Functional Genomics and Single-Cell Genomics

Christine Queitsch
Department of Genome Sciences
queitsch@uw.edu
Outline

• Challenges

• Combining methods for better genome assembly

• Functional genomics – let’s sequence it

• Principles and examples of major findings

• Single cell genomics – why and how

• Applications and examples

• What’s next?
Challenges: variants discovery and characterization

Structural variation and repetitive DNA pose unsolved challenges!

patterns of intrachromosomal (blue) and interchromosomal segmental duplication (red, ≥10 kb)
Why do these regions matter?

- segmental duplications preceded divergence of humans
- segments encode human/great ape-specific gene families
- genes with functions in neurodevelopment, cell proliferation
- CNVs in segmental duplication regions are implicated in many human disease
How to assemble such regions?

• reduce complexity by sequencing haploid genomes ->
  - hydatidiform moles
  - HAP1 and hapESC cells

• longer reads -> PacBio, nanopore and short-read sequencing for error corrections

• adding proximity context -> Hi-C

Huddleston et al. 2017, Genome Research
Bickart, 2017, Nature Genetics
Burton, 2013, G3
Context allows deconvolution of complex samples
Variants are everywhere – what do they mean?

The challenge is enormous…this is just representing the coding portion of the genome (<1%).
Functional genomics – massively parallel assays

Saturation mutagenesis

…any variant associated with a selectable function can be interrogated...

cell growth
drug resistance
aggregation
expression of reporter

neutral
beneficial
deleterious
possibly beneficial

Fowler & Fields Nat. Methods 2010
Examples

Yeast mating – a canonical MAPK pathway with several phenotypic outputs

Mating and invasive growth converge on a single transcription factor: Ste12.

Dorrity et al., PNAS, 2018
Ste12 variant performance in mating
Are there Ste12 DBD mutants that separate traits?

Phenotyping thousands of variants by sequencing

Deep mutational scanning allows assessing functional consequences of all (most) possible Ste12 DBD variants for mating and invasion under different conditions.
Single mutation suffices to shift between traits

Certain Ste12 variants confer hyperinvasion at the cost of mating.

Dority et al. PNAS 2018
From yeast to human genes – within genomic context

genomic locus for saturation genome editing

high-content and high-throughput phenotyping:
molecular phenotypes
cellular phenotypes
BRCA1 – the breast cancer gene

BRCA1 saturation genome editing yields functional information on variants of previously unknown effects.

Challenge – how to scale up???

generating all possible variants:

• efficient insertion of variant libraries in genome OR efficient editing

selection regimes and relevant phenotypes:

• protein aggregation holds functional information → VAMP-seq
• cell shape → advanced microscopy compatible with selection of cells
• nucleoli shape and size
• assessing protein surfaces and interactions through sequencing

NEXT Frontier: assessing function of proteins in similar fashion with mass spectrometry
How does cellular gene expression differ for each cell carrying a particular variant? Or accessibility? Or TF occupancy.
Nucleolar morphology and size are aging and disease markers

Hasle et al. unpublished
Dominant negative peptides for every protein

competing for a protein interface

titrating a ligand

interfering with co-translational folding

Dominant negative action is protein-based:
• rapid
• highly specific
• reversible

…but dominant negative mutants are scarce…

Betz and Fall, Gene 1988
Michaels et al, PNAS 1996
Can we identify dominant negative fragments of Hsf1 that inhibit cell growth?
Dominant negative fragments reflect function

...and reveal novel functional aspects.

- >12,000 fragments tested
- 237 dominant negatives in 30C, 434 at 37C
- overlapping fragments define minimal regions
- dominant negatives are context-specific
- high reproducibility between replicates (r²>0.90)

Dorrity et al. Nature Methods, 2019
Genome-wide screen for DN fragments

Selection for growth and cDNA-based approach limit possible targets.

- 6,713 annotated yeast genes
- >172,000 cDNA fragments
- 3,311 with at least one in-frame fragment at required sequencing depth
- 1,505 with at least one strongly depleted fragment
- 20,000 most depleted fragments map to 157 genes (and an additional 257 genes at lower stringency)
Large complexes can be inhibited with DN fragments

...inhibiting ribosome function

Dominant negative polypeptides can inhibit paralogs or distinguish among them.
DN activity changes with altered cellular conditions

Loss of dominant negative activity in heat stress conditions

Gain of dominant negative activity in heat stress conditions

Large overlap in depleted fragments (97.8%) but 508 peptides were specifically depleted in heat stress.
DN activity changes with altered cellular conditions

The fragment with the greatest DN activity specific to heat stress corresponds to Hsf1 linker region.

Dorrity et al. Nature Methods, 2019
Single-cell genomics – when?

Timeline of Single Cell Sequencing Milestones

- Development of Next Generation Sequencing Platforms (2005)
- First DNA Genome Sequencing of Single Human Cells (Navin et al., 2009)
- Exome Single Cell DNA Sequencing (Xu et al., Hou et al., 2011)
- DNA SCS of Neurons (Evrory et al., 2012)
- MALBAC DNA SCS Method (Zong et al., Lu et al., 2013)
- Microbial Tree of Life DNA SCS (Rinke et al., 2013)
- First SCS of Epigenomes (Nagano et al., 2014)
- DNA SCS of Human Oocytes (Hou et al., 2014)
- RNA SCS using Unique Molecular Indexes (Islam et al., 2014)
- Organ lineage tracing (Trueltlein et al., 2014)

- First RNA Transcriptome Sequencing of a Mammalian Cell (Tang et al., 2005)
- DNA SCS of Sperm Cells (Wang et al., 2009)
- RNA SCS of CTCs using Template-Switching (Ramskold et al., 2013)
- RNA SCS of Immune Cells (Shalek et al., 2013)
- DNA SCS Tissue Mocaicism (McConnell et al., 2014)
- DNA SCS of CTCs (Ni et al., 2014)
- In Situ RNA SCS in Tissues (Lee et al., 2014)
- G2/M DNA SCS of Breast Tumors (Wang et al., 2014)

Degenerate oligonucleotide primed PCR (DOP-PCR, Telenius et al. 1992)

- Single cell transcriptome sequencing (Tang et al., 2009)
- scRRBS (Guo et al., 2013) scHIC (Nagano et al., 2013)
- Drop-seq (Macao et al. 2015) scATAC-seq (Buenrostro et al., 2015)
- scChIP-seq (Goren et al., 2015)
- Linear Amplification via Transposon Insertion (LIANTI, Chen et al. 2017)

- Multiple displacement amplification (Dean et al., MDA, 2001)
- Smart-seq (Ramskold et al., 2012) Single cell Exome sequencing (Li et al., 2012)
- Multiple Annealing and Looping Based Amplification Cycles (MALBAC, Zong et al., 2012)
- Co-detection of DNA and RNA of a single cell (Han et al., 2014) scBS-seq (Smallwood et al., 2014)
- scNucleosome positioning (Small et al., 2014)
- Single cell small RNA seq (Faridani et al., 2016)


Hu et al. Frontiers in Cell and Developmental Biology 2018
Single-cell genomics – why?

- discover new cell types
- reveal developmental trajectories
- understand heterogeneity
- cancer
- neurodevelopment and memory
Single-cell genomics – technical principles

Drop-seq

Macosko et al. 2015 Cell
Sci-seq – combinatorial indexing

Vitak et al 2017 Nature Methods
Exploring single biology in well-studied models

*C. elegans* - 959 cells

John Sulston - 1983 – embryonic development
Nobel Prize in 2002 with Sydney Brenner and Bob Horvitz
What do we find – rare neuronal cells
GFP marker lines:

- Extensive expression profiling of GFP-sorted cells
- Some lines express GFP in more than one cell type
- Several cell types lack marker genes

Extensive prior knowledge facilitates annotation.

*De Smet et al., Int. J. Mol. Sci. 2015  Brady et al., 2007; Cartwright et al., 2009, Li et al., 2016*
Annotation with Monocle 3 and UMAP

3,121 cells
22,419 total genes expressed
2,445 median genes per cell expressed
1,500 ordering genes

Jean-Baptiste et al., Plant Cell 2019  Brady et al., 2007; Cartwright et al., 2009, Li et al., 2016
Root hair cell trajectory and drivers

Jean-Baptiste et al., Plant Cell 2019
Root hair trajectory reflects differentiation

Waddington’s epigenetic landscape

Bhosale et al., The Plant Cell, 2018

Waddington, 1957

Jean-Baptiste et al., Plant Cell 2019
A surprising number of differentially expressed genes are cell-specific.

Jean-Baptiste et al., Plant Cell 2019
What’s next?

Combine functional genomics with single-cell genomics

How does cellular gene expression differ for each cell carrying a particular enhancer or gene variant? Or accessibility? Or TF occupancy? Or protein content and interactions?
WE ARE DEVELOPING HIGHLY GENERALIZABLE, REPRODUCIBLE, AND SCALABLE TECHNOLOGIES FOR RELIABLY ASSESSING THE FUNCTIONAL IMPACT OF ALL POSSIBLE SINGLE NUCLEOTIDE VARIANTS IN HUMAN GENES AND THEIR REGULATORY ELEMENTS

https://www.cmap.gs.washington.edu

Team: Doug Fowler  Judit Villen
    Stan Fields  Fritz Roth
    Jay Shendure  Lea Starita