Matching identities of iPSCs and donors using CLA IdentiFiler STR profiling kits

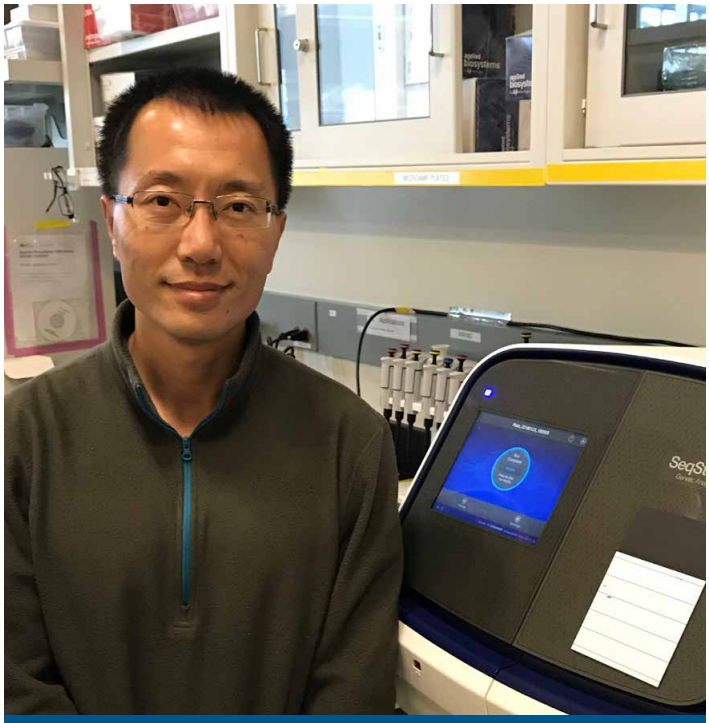
In this report, we show:

- Induced pluripotent stem cells (iPSCs) and donor blood draws can be matched using short tandem repeat (STR) profiling
- Applied Biosystems™ Cell Line Authentication (CLA) IdentiFiler™ Direct kits and NUCLEIC-CARD™ devices can simplify authentication by bypassing DNA purification steps
- The Applied Biosystems™ SeqStudio™ Genetic Analyzer provides a "right-throughput" option for iPSC researchers

Introduction

Research on stem cells as potential treatments for debilitating syndromes is progressing at a rapid rate. The discovery that differentiated adult cells could be induced to become iPSCs by treating with a defined set of genes (e.g., by transforming/transducing with DNA or mRNA) [1,2] opened the door to novel opportunities for basic research, disease modeling, and drug discovery. One exciting outcome of these studies is the potential development of *ex vivo* cellular therapies for some debilitating syndromes.

An important step in developing *ex vivo* therapies is ensuring that the cells derived from the iPSCs are matched to the original donor. Matching iPSCs to the original donor ensures that the appropriate cells are utilized and potential immunological rejection is minimized. Several methods are used to authenticate the iPSC samples: one of these, single-nucleotide polymorphism (SNP) genotyping, is performed with either next-generation sequencing (NGS) or hybridization arrays. Although SNP genotyping methods provide high resolution, the workflows are often tedious and time-to-results can be long. Another method is based on the analysis of STR fragments with highly variable lengths. This method is widely used in forensic analysis since a unique molecular fingerprint of alleles at different genomic loci is generated.



Guangwen (Gavin) Wang is the Director of the Stem Cell Core in the Department of Genetics, Stanford University. In 2012, he initiated the Stem Cell Core at Stanford University. The major focus of the facility is to collaborate with different labs for derivation, characterization, and differentiation of iPSCs. His group is now working to build an iPSC biobank from 300 donors with different types of heart disease. With the support of the Grace Science Foundation, the lab is also using iPSCs as the tool to study the mechanism underlying the ultra-rare *NGLY1* deficiency.

ThermoFisher SCIENTIFIC

An added benefit of short tandem repeat (STR) genotyping is that the workflow and analysis are simple, and can be performed in a matter of hours. Analysis of STRs is performed by capillary electrophoresis (CE) of fragments amplified from microsatellite loci with variable numbers of repeats. Thermo Fisher Scientific offers instruments that are optimized for researchers' needs in both sensitivity and throughput. Furthermore, the Applied Biosystems™ product portfolio has several different kits for PCR-based STR fingerprinting for use on CE instruments. The Applied Biosystems™ CLA IdentiFiler™ Plus PCR Amplification Kit has been optimized to analyze 16 highly variant human STRs over a wide range of purified genomic DNA (gDNA) preparations. The Applied Biosystems™ CLA IdentiFiler™ Direct PCR Amplification Kit was developed to analyze the same 16 STR loci, starting from dried blood, buccal spots (for example, on NUCLEIC-CARD devices), or buccal swabs. For the NUCLEIC-CARD device, a 1.2 mm punch from the card is placed directly into a PCR tube or well, and amplified without any further purification. An illustration of the complete workflows for STR analysis is shown in Figure 1 (see also reference 3).

Figure 1. Workflows for human sample authentication. Two methods can be used to match human samples. **(A)** Samples containing intact cells can be spotted onto NUCLEIC-CARD devices, punches of the cards amplified directly using the CLA IdentiFiler Direct kit, and fragments analyzed on Applied Biosystems™ CE instruments using Applied Biosystems™ GeneMapper™ Software 6. **(B)** Alternatively, gDNA can be purified from cell lines and amplified using the CLA IdentiFiler Plus kit, and fragments analyzed by CE using GeneMapper Software 6.

Background

Dr. Guangwen (Gavin) Wang is the Director of the Stem Cell Core facility at Stanford University. Donor research samples were sent to his laboratory, and subsequently used to generate iPSCs and their derivatives. In his laboratory Dr. Wang was using a SNP genotyping strategy for sample matching, but he found that the turnaround time was long, and data analysis was complicated. He was therefore interested in an STR-based solution.

Materials and methods

To test how well the CLA IdentiFiler kits could verify samples that came from the same individual, Dr. Wang collected blood samples from 12 donors. He generated iPSC cultures from those samples, randomized and blinded pairs, and sent purified DNA from the randomized samples to Thermo Fisher Scientific. In our laboratories, we used the CLA IdentiFiler Plus kit to generate STR profiles from all of the samples, and pairs were identified by matching STR profiles.

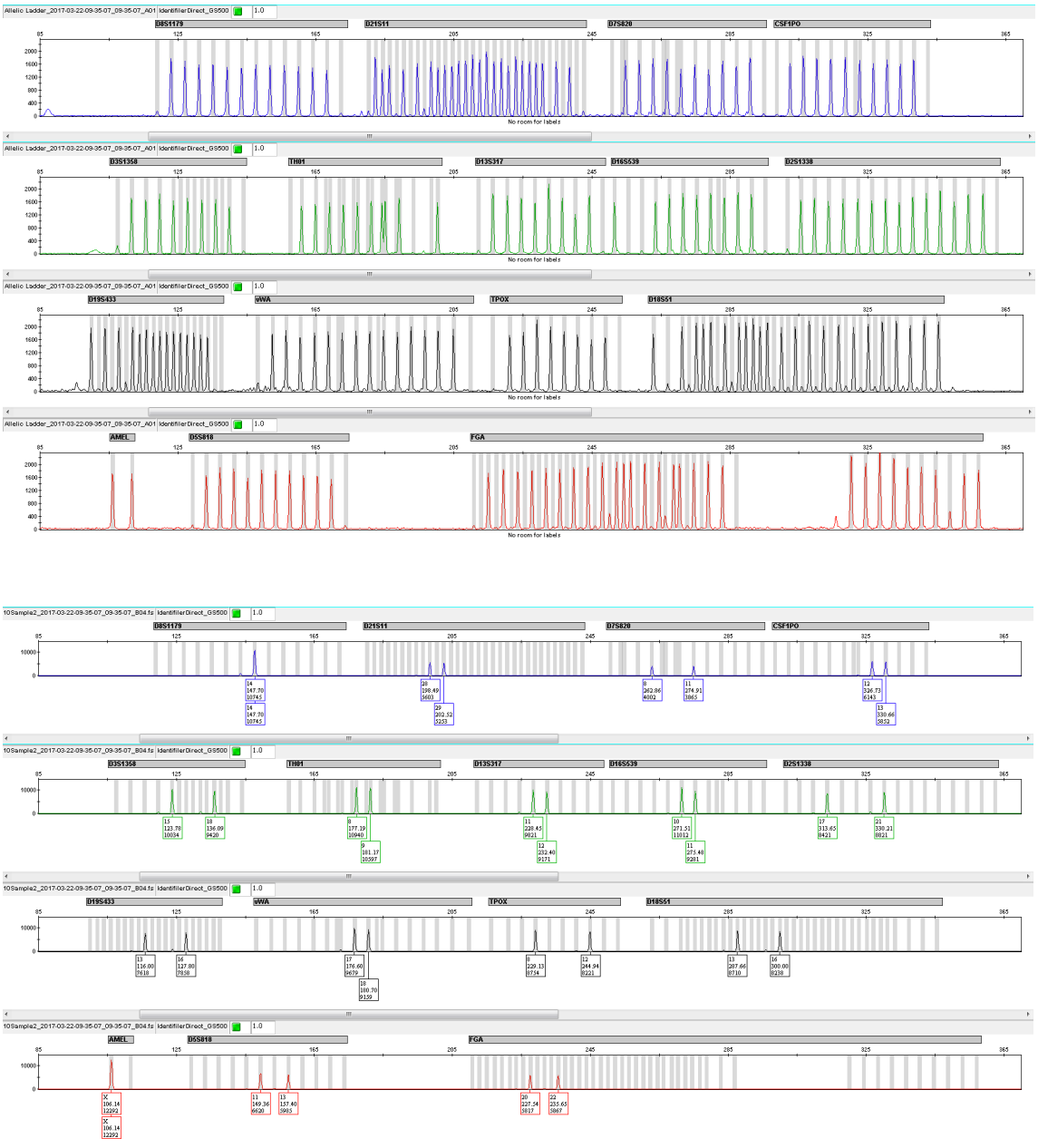


Figure 2. Analysis of CLA IdentiFiler™ Plus Allelic Ladder (top) and purified genomic DNA from an iPSC sample (bottom) by GeneMapper Software 6. GeneMapper Software 6 uses the allelic ladder provided in the CLA IdentiFiler kits to assign the alleles present in an unknown sample to known STR alleles. The boxes below the peaks show the allele number, the height of the peak, and the size of the fragment in base pairs.

Results

Although all of the samples produced good STR results (example in Figure 2), finding matching alleles at each locus among the 24 samples proved to be tedious. To facilitate this analysis, we determined a sample allele score that was simply the sum of the allelic designations for all the autosomal STRs of a sample (Figure 3A; for details, see appendix). Finding matching scores in sample allele calls among the 24 unknown samples then became very simple, since each sample was reduced to a single representative numerical value. Of course, once putative matching pairs were identified, the results were validated by confirming that the individual allele calls at each locus were identical. Once the samples were unblinded and checked, the results were found to be 100% accurate in matching the donors to their iPSCs (Figure 3B). This procedure was repeated once with another set of blinded iPSC samples from 12 donors, and again, exact matching of donor to iPSC sample was confirmed.

A

Sample A	Panel	Locus	Dye	Allele 1	Allele 2	Sample B	Panel	Locus	Dye	Allele 1	Allele 2	Sample A allele score	Sample B allele score
Sample 1	Identifiler_v2	AMEL	R	X	Y	**Sample10**	Identifiler_v2	AMEL	R	X	Y		
Sample 1	Identifiler_v2	CSF1PO	B	10	12	**Sample10**	Identifiler_v2	CSF1PO	B	10	12		
Sample 1	Identifiler_v2	D13S317	G	13	13	**Sample10**	Identifiler_v2	D13S317	G	13	13		
Sample 1	Identifiler_v2	D16S539	G	9	12	**Sample10**	Identifiler_v2	D16S539	G	9	12		
Sample 1	Identifiler_v2	D18S51	Y	14	15	**Sample10**	Identifiler_v2	D18S51	Y	14	15		
Sample 1	Identifiler_v2	D19S433	Y	12	14	**Sample10**	Identifiler_v2	D19S433	Y	12	14		
Sample 1	Identifiler_v2	D21S11	B	30.2	31.2	**Sample10**	Identifiler_v2	D21S11	B	30.2	31.2		
Sample 1	Identifiler_v2	D2S1338	G	23	25	**Sample10**	Identifiler_v2	D2S1338	G	23	25		
Sample 1	Identifiler_v2	D3S1358	G	16	18	**Sample10**	Identifiler_v2	D3S1358	G	16	18		
Sample 1	Identifiler_v2	D5S818	R	11	12	**Sample10**	Identifiler_v2	D5S818	R	11	12		
Sample 1	Identifiler_v2	D7S820	B	11	12	**Sample10**	Identifiler_v2	D7S820	B	11	12		
Sample 1	Identifiler_v2	D8S1179	B	13	13	**Sample10**	Identifiler_v2	D8S1179	B	13	13		
Sample 1	Identifiler_v2	FGA	R	23	24	**Sample10**	Identifiler_v2	FGA	R	23	24		
Sample 1	Identifiler_v2	TH01	G	6	6	**Sample10**	Identifiler_v2	TH01	G	6	6		
Sample 1	Identifiler_v2	TPOX	Y	8	11	**Sample10**	Identifiler_v2	TPOX	Y	8	11		
Sample 1	Identifiler_v2	vWA	Y	17	19	**Sample10**	Identifiler_v2	vWA	Y	17	19	453.4	453.4
Sample11	Identifiler_v2	AMEL	R	X	Y	**Sample15**	Identifiler_v2	AMEL	R	X	Y		
Sample11	Identifiler_v2	CSF1PO	B	12	13	**Sample15**	Identifiler_v2	CSF1PO	B	12	13		
Sample11	Identifiler_v2	D13S317	G	11	11	**Sample15**	Identifiler_v2	D13S317	G	11	11		
Sample11	Identifiler_v2	D16S539	G	10	15	**Sample15**	Identifiler_v2	D16S539	G	10	15		
Sample11	Identifiler_v2	D18S51	Y	13	15	**Sample15**	Identifiler_v2	D18S51	Y	13	15		
Sample11	Identifiler_v2	D19S433	Y	15	15.2	**Sample15**	Identifiler_v2	D19S433	Y	15	15.2		
Sample11	Identifiler_v2	D21S11	B	29	33.2	**Sample15**	Identifiler_v2	D21S11	B	29	33.2		
Sample11	Identifiler_v2	D2S1338	G	19	24	**Sample15**	Identifiler_v2	D2S1338	G	19	24		
Sample11	Identifiler_v2	D3S1358	G	16	17	**Sample15**	Identifiler_v2	D3S1358	G	16	17		
Sample11	Identifiler_v2	D5S818	R	10	12	**Sample15**	Identifiler_v2	D5S818	R	10	12		
Sample11	Identifiler_v2	D7S820	B	11	12	**Sample15**	Identifiler_v2	D7S820	B	11	12		
Sample11	Identifiler_v2	D8S1179	B	10	14	**Sample15**	Identifiler_v2	D8S1179	B	10	14		
Sample11	Identifiler_v2	FGA	R	24	25	**Sample15**	Identifiler_v2	FGA	R	24	25		
Sample11	Identifiler_v2	TH01	G	7	7	**Sample15**	Identifiler_v2	TH01	G	7	7		
Sample11	Identifiler_v2	TPOX	Y	8	11	**Sample15**	Identifiler_v2	TPOX	Y	8	11		
Sample11	Identifiler_v2	vWA	Y	18	19	**Sample15**	Identifiler_v2	vWA	Y	18	19	456.4	=SUM(L20:M34)

B

Sample (DNA from blood or iPSC)	Well position	Allele score
29	E1	415.2
31	G1	415.2
36	D2	417.5
38	F2	417.5
30	F1	431.2
40	H2	431.2
26	B1	435.3
47	G3	435.3

Sample (DNA from blood or iPSC)	Well position	Allele score
42	B3	436.2
45	E3	436.2
39	G2	436.4
48	H3	436.4
25	A1	440.4
44	D3	440.4
28	D1	444.5
32	H1	444.5

Sample (DNA from blood or iPSC)	Well position	Allele score
43	C3	448.3
46	F3	448.3
33	A2	448.5
41	A3	448.5
35	C2	451.5
37	E2	451.5
34	B2	459.2
27	C1	459.2

Figure 3. Method for quickly finding matching samples among unknowns. (A) Part of the data showing STR genotypes of four different blinded samples. A sample allelic score can be calculated by summing the numerical designation of each allele present. Matched pairs among a small set of unknown samples can then be determined quickly by finding identical allele scores. Pairing should be verified for matched samples at each allele, since there are several ways to produce an identical allele score. **(B)** Sample allelic scores were successfully used to match 12 blinded blood donor–iPSC sample pairs.

Sample authentication using STR profiling was also verified using NUCLEIC-CARD devices and the CLA IdentiFiler Direct kit. For these experiments, Dr. Wang again collected blood samples from the donors, but instead of purifying gDNA he spotted a fraction of the blood cells (40,000 cells) onto a NUCLEIC-CARD device and dried it at room temperature. iPSCs were obtained from the same samples, and aliquots of those cells were also spotted onto cards and dried at room temperature overnight. Once all samples were collected, the cards were sent to our laboratories for analysis. Punches (1.2 mm) were taken from the cards and analyzed using the CLA IdentiFiler Direct kit without further manipulation. The PCR products were analyzed on Applied Biosystems™ SeqStudio™ and 3500 Genetic Analyzers. The profiles obtained on these two instruments were identical; in this case, the samples were not blinded, so donor–iPSC matching pairs could be validated (Figure 4). STR genotyping on the SeqStudio and 3500 Genetic Analyzers confirmed that the samples were indeed collected from the same individual; however, the 4-capillary injection throughput of the SeqStudio Genetic Analyzer means that the samples can be analyzed immediately, without having to wait for enough samples to fill larger-capacity capillary arrays.

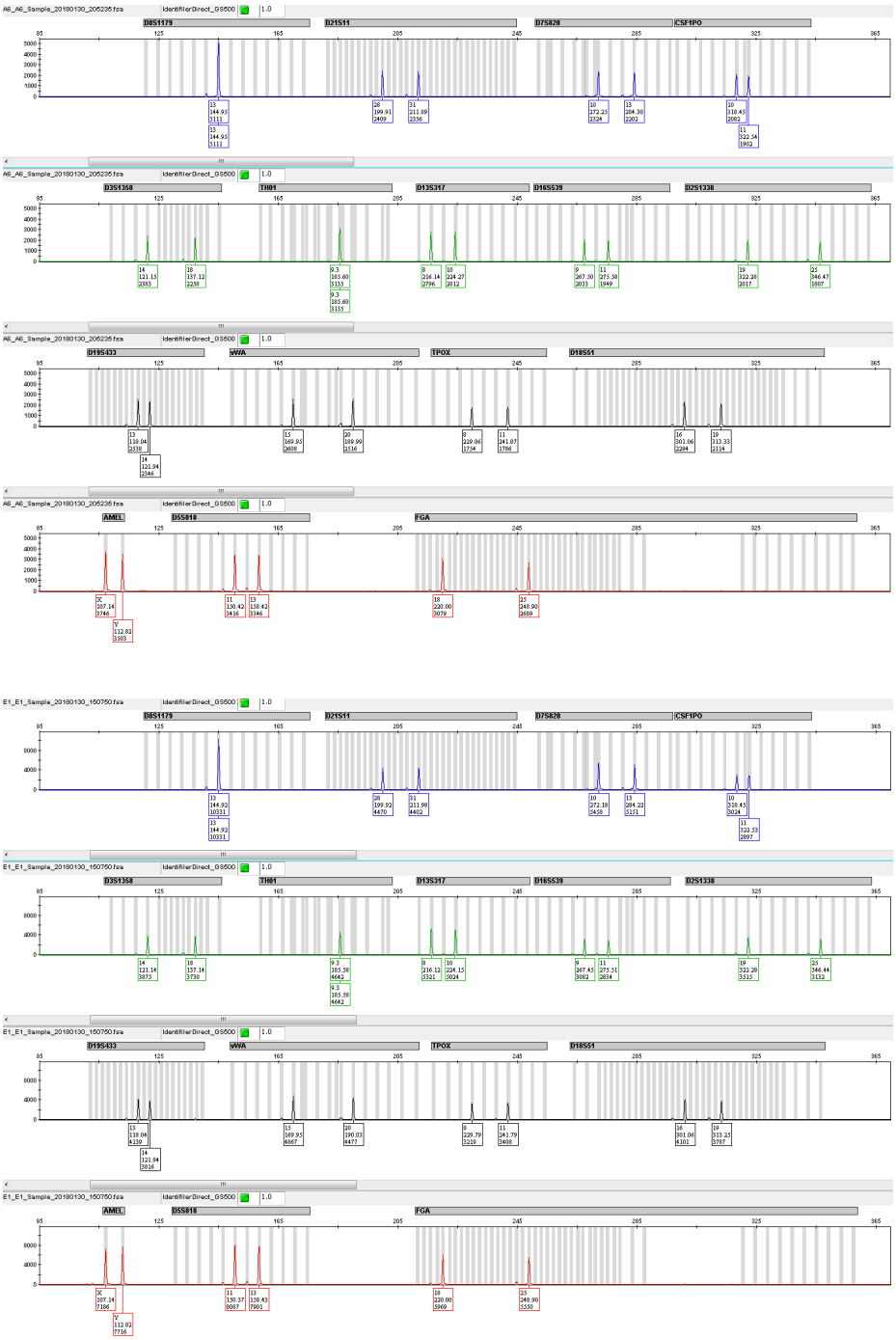


Figure 4. Analysis of matching samples on the NUCLEIC-CARD device and SeqStudio Genetic Analyzer. Blood (upper four sets of traces) and iPSCs (lower four sets of traces) were analyzed using the NUCLEIC-CARD device and CLA IdentiFiler Direct kit. The allelic profile is identical in the two samples, verifying that the samples came from the same individual. Note that these traces were collected using the SeqStudio Genetic Analyzer, but samples run on other instruments produced data of identical quality.

Conclusions

STR genotyping can rapidly confirm that samples of donor blood and iPSCs are from the same individual. Investigators can choose between analyzing purified gDNA or samples dried onto the NUCLEIC-CARD device, depending on the workflow that is most convenient for their lab. In addition, Thermo Fisher offers CE instruments to analyze these samples to suit the throughput of all laboratories—from the high-throughput Applied Biosystems™ 3730 DNA Analyzer and the flagship 3500 Genetic Analyzer, to the convenient and easy-to-use SeqStudio Genetic Analyzer. Together, these options provide choices that are designed to meet the needs of any stem cell research laboratory.

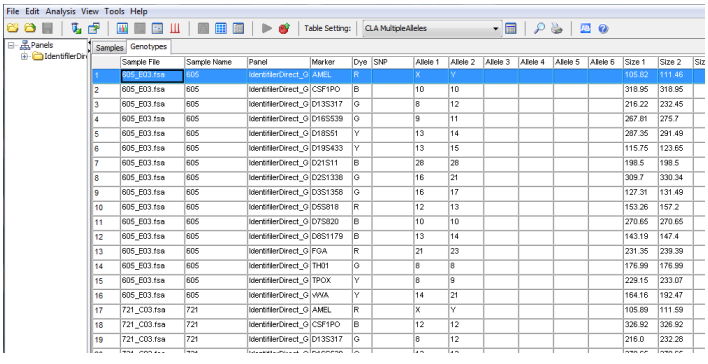
References

1. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryo and adult fibroblast cultures by defined factors. *Cell* 126:663-676.
2. Shi Y et al. (2017) Induced pluripotent stem cell technology: a decade of progress. *Nat Rev Drug Discov* 16:115-130.
3. Thermo Fisher Scientific Inc. (2017) Authenticating human cell lines using the IdentiFiler kits and capillary electrophoresis platforms. (Application note) Request download at **thermofisher.com/us/en/home/global/forms/life-science/seqstudio-application-notes.html**

Appendix—method for determining sample allele score using Microsoft™ Excel™ format

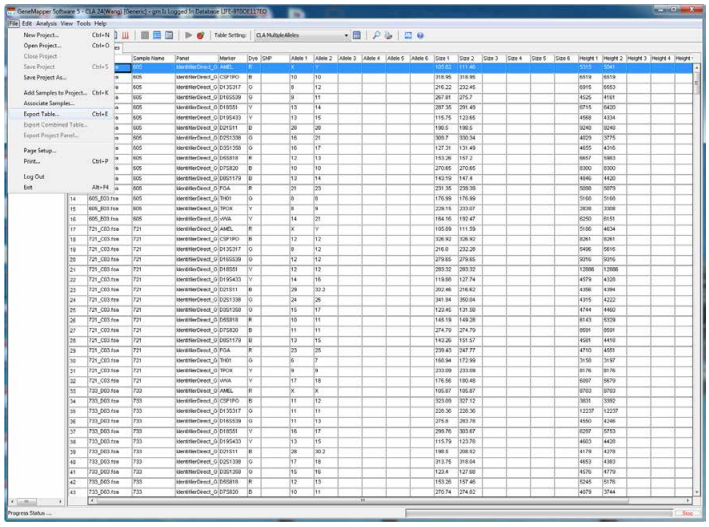
1. Using GeneMapper Software 6, process data to identify alleles present in all samples.

2. Click on the “Genotypes” tab in the results table.



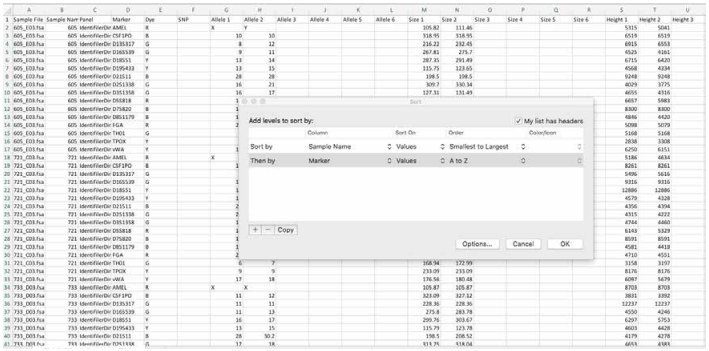
	Sample File	Sample Name	Panel	Marker	Dye	SNP	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 6	Size 1	Size 2
1	605_E03.fsa	605	IdentifilerDirect_G	AMEL	R		X	Y					105.82	111.46
2	605_E03.fsa	605	IdentifilerDirect_G	CSF1PO	B		10	10					318.95	318.95
3	605_E03.fsa	605	IdentifilerDirect_G	D13S317	G		8	12					216.22	232.45
4	605_E03.fsa	605	IdentifilerDirect_G	D16S539	G		9	11					267.81	275.7
5	605_E03.fsa	605	IdentifilerDirect_G	D18S51	Y		13	14					287.35	291.49
6	605_E03.fsa	605	IdentifilerDirect_G	D19S433	Y		13	15					115.75	123.65
7	605_E03.fsa	605	IdentifilerDirect_G	D21S11	B		28	28					198.5	198.5
8	605_E03.fsa	605	IdentifilerDirect_G	D2S1338	G		16	21					309.7	330.34
9	605_E03.fsa	605	IdentifilerDirect_G	D3S1358	G		16	17					127.31	131.49
10	605_E03.fsa	605	IdentifilerDirect_G	D5S818	R		12	13					153.26	157.2
11	605_E03.fsa	605	IdentifilerDirect_G	D7S820	B		10	10					270.65	270.65
12	605_E03.fsa	605	IdentifilerDirect_G	D8S1179	B		13	14					143.19	147.4
13	605_E03.fsa	605	IdentifilerDirect_G	FGA	R		21	23					231.35	239.39
14	605_E03.fsa	605	IdentifilerDirect_G	TH01	G		8	8					176.99	176.99
15	605_E03.fsa	605	IdentifilerDirect_G	TPOX	Y		8	9					229.15	233.07
16	605_E03.fsa	605	IdentifilerDirect_G	vWA	Y		14	21					164.16	192.47
17	721_C03.fsa	721	IdentifilerDirect_G	AMEL	R		X	Y					105.89	111.59
18	721_C03.fsa	721	IdentifilerDirect_G	CSF1PO	B		12	12					326.92	326.92
19	721_C03.fsa	721	IdentifilerDirect_G	D13S317	G		8	12					216.0	232.28

3. From the “File” menu, select “Export Table.”

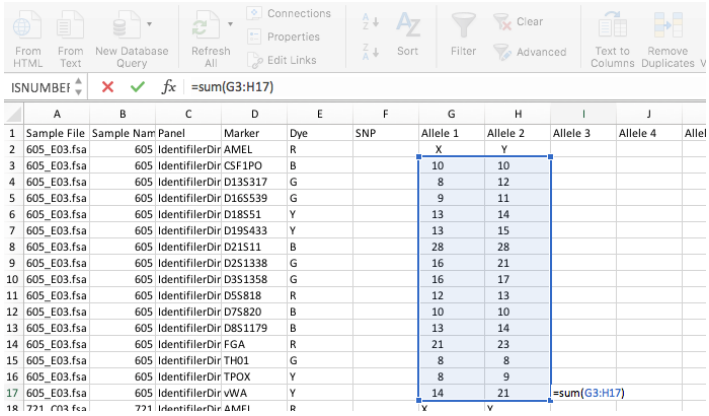


4. Find the .txt file for the table and open it in Excel software.

5. Select all rows. Sort rows by “Sample Name” and “Marker.”



6. For each sample, calculate the sum of all numerical values in the “Allele 1” and “Allele 2” columns. This sum is the sample allele score.



ISNUMBEI =sum(G3:H17)

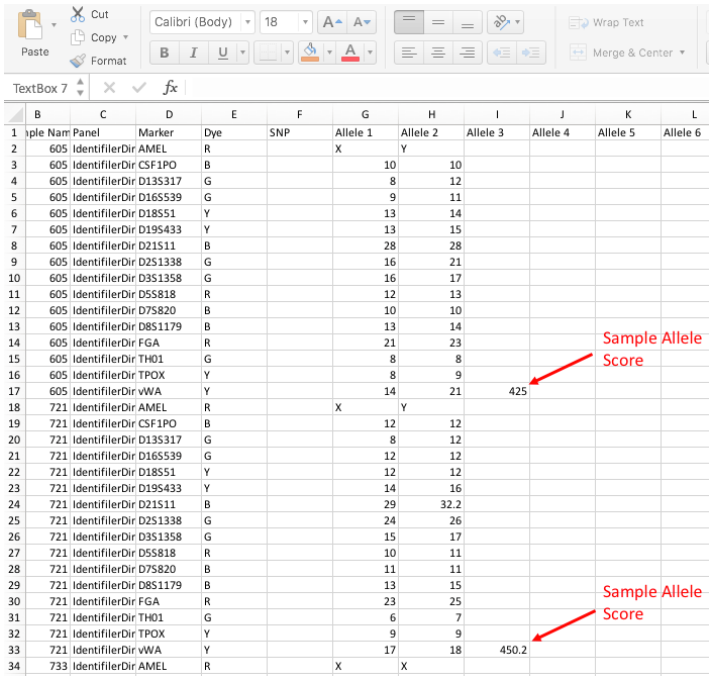
	A	B	C	D	E	F	G	H	I	J
1	Sample File	Sample Nam	Panel	Marker	Dye	SNP	Allele 1	Allele 2	Allele 3	Allele 4
2	605_E03.fsa	605	IdentifilerDir	AMEL	R		X	Y		
3	605_E03.fsa	605	IdentifilerDir	CSF1PO	B		10	10		
4	605_E03.fsa	605	IdentifilerDir	D13S317	G		8	12		
5	605_E03.fsa	605	IdentifilerDir	D16S539	G		9	11		
6	605_E03.fsa	605	IdentifilerDir	D18S51	Y		13	14		
7	605_E03.fsa	605	IdentifilerDir	D19S433	Y		13	15		
8	605_E03.fsa	605	IdentifilerDir	D21S11	B		28	28		
9	605_E03.fsa	605	IdentifilerDir	D2S1338	G		16	21		
10	605_E03.fsa	605	IdentifilerDir	D3S1358	G		16	17		
11	605_E03.fsa	605	IdentifilerDir	D5S818	R		12	13		
12	605_E03.fsa	605	IdentifilerDir	D7S820	B		10	10		
13	605_E03.fsa	605	IdentifilerDir	D8S1179	B		13	14		
14	605_E03.fsa	605	IdentifilerDir	FGA	R		21	23		
15	605_E03.fsa	605	IdentifilerDir	TH01	G		8	8		
16	605_E03.fsa	605	IdentifilerDir	TPOX	Y		8	9		
17	605_E03.fsa	605	IdentifilerDir	vWA	Y		14	21	=sum(G3:H17)	
18	721_C03.fsa	721	IdentifilerDir	AMEL	R		X	Y		

applied
biosystems

7. Finding the identical sample allele score can identify samples from the same individual.

8. Verify that the samples are identical by confirming genotypes of all alleles at all loci.

Note: The same method described above can be used with the new GeneMapper Software 6.



	B	C	D	E	F	G	H	I	J	K	L
1	ple Nam	Panel	Marker	Dye	SNP	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 6
2	605	IdentifilerDir	AMEL	R		X	Y				
3	605	IdentifilerDir	CSF1PO	B		10	10				
4	605	IdentifilerDir	D13S317	G		8	12				
5	605	IdentifilerDir	D16S539	G		9	11				
6	605	IdentifilerDir	D18S51	Y		13	14				
7	605	IdentifilerDir	D19S433	Y		13	15				
8	605	IdentifilerDir	D21S11	B		28	28				
9	605	IdentifilerDir	D2S1338	G		16	21				
10	605	IdentifilerDir	D3S1358	G		16	17				
11	605	IdentifilerDir	D5S818	R		12	13				
12	605	IdentifilerDir	D7S820	B		10	10				
13	605	IdentifilerDir	D8S1179	B		13	14				
14	605	IdentifilerDir	FGA	R		21	23				
15	605	IdentifilerDir	TH01	G		8	8				
16	605	IdentifilerDir	TPOX	Y		8	9				
17	605	IdentifilerDir	vWA	Y		14	21	425			
18	721	IdentifilerDir	AMEL	R		X	Y				
19	721	IdentifilerDir	CSF1PO	B		12	12				
20	721	IdentifilerDir	D13S317	G		8	12				
21	721	IdentifilerDir	D16S539	G		12	12				
22	721	IdentifilerDir	D18S51	Y		12	12				
23	721	IdentifilerDir	D19S433	Y		14	16				
24	721	IdentifilerDir	D21S11	B		29	32.2				
25	721	IdentifilerDir	D2S1338	G		24	26				
26	721	IdentifilerDir	D3S1358	G		15	17				
27	721	IdentifilerDir	D5S818	R		10	11				
28	721	IdentifilerDir	D7S820	B		11	11				
29	721	IdentifilerDir	D8S1179	B		13	15				
30	721	IdentifilerDir	FGA	R		23	25				
31	721	IdentifilerDir	TH01	G		6	7				
32	721	IdentifilerDir	TPOX	Y		9	9				
33	721	IdentifilerDir	vWA	Y		17	18	450.2			
34	733	IdentifilerDir	AMEL	R		X	X				

Ordering information

Product	Quantity	Cat. No.
SeqStudio Genetic Analyzer	1 instrument	A34274
SeqStudio Analysis Software	1	44443764
SeqStudio Cartridge v2	1,000 reactions	A41331
SeqStudio Starter Kit	1 kit	A35000
SeqStudio Full-Day SmartStart Training	1	A34684
SeqStudio 8 Flex Genetic Analyzer	1 system	A53627
SeqStudio 24 Flex Genetic Analyzer	1 system	A53630
CLA IdentiFiler Plus PCR Amplification Kit	50 reactions	A47624
CLA IdentiFiler Plus PCR Amplification Kit	200 reactions	A65672
CLA IdentiFiler Direct PCR Amplification Kit	200 reactions	A65908
CLA GlobalFiler PCR Amplification Kit	200 reactions	A65909
NUCLEIC-CARD Matrix, 1 spot	100 cards	4473973
NUCLEIC-CARD COLOR Matrix, 4 spots	50 cards	4473978
RecoverAll Total Nucleic Acid Isolation Kit for FFPE	40 reactions	AM1975
GeneMapper Software 6	1 license	4475074
GeneScan 600 LIZ Dye Size Standard v2.0	800 reactions	4408399

Find out more at
thermofisher.com/seqstudioapplications

ThermoFisher SCIENTIFIC