



评述与展望

Review and Progress

CDRH: 人类复杂疾病相关单倍体数据库

Ruijie Zhang[✉], Yongshuai Jiang[✉], Hongchao Lv[✉], Xuehong Zhang[✉], Peng Sun[✉], Yan Zhang[✉], Jin Li[✉],
Mingming Zhang[✉], Zhenwei Shang[✉], Xia Li[✉]

哈尔滨医科大学, 生物信息科学与技术学院, 哈尔滨, 150086

✉ 通讯作者: zhangruijie2013@gmail.com; ✉ 作者

计算分子生物学, 2012 年, 第 1 卷, 第 10 篇 doi: 10.5376/cmb.cn.2012.01.0010

收稿日期: 2012 年 12 月 01 日

接受日期: 2012 年 12 月 01 日

发表日期: 2012 年 12 月 02 日

本文首次以英文发表在 *Computational Molecular Biology* 上。现依据版权所有人授权的许可协议, 采用 Creative Commons Attribution License 协议对其进行授权, 用中文再次发表与传播。只要对原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。如果读者对中文含义理解有歧义, 请以英文原文为准。

建议最佳引用格式:

Zhang et al., 2011, CDRH: A Database of Complex Disease-related Haplotypes in Human, *Computational Molecular Biology*, Vol.1, No.3 12-19 (doi: 10.5376/cmb.2011.01.001)

摘要 许多常见 DNA 序列及其特定组合(单倍体)的变异可能是导致个体对复杂疾病的易感性差异的根本原因。在积累大量与复杂疾病相关的单倍体资源方面取得了巨大进展。然而, 这些资源分散在不同的文献中, 导致信息的利用率降低。因此, 我们开发了人类复杂疾病相关单倍体数据库(CDRH)。到目前为止, 已经从 274 篇论文中手动提取了总共 1 125 个单倍体, 涉及到 114 种复杂疾病如乳腺癌, 2 型糖尿病和类风湿性关节炎。经仔细阅读文献后, 我们获得了关于单倍体和疾病的详细信息。此外, 我们整合了来自外部数据库的基因和 SNP(和/或微卫星)的相关信息, 以促进进一步分析。通过一个用户容易掌握的界面, 用户可以通过疾病名称, 基因名称, 染色体数或 SNP ID(rs#)查询 CDRH。我们希望 CDRH 将丰富我们的单倍体知识, 并促进单倍体与可遗传的复杂疾病风险之间的关系的研究。CDRH 数据库可从 <http://bioinfo.hrbmu.edu.cn/cdrh> 免费获取。

关键词 CDRH; 单倍体; 复杂疾病

CDRH: A Database of Complex Disease-related Haplotypes in Human

Ruijie Zhang[✉], Yongshuai Jiang[✉], Hongchao Lv[✉], Xuehong Zhang[✉], Peng Sun[✉], Yan Zhang[✉], Jin Li[✉],
Mingming Zhang[✉], Zhenwei Shang[✉], Xia Li[✉]

哈尔滨医科大学, 生物信息科学与技术学院, 哈尔滨, 150086

✉ Corresponding author, xmin@ysu.edu; ✉ Authors

Abstract Many common variations in DNA sequences and their specific combinations (haplotypes) may be the underlying causes of differences in individual susceptibility to complex diseases. Great progress has been made in accumulating abundant resources relating to complex disease-related haplotypes. However, these resources are scattered among different literature, resulting in reduced utilization of the information. Therefore, we developed a database of complex disease-related haplotypes in human (CDRH). To date, a total of 1 125 haplotypes involved in 114 complex diseases, such as breast cancer, type 2 diabetes, and rheumatoid arthritis, have been manually extracted from 274 papers. After careful review of the literature, we obtained detailed information on haplotypes and diseases. Furthermore, we integrated gene- and SNP- (and/or microsatellite-) related information from external databases to facilitate further analysis. Via a user-friendly interface, users can query the CDRH by disease name, gene name, chromosome number, or SNP ID (rs#). We hope that CDRH will enrich our knowledge of haplotypes and promote research into the relationship between haplotypes and heritable risk for complex diseases. The CDRH database is freely available at <http://bioinfo.hrbmu.edu.cn/cdrh>.

Keywords CDRH; Haplotypes; Complex diseases

引言

单倍体包括在单个染色体或染色体的一部分上观察到的特定等位基因集合(HapMap, 2003;

LIN and ZENG, 2006)。单倍体可以为复杂性状, 群体历史和自然选择提供关键的见解(Tishkoff et al., 2000; Daly et al., 2001; Gao et al., 2009)。重要



的是, 越来越多的来自经验和模拟研究的证据表明, 在一些情况下, 在染色体的染色体区中的单倍体可能比使用单个标记鉴定复杂疾病的易感性更有效(Zhao et al., 2003; Gabriel et al., 2002)。许多基于单倍体的研究已成功检测到复杂人类疾病的遗传易感性(Berger et al., 2008; Soma et al., 2008), 如前列腺癌(Yaspan et al., 2008), 乳腺癌(Slattery et al., 2008), 1 型糖尿病(Santiago et al., 2008)和类风湿性关节炎(Hung et al., 2007)。

随着遗传变异数据集的规模和密度的指数增加, 单倍体分析在人类疾病的遗传研究中变得更加重要, 并且已经积累了大量的单倍体数据。在过去的几十年里, 已经开发了一些单倍体相关的数据库用于收集和保存的单倍体信息。D-HaploDB (Higasa et al., 2007) 是由全基因型的葡萄胎样本集合构建的全基因组定位单倍体数据库。YHRD (Kayser et al., 2002) 旨在为美国人群储存 Y 染色体短串联重复序列单倍体。mtDB (Ingman and Gyllensten, 2006) 为医学和人群遗传研究者提供线粒体单倍体搜索功能。然而, 没有特定的与复杂疾病相关的单倍体数据库的收集研究。

为了满足分子生物学家, 遗传学家和病理学家的要求, 我们通过整合大量文章中关于单倍体和疾病的信息, 开发了一个手工策划的人类复杂疾病相关单倍体数据库 (CDRH, <http://bioinfo.hrbmu.edu.cn/cdrh>)。CDRH 是一个综合的和注释的数据库, 是研究人员在单倍体水平了解复杂疾病的有用资源。

1 结果

1.1 数据收集和数据库内容

文本挖掘是用于收集复杂的疾病相关单倍体和其他数据库构建的详细信息。我们使用一系列关键词在 PubMed 数据库中搜索 (<http://www.ncbi.nlm.nih.gov/pubmed>), 例如“复杂疾病单倍体”, “癌症单倍体”, “糖尿病单倍体”, 将结果限制到当前版本 CDRH 的 2010 年 5 月之前的出版物。为了系统和可靠的数据收集, 我们手动检查了重要信息, 并执行以下标准: (i) 文章必须提出和阐述复杂疾病与易感性(或保护性)单倍体之间的关系; 和(ii) 易感性(或保护性)单倍体必须具有统计学检验的某个阈值或 p 值。最终, 总共 1125 个

与 114 种复杂疾病相关的单倍体存储和管理在当前 CDRH 数据库中。数据库中的大多数归档信息用于 SNP 单倍体, 其余的由微卫星组成。

在 CDRH 数据库中, 每个条目包含关于单倍体和疾病的详细信息。收集的信息包括疾病名称, 与疾病相关的单倍体, 单倍体频率, 单倍体的风险情况, 统计检验的 p 值, 单倍体所在的染色体, 单倍体相关的基因符号, 构成单倍体的 SNP(或微卫星), 以及引用文献的参考文献信息。我们不仅收集了广泛的风险单倍体, 还收集了被认为是保护性的单倍体, 这两者都为复杂疾病的未来遗传研究提供了有价值的信息。

我们还整合了来自外部数据库的明确的生物注释来补充和扩展文献信息。从 NCBI 检索与单倍体鉴定相关基因的基本信息, 包括 Entrez Gene ID, Unigene ID, 全基因名称, 基因的染色体位置和基因功能的简要描述。CDRH 中的大多数单倍体是一系列 SNP 构成; 因此, 我们从 dbSNP 收集了单倍体相关 SNP 的信息, 包括 SNP ID, 物理位置和每个 SNP 的等位基因。此外, 还提供了许多方便的有助于复杂疾病相关单倍体的未来调查的外部数据库链接, 如 dbSNP, PubMed, D-HaploDB 和 HapMap。表 1 是 CDRH 数据库中的统计信息。

1.2 数据库启用和 web 界面

CDRH 数据库用 MySQL 5.0 存储和管理数据, 并在 Apache/PHP 环境中运行的 PHP 脚本中启用。

表 1 CDRH 数据汇总

Table 1 Summary of the data in CDRH

Resources	Number
Haplotype	1125
Disease	114
SNP	954
Microsatellite	288
Popoulation	66
Chromosome	22 autosomes, X Y and mitochondrion
Gene	218

1.3 搜索页面

CDRH 数据库可在线访问, 并允许用户通过疾病名称, 基因名称, 染色体数或 SNP ID (rs#) 检索与复杂疾病相关的人类单倍体有关的详细信息。我



们首先介绍疾病名称的搜索, 它在下拉列表框中按字母顺序排序。例如, 如果用户选择“结肠直肠癌”

作为查询项(图 1a), 则搜索和浏览结果将显示在新页面中(图 1c)。详细信息包括三个部分: 疾病, 文

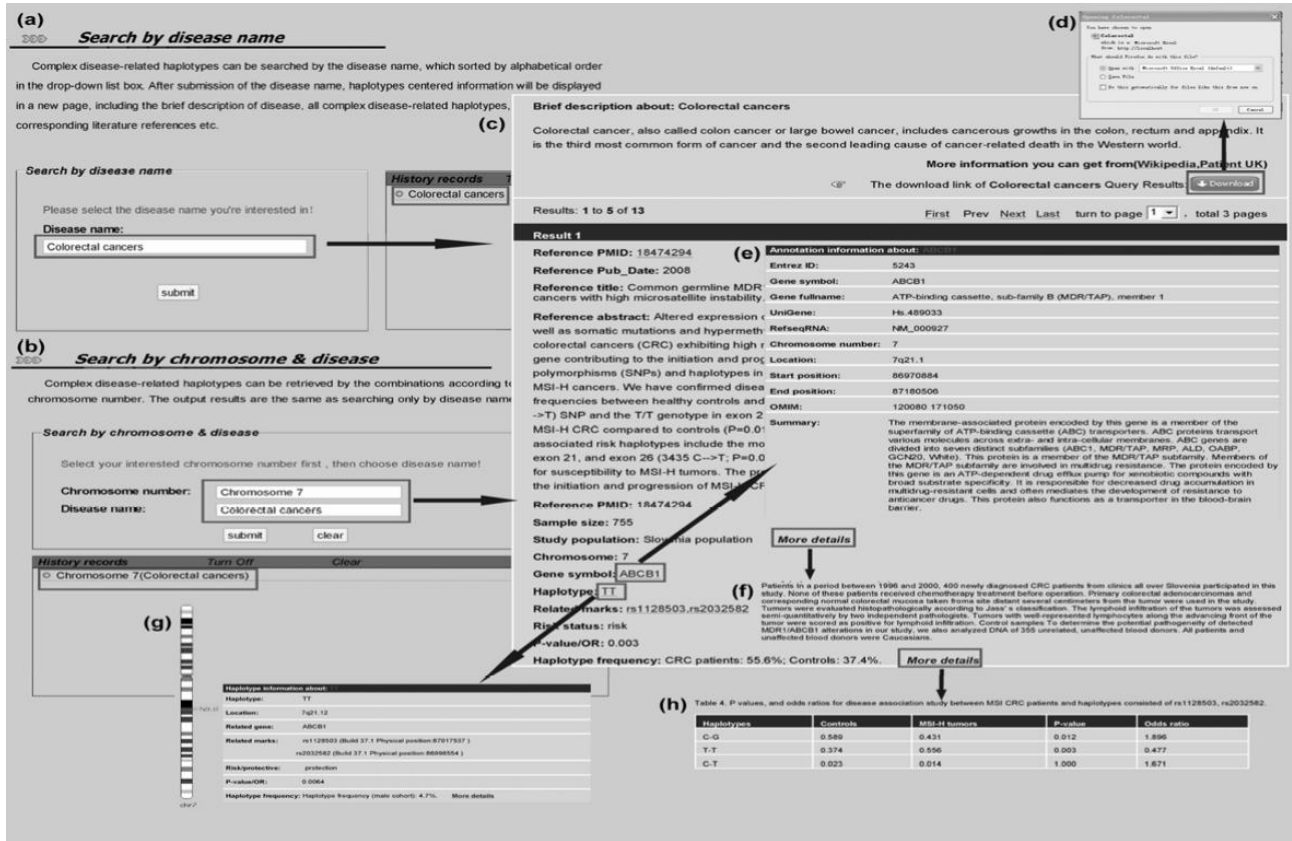


图 1 结肠直肠癌搜索结果

注: (a) 搜索结肠直肠癌(b) 搜索结肠直肠癌和 7 号染色体(c) “结肠直肠癌”的检索结果列表 (d) 检索结果的下载页面 (e) 通过点击“ ABCB1”基因符号链接获得的详细信息 (f) 斯洛文尼亚人人群的描述(g) 通过单击“ TT”单倍体超级链接获得的详细信息. (h) 相由同一研究人群和病发率的同一 SNP 定义的其他单倍体

Figure 1 The results of searching by 'colorectal cancer'

Note: (a) Searching with 'colorectal cancer'. (b) Searching with 'colorectal cancer' and 'chromosome 7'. (c) The list of results retrieved by searching with 'colorectal cancer'. (d) The download page of retrieved results. (e) The detailed information obtained by clicking the hyperlink of the gene symbol 'ABCB1'. (f) The description of the 'Slovenia population'. (g) Detailed information obtained by clicking the hyperlink for haplotype 'TT'. (h) All the other haplotypes defined by the same SNPs in the same study populations and their frequencies.

献和单倍体。疾病部分集中于结肠直肠癌的发病机理和临床特征的简要概述。如果用户希望更全面地了解疾病及其影响, 他们可以通过超链接进入 Patient UK 或 Wikipedia 网站。文献部分列出了关于结肠直肠癌易感性(或保护性)单倍体的所有文献, 包括 PubMed ID, 出版日期, 标题和摘要。该信息提供了基于单倍体分析的结肠直肠癌的检测和治疗进展的初步见解。单倍体部分呈现所有结肠直肠癌相关的单倍体, 单倍体频率, 相关染色体数目和

基因符号, SNP(或微卫星), 包含单倍体, 单倍体的风险情况, 统计检验的 p 值和研究群体(图 1f)。有关基因或单倍体的更多详细信息, 用户可以单击相关链接, 将出现一个新页面, 如图 1e 和图 1g 所示。表示染色体带上的单倍体位置的图像在左边, 给予用户对单倍体位置的直观指示。

图 1c 表示“风险状况”的查询结果。它具有四个不同的值: 如文献中所述, “风险”和“保护”分别代表单倍体增加或减少疾病风险; “统计推断



风险”和“统计推断保护”分别代表增加或减少疾病风险的单倍体,它们仅存在于关联测试的结果表中。

类似于疾病名称的搜索,用户可以通过基因名称(目前支持 Entrez 基因 ID 和基因符号)搜索数据库。这有助于帮助用户直接识别感兴趣的基因的相关的单倍体。用户还可以通过染色体数目搜索数据库。以复杂疾病相关单倍体为中心的信息按照文章在线公开日期的顺序显示。用户可以了解到在此染色体上的复杂人类疾病单倍体的设计和分析研究进程。此外,用户可以通过 SNP ID (rs#)检索信息。如果查询 SNP 已被确定为我们数据库中的单倍体,搜索结果将在新页面中返回。基本 SNP 信息和相关参考文献的简要描述将帮助用户更好地了解复杂疾病的遗传易感性。用户可以通过点击超文本链接查看感兴趣项目的详细信息。我们的数据库还保留每个查询模型的搜索历史记录,允许用户调用以前的搜索结果。

以不同方式获得的查询结果可以通过视图顶部的下载链接(图 1d)直接下载为 Excel 文件。此外,复杂疾病相关单倍体的所有数据以及相应的分析软件可在下载页面上免费获得。

1.4 提交页面

我们鼓励用户提交未记录的复杂疾病相关单倍体的信息。数据可以通过提交网页直接提交给 CDRH。所需的提交信息包括疾病名称,人群,染色体数,基因符号,单倍体, PubMed ID 和提交者的详细通信信息。所有提交的材料都将接受系统的质量保证审查。

如果提交的文件通过上述检查,提交的记录和其他必要信息将尽快添加到 CDRH。通过从 PubMed 的文献数据库检索的出版物中手动提取相关信息,定期更新 CDRH 中包含的数据。每次更新后,新的和改进的项目的集合将显示在浏览页面的顶部。

2 讨论

了解复杂人类疾病的遗传变异和可遗传风险之间的关系是现代人类遗传学的一个巨大的挑战。这也是发现影响复杂人类疾病的基因的重要一步。为了给研究复杂疾病相关单倍体的分子生物学家和遗传学家提供中心资源,我们收集了大量的信息,这些信息散布在现有研究中,并且已经开发了

一个复杂疾病相关单倍体(CDRH)的数据库。它不仅提供了一个易于使用的界面来查询从文献中提取的有关单倍体和疾病的有价值的参考信息,而且还从外部数据库集成了大量的补充注释。CDRH 数据库清楚地反映了单倍体与复杂疾病之间的关系。因此,它便于收集关于复杂疾病相关单倍体的更全面的的信息,同时,节省研究人员搜索多个数据库和大量文献的麻烦。

目前,在 CDRH 数据库中记录了 1125 个单倍体,涉及 22 个常染色体,染色体 X,染色体 Y 和线粒体。图 2a 是每条染色体上的复杂疾病相关单倍体的数目的直方图。图 2b 是每条染色体上的复杂疾病相关基因的数量直方图。从图 2 可知,绝大多数单倍体(431 个单倍体)和基因(39 个基因)位

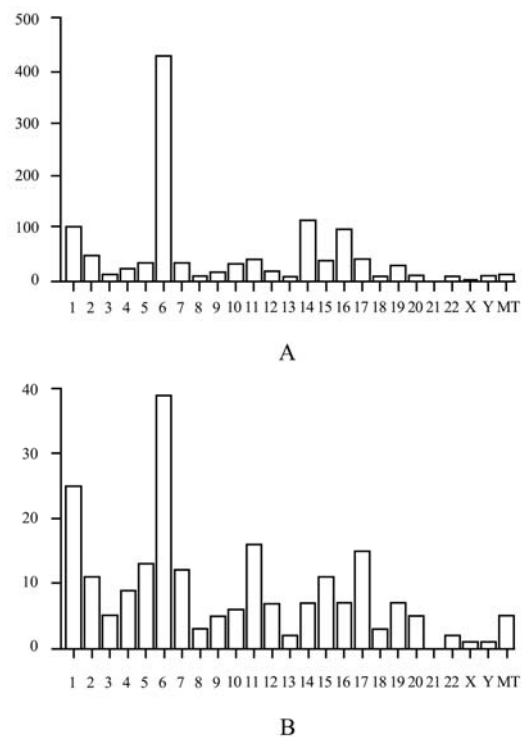


图 2 复杂疾病相关单倍体 CDRH 数据库和基因在染色体上的分布

注: MT; 线粒体 (a) 每条染色体上的复杂疾病相关单倍体数的直方图。(b) 每条染色体上复杂疾病相关的基因数的直方图

Figure 2 The chromosomal distribution of complex disease-related haplotypes and genes in the CDRH database

Note: MT; mitochondrion. (a) Histogram of the number of complex disease-related haplotypes on each chromosome. (b) Histogram of the number of complex disease-related genes on each chromosome



于染色体 6 上。尤其是这些单倍体和基因主要集中在 6p21.3(74.36%)区域。一些以前的研究表明, 该区域与许多复杂的免疫疾病有关, 如 1 型糖尿病 (Noble et al., 1996; Hermann et al., 2003), 类风湿性关节炎 (Newton et al., 2004), 风湿性心脏病 (Hernandez-Pacheco et al., 2003) 和系统性红斑狼疮 (Vargas-Alarcon et al., 2001)。这些结果意味着某些复杂的疾病有一些共同的生物标志物, 可能有潜在的易感基因之间的功能性相互作用。在未来, 更多的研究将使我们对于 6p21.3 区有更深入的理解。

图 2a 还表明在 21 号染色体上没有复杂的疾病相关单倍体。这种现象归因于文献中没有 21 号染色体的精确单倍体信息。

到目前为止, CDRH 数据库有 114 种复杂疾病的记录。表 2 按照单倍体数目的顺序显示了前六种复杂疾病的统计信息。这些疾病涉及至少两个群体, 以及一个以上染色体和基因, 这意味着这些疾病与其他疾病相比更常见, 并且可能由多个基因引

起。在我们的数据库中多发性硬化 (Rosati, 2001) 和类风湿性关节炎 (Harris, 1990) 都至少有两项研究, 这可能意味着研究人员应该更加注意这些疾病。

单倍体可以比单个标记包含更多的信息, 并且可以揭示 SNP 之间的协同效应。因此, 负责某些遗传疾病的单倍体正在开发用于遗传疾病 (尤其是常染色体隐性遗传疾病) 的分子诊断。一些研究 (Basel 等人, 2004; Sossenheimer et al., 1997; Repiso et al., 2005; Lian et al., 2004) 已经表明对于分子疾病诊断和携带者的单倍体分析的重复性好。因此, 通过提供关于复杂疾病相关单元型的详细信息, CDRH 可能有助于设计未来的实验和计算生物学研究。

3 结论

CDRH 是第一个通过收集和记载大量文献来强调单倍体水平的复杂人类疾病数据库。它提供了一个用户容易掌握的界面, 以搜索有关单倍体和疾病的详细信息。我们鼓励研究人员提交有趣的新数据

表 2 医疗数据库中六大复杂疾病的统计信息

Table 2 The statistical information of the top six complex diseases in the CDRH database

Disease name	Haplotypes (n)	SNP/micro ^a (n)	References (n)	Populations (n)	Genes (n)	Chr. num. ^b
Type 1 diabetes mellitus	247	147	15	15	17	4
Multiple sclerosis	64	81	16	15	15	4
Bipolar disorder	58	30	2	2	2	2
Schizophrenia	57	49	8	9	8	8
Age-related macular degeneration	48	61	12	13	13	5
Rheumatoid arthritis	40	58	11	12	12	6

注: SNP/micro: 单核苷酸多态性和微卫星; CHR. Num: 染色体数目

Note: SNP/micro: SNPs and microsatellites; Chr. Num: chromosome number

并提供下载功能。我们致力于维护和更新 CDRH 数据库, 并希望它将指导研究人员更全面地了解复杂的人类疾病。

4 未来展望

随着 SNP 基因分型技术和单倍体分析方法的快速改进, 我们可以方便地获得全基因组 SNP 数据。因此, 基于单倍体的全基因组关联研究可能是一种有效的方式来识别与复杂疾病相关的遗传区域或基因。

致谢

本研究由国家自然科学基金(批准号 81172842, 31200934) 和黑龙江省自然科学基金(批准号 C201206)共同资助。感谢哈尔滨医科大学生物信息学与技术学院统计遗传学研讨会的所有成员。

参考文献

- Basel D., Kilpatrick M.W., and Tsipouras P., 2004, Haplotype analysis enables the diagnosis of Marfan syndrome, *Conn Med*, 68(6): 363-366
- Berger M., Moscatelli H., Kulle B., Luxembourg B., Blouin K., Spannagl M., Lindhoff-Last E., and Schambeck C.M., 2008, Association of ADAMDEC1 haplotype with high



- factor VIII levels in venous thromboembolism, *Thromb Haemost*, 99(5): 905-908
- Daly M.J., Rioux J.D., Schaffner S.F., Hudson T.J., and Lander E.S., 2001, High-resolution haplotype structure in the human genome, *Nat Genet.*, 29(2): 229-232
<http://dx.doi.org/10.1038/ng1001-229>
- Gabriel S.B., Schaffner S.F., Nguyen H., Moore J.M., Roy J., Blumenstiel B., Higgins J., Defelice M., Lochner A., Faggart M., Liu-Cordero S.N., Rotimi C., Adeyemo A., Cooper R., Ward R., Lander E.S., Daly M.J., and Altshuler D., 2002, The structure of haplotype blocks in the human genome, *Science*, 296(5576): 2225-2229
<http://dx.doi.org/10.1126/science.1069424>
- Gao G., Allison D.B., and Hoeschele I., 2009, Haplotyping methods for pedigrees, *Hum Hered*, 67(4): 248-266
<http://dx.doi.org/10.1159/000194978>
- Hapmap, 2003, The International HapMap Project, *Nature*, 426(6968): 789-796
- Harris E.D. Jr., 1990, Rheumatoid arthritis. Pathophysiology and implications for therapy, *N. Engl. J. Med.*, 322(18): 1277-1289
<http://dx.doi.org/10.1056/NEJM199005033221805>
- Hermann R., Turpeinen H., Laine A.P., Veijola R., Knip M., Simell O., Sipila I., Akerblom H.K., and Ilonen J., 2003, HLA DR-DQ-encoded genetic determinants of childhood-onset type 1 diabetes in Finland: an analysis of 622 nuclear families, *Tissue Antigens*, 62(2): 162-169
<http://dx.doi.org/10.1034/j.1399-0039.2003.00071.x>
- Hernandez-Pacheco G., Aguilar-Garcia J., Flores-Dominguez C., Rodriguez-Perez J.M., Perez-Hernandez N., Alvarez-Leon E., Reyes P.A., and Vargas-Alarcon G., 2003, MHC class II alleles in Mexican patients with rheumatic heart disease, *Int. J. Cardiol.*, 92: 49-54
[http://dx.doi.org/10.1016/S0167-5273\(03\)00040-8](http://dx.doi.org/10.1016/S0167-5273(03)00040-8)
- Higasa K., Miyatake K., Kukita Y., Tahira T., and Hayashi K., 2007, D-HaploDB: a database of definitive haplotypes determined by genotyping complete hydatidiform mole samples, *Nucleic Acids Res.*, 35: D685-689
<http://dx.doi.org/10.1093/nar/gkl848>
- Hung H.C., Lin C.Y., Liao Y.F., Hsu P.C., Tsay G.J., and Liu G.Y., 2007, The functional haplotype of peptidylarginine deiminase IV (S55G, A82V and A112G) associated with susceptibility to rheumatoid arthritis dominates apoptosis of acute T leukemia Jurkat cells, *Apoptosis*, 12(3): 475-487
<http://dx.doi.org/10.1007/s10495-006-0005-0>
- Ingman M., and Gyllensten U., 2006, mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences, *Nucleic Acids Res.*, 34: D749-751
<http://dx.doi.org/10.1093/nar/gkj010>
- Kayser M., Brauer S., Willuweit S., Schadlich H., Batzer M.A., Zawacki J., Prinz M., Roewer L., and Stoneking M., 2002, Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for U.S. populations, *J. Forensic Sci.*, 47: 513-519
- Lian J.F., Cui C.C., Xue X.L., Huang C., Cui H.B., and Zhang H.Z., 2004, Long QT syndrome gene diagnosis by haplotype analysis, *Zhonghua Yi Xue Yi Chuan Xue Za Zhi*, 21: 272-273
- Lin D.Y., and Zeng D., 2006, Likelihood-Based Inference on Haplotype Effects in Genetic Association Studies, *Journal of the American Statistical Association*, 101: 104-106
<http://dx.doi.org/10.1198/016214505000000808>
- Newton J.L., Harney S.M., Timms A.E., Sims A.M., Rockett K., Darke C., Wordsworth B.P., Kwiatkowski D., and Brown M.A., 2004, Dissection of class III major histocompatibility complex haplotypes associated with rheumatoid arthritis, *Arthritis Rheum*, 50: 2122-2129
<http://dx.doi.org/10.1002/art.20358>
- Noble J.A., Valdes A.M., Cook M., Klitz W., Thomson G., and Erlich H.A., 1996, The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families, *Am. J. Hum. Genet.*, 59(5): 1134-1148
- Repiso A., Corrons J.L., Vulliamy T., Killeen N., Layton M., Carreras J., and Climent F., 2005, New haplotype for the Glu104Asp mutation in triose-phosphate isomerase deficiency and prenatal diagnosis in a Spanish family, *J. Inher. Metab. Dis.*, 28(5): 807-809
<http://dx.doi.org/10.1007/s10545-005-0098-6>
- Rosati G., 2001, The prevalence of multiple sclerosis in the world: an update, *Neurol Sci*, 22(2): 117-139
<http://dx.doi.org/10.1007/s100720170011>
- Santiago J.L., Martinez A., Nunez C., De La Calle H., Fernandez-Arquero M., De La Concha E.G., and Urcelay



- E., 2008, Association of MYO9B haplotype with type 1 diabetes, *Hum. Immunol.*, 69(2): 112-115
<http://dx.doi.org/10.1016/j.humimm.2008.01.003>
- Slattery M.L., Curtin K., Sweeney C., Wolff R.K., Baumgartner R.N., Baumgartner K.B., Giuliano A.R., and Byers T., 2008, Modifying effects of IL-6 polymorphisms on body size-associated breast cancer risk, *Obesity (Silver Spring)*, 16(2): 339-347
<http://dx.doi.org/10.1038/oby.2007.44>
- Soma H., Yabe I., Takei A., Fujiki N., Yanagihara T., and Sasaki H., 2008, Associations between multiple system atrophy and polymorphisms of SLC1A4, SQSTM1, and EIF4EBP1 genes, *Mov. Disord*, 23(8): 1161-1167
<http://dx.doi.org/10.1002/mds.22046>
- Sossenheimer M.J., Aston C.E., Preston R.A., Gates L.K., Jr., Ulrich C.D., Martin S.P., Zhang Y., Gorry M.C., Ehrlich G.D., and Whitcomb D.C., 1997, Clinical characteristics of hereditary pancreatitis in a large family, based on high-risk haplotype. The Midwest Multicenter Pancreatic Study Group (MMPSG), *Am. J. Gastroenterol*, 92(7): 1113-1116
- Tishkoff S.A., Pakstis A.J., Ruano G., and Kidd K.K., 2000, The accuracy of statistical methods for estimation of haplotype frequencies: an example from the CD4 locus, *Am. J. Hum. Genet.*, 67(2): 518-522
<http://dx.doi.org/10.1086/303000>
- Vargas-Alarcon G., Salgado N., Granados J., Gomez-Casado E., Martinez-Laso J., Alcocer-Varela J., Arnaiz-Villena A., and Alarcon-Segovia D., 2001, Class II allele and haplotype frequencies in Mexican systemic lupus erythematosus patients: the relevance of considering homologous chromosomes in determining susceptibility, *Hum. Immunol.*, 62(8): 814-820
[http://dx.doi.org/10.1016/S0198-8859\(01\)00267-1](http://dx.doi.org/10.1016/S0198-8859(01)00267-1)
- Yaspan B.L., McCreynolds K.M., Elmore J.B., Breyer J.P., Bradley K.M., and Smith J.R., 2008, A haplotype at chromosome Xq27.2 confers susceptibility to prostate cancer, *Hum. Genet.*, 123(4): 379-386
<http://dx.doi.org/10.1007/s00439-008-0486-8>
- Zhao H., Pfeiffer R., and Gail M.H., 2003, Haplotype analysis in population genetics and association studies, *Pharmacogenomics*, 4(2): 171-178
<http://dx.doi.org/10.1517/phgs.4.2.171.22636>