

测序技术研发部

提供基于三代纳米孔测序平台的全长转录组测序、微生物宏基因组和全长16S测序等服务

网址: <https://kyzx.gmu.cn/jsyfb/cxjsyfb.htm>

地址: 第五实验楼5楼

负责人: 万绍贵 博士 (18107977707)

临床数据挖掘: 郭有 博士 (19979711221)

纳米孔基因测序平台（ONT）简介



MinION

MinION Mk1C

Portable, biological analysis
USB, or integrated computing

2048 (512* channels)

40G/run



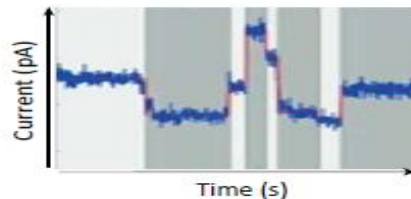
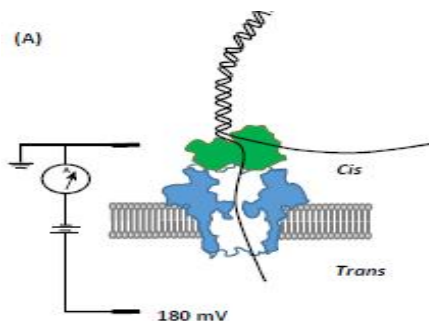
GridION

Five flow cell capacity and integrated GPU computing

5x 2048 (5x 512* channels)

5x40G/run

- ✓ 微生物等小型基因组快速测序
- ✓ 融合基因快速测序
- ✓ 靶向基因突变深度测序
- ✓ 甲基化修饰DNA和RNA直接测序
- ✓ 全长转录组测序



基于PCR技术的cfDNA研究平台和单细胞分析平台

数字PCR仪



Bio-Rad QX200

1. 痕量核酸定量检测
2. 病原体检测
3. 拷贝数变异检测
4. 稀有突变体检测
5. NGS文库质控及验证
6. 无创产前诊断

全自动化毛细管核酸分析仪



Qsep 100 Advance

1. 血浆游离DNA的样本质量检测
2. RNA样本完整度分析
3. 多重PCR产物检测
4. 基因组缺失/扩增分析等
5. NGS文库质控

单细胞文库制备仪



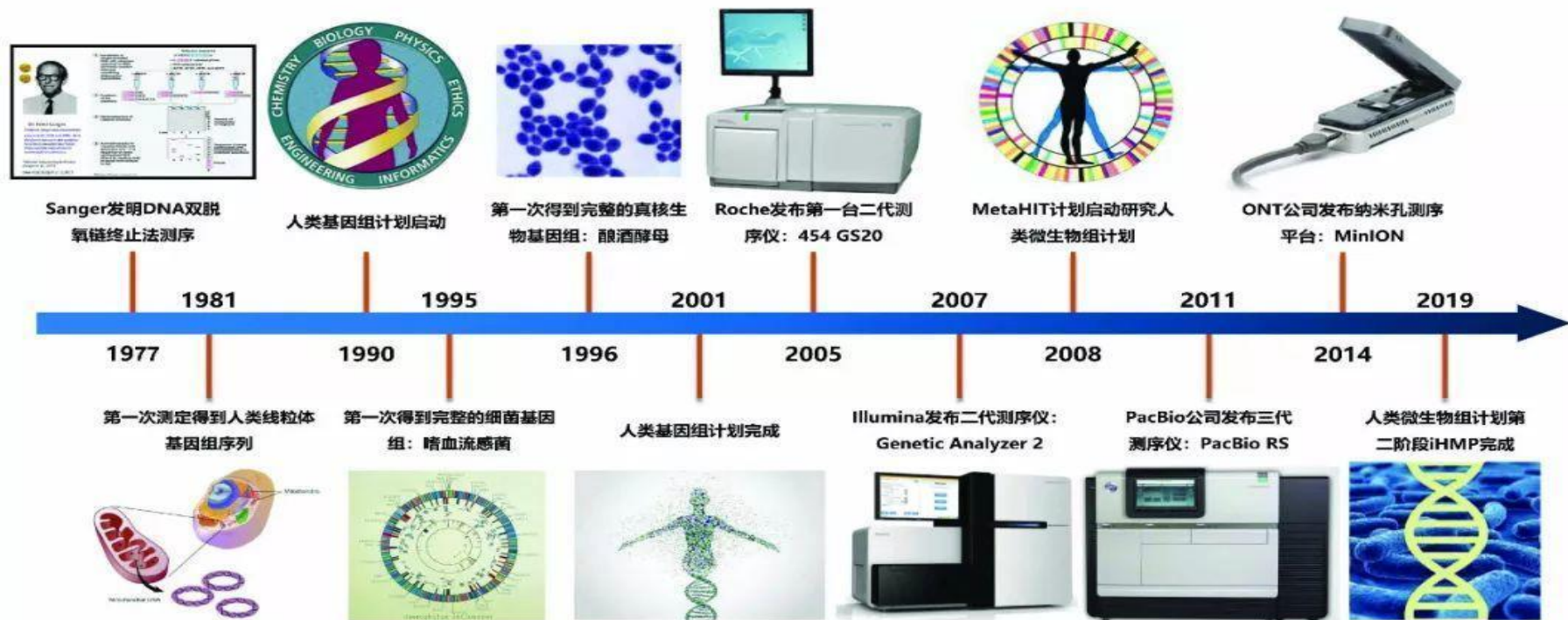
微滴法进行单个细胞标记，用于构建单细胞测序文库

纳米孔第三代高通量基因测序技术原理 及其在生物医药研究中的应用

万绍贵

赣南医学院 科研中心组学技术平台

基因测序技术的发展历史



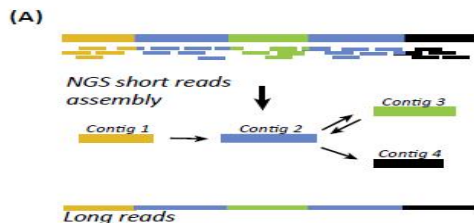
Low throughput → High Throughput

Short read → Long read length

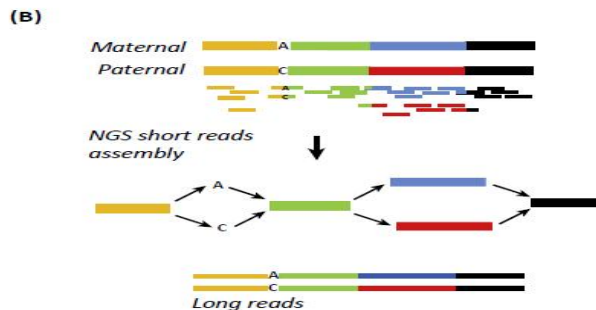
ONT 特性之一：长读长测序

Short-read (二代) vs long-read sequencing(三代)

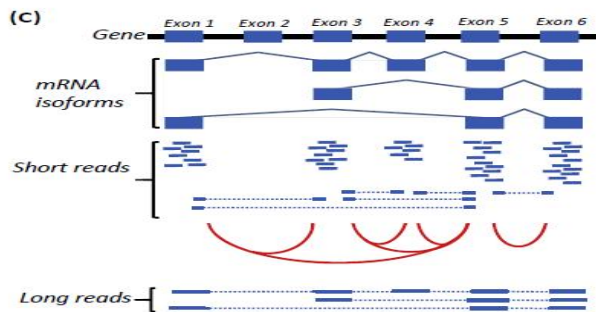
序列组装优势



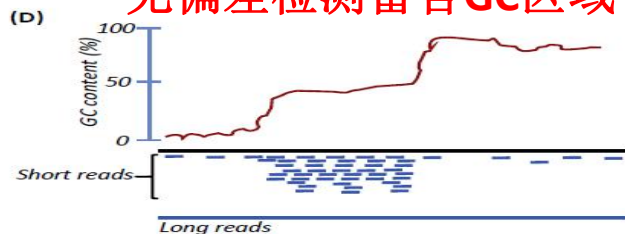
单倍体分型优势



同源异构剪切体拼接优势

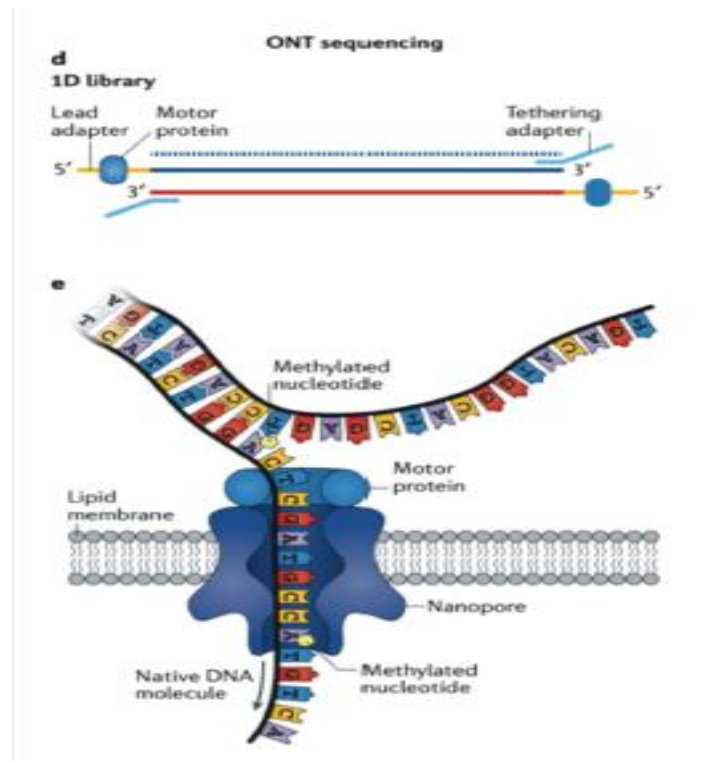
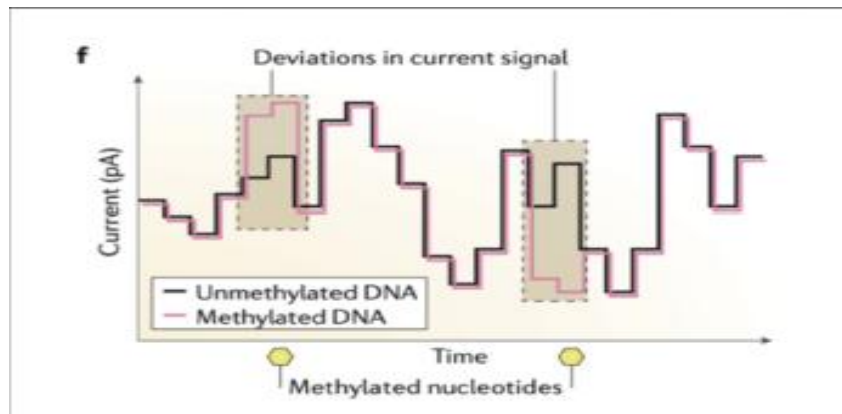
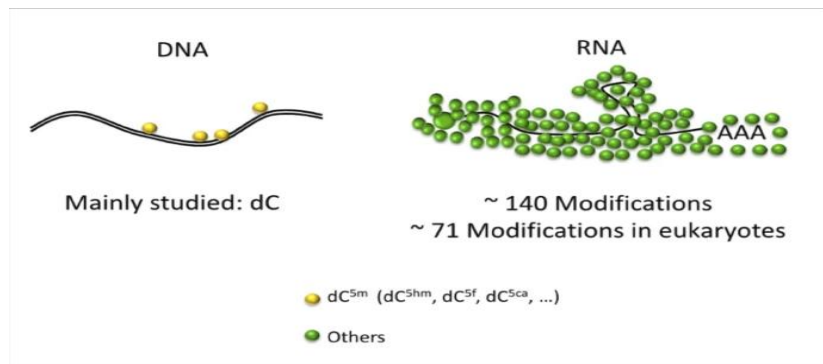


无偏差检测富含GC区域优势



ONT 特性之二：直接测序

DNA/RNA directing sequencing 碱基修饰与序列信息同时检测



ONT 新特性: Adaptive sampling Software-controlled enrichment靶向富集测序

Adaptive sampling

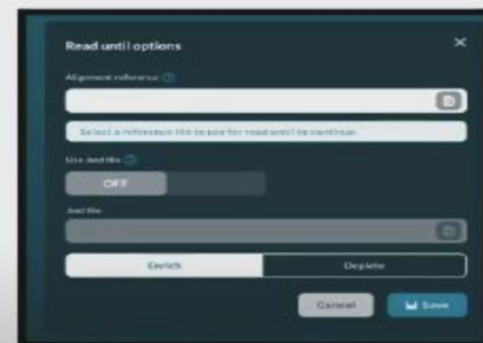
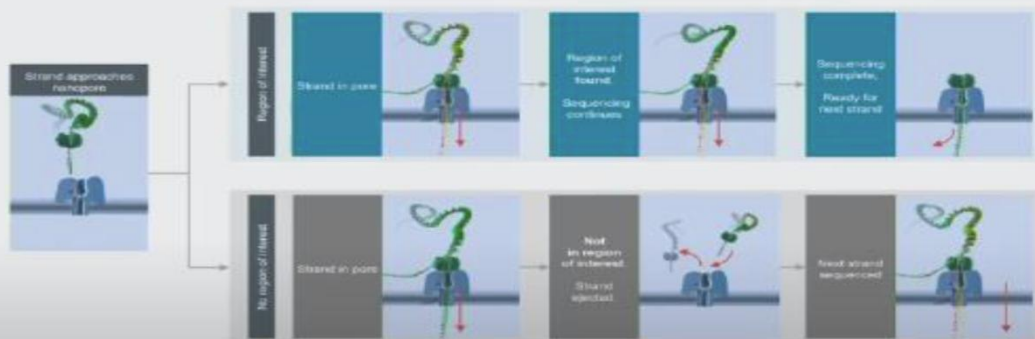
On flow cell targeting

Read-until...

- Accept or reject molecules based on signal or sequence
- Enrich/deplete for specific targets of interest
- Increase on-target data per flow cell, reduce time-to-answer

Integrated solution

- Run via the UI in Q3 software release
- Runs on GridION & Mk1C
- PromethION version in development



文章一：Readfish——利用纳米孔进行Gb级基因组的靶向测序



Readfish enables targeted nanopore sequencing of gigabase-sized genomes

Alexander Payne¹, Nadine Holmes, Thomas Clarke, Rory Munro, Bisrat J. Debebe and Matthew Loose¹

Nanopore sequencers can be used to selectively sequence certain DNA molecules in a pool by reversing the voltage across individual nanopores to reject specific sequences, enabling enrichment and depletion to address biological questions. Previously, we achieved this using dynamic time warping to map the signal to a reference genome, but the method required substantial computational resources and did not scale to gigabase-sized references. Here we overcome this limitation by using graphical processing unit (GPU) base-calling. We show enrichment of specific chromosomes from the human genome and of low-abundance organisms in mixed populations without a priori knowledge of sample composition. Finally, we enrich targeted panels comprising 25,600 exons from 10,000 human genes and 717 genes implicated in cancer, identifying *PML-RARA* fusions in the NB4 cell line in 15k sequencing. These methods can be used to efficiently screen and target panels of genes without enrichment of reads.

文章二：UNCALLED——利用原始电信号实时比对进行纳米孔靶向测序



Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED

Sam Kovaka¹, Yunfan Fan², Bohan Ni¹, Winston Timp² and Michael C. Schatz^{1,3,4}

Conventional targeted sequencing methods eliminate many of the benefits of nanopore sequencing, such as the ability to accurately detect structural variants or epigenetic modifications. The ReadUntil method allows nanopore devices to selectively eject reads from pores in real time, which could enable purely computational targeted sequencing. However, this requires rapid identification of on-target reads while most mapping methods require computationally intensive basecalling. We present UNCALLED (<https://github.com/skovaka/UNCALLED>), an open source mapper that rapidly matches streaming of nanopore current signals to a reference sequence. UNCALLED probabilistically considers k-mers that could be represented by the signal and then prunes the candidates based on the reference encoded within a Ferragina-Manzini index. We used UNCALLED to deplete sequencing of known bacterial genomes within a metagenomics community, enriching the remaining species 4.46-fold. UNCALLED also enriched 148 human genes associated with hereditary cancers to 29.6x coverage using one MinION flowcell, enabling accurate detection of single-nucleotide polymorphisms, insertions and deletions, structural variants and methylation in these genes.

Adaptive sampling options

Alignment reference ?

Select a reference file to use for adaptive sampling to continue.

Start channel ?

End channel ?

—

1

+

—

512

+

Use .bed file ?

OFF

.bed file

Enrich

Deplete

Cancel

Save

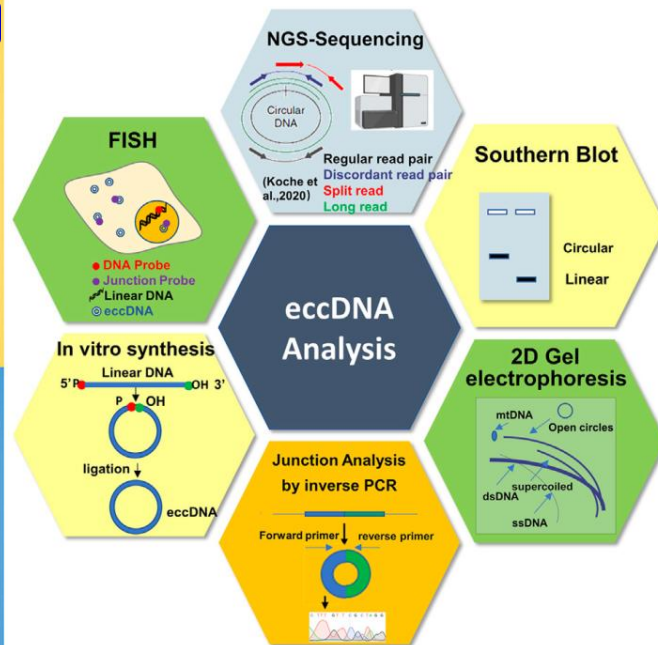
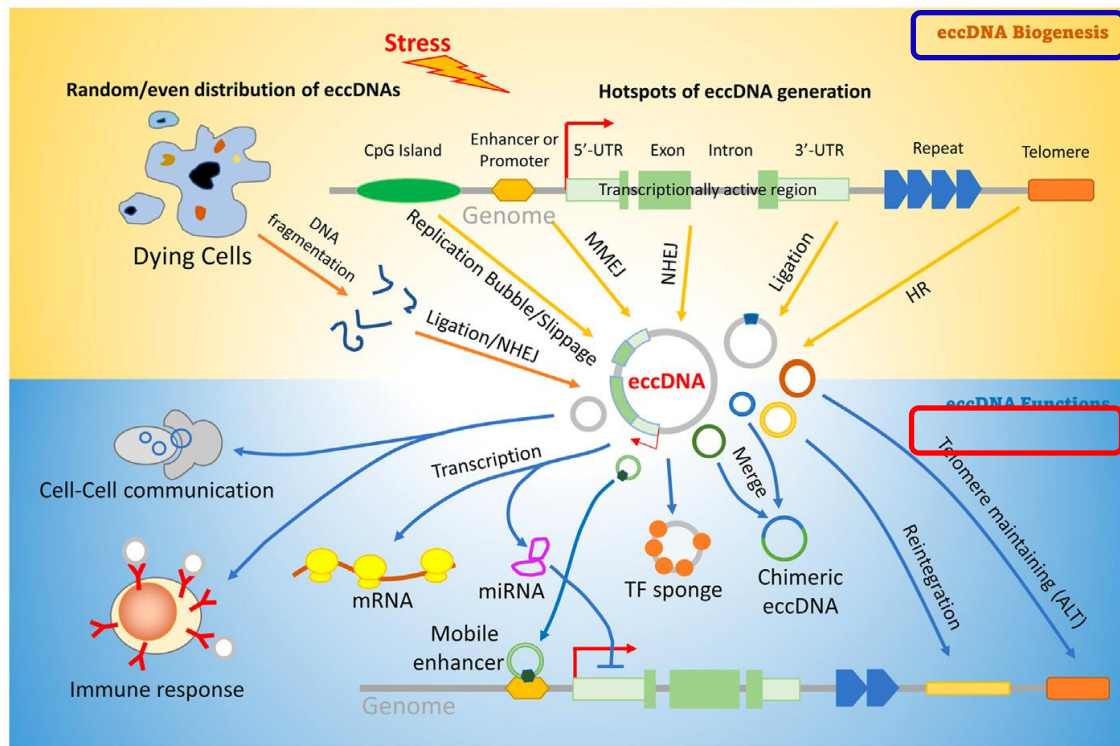
已开展的基于纳米孔测序技术的主要应用

- ◆ 染色体外环状DNA (eccDNA) 分子测序 **已测临床样本数量：36**
- ◆ Cas9靶向富集的融合基因与甲基化测序 **已建立检测与分析流程**
- ◆ 全长转录组测序 **已测细胞系样本数量：12**
- ◆ 微生物组测序 (16S和宏基因组学) **已测临床样本数量：60+**

基于滚环扩增（RCA）的纳米孔测序的应用

- ▣天然存在的环状分子：eccDNA，mtDNA，HBV DNA，circRNA
- ▣人工制备的环状分子：cfDNA，FFPE样本DNA，PCR 扩增子.....

Extrachromosomal circular DNA (eccDNA)



Schematic representation of the different approaches and tools to study circular DNA.

Pathways involved in eccDNA biogenesis and function

HR: homologous recombination; NHEJ: Non-homologous end joining; MEJ: microhomology mediated end-joining

Nathanson DA, Gini B, et al. Targeted therapy resistance mediated by dynamic regulation of **extrachromosomal** mutant EGFR DNA. **Science**. 2014;343:72–6

Turner KM, et al. **Extrachromosomal oncogene** amplification drives tumour evolution and genetic heterogeneity. **Nature**. 2017;543:122–5

Morton AR, Dogan-Artun N, Faber ZJ, et al. Functional enhancers shape **extrachromosomal oncogene** amplifications. **Cell**. 2019;179:1330–41

Wu S, Turner KM, et al. **Circular ecDNA** promotes accessible chromatin and high oncogene expression. **Nature**. 2019;575:699–703

Verhaak RGW, Bafna V, Mischel PS. **Extrachromosomal oncogene** amplification in tumour pathogenesis and evolution. **Nat Rev Cancer**. 2019;19:283–8

Kim H, Nguyen NP, Turner KM, Wu S, et al. **Extrachromosomal DNA** is associated with oncogene amplification and poor outcome across multiple cancers. **Nat Genet**. 2020;52:891–7

Koche RP, Rodriguez-Fos E, et al. **Extrachromosomal circular DNA** drives oncogenic genome remodeling in neuroblastoma. **Nat Genet**. 2020;52:29–34

Mehta D, et al. Full-length sequencing of circular DNA viruses and **extrachromosomal circular DNA** using CIDER-Seq. **Nat Protoc**. 2020;15(5):1673-1689

Zhu Y, Gujar AD, et al. **Oncogenic extrachromosomal DNA** functions as mobile enhancers to globally amplify chromosomal transcription. **Cancer Cell**. 2021;39:694-707

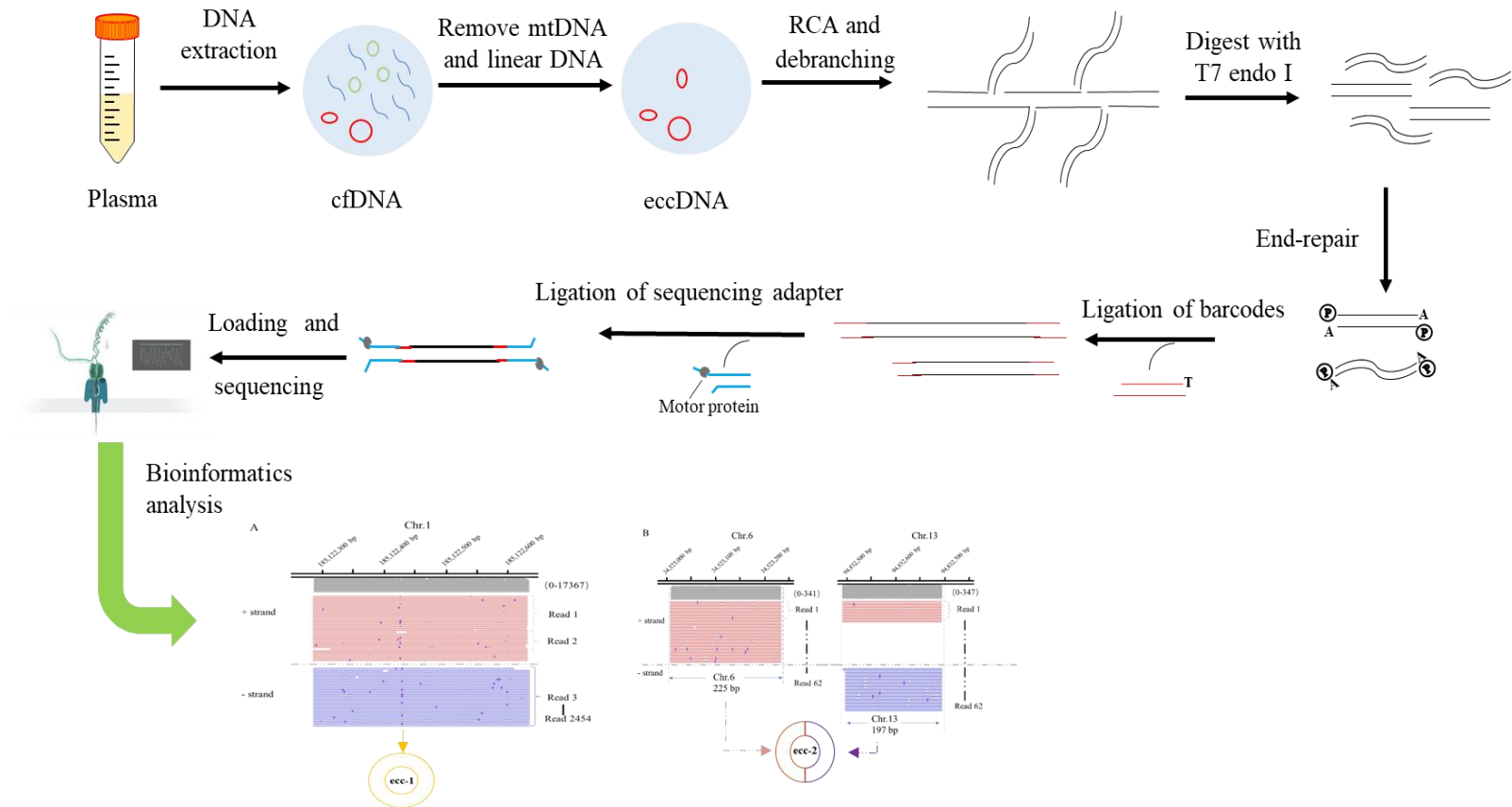
Wang Y, Wang M, et al. **eccDNAs** are apoptotic products with high innate immunostimulatory activity. **Nature**. 2021;599:308-314

Shoshani O, Brunner SF, et al. **Chromothripsis** drives the evolution of gene amplification in cancer. **Nature**. 2021;591:137-141

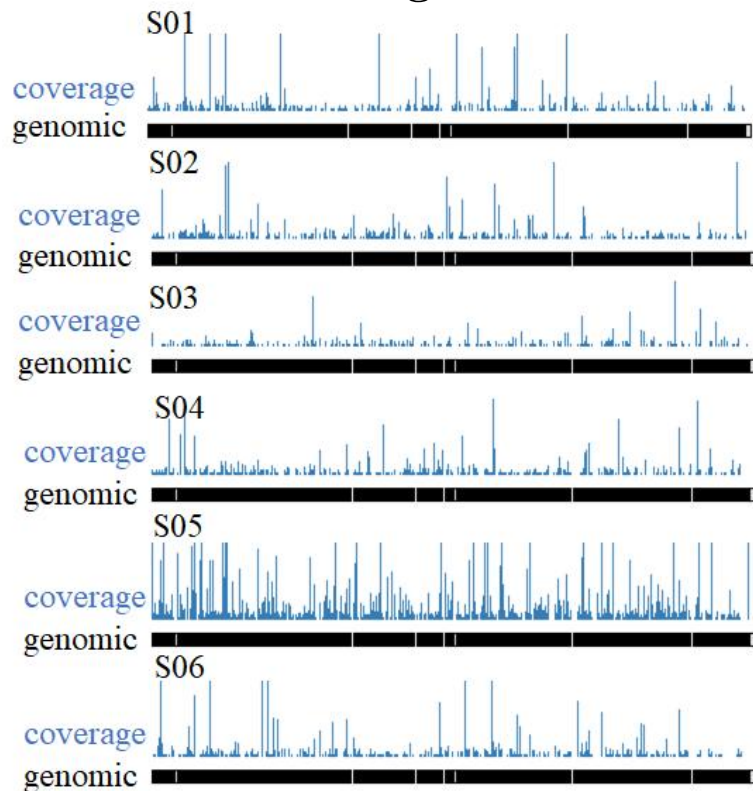
Rosswog C, et al. **Chromothripsis** followed by circular recombination drives oncogene amplification in human cancer. **Nat Genet**. 2021;53:1673-1685

van Leen E, Brückner L, Henssen AG. The genomic and spatial mobility of **extrachromosomal DNA** and its implications for cancer therapy. **Nat Genet**. 2022;54:107-114

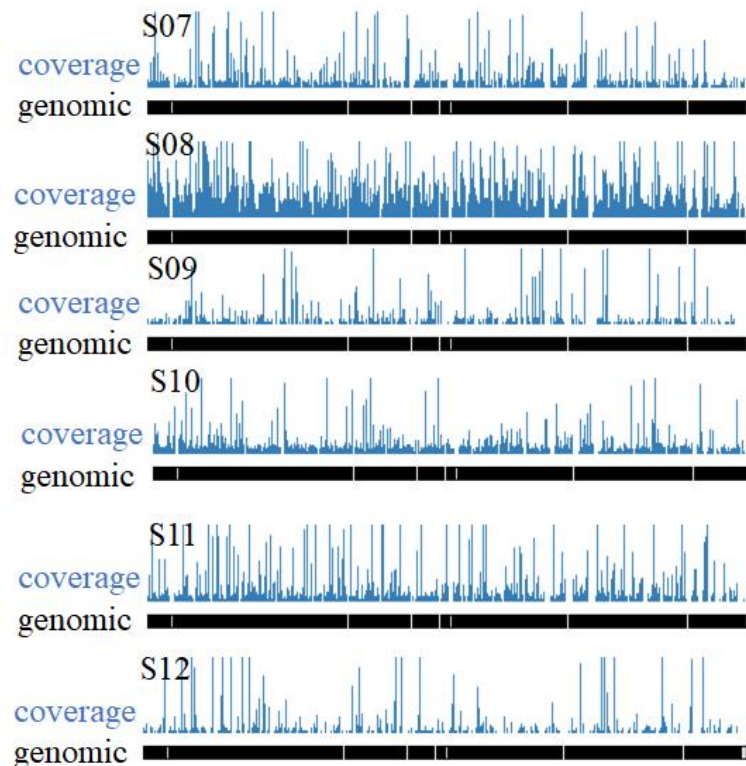
血浆样本中eccDNA分子测序鉴定流程



Coverage of eccDNA in genomic region



HBV samples

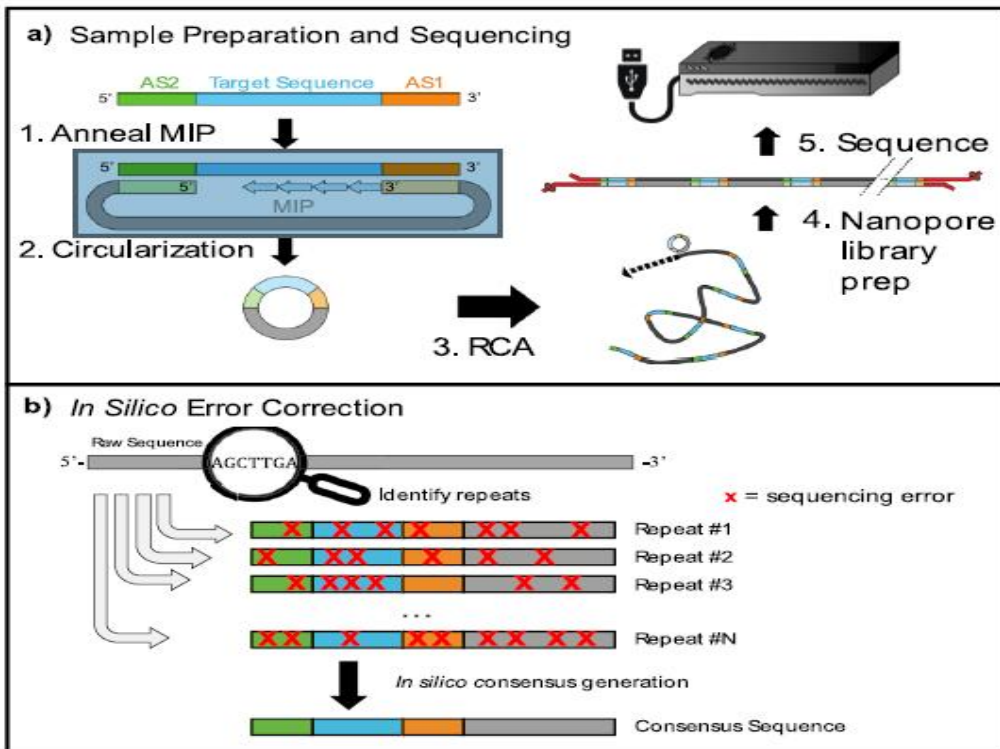


HBV-HCC samples

High-Fidelity Nanopore Sequencing of Ultra-Short DNA Targets

Brandon D. Wilson,[†] Michael Eisenstein,^{‡,§} and H. Tom Soh^{*,‡,§,||}

[†]Department of Chemical Engineering, Stanford University, Stanford, California 94305, United States

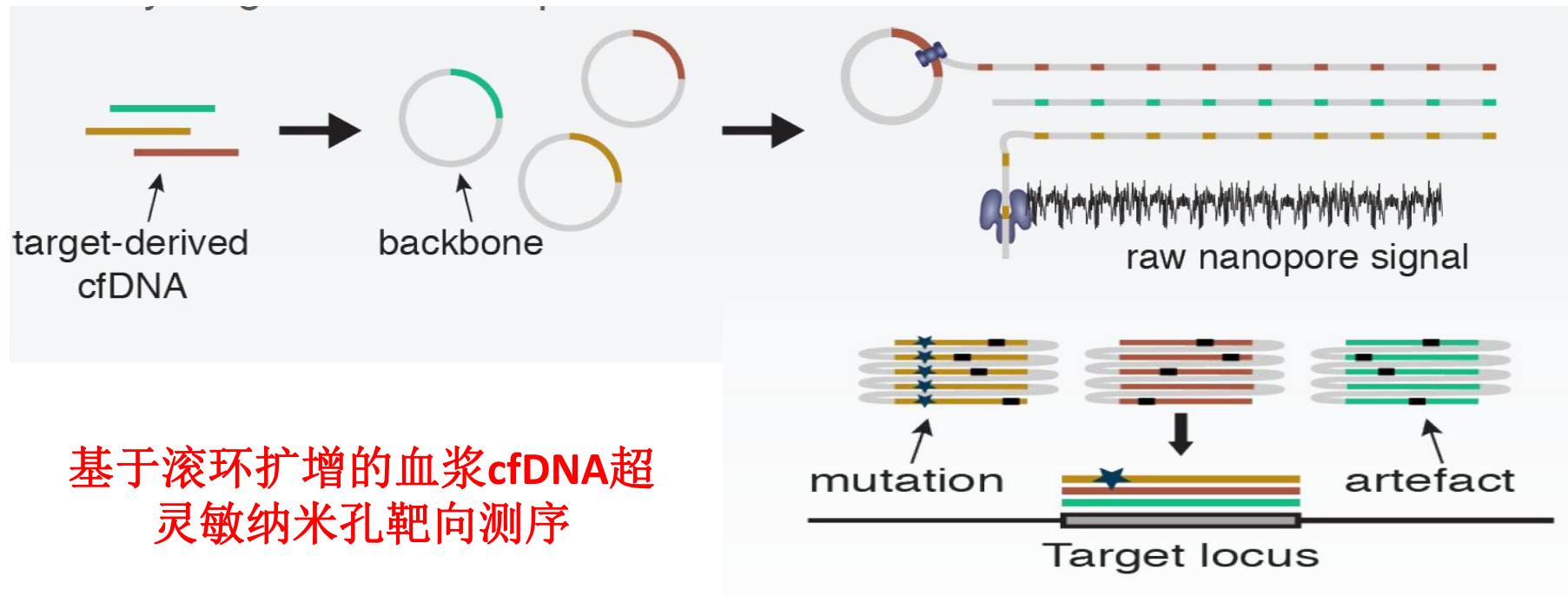


基于滚环扩增富集
超短序列进行准确
测序



CyclomicsSeq a sensitive liquid biopsy genetic test
Real-time and cost-efficient cancer monitoring in blood

荷兰乌特勒支大学医学中心分子医学中心



基于滚环扩增的血浆cfDNA超
灵敏纳米孔靶向测序

基于Cas9靶向切割富集的纳米孔测序（nCATS）的应用

- 融合基因检测
- 热点突变片段深度测序：SNV+SV
- 靶向富集的CpG岛甲基化修饰检测

肿瘤融合基因靶向测序

**慢性淋巴细胞白血病(CLL) 相关的5个高频突变基因靶向测序:
TP53, NOTCH1, BIRC3, SF3B1 and MYD88.**

Sci Rep. 2018 Aug 7;8(1):11798. doi: 10.1038/s41598-018-30330-y.

Design and MinION testing of a nanopore targeted gene sequencing panel for chronic lymphocytic leukemia.

Orsini P¹, Minervini CF¹, Cumbo C¹, Anelli L¹, Zagaria A¹, Minervini A¹, Coccaro N¹, Tota G¹, Casieri P¹, Impera L¹, Parciante E¹, Brunetti C¹, Giordano A¹, Specchia G¹, Albano F².

慢性髓性白血病(CML) 费城染色体阳性融合基因: BCR-ABL1

Exp Mol Pathol. 2017 Aug;103(1):33-37. doi: 10.1016/j.yexmp.2017.06.007. Epub 2017 Jun 27.

Mutational analysis in BCR-ABL1 positive leukemia by deep sequencing based on nanopore MinION technology.

Minervini CF¹, Cumbo C¹, Orsini P¹, Anelli L¹, Zagaria A¹, Impera L¹, Coccaro N¹, Brunetti C¹, Minervini A¹, Casieri P¹, Tota G¹, Russo Rossi A¹, Specchia G¹, Albano F².

肺癌靶向基因测序: EGFR, KRAS, NRAS and NF1, EML4-ALK等

DNA Res. 2017 Dec 1;24(6):585-596. doi: 10.1093/dnares/dsx027.

Sequencing and phasing cancer mutations in lung cancers using a long-read portable sequencer.

Suzuki A¹, Suzuki M², Mizushima-Sugano J^{2,3}, Frith MC^{2,4}, Makalowski W⁵, Kohno T⁶, Sugano S², Tsuchihara K¹, Suzuki Y².

nCATS实验与分析流程

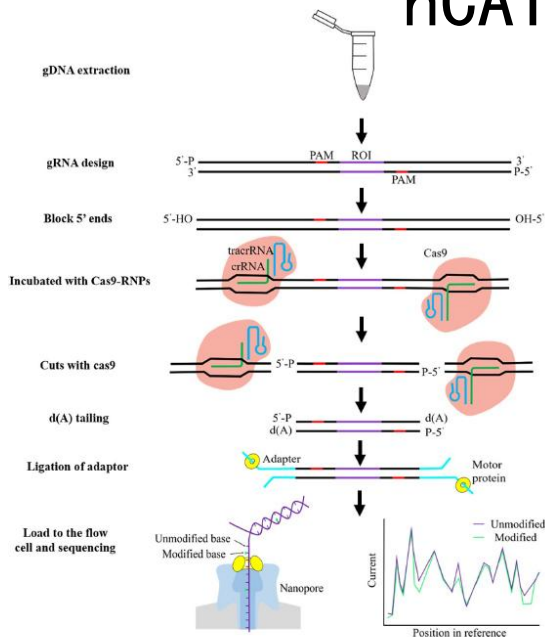
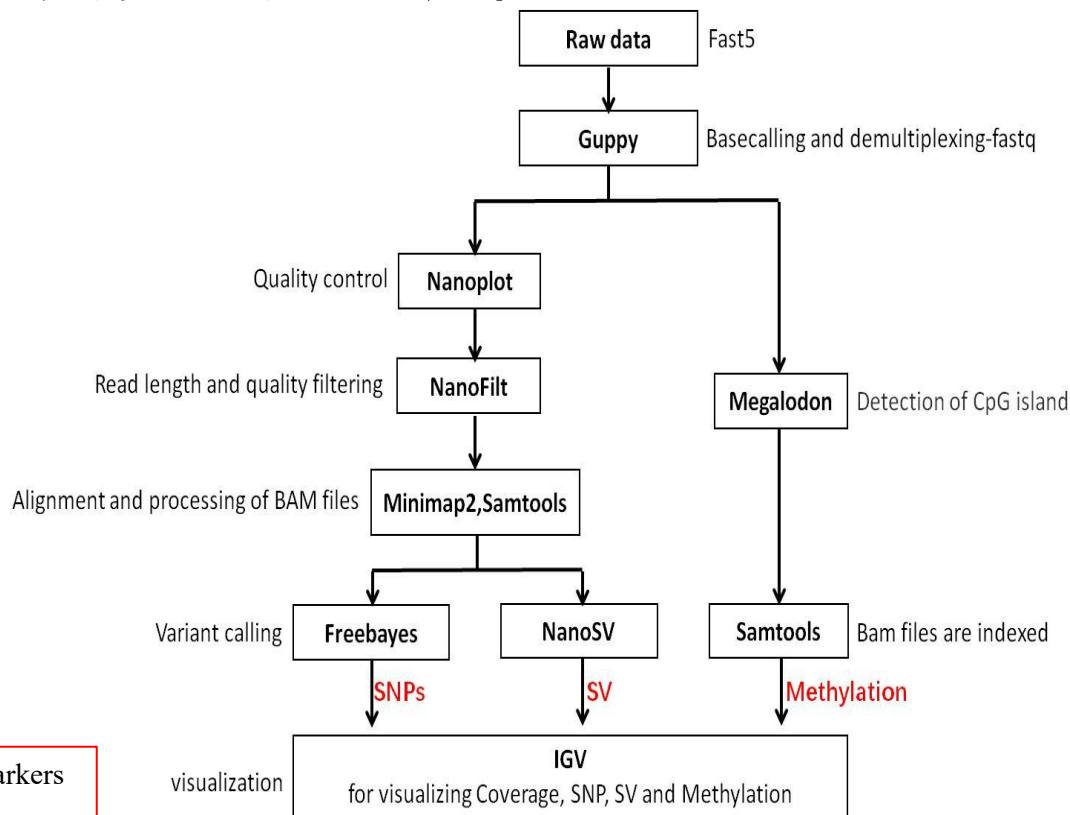
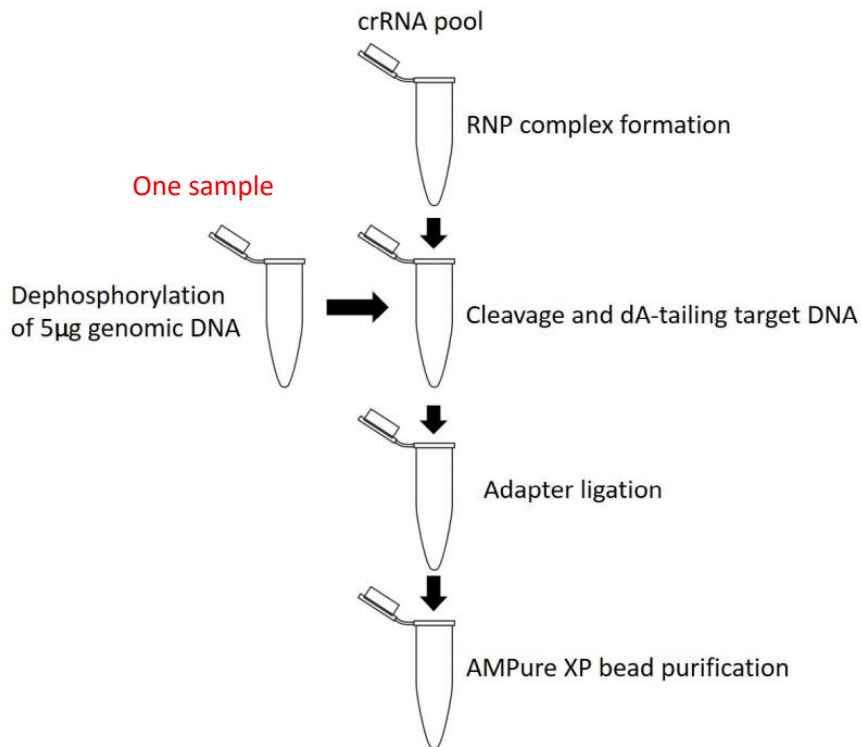


FIGURE 1 | Schematic overview of Cas9 enrichment workflow. Genomic DNA (gDNA) is first extracted from cells or tissue, guide RNA ends are dephosphorylated, and then Cas9 ribonucleoprotein complexes (Cas9 RNPs) were used for cleaving the region of interest (ROI) rapidly added to the 3' ends of cut DNA fragments. ONT adapters are ligated to cut end around the ROI. The sequencing library is can excess adapters using AMPure XP beads. The prepared library is loading onto the flow cell for sequencing. DNA modifications can be sequencing. As the single-stranded DNA moves through the pore, it causes disruptions to the ionic current in a sequence-dependent manner known as a "squiggle." The current signal highlighted with green lines indicates small changes due to the methylated cytosine.

Zhang J, Xie S, Xu J, Liu H and Wan S* (2021) Cancer Biomarkers Discovery of Methylation Modification With Direct High-Throughput Nanopore Sequencing. Front. Genet. 12:672804



BCR_23177704 + BCR_23285486 + BCR_23286143 + ABL1_130888788

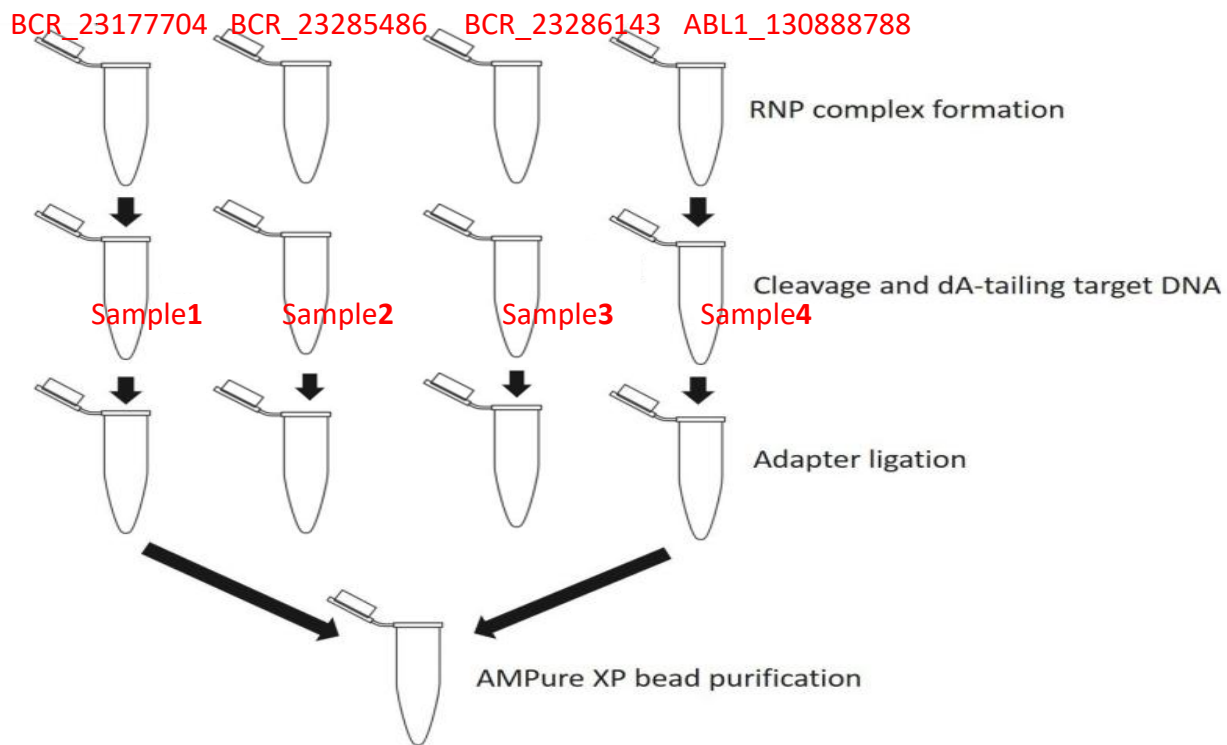


Cas9 targeted sequencing in 1 sample with 4 crRNAs

Detected target reads:

crRNA	BCR_23177704	BCR_23285486	BCR_23286143	ABL1_130888788	Total
Targeted reads	507	496	381	468	1852
% target	0.48%	0.47%	0.36%	0.44%	1.75%

Cas9 targeted sequencing in 4 identical sample with 4 crRNAs respectively



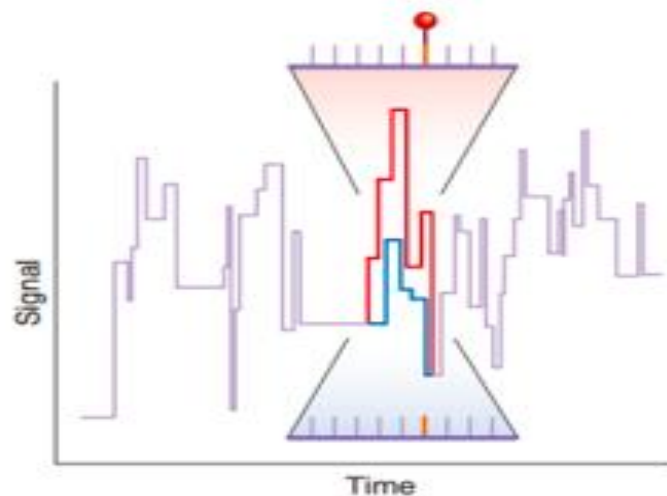
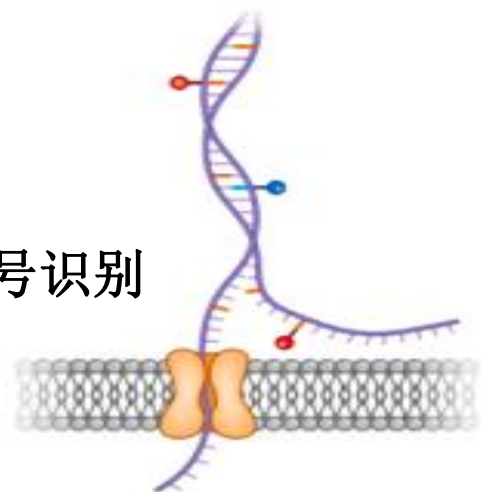
Cas9 targeted sequencing in 4 identical samples with 4 crRNAs respectively

Detected target reads:

Sample + crRNA	Target reads	% target
S1+BCR_23177704	124	0.16%
S2+BCR_23286143	94	0.12%
S3+BCR_23286143	61	0.08%
S4+ABL1_130888788	98	0.12%
Total	377	0.48%

纳米孔基因测序检测碱基修饰原理

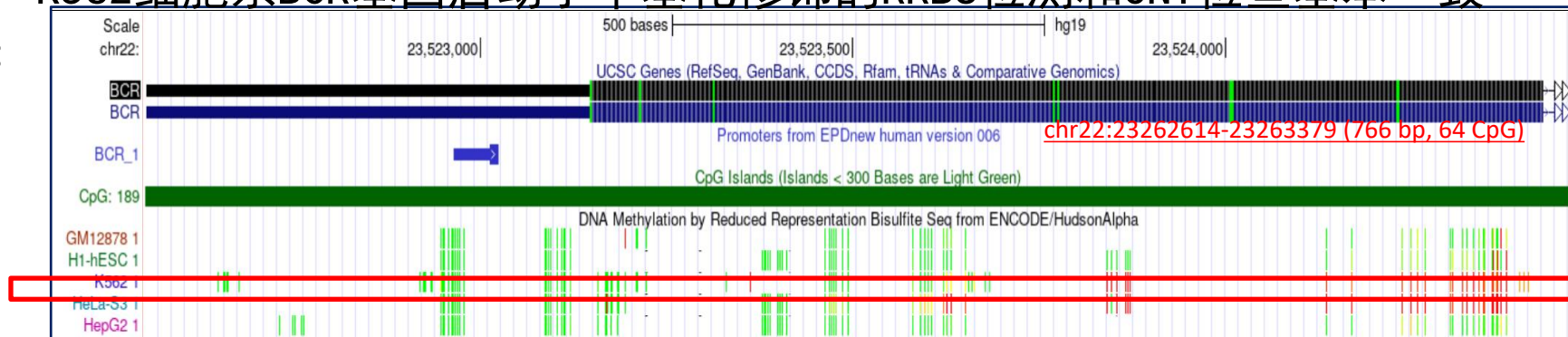
ONT电信号识别



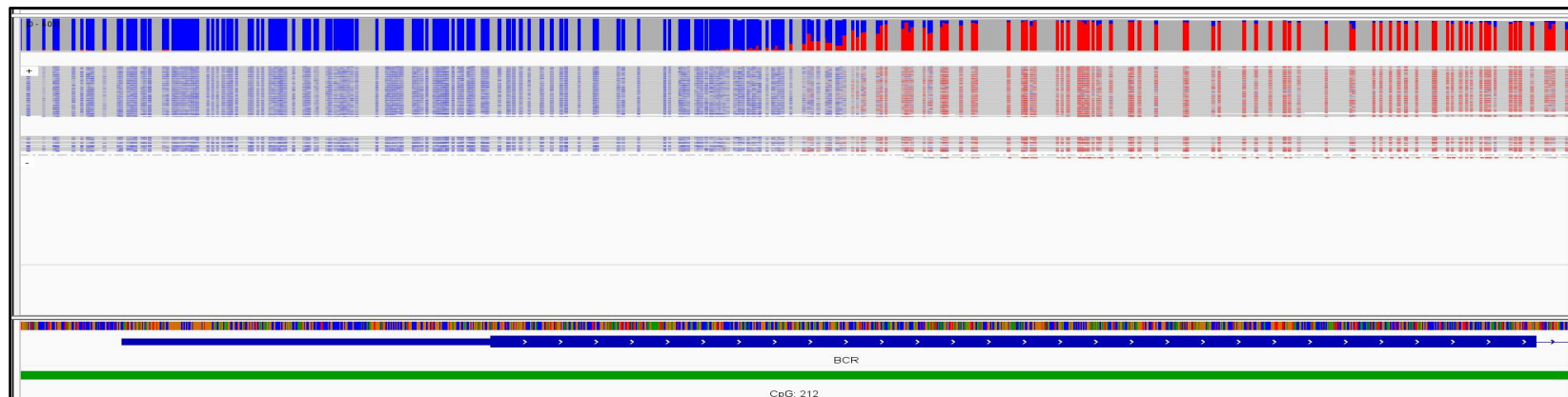
- ✓ 单碱基水平检测，无需任何处理，真实还原碱基修饰信息
- ✓ 测序读长长，避免短读长带来的拼接错误
- ✓ 鉴定种类丰富，比如5mC，5hmC，6mA等多种碱基修饰
- ✓ 可进行等位基因特异性碱基修饰检测

K562细胞系BCR基因启动子甲基化修饰的RRBS检测和ONT检查基本一致

UCSC:



IGV_
Megalon:



DNA Methylation by reduced Representation Bisulfite Seq from ENCODE/HudsonAlpha:

- red = 100% of molecules sequenced are methylated
- yellow = 50% of molecules sequenced are methylated
- green = 0% of molecules sequenced are methylated

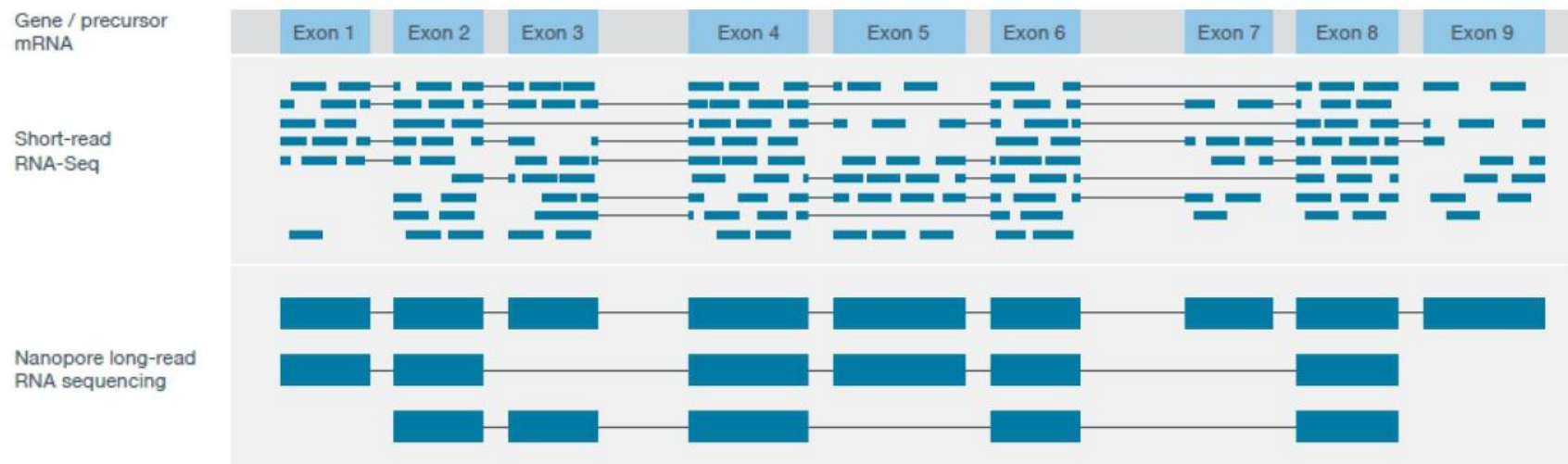
- blue unmethylated
- red methylated

全长转录组测序的应用

- mRNA的定量检测：转录本、基因的表达差异
- mRNA的结构检测：可变剪切、融合基因
- mRNA的修饰检测：6mA、APA、PolyA长度

全长转录组概念

全长转录组测序 (full-length transcriptome sequencing), 又称 Isoform Sequencing (Iso-Seq), 全长转录组是基于第三代测序平台, 无需打断拼接, 直接获得包含5', 3' UTR, polyA 尾的完整转录本, 从而准确分析有参考基因组物种可变剪接及融合基因等结构信息, 克服无参考基因组物种转录本拼接较短、信息不完整的难题。



为什么要选择全长转录组测序

◆ 二代 vs 三代

检测内容	二代转录组	ONT 全长转录组	ONT Direct RNA
定量	+++	+++	+++
差异分析	+++	+++	+++
可变剪切	+	+++	+++
融合基因	+	+++	+++
新转录本预测	+	+++	+++
等位基因分析		+++	+++
APA		+++	+++
PolyA 长度			+++
修饰分析			+++
正负链分析			+++

Isoform定量全长转录组可实现转录本定量，对基因的每一个Isoform进行功能注释。

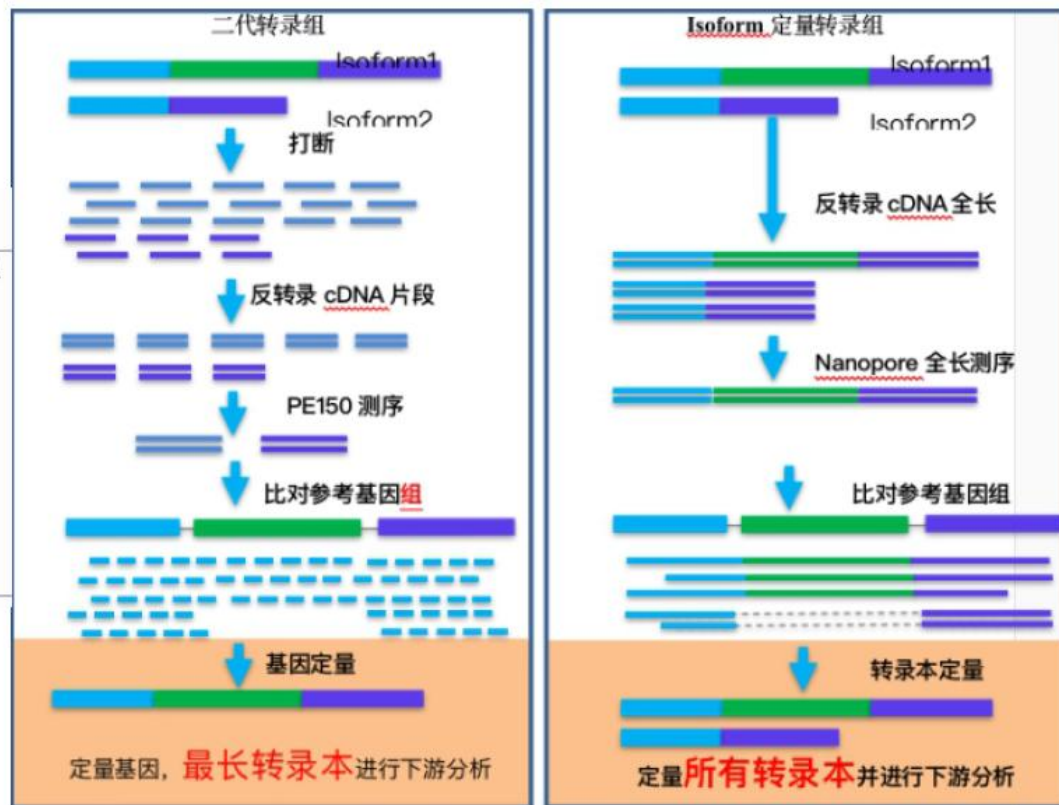
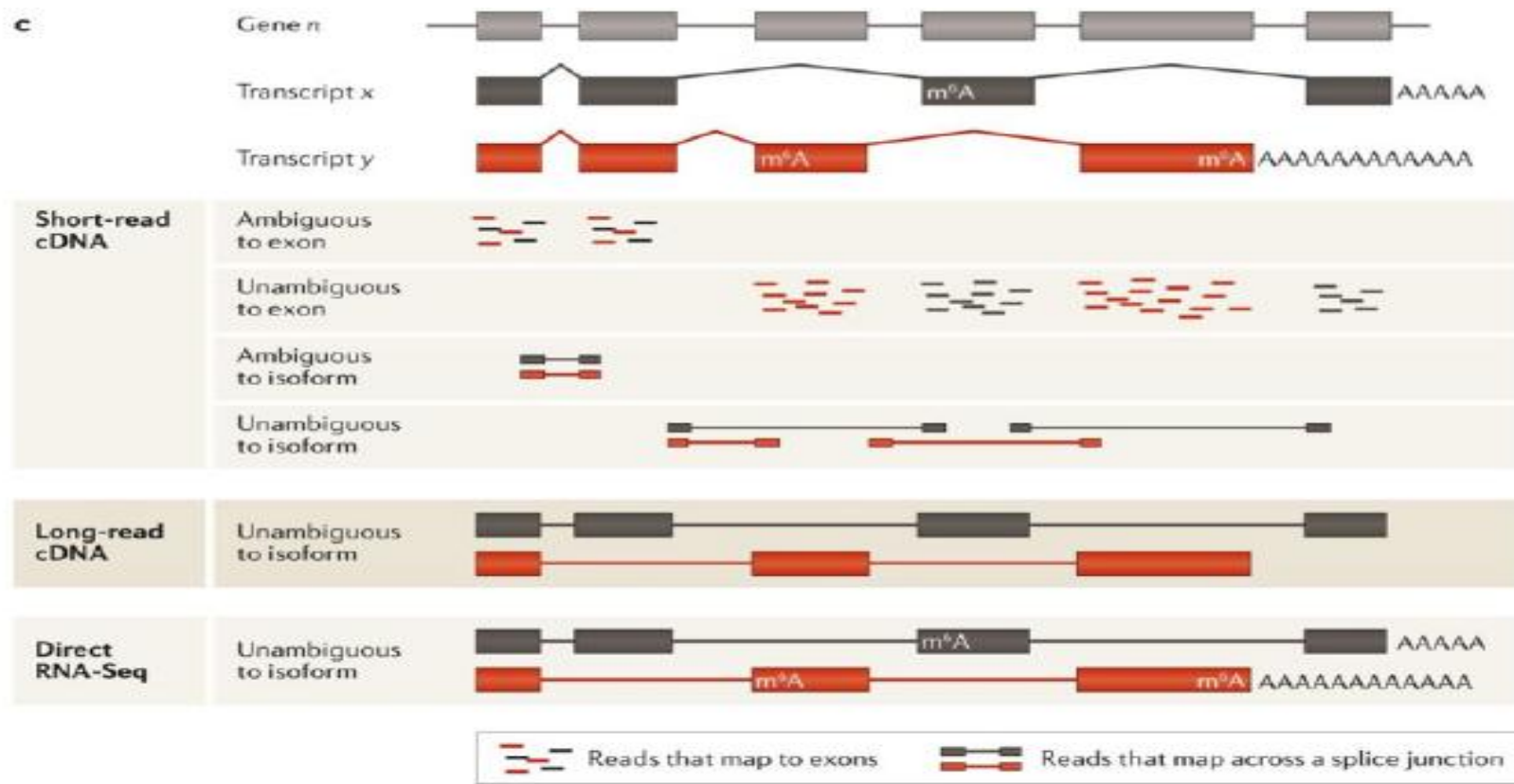


图1 Isoform定量全长转录组和二代转录组建库和比对分析和定量示意图

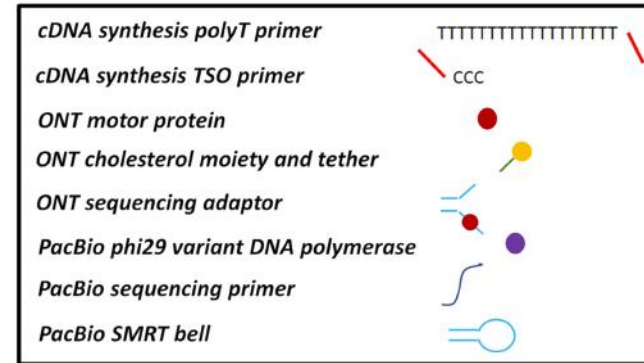
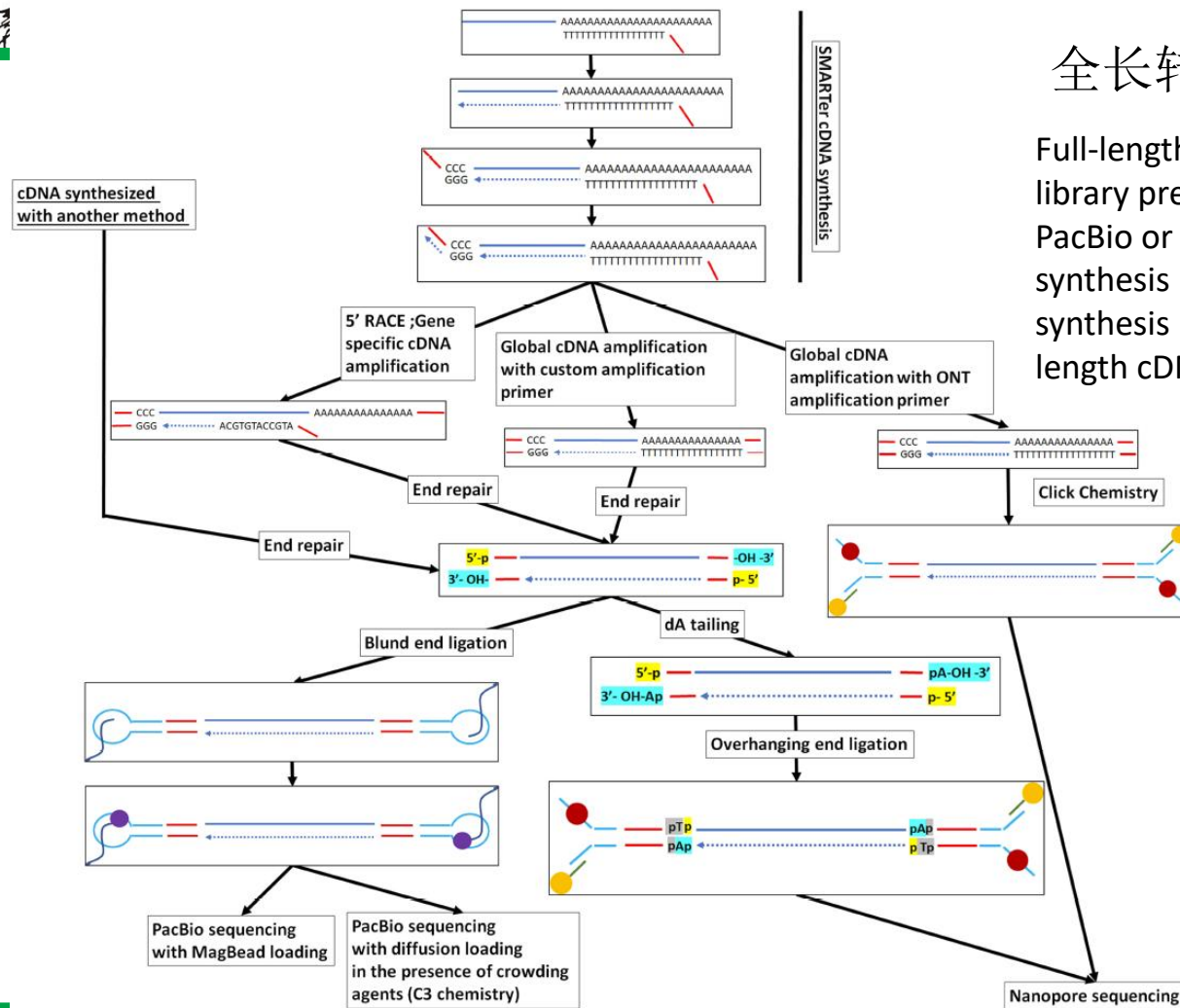
纳米孔全长转录组测序进行结构变异的检测



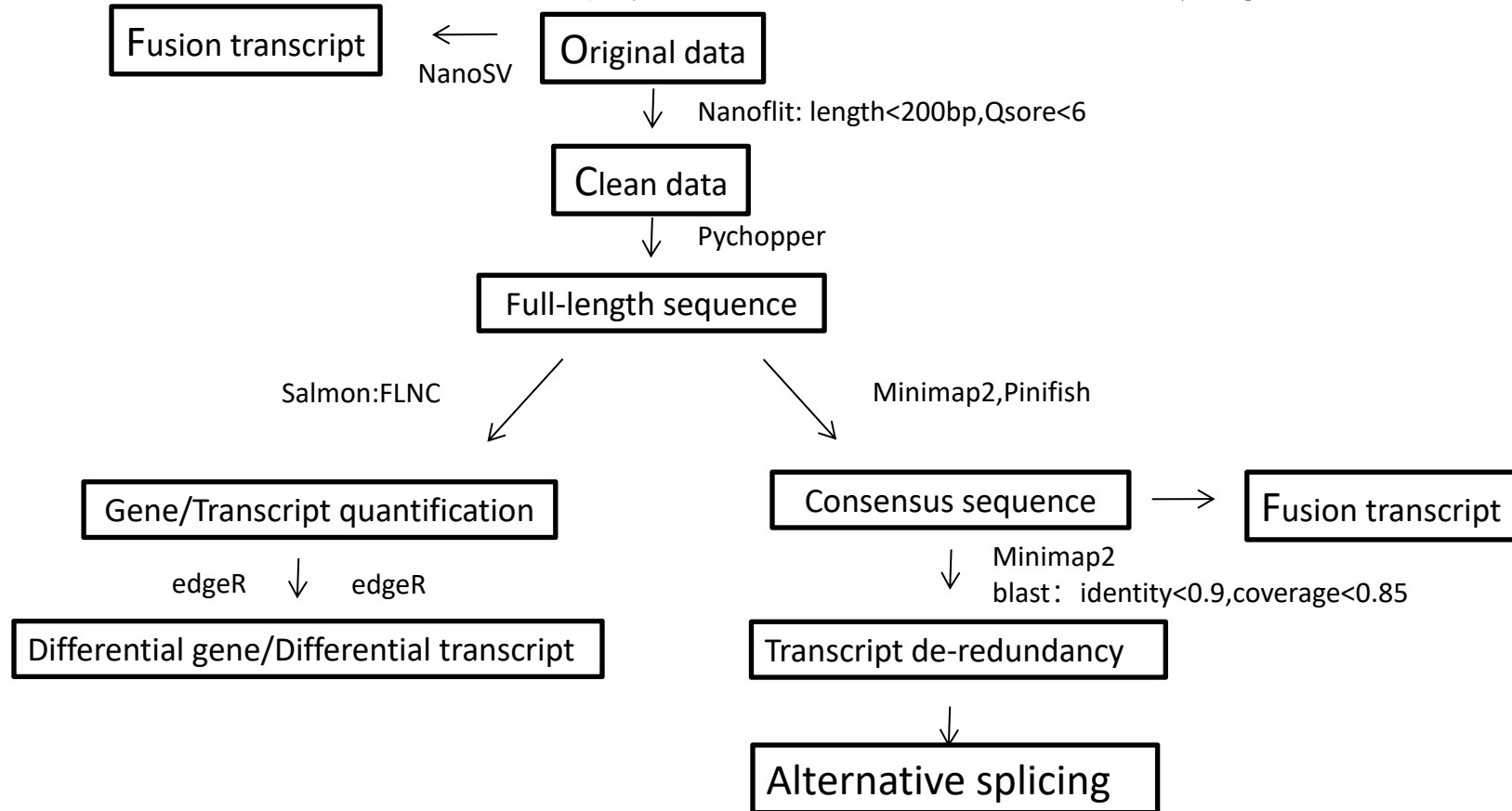
Stark, R., et al. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics*.

全长转录组测序基本实验流程

Full-length cDNA synthesis and long-read cDNA library preparation and sequencing on either the PacBio or the ONT platforms. The SMARTer cDNA synthesis method is presented here but other cDNA synthesis methods can be also used to create full-length cDNA from RNA molecules or RNA fragments.



全长转录组测序基本分析流程



全长转录组测序技术解析不同转移性能 肝癌细胞系的转录本表达谱与结构变异

杨淑欣, 彭民武, 罗娟娟, 徐景祥, 万绍贵
(赣南医学院基础医学院分子病理中心, 江西 赣州 341000)

摘要:为了更好地研究肝细胞癌的转移机制,并筛选出与肝癌预后密切相关的潜在靶点,本研究利用纳米孔长读长测序技术对两株不同转移潜能的肝癌细胞(MHCC97H与MHCC97L)进行全长转录组比较分析。通过生物信息学分析发现643个新基因(Novel gene)和2 471个新转录本(New transcripts)。在这两株细胞中有293个差异表达基因及74个差异表达转录本。在结构变异分析中鉴定出1 008个可变剪切事件,其中外显子跳跃占比最多。本研究通过全长转录组测序进一步完善了肝癌细胞的基因注释及基因表达分析,为揭示肝细胞癌的转移分子机制提供了新的研究思路与线索。

关键词:肝细胞癌;转录本表达谱;mRNA可变剪切;全长转录组测序;纳米孔测序

中图分类号:R735.7 文献标志码:A 文章编号:1001-5779(2021)08-0001-09

DOI:10.3969/j.issn.1001-5779.2021.08.001

开放科学(资源服务)标识码(OSID):



微生物等小型基因组快速测序的应用

- 宏基因组快速检测
- 细菌16S和真菌18S的全长扩增子测序进行菌群分析
- 其他PCR扩增的深度测序

SARS-CoV-2测序鉴定

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES



BMJ Yale

HOME | AE

武汉大学 刘天罡教授团队

[Comment on this paper](#)

Nanopore target sequencing for accurate and comprehensive detection of SARS-CoV-2 and other respiratory viruses

Ming Wang, Aisi Fu, Ben Hu, Yongqing Tong, Ran Liu, Jiashuang Gu, Jianghao Liu, Wen Jiang, Gaigai Shen, Wanxu Zhao, Dong Men, Lilei Yu, Zixin Deng, Yan Li, Tiangang Liu

doi: <https://doi.org/10.1101/2020.03.04.20029538>



 **protocols.io**

Josh Quick / Publications / nCoV-2019 sequencing protocol



PCR靶向富集 病毒基因组



Jan 22, 2020

 Run



nCoV-2019 sequencing protocol ▼

Forked from Ebola virus sequencing protocol

Josh Quick¹

¹University of Birmingham

7

Works for me

[dx.doi.org/10.17504/protocols.io.bbmuik6w](https://doi.org/10.17504/protocols.io.bbmuik6w)

ARTIC

Coronavirus Method Development Community

[view 1 r](#)

快速宏基因组学测序

ICU应用示例：心内膜炎感染病人快速宏基因组检测

Cheng et al. *Ann Clin Microbiol Antimicrob* (2018) 17:43
<https://doi.org/10.1186/s12941-018-0294-5>

Annals of Clinical Microbiology
 and Antimicrobials

RESEARCH

Open Access

Identification of pathogens in culture-negative infective endocarditis cases by metagenomic analysis



Jun Cheng^{1†}, Huan Hu^{1,2†}, Yue Kang^{2†}, Weizhi Chen², Wei Fang², Kaijuan Wang¹, Qian Zhang³, Aisi Fu⁴,
 Shuilian Zhou², Chen Cheng², Qingqing Cao², Feiyan Wang^{1*}, Shela Lee^{2*} and Zhou Zhou^{1*}

感染检验科应用示例：下呼吸道感染病人快速宏基因组检测

nature
 biotechnology

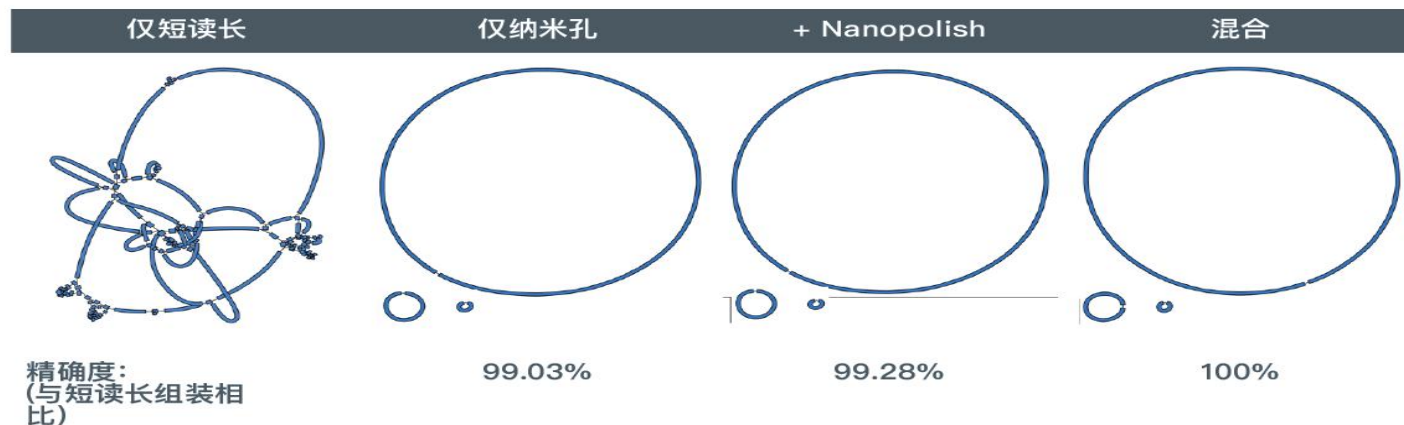
ARTICLES

<https://doi.org/10.1038/s41587-019-0156-5>

Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection

细菌基因组和质粒组装

采用低深度混样建库法，使用纳米孔技术 可以对**12**个细菌分离株进行测序，每个样 本仅需**80**美元



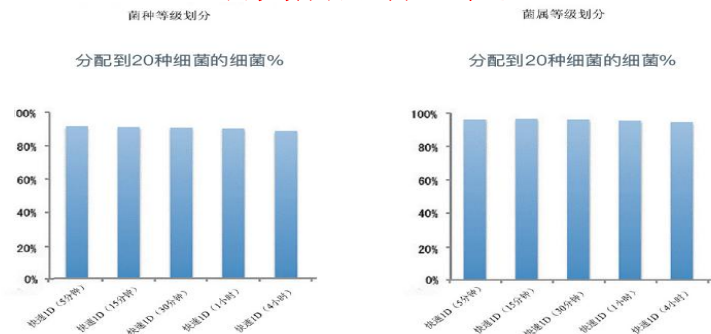
Wick, R. Hybrid。纳米孔和依诺米那组装：致力于完善细菌基因组。 介绍。可在以下地址查看：
<www.vimeo.com/219736365> [访问时间：2018年2月15日]

病原体快速鉴定

20种细菌等摩尔比值汇集16S rDNA, 制备细菌模拟群落。纳米孔测序能在5分钟内鉴定出所有20种细菌, 在菌种和菌属水平上鉴定精度分别为91%和97%。

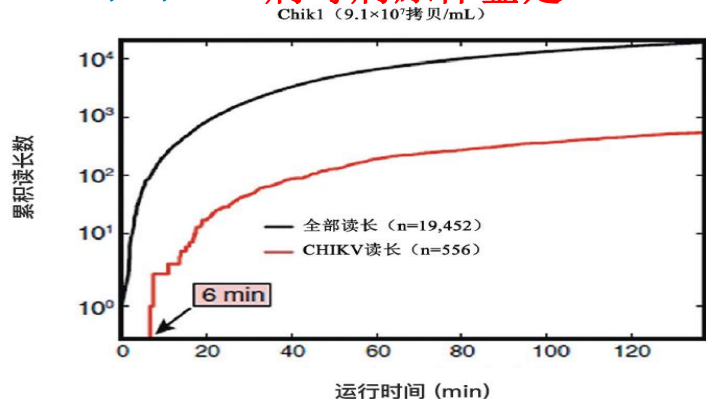
1D纳米孔测序过程中, 5分钟内实现了细菌模拟群落菌种和菌属级别分类, 精度达90%以上 (BLASTN)。图片由日本神奈川县东海大学医学院Imanishi教授提供⁷。

细菌病原体鉴定



在BCG鉴定和吡嗪酰胺耐药性检测中, 从患者采样到获得结果所用时间为7.5小时

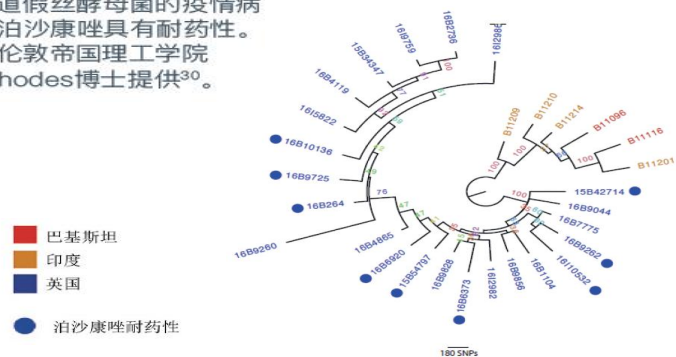
病毒病原体鉴定



临床样本中首个CHIKV读长在6分钟内完成测序, 进入测序运行。图片由美国加州大学旧金山医学院Charles Chiu教授提供; 选自发表于生物医学中心的一篇文章¹⁰中的图片。

取自不同国家的耳道假丝酵母菌分离株之间的进化关系。纳米孔测序显示部分耳道假丝酵母菌的疫情病源分离株对泊沙康唑具有耐药性。图片由英国伦敦帝国理工学院Johanna Rhodes博士提供³⁰。

真菌病原体鉴定





赣南医学院第一附属医院
First Affiliated Hospital of Gannan Medical University

以临床问题为导向的数据挖掘研究

Medical Big-data Center

郭有

Data-deriving Clinical
Research System, DCRS

支部书记、医学博士、副主任；
赣南医学院创新团队负责人；
赣州市医药大数据重点实验室主任；
双聘硕导（计算机科学与技术、公共卫生）；
江西研究型医院学会生物样本分会副主任委员；
江西省医学会医学信息学分会常委；
中国抗癌协会肿瘤流行病学委员会委员；
中国研究型医院学会临床数据与样本资源专业委员会委员。

Part 1	为什么做数据挖掘？	03
Part 2	如何申请临床数据？	07
Part 3	能解决哪些问题？	13
Part 4	如何确定数据集？	14
Part 5	常见问题有哪些？	14

为什么做数据挖掘？

1

为什么做数据挖掘？

李恭进书记8.19指示

2018年8月19日暑期干部学习班：

“建立医学生物信息大数据库，建立3-5个重大疑难重症疾病样本库”。

人事处2018.11.7批复同意成立

赣南医学院人事处

赣医人通字[2018]36号

关于同意第一附属医院（第一临床医学院）设置
精准医学中心等4个业务机构的通知

第一附属医院：

经2018年11月7日学校校长办公会研究，同意第一附属医院（第一临床医学院）设置以下4个业务机构：

一、设置赣南医学院第一附属医院（第一临床医学院）精准医学中心，为科研业务机构，挂靠第一附属医院检验科管理；

二、设置赣南医学院第一附属医院（第一临床医学院）医药大数据与生物信息研究中心，为科研业务机构；

三、设置赣南医学院干细胞临床转化中心第一附属医院（第一临床医学院）分中心，为科研业务机构，日常管理归口第一附属医院临床医学研究中心，业务指导归口赣南医学院干细胞临床转化中心。

四、设置药物/医疗器械临床试验机构（GCP），为一级临床业务机构。

专此。

赣南医学院人事处
二〇一八年十一月七日

医药大数据中心的职能

数据提取：是从我院电子病历系统、检查检验系统、病理系统等数据库提取相关临床数据的操作；

数据分析：是针对公共数据库中的高通量检测数据或本院的临床数据进行医学统计分析或数据挖掘；

计算资源：是指本中心的高性能计算集群的计算能力和存储能力。

样本保藏：略。

加强校本部与附属医院多学科研究合作

张小康校长指示：

学校科研人员联合附属医院职工组成课题组，由附属医院职工代表课题组按照审批程序向医院申请数据支持服务，科研产出由课题组共享。

基金委支持临床科学问题研究

新阶段 新理念 新格局

首届中国血液学科发展大会



孙瑞娟

国家自然科学基金委医学部常务副主任
研究员

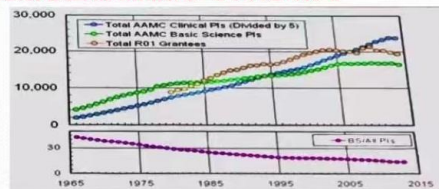
中国·天津

2021年1月29日-1月30日

加大支持临床科学研究

■ 以临床科学问题为导向、以服务临床实践为目标的临床与转化研究

- 医学研究源于临床实践，服务于临床需求
 - 美国NIH一半经费资助应用研究
 - 注册的临床研究约占11%
 - 我国转化和临床研究投入不足
 - 设立“源于临床实践的科学问题探索研究”专项项目（2021指南）
- 支持基于临床现象、临床问题、临床样本及基础研究成果临床转化的研究



近年美国基础研究的PI比例下降，临床研究的PI比例超过基础研究。



中国医学科学院



北京协和医学院

如何申请临床数据？



数据支持服务申请流程



无纸化单据，减少组织碳排放

提交

数据支持服务申请注意事项

赣南医学院第一附属医院（第一临床医学 院）医学科研数据使用审批表				用钉钉扫码
申请部门	临床科室-外科-普外一科(肝胆 胰外科)-护理单元	申请人		
联系电话	15870715179	申请时间	2021-10-20	
电子邮箱	gbb456@126.com			
支撑项目	项目类型	课题		
	项目编号	20195415		
	项目名称	加速康复外科理念在胆结石患者围手术期护 理中的应用		
需求类型	数据提取			
需求概况	提取内容: 2019.01月至2021.09月胆结石手术的信息, 性别, 年龄, 住院天数, 住院总费用, 术后并发症等数据			
科室负责人意见	同意 签字: 何晓 2021-10-25 09:13:49			
科研科负责人意见	同意 签字: 曾陈 2021-10-25 09:20:29			
分管院领导意见	同意 签字: 张胜 2021-10-26 05:46:27			
医药大数据中心 任务安排情况	数据分析者签字:	负责人签字: 郭存	2021-10-26 11:06:42	
注: 本表一式二份, 医药大数据中心、数据使用部门各一份留存备查。 联系电话: 0797-8687768。				

赣南医学院第一附属医院（第一临床医学 院）医学科研数据使用审批表				用钉钉扫码
申请部门	临床科室-外科-泌尿外科(江西 省尿路结石现代治疗中心)-病 区	申请人		
联系电话	17607247636	申请时间	2022-04-08	
电子邮箱	1510471112@qq.com			
支撑项目	项目类型	论文		
	项目编号			
	项目名称	输尿管息肉对内镜下处理上尿路结石的影响		
需求类型	数据分析			
需求概况	1、重点输尿管息肉分级(0级、1级、2级、3级) 与 B类中每一项 关系与影响, 研究价值, 以表格的形式呈现。 2、参考以下文献标记处表格 Title: Ureteral Wall Thickness as a Preoperative Indicator of Impacted Stones in Patients with Ureteral Stones Undergoing Ureteroscopic Lithotripsy (Table 2) DOI: 10.1016/j.urolgy.2017.04.047 Prospective evaluation and classification of endoscopic findings f or ureteral calculi (Table 4, Table 5, Table 6) DOI: 10.1038/s41598-020-69158-w			
科室负责人意见	同意 签字: 钱周 2022-04-08 15:33:56			
科研科负责人意见	同意 签字: 黄文册 2022-04-08 15:35:08			
分管院领导意见	同意 签字: 张胜 2022-04-08 22:54:55			
医药大数据中心 任务安排情况	数据分析者签字:	负责人签字: 郭存	2022-04-11 07:01:00	
注: 本表一式二份, 医药大数据中心、数据使用部门各一份留存备查。 联系电话: 0797-8687768。				

赣南医学院第一附属医院（第一临床医学 院）医学科研数据使用审批表				用钉钉扫码
申请部门	临床科室-外科-泌尿外科(江西 省尿路结石现代治疗中心)-病 区	申请人		
联系电话	15070760950	申请时间	2022-04-19	
电子邮箱	ydzjr@gmu.edu.cn			
支撑项目	项目类型	论文		
	项目编号			
	项目名称	甲基化表观遗传学修饰在嗜铬细胞瘤的作用		
需求类型	数据分析			
需求概况	数据提取源 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE67066 。 提取该研究所列良性嗜铬细胞瘤与恶性嗜铬细胞瘤, 分析其中GO注释包含 methylation的变异基因的表达热图。要求: 1、把表达数据提取出来了, 筛选出里 面的包含了methylation注释的基因做表达热图就行; 2、上述methylation相关基 因, 有表达差异的基因, 以2倍进行筛选, 如果还剩下比较多(>50), 则用4倍差异 进行筛选。即log2(FPKM)>2。3、其他的GO分析等就不需要了, 只要表达热图即可。			
科室负责人意见	同意 签字: 钱周 2022-04-19 15:50:39			
科研科负责人意见	同意 签字: 黄文册 2022-04-19 15:56:49			
分管院领导意见	同意 签字: 张胜 2022-04-19 22:54:30			
医药大数据中心 任务安排情况	数据分析者签字:	负责人签字: 郭存	2022-04-20 07:08:40	
注: 本表一式二份, 医药大数据中心、数据使用部门各一份留存备查。 联系电话: 0797-8687768。				

数据支持服务需求：目的明确、要求清晰

支撑项目	项目类型	课题
	项目编号	81760230
	项目名称	原位注射bFGF-ECM-HP温敏型水凝胶治疗脊髓损伤的作用和机制研究
需求类型	数据分析	
需求概况	1. 请分析GSE19890数据集中： SCI-1d组 (GSM497201 RatSpinalCord-Injury 1 day-1, GSM497202 RatSpinalCord-Injury 1 day-2, GSM497203 RatSpinalCord-Injury 1 day-3, GSM497204 RatSpinalCord-Injury 1 day-4, GSM497205 RatSpinalCord-Injury 1 day-5) 与Sham-1d组 (GSM497206 RatSpinalCord-Sham 1 day-1, GSM497207 RatSpinalCord-Sham 1 day-2, GSM497208 RatSpinalCord-Sham 1 day-3, GSM497209 RatSpinalCord-Sham 1 day-4, GSM497210 RatSpinalCord-Sham 1 day-5) ; SCI-3d组 (GSM497211 RatSpinalCord-Injury 3 day-1, GSM497212 RatSpinalCord-Injury 3 day-2, GSM497213 RatSpinalCord-Injury 3 day-3, GSM497214 RatSpinalCord-Injury 3 day-4, GSM497215 RatSpinalCord-Injury 3 day-5) 与Sham-3d组 (GSM497216 RatSpinalCord-Sham 3 day-1, GSM497217 RatSpinalCord-Sham 3 day-2, GSM497218 RatSpinalCord-Sham 3 day-3, GSM497219 RatSpinalCord-Sham 3 day-4, GSM497220 RatSpinalCord-Sham 3 day-5) ; SCI-7d组 (GSM497221 RatSpinalCord-Injury 7 day-1, GSM497222 RatSpinalCord-Injury 7 day-2, GSM497223 RatSpinalCord-Injury 7 day-3, GSM497224 RatSpinalCord-Injury 7 day-4, GSM497225 RatSpinalCord-Injury 7 day-5) 与Sham-7d组 (GSM497226 RatSpinalCord-Sham 7 day-1, GSM497227 RatSpinalCord-Sham 7 day-2, GSM497228 RatSpinalCord-Sham 7 day-3, GSM497229 RatSpinalCord-Sham 7 day-4, GSM497230 RatSpinalCord-Sham 7 day-5) 这三组组间两两比较差异表达的miRNA分别有哪些并绘制火山图和热图，然后基于三组组间中的差异 miRNA 的靶基因 (Rattus norvegicus)，识别三组中可调控 Notch1 基因的 miRNA，并取交集，绘制韦恩图。	

3 数据挖掘能解决哪些问题？

临床数据挖掘能解决哪些问题？

- 1、揭示人群疾病负担分布和发展趋势；
- 2、揭示发病机制和药物作用提供分子互作网络；
- 3、制订医学参考值范围；
- 4、疾病精准诊断和预后预测提供精准模型；
- 5、药效结构预测与新药发现；
- 6、发现联合用药方案；
- 7、医院运行效率分析。

4

如何确定临床数据集？

如何确定所需临床数据集？

定义临床问题是确定目标数据集前提、是数据挖掘主要任务，采用 PIOC 法则（Participants, interventions, outcomes, comparisons）。

“肝癌患者术后早期是否复发”，则： Participants：指明确诊断的肝癌患者； Interventions：指手术切除或进一步明确为具体术式； Outcomes：指早期复发或进一步明确为术后特定时间段的复发； Comparisons：指肝癌术后早期有复发和无复发的患者所形成的两个可比组，即复发组 vs 无复发组。

如何确定所需临床数据集？

临床科学问题是：

1、影响肝癌术后早期复发的因素有哪些？

生存分析；

探索新的影响因素；

特定因素效应比较。

2、哪些患者肝癌术后会早期复发？

预测模型；

在术前预测、在术后预测、在出院时预测。

5

临床数据挖掘常见问题？

临床数据挖掘中常见的问题

1、研究方向不明，临床价值不大。

1

研究方向不明

比如建专病数据库时研究问题不明确。

2

临床价值不大

低水平的重复性的统计描述研究；
临床应用场景缺如。

临床数据挖掘中常见的问题

2、研究问题有价值，但研究方案设计存在缺陷。



研究方案设计缺陷

1

肿瘤与过敏性皮炎

2

病例对照研究

记录缺乏

3

队列研究

人员流动

临床数据挖掘中常见的问题

3、研究前的探索周期过长而导致项目成功率不高。

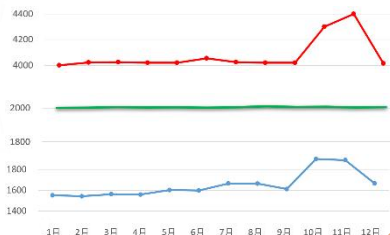
- 踩坑：研究前的可行性预判（如抽取病例看一看数据内容完整与否）要花费大量时间（抽取病历和筛选数据、清洗等数据预处理）。
- 超声引导下甲状腺消融术疗效研究；
- 腰椎融合术前注射封闭对术后感染有没有影响。

前期花了大量功夫后发现往往没法做下去、前功尽弃了。

临床数据挖掘中常见的问题

4、数据筛选所致的选择偏倚。

帕金森高发月份



1

帕金森病分布

2

分布曲线不均匀

3

对照选？

新生儿、其他病例？

4

医院就诊人群特点

临床数据挖掘中常见的问题

5、数据清洗工具灵活适用性不足

任务：数据筛选、结构化特征提取、数据转换

公司：承诺太多（利用机器学习实现结构化、特征提取）

。实际：人工标注时间长但机器学习受限，没有太大价值

○

检查所见

前列腺较大，最大约54毫米，中央叶与外周带分界尚清楚，外周带见小条片状稍短T2异常信号影，DWI呈等信号，中央带左侧见大小约7毫米等T1略长T2结节影，DWI呈高信号，超高B值仍呈高信号，动脉期中度强化，实质期及延迟期持续轻度强化。前列腺包膜完整，双侧精囊腺、膀胱和直肠未见异常，盆腔未见异常肿大淋巴结，所见骨质未见破坏征象。

印象

1.前列腺异常改变，考虑：前列腺增生，前列腺炎可能性大。2.前列腺中央带左侧异常结节，考虑增生结节可能，前列腺癌不排除，建议结合PSA检查，必要时超声引导下穿刺活检。

➡

前列腺大小	54毫米
实质期	持续轻度强化
病灶部位	外周带
延迟期	持续轻度强化
动脉期	中度强化
病灶边缘	尚清楚
超高b值DWI	
高b值DWI	
DWI	等信号
ADC	
T2WI	稍短
PI-RADS	

临床数据挖掘中常见的问题

• 6、数据选择不当与数据处理弯路。

- 数据来源多：首页、病例、医嘱；
- 药物商品名多：厂家不一样商品名不一样，但是又是同类药；
- 药品代码：

患者ID	医疗项目名称	项目代码	费用
	奥西康	F0101J01	
	注射用奥美拉唑钠	F0101J00	
	硝苯地平缓释片 (II)	N0201TA01	
	(德州博诚)		

药品代码	通用名	使用日期区间
F0101J01	注射用奥美拉唑钠	2000-2016
F0101J00	注射用奥美拉唑钠	2017-
N0201TA01	硝苯地平缓释片 (II)	2000-2015

● 感谢聆听 敬请指导 ●