



图片来源:NIH哈钦森美国国家标准与技术研究院

走近最难以捉摸的染色体

◎本报记者 张佳欣

为了揭开Y染色体的“神秘面纱”，国际科研团队刚刚完成了对完整人类Y染色体的首次测序。

Y染色体是人类46条染色体中最小的染色体之一，但却是科学家最难读取的染色体。此次研究掌握的这条独特染色体的全部基因“密码”，至少可以帮助解开部分关于Y染色体的谜团。

根据Y染色体寻找祖源

人们普遍认为，Y染色体“形如其名”，但事实上，它的形状并不像字母Y。X染色体首先被发现并被命名为“X”，大约15年后，Y染色体被发现，研究人员只是沿用了字母表中X的后一位为其命名。

深圳华大生命科学研究院研究员周旻告诉科技日报记者，人类的Y染色体是雄性中性染色体对的其中一条。虽然X和Y染色体都来源于一对相同的祖先常染色体，但是目前的Y染色体在长度上只有X染色体的三分之一，蛋白编码基因也只剩下X染色体的约5%。

Y染色体上具有哺乳动物的性别决定基因SR Y，它能触发睾丸的生长，并由此决定雄性性状。此外，Y染色体上还有其他一些参与精子形成，甚至神经系统发育等过程的基因。

Y染色体的遗传方式非常独特，包括X染色体在内的所有其他人类染色体在繁殖过程中会重新组合并共享遗传物质，但Y染色体会直接从父辈传递给子代，几乎不会重新组合。基于这种直接遗传方式，能够完全测序不同人的Y染色体将有助于更好地理解基因是如何在世代间传承和变化的。

目前，国外已经有许多热门的基因检测公司，如23andMe和FamilyTreeDNA，可以提供关于Y染色体的特定报告。Y染色体独特的遗传方式使得这类公司能够判定某个人处于何类“单倍群”，处于同一“单倍群”的人的DNA有共同的突变，表明他们可能有着共同的祖先。研究人员已经能够追踪许多“单倍群”的地理起源，让人们对自己的祖源有所了解。

科学认识“超雄基因”

我们的祖先遗传给我们的DNA决定了我们的外貌、身体机能，甚至能影响我们的性格。在一些情况下，人体可能会出现染色体异常，其中一种是Y染色体的异常。

此前，有新闻媒体报道称，孕妇在孕检中查出胎儿有“超雄基因”。所谓的“超雄基因”也称超雄综合征，又称XXY染色体综合征，它指的是男孩出生时体内多了一条Y染色体。其本质就是一种染色体异常导致的疾病。

周旻指出，“超雄基因”形成的原因是，父亲产生精子时，在减数第二次分裂后期，两条Y染色体未移向细胞两极，使个别精子细胞含有两条Y染色体。这样的精子与正常的卵子（携带一个X染色体）结合形成受精卵后就发育成了携带XXY染色体的个体。

深圳华大生命科学研究院副院长金鑫向记者介绍说：“除了‘超雄基因’外，还存在其他染色体异常，其中一些与染色体数目的变异相关。例如，我们熟知的21三体综合征（唐氏综合征）就是由第21号染色体数目异常导致的。染色体异常可能会导致个体出现一些疾病表现或特殊情况。”

据国外物理治疗百科全书网站verywellhealth介绍，大多数出生时患有XXY染色体综合征的人智力一般，可能会出现发育迟缓或学习障碍等情况。

金鑫认为，从遗传学角度来看，应该把XXY染色体综合征看作一种遗传病。作为一种疾病，它可能表现出一些特定的外部特征，例如21三体综合征可能表现为智力发育滞后、生长问题、器官疾病等，“超雄基因”携带者也可能会有一些特殊的表现，这些特点可能是疾病本身导致的。然而，需要更多的研究来确定这些特点的可能原因，以及是否有相关的干预方法。

金鑫补充说，Y染色体完整序列的绘制对研究这些机理和进行干预具有重要意义。它能够帮助我们更好地了解染色体异常以及与之相关的生理和心理特点，从而能更好地对症下药。

Y染色体不会真的消失

2011年，澳大利亚科学家珍妮弗·格雷夫斯表示，按照现在的速率，估计Y染色体将在几百万年内消失。这一说法在学术界掀起了轩然大波。

金鑫告诉记者，现在有几个不同的新发现，表明这一认知不一定是完全正确的。首先，通过比较人类与黑猩猩的基因组，科学家发现，自人类与黑猩猩的共同祖先分离后，人类的Y染色体在漫长的时间里没有发生新的基因丢失，特别是与生存和繁殖密切相关的基因，能相对稳定地传递给下一代。

周旻说：“近期的研究发现，Y染色体的退化速率正在逐渐降低。以人为例，人的Y染色体上的基因虽然相较于X染色体发生了接近95%的丢失，但在最近的2500万年内其实并没有发生进一步的基因丢失。”

其次，Y染色体上基因的功能和独特的“回文”序列也能有效避免Y染色体真正消失。

周旻介绍，目前仍然保留在Y染色体上的基因，除了与睾丸发育、精子形成的过程有关外，还参与到一些更广泛的细胞生物学过程（如泛素化、转录翻译起始等）中。相信这些重要的功能在一定程度上阻止了Y染色体的进一步退化。虽然Y染色体在演化过程中无法像常染色体一样和它的同源染色体X进行染色体对之间的同源重组，但是Y染色体上存在“回文”序列，能使基因在“回文”序列间进行同源重组，可以在一定程度上解决有害突变积累的问题。

当然，其实自然界中也确实存在Y染色体完全丢失的物种，如奄美刺鼠。周旻说：“这个物种的Y染色体已经完全丢失，但其实它们仍然存在雄性和雌性两种性别。科学家在这个物种的雄性个体中发现，其3号染色体的Sox9基因的上游产生了一个雄性特异的增强子调控元件，可能可以在SR Y缺失的情况下促进Sox9的表达，进而促进睾丸的形成。”



图为人类(X)和Y染色体。图片来源:英国《新科学家》杂志网站

编者按 8月23日，两篇发表在《自然》杂志的论文宣告：科学家组装了人类Y染色体的第一个完整序列，补齐了人类基因组这部“生命天书”。多年以来，“生命天书”初稿被科学家反复打磨、拾遗补阙。等待了20多年后，Y染色体的完整“章节”姗姗来迟。而这还只是一个开始，未来，有关遗传、生育、疾病、进化的诸多密码有望被一一破解。

科学家首次完成对人类Y染色体的完整测序 最后一块基因“拼图”揭开哪些生命谜团

◎本报记者 张佳欣

好消息传来。美国国家人类基因组研究所领导美国约翰斯·霍普金斯大学、加州大学圣克鲁斯分校等多家机构的科学家组装了人类Y染色体的第一个完整序列。这是最后一个被完全测序的人类染色体，新序列填补了Y染色体长度的一半以上的空白。两篇相关论文于8月23日发表在《自然》杂志上。

几十年来，Y染色体作为人类两性染色体之一，由于其结构的复杂性，一直是基因组学界难啃的“硬骨头”。现在，这一难以捉摸的基因组区域已经被完整测序，这意味着，人类泛基因组的最后一块“拼图”被补上了。

这次测序为当前人类参考基因组增加了逾3000万个碱基对，揭示了多个基因家族的完整结构，并确认了41个新的蛋白质编码基因，为研究生殖、进化和人口变化相关的重要问题提供了关键信息。

Y染色体测序困难重重

深圳华大生命科学研究院副院长金鑫在接受科技日报记者采访时说：“人类基因组测序是理解人类疾病、健康和进化的关键。通过将个体基因序列与参考基因组进行比较，科学家可以找到与疾病相关的基因，进而为个体提供更精准的诊断和治疗方案。”

深圳华大生命科学研究院研究员周旻介绍说，完整人类基因组中约有30亿个碱基对，科学家其实很难通过直接测序把序列测通，所以在测序的时候，需要将大片的DNA进行随机打断再进行测序。对于常染色体是如此，对于X和Y染色体亦然。所以在测序结束后，科学家需要通过测序数据进行拼接来得到完整真实的基因组序列。

1999年，人类首次成功完成对人体染色体完整基因序列的测定。科学家使用“短读长”测序技术来做到这一点，将DNA切成大约一百个碱基的小片段，然后像拼图一样重新组装它们。

2022年，最新版人类参考基因组(T2T-CHM13)问世。该基因组包含22条常染色体和X染色体的无间隙序列，共有30.55亿对碱基，比之前的人类参考基因组(GRCh38)增加了近2亿碱基对的遗传信息，约等于整个基因组8%的序列信息，达到了前所未有的完整程度。然而，人类参考基因组中的Y染色体，仍有约一半序列是缺失的。

Y染色体的测序为什么如此棘手？周旻说：“这些未知区域主要集中在异染色质区域。该区域主要由卫星序列组成。要通过测序数据确定其序列是一件非常困难的事情。”

卫星序列通常属于重复序列，是一种在基因组中广泛存在的特殊DNA序列重复单元。金鑫表示，由于Y染色体包含大量重复或“回文”序列，共有多达百万个碱基对。而“短读长”技术无法跨越这么长的重复序列，从而无法确定正确的序列顺序，也就无法绘制出完整序列。

组装测序数据就像尝试阅读一本被切成条状的长书，每一条都是书中的一句话。如果书中的所有句子都是唯

一的，那么就更容易确定句子的顺序。但是，如果同一句话重复了数千或数百万次，那么这些句子的原始顺序就没有那么清楚了。Y染色体上约有3000万个碱基字母是重复序列，就好像这本书的一半篇幅都重复着同样的几句话。

技术“组合拳”助力测序

如果把染色体比作一本书，那么被称为“短读长”的DNA测序技术一次只能读取相对较短的序列，但现在，“长读长”DNA测序技术可在不影响准确性的情况下生成更长的DNA序列读数，甚至可一次阅读整个“句子”或“段落”。这样，研究人员能够更准确地判断“句子”“段落”之间的排列顺序，从而更有助于完成完整的测序。

得益于“长读长”DNA测序技术和创新的序列组装方法，科学家们终于实现了对Y染色体的无间隙读取。

周旻说：“科学家利用‘长读长’DNA测序技术，一定程度上降低了组装Y染色体的难度。另外，科学家也针对新技术的测序数据开发了新的组装算法进行自动化组装，提升了组装效率。”

今年2月，美国国立卫生研究院研究人员发布了一种创新的软件工具Verkko，用于组装来自各种物种的真正完整的基因组序列。研究人员称，有了Verkko，研究人员现在只需按一下按钮就能自动获取完整的基因组序列。

英国《新科学家》杂志网站8月23日报道，研究人员还运用了英国牛津纳米孔公司所开发的纳米孔(Nanopore)测序技术。当单个DNA分子穿过一个纳米孔时，就能读取该分子的序列，产生数百万个DNA碱基字母的片段，而不仅仅是几百个。

在本次研究中，得益于上述方法，团队能够应对Y染色体测序中的种种挑战。

Y染色体上的新发现

此次发表的一篇论文称，Y染色体包含62460029个碱基对序列。该数字比参考基因组GRCh38里的Y染色体数据多了3000万个碱基对，标志着测序完整性的显著进步。

通过改进GRCh38中的错误并阐明DAZ和RBM Y等基因家族的结构(这两个基因家族都有助于精子的产生)，该团队增进了对Y染色体遗传结构的理解。

此外，这条完整的Y染色体有106个蛋白质编码基因，比参考基因组多了41个。但几乎所有这些额外的基因都只是一种被称为TSPY的基因的副本。他们还辨识出以前被误解为细菌DNA的序列，这些序列其实是之前未知的人类Y染色体序列。

在另一篇论文中，包括美国杰克逊基因组医学实验室在内的联合团队组装了代表世界21个不同人群的43名男性的Y染色体。这些组装结果更详细地阐释了Y染色体在人类演化历史中的遗传差异。研究结果揭示了新的DNA序列、保守区域的特征，以及促进Y染色体复杂结构的分子机制。

在接受科技日报记者采访时表示，之前的一些研究将Y染色体DNA误认为是细菌DNA，部分原因可能在于样本来源和污染控制。细菌DNA的提取通常来自人类皮肤、肠道等样本，而在提取过程中可能会混入人类的DNA。在没有足够参照序列进行比较的情况下，很难准确区分细菌DNA和其他DNA。

美国加州大学圣克鲁斯分校在官网发布的信息称，由于细菌DNA通常取自人类皮肤，因此，在细菌的基因组样本中，或许会不可避免地混入对研究造成污染的人类DNA。在研究细菌DNA时，科学家借助当前的人类参考基因组来识别哪些序列来自人类，从而删除这些序列，只留下细菌DNA进行研究。

但是，又因为Y染色体此前未完成测序，是人类基因组拼图中的空白，这导致在

分析时，研究人员可能没有足够的参照标准来进行比对和确认，难以确定某些序列的来源，从而误以为这部分DNA是他们正在研究的细菌DNA的一部分。

将推动人类对细菌基因组的研究

“细菌DNA和人类DNA都由一系列核苷酸碱基组成，包括腺嘌呤(A)、胸腺嘧啶(T)、胞嘧啶(C)和鸟嘌呤(G)。绝大多数生物的基因组都是由这些碱基按照不同的顺序组成。细菌的基因组同样由这些碱基组成，即使在细菌中也存在着与人类DNA相似的碱基序列。”金鑫介绍说，“然而，细菌DNA与人类DNA在整体上存在差异。细菌是单细胞微生物，其基因组相对较小，只含有少

量基因，而人类拥有复杂得多的基因组，包含大量的基因。”

加州大学圣克鲁斯分校官网公布的信息称，此次将细菌DNA和Y染色体DNA区分开来的新发现有望改善细菌基因组的相关研究。这篇论文指出，有证据表明，在一个公共数据库中，大约5000个细菌基因组可能含有与人类Y染色体基因序列匹配的污染部分。

今后，科学家可以使用更新后的人类参考基因组来研究这些细菌物种的基因组。他们能够正确地删除细菌基因组中删除所有人类污染，包括Y染色体DNA，从而对细菌基因组有更清晰的了解。

而更好地理解细菌DNA可能会对医学产生重大影响，因为人体充满了微生物。目前，细菌的基因序列已经在从肠道健康到癌症的各种研究中得到应用。

Y染色体完整测序或改善细菌DNA研究

◎本报记者 张佳欣

8月23日，美国国家人类基因组研究所在一篇发表在《自然》杂志的论文中表示，美国研究人员领衔的团队首次发布了完整的人类Y染色体基因序列，这是最后一个被完全测序的人类染色体。在此次研究中，一个意想不到的有趣发现是，在过去的研究中，Y染色体DNA曾多次被错误地认为是细菌DNA。如今，科学家终于能够将二者区分开来。

那么，细菌DNA和人类DNA有何不同？为什么二者会被混淆呢？

Y染色体基因序列缺失带来的“误会”

深圳华大生命科学研究院副院长金鑫