



Where is the friend's home?

Guo-Bo Chen*

Centre of Neurogenetics and Statistical Genomics, Queensland Brain Institute, The University of Queensland, St Lucia, QLD, Australia
*Correspondence: chenguobo@gmail.com

Edited by:

Marshall Abrams, University of Alabama at Birmingham, USA

Reviewed by:

Sergio Tofanelli, Università di Pisa, Italy

Keywords: friendship, nature selection, Framingham heart study, FST statistics, GWAS

A commentary on

Friendship and natural selection

by Christakis N., and Fowler, J. (2014). *Proc. Natl. Acad. Sci. U.S.A.* 111, 10796–10801. doi: 10.1073/pnas.1400825111

Using Framingham Heart Study (FHS) data (accession no. phs000153.v7.p6), Christakis and Fowler (2014) showed that friends had more extended genetic correlation than strangers had. This finding may suggest that friends are “functional kin.” Depending on the context, functional kin may have high genetic similarity (homophilic) if people have similar functional kin as friends; conversely, functional kin may have low/negative genetic similarity (heterophilic) if people have complementary functional kin as friends. Christakis and Fowler reported that between friends there is positive/negative genetic correlation at the single nucleotide polymorphism (SNP) level, indicating heritability of friendship. This result is very interesting, particularly for the social sciences (Skyrms et al., 2014), but deserves scrutiny. This note demonstrates that (1) high genetic similarity between friends is possible if a person finds friends from his/her own cultural background, which is a surrogate for genetic similarity, or (2) low genetic similarity is possible if a person finds friends from a different cultural background. As illustrated below, the mechanism involved in inflating/deflating genetic similarity between friends is analogous to the way population stratification raises the false positive rate in case-control studies.

In Christakis and Fowler’s study, a single-marker regression was conducted

to test the association for friendship. The regression is defined as

$$g_{e.m} = \mu_m + b_m g_{f.m} + e_m$$

in which $g_{e.m}$ and $g_{f.m}$ are the genotypes of “ego” and his/her friend, respectively, at the m^{th} locus ($m = 1, 2, 3 \dots M$). The regression coefficient is estimated as $\hat{b}_m = \frac{cov(g_{e.m}, g_{f.m})}{var(g_{f.m})}$. In a conventional GWAS regression, the phenotype is the same for each locus, whereas in this regression, the $g_{e.m}$ phenotype is updated to match each $g_{f.m}$ locus. However, this actually models F_{st} under the circumstances as discussed below (blue box in **Figure 1**). For clarity, the subscript m is hereafter omitted.

Assume that in the sample there are S subgroups (e.g., $S = 2$), each of which has n_s individuals. The proportion of the s^{th} subgroup to the total sample size is $w_s = \frac{n_s}{n}$, in which $n = \sum_{s=1}^S n_s$. Of the total n ego-friend pairs, n_s are from the s^{th} subgroup. The variance of g_f can be written as the within-group variance and the between-group variance (http://en.wikipedia.org/wiki/Mixture_distribution)

$$var(g_f) = \sum_{s=1}^S w_s [\sigma_{f(s)}^2 + (\bar{p}_{f(s)} - \bar{p}_f)^2]$$

in which $\bar{p}_{f(s)}$ is the reference allele frequency (RAF) of the friends from the s^{th} subgroup. $\bar{p}_f = \sum w_s \bar{p}_{f(s)}$ is the RAF of the friends, $\sigma_{f(s)}^2 = 2\bar{p}_{f(s)}(1 - \bar{p}_{f(s)})$ is the within-sub-population variance of the friends, and $(\bar{p}_{f(s)} - \bar{p}_f)^2$ is the between-sub-population sampling variance. Similarly, the covariance between g_e and g_f can be written as

$$cov(g_e, g_f) = \sum_{s=1}^S w_s [\sigma_{ef(s)}^2 + (\bar{p}_{e(s)} - \bar{p}_e)(\bar{p}_{f(s)} - \bar{p}_f)]$$

in which $\sigma_{ef(s)}^2$ is the covariance between the friends’ genotypes within the s^{th} sub-population, and $(\bar{p}_{e(s)} - \bar{p}_e)(\bar{p}_{f(s)} - \bar{p}_f)$ is the between-sub-population covariance. $\sigma_{ef(s)}^2 \neq 0$ if the pair of friends is related or there is heritability for friendship at the i^{th} locus.

In Christakis and Fowler’s study, the null hypothesis is that $H_0 : \sigma_{ef(s)}^2 = 0$ (i.e., no heritability for friendship). Even if friendship is not heritable and relatives are not included, friends can still have inflated/deflated genetic correlation, which raises the regression coefficient from zero, if an ego finds friends from the same/different cultural background.

If an ego finds friends from the same cultural background, $\bar{p}_{e(s)}$ will be similar to $\bar{p}_{f(s)}$. Assuming that $\bar{p}_{e(s)} = \bar{p}_{f(s)}$, the regression coefficient can be written as

$$\begin{aligned} \hat{b} &= \frac{cov(g_e, g_f)}{var(g_f)} \\ &= \frac{\sum_{s=1}^S w_s (\bar{p}_{e(s)} - \bar{p}_e)(\bar{p}_{f(s)} - \bar{p}_f)}{\sum_{s=1}^S w_s [\sigma_{f(s)}^2 + (\bar{p}_{f(s)} - \bar{p}_f)^2]} \\ &\approx \frac{\sum_{s=1}^S w_s (\bar{p}_{f(s)} - \bar{p}_f)^2}{\sum_{s=1}^S w_s [\sigma_{f(s)}^2 + (\bar{p}_{f(s)} - \bar{p}_f)^2]} \end{aligned}$$

By definition, $F_{st} = \frac{\sum_{s=1}^S (\bar{p}_{f(s)} - \bar{p}_f)^2 var(g_f)}{\sigma_{g_e}^2} \geq 0$. The standard deviation of \hat{b} is $\hat{\sigma}_b = \sqrt{\frac{\sigma_{g_e}^2}{n \sigma_{g_f}^2}} \approx \sqrt{\frac{1}{n}}$.

How an ego and a friend from the same subgroup (cultural background, in same color) can inflate genetic similarity, even when there is no heritability for friendship.

Ego Friend
 $\begin{matrix} \text{ego}_1 & \text{friend}_1 \\ \text{ego}_2 & \text{friend}_2 \\ \vdots & \vdots \\ \text{ego}_{n-1} & \text{friend}_{n-1} \\ \text{ego}_n & \text{friend}_n \end{matrix}$, 50% ego – friend pairs from subgroup 1, 50% from subgroup 2

Reference allele frequencies (RAFTs) of subgroup 1, subgroup 2, and all samples.

	Subgroup 1 (s = 1)	Subgroup 2 (s = 2)	All samples
Ego RAF	$p_{e(1)} = 0.6$	$p_{e(2)} = 0.4$	$\bar{p}_e = 0.6 \times 0.5 + 0.4 \times 0.5 = 0.5$
Friend RAF	$p_{f(1)} = 0.6$	$p_{f(2)} = 0.4$	$\bar{p}_f = 0.6 \times 0.5 + 0.4 \times 0.5 = 0.5$

Inflated regression coefficient under the null hypothesis of $\sigma_{ef(i)}^2 = 0$ (indicating neither the inclusion of relatives nor heritability of friendship).

$$\hat{b} = \frac{\text{cov}(g_e, g_f)}{\text{var}(g_f)} = \frac{0.5[\sigma_{ef(1)}^2 + (0.6 - 0.5)^2] + 0.5[\sigma_{ef(2)}^2 + (0.4 - 0.5)^2]}{2 \times 0.5 \times 0.5} = F_{st} = 0.02$$

Inflation of genomic control factors (λ_{GC}) $\chi_1^2 = \left(\frac{\hat{b}}{\hat{\sigma}_b}\right)^2 = nF_{st}^2$

When $n = 100$, noncentrality parameter (NCP) is $100 \times (0.02)^2 = 0.04$, $\lambda_{GC} = \frac{0.473}{0.455} = 1.04$

When $n = 1000$, NCP is $1000 \times (0.02)^2 = 0.4$, $\lambda_{GC} = \frac{0.669}{0.455} = 1.47$

How an ego and a friend from different subgroups (cultural backgrounds) can deflate genetic similarity, even when there is no heritability for friendship.

Ego Friend
 $\begin{matrix} \text{ego}_1 & \text{friend}_1 \\ \text{ego}_2 & \text{friend}_2 \\ \vdots & \vdots \\ \text{ego}_{n-1} & \text{friend}_{n-1} \\ \text{ego}_n & \text{friend}_n \end{matrix}$, 50% ego – friend pairs from subgroup 1 vs subgroup 2, 50% from subgroup 2 vs subgroup 1

Reference allele frequencies (RAFTs) of subgroup 1, subgroup 2, and all samples.

	Subgroup 1 (s = 1)	Subgroup 2 (s = 2)	All samples
Ego RAF	$p_{e(1)} = 0.6$	$p_{e(2)} = 0.4$	$\bar{p}_e = 0.6 \times 0.5 + 0.4 \times 0.5 = 0.5$
Friend RAF	$p_{f(1)} = 0.4$	$p_{f(2)} = 0.6$	$\bar{p}_f = 0.4 \times 0.5 + 0.6 \times 0.5 = 0.5$

Inflated regression coefficient under the null hypothesis of $\sigma_{ef(i)}^2 = 0$ (indicating neither the inclusion of relatives nor heritability of friendship).

$$\hat{b} = \frac{\text{cov}(g_e, g_f)}{\text{var}(g_f)} = \frac{0.5[\sigma_{ef(1)}^2 + (0.6 - 0.5)(0.4 - 0.5)] + 0.5[\sigma_{ef(2)}^2 + (0.4 - 0.5)(0.6 - 0.5)]}{2 \times 0.5 \times 0.5} = F_{st} = -0.02$$

in which F_{st} measures reduced genetic similarity.

Inflation of genomic control factors (λ_{GC}) $\chi_1^2 = \left(\frac{\hat{b}}{\hat{\sigma}_b}\right)^2 = nF_{st}^2$

When $n = 100$, noncentrality parameter (NCP) is $100 \times (-0.02)^2 = 0.04$, $\lambda_{GC} = \frac{0.473}{0.455} = 1.04$

When $n = 1000$, NCP is $1000 \times (-0.02)^2 = 0.4$, $\lambda_{GC} = \frac{0.669}{0.455} = 1.47$

FIGURE 1 | Illustration of how an ego-friend from the same/different subgroup(s) may inflate/deflate genetic similarity as indicated by λ_{GC} .

Given \hat{b} and $\hat{\sigma}_b$, a z-score test can be constructed as $z = \frac{\hat{b}}{\hat{\sigma}_b} = \sqrt{n}F_{st}$ and $z^2 \sim \chi_1^2$, with the non-centrality parameter (NCP) $\Delta = nF_{st}^2$. Christakis and Fowler reported λ_{GC} , the genomic

inflation factor (Devlin and Roeder, 1999), for their GWAS analysis. In their study, $\lambda_{GC} = \frac{\text{median}(\chi_1^2)}{0.455} = 1.04$, which meant that the median of the observed χ_1^2 value was 0.473, with NCP

$\Delta = nF_{st}^2 = 0.04$. Thus, the median of F_{st} was $\hat{F}_{st} = \sqrt{0.04/n} = 0.0066$ between the egos and the friends ($n = 907$ in Christakis and Fowler's study).

An early study (Cavalli-Sforza et al., 1996) indicated that $F_{st} = 0.016$ for European descendants, and a recent study (Novembre et al., 2008) using high-density SNPs showed that $F_{st} = 0.004$ between European nations (the estimated mean of $F_{st} = 0.0042$ over nearly 900 thousand common HapMap SNPs between CEU and TSI). Although the estimate of F_{st} depends on other factors, such as ascertainment and the sample proportion for each subgroup (Bhatia et al., 2013), the estimated \tilde{F}_{st} between the egos and the friends seems to fall in-between the aforementioned values. For a loosely defined trait such as friendship, the heritability may be low, if not zero. However, as long as friendship is more frequently established within one's cultural background, F_{st} will inflate genetic similarity among friendships even in the absence of heritability/“functional kin.” With an increased sample size, λ_{GC} will be even larger (blue box in **Figure 1**), a phenomenon that resembles the manner in which population stratification inflates type I error rate in GWAS.

In the discussed study, the negative genetic correlation between friends was also highlighted. Similarly, if an ego finds friends from a different cultural background, the regression coefficient will tend to turn negative (denoted as \tilde{F}_{st} , indicating reduced genetic similarity between friends, see the yellow box in **Figure 1**).

In practice, both patterns, not as extreme as demonstrated though, will be possible between friends and lead to homophily and heterophily as observed.

In their GWAS-like analysis, principle components were used as covariates to control for F_{st} . However, covariates may not completely eliminate the background effects, such as F_{st} . When covariates reduce F_{st} , the heritability of friendship will also be reduced, potentially to zero. Although the analysis in this note does not entail the rejection of the conclusion drawn by Christakis and Fowler, it warns against misleading interpretations of the results.

ACKNOWLEDGMENTS

This note was benefited greatly from the discussion on the journal club held at Centre of Neurogenetics and Statistical Genomics, Queensland Brain Institute, The University of Queensland (http://cnsngenomics.com/journal_club/). None of discussants are responsible for errors, nor should it be assumed that they agree with the analysis and conclusions presented in this note. The title of the note is inspired by the movie “Where Is Friend's Home,” directed by Abbas Kiarostami (http://en.wikipedia.org/wiki/Where_Is_the_Friend%27s_Home%3F).

REFERENCES

Bhatia, G., Patterson, N., Sankaraman, S., and Price, A. L. (2013). Estimating and interpreting

FST: the impact of rare variants. *Genome Res.* 23, 1514–1521. doi: 10.1101/gr.154831.113

Cavalli-Sforza, L. L., Menozzi, P., and Piazza, A. (1996). *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.

Christakis, N. A., and Fowler, J. H. (2014). Friendship and natural selection. *Proc. Natl. Acad. Sci. U.S.A.* 111, 10796–10801. doi: 10.1073/pnas.1400825111

Devlin, B., and Roeder, K. (1999). Genomic control for association studies. *Biometrics* 55, 997–1004.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., et al. (2008). Genes mirror geography within Europe. *Nature* 456, 98–101. doi: 10.1038/nature07331

Skyrms, B., Avise, J. C., and Ayala, F. J. (2014). In the light of evolution VIII: Darwinian thinking in the social sciences. *Proc. Natl. Acad. Sci. U.S.A.* 111, 10781–10784. doi: 10.1073/pnas.1411483111

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 August 2014; accepted: 29 October 2014; published online: 13 November 2014.

Citation: Chen G-B (2014) Where is the friend's home? *Front. Genet.* 5:400. doi: 10.3389/fgene.2014.00400

This article was submitted to *Evolutionary and Population Genetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.