Expanding the Non-Human Primate Reference Transcriptome Resource (NHPRTR): Tissue-Specific and Immune Cell-Specific Transcriptomes

Xinxia Peng1, Hao Xiong1, Lenore Pipes2, Sara Kelly1, Matthew Thomas1, Lynn Law1, Shujun Luo3, Stanley A. Langevin1, Steven Bosinger4,5, Guido Silvestri4,5, Muthuswamy Raveendran6, Donna Muzny6, Richard A. Gibbs6, Jeffrey Rogers6, Gary P. Schroth3, Christopher E. Mason2, and Michael G. Katze1

1University of Washington and Washington National Primate Research Center, 2Cornell University, 3Illumina, Inc., 4Yerkes National Primate Research Center, 5Emory University, 6Baylor College of Medicine Human Genome Sequencing Center

The Nonhuman Primate Reference Transcriptome Resource (NHPRTR, nhprtr.org) is a project that was initiated in mid-2010. The concept was to develop a NHP species-specific reference transcriptome resource by deep sequencing complete transcriptomes (RNA-seq) from multiple NHP species. Since then NHPRTR has successfully produced and released reference transcriptome data for 15 NHP species/subspecies, which were collected for each species using pools of RNA from about 20 comparable tissues. Here we describe three major new developments of NHPRTR: a) Updated macaque genome annotations. For rhesus and cynomolgus macaques, two of the most commonly used NHP models, we uncovered thousands of novel isoforms for annotated genes and thousands of unannotated intergenic transcripts enriched with noncoding RNAs. We also identified thousands of transcript sequences that are partially or completely ‘missing’ from current macaque genome assemblies. We showed that many newly identified transcripts were differentially expressed during simian immunodeficiency virus infection of rhesus macaques or during Ebola virus infection of cynomolgus macaques. b) RNA-seq analysis of NHP tissue-specific transcriptomes. For this effort, we selected 11 of the original 15 NHP species/subspecies and collected RNA-Seq data from ~15 tissues from each species. The completed tissue-specific RNA-seq dataset consists of over 10 billion paired-end raw reads. c) RNA-seq analysis of immune cell specific transcriptomes. This upcoming RNA-seq dataset will cover sorted immune cell subsets (B cells, Monocytes, NK cells, total CD4+ T cells, naïve CD4+ T cells, CD4+ central memory T cells, CD4+ effector memory T cells, and CD8+ T cells) from rhesus macaque, African Green Monkey, sooty mangabey, and human peripheral blood mononuclear cells. Together these expanded data sets serve as an invaluable NHP community resource and provide comprehensive reference transcriptomes for each species.