AGE (2013) 35:501–518
DOI 10.1007/s11357-011-9374-6

Inter-chromosomal level of genome organization and longevity-related phenotypes in humans

Alexander M. Kulminski · Irina Culminskaya · Anatoli I. Yashin

Received: 7 November 2011 / Accepted: 15 December 2011 / Published online: 27 January 2012

Abstract Studies focusing on unraveling the genetic origin of health span in humans assume that polygenic, aging-related phenotypes are inherited through Mendelian mechanisms of inheritance of individual genes. We use the Framingham Heart Study (FHS) data to examine whether non-Mendelian mechanisms of inheritance can drive linkage of loci on non-homologous chromosomes and whether such mechanisms can be relevant to longevity-related phenotypes. We report on genome-wide inter-chromosomal linkage disequilibrium (LD) and on chromosome-wide intra-chromosomal LD and show that these are real phenomena in the FHS data. Genetic analysis of inheritance in families based on Mendelian segregation reveals that the alleles of single nucleotide polymorphisms (SNPs) in LD at loci on non-homologous chromosomes are inherited as a complex resembling haplotypes of a genetic unit. This result implies that the inter-chromosomal LD is likely caused by non-random assortment of non-homologous chromosomes during meiosis. The risk allele haplotypes can be subject to dominant-negative selection primary through the mechanisms of non-Mendelian inheritance. They can go to extinction within two human generations. The set of SNPs in inter-chromosomal LD (N=68) is nearly threefold enriched, with high significance ($p=1.6\times10^{-9}$), on non-synonymous coding variants (N=28) compared to the entire qualified set of the studied SNPs. Genes for the tightly linked SNPs are involved in fundamental biological processes in an organism. Survival analyses show that the revealed non-genetic linkage is associated with heritable complex phenotype of premature death. Our results suggest the presence of inter-chromosomal level of functional organization in the human genome and highlight a challenging problem of genomics of human health and aging.

Keywords Linkage disequilibrium · Non-Mendelian inheritance · Epistasis · Complex phenotypes · Longevity · Evolution · Quasi-linkage · Dominate-negative selection · Transmission ratio distortion

Introduction

A fundamental problem for humans is unraveling the biology of complex, non-Mendelian diseases, longevity, and, ultimately, aging-related processes. Despite recent discoveries of hundreds of associations between these complex phenotypes and particular genetic variants, the progress in the field is modest (Manolio et al. 2008), explaining only a small fraction of genetic

Electronic supplementary material The online version of this article (doi:10.1007/s11357-011-9374-6) contains supplementary material, which is available to authorized users.

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

susceptibility. The problem is complicated by heritability of complex phenotypes. Apart from rare exceptions, current efforts resulted in explaining less than 5–10% of phenotype heritability (Schork et al. 2009). Such a modest progress is not entirely surprising given the inherent complexity of polygenic phenotypes especially those with post-reproductive manifestation (Kulminski et al. 2011; Martin 2005; Finch and Tanzi 1997; Rockwood et al. 1994) because virtually all of them are a result of a complex aging-sensitive interplay of interacting biological processes regulated by genetic networks (Goh et al. 2007; Tyler et al. 2009).

Current genetic strategies, especially those implemented in large-scale genome-wide association studies (GWAS), are mainly focused on the effect of individual single nucleotide polymorphisms (SNPs) assuming that a polygenic, aging-related trait can be inherited through the classical (Mendelian) mechanism of inheritance of individual genes (Fisher 1918). These studies provide information on the statistical association of a certain tag SNP with a phenotype which might not necessarily be in causative pathway. Given this uncertainty, analysis of linkage disequilibrium (LD) among nearby SNPs becomes an essential tool to more effectively map genes predisposing to complex phenotypes (Lander and Schork 1994; Slatkin 2008).

The term LD refers to statistical associations among genetic variants at different loci. It is typically assumed that useful levels of LD in humans should be within a relatively narrow range of about 3 kb on the same chromosome (Kruglyak 1999). Studies of LD in various human and non-human species show, however, that LD can extend over far more wider ranges on the same chromosome (Sabeti et al. 2007; Dickson et al. 2010), including the entire chromosome (Dyer et al. 2007). More importantly, however, is that modest non-stochastic *inter-chromosomal* LD was observed in plants (Rostoks et al. 2006) and animals (Farnir et al. 2000). Major forces driving LD include natural selection, genetic drift, population structure, inbreeding, inversions, and gene conversion (Slatkin 2008). The most frequent explanation of the observed genome-wide LD in non-human species is their high inbreeding rate (Rostoks et al. 2006; Farnir et al. 2000; Flint-Garcia et al. 2003).

Dickson et al. (2010) also suggested that rare variants might create "synthetic associations" with common variants if they are associated with the same phenotype, i.e., rare variants can occur more often in association (i.e., LD) with one of the alleles of the common variant linked to the same phenotype. Such synthetic associations may span over a 2.5-Mb interval on a chromosome and they were attributed to a stochasticity. Inter-chromosomal LD in humans was observed in the Phase II HapMap but it was attributed to mis-mapping (Frazer et al. 2007).

Epistatic coevolution of alleles at loci on non-homologous chromosomes was recently reported in the human genome (Rohlfs et al. 2010). Inter-chromosomal LD has been recently observed in mouse; it can be associated with a large-scale functional organization of genome in mammals (Graber et al. 2006; Petkov et al. 2005, 2007). Apparently, non-random transmission of parental compound genotypes composed of variants at loci on non-homologous chromosomes to progeny was extensively documented in numerous experiments in non-human species since as early as the 1900s (see, e.g., Sivagnanasundaram et al. (2004), Sapre and Deshpande (1987), Korol et al. (1994), Clegg et al. (1972), and Malinowski (1927) and references therein).

Our recent findings suggest that inter-chromosomal LD can be caused by bio-genetic mechanisms possibly associated with favorable or unfavorable epistatic evolution in humans (Kulminski 2011). In this work, we address two research questions. The first is to examine whether non-Mendelian mechanisms of inheritance can drive (non-physical) linkage of loci on non-homologous chromosomes in the human genome. The second is to elucidate whether such mechanisms can be relevant to heritable longevity-related phenotypes. We use phenotypic and genotyping data available for three successive generations of the Framingham Heart Study (FHS) participants (Govindaraju et al. 2008). The focus of the FHS on large pedigrees followed up for about 60 years provides a unique opportunity to directly address these research questions by using a classical method of genetic analysis based on Mendelian segregation of alleles in families and empirical survival analyses.

Materials and methods

The FHS data and quality control

The FHS data (available from SHARe through the dbGaP) are comprised of three cohorts of successive

generations, i.e., the FHS original cohort (launched in 1948), the FHS Offspring (FHSO) cohort (launched in 1971), and the Third Generation Cohort (launched in 2002). Phenotypic data have been previously described (Govindaraju et al. 2008; Splansky et al. 2007). The released data have information on phenotypes and family structure for 14,174 FHS participants. Of those, there were 1,475 singleton and 12,699 members of 1,537 extended families. Genotyping data are available for 9,274 participants in FHS SHARe (Cupples et al. 2009). Affymetrix performed genotyping using Gene-Chip Human Mapping 500K (250K Sty and 250K Nsp) and the 50K Human Gene Focused arrays. The analyses were primarily focused on genotyping data from the Affymetrix 50K array. After quality control (retaining if: missingness for SNPs and individuals ≤10%, Hardy–Weinberg equilibrium p-values $\geq 10^{-2}$, Mendel errors ≤2%, and minor allele frequency (MAF) ≥2%) and exclusion of non-autosomal chromosomes, we were left with about 38K SNPs in a sample of N=9,016 subjects. Of those, phenotypic information was released for N=8,960 participants. The Affymetrix 500K array was used to ensure that the observed LD is not an array-specific artifact in "Extensive inter-chromosomal and chromosome-wide LD" and "Extensive dominant-negative transmission" and in "Online resource 9". The same constraints as for the Affymetrix 50K chip were applied (resulting in about 38K SNPs).

Pre-selection of SNPs for the LD analyses

We first pre-selected SNPs from the 50K array and then evaluated LD in this pre-selected set. The pre-selection procedure is motivated as follows. Because complex human phenotypes are of polygenic nature (Gibson 2009), we assume that their regulation likely requires coherent action of multiple genes. Accordingly, we hypothesize that enrichment of SNPs for LD can be achieved by selecting SNPs involved in the regulation of a complex phenotype. Representatively, cardiovascular diseases (CVD—diseases of heart and stroke, N=1,751 cases) were considered as a complex phenotype.

We screened all qualified SNPs for their potential associations with CVD (note, these analyses were not intended to ascertain if these associations were true) using logistic regression model implemented in *plink* 1.06 (Purcell et al. 2007) with a SNP as an explanatory variable. Given no prior evidences on the associations, we assumed additive model with minor allele as a risk allele (Omholt et al. 2000). These analyses revealed 69 SNP–CVD associations at the genome-wide significance level (for 38K SNPs) $p < 5 \times 10^{-7}$. These 69 SNPs were pre-selected for LD analysis.

Analysis of LD

LD was evaluated by calculating pair-wise r^2 statistics implemented in *plink* 1.06 (Purcell et al. 2007). Frequencies of alleles for all LD analyses were evaluated using the exact test for founders only.

Survival analysis

This was focused on the FHS/FHSO cohorts only (there was no information on deaths in the third-generation cohort) and was performed using Kaplan–Meier empirical estimates. An individual age, i.e., age at baseline plus chronological time elapsed since the baseline examination, was used as a time variable. Thus, survival curve shows age at death or at the end of follow-up in 2007.

Annotation

SNPs were annotated using WGAViewer (Ge et al. 2008) and ArrayTrack tool (free at http://edkb.fda.gov/webstart/arraytrack/).

Analysis of genetic inheritance

We used a classical method of genetic analysis of inheritance (Morgan 1911) based on Mendelian segregation in families. For this analysis, we used information on all the FHS 1,537 extended families (12,699 family members) irrespective of information on phenotypes/genotypes. First, we classified all extended families into families comprised of parent(s) and offspring given that at least one parent was a participant of the FHS (FamTypes software, free at http://pngu.mgh.harvard.edu/~purcell/famtypes/). This reveals 4,944 families with 15,960 family members (the latter number is larger than the total number of the extended family members, N=12,699, because the same individual can be a child and a parent in the FHS). Then, we excluded families with no genotyped offspring as

irrelevant for the analysis of inheritance. Because certain individuals were either not genotyped at all or genotyping for a particular SNP failed (methods of imputation of genotyping information have essential limitations (de Bakker et al. 2008) and, thus, were disregarded for these analyses), we further selected relevant families (including all family members) if at least one:

(a) Parent in the family had heterozygous genotype of at least one SNP and all SNPs of interest were successfully genotyped (including compliance with the quality control tests)
(b) Child in the family had qualified (after quality control) explicitly genotyped complete heterozygous genotype composed of all SNPs of interest

This procedure ensures selection of all families relevant for the analysis of inheritance in "Genetic analysis based on Mendelian inheritance and segregation". This selection resulted in 210 families (922 members) with carriers of heterozygous genotypes composed of four SNPs (rs1390694, rs7729495, rs11574358, and rs2292664) for the Y set, whereas the G set included only 44 families (200 members). Although analysis of the G set could provide some insights, the results were not included because of the small number of subjects. The analyses were focused on progeny of mating complete heterozygous and recessive (major-allele) homozygous parents (see "The test-cross"). Families in which one parent was complete major-allele homozygote and the other parent was complete heterozygote were aggregated with families in which at least one child was complete heterozygote, one parent was complete major-allele homozygote, and the other parent was not genotyped. Although the focus on mating heterozygous and homozygous parents limits the analyses only to a fraction (232 family members) of the available family data (i.e., 922 family members), this mating variant determines all possible genotypes in progeny with minimal sample size (Morgan 1911), making the analysis of inheritance strikingly convincing in these data.

The test-cross

This is the fundamental genetic tool in analysis of the progeny genotypes. This is a cross of an individual with a genotype of interest to an individual with a homozygous recessive genotype. The progeny genotypes produced by this mating variant reveal all gametes formed by the parental genotypes under the test. Specifically, in the case of no linkage among loci, a monohybrid test-cross gives a genotypic ratio as 1:1. For a dihybrid test-cross, this ratio is 1:1:1:1. For multihybrid test-cross, the genotypic ratio is given by 2^n, where n is the number of heterozygous gene pairs. This implies that in the case of a tetrahybrid test-cross (which we consider), we should have $2^4=16$ different genotypes with equal frequency (1/16=0.0625). This indicates that four pairs of alleles are segregating (assorting) independently. However, in the case of perfect linkage among loci, we have only two types of gametes irrespective of the number of heterozygous gene pairs. This implies that only two parental genotypes are present in progeny, i.e., that offspring inherit only parental genotypes.

Inheritance of compound genotypes was analyzed by evaluating the significance of deviations from theoretical Mendelian expectations using chi-square statistics (Hartl and Clark 2007). In the case when expected counts of genotypes were small (less than five), we used statistics adapted for a small sample size (Hartl and Clark 2007), i.e.:

$$\chi^2 = \sum_i \frac{(|\text{observed}_i - \text{expected}| - 0.5)^2}{\text{expected}},$$

where observed (expected) denotes observed (expected) counts of genotypes and i runs over theoretically expected genotypes.

Pregnancies and live births

The FHS collected data on the number of pregnancies at the baseline examination along with interim history of pregnancies at examinations 2, 4, and 5 (for later examinations, there were no more pregnancies). Because pregnancies in the interim were classified as one or more (with or without toxemia) that represent unnecessary uncertainty, 12 of 58 families with uncertain information on pregnancies were disregarded in arrangement 2 ("Online resource 5"). The FHSO collected information on the number of live births at each examination. For 18 of 50 families, information was missing. Consequently, this part of the analyses was focused on the subset of 32 of 50 families ("Online resource 5", arrangement 4).

Post-zygotic differentiation

The minimal number of the post-zygotic events (N_{PZ}) which is required to make the null hypothesis to be

true (i.e., to make deviation from the expected monohybrid genotypic 1:1 ratio to be insignificant) was evaluated from the following chi-square statistics:

$$\frac{(N_1 - N/2)^2}{N/2} + \frac{(N_2 - N/2)^2}{N/2} = 3.84.$$

Here $N = N_1 + N_2$, $N_1 = N_{11444433} + N_{PZ}/2$, and $N_2 = N_{31243423} + N_{PZ}/2$.

In the most favorable situation (i.e., the least significant deviation from the expected 1:1 ratio; see "Online resource 5", arrangement 4, row "In complete linkage"), we have $N_{11444433} = 48$ and $N_{31243423} = 9$. The minimal number of the post-zygotic events is $N_{PZ} = 339$. In the less favorable case ("Online resource 5", arrangement 3, row "In complete linkage"), we have $N_{11444433} = 57$ and $N_{31243423} = 1$. The minimal number of the post-zygotic events is $N_{PZ} = 758$.

Results

Inter-chromosomal LD

SNPs for the analysis of LD were pre-selected from the Affymetrix 50K array on the basis of their potential association with a complex phenotype (cardiovascular disease; see "Materials and methods"). The pre-selection procedure has resulted in 69 SNPs at loci on all autosomal chromosomes in the human genome ("Online resource 1"). Analysis of LD (see "Materials and methods") among these 69 SNPs shows that 68 of them are in LD with other SNPs at loci on homologous and on non-homologous chromosomes (Fig. 1; "Online resource 1"). The observed inter-chromosomal LD pattern spans the entire human genome. It was found to be a phenotype but not population specific ("Online resource 2").

Inspection of Fig. 1 reveals two sets of highly correlated SNPs (Fig. 2). One set (denoted as Y set) includes four SNPs at loci on chromosomes 1, 5, 8, and 12 ($0.8 \leq r^2 \leq 1.0$). The other set (denoted as G set) includes five SNPs at loci on chromosomes 5, 6, 9, 11, and 17 ($0.77 \leq r^2 \leq 0.87$). SNPs from the Y set show no LD to SNPs in the G set ($r^2 \leq 0.015$). Within and between SNP sets, LD patterns were not altered by sex ("Online resource 3"). Most importantly, the same relationship among SNPs in each set was observed in the parental (the FHS) and the offspring (the FHSO)

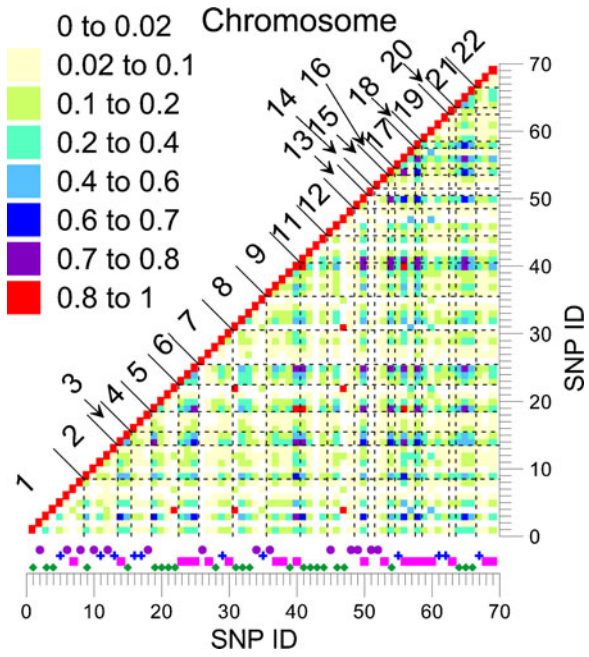

Fig. 1 Map of inter-chromosomal LD among 69 SNPs pre-selected on the basis of association with CVD. SNPs are ordered along the *x*- and *y*-axes by chromosome and base pair number. Pixels report pair-wise LD value (r^2) using colored scale shown in the *inset*. *Colored symbols* just above the *x*-axis show categorized MAF as follows: *green rhombus* ≤5%, *magenta squares* 5–10%, *blue crosses* =10–20%, and *purple-filled dots* >20%

generations ("Online resource 4"), suggesting that these SNPs could be in quasi-linkage (referring to maintenance of the associations among genetic variants located on non-homologous chromosomes in different generations of subjects (Mike 1977)), i.e., they can be inherited through generations as complexes.

Dominant-negative transmission

There are almost no homozygotes for the minor allele for each SNP in the Y and G sets. Analysis of MAF in each FHS generation reveals rapid inter-generational decrease in MAF for these SNPs (Fig. 3) that can be explained by negative selection. Given the lack of minor risk allele homozygotes, this could be a dominant-negative mode. The characteristic feature of this mode is that the risk allele can be eliminated from a population within one generation (Kowles 2001; Strickberger 2000). Although selection affects all SNPs in each set alike, the transmission rates of the minor alleles between the first (MAF_{1st}) and the

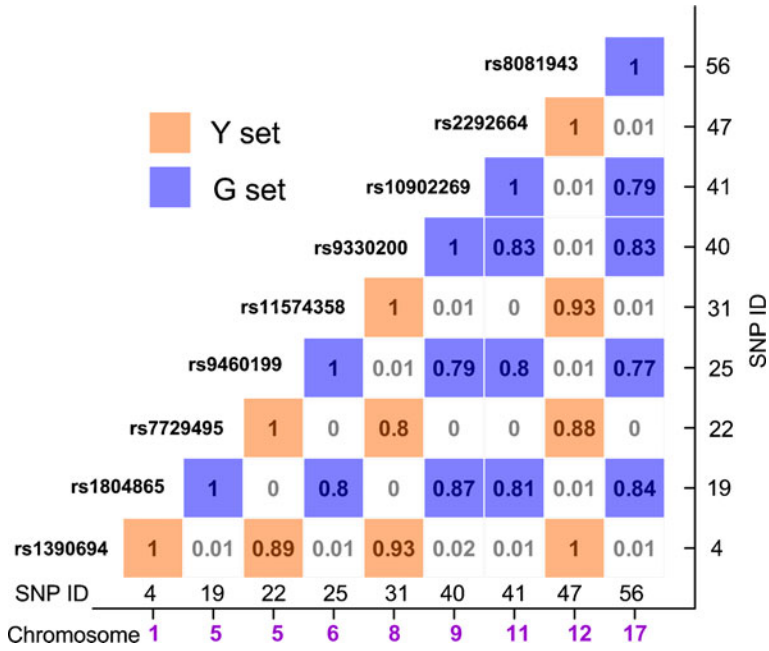

Fig. 2 Map showing perfect inter-chromosomal LD among SNPs in the Y and G sets. SNPs IDs correspond to those in Fig. 1. The *x*-axis shows chromosome for each SNP. *Numbers in squares* represent the magnitude of LD measured by r^2 statistics

second (MAF_{2nd}) generations are qualitatively different. Specifically, the mean transmission rate for the G set is of one order of magnitude smaller (MAF_{2nd}/MAF_{1st}=0.035) than that in the Y set (MAF_{2nd}/MAF_{1st}=0.275). These results along with those shown in "Online resource 4" suggest that the entire complexes of SNPs at loci on non-homologous chromosomes can be involved in dominant-negative selection rather than individual SNPs.

Genetic analysis based on Mendelian inheritance and segregation

Can the observed phenomena (i.e., the inter-chromosomal LD, quasi-linkage, and dominant-negative transmission of SNP complexes) be driven by non-Mendelian mechanisms? The focus of the FHS on large pedigrees from three generations of humans provides a unique opportunity to directly address this question by using a classical method of genetic analysis based on Mendelian inheritance and segregation of alleles in families (see "Materials and methods").

For this analysis, we selected the Y complex of four SNPs and all relevant families bearing the risk alleles. For convenience, alleles of the selected SNPs were coded as 1=A, 2=C, 3=G, and 4=T. Here, we present the results of the analysis of inheritance of genotypes based on the test-cross method (see "Materials and methods").

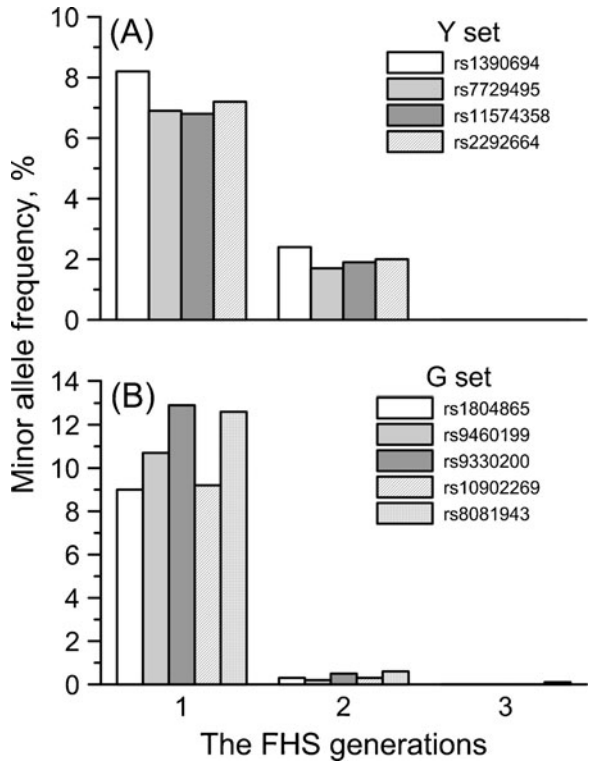

Fig. 3 Frequency of minor allele in each generation of the Framingham Heart Study participants. MAF in this figure was evaluated disregarding family structure, i.e., for all participants of the study. **a** Y set of SNPs, **b** G set of SNPs

These analyses were focused on progeny of mating complete heterozygous (31243423) and recessive (major-allele) homozygous (11444433) parents with genotypes composed of four SNPs of the Y set (see "Materials and methods") located on four non-homologous chromosomes and ordered as rs1390694, rs7729495, rs11574358, and rs2292664. The FHS data allow us to select all offspring of mating heterozygous and major-allele homozygous parents who participated in this study. However, not all offspring were completely genotyped, which resulted in some questionable genotypes. Furthermore, not all offspring could participate in the FHS. Nevertheless, we show below that, despite these limitations, the analyses are entirely feasible and strikingly convincing.

The main results are presented in Table 1. We distinguish explicitly observed genotypes (columns "Genotypes") and those with missing alleles ("Missing"). The latter are used to test if missing information (shown by zeros) can bias the estimates. Numbers in the respective cells show the counts of children in the selected families (see other notations in Table 1 footnotes). To test if incompleteness of genotyped families makes any difference to our conclusions, we also included in the analyses information on the number of live births or pregnancies (see "Pregnancies and live births"). The results are presented in "Online resource 5".

Non-random assortment of non-homologous chromosomes

In the selected mating type (i.e., 31243423×11444433), according to Mendel's second law of independent assortment, we should expect in progeny *16 different genotypes* with about equal frequency, i.e., 1/16= 0.0625 (see "Materials and methods"). Table 1 shows, however, that *virtually all offspring inherit only parental genotypes*. For instance, in 29 families (58 parents) selected by this mating type in the first (parents) and second (offspring) generations, there were 66 children.

Table 1 Genetic analysis of progeny of mating heterozygous (31243423) and recessive (major allele) homozygous (11444433) parents

Mode	Genotypes			Missing			Random		Complex	
	11444433	31243423	Distortion	00444433	31003423	00000000	χ^2	p	χ^2	p
First–second generations, N=66 offspring (second generation) of 58 parents (first generation)										
Actual	33	11	1	12	2	7				
Observed	33	11	1				359.7	2.0×10^{-67}	11.0	9.1×10^{-4}
Under independent assortment	38	12	16				338.4	5.6×10^{-63}	13.5	2.4×10^{-4}
In complete linkage	50	15	1				568.3	1.9×10^{-111}	18.8	1.4×10^{-5}
Second–third generations, N=58 offspring (third generation) of 50 parents (second generation)										
Actual	50	0	0	4	0	4				
Observed	50	0	0				721.3	5.4×10^{-144}	50.0	1.5×10^{-12}
Under independent assortment	52	0	6				735.4	5.2×10^{-147}	52.0	5.6×10^{-13}
In complete linkage	57	1	0				810.2	5.6×10^{-163}	54.1	1.9×10^{-13}

Alleles are denoted using conventional notations as 1=A, 2=C, 3=G, and 4=T and are ordered according to the order of SNPs from the Y set, i.e., rs1390694, rs7729495, rs11574358, and rs2292664

Genotypes: parental (i.e., 11444433 and 31243423) and non-parental or "recombinant" (e.g., 11244433) genotypes (column "distortion") explicitly observed (i.e., no missing data) in the data, *Missing*: genotypes with missing alleles (denoted by zeros), *Random*: the results of chi-square test (see "The test-cross") assuming independent assortment of alleles in the case of four SNPs from the Y set (i.e., in the case of a tetrahybrid test-cross), *Complex*: the results of chi-square test assuming Mendelian monohybrid segregation of "haplotypes" of four SNPs from the Y set (i.e., when alleles of these SNPs are in complete linkage; only parental genotypes (i.e., 11444433 and 31243423) were considered for these analyses), *Actual* counts of *all* compound genotypes observed in offspring in the selected families, *Observed* counts of the explicitly observed genotypes in progeny and the results of chi-square tests with these genotypes, *Under independent assortment* counts of genotypes in progeny with missing alleles imputed based on independent assortment of alleles (detailed in "Online resource 6") and the results of chi-square tests with these genotypes, *In complete linkage* counts of genotypes in progeny with missing alleles imputed based on independent monohybrid segregation (detailed in "Online resource 6") and the results of chi-square tests with these genotypes

Among them, 44 children were explicitly genotyped with parental genotypes (33 homozygotes (11444433) and 11 heterozygotes (31243423)) and only one child was with "recombinant" genotype (11244433) of 14 possible non-parental combinations. Deviation from random assortment in this case is so significant (χ^2=359.7, p=2.0×10^{-67}) that neither viability selection nor incomplete information on non-genotyped children can explain the lack of similar numbers of non-parental genotypes.

For instance, let us assume that all non-explicitly observed genotypes (Table 1, columns "Missing") were *theoretically expected genotypes resulting from an independent assortment of non-homologous chromosomes* ("Online resource 6"). However, even such best-case-scenario imputation of recombinant genotypes makes no difference to our conclusions. For instance, in the case when parents are from the first generation and the offspring are from the second one, we have 16 recombinant genotypes but deviation from random assortment is still tremendously significant (χ^2=338.4, p=5.6×10^{-63}) (Table 1, column "Distortion"). The conclusion is the same if all pregnancies/live births are included ("Online resource 5").

Therefore, by analogy with genetic linkage, we have to assume that four SNPs in the Y set do not assort independently. Unlike classical genetic linkage (Morgan 1911), however, we document non-random assortment for SNPs at loci *on non-homologous chromosomes*. This phenomenon is known as quasi-linkage (Mike 1977) and it can be genetic rationale for the observed inter-chromosomal LD.

Multi-locus inter-chromosomal complex

The observed strong non-random assortment implies that alternative alleles of four SNPs in the Y set tend to be transmitted through generations together, resembling haplotypes of a genetic unit. This can be tested assuming Mendelian monohybrid segregation of haplotypes of the hypothesized genetic unit rather than independent assortment of four alleles, i.e., that genotypic ratio in progeny of mating heterozygous and homozygous parents is closer to monohybrid 1:1 (as in the case of perfect linkage) ratio than to random expectation for alleles in the tetrahybrid test-cross (see "Materials and methods").

Column "Complex" in Table 1 shows that, for the explicitly observed genotypes (Table 1, rows "Observed") and the genotypes with imputed information (Table 1, rows "Under independent assortment"), deviation from the 1:1 ratio is significant. The segregation ratio is, however, strikingly closer to the 1:1 ratio than to the random expectation for four alleles (compare statistics in Table 1, columns "Random" and "Complex"). This result implies that the significance of the deviation from the expected 1:1 ratio is largely attributed to disproportion among frequencies of parental genotypes in progeny rather than to random assortment of non-homologous chromosomes. This means that the inter-chromosomal complex of quasi-linked SNPs resembles a genetic unit. By similarity with inheritance of linked genes (Morgan 1911), the observed distortions should be attributed to some sort of "recombination" events (see "Discussion and conclusions" for potential mechanism).

To show that our conclusion on a genetic unit, which we call here as a multi-locus inter-chromosomal (MLIC) complex, is not altered by missing information, we have imputed missing genotypes (Table 1, column "Missing") assuming that they are theoretically expected *parental genotypes* resulting from random segregation of parental haplotypes of a MLIC complex according to Mendel's first law ("Online resource 6"). Statistics in this case (Table 1, rows "In complete linkage" and columns "Complex") still resemble those in rows "Observed" and "Under independent assortment" (Table 1, columns "Complex"). This result ensures that our conclusion on a genetic unit is not altered. The same is observed if all pregnancies or live births are included ("Online resource 5").

Transmission ratio distortion

If MLIC complex can resemble a genetic unit, a significant deviation from the expected 1:1 ratio should be driven by certain mechanisms. Theoretically, this deviation could be the result of disproportional viability selection, post-zygotic differentiation, and/or non-Mendelian differentiation at the pre-zygotic stage (e.g., meiotic drive, gametic selection (Pardo-Manuel de Villena and Sapienza 2001)).

Viability selection alone cannot explain the observed deviation because it is significant even if we include all live births ("Online resource 5", arrangement 4). To elucidate if this deviation can be explained by post-zygotic differentiation, we estimated the minimal number of post-zygotic deaths which are required in the second–third generations to make the deviation

from the expected 1:1 ratio insignificant. Our analysis (see "Post-zygotic differentiation" in "Materials and methods") shows that there should be at least 339 post-zygotic deaths even in the most favorable (i.e., the least significant) case (i.e., "Online resource 5", arrangement 4, row "In complete linkage"). This number means that each mother should experience at least ten (i.e., 339/32) more *conceptions* in addition to those which resulted in live births, which is apparently unrealistic in developed societies in the twentieth century.

These results imply that a major selection mechanism working against *harmful* minor-allele (recall that minor allele can favor CVD; see also "Survival analysis") haplotype of a MLIC complex is largely relevant to the non-Mendelian differentiation at the pre-zygotic stage.

Biological role

Annotation of the tightly linked SNPs in the Y (four SNPs) and G (five SNPs) complexes highlights the remarkable fact that about 67% of them (i.e., six SNPs; three SNPs in each complex) are non-synonymous coding polymorphisms and eight of nine SNPs are within gene regions (Table 2). Clustering of non-synonymous coding SNPs in these complexes is contrasted by a modest proportion of non-synonymous coding SNPs in the entire qualified set of about 38K SNPs (from the Affymetrix 50K array; see "Materials and methods"), i.e., about 15%, and by their enriched proportion in the set of 68 SNPs in inter-chromosomal LD ("Online resource 1"), i.e., 41.2% (28 of 68 SNPs). This nearly threefold enrichment of non-synonymous coding SNPs, i.e., SNPs which can alter amino acid sequence of proteins, is highly significant with two-sided p-value for the difference in these two proportions (i.e., 15% and 41.2%) $p=1.6\times10^{-9}$. This striking result offset the chance nature of stochastic clustering of the non-synonymous coding SNPs in the pre-selected set because this probability of such event is far below the 5% (i.e., $p=0.05$) level for significance in this analysis. Genes for SNPs from the Y and G complexes (Table 2) and other selected SNPs showing inter-chromosomal LD (Kulminski and Culminskaya 2012) have been extensively studied in relation to multiple biological processes playing a fundamental role in functioning of an organism and to major human diseases.

Table 2 Summary for genes for SNPs that constitute the Y and G multi-locus inter-chromosomal complexes

Set	SNP	Gene	Gene name	Cytolocation	Selected biological processes
Y	*rs1390694*[a]	*FGGY*	Carbohydrate kinase domain containing	1p32.1	Carbohydrate metabolism, neuron homeostasis
	rs7729495[b]	*GPR98*	G protein-coupled receptor 98	5q13	G-protein coupled receptor protein signaling pathway, cell–cell adhesion, nervous system development, response to stimulus
	rs11574358[b]	*WRN*	Werner syndrome, RecQ helicase-like	8p12-p11.2	DNA replication, recombination, and synthesis; aging; regulation of apoptosis; response to UV-C, oxidative stress, and DNA damage stimulus
	rs2292664[b]	*RIMBP2*	RIMS binding protein 2	12q24.33	Synaptic transmission
G	rs1804865[b]	*TRIP13*	Thyroid hormone receptor interactor 13	5p15.33	Transcription from RNA polymerase II promoter, double-strand break repair, meiosis I
	rs9460199[c]	*PSMB1*	Proteasome (prosome, macropain) subunit, beta type, 1	6q27	APC-dependent proteasomal ubiquitin-dependent protein catabolic process, regulation of ubiquitin–protein ligase activity involved in mitotic cell cycle, interspecies interaction between organisms
	rs9330200[b]	*TUBB2C*	Tubulin, beta 2C	9q34	Natural killer cell-mediated cytotoxicity, cellular component movement, protein polymerization
	rs10902269[b]	*MUC6*	Mucin 6, oligomeric mucus/gel-forming	11p15.5	Maintenance of gastrointestinal epithelium
	rs8081943[a]	*RAI1*	Retinoic acid induced 1	17p11.2	Regulation of transcription from RNA polymerase II promoter, skeletal system development, negative regulation of multicellular organism growth

Information is assessed at http://www.ncbi.nlm.nih.gov and http://www.genecards.org, March 10, 2011

[a] Intronic variant

[b] Non-synonymous coding

[c] Intergenic; information is provided for the nearest protein-coding gene

Extensive inter-chromosomal and chromosome-wide LD

To elucidate if apparently coupled inheritance of the quasi-linked alleles can be just the result of mis-mapping, we evaluated LD within ±1-Mb region of each of nine SNPs in both complexes to all qualified SNPs in the Affymetrix 500K and 50K arrays (see "Materials and methods"). If mis-mapped, these quasi-linked SNPs should not show LD to SNPs at homologous chromosomes (Frazer et al. 2007). This analysis shows that each SNP in the Y and G complexes can be in LD to other SNPs within ±1-Mb region ("Online resource 7").

Furthermore, we further tested if these nine quasi-linked SNPs from the Y and G complexes could exhibit LD not only to nearby but also to distant SNPs on homologous chromosomes as well as to SNPs on non-homologous chromosomes. For these analyses, we again used a combined set of SNPs from the 500K and 50K arrays. These analyses revealed 30,766 SNP pairs with $r^2 \geq 0.1$ (this cutoff is fivefolds larger than that for stochastic noise; see Fig. S1 in "Online resource 2"). Figure 4a shows an example of extensive chromosome-wide LD among SNPs genotyped using *independent* Affymetrix chips, i.e., the 500K (250K Sty and the 250K Nsp) and the 50K chips. SNPs from the G complex show more extensive LD than those from the Y complex. The density pattern of the quasi-linked SNPs is remarkably similar within each complex and remarkably dissimilar between complexes. Figure 4b shows for comparison that each of these nine SNPs is also in LD to SNPs on non-homologous chromosomes. The MLIC-complex-specific density patterns of the quasi-linked SNPs are apparently similar to those in Fig. 4a. The quasi-linked SNPs tend to spread over the entire chromosomes irrespective of the LD strength (see "Online resource 8" for $r^2 \geq 0.3$). Interestingly, unlike SNPs from the Y complex, four of five SNPs from the G complex are at the chromosome edges.

Extensive dominant-negative transmission

The risk alleles of the quasi-linked SNPs shown in Fig. 4 are transmitted through generations in a coherent fashion similarly as in Fig. 3 irrespective if they are on homologous or non-homologous chromosomes. Figure 5a documents this transmission pattern for SNPs showing strong intra- and inter- chromosomal LD ($r^2 \geq 0.5$) to rs1804865 from the G complex. To ascertain that this dynamics is not a general property of SNPs in the FHS, we selected SNPs showing no LD ($r^2 \leq 10^{-3}$) to the tag SNPs within ±1-Mb region. Figure 5b shows the lack of dynamics of MAF in the representative example for SNPs not showing LD to rs1804865 from the G complex. The results are similar for SNPs from the Y complex.

Phenotypic structure

The analyses based on the Mendelian segregation provide compelling evidences that the observed inter-chromosomal LD, quasi-linkage, and dominant-negative selection can be caused by non-Mendelian mechanisms of inheritance. Here we explicitly show that these phenomena are associated with heritable longevity-related *phenotypes*.

Genotyping phases

Genotyping of the FHS SHARe sample (N=9,172) was accomplished in four phases ("Online resource 9"). Virtually all minor alleles of the quasi-linked SNPs from the Y and G complexes were genotyped in phase 3 (Ph3) and phase 4 (Ph4), respectively. No apparent clustering of minor alleles or potentially problematic DNA samples on genotyping plates was identified (Pluzhnikov et al. 2010). Importantly, SNPs showing LD to SNPs from both (i.e., the Y and G) complexes were genotyped predominantly in Ph3 and Ph4 ("Online resource 9"). These results imply that massive bias in the allele frequencies between phases 1 and 2 (Ph1&Ph2) and Ph3&4 can only be explained if both technologies of whole genome amplification (Ph3) and genotyping of genomic DNA extracted from blood frozen for a long time (about 20 years; Ph4) failed. However, numerous studies of fidelity of WGA methods using DNA extracted from frozen (for about 10 to 20 years) blood, plasma, and serum did show excellent concordance rates and no systematic biases among WGA and genomic DNA samples (see Xing et al. (2008) and Croft et al. (2008) and references therein). Therefore, it is unlikely that clustering of the risk (minor) alleles in Ph3 and Ph4 was due to failed technologies.

Longevity-related phenotypes

An alternative explanation of the observed differences in the allelic frequencies is specific of individuals but

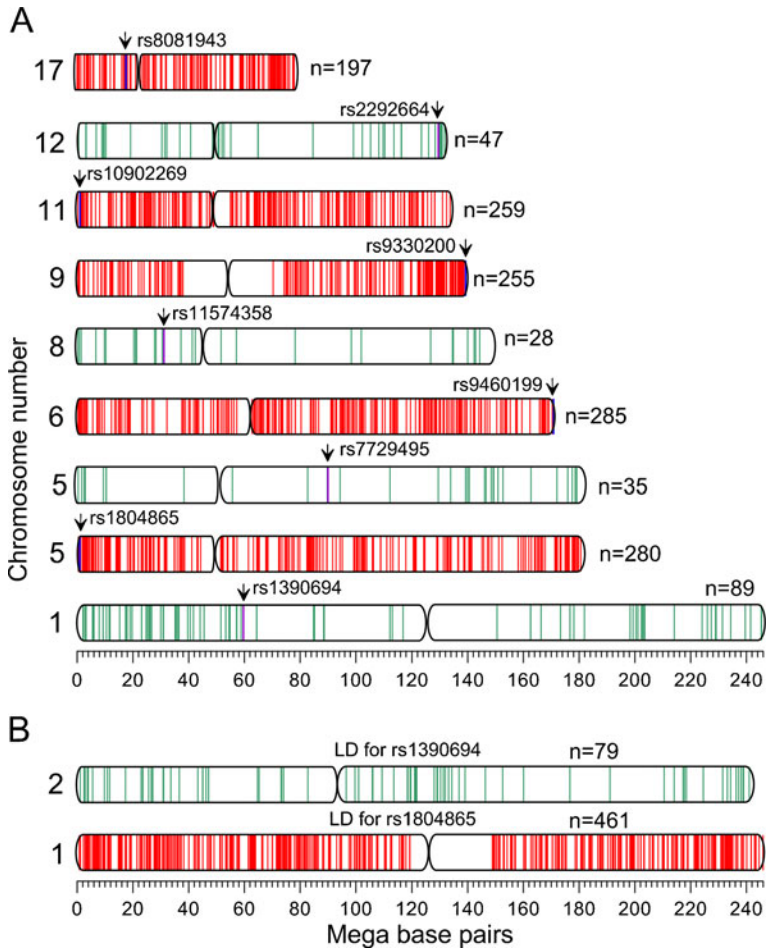

Fig. 4 Maps of extensive chromosome-wide LD. The maps show chromosome-wide LD with $r^2 \geq 0.1$ for nine SNPs from the Y (*green*) and G (*red*) complexes. **a** Map of intra-chromosomal LD for each SNP. **b** Map of LD on chromosome 2 for rs1390694 (Y complex) located on chromosome 1 and map of LD on chromosome 1 for rs1804865 (G complex) located on chromosome 5. *Numbers on the right* denote the number of SNPs showing LD to SNPs from the Y and G complexes depicted by *arrows*

not the DNA samples. Virtually all individuals genotyped in Ph4 and majority of those in Ph3 were deceased, whereas majority of those from Ph1&2 were living as of 2007. The difference in survival of these individuals can be conveniently characterized by empirical Kaplan–Meier age patterns (see "Materials and methods"). Figure 6a, b show that individuals from Ph3&4 enjoyed a tremendously shorter life in each cohort compared to their age peers from Ph1&2 both on individual level (shown by curves) and on average (see life expectancy [LE] estimates). This is not explained by the age-cohort-specific differences in longevity ("Online resource 10"). Therefore, Ph3&4 represent naturally selected populations of individuals who died prematurely, whereas Ph1&2 for the FHS/FHSO represent the population of individuals who outlived their age peers. In other words, the Ph3&4 and Ph1&2 populations have qualitatively different longevity-related phenotypes.

Inheritance of the longevity-related phenotypes

A long life is a heritable phenotype (Christensen et al. 2006; Martin et al. 2007; Bergman et al. 2007) which is typically associated with good health (Evert et al. 2003; Barzilai et al. 2003; Willcox et al. 2008). Contrarily, the phenotype of premature death is associated with diseases ("Online resource 1") which themselves are heritable traits. To explicitly show that this phenotype (i.e., Ph3&4) is a heritable trait, we followed the strategy presented in Perls et al. (2002) and evaluated the survival of all family members, who participated in the FHS, in which at least one child was in Ph3 or Ph4. Figure 6c, d show that the *parents and their children* who are in Ph3 or Ph4 experience survival disadvantage throughout their lives relative to parents and their children who are in Ph1 or Ph2 *and* to the general Ph1&2 population of the FHS and the FHSO participants. This result implies that the

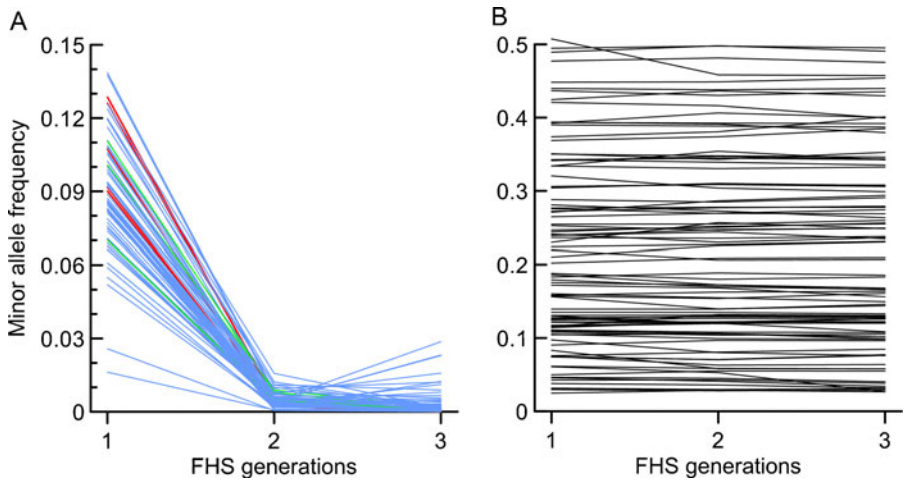

Fig. 5 Frequency of minor allele in each generation of the Framingham Heart Study participants. MAF in this figure for SNPs from the Affymetrix 500K and 50K arrays was evaluated disregarding family structure. **a** MAF of 68 SNPs showing intra- (*green*) and inter-chromosomal (*blue*) LD with $r^2 \geq 0.5$ to rs1804865 (G complex) located on chromosome 5. *Red color* denotes five SNPs from the G complex. **b** MAF of 94 SNPs showing no LD ($r^2 \leq 10^{-3}$) to rs1804865 (G complex) within ±1 Mb region on chromosome 5

phenotype of premature death is a heritable trait (Perls et al. 2002).

Figure 6d shows a striking result that non-genotyped children experience the same survival disadvantage as those who were genotyped in Ph3&4. Furthermore, non-genotyped parents (Fig. 6c) experienced the same survival disadvantage as their non-genotyped and genotyped in Ph3&4 children. These results suggest that the non-genotyped parents and their children in these families were carriers of the disadvantageous genotypes.

Therefore, because individuals from Ph1&2 and Ph3&4 have qualitatively different heritable phenotypes, they inevitably have to have different genetic profiles. Furthermore, because the Ph4 population has lived significantly shorter lives relative to the Ph3 population (see LE estimates in Fig. 6a, b), they have different phenotypes of premature deaths and, respectively, their genetic profiles have to be different. However, because both of these populations lived significantly shorter lives than the Ph1&2 populations, they should also have overlapping genetic component. This is in agreement with our observation of partly overlapping genetic profiles associated with the G and Y complexes (Fig. 1).

Discussion and conclusions

The results of our analyses complement recent findings in Kulminski (2011) and provide compelling evidence that the observed phenomena of inter-chromosomal and chromosome-wide intra-chromosomal LD, quasi-linkage, and transmission ratio distortion of "haplotypes" of the quasi-linked SNPs are real in the FHS data and that they should be based on intrinsic biological and genetic mechanisms relevant to complex, non-Mendelian phenotypes.

In general, according to Mendel's Second Law, non-homologous chromosomes assort (segregate) independently. Accordingly, it is typically believed that alleles at loci on non-homologous chromosomes assort randomly. Nevertheless, *non-random* segregation of chromosomes during meiosis was experimentally observed. Ironically, the first cytological proof of this phenomenon was provided by F. Payne (1912, 1916) even before the first cytological proof of random segregation of chromosomes was credited to E. Carothers (1913, 1917). Since that time, non-random segregation of chromosomes during meiosis was observed in various non-human species (Metz et al. 1926; Metz 1926; Schrader 1921, 1923; Camenzind and Nicklas 1968; White 1973; Kubai and Wise 1981; Wang et al. 2010; Sillers et al. 1983; Wise et al. 1984). Due to difficulties in distinguishing homologous chromosomes, these experiments were limited to sex chromosomes or co-segregation of sex chromosomes and autosomes with apparently different size of homologues. Wang et al. (2010) suggest that non-random segregation of chromosomes during meiosis can be attributed to the chromosome size differences.

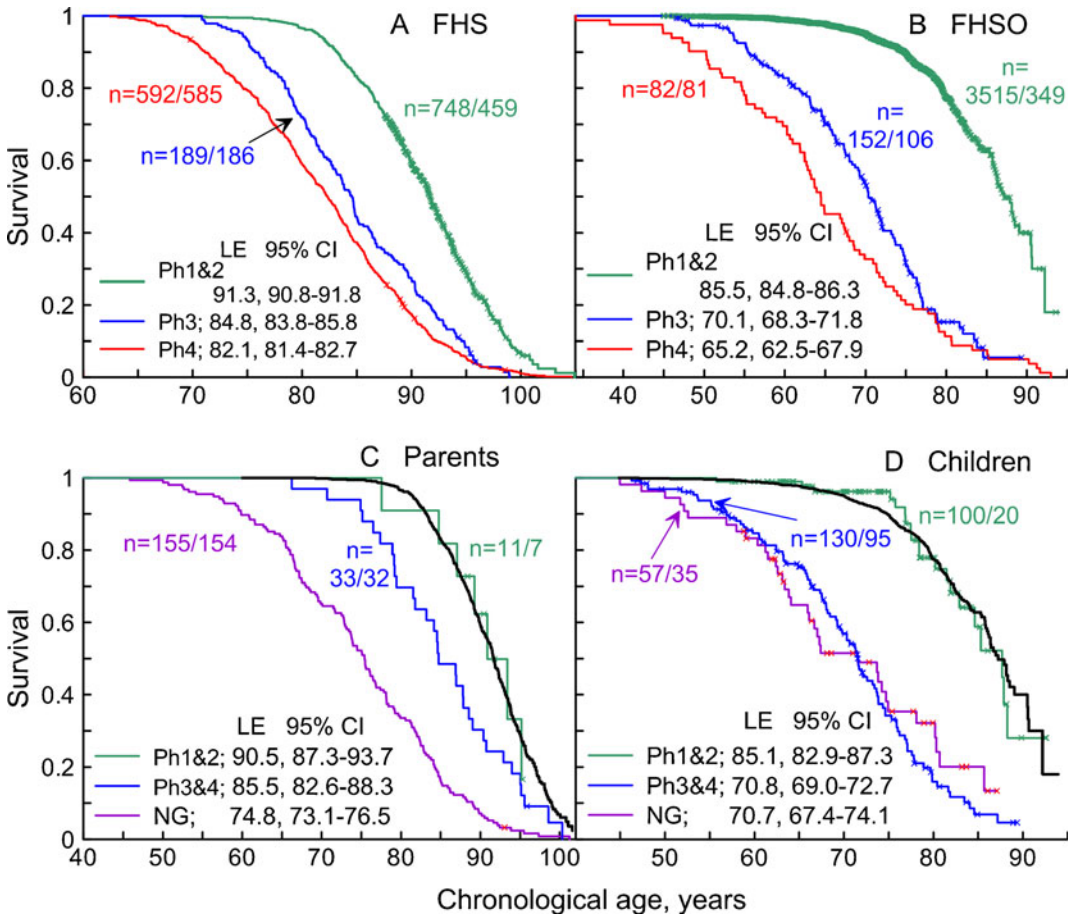

Fig. 6 Survival of genotyped and non-genotyped FHS/FHSO participants. **a**, **b** Survival of genotyped individuals selected for four genotyping phases from the FHS and FHSO cohorts. **c**, **d** Survival of all members of the families in which at least one child was selected for phase 3 (Ph3) or phase 4 (Ph4) (combined due to the small sample size of Ph4). Kaplan–Meier survival curves show the chronological age at death (*line*) or the end of follow-up in 2007 (*crosses*). *Black curves* in **c** and **d** duplicate those in **a** and **b** for individuals genotyped in phases 1 and 2 for the sake of comparison. Sample size (*n*) is for all/dead individuals. *LE* life expectancy, *CI* confidence interval, *NG* not genotyped individuals

Numerous exceptions to Mendel's Second Law were also revealed in the analysis of progeny in backcrosses in a variety of organisms (see, e.g., Sivagnanasundaram et al. (2004), Sapre and Deshpande (1987), Korol et al. (1994), Clegg et al. (1972), Malinowski (1927), and references therein). These exceptions either resemble genetic linkage resulting in an excess of parental genotypes in progeny (called quasi-linkage) or document excess of non-parental genotypes in progeny (called super-recombination) (Robinson 1971). These phenomena were extensively studied in relation to population structure (Malinowski 1927; Mike 1977) and inheritance of phenotypes (Michie 1953) and oncogenic viruses (Boyse 1977). A key feature of these exceptions is that *chromosomes of phenotypically discordant parents* (particularly from different inbred strains) can sense each other, i.e., every chromosome should have some sort of affinity. Consequently, the same-parent chromosomes can go to the same pole during meiosis as a complex (Malinowski 1927). The mechanism of affinity (i.e., when the same-parent non-homologous chromosomes tend to "stick together") is, however, unclear (Boyse 1977; Michie 1953; Mike 1977).

The concept of affinity is relevant to crosses of the parents with qualitatively different heritable phenotypes. This is the case in the FHS (see "Phenotypic structure"). Accordingly, the observed quasi-linkage can be due to some sort of affinity when *non-homologous chromosomes* get assorted non-randomly during meiosis, resulting in an excess of parental

genotypes in progeny (Table 1). By definition, the mechanism of affinity should be relevant to the entire genome that implies an inter-chromosomal level of genome organization. A "side effect" of such a mechanism is that complexes of quasi-linked SNPs should resemble a genetic unit. Consequently, such complexes can complement the gene as a unit of inheritance.

An apparent selection (Figs. 3 and 5; Table 1; "Online resource 5") against the harmful dominant alleles of the quasi-linked SNPs working within two generations in a modern environment is neither the result of viability selection nor the result of post-zygotic differentiation alone. Our analysis (see "Transmission ratio distortion") suggests that the observed extinctive transmission of the risk allele haplotypes through generations should be largely relevant to the non-Mendelian mechanisms at the pre-zygotic stage. Thus, complexes of the quasi-linked SNPs can complement the gene as a unit of evolution that provides substantial rationale for prior inferences (Clegg et al. 1972; Slatkin 1972, 2008).

What can the mechanism of quasi-linkage be? Theoretically, this can be either physical or functional linkage. Prior studies documented physical linkage of mitotic non-homologous chromosomes with nucleolar organizing regions, i.e., chromosomes 13, 14, 15, 21, and 22 (Bobrow and Heritage 1980). This is apparently not the relevant mechanism here (e.g., because neither one of the SNPs of interest is at loci on those chromosomes). Recent studies discovered, however, the phenomenon of *functional linkage* when transcription of genes located on one chromosome may be controlled by a regulatory element located on another non-homologous chromosome (Williams et al. 2010). The mechanism of functional linkage can explain the fact that one allele can "sense" the other allele (Spilianakis and Flavell 2006), i.e., what is known as affinity. Recently revealed large-scale functional organization in mammals extending to different domains on non-homologous chromosomes was specifically attributed to co-adaptation of *functionally linked* genes participated in common biological processes (Graber et al. 2006; Petkov et al. 2005, 2007).

Potential functional mechanisms driving non-random segregation of chromosomes can be, for instance, associated with functional asymmetry of the meiotic spindle poles (Pardo-Manuel de Villena and Sapienza 2001) or interaction of spindle fibers from different chromosomes (Camenzind and Nicklas 1968) through, in particular, specific modifications at the kinetochores (Kubai and Wise 1981; Wise et al. 1984) and connections between the kinetochores of different chromosomes (Kubai and Wise 1981). The latter might be the result of functional connection of chromosomes (Wise et al. 1984).

A key factor in co-inheritance mechanisms is functional connections (e.g., through common biological process; see "Biological function") which are likely environment sensitive and, thus, are involved in evolutionary selection (Graber et al. 2006; Petkov et al. 2005; Rohlfs et al. 2010). Therefore, these mechanisms can be modulated by physiological processes responsible for organismal functioning in a given environment. This is in line with the remarkable similarity of transmission rates of alleles of SNPs within MLIC complexes and their dissimilarity between complexes (see Figs. 3 and 5). Consequently, mechanisms of functional connections can potentially be relevant to a generation of LD and extinction of the risk allele haplotypes of the MLIC complexes.

For instance, genome-wide LD could be generated by epistatic selection (Phillips 2008; Slatkin 2008; Hurst 2009; Graber et al. 2006; Petkov et al. 2005; Rohlfs et al. 2010), viability selection, non-Mendelian differentiation (Pardo-Manuel de Villena and Sapienza 2001), etc., in a given environment. Although the exact mode which generated genome-wide LD is unknown, the loss of the risk alleles of quasi-linked SNPs likely occurs through non-Mendelian mechanisms. Change in the environment and human behaviors that occurred during recent centuries may affect organismal physiology fitted to a markedly distinct environment (Kuningas et al. 2008), making formerly beneficial alleles of the quasi-linked SNPs to be highly disadvantageous in the new environment. This might result in super-fast extinction of the risk alleles through primarily non-Mendelian mechanisms.

Whatever mechanisms of the quasi-linkage and non-random segregation of chromosomes in humans are, these phenomena have far-reaching implications because they suggest that human post-reproductive health and lifespan can be controlled by genome-wide networks of allelic variants which can work in a highly coherent fashion. These findings highlight a challenging problem of genomics of human health and aging that is in line with the real complexity of traits with post-reproductive manifestation (Kirkwood 2011; Goh et al. 2007; Kirkwood et al. 2011). Addressing this

problem may require thorough methods well beyond those offered by standardized guidelines for traditional GWAS. The results suggest that future progress in unraveling genetic origin of human health and aging will require the collaborative multidisciplinary and interdisciplinary input of specialists from diverse disciplines, including bio-genetics, evolutionary biology, epidemiology, bio-demography, and aging.

Acknowledgments AMK contributed to the study conception, design, statistical analysis, interpretation of the results, and writing the manuscript. IC contributed to the study design, conception, genetic analyses of inheritance, analysis of biological role, interpretation of the results, and writing the manuscript. AIY contributed to discussion and interpretation of the results and drafting the manuscript.

The research reported in this paper was supported, in part, by Award Number R01AG030612 from the National Institute on Aging. The FHS and the Framingham SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University. The Framingham SHARe data used for the analyses described in this manuscript were obtained through dbGaP (accession numbers phs000007.v7.p4 and phs000007.v14.p5). This manuscript was not prepared in collaboration with investigators of the FHS and does not necessarily reflect the opinions or views of the FHS, Boston University, or the NHLBI. We thank G. Martin for discussion of the results and drawing our attention to yeast-, worm-, and plant-based models of human diseases.

References

Barzilai N, Atzmon G, Schechter C, Schaefer EJ, Cupples AL, Lipton R, Cheng S, Shuldiner AR (2003) Unique lipoprotein phenotype and genotype associated with exceptional longevity. JAMA 290(15):2030–2040

Bergman A, Atzmon G, Ye K, MacCarthy T, Barzilai N (2007) Buffering mechanisms in aging: a systems approach toward uncovering the genetic component of aging. PLoS Comput Biol 3(8):e170

Bobrow M, Heritage J (1980) Nonrandom segregation of nucleolar organizing chromosomes at mitosis? Nature 288 (5786):79–81

Boyse EA (1977) The Gx system in relation to C-type viruses and heredity. Immunol Rev 33:125–145

Camenzind R, Nicklas RB (1968) The non-random chromosome segregation in spermatocytes of *Gryllotalpa hexadactyla.* A micromanipulation analysis. Chromosoma 24 (3):324–335

Carothers EE (1913) The Mendelian ratio in relation to certain orthopteran chromosomes. J Morphol 24:487–511

Carothers EE (1917) The segregation and recombination of homologous chromosomes as found in two genera of Acridae (Orthoptera). J Morphol 28:445–520

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

Clegg MT, Allard RW, Kahler AL (1972) Is the gene the unit of selection? Evidence from two experimental plant populations. Proc Natl Acad Sci U S A 69(9):2474–2478

Croft DT Jr, Jordan RM, Patney HL, Shriver CD, Vernalis MN, Orchard TJ, Ellsworth DL (2008) Performance of whole-genome amplified DNA isolated from serum and plasma on high-density single nucleotide polymorphism arrays. J Mol Diagn 10(3):249–257. doi:10.2353/jmoldx.2008.070155

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 Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF (2008) Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet 17(R2):R122–R128. doi:10.1093/hmg/ddn288

Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB (2010) Rare variants create synthetic genome-wide associations. PLoS Biol 8(1):e1000294. doi:10.1371/journal.pbio.1000294

Dyer KA, Charlesworth B, Jaenike J (2007) Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. Proc Natl Acad Sci U S A 104(5):1587–1592. doi:10.1073/pnas.0605578104

Evert J, Lawler E, Bogan H, Perls T (2003) Morbidity profiles of centenarians: survivors, delayers, and escapers. J Gerontol A Biol Sci Med Sci 58(3):232–237

Farnir F, Coppieters W, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M (2000) Extensive genome-wide linkage disequilibrium in cattle. Genome Res 10(2):220–227

Finch CE, Tanzi RE (1997) Genetics of aging. Science 278 (5337):407–411

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb 52:399–433

Flint-Garcia SA, Thornsberry JM, Buckler ESt (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374. doi:10.1146/annurev.arplant.54.031902.134907

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T,

Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Yakub I, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449(7164):851–861. doi:10.1038/nature06258

Ge D, Zhang K, Need AC, Martin O, Fellay J, Urban TJ, Telenti A, Goldstein DB (2008) WGAViewer: software for genomic annotation of whole genome association studies. Genome Res 18(4):640–643. doi:10.1101/gr.071571.107

Gibson G (2009) Decanalization and the origin of complex disease. Nat Rev Genet 10(2):134–140. doi:10.1038/nrg2502

Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL (2007) The human disease network. Proc Natl Acad Sci U S A 104(21):8685–8690. doi:10.1073/pnas.0701361104

Govindaraju DR, Cupples LA, Kannel WB, O'Donnell CJ, Atwood LD, D'Agostino RB Sr, Fox CS, Larson M, Levy D, Murabito J, Vasan RS, Splansky GL, Wolf PA, Benjamin EJ (2008) Genetics of the Framingham Heart Study population. Adv Genet 62:33–65. doi:10.1016/S0065-2660(08)00602-0

Graber JH, Churchill GA, Dipetrillo KJ, King BL, Petkov PM, Paigen K (2006) Patterns and mechanisms of genome organization in the mouse. J Exp Zool A Comp Exp Biol 305(9):683–688. doi:10.1002/jez.a.322

Hartl DL, Clark AG (2007) Principles of population genetics, 4th edn. Sinauer Associates, Sunderland, Mass

Hurst LD (2009) Fundamental concepts in genetics: genetics and the understanding of selection. Nat Rev Genet 10 (2):83–93. doi:10.1038/nrg2506

Kirkwood TB (2011) Systems biology of ageing and longevity. Philos Trans R Soc Lond B Biol Sci 366(1561):64–70. doi:10.1098/rstb.2010.0275

Kirkwood TB, Cordell HJ, Finch CE (2011) Speed-bumps ahead for the genetics of later-life diseases. Trends Genet 27(10):387–388. doi:10.1016/j.tig.2011.07.001

Korol AB, Preygel IA, Preygel SI (1994) Recombination variability and evolution: algorithms of estimation and population-genetic models. Chapman & Hall, London

Kowles RV (2001) Solving problems in genetics. Springer, New York

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22(2):139–144. doi:10.1038/9642

Kubai DF, Wise D (1981) Nonrandom chromosome segregation in *Neocurtilla* (*Gryllotalpa*) *hexadactyla*: an ultrastructural study. J Cell Biol 88(2):281–293

Kulminski AM (2011) Complex phenotypes and phenomenon of genome-wide inter-chromosomal linkage disequilibrium in the human genome. Exp Gerontol 46(12):979–986. doi:10.1016/j.exger.2011.08.010

Kulminski AM, Culminskaya IV (2012) Genomics of human health and aging. Age (Dordr). doi:10.1007/s11357-011-9362-x

Kulminski AM, Culminskaya I, Ukraintseva SV, Arbeev KG, Arbeeva L, Wu D, Akushevich I, Land KC, Yashin AI (2011) Trade-off in the effects of the apolipoprotein E polymorphism on the ages at onset of CVD and cancer influences human lifespan. Aging Cell 10(3):533–541. doi:10.1111/j.1474-9726.2011.00689.x

Kuningas M, Mooijaart SP, van Heemst D, Zwaan BJ, Slagboom PE, Westendorp RG (2008) Genes encoding longevity: from model organisms to humans. Aging Cell 7(2):270–280. doi:10.1111/j.1474-9726.2008.00366.x

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265(5181):2037–2048

Malinowski E (1927) The hypothesis of chromosome affinity and the phenomenon of suppression of characters on crossing. J Genet 18(2):223–231

Manolio TA, Brooks LD, Collins FS (2008) A HapMap harvest of insights into the genetics of common disease. J Clin Invest 118(5):1590–1605. doi:10.1172/JCI34772

Martin GM (2005) Genetic modulation of senescent phenotypes in *Homo sapiens*. Cell 120(4):523–532

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

Metz CW (1926) Genetic evidence of a selective segregation of chromosomes in Sciara (Diptera). Proc Natl Acad Sci U S A 12(12):690–692

Metz CW, Moses MS, Hoppe EN (1926) Chromosome behavior and genetic behavior in Sciara (Diptera) I. Chromosome behavior in the spermatocyte divisions. Mol Gen Genet MGG 42(1):237–270. doi:10.1007/BF01741805

Michie D (1953) Affinity; a new genetic phenomenon in the house mouse; evidence from distant crosses. Nature 171 (4340):26–27

Mike V (1977) Theories of quasi-linkage and "affinity": some implications for population structure. Proc Natl Acad Sci U S A 74(8):3513–3517

Morgan TH (1911) Random Segregation Versus Coupling in Mendelian Inheritance. Science 34(873):384. doi:10.1126/science.34.873.384

Omholt SW, Plahte E, Oyehaug L, Xiang K (2000) Gene regulatory networks generating the phenomena of additivity, dominance and epistasis. Genetics 155(2):969–980

Pardo-Manuel de Villena F, Sapienza C (2001) Nonrandom segregation during meiosis: the unfairness of females. Mamm Genome 12(5):331–339. doi:10.1007/s003350040003

Payne F (1912) The chromosomes of *Gryllotalpa borealis* Burm Archiv für Zellforschung. 9:141–148

Payne F (1916) A study of the germ cells of *Gryllotalpa borealis* and *Gryllotalpa vulgaris*. J Morphol 28:287–327

Perls TT, Wilmoth J, Levenson R, Drinkwater M, Cohen M, Bogan H, Joyce E, Brewster S, Kunkel L, Puca A (2002) Life-long sustained mortality advantage of siblings of centenarians. Proc Natl Acad Sci U S A 99 (12):8442–8447

Petkov PM, Graber JH, Churchill GA, DiPetrillo K, King BL, Paigen K (2005) Evidence of a large-scale functional organization of mammalian chromosomes. PLoS Genet 1(3): e33. doi:10.1371/journal.pgen.0010033

Petkov PM, Graber JH, Churchill GA, DiPetrillo K, King BL, Paigen K (2007) Evidence of a large-scale functional organization of mammalian chromosomes. PLoS Biol 5(5): e127. doi:10.1371/journal.pbio.0050127, author reply e128

Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet 9(11):855–867. doi:10.1038/nrg2452

Pluzhnikov A, Below JE, Konkashbaev A, Tikhomirov A, Kistner-Griffin E, Roe CA, Nicolae DL, Cox NJ (2010) Spoiling the whole bunch: quality control aimed at preserving the integrity of high-throughput genotyping. Am J Hum Genet 87(1):123–128. doi:10.1016/j.ajhg.2010.06.005

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559–575. doi:10.1086/519795

Robinson R (1971) Gene mapping in laboratory mammals. Plenum, London

Rockwood K, Fox RA, Stolee P, Robertson D, Beattie BL (1994) Frailty in elderly people: an evolving concept. CMAJ 150(4):489–495

Rohlfs RV, Swanson WJ, Weir BS (2010) Detecting coevolution through allelic association between physically unlinked loci. Am J Hum Genet 86(5):674–685. doi:10.1016/j.ajhg.2010.03.001

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci U S A 103(49):18656–18661. doi:10.1073/pnas.0606133103

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Sham PC, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Johnson TA, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Yakub I, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J (2007) Genome-wide detection and characterization of positive selection in human populations. Nature 449(7164):913–918. doi:10.1038/nature06250

Sapre AB, Deshpande DS (1987) Spontaneous emergence of parents from the F1 interspecific hybrids of Coix-L. J Hered 78(6):357–360

Schork NJ, Murray SS, Frazer KA, Topol EJ (2009) Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 19(3):212–219. doi:10.1016/j.gde.2009.04.010

Schrader F (1921) The chromosomes of *Pseudococcus nipae*. Biol Bull-Us 40(5):259–270

Schrader F (1923) A study of the chromosomes in three species of *Pseudococcus*. Arch Zellforsch 17:45–62

Sillers PJ, Wise D, Forer A (1983) Prometaphase forces towards opposite spindle poles are not independent: an on/off control system is identified by ultraviolet microbeam irradiations. J Cell Sci 64:69–88

Sivagnanasundaram S, Broman KW, Liu M, Petronis A (2004) Quasi-linkage: a confounding factor in linkage analysis of complex diseases? Hum Genet 114(6):588–593. doi:10.1007/s00439-004-1109-7

Slatkin M (1972) On treating the chromosome as the unit of selection. Genetics 72(1):157–168

Slatkin M (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nat Rev Genet 9(6):477–485. doi:10.1038/nrg2361

Spilianakis CG, Flavell RA (2006) Molecular biology. Managing associations between different chromosomes. Science 312(5771):207–208. doi:10.1126/science.1126689

Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB Sr, Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D (2007) The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol 165(11):1328–1335. doi:10.1093/aje/kwm021

Strickberger MW (2000) Evolution, 3rd edn. Jones and Bartlett, Sudbury

Tyler AL, Asselbergs FW, Williams SM, Moore JH (2009) Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. Bioessays 31(2):220–227. doi:10.1002/bies.200800022

Wang J, Chen PJ, Wang GJ, Keller L (2010) Chromosome size differences may affect meiosis and genome size. Science 329(5989):293. doi:10.1126/science.1190130

White MJD (1973) Animal cytology and evolution, 3dth edn. University Press, Cambridge

Willcox DC, Willcox BJ, Wang NC, He Q, Rosenbaum M, Suzuki M (2008) Life at the extreme limit: phenotypic characteristics of supercentenarians in Okinawa. J Gerontol A Biol Sci Med Sci 63(11):1201–1208

Williams A, Spilianakis CG, Flavell RA (2010) Interchromosomal association and gene regulation in trans. Trends Genet 26(4):188–197. doi:10.1016/j.tig.2010.01.007

Wise D, Sillers PJ, Forer A (1984) Non-random chromosome segregation in *Neocurtilla hexadactyla* is controlled by chromosomal spindle fibres: an ultraviolet microbeam analysis. J Cell Sci 69:1–17

Xing J, Watkins WS, Zhang Y, Witherspoon DJ, Jorde LB (2008) High fidelity of whole-genome amplified DNA on high-density single nucleotide polymorphism arrays. Genomics 92(6):452–456. doi:10.1016/j.ygeno.2008.08.007