



WORKSHOP PROGRAM

National Institute on Aging Workshop

Genomic Determinants of Differences Among Humans and Other Primates in Species Life Spans and Progression of Age-Related Conditions

April 14-16, 2025

Day 1 | Monday, April 14, 2025

9:00 a.m. – 9:10 a.m. | Welcome and Introductions/Goals of the Workshop, *Evan Hadley, Director, Division of Geriatrics and Clinical Gerontology, National Institute on Aging (NIA)*

9:10 a.m. – 9:15 a.m. | Workshop Proceedings Communications, *Carol Nguyen, Division of Geriatrics and Clinical Gerontology, NIA*

Session 1, Part 1 | Characterization of genomic regions/features of interest and potential relationship to species longevity, Moderator, Michael Bennani, Division of Neuroscience, NIA

9:15 a.m. Coding regions: Transcriptomic and proteomic evidence, *Vadim Gladyshev, Harvard University*

9:30 a.m. The functional and evolutionary impacts of human-specific deletions in conserved elements, *Steve Reilly, Yale School of Medicine*

9:45 a.m. Alternative splicing in primate brain evolution and a curious case of MAPT/tauopathy, *Chaolin Zhang, Columbia University*

10:00 a.m. The search for primate longevity genetics: with special emphasis on *FOXO3* gene-gene interactions, *Tim Donlon, University of Hawaii*

10:15 a.m. Epigenetic predictors of primate species life spans, beyond the epigenetic clock, *Chris Faulk, University of Minnesota*

10:30 a.m. – 11:15 a.m. Q&A and Discussion from Session 1, Part 1

11:15 a.m. – 12:15 p.m. | LUNCH BREAK

Session 1, Part 2 | Characterization of genomic regions/features of interest and potential relationship to species longevity, Moderator, Nalini Raghavachari, Division of Geriatrics and Clinical Gerontology, NIA

12:15 p.m. Centromere variation among human and non-human primates, *Glennis Logsdon, University of Pennsylvania Perelman School of Medicine*

12:30 p.m. Repeat dynamics in primate genomes, *Rachel O'Neill, University of Connecticut*

12:45 p.m. Aging alters genomic instability at endogenous mutation hotspots in mice, *Karen Vasquez, University of Texas at Austin*

1:00 p.m. New genes, duplications, and complex rearrangements of ape telomere-to-telomere genomes, *DongAhn Yoo, University of Washington*

1:15 p.m. Complete sequences and comparative analysis of great ape sex chromosomes, *Kateryna Makova, Penn State University*

1:30 p.m. – 2:15 p.m. Q&A and Discussion from Session 1, Part 2

2:15 p.m. – 2:30 p.m. | BREAK

Session 4 Preview | Functional study strategies to characterize phenotypic effects of variants under selection for longevity, Moderator, Evan Hadley, Division of Geriatrics and Clinical Gerontology, NIA

2:30 p.m. Comparative studies of genomic functional responses to environmental factors (GxE), *Amanda J Lea, Vanderbilt University*

2:45 p.m. – 3:30 p.m. | Q&A and Further Discussion

Day 2 | Tuesday, April 15, 2025

9:10 a.m. – 9:15 a.m. | Daily Workshop Introduction & Workshop Proceedings Communications, Manuel Moro, Division of Aging Biology, NIA

Session 2 | Evolutionary Analyses, Moderator, Carol Nguyen, Division of Geriatrics and Clinical Gerontology, NIA

Unique features in hominin evolution vs. other hominids

Methods to identify positive or negative selection for differing genetic and epigenetic characteristics related to primate life span (e.g., nonsynonymous/synonymous codon variation, strength of cis-regulatory promoters, identifying clusters of multiple genes or methylation sites correlated with life span).

9:15 a.m. Evolutionary action and evolutionary trace methods to identify functional and fitness effects of genetic divergence. *Olivier Lichtarge, Baylor College of Medicine*

9:30 a.m. Olduvai domains, human brain evolution and neoteny, *James Sikela, University of Colorado Anschutz Medical Campus*

9:45 a.m. Accelerated Regions, *Katie Pollard, University of California San Francisco*

10:00 a.m. TOGA 2.0: Gene annotation, classification and orthology inference, *Michael Hiller, Senckenberg Research Institute*

10:15 a.m. Identifying functional gene duplications important in human evolution, *Megan Dennis, University of California – Davis Health*

10:30 a.m. – 11:15 a.m. | Q&A and Discussion from Session 2

11:15 a.m. – 11:30 a.m. | BREAK

Session 3 | Relationship between genetic and/or epigenetic features and phenotypes, Moderator, Melissa Gerald, Division of Behavioral and Social Research, NIA

Studies for selecting functional “candidates” from genes with evidence of selection related to longevity, rates of chronic aging disease progression or relationships to life history milestones that are correlated with species life span.

11:30 a.m. Effects of proposed aging-related mechanisms or correlates of life span, *Vadim Gladyshev, Harvard University*

11:45 a.m. Biological processes modulating longevity across primates, *Arcadi Navarro, Universitat Pompeu Fabra, Barcelona*

12:00 p.m. Causes and consequences of stress in aging, somatic mutation, and genetic adaptation in long- and short-lived species, *Peter Sudmant, University of California, Berkeley*

12:15 p.m. Epigenetic regulation of timing of neuronal maturation, *Lorenz Studer, Memorial Sloan Kettering Cancer Center*

12:30 p.m. – 1:15 p.m. | Q&A and Discussion from Session 3

1:15 p.m. – 2:15 p.m. | LUNCH BREAK

Session 4 | Functional study strategies to characterize phenotypic effects of variants under selection for longevity, Moderator, Sarah Craig, Division of Geriatrics and Clinical Gerontology, NIA

2:15 p.m. Heterogenous differentiated cultures as models for studying human variation and disease, *Yoav Gilad, University of Chicago*

2:30 p.m. Methods for studying regulatory variation, *Ryan Tewhey, Jackson Laboratory*

2:45 p.m. – 3:30 p.m. | Q&A and Discussion from Session 4

Day 3 | Wednesday, April 16, 2025

9:10 a.m. – 9:15 a.m. | Daily Workshop Introduction & Workshop Proceedings Communications, *Sarah Craig, Division of Geriatrics and Clinical Gerontology, NIA*

Session 5 | Integrative Genomics and Other ‘Omics, Moderator, Julie Mattison, Intramural Research Program, NIA

Strategies for integrative analyses combining multiple genomic and gene functional databases and current literature (including but not limited to machine learning or artificial approaches) to identify novel features related to primate species longevity and rate of progression of aging conditions.

9:15 a.m. Methodological issues in integrating diverse omics and phenotype data, *Nicholas Schork, The Translational Genomics Research Institute*

9:30 a.m. Combining comparative genomics and machine learning to study the genetics of lifespan, *Andreas Pfenning, Carnegie Mellon University*

9:45 a.m. Integrating intra- and interspecies genetic variation in comparative primate studies, *Jeff Rogers, Baylor College of Medicine*

10:00 a.m. Use of integrated omic methods and *in vivo* imaging to understand molecular mechanisms underlying metabolic health, *Laura Cox, Wake Forest University School of Medicine*

10:15 a.m. – 11:00 a.m. | Q&A and Discussion from Session 5

11:00 a.m. – 11:15 a.m. | BREAK

Session 6, Part 1 | The data! What is available? What is missing? Tools? Moderator, Luci Roberts, Division of Neuroscience, NIA

What can NIA and or NIH do regarding informatics and data resources to make integrative analyses easier/faster/more accessible?

11:15 a.m. What we can learn from complete genomes, *Adam Phillippy, National Human Genome Research Institute (NHGRI)*

11:30 a.m. Expanding comparative genomics to hundreds of primates to explore genome function, *Elinor Karlsson, Broad Institute*

11:45 a.m. Primate Sequencing and Conservation Initiative, *Jeff Rogers, Baylor College of Medicine*

12:00 p.m. – 12:45 p.m. | Q&A and Discussion – Primate Genome Projects and BIG data

12:45 p.m. – 1:45 p.m. | LUNCH BREAK

Session 6, Part 2 | The data! What is available? What is missing? Tools? Moderator, Manuel Moro, Division of Aging Biology, NIA

1:45 p.m. Introduction to the Primate Aging Database, *Gergens Polynice, CleMetric*

2:00 p.m. Comparative genomics of ageing: Databases and approaches, *JP de Magalhaes, University of Birmingham (UK)*

2:15 p.m. UCSC Genome Browser: Conservation tracks, *Max Haeussler, UC Santa Cruz Genomics Institute*

2:30 p.m. Using NCBI's Toolkit from the NIH Comparative Genomics Resource to Unlock the Potential of Research Organisms, *Kim Pruitt, National Center for Biotechnology Information (NCBI)*

2:45 p.m. – 3:15 p.m. | Q&A and Discussion – Genomic Resources

3:15 p.m. – 4:00 p.m. | WRAP UP discussion of workshop, *NIA Staff and Invited Speakers*

Index of Speaker Abstracts

LAURA A. COX, WAKE FOREST UNIVERSITY SCHOOL OF MEDICINE	3
INTEGRATION OF MULTIPLE-‘OMICS (TRANSCRIPTOMICS, METABOLOMICS, GENOMICS) TO IDENTIFY MOLECULAR PATHWAYS AND NETWORKS ALTERED BY DEVELOPMENTAL PROGRAMMING	
MEGAN DENNIS, UNIVERSITY OF CALIFORNIA - DAVIS HEALTH.....	3
ROLE OF GENE EXPANSION AND STRUCTURAL VARIATION IN PRIMATE EVOLUTION	
TIM DONLON, PHD, UNIVERSITY OF HAWAI’I.....	4
THE SEARCH FOR PRIMATE LONGEVITY GENETICS: WITH SPECIAL EMPHASIS ON FOXO3 GENE-GENE INTERACTIONS	
MAXIMILIAN HAEUSSLER, UC SANTA CRUZ GENOMICS INSTITUTE	5
UCSC GENOME BROWSER	
AMANDA J. LEA, VANDERBILT UNIVERSITY.....	5
COMPARATIVE STUDIES OF GENOMIC FUNCTIONAL RESPONSES TO ENVIRONMENTAL FACTORS (GxE)	
OLIVIER LICHTARGE, BAYLOR COLLEGE OF MEDICINE	5
EVOLUTIONARY ACTION AND EVOLUTIONARY TRACE METHODS TO IDENTIFY FUNCTIONAL AND FITNESS EFFECTS OF GENETIC DIVERGENCE	
GLENNIS LOGSDON, UNIVERSITY OF PENNSYLVANIA PERELMAN SCHOOL OF MEDICINE.....	6
CENTROMERE VARIATION AMONG HUMAN AND NON-HUMAN PRIMATES	
JOA PEDRO DE MAGALHAES, UNIVERSITY OF BIRMINGHAM (UK)	6
COMPARATIVE GENOMICS OF AGEING: DATABASES AND APPROACHES	
ARCADI NAVARRO, UNIVERSITAT POMPEU FABRA, BARCELONA.....	7
BIOLOGICAL PROCESSES MODULATING LONGEVITY ACROSS PRIMATES	
RACHEL J. O’NEILL, UNIVERSITY OF CONNECTICUT	7
REPEAT DYNAMICS IN PRIMATE GENOMICS	
ANDREAS PFENNING, CARNEGIE MELLON UNIVERSITY	8
COMBINING COMPARATIVE GENOMICS AND MACHINE LEARNING TO STUDY THE GENETICS OF LIFESPAN	
KATIE POLLARD, UNIVERSITY OF CALIFORNIA SAN FRANCISCO.....	8
ACCELERATED REGIONS	
GERGENS POLYNICE, CLEMETRIC.....	9
PRIMATE AGING DATABASE	
NICHOLAS SCHORK, THE TRANSLATIONAL GENOMICS RESEARCH INSTITUTE	9
METHODOLOGICAL ISSUES IN INTEGRATING DIVERSE OMICS AND PHENOTYPE DATA	
JAMES SIKELA, UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS.....	10

OLDUVAI DOMAINS, HUMAN BRAIN EVOLUTION AND NEOTENY	
PETER SUDMANT, UNIVERSITY OF CALIFORNIA, BERKELEY	10
CAUSES AND CONSEQUENCES OF STRESS IN AGING, SOMATIC MUTATION, AND GENETIC ADAPTATION IN LONG- AND SHORT-LIVED SPECIES	
KAREN VASQUEZ, UNIVERSITY OF TEXAS AT AUSTIN	10
THE IMPACT OF AGING ON GENETIC INSTABILITY AT ENDOGENOUS MUTATION HOTSPOTS IN MICE	
CHAOLIN ZHANG, COLUMBIA UNIVERSITY	11
ALTERNATIVE SPLICING IN PRIMATE BRAIN EVOLUTION AND A CURIOUS CASE OF MAPT/TAUOPATHY	

Laura A. Cox, Wake Forest University School of Medicine

Integration of multiple-‘omics (transcriptomics, metabolomics, genomics) to identify molecular pathways and networks altered by developmental programming

It is well-established that a maternal obesogenic diet (MO) during pregnancy impacts offspring development, metabolism, and health, increasing risk of metabolic disorders later in life. However, underlying mechanisms altered in MO offspring leading to age-related metabolic changes are poorly understood. We studied a group of young adult baboons, a translational nonhuman primate (NHP) model of metabolism and aging to determine molecular and metabolic changes in MO offspring versus matched controls (CON). Animals were maintained on a healthy chow diet post-weaning, to identify liver molecular changes associated with liver lipid content and metabolic status due to MO. We generated untargeted liver biopsy omics data (transcriptomics, proteomics and metabolomics) and used cluster analysis to identify co-correlated molecules associated with MRS liver lipids and circulating metabolic measures. In CON, we found enrichment of molecules for hepatic fibrosis repair, serotonin, and glycation signaling pathways associated with liver lipid content. In MON, we found enrichment of TP53 activity, CLEAR signaling, thrombin, and chemokine signaling pathways. Less than 5% of MON lipid-associated molecules overlapped with CON. Interestingly, modules of molecules correlated with circulating lipids did not overlap with modules for liver lipid content in CON or MON, indicating different hepatic molecular networks involved with lipid within the liver compared to those in the circulation. In addition, our results show that developmental programming by MO significantly alters liver molecular signaling, which is not reflected in circulating measures of lipids. These findings raise the question of how well blood measures inform and guide diagnosis and treatment related to liver and metabolic disease for MO offspring. Future studies will investigate how these molecular alterations impact metabolic aging in these NHP as middle age and older adults. Grant Support: U19AG057758, P51OD011133.

Megan Dennis, University of California - Davis Health

Role of gene expansion and structural variation in primate evolution

Duplicated genes expanded in the human lineage likely contributed to evolutionarily unique traits in our species, such as brain evolution. As challenges exist in their discovery due to sequence-assembly errors, we used a complete telomere-to-telomere genome sequence to identify 213 human-specific gene families. From these, paralogs were found in all modern human genomes tested and brain transcriptomes, making them top candidates contributing to human-universal brain features. More globally we observe expression of human duplicated genes broadly across tissues and lifespan. Choosing a subset of paralogs, we used long-read DNA sequencing of hundreds of modern humans to reveal previously hidden signatures of selection. To characterize noncoding regulatory features of human duplications, we also devised

a new genomic approach coupling chromatin-conformation capture (3C) with HiFi long-read sequencing (CiFi). Our holistic approach provides new insights and a comprehensive resource for studying gene expansion drivers of human evolution.

Tim Donlon, PhD, University of Hawai'i

The Search for Primate Longevity Genetics: with special emphasis on FOXO3 gene-gene interactions

FOXO3 is a stress-response gene on chromosome 6. We have identified at least 18 SNPs (single nucleotide polymorphisms) that likely modify transcription factor binding in combination, that make of a “longevity haplotype”. Using *in silico* analyses, we show that the *FOXO3* promoter is physically connected via RNA polymerase II (RNAPII) and CTCF (CCCTC-Binding Factor) connections, with locations to these 18 or more SNPs in a cis-regulatory fashion. Following stress (200 μ M H₂O₂ & serum deprivation) we show that *FOXO3* expression is increased, with the longevity haplotype being activated at a higher rate. The *FOXO3* promoter is also connected to other distant genes on chromosome 6. This 7.3 Mb region is early replicating and has many sites of transcription initiation indicating that DNA synthesis/transcription machinery integrate this region as a “*FOXO3* interactome” or “gene factory”. Several genes, including *HACE1* (3 Mb proximal to *FOXO3*) were expressed at higher levels following cellular stress but were not correlated to the *FOXO3* haplotype. In FISH (fluorescence in situ hybridization) experiments, stress-induced activation of *FOXO3* caused it to move towards its neighboring genes *HACE1* and *LAMA4*, which are located at either distal flanking end of the 53+ gene neighborhood. This movement was confirmed with other genes closer to *FOXO3*. This chromatin movement is consistent with genes moving from a “B” compartment, where quiescent genes are located, to an “A” compartment, that is rich in transcription machinery.

This Interactome is highly conserved in other primates. Since the DNA in other primates has been sequenced the *FOXO3* gene can be directly compared between primate species, with regards to sequence variation and genome organization. *FOXO3* sequence comparisons can be made, at the base-pair level, across many primate species, comparing *FOXO3* promoter and other regulatory variations with differences in lifespan between colonies/sub-species.

Stress-response binding sites may enrich for the identification of additional longevity genes. The active *FOXO3* promoter is marked by regions that are rich in open chromatin, namely 1) DNase I hypersensitivity, 2) rich in histone 2aFZ (H2AFZ), 3) rich in histone 3 lysine 4 trimethylation (H3K4me3), 4) connections to polymerase II (POLR2A) binding, 5) NR2F1 binding sites, and 6) ATF7 binding sites. It appears that most regulatory regions are located in non-coding regions of the genome. These regulatory sites are unlikely to be conserved at the DNA sequence level in primates, however, it should be possible to identify/confirm longevity genes in non-human primates based on their patterns of transcription factor binding, as noted above.

Maximilian Haeussler, UC Santa Cruz Genomics Institute

UCSC Genome Browser

The UCSC Genome Browser provides a 30-way primate alignment, and not only the multiple alignments but also "chains" (alignable) and "net" (syntenic) tracks. During the talk, I explain how these are made and how the tracks can be used to find out what is alignable in a given human locus to primates and what is syntenic.

In addition, we serve two versions of the Zoonomia alignment, one that includes more than 200 primate genomes. I give a quick overview of how to use these.

Amanda J. Lea, Vanderbilt University

Comparative studies of genomic functional responses to environmental factors (GxE)

Genetic variation in non-coding regions of the genome plays a key role in complex trait variation within and between primate species. Further, there is increasing appreciation that genetic effects on gene regulation can vary across environments, cell types, and developmental states. These "context-dependent" genetic effects may be especially important for both evolution and disease, as they 1) may have reduced potential for pleiotropic, deleterious impacts relative to ubiquitous genetic effects and 2) are often associated with genes that serve developmental-, tissue-, or environment-specific functions. One powerful approach for mapping environmentally-dependent genetic effects is via a molecular quantitative trait loci framework, which identifies associations between genetic variants and gene regulatory traits (e.g., DNA methylation, gene expression) across different conditions—either experimentally imposed or naturally occurring. While this framework has been widely applied in humans, considerably less work has incorporated non-human primate species. I will discuss challenges to doing so and approaches we are taking to map environmentally-dependent genetic effects in both humans and non-human primates. Expanding knowledge in this area will enhance our ability to interpret both conserved and lineage-specific functional divergence across primates, particularly in perturbed cell states that may be especially relevant to disease.

Olivier Lichtarge, Baylor College of Medicine

Evolutionary action and evolutionary trace methods to identify functional and fitness effects of genetic divergence

Variation and selection shape the molecular genetics of every organism, yet remain difficult to model quantitatively as they are random, context-dependent, complex, and species-specific. We address this gap through a first-principles model based on statistical physics that defines the fitness effect of coding variants as mutational energy, derived from evolutionary information. Across diverse species and populations, we show that mutational states obey the laws of statistical mechanics, enabling us to quantify the relative mutational sensitivity—or

importance—of genes in populations of interest. In practice, shifts in gene importance and mutational energy burden between populations suggest genes and pathways associated with complex diseases, and, in preliminary data, with lifespan and primate divergence.

Glennis Logsdon, University of Pennsylvania Perelman School of Medicine

Centromere variation among human and non-human primates

Centromeres are essential chromosomal regions that ensure the faithful inheritance of genetic information. In primates, centromeres are composed of near-identical tandem repeats known as α -satellite, which are organized into arrays that can span up to several megabase pairs on our chromosomes. The repetitive nature and size of these arrays have made them challenging to resolve at the sequence level, and as a result, all centromeres have remained missing from current non-human primate reference genomes. In this talk, I will present the first complete map of all centromeres from five non-human primate species, including bonobo, chimpanzee, gorilla, Sumatran orangutan, and Bornean orangutan. I will reveal their variation in size, sequence, and structure within and among species, and I will debut the location of the kinetochore, a proteinaceous complex that serves as the site of microtubule attachment during cell division. Finally, I will show how the rapid evolution of centromeres leads to the emergence of new α -satellite repeats and the loss of others, resulting in unique centromere architectures specific to each species.

Joa Pedro de Magalhaes, University of Birmingham (UK)

Comparative genomics of ageing: Databases and approaches

In this talk, I will present our Human Ageing Genomic Resources (HAGR, <https://genomics.senescence.info>) a collection of online databases and tools designed to assist basic research on ageing. HAGR features several databases with high-quality manually-curated data: (i) GenAge, a database of genes associated with ageing in humans and in model organisms; (ii) LongevityMap, a database of human genetic variants associated with longevity; (iii) CellAge, a database of cell senescence-associated genes; (iv) DrugAge, a database of drugs, compounds and supplements that extend longevity in model organisms; (v) GenDR, a database of genes associated with the life-extension effects of dietary restriction; and (vi) AnAge, a collection of longevity records and complementary life-history traits for >4,000 species. Moreover, I will discuss genome sequencing of long-lived species and comparative genomics methods to identify genomic features and specific genes and pathways associated with the evolution of longevity in primates. The molecular and genetic basis of species differences in ageing remains a major mystery but one that if solved would provide important insights about the roots of the ageing process and human age-related diseases.

Arcadi Navarro, Universitat Pompeu Fabra, Barcelona

Biological processes modulating longevity across primates

Understanding the genomic foundations of longevity across primates can illuminate key biological mechanisms relevant to aging and age-related diseases in humans. In his talk I will present a comprehensive overview of my group's integrative research into the genetic basis of lifespan variation across primate species. I will detailed the development and application in primates of a powerful evolutionary framework to detect Convergent Amino Acid Substitutions (CAAS) in orthologous proteins, identifying over 2,000 genes associated with longevity quotient (LQ) —a metric correcting maximum lifespan for body mass— across more than 200 primate species. These genes, collectively referred to as CAAS+ genes, show statistically significant overlap with previously known longevity-associated genes and are enriched in pathways involved in neurodegenerative diseases, such as Alzheimer's and Huntington's disease.

My talk will highlight a few case studies, such as CUBN, which plays a role in vitamin B12 absorption and neuroprotection, and harbors a CAAS potentially altering its function. Also, I will present preliminary functional validations through overexpression and CRISPR-based editing in cell models in some genes, which confirmed differential apoptotic responses under stress, supporting the relevance of candidate mutations an supporting the notion of non-random, convergent evolution of longevity-related traits.

Rachel J. O'Neill, University of Connecticut

Repeat Dynamics in Primate Genomics

Mobile elements and highly repetitive regions are potent sources of lineage-specific genomic innovation and evolve rapidly, yet are integral to the structure and function of eukaryotic cells. However, our ability to discern the full repeat repertoire of any given genome has been hampered by technological limitations that often relegated repeats to unassembled contigs. Recent advances in long-read based genome assembly approaches now provide gapless reference genomes that include previously unknown and unexplored repeats, necessitating the development of new approaches in repeat identification, curation and annotation. As part of the Human Telomere-to-Telomere (T2T) consortium, we developed a framework to derive comprehensive repeat annotations for human genomes. Building on this success, we have applied these tools to several recently derived T2T ape genomes. In doing so, we have cataloged and modeled repeats across all repeat classes, identified elements still capable of mobilization, and determined the type and number of lineage specific mobile element insertions (MEIs) and expansions. Within apes, the gibbon lineage comprises ~20 species across four genera that have experienced rapid chromosome evolution. Targeting the eastern hoolock gibbon, we built a genome assembly and complete repeat annotations. Using several short and long-read sequencing approaches, including CUT&RUN, Fiber-seq, CpG methylation, DiMelo-seq, PRO-seq, and repli-seq, we identified a collection of genomic attributes with propensities for chromosome instability that shaped gibbon centromere and chromosome evolution. This type

of comparative multi-omics approach can provide insight into the diversity, distribution and evolution of repetitive regions that shape primate genomes and demarcate regions of species-specific genomic innovation.

Andreas Pfenning, Carnegie Mellon University

Combining comparative genomics and machine learning to study the genetics of lifespan

Brain aging is the biggest risk factor for a variety of other neurodegenerative disorders. The multi-faceted nature of aging, which involves epigenetic alterations, mitochondrial dysfunction, cellular senescence, and genomic instability, poses a challenge to identifying which pathways are driving the process. Here, applied comparative genomics and machine learning to identify loci, pathways, and cell types are driving brain aging by linking differences in lifespan across mammals to differences in their genetics. Rather than look for differences directly in genes, which only have limited variation, we have applied a new artificial intelligence approach developed by the Pfenning laboratory to find longevity-associated enhancers, distal regions that control the cell type-specific expression levels of genes. Our estimates of cell type-specific open chromatin conservation have been applied to find candidate enhancers associated with longevity

across the 240 mammals of the Zoonomia project. We have identified thousands of open chromatin regions associated with the evolution of longevity. In astrocytes, the implicated loci tend to be found near DNA damage response genes. In oligodendrocytes, the implicated loci tend to be found near genes involved in polycomb repression. Our results suggest that different neural cell types are under different types of selective pressures associated with longevity.

Katie Pollard, University of California San Francisco

Accelerated Regions

Human accelerated regions (HARs) are sequences that have been highly conserved through millions of years of vertebrate evolution and then changed dramatically in the human genome since divergence from our common ancestor with chimpanzees. This evolutionary signature suggests that HARs play important roles and that their functions may have been lost or changed in our ancestors, making HARs exciting candidates for understanding the genetic basis for what makes us human. However, it has been challenging to determine what HARs do and why the evolutionary forces constraining HAR sequences in other species suddenly changed in our lineage. In this talk, I will describe methods for identifying accelerated regions in any lineage using large multiple sequence alignments and machine learning approaches that have shed light on the evolutionary histories of HARs. These modeling approaches are generating new hypotheses about the fastest evolving regions in the human genome, which we are testing using high-throughput genomic tools for functional characterization of non-coding sequences. This machine-guided strategy exemplifies my vision for a proactive, rather than reactive, role for data science in biomedical research.

We now aim to apply our machine-guided strategy to identify genomic elements that regulate species-specific aging. Our initial project is a collaboration with the laboratory of Alex Pollen focused on the response of substantia nigra dopaminergic neurons to oxidative stress. I will introduce how we are using sequence-based deep learning models to identify human-specific oxidative stress responsive regulatory elements.

Gergens Polynice, CleMetric

Primate Aging Database

The Primate Aging Database (PAD) is a comprehensive collection of physiological and biological aging data obtained from non-human primates. It serves as a valuable tool for various research areas, including aging research, comparative biology, and longevity studies, providing easy access to data for the research community through its website, www.primatedatabase.org. The database is funded by the National Institutes of Health/National Institute on Aging (NIH/NIA) and managed by CleMetric.

PAD facilitates the study of aging patterns in primates and offers valuable data that contributes to human aging and biomedical research. It aids in the understanding of longevity, healthspan, and age-related diseases across different species and supports conservation efforts by monitoring primate health over extended periods. The database gathers data from various sources, including primate research centers, zoological institutions, and universities, utilizing both longitudinal studies and retrospective data analysis with standardized methodologies to ensure consistency and reliability.

Nicholas Schork, The Translational Genomics Research Institute

Methodological issues in integrating diverse omics and phenotype data

There are many issues that complicate association analyses using omics data gathered on different species. For example, accommodating phylogenetic relationships between species, including covariates in an analysis, handling outlying values, and dealing with clade specific effects all require assumptions that can be hard to verify. One particularly complicated issue concerns handling missing omics values on different species. Missing values can arise from problematic assays or reflect important underlying biology (e.g., gene loss events or lack of exposure to a particular exogenous metabolite). Differentiating the explanations for missing data is complicated. We discuss methods and applications for accommodating biologically-informative missing data. We also provide examples of the search of human-specific DNA sequences and genes, arguing that some data that are ‘missing’ in different species could shed light on species-specific phenotypes.

James Sikela, University of Colorado Anschutz Medical Campus

Olduvai domains, human brain evolution and neoteny

Found primarily in primates on NBPF genes, Olduvai (formerly DUF1220) protein domains are increasingly amplified generally as a function of a species evolutionary proximity to humans, where the greatest number of copies (~300) is found. The human Olduvai hyper-amplification was due to tandem additions of a three-domain unit called the Olduvai triplet. Most human copies map to the 1q21 region, and variations in Olduvai copy number have been linked to both brain evolution and cognitive disease.

Recent functional studies indicate that Olduvai expression down-regulates mitochondrial pathways. These and other findings suggest that Olduvai may act as a dosage-dependent driver of neoteny, prolonging developmental programs among primates, and most extremely in humans. In support of this possibility, Olduvai copy number among primates shows a strong linear correlation with increases in both brain size and lifespan. In agreement with such a role in delayed development, removal of Olduvai (e.g., in Olduvai-minus transgenic mice) produces a hyperactive phenotype. Recently reported advances in primate genomic sequencing and functional genomics should provide new opportunities to gain more insight into Olduvai's role in human developmental processes including brain function and aging.

Peter Sudmant, University of California, Berkeley

Causes and consequences of stress in aging, somatic mutation, and genetic adaptation in long- and short-lived species

What are the genetic causes and consequences of extreme variation in lifespan among vertebrates? We apply long-read sequencing approaches to assemble high-quality (near-T2T) genomes of non-model organisms with extreme lifespans, such as rockfish and bats, to understand how lifespan and associated life history traits have evolved. We use comparative genomics to identify putative genes, pathways, and regulatory elements, and characterize these using cellular models. Finally, we use population genetics approaches to understand how age structure in populations impacts patterns of selection and diversity. These types of studies have led us to uncover novel structural and single nucleotide variants that contribute to lifespan as well as ways in which patterns of genetic diversity have been impacted by these life history shifts.

Karen Vasquez, University of Texas at Austin

The Impact of Aging on Genetic Instability at Endogenous Mutation Hotspots in Mice

Genetic instability is a known hallmark of both aging and cancer, and cancer susceptibility increases with age. Alternative DNA-forming sequences (i.e. non-B DNA; e.g. H-DNA) are

endogenous sources of genetic instability and co-localize with mutation hotspots in human cancer genomes. For example, we have found that H-DNA-forming sequences colocalize with an are enriched at mutation hotspots in human cancer genomes, and they are intrinsically mutagenic, implicating them in cancer etiology. Here, we utilized a transgenic mutation-reporter mouse model containing either an H-DNA-forming sequence from a mutation hotspot in the c-MYC gene in Burkitt lymphoma or a control B-DNA-forming sequence to determine the impact of aging on H-DNA-induced genetic instability. We found tissue-specific effects of aging on H-DNA-induced mutagenesis, with mutation frequencies increasing in spleen and liver and decreasing in the brain. Mutation spectra analysis demonstrated an increase in large deletion mutations with age. We also identified tissue-specific modulation of H-DNA processing, including decreased nucleotide excision repair activity in the brain and increased cleavage of H-DNA structures in the spleen as a function of age. These findings provide insight into the link between aging and cancer, and into the role of DNA structure-induced genetic instability in age-associated cancer development.

Chaolin Zhang, Columbia University

Alternative splicing in primate brain evolution and a curious case of MAPT/tauopathy

Alternative splicing (AS) of precursor messenger RNA is a major driving force of the molecular diversity in mammals, but also a key contributor to accelerating evolution. However, molecular mechanisms leading to divergent AS events and their functional impact remain poorly understood. We systematically investigated AS of 42,000 cassette exons in primate brains as a quantitative trait and identified 1,170 exons with lineage-specific splicing shifts under stabilizing selection. Among them, we identified microtubule binding protein tau (MAPT) exon 10 which showed a two-step lineage-specific shift in primates. MAPT encodes the protein tau, which is frequently found in neurofilament tangles (NFT), a pathological hallmark of multiple neurodegenerative diseases including Alzheimer's, frontotemporal dementia (FTD), and Progressive supranuclear palsy (PSP), which are collectively referred to as tauopathies. MAPT exon 10 splicing is tightly regulated during brain development and maintains a balance of the inclusion (i.e., 4R) and skipping (i.e., 3R) isoforms in the adult human brain, which if disrupted, is sufficient to cause FTD. Intriguingly, our lineage-specific AS analysis found that MAPT exon 10 has nearly 100% inclusion in new world monkeys, ~75% inclusion in old world monkeys, and ~40-50% inclusion in hominoids including humans. This evolutionary shift cannot be explained by the previously identified splicing regulatory element, a stem loop structure, at the border of MAPT exon 10/intron 10 that has been implicated in the dysregulation of exon 10 splicing in FTD. Instead, we found a key splicing factor family, muscleblind-like (MBNL) proteins, as a major contributor to the splicing pattern divergence observed in MAPT exon 10 throughout the primate evolution. MBNL proteins are critical regulators of MAPT exon 10 developmental splicing switches by binding distal regulatory sequences in the downstream intron, which has

diverged across primate species. Importantly, we showed that by targeting the distal Mbnl binding sites using CRISPR/dCas13 or antisense nucleotide oligonucleotides (ASOs), we can effectively modulate exon 10 inclusion. Our results suggest an adaptation of MAPT function in primates during neurodevelopment and, on the other hand, may also have implications in neurodegeneration. Targeting the lineage-specific regulatory elements may provide new therapeutic avenues for FTD and other tauopathies.