

· 临床研究 ·

人类线粒体基因控制区序列的变异频率与碱基构成特点分析

谭端军 刘玲玲 文毅 徐斌 赵玉生 LeeJun. C. Wong

【摘要】 目的 探讨人类线粒体基因控制区的多态频率和变异类型。方法 搜集基因文库记录的 86 个人类线粒体基因 D 环控制区序列,并与标准剑桥序列进行比对分析。结果 86 个序列中共在 181 个位点发现 1 253 个碱基变异,主要变异形式为碱基转换(占 71.75%),其次为碱基插入(16.04%)和缺失(8.62%),碱基颠换较少,仅占 3.59%。核酸位点 16 189、303~315、514~523 是变异频率较高,且具有基因长度不稳定的区域。结论 线粒体基因 D 环区为高度多态变异区,其变异频率的确定可为针对线粒体基因多态性与疾病发生风险评估的其他研究提供参考。

【关键词】 线粒体基因;D 环区;多态性;变异频率

Variation frequency and base composition characterization types of human mitochondrial DNA control region

TAN Duanjun¹, LIU Lingling¹, WEN Yi¹, XU Bin¹, ZHAO Yusheng¹, LeeJun. C. Wong²

1. Institute of Geriatric Cardiology, Chinese PLA General Hospital, Beijing 100853, China

2. Institute for Molecular and Human Genetics, Georgetown University Medical Center, Washington, DC, USA

【Abstract】 Objective To explore the variation frequency and base composition characterization of mitochondrial DNA (mtDNA) control region in human beings. Methods Eighty-six sequences of human mtDNA D-loop region from GenBank were collected and compared with Cambridge sequence. Results A total of 1 253 variations in 181 sites were found in 86 sequences. The dominant type of variation was nucleotide transition (71.75%), followed by insertion (16.04%) and deletion (8.62%). The nucleotide transversion only accounted for 3.59%. The nucleotide sites 16 189, 303-315, and 514-523 were the ones with high variation frequency and length instability. Conclusion The mitochondrial control region is a hypervariable area with high frequency of polymorphism. The evaluation of variation frequency in this region provides a good reference for the studies aiming at the relationship between mitochondrial DNA polymorphism and pathogenetic risk of diseases.

【Key words】 mitochondrial DNA; D-loop; polymorphism; variation frequency

人类线粒体 DNA (mitochondrial DNA, mtDNA) 位于细胞质的线粒体中,是细胞核外唯一存在的多拷贝 DNA 分子^[1]。它由 16 569 个碱基对组成双链环状分子,其控制区(D 环区)是存在于转运 RNA 脯氨酸(tRNA^{Pro})和苯丙氨酸(tRNA^{Phe})之间的一段线粒体非编码基因,片段大小为 1 122 bp,含有两个高变区,即高变区 I 和高变区 II,无关个体 mtDNA 在此区间存在着大量的碱基差异,核酸替代的发生率显著高于单拷贝的核 DNA (nuclear DNA,

mtDNA)^[2]。mtDNA D 环控制区具有更高的序列多态性,这已经通过各种人类 mtDNA 变异分析予以证实。由于它的无重组遗传和碱基的高突变频率,对法医学^[3,4]、群体遗传学^[5]、分子进化^[6-8]、人类生态学和考古学^[9]研究都具有重要意义。通过对线粒体基因限制性内切酶片段长度分析或控制区的序列对比并建立地理特异性的单体群从而有助于疾病的研究。目前的研究已发现有一些老年退行性疾病、心血管病与线粒体 DNA 的多态性具有一定的相关性^[10,11],但目前全面阐述正常人群线粒体基因各位点的变异频率的大样本研究较少,因此,在评估这些多态性与疾病遗传背景的关系时存在一定的局限性。为更好地应用 mtDNA 序列多态性进行个体识别和亲子鉴定,分析这些多态性与老年疾病的关系,明确变异频率和碱基构成特点就显得至关重要。

基金项目:中国人民解放军总医院院长基金(YZZZ005)资助

作者单位:100853 北京市,中国人民解放军总医院老年心血管病研究所(谭端军、刘玲玲、文毅、徐斌、赵玉生、王士雯);美国华盛顿特区乔治城大学医学中心分子与人类遗传研究所(LeeJun. C. Wong)

作者简介:谭端军,男,医学博士,副主任医师

通讯作者:谭端军,电话:010-66936762

本文对从基因文库搜集到的 86 个线粒体基因控制区的序列的多态频率和变异类型进行了多方位的分析和探讨,目的在于为采用线粒体基因多态性来判断疾病发生风险时提供参考。

1 材料和方法

1.1 数据来源 所有 86 个人类线粒体基因控制区全序列数据均来自于美国国立生物技术信息中心(National Center for Biotechnology Information, NCBI)所建立的基因文库(GenBank; www. ncbi. nlm. nih. gov),其中来自非洲的序列 38 个、欧洲 16 个、亚洲 14 个、大洋洲和太平洋群岛 8 个、中东 6 个、美洲 4 个,共涵盖 46 个国家和地区的不同种族人群。其中 4 例中国人的序列基因文库进入号分别为 AF346972、AF346973、AF346972、AF346973,其他人种文库进入号未列出。

1.2 分析方法 对所获得的序列均采用美国 Informax 公司 Vector NTI 7.1 软件包中的 Align 软件与国际上普遍采用的标准线粒体基因剑桥序列^[1](GenBank 进入号 J01415)进行计算机比对分析,确定 86 个序列的变异位点和类型。继之分析这些序列全部 D 环区 1 122 bp 的构成情况,确定多态性变异的平均转换率与颠换率、碱基替换与插入缺失的分布、各位点和各功能区域的平均变异频率等。结果统计采用 χ^2 检验。

2 结果

2.1 碱基构成 本研究结果显示剑桥序列的碱基构成 A、G、C、T 分别为 339(占 30.21%)、151(13.46%)、372(33.16%)和 260(占 23.17%),基因文库中 86 个线粒体基因序列的平均碱基构成 A、G、C、T 则分别为 335.79(占 29.93%)、152.85(占 13.63%)、373.88(占 33.33%)和 259.30(占 23.11%),剑桥序列和基因文库记录的 86 个序列碱基构成情况没有明显的统计学差异,两组均以 C 和 A 含量较大,而 G 的含量最少。

2.2 碱基变异类型 在 86 个线粒体 D 环基因序列中共发现 181 个位点共具有 1 253 个碱基变异,按照线粒体 D 环区基因长度(1 122 bp)进行校正,100 个碱基约有 17.37 个位点发生变异,其平均变异密度约为 1.29 个变异/每人每 100 碱基。可变碱基的主要变异形式为碱基转换(71.75%),其次为插入(16.04%)和缺失(8.62%),明显高于碱基颠换

(3.59%),最常见的碱基转换类型为 T/C→C/T 互换(表 1)。

表 1 86 个人类 mtDNA 的变异类型

碱基变异类型	变异位置数(%)*	碱基变异数(%)
转换	159(81.54)	899(71.75)
A→G	37	242
G→A	20	85
C→T	59	254
T→C	43	318
颠换	21(10.77)	45(3.59)
A→C	4	15
A→T	2	3
C→A	7	14
C→G	4	6
G→C	1	2
G→T	1	2
T→A	2	3
T→G	0	0
插入	6(3.08)	201(16.04)
缺失	9(4.62)	108(8.62)

注: * 变异位点数为 181 个,表中显示的部分位点因同时出现多种变异而重复

2.3 各功能区变异频率与密度 表 2 显示线粒体 D 环基因的变异主要存在于两个高变区和重链起始区内,变异密度以保留序列区 2 最高,达 9.78 个/每人每 100 碱基,线粒体转录因子结合位点(np 523~550)的变异密度也较高,达 1.91 个/每人每 100 碱基,而轻、重链控制成分、重链主要启动子的变异频率较低。当将各功能区变异位置数在该区域所占比率与整个 D 环区的环区变异位置数所占比率比较时,可见高变区 I 和终止相关序列具有变异的位点所占比率分别为 27.2% 和 37.5%,明显高于平均水平(16.13%),具有显著的统计学差异($P<0.05$)。

2.4 高频率变异位点 表 3 显示 45 个发生频率在 5% 以上的碱基替换的分布情况,其中以 16 189 变异频率最高,达 41.86%,其次分别为 16 223、152、16 519 和 195 位点。此外,高频率的缺失突变分别为 248 缺失 A(6.98%)、286~287 缺失 A(3.49%)、303~309 缺失 C(8.14%)、514~523 缺失 CA(72%)、16 182~16 183 缺失 A(15.1%)和 16 193 缺失 C(4.65%),高频插入变异包括 303~309 插入 C 或 CC(46.5%)、311~315 插入 C(98.84%)、514~523 插入(CA)_{1~2}(46.5%)、16 193 插入 C 或 CC(30.23%)。显示出 303~315 以及 514~523 是基因长度最不稳定的区域。

表 2 各功能区的变异频率与变异密度

功能区	起止位点	基因长度(bp)	变异位置数(%)	碱基变异数	变异密度
整个 D 环控制区	16 024~576	1 122	181(16.13)	1 253	1.29
高变区 II	57~372	316	53(16.77)	549	2.02
重链起始部	110~441	332	50(15.06)	458	1.60
保留序列区 1	213~235	23	7(30.43)	8	0.40
转录因子 1 结合位点	233~260	28	6(21.43)	31	1.29
转录因子 1 结合位点	276~303	28	6(21.43)	23	0.96
保留序列区 2	299~315	17	4(23.53)	143	9.78
复制引物	317~321	5	1(20.0)	1	0.23
保留序列区 3	346~363	18	1(5.56)	2	0.13
重链控制成分	371~379	9	1(11.11)	1	0.13
重链控制成分	384~391	8	0(0)	0	0.00
轻链启动子	392~445	54	2(3.70)	2	0.04
转录因子 1 结合位点	418~445	28	1(3.57)	1	0.04
转录因子 1 结合位点	523~550	28	2(7.14)	46	1.91
重链主要启动子	545~567	23	0(0)	0	0.00
高变区 I	16 024~16 383	360	98(27.2)*	494	1.60
终止相关的序列	16 157~16 172	16	6(37.5)*	18	1.31
控制成分	16 194~16 208	15	1(6.67)	1	0.08
轻链控制成分	16 499~16 506	8	0(0)	0	0.00

注:变异密度:平均每每人 100 碱基中的变异个数; * $P < 0.05$, ** $P < 0.01$, χ^2 检验,与整个 D 环区变异位置数所占比率比较

表 3 高频率的碱基替换位点

核酸位点	剑桥序列	基因文库序列(例数)				变异频率(%)		核酸位点	剑桥序列	基因文库序列(例数)				变异频率(%)
		A	G	C	T					A	G	C	T	
73	A	11	75			12.79		16 148	C		80	6		6.98
93	A	78	8			9.3		16 172	T		10	76		11.63
146	T			13	73	15.12		16 187	C		73	13		15.12
150	C			67	19	22.09		16 189	T		36	50		41.86
151	C			81	5	5.81		16 223	C		34	52		39.53
152	T			31	55	36.05		16 230	A	80	6			6.98
182	C			77	9	10.47		16 234	C			81	5	5.81
185	G	3	81		2	5.81		16 235	A	81	5			5.81
186	C	7		79		8.14		16 249	T			5	81	5.81
189	A	71	8	7		17.44		16 270	C			81	5	5.81
195	T	2		25	59	31.39		16 274	G	5	81			5.81
198	C			80	6	6.98		16 278	C			66	20	23.26
204	T			5	81	5.81		16 293	A	80	6			6.98
236	T			6	80	6.98		16 294	C			72	14	16.28
247	G	15	71			17.44		16 298	T			8	78	9.30
297	A	80	6			6.98		16 311	T			25	61	29.07
316	G	7	79			8.14		16 320	C			80	6	6.98
489	T			20	66	23.26		16 327	C			80	6	6.98
16 051	A	81	5			5.81		16 360	C			79	7	8.14
16 111	C	1		80	5	6.98		16 362	T			16	70	18.60
16 126	T			9	77	10.47		16 390	G	10	76			11.63
16 129	G	15	69	2		19.77		16 519	T			56	30	34.88
16 145	G	5	81			5.81								

3 讨论

人类 mtDNA 作为存在于胞质细胞器线粒体中

的一个环状 DNA 分子,编码 13 种线粒体氧化磷酸化呼吸链组成酶所必需的多肽、12S rRNA、16S rRNA 和 22 个 tRNA,一般认为 mtDNA 具有比

nDNA高10~20倍的进化率^[6,7],而其中D环控制区是mtDNA基因进化组中进化速度最高、最具有多态性的区域。由于具有单亲(母系)遗传、不存在类似nDNA的重组现象,而且群体变异大等特点,使这一特殊的基因系统发育形成了探索种群遗传变异,研究人类起源、演化和亲缘关系的基础,对于理解种群遗传结构和进化关系、探讨形成特定遗传结构的历史、地理原因以及疾病发生的风险非常有价值,目前已被作为遗传标记越来越广泛地应用于研究群体遗传和系统进化^[8,12~14]、鉴定种群的进化趋势,并可从进化或系统发生的前景评估一个种群发生疾病的风险及可能与疾病有关的遗传基因和标志。然而,由于对正常人群中的各位点变异发生频率、密度和碱基构成缺乏大样本的研究作为参照,因此,在探讨线粒体基因变异作为遗传背景在疾病发生中的作用时其应用一直受到很大限制。

在本研究中我们分析了基因文库已经记载的86个人类线粒体基因D环区序列,结果显示碱基多态性分布在整个D环控制区中,其中57~372、16 024~16 383是两个高变异区域,其变异密度高出整个D环的平均变异密度的1倍。可变碱基的变异形式主要是碱基替代、插入和缺失,碱基转换率明显高于颠换率。一般认为保守序列区是在人类和动物进化过程中所必须的,而本研究对各功能区的分析显示,保守序列区2(299~315)具有高变异密度的特点,这与以前所确定该区域作为保守序列是矛盾的^[15]。此外,在该区域内的303~309位点具有高度的基因长度不稳定,因此我们认为该区域可能不是人类进化所必须的保守序列。重链起始部与复制终止相关序列的变异位点数和变异密度均相对较高,虽然未能直接检测这些受试人群的mtDNA拷贝数目,但DNA复制与终止相关序列在mtDNA复制过程中具有重要的作用,这些多态性变异或多或少将影响DNA的复制,这可能是不同个体之间细胞内mtDNA拷贝数目具有较大差异的原因之一。此外,本研究显示最具高度多态性的碱基替换包括C150T、T152C、T195C、G247A、T489C、T16 189C、C16 223T、C16 278T、T16 311C、T16 519C。值得提及的是有41.86%的个体具有16 189T→C改变,由于该区域(np16 184~16 193)是一个微卫星区(CCCCTCC-CC),处于中间的16 189T转为C后,使该区域变为连续10C,从而导致了该区域的DNA复制不稳定,许多受试者发生该区域基因长度不稳定。

有研究显示上述部分高度多态性的碱基替换和

微卫星区的不稳定状态(如T16 189C,303~309C插入和缺失等)在不同肿瘤患者的正常组织中未见表达,但在同一个体的肿瘤组织中却呈现出高度获得性变异^[16~19],这种高发的体细胞性突变进一步证实了线粒体基因控制区的易变异性。下列原因可能可以解释线粒体基因易变异性和多态性特点,首先,线粒体基因缺乏组蛋白的保护;第二,mtDNA修复机制十分有限;第三,线粒体是活性氧产生的主要场所,它对氧化性DNA损伤十分敏感。因此,导致了线粒体基因在进化过程中呈现的高度变异。而线粒体基因的变异主要集中在D环控制区的原因可能与该区域不参与氧化磷酸化各复合物的基因编码有关,除D环区以外的线粒体基因为紧密排列的编码序列,其突变将造成细胞内呼吸链氧化磷酸化过程中的能量代谢障碍而使具有突变的细胞无法在进化过程中幸存下来。相反,D环控制区即使发生变异,却只在一定程度上影响mtDNA拷贝数目,而不影响氧化磷酸化过程,使这些细胞得以生存并将这些变异遗传给子代。

本研究阐述了mtDNA控制区的多态变异特征,可为进行法医学个体识别和亲缘关系鉴定提供一定理论依据,并为利用mtDNA控制区变异来判断多种疾病,如老年退行性疾病和心血管病的遗传背景提供参照。

参考文献

- 1 Anderson S, Bankier AT, Barrell BG, et al. Sequence and organization of the human mitochondrial genome. *Nature*, 1981, 290:457-465.
- 2 Calloway CD, Reynolds RL, Herrin GL, et al. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am J Hum Genet*, 2000, 66:1384-1397.
- 3 Parsons TJ, Coble MD. Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. *Croat Med J*, 2001, 42: 304-309.
- 4 Szibor R, Michael M, Plate I, et al. Efficiency of forensic mtDNA analysis. Case examples demonstrating the identification of traces. *Forensic Sci Int*, 2000, 113:71-78.
- 5 顾明波,柳杰,杜庆新,等.中国汉族人群的线粒体DNA控制区多态性研究. *中国法医学杂志*, 2001, 16:6-9.
- 6 Vigilant L, Stoneking M, Harpending H, et al. African populations and the evolution of human mitochondrial DNA. *Science*, 1991, 253:1503-1507.

(下转第274页)

均检查费分别是后者的1.80倍、1.48倍、1.19倍和1.02倍,药费是2.01倍、2.12倍、1.41倍和1.59倍。尽管有否高血压患者的平均检查费达到统计学意义外,其余各项费用无明显差异,但总费用差异显著,这说明并发症造成了患者医疗费用明显增加。

糖尿病足是完全可以预防的。一些国家对有足溃疡的患者采取多学科联合防治,加强有关教育与预防并采取多病因的治疗,使截肢率下降了43%~85%。我国至今还没有一支受过专门训练的糖尿病足防治专业队伍。我国糖尿病患者,尤其是老年患者中,有关预防糖尿病足的教育开展得非常不够。胡景胜等^[8]报道了北京某医院10年的2306例住院患者中,因糖尿病足住院的患者为2.5%。在这些糖尿病足患者(平均年龄为65岁)中,从来没有接受过糖尿病教育的为88%,不活动的为91%,足溃疡为76%,截肢率为22%。本组糖尿病足病患者的血管病变危险因素非常强烈,临床上必须加强糖尿病治疗的全面达标。对这些糖尿病足病患者应该加强糖尿病教育以提高治疗的顺从性和达标率,如此,才能从根本上防治糖尿病足和降低有关的医疗费用。

参考文献

- 1 Xu ZR, Wang YZ, Hu CW, et al. Analysis on medical cost in diabetic patients with foot problems in China. *Diabetes Res Clin Pract*, 2002, 56(Suppl 1):S37-S38.
 - 2 许樟荣,王玉珍,王先丛,等. 糖尿病慢性并发症与糖尿病治疗关系的调查. *中华医学杂志*, 1997, 77:119-122.
 - 3 Boulton AJM. Foot problems in patients with diabetes mellitus. In: Pickup J, Williams G, eds. *Textbook of Diabetes*. 2nd ed. London: Blackwell, 1997:58.
 - 4 陈兴宝,唐玲,陈慧云,等. 2型糖尿病并发症对患者治疗费用的影响评估. *中国糖尿病杂志*, 2003, 11:238-241.
 - 5 中华医学会糖尿病学分会慢性并发症调查组. 1991~2000年全国住院糖尿病患者慢性并发症及相关大血管病回顾性分析. *中国医学科学院学报*, 2002, 2:447-451.
 - 6 Ouriel K. Peripheral arterial disease. *Lancet*, 2001, 358:1257-1264.
 - 7 潘长玉,高妍,袁申元,等. 2型糖尿病下肢血管病变发生率及相关因素调查. *中国糖尿病杂志*, 2001, 9:323-328.
 - 8 Hu JS, Ma XY, Jing DQ, et al. Investigation of the incidence and complication of the diabetic foot in inpatients from 1991-2001. *Diabetes Res Clin Pract*, 2002, 56(Suppl 1):S104.
- (收稿日期:2003-8-11)
(本文编辑 张和起)
- (上接第263页) •
- 7 Brown WM, George M Jr, Wilson AC. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA*, 1979, 76:1967-1971.
 - 8 Avise JC, Bowen BW, Lamb T. DNA fingerprints from hypervariable mitochondrial genotypes. *Mol Biol Evol*, 1989, 6:258-269.
 - 9 Montiel R, Malgosa A, Francalacci P. Authenticating ancient human mitochondrial DNA. *Hum Biol*, 2001, 73:689-713.
 - 10 Chang SW, Zhang D, Chung HD, et al. The frequency of point mutations in mitochondrial DNA is elevated in the Alzheimer's brain. *Biochem Biophys Res Commun*, 2000, 273:203-208.
 - 11 Arbustini E, Diegoli M, Fasani R, et al. Mitochondrial DNA mutations and mitochondrial abnormalities in dilated cardiomyopathy. *Am J Pathol*, 1998, 153:1501-1510.
 - 12 Wallace DC. Mitochondrial diseases in man and mouse. *Science*, 1999, 283:1482-1488.
 - 13 Schon EA. Mitochondrial genetics and disease. *Trends Biochem Sci*, 2000, 25:555-560.
 - 14 Avise JC. Mitochondrial DNA and the evolutionary genetics of higher animals. *Philos Trans R Soc Lond (Biol)*, 1986, 312:325-342.
 - 15 MITOMAP: A Human Mitochondrial Genome Database. Center for Molecular Medicine, Emory University, Atlanta, GA, USA. <http://www.mitomap.org>, 2001.
 - 16 Tan DJ, Bai RK, Wong LJC. Comprehensive scanning of somatic mitochondrial DNA mutations in breast cancer. *Cancer Res*, 2002, 62:972-976.
 - 17 Penta JS, Johnson FM, Wachsman JT, et al. Mitochondrial DNA in human malignancy. *Mutat Res*, 2001, 488:119-133.
 - 18 Tan DJ, Chang J, Chen WL, et al. Novel heteroplasmic frameshift and missense somatic mitochondrial DNA mutations in oral cancer of betel quid chewers. *Genes Chromosomes Cancer*, 2003, 37:186-194.
 - 19 Bianchi NO, Bianchi MS, Richard SM. Mitochondrial genome instability in human cancers. *Mutat Res*, 2001, 488:9-23.
- (收稿日期:2003-05-30)
(本文编辑 张和起)