7 蛋白质组学在肾病研究中的应用

7.1 肾癌的分型与发病机制

7.1.1 肾癌的分型:

在诊断和治疗上， 2004年WHO提出了肾癌的分类标准，表7.1列出了较为常见的肾癌类型。80%以上成年人患的肾癌属于肾细胞癌，肾细胞癌中透明细胞癌、乳头状肾细胞癌和嫌色细胞癌是最常见的三种类型。约4%的肾癌是因为遗传或家族基因倾向引起的[8]，表格7.2中列出了和遗传综合症相关的的肿瘤类型，每一种综合症都引起一种从组织形态学上特异的肾细胞癌或其他肾癌，从临床角度分析，相比于非家族遗传性肾肿瘤，遗传性的肾癌表现为发病年龄提前、多病灶和双侧肾[8]。

7.1.2 肾癌的发病机制:

主要的肾细胞癌，包括肾细胞癌，乳头状肾细胞癌和嫌色肾细胞癌发生和发展的基因分子机理已有大量的报道。

1. 肾透明细胞癌

肾透明细胞癌（clear cell renal cell carcinoma，ccRCC）占所有肾细胞癌的70%，组织学上，该肿瘤具有透明的细胞质，细胞被浓稠的内皮细胞网络所包裹，组成网状结构。ccRCC经常发生在60岁左右的人群中，并且病人大部分是男性，男女比例为2:1。大部分ccRCC是散发性的，只有2-4%的病例来自家庭遗传，包括VHL综合症，BHD（Birt-Hogg-Dube）综合症和3号结构染色体易位综合症。

表7.1 肾癌分类（来自2004年WHO分类标准）

|  |  |  |  |
| --- | --- | --- | --- |
| **肾癌亚类** | **肾癌具体类别** | **恶性程度** | |
| 肾细胞癌  （Renal cell tumors） | 肾透明细胞癌 | | 恶性肿瘤 |
| 多房囊性肾细胞癌 | |
| 乳头状肾细胞癌 | |
| 嫌色肾细胞癌 | |
| Bellini收集管细胞癌 | |
| 肾髓质癌 | |
| Xp11易位肾细胞癌 | |
| 与成神经细胞瘤相关的细胞癌 | |
| 黏液性管状梭形细胞癌 | |
| 未分类肾细胞癌 | |
| 乳头状腺瘤 | | 良性肿瘤 |
| 大嗜酸粒细胞瘤 | |
| 后肾肿瘤  （Metanephric tumor） | 后肾腺瘤 | | 良性肿瘤 |
| 后肾腺纤维瘤 | |
| 后肾间质瘤 | | 恶性肿瘤 |
| 混合性间质细胞和上皮细胞肿瘤 | 囊性肾瘤 | | 恶性肿瘤 |
| 混合性上皮细胞和间质肿瘤 | |
| 滑膜肉瘤 | |
| 肾母细胞瘤（Nephroblastic tumor） | 肾源性残余肿瘤 | | 恶性肿瘤 |
| 肾胚细胞瘤 | |
| 囊性部分分化的肾胚细胞瘤 | |
| 神经内分泌瘤 | 良性肿瘤 | | 恶性肿瘤 |
| 神经内分泌癌 | |
| 原始神经外胚层肿瘤 | |
| 成神经细胞瘤 | |
| 嗜铬细胞瘤 | |
| 其他类型肾肿瘤 | 间叶细胞肿瘤 | | 恶性肿瘤 |
| 造血和淋巴细胞肿瘤 | |
| 生殖细胞肿瘤 | |
| 转移瘤 | |

70-90%的ccRCC病人的染色体3p上都会出现改变，包括一些重要基因的缺失、突变或者甲基化，如位于染色体3p25-26位置上的*VHL*基因，位于3p21上的*RASSF1*基因和位于3p14.2上的*FHIT*基因；重复的5q22是ccRCC病人中另一个普遍存在的现象；另外一些体细胞基因上的变化包括染色体6q、8p12、9p21、9q22、10q、17p和14q的缺失[9]。

表7.2 家族性肾癌与相关遗传综合征[33]

|  |  |  |
| --- | --- | --- |
| **综合征** | **涉及的基因** | **肿瘤类型** |
| Von Hippel-Lindau (VHL) | *VHL* （3p25） | 肾透明细胞癌 |
| 结节状硬化 | *TSC1*，*TSC2* | 血管肌脂肪瘤，肾透明细胞癌，其他 |
| 组成型3号染色体易位 | 相关基因未鉴定出来 | 肾透明细胞癌 |
| 家族性肾癌 | 相关基因未鉴定出来 | 肾透明细胞癌 |
| 遗传性乳头状肾细胞癌 | *C-MET* | 乳头状肾细胞癌类型1 |
| Birt-Hogg-Dube (BHD) | *BHD* | 嫌色肾细胞癌 |
| 家族性大嗜酸粒细胞瘤 | 部分或全部多染色体缺失 | 大嗜酸粒细胞瘤 |
| 遗传性平滑肌瘤肾细胞癌 | *FH* | 乳头状肾细胞癌类型2 |

18-82%的散发性ccRCC病人具有*VH*L基因的体细胞突变， 而98%的病人具有*VHL*位点杂合性缺失[10]。在5-20%的*VHL*未发生基因突变的ccRCC病人中，存在*VHL*基因启动子高度甲基化从而导致VHL基因失活。VHL蛋白是一种E3泛素化连接酶复合物的一个组分，在细胞缺氧应激的调节中起着关键性作用，其在细胞内的作用机制如图7.2所示。缺氧诱导因子HIF（hypoxia-inducible factor）是一个转录因子，它在细胞内的浓度受到VHL的调节。在氧气充足的情况下，HIF被羟基化，VHL蛋白和羟基化的HIF结合，促进HIF在蛋白酶体内的降解。因此，在正常情况下，VHL调控HIF的降解使其保持很低的浓度。在缺氧情况下，VHL无法识别非羟基化HIF，导致HIF的积累，激活下游的缺氧驱动基因，包括一些促血管生成基因如促血管生成因子和血小板源生长因子，及细胞生长和抗凋亡基因，红细胞生成基因等。致使细胞内多个信号转导通路的激活，包括PI3K-Akt-mTOR和Ras-raf-erk-mek信号通路，这些通路参与了细胞增殖、生存和分化的调节、促进血管生成以及代谢异常[11]。*VHL*基因突变或者其启动子的高度甲基化影响了蛋白的稳定性和表达，使得VHL不能介导降解HIF，而HIF持续激活下游通路，增加细胞增殖和分化，并上调参与血管生成、细胞迁移和代谢的基因。

(2) 肾乳头状细胞癌

肾乳头状细胞癌（Papillary RCC，PRCC）是第二大类高发的肾细胞癌，其比例占肾细胞癌的10-15%，男女发病比例和ccRCC类似，PRCC的五年生存率可达到90%。现在已经确定的PRCC有两种，PRCC1和PRCC2。PRCC的细胞质一般呈现嗜碱性，呈现泡沫状组织形态。大部分PRCC都是散发性的，极个别属于遗传性的。

图7.1 VHL参与的细胞内分子通路

注：在氧气充足的条件下，VHL通过识别HIF1-α上脯氨酸的羟基化，介导其泛素化途径的降解；在低氧条件下，HIF1-α不能发生羟基化修饰，从而在在细胞中积累并持续发挥作用，激活下游的信号转导通路；而发生基因突变的VHL失去了降解HIF1-α的功能，导致下游通路的持续激活。

目前对PRCC发生的分子机理了解得很少，家族遗传综合症可能会增加PRCC的风险。*MET*原癌基因的突变使患者倾向于发生多病灶的PRCC1，同时延胡索酸脱氢酶基因的突变也会增加家族性PRCC2发生的风险[12]。PRCC的基因组学还需要大规模的基因测序和分析，以确定PRCC的发病机理，并发现合适的诊断标志物和治疗靶标。

## 7.2蛋白质组数据分析方法

### 7.2.1蛋白质组大数据的定性方法

要分析和挖掘蛋白质组大数据，首先要弄清楚数据中包含了哪些蛋白质和肽段，发生了哪些翻译后修饰，也就是说首先要解决数据的定性问题。在基于质谱的蛋白质组数据鉴定分析方法中，有两种常用的鉴定策略：

1. 基于bottom-up（自底向上）的鸟枪法策略：通常将蛋白质复杂样本进行酶切，得到以小肽段（5kDa以下）为主的混合物，再经液相色谱等技术分离后，送入质谱进行碎裂。这里的鉴定问题首先是要解决质谱产出的实验谱图如何匹配上相应的肽段序列。常用的谱图-序列匹配算法可以分为三大类：一是蛋白质序列数据库搜索。作为最常用的一种鉴定方法，该方法是将序列数据库进行理论酶切和碎裂后生成理论谱图，与实验谱图进行匹配，如图10-1所示。二是从头测序（*de novo*）法，直接通过实验谱图中碎片离子的质量数组合来推测实验谱图所对应的肽段序列，不需要与数据库进行匹配。这种方法要求质谱数据具有较高的分辨率和较完整的碎裂状态，但实际情况中实验谱图往往包含不完全碎裂离子以及干扰噪声，因此该方法目前还有待进一步发展。三是序列标签查询的方法，可以看做是前两种方法的结合。该方法首先利用实验谱图中的部分碎片离子得到几个氨基酸组成的肽段片段，然后利用序列数据库搜索该片段，从而得到肽段的全序列。

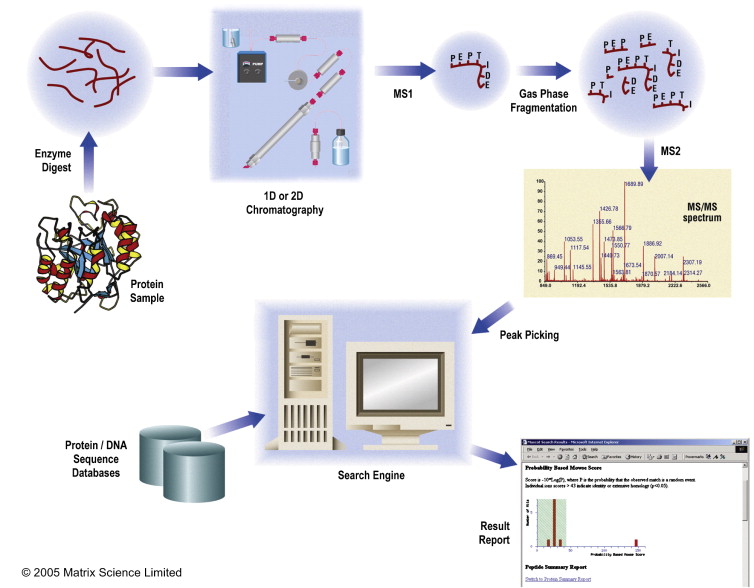


图7-2基于鸟枪法的质谱鉴定流程[[9](#_ENREF_9" \o "Cottrell, 2011 #2426)]

基于top-down（自顶向下）的策略：该策略不需要酶切，直接以完整蛋白质作为质谱检测对象，该策略所对应的蛋白质分离技术、质谱检测技术以及生物信息学分析方法都与上述的鸟枪法策略不尽相同。这里我们只阐述基于top-down策略的生物信息学分析方法。与基于鸟枪法的质谱数据相比，基于top-down策略的质谱数据产生的谱图要更复杂：同位素峰所带电荷更大，离子质量范围更宽、大质量离子的同位素峰数也更多，谱峰混合叠加的现象也更加严重。因此，在鉴定之前首先需要对谱图进行预处理，包括去噪、去卷积等操作。然后再采用与上述一样的三类搜索策略对top-down策略的质谱数据进行鉴定。该策略能够直接针对蛋白质以及蛋白质变体（proteoform）进行分析，适用于研究多种翻译后修饰条件下的蛋白质。

随着PCOS代谢组学的研究深人，在今后的PCOS病因学研究中，将代谢组学数据与以往其他组学(如基因组学、转录组学、蛋白质组学等)数据进行综合与整合，将可以获得更完整的机体生物学图像和功能性的生物标志物，对PCOS的早期预防及针对性临床治疗具有重要意义。

表7-3 GC-TOF-MS方法检测得多囊卵巢综合征组和对照组血浆中代谢物的水平差异[62]

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 保留时间, min | 代谢物 | 对照组 | | A 组(HA+AO+PCO) | | | | | | B 组 (AO+PCO) | | | | | C 组 (AO+HA) | | | | | D 组 (HA+PCO) | | | |
| 中位值± SD | | P adjusted | | | | 中位值± SD | | P adjusted | | | 中位值± SD | | P adjusted | | | 中位值± SD | | P adjusted | |
| 6.249 | 乳酸 | 2.17 | ± 0.27 | 2.29 | ± 0.46 | 0.039 | | | U | 2.5 ± 0.54 | | 0.026 | U | | 2.03 | ± 0.61 | 0.079 | - | | 2.82 | ± 0.77 | 0.035 | U |
| 25.818 | 葡萄糖 | 4.96 | ± 0.78 | 4.03 | ± 0.32 | 0.019 | | | D | 4.02 | ± 0.41 | 0.034 | D | | 4.29 | ± 0.62 | 0.054 | - | | 4.27 | ± 0.78 | 0.04 | D |
| 26.184 | 乳糖 | 1.94 | ± 0.29 | 1.76 | ± 0.13 | 0.582 | | | - | 1.69 | ± 0.13 | 0.45 - |  | | 1.87 | ± 0.43 | 1 | - | | 1.87 | ± 0.4 | 1 | - |
| 28.818 | 软脂酸 | 0.72 | ± 0.12 | 0.76 | ± 0.12 | 0.876 | | | - | 0.91 | ± 0.11 | 0.001 | U | | 0.95 | ± 0.15 | <0.001 | U | | 0.82 | ± 0.13 | 0.014 | U |
| 29.751 | 尿酸 | 0.27 | ± 0.13 | 0.58 | ± 0.17 | <0.001 | | | U | 0.3 | ± 0.21 | 1 | - | | 0.16 | ± 0.15 | 0.088 | - | | 0.31 | ± 0.18 | 1 | - |
| 31.785 | 亚油酸 | 0.05 | ± 0.03 | 0.1 ± 0.09 | | 0.003 | | | U | 0.36 | ± 0.18 | 0.001 | U | | 0.5 ± 0.21 | | <0.001 | U | | 0.15 | ± 0.16 | 0.019 | U |
| 32.418 | 硬脂酸 | 0.85 | ± 0.14 | 0.91 | ± 0.08 | 0.17 - | | |  | 0.98 | ± 0.08 | 0.023 | U | | 0.96 | ± 0.12 | 0.067 | U | | 0.94 | ± 0.10 | 0.01 | U |
| 45.636 | 胆固醇 | 0.72 | ± 0.35 | 0.69 | ± 0.33 | 0.043 | | | D | 0.68 | ± 0.26 | 0.038 | D | | 0.8 | ± 0.34 | 0.805 | - | | 0.96 | ± 0.73 | 0.342 | - |
| 7.149 | 丙氨酸 | 0.43 | ± 0.14 | 0.56 | ± 0.1 | <0.001 | | | U | 0.63 | ± 0.11 | 0.001 | U | | 0.61 | ± 0.6 | <0.001 | U | | 0.48 | ± 0.17 | 0.813 | - |
| 9.866 | 缬氨酸 | 0.46 | ± 0.07 | 0.52 | ± 0.1 | <0.001 | | | U | 0.57 | ± 0.07 | 0.001 | U | | 0.53 | ± 0.11 | 0.001 | U | | 0.5 | ± 0.08 | 0.042 | U |
| 11.3 | 亮氨酸 | 0.28 | ± 0.09 | 0.28 | ± 0.06 | 1 | | | - | 0.29 | ± 0.09 | 1 | - | | 0.32 | ± 0.07 | 0.399 | - | | 0.37 | ± 0.11 | 0.002 | U |
| 11.85 | 异亮氨酸 | 0.19 | ± 0.08 | 0.18 | ± 0.06 | 0.195 | | | - | 0.19 | ± 0.05 | 1 | - | | 0.18 | ± 0.07 | 1 | - | | 0.12 | ± 0.05 | 0.001 | D |
| 11.933 | 脯氨酸 | 0.59 | ± 0.07 | 0.28 | ± 0.08 | <0.001 | | | D | 0.25 | ± 0.08 | 0.001 | D | | 0.26 | ± 0.08 | 0.001 | D | | 0.19 | ± 0.07 | 0.001 | D |
| 12.15 | 甘氨酸 | 0.59 | ± 0.09 | 0.49 | ± 0.08 | 0.004 | | | D | 0.53 | ± 0.08 | 0.026 | D | | 0.5 | ± 0.09 | 0.008 | D | | 0.47 | ± 0.11 | 0.001 | D |
| 13.6 | 丝氨酸 | 0.29 | ± 0.08 | 0.38 | ± 0.07 | <0.001 | | | U | 0.38 | ± 0.07 | <0.001 | U | | 0.35 | ± 0.09 | 0.012 | U | | 0.24 | ± 0.08 | 0.015 | D |
| 14.233 | 苏氨酸 | 0.35 | ± 0.11 | 0.42 | ± 0.08 | <0.001 | | | U | 0.42 | ± 0.08 | 0.001 | U | | 0.39 | ± 0.1 | 0.085 | U | | 0.27 | ± 0.09 | 0.006 | D |
| 16.317 | 天冬氨酸 | 0.13 | ± 0.09 | 0.13 | ± 0.09 | 1 | | | - | 0.2 ± 0.14 | | 0.005 | U | | 0.12 | ± 0.06 | 1 | - | | 0.08 | ± 0.07 | 1 | - |
| 19.867 | 苯丙氨酸 | 0.1 | ± 0.05 | 0.24 | ± 0.08 | <0.001 | | | U | 0.21 | ± 0.08 | 0.001 | U | | 0.19 | ± 0.08 | 0.001 | U | | 0.11 | ± 0.05 | 1 | - |
| 24.118 | 鸟氨酸 | 0.13 | ± 0.08 | 0.25 | ± 0.09 | <0.001 | | U | | 0.23 | ± 0.09 | 0.001 | U | | 0.2 ± 0.1 | | 0.002 | U | | 0.12 | ± 0.08 | 1 | - |
| 26.334 | 赖氨酸 | 0.28 | ± 0.9 | 0.5 ± 0.18 | | <0.001 | | | U | 0.39 | ± 0.11 | 0.126 | - | | 0.34 | ± 0.10 | 1 | - | | 0.33 | ± 0.26 | 1 | - |
| 26.634 | 酪氨酸 | 0.09 | ± 0.04 | 0.24 | ± 0.06 | 0.001 | | | U | 0.2 ± 0.07 | | 0.001 | U | | 0.17 | ± 0.07 | 0.036 | U | | 0.14 | ± 0.07 | 1 | - |
| 31.7 | 色氨酸 | 0.02 | ± 0.01 | 0.17 | ± 0.04 | 0.001 | | | U | 0.09 | ± 0.06 | 0.003 | U | | 0.05 | ± 0.04 | 0.03 | U | | 0.06 | ± 0.05 | 0.002 | U |
|  | 内源性氨基酸 | 3.8 | ± 0.71 | 4.37 | ± 0.62 | 0.001 | | | U | 4.32 | ± 0.64 | 0.001 | U | | 4.0 | ± 0.73 | 0.294 | | - | 3.38 | ± 0.78 | 0.153 | - |
|  | 糖异生氨基酸 | 3.4 | ± 0.62 | 3.93 | ± 0.52 | 0.003 | | | U | 4.14 | ± 0.67 | <0.001 | U | | 3.6 | ± 0.70 | 1 | | - | 2.92 | ± 0.65 | 0.002 | D |
|  | 支链氨基酸BCAA | 0.87 | ± 0.2 | 0.95 | ± 0.22 | 1 | | | - | 1.03 | ± 0.17 | 0.024 | U | | 1.04 | ± 0.20 | 0.017 | | U | 0.98 | ± 0.17 | 0.865 | - |
|  | 氨基酸AAA | 0.21 | ± 0.1 | 0.66 | ± 0.17 | 0.001 | | | U | 0.49 | ± 0.17 | 0.001 | U | | 0.4 ± 0.16 | | 0.019 | | U | 0.31 | ± 0.17 | 1 | - |
|  | BCAA/AAA | 4.82 | ± 1.8 | 1.51 | ± 0.37 | <0.001 | D | | | 2.35 | ± 0.91 | <0.001 | | D | 3.09 | ± 1.62 | 0.002 | | D | 4.3 | ± 2.50 | 0.298 | - |

数据侧的U (up) and D (down)表示代谢物浓度的升高和降低。内源性氨基酸包括除鸟氨酸外的所有氨基酸；糖异生氨基酸包括丙氨酸、缬氨酸、异亮氨酸、脯氨酸、甘氨酸、丝氨酸、苏氨酸、天冬氨酸、苯丙氨酸、赖氨酸和酪氨酸；支链氨基酸包括亮氨酸、缬氨酸和异亮氨酸；芳环氨基酸包括苯丙氨酸、酪氨酸和色氨酸。AO=无排卵，HA=高雄激素，PCO=多囊卵巢样

## 7.3 第四代DNA测序技术

虽然第三代测序技术已从研发走向了应用，但其测序通量和成本依然和100美元测一个人基因组的宏伟目标存在很大距离。提高测序的同类和降低测序的成本依然是技术研发人员不懈的追求，第四代测序技术既固态纳米孔测序技术的研究应运而生。

### 7.3.1 可能利用的测序原理[34]

作为单分子测序技术，生物纳米孔在稳定性、电流噪声等方面的缺陷一定程度上限制了其发展。相比于生物纳米孔，固态纳米孔测序在稳定性、电流噪声、工艺集成方面有着显著的优势，因而被认为是最具前景的DNA测序技术。固态纳米孔主要是利用硅及其衍生物制造而成，一般使用离子束或电子束在硅或其他材料薄膜表面钻出纳米尺度的孔洞，再进一步对孔的形状和大小进行修饰而成。固态纳米孔测序原理与生物纳米孔相似，以电阻脉冲技术为原型，纳米孔测序法通过在一定电压下检测分析物通过纳米孔时产生的电流变化来判断分析物的结构特性。固体薄膜可以起到很好的绝缘作用，可以将密封容器里的电解质溶液分隔成两部分，仅通过镶嵌在薄膜上的纳米孔将薄膜两边的电解质溶液连接起来。在一定电压下，DNA通过纳米孔时，C、A、T、G四种碱基化学性质的差异会使得他们引发不同的电学参数变化量。通过收集DNA过孔期间的离子电流信号则可以判断不同时刻过孔碱基的类型，从而推断出此该DNA分子的序列。

常用的利用电学信号实现纳米孔技术对DNA序列进行检测有三种方法。第一种是离子电流检测法。当DNA等分子在电压作用下单向通过纳米孔时，会阻塞部分纳米孔，这时电解质溶液中的离子通量降低，离子电流减弱，因此可以根据离子电流的大小检测DNA。第二种是隧道电流检测法。对于传统的纳米孔，离子电流法不能直接检测DNA上单个碱基的信息，所以可以在固态纳米孔上嵌入电极并施加电压，根据检测到的隧道电流来识别单个碱基。第三种方法是电容检测法，即通过碱基通过纳米孔时引起的纳米电容的电荷量变化来进行DNA测序。此类应用在国际上取得了一定的的进展，如Cees Dekker 等利用光学钳子来控制DNA 进入和在纳米孔内的移动；Xinsheng Sean Ling使用磁学镊子控制DNA在纳米孔内的运动。与生物纳米孔类似的是，固态纳米孔也需要解决DNA过孔速率太快导致信号难以分析的困扰。

### 7.3.2固态纳米孔加工工艺[35]

随着半导体工业技术的飞速发展，固态纳米孔逐渐展露出其在DNA测序领域的巨大的优势。固态纳米孔的基底材料主要为硅及其衍生物，其制作方法一般通过电子束刻蚀和击穿电压打孔两种方式，利用离子束或者电子束在薄膜材料上制备出纳米尺度的空洞，再进一步对孔的形状和大小进行修饰。虽然纳米孔结构简单，但要实现纳米孔的直径大小的精度控制，并实现高重复性、规模化制造的要求，是学术和企业界双方都迫切寻求的答案。对于高校研究所的基础前沿探索研究来说，利用透射电子显微技术（TEM），氦离子束显微技术（HIM）等前沿技术可以实现灵活多样的实验需求。在2001 年，Jene Golovchenko课题组利用3 KeV的氩离子束和Ar修饰在氮化硅薄膜上制作出第一个直径为1.8 nm的纳米孔。2003年，Arnold Storm等利用高能电子束在SiO2薄膜上制作出了直径 2 nm 的纳米孔。Gregory Timp和Cees Dekker 几乎同时提出利用透射电子显微镜在氮化硅和二氧化硅薄膜上制作纳米孔的方法。中科院物理所陆兴华小组和王鹏业课题组也对利用聚焦离子束制备基于氮化硼薄膜的纳米孔进行了研究。然而对于未来的市场化工业化应用而言，需要和大规模CMOS半导体生产技术相兼容，从而降低生产成本。中国科学院重庆绿色智能技术研究院王德强研究员所在的课题组率先制作出CMOS工艺相兼容多层纳米孔结构-DNA晶体管，并在8英寸的半导体生产线上流片成功，纳米孔的直径控制在18±2纳米（参见图7-3）。值得一提的是，由于纳米孔形成在几十纳米的绝缘薄膜上面，薄膜的寄生电容将会严重影响纳米孔的相应速度，降低纳米孔的检测灵敏度，王德强课题组利用电学模型对于单层和多层的纳米孔结构进行了不同频率的噪音测量，提取纳米孔的特性参数，提出了优化的模型结构，并进行了实验制作和验证。纳米孔制备技术的长足发展使得在不同材料上制备各种尺寸纳米孔的工艺得以实现，新型材料的研发也让纳米孔薄膜材料的选择更加多样化。如今，人们已经可以在很多材料上制作出亚10纳米尺度的固态纳米孔，例如，SiNx，SiO2，SiC，Al2O3等。此外，石墨烯因其本身超薄的结构和特殊的电子特性也作为薄膜材料的一种新选择，它的超薄的单原子层结构十分适合隧道电流的测量。

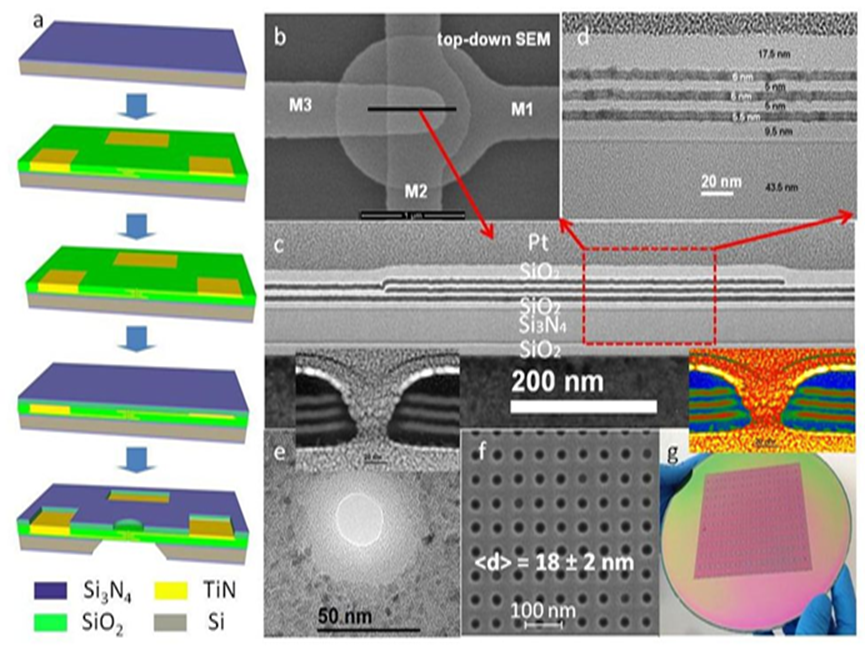


图7-3 规模化制作多层纳米孔整列。（a）工艺流程图；（b）含有三个金属电极的扫描电子显微镜俯视图；透射电子显微镜的截面图（c）和（d），沿着图b中的黑线部分；（e）单个纳米孔的透射电子显微镜俯视图和截面图；（f）纳米孔阵列的扫描电子显微镜照片，纳米孔的直径为18±2纳米；（g）在工艺线上制作的含有纳米孔阵列的8英寸硅片照片。

### 7.3.3面临的挑战和发展的前景

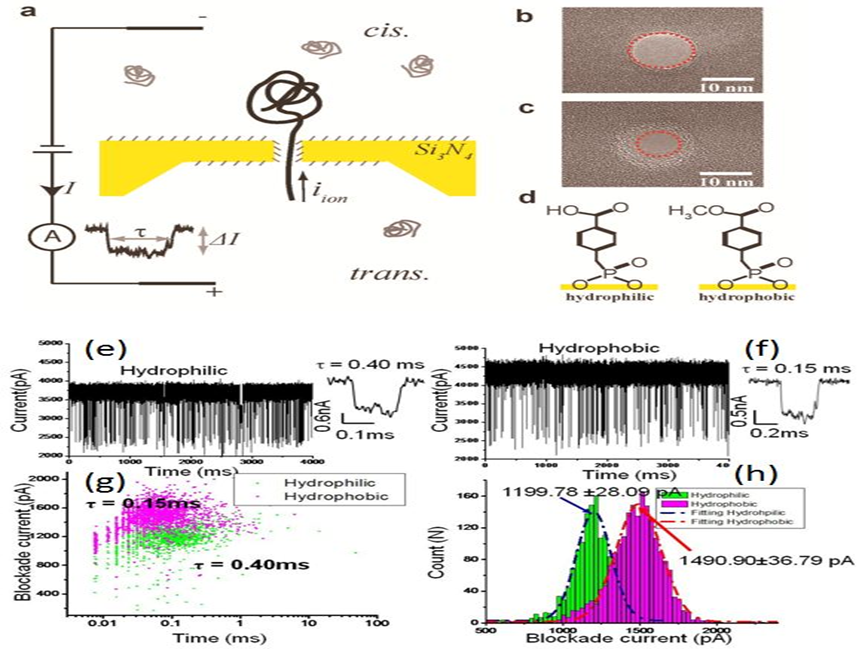
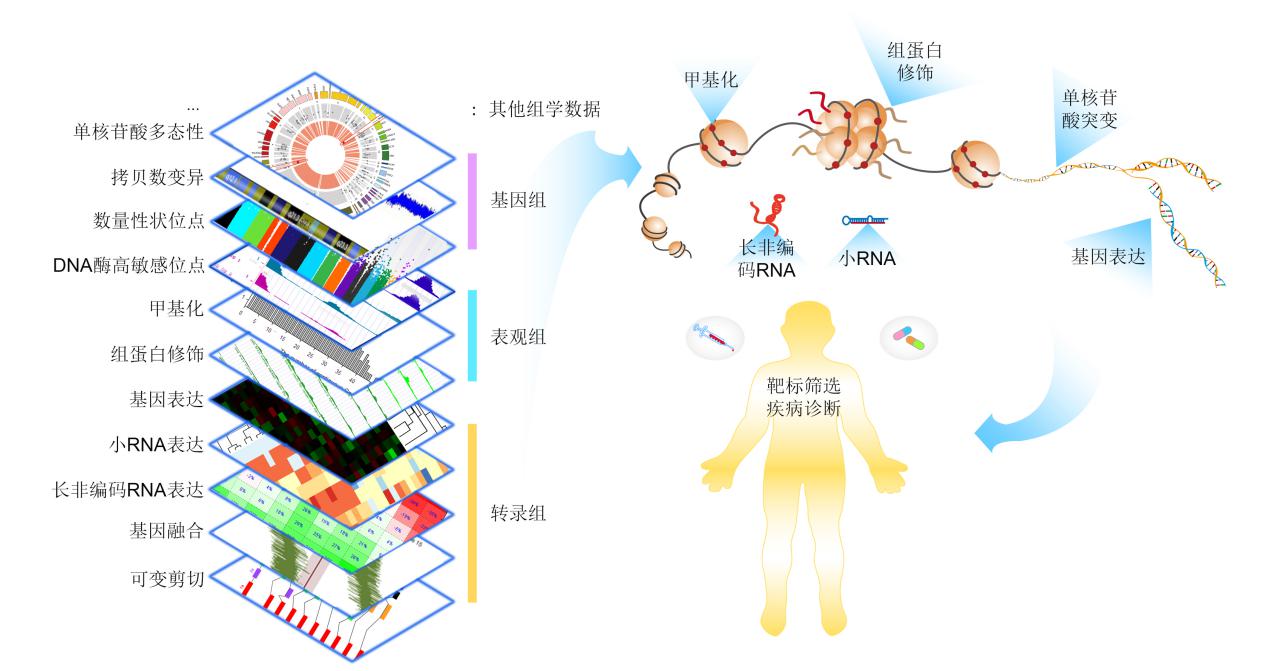
虽然纳米孔测序的优点十分明显, 与前几代技术相比在成本、速度方面有着很大优势, 但是目前还处在起步阶段, 从测序原理到制造工艺都存在有许多问题, 许多技术也都只停留在理论阶段，尚未运用到商业中。要实现固态纳米孔在DNA 测序领域上的应用需要攻克减慢DNA过孔速率、提高纳米孔的空间灵敏度等一系列的技术瓶颈，其最终的目的是实现单碱基识别。研究表明，通过对纳米孔表面作出一定的化学修饰或者改变缓冲溶液的粘稠度可以有效地减缓DNA的过孔速率。例如，将一种特殊设计的化学分子通过其本身的自组装性能修饰在纳米孔的内表面，形成单分子层。通过这个单分子层改变纳米孔内表面的疏水和亲水特性来减慢DNA通过纳米孔的速度。在亲水和疏水的纳米孔内，DNA平均的传输时间分别为 0.40毫秒和 0.15毫秒，得到亲水的状态比疏水的状态减慢DNA通过纳米孔的速度大约3倍（参见图7-4）。此外，在保证溶液盐浓度不变的情况下，通过加入不同比例的甘油来改变溶液的粘度也可以有效地对DNA过孔速度进行调整。对于在同一个纳米孔下，同一种单链DNA（Poly(dA)5000）通过三种不同粘度（0%，20%，50%）的1M KCL溶液,我们获取DNA的传输时间分别为0.24毫秒、1.12毫秒和6.43毫秒。相比于0%的溶液，50%的甘油溶液可以将DNA的平均传输速度从1碱基/48纳秒减慢为1碱基/1290纳秒，减慢大约27倍（参见图2-13）。  


图7-4 自组装分子对DNA传输速度的影响。（a）纳米孔DNA测量系统的装置图; 氮化硅纳米孔内无（b）和有（c）自组装分子的透射电子显微镜照片；（d）给出自组装分子在亲水和疏水两种状态下的结构图；长度为2千碱基对的双链DNA分别通过亲水（e）和疏水（f）两种状态的典型电流信号；获得DNA在两种状态下的时间和振幅的散点图（g）和对应的振幅柱状图（h）。

肿瘤的发生与发展涉及到基因组、转录组、表观组、蛋白组及代谢组等多个不同层次的病理过程，多组学数据的整合分析（图7‑5）是医学数据挖掘的趋势，也是个体化诊疗的必要前提。本章将以转录组为例来介绍组学数据在肿瘤分型、预后、药效预测、治疗靶点等方面的应用。



##### 图7‑5多组学数据分析流程

###### 现有的组学数据主要包括：基因组(数量性状位点、拷贝数变异和单核苷酸多态性)，转录组(基因表达、小RNA表达、长非编码RNA表达、基因融合和可变剪接)和表观组(DNA酶高敏感位点、甲基化和组蛋白修饰)。