Limits of Data and Interpretation in Genetics, Genetic Medicine, Evolution, & the Origin of Life

David Thaler

Collaborations & Inspirations

Marcelo Magnasco

Albert Libchaber

Joshua Lederberg

Bradley T. Messmer

Gerald Peretz

Fiona Doetsch

Current Limits:

1. Confusion

2. Technologies

Deep Limits: Likely

- Two Definitions of Genetics
- Both Necessary
- Partial Overlap, Amount Unknown

Classical Definition of a Gene Still Relevant and Essential

- Gene: Unit of inheritance
  - Based on Heredity of Phenotype
    - Historically independent of material basis
    - Genes linked to chromosomes
  - Exceptions Then and Now
    - Genes not on chromosome
      - Cytoplasmic/Maternal inheritance/Organelle DNA
    - Other (Prions , imprinting)

Second and Pseudo-modern Definition of Gene

- Gene is a sequence of nucleotide bases in DNA
  - Sometimes limited to protein-encoding nucleotides
- Avery, McLeod, McCarty: Proved that the material basis for one gene is DNA
- Now many examples of genes as DNA sequence
- Are all classical genes DNA sequence?
  - Logically No
  - Factually No
  - The two definitions of genes have become blurred.

+ De-Babelization of Biology - Light Under the Lamppost

- Power of universal databases & metrics
  - Genomics allows clear communication but at expense of jamming everything into sequence.
  - Non-sequence Biology is hard to communicate.
- Genomics => Expression Genomics and Proteinomics. Limitations may hitchhike or may be transcended. Need to frame as paradox search, i.e. to autofind apparent contradictions between theory and experiment, e.g. gene vs protein vs enz.

Are Non-Sequence Genes Trivial?

- Prions
  - Alternative Protein Conformation (S. Prusiner)
  - Ice Nine of Vonnegut
- Histone modification: acetylation, ribosylation, methylation, phosphorylation
  - http://www.cstone.net/~jrb7q/chrom.html
- Methylation of DNA
  - Imprinting e.g. of X chromosome in female mammals
  - (R. Holliday, M. Lyons X)
- Stable Physiol. States (A. Novick, L. Szilard, M. Ptashne)
- Cytoplasmic Continuity of Life

Current Technical Limits of Sequence-Based Genetics

- Data quality often not high & not known
- Cell to cell heterogeneity
  - Expression and sequence
- Purification and unbiased amplification
  - small samples
  - Destroy in order to study. In vivo veritas.
- Intracellular heterogeneity: Distribution/3D structure
- Solving these problems in nomics ways will lead to progress in mechanistic understanding akin to new types of microscope or telescope. Can happen.
- Cost, Centralization

Current Limits of Non-Sequence-Based Genetics are Severe

- No universal way to communicate Babel Biol.
- Proportional contribution Evo & Devo unknown
  - X chromosome in female mammalian somatic cell
  - Transgenes and native somatic genes undergo silencing (Yeast, Mice, Drosophila, Human)
  - Stable through meiosis in Yeast and Mice
- Range of stability more variable than sequence
  - But some sequences are highly cis-mutable

A Conjecture Concerning Sequence Analysis

Given

- the genomic sequence of an organism
- the expression profiling of the organism
- any possible algorithm for relating sequence to expression pattern
- an unknown amount of expression/differentiation is controlled in a non-sequence-dependent manner

Some sequence-based interpretations will be wrong

- It will be difficult to know which are wrong

Conjecture

The minimum complexity of any model that can account for the degree of expression on the basis of sequence that is indisputably the case will also be complex enough to automatically, seamlessly and wrongly rationalize non-sequence based aspects of gene control.

I've always found theology a certain kind of delightful titillation. Theology or religious speculation bear the same relationship to real experience as pornography does to lovemaking. They're not entirely unconnected. I mean, you can get turned on. One of the reasons that they're both powerful is that they ignore a lot of other material and they both focus in on something very specific. In these days of overload, it's very restful to know, at last, what you're talking about. -Leonard Cohen

Genetic Medicine

IF

- **Medicine follows science ~years behind**
- **Narrow definition of DNA sequence as gene becomes dogma in medicine**

THEN

- **Non-sequence phenotypes and genotypes remain under-appreciated**
- **Relative importance of what is ignored remains unknown & unstated**
- **Somatic genetics (Cancer and aging research & medicine) at risk. Non-Sequ. has > role in development than evolution.**

The Generation of Evolutionary Diversity

- Generation of Bio. Diversity has both random and non-random aspects in both evolution & development
- Mechanisms that generate evolutionary diversity are themselves encoded by genes and are subject to mutation and inheritance just like any other gene
  - Marcelo Magnasco formalized & de-glassed a glass
- G.O.D. responds to environment
- Inter-relatedness of stochastic/random and algorithmic/deterministic in generating evolutionary and developmental diversity is unknown & rich

From Meeting Program

- What we see, indeed, may reflect the influence of frozen historical accidents, and represent only one realization of a stochastic process that admits many theoretical possibilities. A modern example is the question of whether the genetic code is a frozen accident, or the only code possible in some broad sense.
- Consider Q from two points of view
  - Can a modern organism function with a different genetic code? What would have to be adjusted?

Can life with a different Genetic Code work?

- Which codes would be nice to try? How many?
- How to do it? What might be learned from trying?
  - Synthesize genomes with different coding and a set of cognate tRNAs Doable lab in 5-10 years.
  - Bacteriophage easier Technically at Edge Now
  - Code could be altered one (or two) AA at a time
    - What works, what needs to mutate and adjust?
    - (re StrepR, Bruce Levin)

Is Evolutionary drive for genetic code knowable?

- Is history a final block? Not Necessarily
  - Need some hypothesis & insight
- A current hypothesis being tested by Laura Landweber (Princeton) among others
  - Random pool of RNA sequences-> selection for binding to a particular AA
  - For some AA the cognate codons are frequent among RNA selected for binding to that AA
  - Other AA: no relation to codons.
  - Possibility of multiple origins of the code or parts of the code

Symbiosis at the Origin

- Another Scenario
  - Manuscript Available
- Key Points in Contemporary Life
  - Ribozymes Synthesize Proteins
  - Peptides Synthesize Polynucleotides
  - Neither is Chemical Necessity
- Origin Hypothesis: epicycle in which ribozyme and peptides synthesize each other. Catalysis precedes templates. Selection early. Predict: patterns of sequence similarity in ribosome ribozyme and polymerase catalytic region.
    - Inspired and adapted from F. Dyson

Summary

- Two Definitions of Genetics often confused.
- Arrogance of sequence may be efficient today (trains run on time) and be overthrown later
  - Having said that, as intellectuals, universal & public claims of sequence as genetics are questionable
  - Sequence-based Medicine may also focus narrowly
- The Generation of Diversity in living lineages occurs in both evolution and development. Their intermixture is subtle and consequential.
- Origin & Maintenance of the Genetic Code knowable.
  - Requires new insight & experiment on the origin of life.

- Thaler, D. S. (2000). Symbiosis at the One Origin of Life. MS only
- Thaler, D. S.(1999). Hereditary stability and variation in evolution and development. Evolution & Development 1(2): 1-10.
- Field, D, M. O. Magnasco, et al. (1999). Contingency loci, mutator alleles, and their interactions. Synergistic strategies for microbial evolution and adaptation in pathogenesis. Ann N Y Acad Sci 870: 378-82.
- Kaplan, P.D., Q. Ouyang, et al. (1997). Parallel overlap extension for construction of computational DNA libraries. J. Theor. Biol. 188: 333-341.
- Magnasco, M. and D. S. Thaler (1996). Changing the pace of evolution. Physics Letters A 221: 287-292.
- Moxon, E.R. and Thaler, D.S. 1997. The Tinker s Evolving Toolbox Nature 387: 659-662
- Thaler, D.S. 1994. The evolution of genetic intelligence. Science 264: 224-225.
- Thaler, D. S. (1996). Paradox as path: Pattern as map. The Philosophy and History of Molecular Biology: New Perspectives. S. Sarkar. Klewer Academic, Klewer Academic: 233-248.