaborating programmer Alex Frase in Marylyn Ritchie's lab), and I assist with the programming and administration. These servers were developed in Django using libraries that I developed for PILGRM. These servers have become heavily used. Google Analytics calculates that webserver and websites that we developed served 28,513 page views to 8885 different users in a recent quarter.

1. PILGRM: Platform for Interactive Learning by Genomics Results Mining. <http://pilgrm.princeton.edu>
2. IMP: Integrative Multi-species Prediction. <http://imp.princeton.edu>

4. Tribe: The collaborative platform for mining big data in biology. <http://tribe.greenelab.com>

PubMed Indexed work on My Bibliography:

<http://www.ncbi.nlm.nih.gov/myncbi/browse/collection/40332249/?sort=date&direction=descending>

Full list of published work (includes computer science publications) on Google Scholar:

<https://scholar.google.com/citations?user=ETJoidYAAAAJ>

D. Research Support

Ongoing Research Support

Moore Investigator	Greene (PI)	12/01/2014-11/30/2019
--------------------	-------------	-----------------------

Gordon and Betty Moore Foundation, Moore Investigators in Data-Driven Discovery

Learning the context of publicly available genome-wide data

We are developing algorithms that summarize publicly available assays of gene expression into models that represent important contextual information, e.g. the environment of the assayed cells and their responses to it.

Role: Principal Investigator

NSF #1458390	Greene (PI)	09/01/2015-08/30/2018
--------------	-------------	-----------------------

National Science Foundation

Bilateral BBSRC-NSF/BIO Collaborative Research: ABI Development: A Critical Assessment of Protein Function Annotation

We are developing a biologically grounded and comprehensive assessment panel for methods that aim to predict protein function and continuing the Critical Assessment of Function Annotation challenges.

Role: Principal Investigator

NIH R01 CA200854 Doherty, Schildkraut (PIs) 12/01/2015-11/30/2020

NIH/NCI

Characterizing Molecular Subtypes of Ovarian Cancer in African-American Women

We are characterizing the distribution of high-grade serous ovarian cancer subtypes in African-American women.

Role: Co-Investigator

NIH R01 NS095411 Sanchez, Ratner, Hoopes (PIs) 09/30/2015-07/31/2020

NIH/NINDS

Targeting tumors with NF1 loss.

We are developing a classifier that identifies cancers with genetic alterations that render them susceptible to treatment with drugs that are synthetic lethal with the loss of NF1.

Role: Co-Investigator

ALSF CCDL	Greene (PI)	07/01/2017-06/30/2018
-----------	-------------	-----------------------

Alex's Lemonade Stand Foundation

Childhood Cancer Data Lab

The major goal of this project is to establish a Childhood Cancer Data Lab within Alex's Lemonade Stand Foundation to build software infrastructure and enabling technologies for researchers studying childhood cancers.

Role: Principal Investigator

Digital Innovation Grant	Greene (PI)	12/01/2017-06/30/2018
--------------------------	-------------	-----------------------

Pfizer

“Digital Data Innovation Project”

The goal of this project is to develop software to construct hetnets that can be incorporated into Pfizer's research workflows.

Role: Principal Investigator

Conflict of Interest Disclosure

The investigators and their staff as well as the University of Pennsylvania have no potential conflicts of interest or sources of bias to report. This includes no financial interests, managerial affiliations, consulting interests, or patent interests to report.

Attention to Diversity Statement

Manubot is an open source project, which accepts contributions and welcomes users regardless of their race, sex, or any other characteristics. Unfortunately, there are well-documented gender and racial disparities in open source software contribution and usage. As such, we strive to lead a project where diverse persons and perspectives are welcome.

The Greene Lab has a public code of conduct, which every member is required to abide by, as part of the lab's [onboarding documentation](#). The code of conduct forbids "harassment of lab members in any form". While explicit harassment certainly will not be accepted on any Manubot events or discussion venues, we strive to do more to ensure a positive experience for all users and contributors. As such, we are devoted to civil and respectful online interactions and aim to lead by example.

In online issues related to open source projects, it's common for individuals to disagree on technical issues and for these disagreements to occasionally escalate into personal arguments or insults. We go to great lengths to keep discussion technical and focused, while avoiding digressing to personal discussion which could lead to a less inclusive environment. Daniel Himmelstein has a large history of civil participation in open source projects, as evidenced by the [640+ GitHub Issues](#) he has participated in.

Mentoring is an important way to allow new individuals to reach their potential as researchers. Daniel has mentored several individuals including:

- [María Teresa Chávez](#) who moved to San Francisco from Guatemala while still a high school student to pursue her goal of curing ALS. Daniel mentored her for several months at UCSF.
- [Sabrina Chen](#) for the summer before she entered college at John Hopkins University.
- [Ariel Rodriguez](#) on the Sci-Hub Coverage Study helping Ariel decide that his longterm goal was to pursue a PhD in computational science.
- Additional mentees of Daniel include Gregory Way, Leo Brueggeman, and David Nicholson.

María, Ariel, Gregory, and David are members of groups identified as underrepresented in STEM fields by the National Action Council for Minorities in Engineering (NACME). Accordingly, Daniel has mentored a diverse group of individuals, and helped provide them with positive experiences and excitement to be part of the open science community.

As a project, Manubot aims to reduce financial barriers to scholarly publishing. Currently, the cost of open access publication causes inequities and differential access to premier publishing venues. If Manubot is successful in shaping the future of publishing, the project could reduce financial barriers that privilege certain researchers, especially those from wealthy countries and institutions.

Information Product Appendix

Below is our general information product policy, which is based on the [blueprint resource sharing plan](#) (a community resource we created to help guarantee information products resulting from philanthropic funding are released openly). In addition, we provide a Manubot specific details describing the specific information products produced by Manubot and how we intend to release and license them.

General Information Product Policy

All outputs from this project — including code, data, figures, and documentation — will be made publicly available under an open license as soon as a stable version is available or an associated manuscript is published, whichever is earlier. Data will be released under the Creative Commons Public Domain Dedication (CC0, [version 1.0](#) or later). Code will be released under a [BSD 3-Clause License](#), a permissive open source software license, or dedicated to the public domain via CC0. Figures, documentation, and writing (including preprints) will be released under a Creative Commons Attribution License ([version 4.0](#) or later).

In addition to the aforementioned licensing for project outputs, creators of specific project content may release any such content as CC0, at their individual discretion. The principal investigator of this project may release any project content as CC0, at his individual discretion.

In instances where upstream inputs are used that restrict the licensing of project outputs beyond the aforementioned guidelines, the most permissive licensing option possible will be applied. However, no inputs will be incorporated that prevent original software from being released under an Open Source Initiative ([opensource.org](#)) approved license or prevent original non-code content from being released under an Open Definition ([opendefinition.org](#)) conformant license.

Manubot Specific Information Product Policy

All manubot software is developed in public repositories which are openly licensed. Currently, the Manubot codebase consists of two repositories: <https://github.com/greenelab/manubot> which is licensed under BSD 3-Clause and <https://github.com/greenelab/manubot-rootstock> whose code files are released under CC0 1.0. Any additional repositories we create will be immediately public and openly licensed. Since Manubot is a collaborative open source project, it's development workflow necessitates openness from the start.

Manubot defaults to applying a CC BY 4.0 license for manuscripts (i.e. the prose of a document). This is the license that manuscripts describing Manubot have retained, including the following manuscripts: <https://greenelab.github.io/meta-review>, <https://greenelab.github.io/manufund-2018>, <https://greenelab.github.io/deep-review>, and <https://greenelab.github.io/scihub-manuscript>. Any manuscript published in a peer-reviewed journal that describes the Manubot project and that was funded by this proposal will be published under an open license. As such, we will ensure that any publications describing the Manubot project are open access and widely accessible.

ALFRED P. SLOAN FOUNDATION
630 Fifth Avenue, Suite 2200
New York, NY, 10111
Phone: (212) 649-1649 | Fax: (212) 757-5117
sloan.org

PROPOSAL ADMINISTRATION FORM

Please complete and sign this form. It must also be signed by the head of your organization or another official authorized to sign on its behalf, endorsing this application and verifying that the information below is correct.

Applying Organization (Universities: specify if applying organization is a supporting foundation)

Legal Name:	Trustees of the University of Pennsylvania (Will be Grantee Organization if funded)		
Address Line 1:	Office of Research Services		
Address Line 2:	3451 Walnut Street, 5th Floor	U.S. Tax ID (EIN#):	23-1352685
City, State, Zip:	Philadelphia, PA 19104-6205	Tax Status:	Not for profit
Country:	USA		
Phone:	215-898-7293	Twitter Handle:	
Fax:	215-898-9708	Facebook Page:	
Email:	PennaORS@lists.upenn.edu	Org URL:	

Required Documents

U.S. Organizations (Universities are not required to submit these documents)

- | | |
|---|---|
| <input type="checkbox"/> Tax-exempt determination letter from the IRS | <input type="checkbox"/> Most recent Annual Report (or URL) |
| <input type="checkbox"/> Current operating budget | <input type="checkbox"/> Most recent audited financial report |
| <input type="checkbox"/> Explanation of the nature of the relationship your organization has with its fiscal agent or sponsor (if applicable) | |

(If not contained in the annual report, provide a brief description of your organization's mission, founding date, major programs, and size of staff. List major financial contributors and board members.)

U.S. Universities: Submit A 133 Audit Report

Canadian Organizations: Submit W-8BEN Form if available

Other Non-U.S. Organizations: Contact Sloan Program Director or email grantsadmin@sloan.org for requirements

Contact Information (Fill in Address only if different from Organization's address)

Project Director/Principal Investigator

Name:	Daniel Himmelstein
Title:	Postdoctoral Researcher
Organization:	University of Pennsylvania
Department:	Pharmacology
Address Line 1:	3400 Civic Center Boulevard
Address Line 2:	10th Floor
City, State, Zip:	Philadelphia, PA 19104-5158
Country:	USA
Work Phone:	215-898-1189
Mobile Phone:	
Email:	dhimmel@pennmedicine.upenn.edu
URL:	

Co- Project Director/Principal Investigator (If applicable)

Name:	
Title:	
Organization:	
Department:	
Address Line 1:	
Address Line 2:	
City, State, Zip:	
Country:	
Work Phone:	
Mobile Phone:	
Email:	
URL:	

Continued on next page →

Contact Information (continued)**Additional Contact**

Name: Andrea Albelda
Title: Grants Manager
Organization: University of Pennsylvania
Department: Pharmacology
Address Line 1: 3400 Civic Center Boulevard
Address Line 2: 10th Floor
City, State, Zip: Philadelphia, PA 19104-5158
Country: USA
Work Phone: 215-898-5270
Email: aalbelda@penntmedicine.upenn.edu

Admin/Financial Officer responsible for financial reporting

Name: Elizabeth D. Peloso
Title: Assoc. Vice President/Assoc. Vice Provost, Research Services
Organization: University of Pennsylvania
Department: Office of Research Services
Address Line 1: Franklin Building, 5th Floor
Address Line 2: 3451 Walnut Street
City, State, Zip: Philadelphia, PA 19104-6205
Country: USA
Work Phone: 215-898-7293
Email: PennaORS@lists.upenn.edu

Payment Address (This information will only be used if your project is funded)**Mail to**

Org Name: (must be same as Applying Org)
Department:
Address Line 1:
Address Line 2:
City, State, Zip:
Country:

Payee Contact (only if name should appear on mailing label)

Name:
Work Phone:
Email:

Form Completed by:

Name: _____ Title: _____
(Print name)
Signature: _____ Date: _____

Endorsement and Verification


To be signed by head of the organization or another official authorized to sign on its behalf

Head or Official authorized to sign

Name: Mark Nave
Title: Assistant Director
Department: Office of Research Services
Address Line 1: 5th Floor, Franklin Building
Address Line 2: 3451 Walnut Street
City, State, Zip: Philadelphia, PA 19104-6205
Country: USA
Work Phone: 215-898-7293
Email: PENNAORS@lists.upenn.edu

Director of Sponsored Research or equivalent (if University)

Name: _____
Title: _____
Email: _____

Signature: 
Date: 6/14/18