Online Methods

Design and synthesis of padlock probes. For each SNP, we used 20 bp upstream and downstream mRNA sequences as the capturing arms, which are connected by a common linker: GGCTCATCGTTCCTATTCAGCTGCAGATGTTATCGAGGTCCGAC. We added a 5'-adaptor (TTGGGTCATATCGGTCACTGTT) and a 3'-adaptor (GATCAGGATACACACTACCCGTG) to each probe for amplification. Illumina's sequencing adaptors were included in three libraries (CES22k-2, CES22k-3.1, CES22k-3.2) in order to verify them by sequencing without any enzymatic manipulation.

Oligonucleotide libraries synthesized by Agilent were diluted by nuclease-free water to 20 nM (total probe concentration), and amplified in 100 µl reactions in 96-well plates. Each reaction contains 2 pM template oligonucleotides, 1.5 mM MgCl2, 200 nM dNTP, 0.4 × SYBR Green I, 2.5 U JumpStart Taq, 400 nM primers (AP1V6: G*G*GTCATATCGGTCACTGTU; AP2V6: /Phos/CACGGGTAGTGTGTATCCTG) in 1X PCR buffer. Amplifications were performed with a Bio-Rad Chromo4 real-time thermalcycler: 94 ^{o}C 2-minutes, follow by <22 cycles of 94 ^{o}C 30 seconds, 58 ^{o}C 1 minutes, 72^{o}C 30 seconds. The amplification was terminated when the amplification curves increased close to the plateau stage. We found that over-amplification tends to result in single-stranded chimeric sequences that appear as a smear in gel electrophoresis.

Amplicons were pooled and purified with three DNA concentrator-100 columns (Zymo Research). Purified DNA (50~70 µg) was digested with 200U λ exonuclease in 400 µl reaction for 45 minutes at 37 ^{o}C, followed by heat inactivation at 75 ^{o}C for 15 minutes. After purification with four QiaQuick PCR purification columns (Qiagen), the single-stranded DNA was mixed with 8 pmole guide oligonucleotide (GTGTATCCTGATC) in 220 µl 1x *Dpn* II buffer, denatured at 94 ^{o}C for 5 minutes, 60 ^{o}C for 10 minutes, 37^{o}C for 1min, then digested with 200 U *Dpn* II at 37 ^{o}C for 2 hours, and 10 U USER Enzyme for one additional hour. The digestion was repeated once if it is not complete. Digested probes were purified with 6% TB-Urea polyacrylamide gel (Invitrogen), and quantified by comparing the band intensity with a dilution series of DNA ladders of known concentration (Invitrogen Low Mass DNA ladder).

Tissue culture, DNA/RNA preparation and ds-cDNA synthesis. Blood and skin samples were obtained from a Personal Genome Project donors (PGP1 and PGP9) following the protocol approved by Harvard Medical School's Institutional Review Boards. The EB virus transformed B-lymphocyte cell line (PGP1L) was derived and distributed by Coriell Cell Repository. The primary fibroblast lines (PGP1F, PGP9F) and keratinocyte line (PGP1K) was derived in the Brigham Women's Hospital. PGP1L was cultured in RPMI-1640 medium (Invitrogen) with 10% FBS (Invitrogen) and 2 mM L-Glutamine (Invitrogen). PGP1F/PGP9F was cultured in DMEM/F12 medium with 15% FBS and 10 ng/µl EGF. PGP1K was cultured in the Keratinocyte SFM medium (Invitrogen) with 25 µg/ml bovine pituitary extract, 0.2 ng/µl EGF, 0.3 mM CaCl2 and 1x penicillin/streptomycin. Human embryonic stem cell lines (HUES37, HUES38, HUES56, HUES58,) were grown on a feeder layer of mouse embryonic fibroblast in hES media, and mechanically separated from mouse cells prior to DNA/RNA extraction.

Genomic DNAs were extracted with DNeasy kit (Qiagen). Total RNAs were extracted with Rneasy Plus columns (Qiagen) or Trizol (Invitrogen). Genomic DNA contamination in RNA was removed with RNA Clean-up Kit (Zymo Research). First-strand cDNA was synthesized from 10~20 µg total RNA with the SuperScript III First Strand Synthesis System (Invitrogen) using the oligo-dT(12-18) primer, then cleaned up with a Sephadex G-25 column. Double-stranded cDNA was synthesized at 16 ^{o}C for 2 hours in 100 µl reaction containing 0.5 µM dNTP, 20 U DNA polymerase I (Invitrogen) and 4 U Rnase H (Invitrogen), and purified with QiaQuick PCR columns (Qiagen).

SNP capture and sequencing. Circularization were performed in 10 µl reactions (covered with mineral oil) containing 200 ng genomic DNA or 100 ng ds-cDNA, 0.5 pmole padlock probes (total concentration), 0.5 U AmpLigase (Epicenter), 2 U AmpliTaq Stoffel fragment (Applied Biosystems), 1 µM dNTP in 1x AmpLigase buffer. The reactions were incubated at 95 ^{o}C for 5 minutes, 60 ^{o}C for 40 hours. The reactions were then denatured at 94 ^{o}C for 1 minutes, cooled down to 37 ^{o}C, then digested with Exonuclease I (10 U) and Exonuclease III (100 U) for 2 hours at 37^{o}C, and finally heat inactivated at 94 ^{o}C for 5 minutes. Five reactions were performed for each sample, and the resulting products were pooled for PCR and sequencing.

Post-capturing PCR reactions were performed in 100 ul reactions including 10ul circularization product, 200 nM dNTP, 0.4 x SYBR Green I, 0.4 µM forward and reverse PCR primers (AmpF6.2Sol: AATGATACGGCGACCACCGACACTCTCTGCAGATGTTATCGAGGT; AmpR6.2Sol: CAAGCAGAAGACGGCATACGAGCTCTTCACGCAGCTGAATAGGAACGAT) in 1x iProof PCR master mix. Thermal cycling were performed on Chromo4 real-time PCR thermalcycler (Bio-Rad) :98C 30 seconds; followed by 8 cycles of 98^{o}C 15 seconds, 60^{o}C 20 seconds, 72^{o}C 10 seconds; then <15 cycles of 98^{o}C 15 seconds and 72^{o}C 20 seconds. Similar to the amplification of oligonucleotide libraries, we terminated the reactions when the amplification curves went up close to the plateau stage. The amplicons typically contain three or more fragments representing DNA amplified from one, two or more rounds of the circular templates. The smallest amplicon (145 bp) were purified with 6% TBE polyacrylamide gel (Invitrogen), and sequenced with Illumina Genome Analyzer.

Data analysis. We mapped sequencing reads (25-41 bp) to the SNP flanking sequences by NCBI BLAST using the word size of 8-10 depending on the read length. We made genotyping calls using the “best-P” method on SNPs that were sampled at least 20 times. For each SNP we performed both the test of homozygosity (assuming the allelic ratio of 1-e:e where e is the

sequencing error) and the test of heterozygosity (assuming 50:50 allelic ratio), and determined the genotype based on the one that gives a higher p-value (less likely to reject the null hypothesis).

We used a QC metrics for the assay based on the consistency of multiple SNPs in the same exons or the same genes. Since the haplotypes are unknown, we first calculated the absolute deviations of allelic ratios from 0.5. The standard deviation of this absolute deviation was calculated for the SNPs in the same exons/genes, and then was averaged over all exons/genes in assayed. The typical value for a successful assay is approximately 0.05, which is consistent with the theoretical expectation of the level of stochastic drift when sampling from 50 independent events. A value of > 0.08 indicates sub-optimal experimental conditions, such as poor RNA/cDNA quality, or impurity of padlock probes.

We used $\chi 2$ test to identify SNPs that exhibit RNA allelic ratios significantly different from the genomic allelic ratios. Hierarchical clustering and ANOVA analyses were performed with the R package.

The sequences and other information of the CES27k probe set are listed in Supplementary Table 1.

Illustration	Title
Supplementary Figure 1	Comparison of allelic ratios quantified by digital allelotyping and quantitative Sanger sequencing.
Supplementary Figure 2	False positive in technical replicates and variability in biological replicates.
Supplementary Figure 3	Histogram of allelic ratio distribution.
Supplementary Figure 4	Distribution of allelic ratios as a function of sequencing depth.
Supplementary Figure 5	Distribution of SS/DS ratios for SNPs captured from the sense strands and the anti-sense strands.
Supplementary Figure 6	Line-specific ASE in sibling hES cells.
Supplementary Table 1	Detailed information of the CES27k probe set (included in a separated Excel file)
Supplementary Note	

AOP
Padlock probes, synthesized in large scale on programmable microarrays, capture expressed SNPs for high throughput sequencing in this method for RNA allelotyping. The approach combines the sensitivity

of digital expression measurements with the efficiency of targeted re-sequencing to quantify allele specific gene expression in various tissues across several individuals.

Issue

Padlock probes, synthesized in large scale on programmable microarrays, capture expressed SNPs for high throughput sequencing in this method for RNA allelotyping. The approach combines the sensitivity of digital expression measurements with the efficiency of targeted re-sequencing to quantify allele specific gene expression in various tissues across several individuals.