

疾病与生命科学前沿研究丛书

孙树汉 胡振林 颜宏利 主编

染色体、基因与疾病

Chromosome, Gene and Disease



科学出版社
www.sciencep.com

疾病与生命科学前沿研究丛书

染色体、基因与疾病

Chromosome, Gene and Disease

孙树汉 胡振林 颜宏利 主编

科学出版社

北 京

内 容 简 介

本书以人类染色体为主线,依次介绍了每一条染色体的主要特征、染色体上的致病或易感基因及其与疾病的关系,尤其对该染色体上的主要基因的名称、定位、生理功能及其与人类疾病的关系进行了归纳总结,以期对人类基因组计划及其后续的功能基因组和疾病基因组研究的最新成果进行梳理,给研究者一个比较清晰的脉络。

本书是国内第一本系统介绍染色体、基因与疾病的书籍,对于生命科学研究人员或临床医生具有很好的参考价值。

图书在版编目(CIP)数据

染色体、基因与疾病/孙树汉,胡振林,颜宏利主编. —北京:科学出版社, 2008

(疾病与生命科学前沿研究丛书)

ISBN 978-7-03-022507-8

I. 染… II. ①孙…②胡…③颜… III. ①染色体-遗传病②基因-医学遗传学-关系-疾病 IV. R596.1 R394

中国版本图书馆 CIP 数据核字 (2008) 第 102849 号

责任编辑:夏 梁 席 慧/责任校对:钟 洋

责任印制:钱玉芬/封面设计:王 浩

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

新 蕾 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2009 年 1 月第 一 版 开本:787×1092 1/16

2009 年 1 月第一次印刷 印张:37

印数:1—2 000 字数:857 000

定价:98.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

《染色体、基因与疾病》编委会名单

主 编 孙树汉 胡振林 颜宏利

编 委 (按姓氏笔画排序)

丁飞翔	王 芳	王凯慧	孙树汉
刘善荣	朱海英	陆一鸣	陈蕊雯
张 毅	胡振林	郭瀛军	颜宏利

前 言

基因组是一个物种所有染色体及其位点的总和。组成人类核基因组的染色体有 24 条, 分别由全长约 1m、约 32 亿对核苷酸的 DNA 分子组成, 而基因的数目有 3 万~4 万个, 携带了人类任何一个个体的有关生、长、育、病、老、死的全部遗传信息。随着人类基因组计划 (HGP) 的顺利进行, 生物医学研究已进入后基因组时代 (post-genome era)。人类基因组计划是由来自美国、英国、日本、法国、德国、中国的科学家共同执行的一项计划。1990 年, 人类基因组计划在美国正式启动, 目标是用大约 15 年的时间完成人类基因组 3×10^9 碱基对的测序。1999 年, 中国获准加入人类基因组计划, 承担 1% 的测序任务, 成为参与这一计划的唯一发展中国家。2000 年 6 月 26 日, 中、美、日、德、法、英 6 国科学家宣布首次绘成人类基因组“工作框架图”。2003 年 4 月 14 日, 6 国科学家宣布人类基因组序列图绘制成功, 人类基因组计划的所有目标全部实现。2006 年 5 月 18 日, 英、美科学家公布了人类 1 号染色体的基因测序图, 这条染色体是人类“生命之书”中最长也是最后被破解的一章。公布最后一个人类染色体的基因测序图, 不仅标志着人类基因组计划已经完成, 而且也标志着建立在人类基因组测序图基础上的生物和医学研究的浪潮将日益高涨。

1990 年启动人类基因组计划和其他的生物基因组研究以来, 产生了大量的数据。而事实上不仅仅是核苷酸和蛋白质序列的数据在迅速增长, 各种生物的特征基因和蛋白质结构的数据量以及对致病基因和复杂疾病相关基因的结构和功能的研究资料也飞速膨胀。面对浩繁的数据信息, 如何能将人类基因组计划及其后续的功能基因组、疾病基因组研究的最新成果进行归纳、梳理, 给研究者一个比较清晰的脉络, 并能为生命科学研究人员或临床医生提供有用的精华信息, 是我们思考、讨论了很长时间的问题。经过两年多的酝酿, 参考近年来最新的文献, 本书从一个新的角度, 以人类染色体为主线, 系统介绍每一条染色体的特点、染色体上的致病或易感基因及其与疾病的关系。尤其对染色体上主要基因的名称、定位、生理功能及其与人类疾病的关系进行了归纳总结, 包括孟德尔遗传疾病、染色体病、复杂性状疾病 (包括高血压、糖尿病、哮喘、肿瘤等)。本书的目的是为读者提供一份人类染色体的“精华本”, 使读者读后能够对每条染色体的概况有一个清晰的认识; 另外也是为读者提供一份“地图”, 使读者能够常放案头, 需要时能按图索骥, 准确有效地检索到有用的信息。因此, 本书是国内第一本系统介绍染色体、基因与疾病的书籍, 具有较高的学术参考价值。可供生命科学研究人员或临床医生参考使用。

由于篇幅、时间和水平的限制, 加之编写过程略显仓促, 本书难免有疏漏和不足之处, 恳请同道们不吝赐教, 并期望读者提出宝贵意见, 以利再版时修正。

编 者

2008 年 4 月于第二军医大学

• i •

目 录

前言

第一章 人类最大的 1 号染色体	1
第一节 DNA 序列和生物学特征	1
一、染色体图谱和 DNA 序列特征	1
二、基因注解	3
三、序列重复、变异和染色体重组及其生物学意义	3
第二节 疾病相关基因与人类疾病	5
一、疾病相关基因	5
二、孟德尔遗传疾病	23
三、多基因遗传病	29
四、1p36 微缺失综合征	34
五、1p 结构异常与肿瘤	34
主要参考文献	36
第二章 2 号染色体与非胰岛素依赖型糖尿病	37
第一节 2 号染色体结构与疾病相关基因	37
一、2 号染色体结构特征	37
二、2 号染色体上已定位的疾病相关基因	39
第二节 2 号染色体与人类疾病	53
一、单基因病	53
二、非胰岛素依赖型糖尿病	55
三、遗传性非息肉病性结直肠癌	63
主要参考文献	66
第三章 拥有蛋白质编码基因最多的 3 号染色体	67
第一节 DNA 序列和生物学特征	67
一、染色体图谱和 DNA 序列特征	67
二、基因注解	69
三、染色体重组	70
第二节 疾病相关基因与人类疾病	70
一、疾病相关基因	70
二、孟德尔遗传疾病	81
三、多基因遗传病易感基因	86
四、3p 和 3q 部分三体综合征	87
五、肿瘤相关基因——FHIT 基因	87
六、CCR5 基因突变与抗艾滋病自然免疫力	89

主要参考文献	90
第四章 4号染色体与亨廷顿舞蹈病	92
第一节 4号染色体结构与疾病相关基因	92
一、4号染色体结构特征	92
二、4号染色体上已定位的疾病相关基因	93
第二节 4号染色体与人类疾病	101
一、亨廷顿舞蹈病	101
二、复杂性状遗传病	104
三、鼻咽癌	106
主要参考文献	110
第五章 5号染色体与猫叫综合征	111
第一节 5号染色体结构与特征	111
一、染色体图谱和碱基测序	111
二、生物学比较	113
第二节 猫叫综合征	114
一、5号染色体与疾病	114
二、5p ⁻ 综合征	123
主要参考文献	133
第六章 6号染色体与自身免疫病	135
第一节 6号染色体结构与特征	135
一、染色体图谱和碱基测序	135
二、染色体的特性	137
三、生物学比较	138
第二节 6号染色体与自身免疫病	138
一、6号染色体与疾病	138
二、自身免疫病	149
主要参考文献	161
第七章 7号染色体与囊性纤维化	163
第一节 7号染色体的基本特征	163
一、基因组序列	163
二、人类7号染色体的基因注释	165
三、7号染色体DNA序列的医学注释	166
第二节 7号染色体与人类疾病	177
一、孟德尔遗传疾病	177
二、微结构异常染色体病	183
三、多基因遗传病	186
四、染色体病	187
主要参考文献	188
第八章 8号染色体与早老症	190

第一节 8号染色体的基本特征	190
一、序列特征	190
二、8号染色体远端高突变区及致病基因	191
第二节 8号染色体与人类疾病	199
一、孟德尔遗传疾病	199
二、Neuregulin 基因与精神分裂症	204
三、染色体病	207
四、MYC 基因、癌基因与肿瘤	208
主要参考文献	209
第九章 9号染色体与半乳糖血症	211
第一节 9号染色体的结构与基因特点	211
一、染色体图谱和序列特征	211
二、基因注解	213
第二节 9号染色体与人类疾病	215
一、孟德尔遗传疾病	220
二、复杂性状疾病	229
三、染色体异常相关疾病	231
主要参考文献	234
第十章 10号染色体与家族性癌综合征	236
第一节 10号染色体的结构与基因特点	236
一、染色体图谱和序列概貌	236
二、基因注解	238
第二节 10号染色体与人类疾病	241
一、家族性癌综合征	250
二、其他孟德尔遗传疾病	253
三、复杂性状疾病	255
主要参考文献	259
第十一章 11号染色体——嗅觉	260
第一节 11号染色体的结构和基因特征	260
一、11号染色体的结构和基因特征	260
二、11号染色体基因定位与疾病	264
三、人类11号染色体与肿瘤	276
第二节 11号染色体与人类疾病	280
一、 β 地中海贫血	280
二、肾母细胞瘤	284
主要参考文献	287
第十二章 12号染色体与癌症	289
第一节 12号染色体结构和基因特征	289
一、12号染色体结构和基因特征	289

二、12 号染色体基因定位与疾病	299
第二节 12 号染色体与人类疾病	299
一、齿状核红核苍白球路易体萎缩症	299
二、血管性假血友病	303
三、胰腺癌	306
主要参考文献	317
第十三章 13 号染色体与 13 三体征	320
第一节 染色体结构与基因特点	320
一、染色体图谱和 DNA 序列特征	320
二、基因索引	320
三、染色体全景图	322
四、与遗传连锁图谱的比较	322
五、同源性比较	323
第二节 13 号染色体与人类疾病	323
一、定位于 13 号染色体的疾病相关基因	323
二、孟德尔遗传疾病	327
三、多基因病：精神分裂症	331
四、染色体病	332
五、肿瘤	334
主要参考文献	340
第十四章 14 号染色体与原发性视网膜色素变性	342
第一节 染色体的 DNA 序列和生物学特征	342
一、染色体结构和 DNA 序列特征	342
二、基因注释	344
三、比较基因组学	345
第二节 14 号染色体与人类疾病	345
一、单基因病	345
二、多基因遗传病	351
三、染色体病	352
四、肿瘤	352
五、14 号染色体的疾病相关基因	353
主要参考文献	358
第十五章 15 号染色体与 Prader-Willi 综合征	360
第一节 染色体结构特征与基因注释	360
一、染色体序列特征概述	360
二、染色体片段复制	361
三、15 号染色体基因注释	362
第二节 15 号染色体与疾病	365
一、马凡氏综合征	365

二、染色体病	367
三、多基因病	375
四、肿瘤	376
主要参考文献	380
第十六章 16 号染色体与 α 地中海贫血	381
第一节 染色体结构特征与基因注释	381
一、染色体序列结构概述	381
二、高度复制染色体和结构片段多态现象	383
三、基因注释	384
第二节 16 号染色体与疾病	389
一、孟德尔遗传疾病	389
二、多基因遗传病	400
主要参考文献	404
第十七章 17 号染色体与抑癌基因 <i>p53</i>	406
第一节 17 号染色体的基本特征	406
第二节 17 号染色体与人类疾病	409
一、孟德尔疾病	416
二、多基因病	422
三、17 号染色体与肿瘤	423
主要参考文献	433
第十八章 18 号染色体与 Edwards 综合征	434
第一节 18 号染色体的基本特征	434
一、染色体的结构特征	434
二、基因注释	436
第二节 18 号染色体与人类疾病	437
一、孟德尔遗传疾病	439
二、肥胖症相关基因	445
三、Edwards 综合征	451
四、18 号染色体杂合性缺失 (LOH) 引起的肿瘤	455
主要参考文献	459
第十九章 基因密度最高的 19 号染色体	461
第一节 19 号染色体结构与疾病相关基因	461
一、染色体的结构特征	461
二、染色体上已定位的疾病相关基因	463
第二节 19 号染色体与人类疾病	471
一、孟德尔遗传疾病	471
二、复杂性状疾病	476
三、染色体异常与肿瘤	477
主要参考文献	477

第二十章 20 号染色体与 ADA 缺乏症	478
第一节 20 号染色体结构与疾病相关基因	478
一、染色体的结构特征	478
二、染色体上已定位的疾病相关基因	481
第二节 20 号染色体与人类疾病	485
一、腺苷酸脱氨酶缺乏症	485
二、Alagille 综合征	486
三、哮喘	486
四、染色体病	489
五、原发性脑瘤	489
主要参考文献	492
第二十一章 21 号染色体与唐氏综合征	493
第一节 21 号染色体结构与基因特点	493
一、21 号染色体序列多态性	493
二、基因定位与功能	496
第二节 21 号染色体与人类疾病	498
一、唐氏综合征	500
二、孟德尔遗传疾病	511
三、多基因病	516
四、肿瘤	517
主要参考文献	517
第二十二章 22 号染色体与微结构异常	518
第一节 染色体结构与序列特点	518
一、染色体上的低拷贝重复	519
二、22 号染色体负载的基因量	521
第二节 22 号染色体与人类疾病	523
一、微结构异常染色体病	526
二、其他孟德尔遗传疾病	532
三、多基因病	537
四、肿瘤	537
主要参考文献	540
第二十三章 X 染色体与血友病	541
第一节 染色体结构特征与基因注释	541
一、染色体序列特征概述	541
二、两类特殊基因：非编码 RNA 和肿瘤-睾丸抗原相关基因	542
三、X 染色体失活	544
四、基因注释	545
第二节 X 染色体与人类疾病	554
一、孟德尔遗传疾病	554

二、染色体病	560
三、肿瘤	562
主要参考文献	563
第二十四章 Y 染色体与男性不育	564
第一节 染色体结构与序列特征	564
一、染色体序列结构概述	564
二、MSY 常染色质的三种序列	566
三、人类 Y 染色体上的两种生殖性重组的模式	568
四、基因注释	569
第二节 Y 染色体与人类疾病	571
一、孟德尔遗传疾病	571
二、前列腺癌	571
三、Y 染色体微缺失与男性不育	574
主要参考文献	578

第一章 人类最大的 1 号染色体

1 号染色体是人类染色体中最大的, 包含了大约 8% 的人类遗传信息。它的大小决定了它在人类基因组遗传图谱和遗传特性研究方面比其他的染色体更具有代表性。1 号染色体具有重要的医学意义: 超过 350 种人类疾病与这条染色体序列的异常相关, 这些疾病包括癌症、帕金森病、早老性痴呆以及许多单基因病, 而它们对应的基因很多至今还不清楚。1 号染色体也有重要的生物学意义, 它大约比人类染色体中最小的 21 号、22 号和 Y 染色体大 6 倍, 因而它对于研究人类遗传信息在细胞分裂之前协调的复制方式就非常重要。本章将基于人类 1 号染色体中已完成的序列, 对染色体图谱、基因索引和序列多样性及其与人类疾病的关系提供较为详细的注解。希望这些注解可以提供更广泛的遗传学和生物学的研究信息。

第一节 DNA 序列和生物学特征

一、染色体图谱和 DNA 序列特征

来自英国 Wellcome Trust Sanger 学院 (the Wellcome Trust Sanger Institute) 以及美国杜克大学人类遗传学中心等研究机构的研究人员在 2006 年 5 月的 *Nature* 杂志上联合公布了人类基因组第 1 号染色体序列, 完成了人类基因组染色体最长以及最后一个测序。测序工作组确定了一组代表 1 号染色体常染色质的 2220 个最小重叠克隆的序列, 这个序列包含 223 875 858 个碱基对, 准确度大于 99.99%; 其中, 120 405 438bp 位于短臂 (1p) 的 14 个重叠克隆群上, 103 470 420bp 位于长臂 (1q) 的 13 个重叠克隆群上。这个序列延伸到了染色体两臂的端粒重复基序 (TTAGGG)_n 和短臂近端 (1pcen) 的环着丝粒 α 卫星序列。还有位于 1q 邻近着丝粒的异染色质区域的 18Mb 的碱基仍未被测序。图 1-1 显示 1 号染色体的结构特征略图, 更详细的每一区域的特征和注解可通过 Vega Human MapView 浏览器网上追踪查看 (http://vega.sanger.ac.uk/Homo_sapiens/mapview?chr=1)。

对覆盖了 90 个基因当量基因组的细菌和酵母菌衍生的克隆库进行详尽的筛查后发现仍存有 26 个缺口。其中, 8 个聚集于 1p36, 另有 8 个聚集于 1q21.1。这些区域富含 GC 而且含有低拷贝重复, 推测这些特点导致了这些区域的克隆缺失。运用荧光素原位杂交技术对染色体 DNA 进行分析发现 17 个缺口中包含的碱基总量为 0.8Mb, 通过将人类重叠群与鼠、兔和猩猩的基因序列进行比对, 估计剩余的 9 个缺口总计有 0.53Mb。因此 1 号染色体的常染色质部分长度为 225.2Mb, 其中 99.4% 是已完成的序列。

测序工作组通过 3 个独立的分析来评价序列的完整性。首先, 目前位于 1 号染色体的 2580 个参考基因 (Refseq) 中, 除了 1 个以外, 其余全部存在于序列中, 缺失的基因 *RAB7B* 在 GB4 放射杂交图谱上被定位于 1q32, 这一位置应处于缺口 23 或 24 上。另有两个基因 *IPP* 和 *PHACTR4* 只有部分存在于报道的序列中, 但随后已被完全测

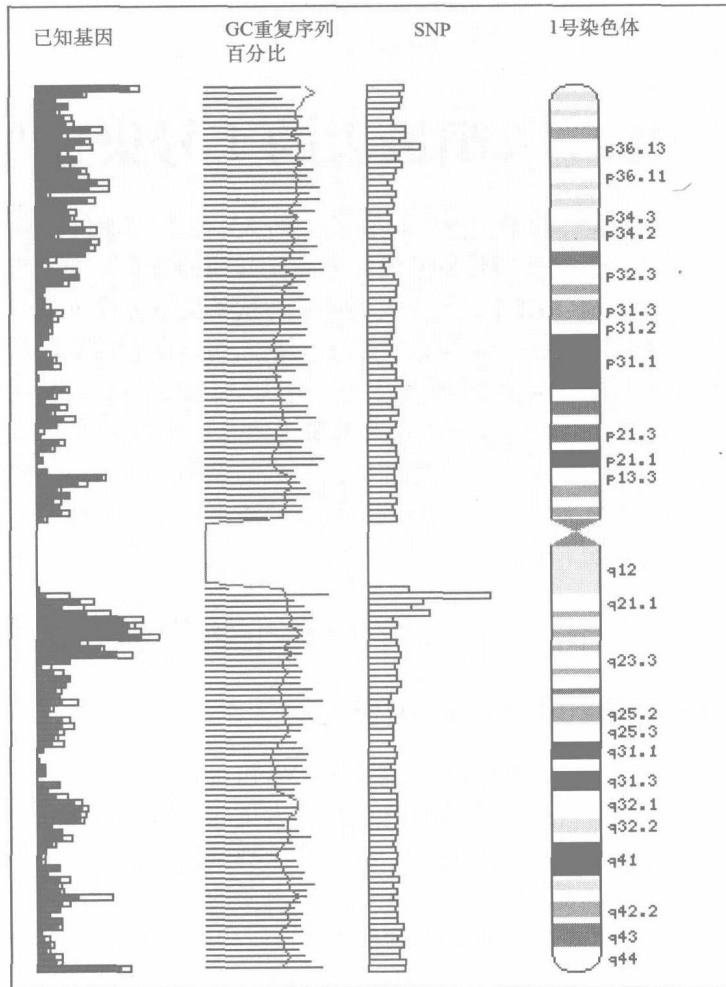


图 1-1 1 号染色体的结构特征

(引自 http://www.ensembl.org/Homo_sapiens/mapview?chr=1)

序。其次，对比 467 个遗传标记在已完成序列和 deCODE 遗传图上的顺序，未发现偏差。最后，通过排列 32 984 对黏粒克隆的末端序列在已完成序列上的特定位置，发现了 11 处不一致，其中，3 处是由于低拷贝重复导致的错误装配，现已被校正。其余的可能是已完成序列中的缺失，或是由自然发生长度多态性所致。例如，*GSTM1* 基因在 5% 的人中都没有，这个基因在参考序列中可以找到，但是黏粒克隆图中却没有。

1 号染色的 GC 含量、重复序列和 CpG 岛密度（分别为 41%、48% 和 8.9 岛/Mb）与整体基因组的平均值相当。富含 G+C 区、基因密度、吉姆萨浅染色区、短散布元件（SINE）和长散布元件（LINE）重复的密度之间均相关。1 号染色体的基因密度为 14.2 个基因/Mb——几乎为基因组平均值（7.8 个基因/Mb）的 2 倍，因此 1 号染色体是基因最密集的染色体之一。位于 1p36.33 邻近 1p 的端粒的 2Mb 吉姆萨浅染色区是染色体末端序列区域的典型代表（G+C 摩尔百分比含量为 58.4%，98 个预测的 CpG 岛和 104 个基因）。

二、基因注解

通过分析所有可用的能够为基因特征提供证据的 cDNA 和蛋白质信息,测序工作组共注解了 3141 个结构,已收录于脊椎动物基因组注解数据库中 [Vertebrate Genome Annotation database (VEGA) (http://vega.sanger.ac.uk/Homo_sapiens/index.html)]。

基因索引包含 2153 个已知基因、85 个新蛋白编码序列、720 个新转录子和 420 个预测转录子。1189 个基因 (39%) 在反义链上有重叠,655 个位点 (21%) 在同义链上有编码区重叠。另外还识别出 114 个假基因。有 56% 的已知基因和 40% 的新编码序列与 CpG 岛相关 (表 1-1)。

表 1-1 1 号染色体的基因汇总

长度	247 249 719 bp
已知蛋白编码基因	2153bp
新蛋白编码基因	85bp
假基因	114bp
miRNA 基因	68bp
rRNA 基因	44bp
snRNA 基因	167bp
snoRNA 基因	63bp
Misc RNA 基因	104bp
SNP	1 059 468bp

通过将 1 号染色体的序列与鼠、兔、斑马鱼和两种河豚属物种的基因序列进行比对来识别进化保守区域 (evolutionarily conserved region, ECR)。共发现在所有 6 种基因组保守的 10 971 个 ECR 中有 10 669 个与已注解的外显子重叠,表明外显子注解已完成了 97.2%,其余的 302 个 ECR 可能是尚无支持证据的外显子或是高度保守的调控或结构元件。

运用 Rfam 数据库的结构 RNA 比对,预测了 459 个非编码 RNA 基因或假基因。通过对收录入 miRNA Registry (<http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml>) 的 1345 个 miRNA 进行 BLAST 分析,新鉴别出了 22 个 miRNA,miRNA 是一组由大约 22 个核苷酸组成的非编码 RNA,它们可以通过转录后调控机制调节基因表达。

三、序列重复、变异和染色体重组及其生物学意义

序列重复导致新的遗传物质产生,并演化产生新的功能。通过分析长度 $\geq 10\text{kb}$ 、一致性 $\geq 90\%$ 的已鉴定的序列中的所有重复,发现 3.49% 的染色体内和 1.64% 的染色体间的重复。1q21.1 的 5Mb 区域含有一个染色体内重复的复杂模式。之前,一个由 1q21.1 衍生的细菌人工染色体克隆可在 1p36.13 和 1p12 上显示 FISH 杂交信号,并且在 1q21.1 上产生宽杂交带。序列分析发现了一个重复序列单元存在于 1q21.1 上的 11 个位点、1p36.13 的 3 个位点和 1p12 的 1 个位点。每个拷贝包含一串由 1.5kb 重复序列组成的不同大小 ($\leq 75\text{kb}$) 串联重复序列,该重复序列由不同的神经母细胞瘤断裂

家族成员 (NBPF) 的外显子复制倍增而形成, 该基因家族的成员之一 NBPF1 被发现在一个神经母细胞瘤患者中因移位而发生断裂, 该基因家族的名称由此而来。基因间的和基因内的复制倍增可能产生异常重组导致 NBPF 基因的变异。

1q21.1 的 NBPF 重复区域的近侧为 NOTCH2NL 基因, 远侧为 FCGR1C 和 FCGR1A 基因。它们的同源基因 NOTCH2 和 FCGR1B 位于着丝粒另一侧的 1p12, 似乎是人类和黑猩猩进化分离后的突发臂间倒位所致, NOTCH2NL 是 NOTCH2 的部分缺失体, 包含了 NOTCH2 基因 5' 端到第 4 内含子的 8kb 长的片段, 它编码了一个与中性粒细胞弹性蛋白酶相互作用的功能蛋白, 被认为与遗传性中性粒细胞减少症相关。NOTCH2NL 蛋白包含几个 NOTCH2 上的表皮生长因子重复, 加上一个新的与中性粒细胞弹性蛋白酶的相互作用的 25 个氨基酸组成的羧基末端。分析表明 C 端区域由外显子 4 与内显子 4 某区域的剪接产生。

同源基因对是由某一基因座的区段复制随后累积突变产生的, 1 号染色体上发现 56 个基因簇是人鼠进化分离后复制产生的。

非同义和同义核苷酸置换率的比值 (K_a/K_s) 可以衡量每个基因对的分化率。最高的 K_a/K_s 率 (1.8) 在 SPRR2A 和 SPRR2F 间 (染色体间平均值为 0.4)。SPRR 基因编码富含脯氨酸的小分子蛋白质, 是角质化细胞膜的最主要的构成部分——交联蛋白支架, 在人体与外界接触中起保护作用。角质化细胞膜的许多其他的蛋白质是由 1q21.3 的 SPRR 基因簇编码的, 共同构成表皮分化复合物 (epidermal differentiation complex, EDC)。最近, 在人类和黑猩猩的对比中发现 EDC 为进化最快的基因簇。

拷贝数多态现象 (copy number polymorphism, CNP) 可发生于表型正常的个体。目前已确定了 1 号染色体序列中 CNP, 并在这些区域定位了一组基因家族, 含 CNP 的基因显示的结构多态性可能会导致疾病的易感性。例如, GSTM1 基因的多态性可能与癌症患病率增高相关, FCGR3 多态性最近被认为与人类肾小球肾炎易患相关。

公共数据库中的 1 059 468 个 SNP 已被确定它们在 1 号染色体序列中的位置。7917 个在蛋白质编码区域, 其中, 4471 个为非同义的, 因此推测可能导致功能变化。另外还鉴别出 90 个 SNP 导致编码序列的提前终止, 这些突变可导致 88 个基因编码的蛋白质的缺失突变, 其中, 15 个与遗传疾病相关, 包括 COL11A1 在 II 型 Marshall 和 Stickler 综合征 [OMIM: 120280]、FY/DARC 在疟疾易感性 [OMIM: 110700] 和 UROD 在迟发性皮肤卟啉症 [OMIM: 176100] 中的作用。许多 1 号染色体中的 SNP 被用来判断遗传变异的模式, 提供关于分子和进化过程的重要新信息。

1 号染色体近端粒区域的重组率最高。最低重组率位于短臂近着丝粒区域 (男性: 0.04cM/Mb, 女性 0.77cM/Mb) 和长臂的异染色质区域 (男性: 0.04cM/Mb, 女性 0.31cM/Mb)。1 号染色体的男女平均重组率是 1.13cM/Mb, 与基因组重组率相当。除 1ptel 和 1qtel 外, 重组在女性中高于男性 (1.43 : 0.82cM/Mb)。

作为 HapMAP 计划的一部分, 已确定 269 个个体的 1 号染色体的 60 000 个 SNP 的基因型。邻近的 SNP 等位基因间的高度关联表示低水平的重组, 反之亦然。据此可以观察到一个高度非随机的重组分布, 80% 的重组发生在 15% 的序列中。越靠近端粒重组率越高。大部分的重组发生在密度不同的热点 (<2kb 的不同区段), 越趋向于端粒热点密度越高。在一些区域 (如在 107.5Mb 和 156.5Mb), 总体重组率升高是因为

高密度的热点所致，而另一些区域（如在 11.5Mb 和 151Mb）则是由于少数极度活跃的热点所致。重组率提高与基因密度和 G+C 摩尔百分比含量明确相关。但进一步的小尺度分析显示高密度重组热点实际位于基因外围，而非基因内部，且被编码区域所抑制。这是由于重组发生时双链的断裂常伴随着基因突变，因此有时比在非必需的侧翼 DNA 重组更加有害。此外，重组率与基因表达水平和 G+C 摩尔百分比含量之间有关联，但出乎意料的是，高重组率和 G+C 摩尔百分比含量与基因的低水平表达相关。

运用 HapMAP 计划的信息，对西欧和北欧、东亚（JPT+CHB）、西非人（YRI）的祖先的 1 号染色体的人群特异性的 SNP 等位基因频率进行分析，显示 68 个 SNP 提供了地域限制性选择的证据。这些 SNP 包括三个非同义的变异体，其中，最著名的是 Duffy 基因 [OMIM: 110700] 的 FYa 突变，可以保护人不被间日疟原虫侵害。以此为基线，其他的 67 个 SNP 显示了与 FYa 同等或更大程度的等位基因分化。其他的两个非同义变异体分别存在于 ACOT11 基因（低温诱导的硫酯酶，表达于脂肪组织，与鼠肥胖相关）和 OR9H1R（嗅觉受体，可能为假基因）。对 ACOT11 的高分辨率单倍体分析显示西非人人群的祖先单倍体有近距离固定（near-fixation）（三个 SNP 的类型：脯氨酸-天冬氨酸-甲硫氨酸），然而东亚人群有更大的多样性。这可能归因于非洲人群的强单一性选择和亚洲人群的松弛约束。其他最值得注意的 SNP 包括聚集分布在 NOTCH2、ATP1A1 和 SLC35F3 等基因簇内和周围以及某些单个非编码 SNP，如胆碱能受体基因 CHRM3 和一些嗅觉受体基因。

第二节 疾病相关基因与人类疾病

一、疾病相关基因

美国国家生物技术信息中心（NCBI）基因序列数据库（<http://www.ncbi.nlm.nih.gov>）收录的位于 1 号染色体的基因总数为 2980 个，其中，300 多个基因已有证据表明与某种或多种人类疾病相关。表 1-2 总结了大部分已知的疾病基因在 1 号染色体上的位置、基因代号、名称以及与之相关的人类疾病。

表 1-2 1 号染色体上的疾病相关基因

位置	代号	基因名称	疾病
1pter-p36.13	CCV	cataract, congenital, Volkmann type	cataract, congenital, Volkmann type
1pter-p36.13	ENO1	enolase 1 (alpha)	Enolase deficiency
1pter-p36.1	CPP	ceruloplasmin (ferroxidase) pseudo-gene	cataract, posterior polar
1pter-p36.1	CTPP	cataract, posterior polar	Cataract, posterior polar
1pter-p33	HMGCL	3-hydroxymethyl-3-methylglutaryl-coenzyme A lyase (hydroxymethyl-glutaricaciduria)	HMG-CoA lyase deficiency
1pter-p22.3	SDC3	syndecan 3	Obesity, association with [MIM: 601665]