



基因组、大数据与精准医学

Genome, Big Data and Precision Medicine

报告人：陈润生



中国科学院生物物理研究所

INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

How many characters are in the “Heaven Book”?

3×10^9 10,000 books

1 book 100 pages

1 page 3,000 characters

CCGGTCTCCCCGCCCCGCGCGCGAAGTAAAGGCCAGCGCAGCCCCGCGCTCCTGCCCTGGGGCCT
CGTCTTTCTCCAGGAAAACGTGGACCGCTCTCCGCCGACAGTCTCTTCCACAGACCCCTGTCGCCT
TCGCCCCCGGTCTCTTCCGGTTCTGTCTTTTCGCTGGCTCGATACGAACAAGGAAGTCGCCCCCA
GCGAGCCCCGGCTCCCCAGGCAGAGGCGGCCCCGGGGGCGGAGTCAACGGCGGAGGCACGC
CCTCTGTGAAAGGGCGGGGCATGCAAATTCGAAATGAAAGCCCCGGGAACGCCGAAGAAGCACG
GGTGTAAGATTTCCCTTTTCAAAGGCGGGAGAATAAGAAATCAGCCCGAGAGTGTAAGGGCGTCA
ATAGCGCTGTGGACGAGACAGAGGGAATGGGGCAAGGAGCGAGGCTGGGGCTCTCACCGCGAC
TTGAATGTGGATGAGAGTGGGACGGTGACGGCGGGCGCGAAGGCGAGCGCATCGCTTCTCGGCC
TTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTAATATCTGATACGTCCTCTATCCGAG
GACAATATATTAAATGGATTGATCAATCCGCTTCAGCCTCCCGAGTAGCTGGGACTACAGACGGT
GCCATCACGCCCAGCTCATTGTTGATTCCCGCCCCCTTGGTAGAGACGGGATTCCGCTATATTGCC
TGGGCTGGTGTGCAACTCATAGAACAAGGATCCTCCCTCCTGGGCCTGGGCGTGGGCTCGCAAA
ACGCTGGGATTCCCGGATTACAGGCGGGCGCACCCACACCAGGAGCAAACACTTCCGGTTTTAAA
AATTCAGTTTGTGATTGGCTGTCATTGAGTATTATGCTAATTAAGCATGCCCGGTTTTAAACCTCTTA
AAACAACTTTTAAATACCTTTCCACCTAAAACGTTAAATTTGTCAAGTGATAATATTCGACAAG
CTGTTATTGCCAAACTATTTTCTATTTGTTTCCTAATGGCATCGGAACTAGCGAAAGTTTCTCGCC
ATCAGTTAAAAGTTTGCGGCAGATGTAGACCTAGCAGAGGTGTGCGAGGAGGCCGTTAAGACTAT
ACTTTCAGGGATCATTCTATAGTGTGTTACTAGAGAAGTTTCTCTGAACGTGTAGAGCACCGAAA
ACCACGAGGAAGAGAGGTAGCGTTTTTCATCGGGTTACCTAAGTGCAGTGTCCCCCTGGCGCGCA
ATTGGGAACCCACACGCGGTGTAGAAATATATTTTAAGGGCGCG

(1250 characters)



中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS, CHINESE ACADEMY OF SCIENCES

Jan. 20, 2011



“I want the country that **eliminated polio and mapped the human genome** to lead a new era of medicine—one that delivers ***the right treatment at the right time***. In some patients with **cystic fibrosis**, this approach has reversed a disease once thought unstoppable. Tonight, I’m launching a new ***Precision Medicine*** Initiative to bring us ***closer to curing diseases*** like **cancer and diabetes**—and to ***give all of us access to the personalized information*** we need to keep ourselves and our families healthier.”

“Let’s Call it Precision Medicine!”



中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

一、精准医学的核心是什么？



中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

组学大数据与医学的结合



基因组

转录组

蛋白组

代谢组

表观遗传组.....



健康科学的发展:

转化医学

个体化医学

精准医学 精准诊断和精准治疗



Angelina Jolie

美国著名影星安吉丽娜·朱莉有家族性乳腺癌史，曾祖母、祖母和姨妈都是因乳腺癌去世。她非常担心自己会重蹈家庭悲剧，因此她去做了基因检测。结果，她的基因检测显示体内携带乳腺癌基因BRCA1突变，患乳腺癌的风险高达87%，在此情形下，她与医生商量后进行了预防性乳腺切除，把乳腺癌风险降低到了5%以下。



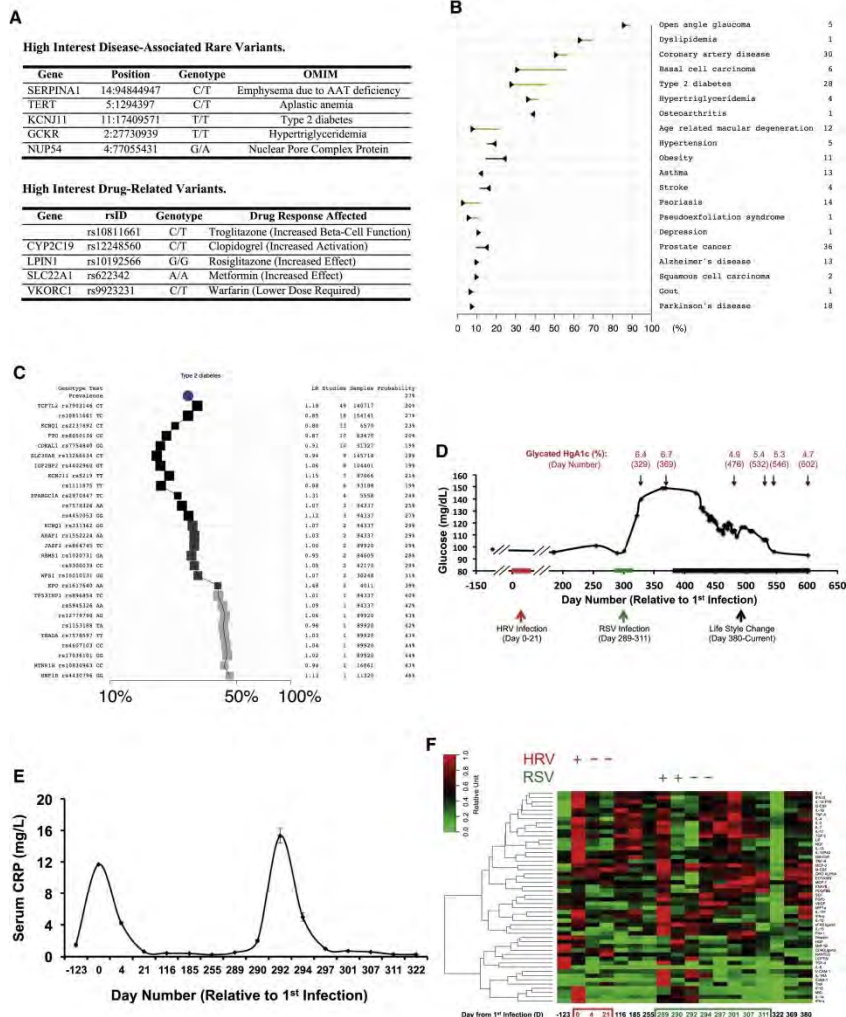
Michael Snyder

来自斯坦福大学医学院的Michael Snyder领导完成了一项历时两年半的跟踪实验，他定期做抽血检查，跟踪监测细胞内的4万余种不同的分子的起伏变化情况：从激素到血糖，到免疫系统蛋白和突变的基因，无所不包。他也见证了遗传基因上易患糖尿病的自己，不久之后真的罹患此病。

在这项研究中，Snyder对自己的基因组进行了测序。DNA检测表明，他罹患II型糖尿病的危险很高。虽然他的医生没有发现他正处于病情发展期的任何表征，他的自检还是查出了早期迹象。不久，他就患上了糖尿病。在确诊后，Snyder调节了饮食，加大了运动量进行减肥，控制住了病情。

Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes

Cell 148, 1293–1307, March 16, 2012





二、精准医学能使医学发生什么本质变化？



从诊断治疗到健康保障

基于精准医学理念的个体化治疗市场规模日益扩大，**2018** 年前全球市场规模将达到**2238** 亿美元



精准医学研究已成为新一轮国家科技竞争和引领国际发展潮流的战略制高点。

美国积极推动精准医学的发展，将其上升为国家战略；欧盟以精准医学理念指导其创新药物二期计划；英国开展十万人测序计划并成立精确癌症研究所；日本将精准医学相关内容列入科技创新计划中。这标志着国际上在基因资源利用、新药靶点发现、新的诊断治疗方法开发、生物医药新产品研发等的竞争进入新的阶段，对我国生物医药与健康产业的发展形成严峻挑战。



产业化前景

精准医学的发展将带动相关产业的快速发展，孕育巨大市场空间

首先，**生物样本和数据**本身就具有市场价值。据统计，2015 年生物样本库市值将超过22.5 亿美元。至2018 年生物大数据的市场总额将增长至76 亿美元，年复合增长率达到71.6%。生物数据的商业价值已经初步体现，2015 年1 月，罗氏制药子公司Genentech 向美国基因测序公司23andMe 注资 6000 万美金，用于共享23andMe 收集的帕金森病患者的基因数据，并基于数据信息研发帕金森病的治疗方案。



基因测序是精准医疗产业的重要组成部分。来自BBC research的数据显示，全球基因测序市场总量从2007年的794.1 万美元增长至2013年的45 亿美元，预计未来几年全球市场仍将继续保持快速增长，2018 年达到117 亿美元。

分子诊断是精准医疗的另一重要子行业，已经成为生物医药行业新热点，据Markets and Markets 公司估测，2018 年的全球市场市值将达到79 亿美元，2013-2018 年间的复合年增长率为9.7%。

基于精准医学理念的个体化治疗市场规模日益扩大，2018 年前全球市场规模将达到2238 亿美元。美国十大商业保险公司已将50 余项疾病个体化诊疗分子检测项目列入医疗保险。巨大市场空间吸引众多医药公司开展研发，目前已有多种个体化诊疗产品上市。截至2013 年，美国FDA 已批准了100 多种个体化药物，重点关注癌症和慢性疾病的个性化治疗。



Illumina推出NovaSeq测序仪 检测成本降低至100美元



在今天的摩根大通**医疗**健康年会上，Illumina公司宣布推出NovaSeq系列测序仪。在新闻通稿中，Illumina指出这个全新的测序框架将重新定义高通量测序。其无可比拟的通量、简捷的操作、低成本以及灵活性有望将基因组测序的成本降低至100美元。

NovaSeq是Illumina至今推出的最为强大的测序仪。设计它的科学家希望它能够满足研究人员对下一代测序技术的需求，并且降低大规模测序项目的难度。它一次运作能够对3-48人的全基因组进行测序，并为比对肿瘤-健康组织的超级深度测序（ultra-deep sequencing）、与复杂疾病相关的大规模基因变异分析等领域提供全新的市场。



除了NovaSeq系统，illumina联手两大科技巨头又出两招

近日，测序巨头Illumina宣布了与皇家飞利浦、IBM的合作协议，旨在共同推进癌症基因组数据的分析和解释。从合作的内容来看，双方的最终目的是“快”——快速解读数据，帮助研究人员快速实现肿瘤精准医疗。

Illumina联合飞利浦提供整合的肿瘤学解决方案

科研人员可以使用BaseSpace Sequence Hub（Illumina的NGS数据管理和分析的云基因组计算环境）获取Illumina测序仪产出的基因组数据，并通过飞利浦的IntelliSpace Genomics肿瘤学解决方案进行处理。该解决方案结合了多个来源的数据，如放射学、免疫组织化学、数字病理学、医疗记录和实验室测试。采用该解决方案的实验室将在一个视图中获得高级分析、深度学习技术、文献、指南和其他证据。

Illumina联手IBM解读癌症基因谱

这两家公司将应用部署Watson for Genomics工具，帮助研究人员解读Illumina的实体肿瘤谱分析结果。科研人员将把Watsonfor Genomics整合到Illumina公司的BaseSpace和肿瘤测序过程中，于是这将允许科学家们使用Illumina的实体肿瘤分析产品TruSight Tumor 170检测多达170个基因的变异，更快速和全面地诠释相关数据。

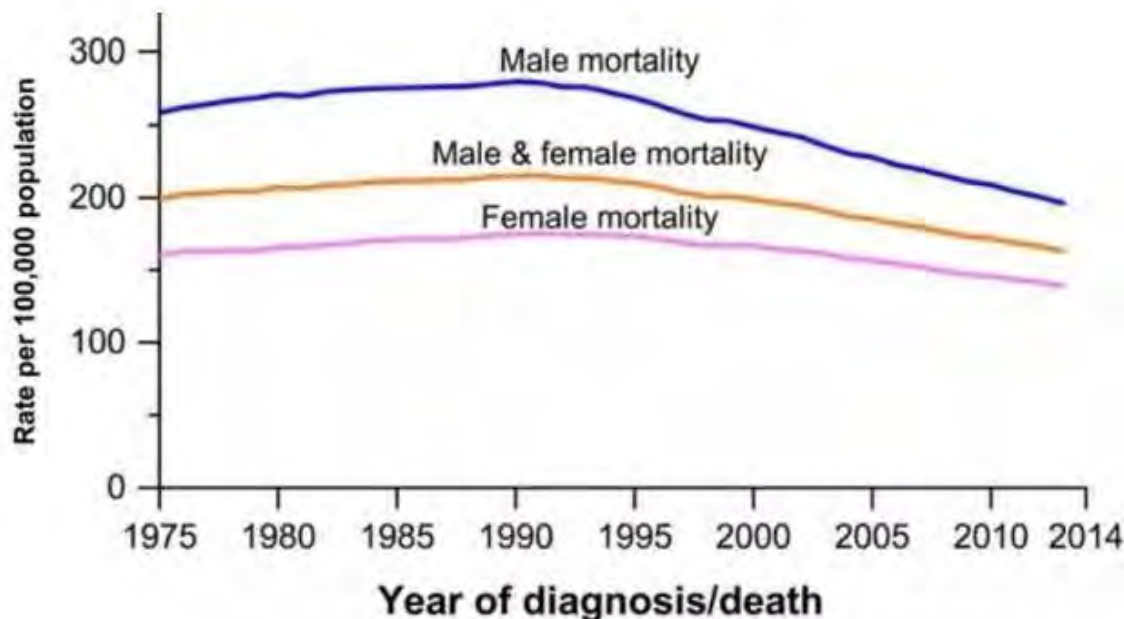


美国癌症协会发布2017癌症年度统计报告



2017年1月5日，美国癌症协会在CA: A Cancer Journal for Clinicians 杂志（影响因子138）上发表2017年癌症统计，对癌症的发病率，死亡率和生存率做了全面的最新年度报告，并出版了纸质版2017年癌症事实和数据。统计表明，美国整体癌症死亡率经过二十多年来的稳步下降，终于使这一数据下降了25%，意味着1991年至2014年期间癌症死亡人数减少210万。

报告估计，在2017年，美国将有1,688,780例新增癌症病例以及600,920例癌症死亡病例。过去十年的可用数据显示，女性的总体癌症发病率是稳定的，男性每年下降约2%，而男性和女性的癌症死亡率每年下降约1.5%。





三、实现精准医学要有哪些基础？



1、 组学: 基因组, 转录组, 蛋白质组, 代谢组

.....

大数据: 人群、队列

是当前国际两大科学前沿的交汇

2、 基因型与表型的关联

建立表型组数据库

生物信息学 生物网络 系统生物学

在此基础上融合临床检验、影像学等指标



中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

四、精准医学才刚刚上路！！



中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

基因组、转录组中的挑战—基因组 中的暗信息



How many characters are in the “Heaven Book”?

3×10^9 10,000 books

1 book 100 pages

1 page 3,000 characters

CCGGTCTCCCCGCCCCGCGCGCGAAGTAAAGGCCAGCGCAGCCCCGCGCTCCTGCCCTGGGGCCT
CGTCTTTCTCCAGGAAAACGTGGACCGCTCTCCGCCGACAGTCTCTTCCACAGACCCCTGTCGCCT
TCGCCCCCGGTCTCTTCCGGTTCTGTCTTTTCGCTGGCTCGATACGAACAAGGAAGTCGCCCCCA
GCGAGCCCCGGCTCCCCAGGCAGAGGCGGGCCCGGGGGCGGAGTCAACGGCGGAGGCACGC
CCTCTGTGAAAGGGCGGGGCATGCAAATTCGAAATGAAAGCCCCGGGAACGCCGAAGAAGCACG
GGTGTAAGATTTCCCTTTTCAAAGGCGGGAGAATAAGAAATCAGCCCGAGAGTGTAAGGGCGTCA
ATAGCGCTGTGGACGAGACAGAGGGAATGGGGCAAGGAGCGAGGCTGGGGCTCTCACCGCGAC
TTGAATGTGGATGAGAGTGGGACGGTGACGGCGGGCGCGAAGGCGAGCGCATCGCTTCTCGGCC
TTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTAATATCTGATACGTCCTCTATCCGAG
GACAATATATTAAATGGATTGATCAATCCGCTTCAGCCTCCCGAGTAGCTGGGACTACAGACGGT
GCCATCACGCCCAGCTCATTGTTGATTCCCGCCCCCTTGGTAGAGACGGGATTCCGCTATATTGCC
TGGGCTGGTGTGCAACTCATAGAACAAAGGATCCTCCCTCCTGGGCCTGGGCGTGGGCTCGCAAA
ACGCTGGGATTCCCGGATTACAGGCGGGCGCACCCACACCAGGAGCAAACACTTCCGGTTTTAAA
AATTCAGTTTGTGATTGGCTGTCATTGAGTATTATGCTAATTAAGCATGCCCGGTTTTAAACCTCTTA
AAACAACTTTTAAATACCTTTCCACCTAAAACGTTAAATTTGTCAAGTGATAATATTCGACAAG
CTGTTATTGCCAAACTATTTTCTATTTGTTTCCTAATGGCATCGGAACTAGCGAAAGTTTCTCGCC
ATCAGTTAAAAGTTTGCGGCAGATGTAGACCTAGCAGAGGTGTGCGAGGAGGCCGTTAAGACTAT
ACTTTCAGGGATCATTCTATAGTGTGTTACTAGAGAAGTTTCTCTGAACGTGTAGAGCACCGAAA
ACCACGAGGAAGAGAGGTAGCGTTTTTCATCGGGTTACCTAAGTGCAGTGTCCCCCTGGCGCGCA
ATTGGGAACCCACACGCGGTGTAGAAATATATTTTAAGGGCGCG

(1250 characters)



The 17 December 2010 issue of *Science* includes special sections highlighting the **Breakthrough of the Year** and **Insights of the Decade**.

Insights of the Decade

Shining a Light on the Genome's 'Dark Matter'

The scope of this “dark genome” became apparent in 2001, when the human genome was first published. Scientists expected to find as many as 100,000 genes packed into the 3 billion bases of human DNA; they were startled to learn that there were fewer than 35,000. (The current count is 21,000.) Protein-coding regions accounted for just 1.5% of the genome. Could the rest of our DNA really just be junk?





中國科學院生物物理研究所

INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

Some examples about noncoding RNAs and diseases



Elevated expression of **PCGEM1**, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients

Gyorgy Petrovics*,¹, Wei Zhang², Mazen Makarem¹, Jesse P Street¹, Roger Connelly¹, Leon Sun¹, Isabell A Sesterhenn², Vasantha Srikantan¹, Judd W Mou^{1,3} and Shiv Srivastava¹

Oncogene (2004) 23, 605–611

PCGEM1 is a novel, highly prostate tissue-specific, androgen-regulated gene. Here, we demonstrate that PCGEM1 expression is significantly higher in prostate cancer (CaP) cells of African-American men than in Caucasian-American men.

PCGEM1 appears to be a noncoding functional RNA gene (Srikantan et al., 2000).



***His-1*: A noncoding RNA implicated in mouse leukemogenesis**

Fan Xu, Molly McFarland and David S. Askew*

Histol. Histopathol., 1999, 14, 235–241.

The *His-1* gene is highly conserved among vertebrate species and is transcribed as a single spliced and polyadenylated cytoplasmic RNA that shares several features in common with the emerging class of untranslated RNAs. A role for the *His-1* gene in neoplastic transformation was first indicated by the identification of transcriptionally activated *His-1* genes in a series of mouse leukemias, and more recent studies with antisense *His-1* RNAs suggest that *His-1* is involved in an oncogenic pathway that controls cell cycle progression.



中國科學院生物物理研究所

INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

MALAT-1, a novel noncoding RNA, and thymosin b4 predict metastasis and survival in early-stage non-small cell lung cancer

Ping Ji^{1,5}, Sven Diederichs^{1,5}, Wenbing Wang¹, Sebastian Boßing¹, Ralf Metzger², Paul M Schneider², Nicola Tidow³, Burkhard Brandt³, Horst Buerger⁴, Etmar Bulk¹, Michael Thomas¹, Wolfgang E Berdel¹, Hubert Serve^{*,1} and Carsten Müller-Tidow^{*,1}

Oncogene (2003) 22, 8031–8041



中国科学院生物物理研究所

INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

Evidence for evolutionarily conserved secondary structure in the *H19* tumor suppressor RNA

Veronica Juan, Chad Crain and Charles Wilson

Nucleic Acids Research 2000 Vol. 28 1221-1227



中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

五、无限的创新机遇！！



中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

非编码RNA的功能



基于表达的临床lncRNA signature, *Gut* 2014

Downloaded from gut.bmj.com on March 10, 2014 - Published by gut.bmj.com

Gut Online First, published on February 12, 2014 as 10.1136/gutjnl-2013-305806

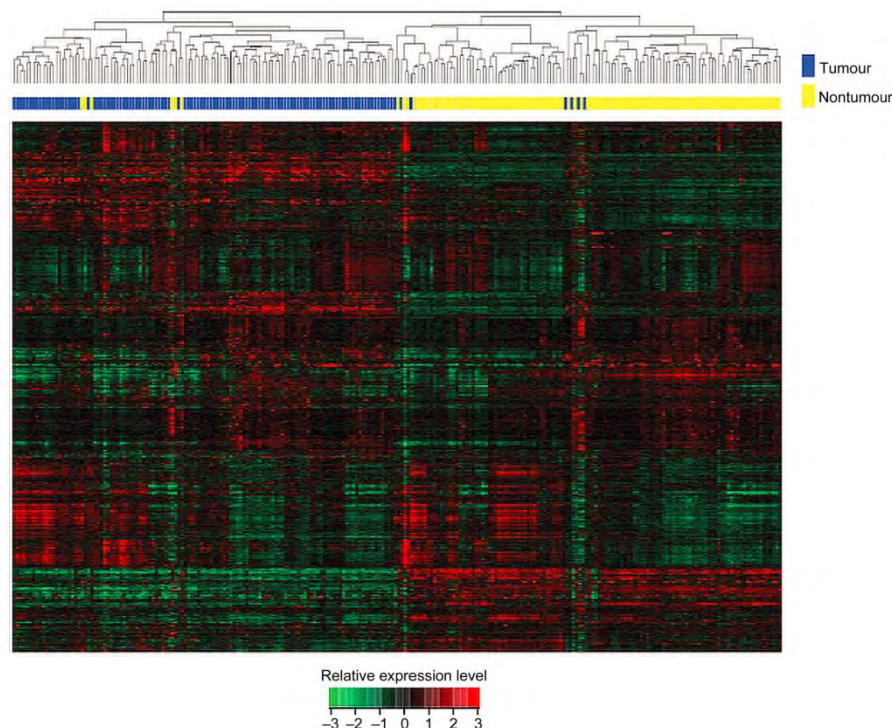
Upper GI cancer

ORIGINAL ARTICLE

lncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma

Jiagen Li,¹ Zhaoli Chen,¹ Liqing Tian,² Cheng Yibo Gao,¹ Suyu Wang,¹ Fang Zhou,¹ Sushen Ziyuan Liu,¹ Geir Skogerboe,² Jingsi Dong,¹ R Baihua Zhang,¹ Yue Yu,¹ Xuejiao Shi,¹ Mei Li Fengwei Tan,¹ Runsheng Chen,² Jie He¹

IBP 陈润生课题组 &
CIH, CAMS 赫捷课题组



normalised expression data of the 6389 lncRNAs with coefficient of variance >0.10 was used for clustering analysis



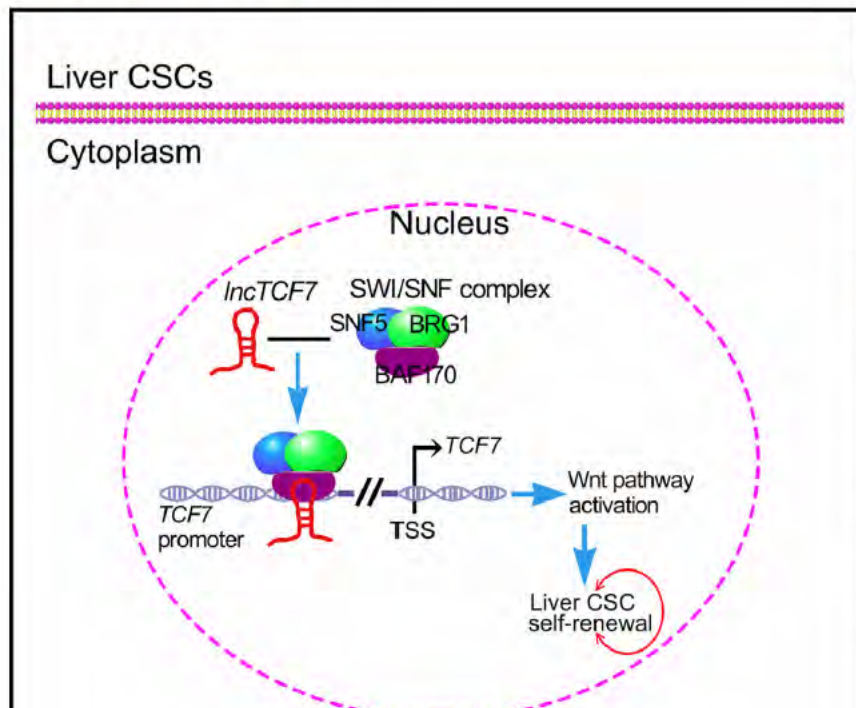
肿瘤干细胞中lncRNA的功能机制研究-lncTCF7

Article

Cell Stem Cell

The Long Noncoding RNA *lncTCF7* Promotes Self-Renewal of Human Liver Cancer Stem Cells through Activation of Wnt Signaling

Graphical Abstract



Authors

Yanying Wang, Lei He, ...,
Runsheng Chen, Zusen Fan

Correspondence

crs@sun5.ibp.ac.cn (R.C.),
fanz@moon.ibp.ac.cn (Z.F.)

In Brief

Wang et al. have identified a long noncoding RNA, *lncTCF7*, that activates Wnt signaling to promote liver cancer stem cell self-renewal and tumor propagation. Targeting this pathway could help address the high recurrence and heterogeneity of liver cancer.



CrossMark

Wang et al., 2015, Cell Stem Cell 16, 413–425
April 2, 2015 ©2015 Elsevier Inc.
<http://dx.doi.org/10.1016/j.stem.2015.03.003>

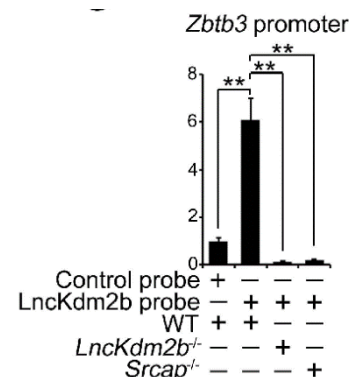
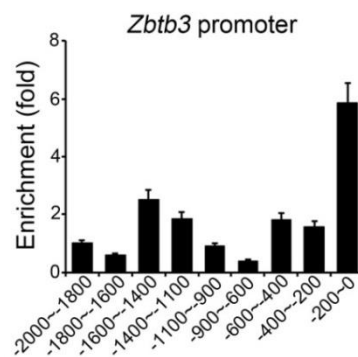
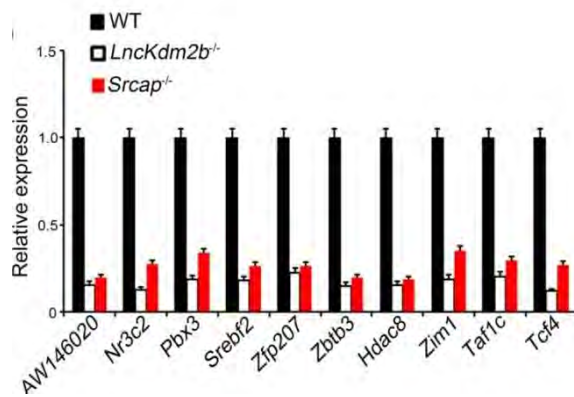
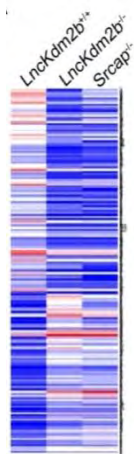
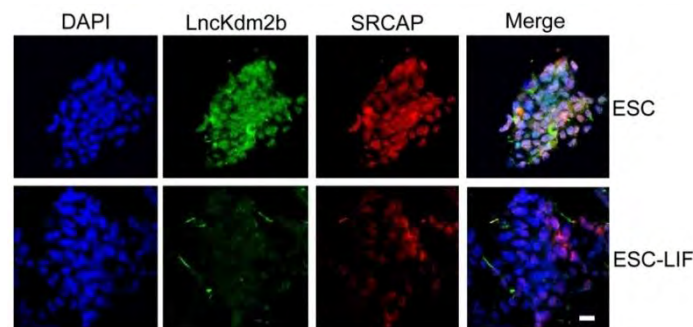
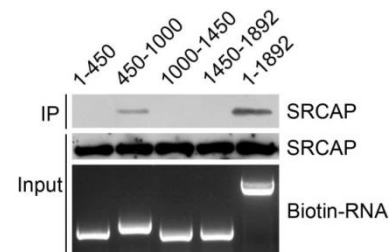
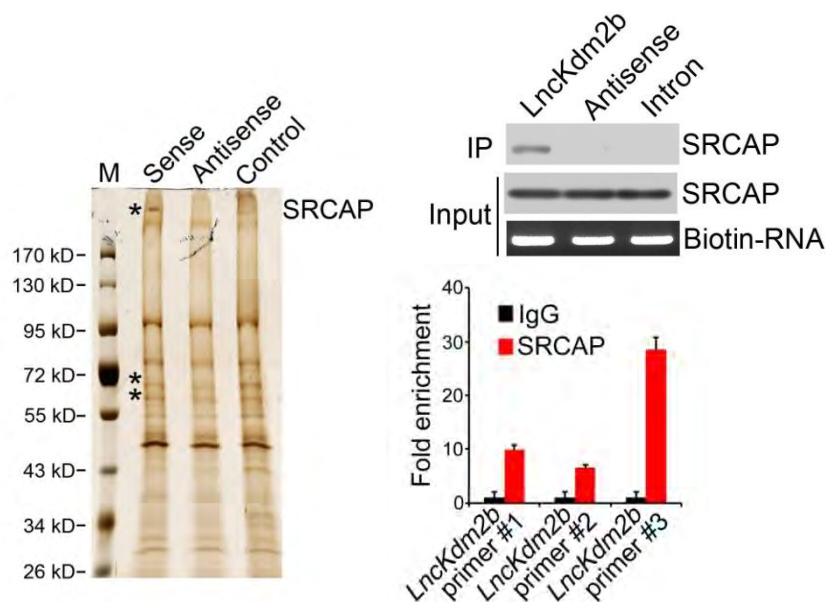
CellPress



长非编码核酸lncKdm2b调控淋巴细胞机制研究

Nature Immunology 2017

Long non-coding RNA *lncKdm2b* is required for the maintenance of group 3 innate lymphoid cells by initiating *Zfp292* expression





有多少非编码基因还没被发现？

几年前，日本的遗传研究所（RIKEN）在小鼠中获得了约181,000个全长的RNA转录本，其中编码蛋白质的转录本仅有约20,000个，其余约161,000个转录本全部归属于非编码RNA。



中国科学院生物物理研究所

INSTITUTE OF BIOPHYSICS, CHINESE ACADEMY OF SCIENCES



XINHUA/AFP



XINHUA/AFP



中国科学院生物物理研究所

INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

***Nature* 391, 806 - 811 (19 February 1998)**

Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*

ANDREW FIRE*, SIQUN XU*, MARY K. MONTGOMERY*,
STEVEN A. KOSTAS*†, SAMUEL E. DRIVER‡ & CRAIG C. MELLO‡

* Carnegie Institution of Washington, Department of Embryology, 115 West University Parkway, Baltimore, Maryland 21210, USA

† Biology Graduate Program, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA

‡ Program in Molecular Medicine, Department of Cell Biology, University of Massachusetts Cancer Center, Two Biotech Suite 213, 373 Plantation Street, Worcester, Massachusetts 01605, USA



疾病的诊断与治疗

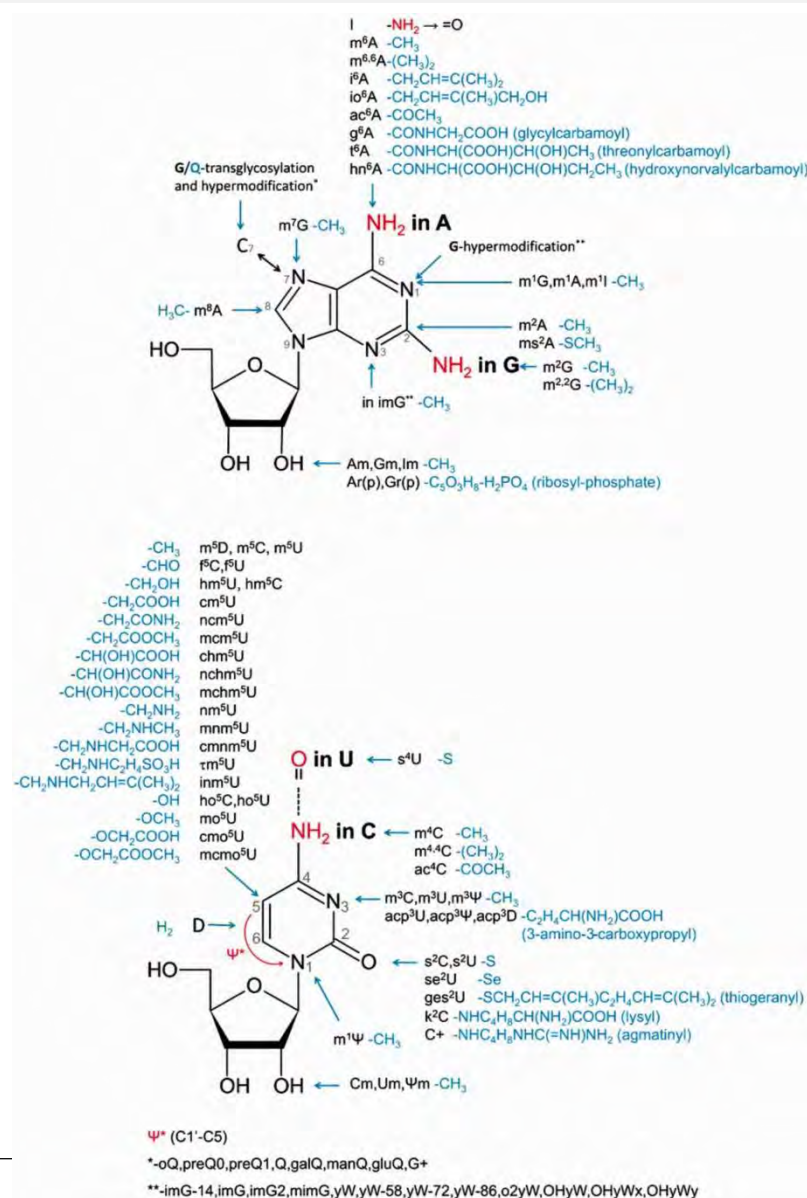
全新的药物设计与研发

动植物新品种、新性状的培育



A scenic landscape photograph of a calm lake. The water is clear and reflects the surrounding environment. In the foreground, a rocky shoreline with smooth, rounded stones in various shades of grey, brown, and green is visible. The right side of the lake is bordered by a dense forest of tall, green evergreen trees. In the background, majestic mountains with rugged peaks and some snow patches rise against a bright blue sky with wispy white clouds. The overall scene is peaceful and picturesque.

Thank you!



From: MODOMICS: a database of RNA modification pathways—2013 update

Nucleic Acids Res. 2012;41(D1):D262-D267. doi:10.1093/nar/gks1007

Nucleic Acids Res | © The Author(s) 2012. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial reuse, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.