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

Ajay Pillai, Adam Felsenfeld, Colin Fletcher, & Riley Wilson

National Human Genome Research Institute, Division of Genome Sciences

September 2022

National Human Genome Research Institute

The Forefront of Genomics®

Relevant Links

Current FOA: RFA-HG-22-019: Molecular Phenotypes of Null Alleles in Cells (MorPhiC) Phase I: Data Analysis and Validation Centers grants.nih.gov/grants/guide/rfa-files/RFA-HG-22-019.html

Email Contact: morphicprogram@nih.gov

Webinar Link: genome.gov/event-calendar/MorPHiC-data-analysis-validation-centers-pre-application-webinar

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

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

2020 NHGRI Strategic Vision: genome.gov/2020SV

Past FOAs:

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

NIH NHGRI

Format

Overall Introduction (Adam Felsenfeld) (10 min)

RFA-HG-22-019: Data Analysis and Validation Centers (Ajay Pillai) (15 min)

Questions: 30 mins

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: 4 awards by Oct 2022
- Data resource and admin coordination: 1 award by Oct 2022
- **Data analysis and validation—this FOA**

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 (Oct 1).

NIH NHGRI

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 month on the website genome.gov/event-calendar/MorPhiC-pre-application-webinar/FAQ

It will be updated.

NIH
NHGRI

Part 2: RFA HG-22-019

Data Analysis & Validation Centers

RFA HG-22-019: 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 with high potential to illuminate strengths and weaknesses

Community utilization & feedback

NIH
NHGRI

RFA HG-22-019: 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-22-019: Challenges

Integrating data: between MorPhiC labs and ‘related datasets’

- Identify & Correct technical bias
- Exptl design
- Effect of KO vs background changes
- Constructing representative GRNs
- Data integration: Multiome measurements

Sufficient metadata reflecting biology & data generation.

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

NIH NHGRI

RFA HG-22-019: Consortium responsibilities

How good/useful is the data?

Metadata and APIs and data access

Merits of common pipelines (leadership role)

RFA HG-22-019: Application & Review

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

Budget: 350K direct cost/yr (max); 5 years

Separate Data Sharing section & it is reviewed.

Specific review criteria in the FOA.

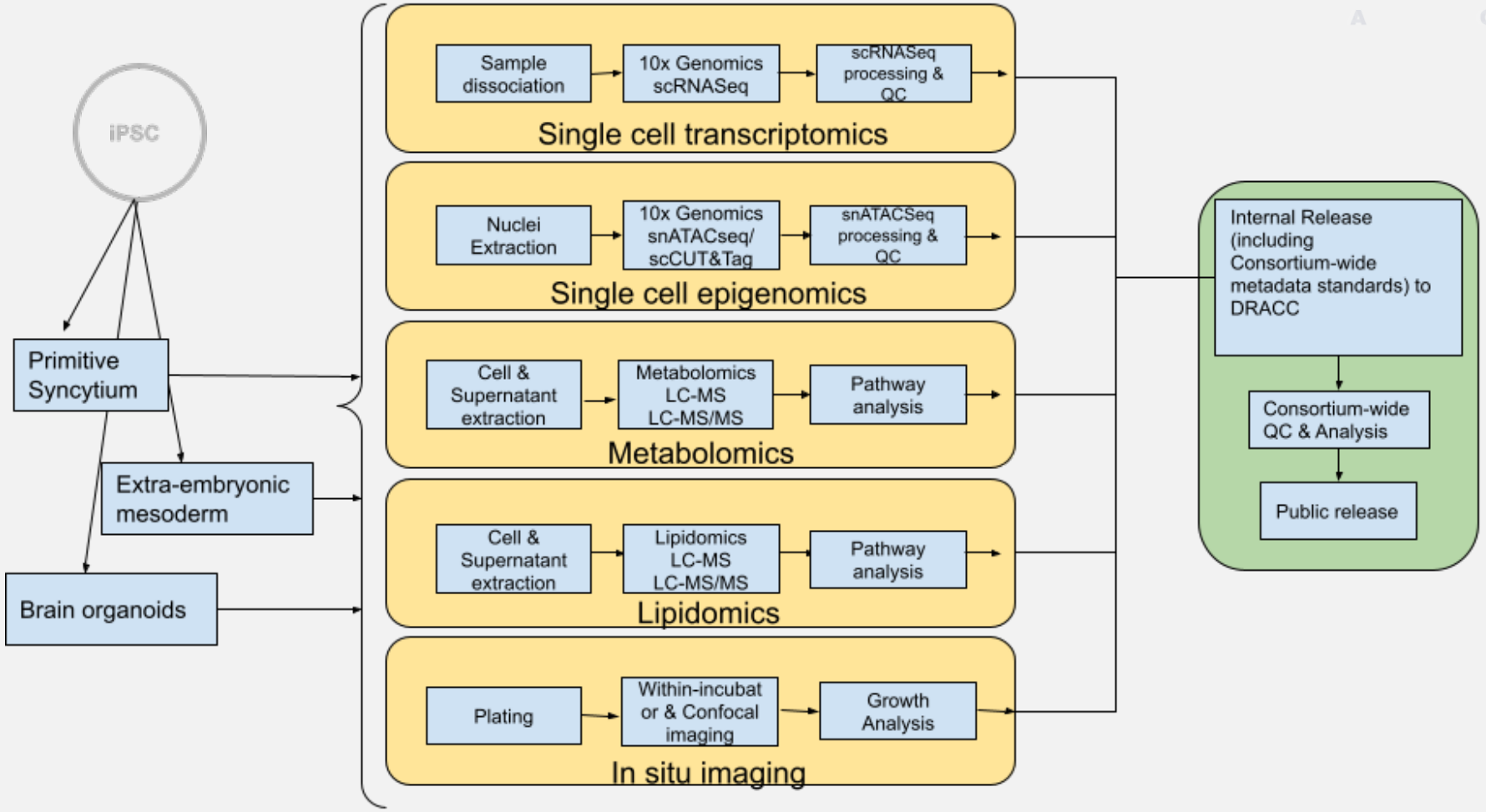
Due Date: Nov 1, 2022

NIH
NHGRI

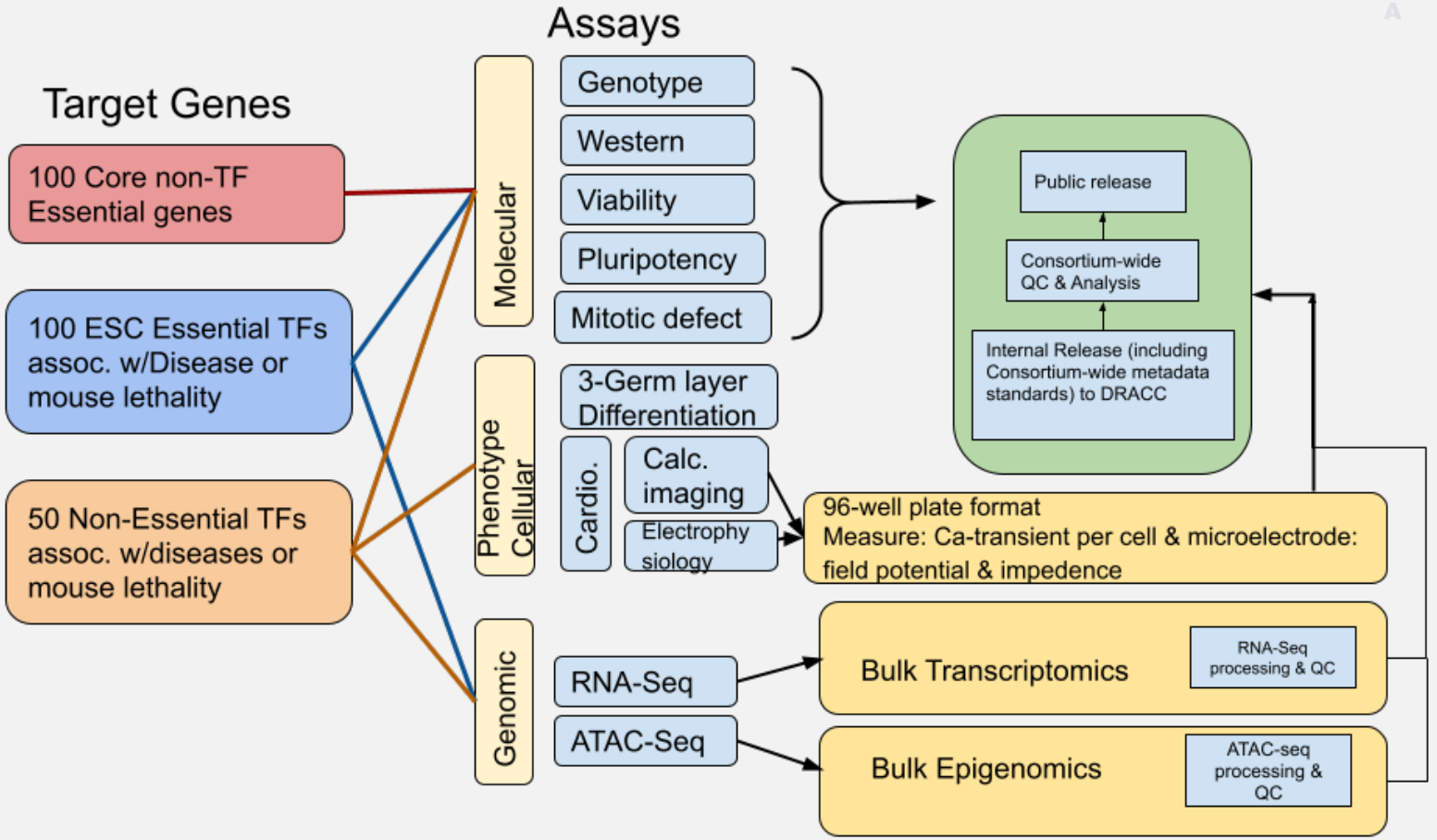
The MorPhiC: Data Production Centers (DPC)

- Each DPC proposes to analyze
  - their own data (single modality)
  - undertake some limited integrative analysis across their datasets to discover biology

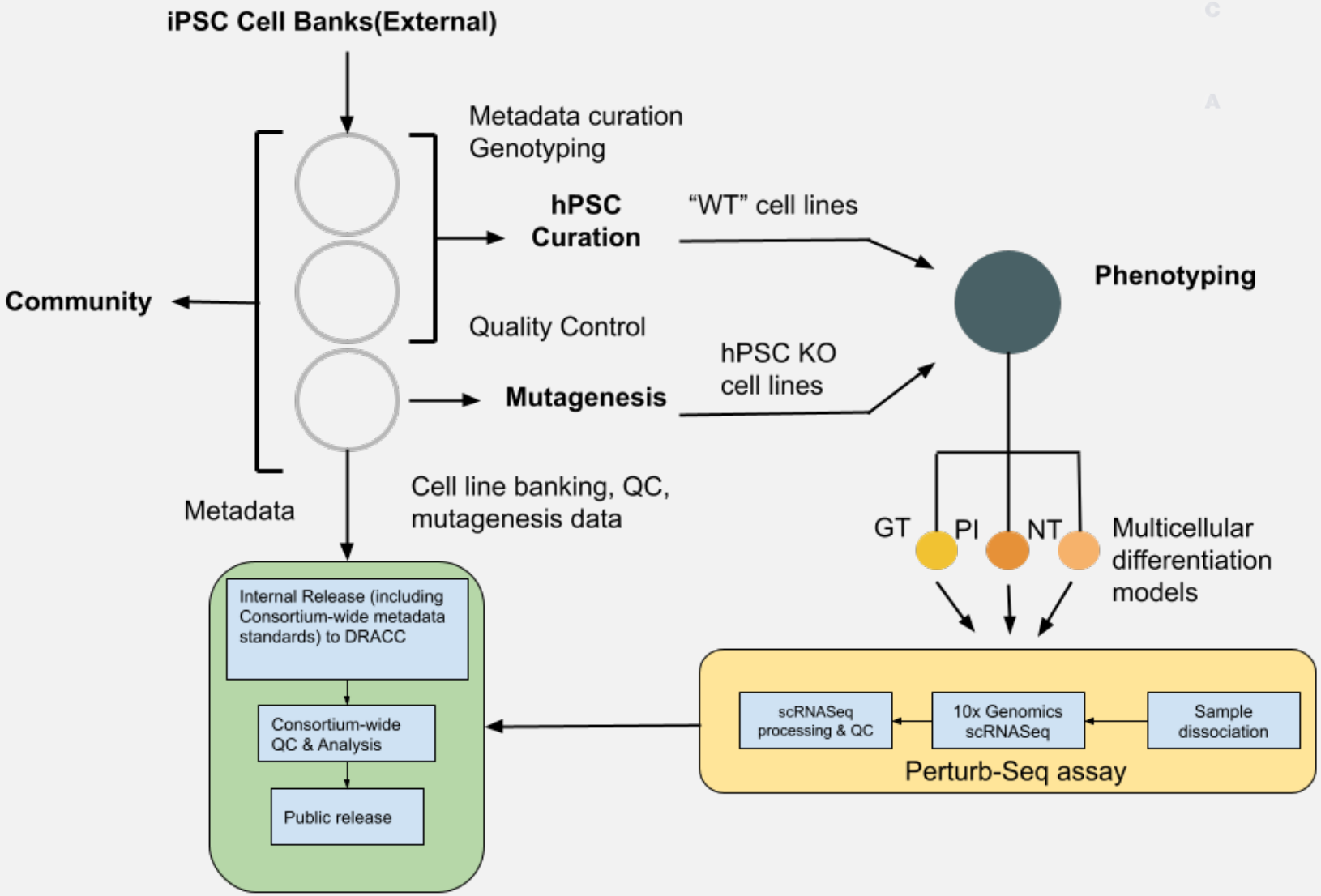
Project 1


iPSC
Primitive Syncytium
Extra-embryonic mesoderm
Brain organoids
Sample dissociation
10x Genomics scRNASeq
scRNASeq processing & QC
Single cell transcriptomics
Nuclei Extraction
10x Genomics snATACseq/ scCUT&Tag
snATACSeq processing & QC
Single cell epigenomics
Cell & Supernatant extraction
Metabolomics LC-MS LC-MS/MS
Pathway analysis
Metabolomics
Cell & Supernatant extraction
Lipidomics LC-MS LC-MS/MS
Pathway analysis
Lipidomics
Plating
Within-incubator & Confocal imaging
Growth Analysis
In situ imaging
Internal Release (including Consortium-wide metadata standards) to DRACC
Consortium-wide QC & Analysis
Public release

Project 2

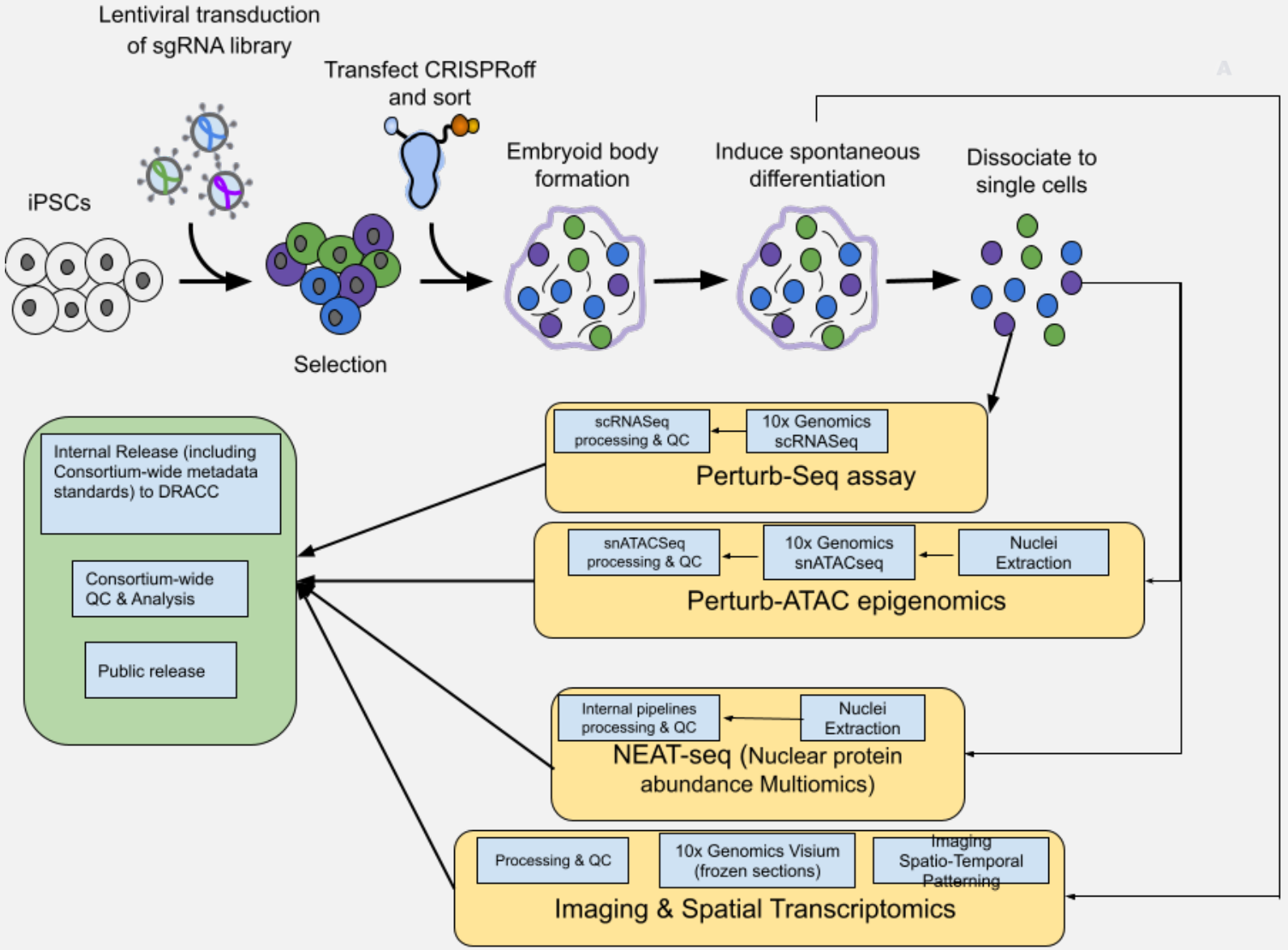

Assays
Target Genes
100 Core non-TF Essential genes
100 ESC Essential TFs assoc. w/Disease or mouse lethality
50 Non-Essential TFs assoc. w/diseases or mouse lethality
Molecular
Phenotype Cellular
Genomic
Genotype
Western
Viability
Pluripotency
Mitotic defect
3-Germ layer Differentiation
Cardio.
Calc. imaging
Electrophysiology
RNA-Seq
ATAC-Seq
Public release
Consortium-wide QC & Analysis
Internal Release (including Consortium-wide metadata standards) to DRACC
96-well plate format
Measure: Ca-transient per cell & microelectrode: field potential & impedence
Bulk Transcriptomics
RNA-Seq processing & QC
Bulk Epigenomics
ATAC-seq processing & QC

Project 3



GT: 2D gastruloid; PI: Pancreatic Islet Organoids; NT: neuron/astrocyte/microglia tri-culture

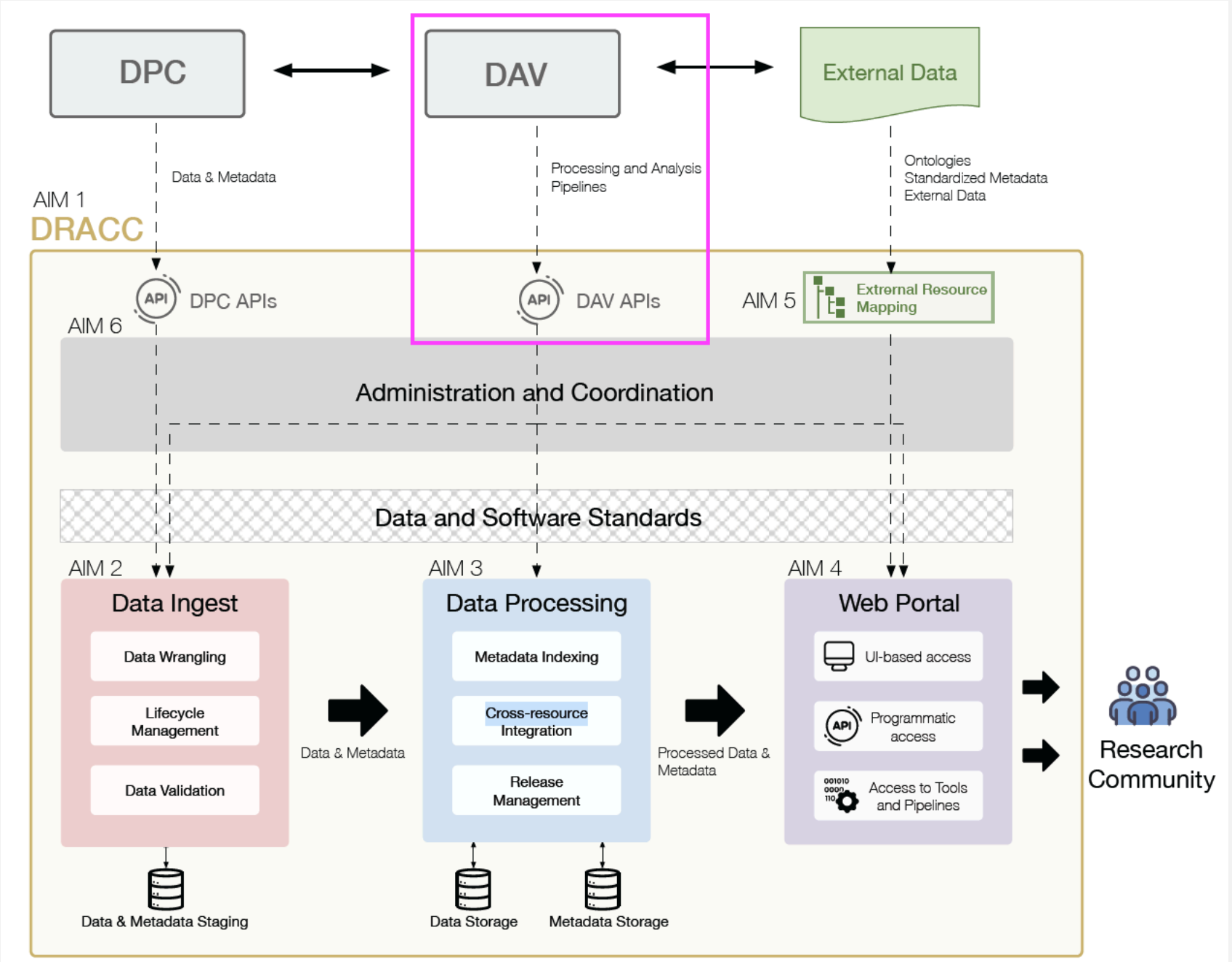
Project 4


Lentiviral transduction of sgRNA library
Transfect CRISPRoff and sort
iPSCs
Selection
Embryoid body formation
Induce spontaneous differentiation
Dissociate to single cells
Internal Release (including Consortium-wide metadata standards) to DRACC
Consortium-wide QC & Analysis
Public release
scRNASeq processing & QC
10x Genomics scRNASeq
Perturb-Seq assay
snATACSeq processing & QC
10x Genomics snATACseq
Nuclei Extraction
Perturb-ATAC epigenomics
Internal pipelines processing & QC
Nuclei Extraction
NEAT-seq (Nuclear protein abundance Multiomics)
Processing & QC
10x Genomics Visium (frozen sections)
Imaging Spatio-Temporal Patterning
Imaging & Spatial Transcriptomics

The MorPhiC: Data Coordination Center

- The DRACC will be doing many things
  - A generic summary follows

Data Coordination



RFA HG-22-019

Questions?

Part 4: Tabular formatted Morphic Experimental Data

Project 1	Classes of Targeted Genes	Cellular Models	Types of Assays	Other General Scientific Questions to be Explored
			✓sc-transcriptomics at 4 different time points ✓sn-epigenetics ✓detailed high-resolution bulk proteomics & lipidomics (75 genes over 5 years): LC-MS and LC-MS/MS ✓ In biological triplicate (replication)	✓Replicate a number of KOs (18 genes with robust cellular phenotypes) in three additional iPSC lines to address impacted by sex and/or genetic background (1 European female, 2 lines of non-European ancestry male & female) ✓Knockout clones will also be reverted to wild-type by a second round of CRISPR-based editing to investigate: reversion of phenotype; evidence for cellular adaptation responses; contribution of genetic drift? ✓generate protein-tagged 'degron' alleles for the same set of five genes to test an orthogonal method for conditional, reversible gene knockouts
Number of Knockouts	✓ 250 genes over 5 years ✓ 391 alleles over 5 years ✓ Main method: CRISPR-Cas9 KO	Main workhorse cellular system for large scale data generation	✓ differentiated cells: KOLF2.1 (46;XY)	Main workshorse assays
Classes of Targeted Genes	✓ gene selection criteria will focus on genes expressed at these early time points. ✓ Following classes will be included (and others based on Consortium discussion): developmental regulators; genes with overt transcriptional effects; others with no known transcriptional/chromatin regulation ✓ separate gene selection criteria for neuroectoderm lineage and extra-embryonic lineage	Other cellular systems that will be tested	✓ iPSC's will subsequently be differentiated into two cell lineages: (1) the extra-embryonic, trophoblast and extra-embryonic mesoderm in monolayer cultures and (2) the neuroectodermal (early embryonic lineages in organoid cultures along the neuroectoderm lineage)	Other assays that will be used in a targeted manner
			✓ differentiation schemes will be temporally imaged to capture phenotypes & morphology ✓ Stable isotope labeled spike-in standards will be used NIST (SRM 1950) & QStd in all metabolomics and lipd assays	✓phenotyping efforts, in Phase I of MorPhiC, will focus on relatively early events in development ✓compare KO strategies—editing biallelic premature termination codon, exon and gene deletion alleles for 24 genes assay based on transcription changes to identify reliable strategy for null allele generation

Project 2

Classes of Targeted Genes		Cellular Models		Types of Assays		Other General Scientific Questions to be Explored
Number of Knockouts	✓ 250 genes in iPSC ✓ Main method: conditional null phenotype by using the Auxin inducible degron technology (AID)	Main workhorse cellular system for large scale data generation	✓ Female hiPSC line ✓ After selection of single cell clones (after transfection) a number of QC steps will be undertaken	Main workshorse assays	✓ bulk RNA-Seq at 0, 6 & 48 hrs after null allele induction for 100 TFs ✓ bulk ATAC-Seq at 0, 6, & 48 hrs after null allele induction for 100 TFs ✓ IncuCyte live cell imaging platform to study viability, proliferation & mitotic defects after induction of null phenotype with auxin treatment	✓ Fitness & Stemness phenotype after null allele using FACS ✓ Differentiation propensity of 250 alleles into 3 germ layers. Mesoderm—cardiomyocytes, endoderm—hepatocyte-like cells, ectoderm—sensory neurons in a 10-day timeline. ✓ Cardiomyocytes: mature d30 cardiomyocytes will be studied for a limited number of KO alleles using (1) high-throughput Ca-transient measurement, and (2) microelectrode array measurement of field potential duration, beat rate etc.
Classes of Targeted Genes	✓ 200 essential genes ✓ 50 non-essential TFs	Other cellular systems that will be tested		Other assays that will be used in a targeted manner	✓ Genotype ✓ Western	✓ Compare ~50 clonal CRISPR KO with AID-mediated protein depletion (RNA and ATAC) ✓ AID-knock-in strategy allows controlling the timing and the level of the target protein degradation and it is reversible (multiple downstream questions can be addressed).

Project 3	Classes of Targeted Genes	Cellular Models	Types of Assays	Other General Scientific Questions to be Explored
Number of Knockouts	✓ Generate clonal KO lines for ~250 genes ✓ generate “population KO” for 3-5 selected genes using all 96 curated hPSC lines ✓ Method: CRISPR-Cas9 based HDR	Main workhorse cellular system for large scale data generation ✓ 96 curated (existing) hPSC lines; equal male, female; emphasize diverse ancestry; healthy ✓ array individual lines in 96-well format, and then pool the lines ✓ clonal KO lines for ~250 genes, select 3-5 diverse and representative hPSC lines, including RUES2 ✓ Differentiation (2D gastruloid, neuron-glial tri-culture, and islet-like organoids) starting from 96 curated hPSC lines in a pooled format	Main workshorse assays ✓ scRNA-seq (Perturb-Seq; 10x) ✓ Perturb-Seq on pooled clonal KO 5-10 diverse hPSC lines ✓ (Custom scripts) for Perturb-Seq demultiplexing and assignment of reads and sgRNA.	✓ Phenotyping in diverse genetic backgrounds ✓ Variability of differentiation behavior (including proliferation related issues in complex cell systems) ✓ Some genes (due to pleiotropy) may need KO after differentiation
Classes of Targeted Genes	✓ 161 genes required for normal embryonic development in mice (distribution of genes with various mouse phenotypes)✓ 50 genes linked to T2D and Alzheimer’s disease based on human genetic studies ✓ Set of +ve & -ve control genes to QC phenotype assays	Other cellular systems that will be tested ✓ Primary human islets	Other assays that will be used in a targeted manner	✓ Improvements in power calculations to determine # of cells screened/sample (extensions into pooled screens).

	Classes of Targeted Genes		Cellular Models		Types of Assays	Other General Scientific Questions to be Explored
Project 4	Number of Knockouts ✓ Method: Use CRISPRoff (directly target mRNA and protein levels) ✓ 250 genes	Main workhorse for large scale data generation	✓ iPSC cells ✓ Embryoid body (all 3 germ lineages at equal representation) ✓ iPSC from 2 Asian, 2 Hispanic, 2 African American, 4 European (1 male and 1 female) ages: 24-72.	Main workshorse assays	✓ Perturb-Seq (pooled screening) ✓ Perturb-ATAC (pooled screening) ✓ (Custom scripts) for Perturb-Seq demultiplexing and assignment of reads and sgRNA.	✓ Demonstrate that CRISPRon can functionally reverse CRISPRoff null allele phenotypes in a polyclonal population of iPSCs and in embryoid bodies.
	Classes of Targeted Genes ✓ Priority: transcription factors and epigenetic regulators	Other cellular systems that will be tested	✓ Each CRISPRoff will be followed by CRISPRon for validation studies.	Other assays that will be used in a targeted manner	✓ NEAT-seq (simultaneouly measure RNA, Chromatin, intracellular proteins—pooled screening) ✓ Non-pooled EB generation with null alleles in each cell (20-30 prioritized genes) follwed by serial confocal imaging every 2hrs from 24-72hrs. ✓ Followed by freezing/sectioning for 10x Visium (spatial profiling).	✓ Explore genetic compensation ✓ Explore the influence of paralogs on 'null phenotypes after KO'


The Forefront
of Genomics®