Guidelines for Approach Section of CIDR Applications

Requests for CIDR services are made via an X01 application (PAR-14-207). Please use the headings in ***bold,*** *and provide the information described below for the Approach section of your application.*

Please note that the application as a whole should reflect the commitment and involvement of key personnel for the CIDR request, including the genetic analyst. It is important to have the lead genetic analyst directly involved in both the design of the project and in the CIDR application.

Sample Information: Provide a table describing all of the samples to be genotyped or included in the overall analysis. You should include the number of samples, DNA source (tissue), the requested genotyping or sequencing service and information about previous genotyping if applicable. If subjects were collected from multiple sites or represent different population groups, this should be clearly indicated on the table. If data from previously genotyped or sequenced samples are included in the analysis, please also include those in the table. In your supporting documentation, provide evidence that you have access to these data.

Example Table

Sample Set (include reference name, case/control status if relevant)	**Number of Samples**	**DNA Source**	**Service Requested**	**Previous Genotyping or Sequencing**
ABC probands	3,000	blood	MultiEthnic Global array	Illumina Linkage array on all
ABC controls	3,000	blood	MultiEthnic Global array	Illumina Linkage and custom arrays
EFG probands	2,000	NA	None	Illumina Omni2.5 array
ABC probands	200	NA	80X Whole Exome Sequencing, average target coverage 80X	Illumina Linkage array

Provide supporting text that justifies the choice of samples. Include the origin of each group of samples, and details on the previous genotyping and sequencing that was done. Also describe the extraction methods used for each DNA source and the approximate DNA concentrations. Indicate when the samples will be ready to ship to CIDR. Please see: https://cidr.jhmi.edu/ArraysSampleRequirements.html and https://cidr.jhmi.edu/SequencingSampleRequirements.html for sample requirements.

Description of Disease/Trait: Include detailed information about the phenotypic characterization of the subjects and describe any relevant endophenotypes or secondary phenotypes that have been measured. Describe relevant environmental factors that have been measured. Explain characteristics of the trait that

suggest the service requested will provide important information.

Study Population: Describe the study population and the method of selection. For case-control studies, provide inclusion and exclusion criteria for cases and controls. If applicable, describe how cases were identified and sampled, whether controls were matched to cases, and if so, how they were matched and how effective the matching has been. For trios and families define how offspring were selected and the completeness of pedigrees. Indicate whether parentage was confirmed by genotyping. Describe any special features of the population that would enhance its value for the study proposed. For mouse studies, provide reasons for strain choices.

For sequencing studies, explain the basis for selection of individuals for sequencing. For example, if individuals from families were included, explain the basis for the selection of families and choice of specific individuals within those families. Attachment of actual pedigrees with included individuals marked is encouraged. For individuals not from families, give details of why specific cases or controls are selected for inclusion, for example, based on unique aspects of their family history or disease.

Justification of Services Requested: Provide a clear justification for the particular service requested, whether genotyping or sequencing. For GWAS, include a justification for a GWAS study design. Also provide the justification for the array chosen and why other genotyping arrays are less appropriate. For custom SNP projects, include a justification for the choice of service and number of SNPs requested.

If targeted sequencing or custom genotyping is requested to follow-up previous GWAS or linkage findings, describe these findings in detail, including the strength of the evidence for localization, any known replications, and whether the sample to be sequenced is drawn from the original localization sample.

Power and Effect Size: Using power analyses, describe the range of effect sizes detectable by the study. In the analyses address relevant features of the analytic plan, such as genetic models to be tested, the extent of multiple testing, and what significance level would be used for testing; important parameters such as risk allele frequencies; as well as expected patterns of linkage disequilibrium. If the study design requires separate analysis of groups of subjects (e.g., phenotypic classes or ancestry groups) provide power analyses for each category. If there is a plan to test for gene-gene interaction effects, address power for that particular design (e.g., testing gene-gene interactions separately from main effects and/or jointly).

When describing the range of detectable effect sizes, include a brief discussion of the likelihood that there are relevant loci within the detectable range. Be sure to put this discussion in the context of trait heritability and results from previous studies that attempted to localize risk loci. If the plan is to incorporate results from other studies in the analysis to increase study power, explain how this will be accomplished.

For targeted sequencing provide power calculations that include correction for the number of markers estimated in each region. If you plan to sequence a discovery set and then genotype the identified markers via custom genotyping for validation, provide association power based on the size of the validation set.

Data Analyses:
Provide a thorough and detailed plan for data analyses, including statistical analyses that are planned. Examples of the expected elements in this section include the analytical approaches to be used and their justification; plans for quality control analyses of genotypic data; methods to account for genotypic and

phenotypic uncertainty, if relevant (e.g., diagnostic uncertainty would be an example of phenotypic uncertainty); plan to control for possible confounding of genotype effects on phenotype due to ancestry; how false positive rates will be controlled in light of multiple testing.

For sequencing studies, explain plans for analysis of the resulting data. For example, filtering SNPs to discard SNPs observed in public data resources.

Summarize the team's expertise and experience with the kind of analysis proposed. Be explicit about who will lead the analysis effort and document their availability to contribute the effort that will be required.

Data Management: Provide a description of the institutional computing resources that will be available to this study. Describe how the data are to be managed, such as type of data base, who will maintain and update it, and who will have access to it. Highlight experience with data management for large data sets, such as those to be produced by the proposed project. Describe strategies for data security.

Plans for the Next Phase: Carefully describe plans for confirmation and/or functional validation for elements detected in the requested sequencing. If collaborations have been established for follow up, include these letters of collaboration.

If you have questions, please contact Barbara Thomas at: barbara.thomas@nih.gov

Updated: October 14, 2016