

# 绘制基因组的“谷歌地图”

## 新工具帮科学家实现染色体复杂构象可视化

染色体的功能远不止保持DNA整齐有序。这种基因组DNA和蛋白质组成的复合物有许多不同的结构和构象。这些结构和构象可能会影响周围基因的表达。在某些构象中,线性DNA中相距较远的两个序列可能实际上非常靠近,并影响彼此的活性;而在其他形式中,这两个序列又可能相距甚远。

美国麻省理工学院研究生 Erez Aiden 与同事共同开发了一种能在基因组水平上揭示染色体折叠方式的新工具。这种名为 Hi-C 的技术不仅能详细描述影响基因表达的 DNA 环和结构域,甚至还能将复杂的基因组拼接在一起。

### 灵感激发

虽然它只是以 2D 矩阵呈现的数据,记录了染色质的交互信息,但在 2009 年,Aiden 还没能找到一种简单方法探索这些空间构象。所以,他自己开发了一种技术。

Aiden 回忆道:“当时我只能打印出多个分辨率的 Hi-C 矩阵,而且需要用上百张纸。我还要找来最大的会议桌,把打印的所有矩阵都摆上去,以查看大规模的空间构象。”

即便如此,Aiden 也认为 Hi-C 是一个很好的界面。不过,他也承认需要一种更环保、可持续和共享的方法观察染色体构象。

最后,Aiden 开发了 Juicebox——一个基于 Java 的桌面应用程序。它可以提供 Google Maps 样式的染色质交互数据集,允许研究人员从基因组水平放大或缩小观察其结构特征。

2014 年发布的 Juicebox 大约被下载了 1.4 万次,今年研究人员又推出了一个基于浏览器的版本。Juicebox 只是一系列探索 2D 基因组交互数据的免费程序中的一个;一些程序专注于相对狭窄的染色体位点,而另一些则可以探索基因组。但这些程序的出现反映了染色质相互作用数据集的日益增长。事实上,4D 核组项目等大项目更是大规模地促进了染色体交互数据集的爆炸式增长。

马萨诸塞州波士顿哈佛医学院生物信息学家 Peter Park 指出,“因为数据变得如此复杂,所以可视化尤为重要。”

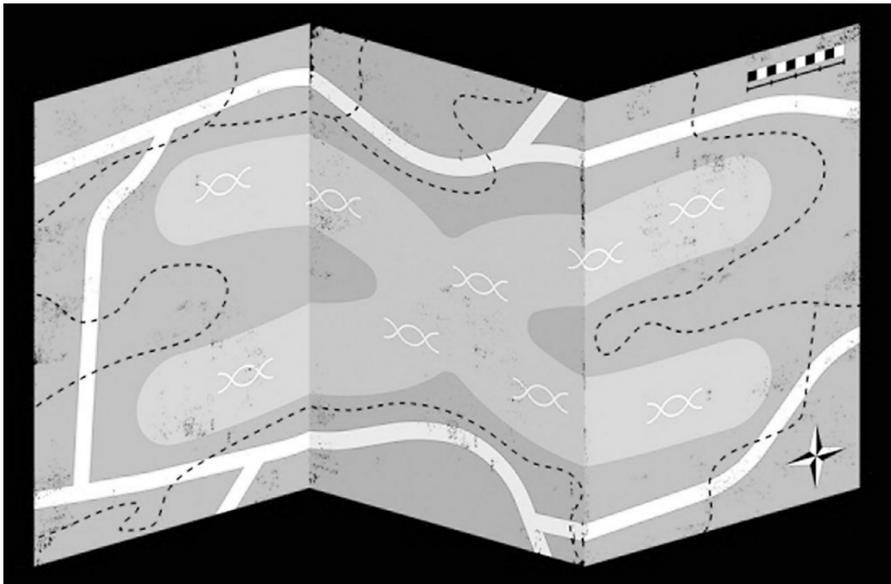
目前,加州大学圣克鲁兹分校开发的 Genome Browsers 是最受欢迎的探索基因组数据的门户之一。像大多数基因组浏览器一样,它将序列数据呈现为一维“轨迹”,显示为表观遗传特征(如组蛋白修饰和甲基化位点)的线性字符串列。

Aiden 等人都从 Google 地图中获得了灵感。加州大学圣迭戈分校的 James Robinson 表示,有了 Google 地图,用户就可以从全球视图无缝切换到街道级视图。这样一来,整个数据集非常巨大,但 Google 并没有一次性提供所有数据。相反,软件“将世界划分成不同分辨率的瓦片”。在任何时间,用户只能查看少量的瓦片。而这些瓦片被组织起来,使相邻的瓦片更易被获取。

类似的,Juicebox 的 hic 文件能以多种分辨率存储每个可能的染色体的图块集。软件的查询表可以直接检索数据,从而加快访问速度。因此,Juicebox 用户可以无缝探索整个基因

“对于基因组生物学而言,可视化是关键因素。”

图片来源: Project Twins



组的交互作用,然后放大以查看精细的功能。

### 基因组同步

今年 3 月,哈佛医学院生物医学信息学家 Nils Gehlenborg 开发了基于网络的 2D 基因组交互可视化工具——HiGlass, 它也提供了类似 Google 地图的体验。

与 Juicebox 一样,在 HiGlass 中,研究人员可以导入基因组轨迹以便了解所看到的内容。此外,HiGlass 还允许用户在一个浏览器窗口中打开多个 HiGlass 视图,并将它们同步起来,使其始终显示相同的区域。

Gehlenborg 表示,这样研究人员就可以比较不同条件或实验中的染色体构象。“我们在为研究者和分析师提供新想法的灵感。”他说。

该团队已经建立了一个 HiGlass 服务器,以挖掘公开的数据。由于需要分析自定义数据集的研究人员必须在本地安装该软件,Gehlenborg 团队为此提供了一个 Docker 容器。

实际上,Juicebox 的 Web 版本和 HiGlass 都允许用户创建可分享的 URL——指向数据的特定视图。Aiden 把这个功能称为软件的“杀手级应用”。他认为,如果用户注意到基因组结构与特定的 1D 轨道完全重叠,“那么点击那个 URL,复制它,就可以推送它了,而所有接收到该分享的人都可以点击它,随后便会得到相同的参数设置”。

另外两个可视化软件——3D 基因组浏览器和 WashU EpiGenome 浏览器均能提供更多的本地化视图。用户可以选择感兴趣的区域,浏览器会显示该区域的基因组交互信息。

此外,Juicebox 和 HiGlass 将热图映射成矩形的镜像,而这些浏览器则将热图显示为三角

形。加州大学圣迭戈分校基因组生物学家 Bing Ren 说:“我们去掉了一半的冗余信息。”

也许,这种变化可能听起来不大,但宾夕法尼亚州立大学的 Feng Yue 认为,这种变化能让研究人员更容易识别功能区域。例如,3D 基因组浏览器允许用户来自两个物种的热图相叠,以评估折叠体系结构的情况。“虚拟 4C”模式允许用户查询与特定基因组位点相互作用的序列的 Hi-C 数据集,从而方便研究者观察基因调控区域之间的相互作用。

另一个流行的基因组交互可视化软件是由加州大学圣迭戈分校的 Sheng Zhong 等人开发的 GIVE。GIVE 允许研究人员使用几行 HTML 代码,将完整功能的基因组浏览器(包括 2D 交互数据查看器)纳入其个人或实验室网页。Zhong 指出,研究人员可以与同事分享数据,发表文章时也可以附上链接,整个操作时间大约为 20 分钟。

意大利米兰 FIRC 分子肿瘤学研究所计算生物学家 Francesco Ferrari 则使用 R 编程语言和 Bioconductor 软件库显示其 Hi-C 数据。虽然这难与其他软件交互,但是由于该团队一直使用 R 和 Bioconductor 进行数据分析,“所以这样更方便”。

### 实现 3D

最终,2D 互动矩阵可以提示 3D 结构。毕竟,如果两个区域相互作用,它们可能距离非常接近。越来越多的研究人员正在使用他们的 2D 数据直接计算和可视化 3D 结构。

英国剑桥巴拉汉研究所博士后 Csilla Vátnai, 参与了今年早些时候单细胞 Hi-C 研究的 3D 模型构建工作。她使用一个名为

Gromacs 的通用分子建模包将染色体模拟成一条串珠——每个珠代表约 10 万个碱基,然后将串珠进行折叠,而 Hi-C 的交互数据则是折叠时的“约束条件”。

某些软件则专门被设计用于染色体结构的建模。挪威奥斯陆大学生物信息学家 Jonas Paulsen 开发的 Chrom3D 软件将 Hi-C 数据与核包膜距离的信息相结合,以模拟染色体在细胞核中的位置。Paulsen 提到,“这对基因调控非常重要”。核外周附近的基因倾向于被抑制,而位于中心的基因通常是有活性的。

西班牙巴塞罗那基因组调控中心基因组分析中心的 Marc Martí-Renom 和 Mike Goodstadt 开发了另一个 3D 工具——TADkit。TADkit 允许用户在相应的 2D 热图和 1D 轨迹旁边查看 3D 染色体模型。只要选中一个视图中的一个特征,那么软件就会自动点亮其他视图中的同一特征。

不过,由于大多数 Hi-C 数据集包含数百万个细胞,到底 3D 视图比 2D 视图能多提供哪些信息还有待观察。

对此,麻省理工学院生物信息学家 Leonid Mimy 打了个比方,你拍了一堆人的照片,然后将它们平均化,最后得到的照片会跟谁都不像。3D 视图可能也会存在这种问题。Zhong 指出,目前还不清楚哪个工具(如果有的话)将成为基因组可视化的黄金标准,现在这方面的争论已经很激烈了。

Ren 则表示,对于基因组生物学而言,可视化是关键因素。他解释称,分析工具是在统计数据的基础上设计而成的。有时候它们会错过一些东西,有时它们会推断出一些压根不存在的功能。“因此,科学家还是要谨慎,自己检查分析数据非常重要。” (唐一尘编译)

### 科学线人

全球科技政策新闻与解析

## 墨西哥地震或影响地区地质活动



墨西哥地震遇难人数升至 90 人。

图片来源:Ronaldo Schemidt

近日,墨西哥南部恰帕斯州托纳拉西南 137 公里的太平洋海域发生 8.2 级地震。当 Vlad Manca 听到这一消息时十分震惊,但并不意外。这位墨西哥国立大学的地球物理学家是研究该地区地震活动的为数不多的专家之一。一个多世纪以来,这里几乎没有有什么地震活动可以研究,但也恰恰是这个原因,让 Manca 认为该地区可能会出现大地震。

这次地震发生在太平洋东部的科科斯板块与北美板块交界附近。目前,科科斯板块以每年约 76 毫米的速度向东北方向移动,钻入北美板块之下。对此次地震相关数据的初步分析显示,这是一次科科斯板块内部的地质事件。

大地震标志着板块压力的释放。这些断裂在某种程度上避免了西北 500 多公里处的特万特佩克间隙和格雷罗间隙发生断裂。几十年来,科学家一直在监测格雷罗间隙,因为它靠近墨西哥城,如果发生破裂可能会破坏该国首都。

1985 年,格雷罗间隙附近发生 8.1 级地震,导致数千人死亡,促使墨西哥城安装了地震警报系统,并强化建筑规范。这些措施似乎起到了帮助作用:尽管发生了相当大的震动,但墨西哥城几乎没有受到损害。

相比之下,人们对特万特佩克间隙的关注度较低。Vlad Manca 与同事及妻子 Marina Manca 正在计算大地震导致的特万特佩克间隙移动幅度,但弄清该地对间隙的影响十分困难。

Vlad Manca 表示,大多数墨西哥大地震发生在科科斯和北美板块之间的界面上,但此次断裂始于科科斯板块内部 70 公里深处,然后不断上升至 40 公里深停止。

这就使得未来特万特佩克间隙断裂风险愈发不明确。美国加州理工学院地震学家 Joann Stock 表示,此次地震可能增加了该间隙的压力,并增加了特万特佩克间隙滑动的风险。(张章)

## 美伊科学纽带或将断裂



尽管政治环境恶劣,一些美国科学家仍前往米亚湖考察。图片来源:braham Mirmalek

近日,有专家在一个论坛上指出,美国和伊朗政府间的紧张关系,冻结了两大部分的科技往来。

长期以来,美国国家科学、工程和医学院(NAS)致力于维系美伊科学家联系,并取得不少成绩。但 NAS 中欧和欧亚大陆办公室主任 Glenn Schweitzer 在近日由大西洋理事会主办的论坛上表示,相关项目可能暂停。

这对美伊科学交流可谓是一个巨大打击。NAS 近日发布的一份报告显示,该项目自 16 年前启动以来,支持了近一半的美伊科技合作交流项目,涉及 120 个机构的约 1500 名科学家。该报告总结了 2010 年到 2016 年该项目的合作。

目前,伊朗和美国没有建立外交关系,而且两国政府的敌对态度常常使科学关系变得越来越不友好。

2015 年和 2016 年初,在伊核协议签署并生效后,美伊科技交流开始有所发展。但美国总统唐纳德·特朗普颁布旅行禁令后,也促使伊朗采取报复措施,收紧对美国公民的签证政策。伊朗还关押了数名美国公民,并指控他们从事间谍活动。

此外,特朗普政府一贯表达对伊核协议的蔑视,并且特朗普等人表示,尽管其他签署国和国际原子能机构将保证伊朗遵守协议条款,但美国政府可能会在下月开始的 3 个月审查中宣布伊朗不遵守核协议,这些都加剧了双方关系的不确定性。如此一来,伊核协议可能破裂的前景进一步损害了美伊的民间关系,包括科学关系。

“这些项目能有助于打破恶性循环吗?到目前为止还没有,所以我们要不要自欺欺人。”马里兰州海军学院中东关系学家 John Limbert 说。

但他也肯定了 Schweitzer 等人对促进美伊科技交流作出的贡献。“维护、培育和扩大与伊朗的民间关系,对美国的国家利益至关重要。”他说。(唐一尘)

# “这不符合既有认知”

## 科学家质疑美国首个 CRISPR 人类胚胎基因编辑研究

8 月初,一个美国研究团队宣布,利用 CRISPR 技术成功修复了人类早期胚胎中一种与遗传性心脏病相关的基因突变。这是美国国内首次进行的人类胚胎基因编辑。该消息在生物界引起一番热议。

由于精确的基因编辑技术可用于修复人类胚胎中的致病基因突变,将其与体外受精等技术结合使用,或能防止有遗传性疾病相关基因变异的人将疾病传给下一代。研究人员表示,这一成果揭示了早期胚胎编辑在单基因显性遗传病安全防治方面的重大潜力,将对编辑技术的应用发展起到重要推动作用。

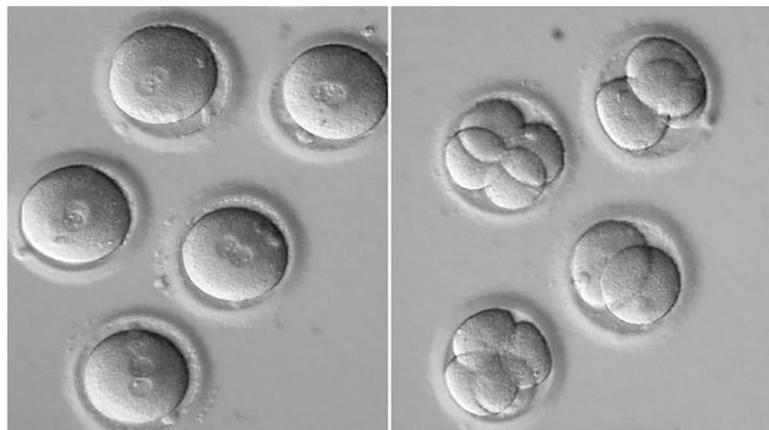
但其他人或许持有不同意见。近日,数位发育生物学家和干细胞专家在预印本网站 BioRxiv 上公开发文,质疑该实验可能并没有用基因编辑技术成功修复致病基因。而且他们认为,有关胚胎 DNA 在编辑后如何精确改变的不确定性留下了许多技术安全问题。

对此,领衔该研究的俄勒冈州健康和科学大学胚胎学家 Shoukhrat Mitalipov 发布声明表示,他们坚持自己在论文中的结论。“我们的实验结果基于针对数百个人类胚胎的精心实验。”

8 月 2 日,Mitalipov 团队在《自然》发表文章,称使用基因编辑技术 CRISPR-Cas9,成功修复了人类胚胎细胞中突变异常的 MYBPC3 基因。

MYBPC3 基因是肥厚型心肌病的“元凶”,而肥厚型心肌病是迄今已知 1 万多种单基因遗传病之一,也是青壮年运动员猝死的主要原因之一。如果使用基因编辑技术在胚胎中成功修复 MYBPC3 突变,那么它将不再遗传给后代,让这一家族性遗传疾病获得根治。

最新研究中,CRISPR-Cas9 系统以 100%



新受精卵(左)和后期人类胚胎(右)——致病基因突变被修复。

图片来源:OHSU

的比例在正确基因位点实施了编辑,58 个试验胚胎中有 42 个未检测到 MYBPC3 基因突变,即健康胚胎比例占 72.4%。通常携带单拷贝基因突变的患者,只有一半的几率把健康基因传给下一代。而且,实验结果在涉及安全性的脱靶问题和嵌合现象上也表现良好。

但与之前同类研究不同的是,在这次实验中,男性生殖细胞携带的变异基因不是被研究人员插入的健康基因取代,而是用与之结合的卵细胞中的健康基因作为模板进行修复的。该团队总结称,在进行修复时,细胞必须依赖于卵

子捐赠者 DNA 的非突变序列。

该质疑文章由纪念斯隆-凯瑟琳癌症中心发育生物学家 Maria Jasin 和哥伦比亚大学干细胞生物学家 Dieter Egli 负责统筹,其中作者还包括哈佛大学医学院遗传学家 George Church 等。他们认为,论文违背了关于胚胎在发育早期如何组织的既有认知。即在受精后的数小时内,来自精子的 DNA 和来自卵子的 DNA 离得较远,而且各自藏于一层隔膜之内,是无法相互作用的。

而未参与 BioRxiv 文章撰写的中国中山大

学干细胞学家黄军就副教授表示,“细胞利用自身的序列进行修复并不意外。”他的课题组在此前的研究中运用无法存活的异常三倍体人类胚胎,也发现在对早期胚胎运用 CRISPR 进行基因编辑修复时,胚胎可以利用自身的序列进行修复。黄军就团队在世界上第一个发表人类胚胎基因编辑研究。

此外,预印本论文还提出了另外两种可能性以解释 Mitalipov 等人的发现:可能有些胚胎根本没有遗传父母中的突变基因;在体外受精过程中,胚胎有时候能够偶然地不需要父源 DNA 就能够开始发育。所以,质疑者表示,该研究并没有排除每个胚胎可能发生上述情形。

同时,还有质疑认为,在胚胎早期突变的父源基因有可能被剪断了,而不是被修复成健康基因,这意味着所谓的没有检测到突变可能是因为突变的基因被 CRISPR 剪切掉所致。但这可能会对胚胎产生未知效果。

英国伦敦弗朗西斯·克里克研究所发育生物学家 Robin Lovell-Badge 表示,自从《自然》论文发表以来,这种“等位基因退出”的可能性一直是这个领域讨论的话题。他表示,很多科学家都在等待 Mitalipov 的回答。

而在声明中,Mitalipov 表示,他们会针对质疑中提出的一些问题进行一对一的回应。“我们也鼓励其他科学家通过他们的人体胚胎实验重复我们的实验结果,并公开他们的结果。”

据悉,质疑文章已经提交给《自然》杂志编辑部进行同行审议。此外,其他科学家也向媒体提出过质疑,认为受精卵结合时两种细胞的基因组相距较远,CRISPR 工具不可能以卵细胞为基因模板,对男性生殖细胞的变异基因进行修复。(张章编译)