Supplementary Figures and Tables

Continental_difference.tif

Figure S1: Pairwise FST for populations grouped by continent (A) between European and African groups (B) between African and Asian groups (C) between European and Asian groups. The 95th percentile and the 99th percentile are represented by blue and dotted red lines, respectively, in each group.

FST_pvalues.tif

Figure S2: Distribution of p-values for pairwise FST’s measured between African and non-African populations. The 95th percentile and the 99th percentile are represented by blue and dotted red lines, respectively. The black dotted line represents p-values below (p < 0.05). PD-miRNA variants are highlighted in orange and HPD-miRNA variants in green. (Insert) Blow-up of all significant FST’s (p <0.05)

S1_seed_sequence.tif

Figure S3: Allele frequencies for the 4 PD-variants in miRNA seed sequences, African populations (green) and non-African populations (blue).

Sig_Targets2.tif

Figure S4: Biological functions of the 72 significant gene targets of the 8 HPD miRNA.

F5_Clustering.tif

Figure S5: Clustering analysis of enriched biological processes. HPD-miRNA significantly enriched for regulation of 55 biological processes (Supplemental Table S8). Using the REVIGO clustering-like algorithm [[113](#_ENREF_113)], biological processes were clustered based on previously defined measures of semantic similarity in gene ontology [[113](#_ENREF_113)]. Clustering identified 26 semantically unique biological processes that represented the minimum set necessary to describe the entire list. Connected biological processes have similar GO classifications with line thickness representing the strength of the semantic connection. Significance levels of biological process clusters are indicated with bubble color (clusters shown in red have p-values that remain significant after Bonferroni correction).

LD_plot_TOTAL_BIS_AR.MC_Final.tiff

Figure S5. LD plots for HPD-SNPs in African Population from West Central, Central and East Africa. Pairwise plots of LD among genotyped loci from the Illumina 1M Duo SNP array that flank HPD-miRNAs on chromosomes 4,5,10, 11,13 and 14 were constructed using the Haploview program (Barrett et al. 2005). The color of each square signifies the strength of the relationship between SNP alleles. For example, bright red squares indicate statistically significant LD between a given pair of SNP loci (D’=1; LOD > 2), while shades of pink/white squares indicate little evidence for LD. Purple squares indicate high LD but with little statistical support (low LOD). The rs numbers for the 1M Illumina SNPs are given across the top of each plot. The position of the 8 HPD-miRNA SNPs are indicated with a black arrows.

mir202_bodymap.tiff

Figure S6: Tissue specific expression of hsa-mir-202. Tissue specific expression was obtained through miRNAbodymap.