PhenX Steering Committee Meeting Minutes

January 29, 2021

Attendees

Steering Committee (SC) Members
Mary Marazita, Co-chair
Cathy McCarty, Co-chair
Lindsay Farrer
Elaine Faustman
Jonathan Haines
Tabitha Hendershot
Erin Ramos
Marylyn Ritchie
Sharon Terry
Marsha Treadwell
Rosalind Wright

National Human Genome Research Institute (NHGRI)
Jyoti Dayal
Madison Goldrich

National Institutes of Health (NIH)
Christopher Barnhart
Karen Parker
Jonathan Pollock
Mariela Shirley
Phil Tonkins
Kay Wanke

Guest
Ian Terry

Smoking Cessation Working Group (WG) Co-Chairs
Laura Bierut
Gary Swan

RTI Staff
Carol Hamilton
Lisa Cox
Steve Edwards
Michelle Engle
Lisa Gehtland
Lauren Gridley
Wayne Huggins
Stephen Hwang
Cataia Ives
Michelle Krzyzanowski
Debbie Maiese
Mark Nelms
Helen Pan
Mike Phillips
Amanda Riley
Thien Lam

Action Items	Responsible Person
Annual email or newsletter reminders to Toolkit users to cite PhenX when publishing	RTI
Provide Erin with COVID-19 Toolkit statistics for the February presentation to NHGRI Council	Stephen Hwang
Analyze use of COVID-19 measurement protocols	Stephen Hwang
Distribute the data sheets for Smoking Cessation to the SC for review	Lisa Cox
Retire the Sexual Identity protocol currently in SI	RTI
Schedule the next SC meeting in May/June	Lisa Cox
Circulate the ClinGen Polygenic Risk Score reporting framework to the Smoking Cessation WG when it is published	Erin Ramos

I. Welcome

Mary Marazita started the meeting at 1:02pm Eastern Time. She welcomed SC members and others joining the meeting.

II. Updates and Planning

Citation Analysis

Steve Edwards gave an update on ~~routine~~ the PhenX citation analysis. Information is pulled from the Web of Science and PubMed. There is also a manual curation step where programmatic annotations are reviewed. Jonathan Haines pointed out that if the number of measures in the Toolkit increases each year, while the number of citations is high and staying stable, then the number of references per measure is potentially going down. Steve said he had planned to touch upon that during the presentation.

There were 41 PhenX citations in 2020. To date, 127 articles include PhenX measures, 45 recommend and/or provide a link to the Toolkit, and there are 163 articles verified as citation of concept only. Steve presented a list of MeSH study types and disease/phenotypes associated with at least 3 publications that include PhenX measures. Based on study location reported in PhenX citations, 70 are studies in the US with publications in 9 other countries citing use of PhenX. An analysis of funding sources for publications citing PhenX show the most from NIDA, followed by NCATS and NHLBI. Funding Opportunity Announcements (FOAs) and notices that mentioned PhenX were reviewed. The years 2011 and 2012 saw a steady increase, then it leveled off in 2015. Steve illustrated how FOAs can be connected to award and publication, noting it can take many years between award of funding and publication of results; half of the publications came four years after the initial award. Can the impact of FOAs recommending PhenX measures be seen? Jonathan said his earlier question was answered. He had not thought about the lag time, so it makes sense.

Erin Ramos said it is incredible to see all the citation information and she is appreciative. One of the challenges is that cumulative data from Toolkit downloads continue to grow. There is not yet comparable growth in publications. Part of it may be the fact that people are not thinking of including PhenX citations in their publications. Steve mentioned it would make the analyses a lot more accurate. Erin noted that NHGRI staff has been talking with journal editors.

Jonathan asked how often PhenX reminds people that have downloaded one of the protocols about the need to cite PhenX when something is published. Carol replied that there is a message at the top of the download, but PhenX does not follow up. Jonathan suggested that emailing people once a year as a reminder (to cite PhenX when publishing) may increase the numbers. Carol thought that is an interesting idea that needs to be looked at. Lisa Cox suggested putting a reminder in the newsletter. Cathy McCarty loved that idea. Once a year, PhenX could ask people whether they used something in their studies.

Lindsay Farrer noted that people who are downloading measures are getting them into their data collection feed. The time from starting to collect data to publishing could be three years. Jonathan thought that is a valid point. Tracking that lag would be interesting to see whether it gets longer or shorter. Mary said the lag is an argument for sending periodic reminders. Carol gave the National Institute on Drug Abuse as an example, which publishes about 3 years after issuing FOAs. The rate of publishing may be driving by specific programs within the Institute.

Ian Terry wondered whether there is any data about people landing on URLs instead of downloading files. Carol stated that landing on pages is tracked to provide insight on navigation and that maybe PhenX could do comparison. Steve said a related thing that was noticed but was not yet folded into the analysis was that sometimes one person will take PhenX protocols and build a data collection instrument. Then another person will reuse that instrument and reference the paper where it was assembled. The second person does not reference PhenX but references the measures pulled together in the previous citation.

Toolkit Statistics

Stephen Hwang presented on usage statistics for the Toolkit. There have been 2,200 new report downloads in the past year. In 2020, there were 700 new registered users. The PhenX team has produced eight new publications since last January. 41 new publications cite PhenX measures and 40 new FOAs mentioned or encouraged the use of PhenX measures. He showed the top 10 protocols added to My Toolkit in the last quarter of 2020. All are part of the Social Determinants of Health (SDOH) core Collection. There was a surge in registered users starting in April when the COVID-19 protocol library was first shared in the Toolkit (as PDF files). New protocols continue to be shared via the COVID-19 library; there have been not additional COVID-19 protocols submitted to PhenX since January 2021. There was a large spike in the number of registered users in January 2017 when the Early Psychosis collection and Pregnancy domain were released.

Erin inquired whether the top 10 protocols were per quarter. Stephen replied they were. The switch to quarterly was made because cumulative counts could not tell the frequency of use of more recent releases and gave more weight to protocols which had been in the Toolkit longer. Erin said last night she was informed she has the green light to give a PhenX update at the next NHGRI advisory council meeting. It would be helpful to include some of these statistics. She will follow up with Jonathan too as an advisory council member. She noted that this will be his last NHGRI Council meeting.

Jonathan asked what the bottom 10 domains and protocols were. For example, is there a reason not to bother with C-reactive proteins anymore? These could be domains that don't need to be reviewed by ERPs. Carol thought that is an interesting point and the bottom protocols should be tracked. Sharon Terry noted that infrequent use doesn't necessarily mean that it is not a good protocol; she cautioned us not to make that assumption. Elaine Faustman asked whether the use of protocols tracks with specific research initiatives like ECHO and other large initiatives that have come out besides the COVID-19 collection. Is there an uptick when those come out? Steven has not looked at correlations with specific studies. Elaine wondered whether when people download protocols it is related to NIH announcements. Carol responded that we look at FOAs and PhenX citations and where they came from, at the institute level rather than by project. For example, ECHO picked up some PhenX measures, but they can only be identified by source. Elaine encouraged us to track disaster response and the use of COVID measures.

Needs Assessment

Helen Pan sought SC input on planning the Needs Assessment survey in support of the PhenX renewal application. The goal of the assessment is to get community feedback and ensure that the renewal proposal is addressing the needs of the scientific community. The first set of questions are about the PhenX project and the Toolkit. What Is the difference?

Erin said it would be fine to ask about the PhenX project and the PhenX Toolkit. But when thinking about users, she thinks they would get confused about how to differentiate the program—PhenX—and the resource—Toolkit. The Toolkit is the programmatic resource. So many people think about PhenX as the Toolkit. Mary said she thinks of the Toolkit as the resource and whatever it takes to get there the is PhenX, the project. Erin said that would include things like the shift towards seeking input from the community. Mary added coordinating with other databases. Jonathan asked who are we asking in this survey? Sharon Terry said if PhenX could be called a leader in standards that would be aspirational and true and visionary.

Helen continued with the presentation. Who should PhenX reach out to, and how? Registered users, WG members, SC members, newsletter subscribers, Twitter followers, NIH liaisons—maybe they can push the survey to their NIH-funded investigators. Put a banner on other NIH institute websites? Get into other project newsletters and conferences? In the zoom chat, Karen

Parker suggested the Trans-NIH Measurement Coordination Group, and noted that the Sexual & Gender Minority Research Office (SGMRO) links to the PhenX Toolkit from their measurement webpage. Jonathan thought all those people, at some level, could provide input. But what they know and care about are very different. These sets of people will have different interests and what they want to know about PhenX. For users, the focus should be more specific on the Toolkit (vs. project). What they do use and what is missing. But for the NIH CDE task force and some of those people, they would be more interested in what the PhenX project should be doing. They will be less focused on the usability of the Toolkit itself than the methods used by the project to put content into the Toolkit. The survey could go in two different directions. Helen confirmed the design would include a brief introduction then ask about use of the Toolkit. The survey will branch into different sets of questions for those who answered yes and no to Toolkit use.

How much should be asked about Toolkit usability? Cathy inquired whether formal usability analyses were being considered. Give the users a task to do something in the Toolkit and see how much time it takes and analyze their keystrokes and eve movements. Ian Terry asked what the definition of usability is. Is it the formal or colloquial definition? Helen recalled that formal testing had been done in the past but that was a while ago, and the Toolkit has evolved since then. Carol said that formal usability testing is being considered for inclusion in the renewal. PhenX may also want to consider a focus group. Since we are doing a needs assessment – take advantage of the situation to see if we can get volunteers for usability testing and/or a focus group.

Jyoti Dayal asked how much time is anticipated for people to fill out the needs assessment. She suggested using skip logic for people not familiar with PhenX and having a set of questions for people familiar with PhenX to give more detailed answers. Helen confirmed the survey design would use skip logic to branch into two sets of questions depending on familiarity with PhenX. The desire is to not make the survey too long to increase response rate.

ClinGen recently did a needs assessment. Their approach was to ask how important each of the five specific aims were, ask about tools and resources. The survey asked about the importance of tools and resources to the individual's research, and about the importance to the field of genomic medicine. Sharon Terry mentioned that "importance" is not the best language to use to pose the question. Helen showed a question that ClinGen asked about the respondents' roles. Jonathan stated that the ClinGen approach worked for ClinGen because the user group is narrow. PhenX is aiming at a broader set of users. It is better to focus more on how the user uses the Toolkit than the user's role. Erin agreed. It might be interesting to know whether users are graduate students or postdocs to help target additional outreach opportunities. But the ClinGen user base is well defined. Users know the ins and outs of ClinGen well. It is different. Ian said, from a crowdsourcing perspective, it depends on what the data will be used for. Does PhenX want to conduct a needs analysis on the entire possible market segment or improve usability for existing users?

Mary inquired whether this is the first needs assessment PhenX has done. Are there earlier ones to base the new one on? Carol replied that a needs assessment has not been done before. Sharon asked whether PhenX could build on the immense response received using card sort for COVID-19 and go with more modern technology to broaden reach at the same time as getting a sense of needs. The assessment needs to be less use-based and more needs-based. In the chat, Ian noted there is a well-supported methodology around asking people how important aspects are to them. If questions are asked about what is important, answers will be received about what is already satisfying. If users are already satisfied with the highly important areas, then they don't need attention. Lindsay mentioned that we should ask—what have they used, why they used it and how satisfied the were with it.

Helen continued, speaking about features. The desire is to find out whether people used any of the features listed. For resources, have people used tutorials, education modules, Twitter, or

relevant non-PhenX resources? Jonathan suggested that in addition to knowing whether people have used the resources, the question of why they used them should also be asked. It might be discovered that someone is not using tutorials because they are unaware of their existence or because they were not helpful. He suggested that asking these questions might also serve the dual purpose of making users aware of the resources. There should be additional information gathered about why someone is or is not using resources. Lindsay suggested asking questions about people's interests in particular areas related to features that PhenX has, then asking if they have used these features. This will show them that PhenX has resources available – that they are not using. Elaine suggested adding links to PhenX tutorials in appropriate places. That we could (also) make the needs assessment promotional and educational.

Helen continued, speaking about the desire to ask whether the Toolkit is easy to use. In an exit survey, respondents will be asked about their roles. What else would be helpful to know? Should people be asked about their countries? Mary said understanding the user community is important. The details do not have to be very granular but knowing users' roles and whether they are international would be good.

Helen continued, speaking about nonusers. For this group, the aim is to keep questions at the scientific community level to find out research areas that are important. What is important to their research? Cross-study analysis, data harmonization, data operability, or other things? How important are general NIH resources? Elaine mentioned that sometimes when she does surveys of this type, she also uses them to introduce people to something they do not know about. It is a good opportunity for nonusers and users who do not know about other available features. Helen noted such things had not yet been considered but they were great points. Elaine added that people cannot remember whether they have done a tutorial or YouTube video—the opportunities blend together. Cathy said that because of the unique terminology used by PhenX, a few things will need to be explained, for example, distinguishing measures in SDOH collections and the definition of protocol.

III. Smoking Cessation Working Group Update

Mary introduced Laura Bierut and Gary Swan, the co-chairs of the Smoking Cessation WG, to give an update. Laura stated that WG was nearing the end of its process. The WG researched and recommended different measures. It completed community outreach and had a post-outreach meeting. Overall, it has been extremely successful and the WG is proposing 14 new measures.

There were three areas of greatest discussion. The first was the Morisky medication adherence questionnaire. There were issues about the author talking about copyright infringement and retracting articles. There was an article in *Science* about this instrument. Therefore, based on outreach, a different questionnaire was selected. The WG did not want users of the PhenX Toolkit to encounter potential difficulties in using the protocol.

The second area of discussion was what would happen with polygenic risk score (PRS) and DNA methylation. By definition, PhenX wants validated, reliable measures. PRS and DNA methylation have an established foundation, but they are changing and moving forward. WG compromised by proposing protocols that are currently being used be added to Supplemental Information (SI). The aim is to standardize the recipe and refer to genome-wide data and the genome-wide association study (GWAS) catalogue, to standardize the approach to developing PRS knowing that the GWAS catalogue would be changing the reference panel. The WG will recommend PRS and DNA methylation for Supplemental lnformation. It is not quite ready for prime time. Nonetheless, it is important to put out there, because if not then where does one start? It is important for the field to start somewhere and have a foundation.

The final area of decision was renaming the domain to Smoking Cessation, Harm Reduction, and Biomarkers. The desire to add harm reduction was to recognize that abstinence is not the only desired outcome. Biomarkers was added to show the field that biomarkers were being recommended. There are already many in Supplemental Information, but there are other biomarkers for smoking cessation. Laura, Gary, and Kay Wanke will present at the Society for Research on Nicotine and Tobacco in February to disseminate information about the Toolkit. Gary added they will publish a paper based on this WG and its outcome. Another suggestion came from WG member Neal Benowitz who believed some measures needed to be highlighted as Core. There will be an effort to come up with a core set of measures considered to be the basics, the absolute minimum to set up a clinical trial. He does not anticipate it will be difficult to do.

Mary said WG is breaking new ground with the idea of core items for a specific purpose and delving further into PRS which is moving very fast. Laura noted that PRS was the most interesting component. Given that this is with NHGRI, WG was happy being a pioneer in proposing something. It is a starting point. Erin said it is good to see open-source protocols. NHGRI has several programs working on improving PRS, including a new PRS Diversity Consortium that will work with the community to expand current PRS methods. ClinGen has a PRS reporting framework that is in press. The goal is to get the community to make sure that when it is doing PRS to address to standard elements so samples can be evaluated. NHGRI is doing a lot in PRS. Laura stated that the more standardization NHGRI could give the better. What are the steps? How should PRS be selected? How many slips go into it? How do you weight them? Erin said she will circulate the reporting framework to the Smoking Cessation WG when it is published. Erin recognized Kay Wanke and Jonathan Pollock for their leadership on this domain. Jonathan stated the desire to push the field in the direction of quantifiable biomarkers. He is glad to have cutting-edge biomarkers in Supplemental Information and hopes to add more moving forward.

Lisa will distribute the link to the Smoking Cessation protocols today for the SC to review and approve. Carol said that the goal is to release the domain before the SRNT meeting in February.

Action item: Erin said she will circulate the ClinGen PRS reporting framework to the Smoking Cessation WG when it is published.

IV. Emerging Issues

Sexual and Gender Minority Collection

Tabitha Hendershot announced that Karen Parker from the NIH Sexual & Gender Minority Research Office (SGMRO) is in the meeting. PhenX worked with her a few years ago on updating the sex and gender measures in the Toolkit. She and Erin were in touch recently about pulling together a sex and gender minority collection for the Toolkit. Karen and Christopher Barnhart, also in the meeting, from SGMRO identified related content already in the Toolkit. Karen is proposing two parts of the project. First, they drafted a list of existing Toolkit protocols related to sex and gender minority research. There are 31 individual protocols and 14 SDOH protocols. The idea is to bring them together into an SGM collection. The content is already in the Toolkit so it will require a relatively low level of effort. Mike Phillips will lead that effort with Karen. There will be a National Academy of Sciences workshop and report expected later this year or early next year, and the collection may need to be updated accordingly. The next step is to engage a WG to identify additional SGM measures for release in the Toolkit.

Marylyn Ritchie suggested choosing a word other than minority to name the collection. Minority can have a specific connotation, especially in the space of racial minority. Marsha Treadwell agreed that minority should not be used and suggested, gender identity. Karen acknowledged the negative connotations associated with the word minority. However, the literature

talks about sex and gender minorities, and sex and gender identity may be seen as limiting. The National Academies of Science put out a report in October using the term sex and gender-diverse populations. Karen added that she knows a lot of people like the term minority. There are institutions starting up that tend to use the nomenclature of sex and gender minority. Sharon suggested calling it sex and gender measurement instruments or protocols. It will be mostly about diversity/minority, but it also says something about the majority population. The tools will not just define or study one set of people. The tools are offering new information about everyone and every sex preference. Karen said she is open to the idea. To be clear about the collection being envisioned, a majority of the measures are not about identity. Sharon reiterated that is why she recommends just saying tools. Karen suggested more discussion after figuring out the list of measures being proposed. Tabitha confirmed that naming the collection will be part of the discussion moving forward.

Sexual Identity Protocol

The next issue is related. There is a protocol called Sexual Identity that is currently in Supplemental Information for the Demographics domain. It was added to the Toolkit in 2016 by an ERP. It is a brief series of questions about feelings toward the same or other genders. Karen and Chris expressed that it should not be included in Supplemental Information. The protocol is a mix of gender identity and sexual orientation—two concepts conflated within a single protocol. There are already protocols for gender identity and sexual orientation. This protocol could cause confusion about what should be used. Tabitha showed the protocol. The first question asks about sexual orientation, and if a person selects something else, they go to the next series of questions where options are gender identity. Should the sex identity protocol be retired from Supplemental Information? When something is retired it is moved to the archive. Protocols in the archive will not be returned when the Toolkit is searched. Sharon though it should be retired. When something gets put in the archive, is a reason given for why it was retired? Tabitha replied yes, not many protocols have been retired, but a reason would be given. Someone asked, when something is retired, does PhenX reach out to people who downloaded it? Tabitha responded that that has not been done in the past. The protocol is kept in the archive so there is a record of what happened with it, and people can still review it. Marsha and others agreed with retiring the protocol. Mary announced that the motion on the floor is to retire the protocol.

Decision/Action item: Retire the Sexual Identity item currently in SI.

COVID-19 Library on the Toolkit Homepage?

In the next topic of business, Carol stated that the COVID-19 collection was released at the end of October. There is still a COVID-19 library, which is a library of PDFs of any COVID-19 surveys or full instruments submitted to NIH as well as modules, which are parts of surveys. The NIH Office of Behavioral and Social Sciences Research proposed modules for instruments, and PhenX is still checking on those. The modules have subtopics that map to the Disaster Research Response (DR2) main topic. It was coordinated with DR2. Since the COVID-19 collection was released, the only way to get to the COVID-19 library is from the PhenX Toolkit home page. People would really have to know what they were doing to find the library. PhenX wants people to use the COVID-19 collection, but the protocol library (of PDF files) is an important resource. A variable compare tool compare the different protocols is in development. Jonathan suggested having it as a subset or adding a link to the COVID-19 collections page. It could be called an additional resource or something to that effect. Naming it the COVID-19 library does not say what it is or why it is different from other collections. Lisa wondered whether it could be put on the COVID-19 research page, so users go to that page first. Carol said that is another option. The banner at top of home page right now takes the user to

the COVID-19 research collections. Jonathan cautioned against having the library on the front page. The aim is for people to use the collection first. Send them there. Mary agreed. Carol summarized the suggestions as: including a link to the library on the COVID-19 collection landing page instead of putting it on the home page. Erin supported that idea. If the library is moved to the COVID-19 resource page, highlighting the DR2 disaster response repository should be considered. The aim is to work with DR2 and make sure there is linking back and forth. Carol said that when one lands on the COVID-19 library page, the very beginning does talk about DR2. And there is a link to upload protocols. A link to : Additional COVID-19 Resources can be placed on the landing page for the COVID-19 collection, but where? Mary stated that Ian, in the chat box, suggested a quick architecture information outreach could help with the decision. Jonathan suggested putting the link to: Additional COVID-19 Resources right underneath the COVID-19 Research Project and COVID-19 Collection links on the COVID-19 collections landing page. In the chat, Marsha agreed with putting Resources below Collection.

Decision/Action item: Add link to Additional COVID-19 Resources right underneath the COVID-19 Research Project and COVID-19 Collection links on the COVID-19 collections landing page.

Cathy asked whether banners were in use before the COVID-19 banner. Carol replied that a banner was put up for the SDOH release because it was a big deal. It was retired recently. A banner is not put up for every release. It is a special occasion. Debbie mentioned (in the chat) that a banner had been put up for the Genomic Medicine Implementation release.

Smoking Cessation

Mary returned to the topic of a name change from Smoking Cessation to Smoking Cessation, Harm Reduction, and Biomarkers. She asked whether anybody objected, and no one did. Carol commented that the Smoking Cessation WG had wanted more biomarkers and wanted to push the envelope with PRS and Epigenetics.(DNA methylation) For PRS, the goal is to make sure people take the right path and get help as they need it. RTI will get data sheets for PRS and Epigenetics together and send to SC for review. Critical comments will be appreciated. There is a desire to put something in the Toolkit that is generally useful but not so detailed that it quickly becomes obsolete.

PhenX Renewal Planning

Carol mentioned the upcoming renewal. Erin chimed in that it was still a U24 but under a different program: Biomedical Knowledgebase with a deadline in late September (instead of late May). Once results from the needs assessment are obtained they will be shared with SC along with an outline for the proposal. These will be discussed in our next SC meeting in May/June.

V. Expanding Sickle Cell Disease (SCD) Collections (National Heart, Lung, and Blood Institute [NHLBI])

Wayne Huggins stated that NHLBI provided co-funding to PhenX in project years 3, 4, and 5 to expand SCD collections. The initial collections were put in through administrative supplements in 2015. A new project scientist, Phil Tonkins, is taking over for Ellen Werner who retired last year. The SCD Research and Scientific Panel are providing overarching guidance and approving WG candidates. They serve as liaisons, reviewing collections. The co-chairs are Jim Eckman and JJ Strouse. Marsha serves as the SC liaison.

There are six WGs planned. Orthopedic and skin were together originally but given comments from the last SC meeting, they will be done separately. Two of the WGs have started:

Curative Therapies and Pain. The aim is to start recruiting for the Psychosocial/SDoH and Genitourinary WGs next.

The Curative Therapies WG is co-chaired by Ross Fasano and Matt Hsieh. Half the members work on transplant and half work on transfusion. Matt is heading up the transplant side, Ross is heading up the transfusion side. Wayne showed the preliminary list of measures. There are eight measures for transplant and six for transfusion. They cover evaluation leading up to therapy and endpoints and adverse events. Supplemental information on liver fibrosis will be added. There will be guidance around specific staging of liver biopsy.

The Pain WG is co-chaired by Amanda Brandow and Pat Carroll. The WG divided the scope into acute and chronic pain. There is parallel construction between the two. Both groups have been going through the PhenX Toolkit closely and seeing what can be reused. The Pain WG is thinking of using cognitive and behavioral aspects already in the Toolkit. Most of the remaining WGs will start in the next six months.

VI. Wrap-Up

Mary suspended the recap because it was not necessary.

VII. Adjourn

The meeting was adjourned at 2:58pm Eastern Time.