

版本号: 2017V1

# METHYL TARGET 项目结题报告

## 上海天昊生物科技有限公司

项目编号:

16B1212A

项目名称:

上海王明华—200 个样品

MethyTarget 富集测序

客户单位:

苏州大学

报告日期:

2017.4.26

上海天昊生物科技有限公司

上海天昊生物



上海市浦东新区康桥路787号9号楼  
400-065-6886  
[www.geneskies.com](http://www.geneskies.com)

[geneskies.com](http://geneskies.com)

## 目录

目录 .....	I
一、服务介绍与说明 .....	1
1、背景介绍 .....	1
2、样本要求 .....	1
3、数据量要求 .....	1
4、服务流程 .....	1
二、项目信息 .....	2
三、实验流程 .....	3
1、建库测序流程 .....	3
2、实验仪器与试剂 .....	3
3、实验步骤 .....	4
四、生物信息学分析流程 .....	5
1、流程图 .....	5
2、分析内容 .....	5
五、生物信息学分析结果 .....	6
1、引物设计与 CpGI 注释 .....	6
1.1 引物设计 .....	6
1.2 CpGI 注释 .....	6
2、原始数据整理及质量评估 .....	6
2.1 原始数据格式 .....	6
2.2 测序错误率分布检查 .....	8
2.3 原始数据碱基质量分布 .....	9
2.4 原始数据碱基组成分布 .....	10
2.5 原始数据整理 .....	11
3、Bisulfite 处理不完全情况统计 .....	12
3.1 富集效率统计 .....	12
3.2 片段有效 reads 数统计 .....	12
3.3 盐化效率统计 .....	12
4、甲基化比例分析 .....	13
5、单倍型分析 .....	13
6、甲基化差异分析 .....	14
6.1 甲基化位点的显著性统计 .....	14
6.2 基因甲基化显著性统计 .....	14

6.3 单倍型显著性统计 .....	15
6.4PCA 和聚类分析.....	15
7. SNV 分析 .....	18
六、参考文献 .....	19
附录 I.....	20
1、结果文件目录列表 .....	20
2、软件信息列表 .....	20
附录 II.....	21
1、上海天昊生物科技有限公司简介 .....	21
2、权责声明 .....	23
3、论文引用或致谢 .....	24

## 一、服务介绍与说明

### 1、背景介绍

甲基化是一种非常重要的表观遗传学标记信息，参与了包括基因调控、生物发育、疾病等方方面面，并且在大多数癌症中广泛发现甲基化异常。获得检测范围内所有 C 位点的甲基化水平数据，对于表观遗传学的时空特异性研究具有重要意义。MethylTarget 以二代测序高通量测序平台为基础，结合 Bisulfite 处理和生物信息数据分析进行低成本、高效率、高准确度的 DNA 甲基化水平图谱绘制，为癌症、自身免疫等疾病的诊疗提供了有力的帮助。

### 2、样本要求

样品类型：基因组 DNA，溶解在 H<sub>2</sub>O 或 TE (pH 8.0) 中。

样品纯度：OD 260/280 值应在 1.8 ~ 2.0 之间。

样品浓度：最低浓度不低于 20ng/μL。

样品总量：DNA 总量 > 1μg。

### 3、数据量要求

Hiseq 2×150，所有样本目的区域的平均有效测序深度的平均值 > 400X，90% 样本的平均有效测序深度 > 280X，对于平均有效测序深度大于 280X 的样本，95% 以上的测序片段测序深度大于 2X（或 85% 以上的测序片段测序深度大于 10X）。

### 4、服务流程

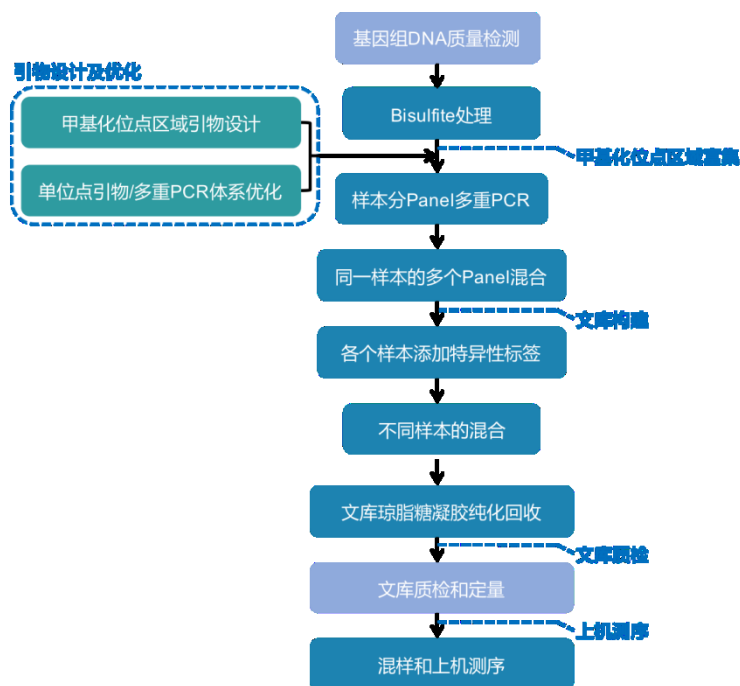
样品接收—样品 DNA 抽提与质检—建库测序—生物信息分析—项目总结报告

## 二、项目信息

项目名称	上海王明华—200 个样品 MethyTarget 富集测序		
项目编号	16B1212A		
客户单位	苏州大学		
项目样品信息			
物种信息	Human	参考基因组信息	Hg19
项目类型	Methyl Target 测序	测序信息	Illumina NextSeq 2 X 150
备注			
天昊生物联系人信息			
实验技术支持	黄泽斌	电话	021-50802060-123
		邮箱	huangzb@geneskies.com
数据分析支持	李才华	电话	021-50802060-117
		邮箱	lch@geneskies.com
项目整体支持	项目专家组	电话	021-50802060-142
		邮箱	
项目售后支持	熊伟明	电话	021-50802060-135
		邮箱	xiongwm@geneskies.com
项目审批			
<p>确认报告显示的内容与合同要求完全一致，同意项目结束，批准本项目总结报告发送。</p> <p style="text-align: right;">签名： <u>李才华</u></p> <p style="text-align: right;">2017 年 4 月 26 日</p>			

### 三、实验流程

#### 1、建库测序流程



#### 2、实验仪器与试剂

##### 主要器材：

器材名称	器材供应商
AB 2720 Thermal Cycler	Applied Biosystems, Waltham, MA, USA
XiangYi H1650-W	XiangYi, Hunan, China
EP600 Gel electrophoresis	上海玉柏实业有限公司
NanoDrop	NanoDrop technologies, Wilmington, DE, USA
Invitrogen Qubit Spectrophotometer	Invitrogen, Carlsbad, CA, USA
<b>NextSeq</b>	Illumina, San Diego, CA, USA

##### 主要试剂：

试剂名称	试剂供应商
<b>NextSeq</b> reagent kit	Illumina, San Diego, CA, USA
EZ DNA Methylation-Gold Kit	ZYMO, CA, USA
TIANGEN Gel Extraction kit	TIANGEN, Beijing, China
10× Reaction buffer	TaKaRa, Dalian, China

### 3、实验步骤

#### 1) 亚硫酸氢盐处理

使用 EZ DNA Methylation-Gold Kit 进行样本处理。

#### 2) 样本目标片段多重 PCR 反应

使用以下 PCR 条件：

A 20 $\mu$ l mixture was prepared for each reaction and included:

1x reaction buffer (TAKARA),

2 mM Mg<sup>2+</sup>,

0.2 mM dNTP,

0.1 $\mu$ M of each primer,

1 U HotStarTaq polymerase (TAKARA) and 2  $\mu$ l template DNA.

The cycling program was:

95°C for 2 min; 11 cycles of 94°C for 20 s, 63°C-0.5 °C per cycle for 40 s, 72 °C for 1 min; 24 cycles of 94°C for 20 s, 65 °C for 30 s, 72 °C for 1 min; 72°C for 2 min.

#### 3) 相同样本的多重 PCR 反应体系混合

#### 4) 样本添加特异性标签序列

使用以下 PCR 条件：

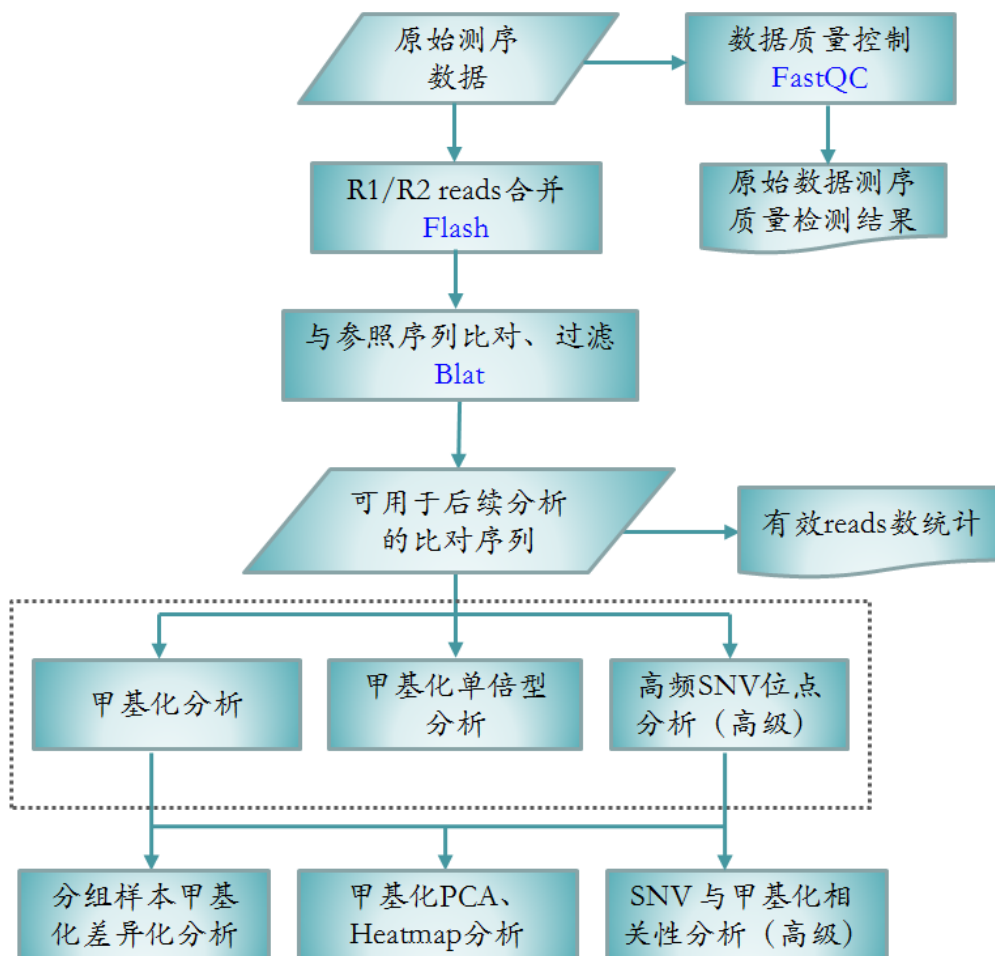
A 20 $\mu$ l mixture was prepared for each reaction and included 1x reaction buffer (NEB Q5™), 0.3 mM dNTP, 0.25 $\mu$ M of F primer, 0.25 $\mu$ M of index primer, 1 U Q5™ DNA polymerase (NEB) and 1 $\mu$ l diluted template.

The cycling program was 98°C for 30s; 11 cycles of 98°C for 10 s, 65°C for 30 s, 72 °C for 30 s; 72°C for 5 min.

#### 5) 定量后上机测序

## 四、生物信息学分析流程

### 1、流程图



### 2、分析内容

生物信息分析内容			
标准分析		高级分析	
引物设计与 CpGI 注释	√	高频 SNV 分析	
原始数据整理及质量评估	√	高频 SNV 与甲基化相关性分析	
Bisulfite 处理不完全情况统计	√		
甲基化比例分析	√		
单倍型分析	√		
差异甲基化位点分析	√		



## 五、生物信息学分析结果

### 1、引物设计与 CpGI 注释

#### 1.1 引物设计

采用 primer3 对 Bisulfite 处理之后的序列进行引物设计，软件 <http://primer3.ut.ee/>

表 1 引物信息

Primer Name	seq
ADCY2_F	GTATTTTAAAGTGTGATTGAAGTTGAG
ADCY2_R	CCCATCCTCTCCACCCTCT
AHRR_F	TTTTGTAATGGGGTTTGGTTGT
AHRR_R	TCTCCCAAATACAAACCCAAAAA

注：“Primer Name”指引物名称，“seq”指引物序列。

◆ 结果目录：methylation\_info.xlsx

#### 1.2CpGI 注释

对引物对应的产物进行位置注释，参考基因组版本:hg19。

表 2 CpGI 注释信息

Target	Chr	Gene	mRNA	mRNA _Strand	TSS	start	end	Length	Target_ Strand	Distance 2TSS
ADCY2_	5	ADCY2	NM_00546	+	7396342	7395281	7395481	201	+	-1061
AHRR_	5	AHRR	NM_001242412	+	304290	374201	374465	265	+	69911

注：“Target”指测序片段名称，“Chr”指染色体号，“Gene”指基因名称，“mRNA”指距离该产物最近的 Mrna，“mRNA\_Strand”指该 mRNA 编码方向，“TSS”指该 mRNA 转录起始位置，“start”指产物起始位置，“end”指产物终止位置，“Length”指产物长度，“Target\_Strand”指产物方向，“Distance2TSS”指产物相对 TSS 的距离。

◆ 结果目录：methylation\_info.xlsx

### 2、原始数据整理及质量评估

#### 2.1 原始数据格式

高通量测序（如 Illumina HiSeq/MiSeq 等测序平台）得到的原始图像数据文

FASTQ 文件，包含了测序序列的序列信息及其对应的测序质量信息。测序样品中真实数据随机截取结果如下图所示：

FASTQ 格式文件中每条 read 由四行描述，其中第一行以“@”开头，随后为 Illumina 测序标识符（Sequence Identifiers）和描述文字（选择性部分）；第二行是碱基序列；第三行以“+”开头，随后为 Illumina 测序标识符（选择性部分）；第四行是对应序列的测序质量。

表 3 测序文件识别标志 (Sequence Identifiers) 详细信息对照表

Read 的质量分数以不同的大写英文字符来表示,其中每个字符对应的 ASCII 值减去 33,即为对应的测序质量值。一般地,碱基质量从 0~40,即对应的 ASCII 码为从“!”(0+33)到“T”(40+33)。如果测序错误率用 E 表示,Illumina HiSeq/MiSeq 的碱基质量值用 Q 表示,则有下列关系:

$$\text{公式 1: } Q = -10 \log_{10}(E)$$

测序错误率与测序质量值简明对应关系如下表：

表 4 测序错误率与测序质量值简明对应表

测序错误率 (E)	测序质量值 (Q)	对应 ASCII 码
5%	13	.
1%	20	5
0.1%	30	?
0.01%	40	I

测序 reads 错误率会随着测序的进行而升高，是由于测序过程中化学试剂的消耗造成，这是 Illumina 高通量测序平台的通病。

## 2.2 测序错误率分布检查

Illumina 测序单次运行能产生数十亿级的 reads，如此海量的数据无法逐个展示每条 read 的质量情况，因此，生物信息分析运用统计学的方法，以图表的方式，从宏观上直观地反映出样品的测序质量和文库构建质量。

测序目前主要为双端（paired-end）测序，简称PE测序，即每个片段会获得2条reads，分别为read 1（R1端序列）和read 2（R2端序列）。通常会采用FastQC（[http:// www .bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)）对原始测序数据的质量进行评估。错误率分布图将R1和R2端的序列合并起来一起作图，而碱基组成和质量评估图则是针对每个样品的R1和R2端序列单独绘制一张图。图1为单个样品原始数据碱基错误率分布图，图2为单个样品原始数据碱基质量分布图，图3为单个样品原始数据碱基组成分布图。

\*\*\*\*示例样品 CASE 的原始数据碱基错误率分布图（CASE\_ErrorRate.png）、原始数据碱基质量分布图（CASE\_R1\_per\_base\_quality.png 和 CASE\_R2\_per\_base\_quality.png）、原始数据碱基组成分布图（CASE\_R1\_per\_base\_sequence\_content.png 和 CASE\_R2\_per\_base\_sequence\_content.png）。结果目录为 Image/fastqc。

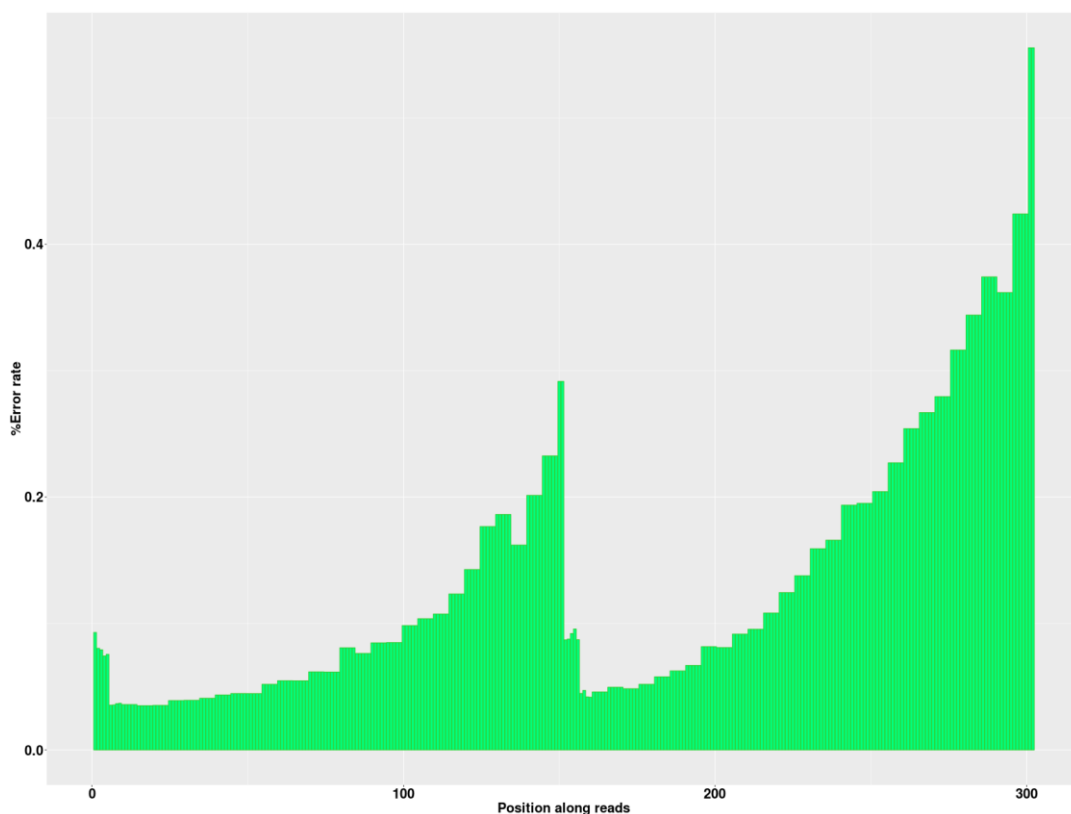


图 1 原始数据碱基错误率分布图

注：横坐标是 reads 碱基坐标；纵坐标是所有 reads 在该位点处碱基的平均错误率（%）。图中绿色阴影对应的是各位点碱基错误率的平均值，反映了测序 reads 碱基错误率的分布情况，一般错误率低于 0.1%即认为在可接受范围内。

### 2.3 原始数据碱基质量分布

碱基质量值  $Q$  的计算方式为  $-10 \cdot \log_{10}(E)$ ， $E$  为测错的概率。测序质量值可用来评估碱基的测序错误率，碱基质量值越高对应的碱基测序错误率越低，通常而言，碱基质量值  $Q_{20}$  就表示碱基的测序错误率为 1%，同理，碱基质量值  $Q_{30}$  则表示碱基的测序错误率为 0.1%。由于测序仪本身、测序试剂、样品等多个因素的共同影响，测序序列 5'端前几个碱基的错误率相对较高，随着序列的延伸，3'端碱基错误率会不断升高，这是由高通量测序的技术特点决定的。

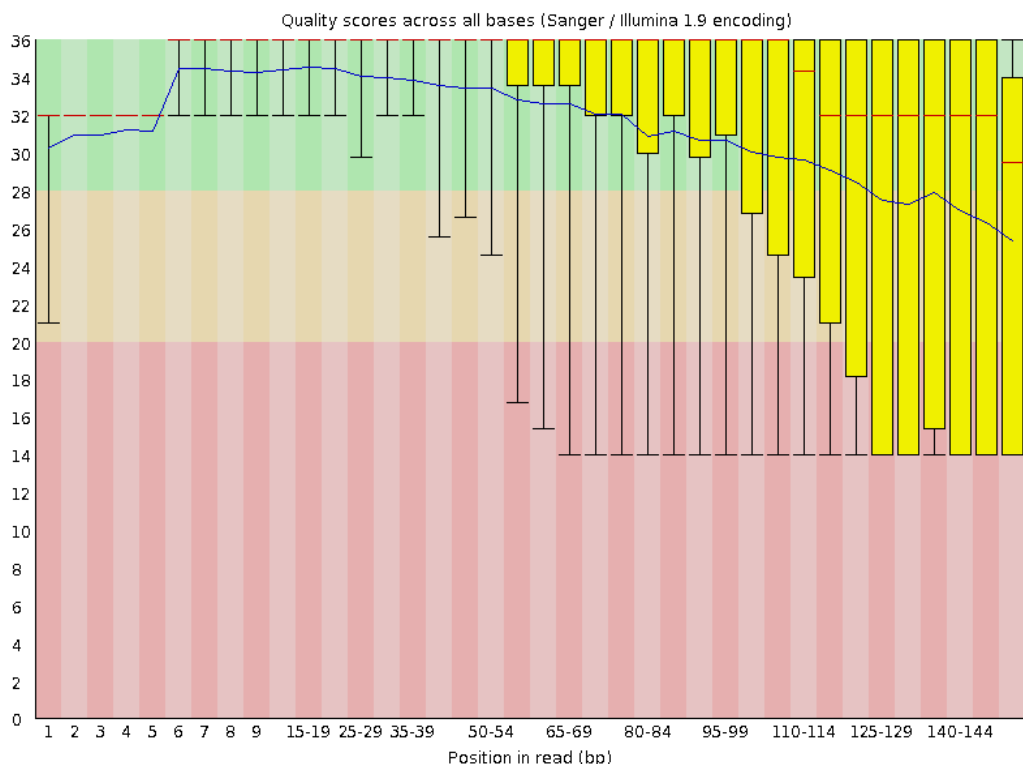


图 2 原始数据碱基质量分布图

注：示例样品 R1 端测序错误率分布图，横坐标为 reads 的碱基位置，纵坐标为 reads 的碱基质量值。图中每个位点有许多个碱基，采用箱线图的方式来展示所有碱基质量的分布，其中黄色方块的上下两条边框分别代表该位点所有碱基质量的上四分位数和下四分位数，红色的线代表该位点所有碱基质量的中位数，蓝色的线代表该位点所有碱基质量的均值，黄色箱子最下方的黑线代表该位点碱基质量的最小值，而最上方的黑线代表该位点碱基质量的最大值。背景色根据碱基质量的大小分成绿色、黄色、红色三个部分，绿色代表碱基质量在 28 以上，处于绿色区间证明该位点碱基质量较高，错误率在 0.01% 以下；黄色代表碱基质量在 20-28 之间，错误率在 0.01%~0.1% 之间，处于黄色区间证明该位点碱基质量稍差，但是也属于可接受范围；红色代表碱基质量在 0-20 之间，错误率在 0.1% 以上，此时的碱基质量比较差，此时测出来的序列可信度不高，可能影响下游分析的准确性，可考虑去除这样的低质量序列。

## 2.4 原始数据碱基组成分布

对于高通量测序平台，因随机性打断及 G/C 和 A/T 含量分别相等的原则，理论上 GC 及 AT 含量在每个测序循环应分别相等，且整个测序过程中稳定不变，呈水平线。但是序列的 5' 端前十几个碱基中会存在明显的偏向性，因此会在碱基分布图中呈现前端波动较大的现象，这属于正常情况。

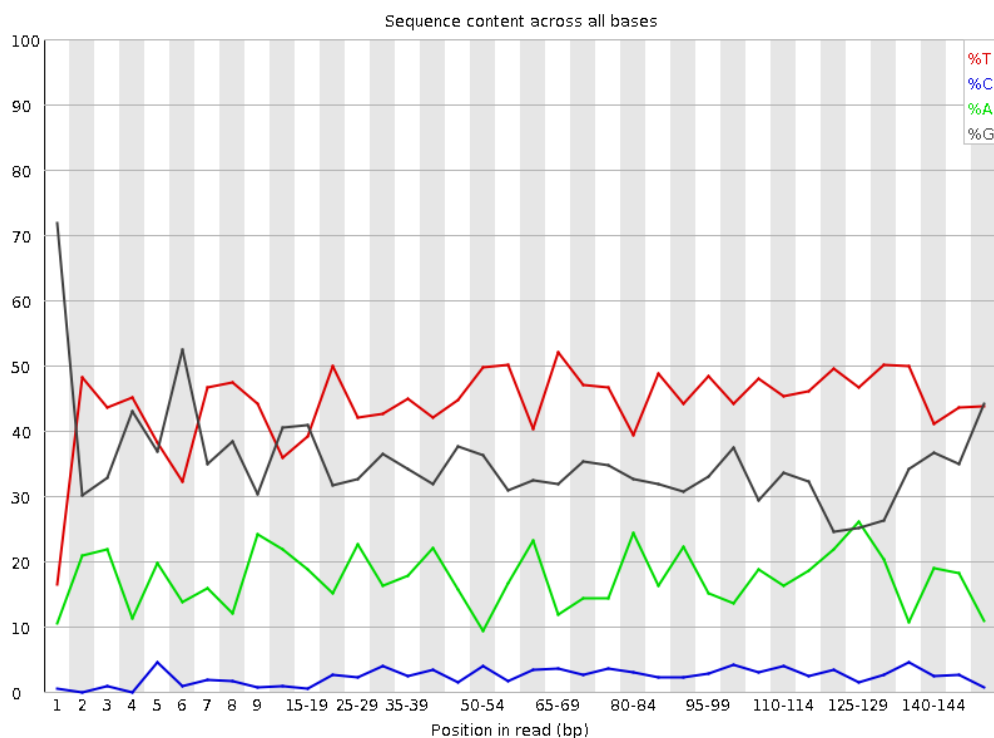


图 3 原始数据碱基组成分布图

注:示例样品 R1 端序列的碱基含量分布图,横坐标为 reads 的碱基位置,纵坐标为单碱基 A/T/G/C 所占的比例,不同颜色代表不同的碱基类型。由于与测序的引物接头相连,因此会有所波动,属于正常波动。一般情况下 A 与 T 相等, C 与 G 相等,各碱基所占百分比会因物种的差异而不同。

## 2.5 原始数据整理

对于原始测序数据,统计各样品的 reads 数目及测序质量。

表 5 原始测序数据统计

Sample	R1				R2			
	Total Reads	Reads length	Q20 (%)	Q30 (%)	Total Reads	Reads length	Q20 (%)	Q30 (%)
100	104632	35-151	99.99%	76.19%	104632	35-151	98.20%	38.46%
101	105360	35-151	99.99%	75.00%	105360	35-151	98.41%	40.23%
102	94476	35-151	99.98%	77.14%	94476	35-151	98.13%	38.02%

注:“Total Reads”和“Read Length”代表原始序列数量和测序读长,“Q20 (%)”和“Q30 (%)”分别表示原始数据中测序质量值 Q 不低于 20 和 30 的序列比例。

### ◆ 结果目录 : Data\_statistics.xlsx

### 3、Bisulfite 处理不完全情况统计

#### 3.1 富集效率统计

测序数据与参考基因组进行比对，统计目标片段富集的效率。

采用方法：Blat。

表 6 reads 富集效率统计

Sample	100	101	102	103
有效 reads 数目	81524	83997	73480	78549
总 reads 数目	104632	105360	94476	99712
富集效率	0.78	0.8	0.78	0.79

◆ 结果目录：Data\_statistics.xlsx

#### 3.2 片段有效 reads 数统计

统计在每个产物片段上的有效 reads 数目。

表 7 样本产物片段统计

Sample	ADCY2_	AHRR_
100	3228	95
101	2664	107
102	2456	98
103	3524	55

◆ 结果目录：Data\_statistics.xlsx

#### 3.3 盐化效率统计

统计每个样本的有效 reads 中，盐处理后碱基 C 转变为 T 的效率。

表 8 盐化效率统计

Sample	100	101	102	103
Transferred(C->T)	3152536	3274077	2884492	3031643
All Base(C)	3192722	3324358	2924695	3071220
Efficiency	98.74%	98.49%	98.63%	98.71%

注：“Transferred(C->T)”指盐处理后碱基 C 转变为 T 的数量，“All Base(C)”指参照序列中碱基 C 总数（非 CpG 位点），“Efficiency”指盐处理转化效率。

◆ 结果目录：methylation.xlsx



## 4、甲基化比例分析

分析每个甲基化位点在样本中甲基化的比例。

表 9 甲基化比例统计

Target	Position	Chr	GenomePosition	Distance2TSS	Type	CT	CU
ADCY2_	29	5	7395309	-1033	CG	0.037748	0.065617
ADCY2_	38	5	7395318	-1024	CG	0.056002	0.115861
ADCY2_	43	5	7395323	-1019	CG	0.085705	0.196475
ADCY2_	47	5	7395327	-1015	CG	0.040223	0.044994
ADCY2_	56	5	7395336	-1006	CG	0.0413	0.057785
ADCY2_	58	5	7395338	-1004	CG	0.045455	0.055515

注：“Target”指产物名称，“Position”指甲基化位点在目标片段上的位置，“Chr”指染色体号，“Genome Position”指该甲基化位点在参照基因组上的位置，“Distance2TSS”指该位点在参照基因组上相对转录起始位置的距离，“Type”指甲基化类型（CG 表示 CG 位点的甲基化）。  
 甲基化程度 = 该位点甲基化的 reads 数目（即检测到碱基 C 的 reads 数目）/ 该位点总的 reads 数目。

◆ 结果目录：methylation.xlsx

## 5、单倍型分析

对每个片段的每种单倍型进行计算。

表 10 单倍型类型、数量统计

Target	Haplotype	Depth	CT	CU	ET
ADCY2_	tttttttttttttt	59686	0.577578	0.307425	0.529412
ADCY2_	ttttttttcttttt	5358	0.052018	0.013921	0.041667
ADCY2_	ttttttttcttttt	3214	0.024215	0.009281	0.047794
ADCY2_	cccccccccccccc	3004	0.003587	0.00116	0.001225
ADCY2_	ttcttttttttttt	2859	0.024215	0.009281	0.017157
ADCY2_	ttttttttttcttt	1921	0.016143	0.00464	0.020833

注：“Target”指产物名称，“Haplotype”指单倍型类型，“Depth”指支持此单倍型的测序 Reads 数量。

单倍型分型示例：



假设参照序列为：ATCATXGATCXGCTAXGCTTTAXGCCTAT XG 即为：CG;

测序的 reads 为：ATCATCGATCTGCTACGCTTTATGCCTAT

该片段的单倍型即为：CTCT;

## ◆ 结果目录：methylation.xlsx

## 6、甲基化差异分析

对所有甲基化结果，根据样本的分组信息，进行差异显著性分析，分析方法为 Ttest 与 Utest。

当样本含量  $n$  较大时，样本均数符合正态分布，故可用  $u$  检验进行分析。当样本含量  $n$  小时（如  $n < 30$ ），若观察值  $x$  符合正态分布，则用  $t$  检验。

### 6.1 甲基化位点的显著性统计

表 11 甲基化位点显著性统计

Target	POS.	Type	P-value(Ttest)	P-value(Utest)
ADCY2_	29	CG	1.53211E-14	2.25033E-19
ADCY2_	38	CG	1.33227E-15	1.15765E-17
ADCY2_	43	CG	0	2.37826E-19
ADCY2_	47	CG	1.1724E-13	1.10486E-16
ADCY2_	56	CG	7.25642E-13	2.22078E-14
ADCY2_	58	CG	2.17937E-13	1.66046E-16

## ◆ 结果目录：Different\_Methylation\_CpG.xlsx

### 6.2 基因甲基化显著性统计

表 12 基因甲基化显著性统计

Gene	P-value(ttest)	P-value(Utest)
ADCY2_	0	2.26271E-21
AHRR_	2.4122E-08	1.5649E-07
ARRB2_	1.62437E-12	8.00108E-18
BCAN_	0	1.14132E-27
DLX5_	1.62448E-12	1.228E-25

注：基因的甲基化程度为该基因上所有的甲基化位点甲基化程度的均值

## ◆ 结果目录：Different\_Methylation\_Gene.xlsx

### 6.3 单倍型显著性统计

表 13 单倍型显著性统计

Target	Haplotype	P-value(ttest)	P-value(Utest)
ADCY2_	ttttttttttttttt	1.62437E-12	1.98673E-19
ADCY2_	ttttttttctttttt	1.0975E-09	2.09356E-14
ADCY2_	ttttttttctttttt	1.86506E-12	6.28596E-17
ADCY2_	ccccccccccccccc	1.52704E-05	1.27127E-06
ADCY2_	ttctttttttttttt	3.9224E-08	2.35536E-14
ADCY2_	ttttttttttctttt	4.5209E-10	7.19514E-13

◆ 结果目录: **Different\_Methylation\_haplotype.xlsx**

### 6.4PCA 和聚类分析

根据甲基化水平，对所有样本进行 PCA 和聚类分析，了解样本间甲基化差异。

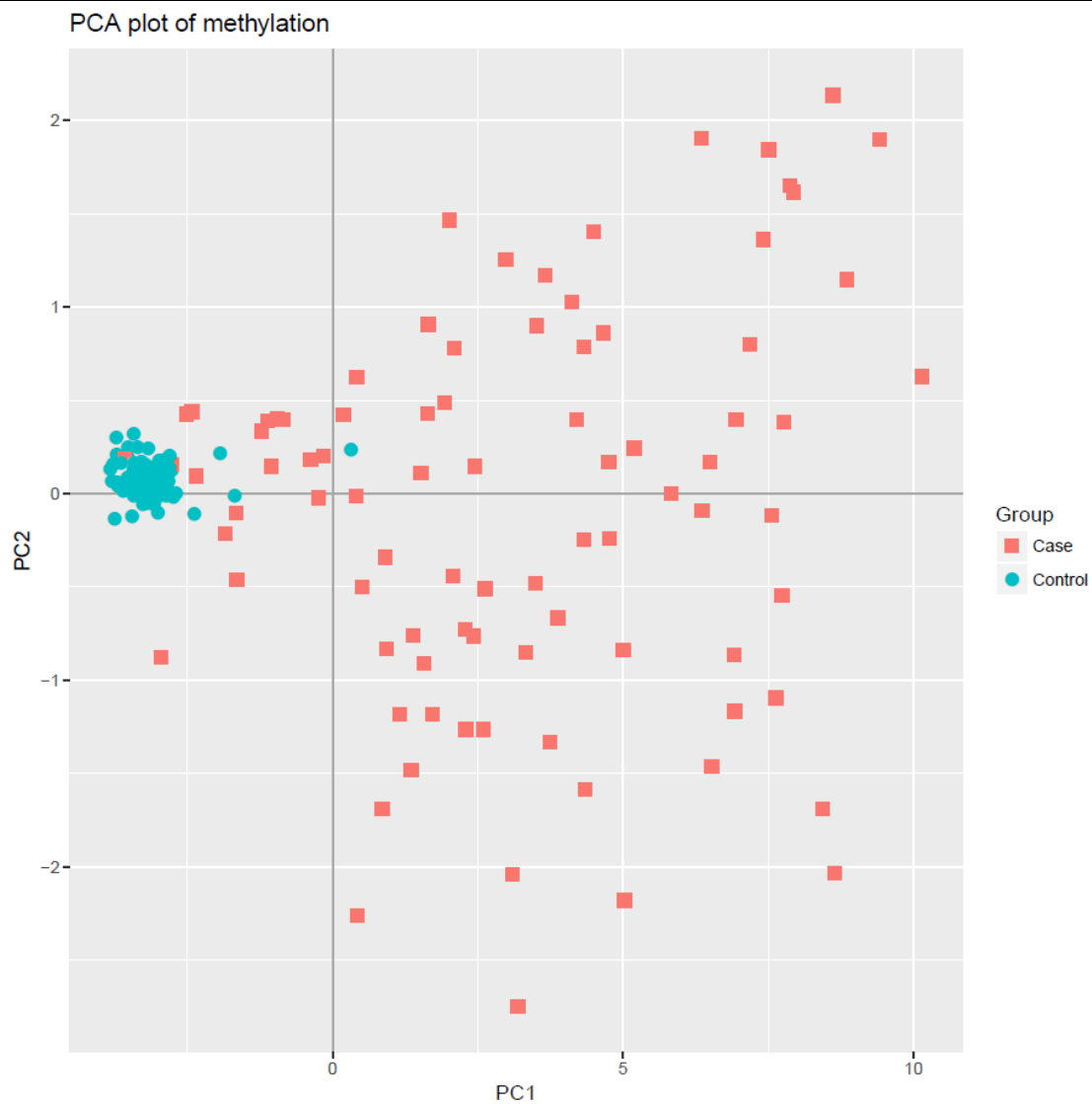


图 4 PCA

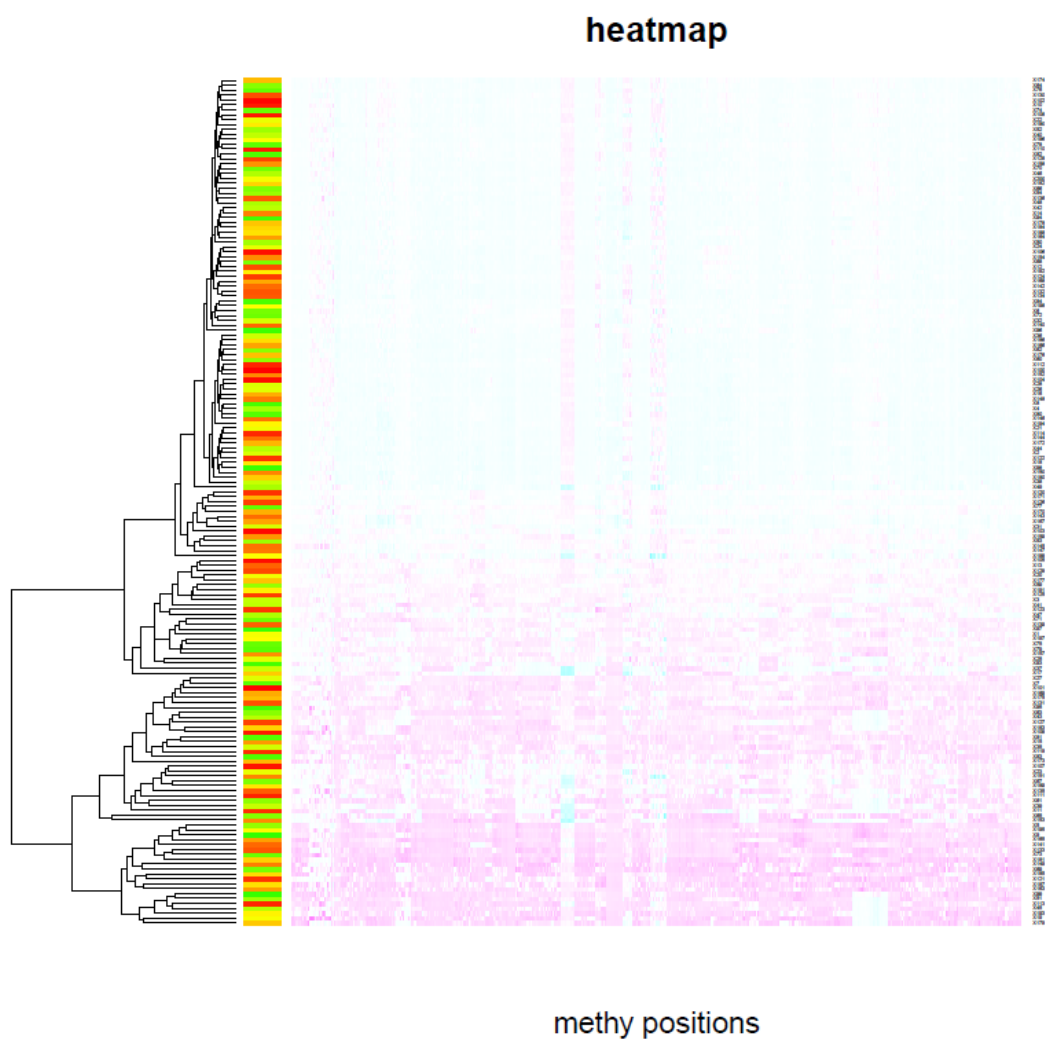


图 5 Heatmap

## 7. SNV 分析

根据目的片段在参照基因组上的位置，以 1000genomes 中整体频率或汉族人群频率高于 0.01 为过滤条件，对目的片段相应位点上的所有高频 SNV 位点进行分型。以突变等位基因频率低于 0.15 为 HOMR（野生型），0.15~0.85 为 HET（杂合突变），高于 0.85 为 HOMA（纯合突变）为标准进行分型。

表 14 SNV 分析结果示例

Target	cg01185433_	GRIN1_	GRIN1_
POS	112	144	165
Strand	+	+	+
RefBase	G	G	A
C2TBase	G	G	A
Chr	4	9	9
GenomePosition	1.75E+08	1.4E+08	1.4E+08
rs_NO.	rs954927	rs141473515	rs6293
Ref	G	G	A
Alt	A	C	G
1000g_AF	0.230431	0.004193	0.140775
1000g_CHBS_F	0.581877	0.012136	0
HWE	0.758818	1	1
MAF	0.554598	0.042553	0.003906
XXXXXXXX			

注：“Target”指目标片段名称，“Strand”指目标片段链方向，“RefBase”指目标片段上该位点的碱基，“C2Tbase”指重硫酸盐处理后目标片段上该位点的碱基，“Chr”指该片段所在染色体，“GenomePosition”指该位点在基因组上的位置，“rs\_NO”指该位点的 rs 号，“Ref”指参照等位基因，“Alt”指突变等位基因，“1000g\_AF”指 1000g 中的频率，“1000g\_CHBS\_F”指 1000g 中汉族人群的频率，“HWE”指哈迪-温伯格平衡(Hardy-Weinberg equilibrium,HWE)的计算值，“MAF”指最小等位基因频率。“XXXXXXXX”指样本的 SNP 分型结果。表“baseCount”中，是每一个高频突变位点的参照与突变碱基测序 Reads 数目、突变碱基比例以及突变类型。另外，可能存在红色标志的 SNP 位点，该类型位点的突变方式为 C->T，无法根据甲基化结果进行分型，故仅列出 Reads 数目。

### ◆ 结果目录：SNP\_genotype.xlsx

## 六、参考文献

- [1]. Feng H, Conneely KN, Wu H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res*. 2014 Apr;42(8): e69. doi: 10.1093/nar/gku154. Epub 2014 Feb 22.
- [2]. Paul DS, Guilhamon P, Karpathakis A, Butcher LM, Thirlwell C, Feber A, Beck S. Assessment of RainDrop BS-seq as a method for large-scale, targeted bisulfite sequencing. *Epigenetics*. 2014 May 1;9(5):678-84. doi: 10.4161/epi.28041. Epub 2014 Feb 11.
- [3]. Holmes EE, Jung M, Meller S, Leisse A, Sailer V, Zech J, Mengdehl M, Garbe LA, Uhl B, Kristiansen G, Dietrich D. Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PLoS One*. 2014 Apr 3;9(4): e93933. doi: 10.1371/journal.pone.0093933. eCollection 2014.

## 附录 I

### 1、结果文件目录列表

- A、Data\_statistics.xlsx 为测序数据量及数据质量的统计。
- B、methylation\_info.xlsx 为引物及注释相关信息。
- C、methylation.xlsx 为甲基化比例及单倍型等分析。
- D、Different\_Methylation\_CpG.xlsx、Different\_Methylation\_Gene.xlsx、  
Different\_Methylation\_Gene.xlsx 为甲基化差异性分析结果。
- E、SNP\_genotype.xlsx 为 SNP 注释信息。

### 2、软件信息列表

Fastqc-v0.11.5	<a href="http://www.bioinformatics.babraham.ac.uk/projects/fastqc/">http://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
FLASH-v1.2.11	<a href="http://www.cbcb.umd.edu/software/flash">http://www.cbcb.umd.edu/software/flash</a>
blat-v.36	<a href="http://genome.ucsc.edu/FAQ/FAQblat.html">http://genome.ucsc.edu/FAQ/FAQblat.html</a>

## 附录 II

### 1、上海天昊生物科技有限公司简介

上海天昊生物科技有限公司，2008 年 4 月创建于上海浦东张江高科技园区。天昊生物目前拥有员工近 100 名，2500 平方的办公与实验场地，建立了整套完备的实验室管理体系和标准流程，主要目标是构建一个具备国际一流水准的标准化多层次分子生物学研究分析平台，为国内外基因生物领域科研机构、医学院校以及生物制药企业提供精确、高效的基因检测分析服务。

2013 年我们在北京和南京分别成立的两个办事处，力求为全国客户提供更全面、更优质的服务。

2014 年天昊生物被评为“国家高新技术企业”，公司总经理姜正文博士获得国家“千人计划”，江苏省“双创人才”、苏州工业园区“领军人才”等荣誉。

2015 年 7 月天昊基因与谱润、山蓝和盛宇 3 家知名基金投资公司，完成了人民币 8000 万元的首轮融资。

自天昊生物成立以来，我们已经建立并完善了包括 ABI 3130xl、ABI 3730xl 测序仪、实时定量 PCR 仪、Illumina GAIIX、Illumina MiSeq、Illumina NextSeq 二代测序平台和细胞生物学分析平台在内的多个分子生物学分析研究平台。不仅先后开展了疾病基因定位、基因突变及多态性测序分析、各种通量 SNV 基因分型分析、基因转录水平分析、甲基化水平分析等服务项目，同时还利用目前世界先进的二代测序技术开展了全基因组测序、全外显子组测序、目的区域富集测序、RNA 测序、micro-RNA 测序、甲基化测序等成熟的科研分析服务。此外，上海天昊在进行高质量技术服务的同时还积极集中优势力量自主创新，独立研发了 AccuCopy® 多重拷贝数检测技术、CNVplex® 高通量拷贝数检测技术、SNVscan® 高通量 SNV 分型方法等具有国际水平的专利技术，另外利用二代测序技术不断开发出大量新的基因分析技术，例如多种策略方法的目的区域富集大样品测序技术 EasyTarget™、超高通量 SNV 分型技术 SNVseq®、超高通量目的区域 CNV 检测技术 CNVseq® 等。截至 2015 年 10 月底，公司申请了 4 项 PCT 专利和 24 项国内发明专利，其中 3 项专利已经授权，另 3 项正在实审阶段，另外获得 5 项软件著作权，申请了 40 商标著作权，已获 14 商标著作权。目前公司可以提供的分子生物学或基因组学相关科研服务项目超过 70 类，迄今已经为国内外近 550 多家科研院校、医疗单位和生物公司提供了超过 6000 项科研技术服务。

在为广大新老客户提供科研技术实验服务的同时，上海天昊集服务与产品于一身，利用自主研发的 AccuCopy®、CNVplex®、SNVscan®、EasyTarget®、SNVseq® 和 CNVseq® 等全新技术为客户提供灵活定制科研项目检测试剂盒服务，为具备



检测实力的大中型科研单位提供更为先进高效的检测方法和试剂。

上海天昊遗传分析中心借助多名长期从事基因及遗传分析的领域专家构成的专家咨询团队，结合准确、高效、经济的科研技术服务体系，致力于长期为分子生物学及医学遗传学领域的研究者提供高质量的科研策略咨询、实验技术服务和遗传数据分析，帮助广大科研人员获得更为优质的科研成果。

上海天昊生物科技有限公司

地址：上海市浦东新区康桥路 787 号 9 号楼

服务热线：400-065-6886

电话：021-50802060/021-50807380；传真：021-50802059

网址：<http://biotech.geneskies.com/index.html>

## 2、权责声明

### 第 1 条

本报告及其附件中披露的由上海天昊遗传分析中心负责设计优化的检测体系（合同中对权利归属已做说明的除外）的知识产权归上海天昊生物科技有限公司（以下简称公司）所有，任何其它单位及个人未经公司许可不得对其进行专利申请或用于商业用途。上述检测体系包括设计合成的各种核酸序列、检测反应组成、反应条件及检测流程等。

### 第 2 条

本报告及其附件中披露的项目内容、实验数据及结果，公司必须严格保密，未经客户单位负责人允许，严禁公司任何人对外宣传。

### 第 3 条

为了保证项目成果的安全，本报告及其附件中披露的实验数据及结果已通过邮件发送，而本报告及其附件中披露的实验及数据分析相关信息、原始实验数据则需客户单位的项目负责人或其联系人通过公司授权的帐号在公司项目管理系统中下载。

### 第 4 条

本报告及其附件内容仅限于项目团队成员、客户单位的项目负责人及其联系人查阅或拷贝，公司内任何其他人员未经项目总监允许不得查阅或拷贝，公司外任何其他人员未经客户单位的项目负责人许可不可查阅或拷贝。

### 3、论文引用或致谢

如果您认可我们的工作，在发表文章时愿意将我们放到致谢名单中，我们将非常感谢，文章中天昊的中英文名称如下：

上海天昊生物科技有限公司（中文）

**Genesky Biotechnologies Inc., Shanghai, 201315 (English)**

天昊生物