Noncanonical microRNAs and endogenous siRNAs in normal and psoriatic human skin

Jing Xia¹,*, Cailin E. Joyce²,*, Anne M. Bowcock²,³,† and Weixiong Zhang¹,²,†

¹Department of Computer Science and Engineering, Washington University, St. Louis, MO, 63130, U.S.A
²Department of Genetics, Washington University School of Medicine, St. Louis, MO, 63110, U.S.A
³Division of Dermatology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, 63110, U.S.A
†Corresponding authors: Weixiong Zhang, Ph.D. Department of Computer Science and Engineering, Washington University in Saint Louis, One Brookings Drive, Saint Louis, MO 63130, Phone: 314-935-8788, Fax: 314-935-7302, Email: weixiong.zhang@wustl.edu. Anne M. Bowcock, Ph.D. Department of Genetics, Washington University School of Medicine in Saint Louis, 4559 Scott Avenue, Saint Louis, MO 63110, phone: 314-747-3261, fax: 314-747-2489, email: bowcock@genetics.wustl.edu
*These authors contributed equally to this work
ABSTRACT

Noncanonical microRNAs (miRNAs) and endogenous small interfering RNAs (endo-siRNAs) are key gene regulators in eukaryotes. Noncanonical miRNAs, which bypass part of the canonical miRNA biogenesis pathway, can originate from a variety of genomic loci, which include snoRNAs, tRNAs and introns, while endo-siRNAs can arise from repetitive elements, some of which are transposable. The roles of noncanonical miRNAs and endo-siRNAs in complex diseases have yet to be characterized. To investigate their potential expression and function in psoriasis, we carried out a comprehensive, genome-wide search for noncanonical miRNAs and endo-siRNAs in small RNA deep sequencing datasets from normal and psoriatic human skin. By analyzing more than 670 million qualified reads from 67 small RNA libraries, we identified 21 novel, noncanonical miRNAs (3 snoRNA-derived and 2 tRNA-derived miRNAs and 16 miRtrons) and 39 novel endo-siRNAs that were expressed in skin. The expression of four novel small RNAs was validated by qRT-PCR in human skin and their Argonaute association was confirmed by co-immunoprecipitation of ectopic small RNAs in HEK293 cells. Fifteen noncanonical miRNAs or endo-siRNAs were significantly differentially expressed in psoriatic involved versus normal skin, including an Alu-SINE-derived siRNA which was 17-fold upregulated in psoriatic involved skin. These and other differentially expressed small noncoding RNAs may function as regulators of gene expression in skin and potentially play a role in psoriasis pathogenesis.
INTRODUCTION

MicroRNAs (miRNAs) and endogenous small interfering RNAs (endo-siRNAs) are classes of small noncoding regulatory RNAs (sncRNAs), which play critical roles in gene regulation for many biological processes in eukaryotic organisms including mRNA cleavage, RNA degradation, translation inhibition, and DNA methylation (1). miRNAs in particular have been heavily studied in recent years and shown to be essential gene regulators that broadly contribute to disease pathogenesis.

miRNAs are produced from short hairpin-forming primary transcripts (pri-miRNAs) which are typically processed through a canonical pathway involving two RNase III enzymes, Drosha and Dicer, and a dsRNA binding protein Dgcr8 (2). Besides canonical miRNAs, there exist noncanonical miRNAs whose biogenesis requires Dicer but neither Drosha nor Dgcr8 (3). The first example of noncanonical miRNA is the class of miRtrons, which arise from short (60-100nt) debranched intron lariats that form stem-loop structures that serve as Dicer substrates (4–7). Noncanonical miRNAs can also arise from local hairpin formation within larger noncoding RNA (ncRNA) species, such as small nucleolar RNAs (snoRNAs) and transfer RNAs (tRNAs) (8, 9). While there have been several snoRNA-derived miRNAs described in eukaryotes, only one example of a tRNA-derived miRNA has been described. In this case, a murine tRNA, IleTAT folds into an alternative secondary structure consisting of a stem-loop hairpin from which mmu-miR-1983 is derived (10). In contrast to miRNAs, endo-siRNAs arise from long double-stranded RNA (dsRNA) transcripts derived from repetitive or transposable elements (11–13), which require Dicer, but not Drosha/Dgcr8 for processing (3, 10). Endo-siRNAs have been described in plants (1, 14), C. elegans (15), D. melanogaster (16–19), and murine embryonic stem cell (mESC) and oocytes (10, 12, 13). There is also evidence suggesting that endo-siRNAs are expressed in murine embryonic skin (20).

Both miRNAs and siRNAs exert their regulatory functions through association with the RNA induced silencing complex (RISC), which contains an Argonaute (AGO) protein (5, 21, 10). However, miRNAs and siRNAs have distinct requirements for mRNA target recognition. The base pairing between miRNAs and their targets is degenerate, while the pairing between siRNAs and their targets requires high fidelity, typically allowing no more than three mismatches (12). Consequently, most miRNAs lower the expression levels of target genes by mRNA destabilization or translational repression (22) while endo-siRNAs elicit direct cleavage of target mRNAs (12).

miRNAs are differentially expressed in myriad human diseases, and in some cases, play a direct role in disease etiologies, but little is known about the expression or function of noncanonical miRNAs and endo-siRNAs in human disease. In the case of the inflammatory skin disorder psoriasis, we have analyzed the expression of the majority of canonical miRNAs with small RNA library construction followed by deep sequencing (23). In the current study, we reanalyzed these deep sequencing datasets, focusing on the discovery and characterization of noncanonical miRNAs and endo-siRNAs, and identified 60 novel sncRNAs. Digital gene
expression analyses revealed differential expression of 15 sncRNAs in psoriatic involved skin versus normal skin. This study broadens our perspective on the diversity of sncRNAs in human skin and reveals a subset of noncanonical miRNAs and endo-siRNAs that are differentially expressed in psoriasis.

RESULTS

NextGen sequencing of small RNAs

Sequencing and pre-processing of 67 small RNA libraries from normal (NN), psoriatic uninvolved (PN) and psoriatic involved (PP) human skin, alignment of reads to the human genome and profiling of canonical miRNAs were described in detail previously (23). Here we systematically analyzed reads that aligned to introns, various types of noncoding RNAs (ncRNAs), repetitive elements, and transposable elements to identify read clusters that were consistent with expression of Dicer-dependent sncRNAs (Fig. S1).

Noncanonical miRNAs derived from snoRNAs and tRNAs

Overall, the small RNA reads derived from ncRNAs account for 26.4%, 30.2% and 26.8% of the qualified reads in PP, PN and NN skin, respectively (Table S1). While many of these reads reflect random degradation products from abundant ncRNAs, several read clusters exhibited characteristic features of Dicer-dependent sncRNA expression. In total, 58 million (8.8%) qualified reads aligned to snoRNAs, including 8/8 of the previously described miRNA-producing snoRNAs in humans (9; Table S2) and three novel miRNA-producing snoRNAs: U60, ACA25 and ACA44 (Table S2, Fig. 1). Folding structure analysis revealed that the snoRNAs could form miRNA-like hairpin structures (Fig. 1, S2A). Alignment of the most abundant small RNA reads to individual arms of the hairpins resulted in 22-nt RNA-RNA duplexes with 3’ overhangs (Fig. 1, S2B). Unexpectedly, ~79% of the reads aligned to snoRNAs aligned to U60, a C/D box snoRNA that has not been previously implicated as a miRNA locus. Of these reads, 98.9% were 21- or 22-nt long. Remarkably, the 22-nt miRNA-like RNA was ranked as the second most highly expressed miRNA in human skin based on read counts. However, qRT-PCR indicated a more moderate expression level in the skin (see below), perhaps indicating that the very high abundance of U60 reads in small RNA libraries was due to a library preparation bias (24, 25).

We next examined small RNA reads that aligned to tRNAs. First, we recognized a cluster of 228 21- or 22-nt sequencing reads aligning to human tRNA-IleTAT, which is consistent with expression of a mmu-miR-1983 homolog in human skin (Fig. S3A,B). A second tRNA candidate with short hairpin-forming potential, PseudoTTA, was associated with a cluster of 14 small RNA reads (Fig. S4C,D). We initially categorized this as a putative miRNA candidate since it was associated with so few sequencing reads and we did not detect miRNA* species, but we were later able to confirm its expression by qRT-PCR (see below). Detailed alignments of small RNAs derived from snoRNAs and tRNAs are provided as File S1.
miRNAs derived from introns – miRtrons

The miRNA subclass of miRtrons includes typical miRtrons which are 60 to 100-nt introns that form miRNA precursors, and tailed miRtrons, whose precursors are generated from one end of a longer intron (5, 26). Fourteen typical miRtrons have been previously described in humans (4) and were also detected in the current study (Table S3). Further, we identified an additional 26 known human miRNAs that are likely to be processed by the Drosha/Dgcr8 independent miRtron biogenesis pathway (2 typical miRtrons and 24 tailed miRtrons; Table S3). Although the currently annotated precursors for the majority of these miRNAs extend beyond their respective host intron regions, we found that, on average, 80% of the reads aligning to the intron/exon boundary terminated within 2-nt of the AG splice acceptor site, which provides strong evidence that these pre-miRNAs were derived directly from debranched introns (Fig. S4).

A search for novel miRtrons led to the discovery of 16 miRtron candidates, which were not reported in our previous study (23) due to their low abundance and/or atypical folding structures (Table S4). The lengths of the host introns varied from 70 to several thousand bases, suggesting that long introns can also give rise to miRtrons. In total, we detected 117 miRtron candidates in human skin, including 75 previously described miRtrons (4, 23), 26 known miRNAs which our new data suggests are processed by the miRtron biogenesis pathway, and 16 novel miRtrons; these can be further classified into 19 typical miRtrons and 98 tailed miRtrons.

In the set of known human miRtrons (4, 23), the most highly expressed in skin was miR-877, with 4,827 reads (Table S3). Four additional miRtrons were expressed at comparable levels to miR-877: miRtrons #80 and #103 (Fig. S2C,D), which we previously discovered (23), and miR-1292 and miR-3605, known miRNAs that we re-classified as miRtrons. Detailed read alignments for novel and re-classified miRtrons are provided as File S2.

Endo-siRNAs from repetitive and transposable elements

The expression of endo-siRNAs in adult human tissues is not well understood. Our dataset of deeply sequenced small RNAs accommodates a genome-wide search for putative endo-siRNAs in human skin. A total of 27 million (4%) qualified small RNA reads from human skin aligned to repetitive and transposable elements, such as SINEs, LINEs, and LTRs (Table S1).

The locus with the most aligned reads was on chromosome 1 within an intron of the gon-4-like (GON4L) gene, where three Alu-SINEs appear in tandem (Fig. 2A). A total of 5,534 reads (53.4% of which were 21-23nt) mapped to this locus; 2,800 of these reads (83.7% of which were 21-23nt) uniquely clustered within the first Alu-SINE element (E1; Fig. 2A), suggesting that this repetitive element may harbor a novel endo-siRNA. Notably, this Alu-SINE/GON4L endo-siRNA was the most strongly upregulated siRNA in psoriatic involved skin compared to normal skin (see below). In contrast, the expression level of the host gene GON4L showed little variation between psoriatic involved and normal skin (data not shown), indicating that the endo-siRNA and the host gene may be transcribed or processed independently. To identify the minimal locus
encoding the siRNA precursor, we ectopically expressed various combinations of the three Alu-SINE elements in HEK293 cells (E1-3, Fig. 2A). We found that transcription of the E1 segment is sufficient for siRNA production (Fig. 2B) while the segment of E1 and E2 together has a slightly enhanced siRNA production (Fig. 2B). Since E1 does not form an intrinsic hairpin structure, the mechanism of siRNA biogenesis from this locus remains unclear. Alternatively, transcription of E1 and E2 together may produce a long hairpin-structured dsRNA for siRNA production (Fig. S5A).

We also uncovered two piRNA loci for which we observed a bimodal distribution of read sizes at 21-23nt and 27-30nt (Fig. S6A). This distribution would be indicative of siRNA and piRNA expression, respectively. Notably, these loci are also annotated as Alu-SINE elements and, like the Alu-SINE/GON4L endo-siRNA locus, no hairpin structures could be detected by folding the corresponding flanking sequences. Interestingly, the siRNAs from these two loci share sequence similarity with miR-1285 and miR-1303 (Fig. S6B), which also overlap Alu-SINE elements. A search for paralogs of miR-1285 and miR-1303 mature miRNAs revealed more than 80 loci with exact sequence identity in the human genome, indicating miR-1285 and miR-1303 may be more accurately classified as repeat-derived siRNAs.

We also identified putative endo-siRNAs derived from repetitive elements with long hairpin-forming potential. In one case, 12 reads (83.3% were 21-23nt in length) uniquely aligned to inverted repeats of tandem L1 elements on chromosome 16 (Fig. S5B). In another case, a cluster of 72 reads (79.2% were 21-23nt in length) aligned to an LTR element on chromosome 4 (75% were uniquely mapped) (Fig. S5C). Transcripts from these loci were predicted to form long hairpin structures of 98nt and 135nt, respectively, which could serve as Dicer substrates. Despite the low number of reads within these clusters, the reads exhibited clear 5’ homogeneity, suggesting that the small RNAs were authentic endo-siRNAs rather than random degradation products.

A striking ~11,000 reads uniquely aligned to Made1 elements. These are a human-specific subclass of miniature inverted repeat transposable element (MITE), which encodes the large miR-548 family (28). We confirmed that 34 Made1 elements (Table S5), which were not previously associated with miRNA production, were predicted to form hairpin structures (Fig. S7). Alignment of small RNA reads to these hairpins revealed Dicer-cleavage signatures: 21-22nt RNA duplexes along the hairpin stem with 3’ overhangs (Fig. S7). These small RNA reads shared extensive sequence identity with the 58 annotated miR-548 family members in miRBase (v17). Thus, based on their genomic origin within a common transposable element, we preferred to classify miR-548 family members as siRNAs. Detailed read alignments for all putative endo-siRNAs are provided as Supplemental File 3.

**Conservation of noncanonical miRNAs and endo-siRNAs**

We performed a comparative genomic analysis of all 130 noncanonical miRNAs (13 derived from tRNAs or snoRNAs and 117 miRtrons) and 39 endo-siRNAs that were expressed
in human skin across eight vertebrate species (File S4). For 48 (37%) noncanonical miRNAs, the mature miRNA sequence was human specific. For 51 noncanonical miRNAs (39%), the mature sequence was present in M. mulatta (macaque) but not in M. musculus (mouse), suggesting that they might be primate specific. The sequences of the remaining 31 (23%) mature noncanonical miRNAs (21 miRtrons, 8 snoRNA- and 2 tRNA-derived miRNAs) were conserved in one or more species beyond M. mulatta. The homologues of two snoRNAs, ACA25 and ACA44, and the PesudoTTA tRNA in the mouse genome are shown in Fig. S8A. Furthermore, small RNA reads derived from the corresponding murine homologues (Fig. S8B) have been found to be Dicer-dependent but Dgcr8-independent (Fig. S8C) (10). The presence of snoRNA- and tRNA-derived miRNAs in non-primate mammals may suggest potential conserved functions for these sncRNAs. The 39 endo-siRNA loci described here were present only in human and M. mulatta, which was largely due to the absence of hosting repetitive elements in non-primates (29). Overall, the majority of the noncanonical miRNAs and endo-siRNAs we studied were not conserved beyond primates, suggesting that these recently evolved sncRNAs may serve primate-specific functions.

qRT-PCR validation of noncanonical miRNAs and endo-siRNAs

We selected three noncanonical miRNAs and one endo-siRNA for validation with qRT-PCR: the U60 snoRNA-derived miRNA, the pseudoTTA-derived miRNA, miRtron #103 (23), and the Alu-SINE/GON4L-derived siRNA. We first attempted to validate their endogenous expression in skin. All four sncRNAs produced consistent qRT-PCR signals that were both RNA and reverse transcriptase dependent (Fig. 3A), and we were able to obtain visible products for the U60- and PseudoTTA-derived miRNAs and the Alu-SINE/GON4L-derived siRNA (Fig. 3B). We next assessed whether the ectopic expression of ~200nt genomic DNA segments harboring these sncRNAs was sufficient for the accumulation of mature sncRNAs in HEK293 cells. For each of the four loci tested, there was >5 fold enrichment in mature sncRNA expression when the appropriate genomic segment was ectopically expressed, suggesting that these loci are indeed sufficient to generate sncRNAs (data not shown). Finally, we assessed whether the ectopically expressed sncRNAs co-immunoprecipitated with AGO2 or AGO3 by performing RNA binding protein immunoprecipitation followed by qRT-PCR (RIP-PCR; Figure 3C). Indeed, ectopically expressed miRtron #103 was significantly enriched in both AGO2 and AGO3 immunoprecipitates compared to IgG; the pseudoTTA-derived miRNA and the Alu-SINE/GON4L siRNA were significantly enriched in AGO2 immunoprecipitates; and the U60-derived miRNA was modestly enriched in AGO2 and AGO3 immunoprecipitates (not significant). This suggested that these novel sncRNAs may function through association with AGO effectors.

Differential expression of noncanonical miRNAs and endo-siRNAs

In order to investigate the potential dysregulation of noncanonical miRNAs and endo-siRNAs in psoriatic skin, we performed a digital gene expression analysis in PP, PN and NN skin samples. A total of 15 out of 169 (8.9%) noncanonical miRNAs and endo-siRNAs exhibited significant (p<0.001) differential expression with at least a 1.5-fold change in one or more
comparisons (PP vs. NN; PP vs. PN; or PN vs. NN). Ten sncRNAs were upregulated in PP vs. PN and NN skin, whereas five were downregulated in PP vs. PN and NN (Table 2). Four of the 15 differentially expressed sncRNAs were expressed at intermediate levels in PN skin versus PP or NN (Table 2). Notably, the Alu-SINE/GON4L-derived siRNA was ~17-fold increased in PP vs. NN skin, and 2.4-fold increased in PN vs. NN skin. The differential expression of this siRNA and miR-937 (re-classified as a typical miRtron) was confirmed by qRT-PCR (Fig. 4A). The clustering of samples based on the expression of the 15 differentially expressed small RNAs distinguished the majority of PP samples from PN and NN samples (Fig. 4B).

Putative functions of targets of differential expressed miRNAs

Our recent study reported 98 differentially expressed canonical miRNAs in PP versus NN skin (23). To further appreciate the roles of these miRNAs as well as the noncanonical miRNAs and endo-siRNAs identified in the current study in regulating the expression of their targets, we searched for putative mRNA targets of differentially expressed miRNAs with TargetScan (30) and the targets of endo-siRNAs with an in-house siRNA target finding method (see Methods). We then identified predicted targets that were differentially expressed in psoriatic skin, based on previously published microarray data (33,28). From this list of psoriasis-relevant targets, we focused on those that were inversely correlated with their respective targeting sncRNAs. This analysis revealed 59 mRNAs that were inversely correlated with 9 differentially expressed noncanonical miRNAs (7 up-regulated and 2 down-regulated) identified in the current study (Fig. 5 Table S6) and 287 mRNAs that were inversely correlated with 55 differentially expressed canonical miRNAs (43-upregulated and 12 down-regulated; Table S6). As for endo-siRNAs, four targets were inversely correlated with two differentially expressed siRNAs, including the Alu-SINE/GON4L siRNA (Fig 5;Table S7). The low number of predicted siRNA targets may reflect an incomplete understanding of their functions; for example, we only queried known human cDNAs as putative targets, but it is possible that siRNAs target unannotated repetitive transcripts, other ncRNAs or even genomic DNA. Taken together, these data suggest that some differentially expressed miRNAs and endo-siRNAs influence psoriasis pathogenesis by regulating the expression of target genes in the skin.

Finally, we performed a gene ontology analysis on the sets of inversely correlated target mRNAs of both noncanonical and canonical miRNAs. Interestingly, the targets of these two types of miRNAs have different enriched biological processes and molecular functions (Table S8), indicating that they may have distinct roles in gene regulation. For the targets of noncanonical miRNAs, the most significantly enriched biological processes included “growth”, “developmental process”, and “multicellular organismal development” (Benjamini corrected P-values <0.01), In contrast, the most significantly enriched biological processes for the targets of canonical miRNAs included “response to cytokine stimulus”, “organic substance” and “estrogen stimulus” (Benjamini corrected P-value <0.02). These results suggest that noncanonical and canonical miRNAs have distinguishable biological functions, some of which are related to the pathogenesis of psoriasis.
DISCUSSION

We have systematically identified novel sncRNAs from diverse origins in the human genome, characterized their expression in normal and psoriatic skin, and interrogated their potential psoriasis-relevant targets. Newly identified sncRNA species include a snoRNA-derived miRNA that was moderately abundant in human skin, the first example of a tRNA-derived miRNA in an adult human tissue, and an siRNA derived from an intronic Alu-SINE that was highly upregulated in psoriatic skin.

Common features of noncanonical miRNAs and endo-siRNAs from diverse genomic origins

The noncanonical miRNAs and endo-siRNAs that we identified exhibit several distinctive characteristics that set them apart from RNA degradation products and other transcription products. First, these sncRNAs exhibited a size distribution centered around 21-23nt, which is distinct from other classes of small RNAs and random degradation products. Second, noncanonical miRNAs and siRNAs both require Dicer but not Drosha/Dgcr8 for their biogenesis. In agreement with this, the murine homologs of the ACA44 and ACA25 snoRNA-derived miRNAs, the pseudoTTA-derived miRNA, and the re-classified tailed miRtron, miR-3064, were present in deep sequencing datasets from Dgcr8-knockout mice but absent from Dicer-knockout mice (20, 21, 10). Other sncRNAs were absent from these datasets, due to shallow sequencing depth or to their lack of conservation in mouse. Third, the 5’ nucleotide identity of noncanonical miRNAs and endo-siRNAs was non-random, with an overrepresentation of A as the 5’ terminal nucleotide (47.1% and 72.8%, respectively). While canonical miRNAs have a stronger bias towards a 5’ terminal U (78.8%), the presence of either a U or A nucleotide has been shown to facilitate interactions within the middle (MID) domain of the human AGO2 protein (32). Another interesting characteristic that distinguished noncanonical miRNAs from most canonical miRNAs was the prevalence of 5’ heterogeneity, which is consistent with previous studies of noncanonical miRNAs (33–37). In particular, the majority of snoRNA-derived miRNAs, including those that were previously described, exhibited 5’ heterogeneity, suggesting that the biogenesis of snoRNA-derived miRNAs is not fully understood.

Association with protein components of the small RNA silencing machinery

miRNAs and siRNAs exert their functions through association with RISC complexes, which include an AGO protein. In humans, there are four AGO family members, but little direct evidence has been collected for the association of specific sncRNA classes with particular AGO family members (38). Here we show that three noncanonical miRNAs and one endo-siRNA associate with AGO2 and AGO3 proteins, indicating that they may function through AGO effector complexes. Our findings are largely consistent with recently published small RNA deep sequencing of AGO1/2/3/4 immunoprecipitates from HEK293 cells (37; GSE21918;). In that dataset, the pseudoTTA-derived miRNA, 28/56 miRtrons, 6/34 MADE1-derived siRNAs, the Alu-SINE/GON4L siRNA, and the piRNA-locus-derived siRNAs all associated with one or more AGO
proteins. Interestingly, sequences derived from the U60 snoRNA were present in the AGO immunoprecipitates, but there were no exact matches to the abundant 22nt species we detected; instead, reads were consistent with the 5’ terminus of the 22nt species but were typically 28nt in length. Also, the abundant miRtron, #103, was not present in AGO1/2/3/4 immunoprecipitates, but we later found that miRtron #103 is not endogenously expressed in HEK293 cells (data not shown). Thus, it is likely that cell-type specific expression underlies the absence of miRtron #103 and perhaps other sncRNA species from the AGO immunoprecipitation datasets as well.

**Genomic origin of siRNAs and miRNAs**

While miRNAs and siRNAs associate with many of the same cellular proteins for their biogenesis and function, they are thought to differ in their genomic origins and their mechanism of target regulation. miRNAs tend to arise from non-repetitive regions of the genome encoding short, intramolecular hairpin sequences whereas siRNAs tend to originate from longer dsRNAs derived from repetitive or transposable elements. This may be one reason that the number of annotated miRNAs greatly exceeds siRNAs; if, by definition, siRNAs originate from repetitive regions, many legitimate siRNA reads may be discarded during analysis due to promiscuous alignment. Although endo-siRNA loci typically exhibit potential to form long RNA hairpins, we identified three putative endo-siRNA loci from which no RNA hairpin is likely. We speculate that precursor transcripts from these repetitive loci may hybridize *in trans* with highly similar repeat-derived transcripts, forming a long dsRNA that could be recognized by Dicer.

Based on the distinct genomic origins of miRNAs and siRNAs, our findings suggested that a small subset of annotated miRNAs should be annotated as siRNAs. One example is the miR-548 family; many of its members are derived from Made1 transposable elements. The characterization of the miR-548 family members as siRNAs would also be consistent with the finding that Tcl1 elements, of which Made1 elements are a human-specific subclass, are known to give rise to siRNAs in *C. elegans* (11). We also acknowledge that distinction of miRNAs and siRNAs based solely on genomic origin is challenging. In mammals, it has been proposed that the insertion of transposable elements into new genomic sites could be one of the driving forces that create new miRNAs during mammalian evolution (40). Thus, we may have captured evolutionarily transient sncRNA species that blur boundaries between the two sncRNA species. Future studies regarding the mechanism of target regulation by individual sncRNAs will likely improve classification.

**Low expression of miRtrons and tRNA-derived miRNAs in skin**

Given that we analyzed 670 million qualified reads, all of the novel noncanonical miRNAs and endo-siRNAs discovered in this study were expressed at low levels, with the exception of the U60-derived siRNA (which was shown to be over-represented in deep sequencing data compared to the result from qRT-PCR). In the case of miRtrons, their low abundance may result from multiple factors, including the presence of bulges and unpaired bases in their precursors that lower Dicer processing efficiency, susceptibility to endogenous lariat degradation pathways,
and, in many cases, their recent evolution. For snoRNA and tRNA-derived miRNAs, the larger ncRNAs that harbor them are typically abundant, but it appears that only a small fraction of those ncRNAs are processed by Dicer. For example, the cluster of reads associated with the novel hsa-miR-1983 derived from tRNA-IleTAT only encompassed 2% of the total number of reads aligning to the tRNA. In contrast, 85% of small RNA reads aligning to murine tRNA-IleTAT in mESCs clustered within mature mmu-miR-1983 (10). Overall, low expression of hsa-miR-1983 and the pseudoTTA-derived miRNA in human skin might indicate that the alternative hairpin structure is less favorable than the tRNA structure in human tissues, or in non-embryonic tissues in general. Differences in sample preparation between the two studies may also account for some of the relative differences in abundance.

**Implications for psoriasis and other human diseases**

For 11 out of the 15 differentially expressed noncanonical miRNAs and endo-siRNAs, we observed inversely correlated expression of a subset of their predicted psoriasis-relevant targets. In several cases, these predicted targets were also predicted to be regulated by canonical miRNAs that are differentially expressed in psoriasis, suggesting that multiple sncRNAs may contribute to psoriasis pathogenesis through their interactions with a single target.

While all of the predicted targets have been implicated in psoriasis pathogenesis by their differential expression in psoriatic involved skin, many are expressed at low levels in human skin, and their fold changes in psoriatic versus normal skin are relatively low. Moreover, these targets have not yet been strongly implicated in psoriasis pathogenesis. Despite this, a few have particular functional relevance. First, the single predicted target for the highly upregulated Alu-SINE-derived siRNA was MDM4. This gene is an inhibitor of p53 that is frequently upregulated in cancers (41, 42), but modestly downregulated in psoriasis (Table S7) (27). Thus, the down-regulation of MDM4 in psoriasis, which may be mediated by the novel Alu-SINE-derived siRNA we identified, might be one factor that distinguishes the benign hyperplasia in psoriatic involved skin from cancerous tumors. A second target with particular relevance to psoriasis pathogenesis was SDC4, which was predicted to be targeted by the tailed miRtron #72. SDC4 knockout mice display delayed wound healing and impaired angiogenesis (43). Thus, down-regulation of SDC4, which may be mediated by miRtron #72, might contribute to the chronic regenerative phenotype observed in psoriatic involved skin. Importantly, both MDM4 and SDC4 are also predicted targets of canonical miRNAs that are upregulated in psoriasis (Table S6) (23).

We have shown that the sncRNAs described here may regulate genes that are differentially expressed in psoriatic lesions. However, many previously described sncRNAs exhibit developmentally-restricted expression patterns (23). Novel sncRNAs described here may also exhibit this property, which would be consistent with their low expression in adult skin. Thus, it is important to consider the possibility that aberrant expression of certain sncRNAs during development may underlie innate skin barrier or immune defects that contribute to disease susceptibility.
Conclusions

In summary, our discovery and analysis of noncanonical miRNAs and endo-siRNAs from a variety of genomic loci have substantially increased our appreciation for the diversity and prevalence of sncRNA expression in adult human tissues, provided insights into their biogenesis, and identified additional small RNAs that are differentially expressed in psoriatic versus normal skin.

MATERIALS AND METHODS

Skin samples and small RNA library preparation and sequencing

4-6 mm punch skin biopsies were collected from healthy controls and the uninvolved and involved skin of psoriasis (PS) patients. PS patients enrolled in this study received no systemic, photo, or topical therapy in the four weeks prior to sample collection. Biopsies were stored in RNA later (Qiagen) at -80 C prior to RNA extraction. RNA was extracted with the miRNeasy Mini Kit (Qiagen), with on-column DNase I digestion. RNA was prepared for sequencing on the Illumina GAIIx platform with the Small RNA Sample Prep Kit (Illumina) according to manufacturer’s instructions (protocol v1.5). This protocol required the use of a proprietary 3’ adapter that has a high affinity for Dicer cleavage products. Briefly, 3’ and 5’ adapters were ligated to 1ug total RNA. cDNA was synthesized with SuperScript II Reverse Transcriptase (Invitrogen) and subjected to 12 cycles of PCR amplification with high fidelity Phusion Polymerase (Finnzymes Oy). Each library was loaded on a single Illumina lane at 20pM and subjected to 36 cycles of sequencing.

Read processing and mapping

Each deep sequencing library was processed independently. Reads with a 3’ adapter substring shorter than 6nt or trimmed sequence length less than 17nt were removed from the analysis. Trimmed reads were mapped to the reference human genome sequences (UCSC human genome, hg19 build) and sequences in many noncoding and small RNA databases, which include UCSC genome browser tables (tRNA and snoRNA/miRNA), microRNA (miRBase v.17, http://www.mirbase.org/ftp.shtml), piwi-RNA (fRNAdb database, http://www.ncrna.org/frnadb/) and repeats regions (UCSC RepeatMasker Table, hg19 build).

Identification of noncanonical miRNA

Qualified reads that aligned to every locus of human snoRNAs, tRNAs and intro-exon boundaries were subjected to our novel miRNA prediction method, which was tailored for identifying noncanonical miRNAs. Any reads that mapped to previously specified miRNA loci were removed. For each snoRNA and tRNA, 50nt extensions on both sides were added for secondary structure analysis with RNAfold and mFold (44, 45). Since the two programs can predict multiple secondary structures, many of which are different, we merged all the predictions from both programs for downstream analysis. Structures lacking stems of at least 18nt and lacking reads that mapped to any of their stems were excluded. To accommodate atypical stem
structures that often appear in noncanonical miRNAs, more unpaired based, in some cases up to 9bp, were allowed. Candidate miRNAs were then prioritized based on 1) occurrence of sequencing reads on the stem of a predicted hairpin structure with minimum free energy less than -18kcal/mol; 2) possible presence of miRNA* reads on the opposite stem of the hairpin; 3) presence of ~2nt 3’ overhangs on the highest likelihood miRNA/miRNA* duplex; and 4) reads length distribution peaks around ~22nt. To identify miRtrons, 160nt intronic sequence segments adjacent to exon/intron boundaries were extracted for introns analyzed. The same procedure as described above, without a sliding window, was then followed to identify miRtrons. In addition, we examined the presence of sequencing reads ending with AG dinucleotides as a characteristic feature of miRtron. We also applied RNALfold (45) to find local secondary structures or manually examined the structures based on the sequencing reads.

**Identification of endogenous siRNAs from multiple genomic origins**

We applied EINVERTED (46) on each 10kb segment of DNA sequences of whole human genome (hg 19 build) to search for inverted repeats and long hairpin structures they can fold into. We chose inverted repeats by setting a minimum score threshold of 70. Next, we filtered out those predicted inverted repeats that have no more than 10 mappable reads. We also analyzed the genomic annotation of the predicted inverted repeats, i.e. checking them if overlap with SINEs, LINEs or pseudogenes. We next blasted these mappable reads allowing mismatches to the annotated human noncoding RNAs to rule out the possibility that such sequencing reads were byproducts of other types of noncoding RNAs. Finally, we examined genomic regions in which the mappable reads have lengths of dominantly 21~23nt, expected for the length of siRNAs. The amount of the normalized reads having unique genomic loci represented the expression levels of such inverted repeats in each skin category.

**Conservation analysis of sncRNAs**

To analyze conservation of sncRNAs, we retrieved from the UCSC genome browser (47) the multiple alignments of nine species – *Homo sapiens* (human), *Macaca mulatta* (rhesus), *Mus musculus* (mouse), *Canis lupus familiaris* (dog), *Loxodonta Africana* (elephant), *Monodelphis domestica* (opossum), *Gallus gallus* (chicken), *Xenopus tropicalis* (frog) and *Danio rerio* (zebrafish) – for each novel miRNA and siRNA. We calculated the number of insertions, deletions and mismatches when each non-human sequence was compared with a human sncRNA (File S4) based on the multiple alignments of the nine species. An sncRNA was considered conserved in one species if the sequence in that species had no variation (insertions, deletions and mismatches) in the seed region for a miRNA or less than 2 variations in the whole region for a siRNA.

**qRT-PCR**

qRT-PCR of novel sncRNAs was performed with custom TaqMan small RNA assays (Life Technologies), as previously described (23). qRT-PCR of the known miRNA miR-937 and the endogenous control Z30 were performed with TaqMan miRNA assays (Life Technologies). Exact primer and probe sequences are proprietary. Relative expression levels were calculated
according to the $2^{-\Delta\Delta Ct}$ method as follows: $100\times2^{(Ct_{Z30} - Ct_{sncRNA})}$. Significance was determined with one way ANOVA and post-hoc two tailed t-tests.

**Cell culture**

The cloning of sncRNAs into pEP-miR cloning and expression vectors (Cell Bio Labs) and transfections of HEK293 cells were performed as previously described (23). Primer sequences were as follows: pEP-miR-E1-F/R: 5'-GAGAAGGAGTTGTGGGAATCCAAT / 5'-CTGTGCCCCAGCTATTCTC; pEP-miR-E2-F/R: 5'-GAGGAATAGCTGGGGCAACAG / 5'-AGACAGAGACAGGGTCTCAGATG; pEP-miR-E3-F/R: 5'-CTGTGTGATCTCTGAGACCCTGTC / 5'-TGCATGCTCTGCTCCAACAAT; pEP-miR-E1-2-F/R: 5'-CTGTGAAAAATAGTCTCTACAAAGTATCC / 5'-GAGCCTAAGGGTGTCAAGATCA; pEP-miR-E2-3-F/R: 5'-CAAATCTTGGAGAGTTGTGG / 5'-TGCATGCTCTGCTCCAACAA; pEP-miR-E1-3-F/R: 5'-GCACACTTGGCTCCTTAGC / 5'-GCCAGGAACCTGAAAGTC; pEP-miR-pseudoloTTA-F/R: 5'-CTGCCTCACCACCAGCT/ 5'-AGGCCACACCACCTCCTTAC; pEP-miRtron#103-F/R: 5'-TTTTCTGTCACTGCTTTTGGGA / 5'-GATCTCCAGTGGGAGAGCAG.

**RIP-PCR**

RIP-PCR was performed as previously described for AGO2 (23). The same protocol was applied with a rabbit polyclonal antibody for AGO3 (Sigma-Aldrich). Significance was determined by one-way ANOVA and post-hoc two tailed t-tests.

**Differential expression of miRNAs and siRNAs**

Reads that aligned perfectly to the miRNAs and siRNAs of interest were subjected to gene expression analysis. Reads that mapped to multiple genomic loci were attributed to all potential derivative small RNAs. Read counts in each skin category (NN, PN, and NN) were normalized to adjust for slight variation in total read count between categories. Let $N_{category}$ be the number of qualified reads that aligned to the human genome (hg19 build) in each category, $C$ the average of $N_{category}$, and $M_{category}$ the number of qualified reads aligned to each miRNA and siRNA in each category. Thus, the normalized number of reads for each miRNA and siRNA in a given category is $(M_{category} \times C)/(N_{category})$. Fold changes were calculated from normalized read counts, and chi-squared test was applied to determine significance.

**Putative targets of miRNAs and siRNAs**

The target mRNA genes of miRNAs of interest were predicted by TargetScan (30). For predicting targets of siRNAs, human cDNA sequences were downloaded from UCSC genome browser (hg19 build) and the reverse complementary sequences of siRNAs were mapped to the cDNA sequences by BLAST (48). Those mapping sites with no more than 3 mismatches were considered as putative targets of siRNAs.
mRNA gene expression analysis

Microarray data of mRNA genes of psoriatic and normal skin of human were downloaded from Gene Expression Omnibus (GEO) accession number GSE13355, which has a total of 122 samples containing 58 psoriatic patients and 64 normal healthy controls. Differentially expressed genes were identified using Rank Product (49). The rationale behind Rank Product is that it is unlikely for a gene to be ranked in the top position in all replicates if the gene was not differentially expressed. It has been shown that RP is less sensitive to noise and has a better performance than other methods when sample size is small (RP_comp). We selected differentially expressed genes with the percentage of false positive predictions (pfp), equivalent to False Discovery Rate, no greater than 0.05 for 1,000 permutations. Differentially expressed probe sets were then mapped to corresponding genes according to the annotation of Affymetrix Human Genome U133 Plus 2.0 Array. We also used the set of differentially expressed genes in PP vs. NN skin previously reported in (31).

GO term and Functional enrichment of target genes

The GO term analysis was performed using the online tool DAVID (50). DAVID provides p-values, before and after multiple-test corrections, based on a modified Fisher’s exact test.

Data deposition

All 76 small RNA sequencing libraries from psoriatic lesion and normal human skin biopsy samples have been deposited into NCBI/GEO databases under GEO31037.

Acknowledgement

We thank the Genome Technology Access Center at Washington University for GAIIx data generation, and Xiang Zhou for preprocessing of sequencing data. This work was supported by the National Institutes of Health (5RC1AR058681 to A.M.B. and W.Z. and R01GM100364 to W.Z.); the National Science Foundation (DBI-0743797 to W.Z.); and the National Human Genome Research Institute (T32HG000045 to C.E.J.).
REFERENCE


Figure 1. Example of snoRNA-derived miRNA. (A) The hairpin structures of ACA25, with the miRNA and miRNA* sequences shown in red. (B) Alignment of reads mapped to ACA25 with the number of sequencing reads, length of read, the number of genomic loci which the read can be mapped to with no mismatches and the percentage of the read counts.

Figure 2. (A) Endo-siRNA derived from tandem Alu SINE transposable elements in an intron of the gon-4-like (GON4L) gene on chromosome 1. Alignment of reads to the Alu SINEs (hg19) with the red box indicating the reads uniquely mapped to this locus. (B) Abundance of siRNAs in six constructs that include six combinations of the three Alu SINEs. The y-axis indicates the normalized expression relative to the wide-type.

Figure 3. Endogenous expression of mature novel miRNAs in skin. (A) qRT-PCR amplification traces of sncRNAs from human skin biopsy RNA (black lines) and negative controls (grey lines). An arbitrary threshold used for relative quantitation of expression levels (not shown) is indicated in green. (B) qRT-PCR product bands from human skin biopsy RNA and negative controls. A band size of ~50bp corresponds to an ~21-nt RNA species due to the addition of arbitrary sequence during the reverse transcription step (see Methods). (C) Abundance of sncRNAs in AGO2 and AGO3 immunoprecipitates relative to IgG. (* p<0.05, **p<0.01, ***p<0.001). RT = reverse transcriptase.

Figure 4. Differential expression of sncRNAs in involved psoriatic skin. (A) qRT-PCR levels of differentially expressed sncRNAs in ten normal (NN), ten uninvolved psoriatic (PN), and ten involved psoriatic (PP) skin samples. Lines indicate matched uninvolved and involved samples from the same patient. Relative expression was calculated with respect to the endogenous snoRNA Z30. (* p<0.05, **p<0.01, ***p<0.001). (B) Heat map of 15 differentially expressed sncRNAs across 67 skin samples.

Figure 5. Anti-correlation of differentially expressed target mRNAs with miRNAs and siRNAs. The color bar indicates the range of fold changes.
Table 1. Summary of noncanonical miRNAs and endo-siRNAs in the current study.

<table>
<thead>
<tr>
<th>Noncanonical miRNA</th>
<th>total</th>
<th>Conserved</th>
<th>Nonconserved</th>
</tr>
</thead>
<tbody>
<tr>
<td>snoRNA-derived</td>
<td>11</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>tRNA-derived</td>
<td>2</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>miRtron typical</td>
<td>19</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>miRtron tailed</td>
<td>98</td>
<td>18</td>
<td>80</td>
</tr>
<tr>
<td>Endo-siRNA</td>
<td>39</td>
<td>-</td>
<td>39</td>
</tr>
</tbody>
</table>
Table 2. Top 15 differentially expressed sncRNAs between psoriatic involved skin (PP), psoriatic uninvolved skin (PN) and normal skin (NN).

<table>
<thead>
<tr>
<th>ID</th>
<th>PP</th>
<th>PN</th>
<th>NN</th>
<th>PP/NN</th>
<th>PP/PN</th>
<th>PN/NN</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alu-SINE-siRNA</td>
<td>1614.78</td>
<td>220.54</td>
<td>91.81</td>
<td>17.59</td>
<td>7.32</td>
<td>2.40</td>
<td>0.00E+00</td>
</tr>
<tr>
<td>miRtron #14</td>
<td>11.58</td>
<td>2.05</td>
<td>1.00</td>
<td>11.58</td>
<td>5.65</td>
<td>2.05</td>
<td>9.41E-04</td>
</tr>
<tr>
<td>miR-937</td>
<td>20.40</td>
<td>2.05</td>
<td>3.19</td>
<td>6.40</td>
<td>9.95</td>
<td>-1.56</td>
<td>4.25E-06</td>
</tr>
<tr>
<td>tRNAile/miR-1983</td>
<td>117.40</td>
<td>25.16</td>
<td>23.98</td>
<td>4.90</td>
<td>4.67</td>
<td>1.05</td>
<td>3.33E-23</td>
</tr>
<tr>
<td>Made1_1-siRNA</td>
<td>1182.67</td>
<td>354.99</td>
<td>293.12</td>
<td>4.03</td>
<td>3.33</td>
<td>1.21</td>
<td>2.72E-176</td>
</tr>
<tr>
<td>miRtron #144</td>
<td>79.48</td>
<td>25.16</td>
<td>22.88</td>
<td>3.47</td>
<td>3.16</td>
<td>1.10</td>
<td>3.24E-11</td>
</tr>
<tr>
<td>miRtron #323</td>
<td>32.75</td>
<td>11.50</td>
<td>17.41</td>
<td>1.88</td>
<td>2.85</td>
<td>-1.67</td>
<td>2.88E-03</td>
</tr>
<tr>
<td>miRtron #375</td>
<td>87.42</td>
<td>34.61</td>
<td>31.63</td>
<td>2.76</td>
<td>2.53</td>
<td>1.09</td>
<td>4.46E-09</td>
</tr>
<tr>
<td>U60 miRNA-5p</td>
<td>1.08E+7</td>
<td>7.64E+6</td>
<td>4.67E+6</td>
<td>2.30</td>
<td>1.41</td>
<td>1.64</td>
<td>0.00E+00</td>
</tr>
<tr>
<td>miRtron #72</td>
<td>143.86</td>
<td>82.93</td>
<td>74.30</td>
<td>1.94</td>
<td>1.73</td>
<td>1.12</td>
<td>6.03E-07</td>
</tr>
<tr>
<td>miRtron #145-5p</td>
<td>10.70</td>
<td>64.03</td>
<td>42.57</td>
<td>-4.00</td>
<td>-5.88</td>
<td>1.50</td>
<td>1.01E-08</td>
</tr>
<tr>
<td>miRtron #252</td>
<td>10.70</td>
<td>20.96</td>
<td>34.92</td>
<td>-3.23</td>
<td>-1.96</td>
<td>-1.67</td>
<td>1.28E-03</td>
</tr>
<tr>
<td>miRtron #80</td>
<td>739.11</td>
<td>1483.14</td>
<td>2179.30</td>
<td>-2.94</td>
<td>-2.0</td>
<td>-1.47</td>
<td>2.84E-154</td>
</tr>
<tr>
<td>miRtron #145-3p</td>
<td>11.58</td>
<td>33.56</td>
<td>28.35</td>
<td>-2.43</td>
<td>-2.86</td>
<td>1.18</td>
<td>4.59E-03</td>
</tr>
<tr>
<td>Made1_47-siRNA</td>
<td>225.87</td>
<td>435.87</td>
<td>340.16</td>
<td>-1.52</td>
<td>-1.92</td>
<td>1.28</td>
<td>4.22E-15</td>
</tr>
</tbody>
</table>
## ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS</td>
<td>Psoriasis</td>
</tr>
<tr>
<td>sncRNA</td>
<td>small non-coding RNA</td>
</tr>
<tr>
<td>dsRNA</td>
<td>double-stranded RNA</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA</td>
</tr>
<tr>
<td>endo-siRNA</td>
<td>endogenous small interfering RNA</td>
</tr>
<tr>
<td>piRNA</td>
<td>piwi-interacting RNA</td>
</tr>
<tr>
<td>snoRNA</td>
<td>small nucleolar RNA</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>RISC</td>
<td>RNA induced silencing complex</td>
</tr>
<tr>
<td>AGO</td>
<td>Argonaute protein</td>
</tr>
<tr>
<td>Dicer</td>
<td>double-stranded RNA-specific endoribonuclease</td>
</tr>
<tr>
<td>Dgcr8</td>
<td>DiGeorge syndrome critical region gene 8</td>
</tr>
<tr>
<td>NGS</td>
<td>Next-generation sequencing</td>
</tr>
<tr>
<td>PP</td>
<td>Involved/lesional psoriatic</td>
</tr>
<tr>
<td>PN</td>
<td>Uninvolved/non-lesional psoriatic</td>
</tr>
<tr>
<td>NN</td>
<td>Normal</td>
</tr>
<tr>
<td>SINE</td>
<td>Short Interspersed Elements</td>
</tr>
<tr>
<td>LINE</td>
<td>Long Interspersed Elements</td>
</tr>
<tr>
<td>LTR</td>
<td>Long terminal repeat</td>
</tr>
<tr>
<td>DNA</td>
<td>DNA transposons</td>
</tr>
<tr>
<td>GON4L</td>
<td>GON-4-like protein</td>
</tr>
<tr>
<td>MDM4</td>
<td>mouse double minute 4 homolog</td>
</tr>
<tr>
<td>SDC4</td>
<td>syndecan 4</td>
</tr>
</tbody>
</table>