Epigenetic Inheritance and the Missing Heritability Problem

Montgomery Slatkin[1]

Department of Integrative Biology, University of California, Berkeley, California 94720-3140

Manuscript received March 13, 2009
Accepted for publication April 20, 2009

Genetics **182:** 845–850 (July 2009)

Modern definition of epigenetics

- The study of heritable changes in gene expression that are not caused by changes in DNA sequence
  - methylation of the cytosine residue in DNA
  - modification of chromatin proteins that package DNA
- In this article: only epigenetic changes that are transmitted to offspring (“*transgenerational epigenetic inheritance*”)

“Missing heritability”

- Inherited causes of risk of complex genetic diseases that have not yet been identified in genomewide association studies (GWAS)

The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

Review

Nature **461**, 747-753 (8 October 2009) | doi:10.1038/nature08494;
Received 25 June 2009; Accepted 11 September 2009

Finding the missing heritability of complex diseases

Teri A. Manolio[1], Francis S. Collins[2], Nancy J. Cox[3], David B. Goldstein[4], Lucia A. Hindorff[5], David J. Hunter[6], Mark I. McCarthy[7], Erin M. Ramos[8], Lon R. Cardon[9], Aravinda Chakravarti[10], Judy H. Cho[11], Alan E. Guttmacher[1], Augustine Kong[12], Leonid Kruglyak[13], Elaine Mardis[14], Charles N. Rotimi[15], Montgomery Slatkin[16], David Valle[9], Alice S. Whittemore[16], Michael Boehnke[17], Andrew G. Clark[18], Evan E. Eichler[19], Greg Gibson[20], Jonathan L. Haines[21], Trudy F. C. Mackay[22], Steven A. McCarroll[23] & Peter M. Visscher[24]

ARTICLE LINKS

- Figures and tables

SEE ALSO

- Editor's Summary

ARTICLE TOOLS

- Send to a friend
- Export citation
- Export references
- Rights and permissions
- Order commercial reprints
- Bookmark in

Goals

- Present a simple model of the inheritance of epigenetic changes.
- Quantify the potential contribution they can make to *average risk* and *recurrence risk*(*).

(*) *recurrence risk* :The likelihood that a trait or disorder present in one family member will occur again in other family members in the same or subsequent generations

Assumptions

- Epigenetic effects are caused by the presence or the absence of epigenetic modifications of specific chromosomal locations.
- Transmission to offspring is the same as the transmission of mutations, except for the possibility that they might be spontaneously lost.
- Appearance of an epigenetic change at a location in the genome is **not** attributable to any particular locus or loci,
- The phenotypic effects of the presence or the absence of an epigenetic change are attributable to the genomic location itself.

Model

- Disease risk is affected by n diallelic genetic loci and by v sites at which epigenetic changes may be present.
- Assume multiplicative interactions across loci and epigenetic sites:
  the average risk and recurrence risks are computed by calculating contributions from each genetic locus and epigenetic site separately and then multiplying.

Multiplicative model

$$X = b \prod_{i=1}^{n} x_i \prod_{j=1}^{v} \xi_j, \qquad (1)$$

where X is the disease risk, b is the background risk, x_i is the contribution to the risk of locus i, and ξ_i is the contribution to the risk of epigenetic site j. The average risk, K, is the average of X taken over all genotypes and epigenetic configurations. This model does not allow for interactions

The contributions of each locus and epigenetic site to risk are assumed to be independent, implying that K is the product of the average contributions,

$$K = E(X) = b \prod_{i=1}^{n} E(x_i) \prod_{j=1}^{v} E(\xi_j). \qquad (2)$$

The recurrence risk ratio for a relative of relationship R is

$$\lambda_R = \frac{1}{K} \frac{E(XX')}{E(X)} = \frac{b^2}{K^2} \prod_{i=1}^{n} E(x_i x_i') \prod_{j=1}^{v} E(\xi_j \xi_j'), \qquad (3)$$

where the prime indicates the risk in a relative with relationship R.

The average contribution of locus i to disease risk is

$$k_i = E(x_i) = \sum_{g=0}^{2} \Pr(g)(1 + r_i)^g = (1 + p_i r_i)^2, \qquad (4)$$

where g is the number of + alleles (0, 1, and 2) and the second equality follows when genotypes are in Hardy–Weinberg equilibrium.

The contribution of locus i to recurrence risk in relatives with relationship R is

$$E(x_i x_i') = \sum_{g,g'=0}^{2} \Pr(g, g')(1 + r)^{g+g'}, \qquad (5)$$

where the joint probability of g and g' depends on relatedness.

Considering the state of an epigenetic site as a two-state Markov chain, the transition matrix in a single generation is

$$\begin{aligned} \Pr(1\to 1) &= 1-\alpha_j, \quad \Pr(1\to 0) = \alpha_j, \\ \Pr(0\to 1) &= \beta_j, \quad \Pr(0\to 0) = 1-\beta_j. \end{aligned} \tag{6}$$

The standard theory of Markov chains tells us that the equilibrium frequency of a mark at site j is $\hat{\pi}_j = \beta_j/(\alpha_j+\beta_j)$ and that the transition probabilities after m generations are

$$\begin{aligned} \Pr(1, t=m \mid 1, t=0) &= (1-\alpha_j-\beta_j)^m + \hat{\pi}_j\left[1-(1-\alpha_j-\beta_j)^m\right] \\ \Pr(0, t=m \mid 1, t=0) &= (1-\hat{\pi}_j)\left[1-(1-\alpha_j-\beta_j)^m\right] \\ \Pr(1, t=m \mid 0, t=0) &= \hat{\pi}_j\left[1-(1-\alpha_j-\beta_j)^m\right] \\ \Pr(0, t=m \mid 0, t=0) &= (1-\alpha_j-\beta_j)^m + (1-\hat{\pi}_j) \\ &\quad \times\left[1-(1-\alpha_j-\beta_j)^m\right]. \end{aligned} \tag{7}$$

The model assumes that presence of an epigenetic mark at site j increases disease risk by a factor $1 + \rho_j$. If the epigenetic sites are at equilibrium under the current rates of gain and loss of marks, the expected contribution of site j to average risk is

$$\kappa_j = E(\xi_j) = \sum_{\gamma=0}^{2} \Pr(\gamma)(1+\rho_j)^{\gamma} = (1+\pi_j\rho_j)^2, \qquad (8)$$

where $\gamma = 0, 1, 2$ is the number of marks at site j (*cf.* Equation 4). Similarly

$$E(\xi_j\xi_j') = \sum_{\gamma,\gamma'=0}^{2} \Pr(\gamma,\gamma')(1+\rho_j)^{\gamma_j+\gamma_j'}. \qquad (9)$$

Can epigenetic sites account for missing causality and heritability? In discussions of missing heritability, there is a tendency to assume genetic and other factors that contribute most to individual risk also contribute most to recurrence risk. But in reality, factors that increase recurrence risk substantially do not necessarily have much effect on average risk and vice versa. The solution to the problem of missing heritability is not necessarily the same as the solution to the problem of missing causality, as has been pointed out by HEMMINKI *et al.* (2008).

These numbers provide a convenient reference point to ask what would have to be assumed about epigenetic sites to account for the same contributions to average risk and recurrence risk. The contributions of each epigenetic site can be calculated from the formulas above and in the APPENDIX. The contribution to the average risk depends on the equilibrium frequency of epigenetic marks, $\hat{\pi}$, and the effect of each mark on risk, ρ. If there were 30 epigenetic sites with $\hat{\pi} = 0.01$ at each and if $\rho = 2$ for each mark, together they would increase average risk by the same factor as above, 3.28. The contribution to λ_S depends on the turnover rate of marks, $\alpha + \beta$. With $\hat{\pi} = 0.01$, $\alpha = 99\beta$. If $\alpha = 0.495$ and $\beta = 0.005$, $\lambda_S = 1.32$ for these 30 sites together, not enough to account for much of the concordance of full siblings. If, instead, $\alpha = 0.0495$ and $\beta = 0.0005$, $\lambda_S = 2.75$. Thus, only if the per generation rate of loss, α, is small can epigenetic marks account for a substantial part of the inherited risk. If $\alpha = 0.0495$, an epigenetic mark would have to persist for a average of slightly more than 20 generations.

If marks are more common at each site, they can contribute substantially more to average risk. If $\pi = 0.2$ and $\rho = 0.25$, then the contribution to average risk is 18.7 for 30 such sites. However, such sites contribute little to recurrence risk. For example, if $\alpha = 0.2$ and $\beta = 0.05$, then together they increase λ_S by only 1.16.

If epigenetic marks do persist for very long times, they are equivalent to mutations and hence have the same opportunity to be in significant linkage disequilibrium with linked marker SNPs as do other mutations. In that case, they would be detected in GWAS to the same extent as other mutations.

Conclusions

- If an epigenetic change and a mutation have the same effect on disease risk and are found in the same population frequency, they will contribute **equally** to average risk but the mutation will contribute **more** to recurrence risk.
- The reason is that the higher rate of loss of epigenetic modifications means that identity by descent does not imply identity in state.
- It will be difficult for inherited epigenetic changes to account for the missing heritability of complex diseases unless they are more common than mutations or have more pronounced effects on disease risk.

Conclusions

- Inherited epigenetic changes must persist for tens of generations or more for them to contribute significantly to similarities of close relatives.
- Until estimates of persistence times of inherited epigenetic changes are available, it will be difficult to draw firm conclusions about their potential role.