TRENDS IN SCIENCES 2026; 23(8): 12823
https://doi.org/10.48048/tis.2026.12823 RESEARCH ARTICLE

Analytical Validation of A 101 Germline miR-SNP Ion AmpliSeq Panel for Breast Cancer-Related Genetic Studies

Thuy Thi Chung Duong[1,3], Hue Thi Nguyen[1,2,3], Nga Thi Nguyen[1,3], Luan Huu Huynh[1,2,3], Thinh Hung Nguyen[4], Tuan Huu Ngoc Nguyen[4] and Thanh Thi Ngoc Nguyen[1,2,3,*]

[1]*Human Genetics Laboratory, Faculty of Biology and Biotechnology, University of Science, Ho Chi Minh, Vietnam*
[2]*Department of Physiology and Animal Biotechnology, Faculty of Biology and Biotechnology, University of Science, Ho Chi Minh, Vietnam*
[3]*Vietnam National University, Ho Chi Minh, Vietnam*
[4]*Medical Biochemistry & Molecular Biology Department, Fundamental Sciences and Basic Medical Sciences, Pham Ngoc Thach University of Medicine, Ho Chi Minh, Vietnam*

([*]Corresponding author's e-mail: ngtnthanh@hcmus.edu.vn)

Received: 27 November 2025, Revised: 26 December 2025, Accepted: 2 January 2026, Published: 10 March 2026

Abstract

Massively parallel sequencing (MPS) technology enables simultaneous genotyping of many SNPs across multiple samples with high depth and accuracy. Such technology has become increasingly valuable for studies investigating genetic variants, including miRNA-related SNPs (miR-SNPs), that may contribute to cancer susceptibility. In this study, we designed a customized Ion AmpliSeq panel targeting 101 candidate breast cancer-associated miR-SNPs and evaluated its sequencing performance and analytical reliability on the Ion S5 XL system. Sequencing metrics, including depth of coverage (DoC), locus coverage balance (LCB), frequency of major allele reads (F_{MAR}), and locus strand balance (LSB), were assessed across all loci. Across 50 DNA samples, the sequencing achieved a mean of 85,194 (95% CI: 79,904 - 101,503), with 91.44% (95% CI: 90.68 - 94.45) on-target reads, an average depth of 515.1× (95% CI: 488.9 - 592.9), and uniformity of 88.77% (95% CI: 88.76 - 89.64), indicating high-quality and consistent performance across libraries. Among 101 miR-SNPs, 97 loci yielded valid genotypes, while four loci resulted in no-calls due to low coverage. Genotypes obtained from the customized miR-SNP panel were fully concordant with those verified by Sanger sequencing. The panel demonstrated stable performance and high reproducibility, with consistent variant calling across replicate runs. Furthermore, reliable and complete genotyping profiles were obtained from as little as 1 ng of input DNA. Collectively, these results indicate that the customized miR-SNP MPS panel provides robust analytical performance and high accuracy, supporting its applicability for large-scale genetic and breast cancer association studies.

Keywords: miRNA-SNP, Breast cancer, Ion AmpliSeq, MPS, Genotyping, Validation

Introduction

MicroRNAs (miRNAs) are small non-coding RNAs that play crucial roles in post-transcriptional regulation of gene expression by binding to target messenger RNAs and influencing their stability or translational efficiency [1]. Increasing evidence suggests that dysregulation of miRNAs is implicated in the initiation and progression of various human diseases, including breast cancer [2,3]. Consequently, miRNAs have attracted growing attention as potential biomarkers for diagnosis, prognosis, and therapeutic response monitoring [4].

Genetic variations within miRNA genes, particularly single nucleotide polymorphisms (miR-SNPs), may influence the maturation, expression, or

binding capacity of miRNAs [5]. Such variants can alter downstream regulatory networks, thereby contributing to interindividual differences in cancer susceptibility [6,7]. Numerous association studies have reported that specific miR-SNPs are linked to breast cancer risk, prognosis, and treatment response [8-10]. However, findings across populations remain inconsistent, and many SNPs in miRNA genes have not been systematically investigated.

Massively parallel sequencing (MPS) technology enables comprehensive profiling of genetic abnormalities, including SNP [11,12]. In breast cancer, MPS has been widely applied to characterize driver mutations, identify susceptibility-associated SNPs, and discover novel biomarkers with prognostic or therapeutic value [13,14]. Compared with traditional molecular methods, MPS offers unparalleled sensitivity and resolution, allowing simultaneous interrogation of hundreds of genes across multiple samples [15]. However, conventional MPS platforms often require large batch sizes, extended turnaround times, and dedicated bioinformatics support, which limit their accessibility in many hospital laboratories [16].

Among available platforms, the Ion Torrent S5 System (Thermo Fisher Scientific) enables rapid and flexible targeted sequencing suitable for both research and clinical applications [17]. Its streamlined workflow allows efficient analysis of multiple samples within 24 - 48 h, making it feasible for hospital-based laboratories to implement molecular testing for cancer-associated variants [18].

Despite the growing application of MPS in somatic mutation profiling [19], the analytical validation of germline miR-SNP panels remains underexplored. Only a few studies have evaluated their performance in accordance with established quality frameworks such as those proposed by the Association for Molecular Pathology (AMP), the Clinical and Laboratory Standards Institute (CLSI), or the College of American Pathologists (CAP) [20]. At the same time, the systematic characterization of germline miR-SNPs, particularly in population-based contexts, has been limited [21] . Most previous investigations have focused on a small set of candidate variants, lacking validated tools capable of parallel genotyping across a broader spectrum of miR-SNPs [22-24].

This gap is even more pronounced in Vietnamese populations, for which comprehensive data on germline miR-SNPs and their association with breast cancer risk are largely unavailable [25, 26]. The absence of standardized analytical validation and population-specific data hinders accurate assessment of cumulative genetic effects, gene-gene interactions, and the development of localized risk prediction models.

In this study, we developed a customized MPS panel targeting 101 miR-SNP loci previously implicated in breast cancer based on our prior evidence [27]. Using the Ion S5 XL System, we conducted analytical validation to evaluate sequencing performance, concordance, reliability, and sensitivity of the panel. These assessments aimed to establish whether the assay could generate consistent and accurate results, thereby providing a validated platform for large-scale germline investigations of miR-SNPs and their contribution to breast cancer susceptibility in the Vietnamese population.

Materials and methods

Marker selection and MPS primer design

Candidate SNPs located within the mature miRNA coding regions were selected through a stepwise approach [26]. First, SNPs were retrieved from the miRNASNP-v3 database and cross-validated with dbSNP (MAF > 1%) and miRBase release 22 to confirm genomic coordinates and sequence annotation. Second, functional relevance was assessed using miRDB, where both reference and alternative miRNA sequences were evaluated, and only predicted targets with scores ≥ 80 were retained. KEGG enrichment analysis of these targets in the breast cancer pathway (hsa05224) was further applied to prioritize biologically significant variants. Third, expression data of target genes were examined in the GENT2 database; genes significantly upregulated in tumors (Log2FC > 0, $p < 0.05$) were defined as oncogenes, whereas those downregulated (Log2FC < 0, $p < 0.05$) were considered tumor suppressors, allowing classification of corresponding miRNAs as oncogenic or tumor-suppressive. Fourth, priority was given to SNPs located in seed regions or predicted to alter miRNA-target interactions. Binding sites were identified with TargetScan 8.0, and changes in minimum free energy (ΔG) between alleles were used as indicators of altered binding affinity [27]. Finally,

only SNPs that allowed successful primer design with amplicon sizes shorter than 200 bp were retained for sequencing.

Based on these criteria, a total of 101 SNPs were chosen to represent functionally relevant variants in mature. miRNA coding regions linked to breast cancer susceptibility. The genomic positions of these SNPs were compiled into a BED file and submitted to the Ion AmpliSeq Designer (Thermo Fisher Scientific, USA; ID: IAD249282, type: 236 accessed on March 18, 2024) for customized panel development. The design was generated using the Ion AmpliSeq Designer database, version 7.81, based on the GRCh38.p2 human reference genome (db = hg38, refGene v99).

Sample preparation

Peripheral blood samples were collected with approval from the Ethics Committee of the Oncology Hospital, Ho Chi Minh City, Vietnam (ethical no: 310/BVUB-HĐĐĐ/2025). Written informed consent was obtained from 50 participants prior to sampling. Genomic DNA was extracted from whole blood using the salting-out method as described by Hue *et al.* [28] DNA concentration and purity were first assessed by spectrophotometry using a NanoDrop 1000 (Thermo Fisher Scientific, USA), and samples with concentrations below 10 ng/μL were further quantified using a Qubit Fluorometer. DNA samples were stored at –20 °C until amplification.

MPS: Library construction, template preparation, and sequencing

For library preparation, 10 - 20 ng of genomic DNA was processed according to Ion AmpliSeq Designer recommendations using the VAHTS® AmpSeq Library Prep Kit V3 (Vazyme, China). PCR amplification was performed on a 96-well thermal cycler (Bio-Rad) under the following conditions: 99 °C for 2 min; 20 cycles of 99 °C for 15 s and 60 °C for 4 min; followed by 72 °C for 10 min. Amplicons were then partially digested with 2.5 μL VAHTS Digest Mix (50 °C for 10 min, 55 °C for 10 min, and 60 °C for 20 min). Adapter ligation was performed with barcode adapters, ligation enhancer, and DNA ligase at 22 °C for 30 min, followed by 68 °C for 5 min and 72 °C for 10 min. Purified libraries were re-amplified (5 cycles) with the HiFi Amplification Kit (Vazyme), cleaned using DNA Clean Beads, and quantified by the Qubit® dsDNA HS Assay kit (Thermo Fisher Scientific).

A total of 36 libraries were normalized to 30 pM and pooled in equal volumes, and 25 μL of the pooled mixture was loaded for automated template preparation using the Ion 510/520/530 Kit on the Ion Chef System (Thermo Fisher Scientific). Sequencing was performed on the Ion S5 XL platform using an Ion 510 chip, which was selected to provide sufficient throughput (~80 Mb per run) and optimal read length (~200 bp) for targeted SNP analysis.

Sequencing data acquisition and analysis

Library preparation and template loading were performed using the Ion Chef System (IC 5.12.3). Sequencing data were processed using Torrent Suite Software V5.12.3 (Thermo Fisher Scientific) on the Ion S5 XL System. Raw sequence reads were aligned to the human reference genome (hg38) using the Torrent Mapping Alignment Program V5.12.3 with default parameters. Variant calling was performed using the Torrent Variant Caller Plugin V5.12.28 under the germline low-stringency configuration. The Coverage Analysis Plugin V5.12.63 was applied to extract run-level sequencing metrics for each sample [29]. Sequencing quality was assessed according to manufacturer recommendations, with acceptance thresholds of on-target rate ≥ 90%, uniformity ≥ 85%, and mean depth ≥ 500× to ensure high analytical performance for subsequent variant analysis [21,30]

Variants were filtered based on the following key quality criteria: minimum read depth (≥ 20×), base quality score (≥ 20), and mapping quality (≥ 30). Genotypes were assigned based on Variant Allele Frequency (VAF) thresholds as follows: homozygous reference (VAF < 0.1), heterozygous (0.1 ≤ VAF ≤ 0.9), and homozygous alternate (VAF > 0.9) [31].

For data validation, the binary alignment map (BAM) and index (BAI) files were manually inspected using Integrative Genomics Viewer (IGV) v2.19.7 [32] and verified in R v4.5.1 with the Rsamtools and GenomicAlignments packages [33].

Sequencing performance and concordance study

To assess the performance and concordance of the customized germline miRNA-SNP panel for breast

cancer risk analysis, Human Reference DNA Female 5190-3797 (Agilent Technologies) together with 50 independent samples were analyzed. Run-level quality metrics, including total mapped reads, on-target percentage, coverage uniformity, mean and median depth of coverage, and the percentage of target regions achieving ≥ 100× and ≥ 2 00× coverage, were extracted.

Sequencing quality and reliability were evaluated using four statistical parameters: depth of coverage (DoC), locus coverage balance (LCB), frequency of major allele reads (F_{MAR}), and locus strand balance (LSB). The DoC, or read depth, was defined as the number of effective reads mapped within each locus. LCB was calculated as the ratio of coverage at a given locus to the average coverage across all loci in a single sample, with an ideal value of 1 indicating uniform amplification. F_{MAR} was calculated as the ratio of reads supporting the major allele to the total reads at each SNP position, expected to range between 90% and 100%. LSB was measured as the proportion of forward strand reads relative to total reads.

To confirm genotype accuracy, genotype calls obtained from the miR-SNP MPS were validated by Sanger sequencing for selected loci showing uncertain or low-confidence calls in the MPS data. Genotype agreement was summarized in a confusion matrix, and analytical accuracy was expressed as percent concordance with corresponding 95% confidence intervals (CI) calculated using the Wilson method.

Analytical performance evaluation

The analytical performance of the customized miR-SNP panel was comprehensively assessed in terms of reliability, sensitivity, stability, and reproducibility.

To evaluate inter-locus reliability, 3 replicates of the reference DNA at an input of 20 ng were sequenced within a single run. The inter-locus balance (ILB) was calculated as the ratio of the lowest to the highest locus coverage, with all replicates processed under identical conditions to minimize run-to-run variation.

Sensitivity was determined using serial dilutions of the reference DNA at input amounts of 20, 5, 1, 0.1 and 0.05 ng, each tested in triplicate. Libraries were constructed using 20 PCR cycles and sequenced in one run to assess the lowest input DNA concentration yielding reliable genotyping results.

Panel stability and reproducibility were further examined using DNA extracted from a single blood sample tested in triplicate. Intra-run repeatability was estimated from ILB values across replicates within the same sequencing run. To assess reproducibility, the same DNA sample was independently processed and sequenced on 2 separate days. All steps, including library preparation and sequencing, were performed using the same instrument and reagent lot. Reproducibility was evaluated as the percentage concordance of genotype calls across all target loci, with a predefined acceptance criterion of ≥ 99%, in accordance with analytical validation guidelines [21].

Results and discussion

Panel design

A total of 101 miR-SNPs were initially selected and submitted to the Ion AmpliSeq Designer (https://www.ampliseq.com/) for custom panel generation. Detailed information for each SNP, including the reference SNP identifier (rsID), chromosomal position based on the hg38 human genome build, and the corresponding mature miRNA symbol, is summarized in **Table S1**.

The final design comprised a 2-pool DNA panel containing 182 amplicons (91 per pool), achieving 100% target coverage across all 101 miR-SNP loci. Amplicon lengths ranged from 67 to 140 bp, with an average of 117.5 ± 12.8 bp, meeting the design requirements for optimal amplification and sequencing uniformity. Of the 101 SNPs, 18 were covered by a single amplicon, 79 were represented by 2 amplicons distributed across both pools, and 4 were covered by 3 amplicons in each pool (**Table S1**). This distribution reflects a balanced and redundant design strategy intended to ensure reliable target representation and minimize potential dropout across loci. Within these, 2 SNP pairs located in the same mature miRNA coding regions shared identical amplicon coverage (rs112489955 with rs149912461 in hsa-miR-501-3p, and rs7162033 with rs7183051 in hsa-miR-806), while another pair located in the same pre-miRNA region (rs115772313 in hsa-miR-3130-3p and rs2241347 in hsa-miR-3130-5p) also shared overlapping amplicons (Table S1). Allele frequency analysis revealed that the minor allele frequencies (MAFs) of the targeted miR-SNPs ranged from 0.005 to 0.995 (**Table S1**). More than

half of these SNPs had MAFs above 0.05, indicating common polymorphisms in the KHV population. The remaining loci showed very low or fixed MAFs, reflecting rare variants or allele fixation.

Hierarchical clustering analysis of 101 miR-SNP loci was conducted using pairwise Euclidean distances to evaluate genetic relatedness across miRNA genes (**Figure 1**). The observed divergence patterns suggest that miRNA-encoding regions harbor varying degrees of polymorphism, which may contribute to their differential roles in gene regulation and disease susceptibility.

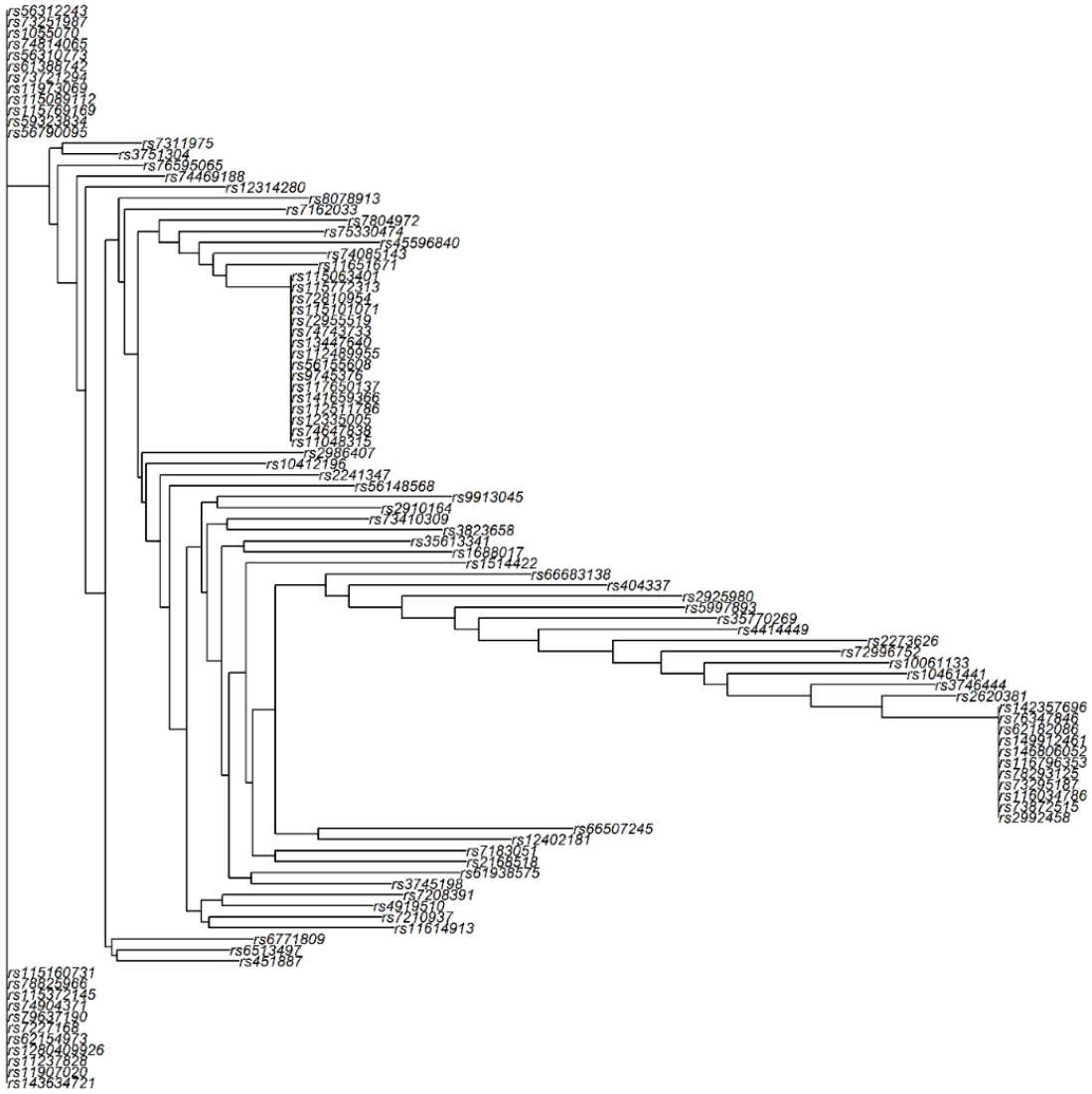


Figure 1 Hierarchical clustering constructed from 101 miR-SNP genotypes.

Performance evaluation of the miR-SNP panel

Sequencing performance of control and sample DNA libraries

Two chips were used to sequence control DNA and 50 sample DNA libraries. Chip 1 generated 6,133,229 reads with 94% ISP loading, while Chip 2 generated 3,475,795 reads with 96% ISP loading (**Figure S1**). Detailed sequencing metrics are summarized in **Table S2**. For the control DNA, the mean number of mapped reads was 170,108 (95% CI: 105,750 - 219,237). The proportion of on-target reads was 97.26% (95% CI: 97.24 - 97.29). The mean sequencing depth reached 1077× (95% CI: 670.1 - 1,395), with a uniformity of 88.96% (95% CI: 88.65 - 89.47). For the 50 DNA samples, the mean number of mapped reads was 85,194 (95% CI: 79,904 - 101,503). On-target reads accounted for 91.44% (95% CI: 90.68 - 94.45). The mean sequencing depth was 515.1× (95%

CI: 488.9 - 592.9), and uniformity was 88.77% (95% CI: 88.76 - 89.64) (**Table 1**). Overall, these results demonstrate that the sequencing performance was consistent and of high quality across both control and sample libraries.

Table 1 Summary of sequencing metrics for control and sample DNA libraries.

Sample	Mean (95% CI)			
	Mapped reads	**On-target**	**Mean depth**	**Uniformity**
Control DNA	170,108 (105,750 - 219,237)	97.26 (97.24 - 97.29)	1,077 (670.1 - 1,395)	88.96 (88.65 - 89.47)
Sample DNAs	85,194 (79,904 - 101,503)	91.44 (90.68 - 94.45)	515.1 (488.9 - 592.9)	88.77 (88.76 - 89.64)

Locus coverage balance

To assess sequencing uniformity, LCB was calculated for all targeted miR-SNPs. As shown in **Figure 2**, LCB values ranged from 0.00017 ± 0.00083 (rs12473206, hsa-miR-4433a-5p) to 1.164 ± 0.0301 (rs115089112, hsa-miR-580-5p). Most loci achieved an LCB close to the ideal value of 1, indicating balanced allelic representation and consistent sequencing performance across the panel. Four loci, rs72810954, rs12473206, rs266435, and rs2155248, did not meet the quality thresholds, resulting in "No Call" assignments. Among these SNPs, rs72810954 showed low coverage (DoC = 26.15 ± 22.10×) but a balanced allelic distribution. This pattern suggests a higher likelihood of no-call rather than inaccurate genotyping, which may be mitigated by optimizing primer concentration in the library pool. In contrast, the remaining 3 loci failed to amplify, reflecting missing data rather than allelic bias or misgenotyping. These findings collectively support the robustness of the panel design, with only a small subset of loci requiring optimization.

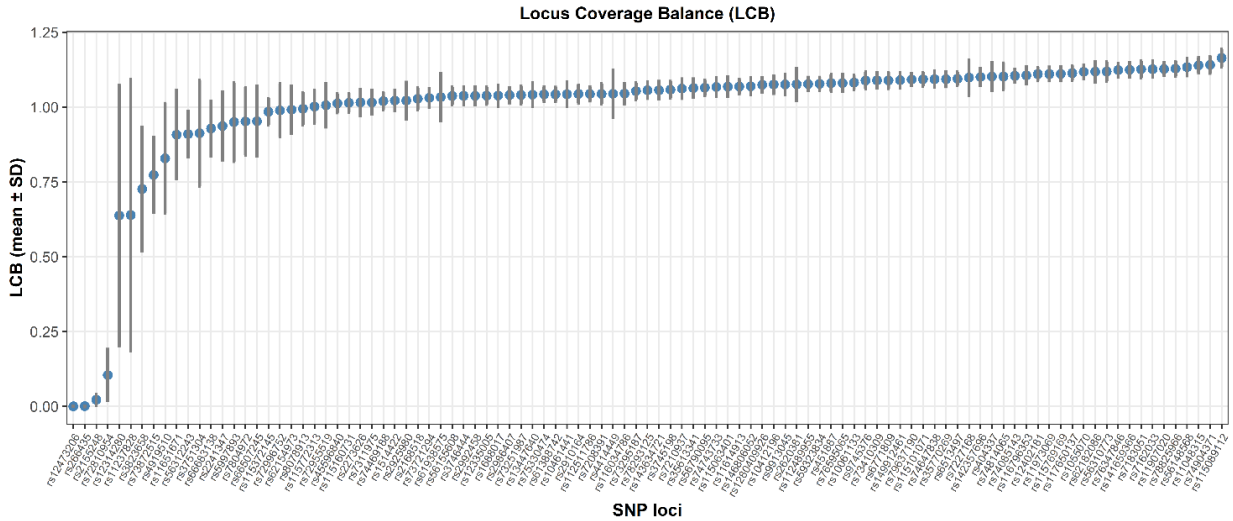


Figure 2 Locus coverage balance of 101 miR-SNP.

Frequency of major allele reads

Given the diploid nature of autosomal miR-SNPs, accurate genotyping was based on F_{MAR} thresholds, with values > 90% required for homozygous calls and ~50% for heterozygous calls. In our dataset, miR-SNPs displayed F_{MAR} values > 90% for homozygous sites and approximately 50% for heterozygous sites, consistent with expected genotypes (**Table S3**). The only exception was rs11237828 (hsa-miR-5579-3p), which showed an F_{MAR} of 88.89%, slightly below the threshold (**Figure 3**). Lower sequencing coverage likely accounted for non-allelic genotype calls, leaving 3 miR-SNPs (rs12473206, rs266435, and rs2155248) unclassified (**Table 2**).

Occasional deviations were observed at loci located near homopolymeric tracts, which are known to affect Ion Torrent sequencing [34]. As summarized in **Table 2**, 5 SNPs (rs1688017, rs2273626, rs404337, rs78293125, and rs12473206) were situated within or near homopolymeric runs [(6G), (6A), (6C), and 6(C), respectively]. Notably, rs12473206 coincided directly with a homopolymeric (6C) stretch at the SNP position itself. The presence of these homopolymers likely contributed to minor read misalignment, particularly evident at rs2273626, as detailed in **Table S3**.

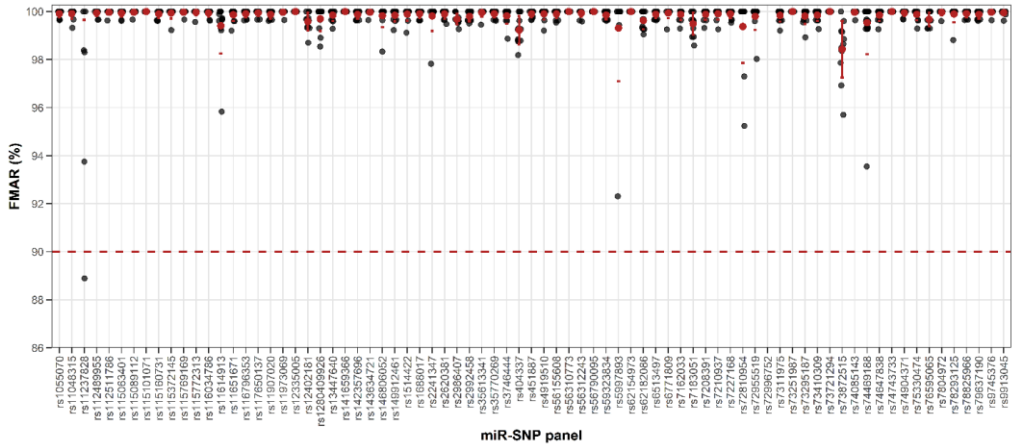


Figure 3 Allelic performance of homozygous miR-SNPs.

Locus strand balance

LSB defined as the proportion of forward strand coverage relative to total coverage, was used to assess the uniformity of sequencing reads between strands. The average LSB across all miR-SNP loci was 0.4858 ± 0.1403, indicating an overall balanced strand representation. Most loci fell within the acceptable range of 0.25 - 0.75, suggesting minimal strand-specific bias.

Apart from 3 loci with low total coverage, five SNPs displayed mild strand imbalance, including rs62154973 (hsa-miR-4772-5p), rs66507245 (hsa-miR-4731-3p), rs73872515 (hsa-miR-548t-3p), rs3823658 (hsa-miR-5090), and rs8078913 (hsa-miR-4520-3p) (**Figure 4**). This imbalance may be partially attributed to the presence of homopolymeric tracts on either the forward or reverse strands, which can interfere with nucleotide extension efficiency during sequencing reactions.

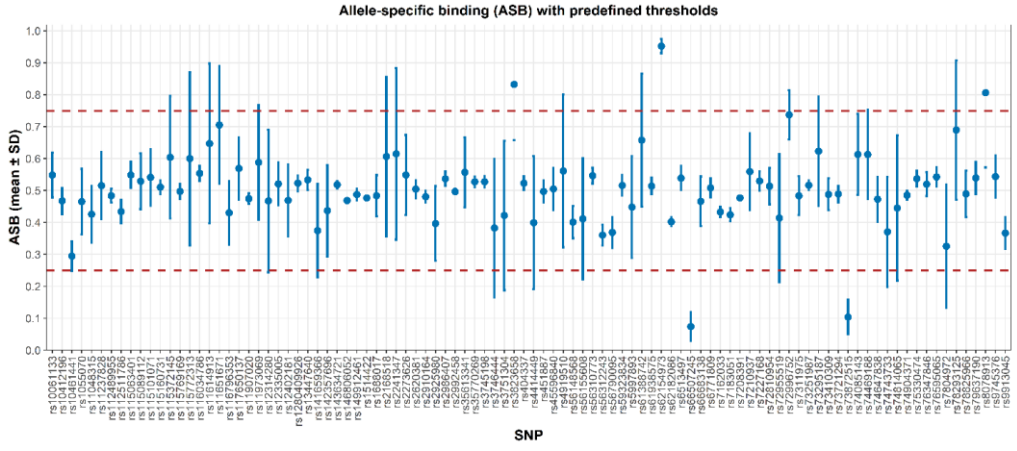


Figure 4 Distribution of Amplicon strand bias values across miR-SNP panel.

Concordance study

Using the customized miR-SNP panel, 97 out of 101 targeted loci were successfully sequenced in both the control DNA and 50 unrelated female samples. Following data processing and variant calling, high-quality genotypes were obtained for these 97 loci, while four loci with low coverage were assigned a no-call status.

To further assess genotyping accuracy, ten representative SNPs showing uncertain allele calls, homopolymeric interference, or mild strand bias were selected for validation by Sanger sequencing. Primer details are provided in **Table S4**. The Sanger sequencing results were fully consistent with those obtained from the customized panel (**Table 2**, **Figure S2**). Based on these findings, both the sensitivity and specificity for these ten loci were calculated to be 100% (95% CI: 72.3% - 100%).

Collectively, these results confirm the analytical reliability and robustness of the customized miR-SNP panel. The panel accurately genotypes multiple loci, including those located in technically challenging regions such as homopolymeric tracts or unbalanced amplicons.

Table 2 Problematic miR-SNPs identified through analyses.

MiRNA	SNP	Flanking Sequence (±15 bp)	Performance	Genotype by MPS	Genotype by Sanger
hsa-miR-1304-3p	rs2155248	AGCCCAGGGGTTCGA[G/A/C/T]GCTACAGTGAGATGT	Low coverage	NN	NA
hsa-miR-4804-5p	rs266435	GTCAGTGTATTTGGA[C/A/G/T]GGTAAGGTTAAGC	Low coverage	NN	NA
hsa-miR-4679	rs72810954	TGTTAGAAACAAAAA[G/A]CAAAGAATCTCTATC	Low coverage	NN	NA
hsa-miR-4433a-5p	rs12473206	TTACGTCCCACCCCC[C/A/G/T]ACTCCTGTTTCTGGT	Low coverage, homopolymer (6C)	NN	NA
hsa-miR-5579-3p	rs11237828	GGTGATTTGATCTGG[T/C]ACTCCTTAAGCTAAT	Uncertain allele call (F_{MAR} < 90%)	NN	TT
hsa-miR-4772-5p	rs62154973	AGGCAAAATTGCAGA[C/T]TGTCTTCCCAAATAG	Strand bias	CC	CC
hsa-miR-4731-3p	rs66507245	AGTGTTGGGGGCCAC[T/A/C]TGTGTGTGGATGACT	Strand bias	TA	TA
hsa-miR-548t-3p	rs73872515	TTTTTAATGACAAAA[A/C]CCACAATTACTTTTG	Strand bias	AA	AA
hsa-miR-5090	rs3823658	CCTTCTGAGGTACCC[G/A/T]GGGCAGATTGGTGTA	Strand bias	GA	GA
hsa-miR-4520-3p	rs8078913	GGTTGATTCCTTCTT[C/A/G/T]CTGCGTGTTTTCTGT	Strand bias	CT	CT
hsa-miR-4707-3p	rs2273626	GAACCTCGGCTGGGG[C/A/T]GGGCTGGCCAGCAGC	Homopolymer (7G)	CA	CA
hsa-miR-6887-5p	rs1688017	ATGGGGGGACAGATG[G/A]AGAGGACACAGGCTG	Homopolymer (6G)	GG	GG
hsa-miR-8084	rs404337	TAAATAGAATACTAA[G/A/T]TAAAAAATCAGTATG	Homopolymer (6A)	AA	AA
hsa-miR-6885-5p	rs78293125	TGGCTTTGCTTGCGC[A/G]GTGCCCCCCTCCAGG	Homopolymer (6C)	AA	AA

NN: No call; NA: Not Applicable.

Reliability study

The performance of the customized SNP panel using MPS was further examined through reproducibility and repeatability testing using Control DNA. Consistent and complete genotype calls were obtained from all sequencing runs with 20 ng input DNA. In the reliability evaluation, except for five SNPs (rs12473206, rs266435, rs2155248, rs11237828 and rs72810954), 3 independent replicates of reference DNA generated concordant results across all analyzed loci, with mean DoC of 846.1 ± 464.7×, 1,491 ± 988.1× and 1,770 ± 970× (**Table S5**). Inter-locus balance, defined as the ratio between the lowest and highest-covered loci, was also assessed among these replicates. Except for 3 non-amplified loci (rs12473206, rs266435 and rs2155248), the locus rs72810954 exhibited the lowest DoC (31.00 ± 8.72×), while rs116034786 showed the highest mean DoC (2,985 ±1,282×). These observations were in line with the general sequencing performance trends described above.

Sensitivity study

Serial dilutions of control DNA (20, 5, 1, 0.1 and 0.05 ng) were analyzed to evaluate the impact of template quantity on genotype calling rate and sequencing depth under standard library preparation conditions (20 PCR cycles). Concordant genotype rates were 96.67% (95% CI: 90 - 100) for 20 ng, 97.78% (95% CI: 96.67 - 100) for 5 ng, 100% (95% CI: 100 - 100) for 1 ng, 76.67% (95% CI: 30 - 100) for 0.1 ng, and 90.00% (95% CI: 80.00 - 96.67) for 0.05 ng (**Figure 5(a)**). Full genotype profiles were achievable at 1 ng, although coverage at some loci was slightly reduced compared with higher inputs (20 and 5 ng). Below 1 ng, both genotype calling rate and mean sequencing depth progressively declined (**Figures 5(a)** and **5(c)**). At 0.1 ng, concordant genotype rates varied substantially among replicates (100%, 100% and 30%), suggesting stochastic amplification and allelic dropout at ultra-low template levels. The apparent recovery observed at 0.05 ng likely reflects random variation rather than true assay robustness. Overall, DNA inputs below 1 ng showed inconsistent performance due to limited template availability and amplification bias.

Logistic regression modeling demonstrated a statistically significant positive association between DNA input and probability of correct genotype calls ($\beta = 0.103$, $p = 0.012$). The model indicated that DNA inputs ≥ 1 ng were sufficient to achieve ≥ 95% callability (**Figure 5(b)**). In contrast, inputs below 1 ng were associated with reduced coverage and an increased risk of false-negative calls, particularly at 0.1 - 0.05 ng. Segmented regression identified a breakpoint at 0.99 ng, with stable genotype concordance above this threshold and a sharp decline below it. Moreover, sequencing coverage showed a strong positive correlation with input DNA amount (Pearson's r = 0.89) (**Figure 5(d)**).

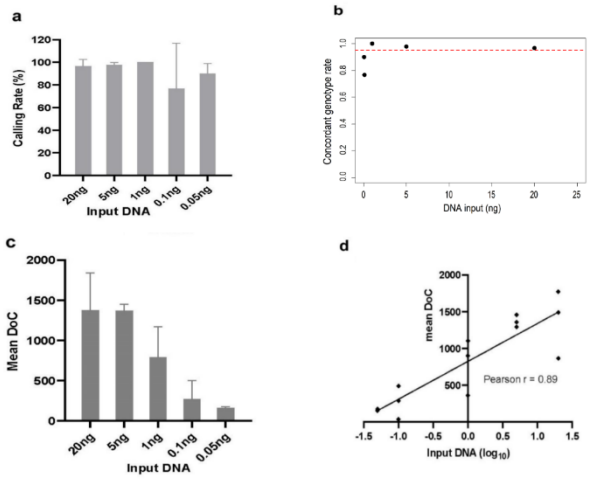


Figure 5 Sensitivity analysis using serial dilutions of control DNA; (a) Variant calling rate, (b) Logistic curve of genotype callability by DNA input, (c) Mean coverage depth and (d) Correlation between input DNA amount and mean coverage depth.

Stability and reproducibility study

To assess the consistency of the customized 101 miR-SNP panel, a blood sample was sequenced in 3 independent replicates. Except for rs12473206, rs266435, rs2155248 and rs72810954, complete genotype profiles were obtained across all runs, and the results were concordant across the 101 loci examined (**Table S6**). The mean DoC was 1,737 ± 1,122×, 1,892 ± 984× and 902.6 ± 493.6× in the 3 replicates. Similarly, the calling rates were comparable between replicates (97.03%, 96.04% and 96.04%), indicating consistent sequencing performance.

The intraclass correlation coefficient (ICC) and Bland-Altman plot were used to evaluate the reproducibility of sequencing depth between 2 independent runs. Complete genotype profiles were successfully obtained for all loci across both sequencing runs, except for five miR-SNPs (rs11237828, rs12473206, rs2155248, rs266435 and rs72810954) that exhibited low coverage (**Table S7**). The mean depth and calling rate were highly comparable between Run 1 (mean DoC = 232.6 ± 62.82×; callability = 95.05%) and Run 2 (mean DoC = 152.1±73.73×; callability = 95.05%) (**Table S7**), suggesting consistent sequencing performance. The single-measure ICC for fixed raters (ICC (3,1) was 0.70 (95% CI: 0.59 - 0.79), indicating good consistency between runs (**Table 3**). When the mean value of both runs was considered (ICC (3,2), the reproducibility improved to 0.83 (95% CI: 0.74 - 0.88) (**Table 3**), demonstrating high reliability of depth measurements across sequencing runs. As shown in **Figure 6**, the Bland-Altman plot illustrates that most data points fell within the 95% limits of agreement, with a minimal mean bias between Run 1 and Run 2, indicating no substantial systematic deviation.

Table 3 Inter-run reproducibility of sequencing depth between 2 independent runs.

Type	ICC	F	df1	df2	*p*_value	Lower bound	Upper bound
ICC(3,1)	0.70	5.77	100	100	3.59E-17	0.59	0.79
ICC(3,2)	0.83	5.77	100	100	3.59E-17	0.74	0.88

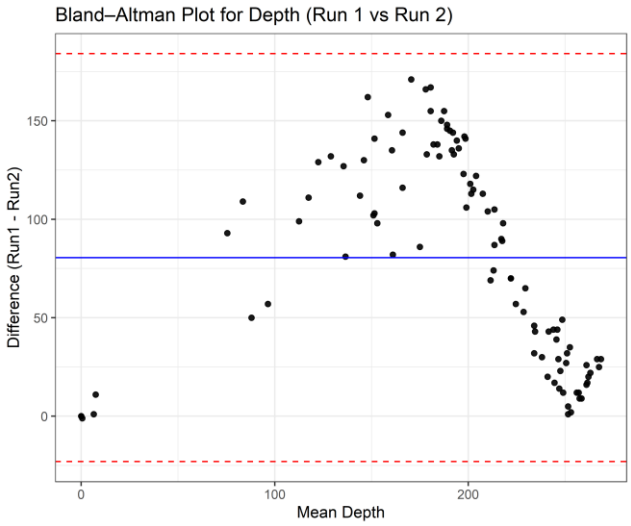


Figure 6 Bland-Altman plot showing agreement of sequencing depth between 2 independent runs.

Together, these findings confirm that the custom 101-miR-SNP AmpliSeq panel yields consistent sequencing performance across runs. The high reproducibility of depth measurements suggests that the assay is technically robust and suitable for downstream applications, including large-scale case–control studies assessing germline miR-SNP associations with breast cancer risk.

Compared with conventional genotyping methods such as SNaPshot or single-plex PCR–based assays, the customized MPS panel allows simultaneous interrogation of a larger number of loci [35], thereby improving throughput and operational efficiency. By

contrast, most commercially available SNP panels are optimized for ancestry inference or somatic mutation profiling [36]. The present panel was specifically designed to target germline miR-SNPs with potential relevance to breast cancer. As such, it fulfills a distinct methodological need in population-based genetic association studies.

Nevertheless, four loci exhibited suboptimal performance. In particular, rs72810954 showed consistently low coverage, suggesting that further optimization may be required. Adjusting primer concentrations within the multiplex library preparation could improve its amplification efficiency. In contrast, 3 loci failed to amplify reproducibly and were therefore excluded from downstream analyses. Future panel optimization may incorporate redesigned primers or alternative amplicon strategies to improve overall locus coverage.

Conclusions

In this study, we developed and validated a custom MPS assay targeting 101 miR-SNPs previously reported to be associated with breast cancer. Sequencing on the Ion S5 XL System demonstrated high accuracy and resolution across nearly all loci, with four showing suboptimal performance. The assay also achieved stable and reproducible results with as little as 1 ng of input DNA. Overall, the customized miR-SNP MPS panel represents a robust and reliable tool for high-throughput genotyping in breast cancer association studies.

Acknowledgements

This research is funded by Vietnam National University, Ho Chi Minh City (VNU-HCM), Vietnam under grant number 562-2024-18-11.

Declaration of Generative AI in Scientific Writing

We acknowledge the use of ChatGPT (OpenAI) to assist in improving the English language and readability of the manuscript. No content generation or data interpretation was performed by AI. The authors take full responsibility for the content and conclusions of this work.

CRediT Author Statement

Thuy Thi Chung Duong: Conceptualization, Data curation, Investigation, Methodology, Software, Writing -original draft. **Hue Thi Nguyen:** Conceptualization, Methodology, Data curation, Software, Supervision, Writing -review & edicting. **Nga Thi Nguyen:** Conceptualization, Investigation, Methodology, Visualization, Writing -review & edicting. **Luan Huu Huynh:** Conceptualization, Investigation, Methodology, Software, Writing -review & edicting. **Thinh Hung Nguyen:** Conceptualization, Investigation, Formal analysis, Software, Writing - review & edicting. **Tuan Huu Ngoc Nguyen:** Conceptualization, Investigation, Formal analysis, Software, Writing -review & edicting. **Thanh Thi Ngoc Nguyen:** Conceptualization, Methodology, Data curation, Software, Resource, Validation, Project administration, Writing - review & edicting.

References

[1] S Ding and P Wang. The life of microRNAs: Biogenesis, function and decay in cancer. *Biomolecules* 2025; **15(10)**, 1393.

[2] A Syeda, SSS Langden, C Munkhzul, M Lee and SJ Song. Regulatory mechanism of microRNA expression in cancer. *International Journal of Molecular Sciences* 2020; **21(5)**, 1723.

[3] HY Loh, BP Norman, KS Lai, N Rahman, NBM Alitheen and MA Osman. The regulatory role of microRNAs in breast cancer. *International Journal of Molecular Sciences* 2019; **20(19)**, 4940.

[4] W Jelski, S Okrasinska and B Mroczko. MicroRNAs as biomarkers of breast cancer. *International Journal of Molecular Sciences* 2025; **26(9)**, 4395.

[5] S Chauhan, R Mathur and AK Jha. The impact of microRNA SNPs on breast cancer: Potential biomarkers for disease detection. *Molecular Biotechnology* 2024; **67(3)**, 845-861.

[6] Y Chhichholiya, AK Suryan, P Suman, A Munshi and S Singh. SNPs in miRNAs and target sequences: Role in cancer and diabetes. *Frontiers in Genetics* 2021; **12**, 793523.

[7] Y Wang and GC Hon. Towards functional maps of non-coding variants in cancer. *Frontiers in Genome Editing* 2024; **6**, 1481443.

[8] T Arancibia, S Morales-Pison, E Maldonado and L Jara. Association between single-nucleotide polymorphisms in miRNA and breast cancer risk:

An updated review. *Biological Research* 2021; **54(1)**, 26.

[9] T Kongton, S Sanguansin, P Saelee, A Thongmee and W Pongstaporn. The association of pri-miR34 b/c gene polymorphism and clinicopathologic data in breast cancer patients. *Asian Pacific Journal of Cancer Prevention* 2024; **25(7)**, 2415-2420.

[10] G Hina, N Khan, ST Muntaha, A Ullah, S Khan, A Alanzi, I Ali and T Chen. Association analysis of miRNA-146a and miRNA-196a polymorphism with breast cancer risk in Khyber Pakhtunkhwa, Pakistan: A preliminary study. *Molecular Biology Reports* 2024; **51(1)**, 827.

[11] AM McInerney-Leo and EL Duncan. Massively parallel sequencing for rare genetic disorders: Potential and pitfalls. *Frontiers in Endocrinology* 2020; **11**, 628946.

[12] F Guo, J Yu, L Zhang and J Li. Massively parallel sequencing of forensic STRs and SNPs using the Illumina® ForenSeq™ DNA Signature Prep kit on the MiSeq FGx™ Forensic Genomics system. *Forensic Science International: Genetics* 2017; **31**, 135-148.

[13] CKY Ng, AM Schultheis, FC Bidard, B Weigelt and JS Reis-Filho. Breast cancer genomics from microarrays to massively parallel sequencing: Paradigms and new insights. *Journal of the National Cancer Institute* 2015; **107(5)**, djv015.

[14] C Kraus, J Hoyer, G Vasileiou, M Wunderle, MP Lux, PA Fasching, M Krumbiegel, S Uebe, M Reuter, MW Beckmann and A Reis. Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2. *International Journal of Cancer* 2017; **140(1)**, 95-102.

[15] JR Bradford, Y Hey, T Yates, Y Li, SD Pepper and CJ Miller. A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. *BMC Genomic*s 2010; **11(1)**, 282.

[16] LJ Semenuk, BA Kartolo, HE Feilotter, SM Lee, CA Savage, AH Boag, GC Digby and M Mates. Implementing next-generation sequencing process changes to increase capacity and improve timeliness of molecular biomarker profiling for lung cancer patients. *The Journal of Applied Laboratory Medicine* 2023; **9(2)**, 284-294.

[17] S Shin, Y Kim, SC Oh, N Yu, ST Lee, JR Choi and KA Lee. Validation and optimization of the Ion Torrent S5 XL sequencer and Oncomine workflow for BRCA1 and BRCA2 genetic testing. *Oncotarget* 2017; **8(21)**, 34858-34866.

[18] M Mehrotra, DY Duose, RR Singh, BA Barkoh, J Manekia, MA Harmon, KP Patel, MJ Routbort, LJ Medeiros, II Wistuba and R Luthra. Versatile ion S5XL sequencer for targeted next generation sequencing of solid tumors in a clinical laboratory. *PLoS One* 2017; **12(8)**, e0181968.

[19] M Ivanov, K Laktionov, V Breder, P Chernenko, E Novikova, E Telysheva, S Musienko, A Baranova and V Mileyko. Towards standardization of next-generation sequencing of FFPE samples for clinical oncology: Intrinsic obstacles and possible solutions. *Journal of Translational Medicine* 2017; **15(1)**, 22.

[20] T Huebner, M Steffens and C Scholl. Current status of the analytical validation of next generation sequencing applications for pharmacogenetic profiling. *Molecular Biology Reports* 2023; **50(11)**, 9587-9599.

[21] LJ Jennings, ME Arcila, C Corless, S Kamel-Reid, IM Lubin, J Pfeifer, RL Temple-Smolkin, KV Voelkerding and MN Nikiforova. Guidelines for validation of next-generation sequencing-based oncology panels: A joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. *The Journal of Molecular Diagnostics* 2017; **19(3)**, 341-365.

[22] JT Bensen, CK Tse, SJ Nyante, JS Barnholtz-Sloan, SR Cole and RC Millikan. Association of germline microRNA SNPs in pre-miRNA flanking region and breast cancer risk and survival: The carolina breast cancer study. *Cancer Causes & Control* 2013; **24(6)**, 1099-1109.

[23] D Chacon-Cortes, RA Smith, LM Haupt, RA Lea, PH Youl and LR Griffiths. Genetic association analysis of miRNA SNPs implicates MIR145 in breast cancer susceptibility. *BMC Medical Genetics* 2015; **16**, 107.

[24] J Gong, Y Tong, HM Zhang, K Wang, T Hu, G Shan, J Sun and AY Guo. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and

biogenesis. *Human Mutation* 2012; **33(1)**, 254-263.

[25] LH Huu, T Nguyen and N Hue. Association between single nucleotide polymorphism rs4919510 (c > g) in miRNA-608 and breast cancer in Vietnamese population. *International Research Journal of Obstetrics and Gynecology* 2019; **2**, 11.

[26] TTH Minh, NTN Thanh, TV Thiep and NT Hue. *Association between single nucleotide polymorphism rs11614913 (C>T) on miR-196a2 and breast cancer in Vietnamese population*. *In*: TV Van, TAN Le and TN Duc (Eds.). 6th international conference on the development of biomedical engineering in Vietnam (BME6). Springer, Singapore, 2018, p. 381-386.

[27] TTN Thi Ngoc Nguyen, THN Nguyen, LH Huynh, HN Phan and HT Nguyen. *Predicting SNPs in mature microRNAs dysregulated in breast cancer*. *In*: L Tutar (Ed.). Recent advances in noncoding RNAs. IntechOpen, London, 2022.

[28] NT Hue, NDH Chan, PT Phong, NTT Linh and NDT Giang. Extraction of human genomic DNA from dried blood spots and hair roots. *International Journal of Bioscience, Biochemistry and Bioinformatics* 2012; **2(1)**, 21-26.

[29] S Roy, C Coldren, A Karunamurthy, NS Kip, EW Klee, SE Lincoln, A Leon, M Pullambhatla, RL Temple-Smolkin, KV Voelkerding, C Wang and AB Carter. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines: A joint recommendation of the Association for Molecular Pathology and the College of American Pathologists. *The Journal of Molecular Diagnostics* 2018; **20(1)**, 4-27.

[30] TF Scientific. Ion AmpliSeq™ designer performance specifications for custom DNA panel, Available at: https://www.ampliseq.com/otherContent/help-content/help_html/GUID-5193AC77-0A80-488E-B26A-1CD138BE1D21.html?utm_source=chatgpt.com, accessed October 2025.

[31] F Muyas, M Bosio, A Puig, H Susak, L Domènech, G Escaramis, L Zapata, G Demidov, X Estivill, R Rabionet and S Ossowski. Allele balance bias identifies systematic genotyping errors and false disease associations. *Human Mutation* 2019; **40(1)**, 115-126.

[32] H Thorvaldsdóttir, JT Robinson and JP Mesirov. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatic*s 2012; **14(2)**, 178-192.

[33] W Huber, VJ Carey, R Gentleman, S Anders, M Carlson, BS Carvalho, HC Bravo, S Davis, L Gatto, T Girke, R Gottardo, F Hahne, KD Hansen, RA Irizarry, M Lawrence, MI Love, J MacDonald, V Obenchain, AK Oleś, H Pagès, A Reyes, …, M Morgan. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* 2015; **12(2)**, 115-121.

[34] LM Bragg, G Stone, MK Butler, P Hugenholtz and GW Tyson. Shining a light on dark sequencing: Characterising errors in Ion Torrent PGM data. *PLoS Computational Biology* 2013; **9(4)**, e1003031.

[35] D Ballard, J Winkler-Galicki and J Wesoły. Massive parallel sequencing in forensics: Advantages, issues, technicalities, and prospects. *International Journal of Legal Medicine* 2020; **134(4)**, 1291-1303.

[36] I Avent, AG Kinnane, N Jones, I Petermann, R Daniel, ME Gahan and D McNevin. The QIAGEN 140-locus single-nucleotide polymorphism (SNP) panel for forensic identification using massively parallel sequencing (MPS): An evaluation and a direct-to-PCR trial. *International Journal of Legal Medicine* 2019; **133(3)**, 677-688.

Supplementary Materials

Table S1 Genetic information of targeted miR-SNPs in the custom MPS-SNP panel.

No.	SNP ID	miRNA	Chromosome	Position	Ref./Alt allels	Flanking Sequence(±15bp)	Number of Amplicon	MAF
1	rs10061133	hsa-miR-449b-5p	chr05	55170716	A/G	AAGCAGCCAGCTAAC[A/G]AT ACACTGCCTACCT	2	0.227
2	rs10412196	hsa-miR-6801-3p	chr19	52222085	T/C	ACTCACCCCTGCCAC[T/C]CAC TGGCCCCCAGGT	2	0.141
3	rs10461441	hsa-miR-548ae-5p	chr05	58530093	A/G	TAAATGACAAAAACC[A/G]CA ATTACTTTTGCAC	2	NA
4	rs1055070	hsa-miR-4700-3p	chr12	120723245	T/G	ATTCACAGGACTGAC[T/G]CCT CACCCCAGTGCA	2	0
5	rs11048315	hsa-miR-4302	chr12	25874055	G/A	AAACTCAGCCACCTC[G/A/T]C TGAGCCACACTGGT	2	0
6	rs11237828	hsa-miR-5579-3p	chr11	79422176	T/C	GGTGATTTGATCTGG[T/C]ACT CCTTAAGCTAAT	2	0.374
7	rs112489955	hsa-miR-501-3p	chrX	50009781	G/A	GAATGCAATGCACCC[G/A/T]G GCAAGGATTCTGAG	2	0
8	rs112511786	hsa-miR-5691	chr11	9090358	G/C	TCTGTCAGCTTTCTC[G/A/C]GA GCTCAGAGCAAGC	2	0
9	rs115063401	hsa-miR-4786-5p	chr2	239943064	G/A	ATCACAGGGAGAGTG[G/A]TG CATCCAGTCCTGG	2	0
10	rs115089112	hsa-miR-580-5p	chr5	36147955	T/C	TTAAATATCTGAGTC[T/A/C/G] GATGAATCATTAGGT	2	0
11	rs115101071	hsa-miR-4467	chr7	102471476	G/A	GGCAGAGGTGGTGGC[G/A/C]G CGGTAGTTATGGGC	2	0
12	rs115160731	hsa-miR-3124-3p	chr1	248826432	C/A	GACTTTCCTCACTCC[C/A/T]GT GAAGTCGGCGGAA	2	0
13	rs115372145	hsa-miR-942-3p	chr1	117094703	C/T	CCCCTCACACATGGC[C/T]GAA ACAGAGAAGTTA	2	0
14	rs115769169	hsa-miR-4457	chr5	1309319	C/T	CTCTTCTGAATACGC[C/T]AGT CAATACCTTGTG	2	0
15	rs115772313	hsa-miR-3130-3p	chr2	206783285	G/A	ACAACAGGCTGCACC[G/A/T]G AGACTGGGTAAGAC	2	0
16	rs116034786	hsa-miR-4514	chr15	80997457	A/G	AAAGATGTTTCCCCA[A/G]TCC TGCCTGTCTCAA	2	0
17	rs11614913	hsa-miR-196a-3p	chr12	53991815	C/T	GGCAACAAGAAACTG[C/G/T]C TGAGTTACATCAGT	2	0.444
18	rs11651671	hsa-miR-548at-5p	chr17	42494785	G/A	TGCAAAAGTTATTGC[G/A/T]G TTTTGGCTGCCAAA	2	0.005
19	rs116796353	hsa-miR-5589-3p	chr19	10038396	A/G	CCTACACAACTGCAC[A/G]TG GCAACCTAGCTCC	2	0
20	rs117650137	hsa-miR-6717-5p	chr14	21023373	G/A	CTACATCCCCACATC[G/A/T]C CTCAAACACCAGGC	2	0
21	rs11907020	hsa-miR-3192-3p	chr20	18470681	T/C	GATCGCCCTCTCAGC[T/A/C]C TTTCCTTCTGACTG	2	0
22	rs11973069	hsa-miR-4284	chr7	73711334	C/T	TCTGTGAGGGGCTCA[C/T]ATC ACCCCATCAAAG	2	0
23	rs12314280	hsa-miR-5700	chr12	94561809	T/C	TAATGCATTAAATTA[T/C]TGA AGGCCCTTGGGC	2	0.096
24	rs12335005	hsa-miR-4661-3p	chr8	91205534	G/T	CAGCCTGATAGACAG[G/T]AT CCACAGAGCTAGT	1	0
25	rs12402181	hsa-miR-3117-3p	chr1	66628488	G/A	AGGGAAGGCATTATA[G/A/T]G ACTCATATAGTGCC	2	0.308
26	rs12473206	hsa-miR-4433a-5p	chr2	64340782	C/G	TTACGTCCCACCCCC[C/A/G/T] ACTCCTGTTTCTGGT	2	0.005
27	rs1280409926	hsa-miR-6077	chr1	148388307	C/T	AAGGGAAGAGCTGTA[C/T]GG CCTTCGCGCTAGT	2	NA
28	rs13447640	hsa-miR-548l	chr11	94466555	G/A	TCTACGACAAAACCC[G/A/T]C AAATACTTTTGCAC	2	0
29	rs141659366	hsa-miR-548ar-3p	chr13	114244551	G/A	AGTGGTAAAACTGCA[G/A/T]T TATTTTTGCACCAA	2	0
30	rs142357696	hsa-miR-4444	chr3	75214531	A/G	AACTCGAGTTGGAAG[A/G]GG CGAGTCCGGTCTC	2	0
31	rs143634721	hsa-miR-888-5p	chrX	145994837	C/A	AATCTAAGTGACTGA[C/A]AG CTTTTTGAGTAGA	1	0
32	rs146806052	hsa-miR-892b	chrX	145997215	A/G	AGCCTTGCTCTACCC[A/G]GAA AGGAGCCAGTGA	2	0
33	rs149912461	hsa-miR-501-3p	chrX	50009773	A/G	TGCTTTCTGAATGCA[A/G]TGC ACCCGGGCAAGG	2	0

No.	SNP ID	miRNA	Chromosome	Position	Ref./Alt allels	Flanking Sequence(±15bp)	Number of Amplicon	MAF
34	rs1514422	hsa-miR-8060	chr3	96360020	G/A	CCCATGAAGCAGTGG[G/A/C]T AGGAGGACAGGAAA	2	0.202
35	rs1688017	hsa-miR-6887-5p	chr19	35122719	G/A	ATGGGGGGACAGATG[G/A]AG AGGACACAGGCTG	2	0.237
36	rs2155248	hsa-miR-1304-3p	chr11	93733700	G/T	AGCCCAGGGGTTCGA[G/A/C/T]GCTACAGTGAGATGT	2	0.04
37	rs2168518	hsa-miR-4513	chr15	74788737	G/A	TATGGGCCTCCAGCC[G/A/C/T] TCAGTCTCCCCACCT	2	0.202
38	rs2241347	hsa-miR-3130-5p	chr2	206783257	C/T	GTCTTACCCAGTCTC[C/A/T]G GTGCAGCCTTGACA	2	0.374
39	rs2273626	hsa-miR-4707-3p	chr14	22956973	C/A	GAACCTCGGCTGGGG[C/A/T]G GGCTGGCCAGCAGC	2	0.162
40	rs2620381	hsa-miR-627-5p	chr15	42199650	A/C	TTTTCTTAGAGACTC[A/C]CTA CCAGTAATAAGT	2	0.091
41	rs266435	hsa-miR-4804-5p	chr5	72878605	C/G	GAACCTCGGCTGGGG[C/A/T]G GGCTGGCCAGCAGC	2	0.157
42	rs2910164	has-miR-146a-3p	chr5	160485411	C/G	GTCAGTGTCAGACCT[C/G]TGA AATTCAGTTCTT	3	NA
43	rs2925980	hsa-miR-7854-3p	chr16	81533949	A/G	TGGGCAGCTGAGGTG[A/C/G/T]CCGCAGATGGGAAGG	2	0.465
44	rs2986407	hsa-miR-1343-5p	chr11	34941869	T/C	GGAGCGGCCCCCGGG[T/A/C]G GGCCTCTGCTCTGG	1	0.172
45	rs2992458	hsa-miR-5087	chr1	148334528	A/G	AACACTTAACATGCC[A/G]GC AAAGCTACAAACC	2	0
46	rs35613341	hsa-miR-5189-3p	chr16	88468999	C/G	TGCCAACCGTCAGAG[C/G/T]C CAGACCCACGTGGC	2	0.414
47	rs35770269	hsa-miR-449c-3p	chr5	55172296	A/T	AGGAGTGCAACTAGC[A/C/T]A CTGTTAAAATCATT	2	NA
48	rs3745198	hsa-miR-6796-3p	chr19	40369893	C/G	AGCCCCTCAGAAGCT[C/G/T]T CCCCTCCCCGCAGG	1	0.475
49	rs3746444	hsa-miR-499a-3p	chr20	34990448	A/G	CCTCTCCACGTGAAC[A/C/G/T] TCACAGCAAGTCTGT	2	0.116
50	rs3751304	hsa-miR-6763-3p	chr12	132582046	C/T	CAACCTCATGCTCCC[C/A/G/T] GGCCTCTGCCCCCAG	1	0.207
51	rs3823658	hsa-miR-5090	chr7	102465754	G/A	CCTTCTGAGGTACCC[G/A/T]G GGCAGATTGGTGTA	2	0.146
52	rs404337	hsa-miR-8084	chr8	93029770	G/A	TAAATAGAATACTAA[G/A/T]T AAAAAATCAGTATG	2	0.48
53	rs4414449	hsa-miR-548ap-5p	chr15	85825667	G/A	TGCAAAAGTAATTGC[G/A/C/T] GTCTTTGTCATTAAA	2	0.394
54	rs451887	hsa-miR-5692b	chr21	42951004	T/C	CCTACTGTGATATTA[T/A/C/G] TCATAATATCCTAGG	2	0.096
55	rs45596840	hsa-miR-4482-5p	chr10	104268396	G/A	TCCATAGCCCACTGG[G/A/C/T] TTGCTCACTATTCAA	3	0.091
56	rs4919510	hsa-miR-608	chr10	100975021	C/G	GGTGTTGGGACAGCT[C/G/T]C GTTTAAAAAGGCAT	2	0.49
57	rs56148568	hsa-miR-7157-3p	chr2	140586631	T/C	CTGAACTCCCACTCT[T/C]CAT CCAGTAGCACAG	2	0.247
58	rs56155608	hsa-miR-6777-5p	chr17	17813539	G/A	ACTGCCTGACTCCCC[G/A/C]T CTTGAGGACAGGGC	1	0
59	rs56310773	hsa-miR-6744-3p	chr11	1256664	C/T	GGCCTCTCTTGTCAT[C/T]CTG CAGGAACCAGAG	2	0
60	rs56312243	hsa-miR-6805-3p	chr19	55388234	C/T	CTTGCTCTGCTCCCC[C/A/G/T] GCCCCCAGGCAGCCA	2	0
61	rs56790095	hsa-miR-6071	chr2	85783659	C/G	AGCCGGACCCAGAGC[C/G]TT GGCCGGCAGCAGA	2	0
62	rs59323834	hsa-miR-548ab	chr3	103524093	C/T	AAGTAATAGCAAAAT[C/T]CA CAATTACTTTTGC	2	0
63	rs5997893	hsa-miR-3928-5p	chr22	31160117	A/G	GAACCTTAGAGCTTC[A/C/G]G CCATGGAAATTAAC	2	0.439
64	rs61388742	hsa-miR-596	chr8	1817259	T/C	CGAAGCCTGCCCGGC[T/C]CCT CGGGAACCTGCC	1	0
65	rs61938575	hsa-miR-3922-5p	chr12	104591665	G/A	GCCAGAGGTCCCACA[G/A/C/T]CAGGGCTGGAAAGCA	2	0.298
66	rs62154973	hsa-miR-4772-5p	chr2	102432320	C/T	AGGCAAAATTGCAGA[C/T]TG TCTTCCCAAATAG	1	0
67	rs62182086	hsa-miR-6810-5p	chr2	218341922	A/G	GGGCCTGGGATGGGG[A/G]CA GGGATCAGCATGG	2	0.061
68	rs6513497	hsa-miR-646	chr20	60308547	T/G	GGAAGCAGCTGCCTC[T/A/G]G AGGCCTCAGGCTCA	2	0.096
69	rs66507245	hsa-miR-4731-3p	chr17	15251649	T/A	AGTGTTGGGGGCCAC[T/A/C]T GTGTGTGGATGACT	2	0.485

No.	SNP ID	miRNA	Chromosome	Position	Ref./Alt allels	Flanking Sequence(±15bp)	Number of Amplicon	MAF
70	rs66683138	hsa-miR-3622a-5p	chr8	27701697	G/A	AGGGTGCACAGGCAC[G/A]GG AGCTCAGGTGAGG	2	0.394
71	rs6771809	hsa-miR-6826-5p	chr3	129272155	T/C	AATCACCTTGGTCAA[T/C]AGG AAAGAGGTGGGA	2	0.121
72	rs7162033	hsa-miR-8063	chr15	36972836	C/G	AAAAGCAGTAAAGCC[C/G/T]C GACTCCTGATTTTG	2	0.258
73	rs7183051	hsa-miR-8063	chr15	36972838	G/A	AAGCAGTAAAGCCCC[G/A]AC TCCTGATTTTGAA	2	0.258
74	rs7208391	hsa-miR-6868-3p	chr17	76098024	C/G	AGAGCAGTTTCTGCA[C/A/G/T] AGACAACAGAAGGAA	1	0.424
75	rs7210937	hsa-miR-1269b	chr17	12917329	G/C	CAGTAGCATGGCTCA[G/A/C]T CCAGAAACCTCAGA	1	0.465
76	rs7227168	hsa-miR-4741	chr18	22933411	C/T	GCCTTAAAGCGCGGG[C/G/T]T GTCCGGAGGGGTCG	1	0.005
77	rs72810954	hsa-miR-4679	chr10	89063382	G/A	TGTTAGAAACAAAAA[G/A]CA AAGAATCTCTATC	3	0.03
78	rs72955519	hsa-miR-4799-5p	chr04	147782619	G/A	ATCTAAATGCAGCAT[G/A]CC AGTCCTGAGATGC	2	0
79	rs72996752	hsa-miR-4999-5p	chr19	8389352	A/G	CTATCACTACCTGAC[A/G]ATA CAGCAGTATGTG	2	0.273
80	rs7311975	hsa-miR-1178-5p	chr12	119713688	T/C	GAGGGCATGCTCAGC[T/C/G]G ACCCTGGACCCTTC	2	0.025
81	rs73251987	hsa-miR-624-3p	chr14	31014677	C/G	TATCTCAAGGTAATA[C/G/T]C AATACCTTGTGTTT	3	0
82	rs73295187	hsa-miR-4727-5p	chr17	38825855	A/C	TCTGCCAGCTTCCAC[A/C/G]G TGGCAGATTTTCCC	1	0
83	rs73410309	hsa-miR-4739	chr17	79707227	G/C	GACAAGGGCCCCTCC[G/A/C/T] CTCCTCCTCCCTTCT	1	0.091
84	rs73721294	hsa-miR-593-5p	chr7	128081882	C/T	GAATCTGTCAGGCAC[C/T]AG CCAGGCATTGCTC	1	0
85	rs73872515	hsa-miR-548t-3p	chr4	173268209	A/C	TTTTTAATGACAAAA[A/C]CCA CAATTACTTTTG	2	0
86	rs74085143	hsa-miR-4781-3p	chr1	54054127	G/A	GCCAATTCCCCCAAT[G/A]TTG GAATCCTCGCTA	2	0.035
87	rs74469188	hsa-miR-6504-5p	chr16	81611365	T/C	AGTCTGGCTGTGCTG[T/C]AAT GCAGTCTGCACC	2	0.015
88	rs74647838	hsa-miR-1302	chr12	112695096	G/A	TTTTAGCATAAGTAT[G/A/T]TC CCAAATATTGTAT	2	0
89	rs74743733	hsa-miR-4257	chr1	150551992	G/A	CCTACAGGGCCCAGA[G/A]GT GGGGACTGAGCCT	1	0
90	rs74814065	hsa-miR-6879-3p	chr11	65018557	C/T	GACCCTCCTGTCACC[C/T]GCT CCTTGCCCAGAT	2	0
91	rs74904371	hsa-miR-2682-3p	chr1	98045291	C/T	ACAGCGCTGAAGAGG[C/T]GT CCCAAAGAGATTG	2	0
92	rs75330474	hsa-miR-323b-5p	chr14	101056252	C/T	TGTCCGTGGTGAGTT[C/T]GCA TTATTTAATGAT	1	0.101
93	rs76347846	hsa-miR-6841-3p	chr8	24953808	A/G	TTCCTGGGGATGCAG[A/G]TG CAAGGTGAGGTTA	2	0.005
94	rs76595065	hsa-miR-4704-3p	chr13	66218307	T/C	TCAATCAGTCACATA[T/C]CTA GTGTCTAGAATG	2	0.035
95	rs7804972	hsa-miR-6839-5p	chr7	64679085	G/A	CTGGATTGAAGAGAC[G/A/T]A CCCAAGCAGGCTTT	2	0.141
96	rs78293125	hsa-miR-6885-5p	chr19	6389688	A/G	TGGCTTTGCTTGCGC[A/G]GTG CCCCCCTCCAGG	1	0
97	rs78825966	hsa-miR-557	chr1	168375591	C/T	GGAGAAGTGTTTGCA[C/T]GG GTGGGCCTTGTCT	2	0
98	rs79637190	hsa-miR-4695-5p	chr1	18883265	C/T	CCTGCTCGCCCACTG[C/A/T]CT CCTGCAGGTACCT	2	0
99	rs8078913	hsa-miR-4520-3p	chr17	6655449	C/G	GGTTGATTCCTTCTT[C/A/G/T] CTGCGTGTTTTCTGT	1	0.227
100	rs9745376	hsa-miR-662	chr16	770249	G/A	CTGAAGGTCTCCCAC[G/A/C/T] TTGTGGCCCAGCAGC	2	0
101	rs9913045	hsa-miR-548h-5p	chr17	13543607	G/A	TCAATGACAAAAACC[G/A]CG ATTACTTTTGCAC	2	0.278

NA: not available

Table S2 Summary of sequencing performance metrics for control and sample libraries.

Sample	Mapped reads	On-target	Mean depth	Uniformity
Control DNA	105750	97.24	670.1	88.77
Control DNA	185337	97.26	1167.1	89.47
Control DNA	219237	97.29	1395	88.65
1	149019	90.13	866.1	83.14
2	112137	95.8	706.8	87.9
3	110927	95.37	695.1	87.6
4	94201	91.63	565.9	90.29
5	97674	92.77	592.9	90.47
6	112453	94.45	694.3	89.02
7	111239	96.93	710.5	89.76
8	98939	94.45	613.1	89.75
9	123669	95.83	778	89.97
10	111211	92.06	669.8	89.63
11	77637	96.82	488.9	87.32
12	42809	92.6	2.605	88.69
13	89348	95.41	562.7	83.95
14	85248	94.31	526	86.85
15	60086	94.42	374.2	87.69
16	108815	93.89	671.1	90.05
17	91638	93.83	563.4	89.71
18	74965	90.68	447.2	90.29
19	137622	96.04	870.4	90.47
20	57976	84.66	322.1	90.06
21	56670	82.68	300.1	88.65
22	17950	93.62	109.7	87.72
23	33637	95.85	211	84.87
24	107922	98.11	701.5	83.34
25	79904	96.63	505.5	87.81
26	23335	94.4	114.3	88.76
27	86642	93.15	528.6	89.46
28	13039	93.69	79.53	89.17
29	22842	93.43	139	89.81
30	80321	96.94	437.4	89.4
31	68925	96.94	437.4	89.4
32	101227	97.29	645.7	89.44
33	83583	97.32	536	89.06
34	88715	97.27	567.6	90.29
35	84971	95.79	536	89

Sample	Mapped reads	On-target	Mean depth	Uniformity
36	101503	85.17	557.8	88.56
37	93268	81.51	500.5	88.73
38	108145	91.75	646.6	89.24
39	123092	88.88	719.3	89.68
40	78231	79.56	409.7	89.68
41	115477	87.85	666.1	90.2
42	68105	79.56	350.4	88.1
43	102.036	74.85	498	90.18
44	79364	90.34	469.4	88.92
45	122565	88.65	710.2	90.09
46	84506	83.51	462.9	89.4
47	104651	85.42	589.3	88.64
48	105639	83.21	577.7	89.64
49	70195	88.91	405.8	89.35
50	107564	87.76	621.6	89.5

Table S3 Results of miR-SNP MPS for the female control DNA 5190-3797.

miRNA	SNP ID	Genotyping	Coverage depth	A	G	C	T	-	F_{MAR} (%)
hsa-miR-449b-5p	rs10061133	AG	1364	669	695	0	0	4	50.95
hsa-miR-6801-3p	rs10412196	TC	1148	1	0	578	569	1	50.35
hsa-miR-548ae-5p	rs10461441	AG	440	197	243	0	0	0	55.23
hsa-miR-4700-3p	rs1055070	TT	1546	0	0	0	1546	0	100.00
hsa-miR-4302	rs11048315	GG	1553	2	1551	0	0	0	99.87
hsa-miR-5579-3p	**rs11237828**	**NN**	108	9	3	0	96	0	88.89
hsa-miR-501-3p	rs112489955	GG	1356	2	1353	1	0	3	99.78
hsa-miR-5691	rs112511786	GG	745	1	744	0	0	0	99.87
hsa-miR-4786-5p	rs115063401	GG	1323	1	1322	0	0	0	99.92
hsa-miR-580-5p	rs115089112	TT	1497	0	0	1	1496	0	99.93
hsa-miR-4467	rs115101071	GG	1296	1	1295	0	0	1	99.92
hsa-miR-3124-3p	rs115160731	CC	1298	0	0	1297	1	0	99.92
hsa-miR-942-3p	rs115372145	CC	540	0	0	540	0	1	100.00
hsa-miR-4457	rs115769169	CC	1381	0	0	1381	0	0	100.00
hsa-miR-3130-3p	rs115772313	GG	137	0	137	0	0	2	100.00
hsa-miR-4514	rs116034786	AA	1667	1666	0	0	1	0	99.94
hsa-miR-196a-3p	rs11614913	TT	975	1	0	2	972	0	99.69
hsa-miR-548at-5p	rs11651671	GG	469	1	468	0	0	0	99.79
hsa-miR-5589-3p	rs116796353	AA	1367	1366	1	0	0	0	99.93
hsa-miR-6717-5p	rs117650137	GG	1522	1	1521	0	0	0	99.93
hsa-miR-3192-3p	rs11907020	TT	1351	0	0	2	1349	0	99.85
hsa-miR-4284	rs11973069	CC	1305	0	0	1304	1	0	99.92
hsa-miR-5700	rs12314280	TC	377	0	0	166	211	0	55.97

miRNA	SNP ID	Genotyping	Coverage depth	A	G	C	T	-	F_{MAR} (%)
hsa-miR-4661-3p	rs12335005	GG	648	0	648	0	0	0	100.00
hsa-miR-3117-3p	rs12402181	AA	1580	1575	4	0	1	0	99.68
hsa-miR-4433a-5p	**rs12473206**	**NN**	0	0	0	0	0	0	0.00
hsa-miR-6077	rs1280409926	CC	436	0	0	436	0	0	100.00
hsa-miR-548l	rs13447640	GG	764	1	762	0	1	0	99.74
hsa-miR-548ar-3p	rs141659366	GG	1219	0	1219	0	0	0	100.00
hsa-miR-4444	rs142357696	AA	974	974	0	0	0	0	100.00
hsa-miR-888-5p	rs143634721	CC	731	0	0	731	0	0	100.00
hsa-miR-892b	rs146806052	AA	778	777	1	0	0	0	99.87
hsa-miR-501-3p	rs149912461	AA	1357	1354	3	0	0	0	99.78
hsa-miR-8060	rs1514422	GG	824	1	823	0	0	0	99.88
hsa-miR-6887-5p	rs1688017	GG	1440	0	1440	0	0	0	100.00
hsa-miR-1304-3p	**rs2155248**	**NN**	1	0	0	0	1	0	100.00
hsa-miR-4513	rs2168518	GA	655	348	307	0	0	0	53.13
hsa-miR-3130-5p	rs2241347	CC	242	0	0	242	0	0	100.00
hsa-miR-4707-3p	rs2273626	CA	537	280	2	255	0	13	52.14
hsa-miR-627-5p	rs2620381	AA	648	648	0	0	0	0	100.00
hsa-miR-4804-5p	**rs266435**	**NN**	0	0	0	0	0	0	0.00
has-miR-146a-3p	rs2910164	CG	684	0	357	326	1	0	52.19
hsa-miR-7854-3p	rs2925980	AG	1231	605	626	0	0	3	50.85
hsa-miR-1343-5p	rs2986407	CC	623	0	3	620	0	1	99.52
hsa-miR-5087	rs2992458	AA	627	626	1	0	0	0	99.84
hsa-miR-5189-3p	rs35613341	CC	669	0	0	669	0	0	100.00
hsa-miR-449c-3p	rs35770269	AA	1195	1195	0	0	0	2	100.00
hsa-miR-6796-3p	rs3745198	CG	795	0	375	420	0	0	52.83
hsa-miR-499a-3p	rs3746444	GG	664	0	663	0	1	0	99.85
hsa-miR-6763-3p	rs3751304	CT	604	0	3	268	333	0	55.13
hsa-miR-5090	rs3823658	GA	282	138	142	2	0	7	50.35
hsa-miR-8084	rs404337	AA	2001	1991	5	2	3	5	99.50
hsa-miR-548ap-5p	rs4414449	GA	1346	605	741	0	0	0	55.05
hsa-miR-5692b	rs451887	CC	938	0	1	936	1	0	99.79
hsa-miR-4482-5p	rs45596840	GA	682	337	344	0	1	0	50.44
hsa-miR-608	rs4919510	CC	220	0	0	220	0	0	100.00
hsa-miR-7157-3p	rs56148568	TC	1318	0	0	638	680	0	51.59
hsa-miR-6777-5p	rs56155608	GG	654	2	652	0	0	0	99.69
hsa-miR-6744-3p	rs56310773	CC	1437	0	0	1437	0	0	100.00
hsa-miR-6805-3p	rs56312243	CC	799	0	0	799	0	8	100.00
hsa-miR-6071	rs56790095	CC	931	0	0	931	0	0	100.00
hsa-miR-548ab	rs59323834	CC	1222	0	0	1221	1	0	99.92
hsa-miR-3928-5p	rs5997893	GG	410	0	410	0	0	0	100.00
hsa-miR-596	rs61388742	TC	521	0	0	260	261	3	50.10
hsa-miR-3922-5p	rs61938575	GA	536	285	250	0	1	1	53.17
hsa-miR-4772-5p	rs62154973	CC	372	0	0	372	0	0	100.00

miRNA	SNP ID	Genotyping	Coverage depth	A	G	C	T	-	F_{MAR} (%)
hsa-miR-6810-5p	rs62182086	AA	1159	1153	3	3	0	0	99.48
hsa-miR-646	rs6513497	TT	1380	0	0	0	1380	1	100.00
hsa-miR-4731-3p	rs66507245	TA	317	130	6	3	178	2	56.15
hsa-miR-3622a-5p	rs66683138	GA	524	279	245	0	0	1	53.24
hsa-miR-6826-5p	rs6771809	TT	1517	0	0	1	1516	0	99.93
hsa-miR-8063	rs7162033	CC	1306	0	2	1304	0	0	99.85
hsa-miR-8063	rs7183051	GG	1294	0	1291	3	0	1	99.77
hsa-miR-6868-3p	rs7208391	CC	710	0	0	710	0	0	100.00
hsa-miR-1269b	rs7210937	GG	595	0	595	0	0	0	100.00
hsa-miR-4741	rs7227168	CC	458	0	1	456	1	0	99.56
hsa-miR-4679	**rs72810954**	**NN**	21	1	20	0	0	0	95.24
hsa-miR-4799-5p	rs72955519	GG	521	1	520	0	0	0	99.81
hsa-miR-4999-5p	rs72996752	AG	410	151	259	0	0	1	63.17
hsa-miR-1178-5p	rs7311975	TT	631	0	0	0	631	0	100.00
hsa-miR-624-3p	rs73251987	CC	696	0	0	696	0	0	100.00
hsa-miR-4727-5p	rs73295187	AA	818	817	1	0	0	0	99.88
hsa-miR-4739	rs73410309	GG	672	0	672	0	0	0	100.00
hsa-miR-593-5p	rs73721294	CC	796	0	0	796	0	0	100.00
hsa-miR-548t-3p	rs73872515	AA	349	345	2	0	2	0	98.85
hsa-miR-4781-3p	rs74085143	GG	514	0	514	0	0	0	100.00
hsa-miR-6504-5p	rs74469188	TT	541	0	0	0	541	0	100.00
hsa-miR-1302	rs74647838	GG	1614	0	1614	0	0	0	100.00
hsa-miR-4257	rs74743733	GG	840	0	839	0	1	0	99.88
hsa-miR-6879-3p	rs74814065	CC	695	0	0	695	0	0	100.00
hsa-miR-2682-3p	rs74904371	CC	1403	0	0	1402	1	0	99.93
hsa-miR-323b-5p	rs75330474	CC	815	1	0	814	0	0	99.88
hsa-miR-6841-3p	rs76347846	AG	1461	753	708	0	0	0	51.54
hsa-miR-4704-3p	rs76595065	TT	1489	1	0	2	1486	0	99.80
hsa-miR-6839-5p	rs7804972	AA	329	329	0	0	0	0	100.00
hsa-miR-6885-5p	rs78293125	AA	676	676	0	0	0	0	100.00
hsa-miR-557	rs78825966	CC	1439	0	0	1438	1	1	99.93
hsa-miR-4695-5p	rs79637190	CC	1103	0	0	1103	0	2	100.00
hsa-miR-4520-3p	rs8078913	CT	361	0	0	181	180	1	50.14
hsa-miR-662	rs9745376	GG	1006	1	1004	1	0	0	99.80
hsa-miR-548h-5p	rs9913045	AA	847	847	0	0	0	1	100.00

Table S4 List of primer sequences used for Sanger sequencing validation of selected miR-SNP loci.

SNP	miRNA	Forward primer	Reverse primer	Size
rs11237828	hsa-miR-5579-3p	AGAGCTGCCAATCATCACCC	CCAAGACCACACAGCAGGAA	502
rs62154973	hsa-miR-4772-5p	GGATTCCAGGTGGGCTAGTT	CCAATCAAGCCTCAATGCCC	542
rs66507245	hsa-miR-4731-3p	AGTTCCCCATCAGTGGTAGGA	GAGATGCTGAGACCCTGGTG	509

SNP	miRNA	Forward primer	Reverse primer	Size
rs73872515	hsa-miR-548t-3p	ACATGAGCTCCACCGGTTG	GCTTTGTTTGGGTGCGAGTT	508
rs3823658	hsa-miR-5090	CGCTATCGTTAGAGGGCCTG	GCGTCTCCAGGTGGTTGTTA	535
rs8078913	hsa-miR-4520-3p	GCAATCCAACTTGTGAGGCG	AGATATGACGGGAGGTGGGT	503
rs2273626	hsa-miR-4707-3p	CAGGGCCCGAGATGAATCAA	AAGCCTTTCCGAGACCATCC	536
rs1688017	hsa-miR-6887-5p	CAGATCGTGTTTGCTGGTGTG	TGAAAAGCGTCCAGTCAGCC	422
rs404337	hsa-miR-8084	CTCTGATCCCAGCAACCGTT	TGTGACCCAGAGACTTCCCA	597
rs78293125	hsa-miR-6885-5p	CTGCCTGATTTGCCGTTGAC	GGAACCAGATGCCTGCTGTA	548

Table S5 Reproducibility of miR-SNP genotyping in three replicates of the female control DNA sample.

SNP	1st						2nd						3rd					
	Genotype	coverage	A	G	C	T	Genotype	coverage	A	G	C	T	Genotype	coverage	A	G	C	T
rs10061133	AG	1364	669	695	0	0	AG	1759	897	862	0	0	AG	2809	1459	1350	0	0
rs10412196	TC	1148	1	0	578	569	TC	2078	0	0	1045	1033	TC	2475	1	0	1283	1191
rs10461441	AG	440	197	243	0	0	AG	1116	585	531	0	0	AG	1125	572	550	2	1
rs1055070	TT	1546	0	0	0	1546	TT	3113	0	0	5	3108	TT	3450	0	0	1	3449
rs11048315	GG	1553	2	1551	0	0	GG	2646	3	2643	0	0	GG	3153	1	3152	0	0
rs11237828	**NN**	108	9	3	0	96	**NN**	62	1	0	0	61	TT	154	5	1	0	148
rs112489955	GG	1356	2	1353	1	0	GG	3093	5	3088	0	0	GG	2925	0	2925	0	0
rs112511786	GG	745	1	744	0	0	GG	1350	1	1348	0	1	GG	1539	1	1538	0	0
rs115063401	GG	1323	1	1322	0	0	GG	2127	0	2126	1	0	GG	2598	3	2595	0	0
rs115089112	TT	1497	0	0	1	1496	TT	3246	0	0	1	3245	TT	3121	0	0	0	3121
rs115101071	GG	1296	1	1295	0	0	GG	1601	5	1596	0	0	GG	2580	1	2579	0	0
rs115160731	CC	1298	0	0	1297	1	CC	2057	0	0	2055	2	CC	2670	0	0	2669	1
rs115372145	CC	540	0	0	540	0	CC	572	0	0	572	0	CC	1467	0	1	1464	2
rs115769169	CC	1381	0	0	1381	0	CC	1518	0	0	1518	0	CC	2883	0	0	2883	0
rs115772313	GG	137	0	137	0	0	GG	352	1	351	0	0	GG	151	0	151	0	0
rs116034786	AA	1667	1666	0	0	1	AA	4227	4224	1	1	1	AA	3060	3059	0	1	0
rs11614913	TT	975	1	0	2	972	TT	1083	0	0	3	1080	TT	1890	0	0	2	1888
rs11651671	GG	469	1	468	0	0	GG	386	0	386	0	0	GG	652	1	651	0	0
rs116796353	AA	1367	1366	1	0	0	AA	1994	1992	2	0	0	AA	2645	2643	1	1	0
rs117650137	GG	1522	1	1521	0	0	GG	3314	0	3314	0	0	GG	3293	0	3292	0	1
rs11907020	TT	1351	0	0	2	1349	TT	2987	0	0	4	2983	TT	3015	0	0	9	3006
rs11973069	CC	1305	0	0	1304	1	CC	1806	0	0	1806	0	CC	2498	0	0	2497	1
rs12314280	TC	377	0	0	166	211	TC	510	1	0	225	284	TC	729	0	0	344	385
rs12335005	GG	648	0	648	0	0	GG	1603	0	1603	0	0	GG	1284	1	1283	0	0
rs12402181	AA	1580	1575	4	0	1	AA	2924	2919	5	0	0	AA	3430	3416	12	2	0
rs12473206	**NN**	0	0	0	0	0	**NN**	1	0	1	0	0	**NN**	1	0	1	0	0
rs1280409926	CC	436	0	0	436	0	CC	1207	0	6	1200	1	CC	983	0	0	983	0

SNP	1st						2nd						3rd					
	Genotype	coverage	A	G	C	T	Genotype	coverage	A	G	C	T	Genotype	coverage	A	G	C	T
rs13447640	GG	764	1	762	0	1	GG	2330	2	2328	0	0	GG	1821	1	1820	0	0
rs141659366	GG	1219	0	1219	0	0	GG	446	0	446	0	0	GG	2489	5	2484	0	0
rs142357696	AA	974	974	0	0	0	AA	1199	1195	4	0	0	AA	1857	1856	1	0	0
rs143634721	CC	731	0	0	731	0	CC	2003	0	0	2001	2	CC	1472	0	0	1472	0
rs146806052	AA	778	777	1	0	0	AA	1870	1868	1	0	1	AA	1791	1788	3	0	0
rs149912461	AA	1357	1354	3	0	0	AA	3119	3114	3	1	1	AA	2927	2924	3	0	0
rs1514422	GG	824	1	823	0	0	GG	1450	2	1447	0	1	GG	1557	0	1556	0	1
rs1688017	GG	1440	0	1440	0	0	GG	2312	1	2311	0	0	GG	2821	0	2821	0	0
rs2155248	**NN**	1	0	0	0	1	**NN**	3	0	0	0	3	**NN**	1	0	0	0	1
rs2168518	GA	655	348	307	0	0	GA	481	296	185	0	0	GA	1159	575	583	1	0
rs2241347	CC	242	0	0	242	0	CC	368	0	0	368	0	CC	252	0	0	251	1
rs2273626	CA	537	280	2	255	0	CA	716	401	1	313	1	CA	1027	527	2	498	0
rs2620381	AA	648	648	0	0	0	AA	1115	1114	1	0	0	AA	1397	1396	1	0	0
rs266435	**NN**	0	0	0	0	0	**NN**	2	0	0	2	0	**NN**	2	0	0	2	0
rs2910164	CG	684	0	357	326	1	CG	2035	28	999	1007	1	CG	1336	2	676	658	0
rs2925980	AG	1231	605	626	0	0	AG	2169	1170	999	0	0	AG	2682	1368	1314	0	0
rs2986407	CC	623	0	3	620	0	CC	1214	0	1	1211	2	CC	1319	1	4	1310	4
rs2992458	AA	627	626	1	0	0	AA	1207	1206	1	0	0	AA	1199	1198	0	1	0
rs35613341	CC	669	0	0	669	0	CC	1031	0	0	1031	0	CC	1357	0	0	1356	1
rs35770269	AA	1195	1195	0	0	0	AA	2392	2391	0	1	0	AA	2458	2455	2	1	0
rs3745198	CG	795	0	375	420	0	CG	719	0	268	451	0	CG	1733	1	807	924	1
rs3746444	GG	664	0	663	0	1	GG	852	1	850	0	1	GG	1331	0	1328	1	2
rs3751304	CT	604	0	3	268	333	CT	191	0	0	87	104	CT	1193	0	0	569	624
rs3823658	GA	282	138	142	2	0	GA	246	123	123	0	0	GA	603	289	314	0	0
rs404337	AA	2001	1991	5	2	3	AA	3760	3750	2	0	8	AA	3898	3876	11	3	8
rs4414449	GA	1346	605	741	0	0	GA	1684	719	964	0	1	GA	2537	1278	1259	0	0
rs451887	CC	938	0	1	936	1	CC	2554	0	0	2548	6	CC	2035	0	0	2032	3
rs45596840	GA	682	337	344	0	1	GA	1072	484	587	0	1	GA	1449	737	711	0	1
rs4919510	CC	220	0	0	220	0	CC	188	0	0	188	0	CC	522	0	0	522	0
rs56148568	TC	1318	0	0	638	680	TC	3027	0	0	1533	1494	TC	2715	0	0	1366	1349
rs56155608	GG	654	2	652	0	0	GG	646	0	646	0	0	GG	1480	0	1479	1	0
rs56310773	CC	1437	0	0	1437	0	CC	2953	1	0	2951	1	CC	2876	0	0	2873	3
rs56312243	CC	799	0	0	799	0	CC	1494	0	4	1489	1	CC	1746	0	3	1743	0
rs56790095	CC	931	0	0	931	0	CC	993	0	0	993	0	CC	1919	0	0	1918	1
rs59323834	CC	1222	0	0	1221	1	CC	1928	0	0	1925	3	CC	2955	0	0	2951	4
rs5997893	GG	410	0	410	0	0	GG	230	0	230	0	0	GG	759	0	759	0	0
rs61388742	TC	521	0	0	260	261	TC	1114	0	0	586	528	TC	940	0	0	469	471
rs61938575	GA	536	285	250	0	1	GA	1505	735	769	0	1	GA	1219	634	585	0	0

SNP	1st						2nd						3rd					
	Genotype	coverage	A	G	C	T	Genotype	coverage	A	G	C	T	Genotype	coverage	A	G	C	T
rs62154973	CC	372	0	0	372	0	CC	976	1	0	975	0	CC	715	0	0	715	0
rs62182086	AA	1159	1153	3	3	0	AA	1646	1644	2	0	0	AA	2301	2295	4	1	1
rs6513497	TT	1380	0	0	0	1380	TT	2596	0	0	2	2594	TT	2763	0	1	2	2760
rs66507245	TA	317	130	6	3	178	TA	742	364	0	0	378	TA	663	296	10	6	351
rs66683138	GA	524	279	245	0	0	GA	669	430	239	0	0	GA	1093	615	478	0	0
rs6771809	TT	1517	0	0	1	1516	TT	2436	0	0	0	2436	TT	3269	0	0	1	3268
rs7162033	CC	1306	0	2	1304	0	CC	1723	0	2	1721	0	CC	2737	1	2	2734	0
rs7183051	GG	1294	0	1291	3	0	GG	1680	0	1675	5	0	GG	2715	0	2712	3	0
rs7208391	CC	710	0	0	710	0	CC	1390	0	0	1389	1	CC	1407	0	0	1406	1
rs7210937	GG	595	0	595	0	0	GG	1116	0	1116	0	0	GG	1192	0	1192	0	0
rs7227168	CC	458	0	1	456	1	CC	820	0	0	820	0	CC	1061	0	1	1060	0
rs72810954	**NN**	21	1	20	0	0	**NN**	38	1	37	0	0	**NN**	35	0	35	0	0
rs72955519	GG	521	1	520	0	0	GG	899	2	891	0	6	GG	1122	0	1122	0	0
rs72996752	AG	410	151	259	0	0	AG	178	47	131	0	0	AG	929	335	594	0	0
rs7311975	TT	631	0	0	0	631	TT	963	1	0	1	961	TT	1095	0	0	0	1095
rs73251987	CC	696	0	0	696	0	CC	2004	0	0	2004	0	CC	1331	0	0	1331	0
rs73295187	AA	818	817	1	0	0	AA	1359	1352	2	5	0	AA	1649	1649	0	0	0
rs73410309	GG	672	0	672	0	0	GG	987	1	985	1	0	GG	1392	0	1392	0	0
rs73721294	CC	796	0	0	796	0	CC	815	0	0	814	1	CC	1520	0	0	1520	0
rs73872515	AA	349	345	2	0	2	AA	904	898	2	1	3	AA	668	659	1	1	7
rs74085143	GG	514	0	514	0	0	GG	899	0	899	0	0	GG	1100	0	1100	0	0
rs74469188	TT	541	0	0	0	541	TT	942	0	0	0	942	TT	1187	0	0	0	1187
rs74647838	GG	1614	0	1614	0	0	GG	2944	1	2936	2	5	GG	3653	3	3650	0	0
rs74743733	GG	840	0	839	0	1	GG	724	0	724	0	0	GG	1883	0	1883	0	0
rs74469188	CC	695	0	0	695	0	CC	1004	0	0	1004	0	CC	1562	0	0	1561	1
rs74904371	CC	1403	0	0	1402	1	CC	3101	0	1	3100	0	CC	3164	0	0	3161	3
rs75330474	CC	815	1	0	814	0	CC	1477	0	4	1472	1	CC	1495	0	0	1495	0
rs76347846	AG	1461	753	708	0	0	AG	3380	1649	1731	0	0	AG	3029	1598	1431	0	0
rs76595065	TT	1489	1	0	2	1486	TT	3369	0	1	2	3366	TT	3103	0	0	0	3103
rs7804972	AA	329	329	0	0	0	AA	504	504	0	0	0	AA	796	796	0	0	0
rs78293125	AA	676	676	0	0	0	AA	828	828	0	0	0	AA	1381	1380	1	0	0
rs78825966	CC	1439	0	0	1438	1	CC	2159	1	0	2156	2	CC	2846	0	1	2842	3
rs79637190	CC	1103	0	0	1103	0	CC	2615	0	0	2614	1	CC	2435	0	0	2433	2
rs8078913	CT	361	0	0	181	180	CT	207	0	0	62	145	CT	680	0	1	310	369
rs9745376	GG	1006	1	1004	1	0	GG	1622	1	1621	0	0	GG	2317	1	2316	0	0
rs9913045	AA	847	847	0	0	0	AA	1463	1461	0	2	0	AA	1899	1898	1	0	0
mean		864.1						1491						1770				
SD		464.7						988.1						970				

Table S6 Reproducibility of miR-SNP genotyping in three replicates of the blood-extracted DNA sample.

SNP	1st			2nd			3rd		
	Coverage	**Genotype**	**Calling rate**	**Coverage**	**Genotype**	**Calling rate**	**Coverage**	**Genotype**	**Calling rate**
rs10061133	2882	AA	97.03	2850	AA	96.04	1354	AA	96.04
rs10412196	1943	TC		2440	TC		1057	TC	
rs10461441	788	AA		1063	AA		474	AA	
rs1055070	3235	TT		3220	TT		1549	TT	
rs11048315	2136	GG		3001	GG		1323	GG	
rs11237828	1384	CC		1104	CC		465	CC	
rs112489955	2351	GG		3064	GG		1291	GG	
rs112511786	1227	GG		1662	GG		844	GG	
rs115063401	1368	GG		2596	GG		1029	GG	
rs115089112	4206	TT		3248	TT		1830	TT	
rs115101071	2842	GG		3003	GG		1516	GG	
rs115160731	2704	CC		3094	CC		1505	CC	
rs115372145	1094	CC		1380	CC		658	CC	
rs115769169	2442	CC		2756	CC		1363	CC	
rs115772313	955	GG		1270	GG		520	GG	
rs116034786	3126	AA		3553	AA		1821	AA	
rs11614913	1213	CT		1584	CT		951	CT	
rs11651671	208	GG		589	GG		283	GG	
rs116796353	1810	AA		2526	AA		1235	AA	
rs117650137	4589	GG		3951	GG		1660	GG	
rs11907020	2633	TT		2919	TT		1331	TT	
rs11973069	1960	CC		2618	CC		1191	CC	
rs12314280	331	CC		532	CC		218	CC	
rs12335005	2137	GG		1376	GG		824	GG	
rs12402181	2179	GG		2970	GG		1522	GG	
rs12473206	0	**NN**		0	**NN**		0	**NN**	
rs1280409926	1783	CC		2165	CC		1077	CC	
rs13447640	1218	GG		1651	GG		813	GG	
rs141659366	3775	GG		2557	GG		1140	GG	
rs142357696	2460	AA		1956	AA		1008	AA	
rs143634721	760	CC		1311	CC		762	CC	
rs146806052	2462	AA		2095	AA		943	AA	
rs149912461	2341	AA		3070	AA		1293	AA	
rs1514422	1352	GG		1517	GG		947	GG	
rs1688017	3538	GG		3250	GG		1620	GG	
rs2155248	4	**NN**		9	**NN**		5	**NN**	
rs2168518	1179	GG		1207	GG		601	GG	

SNP	1st			2nd			3rd		
	Coverage	**Genotype**	**Calling rate**	**Coverage**	**Genotype**	**Calling rate**	**Coverage**	**Genotype**	**Calling rate**
rs2241347	986	CT		1309	CT		545	CT	
rs2273626	626	CA		925	CA		469	CA	
rs2620381	699	AA		1222	AA		532	AA	
rs266435	2	**NN**		3	**NN**		2	**NN**	
rs2910164	2372	CC		1550	CC		781	CC	
rs2925980	2789	AG		2867	AG		1262	AG	
rs2986407	692	CC		1224	CC		491	CC	
rs2992458	676	AA		1116	AA		553	AA	
rs35613341	597	CG		1433	CG		623	CG	
rs35770269	1588	AT		2299	AT		944	AT	
rs3745198	2068	CC		2109	CC		998	CC	
rs3746444	1253	AA		1350	AA		635	AA	
rs3751304	289	TT		557	TT		331	TT	
rs3823658	221	GG		530	GG		225	GG	
rs404337	4860	AA		4016	AA		2311	AA	
rs4414449	2755	AA		2659	AA		1428	AA	
rs451887	3332	CC		2238	CC		1123	CC	
rs45596840	788	GG		1485	GG		758	GG	
rs4919510	369	CC		658	CC		199	CC	
rs56148568	3677	TC		3336	TC		1561	TC	
rs56155608	981	GG		1449	GG		653	GG	
rs56310773	3672	CC		3370	CC		1538	CC	
rs56312243	1232	CC		1635	CC		771	CC	
rs56790095	1343	CC		2148	CC		972	CC	
rs59323834	2520	CC		2404	CC		1108	CC	
rs5997893	755	AA		835	AA		389	AA	
rs61388742	1178	TT		1374	TT		748	TT	
rs61938575	1759	GG		1345	GG		454	GG	
rs62154973	1664	CC		987	CC		507	CC	
rs62182086	1718	AA		2318	AA		921	AA	
rs6513497	1952	TT		2941	TT		1271	TT	
rs66507245	487	TT		719	TT		352	TT	
rs66683138	788	GA		1078	GA		386	GA	
rs6771809	3193	TT		3384	TT		1508	TT	
rs7162033	2538	CC		2752	CC		1538	CC	
rs7183051	2517	GG		2721	GG		1526	GG	
rs7208391	698	CG		1372	CG		600	CG	
rs7210937	1570	GG		1384	GG		635	GG	

SNP	1st			2nd			3rd		
	Coverage	**Genotype**	**Calling rate**	**Coverage**	**Genotype**	**Calling rate**	**Coverage**	**Genotype**	**Calling rate**
rs7227168	849	CC		1075	CC		456	CC	
rs72810954	102	GA		52	**NN**		29	**NN**	
rs72955519	605	GG		794	GG		358	GG	
rs72996752	409	AA		786	AA		360	AA	
rs7311975	811	TT		1027	TT		506	TT	
rs73251987	2180	CC		1477	CC		853	CC	
rs73295187	1293	AA		1444	AA		921	AA	
rs73410309	1906	GG		1420	GG		712	GG	
rs73721294	1621	CC		1701	CC		843	CC	
rs73872515	719	AA		1029	AA		395	AA	
rs74085143	1136	GG		1571	GG		677	GG	
rs74469188	1161	TT		1262	TT		513	TT	
rs74647838	2492	GG		3065	GG		1740	GG	
rs74743733	3707	GG		2385	GG		1163	GG	
rs74814065	2216	CC		1843	CC		905	CC	
rs74904371	3802	CC		3347	CC		1745	CC	
rs75330474	732	CC		1334	CC		603	CC	
rs76347846	2828	AA		3700	AA		1572	AA	
rs76595065	2449	TT		3132	TT		1398	TT	
rs7804972	629	GA		807	GA		336	GA	
rs78293125	1253	AA		1341	AA		516	AA	
rs78825966	3924	CC		3484	CC		1848	CC	
rs79637190	1552	CC		2458	CC		1143	CC	
rs8078913	907	TT		1150	TT		519	TT	
rs9745376	1515	GG		2250	GG		995	GG	
rs9913045	781	GG		891	GG		381	GG	
Mean	1737			1892			902.6		
SD	1122			984.1			493.6		

Table S7 Reproducibility of miR-SNP genotyping in two independent runs of the blood-extracted DNA sample.

SNP	Run 1			Run 2		
	Coverage	**Genotype**	**Calling rate**	**Coverage**	**Genotype**	**Calling rate**
rs10061133	265	AG	0.95	226	AG	0.95
rs10412196	255	TT		243	TT	
rs10461441	202	AA		104	AA	
rs1055070	272	TT		252	TT	
rs11048315	269	GG		253	GG	
rs11237828	7	**NN**		6	**NN**	
rs112489955	251	GG		231	GG	

SNP	Run 1			Run 2		
	Coverage	Genotype	Calling rate	Coverage	Genotype	Calling rate
rs112511786	263	GG		115	GG	
rs115063401	256	GG		213	GG	
rs115089112	281	TT		252	TT	
rs115101071	266	GG		222	GG	
rs115160731	253	CC		236	CC	
rs115372145	177	CC		96	CC	
rs115769169	263	CC		220	CC	
rs115772313	200	GG		88	GG	
rs116034786	254	AA		249	AA	
rs11614913	257	CT		170	CT	
rs11651671	229	GG		67	GG	
rs116796353	253	AA		223	AA	
rs117650137	268	GG		224	GG	
rs11907020	259	TT		236	TT	
rs11973069	255	CC		202	CC	
rs12314280	187	TC		58	TC	
rs12335005	259	GG		126	GG	
rs12402181	262	GG		253	GG	
rs12473206	0	**NN**		0	**NN**	
rs1280409926	265	CC		143	CC	
rs13447640	218	GG		132	GG	
rs141659366	261	GG		232	GG	
rs142357696	269	AA		128	AA	
rs143634721	264	CC		124	CC	
rs146806052	260	AA		142	AA	
rs149912461	264	AA		237	AA	
rs1514422	259	GG		136	GG	
rs1688017	254	GG		240	GG	
rs2155248	0	**NN**		1	**NN**	
rs2168518	264	GG		151	GG	
rs2241347	224	TT		108	TT	
rs2273626	261	AA		95	AA	
rs2620381	222	AA		81	AA	
rs266435	0	**NN**		0	**NN**	
rs2910164	267	CG		169	CG	
rs2925980	253	AG		196	AG	
rs2986407	203	CC		100	CC	
rs2992458	265	AA		110	AA	
rs35613341	260	CG		145	CG	
rs35770269	262	AA		172	AA	
rs3745198	263	CG		118	CG	
rs3746444	253	AA		115	AA	
rs3751304	261	TT		111	TT	
rs3823658	138	GA		59	GA	
rs404337	274	GA		252	GA	
rs4414449	257	GA		211	GA	

SNP	Run 1			Run 2		
	Coverage	Genotype	Calling rate	Coverage	Genotype	Calling rate
rs451887	266	TC		161	TC	
rs45596840	259	GG		124	GG	
rs4919510	122	CC		69	CC	
rs56148568	274	TT		248	TT	
rs56155608	228	GG		93	GG	
rs56310773	270	CC		253	CC	
rs56312243	202	CC		120	CC	
rs56790095	250	CC		176	CC	
rs59323834	250	CC		218	CC	
rs5997893	125	AA		68	AA	
rs61388742	258	TT		103	TT	
rs61938575	252	GA		146	GA	
rs62154973	199	CC		72	CC	
rs62182086	267	AA		235	AA	
rs6513497	273	TT		224	TT	
rs66507245	195	AA		63	AA	
rs66683138	202	GA		100	GA	
rs6771809	263	TT		251	TT	
rs7162033	252	CC		251	CC	
rs7183051	254	GG		252	GG	
rs7208391	251	CG		119	CG	
rs7210937	262	CC		116	CC	
rs7227168	235	CC		82	CC	
rs72810954	13	**NN**		2	**NN**	
rs72955519	238	GG		94	GG	
rs72996752	113	AA		63	AA	
rs7311975	256	TT		85	TT	
rs73251987	262	CC		158	CC	
rs73295187	262	AA		173	AA	
rs73410309	264	GG		120	GG	
rs73721294	258	CC		145	CC	
rs73872515	211	AA		81	AA	
rs74085143	269	GG		127	GG	
rs74469188	251	TT		113	TT	
rs74647838	263	GG		254	GG	
rs74743733	257	GG		187	GG	
rs74814065	245	CC		112	CC	
rs74904371	280	CC		255	CC	
rs75330474	263	GG		127	CC	
rs76347846	283	AA		254	AA	
rs76595065	262	TT		250	TT	
rs7804972	162	GG		63	GG	
rs78293125	264	AA		97	AA	
rs78825966	270	CC		235	CC	
rs79637190	257	CC		187	CC	
rs8078913	173	TT		62	TT	

SNP	Run 1			Run 2		
	Coverage	Genotype	Calling rate	Coverage	Genotype	Calling rate
rs9745376	246	GG		177	GG	
rs9913045	262	AA		197	AA	
Mean	232.6			152.8		
SD	62.82			72.72		

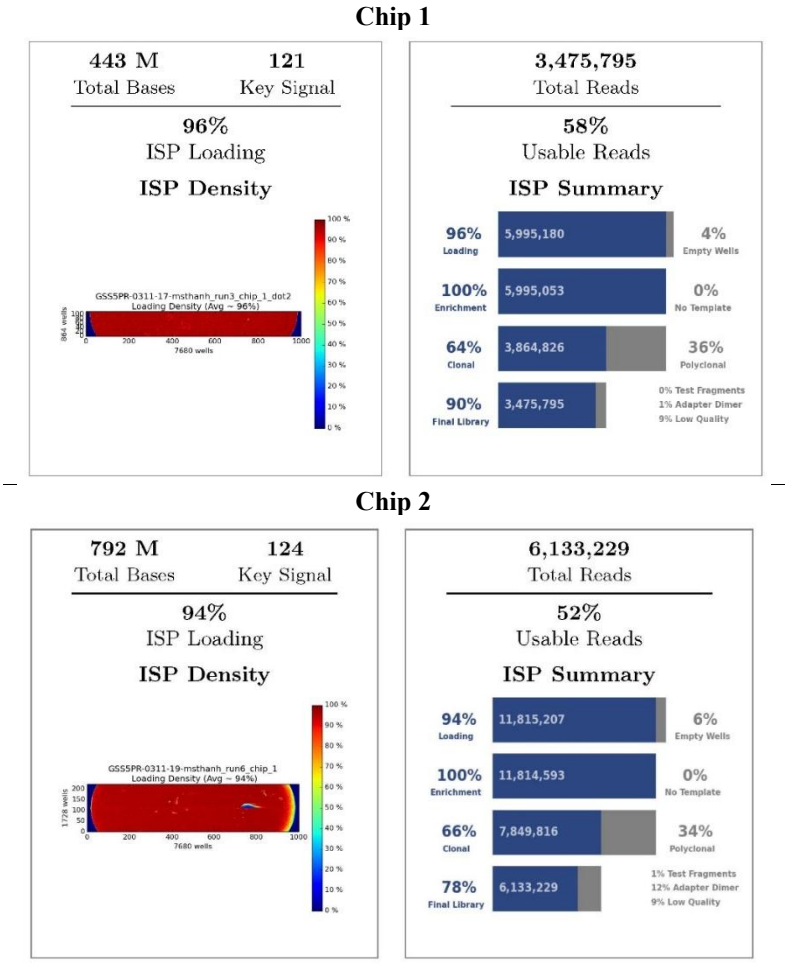
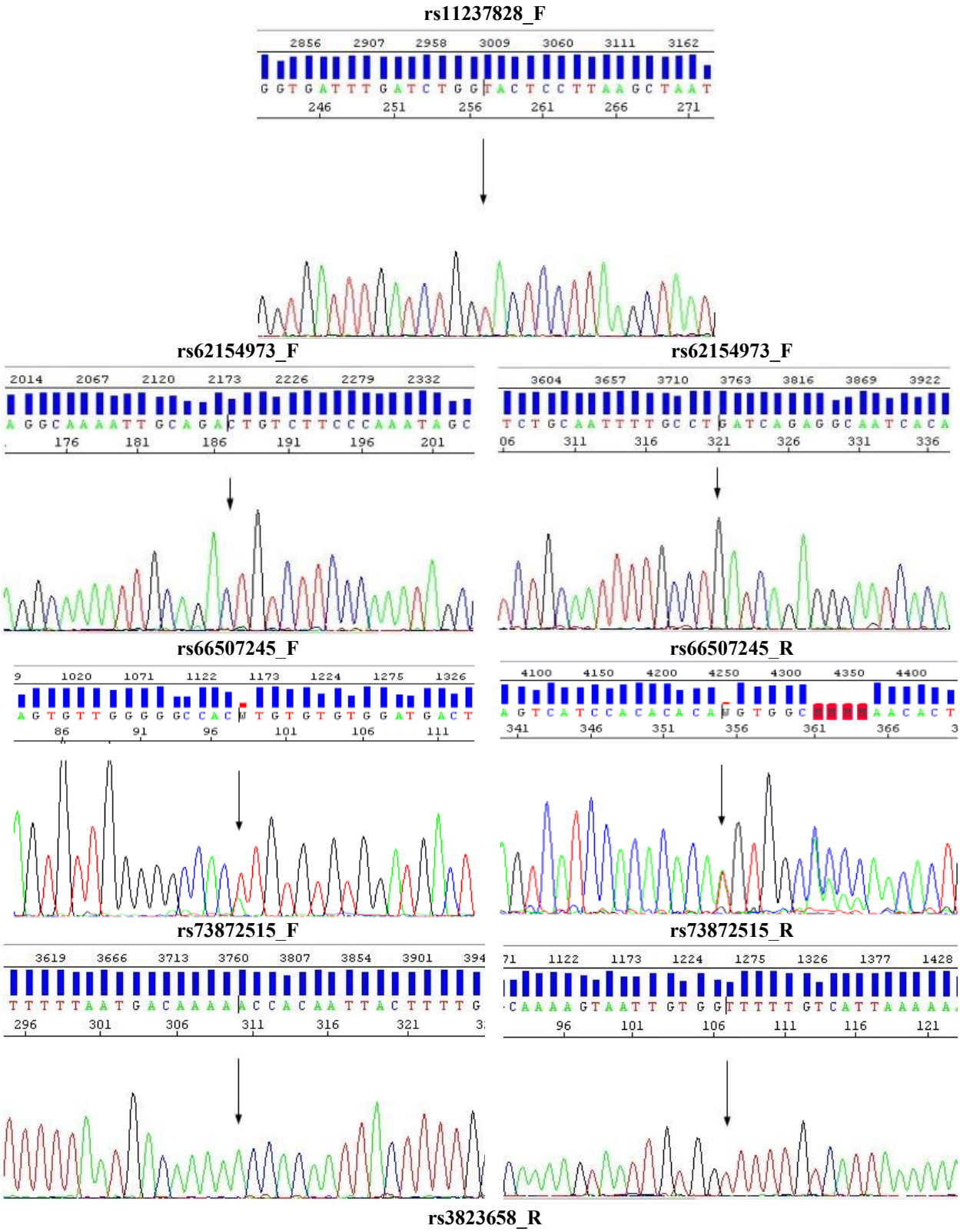
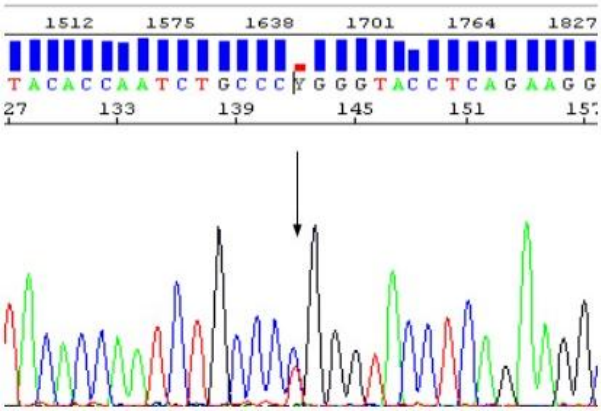
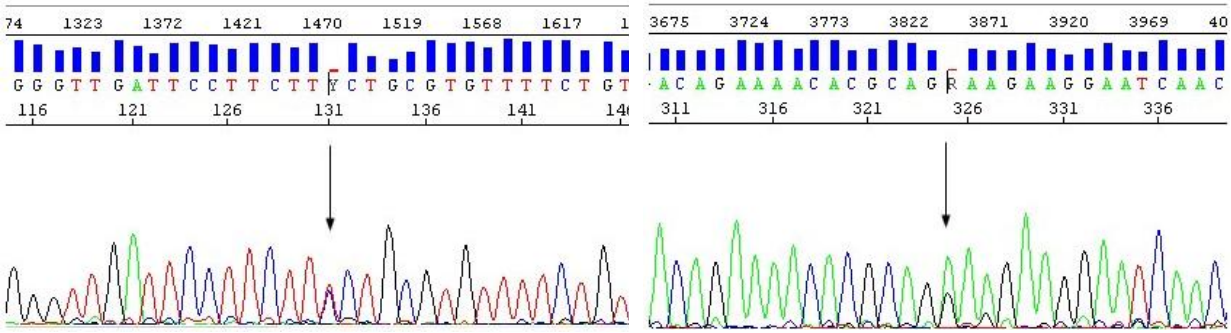


Figure S1 Sequencing performance metrics for control DNA and sample DNAs.

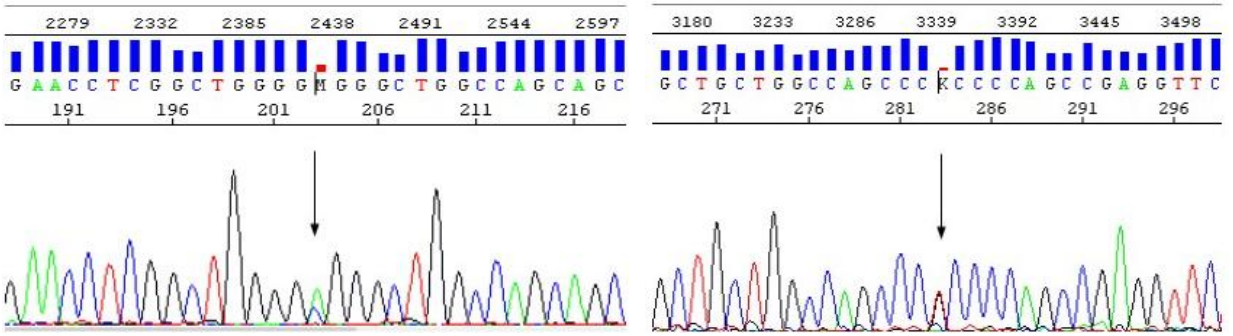

rs11237828_F
rs62154973_F
rs62154973_F
rs66507245_F
rs66507245_R
rs73872515_F
rs73872515_R
rs3823658_R



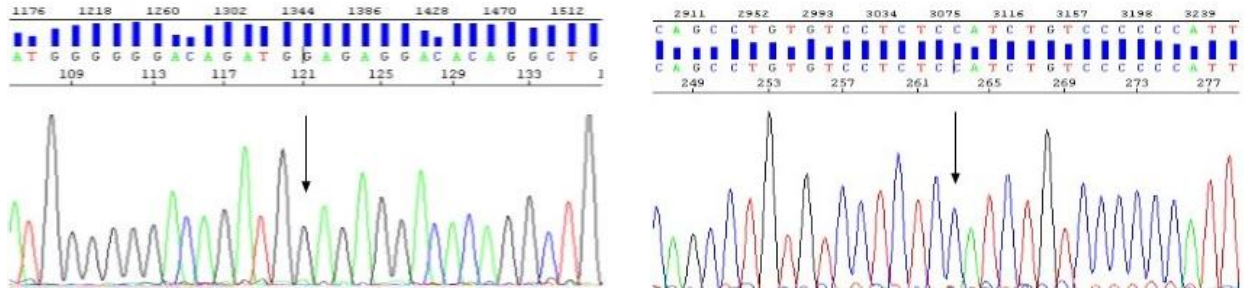
rs8078913_F **rs8078913_R**



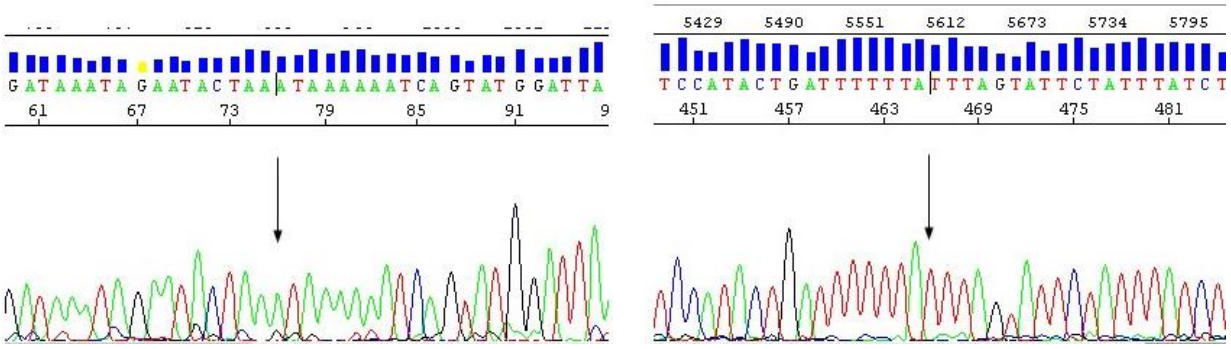
rs2273626_F **rs2273626_F**



rs1688017_F **rs1688017_R**



rs404337_F **rs404337_R**



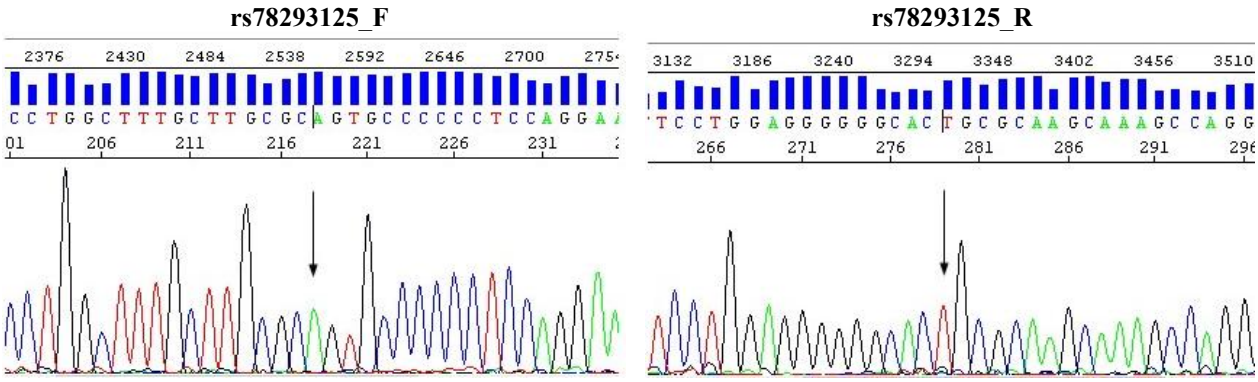


Figure S2 Results of Sanger sequencing for validation of the selected miR-SNPs. Each SNP was sequenced in both forward (F) and reverse (R) directions using specific primers designed for the corresponding locus. Genotypes obtained from Sanger sequencing (Forward and Reverse) were compared with those generated by next-generation sequencing (NGS) to evaluate concordance and genotype calling accuracy. "_F" and "_R" denote chromatogram results from the forward and reverse primers, respectively.