

## Supplementary Methods and Results

### Comparing miRNA expression levels using five different metrics

We assessed the use of five different metrics to summarize miRNA expression levels based on the two sequence libraries. The first three metrics are calculated from sequences with either a perfect match, or sequences with perfect 3' or 5' matches to the miRBase reference sequence. The other two were calculated using either the most common or the sum of all sequences representing a single arm of a particular miRNA.

The pairwise correlation between the hESC and EB libraries using each of the five metrics is included in the first table (below). The most robust discrimination between these libraries, manifested as a low  $R^2$  value, resulted from comparison using the most frequently observed sequence from each arm of every miRNA ( **$R^2=0.87$ , Spearman Correlation**). Using only the sequences matching the miRBase reference sequence gave reasonable, but not as large, a discrimination between libraries ( **$R^2=0.92$** ). The 'sum of all tags' metric showed an even lower correlation, suggesting it may provide better discrimination between libraries. However, this metric did not perform well when used for clustering libraries of different origin in both human and mouse (not shown).

Spearman correlation between hESC and EB libraries using each of the five metrics for miRNA expression:

|                     | hESC miRBase | hESC most common tag | hESC sum of tags | hESC 5prime miRBase match | hESC 3prime miRBase match |
|---------------------|--------------|----------------------|------------------|---------------------------|---------------------------|
| EB miRBase          | <b>0.92</b>  | 0.68                 | 0.57             | 0.85                      | 0.93                      |
| EB most common tag  | 0.76         | <b>0.87</b>          | 0.64             | 0.89                      | 0.85                      |
| EB sum of tags      | -0.16        | -0.09                | <b>0.66</b>      | -0.16                     | -0.12                     |
| EB 5' miRBase match | 0.77         | 0.70                 | 0.29             | <b>0.92</b>               | 0.84                      |
| EB 3' miRBase match | 0.67         | 0.68                 | 0.66             | 0.85                      | <b>0.95</b>               |

Comparison of counts for miRBase sequences with most abundant isomiRs with a different sequence:

| miRNA-arm        | most abundant isomiR       | miRBase count | isomiR count |
|------------------|----------------------------|---------------|--------------|
| hsa-let-7b-5-p   | TGAGGTAGTAGGTTGTGT         | 136           | 284          |
| hsa-mir-101-3-p  | GTACAGTACTGTGATAACTGAA     | 1674          | 3361         |
| hsa-mir-105-5-p  | AAATGCTCAGACTCCTGTGGTG     | 2             | 23           |
| hsa-mir-106b-5-p | TAAAGTGCTGACAGTGCAGATA     | 459           | 1771         |
| hsa-mir-10a-5-p  | TACCCTGTAGATCCGAATTTGT     | 984           | 3887         |
| hsa-mir-10b-5-p  | TACCCTGTAGAACCGAATTTGT     |               | 23           |
| hsa-mir-124-3-p  | TAAGGCACGCGGTGAATGCCAA     | 148           | 607          |
| hsa-mir-125a-5-p | TCCCTGAGACCCTTTAACCTGTG    | 148           | 311          |
| hsa-mir-138-5-p  | AGCTGGTGTGTGAATCAGGCCGT    | 31            | 32           |
| hsa-mir-139-5-p  | TCTACAGTGCACGTGTCTCCAGT    | 5             | 15           |
| hsa-mir-140-3-p  | TACCACAGGGTAGAACCACGGAC    | 1676          | 6218         |
| hsa-mir-141-3-p  | TAACACTGTCTGGTAAGATG       | 3             | 4            |
| hsa-mir-146a-5-p | TGAGAACTGAATTCCAT          | 174           | 556          |
| hsa-mir-146b-5-p | TGAGAACTGAATTCCATAGGCTGT   | 214           | 2075         |
| hsa-mir-151-5-p  | TCGAGGAGCTCACAGTCTAGTA     | 167           | 272          |
| hsa-mir-181a-5-p | AACATTCAACGCTGTCCGGTGAGTTT | 101           | 1220         |
| hsa-mir-181b-5-p | AACATTCAATGCTGTCCGGTGGGTTT | 79            | 588          |
| hsa-mir-181c-5-p | AACATTCAACCTGTCCGGTGAGTTT  |               | 24           |
| hsa-mir-181d-5-p | AACATTCAATGTTGTCCGGTGGGTT  | 37            | 281          |
| hsa-mir-186-5-p  | CAAAGAATTCTCCTTTTGGGCTT    | 216           | 1095         |
| hsa-mir-188-5-p  | CATCCCTTGCATGGTGGAGGGT     | 5             | 6            |
| hsa-mir-18b-5-p  | TAAGGTGCATCTAGTGCAGTT      | 5             | 22           |
| hsa-mir-194-5-p  | TGTAACAGCAACTCCATGTGGAA    | 4             | 31           |
| hsa-mir-195-5-p  | TAGCAGCACAGAAATATTGGCA     | 3             | 25           |
| hsa-mir-200a-3-p | TAACACTGTCTGGTAACGATGTT    | 7             | 136          |
| hsa-mir-200b-3-p | TAATACTGCCTGGTAATGATGAC    | 20            | 221          |
| hsa-mir-205-5-p  | TCCTTCATTCCACCGGAGTCTGT    | 459           | 502          |
| hsa-mir-210-3-p  | CTGTGCGTGTGACAGCGGCTG      | 247           | 311          |
| hsa-mir-214-3-p  | ACAGCAGGCACAGACAGGCAG      | 4             | 10           |
| hsa-mir-215-5-p  | ATGACCTATGAATTGACAGACA     | 8             | 38           |
| hsa-mir-21-5-p   | TAGCTTATCAGACTGATGTTGAC    | 7564          | 39818        |
| hsa-mir-216b-5-p | AAATCTCTGCAGGCCAAATGTG     | 2             | 4            |
| hsa-mir-217-5-p  | ATACTGCATCAGGAAGTATTG      | 16            | 32           |
| hsa-mir-219-5-p  | TGATTGTCCAAACGCAATTCTCG    |               | 6            |
| hsa-mir-222-3-p  | AGCTACATCTGGCTACTGGGTCTC   | 558           | 4719         |
| hsa-mir-224-5-p  | CAAGTCACTAGTGGTTCCGTTTA    | 2             | 109          |
| hsa-mir-23a-3-p  | ATCACATTGCCAGGGATTCCA      | 109           | 335          |
| hsa-mir-23b-3-p  | ATCACATTGCCAGGGATTACCAC    | 55            | 286          |
| hsa-mir-26b-5-p  | TTCAAGTAATTCAGGATAGGTT     | 116           | 2957         |
| hsa-mir-27a-3-p  | TTCACAGTGGCTAAGTT          | 66            | 163          |
| hsa-mir-296-3-p  | AGGGTTGGGTGGAGGCT          | 15            | 97           |

|                  |                           |      |       |
|------------------|---------------------------|------|-------|
| hsa-mir-301b-3-p | CAGTGCAATGATATTGTCAAAGCAT | 6    | 30    |
| hsa-mir-302c-3-p | TAAGTGCTTCCATGTTTCAGT     | 1068 | 3929  |
| hsa-mir-30a-5-p  | TGTAAACATCCTCGACTGGAAGCT  | 419  | 2822  |
| hsa-mir-30c-5-p  | TGTAAACATCCTACACTCTCAGCT  | 167  | 828   |
| hsa-mir-30d-5-p  | TGTAAACATCCCCGACTGGAAGCT  | 548  | 4988  |
| hsa-mir-30e-5-p  | TGTAAACATCCTTGACTGGAAGCT  | 61   | 738   |
| hsa-mir-31-5-p   | AGGCAAGATGCTGGCATAGCTG    | 76   | 235   |
| hsa-mir-323-3-p  | GCACATTACACGGTCGACCTCT    | 16   | 28    |
| hsa-mir-32-5-p   | TATTGCACATTACTAAGTTGCAT   | 4    | 10    |
| hsa-mir-331-3-p  | GCCCCTGGGCCTATCCTAGAA     | 3    | 1129  |
| hsa-mir-335-5-p  | TCAAGAGCAATAACGAAAAATG    | 198  | 331   |
| hsa-mir-338-3-p  | TCCAGCATCAGTGATTTTGTGA    |      | 7     |
| hsa-mir-342-5-p  | AGGGGTGCTATCTGTGATTGAGG   | 2    | 14    |
| hsa-mir-34a-5-p  | TGGCAGTGTCTTAGCTGGTTGTT   | 15   | 44    |
| hsa-mir-34b-3-p  | AATCACTAACTCCACTGCCATC    |      | 9     |
| hsa-mir-367-3-p  | ATTGCACTTTAGCAATGGTGA     | 113  | 125   |
| hsa-mir-378-3-p  | ACTGGACTTGGAGTCAGAAGGC    | 6149 | 16194 |
| hsa-mir-411-5-p  | ATAGTAGACCGTATAGCGTACG    | 3    | 5     |
| hsa-mir-432-5-p  | TCTTGGAGTAGGTCATTGGGTG    | 22   | 70    |
| hsa-mir-448-3-p  | TTGCATATGTAGGATGTCCC      | 60   | 501   |
| hsa-mir-450-3-p  | ATTGGGGACATTTTGCATTCAT    |      | 12    |
| hsa-mir-451-5-p  | AAACCGTTACCATTACTGAGTTT   |      | 6     |
| hsa-mir-452-5-p  | AAGTGTTCGAGAGGAACTGAGA    | 42   | 122   |
| hsa-mir-454-3-p  | TAGTGCAATATTGCTTATAGGGTT  | 85   | 129   |
| hsa-mir-485-5-p  | AGAGGCTGGCCGTGATGAATTCG   | 2    | 6     |
| hsa-mir-491-5-p  | AGTGGGGAACCCCTTCCATGAGGA  |      | 5     |
| hsa-mir-494-3-p  | TGAAACATACACGGGAAACCTCT   |      | 32    |
| hsa-mir-497-5-p  | CAGCAGCACACTGTGGTTTGTA    | 5    | 28    |
| hsa-mir-499-5-p  | TTAAGACTTGCAAGTATGTTTA    | 29   | 59    |
| hsa-mir-500-5-p  | TAATCCTTGCTACCTGGG        |      | 4     |
| hsa-mir-502-3-p  | AATGCACCTGGGCAAGGATTCAG   | 18   | 30    |
| hsa-mir-503-5-p  | TAGCAGCGGGAACAGTTCTG      | 91   | 235   |
| hsa-mir-504-5-p  | AGACCCTGGTCTGCACTCTATCT   | 24   | 49    |
| hsa-mir-505-3-p  | TCAACACTTGCTGGTTTCCTCT    | 4    | 19    |
| hsa-mir-512-5-p  | ACTCAGCCTTGAGGGCACTTTC    |      | 50    |
| hsa-mir-518c-3-p | CAAAGCGCTTCTCTTTAGAGTG    | 8    | 29    |
| hsa-mir-518f-3-p | GAAAGCGCTTCTCTTTAGAGGA    | 63   | 205   |
| hsa-mir-519d-3-p | CAAAGTGCCTCCCTTTAGAGTGT   | 3    | 51    |
| hsa-mir-520b-3-p | AAAGTGCTTCCTTTTAGAGGGT    | 10   | 42    |
| hsa-mir-520e-3-p | AAAGTGCTTCCTTTTAGAGGGT    | 6    | 7     |
| hsa-mir-520f-3-p | CAAGTGCTTCCTTTTAGAGGGT    | 4    | 34    |
| hsa-mir-523-3-p  | AACGCGCTTCCCTATAGAGGGT    | 16   | 38    |
| hsa-mir-556-3-p  | ATATTACCATTAGCTCATCTT     | 2    | 6     |
| hsa-mir-565-5-p  | TGGATAACGCGTCTGACT        |      | 11    |
| hsa-mir-576-3-p  | AAGATGTGGAAAAATTGGAATCC   | 8    | 19    |
| hsa-mir-577-5-p  | GTAGATAAAATATTGGTACCTG    | 2    | 96    |
| hsa-mir-584-5-p  | TTATGGTTTGCCTGGGACTGA     |      | 11    |

|                 |                           |     |     |
|-----------------|---------------------------|-----|-----|
| hsa-mir-589-3-p | TCAGAACAAATGCCGGTTCCTCAGA | 11  | 26  |
| hsa-mir-589-5-p | TGAGAACCACGTCTGCTCTGA     | 11  | 78  |
| hsa-mir-625-5-p | AGGGGGAAAGTTCTATAGTCCT    | 3   | 64  |
| hsa-mir-629-5-p | TGGGTTTACGTTGGGAG         | 21  | 453 |
| hsa-mir-643-3-p | ACACTTGTATGCTAGCTCAGGT    |     | 11  |
| hsa-mir-653-5-p | TTGAAACAATCTCTACTGAACC    |     | 5   |
| hsa-mir-654-3-p | TATGTCTGCTGACCATCACC      | 3   | 4   |
| hsa-mir-7-5-p   | TGGAAGACTAGTGATTTTGTGTT   | 451 | 638 |
| hsa-mir-760-3-p | CGGCTCTGGGTCTGTGGGG       | 2   | 25  |
| hsa-mir-767-5-p | TGCACCATGGTTGTCTGAGCAT    | 7   | 29  |
| hsa-mir-768-5-p | GTTGGAGGATGAAAGTACGGA     | 10  | 151 |
| hsa-mir-801-5-p | TGGCACACGTAGGGCAAC        |     | 185 |
| hsa-mir-874-3-p | CTGCCCTGGCCCCGAGGGACCGAC  | 9   | 19  |
| hsa-mir-877-5-p | GTAGAGGAGATGGCGCAGGGG     | 39  | 101 |
| hsa-mir-887-3-p | GTGAACGGGCGCCATCCCGAGGCT  | 3   | 19  |
| hsa-mir-923-5-p | AGCGGAGGAAAAGAACT         | 4   | 52  |
| hsa-mir-933-3-p | TGTGCGCAGGGAGACCTCTCCCA   | 4   | 9   |
| hsa-mir-935-3-p | CAGTTACCGCTTCCGCTACCGC    | 3   | 52  |
| hsa-mir-98-5-p  | TGAGGTAGTAAGTTGTAT        | 4   | 12  |

### Detecting RNA editing events

The two types of RNA editing that have been characterized in mammals include Adenine and Cytosine deamination reactions. The former is observed in sequence reads as an A to G transition while the latter is observed as a C to T transition. To search for evidence of RNA editing, we aligned all unaligned reads to the human genome using a local alignment tool that allows up to two mismatches (ELAND, unpublished). Taking all the reads that aligned to miRNA loci using this tool, we produced multiple sequence alignments representing ‘modified’ versions of each miRNA. These multiple sequence alignments were converted to sequence logos to facilitate visualization of recurrent editing events. The following figure shows the identification of a known A->I editing site in has-miR-151 (Blow et al, 2006) as well as a potential C->U editing site. It is also possible that the latter results from minor contamination of the mouse fibroblast feeder

cells on which the hESCs were grown, considering the mismatching nucleotide corresponds to the mouse genome sequence.

Sequence logo representing has-miR-151 reads with imperfect genome alignments:

