**重组DNA技术与基因工程-张惠展**

**（2017 张惠展）P53和p21是两个功能相似的蛋白，构建到相同的载体，在大肠杆菌中表达，p21能表达 但是p53不能，可能的原因？(同2015一样)**

**（2015 张惠展）研究显示，临床上几乎所有癌症患者体内均伴有肿瘤抑制因子p53蛋白或其编码基因的缺失突变。因此规模化生产重组人p53蛋白有可能为癌症治疗提供一条新途径。前期相关研究将人类p53蛋白及其功能相似的伙伴p21的cDNA编码序列分别克隆在同一种大肠杆菌的表达质粒上。由此获得的p21目的重组子能表达出相应的p21重组蛋白，但是p53的重组子却不能，试分析可能原因。**

有很多因素能够影响人的基因（外源基因）在大肠杆菌的表达型质粒中的表达。有可能是是转录效率较低，也有可能是翻译效率的影响。由于这两个基因是克隆在同一种质粒上，启动子的影响应该比较小，更有可能是以下因素引起的，例如：

1.核糖体结合位点（核糖体结合位点（ribosomebinding site，简称RBS），是指mRNA的起始AUG上游约8~13核苷酸处，存在一段由4~9个核苷酸组成的共有序列-AGGAGG-，可被16SrRNA通过碱基互补精确识别的序列。）——影响了mRNA翻译起始效率：翻译起始密码子效率低；SD序列（Shine-Dalgarno （SD）是细菌和古细菌中信使RNA中核糖体结合位点序列。通常位于翻译起始密码子AUG上游约8个碱基位置。SD序列帮助招募核糖体RNA，并将核糖体比对并结合到信使RNA（mRNA）的起始密码子（AUG），从而开始蛋白质合成。一旦被招募，tRNA可以按照密码子的指令顺序添加氨基酸，从翻译起始位点向下游移动进行蛋白质合成。）与翻译起始密码子之间的距离及碱基组成影响了核糖体结合；基因编码区5‘端若干密码子的碱基序列不利于核糖体结合

2.密码子（密码子（codon）是指信使RNA分子中每相邻的三个核苷酸编成一组，在蛋白质合成时，代表某一种氨基酸的规律。）——不同生物对密码子的偏爱性不同：tRNA含量少，SD样序列多，稀有密码子（稀有密码子：遗传密码子有64种，但是绝大多数生物使用密码子具有偏好性，倾向于利用这些密码子中的一部分。被最频繁利用的称为佳密码子，那些不被经常利用的称为稀有密码子。）

3.质粒（一类独立于寄主菌染色体之外的小型环状DNA）拷贝数——少，翻译与复制冲突

4.mRNA不稳定——被降解，无法翻译重组蛋白

5. 除了以上影响重组蛋白产生的原因，也有可能是重组蛋白会被快速降解：异源蛋白稳定性低，或者能被大肠杆菌的蛋白酶降解（具有PEST序列）

“N 端 富 含 Pro、Glu、Ser、Thr 的 真 核 生 物 蛋 白 质 在 真 核 或 原 核 细 胞 中 的 半 衰 期 通 常都很短，至少E1A、c Myc、c Fos和p53等蛋白都具有这种特性，特别是由这四种氨基 酸残基构成的Pro Glu Ser Thr四肽序列(即PEST序列)显示出对细胞内蛋白酶系统 的超敏感性。在大多数情况下，PEST序列的两侧拥有一些带正电荷的极性氨基酸，据推测 PEST序列实质上是一个钙结合位点，它能促进那些钙依赖性蛋白酶系统对蛋白质的降解 作用。尽管有些实验结果并不能证实PEST序列对蛋白质稳定性的负面影响，但当某一真 核生物基因在大肠杆菌中不能稳定表达时，注意PEST序列是否存在，并在不影响异源蛋白 生物功能的前提下对之进行适当的改造，不失为一种提高表达产物稳定性的尝试。”——来自张惠展《基因工程》

**（2012 张惠展）简述将基因A构建到载体B上面的方法。**

即重组载体构建策略，切、接—转—增—筛

1.DNA的的体外重组（切与接）

同种内切酶，同尾酶，不同粘性末端，人工粘性末端，平头末端；同源重组infusion，CRISPR系统体内切接

需要先计算重组率来指导实验

2.重组DNA分子的转化与扩增

Ca诱导的，电击转化；还有其他穿刺，微流体剪切，局部加热，激光等

计算转化率（载体本身，受体细胞，方法）

3.转化细胞的扩增

4.转化子的筛选和鉴定-区分转化子，重组子，目的重组子

基于载体的遗传标记检测：抗药性，营养缺陷型，显色筛选

基于克隆DNA序列检测：PCR扩增检测，限制性酶切图谱法，菌落噬菌斑原位杂交法，DNA序列分析法

基于外源基因产物检测：蛋白质生物功能（淀粉酶，抗菌素，DNA结合蛋白），蛋白质生物结构（放射免疫原位杂交鉴定），分子量（聚丙烯酰胺凝胶电泳法）

**基因克隆的方法**

1.鸟枪法

（鸟枪法(Shot gun)是一种直接从生物细胞基因组中获取目的基因的方法。鸟枪法将某种生物体的全基因组或单一染色体切成大小适宜的DNA片段，分别连接到载体DNA上，转化受体细胞，形成一套重组克隆，从中筛选出含目的基因的目的重组子 [1-2] 。此法操作简单，但工作量大）

2.cDNA（cDNA是指互补(有时称拷贝)DNA。特指在体外经过逆转录后与RNA互补的DNA链。与平常我们所称谓的基因组DNA不同，cDNA没有内含子（内含子（Intron）又称间隔顺序，指一个基因或mRNA分子中无编码作用的版段）而只有外显子的序列）法

3.PCR法（PCR是聚合酶链式反应的简称，是一种酶促化学反应，可以在试管里将待测的目的基因在很短的时间内扩增五十万倍乃至上百万倍，大大提高了基因诊断的灵敏度，降低了分析的难度。），基于同源重组（同源重组（Homologous Recombination) 是指发生在姐妹染色单体（sister chromatid) 之间或同一染色体上含有同源序列的DNA分子之间或分子之内的重新组合。同源重组需要一系列的蛋白质催化，如原核生物细胞内的RecA、RecBCD、RecF、RecO、RecR等；以及真核生物细胞内的Rad51、Mre11-Rad50等等。同源重组反应通常根据交叉分子或Holliday结构（Holliday Juncture Structure) 的形成和拆分分为三个阶段，即前联会体阶段、联会体形成和Holliday 结构的拆分。）的infusion法，盒式PCR扩增法，反向PCR扩增法

4.化学合成法

**（2009 张惠展）如何分析一段多少长的序列是不是启动子 ，如何把它克隆出来？**



**（2007 张惠展）根据给出的一段需要表达的肽段的序列，设计载体，宿主，限制性内切酶的选择，写出关键的注意点。**

构建克隆的一般方法及步骤参见下面几题，下面列出过程中的注意点。

1.设计载体：根据目的基因的大小、宿主的不同、宿主中拷贝数的多少、标记基因的不同、是否含有所需的酶切位点、是否需要整合入宿主DNA、目的基因是否需要表达以及转化效率的高低等等需求从而选择合适的载体。

2.宿主的选择：

a.根据研究的目的基因的种类如真核原核或者植物动物等特征进行选择

b.根据研究对象的不同，如细胞内蛋白功能研究或者规模化生产，从而选择合适的宿主。

C．宿主应具备的条件

限制酶缺陷型、重组整合缺陷型、与载体筛选标记互补的表型、感染寄生缺陷型

D．更具体的不同宿主菌株的选择参见http://www.bbioo.com/bio101/2008/23150.htm

3.限制性内切酶的选择：限制性核酸内切酶识别双链DNA分子中的特定序列，并切割DNA双链，常用的都为II型限制性内切酶。根据引物设计过程中加入的不同酶切位点，从而选择想应的限制性内切酶。酶切位点的选择主要由以下角度考虑。

A．引物的需求。

B．一般目的基因用于克隆表达的情况下，目的基因片段内部不含有所选的酶切位点（不然鉴定阳性重组子双酶切时会将目的基因切断）。

C．实验后继应用的所有载体(真核、原核、酵母、昆虫)都尽可能含有所选的酶切位点．并且要保证在载体上的插入方向正确（定向克隆）。（不然换载体表达时，还要重新设计引物，以引进新的酶切位点）。

D．尽可能选比较常用的酶切位点。（常用切点酶的价格比较便宜）

**生物信息学，基因组分析-严军**

**（2017 严军）CpG岛，在基因中存在的概率20%，正常基因CG比例为50%，给一段序列，用贝叶斯方程算出是CpG岛的概率**

该题目争论中

**基因表达与分析-胡苹**

**（2017 胡苹）Fox是一个转录因子，缺失小鼠表现为肥胖。设计实验说明Fox调控靶基因的分子机制。**

（脂肪细胞？）

1.已经明确Fox为转录因子（增强子通过与之结合的蛋白质而发挥作用的蛋白质的称谓），转录因子一般对靶基因的调控机制是通过结合到靶基因的特定序列，招募其他一些转录影响因子来调控靶基因的表达的。常用的研究手法如：ChIP-seq（ChIP-seq，指的是结合位点分析法，作用为研究体内蛋白质与DNA相互作用。染色质免疫共沉淀技术（Chromatin Immunoprecipitation，ChIP）也称结合位点分析法，是研究体内蛋白质与DNA相互作用的有力工具，通常用于[转录因子结合位点](https://baike.baidu.com/item/转录因子结合位点/11013549)或[组蛋白](https://baike.baidu.com/item/组蛋白/604762)[特异性](https://baike.baidu.com/item/特异性/9839939)修饰位点的研究。）。

有较好的Fox抗体时，可以对WT小鼠和Fox KO小鼠的基因组做ChIP-seq(ChIP)实验，找到Fox蛋白在基因组上的结合位点，就可以找出Fox调控的下游靶基因（[起始密码子](https://www.baidu.com/s?wd=起始密码子&tn=SE_PcZhidaonwhc_ngpagmjz&rsv_dl=gh_pc_zhidao)之前的成为上游基因，之后的成为下游基因。下游基因含有增强子原件，所以也会[调节基因](https://www.baidu.com/s?wd=调节基因&tn=SE_PcZhidaonwhc_ngpagmjz&rsv_dl=gh_pc_zhidao)的表达。上有基因中包含启动子操纵子之类的所以也会调控[目的基因](https://www.baidu.com/s?wd=目的基因&tn=SE_PcZhidaonwhc_ngpagmjz&rsv_dl=gh_pc_zhidao)的表达。）。

2.在KO (敲除小鼠)与WT（野生型小鼠）中，在mRNA水平（RTqPCR实时荧光定量PCR）和蛋白水平（WB，免疫印迹法）观察这些靶基因的表达量。从而得知Fox是激活这些靶基因的表达，还是抑制他们。

3.对结合位点构建luciferase报告基因系统（荧光素酶报告基因是指以荧光素(luciferin)为底物来检测萤火虫荧光素酶(fireflyluciferase)活性的一种报告系统。荧光素酶可以催化luciferin氧化成oxyluciferin，在luciferin氧化的过程中，会发出生物荧光(bioluminescence)。），可以通过KO fox或过表达（过表达，即表达过度，当基因表达（转录）的严格控制被打乱时，基因可能不恰当被“关闭”，或以高速度进行转录。高速转录导致大量mRNA产生，大量蛋白质（protein）产物出现。）Fox，研究Fox对靶基因的影响是否直接，并且是何影响。

4….MS IP (免疫沉淀和质谱联用(IP-MS)，可用来确认抗体与预期靶标的特异性相互作用。在现有的抗体验证方法中，这种方法的特别之处在于可以主动识别和定量纯化自样本的实际抗体靶标。与其他评估抗体特异性的技术（如，通过对基因敲除小鼠来源样本进行免疫印迹染色，确定某个消失的条带）不同，质谱法可以直接确定来源于样本蛋白的肽段序列。因此，IP-MS是确认抗体与预期靶标直接结合的首选方法。)

5.in vitro transcription 转录调控的验证（转录调控是指通过改变转录速率从而改变基因表达的水平，其对遗传信息的传递的准确性和多样性具有重要的作用）

**（2016 胡苹) 通过查阅文献和数据库发现，在远端肌肉萎缩患者中转录因子Msx1突变比例增高，同时基因组中的一段目前已知不编码的3kb区域也是突变高发区。请根据上述初步结果设计实验，证实Msx1和突变高发的3kb非编码区域是否与远端肌肉萎缩的发病相关。如果实验结果证明相关，请设计实验说明Msx1和突变高发的3kb非编码区域在远端肌肉萎缩中的功能。**

1.验证患者和正常人的Msx1基因表达差异-蛋白水平，mRNA水平；以及3kb非编码区在患者中的突变概率

2.在正常细胞中引入突变，肌肉细胞的表型出现问题；纠正两种突变，肌肉细胞的表型有所恢复？

3.转录因子：ChIP-seq，报告基因（指一类在细胞、组织/ 器官或个体处于特定情况下会表达并使得他们产生易于检测、且实验材料原本不会产生的性状的基因）

4.非编码区：RNA-IP，RIP-seq？？ enhancer-trap；chip-loop；

5.in vitro transcription 转录调控的验证

**（2014 胡苹）前期实验结果提示常发生于肾脏的疾病甲可能是由于转录因子F发生突变，因而导致对于下游的靶基因调控失灵造成的。F下游的靶基因包括编码蛋白和非编码RNA。请设计是按验证上述假说，并阐述F对靶基因调控失灵诱发疾病甲的可能机制。请写明实验设计的逻辑推理过程，指明每步所需试验的名称和相关的正对照和负对照。**

A对基因B和RNAC的影响。 对照：证明患病时A在mRNA水平和蛋白质水平表达都升高，过表达A敲出A观察影响。 基因修饰（甲基化乙酰化磷酸化泛素化）染色体结构，SNP，转录因子，将编码转录因子的geneA构建成一个质粒，基因B构建成另一个质粒共转染细胞 看蛋白B的表达 利用凝胶电泳迁移率实验，放射性标记RNAC，将纯化的蛋白A和RNAC一同保温，在非变性的聚丙烯凝胶电泳中，结合的跑得慢。

**（2011 胡萍）初步实验证明细胞在生长因子的作用下，可产生核蛋白X，核蛋白X可能影响基因A的表达。请设计实验证明核蛋白X是否直接与基因A的转录因子结合，并证明核蛋白X对基因A的影响，以及如何影响。**

（1）核蛋白X和基因A转录因子互作验证

酵母双杂交（酵母双杂交系统是将待研究的两种蛋白质分别克隆（融合）到酵母表达质粒的转录激活因子（如GAL4等）的DNA结合结构域（DNA-BD）和转录激活域（AD）上，构建成融合表达载体，从表达产物分析两种蛋白质相互作用的系统。），构建核蛋白X基因+GAL4 DB和基因A转录因子编码基因+GAL4 转录因子结合域+基因A两个重组载体，做两个对照，GAL4对基因A的表达影响以及转录因子对基因A的影响，再根据基因A的表达差异性判断是否存在直接相互作用

（2）核蛋白X和基因A

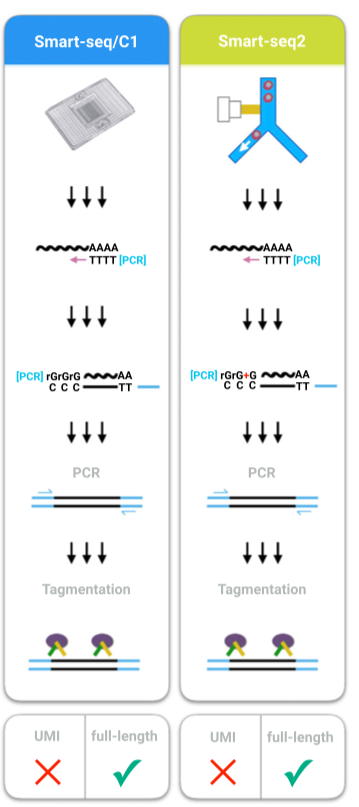
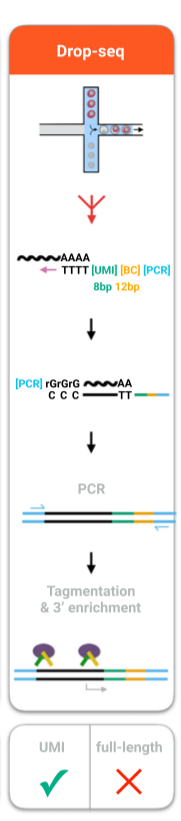
敲除核蛋白X编码基因，方法有crisper CAS 9、RNAi（RNA干扰（RNA interference, RNAi）是指在进化过程中高度保守的、由双链RNA（double-stranded RNA，dsRNA）诱发的、同源mRNA高效特异性降解的现象。）等，观察基因A的表达情况，上调或下调或不变过表达核蛋白X，对比单独表达A，对比

**单细胞测序-何杰（预测）**

Smart-seq和10xGenomics单细胞测序建库扩增的原理。

Smart-seq 1/2 (Switching mechanism at 5’ end of the RNA transcript)

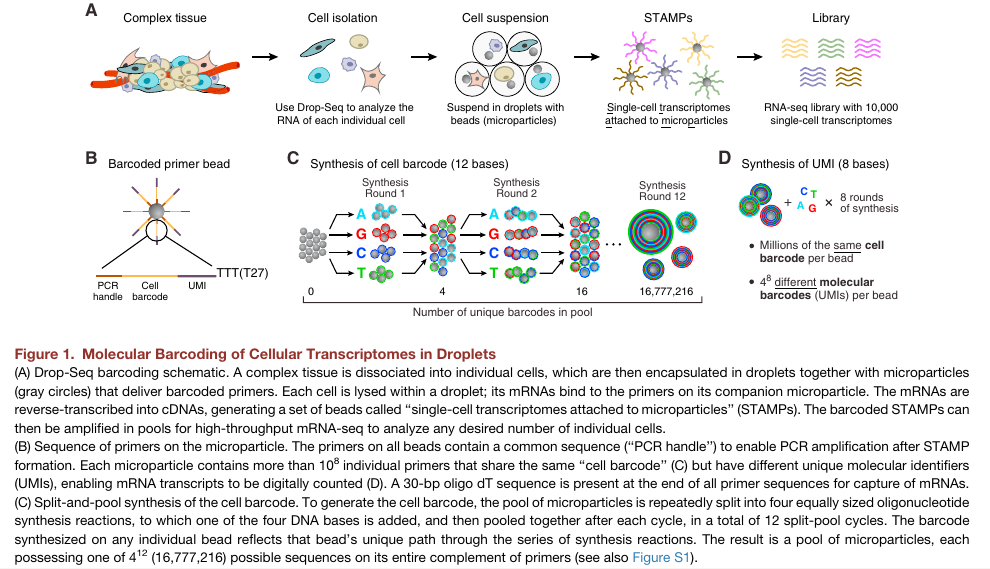
没有UMI（Unique molecular identifiers (UMIs), or molecular barcodes (MBC) are short sequences or molecular "tags" added to DNA fragments in some next generation sequencing library preparation protocols to identify the input DNA molecule. ），可以**测全长**，Fluidigm C1（一种微流体芯片系统，Smart-seq/C1）或者FACS（荧光激活细胞分离技术 (fluorescence-activatedcell sorting, FACS)是高纯度识别 、分离稀有细胞群体的专项技术 ，Smart-seq2）先通过流式分选或者C1微流体芯片技术获得单细胞，裂解细胞使细胞内RNA与一段polyT(poly(dT))的引物（引物，是指在核苷酸聚合作用起始时，刺激合成的一种具有特定核苷酸序列的大分子，与反应物以共价键形式连接，这样的分子称为引物）杂交，逆转录得到cDNA，所用的MMLV逆转录酶（M-MLV反转录酶，无RNase H活性。 具有合成效率高（较多的全长cDNA), 热稳定性好和半衰期长的特点。无RNaseH活性，避免了第一链cDNA合成反应中DNA/RNA杂交体中模板RNA被降解，从而保证第一链cDNA合成量和长度。）会在3‘末端多添加几个C，生成第一条链。再利用有连续G碱基的上游引物（forward primer，也称正向引物，是沿着负链（1.正链RNA介导合成负链RNA,负链RNA再介导合成正链DNA.2.负链RNA直接介导合成正链DNA.图示为：(+RNA) ---> (-RNA) ---> (+DNA)）进行不间断延长的）进行第二链的合成，保证了双链cDNA是全长cDNA。再利用**两端**的PCR引物进行扩增。



10xGenomics基于drop-seq

有UMI，有Barcode，不能测全长3’countin，有微流控系统Droplets，获得单细胞悬液后，通过10X的微流控系统，对每一个细胞形成油包水的微反应体系。“双十字”交叉系统，将含有barcode与UMI的beads（磁珠，Fig1.A的第三个图和第四个图灰色珠子）和酶与细胞装配起来，形成GEM(油滴，油包水，Fig 1.A 第三个图的那些大圈)。在油滴内进行逆转录反应，凝胶珠溶解，细胞裂解释放mRNA，通过逆转录产生用于测序的带条形码的cDNA。通过Barcode与UMI分别作为单细胞和单RNA的标签。然后破坏油包水的结构，混合的cDNA进行后续的文库构建，用PCR进行cDNA扩增。接着是tagmentation（华盛顿大学的Jay Shendure和Andrew Adey针对基因组测序对于测序文库的需要，开发出一种名为 Tagmentation的方法，该方法利用Tn5转座子将DNA片段化，并同时掺入接头。与常规文库构建和连接方法相比，转座子酶方法更加高效，也减少了所需的DNA起始量。随后将该方法优化改进后用于全基因组甲基化的测序分析（常规要求起始DNA量在5ug以上），同样在不减少覆盖度的前提下，大大减少了所需的起始DNA量。）和富集（基因富集分析(gene set enrichment analysis)是在一组基因或蛋白中找到一类过表达的基因或蛋白。一般是高通量实验，如基因芯片，RNA-Seq，蛋白质组学（质谱结果）的后续步骤。）得到3‘端的cDNA ，纯化过程。进行普通的加接头和测序过程。







其实都是template switching-based PCR

10X genomics(Drop-seq)与Smart-seq相比：

敏感度：smart-seq好于10X

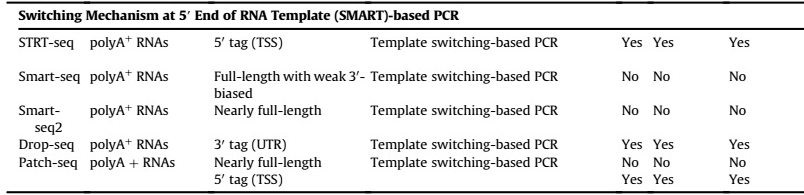
Accuracy：相差无几

Precision：smart-seq好于10X

成本：10X比smart-seq便宜很多

10X并不能像Smart-seq那样测到全长mRNA，但是通量比Smart-seq大很多。Smart-seq 100个细胞，10X 可以测1-10k。





**生物芯片技术及其应用-肖华胜(不可能考了，全村的希望提供的消息)**

**（2014 肖华胜）请简述基因芯片的基本原理，主要制备方法以及在生物医学研究中的主要应用（2013 在生命科学中的应用）。**

概念：是通过微阵列技术microarray将高密度DNA片段通过高速机器人或原位合成方式以一定的顺序或排列方式使其附着在如膜、玻璃片等固相表面，以同位素或荧光标记的DNA探针，借助碱基互补杂交原理（DNA分子杂交的基础是，具有互补碱基序列的DNA分子，可以通过碱基对之间形成氢键等，形成稳定的双链区。在进行DNA分子杂交前，先要将两种生物的DNA分子从细胞中提取出来，再通过加热或提高pH的方法，将双链DNA分子分离成为单链，这个过程称为变性。然后，将两种生物的DNA单链放在一起杂交，其中一种生物的DNA单链事先用同位素进行标记。如果两种生物DNA分子之间存在互补的部分，就能形成双链区。由于同位素被检出的灵敏度高，即使两种生物DNA分子之间形成百万分之一的双链区，也能够被检出。），进行大量的基因表达及监测等方面研究的最新革命性技术。

制作方法：一种为原位合成，常见有光蚀保护，ink-jet，digital micromirror device，分子印章等方法；另一种为合成后点样（微点阵），多用于cDNA PCR产物，oligo，其他形式的基因产物。

基因芯片在生命科学研究中的应用：

1、 基因表达分析

同一组织细胞在不同的发育阶段、不同外界环境因素影响下，其基因的表达模式不同。同一个体的不同组织器官的基因表达模式也不相同。基因芯片技术由于具有高度并行性和高通量的特点正适合于此项研究。

2 、基因突变和多态性分析

在进化过程中同一物种不同种群和个体之间存在着不同的基因型，这些不同的基因型与生物个体间不同的性状有着密切的关系。利用基因芯片可以对基因型与性状之间的关系进行研究。

3、 基因组DNA序列分析

由A、T、C、G4种核苷酸单体组合所形成的所有可能把具体寡核苷酸(寡核苷酸，是一类只有20个以下碱基的短链核苷酸的总称（包括脱氧核糖核酸DNA或核糖核酸RNA内的核苷酸），寡核苷酸可以很容易地和它们的互补链对接，所以常用来作为探针确定DNA或RNA的结构，经常用于基因芯片、电泳、荧光原位杂交等过程)探针共 有65536种。将这些种类的探针全部固定于活化的载体表面，形成寡核苷酸阵列。然后芯片与样品进行杂交，通过计算机对杂交模式进行分析，就可以得出样品的核苷酸序列信息。

4、 核酸和蛋白质相互作用的研究

蛋白质与特定的核酸片段结合对基因的表达起着重要的调控作用，通过对蛋白质与核酸相互作用的研究，人们可以更深入地了解生命活动的内在机制。

5.DNA甲基化(DNA甲基化（DNA methylation）为DNA化学修饰的一种形式，能够在不改变DNA序列的前提下，改变遗传表现。所谓DNA甲基化是指在DNA甲基化转移酶的作用下，在基因组CpG二核苷酸的胞嘧啶5'碳位共价键结合一个甲基基团。)芯片

6.临床诊断芯片