

人免疫球蛋白重链可变区基因的特征分析

石彬^{1,2} 马嘉欣² 涂文静² 吴皓明² 陈先恋¹

(1. 遵义医科大学附属医院, 贵州 遵义 563000; 2. 遵义医科大学检验医学院, 贵州 遵义 563000)

摘要 目的:分析五类免疫球蛋白重链可变区基因的特征. 方法:收集 IMGT/LIGM-DB 数据库中人类五类 Ig 重链可变区的基因序列,利用 IMGT/HighV-QUEST 分析其基因取用、V 区 AA 变化、CDR3 氨基酸长度和构成以及连接多样性. 结果:IgM、IgG 和 IgE 均较多地取用 IGHJ4、IGHV1-69 和 IGHV5-51,且在 IgM、IgG、IgA 与 IgE 中可观察到几个相似的优势 V-J 配对. IgD 具有较高的 V 区突变,主要体现在 FR3 区的高突变(AA 变化值为 7.2 ± 2.4),而 IgM 的每个结构域的突变均明显低于其他类 Ig. 此外,IgD 具有较高 P 插入发生率和插入核苷酸数明显多于其他四类 Ig. 结论:本研究从基因特征上描绘了人类五类 Ig 相互间的相似性与差异性,这些结果可为抗体工程领域提供重要参考数据.

关键词 Ig; 重链可变区基因; CDR3; IMGT/LIGM-DB

中图分类号 R392.9

文献标识码 A

0 引言

Ig(immunoglobulin, Ig)是由 B 细胞合成、分泌的参与介导体液免疫应答重要的效应分子. Ig 重链可变区互补决定区 3(complementarity determining region, CDR3)是其特异性识别和结合抗原的关键区域^[1,2]. 与大多数脊椎动物一样,人类 Ig 基因具有高度的多样性,其多样性产生的机制主要包括重链可变区的 V、D、J 基因片段的重组、各基因片段连接区的核苷酸的插入与删除以及重组后的体细胞超突变^[3,4]. 根据重链的不同,Ig 可划分为 IgG、IgM、IgA、IgD 和 IgE 五大类别. IgG 在血液中的含量最高,是再次免疫应答的主要抗体;IgM 主要出现于初次免疫应答,产生最早且抗体效价高;而 IgA 在局部粘膜免疫应答发挥重要作用;IgD 主要表达于 B 细胞膜表面,发挥细胞识别的作用;而 IgE 则主要参与介导超敏反应^[5]. 尽管过去的研究早已证实五类 Ig 的蛋白结构与功能均有不同程度的差异^[5],然而,对于其 Ig 重链可变区基因的结构和特征,目前尚缺乏系统的描述. 本研究对国际权威数据库 IMGT/LIGM-DB^[6,7] 存储的所有五类 Ig 重链可变区基因数据进行比较,分析它们的可变区基因特征及其相互间的相似性和差异性,这些结果将有助于学者们更好地理解 Ig 基因特性,可为抗体疫苗设计、人工抗体合成及 Ig 免疫组库研究等提供重要参考数据^[8,9].

1 材料与方法

1.1 研究对象

在国际权威免疫遗传信息数据库 IMGT/LIGM-DB 中,储存了数十年来的世界各实验室提交的免疫球蛋白基因数据,通过检索方式获取 IMGT/LIGM-DB 数据库中人类五类免疫球蛋白的重链可变区基因序列. 检索条件选择为:Species: Homo sapiens (human);Molecule type: cDNA;Configuration type: rearranged;Functionality: productive;Chain type: IG-Heavy. 通过该检索,获得了原始序列(IgM 序列 582 条,IgD 序列 132 条,IgG 序列 1839 条,IgA 序列 501 条,IgE 序列 112 条).

收稿日期:2018-11-13

基金项目:国家自然科学基金项目(81760300)资助

通讯作者:石彬,男,汉族,博士,副教授,研究方向:免疫组库与免疫信息,E-mail:shibin_superman@163.com.

1.2 数据处理

将原始序列的 FASTA 格式提交至 IMGT/HighV-QUEST^[10,11] 软件对搜索到的重链可变区基因序列进行比对和鉴定. 通过 IMGT summary 文件, 过滤掉符合以下条件之一的序列: (1)No results; (2)Unknow; (3)Warnings; (4)Unproductive; (5)104/118 氨基酸位点非 C/W; (6)连接区未鉴定; (7)CDR 区未鉴定; (8)重复序列. 最终获取用于分析的 unique productive 的序列为 IgM 序列 547 条, IgD 序列 130 条, IgG 序列 1816 条, IgA 序列 472 条, 以及 IgE 序列 97 条(具体过滤信息见表 1).

1.3 软件与统计

IMGT/HighV-QUEST(版本 1.2.0)用于鉴定序列(VH, DH, JH 和 CDR3), 功能评估和序列数据统计分析; IMGT/V-QUEST(版本 3.2.32)用于鉴定 CDR3 的位置; Microsoft Office Excel(版本 2010)用于序列的存储、过滤和统计; GraphPad Prism7 用于绘图和假设检验. 氨基酸和核苷酸数目用 mean±SD 表示, 其数据的假设检验在两组数据的比较时采用 t 检验, 而在多组数据的比较时采用 ANOVA. 检验水准以 $p < 0.05$ 定义为具有统计学差异, 图中“*”表示 $p < 0.05$.

表 1 数据过滤

| Ig 类别 | 原始序列 | No results | Unknow | Unproductive | 104/118 位点非 C/W | 连接区未鉴定 | CDR 区未鉴定 | 重复序列 | unique productive |
|-------|------|------------|--------|--------------|-----------------|--------|----------|------|-------------------|
| IgM | 582 | 7 | 8 | 2 | 3 | 9 | 2 | 4 | 547 |
| IgD | 132 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 130 |
| IgG | 1839 | 4 | 3 | 1 | 3 | 1 | 1 | 11 | 1 816 |
| IgA | 501 | 9 | 0 | 0 | 0 | 6 | 3 | 11 | 472 |
| IgE | 112 | 4 | 2 | 0 | 0 | 0 | 1 | 8 | 97 |

注: 表中的数值表示序列数目.

2 结果

2.1 基因取用

人类 IGHV 基因包含 71 个等位基因^[12,13], 图 1 显示了五类 IGHV 亚群基因的频率. 与其他 Ig 显著不同的是, IgD 表现为对 IGHV4-34 的偏向取用(84.6%), 而其他类别 Ig 对该基因的取用均不超过 6%. IgA 对 IGHV1-69 的取用频率为 2.8%, 而 IgM、IgG 与 IgE 对其多出了数倍的取用(分别为 13.3%、8.9%、10.3%). 值得关注是, IGHV5-51 在 IgM、IgG、IgA 与 IgE 四类 Ig 中均显示了较高频率的取用(分别为 10.1%、10.2%、9.5%、14.4%). 进一步的 V-J 基因配对的热图显示, 各类 Ig 的配对图谱有明显差别(图 2). IgM、IgG、IgA 与 IgE 对 IGHJ4 的取用频率均接近 50%, 而 IgD 对其取用仅为 2.3%. IgD 对 IGHJ3 的取用高达 78.5%, 而其他四类 Ig 对其取用均低于 23%. 有趣的是, 尽管不同 Ig 间的基因取用差别明显, 但五类 Ig 依然存在三种相似的优势 V-J 配对模式, 包括 IGHV5-51/IGHJ4、IGHV5-51/IGHJ3 以及 IGHV3-30/IGHJ4(表 2).

表 2 五类 Ig 的 top 3 优势 V-J 配对

| Ig 类别 | 1 | 2 | 3 |
|-------|-----------------------------------|-----------------------------------|---|
| IgM | IGHV1-69/IGHJ4(6.0%) | IGHV5-51/IGHJ4(5.3%) ^a | IGHV6-1/IGHJ4(4.4%) |
| IgD | IGHV4-34/IGHJ3(68.5%) | IGHV4-34/IGHJ6(16.2%) | IGHV4-61/IGHJ3(7.7%) |
| IgG | IGHV5-51/IGHJ4(3.8%) ^a | IGHV5-51/IGHJ3(3.7%) ^b | IGHV3-30/IGHJ4(3.4%) ^c |
| IgA | IGHV3-73/IGHJ4(5.1%) | IGHV3-30/IGHJ4(3.8%) ^c | IGHV5-51/IGHJ4(3.6%) ^a |
| IgE | IGHV1-69/IGHJ6(7.2%) | IGHV5-51/IGHJ4(6.2%) ^a | IGHV5-51/IGHJ3(5.2%) ^b IGHV4-30-4/IGHJ4(5.2%) |

注: 1/2/3 分别代表频率排名前三的配对; 括号中数值表示配对频率; 相同模式的配对用相同小写字母上标注释.

2.2 V 区的 AA 变化

Ig V 区的体细胞超突变与抗体亲和力成熟关系密切, 由于其突变水平高, 因此, 一般较少使用核苷酸水平的突变评价, 而是推荐使用 AA 变化(Amino acid changes, 氨基酸变化)对其进行分析^[14,15]. 完整的 V 区包含 FR1、CDR1、FR2、CDR2、FR3、CDR3 和 FR4, 但由于 CDR3 和 FR4 参与了重排进程, 因此, 我们在分析 V 区的体细胞高频突变时未包含这两个结构域. 图 3 显示了五类 Ig 的 V 区 AA 变化为 0 时的 IgM 序列(28.0%)

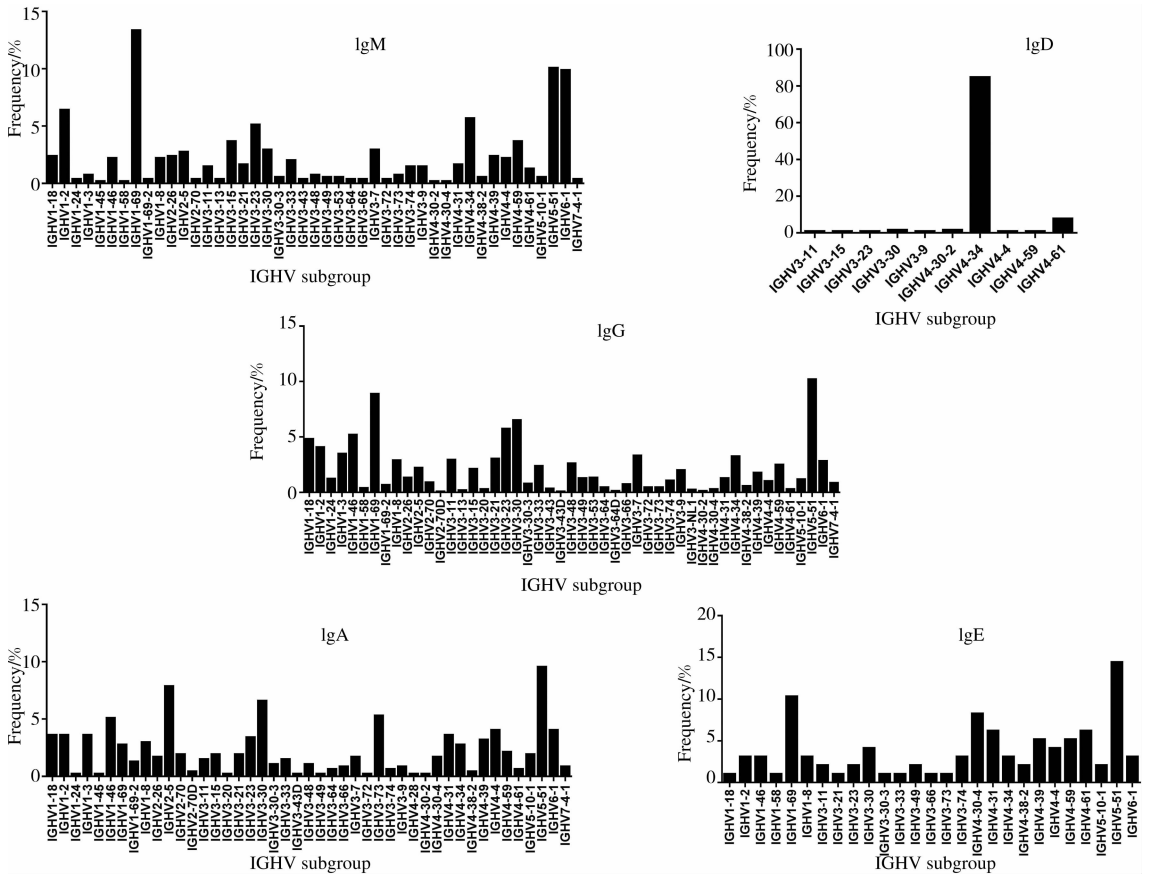


图 1 五类 IGHV 基因亚群的频率

明显多于其他四类 Ig(均低于 3.2%)。值得注意的是,AA 变化数小于或等于 10 时,IgD 序列仅占比 6.9%,IgM 序列比例为 91.4%,而其他 Ig 均在 40%-60%。而当 AA 变化数大于或等于 11 时,IgD 序列比例高达 93.1%。IgM 序列的 AA 变化均值为 3.7 ± 4.6 ,显著低于其他类 Ig(*t* 检验,每个 $p < 0.05$),而 IgD 序列的 AA 变化均值为 14.7 ± 4.3 ,显著高于其他类 Ig(*t* 检验,每个 $p < 0.05$)。IgG、IgA 与 IgE 的 AA 变化均值分别为 11.7 ± 5.9 、 11.0 ± 5.5 以及 9.6 ± 5.2 。因此,这些结果提示 IgD 倾向于较高的 V 区突变水平,而 IgM 具有较低的 V 区突变水平。

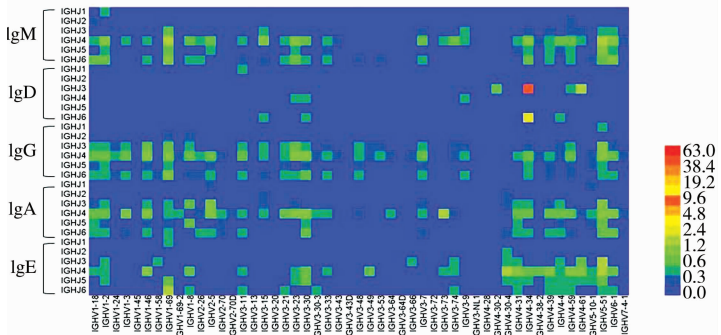


图 2 五类 Ig 的 V-J 基因配对热图

进一步的 V 区各结构域(FR1、CDR1、FR2、CDR2 和 FR3)的 AA 平均变化数表明,IgM 的每个 V 区结构域的突变水平均低于其他类 Ig (ANOVA,每个 $p < 0.05$),而 IgD 则主要体现在 FR3 区的高突变水平(AA 变化值为 7.2 ± 2.4 ,图 4)。

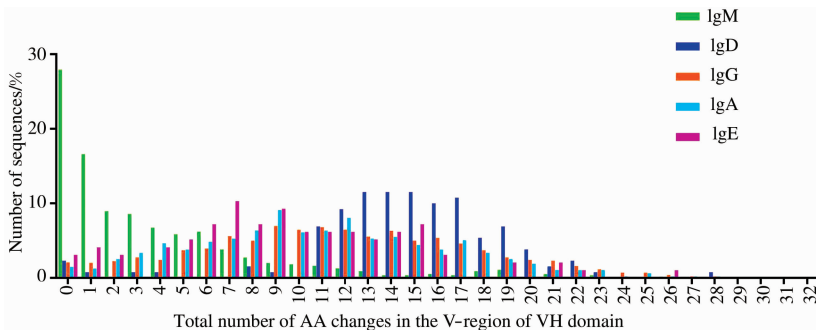


图 3 五类 Ig 的 AA 变化的分布

2.3 CDR3 区构成分析

Ig 可变区的互补决定区 (complementarity determining region, CDR) 包含重/轻链 CDR1、CDR2 以及 CDR3, 其中重链 CDR3 是识别和结合抗原的关键区域^[1,2]. 通过筛选, 我们发现 IgM、IgD、IgG、IgA 和 IgE 的独特型 CDR3 序列数目分别为 483、74、1 656、407 和 85. 图 5 显示五类 Ig 的 CDR3 AA 长度分布谱虽不同, 但 IgM、IgG、IgA、IgE 的 CDR3 平均长度较为相似. 值得注意的是, IgD 的 CDR3 平均长度显著多于其他四类 Ig (ANOVA, $p < 0.05$) (图 6). 图 7 显示了五类 Ig 的 20 种 AA 的取用 (显示了 20 种 AA 的缩写, 从左往右按照疏水性由强到弱排列), 表明 IgD CDR3 更多地取用精氨酸 (R)、脯氨酸 (P)、苏氨酸 (T)、甘氨酸 (G)、丙氨酸 (A), 其相对较少地取用谷氨酸 (E)、组氨酸 (H)、酪氨酸 (Y)、色氨酸 (W)、苯丙氨酸 (F). 而其他四类 Ig CDR3 在 AA 取用上较为相似.

2.4 连接多样性

连接多样性是由于在 V→D 和 D→J 连接处的回文序列“P”核苷酸 (P 区) 和“N”核苷酸 (N 区) 的插入以及核酸外切酶修剪产生的核苷酸删除所致^[16,17]. 值得注意的是, IgD 的 P 区 (P3'V + P5'D + P5'J) 插入发生率为 43.1%, 明显高于其他类 Ig (均值为 $(23.6 \pm 2.1)\%$). 且表 3 显示 IgD 在 P 区添加的平均核苷酸数为 2.9 ± 3.4 , 显著多于其他四类 Ig (t 检验, 每个 $p < 0.05$). 此外, 五类 Ig 的 N 区 (N1 + N2) 插入发生率较为相似 (均超过 98%). IgM 的 N 插入平均核苷酸数相对较少 (11.5 ± 7.6), 而 IgD 与 IgG 的 N 插入平均核苷酸数相对较多 (分别为 14.3 ± 10.6 、 14.0 ± 8.3). 在核苷酸删除方面, IgM 与 IgD 删除的平均核苷酸数相对较少 (均为 18.3 ± 8.2), IgE 的平均核苷酸删除数相对较多 (20.8 ± 9.5 , 表 3). 进一步地调查插入和删除谱 (即不同插入或删除核苷酸数的分布) 显示, IgD 的 P/N 插入谱和删除谱均明显不同于其它类 Ig (图 8). IgD 的 P 插入主要体现在较多数目 (5-7 个核苷酸) 的核苷酸插入比例较多, 而其 N 插入和删除均有明显的峰值. 有趣的是, 其他四类 Ig 的 P 插入谱极为相似.

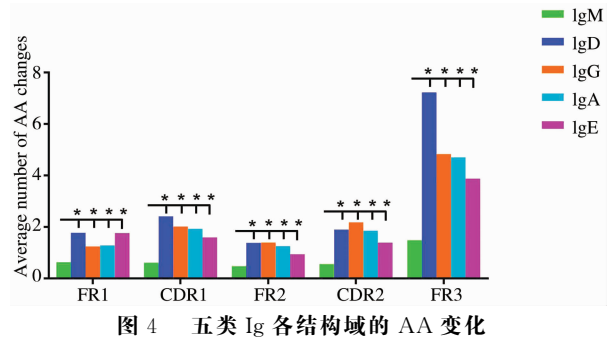


图 4 五类 Ig 各结构域的 AA 变化

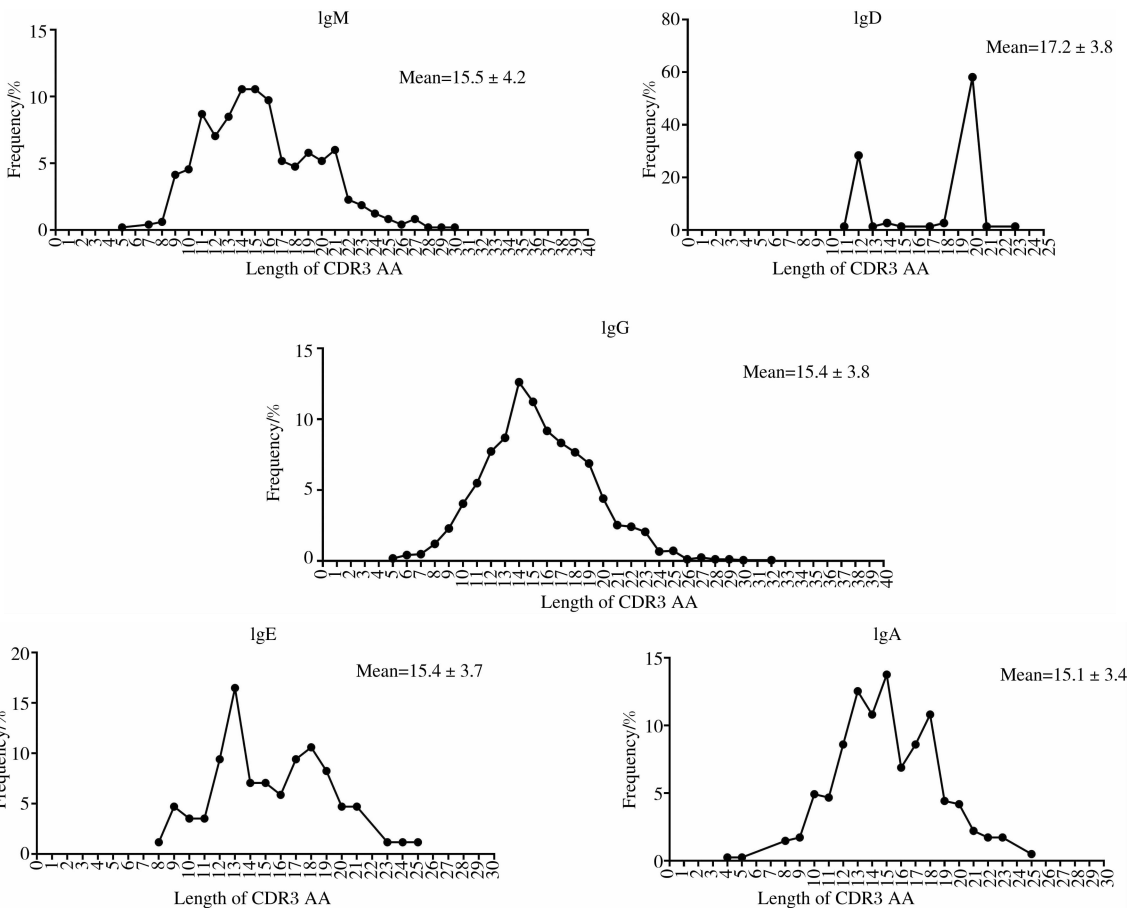


图 5 五类 Ig 的 CDR3 长度分布

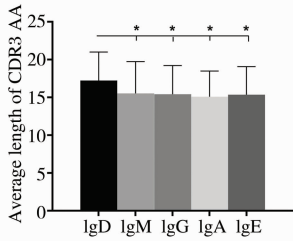


图 6 五类 Ig 的 CDR3 长度比较

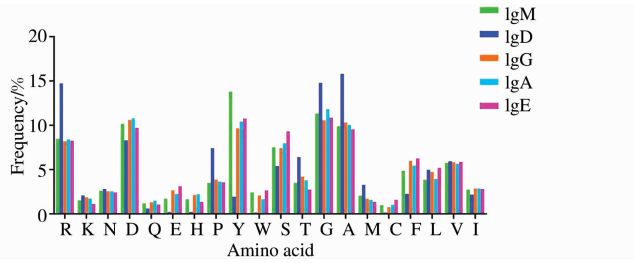


图 7 五类 Ig 的 CDR3 氨基酸构成

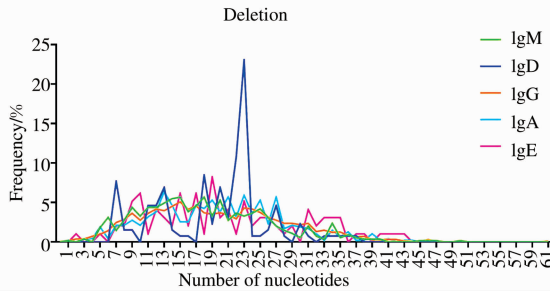
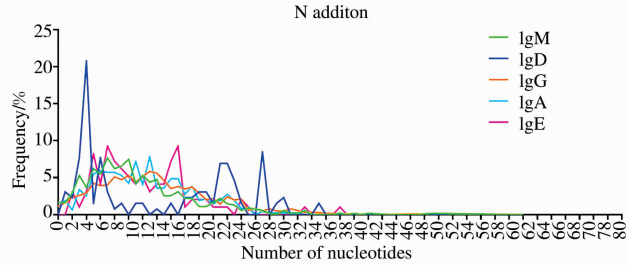
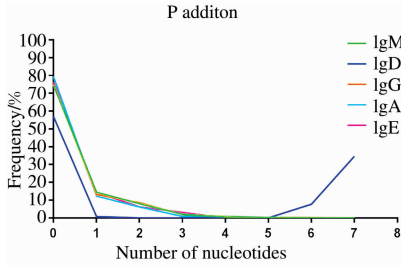


图 8 五类 Ig 的 N/P 插入和删除谱

表 3 五类 Ig 插入与删除的平均核苷酸数

| Ig 类别 | P 插入 | N 插入 | 删除 |
|-------|------------------------|--------------------------|-------------------------|
| IgM | 0.4 ± 0.8 ^a | 11.5 ± 7.6 | 18.3 ± 8.2 |
| IgD | 2.9 ± 3.4 | 14.3 ± 10.6 ^b | 18.3 ± 8.2 |
| IgG | 0.4 ± 0.8 ^a | 14.0 ± 8.3 ^b | 19.4 ± 6.8 |
| IgA | 0.3 ± 0.8 ^a | 12.5 ± 6.8 | 19.5 ± 8.8 |
| IgE | 0.4 ± 0.7 ^a | 12.3 ± 6.9 | 20.8 ± 9.5 ^c |

注: P 插入组中 IgD 与其他 Ig 两两比较有差异的用 a 上标表示, N 插入组中 IgM 与其他 Ig 两两比较有差异的用 b 上标表示, 删除组中 IgM 与其他 Ig 两两比较有差异的用 c 上标表示。

3 讨论

在 Ig 形成的过程中, 多个不连续的 V、D、J 基因会发生随机重排, 并在核酸酶的辅助下组装成完整的 Ig 转录本^[18,19]. 早期研究已证实, 重链可变区的 CDR3 是 Ig 特异性识别和结合抗原的关键区域^[1,2], 其氨基酸含量以及排序将影响 Ig 的功能. 五类 Ig 在功能表现上各有不同已是众所周知, 然而, 目前国内鲜有关于比较人类五类 Ig 重链可变区基因研究报道. 为了更好地理解人类 Ig 形成机制和特征, 本研究对 IMGT/LIGM-DB 数据库中储存的所有人类免疫球蛋白重链可变区基因进行了分析.

基因取用结果表明, 五类 Ig 在 V/J 基因取用偏好性上存在部分相似性. 例如, IgM、IgG 与 IgE 均较多地取用 IGHV1-69, IgM、IgG、IgA 与 IgE 均较多地取用 IGHV5-51, IgM、IgG、IgA 与 IgE 对 IGHJ4 的取用频率均接近 50%, 并且 IgM、IgG、IgA 与 IgE 间可以观察到几个相似的优势 V-J 基因配对; 然而, IgD 与其他四类 Ig 差异明显, 这可能与 IgG、IgA、IgE 来源于 IgM 的类别转换有关. V 区 AA 变化分析显示, IgD 具有较高的 V 区突变水平, 主要体现在 FR3 区高突变水平(AA 变化值为 7.2 ± 2.4). 而 IgM 具有相对较低的 V 区突变水平, 其每个结构域的突变水平均明显低于其他类 Ig, 这与以往的人体疾病与健康 IgM 库的结果相似^[20,21].

此外,连接多样性的结果表明,IgD 具有较高 P 插入发生率,且 P 插入的核苷酸数明显多于其他四类 Ig. 同时,IgD 具有较多的 N 插入和较少的删除,因此,结果显示 IgD 拥有相对较长的 CDR3 长度(约 17 AA). 而其他四类 Ig 长度均较为接近(约 15 AA),与以往小样本实验观察的结果一致^[20]. 此外,IgD 的 CDR3 更多的取用精氨酸(R)、脯氨酸(P)、苏氨酸(T)、甘氨酸(G)、丙氨酸(A). CDR3 区由于其在抗原识别和结合过程中发挥的关键作用,被广泛应用于人工抗体合成和抗体疫苗设计领域^[8,9]. 实际上,抗体是 Ig 的分泌形式,它们的基因转录本是完全一致的. 早期研究证实,CDR3 区单个氨基酸残基的改变可影响抗体的结构和功能^[22],且 CDR3 的长度也与不同类别的抗原识别有关^[23]. CDR3 在应用于合成特定功能的抗体时,需构建大批量的合成库以筛选出高亲和力的抗体,而其中多数低亲和力或无功能的抗体将会被过滤掉^[2]. 因此,我们分析的五类 Ig 的 CDR3 的特征信息可为不同型别抗体相关的合成设计提供参考数据. 当前,IMGT/LIGM-DB 中人的 Ig 重链可变区基因数据仍不充足且缺少细致的分类,因此,利用 NGS 对不同疾病患者和健康志愿者进行大规模的 Ig 组库测序和构建 Ig 大数据专用的分类存储数据库很有必要的,这些工作的开展与积累或将为抗体工程领域带来新的曙光.

4 结论

本研究利用 IMGT/HighV-QUEST 分析了 IMGT/LIGM-DB 来源的人类五类 Ig 重链可变区基因数据,从基因特征上解读了人类五类 Ig 相互间的相似性与差异性,这些结果可作为基于 NGS 结果的 IMGT 标准化比较、抗体工程领域的参考数据.

参 考 文 献

- [1] Flajnik M F. Comparative analyses of immunoglobulin genes: surprises and portents[J]. *Nat Rev Immunol*, 2002, 2(9): 688-698.
- [2] Zemlin M, Klinger M J, Zemlin C, et al. Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures[J]. *J Mol Biol*, 2003, 334(4): 733-749.
- [3] Jung D, Giallourakis C, Mostoslavsky R, et al. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus [J]. *Annu Rev Immuno*, 2006, 24: 541-570.
- [4] Wang Y, Sundling C, Wilson R, et al. High-resolution longitudinal study of HIV-1 Env vaccine-elicited B cell responses to the virus primary receptor binding site reveals affinity maturation and clonal persistence[J]. *J Immunol*, 2016, 196(9): 3729-3743.
- [5] Fahey J L. Antibodies and immunoglobulins I Structure and function[J]. *Jama*, 1965, 194(1): 71-74.
- [6] Lefranc M P, Giudicelli V R, Duroux P, et al. IMGT(R), the international ImMunoGeneTics information system(R) 25 years on[J]. *Nucleic Acids Res*, 2014, 43(Database issue): D413-422.
- [7] Shi B, Ma L, He X, et al. Comparative analysis of human and mouse immunoglobulin variable heavy regions from IMGT/LIGM-DB with IMGT/HighV-QUEST[J]. *Theor Biol Med Model*, 2014, 11: 30.
- [8] Knappik A, Ge L, Honegger A, et al. Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides[J]. *J Mol Biol*, 2000, 296(1): 57-86.
- [9] Zhu H, Luo H, Yan M, et al. Autoantigen microarray for high-throughput autoantibody profiling in systemic lupus erythematosus[J]. *Genomics Proteomics Bioinformatics*, 2015, 13(4): 210-218.
- [10] Li S, Lefranc M P, Miles J J, et al. IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling[J]. *Nat Commun*, 2013, 4: 2333.
- [11] Aouinti S, Malouche D, Giudicelli V, et al. IMGT/HighV-QUEST statistical significance of IMGT clonotype (AA) diversity per gene for standardized comparisons of next generation sequencing Immunoprofiles of Immunoglobulins and T cell receptors[J]. *PLoS One*, 2015, 10(11): e0142353.
- [12] Lefranc M P. Immunoglobulin and T cell receptor genes: IMGT and the birth and rise of immunoinformatics[J]. *Front Immunol*, 2014, 5: 22-26.
- [13] Aouinti S, Giudicelli V, Duroux P, et al. IMGT/Stat clonotype for pairwise evaluation and visualization of NGS IG and TR IMGT clonotype (AA) Diversity or expression from IMGT/HighV-QUEST[J]. *Front Immunol*, 2016, 7: 339-342.
- [14] Giudicelli V, Duroux P, Kossida S, et al. IG and TR single chain fragment variable (scFv) sequence analysis: a new advanced functionality of IMGT/V-QUEST and IMGT/HighV-QUEST[J]. *BMC Immunol*, 2017, 18(1): 35-39.
- [15] Prabakaran P, Chen W, Singarayan M G, et al. Expressed antibody repertoires in human cord blood cells: 454 sequencing and IMGT/HighV-QUEST analysis of germline gene usage, junctional diversity and somatic mutations[J]. *Immunogenetics*, 2012, 64(5): 337-350.

- [16] Giudicelli V, Lefranc MP. IMGT/junctionanalysis: IMGT standardized analysis of the V-J and V-D-J junctions of the rearranged immunoglobulins (IG) and T cell receptors (TR)[J]. Cold Spring Harb Protoc, 2011, 2011(6): 716-725.
- [17] Shi B, Long M, He X, et al. Compositional characteristics of human peripheral TRBV pseudogene rearrangements[J]. Scientific Reports, 2018, 8(1):5926.
- [18] Roy A, Bystry V, Bohn G, et al. High resolution IgH repertoire analysis reveals fetal liver as the likely origin of life-long, innate B lymphopoiesis in humans[J]. Clin Immunol, 2017, 183: 8-16.
- [19] Longo N S, Rogosch T, Zemlin M, et al. Mechanisms that shape human antibody repertoire development in mice transgenic for Human Ig H and L chain loci[J]. J Immunol, 2017, 198(10): 3963-3977.
- [20] Wu Y C, James L K, Vander Heiden J A, et al. Influence of seasonal exposure to grass pollen on local and peripheral blood IgE repertoires in patients with allergic rhinitis[J]. J Allergy Clin Immunol, 2014, 134(3): 604-612.
- [21] Thiele B, Kloster M, Alawi M, et al. Next-generation sequencing of peripheral B-lineage cells pinpoints the circulating clonotypic cell pool in multiple myeloma[J]. Blood, 2014, 123(23): 3618-3621.
- [22] Short M K, Jeffrey P D, Demirjian A, et al. A single H:CDR3 residue in the anti-digoxin antibody 26-10 modulates specificity for C16-substituted digoxin analogs[J]. Protein Eng, 2001, 14(4):287-296.
- [23] Collis AV, Brouwer A P, Martin A C. Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen[J]. J Mol Biol, 2003, 325(2):337-354.

Characterization of the Heavy Chain Variable Region Gene of Human Immunoglobulins

SHI Bin^{1,2} MA Jia-xin² TU Wen-jing² WU Hao-ming² CHEN Xian-lian²

(1. Department of Laboratory Medicine, The First Affiliated Hospital of Zunyi Medical University, Zunyi 563000, China;

2. School of Laboratory Medicine, Zunyi Medical University, Zunyi 563000, China)

Abstract Objective: To analyze the characteristics of five types of immunoglobulin heavy chain variable region genes. Method: The gene sequences of the heavy chain variable regions of human Ig were obtained from the IMGT/LIGM-DB database. An extensive analysis of these genes was implemented using IMGT/HighV-QUEST, including gene usage, amino acid (AA) changes in V region, length distribution and composition of CDR3 amino acid and junctional diversity. Results: IGHJ4, IGHV1-69 and IGHV5-51 were more frequently used in IgM, IgG and IgE. And several similar superior V-J pairings were observed in IgM, IgG, IgA and IgE. There was a higher mutation level of V region in IgD which mainly reflected in FR3 region (number of AA changes= 7.2 ± 2.4), while the mutations in each domain of IgM were significantly lower than those of other Igs. In addition, IgD has a higher incidence of P insertion and a higher number of nucleotides than other Igs. Conclusion: This study describes the similarities and differences among human five types of Ig from the genetic characteristics. These results can provide important reference data for the field of antibody engineering.

Key words Ig; heavy chain variable region gene; CDR3; IMGT/LIGM-DB