

## Diversity of microRNAs in human and chimpanzee brain

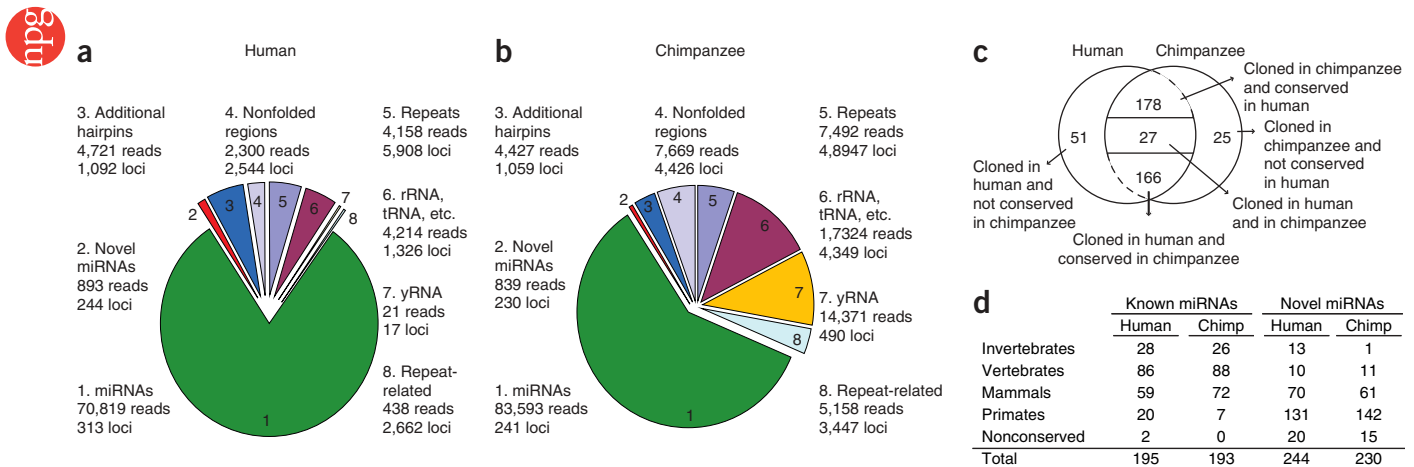
Eugene Berezikov<sup>1</sup>, Fritz Thuemmler<sup>2</sup>, Linda W van Laake<sup>1,3</sup>, Ivana Kondova<sup>4</sup>, Ronald Bontrop<sup>4</sup>, Edwin Cuppen<sup>1</sup> & Ronald H A Plasterk<sup>1</sup>

We used massively parallel sequencing to compare the microRNA (miRNA) content of human and chimpanzee brains, and we identified 447 new miRNA genes. Many of the new miRNAs are not conserved beyond primates, indicating their recent origin, and some miRNAs seem species specific, whereas others are expanded in one species through duplication events. These data suggest that evolution of miRNAs is an ongoing process and that along with ancient, highly conserved miRNAs, there are a number of emerging miRNAs.

miRNAs are ~22-nt RNA molecules that are processed from larger hairpin precursors and can regulate gene expression<sup>1</sup>. The role of miRNAs in diverse developmental processes and disease is increasingly recognized<sup>2</sup>. Several hundred miRNA genes have been identified by sequencing of size-fractionated small RNA libraries in human and other vertebrates<sup>3</sup>, and computational analyses have indicated that there could be substantially more miRNAs in the human genome<sup>4,5</sup>.

To discover miRNAs that have escaped cloning in previous studies, we applied massively parallel sequencing technology<sup>6</sup> to obtain a total

of ~400,000 reads from small RNA libraries prepared from human fetal brain and different regions of adult chimpanzee brain. (All experiments were approved by the ethics committee of University Medical Center Utrecht, and written permission was obtained for the use of human fetal tissue.) We developed a computational pipeline to process the sequencing data and to distinguish miRNA candidates from other types of small RNAs (**Supplementary Methods** online). We mapped more than 87,000 human reads and 140,000 chimpanzee reads to the respective genome sequences and retrieved annotations for the mapped loci from the Ensembl database. Known miRNAs represented 80% and 60% of the reads in the human and chimpanzee library, respectively (**Fig. 1**). The fraction of reads derived from repeat elements, rRNAs, tRNAs, other noncoding RNA and repeat-related regions was three times higher in the chimpanzee library (31% versus 11%). This category includes repeat-associated small interfering RNAs (rasiRNAs) that, in *Drosophila melanogaster*, originate predominantly from the antisense strand of repeats<sup>7</sup>. In our data set, different repeat types showed different strand biases: L2 and CR1 long interspersed nuclear elements (LINEs) had more antisense small RNAs, whereas L1 LINEs and Alu repeats had more small RNAs mapping to the sense strand (**Supplementary Table 1** online). Analysis of the read distribution across the genome shows that a limited number of clusters with more than 1,000 reads within a 100-kb region could be the source of most rasiRNAs in our libraries (**Supplementary Fig. 1** online). However, owing to their repetitive nature, it is often not possible to determine a unique chromosomal location of rasiRNAs, and thus the exact origin and a potential clustered expression of these elements from primate genomes remains undefined.



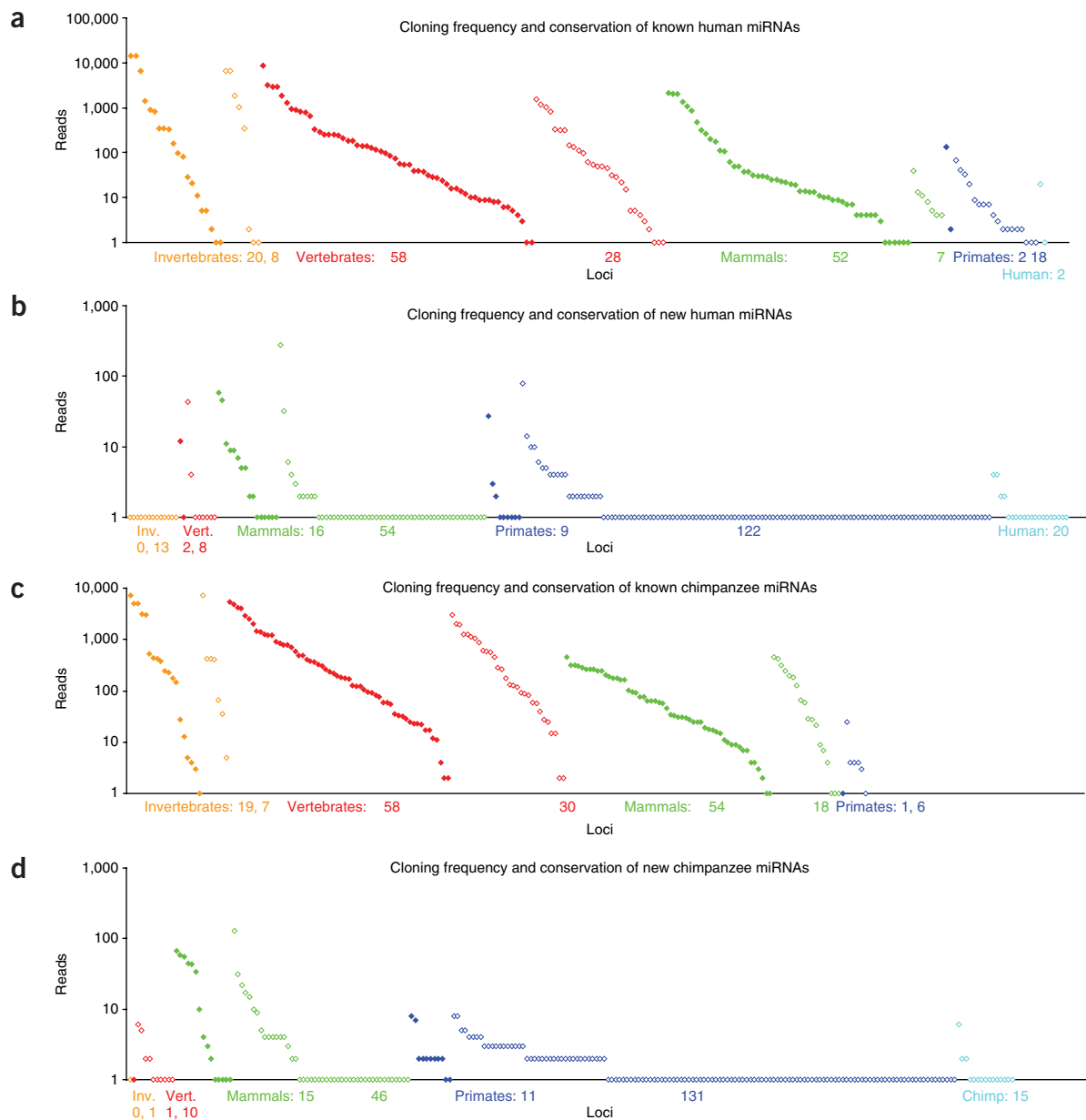
**Figure 1** Human and chimpanzee small RNA libraries. **(a,b)** Composition of **(a)** human and **(b)** chimpanzee brain small RNA libraries. **(c)** Number and overlap between new miRNAs cloned from human and chimpanzee libraries. **(d)** Conservation of cloned miRNAs across different phyla.

<sup>1</sup>Hubrecht Laboratory–KNAW, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. <sup>2</sup>Vertis Biotechnologie AG, Lise-Meitner-Strasse 30, D-85354 Freising-Weihenstephan, Germany. <sup>3</sup>Department of Cardiology, University Medical Center Utrecht, 3508 GA Utrecht, The Netherlands. <sup>4</sup>Biomedical Primate Research Centre, Lange Kleiweg 139, Rijswijk, The Netherlands. Correspondence should be addressed to R.H.A.P. (plasterk@niob.knaw.nl).

Received 27 July; accepted 28 September; published online 29 October 2006; doi:10.1038/ng1914

Noticeably, >10% of the chimpanzee reads corresponded to a family of noncoding RNAs termed 'yRNA' that may be involved in small RNA quality control through regulation of the activity of Ro protein<sup>8</sup>. In human, only 21 reads map to yRNA regions. This difference in expression level of yRNA genes is most likely explained by either species specificity or sampling differences (developmental stages and brain regions) rather than by differences in genome composition or annotation, as both the human and chimpanzee genomes contain a similar number of yRNA genes (815 and 772, respectively). Experimental artifacts, such as sample degradation, are also unlikely, as the fractions of reads in other categories, such as tRNA, were comparable in both samples.

Reads that do not correspond to known miRNAs, other noncoding RNAs or repeats comprised 9% of the reads in both libraries and originate from 3,880 and 5,715 distinct genomic regions in human and chimpanzee genomes, respectively. From these, 1,336 human and 1,289 chimpanzee regions had the predicted hairpin structure characteristic of miRNAs and thus can be considered as new candidate miRNA genes. We used an additional computational filter, randfold<sup>9</sup>, to discriminate miRNAs from other structured RNAs. Whereas 92% of known human miRNAs have randfold values <0.005, only 18% of new candidate miRNAs met this criterion, resulting in 244 human and 230 chimpanzee regions that are very likely to be real miRNAs (Supplementary Table 2 online).



**Figure 2** miRNA cloning frequency and conservation. (a) Known human miRNAs. (b) New human miRNAs. (c) Known chimpanzee miRNAs. (d) New chimpanzee miRNAs. Every miRNA genomic locus is represented by a rectangle. Filled rectangles correspond to miRNAs cloned in both human and chimpanzee; open rectangles represent miRNAs cloned only in one of the species. miRNAs with different degrees of conservation are grouped and shown in different colors.

Cloning frequency of small RNAs generally reflects their relative abundance<sup>10</sup> and can be used to establish miRNA expression profiles. Sampling of nonoverlapping human and chimpanzee brain regions from different developmental stages allowed us to maximize chances of detecting new miRNAs, but it provided only limited opportunity to compare miRNA expression profiles between human and chimp. Overall, the new miRNAs were expressed at low levels, with only a few miRNAs represented by more than one read (Fig. 2). However, intersection of human and chimpanzee data showed that 11% of the new miRNAs cloned from the human library were also cloned from the chimpanzee library (27 of 244 loci, Fig. 1c), even though half of these miRNAs were cloned only once in each library. Although cross-species expression confirmation rate was higher in general for known, highly abundant miRNAs (68%), the number of miRNAs that were cloned only once from both libraries was similar for the known and new miRNAs (5%), providing additional support for the validity of low-abundance miRNAs.

To assess the conservation of cloned miRNAs, we searched the genomes of 17 animal species for homologous hairpins, and we assigned the conservation level based on the most distant species in which a homolog was found (Fig. 1d). Seventy-five percent of known human miRNAs cloned in this study were conserved in vertebrates and mammals, 14% were conserved in invertebrates, 10% were primate specific and 1% are human specific. The new miRNAs have a different conservation distribution: more than half of the human miRNAs were conserved only in primates, about 30% in mammals and 9% in nonmammalian vertebrates or invertebrates; 8% were specific to humans. We saw a similar distribution for the chimpanzee miRNAs.

We used homology searches to establish all orthologous human-chimpanzee miRNA pairs. Some of the new miRNAs were members of existing or new families and were organized in genomic clusters, as is the case for known miRNAs (Supplementary Tables 3 and 4 online). Notably, we identified 14 chimpanzee miRNAs (including four known miRNAs) that had orthologous matches with multiple loci in the human genome, suggesting a species-specific family expansion (Supplementary Table 5 online). We also observed similar expansions in the chimpanzee genome (15 cases, including seven known miRNAs).

In total, we obtained experimental evidence for 447 new miRNAs (Fig. 1c). Although these miRNAs constitute only 1% of the small RNA transcripts in the tissues studied, they more than double the diversity of known miRNAs. Many of the new miRNAs are not conserved beyond primates, indicating their recent origin, and some miRNAs seem to be species-specific, whereas others have been expanded in one of the species through duplication events. These data suggest that evolution of miRNAs is an ongoing process and that along with ancient, highly conserved miRNAs, there is a group of emerging miRNAs, in line with previous observations in plants<sup>11</sup> and animals<sup>4</sup>. The different miRNA repertoire, as well as differences in expression levels of conserved miRNAs, may contribute to gene

expression differences observed in human and chimpanzee brain<sup>12</sup>. Although the physiological relevance of miRNAs expressed at low levels remains to be shown, it is tempting to speculate that a pool of such miRNAs may contribute to the diversity of developmental programs and cellular processes and thus provide evolution's playground for the development of new miRNA-containing regulatory pathways. For example, miRNAs recently have been implicated in synaptic development<sup>13</sup> and in memory formation<sup>14</sup>. As the species-specific miRNAs described here are expressed in the brain, which is the most complex tissue in the human body, with an estimated 10,000 different cell types<sup>15</sup>, these miRNAs could have a role in establishing or maintaining cellular diversity and could thereby contribute to the differences in human and chimpanzee brain evolution and function.

Note: Supplementary information is available on the Nature Genetics website.

#### ACKNOWLEDGMENTS

We thank the various genome sequencing consortia (see Supplementary Methods for full references) for sharing sequence information through Ensembl (<http://www.ensembl.org>) before publication and thank C. Reijmer (Vrelinghuis, Medical Center Biltstraat, Utrecht) for providing tissue material. This work was supported by grants from the Horizon (E.B.) and BioRange (E.C.) programs of the Netherlands Genomics Initiative, and a TOP grant (R.H.A.P.) from the Netherlands Organization for Scientific Research.

#### AUTHOR CONTRIBUTIONS

This study was designed by R.H.A.P., E.C. and E.B.; L.W.L. prepared human tissues; I.K. and R.B. prepared chimpanzee tissues; F.T. generated libraries; E.B. performed the analysis and wrote a first draft of the paper; E.B., E.C. and R.H.A.P. wrote the paper and I.K., R.B. and F.T. contributed critical comments on the manuscript.

#### COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the Nature Genetics website for details).

Published online at <http://www.nature.com/naturegenetics>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. Bartel, D.P. *Cell* **116**, 281–297 (2004).
2. Alvarez-Garcia, I. & Miska, E.A. *Development* **132**, 4653–4662 (2005).
3. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. & Enright, A.J. *Nucleic Acids Res.* **34**, D140–D144 (2006).
4. Bentwich, I. *et al. Nat. Genet.* **37**, 766–770 (2005).
5. Berezikov, E. *et al. Cell* **120**, 21–24 (2005).
6. Margulies, M. *et al. Nature* **437**, 376–380 (2005).
7. Vagin, V.V. *et al. Science* **313**, 320–324 (2006).
8. Stein, A.J., Fuchs, G., Fu, C., Wolin, S.L. & Reinisch, K.M. *Cell* **121**, 529–537 (2005).
9. Bonnet, E., Wuyts, J., Rouze, P. & Van de Peer, Y. *Bioinformatics* **20**, 2911–2917 (2004).
10. Aravin, A. & Tuschl, T. *FEBS Lett.* **579**, 5830–5840 (2005).
11. Allen, E. *et al. Nat. Genet.* **36**, 1282–1290 (2004).
12. Enard, W. *et al. Science* **296**, 340–343 (2002).
13. Schrott, G.M. *et al. Nature* **439**, 283–289 (2006).
14. Ashraf, S.I., McLoon, A.L., Sclarsic, S.M. & Kunes, S. *Cell* **124**, 191–205 (2006).
15. Muotri, A.R. & Gage, F.H. *Nature* **441**, 1087–1093 (2006).