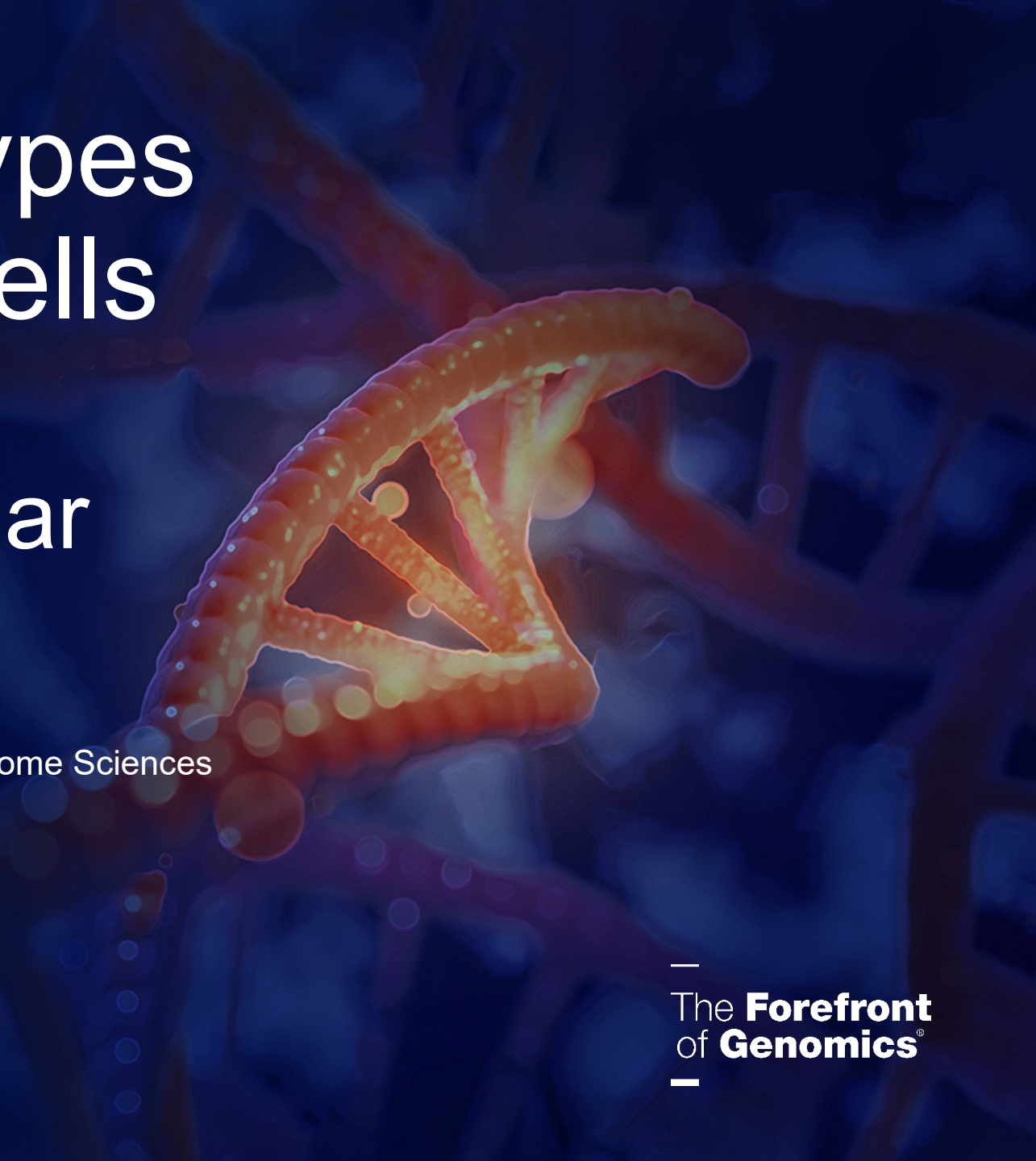
Molecular Phenotypes of Null Alleles in Cells (MorPhiC)
Pre-Application Webinar



Adam Felsenfeld, Colin Fletcher, and Ajay Pillai

National Human Genome Research Institute, Division of Genome Sciences

September 2021

National Human Genome Research Institute

The Forefront of Genomics®

Relevant Links

- **FOAs**:
  - RFA-HG-21-029: Data Production Research and Development Centers (UM1 Clinical trials not allowed)
  - RFA-HG-21-030: Data Analysis and Validation Centers (U01 Clinical trials not allowed)
  - RFA-HG-21-031: Data Resource and Administrative Coordinating Center (U24 Clinical trials not allowed)

- **Email Contact**: morphicprogram@nih.gov

- **FAQ Link**: genome.gov/event-calendar/MorPhiC-pre-application-webinar/FAQ

- **Webinar Link**: genome.gov/event-calendar/MorPhiC-pre-application-webinar

- **Program Webpage**: genome.gov/research-funding/Funded-Programs-Projects/Molecular-Phenotypes-of-Null-Alleles-in-Cells

- **2020 NHGRI Strategic Vision**: genome.gov/2020SV

Format

Overall Introduction (Adam Felsenfeld) (10 min)

RFA-HG-21-029: Data Production Research and Development Centers (Adam Felsenfeld) (10 min+30 mins questions)

RFA-HG-21-030: Data Analysis and Validation Centers (Ajay Pillai) (10 min +20 min questions)

RFA-HG-21-031: Data Resource and Administrative Coordinating Center (Colin Fletcher) (10 min + 20 min questions)

NIH
NHGRI

To Note

- This call will be recorded and posted to the NHGRI MorPhiC Web pages.
- Your questions may be rendered into general FAQ's, with our answers, that will be linked to the MorPhiC Web pages.
- You do not have to identify yourself to ask a question.
- Please ask questions in the Q&A.

NIH
NHGRI

Part 1: Overview

MorPhiC

Long-term (more than 5 years): Develop a catalog of molecular and cellular phenotypes for null alleles in human, across the genome.

IDEAL:

- Consistent (i.e., standardized, well characterized assays)

- Null (or strong l.o.f.) alleles/KOs

- Informative M&C phenotypes; in multicellular systems

- “All” genes

NIH
NHGRI

Why?

- Lack of human KOs; there are mouse and other KOs, but not molecular/cellular phenotypes.
- Strong alleles are useful for interpreting other alleles (incl. noncoding).
- Resource for insight into pathways. Complement to cis-reg initiatives (e.g., IGVF)
- Collection of disease models
- Others…

Purpose of Phase 1

Understand main barriers to the “long term goal”, by getting started at scale of ~1000 loci/5 years

- **Criteria for selecting genes to learn lessons about doing this genome-wide**
- **Optimize making alleles**
- **Selecting cellular systems and assays for informativeness, generalizability, and scalability (tradeoffs!)**
- **Understand scale (costs, throughput)**
- **Raise and address challenges (pleiotropy, cell non-autonomy, compensation, variability)**
- **Understand/improve value of data (management, dissemination, quality, reproducibility, validation, variability, “use cases”, interoperability, etc.) and feed back into data production/design**

Phase 1: To inform feasibility, value, and design of Phase 2

MorPhiC Structure

Three components:

- Data production
- Data analysis and validation
- Data resource and admin coordination

NIH
NHGRI

MorPhiC Overall: General advice (1)

- Cooperative Agreements
  - Substantial NHGRI program management
  - Collaborative tasks (e.g., sample prioritization, QC and data format discussions, data flow between the three component, etc.). Read the FOA *Terms & Conditions* for how this will be managed
  - Flexibility to set, and adjust, milestones (needed in a complex program)
  - There will be a “kick-off” meeting after grants are funded to establish consortium
- Letters of Intent– not required, but encouraged (Sept 15).

MorPhiC Overall: General advice (2)

- Always read the Review Criteria section of any FOA. This is what the reviewers will use to evaluate.
- Please read the instructions to applicants for the “Research Plan” sections.
- FOA’s have a separate Resource Sharing section. *Will be considered in score.*
- Please read the section on “Review and Selection”. It lists criteria that NHGRI may apply in selecting among well-scored applications.
- Read the Budget section (minimum time commitments; consortium meetings).
- Choose letters of support judiciously.

MorPhiC Overall: Diversity & Funding

NHGRI especially encourages applications from

- investigators from demographic groups or institutions that are generally underrepresented in genomic science
- new investigators
- experienced investigators who are new to genomic science
- investigators that have not previously participated in a NHGRI consortium or program

NIH
NHGRI

MorPhiC Overall: FAQ's

Please look out for these in the next few days on the website.

It will be updated.

NIH
NHGRI

Part 2: RFA HG-21-029

Data Production Research and Development Centers

RFA HG-21-029: Outline

- Data Production R&D. Limited scale: 1000 protein coding genes across program
  - Identify genes
  - Make mutant lines
  - Choose appropriate cellular systems
  - Choose and carry out assays
- Focus on generalizability towards addressing the “Main Barriers”

Phase 1 data production needs to answer: Is it feasible to scale this to all human genes? Is it worth doing? How to design?

RFA HG-21-029: Grantees

- Lead the consortium prioritization of genes
- Produce alleles in the chosen cellular system
- Generate data from high-throughput cellular and molecular assays in *in vitro* human system(s). Multicellular complex systems preferred. Molecular and cellular assays considered based on the utility of information, and scale.
- Ensure comparability and reproducibility; metrics and quality standards; standardize allele and assay validations.
- Share data, best practices, cells, protocols, methods, software, etc.
- Collaborate with other consortia or projects developing complementary data sets.
- Develop an approach that will be informative for how we can eventually generalize and scale.

RFA HG-21-029: Responsiveness

- Approaches must be generalizable across multiple cell types and multiple phenotypes
- Not just one disease; not just one cell type (see FAQ's)
- Not cis-reg variation
- Application and demonstration (not tech dev)
- Each application must propose integrated effort (locus selection through molecular and cellular assays; i.e., not just mutagenesis)

NIH
NHGRI

RFA HG-21-029: “Main barriers” (see RFA)

- Allele generation and validation. Do we need alternatives to true null alleles in some cases?
- Prioritize the genes; lessons across the genome and organism.
- What human multicellular systems will be scalable, reproducible, and informative? Are organoids and similar systems currently sufficiently reproducible?
- What are the best assays for measuring molecular and cellular phenotypes? Are assays specific and informative? How do we maximize how informative the assays are? Can assays be made more generalizable?
- How address technical and interpretation issues due to underlying biology, such as: cell- or tissue-type specificity, pleiotropy, cell lethality, haploinsufficiency, cell non-autonomy, natural biological variability?
- What is the ultimate utility of the data for various purposes?
- What quality and consistency of data is needed? What is the best way to manage and organize the data and present it to the community? What is needed for these data to be interoperable and highly useful in combination with other data types?
- What are the major bottlenecks and cost drivers? Where can costs be reduced?

- **You are not limited to what we said the “Main barriers” are. If you think there are ones we overlooked, please raise and address them.**

RFA HG-21-029: Consortium responsibilities

- Develop “final” gene priorities
- Develop quality metrics; data formats; policies
- Share data, plus other products
- QC– same genes through multiple assays possible?
- Characterize variability (experimental and biological)
- Help develop use cases for data

NIH
NHGRI

RFA HG-21-029: Application format

- Please follow instructions in the FOA about the Research Plan/Research Strategy sections.
- Separate Resource Sharing section. (Considered in score.)

RFA HG-21-029: General notes

- Attention needs to be given to both the biology and the potential for scale
- Consortium needs to collaborate. Provide your opinions and justifications about key points, but be flexible

RFA HG-21-029: Some FAQs

1. 1000 genes each, or across the consortium? A: Across consortium, but count on some overlap for e.g., QC.

2. Given that final gene priorities will be decided after funding, how much detail in application? A. Please propose a justification and a rationale, probably with examples. Plus be flexible.

3. Do you really mean null genetic alleles only? A. Should generally be equivalent, justified by e.g., stability/reproducibility/strong effect and or use for addressing e.g., cell lethality etc.

RFA HG-21-029: More FAQs

4. May I propose work on genes affecting a single tissue/disease/pathway/class of genes? A. The key is showing how the approach would be generalizable. E.g.: Applicable to multiple cell types? Learn lessons across multiple classes of genes/proteins? Other?

5. How many samples/assays/cell types? A. See “tradeoffs” and “generalizability” above. Probably multiple, but not that simple.

6. Sample diversity. How important? A. Yes, this is important, but probably underpowered. Need to learn parameters of biological variation to think about better designs.

RFA HG-21-029: More FAQs

7. How much can I spend on tech dev? A. New tech dev is not part of this program. But optimization (e.g., of assays) to scale, or adoption of new tech, would be OK

8. May I use non-human systems? A. OK to use for validation of findings in human cells. Limit 10% of budget. Can you use KOMP or other external data?

9. FOA states: “complex multicellular systems preferred”. Can I propose single cell-type cultures? A. Yes, may be more appropriate/controllable in some cases. Plusses: feasibility for some cell types; cost/throughput; interpretability. Justify in context with rest of FOA.

RFA HG-21-029

Questions?

Part 3: RFA HG-21-030

Data Analysis and Validation Centers

Ajay Pillai

RFA HG-21-030: Outline

Primary goals

- data variability is controllable,
- data is useful to understand basic biological processes,
- data is interpretable for undertaking future hypothesis-driven science by the community.

Projects high potential to illuminate strengths and weaknesses

Community utilization & feedback

RFA HG-21-030: Responsiveness

The FOA has a list of non-responsive criteria:

- Wet-lab data generation
- Do not propose to use MorPhiC data
- Do not address collaborations within Consortia
- Do not have a data sharing plan.

RFA HG-21-030: Key factors

Year 1 is special:

- No MorPhiC data so ‘bring your own/public data relevant to addressing overall aims of MorPhiC Phase 1’
- E.g., address the challenges described in the next slide

RFA HG-21-030: Challenges

Integrating data: between MorPhiC labs and ‘related datasets’

- Identify & Correct technical bias
- Batch correction
- Etc

Sufficient metadata reflecting biology & data generation.

How can small labs use the data and models for downstream experiments?

NIH
NHGRI

RFA HG-21-030: Consortium responsibilities

Algorithmic approaches to ‘selecting the 1000 genes’

How good/useful is the data?

Metadata and APIs and data access

RFA HG-21-030: Application & Review

Please follow instructions in the FOA about the Research Plan/Research Strategy sections.

Budget: 300K direct cost/yr (max); 5 years

Separate Data Sharing section & it is reviewed.

Specific review criteria in the FOA.

RFA HG-21-030

Questions?

Part 4: RFA HG-21-031 Data Resource and Administrative Coordinating Center

RFA-HG-21-031: Outline

The DRACC has 5 tasks:

- Provide a data resource
- Collaborate with DAV Centers
- Analyze, annotate, and disseminate Morphic data
- Integrate external data/information
- Serve as an administrative and coordination center

RFA-HG-21-031

- Data Resource
  - Receive, wrangle, and QC primary data
    - Establish efficient upload procedures
  - Develop a database for storage
    - Adhere to FAIR principles
  - Make data available for consortium and community use
    - Support independent analysis projects

RFA-HG-21-031

- Collaborate with DAV Centers

- Analyze, annotate, disseminate Morphic data
  - Develop methods to identify “mutant” phenotypes
  - Use ontologies for annotation
  - Provide methods for image analysis of 2D and 3D data
  - Provide web portal and APIs

RFA-HG-21-031

- Data Integration
  - Identify complementary information resources
  - Implement interoperability with external datasets
  - Make data available for consortium and community use

- Administrative and Coordinating Center
  - Provide methods/platforms to enable communication
  - Facilitate interactions within the consortium
  - Track consortium activities, experiments, and data
  - Lead outreach efforts to promote consortium resources

RFA-HG-21-031: Application & Review

- Budget: $1.5M total cost/yr; 5 years
- Please follow instructions in the PHS 398 Research Plan: Research Strategy section
  - Note the Resource Sharing Plan instructions
- Check the Application Review Criteria
  - Additional review criteria ‘Specific to this FOA’
- Check the Award Administration Section
  - Cooperative Agreement Terms and Conditions

RFA HG-21-031

Questions?


The Forefront
of Genomics®