



Pure and Confounded Effects of Causal SNPs on Longevity: Insights for Proper Interpretation of Research Findings in GWAS of Populations with Different Genetic Structures

Anatoliy I. Yashin ^{*}, Ilya Zhbannikov, Liubov Arbeeveva, Konstantin G. Arbeev, Deqing Wu, Igor Akushevich, Arseniy Yashkin, Mikhail Kovtun, Alexander M. Kulminski, Eric Stallard, Irina Kulminskaya and Svetlana Ukraintseva ^{*}

Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, Durham, NC, USA

OPEN ACCESS

Edited by:

Elena G. Pasyukova,
Institute of Molecular Genetics of
Russian Academy of Sciences, Russia

Reviewed by:

Sangkyu Kim,
Tulane University, USA
Joan Fibla,
University of Lleida, Spain

*Correspondence:

Anatoliy I. Yashin
aiy@duke.edu
Svetlana Ukraintseva
svo@duke.edu

Specialty section:

This article was submitted to
Genetics of Aging,
a section of the journal
Frontiers in Genetics

Received: 30 August 2016

Accepted: 07 October 2016

Published: 08 November 2016

Citation:

Yashin AI, Zhbannikov I, Arbeeveva L,
Arbeev KG, Wu D, Akushevich I,
Yashkin A, Kovtun M, Kulminski AM,
Stallard E, Kulminskaya I and
Ukraintseva S (2016) Pure and
Confounded Effects of Causal SNPs
on Longevity: Insights for Proper
Interpretation of Research Findings in
GWAS of Populations with Different
Genetic Structures.
Front. Genet. 7:188.
doi: 10.3389/fgene.2016.00188

This paper shows that the effects of causal SNPs on lifespan, estimated through GWAS, may be confounded and the genetic structure of the study population may be responsible for this effect. Simulation experiments show that levels of linkage disequilibrium (LD) and other parameters of the population structure describing connections between two causal SNPs may substantially influence separate estimates of the effect of the causal SNPs on lifespan. This study suggests that differences in LD levels between two causal SNP loci within two study populations may contribute to the failure to replicate previous GWAS findings. The results of this paper also show that successful replication of the results of genetic association studies does not necessarily guarantee proper interpretation of the effect of a causal SNP on lifespan.

Keywords: linkage disequilibrium, population stratification, mortality selection, lack of replication, causal SNP, longevity related traits, population genetic structure

INTRODUCTION

The results of many genome-wide association studies (GWAS) of complex traits suffer from a lack of replication (Shen et al., 2005; Gorroochurn et al., 2007; Greene et al., 2009; Hart et al., 2013; Ioannidis, 2015; Maxwell et al., 2015; Torricco et al., 2016). Differences in population genetic structures among study populations are considered to be possible contributors to this problem (Greene et al., 2009). One aspect of population structure—the differences in genetic frequencies among subgroups of individuals comprising the population—was traditionally linked with the effects of population stratification (Wacholder et al., 2000; Price et al., 2006, 2010). Another one—the presence of linkage disequilibrium (LD) in many parts of the human genome including those that contain causal SNPs—was actively exploited in GWAS of complex traits (Cantor et al., 2010; Moore et al., 2010; Hayes, 2013). Methods of fine mapping following the “discovery” phase are used for evaluating causal SNPs (Clarke et al., 2007; Hassanein et al., 2010; Zapata, 2013; Kichaev et al., 2014; Morris, 2014). One could expect that the non-replication problem due to differences in LD patterns among study populations in GWAS would disappear if the detected marker SNP is a causal one, i.e., if it contributes to the variability of a trait. It turns out that the differences in LD levels around a functional SNP may still contribute to the non-replication problem. The estimated associations in this case depend on whether the detected functional SNP is in LD

with another functional SNP, the effects of these SNPs on the trait in the absence of LD (pure effects), and on the level of LD between corresponding SNP loci. This property has important consequences for interpretation of the results of genetic analyses of complex traits. In the presence of LD the estimated effects of a causal SNP may be spurious and may incorrectly characterize the biological relationships between the SNP and the trait. In contrast the pure effect of a given causal SNP estimated in the absence of LD with other such SNPs may correctly characterize the biological connections between the SNP and the trait. Therefore, for example, performing genetic analyses of African populations (that have lower levels of LD patterns for many SNP pairs than populations of European origin) has the potential to reduce bias in the estimated effects of functional SNPs on a trait caused by the presence of LD between functional loci (Shifman et al., 2003). This condition is, however, not sufficient because of the possible presence of hidden gene/gene interaction effects, gene/environment correlations, and gene/environment interaction effects (Ukrainitseva et al., 2016).

Human lifespan and many other aging, health and longevity related traits are multifactorial phenotypes, that is, they are affected by many genetic and non-genetic factors. The relationships between genes and these phenotypes have special features that distinguish them from other complex traits, influence methods of their genetic analyses, and affect the interpretation of the research results. The genetic variants that influence aging, health, and longevity related traits generate age dependent changes in the population genetic structure, i.e., changes in the frequencies of genetic variants and in the levels of linkage disequilibrium (LD) among them. This feature has important implications for studies focused on the replication of GWAS research findings: Independent populations involved in such studies often have different genetic structures, due in part to the differences in the population age distribution at the time of biospecimen collection. As a result, the frequencies of the genetic variants associated with these traits and their LD patterns may differ even if the genetic structures in the corresponding population cohorts were the same at birth.

Detecting statistically significant associations of genetic variants with complex traits is not the end of the genetic analyses. One reason is that the relationship between a detected marker SNP and the complex trait of interest is not, necessarily, a causal one. More often these relationships serve as proxies for the real effect of some unobserved causal SNPs [due to linkage disequilibrium (LD) between the marker and causal SNPs], and, hence, do not have a direct biological effect on the phenotype. To generate insights about the biological mechanisms responsible for the trait's variability one has to identify the causal SNPs responsible for the association signal. To identify such SNPs a number of efficient fine-mapping procedures have been recommended (Zaitlen et al., 2010; Hormozdiari et al., 2015). The main limitation of existing methods is that they seek to identify a single causal variant which is independent of (not in LD with) other causal variants (Hormozdiari et al., 2014). Since this is not sufficiently realistic, a new approach that allows for efficient detection of multiple causal variants has been proposed (Hormozdiari et al., 2014). The case where two or more causal

SNPs are in LD creates additional problems for interpretation of the results of genetic association studies.

In this paper we show that the estimates of the effects of a causal SNP on lifespan depend on the genetic structure of the population under study (e.g., the level of LD of the SNP with other causal SNPs). Genetic association studies of this trait using data from populations with different LD levels are likely to produce different results. We show that differences in population genetic structures can explain why genetic variants favorable for longevity in one population appear as harmful risk factors in another population. Population structure may also be responsible for the age-specific effects of genetic variants on mortality risk. Differences in genetic structures in distinct populations may be responsible for the low level of replicability of GWAS of human aging, health, and longevity related traits.

DATA AND METHODS

To show how the effects of differences in LD levels and other parameters of the population genetic structure influence the results of genetic association studies we consider lifespan as the trait of interest and two causal SNP loci with minor and major alleles at each locus. Accordingly we will use a simple model of the genetic connections with lifespan that requires only dichotomous genetic variables (e.g., this is the case when the minor alleles at each locus have dominant genetic effects on lifespan). Extensions for cases with SNP genotypes are straightforward. Let $V_1 = (0, 1)$ and $V_2 = (0, 1)$ be the values of the genetic variants at these loci, where "1" denotes the minor allele and "0" corresponds to the major allele (see Figure 1S, Supplementary Materials). The genetic variants affect survival through mortality risks specified for haplotypes (i, j) where $i = 0, 1$ characterizes the presence of major or minor alleles in the first SNP locus, and $j = 0, 1$ describes the presence of such alleles in the second SNP locus. Denote by μ_{ij} , $i, j = (0, 1)$, the mortality risks for each of four haplotypes. We assume for simplicity that these risks are age-independent. Adding age dependence will not qualitatively change our results. Let $m_{00}(t), m_{10}(t), m_{01}(t)$, and $m_{11}(t)$ be the frequencies of corresponding haplotypes at age $t \geq t_0$, and $\mu_{00}, \mu_{10}, \mu_{01}, \mu_{11}$ be the mortality risks for these haplotypes. The minor allele frequencies $m_1(t)$ and $m_2(t)$ at any age $t \geq t_0$ are defined as $m_1(t) = m_{10}(t) + m_{11}(t)$ and $m_2(t) = m_{01}(t) + m_{11}(t)$. We assume that in the absence of LD at age t_0 the minor allele in the first locus has a positive association with lifespan (i.e., it is a "longevity" allele), and the minor allele in the second locus has a negative association with lifespan (i.e., it is a "vulnerability" allele).

To make this possible we assume that mortality risks for the haplotypes satisfy the inequalities:

$$\mu_{10} < \mu_{00} < \mu_{11} < \mu_{01} \quad (1)$$

To minimize the effects of the mathematical representation of the connections between the genetic factors and mortality on the results of our simulation experiment we considered two models

of mortality risks. For Model 1, we assumed that:

$$\mu_{10} = \mu_{00}(1+R_1), \mu_{01} = \mu_{00}(1+R_2), \text{ and } \mu_{11} = \mu_{00}(1+R_1+R_2) \quad (2)$$

where R_1 and R_2 could be called the increments to the haplotypes' relative risks associated with the presence of minor alleles in the first and second loci, respectively. To guarantee Equation (1) the values of R_1 and R_2 have to satisfy inequalities:

$$-1 < R_1 < 0, R_2 > 0, \text{ and } R_1 + R_2 > 0.$$

For Model 2, we assumed that:

$$\mu_{10} = \mu_{00}H_1, \mu_{01} = \mu_{00}H_2, \text{ and } \mu_{11} = \mu_{00}H_1H_2. \quad (3)$$

These relative risks have to satisfy the inequalities:

$$H_1 < 1, H_2 > 1, \text{ and } H_1H_2 > 1.$$

In our simulation experiments, we fix the initial frequencies $m_1(t_0)$ and $m_2(t_0)$, specify values of R_1 and R_2 (or H_1 and H_2), fix the initial levels of $LD(t_0)$, and run the models to calculate the age trajectories of mortality rates for the carriers $\mu_1(t)$ and non-carriers $\mu_0(t)$ of the minor allele in the first locus. The calculation of the age trajectories of mortality rates for the carriers and non-carriers of the minor allele in the first SNP locus, as well as the age trajectories of other variables, are shown in the Supplementary Materials.

RESULTS

The results of four simulation experiments with Model 1 are shown in **Figure 1**. The corresponding model parameters are presented in **Table 1**.

One can see from **Figure 1** and **Table 1** that the graphs of the mortality rates shown in **Figures 1A–C** correspond to the same initial values of the minor allele frequencies $m_1(t_0)$ and $m_2(t_0)$ in the two SNP loci in each of three experiments and the same (constant) values of μ_{00} , R_1 , and R_2 . The latter values are used for calculating mortality risks for haplotypes (Equation 2). The differences in the initial LD levels and in haplotype frequencies are responsible for the radical differences in relationships between mortality rates for carriers and non-carriers of the minor allele at the SNP1 locus (shown in **Figure 1**). **Figure 1A** shows that the mortality rate for carriers of the minor allele of SNP 1 is lower than that for non-carriers of this allele, suggesting that this is a “longevity” allele. **Figure 1B** shows that the mortality rate for carriers of the minor allele of SNP 1 is higher than that for non-carriers of this allele, suggesting that this is a “vulnerability” allele. **Figure 1C** shows that the mortality rates for carriers and non-carriers of the minor allele of SNP 1 intersect: the harmful effect of the allele on mortality risk at the initial age interval changed to a beneficial one later in life. **Figure 1D** shows that mortality rates for carriers and non-carriers of the minor allele of SNP 1 may intersect in the opposite way: the beneficial effect of the allele on mortality risk at the initial age interval changed to a harmful one later in life. The results of analyses using Model 2 are similar to those of Model 1. They can be found in **Figure 2S** in the Supplementary Materials.

DISCUSSION

In this paper we showed that the finding of an association of lifespan with a functional SNP does not exclude the possibility of non-replication due to differences in initial LD patterns among study populations. The detected associations may differ dramatically if the study populations have different LD patterns around the functional SNP locus. This may happen if the LD area includes one or more additional causal loci affecting the same trait. We showed that in this case the difference in the LD value may change the sign of the estimated genetic association with mortality risks to the opposite direction. The possibility of such a phenomenon was highlighted by Lin et al. (2007) who provided examples of genetic variants that show such “flip-flop” effects in genetic association studies. The authors performed comprehensive simulation analyses of “static” cases in which LD levels influence relative risks of corresponding genetic variants. The results of their study provide us with valuable information for proper interpretation of findings from genetic association studies of risks of diseases whose occurrence does not affect mortality risks. Additional analyses motivated by Lin et al. (2007) showed that differences in frequencies of genetic variants may also produce reversals of the effects of genetic variants on disease risks (Zaykin and Shibata, 2008). We also showed that differences in the LD levels and other parameters of population genetic structure can change the signs of the genetic effects on longevity related traits at different age intervals.

Despite their high relevance for many genetic association studies, the “static” cases of LD effects on relative risks described in Lin et al. (2007) and Zaykin and Shibata (2008) may be less informative when studying the risks of chronic conditions (e.g., CVD, cancer) whose occurrences do influence mortality risk. This is because genetic variants affecting such risks are involved in mortality selection processes in which the most vulnerable individuals in a population birth cohort tend to become sick and die first. This process, in turn, generates age-related changes in genotype (allele) frequencies, in values of disease risks for carriers and non-carriers of selected genetic variants, as well as in mortality risks. In particular, non-monotonic age trajectories of the frequencies of selected alleles or genotypes (Yashin et al., 1999, 2000; Atzmon et al., 2006; Bergman et al., 2007) could result from specific LD patterns. The estimated effects of genetic variants on longevity related traits in populations that have different initial LD levels between selected SNP loci or, more generally, different genetic structures may differ dramatically. The age trajectories of LD values between functional SNP loci may also change with increasing age of the population cohorts.

Recently Ukraintseva et al. (2016) investigated possible causes and mechanisms of paradoxical behaviors of genetic risk factors including age-dependence of the effects of genetic variants on disease and mortality risks. Such effects may result from pleiotropic influences of a genetic variant on chronic health disorders and on aging-related phenotypes. They can also be caused by changes in the epistatic effects on mortality risk or in the effects of interactions of genetic factors with environmental conditions with increasing age. Each of these causes may result in non-replication. In this paper we modeled another possible cause

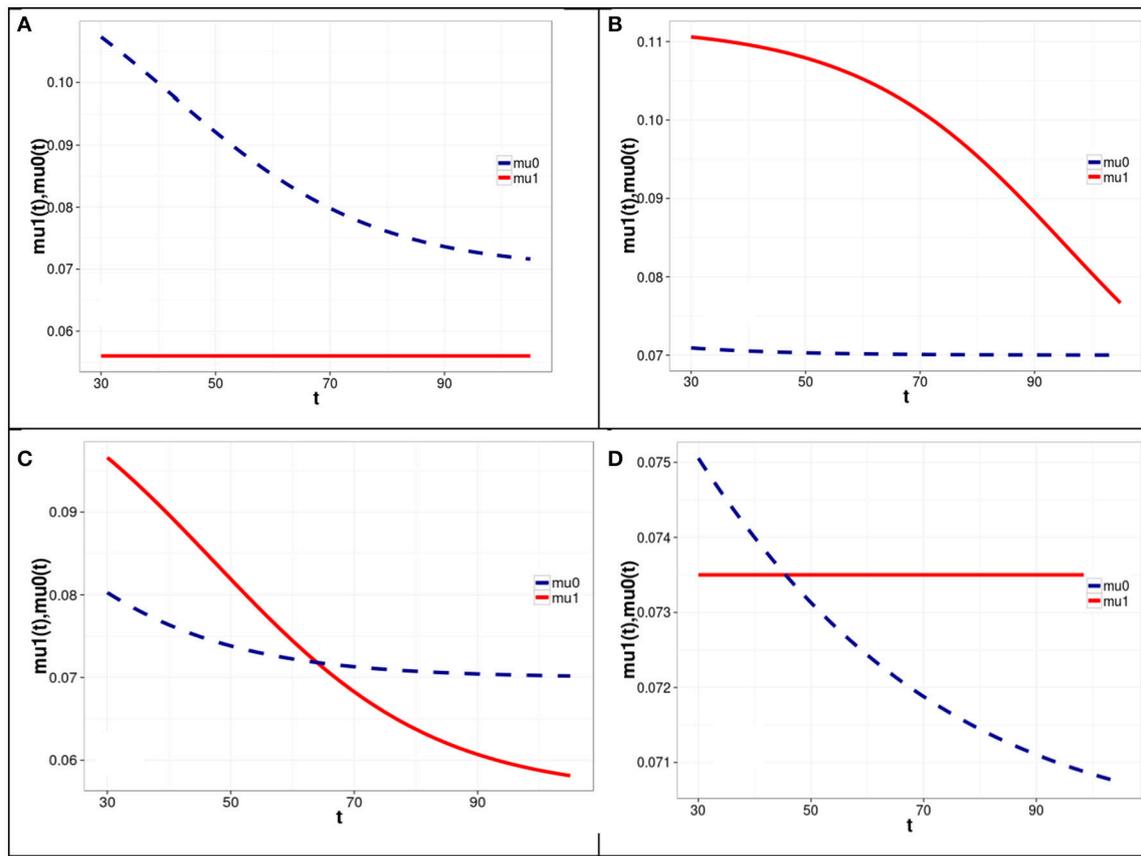


FIGURE 1 | Graphs of mortality rates for carriers (solid line) and non-carriers (dashed line) of the minor allele at the first SNP locus corresponding to different levels of LD between the two loci and different haplotype frequencies in four simulation experiments with Model 1 of the genetic influence on mortality rates. Model parameters corresponding to the graphs shown in (A–D) are represented in Table 1 in rows (A–D), respectively. Equations linking the mortality rates for the carriers and non-carriers of the minor allele of the first SNP (shown in A–D of Figure 1) with the mortality risks for haplotypes are given on last page of the Supplementary Materials.

TABLE 1 | Parameter values used in four simulation experiments with Model 1.

	$m_1(t_0)$	$m_2(t_0)$	$LD(t_0)$	$m_{00}(t_0)$	$m_{10}(t_0)$	$m_{01}(t_0)$	$m_{11}(t_0)$	μ_{00}	R_1	R_2
(A)	0.4	0.4	-0.16	0.2	0.4	0.4	0	0.07	-0.2	0.8
(B)	0.4	0.4	0.23	0.59	0.01	0.01	0.39	0.07	-0.2	0.8
(C)	0.4	0.4	0.13	0.49	0.11	0.11	0.29	0.07	-0.2	0.8
(D)	0.28	0.41	0.165	0.59	0	0.13	0.28	0.07	-0.35	0.4

$m_1(t_0)$ is the initial value of the minor allele frequency at the SNP1 locus; $m_2(t_0)$ is the initial value of the minor allele frequency at the SNP2 locus; $LD(t_0)$ is the initial value of linkage disequilibrium between the SNP1 and SNP2 loci; $m_{00}(t_0)$ is the initial value of the haplotype (0, 0) frequency; $m_{10}(t_0)$ is the initial value of the haplotype (1, 0) frequency; $m_{01}(t_0)$ is the initial value of the haplotype (0, 1) frequency; $m_{11}(t_0)$ is the initial value of the haplotype (1, 1) frequency; μ_{00} is the mortality risk for carriers of haplotype (00); R_1 and R_2 are the increments to the haplotypes' relative risks associated with the presence of minor alleles in the first and in the second loci, respectively. Mortality risks for haplotypes (0, 0), (1, 0), (0, 1), and (1, 1) are calculated using equation (2).

for non-replication resulting from the differences in LD levels between pairs of functional SNP loci, each of which is associated with lifespan, in two populations. Taking the possibility of such effects into account is crucial for proper interpretation of the results of genetic analyses of complex traits.

The results presented in this paper will be highly relevant to genetic studies of human aging, health, and longevity related

traits only if the assumed variability in LD values at different locations of the human genome is common in populations used for genetic association studies. Numerous studies show that LD patterns vary from one population to the next at different areas of the human genome. The differences in LD of lipid-associated loci in different study populations and the implications of these differences for genetic analyses across multiple populations were

investigated in Teo and Sim (2010). Charles et al. (2014) provided an overview of LD measures used in analyses of population genetic structures and discussed their possible use in GWAS of populations with discordant LD patterns. Sawyer et al. (2005) analyzed LD patterns of selected genomic regions in diverse populations and provided evidence of marked differences in haplotype frequencies and in corresponding LD patterns. Liu et al. (2013) emphasized that genetic analyses of admixed populations may result in confounding due to different patterns of genetic structure. Thomson and colleagues found differences in LD patterns between populations of Hutterites and Europeans (Thompson et al., 2010). The results of Koda and colleagues suggest that some differences in the estimates of risk for coronary heart disease can be explained by population differences in haplotype frequencies of the PON1 haplotypes (Koda et al., 2004). Lohmueller et al. (2006) found genetic variation in the LD patterns in the G-protein coupled receptor kinase 4 (GRK4 gene). The product of this gene inhibits the dopamine receptor D1 (DRD1). The LD patterns were found to be different in SNPs related to 81 osteoporosis candidate genes (Kim et al., 2007). Garner and Slatkin found extensive variation in the LD patterns between a disease locus and one or two marker SNP loci—even for closely linked loci (Garner and Slatkin, 2003). They also found that the distribution of LD patterns between common variants is strongly influenced by ancestral population size.

Thus, the differences in the LD patterns among study populations may result in misleading interpretations of the biological mechanisms involved in trait regulation. Differences in LD are likely to be present in study populations sampled in the U.S. Many such populations contain mixtures of subpopulations that migrated to the U.S. from different parts of the world. In cases where the LD patterns in the study populations do not coincide with those of the specific ethnicities in HAP/MAP or other reference datasets, one may expect the results of genetic association studies to be different. Estimation of the LD patterns around specific SNP loci in the targeted study populations can help one to better understand the sources of inconsistency in the GWAS results.

The fact that the second functional SNP may be unobserved (e.g., not included in the list of SNPs for genotyping) in a given study population makes the results of traditional replication analyses unpredictable and may result in a failure to replicate earlier research findings. To reduce the chances of non-replication due to differences in the LD patterns in studies of independent populations, one has to conduct the replication study using populations with similar patterns of LD around specific targeted loci. This means that different study populations may be needed for replication of the research results concerning the effects of different SNPs. Non-replication due to other (non-LD) reasons may also occur. Dealing with these other sources of non-replication will require different approaches (Ukrainitseva et al., 2016).

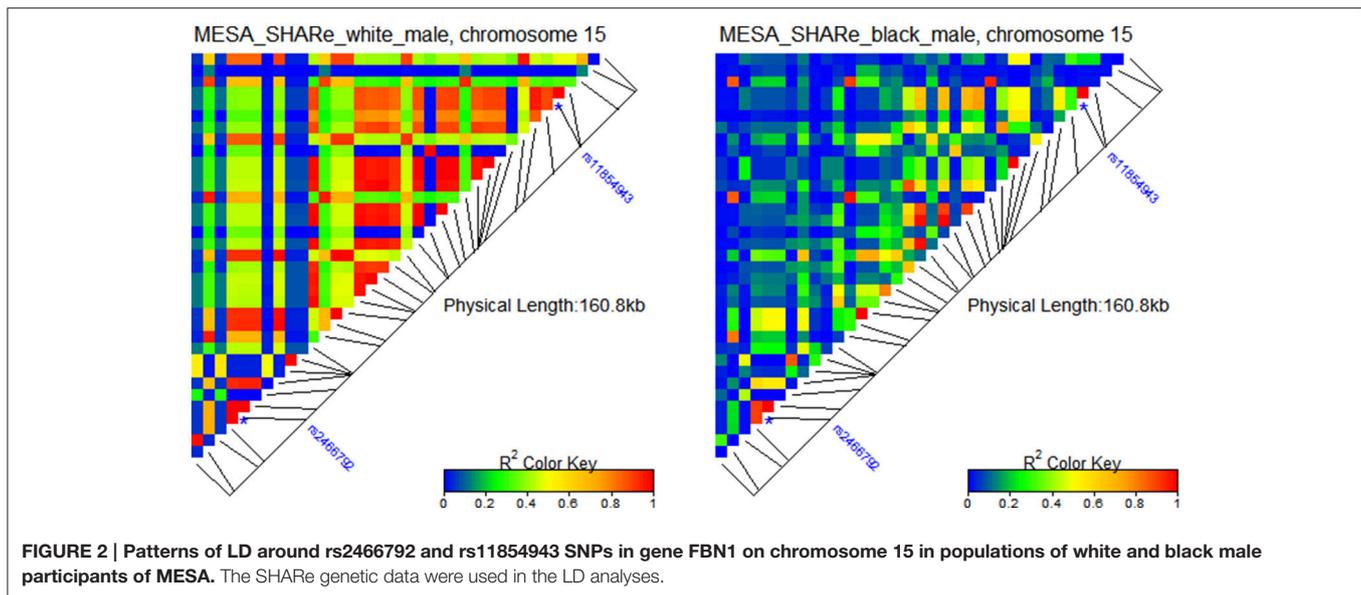
Figure 2 shows two distinct patterns of LD for white and black male participants of the Multi-Ethnic Study of Atherosclerosis (MESA) SNP Health Association Resource (SHARe), available from dbGaP.

One can see from the figure that there are many more SNPs that are in strong LD in the population of white males than in the population of black males from the same study. The two SNPs shown in this figure are rs2466792 and rs11854943 in gene FBN1 on chromosome 15. The minor allele of the rs2466792 SNP (MAF 43.6%) reduces mortality risk (relative risk HR = 0.7; *p*-value 0.03) among black males MESA participants. The minor allele of the rs11854943 SNP (MAF 10.7%) increases mortality risk (relative risk HR = 1.5, *p*-value 0.04) among black males MESA participants. Neither allele shows an association with lifespan among white male MESA participants.

Although, the causes of the different effects of selected SNPs on mortality risks between black and white males may include influences of gene-gene and gene-environment interactions which are not estimated in these analyses, the difference in LD could also make substantial contributions to these differences, as illustrated in Figure 3S. This figure shows age patterns of mortality rates for carriers and non-carriers of minor alleles in populations with different levels of LD. Panels “e, f” of Figure 3S in this figure correspond to populations with higher and lower LD values, respectively. The intersection of age patterns of mortality risks for carriers and non-carriers of the minor allele in panel “e” of Figure 3S indicates that the estimate of relative risk using the Cox regression model over the entire age interval is likely to be close to one (because differences between the curves will tend to compensate each other), i.e., analyses will not show statistically significant associations of this genetic variant with lifespan. The two non-intersecting age patterns of mortality risks for carriers (red line) and non-carriers (blue dashed line) of the minor allele in the population with the lower initial value of LD (panel “f” of Figure 3S) indicate that the differences in hazard rates and hence in survival functions between these groups can be estimated by statistical methods (e.g., using a Cox-type regression model).

It is important to note that differences in the LD patterns may be responsible for non-replication of the results of (non-genetic) epidemiological studies. Such confounding may take place when two or more causal SNP loci influence different phenotypes and are in LD (Aissani, 2014). In this case, one phenotype will always show its association with the other one in epidemiological studies although such a connection may be not causal but induced instead by the LD between the corresponding SNP loci. Further, the use of one such phenotype as a covariate in a GWAS of the second phenotype may substantially modify the estimate of genetic association with the second trait. Different patterns of LD among study populations are expected to produce different confounding effects, jeopardizing proper interpretation of the results of epidemiological studies.

The effect of the minor allele of a causal SNP (SNP1) on lifespan is considered to be “pure” when it is estimated in a population whose genetic structure does not have LD between SNP1 and any other causal SNP. In the absence of other (e.g., non-genetic) confounders this effect characterizes the biological connections between variations in SNP1 and lifespan. If in some other population the causal SNP (SNP1) is in strong LD with another causal SNP (SNP2), then its estimated effect on lifespan in a new genetic association study is confounded and



may be radically different from the pure effect (e.g., have the opposite sign). This confounded effect of SNP1 on lifespan may be successfully replicated in the new genetic association study if the level of LD between SNP1 and SNP2 and other parameters of genetic structure in both populations are about the same. Such replication, however, does not guarantee proper interpretation of the effect of SNP1 on lifespan because it is confounded by the effect of SNP2 and depends on the LD level between two SNPs. This also means that analyses of next (third) population with a different LD level between SNP1 and SNP2 may not replicate the results of the two earlier studies. One, however, cannot say that the results of the first two analyses are “more correct” than the findings from the third study because both are confounded by the LD with another causal SNP. The estimates of the pure effect of SNP1 on lifespan can be obtained from the data on the initial population in which SNP1 and SNP2 were not in LD.

Better understanding of how human genes regulate relationships among aging, health, and longevity related traits will contribute to development of intervention strategies aiming to increase healthy lifespan and reduce the burden of chronic non-communicable diseases. A number of promising genetic associations with human longevity related traits were detected in GWAS and confirmed in other populations (Willcox et al., 2008; Deelen et al., 2011; Nebel et al., 2011; Flachsbart et al., 2013; Soerensen et al., 2013, 2015; Broer et al., 2015). At the same time, numerous attempts to replicate many other detected associations with these traits failed (Novelli et al., 2008). Elbaz and colleagues could not replicate 13 SNPs significantly associated with Parkinson’s disease (Elbaz et al., 2006). Campa and colleagues could not replicate seven loci associated with the risk of pancreatic cancer detected in studies of two Asian populations (Campa et al., 2013). Suarez-Gestal and colleagues were not able to replicate significant associations between 16 SNPs and responses to specific treatments of rheumatoid

arthritis (Suarez-Gestal et al., 2010). Nemr and colleagues could not replicate association of EXT2 genetic variants with the risk of type 2 diabetes in a population of Lebanese Arabs (Nemr et al., 2013). Such persistent non-replication of the results of genetic association studies raises legitimate questions about the factors and mechanisms responsible. Previous research has identified a number of conditions that may contribute to the lack of replication (Ioannidis, 2007; Kavvoura and Ioannidis, 2008; Kraft et al., 2009; Ukraintseva et al., 2016). In this paper, we demonstrated that differences in genetic structures between various study populations can make substantial contributions to the non-replication of the results of genetic association studies of human longevity.

AUTHOR CONTRIBUTIONS

AY and SU conceived the paper and wrote the first draft. IZ developed software for simulations and together with AY performed simulation experiments. LA and KA performed LD analyses. ArY, KA, and ES provide valuable comments and suggestions and together with DW, MK, AK, IK, IA make valuable contributions to the discussion section. All coauthors participated equally in preparing the ultimate version of the manuscript and Supplementary materials.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the NIA/NIH grants R01AG046860, P01AG043352, P30AG034424, and 5U01AG023712-10 (LLFS). The results of this paper were discussed at the conference call of the Research and Development (R&D) Committee of the Long Life Family Study (LLFS) group. The paper does not necessarily reflect the opinions or views of the entire LLFS group. The content is solely the responsibility of the authors and does not necessarily represent

the official views of the National Institutes of Health. MESA and the MESA SHARE project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, and CTSA UL1-RR-024156. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa

Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. The authors thank Debra Fincham for help in preparing this paper for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fgene.2016.00188/full#supplementary-material>

REFERENCES

- Aissani, B. (2014). Confounding by linkage disequilibrium. *J. Hum. Genet.* 59, 110–115. doi: 10.1038/jhg.2013.130
- Atzmon, G., Rincon, M., Schechter, C. B., Shuldiner, A. R., Lipton, R. B., Bergman, A., et al. (2006). Lipoprotein genotype and conserved pathway for exceptional longevity in humans. *PLoS Biol.* 4:e113. doi: 10.1371/journal.pbio.0040113
- Bergman, A., Atzmon, G., Ye, K., MacCarthy, T., and Barzilai, N. (2007). Buffering mechanisms in aging: a systems approach toward uncovering the genetic component of aging. *PLoS Comput. Biol.* 3:e170. doi: 10.1371/journal.pcbi.0030170
- Broer, L., Buchman, A. S., Deelen, J., Evans, D. S., Faul, J. D., Lunetta, K. L., et al. (2015). GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *J. Gerontol. A Biol. Sci. Med. Sci.* 70, 110–118. doi: 10.1093/gerona/glu166
- Campa, D., Rizzato, C., Bauer, A. S., Werner, J., Capurso, G., Costello, E., et al. (2013). Lack of replication of seven pancreatic cancer susceptibility loci identified in two Asian populations. *Cancer Epidemiol. Biomarkers Prev.* 22, 320–323. doi: 10.1158/1055-9965.EPI-12-1182
- Cantor, R. M., Lange, K., and Sinsheimer, J. S. (2010). Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *Am. J. Hum. Genet.* 86, 6–22. doi: 10.1016/j.ajhg.2009.11.017
- Charles, B. A., Shriner, D., and Rotimi, C. N. (2014). Accounting for linkage disequilibrium in association analysis of diverse populations. *Genet. Epidemiol.* 38, 265–273. doi: 10.1002/gepi.21788
- Clarke, G. M., Carter, K. W., Palmer, L. J., Morris, A. P., and Cardon, L. R. (2007). Fine mapping versus replication in whole-genome association studies. *Am. J. Hum. Genet.* 81, 995–1005. doi: 10.1086/521952
- Deelen, J., Beekman, M., Uh, H. W., Helmer, Q., Kuningas, M., Christiansen, L., et al. (2011). Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* 10, 686–698. doi: 10.1111/j.1474-9726.2011.00705.x
- Elbaz, A., Nelson, L. M., Payami, H., Ioannidis, J. P., Fiske, B. K., Annesi, G., et al. (2006). Lack of replication of thirteen single-nucleotide polymorphisms implicated in Parkinson's disease: a large-scale international study. *Lancet Neurol.* 5, 917–923. doi: 10.1016/S1474-4422(06)70579-8
- Flachsbart, F., Möller, M., Däumer, C., Gentschew, L., Kleindorp, R., Krawczak, M., et al. (2013). Genetic investigation of FOXO3A requires special attention due to sequence homology with FOXO3B. *Eur. J. Hum. Genet.* 21, 240–242. doi: 10.1038/ejhg.2012.83
- Garner, C., and Slatkin, M. (2003). On selecting markers for association studies: patterns of linkage disequilibrium between two and three diallelic loci. *Genet. Epidemiol.* 24, 57–67. doi: 10.1002/gepi.10217
- Gorroochurn, P., Hodge, S. E., Heiman, G. A., Durner, M., and Greenberg, D. A. (2007). Non-replication of association studies: “pseudo-failures” to replicate? *Genet. Med.* 9, 325–331. doi: 10.1097/GIM.0b013e3180676d79
- Greene, C. S., Penrod, N. M., Williams, S. M., and Moore, J. H. (2009). Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS ONE* 4:e5639. doi: 10.1371/journal.pone.0005639
- Hart, A. B., de Wit, H., and Palmer, A. A. (2013). Candidate gene studies of a promising intermediate phenotype: failure to replicate. *Neuropsychopharmacology* 38, 802–816. doi: 10.1038/npp.2012.245
- Hassanein, M. T., Lyon, H. N., Nguyen, T. T., Akyzbekova, E. L., Waters, K., Lettre, G., et al. (2010). Fine mapping of the association with obesity at the FTO locus in African-derived populations. *Hum. Mol. Genet.* 19, 2907–2916. doi: 10.1093/hmg/ddq178
- Hayes, B. (2013). Overview of statistical methods for genome-wide association studies (GWAS). *Methods Mol. Biol.* 1019, 149–169. doi: 10.1007/978-1-62703-447-0_6
- Hormozdiari, F., Kichaev, G., Yang, W. Y., Pasaniuc, B., and Eskin, E. (2015). Identification of causal genes for complex traits. *Bioinformatics* 31, i206–i213. doi: 10.1093/bioinformatics/btv240
- Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B., and Eskin, E. (2014). Identifying causal variants at loci with multiple signals of association. *Genetics* 198, 497–508. doi: 10.1534/genetics.114.167908
- Ioannidis, J. P. A. (2007). Non-replication and inconsistency in the genome-wide association setting. *Hum. Hered.* 64, 203–213. doi: 10.1159/000103512
- Ioannidis, J. P. A. (2015). Failure to replicate: sound the alarm. *Cerebrum* 2015:cer-12a-15.
- Kavvoura, F. K., and Ioannidis, J. P. (2008). Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. *Hum. Genet.* 123, 1–14. doi: 10.1007/s00439-007-0445-9
- Kichaev, G., Yang, W. Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A. L., et al. (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* 10:e1004722. doi: 10.1371/journal.pgen.1004722
- Kim, K. S., Kim, G. S., Hwang, J. Y., Lee, H. J., Park, M. H., Kim, K. J., et al. (2007). Single nucleotide polymorphisms in bone turnover-related genes in Koreans: ethnic differences in linkage disequilibrium and haplotype. *BMC Med. Genet.* 8:70. doi: 10.1186/1471-2350-8-70
- Koda, Y., Tachida, H., Soejima, M., Takenaka, O., and Kimura, H. (2004). Population differences in DNA sequence variation and linkage disequilibrium at the PON1 gene. *Ann. Hum. Genet.* 68(Pt 2), 110–119. doi: 10.1046/j.1529-8817.2003.00077.x
- Kraft, P., Zeggini, E., and Ioannidis, J. P. (2009). Replication in genome-wide association studies. *Stat. Sci.* 24, 561–573. doi: 10.1214/09-STS290
- Lin, P. I., Vance, J. M., Pericak-Vance, M. A., and Martin, E. R. (2007). No gene is an island: the flip-flop phenomenon. *Am. J. Hum. Genet.* 80, 531–538. doi: 10.1086/512133
- Liu, J., Lewinger, J. P., Gilliland, F. D., Gauderman, W. J., and Conti, D. V. (2013). Confounding and heterogeneity in genetic association studies with admixed populations. *Am. J. Epidemiol.* 177, 351–360. doi: 10.1093/aje/kws234
- Lohmueller, K. E., Wong, L. J. C., Mauney, M. M., Jiang, L., Felder, R. A., Jose P. A., et al. (2006). Patterns of genetic variation in the hypertension candidate gene GRK4: ethnic variation and haplotype structure. *Ann. Hum. Genet.* 70(Pt 1), 27–41. doi: 10.1111/j.1529-8817.2005.00197.x
- Maxwell, S. E., Lau, M. Y., and Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *Am. Psychol.* 70, 487–498. doi: 10.1037/a0039400
- Moore, J. H., Asselbergs, F. W., and Williams, S. M. (2010). Bioinformatics challenges for genome-wide association studies. *Bioinformatics* 26, 445–455. doi: 10.1093/bioinformatics/btp713
- Morris, A. P. (2014). Fine mapping of type 2 diabetes susceptibility loci. *Curr. Diab. Rep.* 14, 549. doi: 10.1007/s11892-014-0549-2

- Nebel, A., Kleindorp, R., Caliebe, A., Nothnagel, M., Blanché, H., Junge, O., et al. (2011). A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech. Ageing Dev.* 132, 324–330. doi: 10.1016/j.mad.2011.06.008
- Nemr, R., Al-Busaidi, A. S., Sater, M. S., Ehtay, A., Saldanha, F. L., Racoubian, E., et al. (2013). Lack of replication of common EXT2 gene variants with susceptibility to type 2 diabetes in Lebanese Arabs. *Diabetes Metab.* 39, 532–536. doi: 10.1016/j.diabet.2013.05.001
- Novelli, V., Viviani Anselmi, C., Roncarati, R., Guffanti, G., Malovini, A., Piluso, G., et al. (2008). Lack of replication of genetic associations with human longevity. *Biogerontology* 9, 85–92. doi: 10.1007/s10522-007-9116-4
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847
- Price, A. L., Zaitlen, N. A., Reich, D., and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* 11, 459–463. doi: 10.1038/nrg2813
- Sawyer, S. L., Mukherjee, N., Pakstis, A. J., Feuk, L., Kidd, J. R., Brookes, A. J., et al. (2005). Linkage disequilibrium patterns vary substantially among populations. *Eur. J. Hum. Genet.* 13, 677–686. doi: 10.1038/sj.ejhg.5201368
- Shen, H., Liu, Y., Liu, P., Recker, R. R., and Deng, H. W. (2005). Nonreplication in genetic studies of complex diseases—lessons learned from studies of osteoporosis and tentative remedies. *J. Bone Miner. Res.* 20, 365–376. doi: 10.1359/JBMR.041129
- Shifman, S., Kuypers, J., Kokoris, M., Yakir, B., and Darvasi, A. (2003). Linkage disequilibrium patterns of the human genome across populations. *Hum. Mol. Genet.* 12, 771–776. doi: 10.1093/hmg/ddg088
- Soerensen, M., Dato, S., Tan, Q., Thinggaard, M., Kleindorp, R., Beekman, M., et al. (2013). Evidence from case-control and longitudinal studies supports associations of genetic variation in APOE, CETP, and IL6 with human longevity. *Age (Dordr)* 35, 487–500. doi: 10.1007/s11357-011-9373-7
- Soerensen, M., Nygaard, M., Dato, S., Stevnsner, T., Bohr, V. A., Christensen, K., et al. (2015). Association study of FOXO3A SNPs and aging phenotypes in Danish oldest-old individuals. *Aging Cell* 14, 60–66. doi: 10.1111/acel.12295
- Suarez-Gestal, M., Perez-Pampin, E., Calaza, M., Gomez-Reino, J. J., and Gonzalez, A. (2010). Lack of replication of genetic predictors for the rheumatoid arthritis response to anti-TNF treatments: a prospective case-only study. *Arthritis Res. Ther.* 12, R72. doi: 10.1186/ar2990
- Teo, Y. Y., and Sim, X. (2010). Patterns of linkage disequilibrium in different populations: implications and opportunities for lipid-associated loci identified from genome-wide association studies. *Curr. Opin. Lipidol.* 21, 104–115. doi: 10.1097/MOL.0b013e3283369e5b
- Thompson, E. E., Sun, Y., Nicolai, D., and Ober, C. (2010). Shades of gray: a comparison of linkage disequilibrium between Hutterites and Europeans. *Genet. Epidemiol.* 34, 133–139. doi: 10.1002/gepi.20442
- Torricco, B., Chiocchetti, A. G., Bacchelli, E., Trabetti, E., Hervás, A., Franke, B., et al. (2016). Lack of replication of previous autism spectrum disorder GWAS hits in European populations. *Autism Res.* doi: 10.1002/aur.1662. [Epub ahead of print].
- Ukrainseva, S., Yashin, A., Arbeev, K., Kulminski, A., Akushevich, I., Wu, D., et al. (2016). Puzzling role of genetic risk factors in human longevity: “risk alleles” as pro-longevity variants. *Biogerontology* 17, 109–127. doi: 10.1007/s10522-015-9600-1
- Wacholder, S., Rothman, N., and Caporaso, N. (2000). Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J. Natl. Cancer Inst.* 92, 1151–1158. doi: 10.1093/jnci/92.14.1151
- Willcox, B. J., Donlon, T. A., He, Q., Chen, R., Grove, J. S., Yano, K., et al. (2008). FOXO3A genotype is strongly associated with human longevity. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13987–13992. doi: 10.1073/pnas.0801030105
- Yashin, A. I., De Benedictis, G., Vaupel, J. W., Tan, Q., Andreev, K. F., Iachine, I. A., et al. (1999). Genes, demography, and life span: the contribution of demographic data in genetic studies on aging and longevity. *Am. J. Hum. Genet.* 65, 1178–1193. doi: 10.1086/302572
- Yashin, A. I., De Benedictis, G., Vaupel, J. W., Tan, Q., Andreev, K. F., Iachine, I. A., et al. (2000). Genes and longevity: lessons from studies of centenarians. *J. Gerontol. A Biol. Sci. Med. Sci.* 55, B319–B328. doi: 10.1093/gerona/55.7.B319
- Zaitlen, N., Pasaniuc, B., Gur, T., Ziv, E., and Halperin, E. (2010). Leveraging genetic variability across populations for the identification of causal variants. *Am. J. Hum. Genet.* 86, 23–33. doi: 10.1016/j.ajhg.2009.11.016
- Zapata, C. (2013). Linkage disequilibrium measures for fine-scale mapping of disease loci are revisited. *Front. Genet.* 4:228. doi: 10.3389/fgene.2013.00228
- Zaykin, D. V., and Shibata, K. (2008). Genetic flip-flop without an accompanying change in linkage disequilibrium. *Am. J. Hum. Genet.* 82, 794–796; author reply 796–797. doi: 10.1016/j.ajhg.2008.02.001

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

Copyright © 2016 Yashin, Zhbannikov, Arbeeveva, Arbeev, Wu, Akushevich, Yashkin, Kovtun, Kulminski, Stallard, Kulminskaya and Ukrainseva. 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.