**Steering Committee Meeting Minutes: February 2021**

**Friday, February 26, 2021**

* [Announcements, opening remarks | Rex Chisholm (SC Chair, Northwestern)](#7hgm1ipikbsd)
* [PRSs validated in multiple ancestries and populations](#4zbt7lwqm97v)
	+ [Abdominal Aortic Aneurysm](#klfhx2nf5nlb)
	+ [Breast Cancer](#727z1rc8nsqb)
	+ [Colorectal Cancer](#48wq7hubx7n6)
	+ [Coronary Heart Disease](#8ii5kq5fhiys)
	+ [Hypercholesterolemia](#mxjsomrxaugc)
	+ [Obesity](#9m4xec1nbvrx)
	+ [Prostate Cancer](#ei5fd0sr71u4)
	+ [Type I Diabetes](#aswifqn5msjb)
	+ [Type II Diabetes](#l2jn5qffs577)
* [PRSs validated in single populations or not published](#pkwznhi05wjq)
	+ [Asthma](#1gqmxdxx5asv)
	+ [Atrial Fibrillation](#8iy8sxxdo7r2)
	+ [Chronic Kidney Disease](#mcmos5y9og69)
	+ [Depression](#osly4eh6941p) *(Update: Moved to developmental pathway March 11th)*
	+ [Hypertension](#r6mw1gh67pi9) *(Update: Moved to developmental pathway March 11th)*
	+ [Lupus](#926h9t6v7pwm) *(Update: Moved to developmental pathway March 11th)*
	+ [Non-alcoholic Fatty Liver Disease](#yhcv56b8zygf) *(Update: Moved to developmental pathway March 11th)*
	+ [Stroke](#v09a52w9jwla) *(Update: Moved to developmental pathway March 11th)*
* [sIRB: Reports, surveys, & pending questions | Digna Velez Edwards (VUMC), Ingrid Holm (BCH), and Wendy Chung (Columbia)](#kwmo8n13ill)
* [Network decisions & timeline review | Rex Chisholm (SC Chair, Northwestern)](#m34ppvguwe8x)

**Action Items**

* **Action Item:** Surveys need to be finalized prior to submitting the sIRB.
* **Action Item**: Timing of when the MeTree survey is completed needs to be decided.
* **Action Item:** A decision on when the MeTree report will be returned needs to be made.
* **Action Item:** Site leads will need to decide on which data variables will be collected in MeTree and the surveys to limit duplicate answers.
* **Action Item:** The PRS workgroup will discuss SNPs on GDA chips on the next call (*complete*)
* **Action Item**: The PIs will complete a survey to facilitate discussion on the March 11th PI call to determine which conditions should progress with PRS validation. (*complete*)

**Decisions**

* **Decision**: For not at high risk participants, language should read “*does not meet study threshold for high risk*.”
* **Decision:** Multiple workgroups discussed study design mechanisms and concluded:
	+ High monogenic risk will be returned in person (face to face, Zoom, phone).
	+ High PRS will be returned in person (face to face, Zoom, phone).
	+ Decisions about returning high family history risk for patients without high monogenic or polygenic risk is dependent on the phenotype and that may be returned by certified mail or a read receipt in the EHR.
	+ All others would be passive return in the EHR for providers; patients will get notifications via the portal or mail.

**Friday, February 26, 2021, Steering Committee Notes**

### Announcements, opening remarks | Rex Chisholm (SC Chair, Northwestern)

* + The group was welcomed to the February 2021 Steering Committee Meeting. The majority of this meeting will consist of a discussion of the PRS validations.
	+ The goal is for everyone on the call to listen and interpret the information being presented and give feedback regarding the particular PRSs and Phenotypes.
	+ At the end of the meeting, the goal is to determine which PRSs need to be prioritized.
	+ Each set of conditions leads has been asked to complete two template slides to be presented to the group. There will be a 5-minute discussion on each phenotype, and then there will be times for questions.
	+ There will also be a presentation from the sIRB group that will detail the progress made.
	+ The remainder of the meeting will consist of a discussion of the Network decisions that have been made up until this point. There will also be a discussion of the Network timeline progress.
* PRSs validated in multiple ancestries and populations| Sites
	+ Abdominal Aortic Aneurysm (AAA)
		- Expected completion:
			* The AAA phenotype risk score validation in multiple ancestries and populations is expected to be completed in May 2021.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
			* It is very likely for the condition to be successful in the prospective cohort since it is highly actionable and will be delivered to Broad on time.
		- Availability of datasets for clinical validation:
			* The validation pipeline includes RF Detection using the Millions Veterans Program (MVP) literature to detect genetic RFs, Model Building using BioVU to build PRs and GIRA models, and Model Testing.
			* The eMERGE 1-3, VUMC’s Synthetic Derivative (SD), and Million Veterans Program data repositories will be used for validation.
				+ In the GWAS published by MVP, 20 SNP were discovered to be significant based on 7,642 AAA cases and 172,172 controls.
		- Findings:
			* In BioVU, 824 cases and 86,551 controls were evaluated. White: 757 cases and 64,760 controls. Black: 52 cases and 13,848 controls.
			* It has been more difficult to identify cases across multiple ancestries and populations.
				+ The AUC was cross validated and between 0.48-0.64 for transethnic, white, and black ancestries.
				+ When age, sex, BMI, PC1-10, and smoking status were added to PRS the AUC rose to 0.89-0.95.
				+ There is subtle diminishing ability when without PRS and only age, sex, BMI, PC1-10, and smoking status; the AUC was between 0.89-0.95.
		- What additional resources or data would help with development and validation:
			* There is still only based data from white ancestry.
			* There is concern about building the PRS for a small number of cases, when people of African ancestry experience AAA less than other ethnic ancestry groups.
	+ Breast Cancer
		- Expected completion:
			* The PRS is completed for non-African populations.
			* The GWAS for African is likely to be completed by the end of 2021.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
			* It is likely to be successful for the African population by the end of 2021.
				+ This is contingent on the African American GWAS.
			* PRS has been completed in the non-African American population.
			* PRS can stratify breast cancer risk in women of African ancestry with attenuated performance.
			* African-specific breast cancer GWAS is ongoing.
			* It is very likely to be successful for the African population by the end of 2021 and is contingent on the African American GWAS.
		- Availability of the dataset for clinical validation
			* Breast Cancer is one of the most researched phenotypes.
			* The largest GWAS is in the European population.
			* The 3820 SNP model developed by the BCAC consortium is the most widely used.
			* All available models are based on the European model.
			* There may need to be African-specific PRS models for Breast cancer.
			* PRS validation is based on eMERGE III.
			* There are two clinical models available that use PRS.
				+ **BOADICEA** v5 ([https://canrisk.org](https://canrisk.org/))
				+ This model has proven superior to the Tyrer-Cuzick model.
				+ **Tyrer-Cuzick**/IBIS ([https://ibis.ikonopedia.com](https://ibis.ikonopedia.com/))
		- Findings
			* PRS has been completed in the non-African American population.
			* PRS can stratify breast cancer risk in women of African ancestry with attenuated performance.
		- What additional resources or data would help with development and validation?
			* Complete African-specific breast cancer GWAS is needed.
	+ Colorectal Cancer
		- Background and Justification:
			* Modern screening tools have led to the decline in colorectal cancer incidence in people above the age of 50 since 1980. However, for those under 50 incidence has risen.
			* The group discussed the implications of those with FH risk receiving screenings 10 years earlier than usual.
				+ Colorectal cancer phenotype will impact a participant’s FH risk status.
				+ Results will be returned for FH or high risk colorectal cancer.
		- Availability of dataset for clinical validation:
			* AUC analysis was adjusted for age, study, and endoscopy history.
			* E-score is the environmental score based on height, BMI, education, alcohol, physical activity, NSAIDs use, postmenopausal hormone use and dietary factors (intake of fiber, calcium, folate, processed meat, red meat, fruit, vegetable, and total energy.
			* G-score is based on known GWAS loci.
			* There are five published risk prediction models for colorectal cancer: family history (AUC: 0.52-0.53); family history and E-score with 19 factors (0.60); family history and G-score with 63 SNPs (AUC: 0.59); family history, E-score with 19 factors, and G-score with 63 SNPs (AUC: 0.62-0.63); and family history and G-score with 120 SNPs (AUC: 0.63).
			* For early onset colorectal cancer, the PRS is based on 140 known GWAS SNPs (model adjusted for age, sex, and study).
			* The genetic risk factors are strong predictors for early onset colorectal cancer without a family history. This provides great potential for precision prevention.
			* The eMERGE 1-3 of 573 cases and 37, 641 controls will be used for model validation.
			* The GERA cohort of 1,635 cases and 70, 660 controls will be used for model evaluation.
			* The PRS development includes three approaches:
				+ Approach 1 has 140 known GWAS variants; 58,131 cases; and 67,347 controls.
				+ Approach 2 will consist of feature selection and machine learning.

LD Clumping and selecting the top SNPs to maximize AUC.

The machine learning techniques will include penalized regression and XGBoost.

* + - * + Approach 3 will apply the LDpred method using association statistics.
			* There is a 95% confidence interval for the genome-wide risk prediction for colorectal cancer across multiple ancestries.
			* The models were adjusted for age and sex (not included as predictors in the AUC calculation.
			* There is ongoing trans-ancestral PRS development.
			* The AUC for the model is based on known GWAS and LDpred developed in Asians and Europeans.
		- Expected completion:
			* The group expects to complete PRS validation by the end of Summer 2021.
				+ There are not any barriers to completion as data are available. eMERGE data are available with colorectal cancer outcomes.
				+ It is very likely this condition will be successful for the prospective cohort with transfer Broad planned for early summer.
				+ The group’s recommendation is to use OR of 2.2 to identify those between 40-50 years of age to start screening earlier.
	+ Coronary Heart Disease
		- Expected Completion:
			* Coronary Heart Disease is expected to be finalized by the end of May 2021.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
			* There are currently no anticipated barriers to completion and it is very likely that the condition will be successful in the prospective cohort.
		- Availability of datasets for clinical validation:
			* Six cohorts from dbGaP are also being used for clinic validation with an estimated completion of 3 to 6 weeks.
		- Findings:
			* Previous work conducted last year tested the utility of 4 polygenic risk scores. A metaGRS composed of about 1.7M genetic variants performed best.
			* The results of the validation in 3 major ancestry groups were similar to those of the 2018 published paper with very similar hazard ratios per standard deviation.
			* When comparing the top 10% and the rest, the odds ratio was about 3 for Europeans and about 2 in African ancestry.
		- What additional resources or data would help with development and validation?
			* Currently, Mayo’s work is focused on expanding upon methods to improve performance in non-European ancestries, particularly African American.
			* It could be helpful, across all phenotypes, to be explicit when mentioning self-reported race and ethnicity versus genetic ancestry.
	+ Hypercholesterolemia
		- Expected Completion:
			* Hypercholesterolemia is expected to be finalized by the end of May 2021.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
			* Hypercholesterolemia is very likely to be validated and to be included in the prospective cohort for eMERGE.
		- Hypercholesterolemia is a well-studied trait and the Global Lipid Genomics Consortium (GLGC) has developed large-scale GWAS for this condition.
			* There are multiple existing validated PRS scores in the literature.
		- The literature that is currently being used for this condition is Kuchenbaecker et al, “Validated Multi-ethnic PRS”, which was published in 2019.
			* In this study, hypercholesterolemia was validated across African, Asian, and European populations.
		- It is anticipated that the GLGC will release a new PRS that includes over 1.6 million participants that will have trans-ethnic validated PRS across African, Asian, European, and Hispanic ancestries.
		- Two models are being used for testing. One model is being used to see if it is predictive for the trait itself, the other model is used to test if hypercholesterolemia is predictive.
		- The group is producing very similar results to the Kuchenbaecker study.
		- The group is seeing odds ratios around 1.5 per standard deviation of PRS.
			* This score seems to appear stable across groups.
		- The new score will likely have a 30% improvement.
		- The group is wanting to get additional sample sizes for the Asian population.
		- The group is waiting for the GLGC to release the trans-ethnic PRS, which is expected to be released in March 2021.
		- In the case where there is an extended delay in GLGC releasing their new score, the group will use and optimize Kuchenbaecker’s score as backup.
		- As of February 25th, the new GLGC score has now been released, which is much earlier than expected.
		- The new GLGC PRS performs the best in African descent populations, and is very stable across all ancestries.
			* Most of the study is conducted in the adult population.
		- Using the new GLGC PRS, the odds ratio is predicted to increase to 2.5.
		- Statins may be recommended for high PRSs
		- Clinical utility in the pediatric population is likely to be high.
		- There is a concern that hypercholesterolemia will not be detectable in the pediatric population.
	+ Obesity
		- Expected completion:
			* The plan is to transfer the model and data to the Broad by early summer.
			* A barrier to completion is the availability of the GIANT consortium data not being available until Spring 2021.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
			* There are good predictive effects across ancestries, as well as good clinical utility and actionability.
			* The odds ratios are very reasonable across all populations.
			* Validation is mostly available.
		- Availability of datasets for clinical validation:
			* Trans-ancestry summary stats from ongoing GIANT GWAS meta-analysis for BMI was used.
			* There is now access to the more extensive meta-analysis with 1.6 million participants and is more diverse. This data will be available soon.
			* The trans-ancestral score was based on the UK biobank and was validated using the MVP dataset.
		- Findings:
			* The GWAS for this phenotype is based on routinely measured traits.
			* The population included in this study was mainly European.
			* PRS generally performs better than age and sex combined in the UK biobank.
			* Performance of the eMERGE population will likely fall between the UK biobank and MVP performances.
	+ Prostate Cancer
		- Expected Completion
			* From a PRS perspective, this group is confident that this will be ready for implementation if the condition is selected.
		- How likely this condition is to be successful for the prospective cohort considering timing needs
			* This condition is very likely to be successful.
		- Availability of datasets for clinical validation
			* A large GWAS was published with 107,000 cases in January 2021 that is multi-ethnic and has validated new 269-variant PRS for prostate cancer.
		- Findings
			* The top 1% has an OR well above 5.5.
			* Raw summary statistics have not been published on dbGAP yet so the tail probabilities have not been determined at this time.
			* The PRS has been validated in European and African cohorts.
			* Areas under the curve for various groups are:
				+ AUC for age and family history is 0.78.
				+ AUC for age and GRS is 0.83.
				+ AUC for age, family, and GRS is 0.84.
		- The top 10% has OR above 3.5 in all populations and could be challenging with setting the parameter for at high risk results.
		- A screening recommendation was suggested for those who are at high risk should be screened earlier.
		- The actionability for this condition would be PSA screening 5 years earlier than current guideline recommendations in those at high risk.
		- It is scientifically valid to return results to the top 10% but not budgeted to return results to this large of a group.
		- There is thinking in the group that returns are in line with the other cancer phenotypes with communication and assessing risk.
	+ Type I Diabetes
		- Expected completion:
			* Internal validation in CHOP data is complete and validation in eMERGE I-III data will be done by May.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
			* Given how predictive the score is, it is anticipated this condition will highly likely be successful in the prospective cohort.
		- Availability of datasets for clinical validation:
			* There are a couple genetic risk scores for Type I Diabetes developed by the National Genetics Consortium and CHOP is proposing to use the newer one called GRS2.
		- Findings:
			* Previous scores have been developed separately for Europeans and African Americans although overlap exists in the scores and mechanisms.
			* The AUC for just cases and control in Caucasians is 0.86 (specificity = 73.9%, sensitivity = 83.3%) not including any covariates.
			* Relative risks and odds ratios may be very high for tail discrimination while the absolute risk of the disease will be very low so as a pediatric condition, communicating this to parents may be tricky.
	+ Type II Diabetes
		- Expected Completion:
			* PRS validation is expected to be completed in April 2021.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
			* Type 2 diabetes is very likely to be successful, and used in eMERGE.
		- The group started with a training GWAS from Mahajan et al. 2018.
			* The GWAS included a large sample of European ancestry.
		- There is also a DIAMANTE paper from Mahajan et al.. DIAMANTE is a consortium with many diabetes studies.
			* DIAMANTE has a decent sample size of all ancestries.
		- Using this GWAS, the group created a PRS in three methods: P+T, LDpred, PRS-CS.
			* The performance metrics were better in LDpred than PRS-CS.
		- UAB has 4 datasets that have more than 18,000 participants with African ancestry imputed to TopMed.
		- There is a publication from Vanderbilt from the MEDIA consortium, which is a meta-analysis of 17 GWAS of type 2 Diabetes In African Americans.
			* From this study, UAB has been calculating PRS scores.
			* LPpred is performing well in REGARDS cohort
			* UAB is applying this African ancestry PRS to the other three cohorts they have, and they will do meta-analysis. The goal will be to apply this data to diabetes in eMERGE III.
		- The summary statistics from the DIAMANTE paper will be used to make trans-ancestry PRSs.
		- In Seattle, a large number of Asian Amercians are being collected in their biorepository that are expected to enroll in eMERGE IV.
		- Type 2 diabetes is rare in the pediatric population.
			* About 1:50,000 children have type 2 diabetes. For children who are considered to be at a high-risk for type 2 diabetes, about 1:500 of children actually have it.
		- It is being considered to not include type 2 diabetes in pediatrics.
* PRSs validated in single populations or not published| Sites
	+ Asthma
		- Expected Completion:
			* Asthma is likely to be available for prospective trial by May 2021.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
			* Asthma is likely to be successful when using the trans-ancestry approach and TAGC as a reference.
		- For Asthma, the eMERGE-GWAS is being targeted for operations.
		- In pediatrics, there are a sufficient number of cases, which is 1324. This will be used as a target to validate the PRS.
		- Using the Bayesian method for validating PRS, European ancestry only has 15 SNPs and has a AUC of 0.61.
		- The CAAPA (Consortium on Asthma Among African-ancestry populations in the Americas), has an odds ratio of 1.19.
		- Pediatrics was isolated and a stronger result was found (OR 3.67, AUC 0.073)
		- The TAGC does not provide pediatric-only results.
		- The odds ratio for all population groups combined is 3.74.
		- The Network is looking for stratified performance across different ancestries and this should be presented.
	+ Atrial Fibrillation
		- Availability of datasets for clinical validation:
			* The majority of GWAS is in European cohorts with some data in other ancestry.
				+ The PRS was calculated in European cohorts.
				+ There is limited data for other ancestries; the group has tested performance of smaller PRS that includes results from multi-ethnic GWAS (84% European) in the eMERGE 1-3 data set.
			* The PRS AUC crude for African ancestry is 0.588 and 0.606 for European ancestry.
			* The adjusted PRS for AF in African ancestry is 0.772 and 0.696 for European ancestry
			* The group plans to continue testing with additional available PRS across all ancestries.
			* There is no GWAS available with a significant number of non-European participants.
				+ This would be ready to go in early May if the Network agrees that there are sufficient non-European results.
			* ACTION ITEM: The PRS workgroup will discuss SNPs on GDA chips on the next call.
			* In the Roseli paper, groups were meta-analyzed.
			* The top 2% PRS is significantly discriminatory but not at the top 5% PRS.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
			* It is very likely the condition will be successful for the prospective cohort. If the PRS is validated across ancestries,, the group will be ready to transfer to Broad in early summer. AF risk is of high clinical utility.
	+ Chronic Kidney Disease
		- Background
			* A study published in 2019 has over one million participants but mostly in Europeans.
			* There is a gene with two risk variants with a large effect on kidney disease that are exclusive to African ancestry.
			* These variants are practically absent in those with European ancestry.
			* The outcome for this condition is CKD stage III or higher.
		- Expected completion
			* There does not seem to be a major barrier to completion and expect to be ready by May/June 2021.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
			* It is very likely to be successful.
		- Availability of datasets for clinical validation
			* GWAS summary statistics and the 1000 Genome Reference Panel was used.
			* UK Biobank and eMERGE 3 were used for validation.
		- Findings
			* The discovery cohort has an OR of over 3 in the upper range of the tail.
			* The African population in the UKBB has a lower OR of 1.67 while other populations have higher AUC.
			* eMERGE III Odds Ratios are higher in Hispanics than the UKBB.
			* The African American population in eMERGE III has an adequate OR of 1.31.
			* A new tran-ethnic score was developed and is currently being refined.
			* The eMERGE 3 dataset performed better than the UKBB due new imputations of a missing G2 allele.
		- What additional resources or data would help with development and validation?
			* UAB completed their phenotyping and BioMe are close to completion with their phenotyping.
		- The APOL1 is equivalent to 2 standard deviations of the GPS in the eMERGE 3 cohort.
		- Considering returning APOL1 separately to people who self-report as African ancestry was discussed.
		- Standard deviations in non-European populations are very large and more information may be needed.
	+ Depression
		- Expected completion:
			* The expected timing of PRS validation completion and publication is April 2021.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
			* Speaking conservatively, it may be unlikely that the condition is successful in the prospective cohort considering timing needs.
		- Availability of datasets for clinical validation:
			* Depression has been studied in very large samples with GWAS studies including almost 850K individuals with 264K cases. The PRS was tested in the MGB biobank with 4K cases in the European ancestry sample.
			* MGB is in the process of validating a new algorithm and is also validating in MGB biobank and possibly in another network.
		- Findings:
			* P+T, LDpred, and PRS-CS were compared with the best performance occurring in PRS-CS using a threshold of the top 2% versus the bottom 98%.
			* There is not good data for Hispanic or Asian populations.
			* The algorithm used in the European ancestry biobank sample was developed at UW which includes a broad spectrum of clinical depression diagnoses. In the dataset used for training, a relatively stringent ICD based phenotype was used which required an inpatient code plus another outpatient code for depression.
		- What additional resources or data would help development and validation?
			* Although using the large cohort works well, the sample is from the Millions Veteran program and with large sex differences, with Depression being twice as common in women, the best available data may not be the best to use in this case.
	+ Hypertension
		- Background/Justification
			* For hypertension, a genetic risk score from the MVP was published a while back, but the decision was made to develop a genetic polygenic risk score PRScs.
		- Expected completion
			* Unclear from slides and discussion.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
			* It is likely if the Network wants a good discriminative score that predicts outcome.
			* It is unlikely if the Network wants a model for predictive power of the PRS.
		- Availability of datasets for clinical validation
			* The size of the resource data available to do this was just under 500,000. The data from BioVU contained almost 68,000 individuals, and the data was reasonably trans-ethnic compared to some other traits.
			* Sixty features were selected from BioVU with 51 phecodes, and the eMERGE model was then applied.
		- Findings
			* In the UK Biobank, more variants were explained for trans-ethnic models concerning BioVU.
			* For hypertension, it is best to combine the systolic and diastolic polygenic scores.
			* The trans-ethnic model is the best in regards to the odds ratio.
		- The model chosen was the 5e-8 from optimization in BioVU.
		- Models are being developed for both systolic and diastolic BP.
		- The interest is to build models that include polygenic scores and other features.
		- Models are built-in BioVU.
		- Elastic net input includes basic covariates + ancestry proportions + 87 phecodes.
		- Summary statistics from GWAS of East Asians are being applied.
		- The PheWas will also be applied to the Synthetic Derivative at VUMC to help capture clinical risk factors.
	+ Lupus
		- Expected completion:
			* Lupus is likely to be validated by May 2021.
		- eMERGE-GWAS was used as a target for the PRS.
		- Multiple ancestries have an AUC over 70%.
		- The group is expecting an additional paper for Asian ancestry to be released in May.
		- African ancestry may be an area of concern.
		- Lupus is a rare disease that has a .1% prevalence, so there may be a large amount of false positives.
		- There is no preventative strategy for Lupus.
		- Out of 20,000 adults, approximately 420 will be defined as high risk. Out of the 420 adults who are defined as high-risk, 400 of them will likely be false positives.
	+ Non-alcoholic Fatty Liver Disease
		- Availability of datasets for clinical validation:
			* The NAFLD PRS was developed with 77 loci significant in trans-ancestry meta-analysis across two dataset: Million Veterans Program (MVP) and BioMe.
			* The phenotype leads presented the MVP 2020 GWAS and PRS validation as well as BioMe NAFLD case and controls.
			* The group was initially more optimistic about the NAFLD phenotype validation but the current results are not meeting expectations.
				+ The performance of models is not predictive for ethnic groups outside of the Hispanic population.
				+ The validation with local samples has not yet been accomplished.
				+ BioMe and eMERGE-3 are available datasets for clinical validation, however there is a limited sample size for validation and limited summary statistics to build a robust trans-ancestry genome-wide PRS.
				+ The lead and co-site sites feel it is unlikely for this condition to be successful in the prospective cohort due to timing and lack of validation.
	+ Stroke
		- Expected completion
			* PRS validation is expected to be completed in 6 months, depending on REGARDS completion.
			* It is unlikely to be completed for the African ancestry but is completed for non-African ancestries.
		- How likely this condition is to be successful for the prospective cohort considering timing needs:
		- Availability of datasets for clinical validation
			* Datasets used are eMERGE III and REGARDS.
		- Findings
			* The tail odds ratio underperforms in the African population but may be resolved if there are more samples to include.
		- What additional resources or data would help with development and validation?
			* A large limitation is the lack of GWAS data in the African population.
			* Capturing downstream compliance might be difficult due to the actionability being similar for several conditions.
		- The PRS method by Abraham et al in 2019 that had nearly 400,000 participants.
		- The meta-PRS combines several PRS using elastic net regression.
		- This method was tested in the eMERGE 3 dataset and performs well in European, Latin, and Asian populations but not well in the African populations.
		- Completion of the PRS validation is expected to be in about 6 months.
		- This PRS is expected to perform well in the non-African populations.
		- Testing in almost 9000 African Americans in REGARDS is still pending.
		- Subtypes of ischemic stroke can be removed if necessary but not typically done.
		- Hemorrhagic strokes are usually removed from the PRS if needed.
* sIRB: Reports, surveys, & pending questions | Digna Velez Edwards (VUMC), Ingrid Holm (BCH), and Wendy Chung (Columbia)
	+ Submission to the sIRB is not quite ready but several important pieces are finished or close to being finished.
	+ There will be an initial submission with subsequent amendments submitted with Spanish translations, completed surveys, return of results process updates, etc.
	+ Based on focus groups, the Network should aim for a very short time period between enrollment and return of results.
	+ Participant progress along the study timeline can be paused at the Recruitment phase if conditions still need validation but may lead to poor participant good will.
	+ Completion of the MeTree survey can be done between 0-3 months after enrollment.
		- MeTree will provide a report after the survey has been completed and entered.
		- MeTree collects a very detailed family history and will require a significant amount of time for the participant to complete.
		- Timing of when MeTree is being completed is important due to the report that is generated after completion, which will likely be before getting the PRS and other GIRA components.
		- There is concern that participants and providers will receive the information contained in the GIRA report in a piecemeal fashion complicating the impact the GIRA has on outcomes.
			* Expectation setting needs to be done early in the process and offering an incentive should be considered.
	+ There are three times for participant compensation: at completion of the pre-RoR survey, GIRA return, and completion of the post-RoR.
		- There might need to be consideration of another compensation point if MeTree is difficult to complete.
	+ Biological samples will be collected, DNA extracted at the site, and shipped off in batches for The Broad.
	+ Invitae will provide sample collection kits that can be used to send the sample. The process for ordering is currently being worked upon on how to coordinate.
	+ Timing of study completion is variable and dependent on how quickly the participant completes their portions and how long it takes to get the PRS and monogenic results.
	+ Consistency on how results will be returned is important while trying to allow for some variability on who returns the results.
	+ A randomized sample of non-high risk participants that get results returned and monitored similar to the at high risk participants is being considered.
	+ There are flexible enrollment strategies by site, but core inclusion and exclusion criteria need to be followed.
	+ Common recruitment materials for all sites will be harmonized to a degree (language, images) but sites will have site-specific information as well.
	+ There are pediatric versions of the surveys under development.
	+ Talking points and FAQs will be available for all RoR outcomes.
	+ Educational infographic style videos are being prepared to introduce information regarding the study to participants.
	+ Common virtual training for study staff in returning results is being planned.
		- Columbia is working on genomic medicine training modules for PRS and might be able to assist.
	+ Most of the sIRB core pieces are complete but will need to finalize a draft of the baseline survey in order to be able to submit.
	+ The consent is in two parts. Part one is the main consent submitted to the sIRB. Part 2 is site-specific and submitted to the IREx portal for each site.
	+ Assents are **not required** by the sIRB but must be uploaded to the IREx for each site.
	+ Invitae consent language will be merged into the master consent.
	+ Invitae will provide eMERGE with a CLIA PDF through the physician ordering portal.
	+ How to handle missing GIRA data needs to be discussed/decided.
		- Should there be a section in the protocol that outlines steps on how to handle missing surveys etic.
		- Is a GIRA useful if only a PRS is generated with other data missing will need to be discussed.
	+ **Action Item:** Surveys need to be finalized prior to submitting the sIRB.
	+ **Action Item**: Timing of when the MeTree survey is completed/entered needs to be decided.
	+ **Action Item:** A decision on when the MeTree report will be returned needs to be made.
	+ MeTree discussion
		- Due to the amount of detail needed, participant MeTree information gathering can start at baseline.
		- Language regarding classifying participant risk needs to be harmonized across the Network to minimize confusion.
		- MeTree does have a Spanish version.
		- Parent/child records are not linked in the pedigree together so a child record will be separate but can be linked on the back end for analysis.
		- In order to get specific risk reports for the parent and child, the survey may need to be completed separately.
		- MeTree might be customizable to include only the eMERGE conditions but it would depend on the programmers.
			* The easy fix is to provide custom instructions on what the participants need to complete for this project so the changes aren’t implemented across the MeTree network.
		- **Action Item:** Site leads will need to decide on which data variables will be collected in MeTree and the surveys to limit duplicate answers.
		- MeTree can be modified to the style of report provided to the participant but the timing of the report hasn’t been modified before.
	+ The current language for risk stratification is “meets study threshold for high risk” and “does not meet study threshold for high risk.”
		- The current phrasing may be confusing to the participant and should be tested in focus groups/interviews soon.
	+ The consent form should have language that explains the cut-offs that indicates high risk or not at high risk and that this isn’t to diagnose a condition but to help with implementing prevention strategies to prevent or identify the disease early.
	+ **Decision**: For not at high risk participants, language should read “*does not meet study threshold for high risk*.”
	+ The outcomes group discussed that they need to have all elements that go into the GIRA in the baseline survey.
	+ Genotyping could continue if there is missing data and it is not realistic to expect 100% complete data.
		- Family history will be critical to be completed because family history for some conditions can put a participant at high risk.
* Network decisions & timeline review | Rex Chisholm (SC Chair, Northwestern)
	+ The eMERGE Network is recruiting from an unbiased general population ages 3-75, and able to consent in English or Spanish. Underserved populations will be emphasized.
	+ Conditions should recommend clinical actions that are pertinent to the care of the participant and process and health outcomes that can be measured within 1-2 years of return.
	+ High-risk GIRA results will be completed by genetic counselors or physicians and initial contact should be conducted by highly trained study staff. Results under the study threshold for high risk will be returned via email, mail, EHR portal, or mass distribution.
	+ High risk and under threshold criteria for high risk will be utilized for return and that will occur one time.
	+ Structured and PDF reports should contain the same data elements.
	+ The PRS reports will be generated at the Broad from DNA extracted from blood or saliva, stratified by age group (between pediatrics and adults), and run on the Global diversity array.
		- An interim report can be returned to the R4 portal for pipeline generation.
	+ Reports will be aimed to clinical providers at each site and not to participants.
	+ The GIRA will integrate monogenic, polygenic, family history, and clinical risk depending on specific conditions.
		- Quantitative risk scores in the GIRA should only be returned to “high risk” participants.
		- Invitae will provide targeted sequencing on the adult participants.
		- MeTree will be utilized to collect family health history data and results will be returned via non-face-to-face mechanisms, unless part of a larger ‘high risk’ GIRA. The phenotype leads will decide what the mechanism of return will be.
	+ Multiple workgroups discussed study design mechanisms and concluded:
		- High monogenic risk will be returned in person (face to face, Zoom, phone).
		- High PRS will be returned in person (face to face, Zoom, phone).
		- High family history risk for patients without high monogenic or polygenic risk is dependent on the phenotype and that may be returned by certified mail or a read receipt in the EHR.
		- All others would be indirect return (not one-on-one engagement with participants) in the EHR for providers and patients will get notifications via the portal or mail.
	+ Going forward, the network would begin to whittle down the list of phenotypes.
	+ Current pending decisions include:
		- PRS validation and condition selection - how to handle return in populations with less validation or reduced odds ratio impact.
		- Confirm monogenic gene list and initiate contracting with Invitae.
		- GIRA format and content which includes structured data for GIRA and PRS reports and determining how and where to assemble GIRA and confirm **no** signature needed.
		- Data flow of reports from vendors to R4 and from R4 to sites including timing of surveys and report generation.
		- Finalize consents and protocol.
	+ A survey will go out to site PIs that will allow sites to make their assessments/recommendations about which phenotypes they believe should move forward or which should be deprioritized.
		- Votes will occur on the PI call on **Thursday, March 18, 2021** to determine which conditions will continue on the clinical validation pathway.
			* All conditions can be included in the sIRB and amended later if some are removed, however shortening the list now does streamline accessory documentation like surveys.
			* Based on today’s slides, sites are being asked if they think the phenotype should continue down the clinical validation pathway for the PRS or if it should be moved to a developmental pathway. In the developmental pathway, sites would still be free to develop and publish data but it would not be part of the prospective GIRA return.
				+ If a phenotype is kept on the clinical pathway, that does not guarantee it is approved but that it will be reassessed when the site is ready to give an official proposal.
		- The goal is to do around 10 phenotypes (potentially one per site) and considerations such as logistics, budget, and experiences should be considered.
	+ All sites should review the prospective cohort timeline at the end of the Steering Committee meeting slides. Recruitment is targeted to begin late summer 2021, after sIRB approval.