Use of Spike-ins for Sample Tracking in Agilent Array CGH

Application Note

Authors

Srirangan Sampath, Ph.D., FACMG and Heather Klein
PreventionGenetics

Iman Kishawi, Ph.D.
Agilent Technologies, Inc.

Abstract

This Application Note describes a method that allows sample tracking through introduction of DNA spike-ins to human DNA samples before wet lab handling. Adding the spike-ins to samples ensures the identity of the sample throughout the process, and reduces sample mix-up in high-throughput DNA microarrays. In principle, this method can be applied to many other molecular genetic methodologies including NGS.

DNA spike-ins contain one or more PCR-generated amplicons that are approximately 400 bp in length. The amplicons are designed to span sequences of unique and nondisease associated regions in the human genome. Agilent CytoGenomics software identifies sample identity by detecting the signal intensity of spike-in probes in individual arrays.

Agilent Technologies

Introduction

Many laboratories process multiple samples on any given day, and human error can contribute to mistakes in the processing and reporting of results for samples. Sample mix-up errors constitute the most serious pre-analytic mistakes[1-2]. In addition, the chances of a specimen mix-up during the analysis process increases with the number of experimental steps (multistep or multiday experiments).

DNA microarray applications, such as array comparative genomic hybridization (aCGH), are a multistep process. The wet lab portion includes restriction-digestion (for arrays containing SNP probes), fluorescent-labeling of genomic DNA (gDNA), cleanup and quantification of labeled gDNA, and hybridization of labeled gDNA to microarray slides. Sample mix-up at any of these stages can lead to erroneous interpretation of results. Human error can also occur with data analysis if sample attributes are associated with the wrong array.

A spike-in is a calibrated reference standard made up of one or more DNA amplicons of known concentration and identity. Amplicons are added in large excess to each DNA sample being tested. Fluorescent labeling of the spiked amplicon generates a bright fluorescent signal after hybridizing to probes that correspond to the genomic loci of spike-ins. Agilent CytoGenomics software v4 or higher reports the identity of the spike-in in each sample. This reporting is done by calculating the $\log_2$ ratio of the sample signal to the reference signal at any given probe. The software identifies the identity of the spike-in based on a default $\log_2$ ratio equal to, or greater than one. A default threshold of one corresponds to two times more in a spiked DNA sample versus a reference DNA sample where no spike-ins were added. The spike-in identity is entered into the CytoGenomics software by the user, and the software reports if the spike-in identity does or does not match the user-inputted sample attribute.

Methods

Reagents and supplies needed

Material	Vendor	Part number
KAPA HS MasterMix	Kapa Biosystems	KK1606
Oligonucleotide primer mixes	IDT	Custom (25 nmole DNA oligo)
Nanodrop	Thermo Scientific	
Human Reference Genomic DNA	Agilent European Male	5190-4370
1X TE buffer	IDT	N/A
Eppendorf Mastercycler ProS	Eppendorf	950030020
Agilent TapeStation 2200	Agilent Technologies	G2964AA
Agilent D1000 ScreenTape	Agilent Technologies	5067-5582

Spike-in identity

The 50 unique amplicons (Table 1) are PCR-generated using the forward and reverse primers shown in the table. The probes corresponding to each amplicon (Table 1) are printed on the CGH arrays. The software recognizes and identifies them by their probe names. Individual probes that hybridize to the spike-ins are printed 3–5 times on the array. The data presented in Figure 1 show the average $\log_2$ ratio of spike-ins from replicated probes. Spike-in probes are part of the HD Agilent probe database. Spike-in controls can be added to any array by selecting the spike-in probe group in SureDesign, and replicating it 3–5 times depending on available space on the array.

Table 1. Forward and reverse primers for each of the 50 DNA amplicons that are used as spike-ins. The expected sizes for each amplicon, and the Agilent probe names are provided.

Band	Probe_name	FP_seq	RP_seq	Amplicon size
1q32.1	A_16_P35422305	CATTCAATGTTACCTTCCAA	AGCCACTCATTTTTAACACC	487
	A_16_P15405423	AATGTACAAACTGGAAACC	TTTCTCTCTCACAGGAATAA	433
	A_16_P15405605	TTCTCCTTAAGCTGGTTTAA	GGAAAACTGGTTCAGTAGAC	500
2q31.3	A_18_P13533801	TTATCTCTTGCCTTCTTTTC	AGACAATAGACTGCAAATCC	503
	A_16_P15950347	CCTACCACCTTGTGAAGA	GCACTGTACACTGAAAATTG	425
3q13.32	A_16_P16375641	CCAAGATGCTTAGAAAAATG	GGCATCAAATCTCTGGTT	444
	A_16_P16375863	AGCTTCAGGTAAGTTTTTGG	ACACAGAAGCATCTGTTCTT	451
	A_18_P14352811	TTACTCCCAAAGTCTATTGC	TGTAACCAGTGAGATAAGGC	475
4q21.22	A_18_P14767506	GGTTGTGGGTTTACATTCAA	AAGGGTGACCTAGAGCACA	428
	A_18_P14771686	GAATCTGAAGACAGTCCAGT	TTTAACAACCCTTCACTTC	516
	A_16_P36800196	GAACTCAGTGCATGAACAAG	TCTGAGGACAGAGCTGTATC	333
5q14.2	A_18_P15262957	GCTGTTCTCCAAACACTAAT	TTACATCGAGCTAACCTGTC	431
	A_18_P15267322	TTCATTCAAGAGCACAGAAA	ACAAACTAGGGCAAAACAG	429
	A_16_P37250366	GTCATTTAAATTTTCTTGGC	TAATCTAGGTGGTGGTTACC	487
6q14.3	A_18_P15726331	CAGAAGCTCTTTGGTTTAAT	GCTGCTTTACAAAATACTGC	413
	A_18_P15730616	CCTTGGGTTCAATGAAACCT	GGCACAGAAGAAAGGGACAG	395
	A_14_P101769	GAGCCCACCCTATAAACATT	CAGATCTTGGCAAGACTCAG	434
8q23.3	A_16_P18442814	CCTTTAGGAATTACCCAGTA	TTAGGAAGCAGTAATGAGG	465
	A_16_P18442955	GTTGACATGCTGGGAAATC	TGCCCATACAGCTGTTCTCT	364
	A_16_P01996624	AGTTGTTTCAGAAGGCCTCT	AGTGACAGGAATGAGGCTG	366
9q31.2	A_18_P16960445	GATGTGCACACATGGACC	AGACATCTGCTCAGGGAGTT	473
	A_16_P02153618	CAATCTAAAATCAAGGAAAT	ATACAATGTGACAGATGCTA	355
	A_16_P18730465	TTGTTTGCCTAGAAAACAGG	TCACTTTTTCAATCCCACTT	453
10q25.1	A_16_P19058209	GGATATTTCTGTCCATTGAT	CACAGTGGTATTAAACAAGG	472
	A_18_P10889615	GGCATTGTCTCTTCTTGCTA	TAACAGAACCCAAACTCGTG	473
	A_16_P19058540	TCATAATTACCATCATAGCA	AAATCAGCGATATTAATAGC	416
11q22.1	A_16_P19340901	GGCAAAAGAGTGAGACTC	CCTTCCCTTAATAGCCTAAG	468
	A_16_P19341057	AGATAGCCTGCTAGGATTTG	GTGATGGATACCCCATTTAC	413
	A_18_P11204159	CCATTACTTCATTTCTGCCT	TTGGTCACAGAACATTGAAT	411
14q21.3	A_16_P20035587	TTATTTCCTTGCTATAGTGG	GAAACAACTCAGAAAGCTAG	364
	A_16_P40178245	TGGAACAAGAACTTGTGAGC	GCCAAAATGCATAATTTCA	471
	A_18_P11991646	TCTCTGTGCATCACCCTATT	GGCAAAGTTCAAGTCTCTGG	404
15q21.1	A_18_P12205990	GATTCCCGTCTGTGTGGTCT	TCTTATTTTTGACTCTGAGATTGTGG	251
	A_18_P12206420	TGGTGCCTAAGAAAGACATTCA	TTCCTTCCCAAAGTGGCTTA	287
	A_16_P40388779	TCCCAAAGAATTGTCAATGG	TTTGGTTTCCCATCAAGGTC	300
16q21	A_18_P12484550	AACATTGAGGATGGAGCAG	CAAAAATGGGGTCTTATCTG	431
	A_18_P12480335	TGTCAATGTAAAAATGCAAG	ACAGCAAAGGTCTTCTATTC	436
	A_16_P20492408	GTAACAGAGGGCTAGTCTTG	AGTTTGAGTGGTGACTGATT	445
18q12.2	A_18_P12842528	GCCTTTATTGATTCCTGCTT	TGGGGCTCACCTAAAAGTT	522
	A_18_P12844310	GGTGTCTCATCAGCAGGG	GATTCCTATTTTGCTCAGGC	459
	A_18_P12846883	CCTTAGGTCCACATGACATG	GTGCAGTCAAGCTTGAAGTC	452
19q12	A_16_P20991672	AAGTTTCCTTTTGATTAGGT	ACTAGGAGAAAATATTTGCC	422
	A_16_P03441288	AACAAACAAACAAAAAACGC	CACAGCTTCTTCATCTGACC	412
	A_18_P13021425	GAAGAGGTCACATATGTGAG	ATTACCAAACAAGATTGTTG	397
20q12	A_18_P13795069	GCATTCCAGCTTAAAGGA	ACATTTAAAAGATGGTGCAA	408
	A_18_P13792864	GCATAGCCTTGAAAAAATCC	TGTGGAATGGATAAATTGCC	438
	A_16_P21129167	TTCATGGACTCAATAGTCCC	TGAGCTTGAAGAGGAAAGAG	433
21q21.2	A_16_P41413387	CCTGTCAGCAGGTAGAAGGA	TCTCATGAAGTAGGCAGAGGAA	256
	A_16_P21216858	TTTTGCAAGTAAAGCAATTTTAGG	GCTCTGCTGGACACCTTGAT	299
	A_18_P13889716	CAAATTGATGAAATCCTTATTTGC	AAACCATCAAATACAGGAAAAGAAA	286

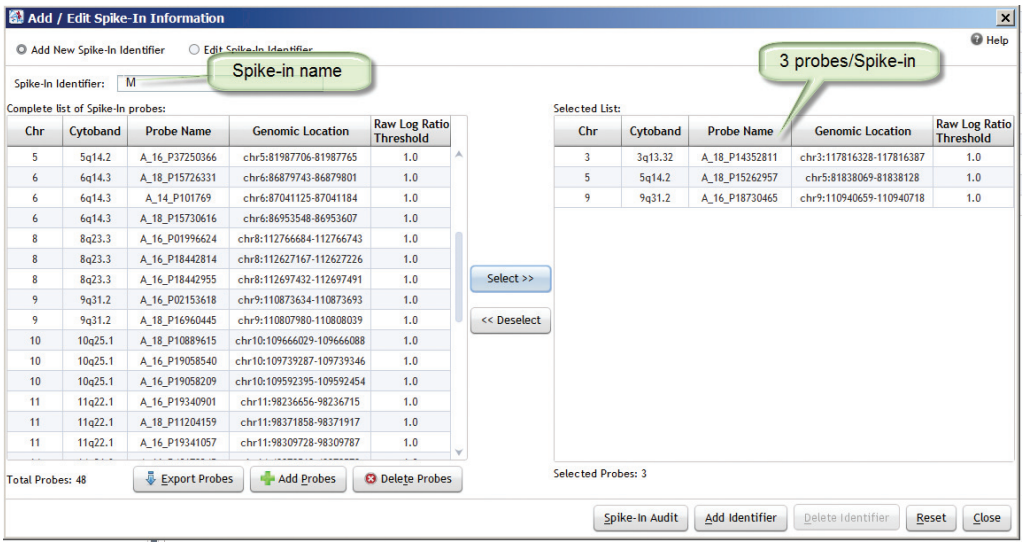

Figure 1. Spike-in M is assigned three probes. Three different amplicons can be spiked into a DNA sample. CytoGenomics software can identify the identity of the three amplicons as Spike-in M.

Spike-in generation

DNA spike-ins are amplicons generated from PCR of human genomic sequences using a reference human genome (for example, Agilent European Male). Tables 2 and 3 provide the conditions of the PCR reactions. The amplification efficiency differs for each of the 50 amplicons, and may need some level of PCR optimization in the hands of users. The PCR efficiency and identity of the amplicons were checked by running 2-µL aliquots of all amplicons on an Agilent TapeStation 2200 using Agilent D1000 ScreenTape[3]. The quality and identity of each amplicon were confirmed by the presence of a correct sized single band. The PCR products were purified using the Qiagen QIAquick PCR Purification Kit (catalog # 28106) using the manufacturer's protocol.

Depending on the amount of spike-in needed, each of the 50 amplicons may be amplified in replicates of four or five PCR reactions. Replicates from each amplicon were pooled before Qiagen PCR Purification. Pooled amplicons can be eluted with 30 µL of elution buffer. The purified amplicons were checked again on the TapeStation 2200, using the D1000 ScreenTape to ensure proper cleanup and size (Figure 2).

Table 2. Setup of PCR reactions.

Reagents	Volume/rxn (µL)
KAPA HS MasterMix	20.25
Amplicon F/R Primers, 5 µM	3.75
Template, 5 ng/µL	6
Total reaction volume	30

Table 3. PCR Conditions are listed below (37 cycles).

Conditions	Time
95 °C	3 minutes
95 °C	30 seconds
55 °C	30 seconds
72 °C	1 minute
72 °C	5 minutes

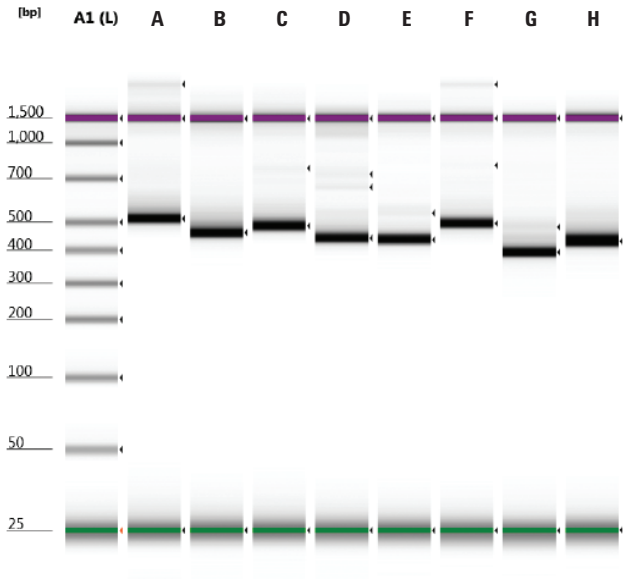

Figure 2. To ensure unique amplification of the target region, PCR amplicons are evaluated on the Agilent TapeStation 2200 using the Agilent D1000 ScreenTape. Each spike-in is given a unique identifier (in this case, alphabets spike-in A-H for eight amplicons).

The purified amplicons were quantified on a Nanodrop. When pools of 4–5 replicate amplicons were eluted in 30 µL of elution buffer, the concentration was usually >100 ng/µL; hence, the spike-in stock solution was normalized to 40 ng/µL in 1X TE buffer (pH 8). A spike-in working solution was made from the stock solution by diluting it ~1:2,000 with 1X TE buffer (pH 8). At this dilution, the spike-in probe fluorescent signal was significantly above background noise in an aCGH experiment. The spike-in working solution dilution can be determined empirically by performing microarray experiments at different dilutions.

Aliquots of the spike-in stock (40 ng/µL) can be stored at –20 °C indefinitely. Working solution (40 ng/µL diluted at ~1:2,000 with 1X TE buffer) can be stored at 4 °C for a week without appreciable degradation. For optimum performance, avoid repeated freeze-thaw cycles.

Using spike-ins in an aCGH experiment

Spike-in identifiers can be associated with the amplicon probe name in the CytoGenomics software. To identify a sample, one or more amplicons can be added to the sample before a wet lab procedure. An identifier can represent one probe name if one PCR amplicon is spiked into the sample. An identifier can represent more than one probe if multiple PCR amplicons are spiked into the sample. For example, Spike-in A represents a single probe, and spike-in M represents more than one probe (Figure 1). Two microliters of spike-in control is added to DNA sample before processing.

During the analysis setup in the CytoGenomics software, the user needs to assign a spike-in identifier to an array (or sample). The spike-in can be assigned from the drop down menu, or as part of the sample attribute file with a column header Spike-In. After data analysis, the CytoGenomics software reports the observed and expected value of each spike-in. Spike-in identity in each sample is reported as match or no match based on the user inputted spike-in identifier in the software (Figure 3). Additionally, the software reports the $\log_2$ ratio corresponding to each spike-in. Note that based on this procedure, 50 spike-ins can be generated using one spike-in amplicon per sample. This number can go much higher if combinations of spike-in amplicons are added to each sample.

This experiment describes using one amplicon per sample. For each experiment, two microliters of DNA spike-in working solution are added to the test sample gDNA before labeling with Cy5 (red signal). A different DNA spike-in is used for each test sample. Spike-ins are not added to the reference DNA (labeled with Cy3). Consequently, probes corresponding to spike-ins have a high red signal and low green signal.

Expected Spike...	Observed Spik...	Spike-In Result
A	A	Matched
B	B	Matched
C	C	Matched
E	E	Matched
H	G	No Match
G	H	No Match
A	A	Matched
B	B	Matched
C	C	Matched
A	D	No Match
B	A	No Match
D	B	No Match
C	C	Matched
D	D	Matched

Figure 3. Agilent CytoGenomics software identifies the identity of the spike-in controls.

Spike-in analysis

gProcessedSignal and rProcessedSignal values from each aCGH experiment were used for plotting.

Four plots are shown in Figure 4, each with a different spike-in added to the test DNA sample. For each sample, the $\log_2$ ratio of rProcessedSignal/gProcessedSignal for all 50 spikes (in trilpicate) are plotted. Each of the 50 spike-ins is represented by a different color. There is very little variation between triplicate values for all spike-ins.

As can be seen in Figure 4, the $\log_2$ ratio values differ for individual spike-ins despite the fact that the amplicons are diluted to the same concentration of 40 ng/µL.

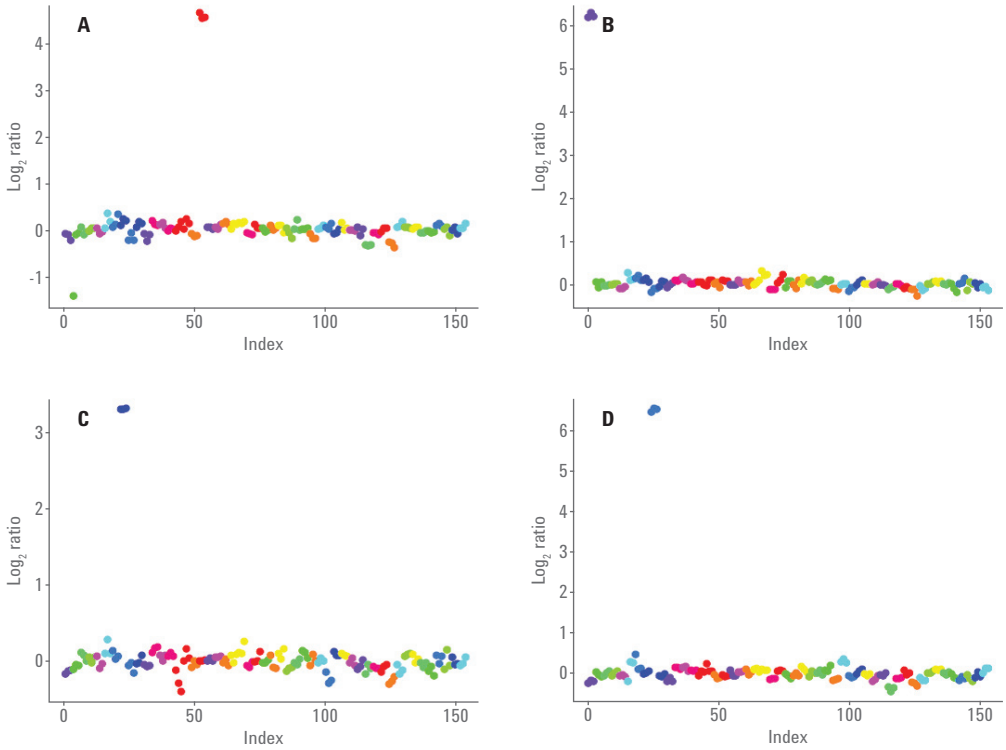

Figure 4. $\log_2$ ratio is plotted for all 50 amplicons. Greater than 1 $\log_2$ ratio value represents the added DNA spike.

Conclusion

In conclusion, 50 unique amplicons can be used as spike-ins for sample tracking. The spike-ins can be one or more amplicons that are introduced into DNA samples before wet lab processing. The $\log_2$ ratio of spike-ins should be equal to or higher than 1, under the dilution conditions described. The data are highly reproducible, and give the user confidence that there was no mix-up of samples during wet lab or analysis processing.

References

1. Douglas, J. A.; Boehnke, M.; Lange, K. A. Multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. *Am. J. Hum. Genet.* **2000**, *66(4)*, 1287-1297. doi: 10.1086/302861. [PMID:10739757]

2. Jun, G.; *et al.* Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* **2012**, *91(5)*, 839-848. doi: 10.1016/j.ajhg.2012.09.004. [PMID: 23103226]

3. Burrell, A.; Foy, C.; Burns, M. Applicability of Three Alternative Instruments for Food Authenticity Analysis:GMO Identification. *Biotechnology Research International* **2011**, *vol. 2011*, Article ID 838232, 8 pages, doi:10.4061/2011/838232

For more information and to learn more

www.genomics.agilent.com

Find an Agilent customer center in your country

www.agilent.com/chem/contactus

U.S. and Canada

800.227.9770
agilent_inquiries@agilent.com

Asia Pacific

adinquiry_aplsca@agilent.com

Europe

info_agilent@agilent.com

www.agilent.com/genomics

PR7000-0131

Published in the USA, March 15, 2016
5991-6661EN

Agilent Technologies