HUMAN EVOLUTIONARY GENOMICS AND THE SEARCH FOR THE GENES THAT MADE US HUMAN

JAMES M. SIKELA
Department of Biochemistry and Molecular Genetics
Human Medical Genetics and Neuroscience Programs
University of Colorado School of Medicine
Aurora CO 80045, USA
Email: james.sikela@ucdenver.edu

1. Introduction:

Due to advances in genome sequencing and analysis we can now carry out comparisons of genomic data with unprecedented scope and detail, both within and between species. As a result gene and genomic changes that have occurred specifically in the human lineage can now be identified. This field is one of the most interesting areas of human biology, and there has now emerged, due in large part to our ability to carry out computational analyses of large genome datasets, a growing list of significant discoveries. This workshop will survey the latest findings in this rapidly advancing field, including both key discoveries and remaining challenges (e.g. complexity of some genomic regions, difficulties in moving from identification of genomic differences to linking them with human-specific traits, etc.).

The types of human lineage-specific genomic changes, and the methods used to detect them, cover a wide range, and speakers have been chosen with this in mind. Speakers include individuals focused on 1) computational analysis at the transcriptional level particularly in brain (D. Geschwind; E. Lein), 2) computational genomics to identify rapidly evolving genomic sequences including non-coding sequences (D. Kostka), 3) computational analysis of protein coding sequences under selection in the human lineage (E. Vallender), 4) identification of gene coding regions that have undergone extreme human lineage-specific copy number expansions (J. Sikela), and 5) computational approaches to identifying genomic changes that distinguish "modern" humans from archaic hominins (J. Wall).

2. Speakers and Abstracts:

Transcriptional Features of Human Brain Development

Ed S. Lein, PhD
Allen Institute for Brain Science

How the genome guides brain formation is still poorly understood, as is our understanding of which features of brain development are conserved across all mammals versus specific to human. To address these questions we have created a series of high anatomical resolution transcriptional atlases of the developing human, non-human primate and mouse brain, and performed systematic analyses of transcriptional dynamics during brain development and between species. We find the transcriptome is tremendously dynamic across brain development, particularly in prenatal and early postnatal stages, with both similarities and many differences between rodents and primates. These differences in gene regulation likely underlie the unique

features of human brain structure and function, and provide clues about the locus of action for genes associated with neurodevelopmental diseases.

Transcriptional Networks in Human Brain
Daniel Geschwind, MD/PhD
University of California, Los Angeles School of Medicine

Understanding the evolution of human cognitive phenotypes has benefited greatly from comparison with non-human primates. Studies of genetic and transcriptional changes in brain genes has identified many candidates, but studies of transcription are often limited by the notion that gene expression differences between species are mostly neutral. We have developed network approaches to analysis of transcription data that allow us to put changes in gene expression in a functional context, permitting us to identify changes that are likely to be relevant to brain evolution on the human lineage. We have also used this framework to connect gene expression networks to other types of networks in human brain to connect different levels of brain function.

Lineage-specific Accelerated Evolution in Five Primates
Dennis Kostka, PhD
University of Pittsburgh School of Medicine

Regulatory evolution has been proposed to explain diversity of closely related species, and there is great interest in identifying the genetic basis of lineage-specific traits. We developed a statistical phylogenetic approach to identify genomic regions with lineage-specific accelerated substitution rates (linARs) and applied it to human and four non-human primates. We find an enrichment of human-specific linARs in non-coding regions with epigenetic marks of regulatory sequences, particularly nearby neurodevelopmental genes. Comparing across primates, similar loci and pathways harbor distinct linARs from multiple species. Thus, shared biological processes may have been independently targeted by adaptive events in multiple primate lineages.

Detecting Signatures of Selection through Interspecific Comparisons
Eric J. Vallender, PhD
Harvard Medical School

Selection on proteins can be identified through the ratio of fixation of nonsynonymous changes to synonymous changes (dN/dS). With the proliferation of genomic sequence, the ability to broadly and without bias survey protein evolution is now easily accessible. Using genomes from forty-five mammalian species, we recently interrogated 2,350 genes associated with neuropsychiatric disease in humans. In doing so, we identified strong signatures of purifying selection across mammals with a moderate elevation, indicative of an apparent relaxation of constraint, in catarrhine primates and a large, pervasive, apparent acceleration in cetaceans. This research not only demonstrates a general lack of selection on coding changes in genes associated with these diseases, but also highlights some of the challenges associated with molecular

evolution in the post-genomic era including ortholog identification, genome error and misannotation, and difficulties in functional attribution of genes and proteins.

DUF1220 Protein Domains: Gene Sequences Showing Extreme Human-specific Copy Number Expansion

James Sikela, PhD
University of Colorado School of Medicine

Gene duplication is thought to be the primary means by which evolution creates new genes and biochemical processes that have facilitated the evolution of complex organisms from primitive ones. Gene duplications (gene dosage increases) that are specific to the human lineage have been identified in well over 130 genes some of which are excellent candidates to underlie human lineage-specific traits. Among these sequences are those encoding DUF1220 protein domains, which show the largest human lineage-specific copy number increase of any protein coding region in the genome and have been linked to human brain expansion.

Admixture at Different Timescales in Human Genomes

Jeff Wall, PhD
University of California, San Francisco

Admixture between diverged populations is a common phenomenon in human history. We review some of the methods for detecting this admixture, and show how it provides evidence for admixture between modern humans and 'archaic' human groups across a wide range of contemporary human populations. The strongest signal of this admixture occurs in sub-Saharan African populations, which is consistent with the extent of presumed opportunities for admixture that have been proposed from archeological research.