

What role do telomeres and centromeres play in genome function?

端粒与着丝粒——染色体上的高度重复序列区域

樊起傅^{1,2}, 付钰^{1,2*}

1. 微生物资源前期开发国家重点实验室, 中国科学院微生物研究所, 北京 100101;

2. 中国科学院大学存济医学院, 北京 100049

* 联系人, E-mail: fuyu@im.ac.cn

2017-02-15 收稿, 2017-05-19 修回, 2017-05-19 接受, 2017-08-14 网络版发表

国家自然科学基金(31371264, 31571288)资助

摘要 端粒与着丝粒是真核生物细胞染色体上的两个特殊结构, 对于维持基因组的稳定具有极其重要的作用. 端粒位于染色体的末端, 保证染色体的独立性和稳定性. 着丝粒与染色体的分离有关, 确保染色体在细胞分裂时被正确地平均分配到两个子细胞中. 端粒和着丝粒都由高度重复DNA序列组成, 并且都与多种蛋白质相结合形成复杂的结构. 但是由于端粒和着丝粒的结构特点, 尤其是着丝粒, 可为上百万碱基对的高度重复序列, 测序存在较大的困难, 所以对它们的功能研究依然有很多未知的领域等待探索. 本文简要综述了端粒与着丝粒的发现过程及其结构特征, 并且介绍了端粒、着丝粒异常与疾病发生之间的特殊关系. 因为端粒的长度维系主要通过端粒酶的合成作用, 重点介绍了端粒酶的结构和功能以及目前端粒酶研究的进展. 最后针对端粒和着丝粒的热点研究进行了总结和展望. 相信随着测序技术的发展, 对端粒和着丝粒的认识会更加深入, 为肿瘤治疗和衰老研究打下基础.

关键词 端粒, 端粒酶, 着丝粒, DNA 重复序列

真核生物的基因组DNA与核小体蛋白相结合, 高度压缩形成染色体. 由于染色体是遗传物质的载体, 所以染色体必须在细胞世代中具有稳定性和正确的传递性, 从而保证遗传信息准确无误地传递给子代. 研究表明, 染色体上的端粒、着丝粒以及DNA复制起始位点, 是维持染色体完整性、稳定性和传递性的三大要素, 在生命活动中行使重要功能. 端粒和着丝粒均为真核生物染色体上存在的由DNA重复序列和大量蛋白共同组成的特殊结构. DNA重复序列是指基因组上具有多拷贝的片段, 依据序列在染色体上的分布位置和序列特点, 重复序列可分为三类: 散在重复(interspersed repeats), 一般来源于转座元件; 片段重复(segmental duplications), 为拷贝数较低且序列不完全相同的重复序列; 串联重复(tandem repeats), 这一类重复序列通常连续排列在基因组上, 端粒和着

丝粒的重复序列均属于串联重复序列^[1,2]. 本文分别对端粒和着丝粒的发现历程、结构和功能, 与人类疾病的关系以及研究热点和进展进行简要的介绍.

1 端粒

在染色体DNA复制的过程中, 由于DNA聚合酶无法复制滞后链5'末端在消除RNA引物后造成的空缺, 随着染色体DNA的每轮复制, 染色体会因无法完全复制而逐渐缩短, 丢失染色体所含有的遗传信息. 为了避免这种情况发生, 真核生物进化出了一种位于染色体上的特殊结构来解决这个问题.

在20世纪30年代, Muller^[3]和McClintock^[4]发现, 染色体的末端存在一种特殊的结构可以阻止染色体互相连接在一起, 这种染色体末端的结构被称作为端粒. 1978年, Blackburn和Gall^[5]在对四膜虫染色

引用格式: 樊起傅, 付钰. 端粒与着丝粒——染色体上的高度重复序列区域. 科学通报, 2017, 62: 3245–3255

Fan Q F, Fu Y. Telomere and centromere—DNA tandem arrays on the chromosome (in Chinese). Chin Sci Bull, 2017, 62: 3245–3255, doi: 10.1360/N972016-01145

体端粒的研究中发现了端粒含有高重复序列5'-CCCCAA-3'。随后1982年,在Szostak与Blackburn^[6]合作研究中,首次证明了端粒对染色体具有保护作用。1988年, Moyzis等人^[7]发现,人类端粒DNA重复序列为TTAGG,说明端粒序列具有种属特异性(表1)。

1.1 端粒的结构特征

端粒的长度在不同的物种中也不尽相同,在酵母中端粒的长度约为300 bp,人类生殖细胞的端粒长度约为15 kb,而烟草细胞的端粒长度可以达到160 kb。进一步的研究发现,端粒的末端可以弯曲形成T-loop结构。端粒末端300 bp的单链DNA区域,可弯曲打折与端粒的DNA双链共同形成D-loop结构(图1)^[9,10]。同时,由于端粒为富含鸟嘌呤的重复序列构成,因此端粒DNA也常形成G四联体等高级结构^[11],以保护端粒免受核酸酶的降解。端粒结合蛋白复合体Shelterin与DNA结合也稳定了这种结构。Shelterin蛋白复合体由TRF1, TRF2, RAP1, POT1, TPP1和TIN2组成^[12]。其中,TRF1, TRF2不仅直接与DNA结合,并且通过与其他参与细胞周期、DNA损伤修复调控蛋白的结合来维持端粒的结构和长度^[13,14]。与TRF1和TRF2不同, POT1结合在单链DNA上,保护端粒末端。RAP1, TIN2和TPP1与这些蛋白结合形成端粒结合蛋白复合体,共同维持端粒结构,调节端粒长度^[12]。除了

Shelterin蛋白复合体,还有很多其他的端粒结合蛋白,如CST蛋白复合体、Myb类似蛋白、OB折叠蛋白等^[15]。这些蛋白直接或者间接与端粒结合,形成“端粒帽子”的结构,保护端粒不被核酸酶降解或者避免端粒被误认为DNA双链断裂位点而进行DNA损伤修复,从而维持和调节端粒的长度。

1.2 端粒酶的结构与功能

端粒的合成不是由DNA聚合酶来完成,而是由一种特殊的酶——端粒酶来完成。Greider和Blackburn^[16]在1984年发现并纯化了负责合成端粒的端粒酶。端粒酶由3个部分组成:端粒酶RNA (TR)、端粒酶逆转录酶(TERT)和催化亚单位^[17-19]。以人端粒酶为例,hTR区包括假结、模板和模板边界成分(template boundary elementary)^[20],是端粒合成的RNA模板,具有与人端粒序列TTAGGG互补的RNA序列5'-CUAACCCUAA-3'^[21]。hTERT区则是催化合成端粒DNA的逆转录酶^[22],是端粒酶的必需结构。端粒酶延伸端粒机制如图2所示,3'末端单链DNA与hTR序列通过碱基互补配对,同时与hTERT通过相互作用结合;hTERT将单个核苷酸催化连接至DNA的3'端;在完成一个重复序列的添加后,DNA与端粒酶分离,结束端粒合成,或者端粒酶发生易位,使得DNA 3'单链末端与hTR序列继续结合,开启新一轮的端粒重复

表1 一些代表真核生物的端粒重复序列(数据来源于Telomerase Database)

Table 1 The DNA tandem repeat sequences in some typical eukaryotic telomeres (Data from Telomerase Database)

生物种类	端粒重复序列(5'→3')	
脊椎动物	Human, mouse, <i>Xenopus</i>	TTAGGG
高等植物	<i>Arabidopsis thaliana</i>	TTTAGGG
	<i>Nicotiana tabacum</i>	TTAGGG
藻类	<i>Chlamydomonas reinhardtii</i>	TTTTAGGG
	<i>Cyanidioschyzon merolae</i>	AATGGGGGG
昆虫	<i>Bombyx mori</i>	TTAGG
	<i>Spondylis buprestoides</i>	TTAGG
丝状真菌	<i>Neurospora crassa</i>	TTAGGG
	<i>Aspergillus fumigatus</i>	TTAGGG
裂殖酵母	<i>Schizosaccharomyces pombe</i>	G ₂₋₈ TTAC(A)
出芽酵母	<i>Saccharomyces cerevisiae</i>	T(G) ₂₋₃ (TG) ₁₋₆
	<i>Debaryomyces hansenii</i>	ATGTTGAGGTGTAGGG
	<i>Candida albicans</i>	GGTGTACGGATGTCTAACTTCTT
	<i>Kluyveromyces lactis</i>	GGTGTACGGATTTGATTAGGTATGT

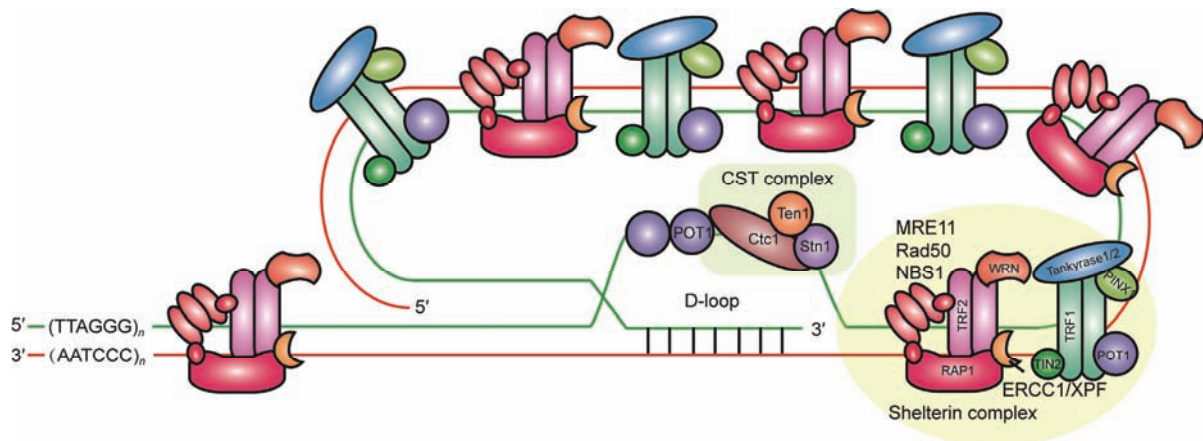


图 1 (网络版彩色)端粒的T-loop结构示意图. 人的端粒由2~30 kb的TTAGGG重复序列组成, 端粒的3'末端为100~200核苷酸的单链DNA, 这段单链DNA侵入端粒重复序列的双链区, 与双链区的一条DNA链配对形成一个置换环(displacement loop, D-loop), 因此整个端粒的DNA呈现为一个被称为T-loop的结构. 端粒的双链DNA和单链DNA都被各种蛋白结合和保护, POT1 蛋白(protection of telomere-1)和CST蛋白复合体(Ctc1/Stn1/Ten1 complex)结合于端粒的单链DNA上; TRF1 和TRF2 (telomere repeat factor-1, -2)结合于端粒的双链重复序列上, 并与各种蛋白相互作用形成Shelterin等复杂蛋白复合体(改编自文献[8])

Figure 1 (Color online) The T-loop structure of telomere. Human telomeres consist of 2–30 kb TTAGGG repeats. The 3' end of the telomere is a 100–200 nt single-stranded DNA that invades the double-stranded DNA region to form a displacement D-loop, thus the entire telomere exists as a T-loop structure. Both double-stranded and single-stranded DNA of telomeres are bound and protected by various proteins. POT1 (protection of telomere-1) and CST complex (Ctc1/Stn1/Ten1 complex) bind to the single-stranded DNA; TRF1 and TRF2 (telomere repeat factor-1, -2) bind to the double-stranded DNA, and interact with other proteins to form protein complexes, such as Shelterin (Adapted from ref. [8])

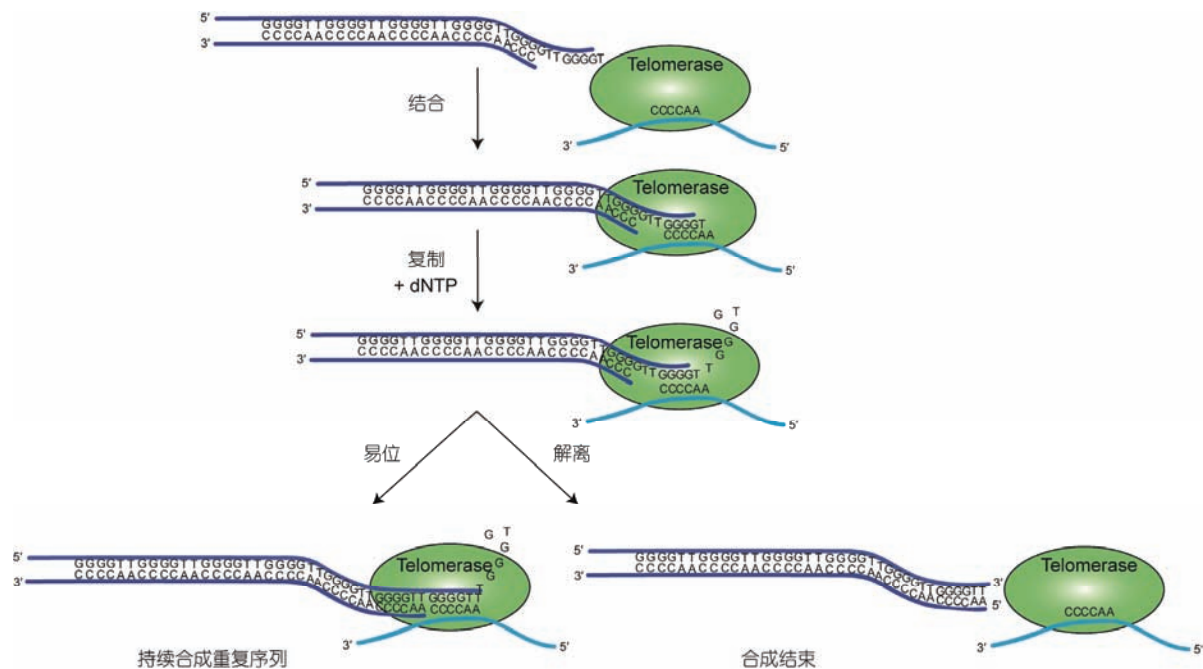


图 2 (网络版彩色)端粒延伸机制^[19]

Figure 2 (Color online) Schematic of the mechanism how telomerase elongates telomere ends

序列合成.

1.3 端粒与衰老以及各种疾病的关系

由于DNA复制的半不连续性、核酸酶以及氧自

由基等因素的影响, 当细胞分裂一次, 染色体的端粒会缩短大概25~200个碱基; 当端粒缩短到一定程度时, 细胞终止复制, 不再分裂. Olovnikov^[23]首次将端粒缩短与细胞衰老联系起来, 认为端粒缩短到一定

程度将终止细胞复制。有研究表明,激活细胞中的端粒酶,可以延长端粒,减缓细胞衰老的速度^[24]。在端粒酶缺陷型小鼠(*Mus musculus*)体内恢复端粒酶的活性,增加端粒的长度可以成功逆转小鼠表现出衰老症状^[25]。针对人类早衰疾病的研究也发现,端粒的异常与早衰症状息息相关^[26~28]。沃纳综合征(Werner syndrome)是一种成人早衰综合征,表现在青春期发育停滞,出现皮肤损害、心血管病变等衰老症状。研究发现,当WRN基因突变后会导致典型的沃纳综合征^[29]。该基因编码的WRN蛋白可与TRF1, TRF2以及POT1等Shelterin复合体组成蛋白相结合,参与到端粒结构的稳定和长度的维持^[30,31]。人们在WRN缺失的细胞中观察到了端粒的丢失,并且这种缺陷可以通过外加端粒酶缓解。研究表明,在WRN基因缺失的成纤维细胞中过表达端粒酶可以缓解细胞的早衰症状^[32]。通常情况下,端粒的缩短与衰老程度呈正相关,但是端粒长短与衰老的关系远不止想象的这么直接、简单。研究表明,在哺乳动物中,体型较小、寿命较短的动物拥有更长的端粒以及更高的端粒酶活性^[33,34]。在对一些长寿鸟类的研究中发现,虽然从幼鸟到成年鸟的过程中端粒长度会缩短,但是在成年鸟中端粒并不随着年龄的增长而缩短^[34,35]。

虽然端粒的增长可以抵抗细胞衰老,但是端粒的异常增长及端粒酶的异常激活会诱发细胞的无限增殖,导致肿瘤的发生^[36]。通常恶性肿瘤细胞具有较高活性的端粒酶,从而端粒的长度能维持在一个相对稳定的长度。因此,端粒酶的激活被认为是肿瘤发病的标志。由于在正常的人体体细胞中,端粒酶活性很低,所以,在对肿瘤的治疗中,采用抑制端粒酶活性的疗法可以达到一定的疗效。目前已有将能够抑制端粒酶活性的hTR序列的互补序列用于临床治疗癌症的例子^[37]。同端粒和衰老的关系一样,端粒和肿瘤的关系也十分复杂。Greider课题组^[38]将小鼠生殖细胞的端粒酶RNA组分(*mTR*)敲除,使之丧失端粒合成能力,发现这些端粒酶功能缺失的细胞具有染色体末端融合的特征,并且转变为肿瘤细胞。另外一个研究指出,染色体不同位点的端粒增长或者缩短均会引起癌症,并且和癌症发生的部位和种类密切相关^[39]。因此,过短和过长的端粒可以作为不同肿瘤发生风险的评价因子。

除了癌症,端粒还同骨髓衰竭、肝脏疾病、糖尿病、皮肤及肺部疾病相关。但是很多研究目前还只是

停留在相关性分析阶段,关于端粒异常引起相关疾病的机制还需要更深入的探索。

1.4 当前端粒研究的问题及方向

端粒作为染色体中重要的结构参与细胞命运的调控,但是其在细胞生命事件的具体作用现在还不清楚,尤其是端粒长度变化如何决定细胞命运的分子细节还需要进一步探索分析。

端粒中的重复序列虽然DNA序列一致,但是不同重复序列之间可能会有表观遗传上的差异,从而决定了每个重复序列的独特性,这将有待进一步发现和了解。端粒的长度在细胞内是一个动态的变化,受到诸多因素的调节,如端粒酶的活性、压力知觉、慢性应激等^[40]。由于端粒是富含鸟嘌呤的高度重复序列,对细胞内端粒长度测量有比较大的难度。目前常用端粒长度测量的实验方法有TRF (terminal restriction fragment)分析、定量PCR (quantitative polymerase chain reaction, qPCR)、定量FISH (quantitative fluorescence *in situ* hybridization, qFISH), STELA (single telomere length analysis)等。近来基于全基因组测序的数据,端粒的长度可以通过对端粒重复序列读数的计算来分析端粒的长度。随着测序技术和各种组学研究技术的发展,端粒功能的研究需要整合基因组、转录组、蛋白组、代谢组以及表观修饰组等高通量的组学数据,动态分析比较端粒长度与各种生理状态的联系,从而帮助理解端粒在决定细胞命运中所起的具体作用。另外,研究发现,端粒的DNA虽然不能编码蛋白,但是可以转录RNA,并且转录的RNA在细胞发育和分化过程中对于维持端粒的长度起了很重要的作用^[41]。这些有意思的现象和内在的机理都将会是未来研究的热点。

端粒的长度维系主要通过端粒酶的合成作用,因此深入研究端粒酶的作用细节对于理解端粒的功能有重大的意义。由于端粒酶表达量低、纯化困难,传统研究方法已经不能满足对端粒酶研究的需要。目前研究者主要采用单分子荧光共振能量转移技术(single molecule fluorescence resonance energy transfer, smFRET)、荧光双色同步响应检测技术(two-color coincidence detection, TCCD)等单分子荧光技术对端粒酶进行实时定量的检测。通过smFRET技术,有研究发现,人端粒内G四联体存在两种稳定的折叠结构^[42]。Wu等人^[43]则利用smFRET确定了一种端粒酶

的活性构象。另一种单分子检测方法——荧光双色同步响应检测技术，利用部分重合的共焦区将两种不同发射波长的荧光分子分离，再对各个荧光分子进行检测，提高信噪比^[20]。利用该方法可以灵敏检测到端粒酶的 κ_{trans} ，从而得到端粒酶的持续合成能力^[44]。另外，端粒与各种端粒结合蛋白相互作用对于端粒功能有极其重要的影响，通过单分子技术研究这些蛋白与端粒的动态作用细节，也将会协助深入理解端粒的生物学功能。Alves等人^[45]就利用TCCD法标记了端粒酶的各个组分，得到人端粒酶核糖核蛋白复合物中hTERT, hTR以及DNA是等比组成的。

2 着丝粒

与端粒相似，着丝粒也是真核生物细胞染色体上的特殊结构，德国生物学家Walter Flemming在19世纪80年代第一次在显微镜下观察到了染色体上的主缢痕(primary constriction)——着丝粒(centromere)。着丝粒位于染色体的异染色质区，富含短DNA重复序列，在不同的染色体上具体位置不尽相同。在真核细胞分裂时，着丝粒作为姐妹染色体(sister chromatids)的连接区域，是动粒(kinetochores)组建的地方，提供微管与染色体连接的位点，保证复制完成的染色体平均分至两个子代细胞中，确保子代细胞拥有同等正确数目的染色体。如果着丝粒的功能丧失，将会导致姐妹染色体无法分开和染色体在子细胞中错误分配。对于多数真核生物的染色体，通常只有一个着丝粒(monocentromere)，着丝粒的位置并非固定在染色体的中心，着丝粒可以分布在染色体上的各个位置，

包括染色体的末端(图3)。有的物种则是整个染色体都可作为着丝粒(holocentricity)，微管可以与整个染色体连接(图3)。很多肿瘤细胞染色体则不具备正确的着丝粒结构，遗传物质无法均等分配^[46,47]。

2.1 着丝粒的结构

与端粒DNA序列具有相对的保守性不同，尽管所有真核生物着丝粒的生物功能完全相同，但是着丝粒的DNA序列和长度存在巨大的种间差异。从着丝粒的长短组成和特性来看，着丝粒主要分为3大类。第一类是以酿酒酵母(*Saccharomyces cerevisiae*)为代表的，可以通过保守的DNA序列辨识的点着丝粒(point centromere)，这种着丝粒含有对于执行着丝粒的功能至关重要的保守DNA序列。酿酒酵母着丝粒长度只有125 bp，包含着丝点结构域、中心结构域和配对结构域3个保守的DNA序列因子，可以招募相关蛋白结合于该序列，该DNA序列的突变将导致着丝粒功能的全部丧失(图4(a))^[49]。第二类着丝粒——区域着丝粒(regional centromere)，普遍存在于大多数真核生物中，其缺乏类似酿酒酵母着丝粒中对于执行着丝粒功能起决定性作用的DNA序列，不能利用保守的DNA序列来辨识着丝粒。区域着丝粒DNA通常为大量相似度较高的重复序列，以串联的方式排布，长度可达4 Mbp。裂殖酵母的着丝粒中央包含一个核心区域 cnt (centromere core domain)，核心区域两侧是序列有所差异的重复序列区 imr (innermost repeats)，在 imr 区域的外侧为 otr 重复序列区(outermost repeat regions)(图4(b))^[50]。人及大多数灵长类动物的着丝粒

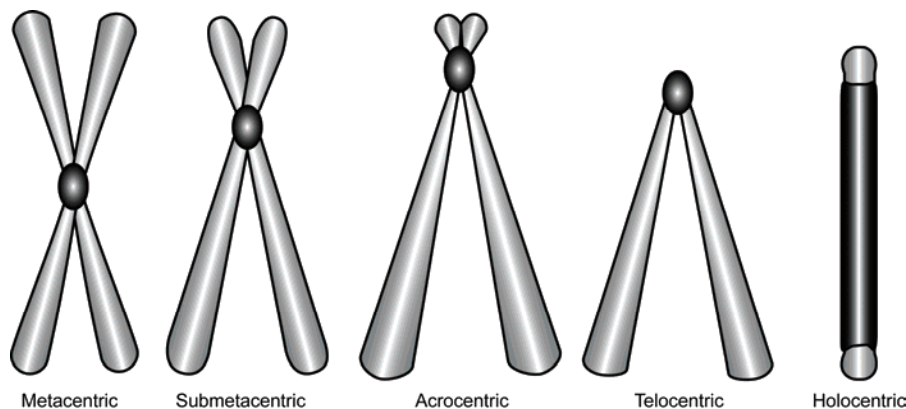


图3 着丝粒在染色体上的位置。图中着丝粒表示为深色部位，可以根据其在染色体上的位置将染色体分为如图所示的5类
Figure 3 The locations of centromere in chromosomes. The dark region represents centromere. According to the position on the chromosome, centromere can be classified into five categories shown in this figure

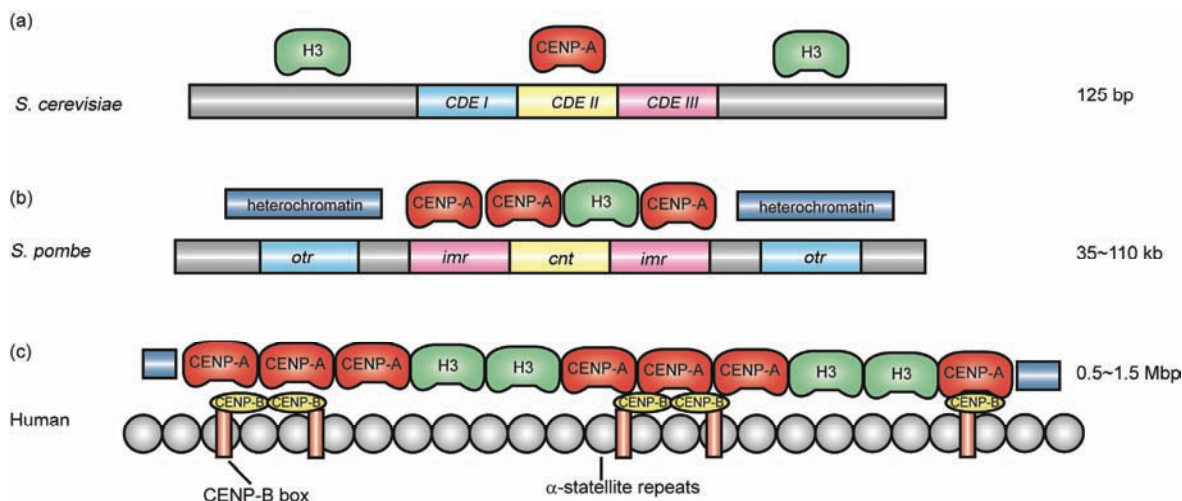


图 4 (网络版彩色)真核生物着丝粒结构. (a) 酿酒酵母的着丝粒结构. 酿酒酵母的点着丝粒长度为 125 bp, 含有 3 个保守的 DNA 序列, 一个 CENP-A (Cse4) 蛋白结合于该序列; (b) 裂殖酵母的着丝粒结构. 裂殖酵母的区域着丝粒位于异染色质中, 含有 *cnt*, *imr* 和 *otr* 区域, 多个 CENP-A 蛋白结合于 *cnt* 和 *imr* 区域; (c) 人的着丝粒结构. 人的着丝粒为很多拷贝的 α 卫星 DNA 重复序列串联而成, CENP-A 和着丝粒核心区中部分 α 卫星 DNA 序列结合, CENP-B 蛋白通常形成二聚体, 结合于 17-bp 的 CENP-B 盒子序列上(改编自文献[48])

Figure 4 (Color online) The centromeric structures among eukaryotes. (a) The structure of *Saccharomyces cerevisiae* centromere. The point centromere of *S. cerevisiae* contains three conserved DNA sequences, and protein CENP-A (Cse4) binds to the sequences; (b) the structure of fission yeast centromere. The regional centromere of fission yeast has *cnt*, *imr* and *otr* three regions, and it is located on the heterochromatin. Several CENP-A proteins bind to the *cnt* and *imr* region; (c) the structure of human centromere. The human centromere is composed of many copies of alpha satellite DNA repeats. CENP-A protein binds to the alpha satellite DNA sequence in the centromeric core region. CENP-B protein usually forms a dimer and binds to the CENP-B box (Adapted from [48])

主要由 α 卫星 DNA (alpha satellite DNA) 组成. α 卫星 DNA 是由大量相似度 50%~70% 的 171 bp 的 DNA 片段头尾串联而成, 长度为 200~5000 kb (图 4(c))^[51]. 序列研究显示, 人类染色体的着丝粒可以在 DNA 序列没有任何改变的情况下完全丧失其功能^[52], 说明区域着丝粒的 DNA 序列与着丝粒的功能相关, 但是并不决定着丝粒的功能. 第三类着丝粒为前文所述的全着丝粒 (holocentromere), 着丝粒分布在整条染色体上. 这类着丝粒多见于线虫, 部分植物和昆虫中^[53].

对于大多数着丝粒而言, DNA 序列本身并不决定染色体上一个区域是否为着丝粒, 而是一些表观遗传因素决定了染色体的某个区域为着丝粒. 着丝粒的 DNA 序列虽然在不同物种间千差万别, 但是所有的着丝粒存在极其保守的特征: 着丝粒的异染色质核小体含有组蛋白的变体蛋白 CENP-A 代替组蛋白 H3 (图 4)^[51]. CENP-A 只存在于着丝粒染色质中, 是着丝粒的重要组成成分, 其在着丝粒的功能中发挥着重要的作用, 缺乏 CENP-A 的细胞无法分离染色体. 甚至人为将 CENP-A 定位于染色体的某个区域也能激活该区域的染色体指向功能, 并且提供动粒蛋白结合位点^[54-57]. CENP-A 由两个重要的区域: 其一

为与 H3 高度同源的组蛋白折叠结构域, 另一为有物种间特异性的 N 端尾巴^[58]. 这两个区域均与动粒蛋白的招募和结合有关^[57,59-63]. 在人、小鼠和鸡 (*Gallus domesticus*) 的染色体中, CENP-A 被证明主要结合于着丝粒核心区中的一个区域, 而不是广泛分布于着丝粒 DNA 上^[64-66]. CENP-A 被装载到 DNA 上需要很多辅助因子的协助, 该过程的具体细节还不是特别清楚. 另外, 对裂殖酵母的着丝粒的分析显示, 其 *otr* 重复序列区能够形成特殊的 3D 结构, 该高级结构可能帮助 CENP-A 定向于着丝粒^[67]. CENP-B 是另一个重要的着丝粒组成蛋白, 在人细胞中, CENP-B 结合于 171 bp α 卫星 DNA 重复序列中的一个 17 bp 的片段, 这个片段也被称为 CENP-B 盒子 (CENP-B box) (图 4). CENP-B 蛋白形成一个二聚体, 因此可以通过分别结合两个 CENP-B 盒子形成着丝粒 DNA 的高级结构^[68].

2.2 着丝粒异常与疾病的关系

由于着丝粒与细胞分裂时染色体平均分配有着密切的联系, 当着丝粒异常时常会造成非整倍体细胞的存在, 从而导致癌症的发生以及其他一些与非整倍体相关疾病, 如唐氏综合征^[69]. 导致着丝粒异

常的原因有很多,包括有丝分裂检查点失效、异常动粒组装、异常纺锤体形成以及异常的DNA甲基化位点等^[70]。目前在针对着丝粒异常与癌症的研究中,通常采用敲除或者敲降与有丝分裂检查点相关基因的小鼠作为模型,如*MAD1*, *MAD2*, *BUB1*, *BUB3*, *BUBR1*以及*CENP-E*等^[71-76]。研究发现,这些基因敲降的小鼠在晚期会发生肺癌、胰腺癌以及肝癌等致命癌症^[71,72,77,78]。尽管有大量证据表明,着丝粒异常与癌症发生有着密切的联系,但是在特定情况下,着丝粒异常导致的非整倍体可以抑制癌症的发生。例如,驱动蛋白基因——*CenpE*单倍剂量不足的情况下会降低致癌物诱导发生癌症的概率,并且可以显著延长敲除了癌症抑制因子*p19^{Arf}*小鼠的存活期^[77]。因此,对于不同的癌症,着丝粒的异常组装既可以作为检测因子,也可以作为治疗手段。

另外一种着丝粒异常相关疾病是ICF综合征(immunodeficiency with Centromeric instability and Facial anomalies, ICF syndrome),其致病原因是DNA甲基转移酶DNMT3B异常,从而导致着丝粒区域重复序列异常。患者表现为严重的免疫缺陷、智力低下、神经系统缺陷,外表多为塌鼻梁、宽眼距以及低位耳^[79]。对染色体形态分析发现,患者常有1号、16号染色体异染色质部分的解离及不完整的9号染色体,这导致大量的染色体异常,包括断裂、重组等^[80]。研究发现,大部分的患者具有功能缺失的DNMT3B。在小鼠中,DNMT3B是着丝粒附近的卫星序列甲基化的必需蛋白。但是在ICF综合症患者的细胞中却发现了卫星序列的超甲基化(hypomethylation)。这种种间差异可能跟卫星序列的差异性有关^[81]。在另一些ICF患者中,虽然没有发现DNMT3B蛋白的突变,但是均能在着丝卫星序列区域发现超甲基化现象^[82]。因此,推测这种着丝粒异常甲基化导致了染色体的异常行为,从而引起一系列ICF症状。

2.3 当前着丝粒研究的问题及方向

着丝粒是真核细胞染色体上的基本结构,在细胞分裂中染色体正确的分配中发挥着重要的功能。尽管不同物种间着丝粒的DNA序列具有极高的特异

性,但是不同的DNA重复序列是否可以形成特定的空间结构还有待进一步探索。由于着丝粒DNA是高度重复串联序列,对于精确的测序和测序后拼接提出了较高的要求,这也是着丝粒研究的瓶颈之一。随着二代测序技术和ChIP-seq技术的发展,会提供着丝粒DNA更丰富的信息,有助于更深入地了解着丝粒。着丝粒组成蛋白如CENP-A等是着丝粒发挥功能的基本成分,这些蛋白仅定位装载于着丝粒部分特定区域的DNA上,既然着丝粒DNA重复序列相似度很高,这种仅特定区域装载肯定与特定的表观遗传信息直接相关,具体分子机制还需要深入了解。同时,进一步了解着丝粒与动粒的相互作用方式以及动粒的组成结构,这会有助于理解染色体分离的发生。随着研究的不断深入,未来对于着丝粒和着丝粒周缘异染色质的形成,着丝粒-动粒形成机制会有更深入的理解,帮助防治癌症和其他非整倍体导致的疾病。

3 端粒和着丝粒的研究展望

端粒和着丝粒作为染色体上由重复DNA序列组成的重要结构,在维持真核生物基因组稳定的过程中发挥着重要的作用。端粒和着丝粒虽然很早就被发现,对于它们的结构和功能也有了相当的研究,但是由于它们是由高度重复的DNA序列组成,尤其是着丝粒,可为上百万碱基对的高度重复序列,测序存在较大的困难,因此在对基因组已经理解较为深刻的现在,依然有很多未知的领域等待探索。例如,特异DNA重复序列间的相互关系,重复序列可能具有的高级空间结构,蛋白与这些序列的相互作用机理,DNA序列是否会因环境变化而发生改变,以及DNA重复序列的各种表观遗传修饰等都是亟待回答的问题。从应用方面来看,针对端粒酶活性抑制的疗法已经进入临床实验阶段;针对诱导着丝粒正确组装的预防肿瘤疗法研究进展较慢,这也是对着丝粒研究不够全面所致。虽然端粒和着丝粒在细胞衰老及肿瘤之间的机制尚未完全阐述清楚,但是相信随着科研的深入、技术的发展,对端粒和着丝粒的认识会更加深入,为肿瘤治疗和衰老研究打下基础。

参考文献

- 1 Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409: 860–921
- 2 Cheung J, Estivill X, Khaja R, et al. Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol*, 2003, 4: R25
- 3 Muller H. The remaking of chromosomes. *Collect Net*, 1938, 13: 181–198
- 4 McClintock B. Cytological observations of deficiencies involving known genes, translocations and an inversion in *Zea mays*. *Univ Missouri Agricul Exp Stat Res Bull*, 1931, 163: 3–30
- 5 Blackburn E H, Gall J G. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in tetrahymena. *J Mol Biol*, 1978, 120: 33–53
- 6 Szostak J W, Blackburn E H. Cloning yeast telomeres on linear plasmid vectors. *Cell*, 1982, 29: 245–255
- 7 Moyzis R K, Buckingham J M, Cram L S, et al. A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci USA*, 1988, 85: 6622–6626
- 8 de Lange T. T-loops and the origin of telomeres. *Nat Rev Mol Cell Biol*, 2004, 5: 323–329
- 9 Palm W, de Lange T. How shelterin protects mammalian telomeres. *Ann Rev GenE*, 2008, 42: 301–334
- 10 Xu Y, Sato H, Shinohara K I, et al. T-loop formation by human telomeric G-quadruplex. In: *Proceedings of the Nucleic Acids Symposium Series*. Oxford: Oxford University Press, 2007
- 11 Henderson E, Hardin C C, Walk S K, et al. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell*, 1987, 51: 899–908
- 12 Xin H, Liu D, Songyang Z. The telosome/shelterin complex and its functions. *Genome Biol*, 2008, 9: 232
- 13 Blastyak A, Pinter L, Unk I, et al. Yeast rad5 protein required for postreplication repair has a DNA helicase activity specific for replication fork regression. *Mol Cell*, 2007, 28: 167–175
- 14 Longhese M P. DNA damage response at functional and dysfunctional telomeres. *Genes Dev*, 2008, 22: 125–140
- 15 Gao H, Cervantes R B, Mandell E K, et al. Rpa-like proteins mediate yeast telomere function. *Nat Struct Mol Biol*, 2007, 14: 208–214
- 16 Greider C W, Blackburn E H. Identification of a specific telomere terminal transferase activity in tetrahymena extracts. *Cell*, 1985, 43: 405–413
- 17 Collins K. The biogenesis and regulation of telomerase holoenzymes. *Nat Rev Mol Cell Biol*, 2006, 7: 484–494
- 18 Greider C W, Blackburn E H. A telomeric sequence in the RNA of tetrahymena telomerase required for telomere repeat synthesis. *Nature*, 1989, 337: 331–337
- 19 Lingner J, Hughes T R, Shevchenko A, et al. Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science*, 1997, 276: 561–567
- 20 Fan X, Li Y Y, Liu Y Y, et al. Application of single molecule fluorescence techniques on telomere and telomerase (in Chinese). *Prog Chem*, 2014, 26: 1987–1996 [范宵, 李艳艳, 刘迎亚, 等. 单分子荧光技术在端粒和端粒酶研究中的应用. *化学进展*, 2014, 26: 1987–1996]
- 21 Feng J, Funk W D, Wang S S, et al. The rna component of human telomerase. *Science*, 1995, 269: 1236–1241
- 22 Moerner W, Orrit M. Illuminating single molecules in condensed matter. *Science*, 1999, 283: 1670–1676
- 23 Olovnikov A M. A theory of marginotomy: The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *J Theor Biol*, 1973, 41: 181–190
- 24 Bodnar A G, Ouellette M, Frolkis M, et al. Extension of life-span by introduction of telomerase into normal human cells. *Science*, 1998, 279: 349–352
- 25 Jaskelioff M, Muller F L, Paik J H, et al. Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice. *Nature*, 2011, 469: 102–106
- 26 Michishita E, McCord R A, Berber E, et al. Sirt6 is a histone h3 lysine 9 deacetylase that modulates telomeric chromatin. *Nature*, 2008, 452: 492–496
- 27 Stewart S A, Ben-Porath I, Carey V J, et al. Erosion of the telomeric single-strand overhang at replicative senescence. *Nat Genet*, 2003, 33: 492–496
- 28 Crabbe L, Jauch A, Naeger C M, et al. Telomere dysfunction as a cause of genomic instability in werner syndrome. *Proc Natl Acad Sci USA*, 2007, 104: 2205–2210
- 29 Burtner C R, Kennedy B K. Progeria syndromes and ageing: What is the connection? *Nat Rev Mol Cell Biol*, 2010, 11: 567–578
- 30 Li B, Jog S P, Reddy S, et al. Wrn controls formation of extrachromosomal telomeric circles and is required for TRF2^{AB}-mediated telomere shortening. *Mol Cell Biol*, 2008, 28: 1892–1904
- 31 Opresko P L, Mason P A, Podell E R, et al. POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. *J Biol*

-
- Chem, 2005, 280: 32069–32080
- 32 Wyllie F S, Jones C J, Skinner J W, et al. Telomerase prevents the accelerated cell ageing of werner syndrome fibroblasts. *Nat Genet*, 2000, 24: 16–17
- 33 Gomes N M, Ryder O A, Houck M L, et al. Comparative biology of mammalian telomeres: Hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging Cell*, 2011, 10: 761–768
- 34 Juola F A, Haussmann M F, Dearborn D C, et al. Telomere shortening in a long-lived marine bird: Cross-sectional analysis and test of an aging tool. *Auk*, 2006, 123: 775–783
- 35 Hall M E, Nasir L, Daunt F, et al. Telomere loss in relation to age and early environment in long-lived birds. *Proc Biol Sci*, 2004, 271: 1571–1576
- 36 Aspect A F. Exploiting tumor cell senescence in anticancer therapy. *BMB Rep*, 2014, 47: 51–59
- 37 Harley C B. Telomerase and cancer therapeutics. *Nat Rev Cancer*, 2008, 8: 167–179
- 38 Blasco M A, Lee H W, Hande M P, et al. Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell*, 1997, 91: 25–34
- 39 Bojesen S E, Pooley K A, Johnatty S E, et al. Multiple independent variants at the tert locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet*, 2013, 45: 371–384
- 40 Epel E S, Blackburn E H, Lin J, et al. Accelerated telomere shortening in response to life stress. *Proc Natl Acad Sci USA*, 2004, 101: 17312–17315
- 41 Azzalin C M, Reichenbach P, Khoriaili L, et al. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, 2007, 318: 798–801
- 42 Ying L, Green J J, Li H, et al. Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc Natl Acad Sci USA*, 2003, 100: 14629–14634
- 43 Wu J Y, Stone M D, Zhuang X. A single-molecule assay for telomerase structure-function analysis. *Nucleic Acids Res*, 2010, 38: e16
- 44 Ren X, Li H, Clarke R W, et al. Analysis of human telomerase activity and function by two color single molecule coincidence fluorescence spectroscopy. *J Am Chem Soc*, 2006, 128: 4992–5000
- 45 Alves D, Li H, Codrington R, et al. Single-molecule analysis of human telomerase monomer. *Nat Chem Biol*, 2008, 4: 287–289
- 46 Zhang W, Mao J H, Zhu W, et al. Centromere and kinetochore gene misexpression predicts cancer patient survival and response to radiotherapy and chemotherapy. *Nat Commun*, 2016, 7: 12619
- 47 Kabisch M, Lorenzo Bermejo J, Dunnebier T, et al. Inherited variants in the inner centromere protein (incenp) gene of the chromosomal passenger complex contribute to the susceptibility of er-negative breast cancer. *Carcinogenesis*, 2015, 36: 256–271
- 48 Yamagishi Y, Sakuno T, Goto Y, et al. Kinetochore composition and its function: Lessons from yeasts. *FEMS Microbiol Rev*, 2014, 38: 185–200
- 49 Clarke L, Carbon J. Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature*, 1980, 287: 504–509
- 50 Chikashige Y, Kinoshita N, Nakaseko Y, et al. Composite motifs and repeat symmetry in *S. pombe* centromeres: Direct analysis by integration of NotI restriction sites. *Cell*, 1989, 57: 739–751
- 51 McKinley K L, Cheeseman I M. The molecular basis for centromere identity and function. *Nat Rev Mol Cell Biol*, 2016, 17: 16–29
- 52 Earnshaw W C, Migeon B R. Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. *Chromosoma*, 1985, 92: 290–296
- 53 Guerra M, Cabral G, Cuacos M, et al. Neocentrics and holokinetics (holocentrics): Chromosomes out of the centromeric rules. *Cytogenet Genome Res*, 2010, 129: 82–96
- 54 Heun P, Erhardt S, Blower M D, et al. Mislocalization of the drosophila centromere-specific histone cid promotes formation of functional ectopic kinetochores. *Dev Cell*, 2006, 10: 303–315
- 55 Mendiburo M J, Padeken J, Fulop S, et al. Drosophila CENH3 is sufficient for centromere formation. *Science*, 2011, 334: 686–690
- 56 Barnhart M C, Kuich P H, Stellfox M E, et al. HJURP is a CENP-A chromatin assembly factor sufficient to form a functional *de novo* kinetochore. *J Cell Biol*, 2011, 194: 229–243
- 57 Logsdon G A, Barrey E J, Bassett E A, et al. Both tails and the centromere targeting domain of CENP-A are required for centromere establishment. *J Cell Biol*, 2015, 208: 521–531
- 58 Goutte-Gattat D, Shuaib M, Ouararhni K, et al. Phosphorylation of the CENP-A amino-terminus in mitotic centromeric chromatin is required for kinetochore function. *Proc Natl Acad Sci USA*, 2013, 110: 8579–8584
- 59 Fachinetti D, Folco H D, Nechemia-Arbely Y, et al. A two-step mechanism for epigenetic specification of centromere identity and function. *Nat Cell Biol*, 2013, 15: 1056–1066
- 60 Carroll C W, Milks K J, Straight A F. Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J Cell Biol*, 2010, 189: 1143–1155

- 61 Carroll C W, Silva M C, Godek K M, et al. Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. *Nat Cell Biol*, 2009, 11: 896–902
- 62 Chen Y, Baker R E, Keith K C, et al. The N terminus of the centromere H3-like protein Cse4p performs an essential function distinct from that of the histone fold domain. *Mol Cell Biol*, 2000, 20: 7037–7048
- 63 Folco H D, Campbell C S, May K M, et al. The CENP-A N-tail confers epigenetic stability to centromeres via the CENP-T branch of the CCAN in fission yeast. *Curr Biol*, 2015, 25: 348–356
- 64 Bodor D L, Mata J F, Sergeev M, et al. The quantitative architecture of centromeric chromatin. *Elife*, 2014, 3: e02137
- 65 Spence J M, Critcher R, Ebersole T A, et al. Co-localization of centromere activity, proteins and topoisomerase II within a subdomain of the major human X alpha-satellite array. *EMBO J*, 2002, 21: 5269–5280
- 66 Zeng K, de las Heras J I, Ross A, et al. Localisation of centromeric proteins to a fraction of mouse minor satellite DNA on a mini-chromosome in human, mouse and chicken cells. *Chromosoma*, 2004, 113: 84–91
- 67 He H, Zhang S, Wang D, et al. Condensin promotes position effects within tandem DNA repeats via the rits complex. *Cell Rep*, 2016, 14: 1018–1024
- 68 Kitagawa K, Hieter P. Evolutionary conservation between budding yeast and human kinetochores. *Nat Rev Mol Cell Biol*, 2001, 2: 678–687
- 69 Verdaasdonk J S, Bloom K. Centromeres: Unique chromatin structures that drive chromosome segregation. *Nat Rev Mol Cell Biol*, 2011, 12: 320–332
- 70 Holland A J, Cleveland D W. Boveri revisited: Chromosomal instability, aneuploidy and tumorigenesis. *Nat Rev Mol Cell Biol*, 2009, 10: 478–487
- 71 Michel L S, Liberal V, Chatterjee A, et al. MAD2 haplo-insufficiency causes premature anaphase and chromosome instability in mammalian cells. *Nature*, 2001, 409: 355–359
- 72 Iwanaga Y, Chi Y H, Miyazato A, et al. Heterozygous deletion of mitotic arrest-deficient protein 1 (MAD1) increases the incidence of tumors in mice. *Cancer Res*, 2007, 67: 160–166
- 73 Babu J R, Jeganathan K B, Baker D J, et al. Rae1 is an essential mitotic checkpoint regulator that cooperates with Bub3 to prevent chromosome missegregation. *J Cell Biol*, 2003, 160: 341–353
- 74 Perera D, Tilston V, Hopwood J A, et al. Bub1 maintains centromeric cohesion by activation of the spindle checkpoint. *Dev Cell*, 2007, 13: 566–579
- 75 Putkey F R, Cramer T, Morphew M K, et al. Unstable kinetochore-microtubule capture and chromosomal instability following deletion of CENP-E. *Dev Cell*, 2002, 3: 351–365
- 76 Wang Q, Liu T, Fang Y, et al. *BUBR1* deficiency results in abnormal megakaryopoiesis. *Blood*, 2004, 103: 1278–1285
- 77 Weaver B A, Silk A D, Montagna C, et al. Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer Cell*, 2007, 11: 25–36
- 78 Jeganathan K, Malureanu L, Baker D J, et al. Bub1 mediates cell death in response to chromosome missegregation and acts to suppress spontaneous tumorigenesis. *J Cell Biol*, 2007, 179: 255–267
- 79 Ehrlich M, Buchanan K L, Tsien F, et al. DNA methyltransferase 3b mutations linked to the ICF syndrome cause dysregulation of lymphogenesis genes. *Hum Mol Genet*, 2001, 10: 2917–2931
- 80 Tuck-Muller C M, Narayan A, Tsien F, et al. DNA hypomethylation and unusual chromosome instability in cell lines from ICF syndrome patients. *Cytogenet Cell Genet*, 2000, 89: 121–128
- 81 Walton E L, Francastel C, Velasco G. Dnmt3b prefers germ line genes and centromeric regions: Lessons from the ICF syndrome and cancer and implications for diseases. *Biology*, 2014, 3: 578–605
- 82 Jiang Y L, Rigolet M, Bourc'his D, et al. Dnmt3b mutations and DNA methylation defect define two types of ICF syndrome. *Hum Mutat*, 2005, 25: 56–63



付钰

中国科学院微生物研究所研究员。于 1997 年获中国科学院遗传研究所硕士学位，于 2008 年获加拿大萨斯喀彻温大学微生物与免疫学系博士学位，2008–2012 年在美国哈佛大学医学院生物化学与分子药理学系从事博士后研究，2012 年入选国家“青年千人计划”，受聘于中国科学院微生物研究所，主要致力于在单细胞单分子水平解析 DNA 复制、DNA 损伤修复以及细胞间通讯的生物化学机制。

Summary for “端粒与着丝粒——染色体上的高度重复序列区域”

Telomere and centromere—DNA tandem arrays on the chromosome

FAN QiFu^{1,2} & FU Yu^{1,2*}

¹ State Key Laboratory of Microbial Resources, Institution of Microbiology, Chinese Academy of Sciences, Beijing 100101, China;

² Savaid Medical School, University of Chinese Academy of Sciences, Beijing 100049, China

* Corresponding author, E-mail: fuyu@im.ac.cn

The eukaryotic genome contains a large amount of tandem repeated sequences. Among these repeats, telomere and the centromere are two unique structures on a chromosome, which play an important role in maintaining the stability and function of the genome. A telomere is repetitive nucleotide sequences that located at the ends of chromosomes, which protects the chromosome from deterioration or fusion with another chromosome. Although the repetitive sequences of telomeres in different species are different, telomeres share the same structure. The typical structure of telomere consists of the D-loop and T-loop, which are protected by Shelterin complex. The telomere length is preserved by the enzyme telomerase which elongates telomeres through its reverse transcription activity. It has been widely accepted that telomere attrition is related to aging and cancer. The centromere is a region of the chromosome, which ensures correct separation of daughter chromosomes during cell division. According to the position on the chromosome, the centromere is classified into metacentric, submetacentric, acrocentric, telocentric and holocentric. It can also be divided into point centromere, regional centromere and holocentromere due to the characteristics of the repetitive sequences, in line with the characteristics of the centromeric repetitive sequences. The damaged centromere can cause a large number of chromosomal anomalies such as chromosome rearrangements and breaks, multi-branched chromosomes, and whole-arm deletions, etc. These abnormal chromosomes will lead to serious diseases, such as cancer, ICF syndrome, even cell death. Both the telomere and centromere are composed of highly repetitive sequences and combined with a variety of proteins to form a complex structure. Because of their numerous repetitive DNA sequences, accurate DNA sequencing of telomere and centromere is technically difficult. So there are still many unanswered questions regarding their functions and their working mechanisms. This paper briefly reviews the history of discovering telomere and centromere, the structural characteristics of the telomere and centromere, including the difference of telomere and centromere DNA sequences among different species, the functions of specific proteins binding in these regions. This review also emphasizes the relationships between the repetitive sequences and aging, ICF syndrome or the occurrence of diseases, especially cancer. At last, the paper highlights and prospects the hot research topics in telomere and centromere studies, such as the methods of telomere length measurement, studies on the characteristics of telomerase, and research about centromere repetitive sequence. Hopefully, the fast development of research technology including sequencing technics will help to further understand the biological roles of telomere and centromere, thus facilitate tumor treatment and aging research.

telomere, telomerase, centromere, DNA repetitive sequence

doi: 10.1360/N972016-01145