Analysis of RNA editing in non-human primates

M. Bozinoski[1,2], C.E. Mason[2]
[1]Department of Pharmacology, [2]Department of Physiology and Biophysics, Weill Cornell Medical College
1600 York Ave, New York, NY 10065, United States

Introduction

RNA editing is a critical, yet poorly understood post-transcriptional RNA modification, since it can modify the information content of a transcript without changing the underlying DNA, and thus is not directly visible in the genome sequence of an organism. It is mediated by the deamination of adenosine (A), turning it into inosine (I) by ADAR (adenosine deaminase acting on RNA) enzymes. Inosine (I) is then recognized as guanosine (G) by both the splicesome and the ribosome[1], and all other Watson-Crick base pairing interactions (i.e. ones with miRNAs).

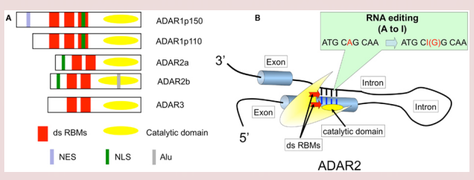

Fig1. A. ADAR isoforms B.RNA Editing
Nishikura K, Annu Rev Biochem 2010

The full biological significance of RNA editing is not yet understood. It is known that there is a good conservation of ADARs in vertebrates[2], and it has been shown that they are essential for normal development in mammals[2]. Aberrant RNA editing has been associated with a wide range of human diseases, including cancer[3], neurological[4] disorders, metabolic diseases[5], viral infections[6] and autoimmune disorders[5]. RNA editing has been shown to contribute to disease pathologies by editing of glutamate receptors[7], editing of serotonin receptors[6], and has roles in microRNA processing[2].

Since RNA editing is known to be a predominantly primate specific phenomenon[10], we focused on exploring the landscape of RNA editing in non-human primates, hoping to elucidate the full biological importance of RNA editing in such valuable animal models for human diseases.

Background

NHPRT Resource

Our lab has helped generate a database of the most comprehensive non-human primate transcriptomic data available to date, having obtained and published RNA-seq data from over a dozen non-human primates[8] as part of the Non-Human Primate Reference Transcriptome Resource (NHPRTR). Our data spans 15 species/ subspecies, including great apes, old world monkeys, new world monkeys and prosimians, spanning approximately 70 million years of primate evolution (Figure 2).

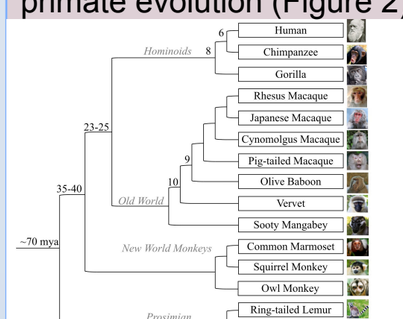

Fig 2. NHPRTR species dendrogram
Pipes, Li, Bozinoski et., al NAR 2012

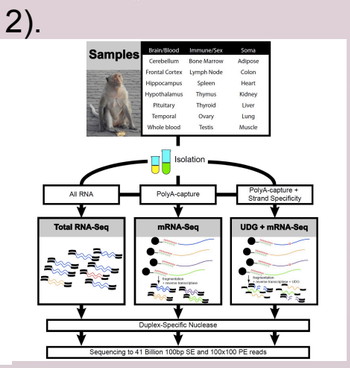

Fig 3. NHPRTR sequencing design
Pipes, Li, Bozinoski et., al NAR 2012

So far, we have published work on 21 pooled tissues from the species, using three different RNA preparation protocols (Total RNA, mRNA and Uracil-DNA glycosylase (UDG) (Fig3). Additionally, In the past year, we have sequenced 14 specific tissues from each of the species (Table 1, right), using the stranded, total RNA protocol, achieving average depth of ~50M per tissue for each animal[9]. Genomic DNA for all of the animals used to obtain the transcriptome data has been sequenced as well.

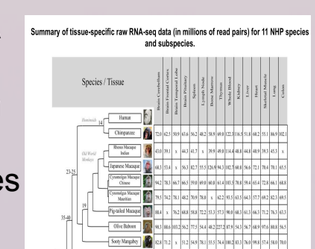

Table 1. Tissue specific sequencing
Peng X et al. Nucl. Acids Res. 2014

Methods

Reads from both RNA-Seq and DNA sequencing have been first trimmed at a fairly stringent threshold for quality (Q>35 and Q>25, respectively), thereby eliminating the low quality base calls at the reads 3' ends. Subsequently, reads were aligned to both human genome (hg19, Gencode V19 annotation) and the genome of the respective primate species (or to the genome of the phylogenetically closest species) . STAR aligner was used in case of RNA reads, while DNA reads were aligned using BWA. Bona fide RNA editing sites show variable portion of A and G reads in the transcriptome reads, but strictly A in the genomic reads (Fig 4) . RNA editing sites (variant calling) were called using two different software packages, Samtools and GATK, and their overlap was taken as RNA editing reliable sites. In order to exclude events that derive false positives, such as SNPs or heterozygosity, we discarded any sites that show bi or polyallelic variants in the genomic alignments.

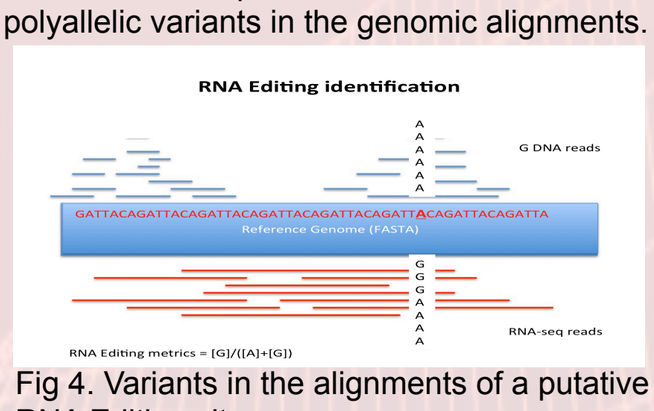

Fig 4. Variants in the alignments of a putative RNA Editing site

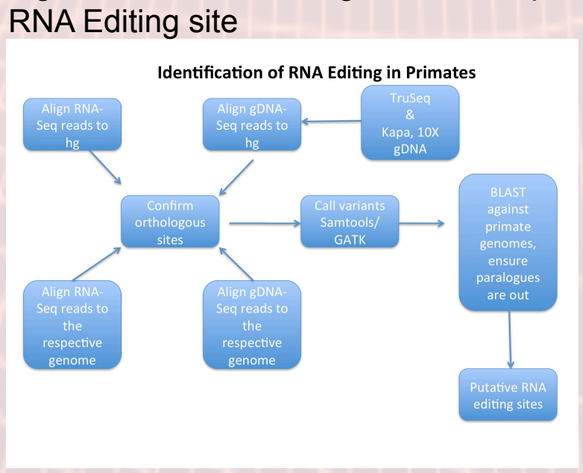

Fig 5. Workflow of obtaining RNA Editing sites from the data.

Having in mind the depth of sequencing that we have, we set a threshold of 10 reads per site(after adjustment for sequencing depth). The criteria for identifying an RNA-editing site was be observance of that least 5% G's in the RNA-reads variants, and exclusively adenosines in the genomic reads. As reads were aligned to both human and primate genomes, orthologs were pinpointed utilizing reciprocal BLAST of alignment from hg19 against already published primate genomes. For sequences with single, high confidence hits (e-value of -20 distance to the next match), we flagged these hits as the likely orthologs. Sequences with reported multiple alignments were be processed through an additional pipeline, in order to discriminate the real ortholog from various types of paralogs (gene duplications, pseudogenes, processed pseudogenes), on the basis of sequence identity, d_N/d_S value (ratio of non-synonymous/synonymous substitutions) and/or presence of truncated 5' UTRs (common for processed pseudogenes).

Putative RNA editing sites found in the orthologs were confirmed with DARNED and RADAR databases of human RNA editing sites, and with sub-sequential usage of data from the Gtex repository for an empirical validation of human editing sites in tissue-specific manner.

Results

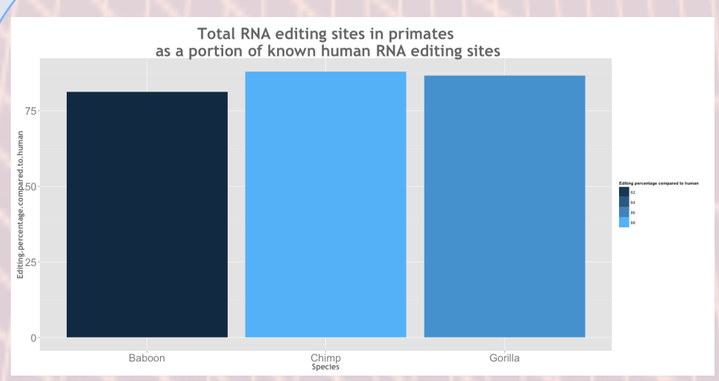

Fig. 6 Total number of primate specific RNA Editing sites, shown as a portion of annotated human RNA editing sites

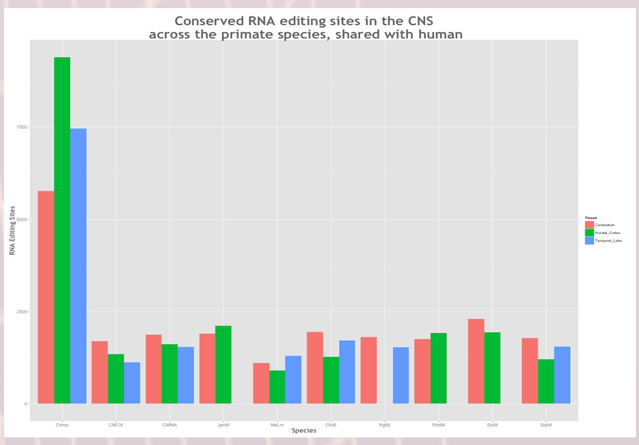

Fig. 7 Number of Primate specific RNA Editing sites per brain region, shared with human

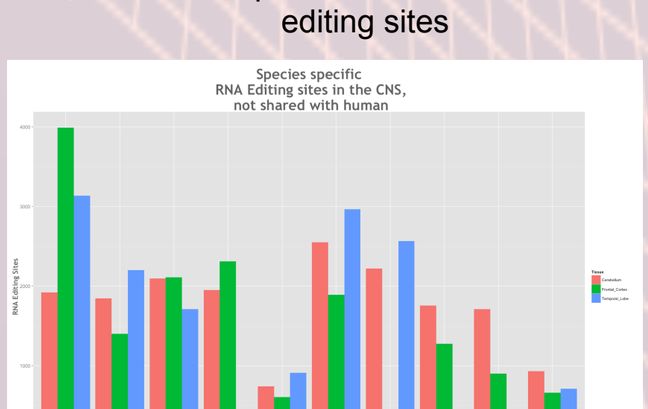

Fig. 8 Number of Primate specific RNA Editing sites per brain region, not observed in human

Abbreviations for Figures 7,8 and 9
Chmp-Chimp,
CMCN-Cynomogus Macaque Chinese
CMMA-Cynomogus Macaque Mauritian
JpnM-Japanese Macaque
MsLm-Mouse Lemur
OlvB-Olive Baboon
PgtM-Pig-tailed Macaque
RhsM-Rhesus Macaque
StyM-Sooty Mangabey
SqrM-Squirrel Monkey

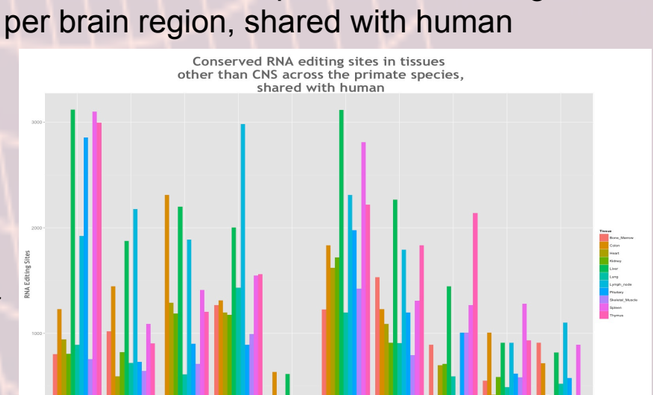

Fig. 9 Number of Primate specific RNA Editing sites per tissue (not CNS), shared with human

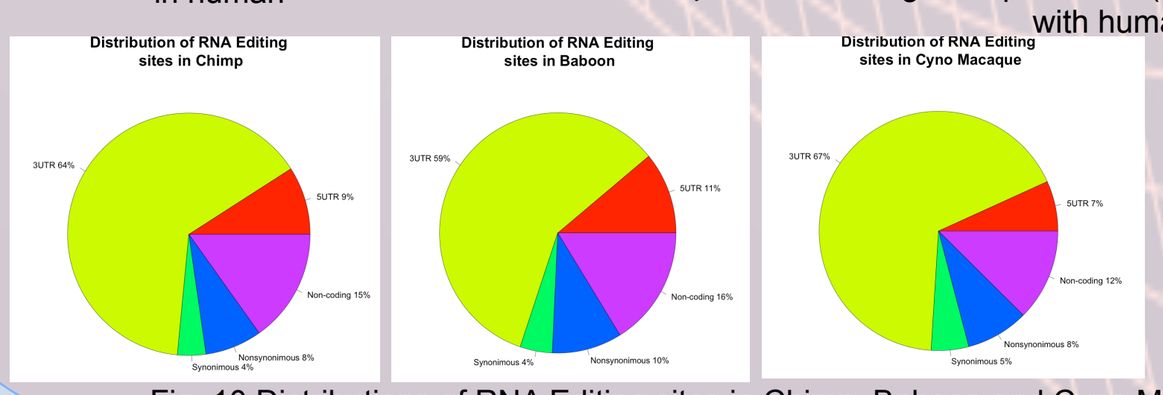

Fig. 10 Distributions of RNA Editing sites in Chimp, Baboon and Cyno Macaque

Conclusions

Non-human primates have less RNA editing than human. Our study shows that primates indeed have less editing in their respective orthologous transcripts. In fact, we see that the total editing sites found in Chimp, Gorilla and Baboon are less numerous than editing sites observed in human. Interestingly, the amount of RNA editing drops down the evolutionary tree, with Chimp having the most, Gorilla less, and Baboon having the least number of editing sites.

Phylogenetically closer species show more human-like RNA editing pattern, especially in CNS. Analysis shows Chimp having the most human-like profile of RNA editing, with a large spike in the cerebral cortex. Other primates show significantly less editing in CNS. Primate specific RNA editing sites, not observed in human, are again the highest in Chimp, followed by Baboon. Editing in other, non CNS tissues follows a species-specific pattern.

RNA editing demonstrably diversifies the transcriptome. Editing in coding regions derives more non-synonymous than synonymous codons (2:1 ratio).

RNA editing has a significant regulatory potential.Consistently large amount of editing in non-coding transcripts accentuates the role of RNA editing in regulation.

References

1. Nishikura K. 2010. Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem 79: 321-349. 9
2. Bass BL (2002) RNA editing by adenosine deaminases that act on RNA. Annu Rev Biochem 71:817–846
3. Galeano, Federica, et al. "A-to-I RNA editing: The "ADAR" side of human cancer." *Seminars in cell & developmental biology*. Vol. 23. No. 3. Academic Press, 2012.
4. Hoopengardner B, Bhalla T, Staber C, Reenan R (2003) Nervous system targets of RNA editing identified by comparative genomics. Science 301: 832–836.
5. Slotkin, William, and Kazuko Nishikura. "Adenosine-to-inosine RNA editing and human." (2013).
6. Kohler M, Burnashev N, Sakmann B, Seeburg PH (1993) Determinants of Ca2+ permeability in both TM1 and TM2 of high affinity kainate receptor channels: diversity by RNA editing. Neuron 10: 491–500
7. Lomeli H, Mosbacher J, Melcher T, Hoger T, Geiger JR, et al. (1994) Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. Science 266: 1709–1713.
8. Pipes, Lenore, et al. "The non-human primate reference transcriptome resource (NHPRTR) for comparative functional genomics." *Nucleic acids research* 41.D1 (2013): D906-D914.
9. Peng, Xinxia, et al. "Tissue-specific transcriptome sequencing analysis expands the non-human primate reference transcriptome resource (NHPRTR)." *Nucleic acids research* (2014): gku1110.
10. Paz-Yaacov N, . , et al Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. Proc Natl Acad Sci U S A. 2010;107:12174–12179