



承启生物
Chi Biotech

基因测序云分析解决方案

专业可靠 触手可及

www.chi-biotech.com
info@chi-biotech.com

新一代测序很棒但难以把握？

新一代测序种类繁多，项目中的测序部分不知该如何下手

您不用对那些复杂专业的名词望而却步。我们专业的团队可以给您充分的科学与技术支援，为您的课题设计最优方案。您只需要配合我们提供样品，剩下的工作请交给我们，从策略设计到分析结果报告，为您提供一站式的服务。

新一代测序数据太难分析，需要各类专业知识：生物分子学、信息学、分析算法、甚至编程……

您不用担心，使用我们的平台，您不需要懂生物信息学，不需要懂Linux，不需要懂算法，不需要懂编程，甚至连参数都不需要设置。您只要知道您测的是什么东西，就可以点点鼠标完成专业分析。

新一代测序分析计算量太大，要配置高性能服务器

您尽管放心，海量的计算任务由我们承担。我们依托于运算速度世界第一的天河二号等超级计算机和特别配置的测序分析计算集群，轻松完成所有已知的分析过程。您可以用普通的PC、平板、智能终端，向云平台上传数据，设定任务，稍作等待即可下载结果。

请某公司做新一代测序分析，却没有过程数据，结果让人看不懂

我们的报告清晰易懂，配有详细的说明以帮助理解。同时，我们会不定期举办一些培训和讲座，手把手地教您如何解读这些数据、如何把这些数据做成一篇论文的重要部分。当然，详尽而专业的数据报表，我们也会毫无保留地提供，供专业用户使用。

新一代测序好是好，就是测序费用太高了

我们极其精确的算法，可以在同一批测序数据中，挖掘出比其他算法更多更准确的信息。使用我们的算法和平台，您可以少测几倍甚至几十倍的数据量，达到同样的甚至更高的精度，显著降低成本，满足您的需求。同时，云平台非常简便易用，您可以自己操作分析，不需要高昂分析费用。

花大价钱去测序，鉴定的突变、差异表达基因却不符合实验结果，感觉被坑……

这不仅仅是您的经验，也是Nature Review作者的肺腑之言。传统算法的不可靠，导致测序项目得拼人品。在我们的云平台上，您无需提心吊胆，也无需赌运气。云平台算法超高的精度和极佳的稳定性，保证了结果的可验证性，并已得到了诸多项目及论文的证实。如果您碰到了不靠谱的公司，我们能以业界权威的经验和技术，独立为您审核测序公司的数据质量和分析结论。如有需要，我们也可以为您推荐靠谱的测序公司。

以往的大规模测序竟然错漏百出！

NATURE REVIEWS | GENETICS

VOLUME 13 | SEPTEMBER 2012 | 667

 APPLICATIONS OF NEXT-GENERATION SEQUENCING — OPINION

Next-generation sequencing data interpretation: enhancing reproducibility and accessibility

Anton Nekrutenko and James Taylor

2012年，权威期刊 Nature 发表综述，揭露了一个令人震惊的真相：

are representative, then most results reported in today's publications using NGS data cannot be accurately verified, reproduced, adopted or used to educate others, creating an alarming reproducibility crisis.

“现在绝大多数的基于测序的结果都不能被验证、重复、采纳或用于教育他人，造成了迫在眉睫的可重复性危机。”

(Nature Reviews Genetics, 2012 (13), 第668页右上)

正因为如此，以往的大规模测序用于医疗和临床是极其危险的：

“基因咨询与诊断极易出现假阳性、假阴性，对接受测序者是极不负责的。”

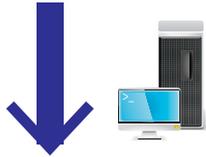
—— 中科院上海生命科学研究院 仇子龙研究员

(2014年5月11日 新华网报道 http://news.xinhuanet.com/2014-05/11/c_1110633108.htm)

不可靠的传统算法：隐形的杀手



NTCAAATAAAAAAAGAAAGACCGTACCACAAAAGGTTACTATAGCCAAAATCCCAGAGCAAAAGAAAGAA
 CCTGGTGCCTGGCCGCGCCGCCCGTGTACCGTATTGTAAGAAACAGCCGTACCCAAAGTCTCG
 TTGACATCTGACGACGTGAAGGAGCAGATTACAACATGGCCAAAGAGGGCTTACTCCTTACAGATCGGTGT
 CTGATTTTGACATGACCTTGAGCAGGTTTGCATATAATGAAAGCAGATGCCCTTCTTACAATATCTGGAT
 GTTTAGAGAATTTTTCTTTAATTTTTTTAACAAGGATATCAGATGGAGGACATAGATTACGCTACC
 TGACCACATGCTGGCTTGAGCAACCCAGAGAATATCGCCACTTGTATTGCTCTCCGGCCCTACCCAAAGAAAG
 CCAGGTTTGGTGACAATAGAAAGAACCCAGCTTAAAGATATGGCTATTGCTACTGGTGGTGCAGTGGTTGGAGA
 GAAACAGGAAAAAATCCTAAACACAAAAGGACCTAGTTCGTGAGAACATTAAGCAAAAATGCAAGCAA
 CTCGAAGCCGTGCGGGACAAGCGGGCAGACATCTGTGAAGCAGGACCTGCTGAAGAGGAGACTTCTATGC
 GTAATAATCTTTGAATTAGAAAGGTTGGGACAGAAAGTACTTTATGTAACAAAGTGGGCTGTCAGAAGC



参考基因组

5' - CCTCTCAACATTGAGGTCGCCAAAATCAGCTCCACAGCCTCATTCTCGACTTTTCAGCAGTGTCTCTTCTTATGATTTCTTCAGTGGGGCCCTAAA - 3'
 3' - GGAGAGTTGTAACCTCCAGGGTTTTAGTCGGAGGTGTCGGAGTAAGAGCTGAAAAGTCGTCACAGGAAAGAACTACAAAAGTCACTCCCGGAATTT - 5'
 3' - GGAGCGTTGTAACCTCCAGGGTTTTAGTCGGAGGTGTCGGAGTAAGAGTT - 5'
 3' - GTTGTAACTCCAGGGTTTTAGTCGGAGGTGTCGGAGTAAGAGTTGAAAA - 5'
 3' - AACTCCAGGGTTTTAGTCGGAGGTGTCGGAGTAAGAGTTGAAAAAGTCGT - 5'
 5' - ctccaggggttttagtcggaggtgtaggagtaagagttgaaaaagtcgtca - 3'
 3' - CCAGGGTTTTAGTCGGAGGTGTCGGAGTAAGAGTTGAAAAGTCGTCAACA - 5'
 5' - ggggttttagtcggaggtgtaggagtaagagttgaaaaagtcgtcacagga - 3'
 3' - TTTTGGTGGGAGGTGTCGGAGTAAGAGTTGAAAAGTCGTCAAGGAAAG - 5'
 3' - TTTAGTCGGAGGTGTCGGAGTAAGAGTTGAAAAGTCGTCAAGGAAAGAA - 5'
 3' - GTCGGAGGCGTCGGAGTAAGAGTTGAAAAGTCGTCAAGGAAAGAACTAC - 5'
 5' - cggaggtgtaggagtaagagttgaaaaagtcgtcacaggaagaactacaa - 3'
 3' - GGGGGGTCGGAGTAAGAGTTGAAAAGTCGTCAAGGAAAGAACTACAAA - 5'
 5' - gaggtgtaggagtaagagttgaaaaagtcgtcacaggaagaactacaaaag - 3'
 3' - GGGTCGGAGTAAGAGTTGAAAAGTCGTCAAGGAAAGAACTACAAAAGAA - 5'
 5' - tcggagtaagagttgaaaaagtcgtcacaggaagaactacaaaagaaagca - 3'
 3' - GAGTAAGAGTAGAAAAGTCGTCAAGGAAAGAACTACAAAAGAAAGTCACTC - 5'
 5' - agagttgaaaaagtcgtcacaggaagaactacaaaagaaagcaactccccgg - 3'
 3' - GTGAAAAGTCGTCAAGGAAAGAACTACAAAAGTCACTCCCGGAAT - 5'

测序所得
短核酸序列

基因测序的关键两步：

第一步：大规模测序实验

将DNA或RNA随机打断，使用二代或三代测序仪进行测序，得到几百万至几亿条短核酸序列(reads)。

这一步所用的试剂、仪器和方法已千锤百炼，获得FDA认证，可靠性高。



第二步：序列比对(mapping)与分析

运用计算机算法将测序得到几百万至几亿条短核酸序列(reads)向参考基因组序列进行比对(mapping)，找到每一个短核酸序列的位置，由此才能找到突变、测定基因表达量，以及做出医学诊断。

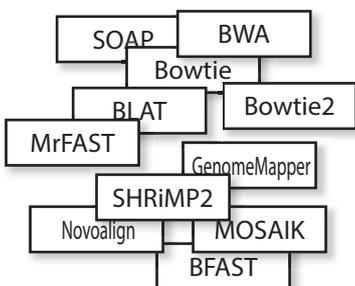
这一步可靠吗？

传统算法的重大缺陷：

- 不准确
丢失许多本来可以比对上的序列
- 不可靠
比对效果时好时坏，计算精度无保障，结果靠运气
- 难验证
由于不准确不可靠，计算结果常常不符合实验事实
- 难操作
通常需在Linux命令行下操作，参数设置复杂、普通生物学家和医生难理解，且非常影响结果
- 耗时间
计算量巨大，需要高性能计算机

不可靠的算法是造成测序结果错漏的隐形杀手！

现在几十种 mapping 算法



同样的测序数据，向同样的基因组比对，不同的算法结果大相径庭！

	Time (s)	% mapped
BFAST	43,775	32.1
BLAT*	68,758	24.3
Bowtie	2,270	13.1
BWA	7,682	16
MAQ	8,607	28.7

PLoS ONE 2009, 4(11):e7767



我们的答案： 承启生物基因测序云分析解决方案

为什么使用 FANSe2算法+云分析平台？

(承启生物基因测序云分析解决方案基于FANSe2算法)

	承启生物云分析解决方案 (FANSe2+云分析平台)	传统解决方案 (采用传统算法)
精度	FANSe2 算法的精度有概率论的理论保证，错误率低至百万分之一以下，在实际应用情况下可基本保证100%准确。	传统算法准确度无保证，丢失许多本可以对上的序列，导致结果失准。
稳健性	FANSe2 算法的精度有概率论的理论保证，错误率可事先估计，便于用户在计算前就调整合适参数以达到实际的精度需求。	其他算法的精度无法有效估计，而且对于不同的测序仪、数据集和物种，精度差别显著。
容错性	FANSe2算法的原理使其可完美应对12%的测序错误率或序列变异度。对更高测序错误率或变异度（高达20%甚至更多），还可以通过逐次累进法完美应对。	传统算法一般在2%的测序错误率或序列变异度下就已经开始失去准确度，大部分算法在4%时已完全失去实用意义。而一般的测序仪测序错误率即高达2~4%，三代测序仪甚至高达13%之多。
实验可验证性	在相关论文中已发表的一系列实验测试结果，全部验证了FANSe2分析结果的正确性，无一例假阳性或假阴性。迄今为止，光是已报告的实验就验证了数百个基因的表达量定量情况和数千个突变鉴定情况，实验验证结果（无论定量还是定性）均与FANSe2分析结果吻合，无一错误。	其他算法的假阳性率和假阴性率相对FANSe2非常高。已有充分的实验证据证明传统算法所鉴定出的突变、基因表达状况的错漏很多，不符合实际情况。
成本	在人类蛋白质组计划的研究中已证明，使用FANSe2只需要100M碱基测序通量即可在人细胞中准确定量超过11000个基因，测序成本比使用传统算法节约50~100倍。	传统算法由于错漏多，只能大幅度提高测序通量来进行有限的弥补。通常人细胞转录组测序需2~4G碱基数据量，相比FANSe2的测序通量高50~100倍，极大地提高了成本。
测序仪兼容性	FANSe2可从容应对所有一代、二代、三代所有品牌所有型号的测序仪。使用不同品牌不同型号的测序仪所测数据均可定量对应，方便做长期的研究以及与其他单位数据的整合分析。	传统算法对某些品牌和型号的测序仪计算结果不佳，而且使用不同品牌和型号的测序仪，其定量结果不可比。一旦该型号测序仪停产，其结果即不可重现。
基因芯片兼容性	使用FANSe2算法处理测序数据，可与基因芯片定量数据线性对应，充分利用文献中报道的和其他研究组所产出的基因芯片数据。	传统算法处理的测序数据无法与基因芯片对应，这使得以前大量基于基因芯片的研究成果无法利用。
非模式物种	利用FANSe2算法极高的精度、稳健性和容错性，可轻松处理各种基因组信息不准或不全的非模式物种，例如各种临床分离的菌株、动物感染性标本、尚无基因组的动植物物种等。	传统算法很难在基因组信息不准或不全时进行分析，难以应对非模式物种。
速度	在专门配置的高性能服务器和超算平台上分析，拥有目前世界上最强的计算能力，速度极快，理论上可在一年内将13亿中国人的基因组全部计算一遍。您只需通过您的电脑、平板等上传数据即可。	自己购买高性能工作站或集群十分昂贵、设置复杂、维护困难，计算速度更是无法与云平台相比拟。测序量大时往往疲于奔命。
易用性	在全可视化界面下，只需轻点鼠标即可完成原本复杂的参数设置。大部分参数自动生成，无需专业背景也能轻松完成全过程，高中生竟能发SCI论文。	Linux命令行界面，大部分生物学家和医学专家难以操作。参数复杂繁冗，需具备很强的生物信息学背景和编程知识才能理解和使用。
自动化	数据分析请求提交后，整个分析过程可无人值守自动完成，平台可以自适应用户数据。批量上传、批量设置、批量分析等功能，满足基因测序公司、大型科研或医疗项目等海量数据用户的使用需求，为用户节省时间与人力，也使得分析标准化。	多步分析操作需要人工干预，针对不同的测序仪和测序设定需人工设置经验参数，费时费力且不易标准化，更谈不上自动化。
结果易读性	在提供专业的数据报表的同时，生成普通用户易懂的报告，甚至可针对特定医疗应用，直接输出老百姓都能懂的检验报告单。我们还会不定期举办讲座和培训班，手把手地教您如何更好地应用结果。	所出结果只是大篇幅的晦涩数字，普通用户几乎无法理解，必须具备非常专业的生物分子学及信息学知识、甚至编程能力才可以理解与应用。

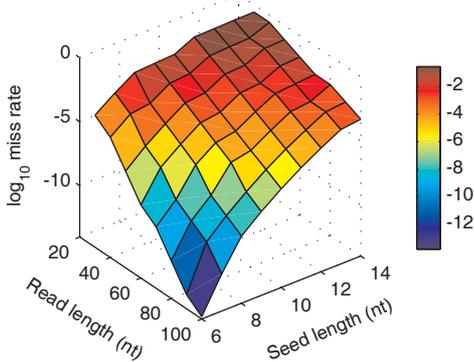
详情请见后页

FANSe2 算法近乎完美的精度

FANSe 是一种超高精度mapping算法，到目前为止发展了两代，最新版本为 FANSe2。

FANSe系列算法以准确度为首要设计目标

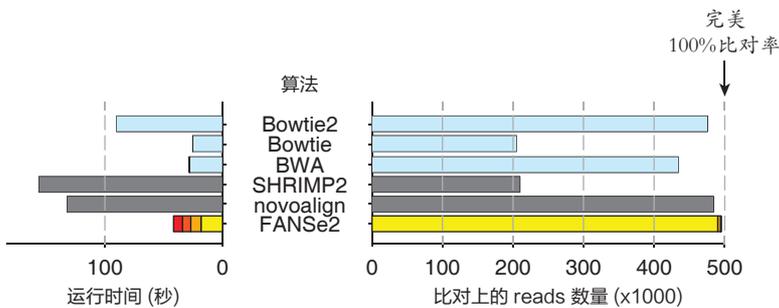
FANSe2的reads丢失率（理论错误率）。



近乎完美的精度，且有理论保证

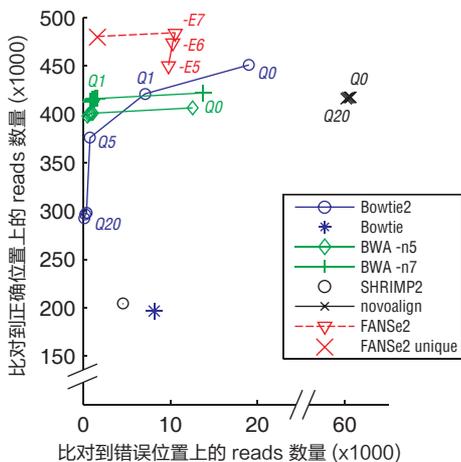
FANSe系列mapping算法的精度极高，在绝大多数情况下，其错误率可低至百万分之一；在实际应用中，基本可保证100%的准确率。这种精度事实上已是当今全球算法的极限，相比其他算法有数量级的精度优势。

而且，FANSe系列算法是迄今为止唯一的精度可由概率论保证的mapping算法，这就保证了其超强的稳健性，不会因为数据集和测序仪的变化而有大幅波动。而且误差率可以在计算之前就预先估计，便于用户调整合适参数以达到实际的精度需求。



测试1：从人的1号染色体序列（2.49亿碱基对）中模仿测序过程，随机产生50万条reads，含4%的错误率。几种常用mapping算法与FANSe2进行比较。

结果：FANSe2比其他任何算法都接近100%的比对率；其他算法可丢失大量本可以比对上的reads。



测试2：同上，评估各种mapping算法是否能将reads比对到基因组正确的位置上。每个点代表一种算法在一种参数设定下的成绩，越靠上表明正确率越高。

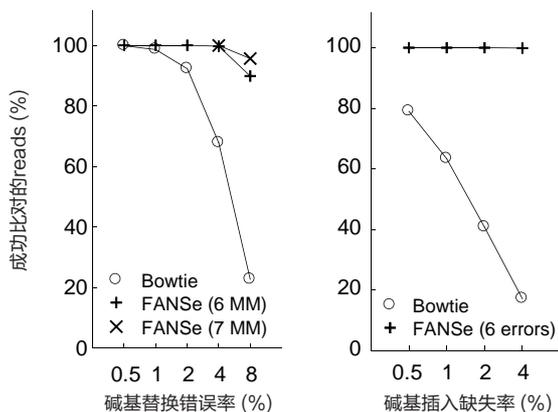
结果：FANSe2比其他任何算法的正确率都高。

更多数据和测试结果，请参见已发表的FANSe、FANSe2算法论文：

Zhang, G.*, Fedyunin, I., Kirchner, S., Xiao, C., Valleriani, A. & Ignatova, Z.*, FANSe: an accurate algorithm for quantitative mapping of large scale sequencing reads. *Nucleic Acid Res.* (2012) 40(11):e83.

CL Xiao, ZB Mai, XL Lian, JY Zhong, JJ Jin, QY He *, G Zhang *, FANSe2: A Robust and Cost-Efficient Alignment Tool for Quantitative Next-Generation Sequencing Applications. *PLoS One* (2014), 9(4):e94250

超强的容错性，轻松应对非模式物种

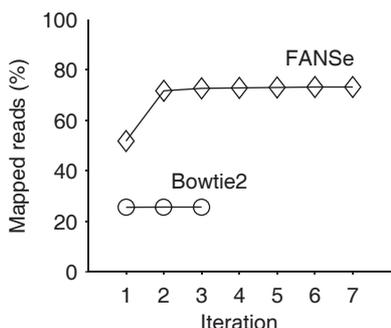


Nucleic Acid Res. (2012) 40(11):e83.

超强的容错能力，无惧高错误率和高变异率

FANSe系列算法在8%错误率时可以保持极高的比准确率，而此时传统算法的比对率已降至20%，已完全丧失实用意义。除此之外，FANSe系列算法对碱基的插入缺失(indel)有着完善的处理机制，因此在高达4%的插入缺失率时仍保持几乎完美的准确率，而此时传统算法已完全不堪使用。

因此，FANSe系列算法可轻松应对非模式物种的研究，同时也可以从容应对高错误率的三代测序仪数据。



案例1：临床分离菌株基因组研究

临床分离一株短小芽孢杆菌(*Bacillus pumilus*)，与标准模式菌株*Bacillus pumilus* SAFR-032基因组差异度高达5%。用Illumina HiSeq-2000测序仪测序其基因组，将reads向标准模式菌株参考基因组进行逐次累进比对，以寻找基因组突变及修正基因组。

结果： FANSe 比传统算法 Bowtie2 多比对上**三倍**数量的reads。

由于Bowtie2的比对率低下，只鉴定了**169**个突变；而FANSe因有极高的容错性，能将差异度高达20%的reads成功比对，共鉴定了**182620**个突变位点，是传统算法的1020倍以上，且得到实验验证（见后）。

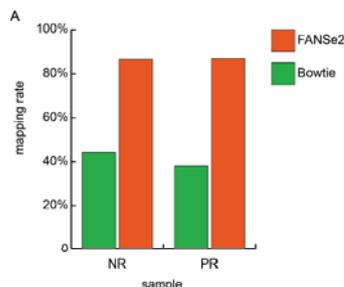
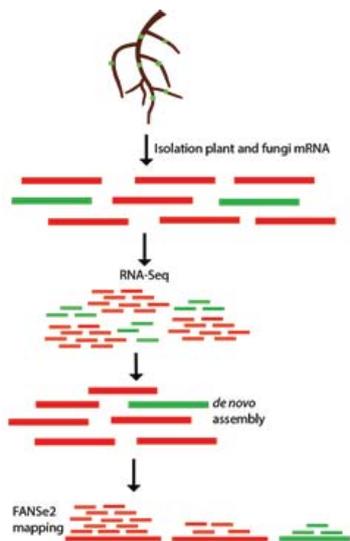
Journal of proteome research (2014), 13 (6), 2724–2734

案例2：无参考基因组植物共生体系宏转录组研究

水生植物凤眼莲迄今尚无参考基因组，其同科植物也无参考基因组。为研究凤眼莲根系在某药物处理时其根系与共生菌群的转录组变化，直接采用宏转录组测序从头拼接的策略，再将reads向拼接出的宏转录组草图上进行比对，进行定量与差异表达基因分析。

结果： FANSe 比传统算法 Bowtie 多比对上**两倍**数量的reads。

由于Bowtie的比对率低下，只鉴定到几十个差异表达的基因拼接片段(contigs)，难以进行后续功能富集分析。而FANSe2比对的结果鉴定了超过2000个差异表达的基因拼接片段，其基因功能分析结果完全符合药物处理后其性状的变化。



FANSe2算法的结果可被实验验证

APPLICATIONS OF NEXT-GENERATION SEQUENCING — OPINION

Next-generation sequencing data interpretation: enhancing reproducibility and accessibility

Anton Nekrutenko and James Taylor

are representative, then most results reported in today's publications using NGS data cannot be accurately verified, reproduced, adopted or used to educate others, creating an alarming reproducibility crisis.



回答Nature之问：使用FANSe2, 可重复性危机一去不返！

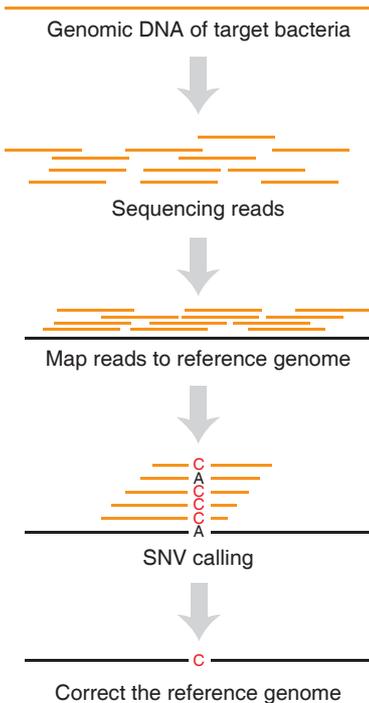
权威期刊2012年刊发综述，指出绝大部分大规模测序结果无法被实验验证和重复。

已发表的论文中已有大量实验证据，证明用传统算法进行分析，有大量的测序计算结果无法被验证，包括碱基突变、基因表达等。而用FANSe2算法得到的测序计算结果，进行实验验证无一错漏，完美符合实验事实。

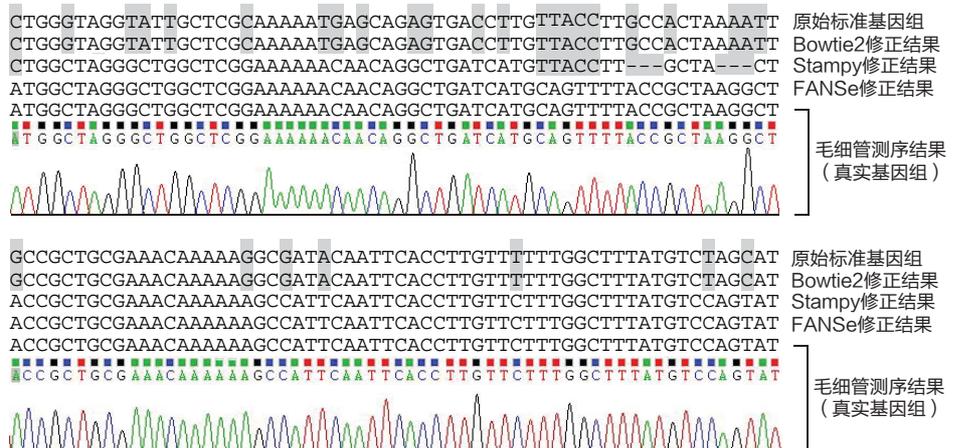
因此，FANSe系列算法解决了Nature所指出的测序重复性危机，您可以得到符合事实的结果。

案例1：临床分离菌株修正基因组

临床分离一株短小芽孢杆菌(*Bacillus pumilus*)，与标准模式菌株*Bacillus pumilus* SAFR-032基因组差异度高达5%。用Illumina HiSeq-2000测序仪测序其基因组，将reads向标准模式菌株参考基因组进行逐次累进比对，以寻找基因组突变及修正基因组。左图是该研究的策略流程图。随机挑选1994个位点，使用毛细管测序(一代测序)方法测定真实的基因组，以检验几种算法检测突变和修正基因组的能力。



结果： 在实验验证的1994个位点中，FANSe 修正基因组的结果无一假阳性、无一假阴性，完美符合真实基因组。而传统算法 Bowtie2 和 Stampy 修正的结果存在大量的假阳性和假阴性，其修正的基因组有许多碱基不符合真实状况，被实验所否定。下图展示了部分实验验证结果，标灰色的碱基为不符合实际的碱基。这一结果显示了FANSe2惊人的准确性和可验证性，并证明传统算法结果不可靠。



Journal of proteome research (2014), 13 (6), 2724-2734



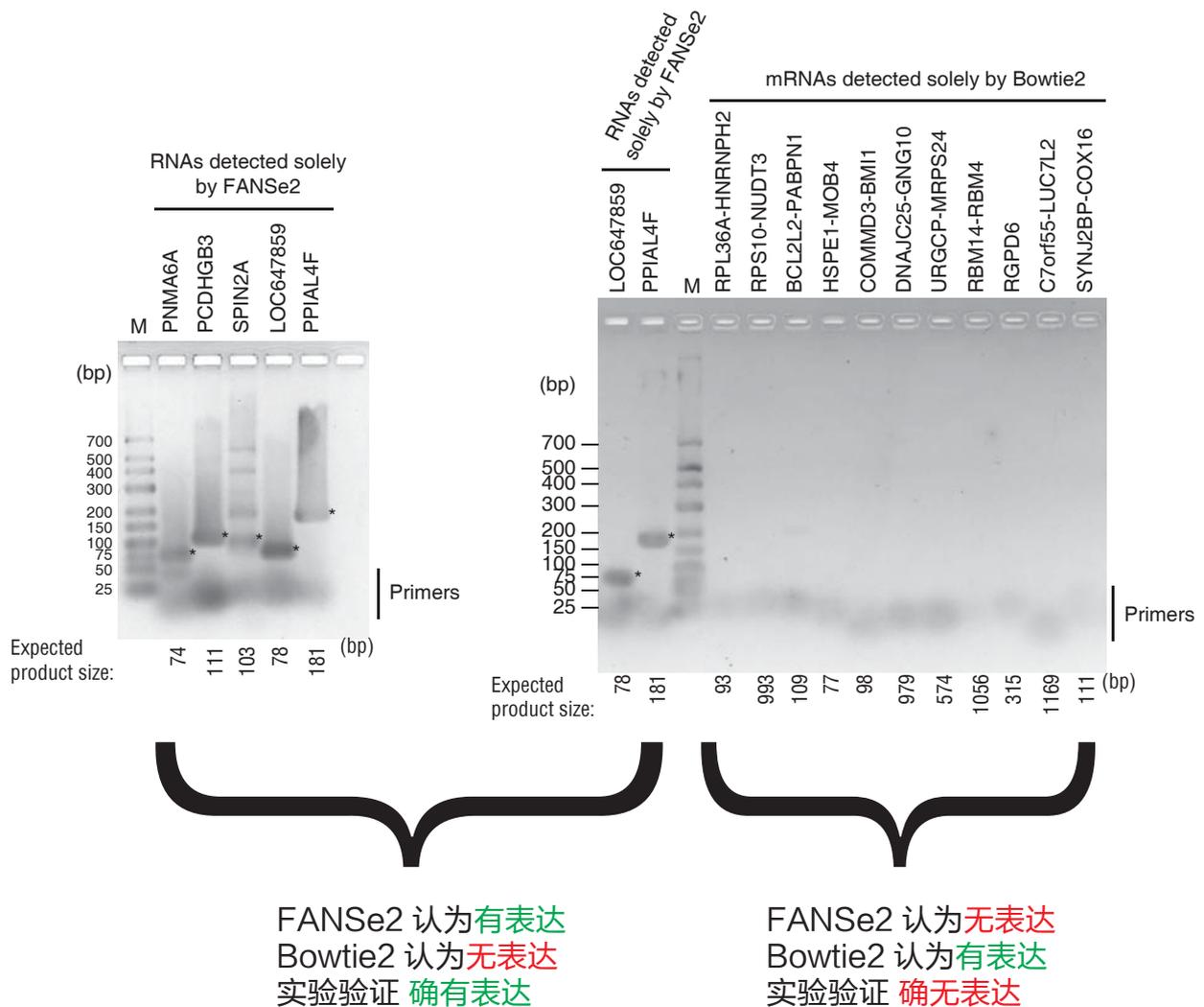
FANSe2算法的结果可被实验验证（续）

案例2：人肺癌细胞系A549转录组基因表达鉴定

提取人肺癌细胞系A549的 polyA+ RNA，用 Illumina GAllx 测序仪做转录组测序，所得的reads向人参考转录组 NCBI RefSeq-RNA库进行比对，使用算法有 FANSe2 和 Bowtie2。使用完全相同的标准，鉴定哪些基因有表达。对一种算法鉴定有表达而另一种算法鉴定无表达的基因，按表达量降序排列以排除人为选择因素，用RT-PCR逐一进行实验验证，检测其是否真有表达。

结果： FANSe2 认为有表达而 Bowtie2 认为无表达的基因，全部被实验证实有表达。Bowtie2 认为有表达而 FANSe2 认为无表达的基因，全部被实验证实无表达。

亦即，**FANSe2 的结果完全符合实验事实，Bowtie2 的结果则完全不符合事实。**



PLoS One (2014), 9(4):e94250

FANSe2算法的结果可被实验验证（续）

案例3：极低通量小鼠转录组差异表达基因分析

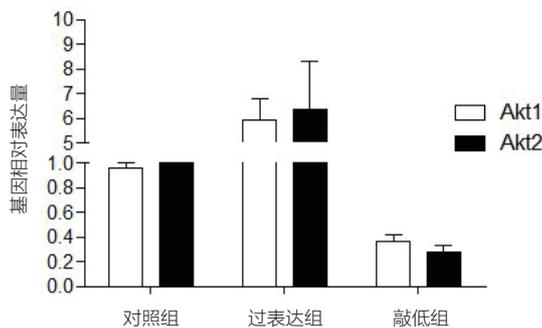
在小鼠细胞内过表达和敲低某基因的表达，要研究这些状况下相对对照组，整个转录组中有哪些重要通路的关键基因的表达量发生改变。为节省成本和尽快拿到结果，使用 ion torrent PGM测序仪进行低通量测序，每个样本中只得到了10多万有效的reads（即0.1M reads），通量仅为通常转录组测序通量的千分之一。

结果：即使仅有0.1M reads，FANSe2的mapping结果依然定量了约2000个基因。通过edgeR软件进行差异表达分析，发现两个与研究项目非常相关的基因Akt1和Akt2，虽然有的样本中只有十几个reads，但依然能检测出其差异表达。通过qRT-PCR进行验证，测序分析结果连上下调倍数在误差范围内都完全符合qRT-PCR的实验验证结果。

云平台测序分析结果：

过表达组：Akt1上调5.0倍
Akt2上调5.1倍
敲低组：Akt1下调至0.43倍
Akt2下调至0.35倍

qRT-PCR验证结果：



案例4：肺癌病人突变基因检测

对5例肺癌病人手术取出的癌组织和癌旁组织进行全外显子测序，将Illumina HiSeq-2000所测得的reads向人参考基因组hg19进行比对，分析其基因组突变位点。国内两家最著名的测序公司分别采用TopHat+Samtools软件和BWA+GATK软件，而我们使用云平台的FANSe2算法。

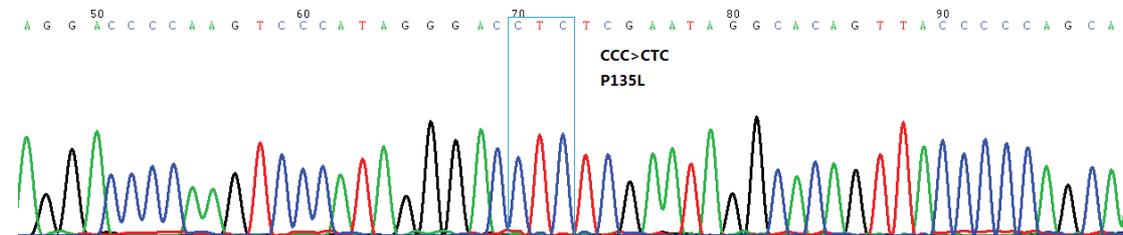
结果：从10个样本中，FANSe2比TopHat+Samtools和BWA+GATK两种方案多鉴定出超过20000个纯合突变。

其中，与肺癌高度相关的FGFR4基因中，FANSe2从10个样品中都检测到了C404T纯合突变，这会导致P135L的氨基酸改变。而其他测序公司的分析未能检测出这一突变。PCR和毛细管测序结果都验证了C404T纯合突变确实存在（见右图）。

此外，随机验证5个我们云平台能检测到但传统算法不能检测到的纯合突变，均能验证出。这说明FANSe2所检测到的临床样品中的突变能被实验验证，传统算法存在大量的错漏。



对FANSe2检测出的C404T纯合突变进行突变特异性PCR验证。10个肺癌病人样本（5个癌组织、5个癌旁组织），全部验证出有此突变。



其中一个样本的毛细管测序结果

C>T纯合突变

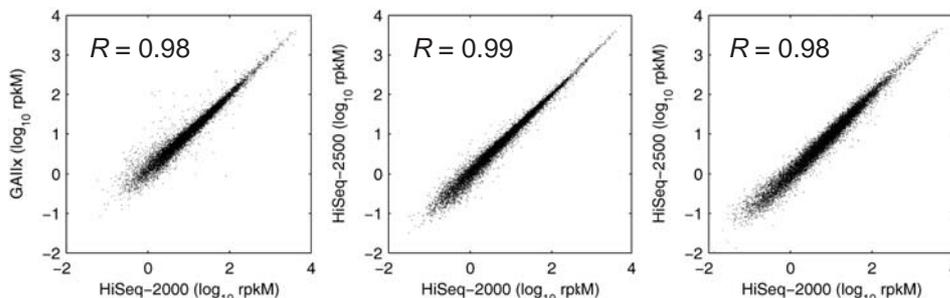
FANSe2能检测出（承启生物云平台）
Tophat+Samtools不能检测出（公司H）
BWA+GATK不能检测出（公司Y）

FANSe2极强的兼容性： 不同测序仪的数据照样可以线性对应 甚至测序和芯片的数据也能线性对应

癌细胞微量转录组测序 定量重复性实验（生物学重复）

- 不同的人来做实验
- 不同的测序建库试剂盒
- 不同的测序公司进行测序
- 不同型号的测序仪
- 相同的FANSe2算法

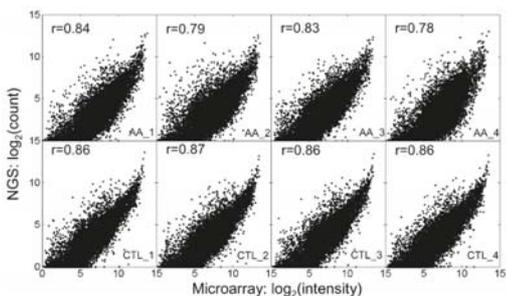
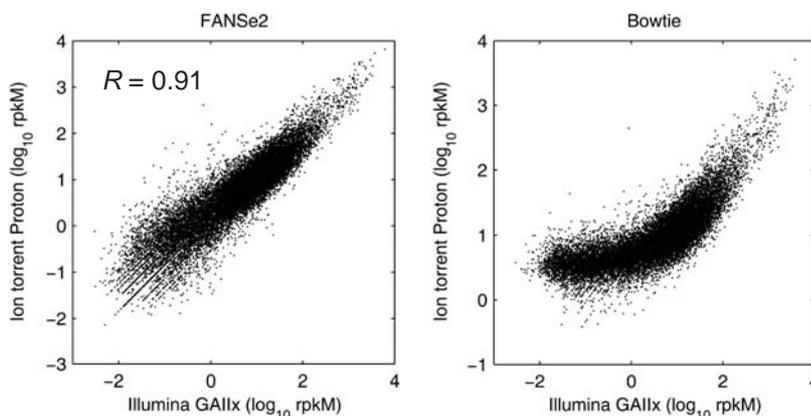
定量结果线性相关度高达
R=0.98~0.99



同一管肺癌A549细胞的mRNA，分别用 Illumina GAlIx 测序仪和 Life Technologies Ion Torrent Proton 测序仪进行测序定量。

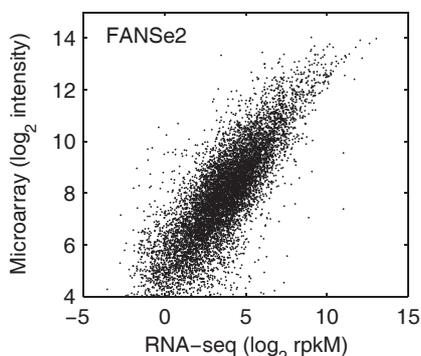
用FANSe2做mapping，两种测序仪的定量结果可线性对应，R=0.91。

用传统的Bowtie算法做mapping，两种测序仪的定量结果不能线性对应。



Comparing next-generation sequencing and microarray technologies in a toxicological study of the effects of aristolochic acid on rat kidneys
Chemical Research in Toxicology (2011), 24(9):1486-93

在这篇文章中，作者用8个大鼠样品同时做了Illumina测序和Affymetrix芯片，用传统算法Bowtie, SOAP2, BWA进行mapping，发现测序对检测低丰度mRNA存在严重的偏差。（左图）

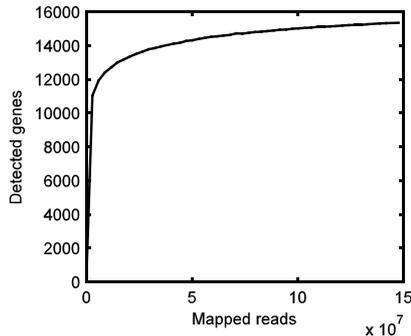


我们下载其数据，发现用FANSe2进行mapping，问题迎刃而解：FANSe2可使测序数据和芯片数据线性对应。（左下图）

PLoS One (2014), 9(4):e94250

无论您用什么测序仪，您的数据都可以和他人的数据进行对应分析，也可以利用前人大量珍贵的芯片数据。

云平台可大幅度降低测序成本



测转录组，多少reads才够用？

一般的测序公司都建议一个样品测2~4G数据量，收费高达上万元。

2013年人类蛋白质组计划(HPP)肝癌细胞翻译组测序的实践表明：
使用云平台的FANSe2算法，在2M reads (100M数据量) 时，
即可准确鉴定和定量11000个编码基因。

**FANSe2可以使测序通量降低50~100倍，
达到同样的定量精度，大大降低测序成本。**

Journal of proteome research (2014), 13 (1), 38-49

不使用云平台

VS

使用承启生物云平台

要完成一个测序项目，您不仅需要付出测序实验的成本，
还需要付出大量的计算和人力成本，包括：



昂贵的高性能计算集群 以及随之而来的维护费用

高速率、大容量、多核数计算集群的购置费用不菲，随之而来是
场地费、电费、空调费……



经验丰富的系统工程师 来运行集群和网络

2014年，初级系统工程师平均月薪1万元。而能配置系统、配置算法、
配置模板的工程师薪水更高。



付出高昂的成本培训工作 人员使用集群和进行基础 分析

培训时间一般不少于三个月，总付出成本约2~3万元，但很多员工
2~3年后就将跳槽。



雇佣优秀的生物信息学分析 人员来分析数据

2014年，普通生物信息学分析工程师的月薪约1.5万元。有能力进行多模块
准确分析的高手更是可遇不可求。



用普通PC、平板电脑上传数据
轻点鼠标/轻点屏幕设定目标
一杯咖啡的功夫
最精确专业的分析结果就到您手上了

左边那些成本都与您无关。

您还能持续享用最新的功能模块、算法、以及数据库

云平台让您更快更容易出成果

郭杰明（右）与指导老师讨论课题



由于使用云平台进行测序数据分析十分简便、廉价和快速，完全无需专业生物信息学知识和编程能力，结果易懂，因此普通人也能快速做出成果。

案例：高中生发表SCI论文

广东实验中学高三学生**郭杰明**，使用承启生物云平台进行分析，以**第一作者**在系统生物学著名期刊 *Molecular BioSystems* (影响因子3.2) 上发表论文。

广东实验中学官方报道链接：

<http://www.gdsyzyx.edu.cn/ifshow.php?id=1415593774>

**Molecular
BioSystems**



PAPER

Length-dependent translation initiation benefits the functional proteome of human cells†

Cite this: DOI: 10.1039/c4mb00462k

Jieming Guo,^{‡,ab} Xinlei Lian,^{‡,a} Jiayong Zhong,^a Tong Wang^{*,a} and Gong Zhang^{*,a}

We previously found that shorter mRNAs are preferably translated in various eukaryotic cells. However, the theoretical basis of this phenomenon is unclear. We hypothesize that shorter mRNA length correlates to the decreased translational error rate to reduce the energy consumption on defective protein degradation. In this study, we established a computational model to explain the length-dependent translation initiation efficiency. We provided mathematical evidence that this translational preference, rather than the protein degradation, is a major factor to shape the genome-wide length-dependent protein abundance. As deduced, we simulated that shorter mRNA length is a determinant of initiation circularization time. Furthermore, our model unveiled that preferentially translating shorter mRNAs benefits the energy efficiency on the proteome functionality. We proposed that cancer cells tend to hijack this evolutionary mechanism by counteracting the higher translational error rate. In conclusion, our model provides insights into the nature of the global length-dependent translational control and its biological significance.

Received 4th August 2014,
Accepted 8th October 2014

DOI: 10.1039/c4mb00462k

www.rsc.org/molecularbiosystems

www.chi-biotech.com
info@chi-biotech.com



承启生物
Chi Biotech

基因测序云分析解决方案
承载生命 启动未来

学界的广泛认可

FANSe系列算法公开发表后，已有多篇论文使用FANSe2算法和云平台分析策略，发表在专业权威期刊上。

藉由FANSe2算法极高的精度和实验可验证性，在这一算法的帮助之下，由中国科学家提出的翻译组测序分析 (Translatome sequencing) 已成为人类蛋白质组计划(HPP)的核心支柱之一，并于2014年3月被列为2014年人类蛋白质组计划的首要突出贡献。目前，FANSe2算法及云平台分析策略负责了2013-2014年世界上所有的翻译组测序的数据分析。

Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific

Nucleic Acids Res (2013) 41(9):4743-54

Resolving Chromosome-Centric Human Proteome with Translating mRNA Analysis: A Strategic Demonstration

Journal of proteome research (2014) 13 (1), 50-59

Systematic analysis of missing proteins provides clues to help define all of the protein-coding genes on human chromosome 1

Journal of proteome research (2014) 13 (1), 114-125

Chromosome-8-Coded Proteome of Chinese Chromosome Proteome Data Set (CCPD) 2.0 with Partial Immunohistochemical Verifications

Journal of proteome research (2014) 13 (1), 126-136

Omics Evidence: Single Nucleotide Variants Transmissions on Chromosome 20 in Liver Cancer Cell Lines

Journal of proteome research (2014) 13 (1), 200-211

Systematic analyses of the transcriptome, translatome, and proteome provide a global view and potential strategy for the C-HPP

Journal of proteome research (2014) 13 (1), 38-49

Iterative Genome Correction Largely Improves Proteomic Analysis of Nonmodel Organisms

Journal of proteome research (2014), 13 (6), 2724-2734

How to discover new proteins—translatome profiling

Science China Life Sciences (2014) 57 (3), 358-360

Length-dependent translation initiation benefits the functional proteome of human cells

Molecular BioSystems (2014), in press

还有更多的文章正在投向 *Nature Medicine, Blood, PLoS Genetics, eLife, ...*

领先全球的基因测序云分析平台

最可靠的定制化测序数据分析

独立评估验证已有的测序结果

专业定制的测序项目规划与设计

敬请咨询：info@chi-biotech.com



承启生物
Chi Biotech

承载生命 启动未来