



# Genetic and environmental factors of pathogeny: a population genetics view

ZHOU Weichen

MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433 China

**ABSTRACT:** The factors of genes and environment are common causes of diseases. With theories of population genetics, we analyzed the possible causes of diseases on the aspects of population resources, familial inheritance, environment, interactions between genes and environment, physiological, biochemical, and cellular level disorders, respectively. The analyses may provide theoretical basis for genetic consulting, disease study, and health evaluation.

**Key words:** Population genetics; Disease genes; Environmental factors; Interaction of genes and environment; Susceptibility

## 群体遗传学角度分析疾病病因中的 基因和环境因素

周玮晨

复旦大学现代人类学教育部重点实验室, 上海 200433

**摘要:** 导致疾病的原因主要是基因因素和环境因素。本文运用群体遗传学知识, 分别从人群群体角度、家族遗传角度、环境角度、基因与环境交互作用、生理生化及细胞等层面讨论疾病的发病原因, 作出相应的分析解释。根据此分析对遗传咨询、疾病研究、健康诊断等提供理论基础。

**关键字:** 群体遗传学; 致病基因; 环境因素; 基因与环境互作; 易感性

### 一、背景介绍

#### 1. 群体遗传学

群体遗传学是研究群体的遗传结构及其变化规律的遗传学分支学科, 应用数学和统计学方法研究群体中基因频率和基因型频率, 以及影响这些频率的选择效应和突变作用, 研究迁移和遗传漂变等过程与遗传结构的关系, 由此探讨进化的机制[1]。

群体指的是孟德尔群体, 即一群相互繁育的个体。一个最大的孟德尔群体是一个物种。一个群体中全部个体所共有的全部基因称为基因库。群体中各种基因的频率, 以及由不同的交配体制所带来的各种基因型在数量上的分布称为群体的遗传结构。获知了不同世代中遗传结构的演变方式就可探讨生物的进化过程并据以培育各种新的生物品系和品种。可是群体遗传学并不等同于进化遗传学。后者探讨的是物种内变异转化为物种间变异的过程, 即物种的形成和绝灭, 而前者则仅仅涉及品系间、品种间和亚种间等的变迁。

#### 2. 群体遗传学分析方法

群体遗传学的分析方法主要包括以孟德尔遗传定律为基础的应用数学, 生物统计学, 其最终目的, 在于阐明生物进化的机制。如果就群体遗传学作历史性地考查, 可以认为它是由生物统计学方法结合达尔文的自然选择与孟德尔遗传定律而形成的。作为群体遗传学基础的数字理论, 主要是在三十年代由 R.A.Fisher, J.B.S.Haldane 和 S.Wright 等的研究才开始建立的体系, 以后大量学者的工作又使之进一步发展和完善[2]。群体遗传学的核心是以群体中的突变基因的动态, 作为概率过程来进行理论处理的, 在实验方面曾以果蝇为材料进行了广泛的研究。在人类及植物群体中的工作亦盛行不衰。群体遗传学不仅对进化机理的发展有很大的影响, 而且其方法现已被采用作为动植物育种学的基础理论, 对育种的现代化作出了重要贡献[3]。另一个不重要的意义就是对人类遗传学的帮助, 在不可能实行自由交配实验的人类遗传学研究中, 辅之以不可缺少的统计学处理, 在阐明人类群体的遗传组成上正在

起着越来越大的作用[4]。此外，由于分子遗传学的发展，终于能够在信息分子(基因或作为其直接产物的蛋白质)水平上来处理变异及进化的问题，在其推动下，群体遗传学的理论也取得了巨大进展。

### 3. 基因与环境互作

基因型决定了生物的根本性状。但更多时候，由于生物生存在的一个大环境里，生物与环境的相互作用也成为了生物分析中必不可少的方面。基因影响性状，而环境也能从一定程度影响性状，两者的相互作用，导致了一个地区内的群体性状的综合表现，以及群体中疾病的发生。由于在疾病中存在的更多的是复杂遗传方式和多基因遗传，使得基因对疾病的影响变得复杂。在群体遗传学研究疾病的层面，环境的因素更起到了举足轻重的作用。而对于没有基因倾向的人群中，环境也可以诱导疾病的产生，因此，从群体遗传学的角度，运用生物统计学来解决疾病中基因与环境互作关系和两者所占比例程度显得尤为重要。

## 二、研究假设与目的

本文旨在运用群体遗传学从基因和环境等层面讨论疾病的发病原因。假设 C 地区的一种疾病发病率显著高于全国的平均发病率，其中有 3 家人，每家祖孙三代中都有一人患有这种疾病(共 9 人)。我们将运用不同的方法解析可能的病因。首先，我们将运用群体遗传学的方法分析和疾病相关的基因，通过 C 地区的正常和患病样本的比较，以及其他地区正常和患病样本的比较等；也可以通过家族疾病的遗传特征分析，对疾病的致病机理和可能相关的基因进行分析。其次，通过对 C 地区的环境分析(如水资源)等，对可能的状况进行评估；再次，通过对疾病的分析，考虑可能的其他人为因素(如吸烟等)，这也可以归为环境影响造成。最后，可以讨论在基因和环境相互作用下，影响和决定该疾病在 C 地区的发生状况，并为该地区的疾病治疗做好理论准备。

本论文在实际背景下，讨论了疾病的发

病原因。通过遗传学的方法，在基因和环境等多角度考虑疾病。其中包括，疾病的产生是否与基因相关，与什么基因相关，是否与环境的作用相关，相关程度多大，基因与环境是以什么状态相互影响而导致了此地疾病的发生，等。通过这些方面的研究，深入了解群体遗传学运用于疾病的研究的方法和手段，对于一种疾病在地区的发生，该疾病的鉴定(环境因素占多还是基因层面占多)，以及治疗等方面，都能够打下良好的理论基础，同时更能对可能致病基因的研究做出铺垫。

## 三、研究思路

从前文中提到的研究内容中，要求我们运用遗传学的方法首先确定疾病的病因。通过不同角度层面的分析，以及缜密的可能性研究，一步一步的推进假设情况，我们能够得到 C 地疾病可能的致病原因，以及不同原因对疾病产生的贡献，最终为解决疾病问题做好理论基础。

我们首先发现的是疾病在 C 地的发生频率要大于全国水平，因此，我们将首先在基因和环境等方面考虑可能的致病原因。我们能用的方法很多，首先也是最为可能的情况是基因引起的疾病，因此我们可以通过群体遗传学分析或是家族遗传分析；如若不然，我们需要鉴定疾病是否是由环境引起的，抑或是基因和环境两者的相互作用。

### 1. 基于群体遗传学分析疾病致病基因

运用群体遗传学的方法我们可以将疾病的致病原因进行分析。首先，若该疾病为基因导致，我们必须对可能的基因进行调研，即在一定范围内选择可能的致病基因(尽量考虑可能的情况)；其次，我们需要在不同的地方考虑此疾病的问题，对于不同的人群(患病和非患病)也需要列入考虑范围，进而进行采样等的分析(采样的方式有很多，如血样，病变组织等)；最后，对于一个有效的群体遗传学分析，一个范围内调查的人数(即样本量)必须大于一定范围(如大于 30 人)。如上所述，在有效群体个数基础上，我们可以得到以下步骤进行分析：I. 对 C 地区患病和非患病

人群的比较,即对 C 地区的两种不同人群进行采样。II.对全国其他地方的正常人群的比较,即对全国其他地区的正常人群进行采样。III.全国其他地方患病人群的比较,即对全国其他地区患病人群进行采样。IV.对采集的样本进行分析,即对采集的病变组织或同等位置正常组织,或者血液等,提取 DNA 序列,对候选的基因的可能表达情况(如上升表达,下降表达)进行分析(可采用基因芯片等方法)。V.运用生物统计学的方法,对得到的结果进行统计学意义上的分析,如运用线性回归或 Ttest 检验等。

我们可以通过统计,在第 IV, V 步的结果上,得到具有统计学意义的结果: a.发现所有调查的个体中所有基因的表达没有显著差异,则该病可能是由于环境因素造成,或某些我们未考虑到的基因或因素造成; b.发现所有患病个体的某个或某几个基因的表达与正常个体存在显著差异,则说明该疾病是由这些基因导致,若 C 地与其他地区患病个体中的基因表达也出现差异,则可能是 C 地的环境促进了这些基因的致病性。c.如发现 C 地区所有人群的某些基因表达与其他地区显著出现差异,则有可能是该地区的人群与其他地区存在基因层面的差异,可以考虑该地区人群在生活,迁徙和演化上的历史原因导致该疾病的致病。

由此,我们能够通过群体遗传学分析出疾病可能的相关致病原因,并且按照统计给出的不同结果确定不同的研究路线:若结果为 a,则我们将会进行下面 2,3,5 条的研究,来探究究竟是基因致病还是环境造成;若结果为 b,则我们会考虑下面 4 的研究,探讨环境和基因互作的情况;若为结果 c,则我们会从人类遗传学角度考虑 C 地区人群的背景以及他们由此而带来的基因差异,包括在下面的第 6 条中。在完成群体遗传统计后,该病的致病原因实质上已经可以大体上了解。

## 2.家族遗传分析疾病致病基因

我们可以通过家族的遗传分析,更加准确的得到该疾病可能的致病基因。根据研究的已知信息,在 3 个家系中,祖孙三代中都有一个患病,则我们可以根据这个信息做很

好的家族遗传分析。通过家族的信息,首先可以很初步的推断出可能的基因的显隐性以及遗传方式。其次,可以通过很多方法(如单体型检测,染色体杂交,全基因组测序等)来确定候选区,从而在尽量小的候选区中定位致病基因的位置。这种方法在遗传学中对于一个家族的遗传疾病特别有效,在遗传学手段不发达时,白化病血友病等的基因都是通过这样检测出来[5],在现在许多疾病如某些抑郁症,智力低下等疾病的基因也是通过这样的方法找出来的。

## 3.考虑可能的环境因素

若运用第 2 步仍不能发现致病的基因,我们则可以考虑是单纯的环境因素作用(实际上这种情况多会反映在通过环境因素作用影响到基因层面而显示出来),若此,家族中的患者就有可能单单是随机患病了。如,有可能是某种寄生虫病(如血吸虫,螨虫导致的疾病),或者在 C 地的环境中(如水,空气,潜在的放射源等),存在了导致疾病的因素。如, C 地的饮水中存在超标的金属离子,如汞离子,铬离子等,如上世纪在日本某些地方发生的水俣病[6],便是饮水中汞离子超标导致。又如,空气中若存在超标的有害成分,也往往会导致一个地区的呼吸道疾病的爆发。因此,在此类疾病的鉴别中,我们可以首先考虑到的是人群在生理上的异常表现,并将生理症状同以前的病例比较和分析;而环境造成的影响往往是后天的且不可遗传的,这些特性也可以用于区分遗传致病和环境致病。

## 4.分析环境和基因互作产生的致病因子

多数情况下,疾病的产生往往是由环境和基因互作导致的。我们的研究重点除了基因和环境是如何互作的外,将它们对疾病的影响比重也是需要研究的重点之一。一种情况是:很多疾病往往是在基因层面上具有病发的倾向性,而这往往是不够的,当某些地方的某些环境因素同样影响到生物个体时,在基因层面上的“倾向”就会被环境所放大,导致了该地区的疾病发病率的升高。或者是生物体中的基因并为显示出选择上的差异,

但是,在某种环境的影响下,对该基因产生了一个正向选择力,致使该环境下的生物群体中该基因的频率上升,使得该疾病的发病率上升。实际上,这类情况的疾病,可以看做是基因和环境两者相互影响相互作用的结果。如,对于一种G6PD(即葡萄糖-6-磷酸脱氢酶)缺陷症的分布和致病,就与环境的因素息息相关。G6PD酶活性缺乏症会直接影响到红细胞的抗氧化能力,可导致溶血性贫血、蚕豆病、新生儿高胆红素血症等病患,是人类最常见的酶缺陷病[7]。而该病主要见于非洲、地中海沿岸、中东、亚洲、太平洋诸岛、南美洲等热带,亚热带地区。与此同时,人们发现,这类地区的环境上存在一个共同的特点:气候潮湿炎热,并且是疟疾发生的主要地区。科学家们通过研究非洲东南亚等地的G6PD缺陷发生率和基因突变数据及各地疟疾流行情况、东亚和东南亚人群历史人类学、分子人类学资料,分析疟疾对东南亚人群G6PD缺陷的影响,提出了疟疾对G6PD缺陷的正选择作用的假说[8]。

### 5.从生理,生化,细胞层次提出可能的病因假说

这部分的研究主要是补充 2, 3 的研究,疾病的致病原因可以多种多样,不同的致病原因,可以导致从个体到细胞直至到分子各个层面上不同的病理反应,因此,我们不能只拘泥于一种方法或是一种表现,而是应该全面的对疾病进行分析。从生理上分析,如水俣病是在上世纪日本爆发的一种疾病,并且可以遗传给子女,病人在生理上显示“运动失调、四肢麻木、疼痛、畸胎”等病症[6]。再通过对人体各指标的研究就可以发现,这些都是因为汞离子在人体内大量积累造成的,而所谓的遗传也是汞通过胎盘毒害到胎儿导致的。从器官,细胞层面,我们实质上可以发现更多,比如肾的多囊化,脑部组织海绵状(疯牛病),染色体的缺失或是大段重复(21三体综合症),细胞内糖原的积累等,这些疾病都可以从这个角度去研究然后深入得到具体的致病机理的。同样,我们也能够从分子层面直接了解致病原因,如通过体内蛋白表达或转录物的功能推测,从而上溯到 DNA 得

到确定的致病基因, Treacher Collins 综合征就是一个通过纯粹转录物作图从而分离得到致病基因的病症。

### 6.从人群背景提出可能的人群差异导致的致病性

不同的人群对于不同的疾病也存在易感性的差别,并且,不同人群中的基因种类和拷贝数的不同,也会导致疾病在不同人群中的发病率不同。因此,从人群的背景上考虑疾病发病率不同的差异,是研究疾病的有一个必要的方向。对于不同的情况,可以有如下的推断: I.C 地区人群祖先是来自外地(或其他民族)迁徙而来的,导致的疾病的易感性不同。如古代的新疆地区就是一个多民族多肤色混杂的交界地带,不同民族对于不同的疾病自然会有不同的易感性[9]。II.C 地区人群不同于别地是由于历史上存在的奠基者效应。如果 C 地区的祖先是几个很少的个体,而这几个个体若存在与其他个体在基因上的不同(如某些遗传疾病等),会导致其由于奠基者效应产生的遗传和基因上与别地区的巨大差别[10]。III.C 地区人群民族文化与其他地区一致,但不能排除其祖先的来源不同的地方。最明显的例子就是 Y 染色体的不同 SNP 单倍型,如对于 N 型 Y 染色体人群,已经有科学家发现其独一无二的与其他单倍型在指定的 CNV 存在差别,而导致了可能的男性不育的倾向[11]。因此,将人群的背景考虑进去,不仅仅是为了更加佐证疾病发病的可能原因,也是在分析疾病发病原因中不可缺少的一环。

## 四、讨论与展望

本文旨在通过群体遗传学的基本方法解决疾病致病原因调查的问题,之后的基因和环境互作的讨论也是基于群体遗传学分析统计方法后做出的相应讨论,因此,得到的结果,希望能够在群体中具有显著效果,体现出整体效应。

对于疾病的可能发病病因的研究,从生物学诞生以来一直是人们关注的重点,而遗传学的产生使得这一研究更加深入。随着遗传学的发展,特别是遗传学同不同领域的融

合，科学家们在对生物的疾病研究有了越来越多，越来越有效的方法，基于生物统计的群体遗传学就是其中之一。在之后的岁月里，随着测序技术进一步的发展，全基因组测序变得简单而有效，对于人们个体之间的不同的个性化诊断和治疗也随之发展起来，以后的人们可以更加清楚的掌握自己的健康状况，以及更加及时有效的应对不同疾病的困扰。

#### 参考文献

1. Daniel H (2000) *Primer of Population Genetics*. Sunderland: Sinauer.
2. Benjamin L (2004) *Gene VIII*. New York: Prentice Hall.
3. 盛祖嘉, 沈仁权 (1988) *分子遗传学*. 上海: 复旦大学出版社.
4. Strachan T, Read AP, 孙开来译 (2007) *人类分子遗传学*. 北京: 科学出版社.
5. Price M (2009) Case Closed: Famous Royals Suffered From Hemophilia. *Science now* 8 October 2009.
6. 原田正纯 (1995) *水俣病と世界の水銀汚染*. 东京: 实教出版.
7. Persico MG, Viglietto G, Martini G, Toniolo D, Paonessa G, Moscatelli C, Dono R, Vulliamy T, Luzzatto L, D'Urso M (1986) Isolation of human glucose-6-phosphate dehydrogenase (G6PD) cDNA clones: primary structure of the protein and unusual 5' non-coding region. *Nucleic Acids Res* 14: 11-22.
8. 潘尚领 (2007) 东亚及东南亚地区 G6PD 缺陷的地域分布以及疟疾的正选择作用. *现代人类学通讯* 1:42-52.
9. Han XS, Han XM, Luo QL (2008) Cancer Incidence Patterns by Race and Sex in the United States, 1992-2005. *Commun Contemp Anthropol* 2:62-71.
10. Ober C, Elias S, Kostyu DD, Hauck WW (1992) Decreased fecundability in Hutterite couples sharing HLA-DR. *Am J Hum Genet* 50:6-14.
11. Zhang F, Lu C, Li Z, Xie P, Xia Y, Zhu X, Wu B, Cai X, Wang X, Qian J, Wang X, Jin L (2007) Partial deletions are associated with an increased risk of complete deletion in AZFc: a new insight into the role of partial AZFc deletions in male infertility. *J Med Genet* 44:437-444.