

# SNP研究全攻略——从临床到动物模型

SNP这一概念首次出现于1994年发表的人类分子遗传杂志上，随后于1996年由美国麻省理工的Lander正式提出。SNP研究被认为是联系基因型和表现型之间关系的桥梁，是研究人类基因组计划走向应用的重要步骤。

生活中你是否会有种疑惑：为什么有的人喝了牛奶长得又高又壮，有的人喝了牛奶就拉稀？为什么有的人吸烟身体变差，有的人却非常长寿？为什么有的人吃药病好了，有的人吃药却不管用？面对这些困扰无数人的差异，大家是否感觉受到了命运的捉弄。今天，小编就来聊一聊造成某些人“天赋异禀”的关键：SNP。注意，小编说的SNP可不是韩国面膜，而是Single Nucleotide Polymorphisms，即单核苷酸多态性。

SNP这一概念首次出现于1994年发表的人类分子遗传杂志上，随后于1996年由美国麻省理工的Lander正式提出。SNP研究被认为是联系基因型和表现型之间关系的桥梁，是研究人类基因组计划走向应用的重要步骤[1,2]。



图1 美国学者Eric S. Lander于1996年正式提出SNP为第三代分子标记

## SNP的定义

SNP主要是指在基因组水平上由单个核苷酸的变异所引起的DNA序列多态性，即在群体中，基因组内特定核苷酸位置上存在两种不同的核苷酸，并且其出现的频率大于1%。

这里，还需注意另外两个概念，点突变（point mutation）及单核苷酸变异（SNV, single nucleotide variant）。如果单个核苷酸的变异在群体中出现频率低于1%，则看作点突变；如果只是在病人体内检测到单个核苷酸的变异，而其在人群中出现的频率未知，则可看作SNV。

从理论上来看，每一个SNP位点都可以有4种不同的变异形式（C、G、T、A）但实际上发生的只有两种，即转换（transition）（C变换为T，在其互补链上则为G变换为A）或颠换（transversion）（C变换为

A, G变换为 T, C变换为 G, A变换为 T) (图2)。因此通常所说的SNP都是二等位多态性的。也就是说, 大多数情况下, C都是变成T, 而变成A和G的概率很小, 所以一般认为SNP是二等位的, 或者是二态性。

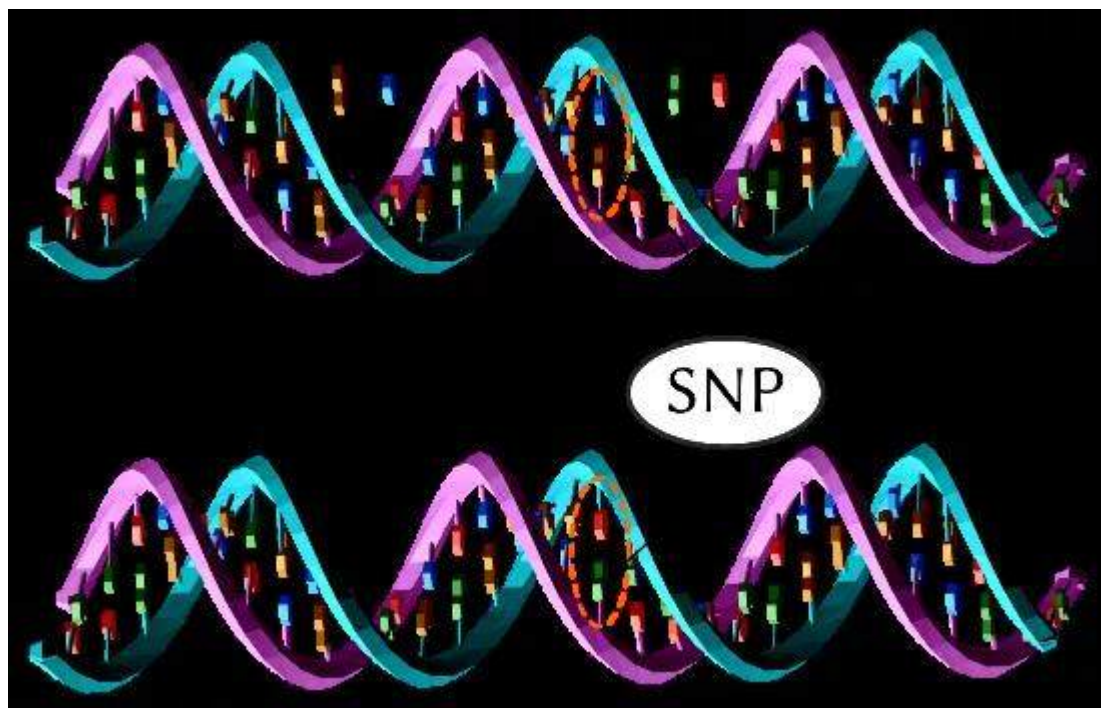


图2 基因组上的SNP位点

## SNP与人类健康

SNP 在人类基因组中的发生频率比较高, 大约平均每1000个碱基对中就有一个多态位点, 估计其总数可达300万个甚至更多[2]。当SNP发生在基因编码区或基因的调节区域时, 它会对基因表达产生巨大的影响, 从而影响基因的功能。有一些SNP位点已被证明在人类健康研究中非常重要, 例如镰状细胞贫血、 $\beta$ 地中海贫血和囊性纤维化的发生主要是由于SNP的出现。同时, 研究人员发现, SNP可能有助于预测个体对某些药物的反应、对毒素等环境因素的易感性以及罹患特定疾病的风险。例如, 携带有APOE(载脂蛋白) $\epsilon$ 4突变的个体表现出患阿尔茨海默病的更高风险[3,4](图3)。

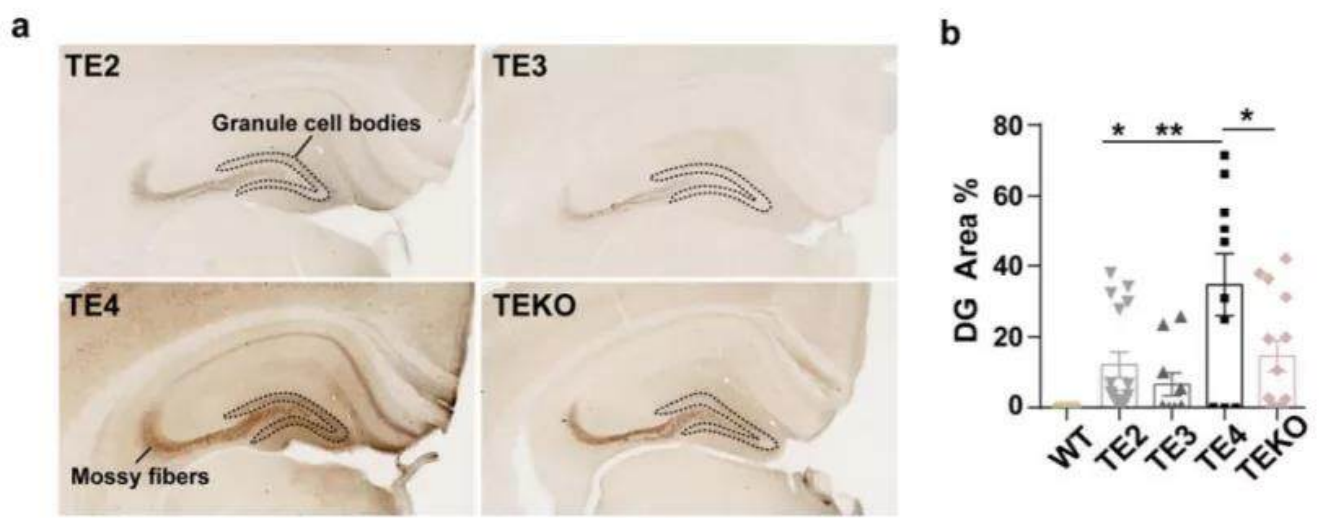


图3 APOE4促进病理性tau蛋白的再分配。

a: 人源APOE敲入小鼠海马体tau蛋白染色结果; b: Tau蛋白阳性区域占海马体齿状回部位的面积 (TE2:APOE2敲入小鼠, TE3: APOE3敲入小鼠, TE4: APOE4敲入小鼠, TEKO: ApoE敲除小鼠)[4]

## 全基因组范围内SNP关联分析

随着测序技术的发展, 全基因组关联分析 (Genome-wide Association Study, GWAS) 已广泛应用于复杂疾病的遗传因素分析 (图4), 可以更高效地获得疾病与SNP之间的相关性。GWAS利用高通量测序技术, 分析数以万计的SNPs以及这些SNPs与临床表型和可测性状的相关性, 从而全面地揭示出不同复杂性状的遗传机制。目前, 全世界范围内已开展了4000多项GWAS研究, 发现了超过10万个与各种疾病 (如癌症, 高血压, II型糖尿病, 类风湿性关节炎等) 以及重要生理性状关联的基因位点。

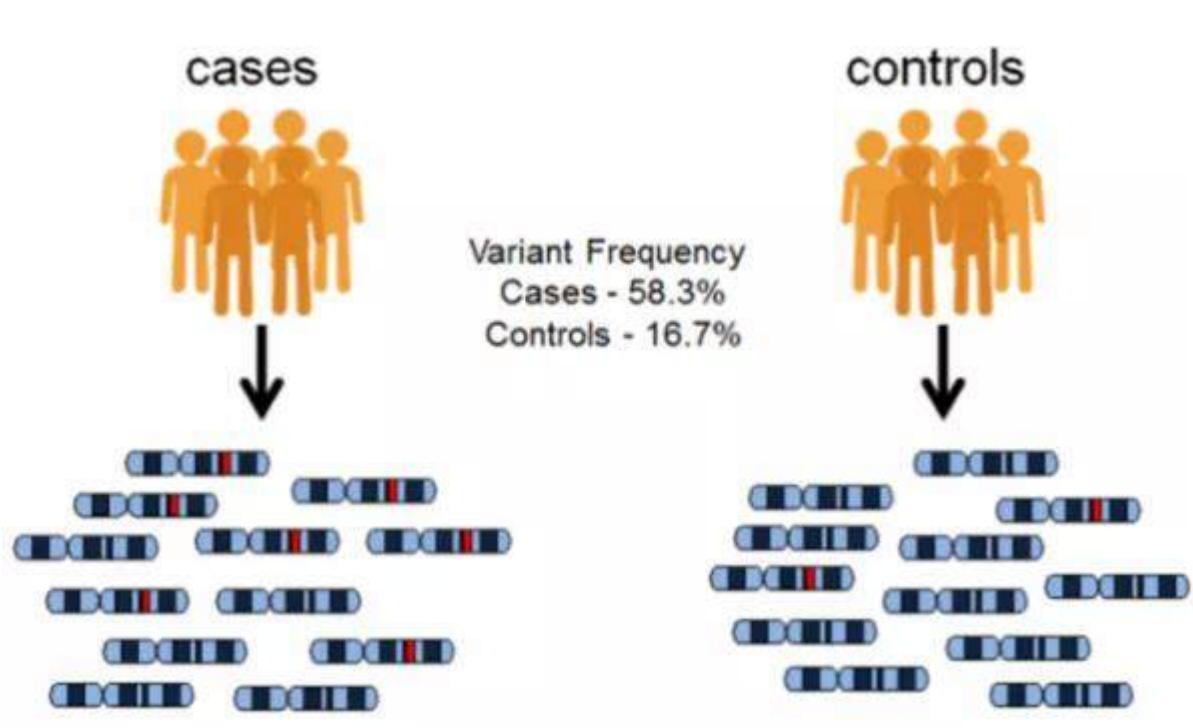




图4 GWAS分析人群中位基因的变异频率 (<https://www.ebi.ac.uk/>)

完成了GWAS分析之后，我们还需要借助下游工具对GWAS的分析结果进一步注释和挖掘，筛选可能致病的SNP位点。相关的注释网站和数据库很多，以下推荐2个网站，供大家参考。GWAS4D是通过组织/细胞类型特异性和非编码调控变异的优先排序，来预测特定复杂疾病和性状的潜在致病机制。我们将GWAS summary数据导入，便可以输出SNP位点的优先排序、SNP位点在基因组区域的可视化以及SNP功能预测和注释等结果（图7）。网址：<http://mulinlab.tmu.edu.cn/gwas4d>。

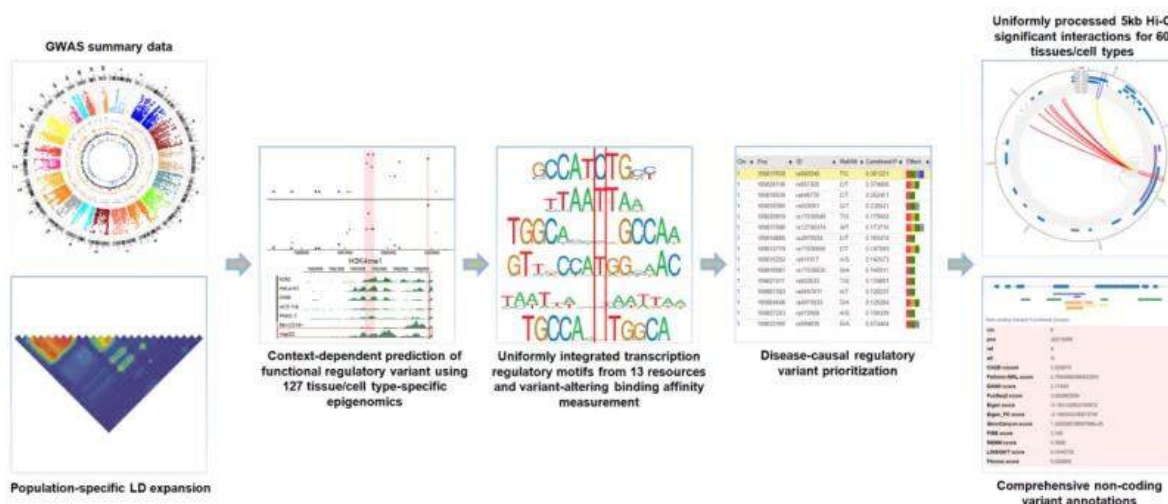


图5 GWAS4D可进行分析的界面

SNPedia是一个SNP百科全书类网站，它引用已经发布的文章或者数据库信息，对SNP位点进行描述，共享着人类基因组变异的信息。我们可以搜索某个SNP位点来寻找与之相关的信息，也可以根据相关疾病和症状来寻找相关的SNP（图6）。网址：<https://www.snpedia.com/index.php/SNPedia>。

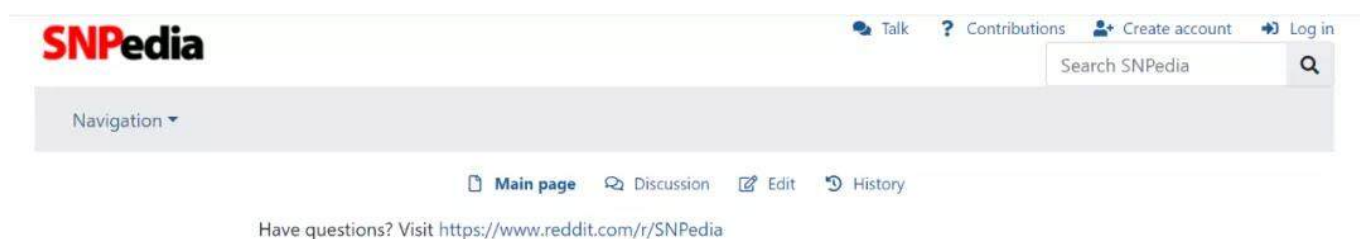


图6 SNPedia首页

## 类人SNP点突变小鼠的构建

受益于二代测序及GWAS的联合应用，研究人员发现了众多和临床表型具有一定联系的SNP位点，然而要验证这些位点与临床表现的关系，必须回到体内进行研究。如果这些突变可以在小鼠体内找到，那么突变小鼠就能为临床研究提供宝贵的信息。因此构建突变小鼠，是有效验证临床检测结果的关键一步。

构建突变小鼠时，首先需要确定的是突变位点的人鼠同源性及保守性。小鼠是解析人类基因功能、研究人类疾病重要的模式生物。99%的人类基因都可以在其基因组中找到相应的基因。但需要注意的是，人鼠编码

基因序列相似性平均为85%，而非编码基因的相似度只有50%[5]，因此找到同源基因还需要判断突变位点的序列保守性。

这里举一个例子，PLD3的一个SNP (rs145999145) 是迟发性阿尔兹海默症的危险因素之一。我们在NCBI-SNP数据库可搜索到该SNP的位点信息，如下：

☐ rs145999145 [Homo sapiens]

1.
 

Variant type: SNV  
 Alleles: G>A [Hide Flanks]

AAAGACAGAGACCAGACTGGGGAGTGTCCCTGTCATCTGTGAGCACTAGG  
 CCGCTATCGCTGAGCTCAGCACTGCCCTCCTACAGGTCAAGGAGCTGGGC  
 [G/A]  
 TGGTCATGTACAACCTGCAGCTGCCTGGCTCGAGACCTGACCAAGATCTTT  
 GAGGCTACTGGTTCCTGGGCCAGGCAGGCAGCTCCATCCCATCAACTTG

Chromosome: 19:40371688 (GRCh38)  
 19:40877595 (GRCh37)

Canonical SPDI: NC\_000019.10:40371687:G:A  
 Gene: PLD3 (Varview)  
 Functional Consequence: coding\_sequence\_variant,missense\_variant  
 Clinical significance: uncertain-significance  
 Validated: by frequency,by alfa,by cluster  
 MAF: A=0.004998/421 (ALFA)  
 A=0.00006/1 (TOMMO)  
 A=0.000599/3 (1000Genomes)

图7 rs145999145在NCBI数据库中的信息页

可见红色框中提供了突变位点的序列信息，而绿色框中提示了突变类型是错义突变（图7）。根据该位点附近的侧翼序列以及给到的基因组位点，我们可以找到其在人源PLD3基因上的位置，转录本NM\_012268.4, c.694G>A, 也即p.232V>M, 位于exon9。通过对人鼠同源基因的同源性分析（图8），可以很快找到相应小鼠的突变位点：NM\_001317355.1, c.688 G>A, p.230V>M, 位于exon10。

```

ENSMUSP000000112942/1-488 DIASFYWTLTNNDTHTQEPsAQGGEEVLQQLQALAPRGVVKVRIAVSKPNGF--LADLQSL
ENSP000000386938/1-490 DIASFYWTLTNNDTHTQEPsAQGGEEVLRLQLTLAPKGVNVRIAVSKPSPGFPQADLQAL
*****
ENSMUSP000000112942/1-488 LQSGAQVRMVDQMQLTHGVLHTKFWVVDQTHFYLGSAAMDWRSLTQVKELGVMYNCSCSL
ENSP000000386938/1-490 LQSGAQVRMVDQMQLTHGVLHTKFWVVDQTHFYLGSAAMDWRSLTQVKELGVMYNCSCSL
*****
ENSMUSP000000112942/1-488 ARDLTKIFEAYWFLGQAGSSIPSTWPRSFDTRYNQETPMEICLNGTPALAYLASAPFFLC
ENSP000000386938/1-490 ARDLTKIFEAYWFLGQAGSSIPSTWPRFYDTRYNQETPMEICLNGTPALAYLASAPFFLC
*****
  
```

图8 人鼠PLD3蛋白同源性对比结果

确定了小鼠的突变位点后，我们便可以针对该突变位点设计Guide RNA，利用Crispr/Cas9技术来构建点突变小鼠（图9）。

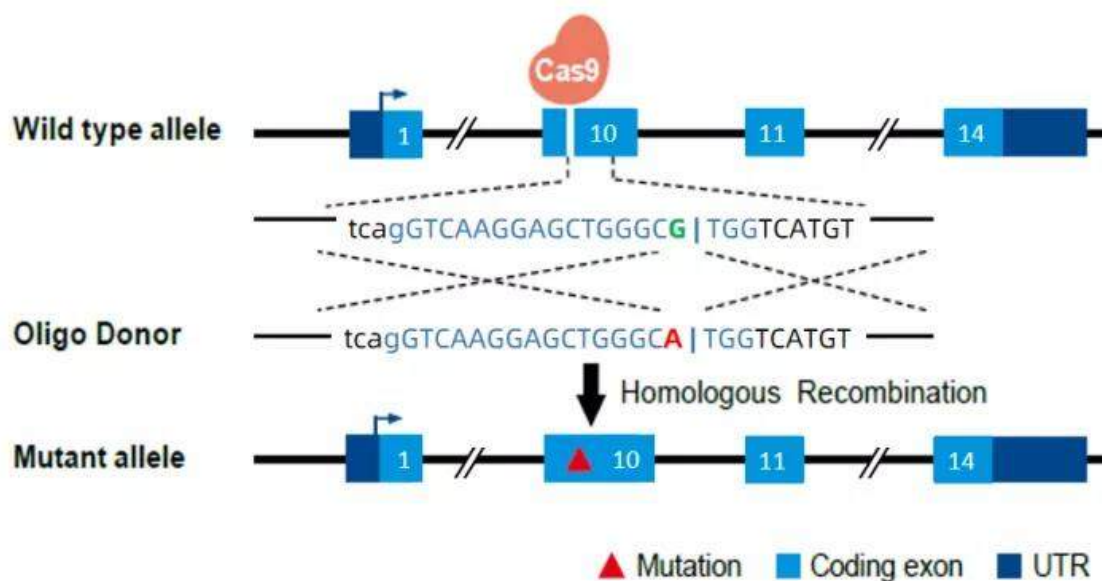


图9 PLD3-V230M突变小鼠构建策略

## 案例分享

接下来和大家分享一个案例来具体说明如何利用基因修饰小鼠验证SNP的临床表型。

【标题】自身免疫疾病相关IgG1变体调控B淋巴细胞活化和分化的新机制

【期刊】Science (IF=41)

Science

REPORTS

Cite as: X. Chen *et al.*, *Science*  
10.1126/science.aap9310 (2018).

## An autoimmune disease variant of IgG1 modulates B cell activation and differentiation

Xiangjun Chen<sup>1</sup>, Sun Xiao-Lin<sup>2\*</sup>, Wei Yang<sup>3\*</sup>, Bing Yang<sup>4\*</sup>, Xiaozhen Zhao<sup>2</sup>, Shuting Chen<sup>1</sup>, Lili He<sup>1</sup>, Hui Chen<sup>4</sup>, Changmei Yang<sup>1</sup>, Le Xiao<sup>1</sup>, Zai Chang<sup>3</sup>, Jianping Guo<sup>2</sup>, Jing He<sup>2</sup>, Fuping Zhang<sup>5</sup>, Fang Zheng<sup>6</sup>, Zhibin Hu<sup>7</sup>, Zhiyong Yang<sup>8</sup>, Jizhong Lou<sup>4</sup>, Wenjie Zheng<sup>9</sup>, Hai Qi<sup>10</sup>, Chenqi Xu<sup>11</sup>, Hong Zhang<sup>12</sup>, Hongying Shan<sup>12</sup>, Xujie Zhou<sup>12</sup>, Qingwen Wang<sup>13</sup>, Yi Shi<sup>14,15</sup>, Luhua Lai<sup>16</sup>, Zhanguo Li<sup>2†</sup>, Wanli Liu<sup>1,17†</sup>

系统性红斑狼疮SLE是最常见的自身免疫疾病之一，是一种由遗传、环境、免疫等多种因素导致的复杂性疾病。其发病机制至今尚不清楚，导致治疗手段有限且无法根治。人体中广泛存在具有自身反应性的IgG型记忆性B细胞，SLE等自身免疫疾病常伴随大量自身抗体的产生。

研究人员猜想IgG型B细胞受体（IgG-BCR）的异常活化可能参与到SLE的发病过程。通过对患者进行基因检

测，研究人员发现，IgG1重链基因IGHG1上的SNP（rs117518546）在SLE患者中显著增加，该SNP导致IgG1第396位甘氨酸突变为精氨酸（IgG1-G396R）（图9）。携带该SNP的SLE患者产生更多更广泛的IgG1型的自身抗体，发生炎症反应的风险增加，疾病活动指数也更高，揭示该SNP为新的SLE易感基因位点。

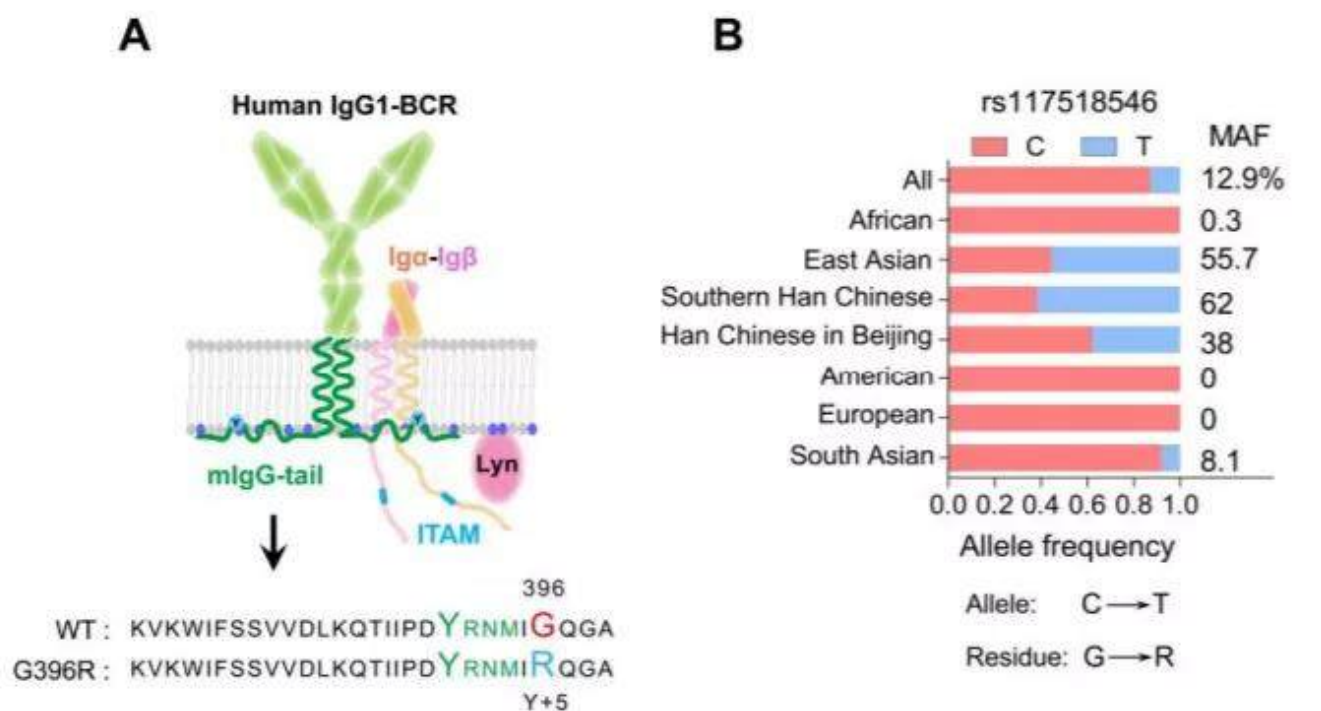


图10 rs117518546突变位点（图A）及在不同地区人群中的比例（图B）

值得注意的是，该SNP在东亚人群中的出现的频率远高于欧美人群（图10），这提示该SNP是以汉族为代表的东亚人群特有的SLE易感基因位点，可能是我国SLE患者与其他人群患者的临床表现和治疗反应存在差异的因素之一。

通过同源性对比，研究者发现该SNP在小鼠上的位点为390位甘氨酸突变为精氨酸，因此构建了IgG1-G390R突变小鼠（图11）。通过诱导野生型小鼠和突变小鼠发生自身免疫疾病，研究者发现携带该SNP的小鼠产生更多、更广泛的自身抗体，而这依赖于B细胞受抗原刺激后大量自身反应性浆细胞的分化（图12）。这一结果验证了在SLE患者中的发现，并从B细胞命运决定层面给出了初步解释[6]。



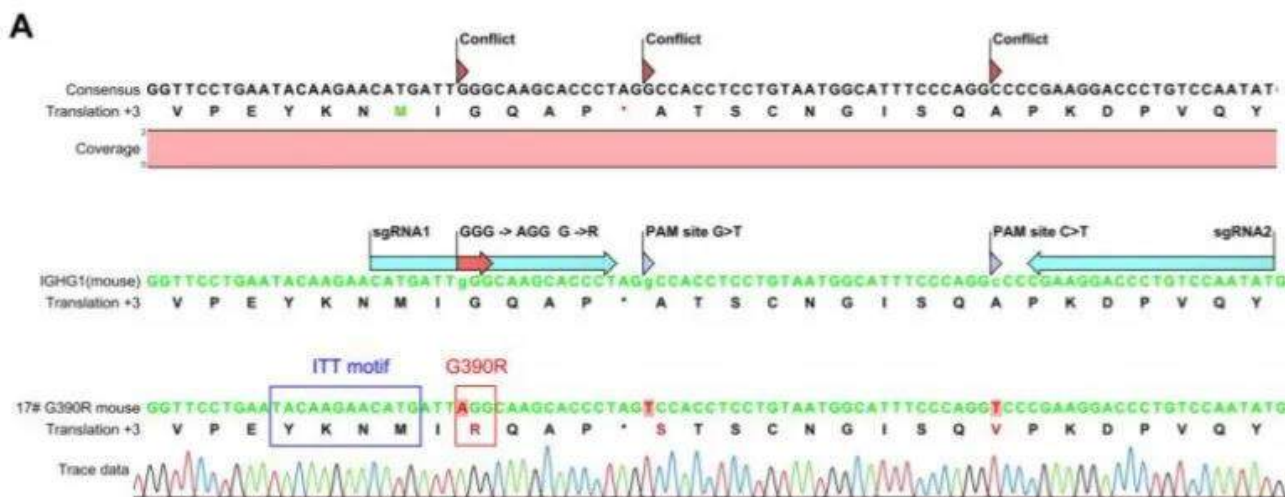


图11 构建IgG1-G390R敲入小鼠

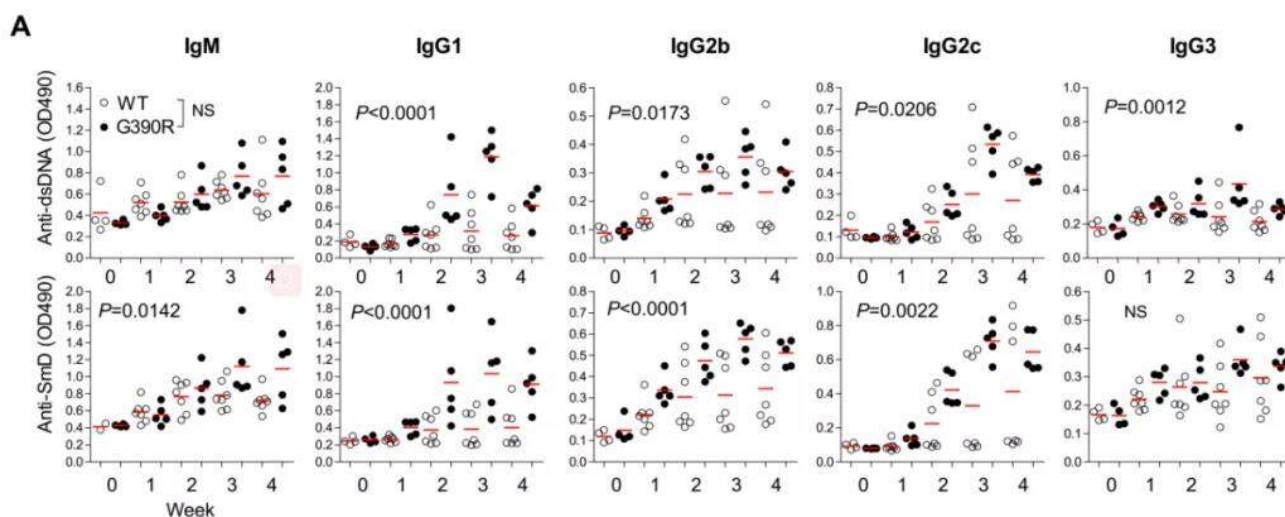


图12诱导自身免疫病发生后，野生型小鼠和G390R小鼠体内dsDNA和SmD抗体动态变化。

该研究从疾病易感基因位点的临床遗传学研究出发，经过基于基因修饰小鼠的动物模型研究，将临床上发现的问题在真实的体内实验中验证，确认了IgG1-G396R在SLE疾病发生发展中的作用。这项研究为SLE研究提供了全新的研究靶点，也为复杂疾病相关易感基因位点的研究提供了创新性的研究模式，最终发表在《Science》杂志上。

南模生物深耕基因修饰动物模型行业二十余年，在构建类人SNP同源突变小鼠方面具有丰富的经验，助力客户发表在《Cell Research》、《Immunity》等多种高水平杂志上，受到广泛认可。欢迎大家咨询，快来定制专属您的点突变小鼠吧。

## 参考文献：

[1]杨焕明. 人类基因组计划及其意义——规模化、序列化、信息化与产业化、医学化、人文化[J]. 科技与法律, 2000(3):2.



[2] Altshuler D, Pollara VJ, Cowles CR. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*. 2000 Sep 28;407(6803):513-6.

[3] <https://www.snpedia.com/index.php/APOE-%CE%B54>

[4] Shi Y, Yamada K, Liddel SA, et al. ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature*. 2017 Sep 28;549(7673):523-527. doi: 10.1038/nature24016

[5] <https://www.nih.gov/news-events/news-releases/new-comprehensive-view-mouse-genome-finds-many-similarities-striking-differences-human-genome>

[6] Chen X, Sun X, Yang W, et al. An autoimmune disease variant of IgG1 modulates B cell activation and differentiation. *Science*. 2018 Nov 9;362(6415):700-705. doi: 10.1126/science.aap9310.