

人都需要做基因检测, 许多疾病发病原因通过常规检测即可确定, 在这样的情况盲目基因检测无疑是加重医疗成本。针对此, 他建议国家层面一定要严格控制基因检测, 举措包括资格许可证的颁布、基因测序市场主体的控制和对检测主体的行为抽查。

另外新型药物层出不穷, 各界皆报以很大程度的期待。但是新型药物研发难度大、毒性控制风险也难以把控, 想要应用在临床治疗上, 还有很长的路要走。“例如基因治疗药物的研发, 由于病毒载体, 怎样控制载体毒性是一个必需解决的问题。目前存在一些观点认为基因药物能够怎样改变未来的药品研发方向, 其实有些夸大。在未来很长的时间内, 基因治疗药物会是药物研发的方向之一, 但不会取代现在药物研发的主体。”

近日, 美国食品和药品管理局 (FDA) 批准了第一种3D打印药物, 这标志着个性化药物又向前迈了一步。看似药物研发正在朝着精准化方向高歌猛进, 但是在安全性和监管方面我们是否也做到了与之匹配的程度呢? 答案是否定的, “传统药物的生产是需要全程控制和监管的, 但是3D打印机目前无法满足全程管控的要求, 用于制备药物存在很多不确定性因素。现在就谈到3D打印改变制药方式更多程度是为了博眼球, 两者之间目前看来发展路径不是很清晰。”

郑维义, 国家“千人计划”专家, 南京百家汇科技创业社区有限公司首席运营官。

他山之石：美国如何建设精准医疗数据平台

——专访国家“千人计划”专家、美国维克森林大学生物信息中心主任周小波

文/本刊记者 魏晶晶



国家“千人计划”入选者周小波

中医讲求辩证施治, 针对不同的病人采取不同的治疗方法。今年医学界的热词“精准医疗”与其有异曲同工之妙: 在正确的时间, 给正确的人, 使用正确的药物和剂量。随着这一概念的提出, 个性化诊断、基因测序等成为精准医疗搜索的关键词, 然而, 这一切都离不开海量数据的处理。Clariant Diagnostic Service 首席医学事务执行官肯尼斯·布鲁姆指

出, 在疾病治疗中, 要进行筛选、诊断、分阶段等流程, 每一步都会面临非常多的信息, 最后我们要从大数据中提取最高质量的信息, 做出最好的决策, “精准医疗”其实就是帮助我们利用患者的数据信息作出最好的决策。

目前, 美国FDA计划建立一个名为“精准FDA (precisionFDA)”的平台, 为研究人员、新一代测序技术开发者等提供存放和共享基因信息的云工具; 在不久前, 美国国立卫生研究院 (NIH) 负责人Francis Collins则公布NIH将着手收集多达1百万美国人的基因信息。基因数据库的建立是美国实行“精准医疗”计划的必要装备, 由于人种基因的差异, 这些数据库对白种人和黄种人并非具备同等的价值。因此, 我们需要以自己的大型基因组学数据为基础来解读数据, 才能做好黄种人自己的“精准医疗”。

近日, 《千人》对话美国维克森林大学生物信息中心主任和首席科学家周小波, 透过他的视野, 为我国建立精准医疗的数据平台提供一些新的思路和对策。

中式盲区：数据到治疗未架桥梁

随着“精准医疗”一词在全国范围内的升温，“精准医疗与基因测序”、“精准医疗与肿瘤临床诊疗”、“精准医疗与大数据”等研讨会遍地开花，周小波教授在参加的几次研讨会中发现，对于“精准医疗大数据平台的建设”，国内似乎还存在着一定的误解。“精准医疗建设中有两个重点：一是患者的基因组数据信息，二是基于患者的电子病历(EMR)为病人设置精准治疗方案。这两点都已展开相应的研究工作，但如何将二者有机地结合仍是国内数据库建设的盲区。”

周小波教授分析，医疗大数据平台通常是指医院的临床信息管理系统，用于个性化诊断和精准医疗治理，主要由计算机学院主导医疗大数据平台的开发。但从事计算机行业的工程师对医疗系统的专有名词和治疗程序缺乏了解，开发出的数据软件往往不是医生和医学研究者真正想用的，这样就使得数据平台的建设难以达到医院治疗的要求。他进一步指出，精准医疗是一个交叉学科，涉及现代遗传、分子影像、临床医学、生物信息、计算机等领域，对患者生活环境和临床数据进行分析，实现精准的疾病分类和诊断，为患者提供个性化的疾病预防和诊疗方案。“所以，精准医疗是对临床电子记录，基因测序，医学图像等数据的整合和分析，这需要计算机工程、生物信息、医生、遗传学等跨领域专家的紧密合作。数据平台的建立远非计算机专业的专家孤军奋战。”

相较于中国，美国早已开始“杂糅”的实践：一些医院开始提取病人的临床样本，并将样本寄到“Foundation Medicine”公司(美国生物科技公司，面向公众提供基因检测服务)，由

“Foundation Medicine”公司负责343(到2014年8月4号)个标志性基因的测序，医院收到测序结果后进行一系列的生物信息统计并对遗传、疾病等各个方面的信息综合评估，制定最佳的治疗方案。周小波教授认为，尽管这个方案还未通过FDA的批准，但保险公司已经正式开始支付所有费用，这也意味着通过社会组织与医疗机构的携手，将基因数据用于精准医疗诊断的模式得到市场的认可。

社会组织与医疗机构的合作模式初见成效，代表着未来的临床研究模式很可能是医生和专业基因学家合作，一同进行实验和分析。但这一过程较为艰难，在培养大量的基因学专业人才的同时，还需引导医生自主学习基因学并用于临床试验。针对此，美国NIH(美国国立卫生研究院)提出“T32、R25”(医疗数据人才培养专项)计划，各医学院和大学都可进行项目的申请，以资助人才的培养。

它山之石，可以攻玉。周小波教授认为，要解决中国建设数据库的盲区问题，我国也需要建立类似基于大数据精准医疗的人才培养计划。他建议在每个大学任命一位在人才培养、研究、基金申请等方面有领导能力的教授做计划的负责人，由30个左右来自计算机、生物信息、遗传学、分子与细胞生物学以及临床医学等学科领域的教授组成委员会，组成人才培养计划的上层建筑；具体到学生能力的培养，周小波教授建议设定必修选修的机制，必修课应包含大数据科学、机器学习、数据挖掘、统计与方法，选修课程包括基因组学和生物信息学、计算系统医学、临床与转化医学信息学、分子遗传学与基因组学等。另外学生可以参加研讨班、国际核心期刊文献的报告会，并要开发一些初级的

医疗大数据平台或出版高水平的文章，以培养其学术研究和数据平台建设的能力。

信息“加密”既是机遇也是挑战

精准医疗是建立在海量的基因数据分析之上，患者的信息采集工作尤为重要，美国医疗中心正尝试采集更多、更高质量的信息。通过与优质的可穿戴设备厂商合作，扩大对用户健康数据的采集范围，并在线下加强社区驿站医疗服务中心的建设，采集更多的居民健康数据；在线上开设移动互联网服务，并进一步提高线上咨询服务的质量，从而提升数据获取的质量。

在数据的采集过程中，样本信息涉及病人的遗传、生活习惯、病史等个人隐私，如何说服患者、让他们放心的提供信息是一大挑战。十年前，美国就通过搜集多种癌症患者的样本进行测序治疗，并有严格的质量检控。为征得病人的同意，在搜集数据前，医院通常会对患者进行“医疗教育”，告诉他们所提取的样本信息只用于诊断研究，并签署保密协议。协议中明确规定不得泄露患者的可识别信息，确保信息的安全性。

相较于美国，周小波教授坦言，我国目前还未形成强有力的医疗教育系统，也缺乏可确保病人信息安全的医疗数据采集中心，说服患者提供样本将是一大难题。“如果在我国能建立类似‘Foundation Medicine’的权威机构，并设定保密的政策和书面协议，最大限度确保患者信息的安全性，我们相信将会有更多的人去提供基因数据，并接受这种诊断与治疗方式。”

采集之后如何实现信息的共享，是数据库建设中又一大难题。周小波教授在采访中提到，“美国NIH对癌症样本数据提取及测序中心给予大量资

金支持，已遍及美国多个州，这些数据根据NIH的规定进行公开，让全世界对医疗感兴趣的科学家和医生使用这些数据。”在精准医疗领域，数据的开放将大大推动医生对患者的靶点病情分析和治疗。通过共享精准医疗的数据，研究人员可更准确地判断出最有效的测试方式，避免不必要的重复试验以及改善临床护理的质量。但也应看到，大数据在传输的过程中，本身就存在信息泄露的缺陷。患者的个人隐私一旦泄露，后果不堪设想。为保证在信息共享中的的安全性，美国实行数据的“加密”开放。制定严格的IRB（Institutional Review Board）伦理审查委员会制度，患者的名字以及敏感的可识别信息不能公开。临床电子记录中有大量与病人隐私有关的信息，在开展研究项目时，一律要求在数据分析前去掉能识别病人的信息。

周小波博士补充到，“如果工作人员不小心将包含患者个人信息的电子记录从医院的邮件系统发到公开的邮件系统，那么该工作人员将被立即解雇。只有当医院要做临床试验时才需用到与患者名字有关的信息，但仅供医院内部使用，而当共享时，这些与名字有关的信息将被去掉。各学校和社会组织通常会设立数据共享委员会，用来确定可共享的数据信息和医疗项目。”

建设符合国情的数据平台

十年前，美国NIH提出National Center for Biomedical Computing (NCBC) 的计划，医疗大数据研究中心的医生就和计算机工程师在此计划的资助下，以哈佛大学（Harvard）领头，共同建立临床数据仓库（Informatics for Integrating Biology & the Bedside），简称I2b2，用于电子医疗和临床研究的数据分析。但很多医生在研究中无

我们需要以自己的大型基因组学数据为基础来解读数据，才能做好黄种人自己的“精准医疗”。

法将病例进行分享和有效的查询，于是他们开发了医学研究信息分享网络（SHRINE），这是一个利用i2b2节点查询医学数据并返回统计数值的网络。换言之，如果一个医院安装了i2b2数据仓库和SHRINE查询软件，同一医院的医生和研究者就可方便地找出符合临床试验条件的患者，从而加快临床试验进程。但同一医院的病人大都取于同一区域，且数量有限。所以美国国会会在2013年通过了一个推动临床与转化研究的提案，将大量的资金拨给临床数据研究网络CDRN（Clinical Data Research Networks），旨在推动不同地区临床试验方面的交流与合作。为了能够应用i2b2的医学研究病理，去辅助临床决策，由Harvard I2b2带领成立支持临床研究的IT基础设施，对筛选的特定病人组进行电话调研(1期)，之后再做临床试验(2期)。这也是目前美国临床与转化研究（CTSA）计划的重点研究之一，周教授也是这几个项目中的主要负责人之一。

美国精准医疗数据共享中心从无到有、从临时的数据库到建立完备的数据共享中心，得益于政府的政策和资金支持、医疗研究人员与计算机工程师们不断的发现问题、共同寻找解决困境的办法，根据形势和医疗内在需求的变化而调整步伐，始终站在精准医疗研究的最前沿。美国社会组织通力合作、不断创新的精神和理念值得我们借鉴，但

不能照搬照抄其发展模式，需强调一点的是，由于人种的基因差异，这些数据库对白种人和黄种人并非具备同等的价值。因此，黄种人需要自己的大型基因组学数据做基础来解读数据，才能做好黄种人自己的“精准医疗”。如何自立更生，建立适合我国精准医疗发展的数据平台，值得我们进一步探究。

对此，周小波博士认为，“美国通过十年建立较为完整的基因数据库，其模式的探索值得而我们借鉴。但白种人的基因数据库对我们中国人的参考价值有限，检测结果的准确性会大打折扣。因此，我们需要建立自己的数据库。我国的基因测序公司以中小型企业为主，而一些拥有先进技术的企业大多以科学研究为核心，面向公众的测序项目较少，且缺乏市场运作经验，导致我国基因检测行业发展较慢。因此，当前形势下应该将医院、科研院所、基因测序公司的资源有机的结合，一起去推动建立自己基因数据库的建立。”

此外，数据研究机构间的竞争非常重要，以美国NIH的运行模式为例，由政府出一定的资金，在全国各地建立大型的医疗数据研究中心，每个研究中心负责一种或几种疾病的研究，包括对不同地域病人基因信息的搜集、测序及分析，同时由政府支持在全国建立多个数据分析中心，便于研究中心将数据进行共享。

数据平台的建设离不开政府的政策引导和支持，更需要医疗机构、大学研究中心的通力合作，社会各方进行“抱团”，我国实现“精准医疗”的蓝图指日可待。