大发现时代的“生命组学”(代序)

贺福初

军事医学科学院, 北京蛋白质组研究中心, 蛋白质组学国家重点实验室, 国家蛋白质科学中心(北京), 北京 100850

E-mail: hefc@bmi.ac.cn

收稿日期: 2013-01-05; 接受日期: 2013-01-10

中澳国际科技合作项目(批准号: 2010DFA31260)和中加健康研究合作计划(批准号: 81010064)资助项目

摘要 自然科学史表明, 当人类对某一领域的认知积累到一定程度时, 常常会出现一位甚至数位科学大家, 促使一系列重大发现纷至沓来, 相关学科因而进入“大发现时代”。数学、地理学、物理学、化学都曾出现过“大发现时代”。而生命科学的“大发现时代”自 16 世纪以来层出不穷、持续不断。随着分子生物学 50 余年日新月异的发展, 当代生命科学的“大发现时代”临近再次爆发, 而最终“点燃”此次爆发的极大可能就是“生命组学”。20 世纪末, 基因组学书写了“生命天书”, 蛋白质组学随即解读这部伟大天书, RNA 组学、糖组学、代谢组学等也相继蓬勃兴起。简言之, “组学”以其特立独行的认识论、方法论, 一经问世便迅速成为推动并主导生命科学再度迈入大发现时代的强劲引擎。集多种组学之大成的“生命组学”研究模式已现端倪, 大发现时代将如影随形。

关键词

大发现时代
生命组学
生命科学

自然科学是发现的“艺术”, 发现是自然科学的基石。发现, 基于人类对已知世界的认识和对未知领域的探索, 既需要扎实的积累, 又需要天才的思想。纵观自然科学史, 常常可以看到“厚积薄发”的现象: 当人类对某个领域的认知积累到一定程度时, 必然会出现一个甚至数个划破历史长空的科学大家, 应承时代的召唤, 指引纪元的更替, 他(们)促使重大发现蜂拥而至、喷薄而出, 并汇聚成滚滚洪流, 冲破已有理论信条的桎梏, 将该领域的理性认识推上全新的高度, 使一个或多个相关学科呈现爆发式成长, 脱胎换骨乃至革命性突变, 如此该学科可谓进入了“大发现时代”。随着分子生物学 50 余年突飞猛进的发展, 尤其是生命组学等领域近 20 年的日新月异, 当代生命科学临近爆发的边缘, 而最终“点燃”此次爆发的极大可能就是“生命组学”(基因组学、RNA 组学、蛋白质组学等组学的集合)。

1 自然科学史上的若干“大发现时代”

人们对于自然科学的了解和学习大多从数学开始, 近代科学正是在追寻自然界的数学规律中取得长足进步的, 许多重大突破, 都是基于数学的发展。古希腊毕达哥拉斯学派的突出贡献即开启了自然科学的第一个大发现时代。该学派最早证明了勾股定理, 提出了奇数、偶数、质数、亲和数、完全数等概念。他们发现, 算术的本质是“绝对的不连续量”, 音乐的本质是“相对的不连续量”, 几何的本质是“静止的连续量”, 天文学的本质是“运动的连续量”。在此基础上, 该学派提出“数即万物”学说^[1]。在此学派“数本主义”哲学的影响下, 柏拉图甚至在学院门前立下牌匾: “不懂数学者不得入内”。在他看来, 数学是通向理念世界的必备工具。数学的发展, 奠定了以后各学科“大发现”的基石。

引用格式: 贺福初. 大发现时代的“生命组学”. 中国科学: 生命科学, 2013, 43: 1-15, doi: 10.1360/052013-5

英文版见: He F C. “Life Omics” in grand discovery era. Sci China Life Sci, 2013, 56, in press

地理学的“大发现时代”爆发于短短的 40 年,却影响了世界数百年的格局. 15 和 16 世纪之交,以地球说为理论指导,借助于指南针和罗盘的发明,地球上不为文明世界所知的地域和航线不断被发现. 1485 年,哥伦布发现北美大陆; 1497 年,达·迦马发现印度洋和印度; 1498 年,哥伦布又发现南美大陆; 1519 年,麦哲伦发现南美大陆最南端海峡,从而找到大西洋直达太平洋的通路; 1521 年,麦哲伦通过此海峡发现了太平洋,从此开启了西方和现代文明的新篇章^[1]. 地理学大发现所引起的观念革命与它所带来的经济后果一样巨大,它大大突破了亚里士多德和托勒密的知识范畴与视野,促使欧洲的知识阶层能从近 2000 年来古典大家的绝对权威与神条中解放出来,从顶礼膜拜中猛醒,昂起头、睁开眼,为近代科学革命开启了批判的理性天窗和革新的精神动力.

现代化学的创立时期也有一段突飞猛进的辉煌. 18 世纪关于气体的大发现、元素的大发现催生了化学革命. 1756 年布莱克发现二氧化碳; 1760 年丹尼尔·卢瑟福发现氮气; 1766 年卡文迪许发现氢气; 1774 年普利斯特发现氧气^[1]. 这些气体的发现,使拉瓦锡于 1789 年把燃烧定义为一个更科学的表述——氧化,推翻了统治化学界上千年的燃素说,奠定了现代化学的科学基础. 在大量科学研究的基础上,拉瓦锡写出了划时代的《化学纲要》^[1],提出了元素学说,开创了化学的新纪元,并将此后的数十年带入化学元素大发现的时期,存在于宇宙亿万年的上百种元素一半以上在此历史瞬间被发现(表 1). 这无疑是科学的辉煌、人类的荣耀. 1869 年,门捷列夫就是在以上发现的 63 种元素及其性质描述的基础上^[1],发现了元素性质与元素的原子量之间存在周期性的变化规律,并发表了第一张元素周期表,使得元素周期律作为化学世界第一个伟大定律永远载入史册. 由此不难看出,化学的大发现同样源于厚积薄发,并成于学科革命.

上述几个大发现时代均处于学科的奠基阶段. 人们不禁要问: 学科成熟后,是否还可能再有大发现时代呢? 答案是肯定的,天文学、物理学即是很好的例证.

自然科学史上,天文学与数学一样是古老而成熟的学科,引领了启蒙时期的科学发展. 在 16 和 17 世纪之交,以革命性的日心说为理论指南,凭借望远镜技术支撑,出现了天文学史上新的发现时代.

1572 年,仙后座一颗新星爆发,第谷在其出现至消失于视野的 16 个月里进行了详细观测,以翔实的观测数据宣告其被首次发现,并创造了“新星”(nova)一词; 1577 年,第谷依据对一颗明亮彗星连续 74 天的观测记录,得出彗星是地球外绕日运行天体的结论,纠正了当时认为彗星是大气层内现象的错误认识^[1]; 1576 年,第谷在丹麦国王腓特烈二世资助下,在汶岛建立了当时最华丽的天文台,此后 21 年如一日地观测恒星、行星、太阳、月亮的运动,积累了大量观测数据,并编制了包括 1000 余颗恒星的第一份完整近代星表. 被誉为“星学之王”的第谷,凭借鹰鹫一般的锐眼,将观测天文学演绎到登峰造极,其观测精度之高,令同时代人望尘莫及. 1600 年,第谷慧眼识中开普勒,为其无私提供了自己经年积累的天象观测数据. 开普勒在这些珍贵资料的基础上,力求“拨云见日”、苦寻宇宙秩序,很快发现了开普勒第一、第二定律,并将两大定律推广到太阳系所有行星. 此后不久,开普勒又发现了第三定律即周期定律,并利用这三大定律将所有的行星运动与太阳紧密地联系于一体,从而使得哥白尼的“日心说”由悬空的假说落地为有实证依据的理论体系,由“感性的天空”降落“理性的大地”. 太阳系的众多行星按照开普勒定律有条不紊地遨游太空,开普勒因而被称为“天空立法者”. 他的墓志铭令人沉思: “我曾测天高,今欲量地深; 思想遨游天际,肉体长眠大地.”可以说,正是基于第谷大量的精确观测、实验发现,开普勒才首次实现了对天体世界运动规律的高度理论概括,开启了科学理性照亮浩瀚宇宙的探照巨灯. 由此可见,成熟的天文学一旦接受革命性假说的理论指导,便可以再度进入实验的大发现时代,而其间的系列大发现又反过来实证并确立了它的革命性理论与学说.

无独有偶,17 世纪初又诞生了一位同开普勒一样功勋卓著的物理学家——伽利略. 他被公认为近代“科学之父”,近代实验科学精神的创造者. 他于 1604 年先后发现了运动的第一(匀速)、第二(匀加速)定律; 1609 年,他借助自己创制的天文望远镜观测天体,发现了月球表面的凹凸不平,并亲手绘制了第一幅月面图^[1],从而使人们对于月球的认识由想象上升为科学和理性可及的实体; 在此基础上,伽利略又发现了木星的 4 颗卫星,为哥白尼学说找到了确凿的证据,标志着哥白尼学说开始从解释存在到走向预见的胜利; 2 年后,伽利略相继发现了太阳黑子、太阳的

表 1 19 世纪为人类发现的化学元素(52 种)

序号	年份	化学元素	地点	发现者
1	1868	氦 He(Helium)	法国	让逊(Pierre Janssen)发现
2	1817	锂 Li(Lithium)	英国	洛克尔(Lockyer S.N.)定名
3	1808	硼 B(Boron)	瑞典	阿尔费德森(Arfvedson J.A.)
4	1898	氖 Ne(Neon)	英国	戴维(Davy H.)
5	1807	钠 Na(Sodium)	德国	科赫(Koch Robert)
6	1808	镁 Mg(Magnesium)	德国	戴维(Davy H.)
7	1825	铝 Al(Aluminium)	法国	戴维(Davy H.)
8	1823	硅 Si(Silicon)	丹麦	厄斯泰德(Oeisted H.C.)
9	1810	氯 Cl(Chlorine)	瑞典	贝采里乌斯(Bertholius J.J.)
10	1894	氩 Ar(Argon)	瑞典	舍勒(Scheele C.W.)发现
11	1807	钾 K(Potassium)	英国	戴维(Davy H.)定名
12	1808	钙 Ca(Calcium)	英国	瑞利(Rayleigh J.W.), 拉姆赛(Ramsay W.)
13	1879	钪 Sc(Scandium)	瑞典	戴维(Davy H.)
14	1831	钒 V(Vanadium)	瑞典	尼尔逊(Nilsson L.F.)
15	1875	镓 Ga(Gallium)	德国	塞夫斯特勒姆(Sefstrom)
16	1885	锗 Ge(Germanium)	法国	布瓦斯博达朗(Boisbaudran L.de.)
17	1817	硒 Se(Selenium)	德国	温克勒(Winkler C.A.)
18	1824	溴 Br(Bromine)	瑞典	贝采里乌斯(Berzelius J.J.)
19	1898	氪 Kr(Krypton)	法国	巴拉德(Balard A.J.)
20	1860	铷 Rb(Rubidium)	英国	莱姆塞(Ramsay W.)
21	1808	锶 Sr(Strontium)	德国	特拉弗斯(Travers M.W.)
22	1801	铌 Nb(Niobium)	德国	本生(Bunsen R.)
23	1827	钌 Ru(Ruthenium)	英国	基尔霍夫(Kirchhoff G.R.)
24	1803	铑 Rh(Rhodium)	英国	戴维(Davy H.)发现并制得
25	1803	钯 Pd(Palladium)	英国	哈切特(Hatchett C.)
26	1817	镉 Cd(Cadmium)	俄国	克劳斯(Klayc K.K.)
27	1863	铟 In(Indium)	英国	武拉斯顿(Wollaston W.H.)
28	1814	碘 I(Iodine)	英国	武拉斯顿(Wollaston W.H.)
29	1898	氙 Xe(Xenon)	德国	F.施特罗迈尔(Strohmeyer F.)
30	1860	铯 Cs(Caesium)	德国	里希特(Richter H.T.), 赖希(Reich F.)
31	1808	钡 Ba(Barium)	法国	库特瓦(Courtois B.)
32	1839	镧 La(Lanthanum)	法国	拉姆塞(Ramsay W.)
33	1803	铈 Ce(Cerium)	英国	特拉弗斯(Travers M.W.)
34	1885	镨 Pr(Praseodymium)	德国	本生(Bunsen R.), 基尔霍夫(Kirchhoff G.R.)
35	1885	钕 Nd(Neodymium)	英国	戴维(Davy H.)
36	1879	钐 Sm(Samarium)	瑞典	莫桑德尔(Mosander C.G.)
37	1896	铕 Eu(Europium)	瑞典	贝采里乌斯(Berzelius J.J.)
38	1880	钆 Gd(Gadolinium)	德国	克拉普罗特(Klaproth M.H.)
39	1843	铽 Tb(Terbium)	瑞典	希新格(Hisinger W.)
40	1886	镝 Dy(Dysprosium)	奥地利	冯·韦尔塞巴赫(Auer Von Welsbach C.F.)
41	1879	钬 Ho(Holmium)	奥地利	冯·韦尔塞巴赫(Auer Von Welsbach C.F.)
42	1843	铒 Er(Erbium)	法国	德·布瓦斯博达朗(de Boisbaudran L.)
43	1879	铥 Tm(Thulium)	法国	德马尔盖(Demarcay E.)
			瑞士	马里纳克(Marignac C.G.)发现
			法国	布瓦斯博达朗(Boisbaudran L.de.)制出
			瑞典	莫桑德尔(Mosander C.G.)
			法国	布瓦斯博达朗(Boisbaudran L.de.)发现
			瑞典	克莱夫(Cleve P.T.)
			瑞典	莫德桑尔(Mosander C.G.)
			瑞典	克莱夫(Cleve P.T.)

(续表 1)

序号	年份	化学元素	地点	发现者
44	1878	镱 Yb(Ytterbium)	瑞士	马里纳克(Marignac C.G.)
45	1802	钽 Ta(Tantalum)	瑞典	埃克伯格(Ekaberg A.G.)
46	1803	锇 Os(Osmium)	英国	台奈特(Tennant S.)等
47	1803	铱 Ir(Iridium)	英国	台奈特(Tennant S.)等
48	1861	铊 Tl(Thallium)	英国	克鲁克斯(Crookes W.)
49	1898	钋 Po(Polonium)	法国	皮埃尔·居里夫妇(Curie P., M.S.Curie)
50	1898	镭 Ra(Radium)	法国	皮埃尔·居里夫妇(Curie P., M.S.Curie)
51	1899	锕 Ac(Actinium)	法国	A.L.德比尔纳(Debierne A.L.)
52	1828	钍 Th(Thorium)	瑞典	贝采尼乌斯(Jons Jakob Berzelius)

自转、金星和水星的盈亏现象等。这一系列的重大发现再一次开辟了天文学的新时代。他通过观测，形成了实验测量与理论分析良性互动、交相映辉的研究模式，并留世了“测量一切可测之物，并变不可测者为可测”的不朽名言，激励着后来所有实验科学的探索者。在实验科学雄厚的基础上，逐步树立起了理论物理学的灯塔。在理论物理学灯塔的指引下，人类因对于物质世界的洞悉而获得“海阔凭鱼跃，天高任鸟飞”的自由。

开普勒、伽利略等人的系列大发现，催生了牛顿的集大成时代。在数学上，他发现了二项式定理并创立了微积分；在天文学上，他发现了万有引力定律，开辟了天文学的新纪元；在物理学方面，他系统总结了三大运动定律，统一了开普勒、伽利略运动定律；在光学中，他发现了太阳光的光谱，发明了反射式望远镜。在取得这一系列伟大成就之后，牛顿于 1686 年发表盖世巨著《自然哲学的数学原理》^[1]，最终统一了他关于动力学和引力问题的系列重大发现，创立了完整的新物理学体系。正如牛顿勋爵自己所言：如果我比别人看得远些，那是因为我站在巨人的肩膀上。人们常将此看作谦虚之词，但认真回顾其理论的形成过程，可以看出它并非空穴来风。因为有了第谷系统精确的观测数据，开普勒才能从中归纳出行星的运动规律；因为伽利略的大量实验研究，他才能发现运动第一、第二定律；随后在与天文学相关规律、物理学相关定律整合、统一的基础上，才有了牛顿力学理论的集大成。所以说牛顿确实是站在第谷、开普勒、伽利略等巨人的肩膀上，才建立起如此盖世之功。而这些巨人的产生，则是基于大量的观测、实验数据的积累和规律的发现。

19 世纪末，牛顿力学统一了声学、光学、电磁学和热学，“万能”的牛顿定律支配着小到超显微粒子、

大到宇宙天体的整个物质世界，物理学似乎已没有继续发展的空间。然而，“美丽而晴朗的天空上仍笼罩着两朵乌云”。正是迈克耳逊-莫雷实验和黑体辐射实验这两个“灾难性”的发现，使相对论、量子论取代牛顿力学成为物质世界更普适的理论，人类物质生活也发生了翻天覆地的改变。因此也出现了一种新的科学研究模式——“假设驱动”的研究模式，以区别于传统的“发现驱动”研究模式。是否任何一个理论上的创见都是源于假设？实际上不尽其然。纵使理论性高的学科，无论是天文学，还是物理学，理论或源于观测、或终于观测。而这些观测，有些是在假设的引领下，有些则完全没有理论的指引。如前所述，在地球说和日心说的引领下所产生的地理学和天文学的大发现，源于假设驱动，终于假说证实。然而开普勒三大定律和伽利略两大定律的发现，是没有假设前提的，纯粹是基于大量的科学观测和海量的数据归纳，它们往往导致“意外的”、原创性、原理性的重大发现。即使是大多数情况下只能进行思维实验的相对论，其最基本的假设也是基于迈克耳逊-莫雷实验的结果。事实上，假设驱动与发现驱动对于现代科技的起源与发展同等重要。先行者伽利略的至理名言“测量一切可测之物，并变不可测者为可测”应当成为现代科学，尤其是实验性科学遵循的基本原则。

综上所述，只有厚积，才能薄发。学说牵引、技术推动，共同驱动大发现。没有大发现，难成“大革命”。“大发现”必至“集大成”。

2 生命科学史上的大发现时代

近代以来，生命科学是一个不断出现大发现时代的古老而青春焕发的学科。因而整体上可以说，生

命科学的大发现时代从 500 多年前持续至今。

16~17 世纪, 近代生命科学肇始于生物学“大发现”。1543 年, 以维萨留斯为代表的解剖学家通过对动物、人体的解剖, 从整体的角度对人体结构有了充分的认识。维萨留斯出版的伟大著作《论人体结构》^[1], 系统地总结了多年来他对解剖学的系列重大发现, 包括骨骼、肌肉、血液、神经、消化、内脏六大系统与脑感觉器官。它们的发现是科学战胜“千年黑暗与愚昧”的盖伦神威、基督神权的伟大胜利! 因而丝毫不逊于在人体这一“太阳系”中发现七大行星。《论人体结构》不仅构成了现代生命科学的开篇, 同时也奠定了现代医学的基石。1616 年, 哈维发现心脏的结构与功能, 建立血液循环理论。17 世纪 60 年代, 马尔比基发现“有机体越低级, 呼吸器官比例就越大”。1665 年, 胡克发表《显微图》, 发现并命名“细胞(cell)”。生命科学从此进入微观世界。1683 年, 列文虎克借助显微镜发现比原生虫小得多的细菌, 于 1688 年又发现红血球^[1]。不难看出, 微观生物学的这些大发现不仅打开了认识生命微观世界的“天眼”, 而且开创了“以微释著”的新型科学认识论模式。

17 世纪, 也是生物物种的大发现时代。在公元前 335 年的亚里士多德时代, 经过科学研究和精确描述的动物已有 500 多种。亚里士多德首创“动物”一词, 并创作了第一本动物学名著, 其学说主宰学界上千年。历经 2000 年, 时钟指向 1600 年, 有科学记载的植物约 6000 种。而后, 仅在 17 世纪的一百年中, 植物学家新发现物种 12000 个^[1], 是前 2000 年发现总数的 2 倍。同时期, 动物新物种的发现也经历了同等规模的“井喷”。因此, 毋庸置疑, 17 世纪, 是生物物种的大发现时代。

正是因为大量新物种的发现, 使得建立一个逻辑自洽的理性物种分类系统的需求变得极为迫切。1735 年, 林奈出版《自然系统》, 首先提出以植物的性器官进行分类的方法。虽然他主观上抗拒进化思想, 但其建立的分类体系客观上推动了进化思想的发展。1749~1788 年, 布丰出版了鸿篇巨著《自然史》^[1], 提出自然界演化图景, 即物种可能具有共同祖先的观点, 成为近代第一个以科学精神对待物种起源问题的学者。1785 年, 赫顿发表生物化石的火成论思想, 使得化石纳入了科学研究的范畴, 成为生物体系建立的重要依据, 尤其是文明史前发生、现存生物界缺如的断代、断链性证据。1801 年, 拉马克出版《无脊

椎动物的分类系统》, 首次提出生物进化的思想。1809 年, 拉马克出版《动物哲学》, 系统阐发了拉马克主义的进化理论, 并提出“生物学”(biology)一词^[1]。1831 年, 达尔文以博物学家的身份开始历时 5 年的环球科学考察, 在动植物和地质方面进行了大量标本分析和采集, 经过综合归纳, 形成了生物进化的概念; 又经过 20 余年的大量研究, 于 1859 年发表了《物种起源》^[1], 从而使萌发了近一个世纪的进化思想, 终于成为宏大而具说服力的革命性进化理论。可以说, 没有物种的大发现, 没有达尔文环球之旅的大发现, 就没有永载科学史册、文明史册的进化论。

19 世纪, 细胞学说的问世标志着生命科学第一次微观层面的系统大发现结出了革命性理论的硕果, 同时还衍生了现代医学的革命性研究模式——实验医学。1831 年, 植物学家布朗发现细胞核。1838 年, 施莱登发表《植物发生论》, 指出“无论怎样复杂的植物体均由细胞组成”^[1], 细胞作为植物体生命的基本单元维持着整个植物体的生命。1839 年, 施旺发表“动植物结构和生长的相似性的显微研究”的论文, 指出“一切动物组织, 均由细胞组成”, 同时推倒了分割动植物界的巨大屏障, 建立生物学中统一的细胞学说。1858 年, 微耳和发表《细胞病理学》, 指出“一切细胞来自细胞”、“所有疾病都只不过是改变了的生命现象”。1875, 斯特拉斯伯格出版《细胞组成与细胞分裂》, 指出“一切细胞均由已存在细胞的均等分裂而成”、“细胞核的分裂先于细胞分裂”。细胞学说从此不再是假说, 进而迅速在生命科学中显示出强大的理论指引性威力。

随着显微镜技术的日渐成熟, 19 世纪下半叶, 生命科学又迎来了微生物的大发现时代(表 2), 成就了生命科学史上几近可与进化论相媲美、现代医学史上引发第一次革命的微生物学的辉煌。1856 年, 巴斯德指出“所有的发酵都是由微生物引起”。1859 年, 巴斯德推翻“自然发生论”。1860~1880 年, 巴斯德以对工业上酒发酸、农业上丝蚕病、医学上传染病等系列惊世难题的成功处置, 奠定“疾病病菌说”至高无上的地位。1870~1890 年, 科赫在细菌学原理和技术方面(纯培养法、染色法)做出开创性工作, 提出“每种传染病都有特定的病原菌”并予以系列确证, 确立“科赫准则”, 即某一种微生物要确定为病原体, 需符合以下四点: 必须在所有病人身上发现病原体; 必须从病人身上分离并培养出病原体; 培养出的病原体接种

表 2 病原微生物的发现年代

年份	病原体	所致疾病	地点	发现者
1877	炭疽芽孢杆菌(<i>Bacillus anthracis</i>)	炭疽病	德国	科赫(Robert Koch)
1878	葡萄球菌(<i>Staphylococcus</i>)	化脓	德国	科赫(Robert Koch)
1879	麻风杆菌(<i>Mycobacterium leprae</i>)	麻风病	瑞典	汉森(Hansen Armaner)
1879	淋病奈瑟氏菌(<i>Neisseria gonorrhoeae</i>)	淋病	德国	奈瑟(Neisser A.L.S.)
1880	金黄色葡萄球菌(<i>Staphylococcus aureus</i>)	中毒性休克综合征	苏格兰	亚历山大·奥斯顿(Alexander Ogston)
1880	伤寒沙门氏菌(<i>Salmonella typhi</i>)	伤寒	德国	艾博斯(Eberth C.J.)
1881	链球菌(<i>Streptococcus</i>)	化脓	苏格兰	阿格斯通(Ogston A.)
1882	结核分枝杆菌(<i>Mycobacterium tuberculosis</i>)	结核病	德国	科赫(Robert Koch)
1884	白喉棒杆菌(<i>Corynebacterium diphtheriae</i>)	白喉	德国	Edwin Klebs 和 Friedrich Löffler
1884	霍乱弧菌(<i>Vibrio cholerae</i>)	霍乱	德国	科赫(Robert Koch)
1884	破伤风梭菌(<i>Clostridium tetani</i>)	破伤风	德国	尼可奈尔(Nicolaier A.)
1885	大肠埃希氏菌(<i>Escherichia coli</i>)	腹泻	德国	欧利希(Ehrlich Paul)
1886	肺炎链球菌(<i>Streptococcus pneumoniae</i>)	肺炎	德国	弗伦克尔(Fraenkel Karl)/巴斯德(Louis Pasteur)
1887	脑膜炎球菌(<i>Neisseria meningitidis</i>)	流行性脑脊髓膜炎	德国	维克塞博姆(Weickselbaum Antou)
1888	肠炎沙门氏菌(<i>Salmonella enteritidis</i>)	食物中毒	美国	格尔特内(Gaertner A.A.H.)
1892	产气荚膜梭菌(<i>Clostridium perfringens</i>)	气性坏疽	美国	伟克(Welch W.H.)
1892	流感嗜血杆菌(<i>Haemophilus influenzae</i>)	脑膜炎和肺炎	德国	普菲费尔(Pfeffer Wilhelm)
1894	鼠疫耶尔森氏菌(<i>Yersinia pestis</i>)	鼠疫	德国	北里(Kitasato S.), 耶尔森(Yersin A.J.E.)分别独立发现
1896	肉毒梭菌(<i>Clostridium botulinum</i>)	肉毒中毒	比利时	埃尔蒙坚(van Ermengem E.M.P.)
1896	副伤寒杆菌(<i>Salmonella paratyphi</i>)	副伤寒	法国	阿夏尔(Achard Charles)
1898	痢疾志贺氏菌(<i>Shigella dysenteriae</i>)	痢疾	日本	志贺·洁(Shiga K)
1911	苍白密螺旋体(<i>Treponema pallidum</i>)	梅毒	日本	野口英世(Hideyo Noguchi)
1906	百日咳博德特氏菌(<i>Bordetella pertussis</i>)	百日咳	法国	博德特(Bordet J.), 根高(Gengou O.)
1909	杆菌状巴尔通体(<i>Bartonella bacilliformis</i>)	巴尔通体病	南美	阿尔伯特·巴通(Albert Barton)
1973	轮状病毒(<i>Rotavirus</i>)	婴儿腹泻	澳大利亚	毕夏普(Bishop R.F.)
1975	甲型肝炎病毒(<i>Hepatitis A virus</i>)	甲型肝炎	德国	Stephen M. Feinstone
1976	埃波拉病毒(<i>Ebola virus</i>)	埃波拉出血热	非洲	波文(Bowone), 帕提尼(Pattyn)
1977	汉坦病毒(<i>Hantaan virus</i>)	肾病综合征出血热	韩国	李镐汪
1977	嗜肺军团菌(<i>Legionella</i>)	军团菌病	美国和菲律宾	麦克戴德(McDade)
1978	丁型肝炎病毒(<i>Hepatitis D virus, HDV</i>)	丁型肝炎	意大利	马里奥·李泽托(Mario Rizzetto)
1981	人嗜 T 细胞白细胞病毒 1 (<i>Human T-lymphotropic virus 1, HTLV-1</i>)	T 细胞淋巴瘤, 白血病	美国	伯纳德·皮尔兹(Bernard Poiesz), 弗兰西斯卢思提(Francis Ruscetti)
1982	肠出血性大肠杆菌(<i>E. coli. O157:H7</i>)	出血性结肠炎	美国	雷利(Riley)
1982	人类 T 细胞性白细胞病毒 2 (<i>Human T-lymphotropic virus 2, HTLV-2</i>)	毛细胞性白血病	日本	Kalyanaraman VS
1982	伯氏疏螺旋体(<i>Borrelia burgdorferi</i>)	莱姆病	美国	博得佛(Burdorfer)
1982	比氏肠细胞内原虫(<i>Enterocytozoon bieneusi</i>)	顽固性腹泻	美国	郝雷(Gour ley)

(续表 2)

年份	病原体	所致疾病	地点	发现者
1983	人类免疫缺陷病毒 (<i>Human immunodeficiency virus</i> , HIV)	艾滋病	美国和法国	罗伯特·加洛(Robert Gallo), 吕·蒙塔利亚(Luc Montagnier)
1983	幽门螺杆菌(<i>Helicobacter pylori</i>)	慢性胃炎, 消化性溃疡	澳大利亚	巴里·马歇尔(Barry J. Marshall), 罗宾·沃伦(Warren J. Robin)
1986	人类疱疹病毒 6 型 (<i>Human herpesvirus 6</i> , HHV-6)	幼儿玫瑰疹	美国	罗伯特·加洛(Robert C. Gallo)
1983	戊型肝炎病毒 (<i>Hepatitis E virus</i> , HEV)	戊型肝炎	俄罗斯	巴娜炎(Balayan)
1989	丙型肝炎病毒 (<i>Hepatitis C virus</i> , HCV)	丙型肝炎	美国	迈克尔·侯顿(Michael Houghton)
1989	查菲埃立克体(<i>Ehrlichia chafeensis</i>)	人埃立克体病	美国	伯特·安德森(Burt E. Anderson)
1990	人类疱疹病毒 7 型 (<i>Human herpesvirus 7</i> , HHV-7)	热性皮疹, 中枢神经系统感染	美国	里加·弗里克(Niza Frekel)
1991	Gunarito 病毒	委内瑞拉出血热	委内瑞拉	萨拉斯(Salas R)
1991	脑细胞内原虫 (<i>Encephalitozoon hellem</i>)	脑膜炎, 弥漫性疾病	美国	迪迪奥(Didier E.S)
1991	巴贝虫新种 (<i>New species of Babesia</i>)	非典型巴贝虫病	美国	奎克(Quick RE)
1992	O139 霍乱弧菌(<i>Vibrio cholerae</i>)	霍乱	亚洲	菲利普·帕齐尼(Filippo Pacini)
1993	辛诺柏病毒(Sin Nombre virus)	成人呼吸窘迫综合征	美国	特里·亚特斯(Terry Yates)
1994	Sibia 病毒(Sibia virus)	巴西出血热	巴西	巴里(Barry M)
1995	人类疱疹病毒 8 型 (<i>Human herpesvirus 8</i> , HHV-8)	艾滋病相关卡波氏瘤	美国	恩瑞克·梅西(Enrique A. Mesri)
1995	庚型肝炎病毒 (<i>Hepatitis G virus</i> , HGV)	丁型肝炎	美国	西蒙斯(Simons J.N.)
1996	朊粒(Prion)	新型克雅氏病	美国	斯坦利·普罗西尼 (Stanley B.P Prusiner)
1997	输血后肝炎病毒 (<i>Transfusion Transmitted virus</i> , TTV)	输血后肝炎	日本	Nishizawa T
1997	禽流感病毒(H5N1)	流感	中国香港	de Jong
1998	西尼罗病毒(West Nile virus)	尼罗河热	乌干达	Smithburn K. C.
1999	尼派病毒(Nipah virus)	脑炎	马来西亚	Chua K. B
2003	SARS 冠状病毒(SARS-CoV)	严重急性呼吸综合征	中国	
2003	高致病性禽流感(H9N2)	流感	中国	
2005	人类博卡病毒 (<i>Human bocavirus</i>)	肺炎	瑞典	奥兰德(allander)
2010	新型布尼亚病毒 (severe fever with thrombocytopenia syndrome bunyavirus)	发热伴血小板减少综合征	中国	于学杰

动物, 使其出现与病人相同的症状; 从出现症状的动物体内能分离培养出同一种病原体. 科赫准则即使在 100 多年后的今天仍是必须遵循的铁则与金标准. 1890 年, 梅切尼科夫发现吞噬细胞与吞噬过程, 提出感染机理的细胞理论. 1890 年, 贝林发现对付白喉和破伤风的抗毒素, 提出体液免疫理论. 短短半个世纪, 在微生物大发现基础上建立的微生物学, 不仅为生物万千世界在传统、直观意义上的动物界、植物界的基础上, 打开了其种类、其作用丝毫不亚于前两界的微生物界, 而且使人类的理性之光让蔓延并横行人间数千年的“瘟疫”、“瘴气”从黑暗与愚昧中现形,

并为现代医学的第二次革命——抗生素革命开启了大门.

20 世纪被称为基因的世纪^[2], “基因”几乎成为 20 世纪生命科学的神话. 1900 年, 德佛里斯、科林斯、切马克各自独立重现孟德尔遗传定律. 1910~1930 年, 摩尔根发现基因连锁定律, 绘制出第一个染色体基因连锁图(果蝇), 出版了《遗传的物质基础》与《基因论》, 建立了完整的基因遗传理论体系. 1928 年, 格里菲斯^[1]发现肺炎双球菌转化因子, 艾弗里^[1]于 1944 年用生化方法证实其为 DNA. 1937 年, 欧洲结构学派的先驱者阿斯特伯里提出了 DNA 的第一个结

构假说. 他发表了一张 DNA 的 X 射线衍射照片, 并指出 DNA 是一个具有较强刚性的纤维结构. 1951 年, 威尔金斯和富兰克林^[1]拍下一张极好的 DNA 的 X 射线衍射照片, 提示其结构极有可能是右手双螺旋. 1952 年, 赫希和蔡斯用同位素标记方法证明噬菌体的遗传物质为 DNA. 同年, 查伽夫^[1]发现 DNA 碱基组成定律(A=T, G=C). Watson 和 Crick^[3]综合上述发现, 于 1953 年提出 DNA 双螺旋模型, 并指出, 碱基特异性配对可能是遗传物质复制的基础, 碱基排列顺序可能就是携带遗传信息的密码. 这些革命性猜想是 DNA 双螺旋模型的精要所在, 一经提出, 旋即掀起了生物科学史上最惊心动魄、人类文明史上最波澜壮阔、划时代的分子生物学的兴起. 它揭示了万古遗传之谜及其遗传密码, 揭示了统一万千生命世界的中心法则, 产生了比“创世上帝”更伟大的基因工程. 这段历史进程, 毫无疑问, 是生命科学领域一个典型的大发现时代.

1955 年, 美国的奥乔亚首次实现 RNA 的人工合成, 1957 年, 科恩伯格分离纯化 DNA 合成酶并人工合成 DNA, 对于解析遗传密码起到核心支柱作用(Nobelprize.org). 1955 年, Sanger 实验室^[4]测定牛胰岛素氨基酸序列; 1958 年, 同一实验室的 Kendrew 等人^[5]测定了肌红蛋白立体结构; Crick^[6]发表“论蛋白质的合成”并提出“连接物假说”, 论述核酸中碱基顺序与蛋白质中氨基酸顺序的线性对应逻辑并详述“中心法则”(从 DNA 到 RNA 到蛋白质). 1960 年, Jacob 等人^[7]提出 mRNA 和“操纵子”概念, 阐明其在遗传信息传递、蛋白质合成中的调节机制. 1961 年, Nirenberg 和 Matthaei^[8]破译遗传密码表. 1970 年, Baltimore^[9]及 Temin 和 Mizutani^[10]发现 RNA 逆转录酶, 修订中心法则, 标志分子生物学完整建立.

20 世纪 70 年代以后, 分子生物学挥舞还原论的神剑, 不断演绎基因的神奇, 几近到达无以复加、出神入化的境界. Linn 和 Arber^[11]及 Smith 和 Wilcox^[12]分别发现两种“工具酶”, 能对 DNA 剪切、连接. 1970 年, Nathans 用之实现 DNA 体外切割. 1971 年, Berg 运用“工具酶”实现不同种属基因重组. 1976 年, 简悦威报告首例单基因遗传疾病的基因诊断. 1977 年, Bishop 实验室^[13]首次发现癌基因 *src*. 1975 年, Sanger 等人^[14]建立 DNA 顺序分析法, 并于 2 年后完成 $\phi\chi$ -174 噬菌体 DNA 全序列测定. 1978 和 1981 年, Altman 和 Cech 分别发现 RNA 自身具有酶的生物催

化作用. 1979 年, 高德尔用基因工程生产人胰岛素. 1982 年, Palmiter 和 Brinster 获超级转基因小白鼠(生长激素). 1985 年, Mullis 的研究团队^[15]发明了 PCR 技术, 实现了神话般的基因体外扩增. 基因全序列测定、基因重组、基因诊断、基因工程、基因转移、基因扩增, 构成了新一轮的“基因神话”.

综上所述, 在当代生命科学史上, 确实存在诸多连续或彼此重叠的“不可思议”的大发现时代. 而每一阶段, 均有一位或几位高瞻远瞩、扭转乾坤、“开天辟地”的旗手性科学大师, 高擎革命的火炬, 划破黑暗的长空, 指引前进的方向.

3 “组学”引领新的大发现时代

生命体, 是迄今已知最为复杂的物质系统. 以人体为例, 从还原论的角度看, 当人们的目光从器官、组织深入到细胞、分子时, 每前进一步, 都意味着研究对象细分为几十个乃至上万个子集, 其相互之间的联系更是呈指数增长; 而从系统论的角度看, 不仅要看到内部多层次动态变化的“生物人”, 还要看到复杂理化因素作用下的“环境人”、与上亿个微生物共生的“生态人”、受众多社会心理因素影响的“社会人”. 对生命的认知就像芝诺的圆, 圆里已知的东西越多, 探索中接触的未知谜题反而也越多. 瑰丽的生命画卷, 在常人眼里, 是波澜壮阔、诗情画意, 但到了科学家的显微镜下, 则成为深浅不一的像素、色彩斑斓的线条和浓墨重彩的色块. 在生命科学新的大发现时代, 更令科学家们困惑的是如何将单调的点、线、面、体再现为艺术的巨幅画卷, 既不是东方形式的“盲人摸象”, 也不是西方模式的“阿尔钦博托肖像”(图 1). 系列生命“组学”就像俄罗斯套娃, 为探索生命奥秘提供了一套神奇的“魔镜”. 还原论上“全元素洞幽的放大镜”, 穷尽生命系统全部构成元件-各类生物大分子全组成及其调控规律, 因此将“还原”进行到底、到边, 散之为“理”, 是为“太”; 整体论上“巨系统揽胜的望远镜”, 汇融万千元素为一体、为系统, 化平实为神奇, 点无机成有机并生机的“魔幻大师”, 因此将整合升华到“际”、到“巅”, 统之为“道”, 是为“极”. 始于“组”, 终于“分”, 即是“还原”; 始于“分”, 终于“组”, 即是“整合”. 基因组学、RNA 组学、蛋白质组学、代谢组学等, 概莫能外. 因此, 生命组学独特、独到的认识论、方法论, 使其一经问世即迅

图 1 阿尔钦博托肖像

速成为启动并主导生命科学再度迈入大发现时代并直奔新世纪自然科学“盟主”的引擎、利器。

组学的发展引领了 20 世纪末至今的生命科学大发现。“碱基的排列顺序就是携带遗传信息的密码”，1953 年，沃森和克里克在提出 DNA 双螺旋结构后又续写了另一空谷绝唱。1958 年，其同一实验室的 Sanger 建立蛋白质氨基酸序列测定方法(http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1958/sanger.html), 70 年代梅开二度, 又建立了 DNA 序列分析方法^[16], 并因此两度荣获诺贝尔化学奖。“序列之王”Sanger 的卓越成就, 使读写基因的信息不再是空想。序列成为对生命的新认知, 人类两大先锋科技生物科技与信息科技, 通过序列实现了历史性的会师。1986 年, 第一代基于荧光测序技术的 DNA 自动测序仪诞生, 26 年来 DNA 测序能力呈指数增长, 当前的日数据产出量已达 Gb 级^[17], 比肩计算机芯片发展的“摩尔定律”。以色谱-质谱为代表的蛋白质大规模测序技术发展之势同样不可小觑, 在生物质谱技术获得诺贝尔奖的 2002 年, 一个样本可鉴定到的蛋白质尚不过几百种, 而现今生物质谱一次运行可鉴定到

的蛋白质竟多达数千种, 总数直逼转录组。大规模自动测序技术破茧而出, 似大鹏振翅高翔, 生命“组学”洞开生命大发现之巨门。

第一代 DNA 自动测序仪诞生后仅 4 年, 美国政府正式启动“人类基因组计划”^[18], 随后德、日、英、法和中國相继加盟。计划拟定之初, 已知的 DNA 序列仅有区区数十万级碱基对, 而 10 年弹指间, 人类不仅豪迈地完成了人类基因组 30 亿碱基对的第一次完整测序, 还成功实现了酵母菌(*Saccharomyces*)、大肠杆菌(*E. coli*)、结核分枝杆菌(*Mycobacterium tuberculosis*)、梅毒螺旋体(*Treponema pallidum*)、线虫(*Nematoda*)、果蝇(*Drosophila*)、拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa*)、小鼠(*Mus musculus*)、疟原虫(*Plasmodium*)、按蚊(*Anopheles*)等一批重要病原体 and 模式生物的基因组测序。该计划的实施还积极推动了一系列后续研究计划的诞生, 如 2002 年启动的“国际人类基因组单体型图计划”、2008 年启动的“国际千人基因组计划”、2009 年启动的“万种微生物基因组计划”、2010 年启动的“千种动植物基因组计划”、“万种脊椎动物基因组计划”。在人类基因组草图完成后 10 年的时间内(到 2011 年 10 月), 1200 多种生物的基因组已获解析。截止 2012 年 6 月 14 日, NCBI (National Center for Biotechnology Information) 的基因组数据达到 2.08 Pb(1 Pb=10¹⁵ b), 与计划前相比增加了 9 个数量级, 中国科学家的贡献超过 0.64 Pb, 占 30% 以上。其中, 基因组学的突出贡献可概括为以下 3 个方面: (i) 遗传信息的大发现。尽管目前只能解读其中部分, 但基因组测序使阅读亿万年生命进化中亘古积累、造化“神就”的奥秘成为可能, 仅此意义就不亚于燧人取火, 点亮文明之光。掌握上千个物种的遗传信息(生命蓝图)固然可喜, 然而这不过是冰山一角, 未来不仅在物种数量上会有更多突破, 而且将通过整体乃至单个细胞的基因组、转录组测序实现更为丰富的基因组构成规律、调控规律的大发现, 从而为合成生命、创造生命提供指南; (ii) 基因功能的大发现。人类基因组计划初步确认人类拥有 2 万余个蛋白编码基因, 而此前一个世纪人类仅发现其基因总数的零头; 它们的系统确认掀开了人类基因组神秘的面纱, 掀起了人类转录组、蛋白质组、比较基因组、模式动物表型组等组学研究的“狂潮”, 人们对基因功能的大规模发现日新月异。更有甚者, 人类基因组中由于编码蛋白质的基因序列不到总序列的 3%,

因而在物理世界发现暗物质、暗能量之后, 在生命世界可能首度革命性地开启“暗信息”——这一人类理性发现的“最新大陆”。过去 20 年里, DNA 测序能力的迅猛增长和测序费用的急剧下降, 使得“旧时王谢堂前燕”的 DNA 测序技术“飞入寻常百姓家”, 成为生物学研究、甚至临床诊断的常用工具; (iii) 致病基因的大发现. 以诞生不到 10 年的全基因组关联分析为例, 目前已完成 700 余种疾病和性状研究, 新发现了 5000 余种致病基因和重要性状基因, 其产出 10 倍于此前 100 年的发现, 大发现时代的突出特征由此可见一斑. 中国在此领域虽然起步较晚, 但成绩也是可圈可点(表 3). 如此辉煌成就足以证明沃森 10 年前那句豪情万丈的预言: “未来所有生物学只有以基因组开始才有希望发展!”

基因组测序的完成只是欲来山雨的“满楼风”, 而这部天书的解读则需要拨云见日、再造乾坤. 就在人类基因组草图公布的同时, 国际人类蛋白质组组织(human proteome organization, HUPO)宣告成立. 次年, 人类蛋白质组计划(human proteome project, HPP)即宣布启动. 鉴于蛋白质组的组分种类多(至少百倍于基因数)、丰度跨度大、翻译后修饰形式广、时空特异性繁复、组分间的网络性等特点, 人类蛋白质组计划先期启动了一系列示范计划: 于 2002 年首批启动肝脏、血浆蛋白质组分计划; 之后又陆续启动脑、肾脏和尿液、心血管等器官/组织蛋白质组分计划; 以及数据分析标准化、抗体、生物标志物等支撑分计划(表 4). 2005 年, “人类血浆蛋白质组计划”发布了 3020 种蛋白质的核心数据集^[51], 这是首个被鉴定的人体体液蛋白质组; 2010 年, “人类肝脏蛋白质组计划”精确鉴定出 6788 种蛋白质^[49], 这是首个被鉴定的人体器官蛋白质组, 其中半数以上为在人类肝脏中首次发现, 约 1/4 是首次在蛋白质层次被发现, 经典代谢通路数以千计的所有成员悉数发现, 因此构成了名副其实的“大发现”; 中国科学家领衔“人类肝脏蛋白质组计划”也显著推进了中国的蛋白质组学研究. 2009 年, 中国科学家通过通量化的蛋白质研究和不同物种的代谢通路研究, 发现了 1000 余个乙酰化蛋白^[52], 而此前仅在人肝细胞中发现 76 个乙酰化蛋白, 尤其重要的是, 该研究首次发现乙酰化修饰对代谢的广泛调控, 撼动了代谢调控的经典认识. 2012 年, 中国科学家选取大肠杆菌、酵母、线虫、果蝇、小鼠

及人类为研究对象, 通过对海量规模化定量蛋白质组数据的分析, 发现了蛋白质丰度整体分布的 3 个普适性规律^[53]: 丰度与其起源时间和序列保守性呈正相关的进化律; 丰度与其结构域数目呈负相关、与覆盖度呈正相关的结构律; 参与“基础物质流”的蛋白质丰度高于调控“精细信息流”的蛋白质的功能律. 蛋白质丰度分布规律的发现指明了一条点石成金的大道, 海量的组学数据将成为生命规律发现的“金矿”. 此外, 对其他器官/组织的分计划也发布了 1000~3000 余种蛋白质构成的数据集. 近年来, 随着生物质谱的鉴定能力和生物信息分析能力迅猛提升, 在细胞系样本中已可鉴定 7000~9000 种蛋白质, 覆盖基因表达产物的 60%~90%, 实现对人类基因编码蛋白的全部覆盖指日可待. 蛋白质相互作用图研究也取得惊人进展. 2005 年, 德国和美国的两个研究小组分别发布了包括 3169 对和 2754 对人类蛋白质组相互作用的图谱^[54]; 2011 年, 中国科学家发布了人类肝脏蛋白质相互作用图谱, 包括 2582 种肝脏蛋白质之间的 3484 种高可信度相互作用^[55]. 上述 3 项研究中重复的相互作用仅有 54 对, 提示每一个大规模的蛋白质相互作用图谱的建立均会导致 90% 以上未知相互作用的规模化发现, 因而称得上名副其实的“大发现”. 2011 年, 对人类内源蛋白质复合体的研究已涵盖 11000 余种基因表达产物, 约占基因组表达产物的 50%. 此外, 蛋白质组支撑分计划同样成绩斐然. 2011 年, 人类抗体计划发布的第 7 版数据库已覆盖 10118 个编码基因产物; 2009 年, 美国科学家 Chan 团队发现的 OVA1 及其标志物群成为第一个被美国食品药品监督管理局(Food and Drug Administration, FDA)批准的源于蛋白质组研究的肿瘤生物标志物(<http://www.fda.gov/NewsEvents/Newsroom/Press-Announcements/ucm182057.htm>). 这一划时代的事件, 标志着蛋白质组学不仅可有大量发现, 而且表明其发现可直接用于临床.

基因组学和蛋白质组学是生命组学中领异标新的二月花, 近年来其他组学也已风生水起, 可谓春色满园. 在人类基因组计划完成后, 人们发现仅有 2% 左右的序列编码蛋白质^[56]; 大量的非编码序列转录产生的非编码 RNA 中, 仅有 1/4 左右是传统的 tRNA 和 rRNA, 而 3/4 则为在染色质重构、转录调控、翻译调控等方面发挥重要作用的 snRNA, snoRNA,

表 3 中国科学家发表的全基因组关联研究(GWAS)

发表年份	GWAS 涵盖的疾病	研究团队	文献
2009	银屑病	张学军等人	[19]
2009	系统性红斑狼疮	张学军等人	[20]
2009	麻风病	张福仁等人	[21]
2010	鼻咽癌	曾益新等人	[22]
2010	白癜风	张学军等人	[23]
2010	肝细胞肝癌	贺福初、周钢桥等人	[24]
2010	食管癌	王立东等人	[25]
2010	水稻的农艺性状	韩斌等人	[26]
2010	多囊卵巢综合征	陈子江等人	[27]
2011	冠状动脉疾病	王擎等人	[28]
2011	食管癌	林东昕等人	[29]
2011	特异反应性皮炎	张学军等人	[30]
2011	肺癌	沈洪兵等人	[31]
2011	格雷夫斯氏病	宋怀东等人	[32]
2011	胃癌	沈洪兵等人	[33]
2011	精神分裂症	贺林等人	[34]
2011	精神分裂症	张岱等人	[35]
2011	麻风病	张福仁等人	[36]
2011	水稻的抽穗期和产量	韩斌等人	[37]
2011	胰腺癌	林东昕等人	[38]
2011	强直性脊柱炎	古洁若等人	[39]
2011	IgA 肾病	余学清等人	[40]
2011	非梗阻性无精子症	沙家豪等人	[41]
2012	川崎症	郭哲源等人	[42]
2012	冠状动脉疾病	顾东风等人	[43]
2012	多囊卵巢综合征	陈子江等人	[44]
2012	甲亢性周期性瘫痪	龚慧慈等人	[45]
2012	食管癌	林东昕等人	[46]
2012	前列腺癌	孙颖浩等人	[47]
2012	肝细胞肝癌	余龙等人	[48]

表 4 人类蛋白质组计划进展

启动年份	子计划	核心数据集	产出年份
2002	人类肝脏蛋白质组计划	ProteomeView, 6847 个蛋白质(6788 个 ^[49])	2010
2002	血浆蛋白质组计划	Human Plasma PeptideAtlas, 1929 个蛋白质(3020 个 ^[50])	2005
2003	人类脑蛋白质组计划	1832(人)/792(鼠)个蛋白质 蛋白质信息: mzML	2010
2003	蛋白质组标准化计划	分子间相互作用: PAR/MIAPAR/PSICQUIC 蛋白质分离: MIAPE-GEL/MIAPE-CC/MIAPE-CE 人类肾脏和尿液蛋白质组计划: 3679 个蛋白质; 非蛋白尿蛋白质组学的最终标准	2010
2005	人类肾脏和尿液蛋白质组计划	第 10 版数据库已覆盖 14079 个编码基因	2009
2005	人类抗体启动计划	O-糖基化全谱分析方法的比较, 肿瘤细胞糖蛋白的全谱分析	2012
2005	人类疾病糖组学启动计划	小鼠分泌蛋白质组已鉴定 1400 个以上蛋白质	2010
2005	人类疾病小鼠模型计划	1333 个蛋白质并附 10000 以上 GO 注释, 其中半数以上来自人的数据	2006
2006	人类心血管蛋白质组启动计划	干细胞标志物、干细胞信号通路、干细胞与疾病	2009
2007	干细胞生理蛋白质组启动计划	肿瘤、心血管疾病、肺病标志物	2009
2009	疾病标志物启动计划	模式生物进展	2010
2010	模式生物蛋白质组启动计划	即将公布(<i>J Proteome Res</i> , special issue)	2010
2010	人类染色体蛋白质组计划		2013

shRNA 和 microRNA, 长链非编码 RNA^[57]等, 影响着生物体的发育、细胞增殖分化, 及肿瘤、代谢性疾病、病毒性疾病等的发生发展. 2000 年“RNA 组学”概念问世, 仅 5 年时间被发现的哺乳动物非编码 RNA 就达 35000 余种. 非编码 RNA 对于 DNA、mRNA、蛋白质的调控作用以及非编码 RNA 之间的相互调控引起了广泛关注. RNAi 现象被发现仅 8 年时间, 就获得 2006 年度诺贝尔生理学与医学奖, 这在此领域诺贝尔奖历史上实属罕见. 无独有偶, 系统研究组织、细胞全部寡糖或聚糖的结构、糖链结构其宏观和微观的不均一性、糖链与糖识别分子的相互作用与其功能的“糖组学”, 以及整体定量描述生物内源性代谢物质及其对内外因应答规律的“代谢组学”等, 最近 10 年均取得长足发展, 在大发现时代的生命科学中发挥了生力军的作用.

“忽如一夜春风来, 千树万树梨花开”. 众多“组学”的蓬勃兴起, 既各领风骚, 又纵横捭阖. 2005 年, 人类蛋白质组计划启动的人类疾病糖组学分计划, 推动了蛋白质组学与糖组学的聚首. 2011 年形成的人类染色体蛋白质组计划(chromosome based human proteome project, Ch-HPP)联盟, 以蛋白质组技术鉴定每条染色体上编码的所有基因产物, 更是促成了人类蛋白质组计划与人类基因组计划的会师. 天地不仁, 以万物为刍狗; 生命不仁, 以纵“组”为刍狗. 综上所述, 整合基因组、RNA 组和蛋白质组、糖组、代谢组等多种组学的生命组学“大成”研究模式已现端倪, 前程未可限量.

回首自然科学史, 尤其是生命科学史上“大发现时代”的兴盛历程以及最近以来组学的勃兴进程, 可

以把“大发现时代”形象地比作“核聚变”, 原子核聚合并瞬间释放出巨能. “核聚变”的发生前提是原子核必须瞬间达到超高温, 才能突破巨大的能垒从而发生聚合, 因此核聚变必须靠核裂变来点燃, 即俗话说说的“氢弹靠原子弹点燃”. 科学史上常常靠思想或/和技术的重大突破, 甚至双重利器并举, 才能撬开新世界的大门、打通新大陆的道路, 从而使人类的理性、科学的光芒似秋风扫叶、长驱直入, 在历史的瞬间照遍新世界、覆盖新大陆, 进入“大发现时代”.

以人类基因组计划发端的生命组学开创了生命科学全新的“大科学、大工程”研究模式. 这种模式人类仅用过 2 次, 均在精确科学领域: 一次是核物理学上的曼哈顿原子弹研制计划, 科学不仅大开微观世界之门, 而且放出人类有文字记载以来所有最伟大的神话、最伟大的预言都远未企及的巨大核能; 一次是天文学上的阿波罗登月计划, 人类不仅终于走出地球——孕育并繁衍人类的摇篮, 而且变宇宙天堑为人间通途, 漫步太空、走向银河, “阿波罗”作为光明之柱照亮了浩瀚的深空——人类未来的远征天路.

万物有生则机, 生物世界是物质世界演进之极; 万生有命则灵, 生命活动是物质运动玄妙之致. 生命, 感性之表为生, 理性之里为命. 生命, 形而下之极为生, 名下实上; 生命, 形而上之致为命, 形玄势妙. 生命, 集万物之理, 不汇江河难为海, 因此, 生命科学必然是自然科学各领域充分发展后的集大成者. 生命, 达万理之成, 千瓦万砖树一楼, 因此生命科学必将是未来自然科学布新局、开新篇、鼓新潮的总设计师. 而集大成者之脊梁、创纪元者之灵魂, 非生命组学莫属.

致谢 感谢中国科学院学部、《中国科学》杂志社、《科学与技术前沿论坛》与中国科学院生命科学与医学学部常委会对“生命组学与转化医学”研讨会的大力支持; 强伯勤、曾益新院士共同组织此次研讨会; 施蕴渝、陈凯先、赵国屏、贺林、曾益新、杨焕明、陈润生、程京、张学敏院士及吴家睿、王俊、韩敬东、朱伟民、李玲等专家的精彩报告; 王东根、甄蓓、傅敏、史宇坤、赵亮等对会议的精心保障; 吴振宇、张雪莉、王琰、曹务春、张红星、孙薇、杨冬等对院士大会报告录音的系统整理与大量资料补充.

参考文献

- 1 吴国盛. 科学的历程. 长沙: 湖南科学技术出版社, 1995
- 2 Gelbart W M. Databases in genomic research. *Science*, 1998, 282: 659–661
- 3 Watson J D, Crick F H C. Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 1953, 171: 964–967

- 4 Ryle A P, Sanger F, Smith L F, et al. The disulphide bond of insulin. *Biochem J*, 1955, 60: 541–556
- 5 Kendrew J C, Bodo G, Dintzis H M, et al. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 1958, 181: 662–666
- 6 Crick F H C. On protein synthesis. *Symp Soc Exp Biol XII*, 1958, 12: 138–163
- 7 Jacob F, Perrin D, Sanchez C, et al. The operon: a group of genes with expression coordinated by an operator. *C R Acad Sci Paris*, 1960, 250: 1727–1729
- 8 Nirenberg M W, Matthaei J H. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci USA*, 1961, 47: 1588–1602
- 9 Baltimore D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, 1970, 226: 1209–1211
- 10 Temin H M, Mizutani S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, 1970, 226: 1211–1213
- 11 Linn S, Arber W. Host specificity of DNA produced by *Escherichia coli*, X. *In vitro* restriction of phage fd replicative form. *Proc Natl Acad Sci USA*, 1968, 59: 1300–1306
- 12 Smith H, Wilcox K W. A Restriction enzyme from *Hemophilus influenzae* *II. Purification and general properties. *J Mol Biol*, 1970, 51: 379–391
- 13 Stehelin D, Fujita D J, Padgett T, et al. Detection and enumeration of transformation-defective strains of avian sarcoma virus with molecular hybridization. *Virology*, 1977, 76: 675–684
- 14 Sanger F, Air G M, Barrell B G, et al. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, 1977, 265: 687–695
- 15 Saiki R, Scharf S, Mullis K B, et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 1985, 230: 1350–1354
- 16 Sanger F, Coulson A R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*, 1975, 94: 441–448
- 17 Gonzaga-Jauregui C, Lupski J R, Gibbs R A, et al. Human genome sequencing in health and disease. *Annu Rev Med*, 2012, 63: 35–61
- 18 Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409: 860–921
- 19 Zhang X J, Huang W, Yang S, et al. Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21. *Nat Genet*, 2009, 41: 205–210
- 20 Han J W, Zheng H F, Cui Y, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet*, 2009, 41: 1234–1237
- 21 Zhang F R, Huang W, Chen S M, et al. Genomewide association study of leprosy. *N Engl J Med*, 2009, 361: 2609–2618
- 22 Bei J X, Li Y, Jia W H, et al. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat Genet*, 2010, 42: 599–603
- 23 Quan C, Ren Y Q, Xiang L H, et al. Genome-wide association study for vitiligo identifies susceptibility loci at 6q27 and the MHC. *Nat Genet*, 2010, 42: 614–618
- 24 Zhang H, Zhai Y, Hu Z, et al. Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers. *Nat Genet*, 2010, 42: 755–758
- 25 Wang L D, Zhou F Y, Li X M, et al. Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies susceptibility loci at PLCE1 and C20orf54. *Nat Genet*, 2010, 42: 759–763
- 26 Huang X, Wei X, Sang T, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*, 2010, 42: 961–967
- 27 Chen Z J, Zhao H, He L, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat Genet*, 2011, 43: 55–59
- 28 Wu C, Hu Z, He Z, et al. Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nat Genet*, 2011, 43: 345–349
- 29 Wu C, Hu Z, He Z, et al. Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. *Nat Genet*, 2011, 43: 679–684
- 30 Sun L D, Xiao F L, Li Y, et al. Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population. *Nat Genet*, 2011, 43: 690–694
- 31 Hu Z, Wu C, Shi Y, et al. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat Genet*, 2011, 43: 792–796
- 32 Chu X, Pan C M, Zhao S X, et al. A genome-wide association study identifies two new risk loci for Graves' disease. *Nat Genet*, 2011, 43: 897–901
- 33 Shi Y, Hu Z, Wu C, et al. A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. *Nat Genet*, 2011, 43: 1215–1218

- 34 Shi Y, Li Z, Xu Q, et al. Common variants on *8p12* and *1q24.2* confer risk of schizophrenia. *Nat Genet*, 2011, 43: 1224–1227
- 35 Yue W H, Wang H F, Sun L D, et al. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at *11p11.2*. *Nat Genet*, 2011, 43: 1228–1231
- 36 Zhang F, Liu H, Chen S, et al. Identification of two new loci at *IL23R* and *RAB32* that influence susceptibility to leprosy. *Nat Genet*, 2011, 43: 1247–1251
- 37 Huang X, Zhao Y, Wei X, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet*, 2011, 44: 32–39
- 38 Wu C, Miao X, Huang L, et al. Genome-wide association study identifies five loci associated with susceptibility to pancreatic cancer in Chinese populations. *Nat Genet*, 2011, 44: 62–66
- 39 Lin Z, Bei J X, Shen M, et al. A genome-wide association study in Han Chinese identifies new susceptibility loci for ankylosing spondylitis. *Nat Genet*, 2011, 44: 73–77
- 40 Yu X Q, Li M, Zhang H, et al. A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nat Genet*, 2011, 44: 178–182
- 41 Hu Z, Xia Y, Guo X, et al. A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. *Nat Genet*, 2011, 44: 183–186
- 42 Lee Y C, Kuo H C, Chang J S, et al. Two new susceptibility loci for Kawasaki disease identified through genome-wide association analysis. *Nat Genet*, 2012, 44: 522–525
- 43 Lu X, Wang L, Chen S, et al. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nat Genet*, 2012, 44: 890–894
- 44 Shi Y, Zhao H, Shi Y, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet*, 2012, 44: 1020–1025
- 45 Cheung C L, Lau K S, Ho A Y, et al. Genome-wide association study identifies a susceptibility locus for thyrotoxic periodic paralysis at *17q24.3*. *Nat Genet*, 2012, 44: 1026–1029
- 46 Wu C, Kraft P, Zhai K, et al. Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. *Nat Genet*, 2012, 44: 1090–1097
- 47 Xu J, Mo Z, Ye D, et al. Genome-wide association study in Chinese men identifies two new prostate cancer risk loci at *9q31.2* and *19q13.4*. *Nat Genet*, 2012, 44: 1231–1235
- 48 Jiang D K, Sun J, Cao G, et al. Genetic variants in *STAT4* and *HLA-DQ* genes confer risk of hepatitis B virus-related hepatocellular carcinoma. *Nat Genet*, 2012, 45: 72–75
- 49 Sun A, Jiang Y, Wang X, et al. Liverbase: a comprehensive view of human liver biology. *J Proteome Res*, 2010, 9: 50–58
- 50 Omenn G S, States D J, Adamski M, et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics*, 2005, 5: 3226–3245
- 51 Deutsch E W, Eng J K, Zhang H, et al. Human plasma peptideAtlas. *Proteomics*, 2005, 5: 3497–3500
- 52 Wang Q, Zhang Y, Yang C, et al. Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science*, 2010, 328: 974
- 53 Zhong F, Yang D, Hao Y, et al. Regular patterns for proteome-wide distribution of protein abundance across species. *PLoS One*, 2012, 7: e32423
- 54 Stelzl U, Worm U, Lalowski M, et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, 2005, 122: 957–968
- 55 Wang J, Huo K, Ma L, et al. Toward an understanding of the protein interaction network of the human liver. *Mol Syst Biol*, 2011, 7: 536
- 56 Pennisi E. Shining a light on the genome's 'dark matter'. *Science*, 2010, 330: 1614
- 57 Wapinski O, Chang H Y. Long noncoding RNAs and human disease. *Trends Cell Biol*, 2011, 21: 354–361

贺福初, 中国科学院院士, 发展中国家科学院院士. 目前主要从事基因组学、蛋白质组学、生物信息学与系统生物学研究. 现任国际“人类肝脏蛋白质组计划”共同执行主席, 这也是中国第一次领导大型国际合作计划, *Nature*, *Science* 等国际著名杂志给予了首肯. 领导揭示了人体首个器官(肝脏)的蛋白质组及其连锁图. 曾获“国家自然科学基金委创新群体基金”、“求是杰出青年实用工程奖”、中国“五四”青年奖章、“何梁何利基金科学与技术进步奖”、“军队专业技术重大贡献奖”、“军队首届科技创新群体奖”、国际人类蛋白质组组织“杰出贡献奖”和“成就奖”、“谈家桢生命科学成就奖”等. 曾主编/主译专著 10 余部. 在 *Science*, *Nature Genetics*,

Nature Cell Biology, *Nature Biotechnology*, *Nature Methods*, *Nature Medicine*, *Nature Protocol*, *Proceedings of the National Academy of Sciences* 等国际核心刊物发表论文 230 余篇. 现任中国共产党第十八届中央候补委员, 国际人类蛋白质组计划执委, 国际人类肝脏蛋白质组计划共同执行主席, 亚太地区人类蛋白质组组织副主席, 国际核心刊物 *Proteomics*, *Proteomics—Clinical Application* 资深编辑, *Journal of Proteome Research* 编委, *Journal Proteomics & Bioinformatics* 执行编辑, 《中国科学: 生命科学》编委, 国务院学位评定委员会学科评审组成员, 国家高技术研究发展计划战略咨询委员会委员, 国家中长期“蛋白质科学重大研究计划”专家组召集人, 中国博士后基金会理事, 中华医学会副会长, 中国遗传学会副理事长, 中国生物物理学会副理事长, 北京市科协副主席等.