Introduction
Misidentified and cross-contaminated cell lines and patient samples are ongoing problems in biomedical research. It is estimated that 5%–35% of human cell lines are misidentified, resulting in a huge waste of resources and publication of false or misleading data.

The current ANSI standard for cell line authentication is short tandem repeat (STR) profiling, a methodology with several acknowledged disadvantages that can lead to misclassification. Here we evaluate a high-throughput panel of 96 single-nucleotide polymorphism (SNP) assays utilizing Fluidigm microfluidics technology for authentication and sex typing of human cell lines and tissues.

The Fluidigm SNP Trace panel was tested on 907 human cell lines previously characterized by 8- or 16-locus STR typing. We report its performance in distinguishing related and unrelated samples, sex typing of samples, and detection of interspecies cross-contamination.

Methods
The SNP Trace panel consists of 96 SNP Type™ allele-specific PCR assays chosen or confirmed by Dr. Andrew Brooks of Rutgers University. Cell mixtures were created at fixed ratios of 100:0, 99:1, 98:2, 95:5, 90:10, 50:50, 9:55, 2:98, 139:1, 0:100. DNA from the 925 cell line and cell mixtures were purified by DNeasy Blood & Tissue Kit (Qiagen, Cat. No. 69506) or 96 Blood & Tissue Kit (Qiagen, Cat. No. 69585), normalized to 50 ng/µL and pre-amplified as described in the Fluidigm SNP Genotyping User Guide. Diploid human male and female DNA were supplied by RUO® Infinite Biologics (Piscataway, NJ), normalized to 60 ng/µL and used to create 20%, 15%, 10%, 5%, and 1% contamination samples by volume. All samples were analyzed by the SNP Trace panel using Fluidigm High-Precision 96.96 Genotyping® (PCRs integrated fluidic circuits) in the Fluidigm Biomark™ HD system. We followed Tanabe et al. and Tanabe, H. et al. Tissue Culture Research Communications, 18 (1999): 329–38 to compute an identity score for sample pairs. Spectral karyotyping (SKY) procedures were carried out as previously described and analyzed using the HISKY® v6.0 software (ASB).

Conclusion
• The SNP Trace panel showed equivalent performance to STR:
  — Discriminating related/unrelated samples
  — Sex typing
  — Interspecies cross-contamination
  — 96-SNPs outperformed 48- and 24-SNPs.
• SNPs can detect as low as 2% to 10% interspecies cross-contamination.
  — Aneuploidy affects results (STR and SNP).
  — 40%–45% of males sampled type as female, independent of assay.
  — Many samples have lost most or all of Chromosome Y.

In summary, the SNP Trace panel is a fast, reliable, accurate and cost-effective method to assess cell line or human biosample identity and intrahuman cross-contamination. The SNP Trace panel can be used alone or in conjunction with established cell line authentication methods to continually monitor cell identity. This study also provides a resource of 907 SNP profiles for future comparisons.

SNP Trace panel profiles
Goal: Evaluate the ability of the SNP Trace panel to identify related and unrelated cell lines.

Method: Pairwise comparisons of the SNP Trace panel results for all samples to calculate identity scores.

Figure 1. Pairwise analysis of all cell line samples broken out into technical and biological replicates, synonymous/derivative sample pairs, and unrelated samples. For unrelated samples, 24- and 48-SNPs were randomly selected for pairwise comparisons and contrasted with the complete 96-SNP set.

Results:
• Biological and technical replicates were highly similar.
  — Cell lines of common origin were highly similar.
  — Highly related samples are defined by an identity score of 95% or greater.
  — 96 SNPs outperform 24- and 48-SNs in identifying related samples.

Goal: Compare the performance of the SNP Trace panel to STR.

Method: Comparison of pairwise comparisons for STR profiles with those of the SNP Trace panel.

Comparison of pairwise scores from the SNP Trace panel and STR profiling. Shaded region is joint distribution of SNP and STR-based identity scores for pairwise comparisons of unrelated lines (n = 467,741). Blue diamonds signify two unrelated pairs with unexpectedly high identity scores. Red and green points correspond to synonymous and replicate pairs, respectively.

Results:
• Related and unrelated samples are distinguished equally by the SNP Trace panel and STR.
  — Two cases of sample handling error were identified.

Sex typing of samples
STRs have a high rate of calling male samples female due to loss of the amelogenin locus. Unlike other SNP profiling panels, the SNP Trace panel has six assays to determine sex and thus loss of one locus does not affect sex calls.

Goal: Compare three different methods (STR, the SNP Trace panel, and Illumina® SNP arrays) of sex identification with annotated sex calls.

Table 1. List of observed discrepancies in SNP/STR comparisons.

Table 2. Comparison of sex calls from Illumina SNP arrays, the SNP Trace panel and STR profiles with annotated sex calls for all cell lines.

Results:
• All three assays performed similarly with 40%–45% of males called as females, but only 3%–7% of female cell lines called as male.
• Spectral Karyotyping (SKY) of 19 cell lines indicated that male samples typed as female had lost the Y chromosome.
• Sex typing is a challenge in aneuploidy samples.

Intraspecies Contamination
Cell line cross-contamination is a major problem in biomedical sciences. STRs can reliably detect 5%–10% contamination.

Goal: Determine the sensitivity of the SNP Trace panel to detect interspecies contamination.

Table 3. Detection of cell line cross-contamination by the SNP Trace panel in diploid cells. Percent identity to each sample is shown. Shaded results indicate contamination.

Results:
• The SNP Trace panel reliably detected 5% contamination in three independent mixes.

Table 4. Detection of cell line cross-contamination by the SNP Trace panel in cancer cells. Percent identity to each sample is shown. Shaded results indicate contamination.

Results:
• Sensitivity of detection ranged from 2% to 10% for cancer cells.
• Results depended on mix ratios, most likely because of genomeric instability/aneuploidy in those samples.