

2013

Novel microRNA families expanded in the human genome

Zhi-Qiang Du
Iowa State University

Cai-Xia Yang
Iowa State University

Max F. Rothschild
Iowa State University, mfrothsc@iastate.edu

Jason W. Ross
Iowa State University, jwross@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/ans_pubs



Part of the [Agriculture Commons](#), and the [Animal Sciences Commons](#)

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/ans_pubs/228. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

This Article is brought to you for free and open access by the Animal Science at Iowa State University Digital Repository. It has been accepted for inclusion in Animal Science Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Novel microRNA families expanded in the human genome

Abstract

Background: Most studies on the origin and evolution of microRNA in the human genome have been focused on its relationship with repetitive elements and segmental duplications. However, duplication events at a smaller scale (<1 >kb) could also contribute to microRNA expansion, as demonstrated in this study. Results: Using comparative genome analysis and bioinformatics methods, we found nine novel expanded microRNA families enriched in short duplicated sequences in the human genome. Furthermore, novel genomic regions were found to contain microRNA paralogs for microRNA families previously analyzed to be related to segmental duplications. We found that for microRNA families expanded in the human genome, 14 families are specific to the primate lineage, and nine are non-specific, respectively. Two microRNA families (hsa-mir-1233 and hsa-mir-622) appear to be further expanded in the human genome, and were confirmed by fluorescence in situ hybridization. These novel microRNA families expanded in the human genome were mostly embedded in or close to proteins with conserved functions. Furthermore, besides the Alu element, L1 elements could also contribute to the origination of microRNA paralog families. Conclusions: Together, we found that small duplication events could also contribute to microRNA expansion, which could provide us novel insights on the evolution of human genome structure and function.

Keywords

Human, Genome, microRNA, Duplication, Evolution

Disciplines

Agriculture | Animal Sciences

Comments

This is an article from *BMC Genomics* 14 (2013): 1, doi:[10.1186/1471-2164-14-98](https://doi.org/10.1186/1471-2164-14-98). Posted with permission.

Rights

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

RESEARCH ARTICLE

Open Access

Novel microRNA families expanded in the human genome

Zhi-Qiang Du^{*†}, Cai-Xia Yang[†], Max F Rothschild and Jason W Ross

Abstract

Background: Most studies on the origin and evolution of microRNA in the human genome have been focused on its relationship with repetitive elements and segmental duplications. However, duplication events at a smaller scale (<1 kb) could also contribute to microRNA expansion, as demonstrated in this study.

Results: Using comparative genome analysis and bioinformatics methods, we found nine novel expanded microRNA families enriched in short duplicated sequences in the human genome. Furthermore, novel genomic regions were found to contain microRNA paralogs for microRNA families previously analyzed to be related to segmental duplications. We found that for microRNA families expanded in the human genome, 14 families are specific to the primate lineage, and nine are non-specific, respectively. Two microRNA families (hsa-mir-1233 and hsa-mir-622) appear to be further expanded in the human genome, and were confirmed by fluorescence *in situ* hybridization. These novel microRNA families expanded in the human genome were mostly embedded in or close to proteins with conserved functions. Furthermore, besides the Alu element, L1 elements could also contribute to the origination of microRNA paralog families.

Conclusions: Together, we found that small duplication events could also contribute to microRNA expansion, which could provide us novel insights on the evolution of human genome structure and function.

Keywords: Human, Genome, microRNA, Duplication, Evolution

Background

Although only ~22 nucleotides long, microRNA are vital to the developmental process of animals and plants through post-transcriptional gene regulation. The efficiency of microRNA transcription and stability is regulated in a tissue-specific manner [1-4]. After processing from precursor microRNA molecules, single-stranded microRNA can bind to 3'-UTR of messenger RNA via its seed region, in turn affecting mRNA translation or stability [1,5]. When aberrantly expressed, microRNA can lead to the progression of certain diseases [6,7]. They also participate in pathogen-host interaction [8], and exert systemic effects through intercellular trafficking [9]. Due to its significant roles in a wide variety of biological processes, genome evolution of microRNA structure and function have been studied extensively [1,2,10-18].

Both duplicated sequence fragments and repetitive elements in the genome could contribute to the expansion of microRNA families [10-18]. Species-specific microRNA expansion could have functional importance, and it has been demonstrated recently that an expanded microRNA cluster is fundamental to the maintenance of embryonic stem cell development in mice [13]. However, most duplication events studied for microRNA evolution are segmental or tandem duplications [15-17]. Segmental duplications are tremendously important in elucidating the evolution of protein-coding genes, and their roles in providing novel genomic mechanisms to cope with selection pressure and environmental changes [19-24]. However, segmental duplications are defined as >1 kb and >90% identity, and computational methods are designed accordingly and limited in studying only these sequences. Thus, small-scale duplication events are not covered, but which could also play important roles in the evolution of small molecules, especially for microRNA. The relevance and importance of these small

* Correspondence: zhqdu@iastate.edu

[†]Equal contributors

Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, 2255 Kildee Hall, Ames, IA 50011, USA

duplication events and their relationship with microRNA evolution has not been reported.

Here, we used a systematic approach, and found that small duplication events in the human genome could also contribute to microRNA expansion. In total, nine novel microRNA families were found to be expanded in the human genome, and additional new genomic regions were discovered to be related to the expansion of microRNA families reported previously. We found that novel microRNA family expanded in the human genome are located close to proteins with conserved function, and confirmed two of the microRNA expansion events by fluorescence *in situ* hybridization. These results could render us novel insights on the evolution of human genome structure and function.

Results

Novel expanded microRNA families in the human genome

To explore the relationship between short duplicated genomic sequences and the expansion of certain microRNA families in the human genome, we used repeat-masked reference genome for pairwise comparison (see Methods), since microRNA associated with repetitive elements have been well characterized [16-18]. We focused on detecting microRNA paralogs enriched in short duplication events (<1 kb) in the human genome, which were overlooked previously by studies on the role of segmental duplications (emerged recently, >1 kb and >90%

identity) to the origin and evolution of microRNA families [10-17].

In total, nine novel microRNA families were found to be enriched in short duplicated fragments (Additional file 1). We detected 26 microRNA families, previously found to be related to segmental duplications [16], in which new microRNA paralogs could also be enriched in short genomic fragments (Additional file 2).

For the nine novel microRNA families, the average identity score is 89.1%, slightly lower compared to microRNA enriched in segmental duplications (90.4%), and far lower than those already deposited in miRBase (99%). These nine novel microRNA families are derived from small duplications, with an average length around 430 bp, which could be the reason they were overlooked by previous segmental duplication analysis. Furthermore, through detailed analyses on local genomic regions, microRNA families previously identified by segmental duplication analyses to be expanded in the human genome could have new paralogs (Additional file 2, Figure 1). For instance, for the microRNA family hsa-mir-1233, multiple isolated and conserved genomic fragments shorter than 1 kb could contain potential microRNA paralogs (Figure 1).

Furthermore, our methods were able to detect 4 mitochondria-related microRNA families (Additional file 3), and 20 other microRNA families already deposited in miRBase (Additional file 4). However, three of the four mitochondria-related microRNA families, hsa-

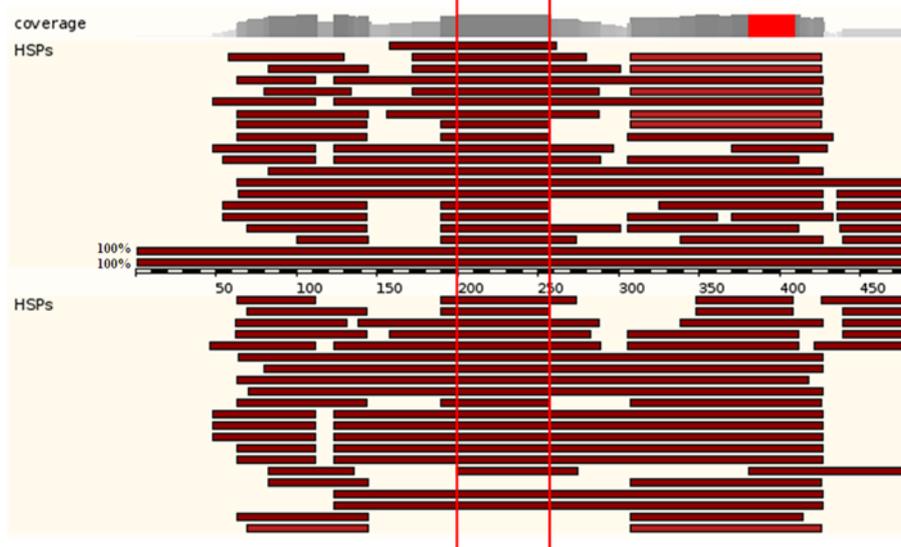


Figure 1 Detailed local genomic sequence analyses reveal new patterns of expanded microRNA families (hsa-mir-1233). Typical Ensembl sequence alignment picture reveals that novel paralogs will not be found if using only segmental duplication analysis, since it will detect only > 1 kb fragments of > 90% identity. In the region around 200–258 bp corresponding to conserved sequences to two microRNAs (in red frame), hsa-mir1233-1 and hsa-mir1233-2 (100% identity as indicated), we can see multiple genomic fragments, which are short and isolated, but highly conserved (percentage of identity can be found in Additional file 2). These genomic regions could contain potential microRNA paralogs. HSPs = high scoring pairs.

mir-1974, hsa-mir-1977, and hsa-mir-1978, were not further curated in miRBase, since they overlap with transfer-RNA (t-RNA) sequences. Despite the exclusion of these three microRNA in the miRBase, experimental evidence suggests that microRNA can be derived from t-RNA, and still demonstrate biological functions [25,26].

To see if these microRNA families expanded in the human genome were specific to the primate lineage, we did further comparison to 14 other mammalian genomes (see Methods). We found 16 primate-specific microRNA families (Figure 2a), and the microRNA cluster on chromosome 19 (C19MC) is the largest primate-specific cluster and may contribute to human reproduction, which has already been examined in detail [10,11,27]. The hsa-mir-1233 family appears to be further expanded in the human genome, compared to other primates. Furthermore, four novel microRNA families detected here were

also specific to primate genomes (hsa-mir-1826, hsa-mir-1827, hsa-mir-3185 and hsa-mir-492) (Figure 2a). In addition, nine microRNA families were found to be non-specific to primate genomes, in which four of them were novel (hsa-mir-622, hsa-mir-220, hsa-mir-199b and hsa-mir-1282) (Figure 2b). Investigation into the hsa-mir-622 family reveals that it further expanded in primates (Figure 2b), but with lower sequence conservation rates (88.0%).

Patterns of emergence of microRNA paralogs

We examined the location of these microRNA paralogs relative to protein-coding genes, the sequence conservation rate, and the flanking genomic sequences as well, to see if there exist certain patterns related to the expansion of these detected microRNA paralog families.

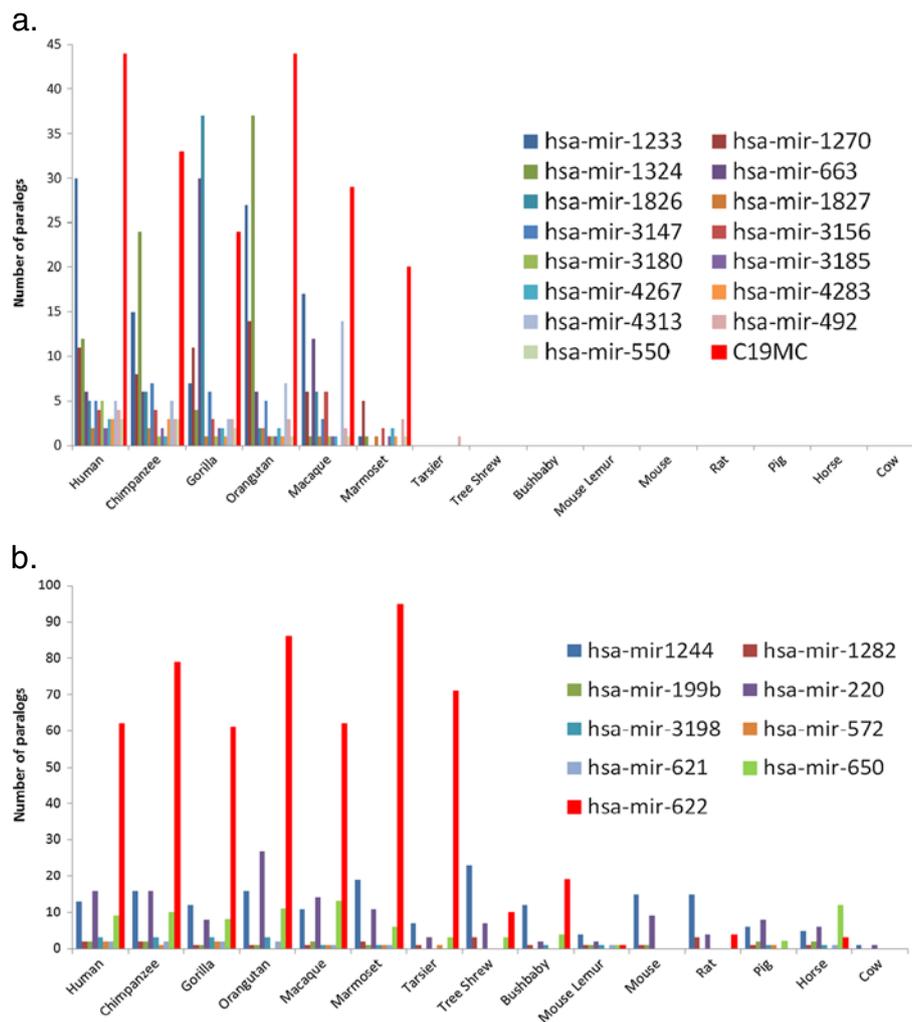


Figure 2 Primate lineage-specifically expanded microRNA families. a. 16 primate-specific families. The microRNA family on human chromosome 19 (C19MC) is the largest cluster identified previously. The hsa-mir-1233 mainly clustered on chromosome 15 expanded further in the human genome. **b.** Primate-nonspecific families (9). The hsa-mir-622 family seems to be further expanded in the primate lineage.

Nearly all of the duplicated genomic fragments containing microRNA paralogs were localized in intronic regions of protein-coding genes or non-coding transcripts, and only two of them were found in the 3'-untranslated region (UTR), which is in agreement with the origin of novel microRNA proposed previously [28] (Additional file 5). Evidence of cDNA expression for microRNA paralogs was found through public database searching (NCBI, see Methods). Furthermore, novel microRNA families expanded in the human genome were found to be embedded in proteins (host genes) mostly with conservative functions, such as tubulins and keratins (Additional file 5). For instance, the hsa-mir-1233 family was found to be related to golgi autoantigen, golgin subfamily a, 6 pseudogenes, which are expressed in fetal brain (hypothalamus) and embryonic stem cells (Additional files 5 and 6). Interestingly, we found that for the hsa-mir-1233 family, the conserved mature microRNA sequences (an 82-bp fragment) in all the paralogs were found to be missing from the cDNA sequences deposited in the database (Figure 3b). These 82-bp fragments in the paralogs could potentially be spliced out after processing precursor mRNAs, and subsequently function as microRNA.

Direct analyses on microRNA paralog sequence features in all expanded microRNA families found that mature

microRNA sequences and the seed regions of microRNA paralogs could be under the influence of different selective forces [29]. For instance, in two microRNA families, hsa-mir-1233 and hsa-mir-1244 (Figure 3, Additional file 7), we can clearly see that most of the microRNA paralogs had nearly identical mature sequences, while others demonstrate dissimilar features. These sequence changes could affect the stability of the stem-loop structure of the microRNA paralogs. Further examination on the flanking genomic regions of these microRNA families found they are highly conserved (Figure 3, Additional file 7).

Furthermore, the analysis on repetitive elements in the flanking genomic regions may contribute to the understanding of the origin of these microRNA families. Repetitive elements, such as Alu elements, were discovered previously to potentially affect the expansion of certain microRNA clusters as well as the genes located in duplicated genomic segments [11,30]. We found that Alu elements are also enriched in the flanking regions of microRNA paralogs (ANOVA) (Additional file 8). Surprisingly, the L1 elements have also been enriched in these sequences, which implied that they could also function in the evolutionary path of formation of certain microRNA paralogs. Furthermore, L1 elements had the highest percentage (21%), but not significantly different with AluS elements (*t*-test, *P* > 0.05).

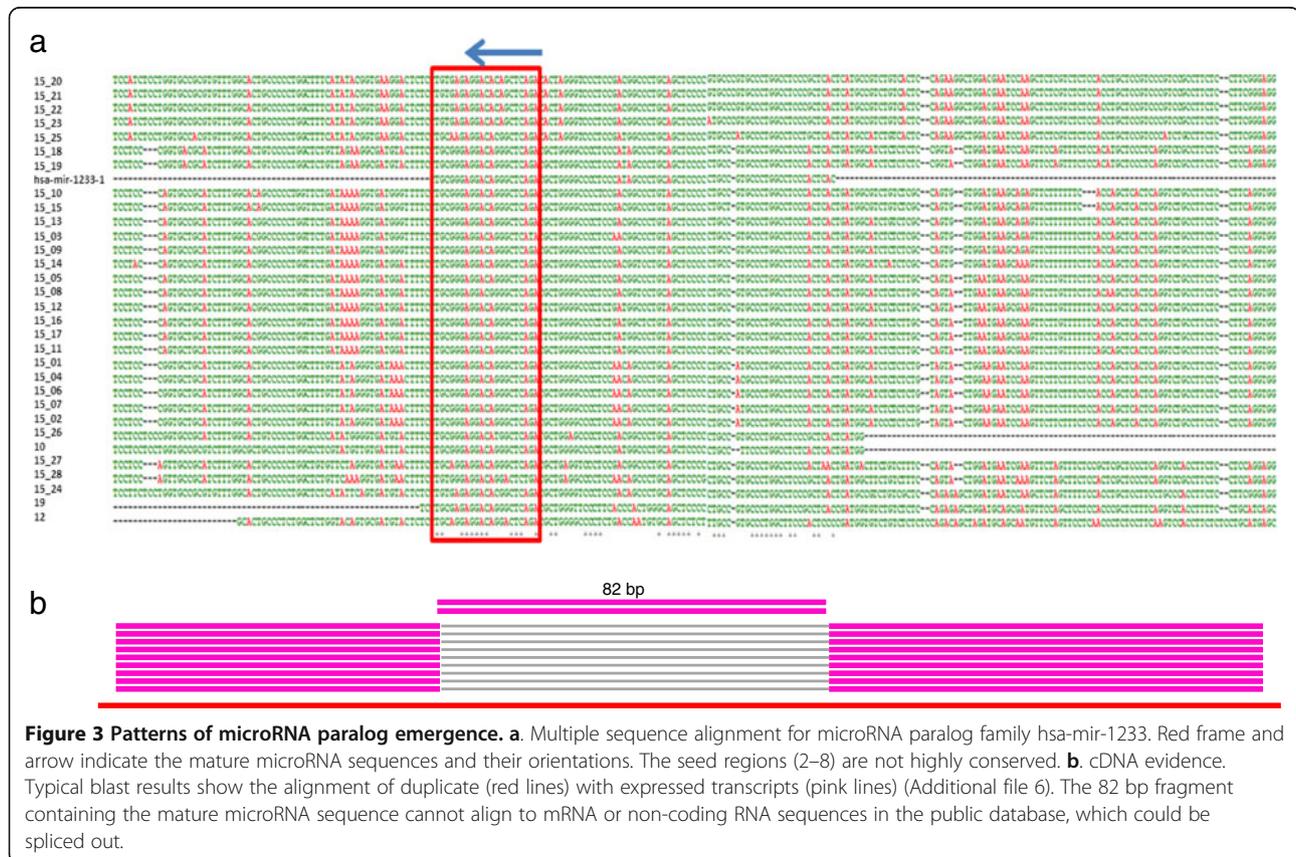
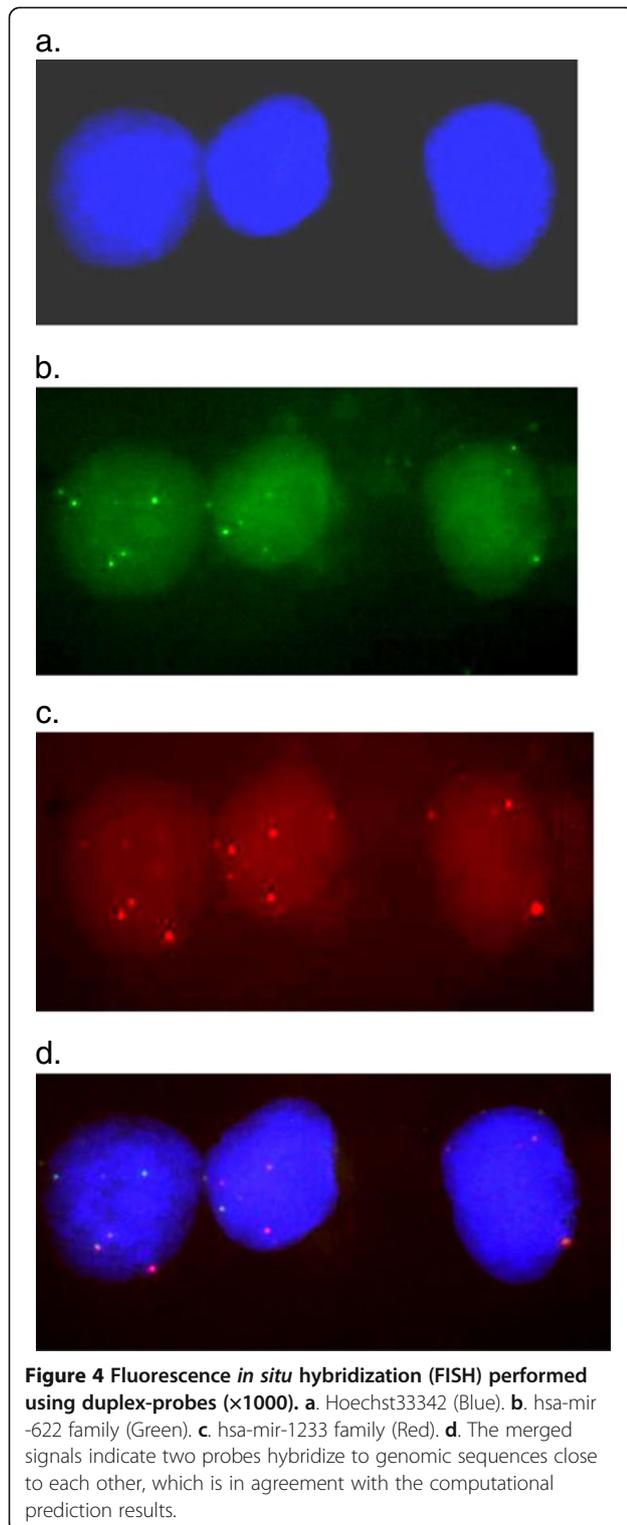


Figure 3 Patterns of microRNA paralog emergence. **a.** Multiple sequence alignment for microRNA paralog family hsa-mir-1233. Red frame and arrow indicate the mature microRNA sequences and their orientations. The seed regions (2–8) are not highly conserved. **b.** cDNA evidence. Typical blast results show the alignment of duplicate (red lines) with expressed transcripts (pink lines) (Additional file 6). The 82 bp fragment containing the mature microRNA sequence cannot align to mRNA or non-coding RNA sequences in the public database, which could be spliced out.



Validation of expansion of two microRNA families

We further confirmed our results for two microRNA families found to be expanded in the human genome, using fluorescence *in situ* hybridization (FISH) (Figure 4). One of them was identified previously (hsa-mir-1233),

and one novel microRNA family (hsa-mir-622) detected in this study. We can see that both probes hybridize to multiple locations in the human genome (Figure 4). Furthermore, our computational methods detected that both miRNA families have paralogs on the same chromosomes, 8, 9, 12 and Y (Additional files 1 and 2), which is evident in the merged hybridization signals being nearby each other (Figure 4).

Discussion

Using information provided by short duplication events in the human genome, we discovered novel microRNA paralogs for different microRNA families, which could have fundamental importance for human biology and the study of complex phenotypic traits. However, since the number of microRNAs deposited in the public database are far less than predicted, new microRNA paralogs and clusters may still exist in the human genome. Furthermore, this method can be extended for other non-coding RNAs, which can potentially reveal interesting patterns of specific important duplications in the human genome.

The study of microRNA from a comparative genomics standpoint is not without some limitations. The majority of the computational prediction methods use either precursor or mature microRNA sequences, combined with the examination of predicted secondary-structure [31]. The drawback of this approach is that only highly conserved potential microRNAs within or across species can be discovered. Experimental methods to date have only sampled limited tissues at few time-points, which will restrict the identification of novel microRNA [32]. Even though high-throughput small RNA sequencing has led to the identification of numerous novel microRNAs, the number and speed of discovery are still limited [33]. Our methods could to some degree help the identification and interpretation of novel microRNA paralogs in the human genome.

Alu elements have been proposed to be involved in the propagation of microRNA clusters and segmental duplications [11,30]. We also found in this study that AluS is the most abundant subfamily (17%), followed by AluJ (8%) and AluY (4%) (Additional file 8), which indicates that microRNA expansion events discovered here could follow a similar evolutionary path as reported in [11]. Furthermore, we provide evidence here that in addition to Alu elements, L1 repetitive elements are also associated with duplication of specific microRNA sequences. While numerous molecular mechanisms could be hypothesized for microRNA expansion, there remains a strong association that specific repetitive elements may have played important roles in the evolution of microRNA containing loci in humans. These duplicated paralogs could obtain partial and/or novel functions or exert dosage effects, acquire

novel regulatory elements for tissue-specific expression, possess modifications in seed sequence resulting in novel microRNAs, and potentially beneficial or detrimental to certain individuals or a specific population [29]. The conserved flanking sequences indicate that similar regulatory mechanisms could be involved in the tuning of these paralogs, but the changed sequences of the mature microRNA paralog could target different sets of genes and affect different gene clusters. However, detailed functional analysis is still needed to understand the outcome of the duplicated microRNA molecules [13].

We found that novel microRNA could potentially be generated through the processing of pseudogene transcripts. These expanded microRNA families in the human genome could provide novel insights on the evolution and regulation of the human genome and its interaction with the environment. Particularly, these expanded microRNA families could add another level of complexity into the regulatory network formed among messenger RNAs, transcribed pseudogenes, long noncoding RNAs, through the amplification of microRNA abundance [34].

Limited amounts of population resequencing data restricted our understanding of intra-species variation regarding one of the structural variations, the expansion of microRNA family, which has important practical implications (e.g. disease genetics). However, the rapid developments in sequencing technology will alleviate this problem in the near future, as already shown in the 1000 genomes project [35-37]. Further detailed analysis on the association of microRNA expansion with features unique to each ethnic group will potentially reveal their biological importance in human diversity.

Conclusions

Taken together, we found that small duplication events in the human genome may contribute to microRNA expansion, which could provide novel insights on the evolution of genome architecture, in addition to the development of human diseases.

Methods

Sequence alignments

The whole procedure for the analysis can be found in Additional file 9. The human genome and 25 other animal genome sequences with repetitive elements masked were downloaded from Ensembl (Additional file 10). We searched for genomic duplications by Megablast using default parameters [38]. All duplicated fragments were kept without restriction on length and identity of the sequence alignments, which is different from the detection methods for segmental duplications (length >1 kb and identity >90%) [30].

Enrichment of duplicated fragments containing microRNAs

Coordinates of human microRNAs were retrieved from miRBase (GRCh37) [39], and used to enrich for those duplicated genomic fragments overlapping with microRNAs using in-house Perl scripts. The extracted final coordinates were compared again to those of microRNAs deposited in miRBase, to find potential novel duplicated microRNAs, as well as their genomic locations with regard to exonic, intronic or untranslated regions of known genes.

Paralogs and orthologs of duplicated microRNAs

To search for paralogs and orthologs of microRNAs duplicated in the human genome, we retrieved the regional genomic sequences with repetitive sequences masked surrounding a representative microRNA deposited in miRBase, 3,000 base pairs (kb) upstream and downstream. The sequences were used to compare to the human genome sequence using Megablast, and an iterative procedure was used to enrich for genomic sequences related to microRNA paralogs. Furthermore, the retrieved genomic sequences were used to detect orthologs in other species using Megablast, including nine primates (*Pan troglodytes*, *Pongo pygmaeus*, *Gorilla gorilla*, *Macaca mulatta*, *Callithrix jacchus*, *Tarsius syrichta*, *Tupaia belangeri*, *Microcebus murinus*, and *Otolemur garnettii*), two rodents (*Rattus norvegicus* and *Mus musculus*), and three domestic livestock species (*Equus caballus*, *Bos taurus* and *Sus scrofa*) (Additional file 10). cDNA evidence was also searched in NCBI by using the retrieved genomic sequences containing microRNAs detected to be expanded in the human genome.

Repetitive elements in the flanking genomic regions and microRNAs

To detect human microRNAs in miRBase overlapping with the repetitive elements, we retrieved the genomic positions of the repetitive elements in the human genome from the UCSC genome table browser (group: variation and repeats, track: RepeatMasker) [40], and compared results to the coordinates of human microRNAs in miRBase. To further explore the relationship of repetitive elements with the origin of duplicated microRNA paralogs, the coordinates of repeats distributed in the regions of microRNA paralogs (upstream and downstream 5 kb regions) were compared and examined. Fractions of repetitive elements to the selected genomic sequences were calculated by their nucleotide lengths, and enrichment test was performed using Chi-square test, by comparing to the fraction of repetitive elements in the whole human genome. Secondary structure of the microRNA precursor was predicted using RNAfold [41].

Fluorescence *in situ* hybridization

To validate the results obtained by computational prediction, we selected two probes for the two microRNA families (hsa-mir-1233 and hsa-mir-622), and labeled them with TAMRA (red) and FITC (green), respectively (Additional file 11). FISH was performed on a normal primary neonatal dermal fibroblast cell line of European origin (PCS-201-010 from ATCC[®] Primary Cell Solutions[™]) following standard procedures (Creative[™] Biolabs) [42]. Due to the short probe length (~1 kb), we optimized the hybridization condition several times, until consistent results were obtained.

Additional files

Additional file 1: Nine novel microRNA families expanded in the human genome.

Additional file 2: Six microRNA families previously found in the human genome.

Additional file 3: Four expanded microRNA families related to mitochondria.

Additional file 4: Twenty expanded microRNA families deposited in the human genome.

Additional file 5: Expanded microRNA families and protein-coding genes.

Additional file 6: cDNA evidence for miRNA paralogs (hsa-mir-1233 family).

Additional file 7: Potential novel mechanism of emergence for microRNA family hsa-mir-1244.

Additional file 8: Repeat elements surrounding duplicated miRNA paralogs.

Additional file 9: Flowchart for computational analysis on animal microRNA expansion.

Additional file 10: 25 Genome sequences screened for miRNA expansion.

Additional file 11: Two probes used for FISH analysis.

Competing interests

The authors declare no competing interests.

Authors' contributions

ZQD, MFR and JWR designed the research. ZQD and CXY performed the experiments. ZQD and CXY analyzed data. ZQD, MFR and JWR wrote the manuscript. All the authors read and approved the final manuscript.

Acknowledgements

This project was supported in part by National Research Initiative Competitive Grant no. 2008-35205-05309 and 2008-35205-18712 from the USDA National Institute of Food and Agriculture. We also appreciate the financial support provided by the State of Iowa and Hatch Funding.

Received: 4 June 2012 Accepted: 31 January 2013

Published: 12 February 2013

References

1. Bartel DP: MicroRNAs: target recognition and regulatory functions. *Cell* 2009, **136**:215–233.
2. Li J, Liu Y, Dong D, Zhang Z: Evolution of an X-linked primate-specific microRNA cluster. *Mol Biol Evol* 2010, **27**:671–683.
3. Fabian MR, Sonenberg N, Filipowicz W: Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* 2010, **79**:351–379.
4. Krol J, Loedige I, Filipowicz W: The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* 2010, **11**:597–610.
5. Lee RC, Feinbaum RL, Ambros V: The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 1993, **75**:843–854.
6. Croce CM: Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* 2009, **10**:704–714.
7. Ryan BM, Robles AI, Harris CC: Genetic variation in microRNA networks: the implications for cancer research. *Nat Rev Cancer* 2010, **10**:389–402.
8. Santhakumar D, Forster T, Laqtom NN, Fragkoudis R, Dickinson P, Abreu-Goodger C, Manakov SA, Choudhury NR, Griffiths SJ, Vermeulen A, Enright AJ, Dutia B, Kohl A, Ghazal P, Buck AH: Combined agonist–antagonist genome-wide functional screening identifies broadly active antiviral microRNAs. *Proc Natl Acad Sci USA* 2010, **107**:13830–13835.
9. Lee YS, Pressman S, Andress AP, Kim K, White JL, Cassidy JJ, Li X, Lubell K, Lim do H, Cho IS, Nakahara K, Preall JB, Bellare P, Sontheimer EJ, Carthew RW: Silencing by small RNAs is linked to endosomal trafficking. *Nat Cell Biol* 2009, **11**:1150–1156.
10. Zhang R, Peng Y, Wang W, Su B: Rapid evolution of an X-linked microRNA cluster in primates. *Genome Res* 2007, **17**:612–617.
11. Zhang R, Wang YQ, Su B: Molecular evolution of a primate-specific microRNA family. *Mol Biol Evol* 2008, **25**:1493–1502.
12. Noguer-Dance M, Abu-Amero S, Al-Khtib M, Lefèvre A, Coullin P, Moore GE, Cavallé J: The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. *Hum Mol Genet* 2010, **19**:3566–3582.
13. Zheng GX, Ravi A, Gould GM, Burge CB, Sharp PA: Genome-wide impact of a recently expanded microRNA cluster in mouse. *Proc Natl Acad Sci USA* 2011, **108**:15804–15809.
14. Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC: Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat Genet* 2004, **36**:1282–1290.
15. Maher C, Stein L, Ware D: Evolution of Arabidopsis microRNA families through duplication events. *Genome Res* 2006, **16**:510–519.
16. Yuan Z, Sun X, Liu H, Xie J: MicroRNA genes derived from repetitive elements and expanded by segmental duplication events in mammalian genomes. *PLoS One* 2011, **6**:e17666.
17. Sun J, Zhou M, Mao Z, Li C: Characterization and Evolution of microRNA Genes Derived from Repetitive Elements and Duplication Events in Plants. *PLoS One* 2012, **7**:e34092.
18. Piriyaopongsa J, Jordan IK: A family of human microRNA genes from miniature inverted-repeat transposable elements. *PLoS One* 2007, **2**:e203.
19. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE: Recent segmental duplications in the human genome. *Science* 2002, **297**:1003–1007.
20. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwork C, Yang F, et al: Global variation in copy number in the human genome. *Nature* 2006, **444**:444–454.
21. Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG: Recent and ongoing selection in the human genome. *Nat Rev Genet* 2007, **8**:857–868.
22. Gao X, Lynch M: Ubiquitous internal gene duplication and intron creation in eukaryotes. *Proc Natl Acad Sci USA* 2009, **106**:20818–20823.
23. Studer RA, Penel S, Duret L, Robinson-Rechavi M: Pervasive positive selection on duplicated and nonduplicated vertebrate protein coding genes. *Genome Res* 2008, **18**:1393–1402.
24. Kassahn KS, Dang VT, Wilkins SJ, Perkins AC, Ragan MA: Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. *Genome Res* 2009, **19**:1404–1418.
25. Reese TA, Xia J, Johnson LS, Zhou X, Zhang W, Virgin HW: Identification of novel microRNA-like molecules generated from herpesvirus and host tRNA transcripts. *J Virol* 2010, **84**:10344–10353.
26. Qi Y, Tu J, Cui L, Guo X, Shi Z, Li S, Shi W, Shan Y, Ge Y, Shan J, Wang H, Lu Z: High-throughput sequencing of microRNAs in adenovirus type 3 infected human laryngeal epithelial cells. *J Biomed Biotechnol* 2010, **2010**:915980.
27. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, Sharon E, Spector Y, Bentwich Z: Identification of

- hundreds of conserved and nonconserved human microRNAs. *Nat Genet* 2005, **37**:766–770.
28. Campo-Paysaa F, Sémon M, Cameron RA, Peterson KJ, Schubert M: **microRNA complements in deuterostomes: origin and evolution of microRNAs.** *Evol Dev* 2011, **13**:15–27.
 29. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations.** *Genetics* 1999, **151**:1531–1545.
 30. Bailey JA, Liu G, Eichler EE: **An Alu transposition model for the origin and expansion of human segmental duplications.** *Am J Hum Genet* 2003, **73**:823–834.
 31. Lai EC, Tomancak P, Williams RW, Rubin GM: **Computational identification of Drosophila microRNA genes.** *Genome Biol* 2003, **4**:R42.
 32. Wu CI, Shen Y, Tang T: **Evolution under canalization and the dual roles of microRNAs: a hypothesis.** *Genome Res* 2009, **19**:734–743.
 33. Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu CI: **The birth and death of microRNA genes in Drosophila.** *Nat Genet* 2008, **40**:351–355.
 34. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP: **A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?** *Cell* 2011, **146**:353–358.
 35. 1000 Genomes Project Consortium: **A map of human genome variation from population-scale sequencing.** *Nature* 2010, **467**:1073.
 36. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, Eichler EE, 1000 Genomes Project: **Diversity of human copy number variation and multicopy genes.** *Science* 2010, **330**:641–646.
 37. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, Li Y, Lin CY, Luo R, *et al*: **Mapping copy number variation by population-scale genome sequencing.** *Nature* 2011, **470**:59–65.
 38. Zhang Z, Schwartz S, Wagner L, Miller W: **A greedy algorithm for aligning DNA sequences.** *J Comput Biol* 2000, **7**:203–214.
 39. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics.** *Nucleic Acids Res* 2008, **36**:D154–D158.
 40. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, Diekhans M, Dreszer TR, Giardine BM, Harte RA, Hillman-Jackson J, Hsu F, Kirkup V, Kuhn RM, Learned K, Li CH, Meyer LR, Pohl A, Raney BJ, Rosenbloom KR, Smith KE, Haussler D, Kent WJ: **The UCSC Genome Browser database: update 2011.** *Nucleic Acids Res* 2010, **39**:D876–D882.
 41. Hofacker IL: **Vienna RNA secondary structure server.** *Nucl. Acids Res.* 2003, **31**:3429–3431.
 42. Tjia WM, Hu L, Zhang MY, Guan XY: **Characterization of rearrangements involving 4q, 13q and 16q in hepatocellular carcinoma cell lines using region-specific multiplex-FISH probes.** *Cancer Lett* 2007, **250**:92–99.

doi:10.1186/1471-2164-14-98

Cite this article as: Du *et al.*: Novel microRNA families expanded in the human genome. *BMC Genomics* 2013 **14**:98.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

