

· 学科进展与展望 ·

国家自然科学基金与我国基因组学研究

谷瑞升 田艳艳 杜生明

(国家自然科学基金委员会, 北京 100085)

[摘要] 基因组学是20世纪末发展和建立起来的新兴学科,在我国发展十分迅速,研究成果令世人关注。本文简要回顾了我国基因组学研究的发展历程,分析了国家自然科学基金(简称科学基金)在推动基因组学发展中的作用,介绍了在科学基金资助下我国科学家获得的代表性学术成果,探讨了基因组学未来应该关注的方向和重视的科学问题。

[关键词] 基因组学发展,项目资助,代表性成果,国家自然科学基金

基因组学(Genomics)是20世纪末发展和建立起来的新兴学科,以测定和解释物种的基因组序列组成和结构为主要研究内容,揭示DNA序列中蕴含的影响生命活动的各类信息,是从最基础层面研究生命本质的学科之一。地球上现存物种无一例外,都具有各自的基因组。基因组作为信息载体,蕴含着个体发育、形态发生、生理功能,乃至接受外界影响后的表现遗传可塑性等各种信息。基因组学研究已经成为人们理解生命现象的基本规律、揭示动植物重要性状的形成机制、人类疾病发生机理的最主要途径之一,成为引领现代生命科学各个学科发展的引擎,也为现代生物技术的发展提供理论基石。

1 基因组学发展历程

基因组学最初是由美国科学家 Thomas Roderick 于1986年提出的,当时其内容是指基因组作图(mapping)和测序(sequencing)。后来随着人类基因组计划的进展,其含义不断发展和更新。之后,科学家们将它定义为研究生物基因组的结构和功能的科学,即从整体水平上来研究一个物种的基因组的结构、功能及调控的一门科学^[1]。

长期以来,由于测定物种全基因组序列价格昂贵和周期长,科学家不得不局限于少数几个模式物种的基因组分析。单纯基于几个少数基因组数据,实际上很难全面理解如同天书的碱基排列顺序所蕴

含信息的微言大义。近年来,新一代测序技术的发展将汇聚成日新月异的基因组学数据的滚滚洪流,使得任何一个物种的基因组数据测定正在变得简便易行,在近缘物种中进行深入仔细的比较将能在前所未有的广度和深度上理解各种表型和性状的遗传基础^[2]。据统计,截止到2010年4月,已经有近4000个物种完成全基因组测序。

2 我国基因组学的研究发展及现状

中国的基因组学研究起步于上世纪80年代末^[3],经过20多年的快速发展,已具有相当的规模和拥有一支优秀的科研队伍,在国际同类研究中已占有一席之地^[4]。在基因组序列解析能力、全基因组序列组装、功能基因定位和克隆以及比较基因组与进化等领域处于世界先进水平,在蛋白组学、表观基因组学以及基因组生物信息学等方面也获得了突出的学术成果^[5]。

在有关基因组学论文方面,据粗略统计,在2005年1月—2010年7月的5年多时间里,我国科学家作为通讯作者在 *Science*, *Nature*, *Nat Genet*, *Nat Med*, *Am J Hum Genet* 等本领域5个国际顶级学术刊物发表论文25篇,其中部分文章作为 *Nature* 封面文章发表,在影响因子前10位的遗传学刊物上发表学术论文60余篇。在5个国际顶级学术刊物发表论文的信息见表1。

本文于2010年11月17日收到。

表1 2005年1月—2010年7月我国科学家在国际顶级学术刊物发表论文数量统计

刊名	2005	2006	2007	2008	2009	2010	论文合计
<i>Am J Hum Genet</i>	1	1	0	3	3	1	9
<i>Nat Genet</i>	0	1	0	1	2	1	5
<i>Nat Med</i>	1	0	0	0	0	0	1
<i>Nature</i>	0	0	0	2	1	2	5
<i>Science</i>	2	1	0	0	1	1	5

在获奖成果方面,据初步统计,近年来在基因组领域获得国家自然科学奖11项,国家科学技术进步奖14项。

我国科学家继参与完成1%的人类基因组测序以来,承担并完成了10%人类HapMap图的绘制,独立开展了“炎黄一号”计划,相继完成了水稻、鸡、黄瓜、家蚕、血吸虫、大熊猫、鲟米鱼等多个大型基因组的测序。“炎黄一号”计划的完成标志着我国基因组学研究进入个体基因组测序的时代。我国水稻全基因组序列图谱绘制工作起步晚于日本,但完成水稻(籼稻)基因组的工作框架序列图却早于日本。2002年水稻全基因草图绘制,标志着我国成为在世界上继美国之后第二个具备独立完成大规模全基因组测序和组装分析能力的国家。中国科学家在基因组学序列解析上站在世界科学的前沿。

在功能基因组研究方面,我国科学家利用图位克隆方法发现了一批控制重要农艺性状和人类遗传疾病的关键功能基因,获得了许多原创性的成果。如率先利用高频耳聋家系克隆了遗传性神经性耳聋致病基因*GJB3*,实现了在我国本土克隆遗传病致病基因零的突破。之后我国科学家相继定位和克隆了遗传性本质乳光牙致病*DSPP*基因、A-1型短指(趾)症致病基因、红斑肢痛症致病*SCN9A*基因、先天性白内障致病*CRYGS*基因、遗传性毛发上皮瘤致病基因*CYLD*、斑秃基因、水稻分蘖形成有关的*MOCI*基因、控制水稻株型、穗粒、产量基因以及水稻细胞质雄性不育基因和恢复基因、植物抗盐、抗虫等等多个重要基因,在单基因遗传病致病基因研究上打破了西方科学界一枝独秀的局面,在多基因复杂性状的研究中取得了令人瞩目的成绩。在*Nature*,*Nat Genet*,*Nat Cell Biol*,*Am J Hum Genet*等本领域国际顶级学术杂志发表了重要的学术成果,产生了重要国际影响,为我国赢得了荣誉。

在比较基因组与进化研究方面,我国科学家在家养动物的人工驯化、生物遗传突变机制、动植物的适应性基因组进化、中国人群起源和演化等方面都取得了重要的进展。我国科学家根据家蚕和野桑蚕

的大规模基因组测序数据,成功获得驯化对家蚕生物学影响的基因组印记,分析出354个蛋白编码基因在家蚕驯化过程中起到重要作用。利用生物信息学技术,我国科学家对人和黑猩猩、恒河猴、小鼠、果蝇、水稻、酿酒酵母等不同类别生物的基因组序列进行比对分析,提出和验证了“Indel诱导自发突变机制假说”。通过大量基因组数据及比较研究,我国科学家在南极鱼适应极端寒冷环境的分子机制、蝙蝠视觉进化的新机制、SARS冠状病毒进化、植物F-box基因序列分化的多种模式等研究中均取得了可喜的进展。另外,我国科学家在稻属的比较基因组学和基因组进化以及葫芦科基因组着丝粒进化研究,在中国人群及周边民族人群的源流和人群的迁移与分化规律研究,在中国人群起源(例如藏族人群起源)等多个方面也取得突出的研究成果。

在基因组学研究的带动下我国的蛋白质组学、表观遗传学以及基因组生物信息学等研究也取得了重要学术成果。我国科学家建立了人胎肝转录组及其蛋白质组,筛选得到了一种可以选择性地干扰抑癌基因的重要新型蛋白质;鉴定和发现多个新表观遗传控制基因,如发现染色质组蛋白H3K79甲基化的调控机制和在白血病发生中的作用,发现*WOR1*的双稳态表达主控白念珠菌的白-灰形态转换。在基因组生物信息学领域,我国的进步十分明显,涌现出了一批优秀人才和创新团队,表现出了良好的发展态势。我国科学家提出并建立了密码学方法;提出将图论中谱分析用于确定蛋白与蛋白相互作用网络的拓扑结构等多种新模型;自主研制和开发了KOBAS和GOEAST等多个基因组数据分析工具;构建成瘾分子通路等多个基因表达调控网络模型;利用贝叶斯网络推测组蛋白各种不同修饰和基因表达之间的因果关系及组合关系;破译处理了复杂的“组蛋白编码”;通过复杂网络挖掘基因表达规律和基因间的表达调控关系。这些工作为推动生物信息学研究发展做出了重要贡献。

3 国家自然科学基金在推动我国基因组学发展中的作用

我国基因组学的快速发展和壮大与国家的重视和支持密不可分。国家自然科学基金委员会(以下简称自然科学基金委)是第一个启动并支持我国基因组学开展系统研究的国家机构。科学基金良好的运行机制和持续固定的经费投入,保证了我国基因

组学研究不断发展壮大,培养和造就出一批高层次的人才队伍,促进了研究氛围和科研环境的不断改善,在推动我国基因组学发展方面发挥了重要作用。随着基因组学发展,特别是其显现出的科学价值和潜在的巨大经济价值,政府其他部门也开始对基因组学研究进行支持。政府与专家形成了共识,抓住时机,加大投入,迎头赶上,发展我国的基因组学,由此大大促进了我国基因组学研究快速发展。

3.1 科学基金在启动支持我国基因组学研究上发挥了重要作用

自然科学基金委在成立的第1年即1986年,也就是国际人类基因组计划实施初期,就利用其有限的经费资源率先支持中国科学家开展基因组学研究,在1986—1990年5年时间里先后立项支持了19个面上项目,奠定了中国基因组学发展的基础。1992年根据国际科研发展态势并结合我国国情,在基因组研究领域设立了第一个重点项目“人类基因组区段作图”,资助经费76万元,积极推动中国科学家参与国际人类基因组计划。1993年,又以重大项目形式在国内启动了我国的人类基因组计划,并根据我国国情和丰富的遗传资源,战略性地将研究定位在功能基因组相关基础研究,开展“中华民族基因组中若干位点基因结构的研究”。1995年之后,又以重大项目的形式延续资助中国人类基因组计划研究,资助金额达1000万元,成为当时科学基金资助经费最多的重大项目。国家自然科学基金对我国人类基因组学研究资助时间之早、强度之大在当时是首屈一指的。

3.2 在科学基金前期资助的基础上,科技部等国家其他部门加大投入,极大地推动了我国基因组研究的快速发展

在科学基金前期支持和推动下,我国的人类基因组计划进展顺利,在引进和建立基因组研究平台、在多民族基因组保存与分析以及若干位点疾病相关基因研究等多方面取得了重要进展,基因组学研究基础不断加强,人才队伍不断壮大,加上国际人类基因组计划和其他基因组学研究凸显的巨大潜在科学和商业价值,科技部和中国科学院等国家其他部门对基因组学研究也加大了投入。

在科学基金前期支持的基础上,1996年国家科技部设立了“重大疾病相关基因的定位、克隆、结构与功能研究”的“863”国家高技术研究发展计划项目

和“人类疾病基因组学”的“973”国家重点基础研究发展计划项目。在完成人类基因组“序列草图”的最后阶段,国家科技部和中国科学院分别投入了3000和1000万元开展1%测序任务,经费的投入保证了我国科学家按时向国际人类基因组数据库提交了全部数据,受到国际同行的一致肯定和尊重。在完成人类基因组草图之后,国际上又启动了人类Hap-Map计划,科技部和中国科学院又给予了大力支持。科技部通过“973”计划连续支持了2个项目“疾病基因组学理论和技术体系的建立”和“多基因复杂形状疾病的疾病基因组学”。之后包括国家科技部、国家发展与改革委员会(以下简称国家发改委)、自然科学基金委、中国科学院以及地方政府都加入对基因组计划研究投入。例如,2003年科技部实施了国家重大科技专项“功能基因组和生物芯片”(国家高技术研究发展计划“863”计划);2006年,深圳市政府、中国科学院、国家发改委、国家科技部、自然科学基金委等单位共同资助开展了全球第一例中国人标准基因组序列图谱绘制研究,该计划命名为“炎黄计划”。国家各部门共同参与,合力支持,极大地推动了我国基因组研究的发展,使我国基因组学研究步入了时代发展的快车道。

3.3 科学基金为我国培养了一支优秀基因组学研究队伍

人才培养是科学基金工作的重要组成部分。科学基金历来把人才培养与支持项目放在同等重要的位置,以项目推动人才培养,以人才促进学科发展,实现两者和谐和共同进步。据不完全统计,至2010年底,科学基金累计直接资助基因组学人才培养项目197项,其中资助青年基金项目162项、国家杰出青年科学基金项目(含外籍)24项,海外青年学者合作研究基金2项,资助创新研究群体9个。青年科学基金成了许多优秀科学家从事基因组学研究的“第一桶金”,为他们发展和成长提供了一次难得机会。一大批年轻优秀科研工作者获得了国家杰出青年科学基金项目资助,为他们在基因组研究领域展示才华和施展抱负提供了良好的平台。以国家杰出青年科学基金获得者为主体的创新团队如人类基因组创新研究群体、分子与进化基因组学创新研究群体、畜禽基因组学创新研究群体、疾病基因组学创新研究群体、人胎肝蛋白质组学创新研究群体、农作物基因组学创新研究群体等,正活跃在我国基础研究第一线,成为我国参与国际竞争和冲击前沿领域的一

支重要力量。

除了直接的人才项目,科学基金在基因组学领域支持了454个面上项目、78个重点项目、38个重大项目(专题),63个重大研究计划项目,也都直接和间接地在人才培养中发挥了重要作用。科学基金全方位多角度支持,为我国基因组研究夯实了全面的人才基础。

3.4 科学基金资助奠定了我国基因组研究快速发展的技术基础

中国的基因组学技术是在不断向美欧等发达国家学习、引进、创新中成长起来的。在这一过程中,科学基金资助发挥了重要的作用。科学基金在基因组学研究初期就十分重视技术发展,1987年就立项资助“人类基因组中基因定位研究”,对染色体高分辨带型分析、染色体原位分析、染色体显微切、PCR扩增、构建区带特异性DNA文库构建等基因组初期分析技术给予大力支持。1993年实施的重大项目“中华民族基因组中若干位点基因结构的研究”将建立和改进人类基因组研究中的新技术列为3个资助方向之一。1995年启动的重大项目又将“发展功能基因组学的新理论、新技术和新方法”作为3个重要研究方向之一。随后基因组大规模测序技术上又经历了从平板电泳、毛细管电泳、单分子测序,走向多重平台技术整合的技术创新;大规模基因型分析理论与技术经历了从微卫星基因组分型、大规模SNP分型、全基因组分型,走向面对临床检测和大规模队列的研究;功能基因组研究的理论与技术也从分子到了系统;生物信息技术研发和服务支撑了基因组学的发展。在上述技术发展过程中都能找到科学基金资助的痕迹。据初步统计,截止到2010年,科学基金支持基因组技术发展的项目达

190项,其中面上项目68项,青年项目44项,重点项目8项,重大研究计划项目20项,重大项目(专题)9项,专项基金10项,国际合作项目21项,其他类项目10项,资助经费达7625万元。目前我国已经建立起了高效和完善的新一代测序技术,拥有独创的序列拼接技术,掌握先进的基因定位和克隆技术,我国在基因研究技术层面已经达到世界水准。

3.5 科学基金作为资助我国基因组学基础研究主渠道之一,在推动学科全面均衡快速发展上发挥重要作用

短短的20年时间,我国基因组学研究经历了从无到有,从弱变强的发展历程。科学基金在国内最先系统资助基因组学研究,在推动学科全面均衡快速发展上发挥重要作用。

在过去的20多年中,随着国家对自然科学基金的投入不断增长,基因组学资助规模不断扩大,资助强度不断提高。科学基金资助基因组学的项目数从1986年5项增加到2010年171项,资助经费由1986年的19.5万元增长到2010年的6769万元。面上项目平均资助强度由1986年3.9万元提高到2010年34.1万元,重点项目资助强度由1996年开始时的90万元提高到2010年的200万元。到2010年,科学基金累计资助基因组学研究项目1025余项,资助经费达56549多万元。在资助的各类项目中,面上项目数最多占44.3%,青年基金项目数占16.3%,国际合作项目数占10.4%,重点项目数占7.6%(图1)^[6]。科学基金在夯实学科基础,促进基因组学全面和均衡发展发挥了重要作用。

基因组学获资助的项目数和经费年度变化见图1、图2和表2。

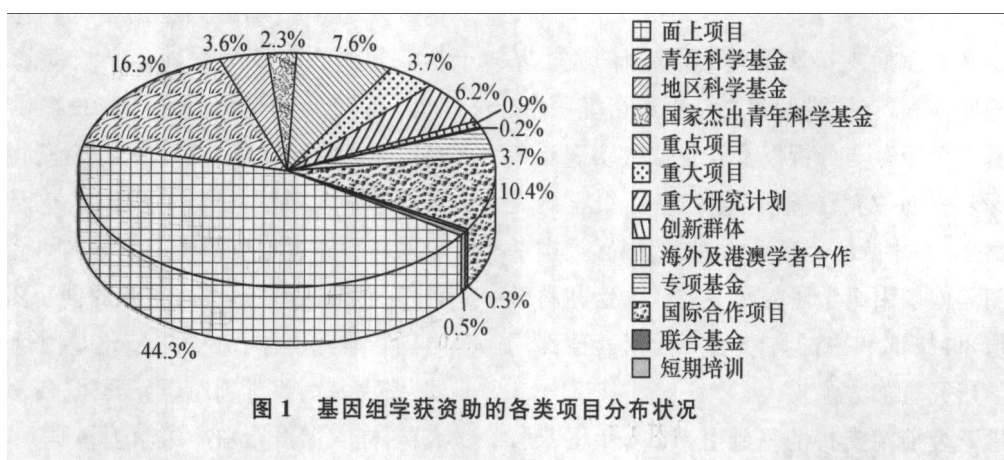


图1 基因组学获资助的各类项目分布状况

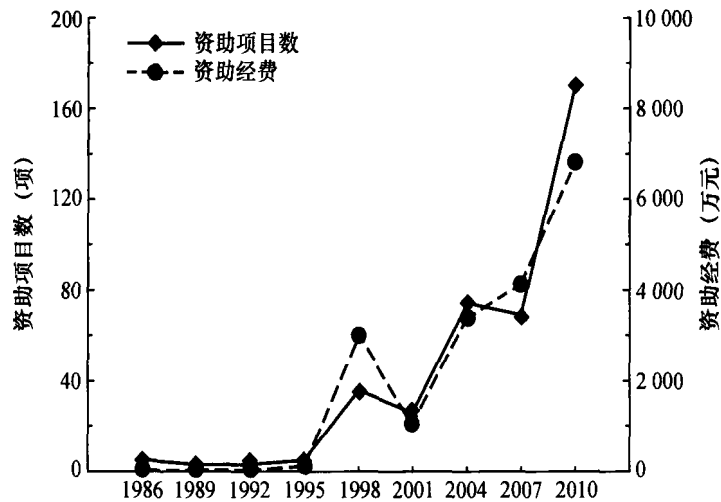


图2 基因组学获资助项目数和资助经费年度变化

表2 基因组学资助项目及经费汇总(1986—2010)

项目类型	项目数(项)	资助金额(万元)
面上项目	454	10 709.1
青年科学基金	167	3 109.5
地区科学基金	37	821.5
国家杰出青年科学基金	24	3 920
重点项目	78	11 604
重大项目	38	12 079.4
重大研究计划	63	3 944
创新群体	9	7 010
海外及港澳学者合作研究基金	2	40
专项基金	38	721
国际合作项目	107	2 051.6
联合基金	3	525
短期培训	5	14
总计	1 025	56 549.1

在我国科研资助体系中,除了科学基金,没有其他项目对基因组学研究进行了长达20多年的持续不断资助,投入了56 549余万元的经费支持。科学基金在促进基因组学发展中真正发挥了主渠道之一的资助作用。

科学基金资助推动了基因组学研究全面均衡发展。科学基金资助特点是科学家自选命题,自由探索。科学基金资助的基因组学研究涉及理论和技术创新,涉及基因组结构与分析、比较基因组与进化、基因组信息学、转录组学、蛋白组学、RNA组学、表观基因组学等基因组学相关的方方面面。科学基金在关注热点领域的同时,也十分注重一些非热点领域的发展,推动学科全面均衡发展。

3.6 基因组学已成为带动我国生命科学发展的重要引擎

在国家自然科学基金资助体系中,当今的基因组学研究已不是仅仅局限于少数几个项目或单一学科,其内涵不断丰富,外延不断扩大。基因组学本身

的发展衍生出一批新型学科和领域如转录组学、蛋白质组学、表观遗传组学和RNA组学等等。基因组学与其他学科交叉催生的一些新的学科,如药物基因组学(pharmacogenomics)、宏基因组学(metagenomics)、病理基因组学(pathogenomics)、生物信息学(bioinformatics)等。基因组的理论与方法已经渗透到生物、农业、医学以及其他自然科学基础研究中,成为理解和认识生命过程和现象的重要工具,也成为推动生命科学研究发展的新动力。基因组学的发展为农业生物技术提供重要的理论和技术基础,医学基因组学的发展使得人们在认识疾病复杂成因上又前进了一步,为新型药物创新提供了新的途径。基因组学发展需求也为数学和信息学科发展提出新的动力。因此从某种意义上讲,我国基因组研究的快速发展,直接和间接推动了诸多学科和研究领域的快速发展,成为引领生命科学基础研究的新引擎,为较短时间内缩短我国与发达国家间差距做出了重要贡献。

4 基因组学发展应该关注的几个问题

基因组测序工作完成后,首先摆在人们面前的是基因组的结构特征与规律的问题^[7],即:物种的基因组构成模式及其规律;物种基因组的比较和演化机制;染色体结构和传递规律;遗传元件的功能分析及其规律;基因组可塑性的机制。目前已经知道不同生物有不同的基因组结构,原核生物基因组序列一般很少重复序列、内含子等传统认为的“无用DNA”(junk DNA),而在真核生物中不同门类的生物基因组结构也十分不同,如植物往往重复序列多、基因数目也多,但哺乳动物则是选择性剪切很多,基因数目却远远少于高等植物。植物和低等脊椎动物往往有较多多倍体物种,但在高等动物中基因组剂

量似乎受到很大制约。除一级序列差异外,不同物种在染色体结构和调控方式上也有很大不同。目前,对基因组的结构特征和变化规律理解尚十分粗浅,是遗传学研究中的一个十分薄弱的环节。如何深刻理解不同生物基因组的结构特征及其变化规律,是解读海量基因组数据的重要研究方向。对基因组结构在3维、4维,也即染色体的结构及其调节的研究,则是理解基因组中遗传信息如何在细胞中存贮、表达和调控的未来热门方向。

基因组测序工作完成后,功能基因组学研究变得越发重要^[7],一些问题如:启动子、增强子、转录因子结合位点等基因表达顺式作用元件的组成和功能;内含子、非翻译区或相近的基因及转座子调控蛋白编码基因转录的作用和机制;假基因的调控功能;转座子在决定基因组大小、基因结构和基因表达中的作用;中心粒和端粒重复序列的组成和拷贝数目的变化调控生物生长发育的机制等等成为研究热点和前沿。

基因组测序工作完成后,“表观基因组学”与非编码核酸日益受到基因组学研究的关注^[7]。表观遗传信息包括基因组中不编码蛋白质的DNA即非编码DNA的信息,也包括DNA分子的修饰(如DNA甲基化等)以及与DNA一起形成染色体的组蛋白修饰所携带的遗传信息。近年来,人们发现小分子非编码RNA在基因的转录和转录后加工、细胞分

化和个体发育等生命活动中发挥重要作用。因此,阐明非编码RNA的产生与调控机制以及发现新的具有调控功能的非编码RNA分子已成为后基因组时代生命科学研究的热点和前沿。

大量组学数据的产生已经改变了传统的针对单个基因或单个信号通路开展研究的模式,也使得生物信息学成为基因组研究密不可分的伙伴。针对各种高通量的组学数据,利用生物信息学的方法和手段,开发新的算法、软件,进行未知基因、新的调控元件与机制以及新的基因功能的预测与发现,开展调控网络构建和生物建模等工作,已经并将继续有力地推动基因组学研究的快速发展。

参 考 文 献

- [1] 段民孝,李宝键.基因组学研究概述. <http://www.hnagri.org.cn/shownews.php?id=7056>.
- [2] Collins F S, Green E D, Alan E. Guttmacher & Mark S. Guyer. A vision for the future of genomics research. *Nature*, 2003, 422: 835—847.
- [3] 王钦南,江虎军,诸嘉佑.中国的人类基因组计划. *生物工程进展*, 1993, 14(8): 52—56.
- [4] 江虎军,王钦南,强伯勤等.我国人类基因组计划的启动、进展及展望. *自然科学进展*, 2001, 11(8): 777—781.
- [5] 黄辛.我国基因组学研究发展令人振奋. <http://www.labweb.cn/html/20100604/422.html>.
- [6] 国家自然科学基金委员会资助项目信息. [Http://isis.nsf.gov.cn](http://isis.nsf.gov.cn).
- [7] 国家自然科学基金委员会.2011—2020年我国生物学学科发展战略.2010年.

DEVELOPMENT OF GENOMIC RESEARCH IN CHINA AND NSFC

Gu Ruisheng Tian Yanyan Du Shengming
(National Natural Science Foundation of China, Beijing 100085)

Abstract Genomics is a novel field emerging at the end of the 20th century. It develops quickly, and some breathtaking achievements have been made in China. This paper gives a brief review of the development and the current status of genomic research in China, presents the unique role of National Natural Science Foundation of China (NSFC) in promoting the genomic development in China, introduces some representative achievements founded by NSFC, and discusses some research areas and scientific questions which should be underlined in future.

Key words genomic development, funding, representative achievements, NSFC