**This document is** [**goo.gl/FZzGuJ**](http://goo.gl/FZzGuJ)

Crisp key points

\* This is the bottleneck for much of genomics, so determining function is critical

-- this is a particularly opportune (variant and genome interpretation, Gene editing [crispr/cas9], synthetic biology, functional genomics in cellular phenotyping: only in the last few years we've had epigenomics & cellular phenotyping; better models [mouse v primates?] [ or should we find ideal diseases to study?])

-- Opportunity for tech standardization and resource sharing

-- Partnering with other IC's / CF for domain-specific problems

\* both are important: top-down (large resources) v bottom up (disease specific variants and personal functional genomics)

- organismal v molecular func.

- combinatorics (multi-variant) & blending

- systems level

\* sweet spot (C) : tech v investment

- what are the correct models , domain experts

- what tech is missing

\* More to functional genomics than variant characterization

--cell type characterization & epigenomics

--healthy vs. pre-disease vs. disease state

--systems level understanding

\* Projects should be foundational for resources and technologies

\* Value in longitudinal studies

1) Combinatorics

- how do you integrate mult alleles to come up with risk (combinatorics)

- Do we need faithful models ?

- We don't understand modifiers.... \

Jim: are gene effects additive ?

Debbie: how does this relate to environmental effects?

2) Functional assays up front

Snyder: Adding functional experiments (e.g. RNA-Seq, methylome, other omes) early into variant discovery for priorization.

- prioritizing variants ? the dichotomy: start with variants or prioritize them . Where should we start?

- - - What of these approaches scale ?

3) "Non-genetics"

Are there large-scale (non allele oriented assays)?

- environmental

- molecular phenotyping that may not be due to allelic variation

What is a cell state ? How to

Carlos: Application of func genomics to other problems in biology ... everything "a genetic" . How many cell types are there?

Carlos: Can we use sequencing as a monitor of cellular state ?

Mark: eg RNAseq after knock-downs. How would long-reads relate to this ?

4) What technologies should we develop to tackle the above ?

- - Gabor: single-cell measurements

- - Debbie: mosaicism

Discussion related to ( 3) How do we validate functional effects of variants in genomics?

Ross: Is NHGRI good a for definitive assignment of function? Shouldn't this be a collaboration with the domain experts? NHGRI can do characterization of function (Aviv’s little “v” validation).

Toby B: NHGRI can partner with other institutes to bring in domain experts.

Rick: But NHGRI has an important role to play.

\*\* IDEAL DISEASES for this

Debbie: To get at function in a cellular way for diseases that are not easily amenable to be studied in a model organism

Len: what would be the ideal disease that avoid above?

 metabolic disease was suggested

Debbie: do we get variants for a disease where we can't test...

\*\* LEVELS OF VALIDATION: V vs v

Aviv: types of validation

"v"alidation : fig. out mol. phenotype (context matters less). Interpretation is weaker . It is more generic

"V"alidation: to the right cell, right condition...

Rick: how do you do this?

Mark: Little v is characterization

Ross: definitive function

Mike: importance of recurrence

\*\* cellular CONTEXT

Carol: We need to validate the cellular context . Cellular context is very important !

Aviv: We should test in contexts that mimic context that it was discovered in...

○ Is it possible to validate thousands (or millions) of

assertions about the genome with one or two small-scale validation

experiments?

○ Is it possible to do validation at a very large scale? Is

medium-scale validation possible and useful? How to think about the

cost of this?

○ How do we incorporate the results of validation into

quantitative error estimates for the functional assertions being made?

● 4) Any other topics that individuals want to bring up will be

discussed in the last 30 minutes. We could poll for these topics at

the beginning.

Discussion related to (2) How do we inter-relate function & variants on a large scale?

\*\* CONSERVATION

Debbie: building an interpretome....

Debbie: can we test everything with a CADD/conservation score over 20

Jay: you don't get this purely with conservation

\*\* DOMAIN EXPERTS

Aviv: categories of tools: readout tools, perturbation tools, bioassay "tools" (most specific)

Elise: Can NHGRI facilitate the many indiv. investigators, by a tool or resource ?

Aviv: Define function with **domain experts**

Mike: we need high-throughput assays

Rick: how do we turn this into a resource ?

\*\* SIMPLICITY

Debbie: making something useable (ENCODE as an ex.). Can we make this simple? Let's not make new resources that are not integrated?

Carlos: should this be done by consortia, not indiv. investigators ? Perhaps for coherence ?

\*\* CELLTYPE & REGION CATALOGS

Pazin: how important is it to have the right cell type? How do we know we've done enough cell types?

Len: can we use WGS to guide this

Joe: we need a fuller catalog of reg. variants & regions

Ross: Individual labs are also doing high quality work in functional genomics, and they have domain experts who can insure relevant cell types are examined. These data should be integrated into the larger annotation efforts. Maintain quality but expand the scope of data acquired.

\*\* Mouse & Models Portfolio

Jim: NHGRI can develop the tools to exploit many different models

Carlos: perhaps Common Fund can be useful here (eg Gtex)?

Joe: emphasizes the importance of organoids. Develop a Common Fund initiative to develop iPS cells -> organoids, model systems for NHGRI and *many* other institutes

Carol: mouse models go beyond KOs....

Aviv: we need a portfolio of models (types of cells, IPS, organoids...). What in genomics fits with each type of model?

\*\* IPS

Debbie: should we have a seq. res. of IPS cells?

Joe: happening in CA....

Mike: eventually we'll all have IPS cells of each person... where we'll go.

\*\* GUIDING EXPTS

Len: We will have more impact if we test variants that have seen to have importance in real world scenarios. Guided by variants found in nature .

How do we get access to patient material?

\*\* USING AVAILABLE FUNCTIONAL INFORMATION FOR INFERENCE IN THE CLINIC

Gabor: What integrative tools need to be developed? Directions for functional prediction?

○ Is this best done by individual investigators pooling together

individual results into a database or is it best done by large-scale,

highly standardized experiments? What is the role of special big data

database architectures for aggregating the knowledge of many

functional assays?

○ Is it more effective to follow up on the many

disease-associated variants uncovered by sequencing in great detail

rather than doing broad genome-wide functional characterization

beforehand?

○ Are there ways for new high-throughput technologies and

computational approaches to significantly help with this endeavor?

○ How do we prioritize those experiments and assays that provide

more functional information compared to others? Is there a particular

way of assessing the information in particular experiments?

Discussion related to (1) What is function in genomics & how do we use it to determine

the effect of variants?

\*\* SWEET SPOT

Carlos: the sweet spot!! impact/resources

\*\* PHENO COMPLICATIONS: BLENDING

Jim: issues with blended phenotypes from simultaneously having mult. conditions. In about 2000 genomes sequenced from the clinic, about 1 in 20 have 2 Mendelian alleles...

\*\* GENE BUNDLES

Mark: context is very important . What is the genomic approach to understanding biology ? Creating "bundles" of genes -- eg from networks.

Ross: there's a hubris in prioritizing. Perhaps the larger community can help with this. Propose that the priority list of loci and diseases be established by proposal, e.g. could require provisional commitment for support from other disease-based Institutes, collaboration with researchers expert in that area.

Debbie: there will always be modifiers. We need to think about this across the....

\*\*\* TECH (CRISPR)

Carlos: we shouldn't limit ourselves to just characterizing the functional impact of each base in the genome. How should we use genome-scale approaches in other contexts? NHGRI is in a very good position to develop this technology.

Debbie disagrees:

Aviv: feels NHGRI is definitely the place to develop these technologies.

Carol : We are at pt. in time when genome editing technologies can be scaled (eg crispr). This has great applicability to Mendelian centers. We do need model organisms here - YES!

Debbie: how do we characterize the function of each of the Mendelian variants, at scale ?

Snyder: the spectrum of things we can do: what every base does (a resource) to deep characterization of individuals ? Different technologies applicable to these

\*\*\* WHAT IS FUNC?

Aviv: maybe function should be called phenotype

\*\*Ross: medical applications are important. Starting with existing variants and providing our best estimate of likelihood of phenotypic consequence is a clear goal, and you can develop metrics of progress toward that goal

 learn from Mendelian loci, but extrapolate/migrate to ~ whole genome

Kelly: is function phenotype ? Is this inherited ?

\*\* IMPORTANCE OF MED REL. & GOALS

Toby: We need to tie basic science to med. applications & personal genomics

Carlos: we need to bring in clinical DBs into this.

Aviv: What is the overall goal of this Breakout Group?

- basic science: “an enhanced genetic code”?

- medical discoveries

- medical application: “interpret variants already associated with (clinical) phenotype

Jim: it might be important to see if all the known variants have an effect on the function

● 1) What is function in genomics & how do we use it to determine

the effect of variants?

○ What are the different aspects of function and why is it hard

to study? For instance, molecular (or biochemical) function vs

cellular role vs organismal phenotype.

○ What are the problems in defining function? Is it meaningful to

localize a function to a single place on the genome so it can be

affected by a single variant? How should one think about the

functional effect of large block variants?

○ Is it possible to quantitatively systematize some aspects of

function so that they can be precisely related and correlated with

genomic variants? In particular, what are the paradigms available to

inter-relate function with variants (eg QTLs & allelic effects and

phenotypes resulting from a single disruption)?

**Additional Topics Proposed:**

- how do you integrate mult alleles to come up with risk (combinatorics)

- Do we need faithful models ?

- We don't understand modifiers....

Snyder: Adding functional experiments (e.g. RNA-Seq, methylome, other omes) early into variant discovery for priorization.

- prioritizing variants ? the dichotomy: start with variants or prioritize them . Where should we start?

- - - What of these approaches scale ?

Are there large-scale (non allele oriented assays)?

- environmental

- molecular phenotyping that may not be due to allelic variation

[This is the bottleneck for much of genomics, so determining function is critical]

What technologies should we develop to tackle the above ?

- single cell measurements

Agenda/Outline for Breakout Session (“Integrating functional genomics

with DNA sequence variants”) – Rick Myers’s edits 7/16/14 - Mark's

edits 3nd 7/21/14

Who we think is here:

Mark Adams

Toby Bloom

James Broach

Carol Bult

Carlos Bustamante

Deborah Colantuoni

Joseph Ecker

Elise Feingold

Kelly Frazer

Mark Gerstein

Ross Hardison

Chanita Hughes-Halbert

?? Stephen Kingsmore

James Lupski

Gabor Marth

Richard Myers

Deborah Nickerson

Mike Pazin

Len Pennacchio

Ulrike Peters

Aviv Regev

Michael Snyder

Simona Volpi

Peter Good

● 2) How do we inter-relate function & variants on a large scale?

○ Is this best done by individual investigators pooling together

individual results into a database or is it best done by large-scale,

highly standardized experiments? What is the role of special big data

database architectures for aggregating the knowledge of many

functional assays?

○ Is it more effective to follow up on the many

disease-associated variants uncovered by sequencing in great detail

rather than doing broad genome-wide functional characterization

beforehand?

○ Are there ways for new high-throughput technologies and

computational approaches to significantly help with this endeavor?

○ How do we prioritize those experiments and assays that provide

more functional information compared to others? Is there a particular

way of assessing the information in particular experiments?

● 3) How do we validate functional effects of variants in genomics?

○ Is it possible to validate thousands (or millions) of

assertions about the genome with one or two small-scale validation

experiments?

○ Is it possible to do validation at a very large scale? Is

medium-scale validation possible and useful? How to think about the

cost of this?

○ How do we incorporate the results of validation into

quantitative error estimates for the functional assertions being made?

● 4) Any other topics that individuals want to bring up will be

discussed in the last 30 minutes. We could poll for these topics at

the beginning.