Title of proposed program: The **Net**work-based **T**axonomy of **D**isease (NetTD)

What is the major obstacle/challenge/opportunity that the Common Fund should address?

The major recommendation from the "Toward Precision Medicine" workshop[1] by the National Academy of Sciences is to develop a Knowledge Network of Disease for a new taxonomy based upon diseases' molecular bases in addition to clinically observable phenotypes. Conceptually this network will provide the foundation for personalized treatments because it will allow revealing the underlying molecular pathways that connect each currently defined disease as a spectrum of subphenotypes consisting of pathways of molecules and genes. For instance, asthma patients may be divided into multiple groups based on their molecular and genomic characterization, with one extreme overlapping with patients suffering from other allergic diseases and the other overlapping with a subset of chronic bronchitis patients. This molecular disease network concept has been tested by investigators, including a few funded by NHLBI's MAPGen[2] project, which has shown that diseases are interconnected by shared genes and biological pathways. Networks will eventually include the data from many on-going projects (e.g., dbGaP, GEO, LINCS, GTEx, ENCODE, EHRs) and the up-coming tools and data common from BD2K. However, in the current disease network the information of intermediate or subclinical phenotypes are missing for most complex diseases, presenting a major obstacle to identifying subphenotypes of disease and developing a new disease taxonomy based on individual patients' molecular makeup. This initiative, Network-based Taxonomy of Disease (NetTD), aims to build a new type of disease network focusing on intermediate, preclinical, and subclinical phenotypes that are more closely linked to the underlying molecular pathophysiology of a target disease.

The current challenge is that, for most diseases, intermediate, or preclinical/subclinical phenotypes are either not identified or not well-defined. The NetTD aims to develop technologies that will facilitate the identification and/or characterization of intermediate and subclinical phenotypes that can provide interconnections between complex diseases or between healthy and disease states. The project will have two phases, with the Phase I of the initiative developing new technologies and tools that can be subsequently used to identify and/or characterize intermediate and subclinical phenotypes, and the Phase II providing the experimental data that will leverage the use of the Phase I technologies and tools to build the framework of subphenotype modules. While the *clinical focus* of the network will be more on pre-disease or cross-disease intermediate phenotypes or endophenotypes, the *biological focus* will be on the fundamental biological processes that may cross multiple diseases and/or organs (e.g., inflammation and fibrosis).

What would the goals of the program be?

The central goal of this initiative is to develop NetTD, a diseases network focusing on intermediate and subclinical phenotypes of complex diseases that will provide molecular bases to discover new disease subtypes, novel biomarkers for early diagnosis, more accurate prognosis, and therapeutic targets, and eventually lead to the development of a new taxonomy of disease. NetTD will develop technologies and tools needed to create a framework for the network of intermediate and subclinical phenotypes of complex diseases. Both tools and the network framework are generalizable and easy to use by the scientific community to identify and characterize the intermediate and subclinical phenotypes of their disease of interest.

Why is a trans-NIH strategy needed to achieve these goals?

It is increasingly evident that many diseases transcend the barrier of a single organ; therefore, a trans-NIH strategy that integrates all ICs might help to elucidate how the molecular pathways that may

be common between diseases lead to clinical phenotypes. By considering diseases as a group of molecularly defined subphenotypes, NetTD will break down these research silos. To do this, NetTD must be built with data from all diseases and all organ systems; therefore the common fund would be the most appropriate funding mechanism to support this initiative. The synergy with all ICs will be further strengthened during Phase II, when the initiative expects cross-disciplinary experts to test the tools and resources under development. This initiative is unique in its vision of creating tools and data resources into an integrated framework that will enable the research community to discover and validate intermediate and subclinical phenotypes of any disease.

What initiatives might form the strategic plan for this topic?

The NetTD initiative will have two phases. Phase I (4-year and $4M/year) will ask applicants to leverage existing clinical and molecular data and tools (e.g., dbGaP, GEO, EHRs, interactome, GRN, LINCS, GTEx, BD2K) to identify potential intermediate, preclinical, and subclinical phenotypes along major biologicals processes and develop high throughput technologies to measure and characterize these phenotypes. For example, such technology could be a genotyping platform designed to identify genetic mutations causing loss- or gain-of-interactions between genes, instead of focusing traditionally on mutations of loss- or gain-of-function in individual genes[3], or a multiplex array platform to profile a small number of proteins, genes, or metabolites representing a major biological process that correspond to intermediate phenotypes of one or many diseases. Phase II may last 4 years ($10M per year) and will support multiple centers for the generation of new data and validation of predictions regarding intermediate and subclinical phenotypes made in Phase I. Outside investigators will also be invited to test usability of the tools and network structures. Centers will build the framework of the network containing intermediate and subclinical phenotypes under a few major biological processes. In addition, Phase II will support a knowledge base center to integrate all data and information generated from other centers into a cohesive database with a user-friendly web-portal to serve the needs of scientific community.

If a Common Fund program on this topic achieved its objectives, what would be the impact?

Implementation of NetTD will create a framework that aims to integrate evolving research on the molecular bases of diseases with clinical data, and will break the artificial boundaries created by decades of organ- or single-disease centric research.

NetTD will eventually drive the development of a new and more precise disease taxonomy that defines disease based on underlying molecular interactions. This more precise classification will provide molecular evidence to discover intermediate phenotypes and biomarkers for early diagnosis, more accurate prognosis, or novel therapeutic targets, and to map existing drugs to different diseases for drug repurposing.

1. http://www.ncbi.nlm.nih.gov/books/NBK91503/
2. http://www.mapgenprogram.org/
3. Sahni N, Yi S, Zhong Q, Jailkhani N, Charloteaux B, Cusick ME, Vidal M. Edgotype: a fundamental link between genotype and phenotype. Curr Opin Genet Dev. 2013 Dec;23(6):649-57.