PhenX (consensus measures for Phenotypes and eXposures) Project
External Scientific Panel Meeting Report
June 23, 2010

The External Scientific Panel (ESP), consisting of Rex Chisholm (Northwestern), Dani Fallin (Johns Hopkins), David Hunter (Harvard), Sudha Seshadri (Boston University, chair), and Ida Sim (UCSF) met for the first time during a conference call. All members of the ESP had received background information on the PhenX program and the agenda.

Open Session – The ESP met with Terri Beaty (PhenX SC member), Lindsay Farrer (PhenX SC member), Jonathan Haines (PhenX SC Chair), Carol Hamilton (PI, RTI), Bill Harlan (PhenX SC Co-chair), Heather Junkins (NHGRI, Program Analyst for PhenX), Teri Manolio (NHGRI, Director, Office of Population Genomics), Destiney Nettles (RTI), Jose Ordovas (PhenX SC member), Joe Pratt (RTI), Erin Ramos (NHGRI, Project Scientist for PhenX), Margaret Spitz (PhenX SC member), and Lisa Strader (Co-I, RTI).

1. NHGRI Perspective on PhenX program: Rationale and charge to ESP

Erin Ramos presented the NHGRI perspective on the PhenX program provided the charge to the ESP. The ESP was asked to provide external input to NHGRI regarding the overall direction of the project. This includes addressing specific scientific and bioinformatics issues as needed, evaluating the adequacy of progress and identifying potential new scientific opportunities.

2. PhenX Overview and brief Toolkit Demonstration

Carol Hamilton, PhenX PI, provided an overview of the project and the role of RTI for leveraging the consensus activities involved with managing 21 domain-specific working groups. She also provided a brief demonstration for navigating the Toolkit and various features.

3. Steering Committee (SC) Perspective and Future Directions

Jonathan Haines, PhenX SC Chair, discussed the role of the SC in shaping the goals and directions of the project. He reviewed several key policy decisions that directly impacted the deliberations of the working groups. These included requiring measures to be well-established and accessible to non-experts, allowing context-dependent (e.g., age, gender-specific) protocols, and prohibiting changes to protocol text or combining more than 2 protocols within one measure.

Discussion:

The ESP opened the discussion inquiring about the 15-measure limit for each domain and the care that was taken to avoid duplication in the Toolkit. Carol commented that the 15-measure limit was in place to encourage the working groups to think critically and carefully about the quality of measures that they chose to include. With 21 domains in the Toolkit, there is a lot to offer users. The PhenX team worked very closely with working groups to

ensure that measures did not overlap across domains. Adding complementary measures was acceptable but the goal was to avoid competing measures.

The ESP was positive about the layout of the Toolkit website and offered feedback for improvements. The Toolkit guidance on the homepage should be updated to reflect the importance of choosing appropriate measures based on a study design; this document should have more of an epidemiology tone rather than legalese and reflect all important caveats of using the Toolkit.

Information about mode of collection, time to administer protocol, and materials and staffing should also be included for each PhenX measure. Carol commented that this information is already included in the header of the individual data sheets. The ESP suggested adding established quality control protocols (e.g., data ranges, data types) to the Toolkit, when available, to ensure standardization. Carol discussed the development of the data collection worksheet (DCW) that will facilitate data collection. The DCW will include data ranges and other relevant QC metrics when appropriate. However, the DCW is not meant to encourage interpreting the data one way or another. Instead, it is meant to ensure that researchers collect all relevant variables associated with each protocol.

The ESP inquired about the consensus process among the working group members and whether the availability of empirical data was a criterion for selecting measures in an attempt to understand how scalable and sustainable the process is for selecting PhenX measures. Carol commented that consensus was challenging at times but the working chairs and SC liaisons were key players in getting the groups onboard. The working groups did review empirical data in the literature when selecting low-burden, high quality, validated measures. The in-person meetings, conference calls, and SC guidance were all essential elements for working group success.

Providing opportunities for users to provide feedback on the Toolkit is important, especially in the event that a user cannot find what they are looking for. Carol commented that there are several places within the Toolkit where users can provide feedback and suggestions; in fact this is widely encouraged. The ESP felt it would be helpful to provide a ‘top 10’ list of the measures that are downloaded the most; this would help users identify the popular measures. Tracking activity and use patterns would be helpful in order to identify areas of improvement.

PhenX Toolkit

The design diagram on the Toolkit homepage should parse out the differences between the notions of phenomena and measures. It is important to tease out the semantics in order to provide guidance to the user for choosing the appropriate measure for a particular study design. For example, myocardial infarction is a particular phenomenon but the actual measure for elucidating the outcome is an electrocardiogram. The addition of conceptual groups to the Toolkit (under development) could very well alleviate this problem but so would offering users more options for searching the measures. Carol commented that PhenX is not limiting, the Toolkit was designed as a resource for non-experts to browse and add a few high-quality measures to their studies.

The ESP commented that PhenX has addressed major areas of research but elements such as pain seem to be lacking. The major contributors to burden of disease in the US should be reviewed to identify any additional gaps in PhenX.

The ESP felt that it was unclear to users how the measures were organized. Instead of the primary browse displaying measures by domain, perhaps we can promote searching through the measures in a variety of ways that make more sense to the user. Carol commented that the conceptual groups of measures might be a more appropriate way to display the measures. Additionally, in order to prevent a null return when searching for a particular measure, a "fuzzy" search and using synonyms would be ideal. If users cannot find what they are looking for (or at least a suggestion) after their first search attempt, Toolkit use and adoption wouldn't be as high as it could be.

The ESP felt it was important to include information about versioning and recording dates of measures added to the Toolkit. This will enable users to see what version of a protocol they are using and how long it has been since the recommended protocol was selected as the best available option.

Promoting the Toolkit

By evolving the Toolkit into a community resource, the ESP felt that this would be a good opportunity to encourage investigators to conduct reproducibility and/or validity studies of PhenX measures within their own studies. This data could then be fed back into the PhenX Toolkit.

The ESP encouraged the re-positioning of the Toolkit as a social networking nexus for communities to be able to connect, blog about their Toolkit experience, and potentially nurture an environment where new measures could be suggested, developed and vetted. By tapping into these existing communities this might alleviate the burden of convening future working groups and improve Toolkit use and sustainability. To start, the ESP suggested that by engaging a small community, within the existing Toolkit framework, this would take advantage of existing group dynamics and in turn promote the project and draw more groups in. The ESP also felt that the Toolkit could offer the ability (and encourage) users to blog about the existing measures; these reviews and recommendations would provide users with an invaluable resource when selecting measures. These blogs would be very useful for non-experts when considering including measures outside of one's area of expertise.

The ESP pointed out that the clinical trials community is lacking in adoption of common measures and standards. The community is a proponent of common measures but there is frustration when outcomes are not comparable. Trainees tend to be more open to adoption and PhenX could engage the younger generation of scientists to use these measures.

The ESP felt strongly about using an existing ontology rather than creating a PhenX-specific ontology. The National Center for Biomedical Ontology is a rich resource and PhenX should consider re-purposing an ontology to fit the needs of the Toolkit. Carol commented that PhenX is not creating a new ontology and is using the NCI Metathesaurus as a starting point.

Evaluation/ Tracking Uptake:

The ESP felt that the ultimate goal was for studies to adopt the PhenX measures. Tracking publications will be useful as a long-term goal. Tracking website statistics can be a useful intermediate step, but this is not the ultimate proxy for measuring uptake. The problem is tracking early adopters; the Toolkit already asks users to register and report any PhenX usage into studies. It would be ideal to track usage via grant proposals and even resources like clinicaltrials.gov; users would be asked to report if an application or study plans to include PhenX measures. The ESP recommended that PhenX contact clinicaltrials.gov about flagging PhenX measures in the database. The ongoing collaboration with dbGaP (mapping measures to PhenX) should be continued, this is another way to track and flag PhenX measures within this GWAS database. Annual reports to NIH could require reporting PhenX usage, although this might not be possible to implement across all NIH ICs.

Collaborations

The ESP is supportive of continuing existing collaborations with groups such as DataSHaPER and dbGaP. PhenX should focus on interfacing well with these other groups without duplicating efforts. The FDA and CDISC have created a resource, CDISC Share, an open source library of standardized data elements. PhenX should consider using this resource within the Toolkit. The REDCAp (supports data capture) resource should be considered to leverage the PhenX data collection worksheets.

The ESP encouraged PhenX to develop collaborations with disease-specific communities.

Continuing the PhenX Program

The ESP believes PhenX has produced a valuable product in a short timeframe and provides one-stop shopping for basic phenotypic and exposure measures. The ESP is supportive of developing an RFA to continue the program and promote its uptake. As is stands, the PhenX Toolkit is a useful resource of common measures and is off to a great start. The real value will come in catalyzing a tipping point in the community for PhenX measures to become commonplace, but that would have been a high bar to reach in its first three years. It seems highly unlikely that this could have been expected this early in the program-- it could take another three to five years before it is clear that PhenX becomes established in the community.

The ESP suggested that when planning for Phase II of the program, attention needs to be given to identifying the Toolkit audience (i.e., whether this is a reference source or a place where the community can meet to develop new measures and discuss harmonization). This will help guide the direction of the project. Although PhenX does target the non-expert (in terms of a particular phenotype), the measures are only useful if they are accepted by the experts as reasonably valid substitutes for 'gold standard' measures. Continued expert interest in PhenX toolkit is important. In addition, the Toolkit should not simply serve as a static reference library. Social networking should be incorporated into the Website, allowing for significant community input including debate about the appropriateness of the selected measures and how the resource should evolve.

Jonathan noted that every working group dealt with the issue of identifying gaps in phenotyping and lack of quality measures within their field. They felt that the Toolkit should facilitate the development and adoption of new measures. The ESP felt that this concept should be considered for the next phase of the Toolkit.

The next Phase of the program should carefully consider issues of scalability, versioning, tracking uptake and community input. It might focus on a moderate expansion into additional domains (particularly those focused on burden of illness), revising and refining the current measures, and improving Toolkit capabilities over a three to five year period. The continuation application(s) should be peer reviewed and the budget could be somewhat modest.

The ESP is supportive of using a small funding initiative administered by NHGRI to support investigators to incorporate PhenX measures into existing studies and assess their experience. Such programs are often the best way to encourage initial uptake, but the real test will come after that support is withdrawn; once investigators are familiar with the measures, will they incorporate them without added financial incentives?

Recommendations

The ESP believes PhenX has produced a valuable product in a short timeframe and provides one-stop shopping for basic phenotypic and exposure measures. It will be important to strike a balance between promoting these measures as options and being prescriptive in their use.

The program is definitely worth some additional resources, otherwise the substantial investment in developing all these domains, measures, and tools risks being wasted. As it stands, the PhenX Toolkit is a useful resource of common measures and is off to a great start. The real value will come in catalyzing a tipping point in the community for PhenX measures to become commonplace, but that would have been a high bar to reach in its first three years, particularly given that several domains will be released only at the end of the funding period. It seems highly unlikely that this could have been expected this early in the program-- it could take another three to five years before it is clear that PhenX has become established in the community. One the other hand, one would know earlier than that if the PhenX measures were not being taken up at all.

PhenX Toolkit

- The Toolkit guidance statement should be updated to reflect 1) the intent of the resource and 2) clear caveats that the PhenX measures are not intended to be "gold standards" and will vary in their feasibility and validity across different study designs (e.g., case-control, population-based etc.). Study-specific scientific considerations must remain paramount and care will be needed when selecting measures for inclusion into a study.

- The Toolkit design diagram should parse out the differences between the notions of phenomena and measures. For example, myocardial infarction could be considered a phenomenon and the measure used to assess this condition is an EKG.

- When relevant, established quality control protocols and acceptable data ranges should be included; this will help to ensure better standardization for data analysis.

- Consider using the Toolkit as a social media website to engage the scientific community (especially younger investigators) and allow them to participate in the evolution of the resource. This could serve as a nexus for identifying gaps within a

particular research area and developing new measures. Adding a well-advertised and monitored blog for each measure would beneficial to the community to see what and how measures are being using and gaining insight about the usefulness and feasibility of implementing a particular measure across a wide spectrum of the research community.

- The Toolkit navigation and search options should be improved. Consider focusing on browsing and searching by measure and not domain because it is not always obvious when a measure belongs to a particular domain. Implementing a "fuzzy" search process and supporting searches by phenomenon, additional synonyms and key words, should be incorporated to prevent null search returns. Displaying prior usage of measures along with indications of related measures commonly selected by the community would also be useful. Null search returns should be stored and periodically audited to identify what the community wants, but is not able to find.

- Provide users a 'top 10' list of the measures that are downloaded the most.

- Instead of creating a new ontology, PhenX should consider re-purposing an existing ontology. The National Center for Biomedical Ontology (NCBO) is an excellent repository of ontologies. UMLS is a widely accepted ontology with a semantic network that might be useful.

- Consider including a 'date added' to each measure in the Toolkit to indicate versioning and expiration, if any.

- Investigators who use PhenX measures should be encouraged to investigate the reproducibility/validity of PhenX measures in their own studies instead of relying on the measurement properties reported in the literature.

Collaborations:

- The ESP is supportive of continuing existing collaborations with groups such as DataSHaPER and dbGaP. PhenX should focus on interfacing well with these other groups without duplicating efforts.

- PhenX should determine how often its measures are actually used in dbGaP, if possible, and consider adding a checkbox for PhenX measures in dbGaP submissions or specific accession numbers.

- CDISC and FDA have developed an electronic standardized library, CDISC Share (http://www.cdisc.org/cdisc-share) that includes standardized data element definitions. PhenX might leverage these existing definitions instead of creating new ones.

- PhenX might consider reaching out to disease-specific groups to form collaborations (e.g., Autism network).

Evaluation/Tracking Uptake:

- PhenX citations should be tracked in the literature. However, it might be years before any studies that incorporated PhenX measures will be published. In the meantime, PhenX should work with ClinicalTrials.gov to try to indicate PhenX measures within registered trials and observational studies. Likewise, the PhenX measures should be flagged in dbGaP. Tracking flagged measures can serve as an intermediate method of evaluating uptake.

- Other intermediate measures, if accessible, might include how often PhenX measures are used in R01 applications, IRB approvals, or downloads to PhenX shopping carts. Acceleration of rate of use of PhenX measures could be a useful intermediate metric.

- PhenX can continue to encourage Toolkit registration. Following up with these users might indicate where and how PhenX measures are being implemented.

Promoting the PhenX Toolkit

- Consider offering PhenX measures as standard data items in the planned REDCap library, a web-based application that was developed to support data capture for research studies. This would complement the PhenX data collection worksheet and promote uptake.

- Consider promoting use of PhenX measures in particular research communities (e.g., interventional studies, population –based epidemiological studies, genetic and non-genetic professional societies) and both national and international collaborations with a targeted approach.

Continuing the PhenX Program:

- The ESP is supportive of developing an RFA to continue the program and promote its uptake.

- When planning for Phase II of the program, attention needs to be given to identifying the Toolkit audience (e.g., experts vs. non-experts vs. trainees). The overall purpose should also be defined, such as whether this is a reference source or a place where the community can meet to develop new measures and discuss harmonization?) This will help guide the direction of the project.

- The Toolkit should not simply serve as a static reference library. Social networking should be incorporated into the Website, allowing for significant community input including debate about the appropriateness of the selected measures, inclusion of additional measures/protocols, and how the resource should evolve.

- The next Phase of the program should carefully consider issues of scalability, versioning, tracking uptake, and community input. It might focus on a moderate expansion into additional domains (particularly those focused on burden of illness), revising and refining the current measures, and improving Toolkit capabilities over a three to five year period. The applications should be peer reviewed and the budget could be somewhat modest.

- The ESP is supportive of using a small funding initiative administered by NHGRI to encourage investigators to incorporate PhenX measures into existing studies and assess their experience. Such programs are often the best way to encourage initial uptake, but the real test will come after that support is withdrawn; once investigators are familiar with the measures, will they incorporate them without added financial incentives?