



STUDY REPORT

STUDY TITLE: Authenticating sample X

SPONSOR: Sponsor.

SPONSOR STUDY NUMBER: **Sample-Study-001**

SPONSOR REPRESENTATIVE: **Sample-Sponsor**

TELEPHONE: (111) 111-11111

EMAIL: email@email.com

TESTING FACILITY:

Crown Bioscience Suzhou
Room 303, Building A6, No. 218
Xinghu Street
Industrial Park
Suzhou City
Jiangsu Province, China

CROWN BIOSCIENCE STUDY NUMBER: 12345

STUDY DIRECTOR: Director Name.

TELEPHONE: (415) 444-4444

EMAIL: email@email.com

STUDY INITIATION DATE: **1 September 2020**

STUDY COMPLETION DATE: **September 15, 2020**

DATE OF DOCUMENT: **September 15, 2020**

DOCUMENT AUTHOR:

VERSION OF DOCUMENT: Version 1.0

TABLE OF CONTENTS

1. OVERVIEW	2
1.1. Specimen Description.....	2
1.2. Services Performed.....	2
2. RESULT SUMMARY.....	1
3. RESULTS	2
3.1. Mouse ratio.....	2
3.2. Matched sample	3
3.3. Heterogeneity ratio	4
3.4. Contaminant detection.....	5
3.5. Viral infection & mycoplasma contamination	6
3.6. Gender identification.....	7
3.7. Genetic admixture analysis	8
4. SIGNATURES.....	9
4.1. Sponsor Approval.....	9
4.2. Crown Bioscience (Suzhou) Approval	9
5. References	10

1. OVERVIEW

1.1. SPECIMEN DESCRIPTION

Species: patient-derived xenograft (PDX) tumor sample dissected from an immunodeficient mouse.
Number of Specimens: 1

QC-ID	Sample ID	Species
20200819-02699	GA9273-R1P2-20191119-R-37790	Human

1.2. SERVICES PERFORMED

Service: Cell Line and Model Authentication - Comprehensive

Service items:

- Sample identification by matching to standard PDX and cell line databases
- Human-mouse interspecies contamination quantification
- Sample genetic heterogeneity quantification
- Sample contamination check and contaminant identification (if the contaminant is already in the standard PDX and cell line databases)
- Mouse syngeneic model identification (if applicable)
- Mouse strain (if applicable)
- Viral infection detection
- Mycoplasma contamination detection
- Gender identification for human samples
- Genetic admixture analysis for human samples

2. RESULT SUMMARY

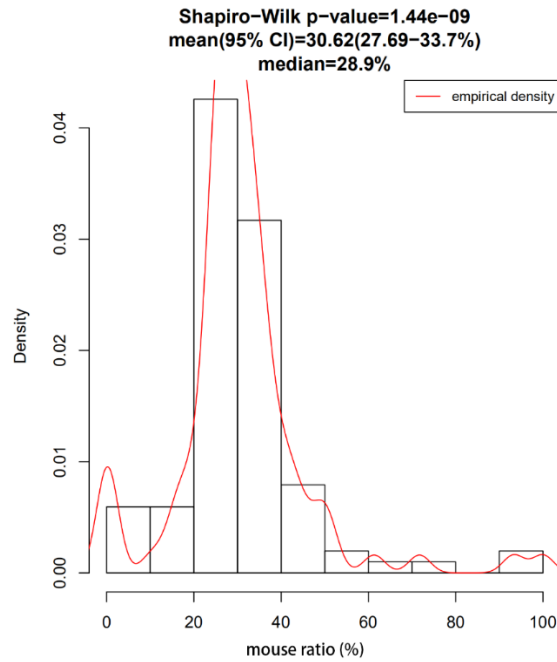
QC Item	Results
Mouse ratio	28.9%
Matched sample	GA9273
Heterogeneity ratio*	1.91%
Contaminant detected	None
Mouse syngeneic model	NA
Mouse strain	NA
Viral infection	EBV
Mycoplasma contamination	Negative
Gender	Male
Genetic admixture (CEU: CHB: YRI) **	0.002: 0.002: 0.996

*: See definition in Reference 1.

** : The three reference populations are: Han Chinese (CHB), Nigeria Yoruba (YRI) and Utah residents with Northern and Western European ancestry from the CEPH collection (CEU).

3. RESULTS

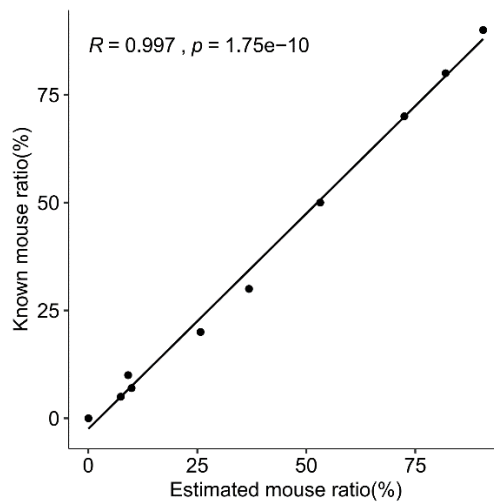
3.1. MOUSE RATIO



Comment: Mouse ratio is estimated based on 108 100-300bp human-mouse homologous segments that share identical flanking sequences and are amplified by same primers.

Conclusion: The mouse ratio in Sample GA9273-R1P2-20191119-R-37790 is 28.9%.

Technical note: This assay can estimate mouse ratio in a human-mouse mix with high accuracy (Figure 4A below, adopted from Reference [1]), and can reliably detect mouse ratio at ~0.1% [1].

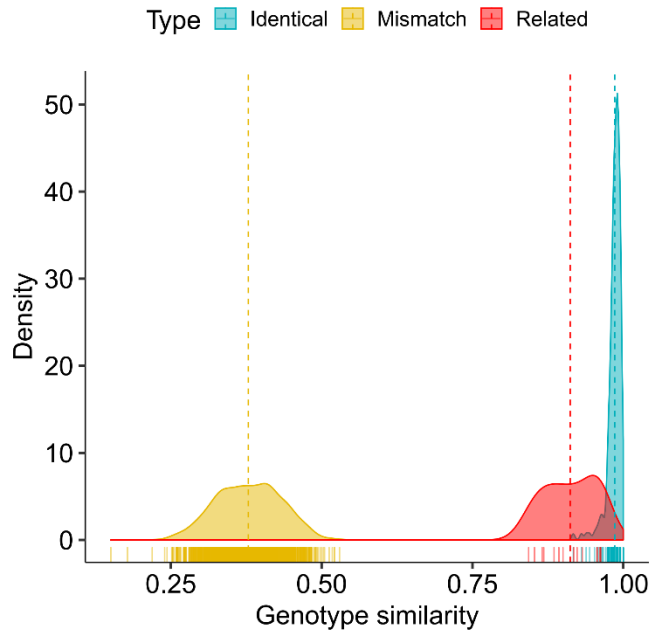


3.2. MATCHED SAMPLE

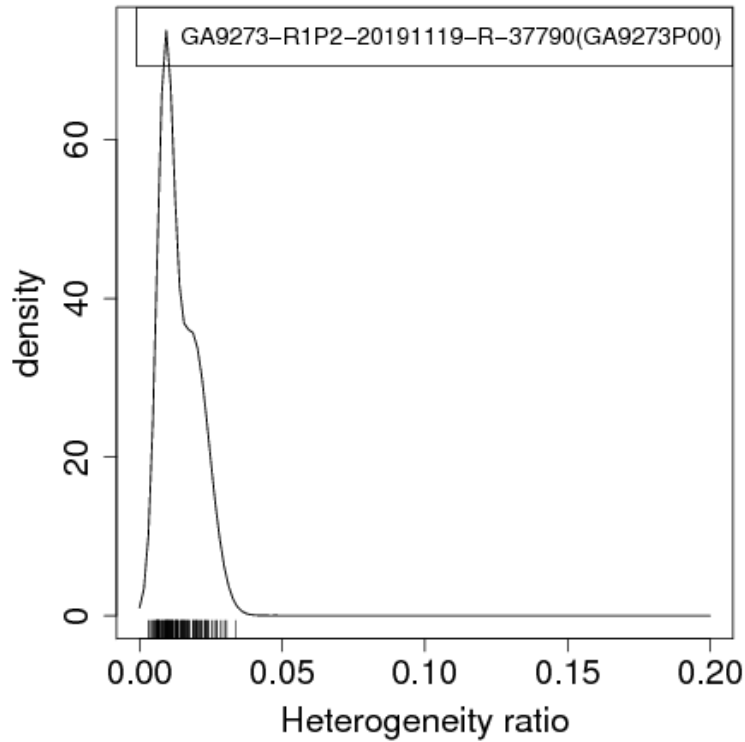
Reference	# Matched SNPs	# Nonmatched SNPs	Total SNPs	Genotype similarity (%)
GA9273P00	222	5	227	97.80
CR5076P00	111	98	209	53.11
LU9342P00	115	103	218	52.75
CL01123	111	100	211	52.61
LU11885P00	109	99	208	52.40

Conclusion: Sample GA9273-R1P2-20191119-R-37790 is GA9273 based on its high identity to GA9273P00—the standard reference sample of PDX model GA9273

Technical note: Genotype similarities calculated from >200 SNPs are always >90% between identical samples even at the presence of minor contaminant. In contrast, genotype similarities between unrelated samples are almost always below 50%. If two samples are related, for example by deriving from same parental sample, their genotype similarity can range from about 80% to close to 100%, depending on their genetic divergence. Below: Figure 1A from Reference [1].



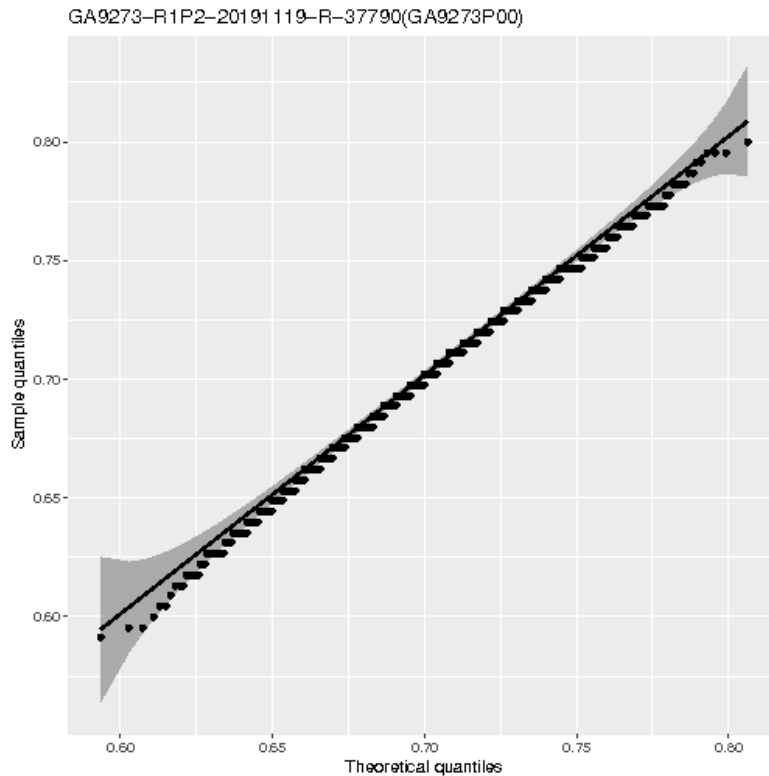
3.3. HETEROGENEITY RATIO



Conclusion: Sample GA9273-R1P2-20191119-R-37790 has a heterogeneity ratio 1.91% with the above distribution of heterogeneity ratios for 227 SNPs. This heterogeneity ratio is within the standard value of PDX models with ~30% mouse ratios (see Figure 1C of Reference 1).

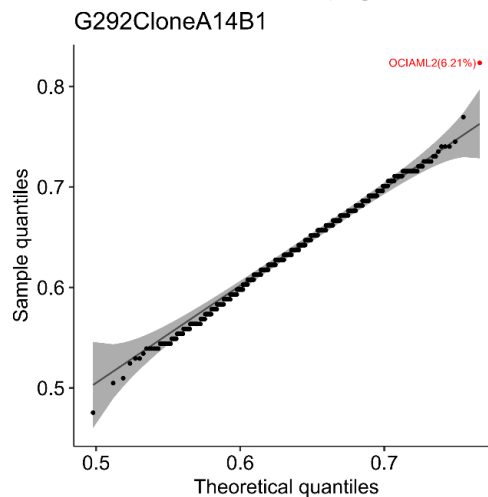
Technical note: Heterogeneity ratio is a quantitative measure for genetic heterogeneity of a sample by using the >200 SNPs and is defined in Table 1 of Reference [1]. Heterogeneity ratio comes from genetic heterogeneity and contamination. High heterogeneity ratio, as well as a two/three-modal distribution, usually indicates possible contamination. See Reference [1] for details.

3.4. CONTAMINANT DETECTION



Conclusion: No contaminant is detected for sample GA9273-R1P2-20191119-R-37790.

Technical note: In this quantile–quantile plot, each dot is a reference cell line; theoretical and sample quantiles were calculated from a beta distribution fitted to genotype similarities between the test sample and all reference samples. The 99% confidence band is shaded. A reference sample was denoted as contaminant if (1) it had the highest genotype similarities, (2) its genotype similarity was above the 99% confidence upper bound in the quantile–quantile graph and (3) its P-value was $<1.0E-6$ in the fitted beta distribution. An example for contaminant is shown below (Figure 3D from Reference [1]).



3.5. VIRAL INFECTION & MYCOPLASMA CONTAMINATION

Virus & Mycoplasma	Number of high coverage bases
CMV	0
EBV	428
HBV	0
HIV	0
HPV16	0
HPV18	0
Mycoplasma	0

Comment: If the read depth is > 1000, a base is accounted as a high coverage one.

CMV: Cytomegalovirus

EBV: Epstein-Barr virus

HBV: Hepatitis B virus

HIV: Human immunodeficiency virus

HPV: Human papillomavirus

Conclusion: Sample GA9273-R1P2-20191119-R-37790 is EBV positive, and CMV/HBV/HIV/HPV16/HPV18/Mycoplasma negative.

Technical note: Multiple pairs of primers were used to detect each virus, and one pair of universal primers were used to detect all mycoplasma species. See Reference [1] for details.

3.6. GENDER IDENTIFICATION

SNP-anchored Y chromosome segment	Read depth
hg19_chrY_14832620	2713
hg19_chrY_15467824	3723
hg19_chrY_15591537	4342

Comments:

(1) If the sum of read depths of the 3 Y-chromosome SNPs >1000, the sample is identified as “Male”; If the sum if <300, the gender is predicted as “Female”; If the estimated mouse ratio is over 95%, or the sum of depths is less than 1000 and more than 300, the gender is not assigned.

(2) Because Y chromosome is frequently lost in cancer genomes, it is advised to use microscope-based observation to confirm the gender prediction.

Conclusion: Sample GA9273-R1P2-20191119-R-37790 is male.

3.7. GENETIC ADMIXTURE ANALYSIS

Based on the genotyping data of 143 SNPs, the genetic admixture of this sample is 0.2% CHB, 0.2% CEU, and 99.6% YRI, where CHB stands for Han Chinese, CEU for Utah residents with Northern and Western European ancestry from the CEPH collection, YRI for Nigeria Yoruba.

Conclusion: Sample GA9273-R1P2-20191119-R-37790 is predominantly of YRI origin.

4. SIGNATURES

4.1. SPONSOR APPROVAL

Sponsor Representative

Date

4.2. CROWN BIOSCIENCE (SUZHOU) APPROVAL

Crown Bioscience Representative

Date

5. REFERENCES

1. Chen,X., Qian,W., Song,Z., Li,Q. and Guo,S. (2020) Authentication, characterization and contamination detection of cell lines, xenografts and organoids by barcode deep NGS sequencing. *NAR Genomics and Bioinformatics*, **2**, 3.
<https://academic.oup.com/nargab/article-abstract/2/3/lqaa060/5893932>