

芯片说明文件

1. 检测结果中给了很多文件夹，每个文件夹都是什么内容呢？

每次检测您将收到一个以“公司+芯片类型+样本数+日期.rar”命名的压缩包，解压后文件中包含 3 个文件夹，“1. QC data”包含的是样本的芯片的内参质控数据，“2. Raw data”保存的是芯片的原始下机的荧光数据，“3. Data result”保存的是被检样本的分型结果文件。

2. 质检不合格能够上机吗？

对于质检不合格的样本不建议上机，假如样本比较珍贵或者无法再次收集时可以安排上机，但是这类样本风险需要用户承担，果壳统计对于质检不合格的样本风险上机检出率合格（Call Rate 大于 97%）的概率在 50%以下。

3. 我使用的这款芯片有多少探针呢，每个探针在染色体的什么位置，探针的序列是什么？

对于 ASA-CHIA 全基因组 SNP 芯片我们提供两种探针信息文件供客户使用，一种是 Illumina 官方的注释文件，详见百度云盘：<https://pan.baidu.com/s/1JINrpFMJ9kCOPXRyQVnpgg> 提取码: 79sh，一种是果壳自己注释的位点信息，详见百度云盘：<https://pan.baidu.com/s/1bDh5RPGkOCBWsuNhqTcFTw> 提取码: t1hd；建议用户采用果壳注释的文件，因为果壳注释过程对于 Illumina 官方的文件进行了矫正，更正了一些位置错误的位点信息，同时也更新了一些位点的注释信息如 ref/alt、rs 等。

4. 果壳提供的芯片的注释文件和官方提供的注释文件有何区别，为什么部分位点注释的染色体及物理位置与官方的不同，果壳是如何进行注释的？

果壳对芯片位点的注释是采用设计探针的 source 序列与参考基因组进行比对，找到 source 序列中 SNP 或 Indel 所在的位置进行注释的，确实部分位点的位置与官方注释存在差异，对于位点信息不一致的位点建议采用果壳注释的结果，因为果壳对于每个位点都是经过核对的并且注释的信息更加完善

（如官方注释标注染色体及物理位置都为 0 的位点果壳准确标注了位点的具体所在的染色体及物理位置等）。

5. 探针的前缀各不相同，分别是什么意思？

IlmnID 是 Illumina 探针数据库的唯一识别号，部分探针保留了其来源信息，可根据探针的命名初步推断位点的信息但是不能定性，具体探针的检测的位点信息需根据果壳提供的位点信息表（2 中的链接：

<https://pan.baidu.com/s/1g-rbX0nmsMoZZ4Om1357iQ> 提取码: j5w2），Illumina 探针命名的几种方式如下：

- 1) 1000genome 来源：a) 染色体位置；b) exm + rs 编号；c) seq + rs 编号；d) kgp 命名；e) GSA + rs 编号命名；
- 2) dbSNP 数据库来源：a) rs 编号；b) exm + rs 编号；c) seq + rs 编号；
- 3) Clinvar 数据库来源：a) seq + rs 编号；b) 基因名称 + 染色体坐标；c) seq + 染色体坐标；d) 染色体坐标；
- 4) PAGE 来源：染色体位置 + 碱基信息；
- 5) BGI 来源：日期 + 编号；
- 6) PGX 来源：染色体位置 + rs 编号；
- 7) eMERGE 来源：a) 染色体坐标；b) seq + rs 编号；
- 8) STAARS 来源：a) indel + 编号；b) variant + 编号；c) new + rs 编号。

6. 分型结果中的 I 和 D 分别代表什么？

ASA-CHIA 芯片中设计有检测插入和缺失的探针，插入用 Insert (I) 表示，缺失用 Delete (D) 表示，无论位点的 reference allele 是什么，D 永远表示短的序列，I 表示长的序列，如位点 seq-rs727504466 的 reference allele 是 AT，alt allele 是 A，分型结果中的 II 表示该样本在在位点是 AT 的纯合插入，假如分型结果为 DD，表示该样本在该位点是纯合的缺失，假如该位点是 ID 表示该位点分型为杂合的插入。

7. 为何芯片上标注的同一个 SNP 位点检测出来的分型不一致？

这类情况可分为两类：

一类是芯片上设计的用于检测同一种突变类型的重复探针，如

1_100715325.1_BOT 和 1_100715325.2_TOP 检测的 allele 类型均为 C/A，理论上同一样本在该位点分型结果应一致，假如出现分型不一致的现象需要具体查看位点的分型聚类图，根据聚类图才能判断哪种分型是准确的。设计检测同一种 allele 重复探针的初衷一般有以下两种情况：一是探针比较重要，要求芯片上必须包含该探针，另一种情况是 SNP 或者 Indel 的上下游序列不好设计探针，为了尽量保证该位点能够被检出芯片上会设计重复探针检测，所

以就会出现某些重复探针检测分型为“--”，有的探针分型成功或者重复探针分型不一致的现象。ASA-CHIA 芯片上设计的重复探针详见百度云盘：链接：<https://pan.baidu.com/s/1BZZKWxi6Wlaxe2VkBx7jw> 提取码：pg7r。

另一类是该位点是三态或者多态，采用多个探针检测多态，如位点 `ilmnseq_rs562215810.1_F2BT` 检测的是 A/C，`ilmnseq_rs562215810.2_F2BT` 检测的是 A/G，两个位点位于染色体的同一位置。这类探针的分型无法通过软件直接判断，需要人工判读，如被检样本在位点 `ilmnseq_rs562215810.1_F2BT` 的分型为 AA，在 `ilmnseq_rs562215810.2_F2BT` 的分型为 GG，那么样本在这个位点的实际分型结果为 AG。ASA-CHIA 芯片上设计的用于检测多态的探针详见百度云盘：链接：https://pan.baidu.com/s/1JPDP_kCqvPPh4YoKgepljw 提取码：r54i。

8. 为何从 GenomeStudio 软件中导出的 forward 链的分型中有些位点的分型结果是反义链的？

首先 GenomeStudio 软件是无法直接导出正义链分型结果的，其 forward 并非我们理解的正义链的分型。Illumina 对于链有三种定义，分别为：Forward/Reverse、plus/minus、Top/Bot，每条链的具体含义详见 Illumina 官网（<https://support.illumina.com/bulletins/2017/06/how-to-interpret-dna-strand-and-allele-information-for-infinium-.html>）。果壳提供基于 top 链分型转换为正义链分型的对应关系，详见 2 中果壳注释文件中的 Top_Strand 列，Top_Strand 为“+”表示这些位点导出的 top 链的分型即为正义链分型，Top_Strand 为“-”表示这些位点导出的 top 链的分型为反义链分型。果壳给客户提供的结果中如客户无特殊说明结果均为转换为正义链的分型结果。

9. 果壳可以提供每个位点对应的 rs 号吗？

果壳提供的注释文件中的 ID 列即为每个位点对应的 rs 号。

10. Illumina 官方提供的注释文件中染色体中标注 0 和 XY 是什么意思？

Chr 为 0 的有两种可能：

- 1) 无法定位到基因组上某个区域，有的探针可能来自早期某个 dbSNP 的侧翼序列，而该 SNP 的记录已经被清除；
- 2) 探针序列对应多个区域。

Chr 为 XY 的位点来自性染色体上的拟常染色体区域（PAR），也是对应多个

区域的情况。

果壳的注释文件中已经对染色体为 0 及 XY 的位点进行了重新注释，具体详见 2 中果壳注释文件中查看。

11. 果壳提供的位点信息的参考基因组版本是什么？

果壳提供的注释文件的参考基因组版本为 hg19/GRCH37，用户想用 GRCH38 版本位点信息可使用 UCSC 的 liftOver 工具进行位置坐标转换：
<https://genome.ucsc.edu/cgi-bin/hgLiftOver>。

12. 每个位点有注释结果吗？如每个位点的 ref/alt allele、位点位于外显子区还是内含子区，位点在千人的频率，位点对应的疾病等等？

果壳生物已经对每个位点都进行了详细的注释，主要包含每个位点对应的染色体、物理位置、ref/alt allele、位点所在的区域、位点在千人的频率、位点可能与哪些疾病相关等等信息，详见百度云盘：

<https://pan.baidu.com/s/15H7EV7cG4E3D2g82IoMC7g> 提取码: qfvg。

13. 果壳的定制芯片 ASA-CHIA 与标准版的 ASA () 芯片有何区别？

ASA-CHIA 是基于标准版的 ASA 基础上增加了 4 万左右的位点。具体区别如下：

- 1) 果壳定制芯片过程中删选了一些标准版芯片上分型不准确及检出率低的位点，所以标准版芯片上 2277 个位点在 CHIA 版本上没有。
- 2) 果壳定制的位点中增加了一些疾病的 pannel，如遗传性耳聋、高血压、糖尿病、乳腺癌、前列腺癌等疾病相关的位点，另外增加了包含 X、Y 和 MT 在内的 3 万多性染色体的位点，主要用于追溯祖源及性染色体疾病的研究。

14. 我想做拷贝数变异研究，ASA-CHIA 芯片可以吗？

ASA-CHIA 芯片是可以做拷贝数变异研究的，但是具体需要看用户关心的 CNV 的大小，果壳可以给用户提供被检测者的 CNV 数据。

15. ASA-CHIA 芯片较市面上的 GSA 芯片或者 PMRA 芯片有什么优势呢？

项目	PMRA	APMRA	GSA	ASA-CHIA
平台	Thermo Fisher	Thermo Fisher	Illumina	Illumina
位点数	920K	800K	700K	740K
应用群体	全球人种	东亚、东南亚人	全球人种	东亚、东南亚人
多态性 (>1%)	50%	60%	55%	80%
低频突变覆盖 (1%-5%)	13% (110K)	20% (160K)	13.6% (95K)	24% (180K)
通量 (每周)	960 samples	960 samples	4608 samples	4608 samples

由上表可看出无论从低频区域覆盖位点还是中高频区域位点覆盖上来看 ASA-CHIA 较同类型芯片均具有显著优势，并且通量很高，一台仪器每周可检测 4608 个样本，非常适合基于亚洲人群的各类研究。

16. ASA-CHIA 芯片适合做什么类型的分析？

1) 疾病易感基因筛查；2) 体检致病基因筛查；3) 药物基因组代谢；4) GWAS 分析；5) 家系连锁分析等

17. 芯片位点检出率达到多少算合格呢？

对于质检合格的样本，果壳承诺样本的位点检出率大于 97%，低于 97% 的样本果壳免费进行芯片二次检测；对于质检不合格的样本上机果壳不保证位点的检出率。

18. 果壳能否导出正义链的 plink 格式数据？

默认果壳仅提供正义链的分型格式数据，假如用户需要果壳可以为户提供正义链的 plink 格式数据，plink 格式数据说明详见：<http://www.cog-genomics.org/plink/1.9/formats#raw>。

19. 果壳是如何计算被检样本性别的？

果壳是通过被检样本性染色体的分型来判断被检样本的性别的，由于该专利果壳正在申报中，现在不便透露，待专利申报完成之后将公开专利号。

20. ASA- CHIA 芯片有没有配套分析软件?

Illumina SNP 芯片有官方的数据分析软件 GenomeStudio, 数据分析软件、软件使用说明及分析需要的文件详见百度云盘:

https://pan.baidu.com/s/1kO-pKY3dMI0tdE05uEn_fw 提取码: xm9g。

北京果壳生物科技有限公司