

Letter to the editor: Standardization of genetic association studies, pros and cons, reaffirmed

Alexander Kulminski · Irina Culminkaya · Anatoli I Yashin

Received: 14 September 2013 / Accepted: 7 November 2013 / Published online: 26 November 2013
© American Aging Association 2013

Dear Dr. Ingram,

We would like to clarify principal limitations of, and erroneous statements in, the comments expressed by Dr. Paterson in his recent letter to the editor published in *Age*.

Paterson A. D. (2013) Letter to the editor: expression of concern, reaffirmed. *Age*.

This letter outlines what the author views as “a series of limitations” in our analyses. Dr. Paterson’s main concern is that our analyses do not follow standard Genome-Wide Association Studies (GWAS) strategies. The author argues that by following standard GWAS majority of SNPs reported in our papers might be excluded from the analyses right away. As a consequence, there would be no matter for “any downstream analysis”.

We agree with Dr. Paterson that following standard GWAS strategies it might be possible to come to such a conclusion. It does not mean, however, that this conclusion is correct because standard GWAS strategies themselves may not be always valid.

Specifically, a principal problem seen by Dr. Paterson is “that there were highly significant differences in genotype frequency” for SNPs from genotyping phases 3 and 4. Then, following the logic of standard GWAS strategies Dr. Paterson concluded that it was “consistent with major

genotyping error.” Table 1 in his letter is intended to illustrate this concern (this table should list rs1390694 but not rs139069).

Unfortunately, standard GWAS strategies are not designed to appropriately handle specific of age-related traits. They are not adapted for longitudinal studies either. Specifically, in longitudinal studies individuals are followed for a certain period of time which, in our case of the Framingham Heart Study (FHS), is very extensive, up to about 60 years. People can die during follow up. Even this obvious bio-demographic aspect of the problem is not within standard GWAS framework. It is, however, evident that if there are genes which influence survival, this selection may result in disproportion in frequencies of the risk allele(s) among individuals with shorter and longer lifespan. If survival is influenced by more than one allele, this selection may cause age-related clustering. Age-related selection is widely documented in gerontology in our studies (De Benedictis et al. 1998) and elsewhere (Atzmon et al. 2006; Christensen et al. 2006). This property is a central tenet of longevity studies (Martin et al. 2007). Clearly, such selection may result in deviation from Hardy-Weinberg equilibrium (HWE) that has nothing to do with genotyping errors. According to standard GWAS strategy pursued by Dr. Paterson, SNPs with deviation from HWE and any type of clustering associated with it are just eliminated. Thus, if we are looking for genes involved in regulation of survival-related traits in an aging study, following standard GWAS strategy will a priori eliminate such effects.

Then, the question is whether survival selection can indeed play a role and can explain the observed

A. Kulminski (✉) · I. Culminkaya · A. I. Yashin
Center for Population Health and Aging, Duke University,
Trent Hall, Room 002, Box 90408, Durham, NC 27708, USA
e-mail: kulminsk@duke.edu

frequencies in phases 3 and 4. Figure 1 (Fig. 6a and b in (Kulminski et al. 2013)) ensures that this is the case. Specifically, virtually all individuals genotyped in phase 4 and major part of individuals genotyped in phase 3 died during follow up in the FHS. This is an important finding which ensures that “major genotyping error” is not a unique explanation of genotype clustering in aging studies. Accordingly, just disregarding strong evidence of mortality selection (which has nothing to do with genotyping errors) is scientifically implausible because the selected groups of individuals with phenotypes of premature death can carry vulnerable alleles which are not carried by more successfully aging individuals.

Next, Dr. Paterson notes that call rate is “so low that for all three SNPs in phase 4, these SNPs would be excluded from any standard analysis.” Phase 4 is associated with the worst survival (Fig. 1). Therefore, we have to understand why we observe low call rate for SNPs in phase 4 rather than merely exclude them given that such SNPs can be characteristic for the very vulnerable subsample of the FHS participants. Phase 4 included so-called legacy samples with DNA extracted from 1,133 frozen blood samples kept in refrigerators since 1970s to 1980s (Cupples et al. 2009). It is a recognized problem that frozen blood is of lower quality than the whole blood. Given this problem all these 1,133 samples were carefully inspected and only well qualified samples were released to the FHS SHARe. This rigorous selection resulted in 674 qualified samples that

implied 59.5 % call rate. Thus, disregarding these samples would be implausible because low call rate is a priori expected for them as a result of quality control selection.

Dr. Paterson states that “phases of genotyping are confounded with generation (i.e., cohorts) in the study” and, accordingly, “this would be expected to demonstrate association with any phenotype that differs by generation, and/or age.” Figure 1 explicitly shows that this statement is not correct. Phases 3 and 4 are not associated with generations (or specific age) because they are present in the FHS original and Offspring cohorts. These phases, however, are associated with phenotypes of premature deaths of different severity. Therefore, by definition we have to expect differences in allele frequencies between these phases due to age-related processes. This is exactly what aging studies are looking for.

Dr. Paterson writes that “since phases are related to the cohorts, one would imagine that drastic differences in genotypes between family members would be expected to result in Mendelian errors.” First, it is incorrect that “phases are related to the cohorts” (see the above paragraph). Next, Mendelian errors imply that offspring could have alleles which are not inherited from their parents. All phases (including phases 3 and 4) are present in parents (i.e., the original FHS cohort) and offspring. Dr. Paterson explicitly shows in Table 1 in his letter that all alleles are concordant in phases 3 and 4. The fact that

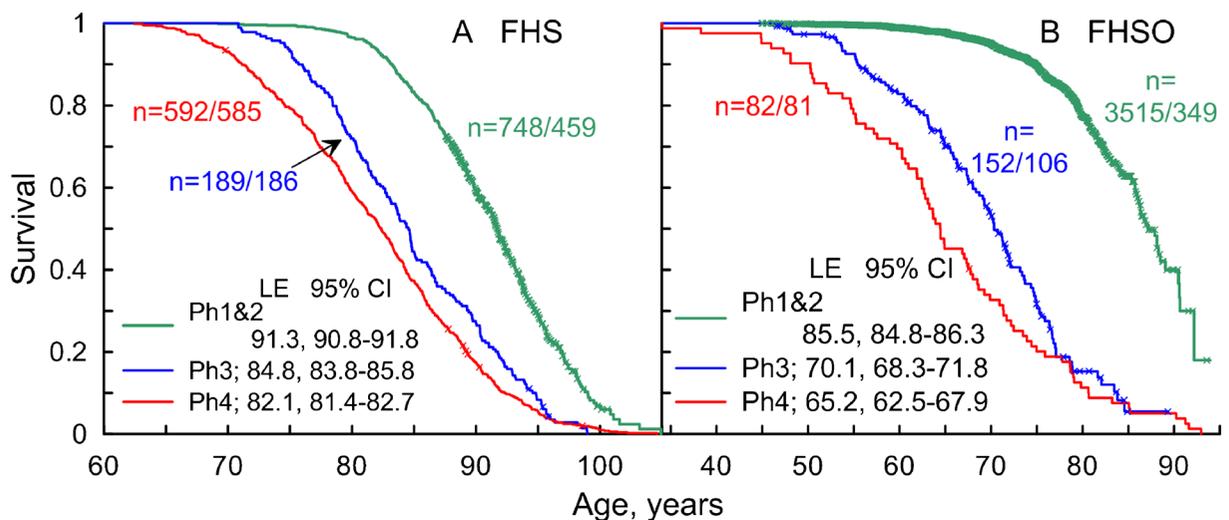


Fig. 1 Survival of the genotyped individuals selected for four genotyping phases from (A) FHS and (B) FHSO cohorts. Kaplan-Meier survival curves show age at death (line) or the end

of follow up in 2007 (crosses). Sample size (n) is for all/died individuals. LE=life expectancy; CI=confidence interval

there is no Mendelian errors in families is evident from Table 1 in (Kulminski et al. 2013). Besides, SNPs and families with Mendel errors more than 2 % were excluded from the analyses. Thus, Dr. Paterson statement contradicts to the presented evidences.

Dr. Paterson suggests that crucial step would be to analyze cluster plots for specific SNPs. It is, however, entirely unclear how the analyses of cluster plots of individual SNPs could help in explaining collective phenomena of inter-chromosomal linkage disequilibrium (LD). In the response to a recent letter of concern by Dr. Paterson published in the *Experimental Gerontology* (Kulminski 2012), it was detailed that neither bad nor good cluster plots of individual SNPs could explain that. Dr. Paterson further suggests that technical errors may not necessarily be randomly distributed that would result in spurious clustering. This statement, however, contradicts to the statement that the analysis of cluster plots of individual SNPs could be crucial step for detecting non-random technical artifacts because the analysis of individual cluster plots cannot explain spurious clustering. What this analysis is intended to suggest is just to exclude SNPs with “bad” plots, if any.

The latter, however, is not reasonable strategy given that SNPs in inter-chromosomal LD are associated with phenotype of premature death. Our analyses provide another relevant result which ensures that “technical problems” unlikely play a role in the observed phenomena of inter-chromosomal LD. This is the observation of highly significant, nearly three-fold enrichment of *non-synonymous coding SNPs*, SNPs which can alter amino acid sequence of proteins, among those in inter-chromosomal LD compared to non-synonymous coding SNPs present in the qualified set of about 38 K SNPs from the Affymetrix 50 K array, i.e., 41.2 % vs. 15 %, $p=1.6\times 10^{-9}$ (Kulminski et al. 2013). This is remarkably important result because it undoubtedly shows that technical artifacts of either random or non-random origin highly unlikely play a role in this phenomenon because

they have to be “smart” enough to selectively indicate functionally significant SNPs. Contrarily to technical artifacts, functional significance of SNPs is in perfect agreement with the nature of phenotype of premature death (Fig. 1).

All these evidences have been presented in our papers cited by Dr. Paterson. The fact that Dr. Paterson disregards them in favor of standard GWAS strategies implies that standardization of GWAS may be not helpful for studies aimed to reveal genetic origin of age-related phenotypes because it detracts attention from the specific of the problem.

References

- Atzmon G, Rincon M, Schechter CB, Shuldiner AR, Lipton RB, Bergman A, Barzilai N (2006) Lipoprotein genotype and conserved pathway for exceptional longevity in humans. *PLoS Biol* 4(4):e113
- Christensen K, Johnson TE, Vaupel JW (2006) The quest for genetic determinants of human longevity: Challenges and insights. *Nat Rev Genet* 7(6):436–448
- Cupples LA, Heard-Costa N, Lee M, Atwood LD (2009) Genetics analysis workshop 16 Problem 2: The framingham heart study data. *BMC Proc* 3(Suppl 7):S3
- De Benedictis G, Carotenuto L, Carrieri G, De Luca M, Falcone E, Rose G, Yashin AI, Bonafe M, Franceschi C (1998) Age-related changes of the 3'APOB-VNTR genotype pool in ageing cohorts. *Ann Hum Genet* 62(Pt 2):115–122. doi:10.1046/j.1469-1809.1998.6220115.x
- Kulminski A (2012) Have to or may? Re: Expression of concern re: Kulminski AM (2011). Complex phenotypes and phenomenon of genome-wide inter-chromosomal linkage disequilibrium in the human genome. *Exp Gerontol*. 46, 979–986. *Exp Gerontol* 47(6):481–482. doi:10.1016/j.exger.2012.03.009
- Kulminski AM, Culminskaya I, Yashin AI (2013) Inter-chromosomal level of genome organization and longevity-related phenotypes in humans. *Age (Dordr)* 35(2):501–518. doi:10.1007/s11357-011-9374-6
- Martin GM, Bergman A, Barzilai N (2007) Genetic determinants of human health span and life span: progress and new opportunities. *PLoS Genet* 3(7):e125