

# Cell Line Authentication Using GeneMarker<sup>®</sup>HID, Human Identity Software

September 2013

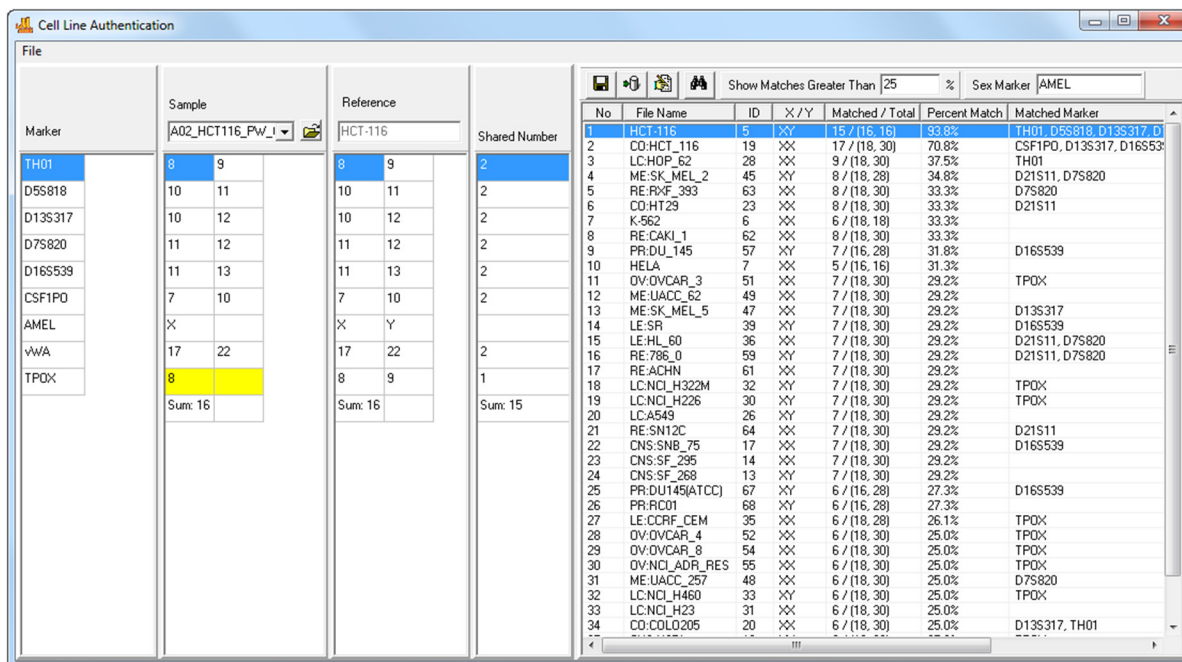
Adam Dillman, Teresa Snyder-Leiby, Xin Li and C.S. Jonathan Liu, SoftGenetics, LLC

## Introduction

The scientific and biomedical research communities have recently been shaken by revelations that a substantial portion of cell lines - as much as 36% - are either contaminated or misidentified entirely<sup>1</sup>. Cell lines are used extensively as in vitro models for a wide range of fields, and researchers often select cell lines because they have specific properties, or because they originate from a specific tissue type. Clearly these findings cast years, if not decades of research into question, and have permanently altered the way research utilizing cell lines is evaluated<sup>2,3</sup>. Many scientific journals now refuse to accept cell line studies unless the cell lines in question have been authenticated.<sup>4</sup>

Fortunately, recent advances in qPCR technology have drastically reduced the cost and time needed to authenticate samples. Kits designed to amplify repetitive loci can generate a unique STR profile for a given cell line, which can then be compared to a database of known cell line genotypes to ascertain the closest match. But while this approach represents a great improvement, it is still severely limited in several ways. After amplification and CE separation, researchers must genotype their samples using one or more analysis programs. Then, after obtaining genotypes, the user must manually query an online database, for example ATCC or CellMiner, usually by painstakingly typing in the STR profile for each locus, and then repeating this process for each sample.

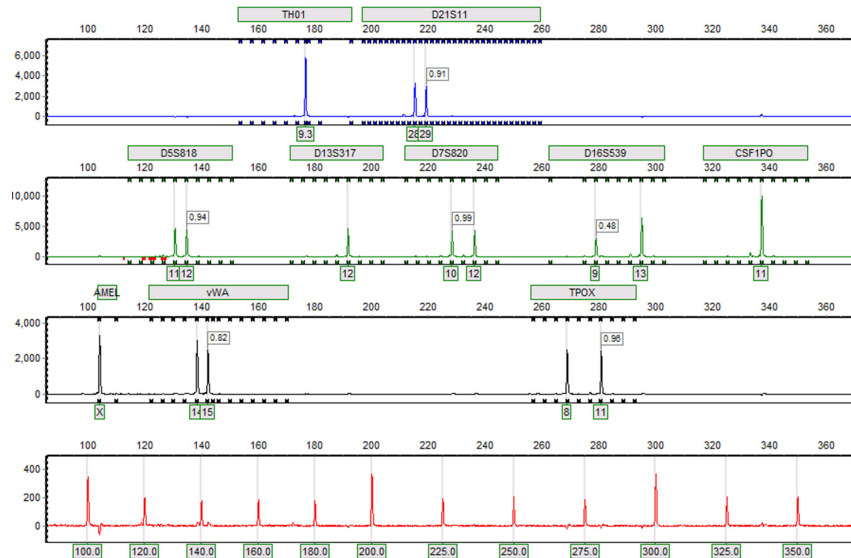
To resolve these problems, GeneMarkerHID is now provided with a Cell Line Authentication tool, which integrates data processing, genotyping, and database searching into a single program. The user may load their raw data files, perform genotyping, and then query the built-in database of cell lines directly - eliminating error prone data transfers, while greatly increasing the speed of the verification process. GeneMarkerHID includes an integral Mixture Analysis tool, allowing the researcher to track down possible contaminants. Finally, GeneMarkerHID is equipped with several customizable reporting and exporting options, and is compatible with all major chemistries and CE instruments.



**Figure 1:** GeneMarkerHID's built-in Cell Line Authentication application allows the user to search for matches between their sample files and a database of preloaded cell line genotypes. Here, the user's sample is most closely matched to HCT-116, although the sample differs from the reference at the TPOX locus.

## Procedure

1. Load raw data files (.fsa, .hid, .sg1, .esd and .rsd).
2. Select presaved template in the Run Wizard and process data.
3. Verify size and allele calls (Figure 2).
4. Proceed to the Cell Line Authentication Application (Applications → Cell Line Authentication).
5. Select percent match threshold or use the default value (80%).
6. Select sample file from the sample dropdown menu to search the database.
7. View automated search results in the report table.
8. Save or export results.



**Figure 2:** Genotypes can be reviewed, printed, and/or saved from GeneMarkerHID’s main analysis window. After genotyping the user can proceed directly to the Cell Line Authentication application. No data transfer is required.

GeneMarkerHID has a built in database with a core of 60 common cell line genotypes, but it can be easily expanded to ultimately include thousands of references. These additional references can be uploaded as a tab-delimited text file, and are permanently saved in the database (Figure 3). Each reference must only be uploaded once - thereafter they can be reviewed, deleted, or modified using GeneMarkerHID’s database management tools.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	AID	AMEL	AMEL	D5S818	D5S818	D13S317	D13S317	D7S820	D7S820	D16S539	D16S539	vWA	vWA	TH01
2	HCT-116	X	Y	10	11	10	12	11	12	11	13	17	22	8
3	K-562	X	X	11	12	8	8	9	11	11	12	16	16	9.3
4	HELA	X	X	11	12	12	13.3	8	12	9	10	16	18	7
5	BR-BT_549	X	X											
6	BR-H55787	X	X											
7	BR-MCF7	X	X											
8	BR-MDA_MB_231	X	X											
9	BR-T47D	X	X											
10	CNS-SF_268	X	Y											
11	CNS-SF_295	X	X											
12	CNS-SF_539	X	X											
13	CNS-SNB_19	X	Y											
14	CNS-SNB_75	X	X											
15	CNS-U251	X	Y											
16	CO-HCT_116	X	X											
17	CO-COLO205	X	X											
18	CO-HCC_2998	X	X											
19	CO-HCT_15	X	Y											
20	CO-HT29	X	X											
21	CO-KM12	X	X											
22	CO-SW_620	X	X											
23	LC-A549	X	Y											
24	LC-EKVP	X	Y											
25	LC-HOP_62	X	X											
26	LC-HOP_92	X	X											
27	LC-NCI_H226	X	Y											
28	LC-NCI_H23	X	X											
29	LC-NCI_H322M	X	Y											
30	LC-NCI_H460	X	Y											
31	LC-NCI_H522	X	X											
32	LE-CCRF_CEM	X	X											
33	LE-HL_60	X	X											

No.	Sample Name	AMEL	D5S818	D13S317	D7S820	D16S539	vWA	TH01	TF
1	HCT-116	X	10	11	11	12	17	22	8
2	K-562	X	11	12	8	9	11	12	16
3	HELA	X	11	12	12	13.3	8	12	9
4	BR-BT_549	X							
5	BR-H55787	X							
6	BR-MCF7	X							
7	BR-MDA_MB_231	X							
8	BR-T47D	X							
9	CNS-SF_268	X							
10	CNS-SF_295	X							
11	CNS-SF_539	X							
12	CNS-SNB_19	X							
13	CNS-SNB_75	X							
14	CNS-U251	X							
15	CO-HCT_116	X							
16	CO-COLO205	X							
17	CO-HCC_2998	X							
18	CO-HCT_15	X							
19	CO-HT29	X							
20	CO-KM12	X							
21	CO-SW_620	X							
22	LC-A549	X							
23	LC-EKVP	X							
24	LC-HOP_62	X							
25	LC-HOP_92	X							
26	LC-NCI_H226	X							

**Figure 3:** Building your database. Genotypes can be easily formatted in a spreadsheet program such as Microsoft Excel, and then uploaded to GeneMarkerHID as a tab-delimited text file. Thousands of samples can be uploaded at a time.

---

## Results & Discussion

High rates of contamination have greatly increased the necessity for cell line authentication. GeneMarkerHID assists with this process by combining raw data analysis, genotyping, and database searching into a single platform – eliminating the need for multiple software packages, unnecessary data transfers, and meticulous manual entry of sample genotypes.

In GeneMarkerHID, sample genotypes are automatically imported directly from the researcher's project into the Cell Line Authentication application. Simply select the desired sample from a dropdown menu to view its genotype and search the database for similar genotypes using a percent match calculation. Results are displayed in an informative table, and clicking on a result will display that reference's genotype, allowing easy side-by-side comparison with the experimental sample. The researcher may choose to display only those results which meet a minimum percent match score. This value can be set to any number between zero and one hundred.

GeneMarkerHID also comes preloaded with relationship testing, database search, and mixture deconvolution applications, allowing analysts to further explore potentially contaminated samples. Finally, GeneMarkerHID supports numerous reporting and saving options, including customizable print reports, excel and text exports, and automated CODIS formatting.

## References

1. Masters et al. "Short tandem repeat profiling provides an international reference standard for human cell lines". PNAS (2001) 14: 8012-8017.
2. Masters, John R. W. "Cell line misidentification: the beginning of the end". Nature Reviews Cancer (2010).
3. Lacroix, Marc. "Persistent used of 'false' cell lines". Int. J. Cancer (2008) 122: 1-4.
4. Lichter, Peter et al. "Obligation for cell line authentication: appeal for concerted action". Int J. Cancer (2010) 126: 1.
5. Capes-Davis et al. "Check your cultures! A list of cross-contaminated or misidentified cell lines". International Journal of Cancer (2010) 127: 1-8.

## Acknowledgements

Special thanks to Laura Hall (Moffitt Cancer Center, Tampa, Florida) for providing sample files.