Human sample authentication using the SNP Trace™ panel: a high-throughput, cost-effective SNP-based platform for fingerprinting human tissue and cell lines

May Min Yian Liang-Chu[1], Mamie Yu[1], Peter M. Haverty[2], Marie M. Lee[3], Richard M. Neve[1], Richard Bourgon[2], and Janet Ziegle[3]

[1]Department of Discovery Oncology, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080
[2]Department of Bioinformatics and Computational Biology, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080
[3]Fluidigm Corporation, 7000 Shoreline Court, Suite 100, South San Francisco, CA 94080

Introduction

Misidentified and cross-contaminated cell lines and patient samples are ongoing problems in biomedical research. It is estimated that 15%–35% or more 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 profiling. We report its performance in distinguishing related and unrelated samples, sex typing of samples, and detection of intraspecies 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, 5:95, 10:90, 2:98, 1:99, 0:100. DNA from the 925 cancer cell lines and cell mixtures were purified by DNeasy® Blood & Tissue (Qiagen®, Cat. No.69506) or 96 Blood & Tissue Kit (Qiagen, Cat. No. 69581), 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 RUCDR® 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™ IFCs (integrated fluidic circuits) in the Fluidigm Biomark™ HD system. We followed Tanabe et al. 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 (ASI).

Conclusion

- The SNP Trace panel showed equivalent performance to STR:

— Discriminating related/unrelated samples
— Sex typing
— Intraspecies cross-contamination

- 96-SNPs outperformed 48- and 24-SNPs.
- SNPs can detect as low as 2% to10% intraspecies cross-contamination.

— Aneuploidy affects results (STR and SNP).

- 40%–45% of males samples 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.

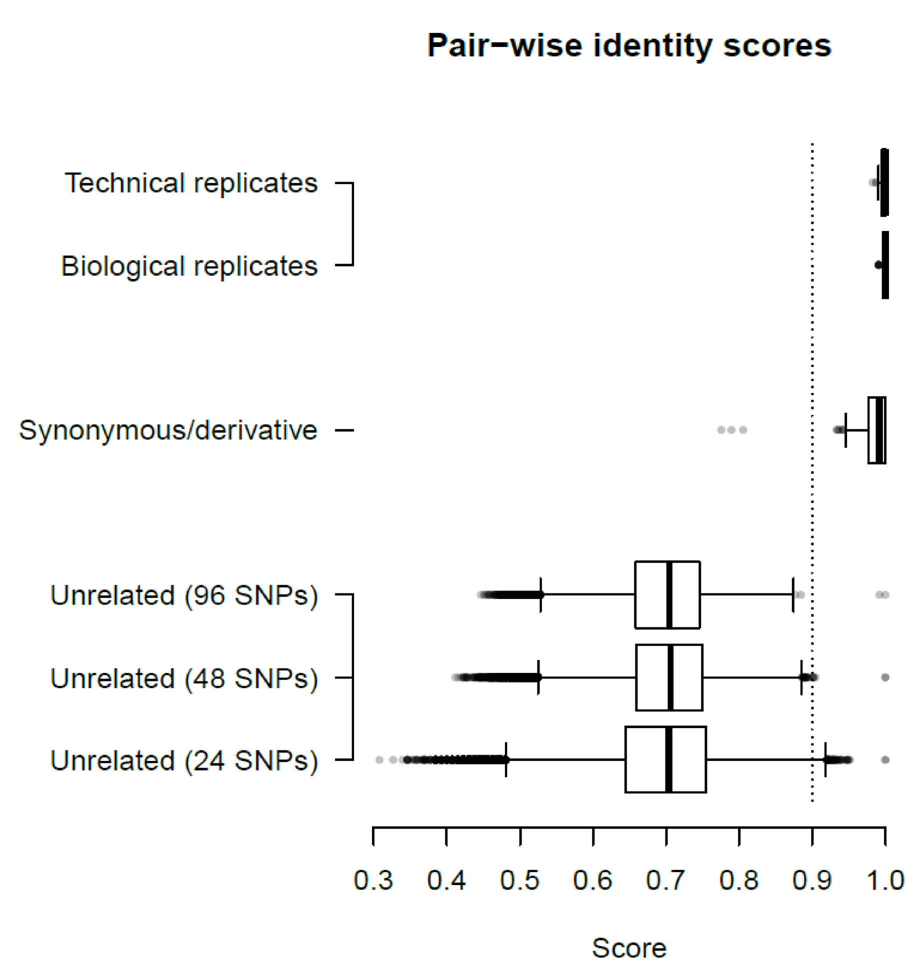


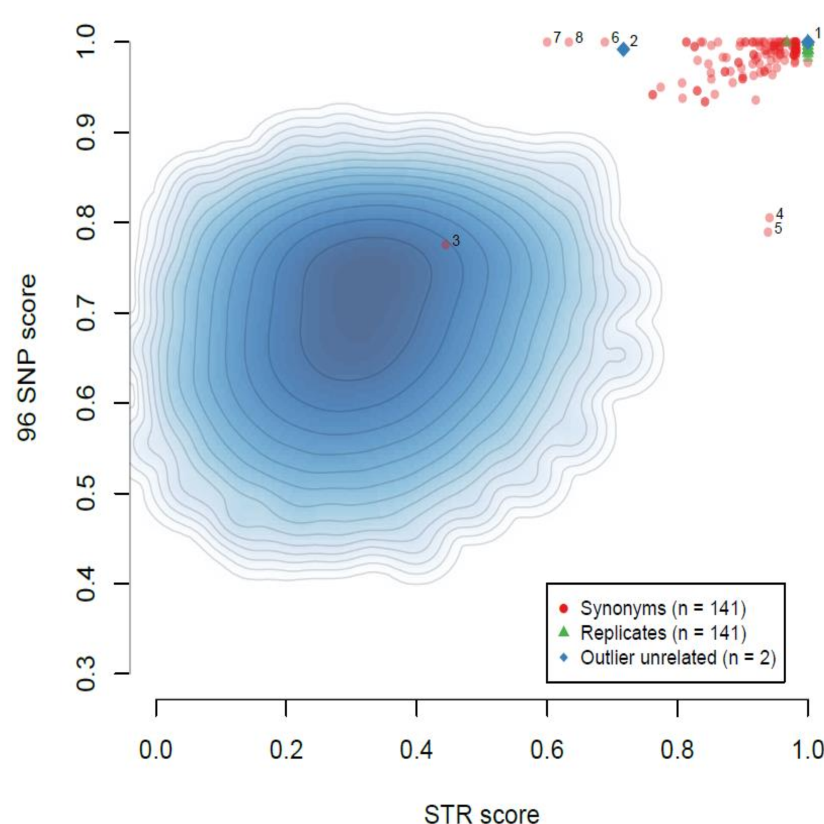
Figure 1. Pairwise analysis of all cell line samples broken out into technical and biological replicates, synonymous/derivate samples, and unrelated samples. For unrelated samples, 24- and 48- SNPs were randomly selected for pairwise comparisons and contrasted with the complete 96-SNP set.

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 >90%.
- 96 SNPs outperform 24- and 48- SNPs in identifying related samples.

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

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



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* = 467,744). 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.
- A previously unknown pair of related cell lines (NCI-H2691/NCI-H2803) was identified.
- 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.

	Analysis Calls					
	STR		SNPTrace		Illumina Array	
Annotated	Male	Female	Male	Female	Male	Female
Male	220 (55%)	180 (45%)	219 (60%)	144 (40%)	194 (60%)	131 (40%)
Female	8 (2%)	382 (98%)	15 (4%)	327 (96%)	23 (7%)	313 (93%)
Total =	790		705		661	

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

Call (based on SNP array)	Cell Line Name	CUD	Sex Calls				
			Annotated	STR	SNPtrace	Illumina	SKY Karyotype
Female Typed as Female	LXFL529	130847	FEMALE	FEMALE	FEMALE	FEMALE	XX
	NCI-H1993	129701	FEMALE	FEMALE	FEMALE	FEMALE	XXXXXX
	NC1-H2009	131971	FEMALE	FEMALE	MALE	FEMALE	X
	NC1-H2073	129695	FEMALE	FEMALE	NO CALL	FEMALE	XXXXXX
	NC1-H2122	130148	FEMALE	FEMALE	FEMALE	FEMALE	XX
	NC1-H292	129184	FEMALE	FEMALE	NO CALL	FEMALE	XX
Male Typed As Female	NC1-H1155	13178	MALE	FEMALE	NO CALL	FEMALE	XX
	NC1-H1299	129461	MALE	FEMALE	NO CALL	FEMALE	XXXX
	NC1-H23	129190	MALE	FEMALE	FEMALE	FEMALE	XX
	NC1-H322T	585717	MALE	FEMALE	FEMALE	FEMALE	XX
	NC1-441	130150	MALE	MALE	MALE	FEMALE	XX
	NC1-H522	129185	MALE	FEMALE	FEMALE	FEMALE	XX
	NC1-H838	129191	MALE	FEMALE	FEMALE	FEMALE	X
Male Typed As Male	A549	130497	MALE	MALE	MALE	MALE	XXY
	NCI-H1703	130536	MALE	MALE	MALE	MALE	XY
	NCI-H226	129544	MALE	MALE	MALE	MALE	XYY
	NCI-H358	129179	MALE	MALE	MALE	MALE	XXY
	NCI-H460	129455	MALE	MALE	MALE	MALE	XXYY
	NCI-H650	129187	MALE	FEMALE	FEMALE	MALE	XXX

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 intraspecies contamination.

	A:B mix		C:D mix		E:F mix	
	Identity comparison with:					
Mix Ratio:	A	B	C	D	E	F
100/0	100%	45%	100%	36%	100%	22%
99.5/0.5	98%	45%	97%	39%	98%	23%
99/1	99%	45%	98%	39%	99%	23%
95/5	88%	45%	88%	39%	80%	25%
90/10	81%	48%	77%	49%	72%	25%
80/20	76%	53%	71%	51%	60%	30%
50/50	71%	56%	57%	58%	48%	45%
50/50	68%	60%	60%	59%	42%	44%
20/80	50%	73%	46%	70%	35%	59%
10/90	51%	74%	43%	78%	28%	72%
5/95	47%	82%	42%	86%	28%	77%
1/99	45%	94%	40%	91%	21%	93%
0.5/99.5	45%	92%	41%	92%	22%	93%
0/100	45%	100%	36%	100%	22%	100%

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.

	MCF-7:LOX-IMVI mix		AU565:SK-MES-1 mix		SK-MEL-1:THP-1 mix	
	Identity comparison with:					
Mix Ratio:	MCF7	LOX-IMVI	AU565	SK-MES-1	SK-MEL-1	THP-1
100/0	100%	55%	100%	53%	100%	38%
99/1	100%	55%	98%	54%	97%	39%
98/2	100%	55%	97%	54%	95%	39%
95/5	93%	55%	95%	54%	80%	39%
90/10	83%	59%	88%	53%	66%	42%
50/50	71%	67%	73%	64%	50%	58%
10/90	61%	75%	61%	73%	40%	71%
5/95	58%	77%	59%	78%	39%	78%
2/98	55%	91%	55%	92%	38%	89%
1/99	55%	95%	54%	95%	38%	96%
0/100	55%	100%	53%	100%	38%	100%

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% to10% for cancer cells.
- Results depended on mix ratios, most likely because of genomic instability/aneuploidy in these samples.

Fluidigm Corporation
7000 Shoreline Court, Suite 100
South San Francisco, California 94080
Toll-free: 1.866.FLUIDLINE • www.fluidigm.com

For Research Use Only.
Not for use in diagnostic procedures.

© Fluidigm Corporation. All rights reserved. Fluidigm, the Fluidigm logo, Biomark, High Precision 96.96 Genotyping, and SNP Trace are trademarks or registered trademarks of Fluidigm Corporation.