

ARTICLE

Human longevity and variation in DNA damage response and repair: study of the contribution of sub-processes using competitive gene-set analysis

Birgit Debrabant^{*1,8}, Mette Soerensen^{*1,2,8}, Friederike Flachsbarth³, Serena Dato⁴, Jonas Mengel-From^{1,2}, Tinna Stevnsner⁵, Vilhelm A Bohr^{5,6}, Torben A Kruse², Stefan Schreiber³, Almut Nebel³, Kaare Christensen^{1,2,7}, Qihua Tan^{1,2} and Lene Christiansen^{1,2}

DNA-damage response and repair are crucial to maintain genetic stability, and are consequently considered central to aging and longevity. Here, we investigate whether this pathway overall associates to longevity, and whether specific sub-processes are more strongly associated with longevity than others. Data were applied on 592 SNPs from 77 genes involved in nine sub-processes: DNA-damage response, base excision repair (BER), nucleotide excision repair, mismatch repair, non-homologous end-joining, homologous recombinational repair (HRR), RecQ helicase activities (RECQ), telomere functioning and mitochondrial DNA processes. The study population was 1089 long-lived and 736 middle-aged Danes. A self-contained set-based test of all SNPs displayed association with longevity (P -value = 9.9×10^{-5}), supporting that the overall pathway could affect longevity. Investigation of the nine sub-processes using the competitive gene-set analysis by Wang *et al* indicated that BER, HRR and RECQ associated stronger with longevity than the respective remaining genes of the pathway (P -values = 0.004–0.048). For HRR and RECQ, only one gene contributed to the significance, whereas for BER several genes contributed. These associations did, however, generally not pass correction for multiple testing. Still, these findings indicate that, of the entire pathway, variation in BER might influence longevity the most. These modest sized P -values were not replicated in a German sample. This might, though, be due to differences in genotyping procedures and investigated SNPs, potentially inducing differences in the coverage of gene regions. Specifically, five genes were not covered at all in the German data. Therefore, investigations in additional study populations are needed before final conclusion can be drawn.

European Journal of Human Genetics (2014) 22, 1131–1136; doi:10.1038/ejhg.2013.299; published online 12 February 2014

Keywords: competitive gene-set analysis; human longevity; SNPs; case-control data; DNA-damage response and DNA repair

INTRODUCTION

It has been estimated that about 25% of the variation in human life span is caused by genetic factors,¹ a contribution which is likely to be minimal before age 60 years and most pronounced from age 85 years onwards.² Candidate longevity genes encode proteins involved in several biological processes including DNA-damage response and DNA repair,³ which are essential for cellular function due to the constant exposure of living cells to various kinds of endogenous and exogenous agents. Such agents can damage diverse cellular components, including the DNA, and potentially induce genomic instability or mutagenesis. To maintain the integrity of the genome, DNA-damage response processes detect the DNA lesions and activate cell cycle arrest and DNA repair processes.⁴ The DNA repair processes each repair a certain subset of DNA lesions:^{5–7} base excision repair removes single-base lesions (eg oxidative lesions induced by reactive oxygen species (ROS)); nucleotide excision repair operates on bulky

lesions (eg induced by UV light); mismatch repair among other repair errors arising during DNA replication, while double-strand break repair (DSB) removes DSB breaks and complex lesions (eg induced by ionizing radiation or chemotherapeutics). DSB is divided into homologous recombinational repair and non-homologous end-joining. The RecQ helicases are among the other things important for ensuring the proper structure of the DNA helix during DNA repair.⁸ Severe mutations in two of the RecQ helicase genes, *WRN* and *RECQL4*, can cause Werner and Rothmund–Thomson premature aging syndromes, respectively.^{9,10} Additional mechanisms affect the stability of the chromosomes' ends, the telomeres. Telomeres undergo shortening with every replication cycle in cells lacking telomerase until the telomeres reach a critical length at which the cell enters cellular senescence.^{11,12} Telomere shortening is associated with organismal aging; in humans, leukocyte telomere length is inversely related to age^{13,14} and is associated with an increased risk of age-

¹Epidemiology, Biostatistics and Biodemography, The Danish Aging Research Center, Institute of Public Health, University of Southern Denmark, Odense, Denmark;

²Department of Clinical Genetics, Odense University Hospital, Odense, Denmark; ³Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany;

⁴Department of Biology, Ecology and Earth Science, University of Calabria, Rende, Italy; ⁵Department of Molecular Biology and Genetics, The Danish Aging Research Center, Aarhus University, Aarhus, Denmark; ⁶Laboratory of Molecular Gerontology, National Institute on Aging, National Institute of Health, Baltimore, MD, USA; ⁷Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark

*Correspondence: B Debrabant or M Soerensen, Epidemiology, Biostatistics and Biodemography, The Danish Aging Research Center, Institute of Public Health, University of Southern Denmark, J.B. Winsloews Vej 9B, 5000 Odense C, Denmark. Tel: +45 65503544 (BD) or +45 65412822 (MS); Fax: +45 65503682 (BD) or +45 65414875 (MS); E-mail: bdebrabant@health.sdu.dk or msoerensen@health.sdu.dk

⁸These authors contributed equally to this work and are mentioned in alphabetical order.

Received 14 November 2012; revised 7 November 2012; accepted 26 November 2013; published online 12 February 2014

related disease and with higher mortality.^{15–18} Finally, mitochondrial DNA (mtDNA) is exposed to specifically high levels of ROS due to its vicinity to oxidative phosphorylation and is thus potentially exposed to high levels of oxidative damage. Specific proteins maintain the function and repair of mtDNA. Accumulation of mtDNA mutations and mitochondrial dysfunction, likely influenced by age-dependent changes in ROS and DNA repair processes, is thought to have important roles in the aging process.^{19–21}

Taken together, these nine damage response and repair processes are crucial to genetic stability, which is considered of major importance for the aging process and longevity from nematodes to humans.^{9,19,22–25} Also, research in animal models shows that overexpression or depletion of some of the genes in the processes affects the life span of the organism (eg *POLB*, *RAD52* and *TP53*^{26–28}), and common genetic variants in humans have been put forward as associated with longevity, for example, in *WRN* and *MLH1*.^{29,30}

An immense number of studies, which report an association of single variants to human longevity, have been published; still, only two genes repeatedly show an association, *APOE* and *FOXO3A*.^{3,31–34} One reason for this may be that the effect of the individual variant is believed to be small,³⁵ and hence may not be easily identified by single variant analysis. Moreover, to avoid false positive findings, there is a need to correct for multiple testing as numerous variants are often tested at the same time. This requires very low *P*-values in order to obtain significance. Investigation of the combined effect of genetic variants of entire pathways is a recent approach,³⁶ which may circumvent such difficulties. From a biological point of view, the rationale for such analysis is that it is the collective functions of the gene products in a pathway, which has a role in the cell, and, thus, the joint effect of the gene products is important to study. From a statistical point of view, a combined analysis of sets of SNPs and genes is also advantageous as it reduces the multiple testing burden and typically increases power.^{37,38}

In this study, we focus on common variation in 77 DNA-damage response and repair genes composing the above mentioned nine sub-processes. First, we investigate the association of the overall pathway with longevity. Next, we apply a gene-set method put forward by Wang *et al.*³⁹ to test whether any of the nine sub-processes is more associated with longevity than the other sub-processes. The analyses were carried out in a study population of 1089 oldest-old and 736 middle-aged Danes, whereas replication was done in a German study population of 763 long-lived individuals and 1085 controls.

MATERIALS AND METHODS

Data set – subjects of the discovery population

The details of the discovery population are described in Soerensen *et al.*⁴⁰ The oldest-old were participants from The Danish 1905 Cohort Study,⁴¹ a nationwide population-based survey of an entire birth cohort, initiated in 1998 when 3600 birth cohort members were still alive; 2262 participated in the survey of which 1651 gave blood samples. The age range for the 1089 individuals included here was 92.2–93.8 years (29% males). The control group was of 736 individuals from the Study of Middle-Aged Danish Twins,⁴² started in 1998 by random selection of 2640 intact twin pairs from 22 consecutive birth years (1931–1952) via the nationwide Danish Central Person Registry. Only one twin from each twin pair was included in the present study, and the age range was 46.0–55.0 years (50% males). Due to minimal immigration into both cohorts, they are considered genetically homogenous, and population stratification is considered minimal. Permission to collect blood samples and usage of register-based information was granted by The Danish National Committee on Biomedical Research Ethics.

Data set – subjects of the replication population

The details of the replication population are described in Nebel *et al.*⁴³ The age ranges of the 763 long-lived individuals and the 1085 controls were 94–110 and 45–77 years, respectively (50% males). All participants were of the German ancestry. Approval was received from the ethics committee of the Christian-Albrechts-University.

Data set – genotype data of the discovery population

The details of the selection of sub-processes, genes and SNPs and the generation of genotype data are described in Soerensen *et al.*⁴⁰ Sub-processes and genes were identified by a comprehensive literature search in five different databases (including <http://www.ncbi.nlm.nih.gov>). To cover the core biological functions of the sub-processes, five different databases were consulted (including <http://www.biocarta.com>). In total, 80 genes were chosen that belonged to one or more of the nine sub-processes. The gene regions plus 5000 base pairs upstream and 1000 base pairs downstream and candidate SNPs in these regions (previously associated SNPs, coding SNPs and SNPs with potential functional impact) were identified using the NCBI (<http://www.ncbi.nlm.nih.gov>) and UCSC (<http://www.genome.ucsc.edu>) genome browsers as well as other SNP databases (eg snpper.chip.org). To cover the common genetic variation within each genomic region, tagging SNPs were obtained from the HapMap consortium database (<http://hapmap.ncbi.nlm.nih.gov>) for the CEU population and selected with the HaploView software (<http://www.broadinstitute.org/haploview>) with the $R^2 > 0.8$ and $MAF > 5\%$ criteria.

DNA was isolated from blood spot cards or full blood using QIAamp DNA Kits (Düsseldorf, Germany), and genotyping of all samples was performed on the Illumina GoldenGate platform (San Diego, CA, USA). Data cleaning of all samples were made according to the manufacturer's recommendations in the GenomeStudio software (http://www.illumina.com/software/genomestudio_software.ilmn). First samples and SNPs with a call rate $< 90\%$ were excluded. Next, further data cleaning was done of SNPs with (a) a call rate 90–95%, (b) close clusters (score < 2.3), (c) clusters of low intensity (score < 0.2 or > 0.8), (d) a heterozygote cluster shifting towards a homozygote cluster (score < 0.13), (e) excess heterozygosity (score < -0.3 or > 0.2) and of (f) SNPs located on the X chromosome. Overall 8.8% of the individuals and 10.4% of the SNPs were excluded, leaving data on 592 SNPs in 77 genes from 1089 oldest-old and 736 controls. The 77 genes are listed by sub-process in Table 1. This data has been analyzed at the single-SNP level for the association to human longevity.⁴⁰

Data set – genotype data of the replication population

Data on 700 SNPs from the same 77 gene regions was extracted from an existing data set of 664 472 SNPs from the German replication sample.⁴³ Genotyping was done using the Affymetrix Genome-Wide Human SNP Array 6.0 (San Francisco, CA, USA). Quality control was carried out in R⁴⁴ and Plink,⁴⁵ SNPs were included if (a) the call rate was $\geq 95\%$ in the cases and controls, (b) the MAF was ≥ 0.02 in the controls or $|MAF(\text{controls}) - MAF(\text{cases})| > 0.02$ and (c) the Hardy–Weinberg equilibrium *P*-value in the controls was > 0.01 . Population stratification was found to be low.⁴³ After visual inspection of cluster plots, 696 SNPs were applied in the replication study.

Statistical analyses – self-contained set-based analysis

A set-based analysis of all SNPs was performed in Plink.⁴⁵ We applied the most inclusive settings, (ie set-max 99999, set- r^2 1 and set- p 1). Single-marker *P*-values from Cochran–Armitage's trend test were summarized by the overall mean of all SNPs, thus estimating the average association of the entire SNP-set with longevity. Significance was determined by permuting the phenotype labels 10 000 times.

Statistical analyses – competitive gene-set analysis

We performed a gene-set analysis using the competitive⁴⁶ gene-set enrichment method published by Wang *et al.*³⁹ which we implemented in R.⁴⁴ The analysis was in four steps: step 1: an association analysis was conducted to calculate a statistic (respectively *P*-value) for each SNP, step 2: based on these, a gene statistic was assigned to each gene choosing the statistic corresponding to the most significant SNP in each gene, step 3: based on these gene statistics, the enrichment score was calculated for each of the nine sub-processes,

Table 1 The 77 genes grouped by biological sub-process

DNA-damage response (no. of genes = 8)	BER (no. of genes = 26)	NER (no. of genes = 18)	MMR (no. of genes = 14)	HRR (no. of genes = 7)	NHEJ (no. of genes = 5)	RECQ (no. of genes = 5)	Telomere functioning (no. of genes = 9)	Mitochondrial DNA processes (no. of genes = 12)
ATM	APEX1 (ape1)	DDB1 (uv- ddb1)	EXO1	LIG1	DCLRE1C (artemis)	BLM	ACD (pip)	C10orf2 (twinkle)
ATR	APTX (aprataxin)	DDB2 (uv- ddb2)	LIG1	MRE11A	LIG4	RECQL (recq1)	PIF1	ERCC6 (csb)
BRIP1 (facj)	ERCC6 (csb)	ERCC1	MLH1	NBN	PRKDC (dna- pkcs)	RECQL4 (recq4)	POT1	ERCC8 (csa)
FANCA (faca)	ERCC8 (csa)	ERCC2 (xpd)	MLH3	RAD50	XRCC4	RECQL5 (recq5)	RECQL4 (recq4)	LONP1 (lonp)
FANCD2 (facd)	FEN1	ERCC3 (xpb)	MSH2	RAD51	XRCC5 (ku80/ku86)	WRN	TERF1	NEIL2
H2AFX (h2a.x)	HMGB1 (hmg1)	ERCC4 (xpf)	MSH3	RAD52			TERF2	POLG (polγ)
TP53 (p53)	LIG1	ERCC5 (xpg)	MSH6	RAD54L (rad54)			TERF2IP (rap1)	POLRMT (mtrpol)
YBX1 (yb1)	LIG3 NEIL1 NEIL2 NTHL1 OGG1 PARP1 PCNA PNKP (pnk) POLB POLD1 (polδ) POLE (polε) RECQL4 (recq4) RECQL5 (recq5) RFC1 (rfc) RPA1 (rpa) UNG WRN XRCC1 YBX1 (yb1)	ERCC6 (csb) ERCC8 (csa) LIG1 PCNA POLD1 (polδ) POLE (polε) RAD23B RFC1 (rfc) RPA1 (rpa) XPA XPC	PCNA PMS1 PMS2 POLD1 (polδ) RFC1 (rfc) RPA1 (rpa) YBX1(yb1)				TERT WRN	RECQL4 (recq4) SUPV3L1 (suv3) TFAM (mttfa) TP53 (p53) YBX1 (yb1)

Abbreviations: BER, base excision repair; HRR, homologous recombinational repair; MMR, mismatch repair; NER, nucleotide excision repair; NHEJ, non-homologous end-joining; RECQ, RecQ helicase activities.
Names of gene products are given in parenthesis if different from gene names.

reflecting the degree of over-representation of associated genes within a chosen sub-process as opposed to the genes outside the sub-process, step 4: significance was determined by permuting the phenotype labels 10 000 times and repeating the steps above, hereby generating the enrichment score P -value for each of the nine sub-processes. Following the recommendations of the authors,³⁹ we used a weight parameter $p = 1$ to calculate the enrichment score.

After deriving the enrichment score P -values for each of the nine sub-processes, we used the false discovery rate (FDR) to take the testing of nine sub-processes into account. As described by Wang *et al*³⁹ for a given sub-process S^* (in our case $S1-S9$) with a normalized enrichment score NES^* ($NES1-NES9$), the FDR was estimated as the proportion of sub-processes with a normalized ES (after permutation) at least as high as NES^* , divided by the proportion of sub-processes with a normalized ES (non-permuted) at least as high as the observed NES^* . We estimated corresponding enrichment score q -values as the minimal FDR obtainable when declaring a test significant.

Moreover, step 1 was implemented in three different ways referred to as setup I, II and III. We applied the three setups to inspect their influence on the results. In setup I, we used χ^2 (2df)-statistics (assumption-free model) assessing the difference in genotype frequencies between cases and controls without any constraints on the genotypic model. In setup II, Cochran–Armitage’s trend test statistic was used (suitably transformed to obtain a (scaled)

χ^2 (2df)-distribution under the null, ie we used $-\log P$ as single-marker statistic, where P is the P -value from the trend test). The trend test was included as it is the most applied model in the literature on the genetics of human longevity. In setup III (best-test model), we inspected all possible modes of inheritance simultaneously by combining Cochran–Armitage’s trend test and χ^2 tests based on assumption-free (2df), recessive (1df) and dominant (1df) models. Thereby we defined the test statistic for each SNP to be the minimum P -value of these four tests (back-transformed so the test statistic equaled $-\log (\min 1 \leq i \leq 4 P_i)$, where $P1-P4$ correspond to the P -values from the four tests, respectively). This procedure is analogous to the MAX test described previously by Freidlin *et al*.⁴⁷ The log-transformation used in setups II and III was applied to make the three setups comparable as the enrichment score used by the method of Wang *et al*³⁹ is sensitive to the scale of the statistics. As the three setups are highly dependent, the enrichment score q -values were calculated (by FDR) individually for each of them.

RESULTS

Self-contained set-based analysis of the entire pathway

The set-based test resulted in an overall P -value of 9.9×10^{-5} for the combined set of SNPs. This result supports that variation in

DNA-damage response and DNA repair genes appears to be of importance for human longevity.

Competitive gene-set analysis of the nine sub-processes

The results of the initial SNP analysis (step 1), using the three different statistical setups, can be seen in Supplementary Table 1. The data can further be found in the GWAS Central data base (www.gwascentral.org). Table 2 shows enrichment score *P*- and *q*-values obtained for the nine sub-processes compared to the respective remaining genes of the entire pathway, under the three different statistical setups. We found the results to be consistent with only slight differences; uniformly over the three setups, sub-processes BER, HRR and RECQ showed *P*-values below 0.05, whereas the remaining sub-processes were non-significant. However, when taking into account the number of sub-processes investigated, only HRR showed a *q*-value below 0.05 and only in setup II (see Table 2).

As the three setups showed consistent results, the subsequent analyses in the discovery data were restricted to setup III. To examine which genes contributed to the nominal significance of BER, HRR and RECQ, the individual gene statistics for all genes in these sub-processes were inspected (see Table 3). These gene statistics enter the calculation of the enrichment score as weights and provide intuitive information about the extent to which the score was mainly determined by specific genes. We found several low-ranking genes in the case of BER, while for HRR the association was mostly due to a single gene, *RAD52*, and to a much lower degree *NBN*, *MRE11A* and *RAD54L*. Similarly, the *P*-value for RECQ was primarily due to *WRN* and to a smaller extent *BLM*. In a *post-hoc* analysis, we then repeated the gene-set analysis of BER, RECQ and HRR removing the respective lowest ranking gene from the data set. Only BER remained nominally significant (*P*-value = 0.037), while loss of significance of RECQ (*P*-value = 0.34) and HRR (*P*-value = 0.88) confirmed that the importance of RECQ and HRR each was induced by only one gene.

Replication study in German long-lived individuals and controls

The single-SNP data from the German replication population can be found in the GWAS Central data base (www.gwascentral.org). Repeating the self-contained set-based analysis of all SNPs as well as the competitive gene-set analysis of BER, HRR and RECQ in the

Table 2 Competitive gene-set analysis of the genetic variation in the nine sub-processes of DNA-damage response and DNA repair in relation to human longevity

Sub-processes	Setup					
	I – χ^2		II – trend test		III – best-test model	
	<i>P</i> -value	<i>q</i> -value	<i>P</i> -value	<i>q</i> -value	<i>P</i> -value	<i>q</i> -value
DmgRes	0.113	0.171	0.475	0.556	0.299	0.447
BER	0.015	0.110	0.029	0.121	0.016	0.111
NER	0.679	0.681	0.556	0.563	0.679	0.681
MMR	0.434	0.553	0.498	0.556	0.686	0.681
HRR	0.035	0.110	0.004	0.030	0.025	0.111
NHEJ	0.099	0.171	0.163	0.286	0.112	0.231
RECQ	0.032	0.110	0.048	0.159	0.049	0.160
TeloFunc	0.076	0.159	0.077	0.163	0.130	0.231
MitoDNA	0.616	0.681	0.439	0.556	0.608	0.681

Abbreviations: BER, base excision repair; DmgRes, DNA-damage response; HRR, homologous recombinational repair; MitoDNA, mitochondrial DNA processes; MMR, mismatch repair; NER, nucleotide excision repair; NHEJ, non-homologous end-joining; RECQ, RecQ helicase activities; TeloFunc, telomere functioning.

P- and *q*-values <0.05 are shown in bold.

German study population did not show significant results: *P*-value (self-contained set-based analysis) = 0.227 and *P*-values (competitive gene-set analysis) = 0.468–0.919 (see Supplementary Table 2).

DISCUSSION

In this study, we aimed to investigate whether genetic variation in any of the sub-processes of DNA-damage response and DNA repair influences human longevity more than others. This is an important question as the individual sub-processes contribute to different aspects of the maintenance of genetic stability, and therefore have diverse biological roles, which are likely to affect longevity in different manners.

Table 3 Ranking of the genes in the BER, HRR and RECQ sub-processes

Sub-process	Gene	<i>p</i> -gene (gene statistic)	Rank
BER	<i>WRN</i>	2.1×10^{-8}	3
	<i>APTX</i>	3.6×10^{-8}	4
	<i>NTHL1</i>	1.4×10^{-7}	5
	<i>POLB</i>	1.8×10^{-5}	6
	<i>RPA1</i>	2.4×10^{-4}	9
	<i>XRCC1</i>	4.9×10^{-4}	13
	<i>YBX1</i>	1.5×10^{-3}	16
	<i>PARP1</i>	2.4×10^{-3}	20
	<i>OGG1</i>	2.8×10^{-3}	21
	<i>PCNA</i>	3.0×10^{-3}	22
	<i>ERCC8</i>	4.9×10^{-3}	26
	<i>UNG</i>	6.0×10^{-3}	28
	<i>LIG3</i>	1.9×10^{-2}	36
	<i>POLD1</i>	2.8×10^{-2}	41
	<i>APEX1</i>	2.9×10^{-2}	42
	<i>HMGB1</i>	3.6×10^{-2}	43
	<i>RFC1</i>	4.1×10^{-2}	46
	<i>NEIL1</i>	5.1×10^{-2}	50
	<i>PNKP</i>	6.3×10^{-2}	51
	<i>LIG1</i>	6.5×10^{-2}	54
HRR	<i>NEIL2</i>	7.3×10^{-2}	57
	<i>ERCC6</i>	8.8×10^{-2}	61
	<i>POLE</i>	1.6×10^{-1}	68
	<i>RECQL4</i>	1.6×10^{-1}	69
	<i>FEN1</i>	1.9×10^{-1}	73
	<i>RECQL5</i>	8.1×10^{-1}	77
	<i>RAD52</i>	4.3×10^{-18}	1
	<i>NBN</i>	2.1×10^{-3}	18
	<i>MRE11A</i>	4.0×10^{-3}	24
	<i>RAD54L</i>	2.7×10^{-2}	39
RECQ	<i>LIG1</i>	6.5×10^{-2}	54
	<i>RAD51</i>	7.8×10^{-2}	59
	<i>RAD50</i>	1.2×10^{-1}	64
	<i>WRN</i>	2.1×10^{-8}	3
	<i>BLM</i>	7.8×10^{-4}	15
	<i>RECQL4</i>	1.6×10^{-1}	69
	<i>RECQL</i>	2.0×10^{-1}	75
	<i>RECQL5</i>	8.1×10^{-1}	77

Abbreviations: BER, base excision repair; HRR, homologous recombinational repair; RECQ, RecQ helicase activities.

Rank: overall rank of each gene among all 77 DNA-damage response and DNA repair genes. *P*-gene values below 0.05 are shown in bold.

The *P*-gene values listed here were based on statistical setup III and the log-transformed *P*-gene values were used as weights in the calculation of the enrichment score. The *P*-gene values for setups I and II can be derived from Supplementary Table 1.

Our initial self-contained test of all SNPs showed significant association, indicating that DNA-damage response and repair overall appears to be important for human longevity. In the subsequent competitive gene-set analysis, we found nominal significant association with longevity of three sub-processes: BER, HRR and RECQ, which indicates that these may be of greater importance compared with the remaining six sub-processes. Notably, the lack of comparative significance for the remaining sub-processes does not necessarily imply that these are irrelevant for longevity, but rather that they appear less important than BER, HRR and RECQ. When correcting for the test of nine sub-processes, only HRR remained significant and only in setup II (Table 2). One reason for the lack of significance for the remaining associations could be reduced power due to an inadequate sample size (1089 oldest-old and 736 middle-aged), which implies that larger study populations may be needed for a more conclusive result. Then again, the lack of significance could well be an inherent part of the competitive analysis as we compare the associations of sub-processes all of which are considered relevant for longevity; thus, it may be difficult to uncover a significant relative difference between such sub-processes.

Our findings indicate that repair of certain types of DNA damage likely could be of particular importance for human longevity; for instance damage induced by ROS or ionizing radiation, which is repaired by BER and HRR, respectively. Moreover, ROS induced damage can give rise to DNA DSBs during DNA replication, which potentially induce replication stress and replication fork collapse, a situation in which HRR is crucial. Finally, the roles of the RECQ helicases for ensuring the proper structure of the DNA during repair could be central. This is especially interesting considering the existence of premature aging syndromes, such as the Werner syndrome, caused by mutations in a RECQ gene.

An interesting aspect of the gene-set analysis is whether the nominal significance of BER, HRR and RECQ was mainly driven by one gene or by the combined effect of several genes. For RECQ and HRR, the associations were driven primarily by *WRN* and *RAD52*, respectively. The latter might indicate that variation in one of the sub-units of the rad51/rad52/rad54 protein complex may be sufficient to modify the function of the entire HRR. Our results could also indicate that *RAD52* and *WRN* have roles beyond HRR and RECQ. *WRN* is a clear example of this, as it takes part in three of nine sub-processes. Similarly, one study indicates that rad52 may affect BER via an interaction with ogg1.⁴⁸ Finally, as our definition of the sub-processes is based on existing knowledge of DNA-damage response and repair mechanisms, we cannot exclude that additional cross-talks between the sub-processes are yet to be discovered.

Contrary to this, the association of BER appears scattered over the entire gene-set, suggesting that variation in different parts of the sub-process may have consequences for longevity. Considering the central role of oxidative damage to several proposed biological theories of aging,^{22,49} the fact that we observe this for BER, the DNA repair process primarily responsible for repair of oxidative DNA lesions, is indeed interesting.

One concern when applying a competitive gene-set analysis strategy is obviously how the details and widths of the sub-processes are identified. This is in general a complicated task as there is no gold standard and available databases may lack sufficient detail³⁶ and accordance in the pathway composition.⁵⁰ Consequently, it is possible that some relevant genetic variation is not covered in the present study.

An intriguing aspect of human longevity is the gender-specific difference in survival, a difference which has been suggested, at least in part, to have a genetic basis.⁵¹ A sex-stratified self-contained set-based analysis of all SNPs investigated showed significance for both

males and females (data not shown). Similarly, stratification of the competitive gene-set analysis yielded consistent significance of BER. Still, HRR and RECQ did not show significance (setup III, data not shown). The latter might possibly be due to the reduction in sample size in the sex-stratified analysis. Altogether, in the present data, gender-specific differences do not appear to be overly relevant.

An issue of increasing interest in genetic association studies is gene-environment ($G \times E$) interactions. It is, though, problematic to consider environmental factors in case-control studies of longevity; these factors may not be constant over time, that is, the longevity cases could have experienced a different environment when they were at the age of the younger controls. Instead, a longitudinal design that investigates survival during old age would be more suitable for $G \times E$ studies. In relation to DNA repair, a $G \times E$ analysis could be especially appealing as the sub-processes repair certain types of DNA damage induced by certain types of environmental sources. An example is the oxidative damage repaired by BER, which, among others, is induced by the general metabolism, which might be affected differently by diverse types and amounts of food intake.

When we performed the self-contained set-based analysis (of all SNPs under study) and the competitive gene-set analysis of BER, HRR and RECQ in the German study population, we did not observe replication. There are a number of possible reasons for this. First of all, different genotyping procedures were used in the two study populations. This may cause differences in the coverage of the variation in the gene regions, which potentially could affect the relative difference between the sub-processes. Specifically, five genes were not covered at all in the German data: *H2AFX* (DNA-damage response), *APEX1* (BER), *ACD* (telomere functioning), *ERCC1* (nucleotide excision repair) and *RECQL4* (BER, RECQ, telomere functioning and mtDNA processes). Moreover, not exactly the same SNPs were genotyped in the two study populations; hence, depending on the LD of the discovery SNPs, these might not all be tagged in the German data. Secondly, we aimed to replicate very modest-sized enrichment score *P*-values ($P = 0.004\text{--}0.048$, Table 2), of which only one in nine remained significant after correcting for the test of nine sub-processes. Thus, it may be difficult to replicate associations with such values as compared to replicating those with smaller *P*-values. Furthermore, it might be difficult to detect a relative significance between sub-processes (competitive gene-set analysis) in the German data set considering that the overall pathway was non-significant (self-contained set-based analysis). Thirdly, the Germans were older than the Danes, and, if the associations are not constant throughout old age, they may not replicate well. Such change in association with age has been reported previously.⁵² Finally, it is possible that the effect of BER, HRR, RECQ and of the pathway overall is population specific, perhaps mediated through population-specific characteristics in environment. Such population-specific effects have previously been suggested for single variants.⁵³ Hence, before final conclusion can be drawn, additional studies in other, preferably larger, study populations are needed.

CONCLUSION

In this study, we investigated the association between human longevity and genetic variation in nine sub-processes of DNA-damage response and repair, all considered to have key roles in maintenance of genetic stability and the aging process. We observed significant support for an association of the entire pathway with longevity. In the competitive analysis, we furthermore found BER, HRR and RECQ to be nominally significantly more associated with longevity than other sub-processes, although this did, in general, not remain significant

when correcting for the number of sub-processes tested. Especially for BER, the association was caused by several genes, which indicates that, of the entire pathway, variation in BER might influence longevity the most. Still, as these findings did not show replication in the German study population, more analyses are needed in additional study populations before final conclusion can be drawn.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by the Max-Planck Institute for Demographic Research, (Rostock, Germany), the INTERREG 4 A programme Syddanmark-Schleswig-K.E.R.N (by EU funds from the European Regional Development Fund), the European Union's Seventh Framework Programme (FP7/2007–2011) under grant agreement no. 259679, the National Institute on Aging (P01 AG08761), the RESOLVE project (FP7-HEALTH-F4-2008-202047) and the Novo Nordisk, the Aase and Ejnar Danielsen, the Brødrene Hartmann, the King Christian the 10th and the Einer Willumsens Mindelegat Foundations. The Danish Aging Research Center is supported by a grant from the VELUX Foundation, while Mette Soerensen is supported by an individual postdoctoral grant from The Danish Council for Independent Research – Medical Sciences.

- Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW: The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870–1900. *Hum Genet* 1996; **97**: 319–323.
- Hjelmborg JVb, Iachine I, Skytthe A *et al*: Genetic influence on human lifespan and longevity. *Hum Genet* 2006; **119**: 312–321.
- Christensen K, Johnson TE, Vaupel JW: The quest for genetic determinants of human longevity: challenges and insights. *Nat Rev Genet* 2006; **7**: 436–448.
- Ciccia A, Elledge SJ: The DNA damage response: making it safe to play with knives. *Mol Cell* 2010; **40**: 179–204.
- Hiom K: Coping with DNA double strand breaks. *DNA Repair (Amst)* 2010; **9**: 1256–1263.
- Hoijmakers JH: Genome maintenance mechanisms for preventing cancer. *Nature* 2001; **411**: 366–374.
- Jeppesen DK, Bohr VA, Stevens T: DNA repair deficiency in neurodegeneration. *Prog Neurobiol* 2011; **94**: 166–200.
- Bohr VA: Rising from the RecQ-age: the role of human RecQ helicases in genome maintenance. *Trends Biochem Sci* 2008; **33**: 609–620.
- Burner CR, Kennedy BK: Progeria syndromes and ageing: what is the connection? *Nat Rev Mol Cell Biol* 2010; **11**: 567–578.
- Croteau DL, Singh DK, Hoh FL, Lu H, Bohr VA: RECQL4 in genomic instability and aging. *Trends Genet* 2012; **28**: 624–631.
- Blackburn EH, Greider CW, Szostak JW: Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. *Nat Med* 2006; **12**: 1133–1138.
- Campisi J, d'Adda de Fagnana F: Cellular senescence: when bad things happen to good cells. *Nat Rev Mol Cell Biol* 2007; **8**: 729–740.
- Lindsey J, McGill NI, Lindsey LA, Green DK, Cooke HJ: *In vivo* loss of telomeric repeats with age in humans. *Mutat Res* 1991; **256**: 45–48.
- Slagboom PE, Droog S, Boomsma DI: Genetic determination of telomere size in humans: a twin study of three age groups. *Am J Hum Genet* 1994; **55**: 876–882.
- Bakaysa SL, Mucci LA, Slagboom PE *et al*: Telomere length predicts survival independent of genetic influences. *Aging Cell* 2007; **6**: 769–774.
- Kimura M, Hjelmborg JV, Gardner JP *et al*: Telomere length and mortality: a study of leukocytes in elderly Danish twins. *Am J Epidemiol* 2008; **167**: 799–806.
- Njajou OT, Cawthon RM, Damcott CM *et al*: Telomere length is paternally inherited and is associated with parental lifespan. *Proc Natl Acad Sci USA* 2007; **104**: 12135–12139.
- Shay JW, Woodring WE: Telomeres and Telomerase in aging and cancer. *Molecular Biology of Aging and Cancer*, In: Guarente LP, Partridge L, Wallace DC (eds). *Molecular Biology of Aging*. Cold Spring Harbor Laboratory Press: New York, 2008, pp 575–597.
- Barja G: Free radicals and aging. *Trends Neurosci* 2004; **27**: 595–600.
- Bohr VA, Anson RM: DNA damage, mutation and fine structure DNA repair in aging. *Mutat Res* 1995; **338**: 25–34.
- Kujoth GC, Bradshaw PC, Haroon S, Prolla TA: The role of mitochondrial DNA mutations in mammalian aging. *PLoS Genet* 2007; **3**: e24.
- Kirkwood TB, Austad SN: Why do we age? *Nature* 2000; **408**: 233–238.
- Kuningas M, Mooijaart SP, van HD, Zwaan BJ, Slagboom PE, Westendorp RG: Genes encoding longevity: from model organisms to humans. *Aging Cell* 2008; **7**: 270–280.
- Lindahl T: Instability and decay of the primary structure of DNA. *Nature* 1993; **362**: 709–715.
- Viig J: Somatic mutations and aging: a re-evaluation. *Mutat Res* 2000; **447**: 117–135.
- Cabelof DC, Ikeno Y, Nyska A *et al*: Haploinsufficiency in DNA polymerase beta increases cancer risk with age and alters mortality rate. *Cancer Res* 2006; **66**: 7460–7465.
- Park PU, Defossez PA, Guarente L: Effects of mutations in DNA repair genes on formation of ribosomal DNA circles and life span in *Saccharomyces cerevisiae*. *Mol Cell Biol* 1999; **19**: 3848–3856.
- Tyner SD, Venkatachalam S, Choi J *et al*: p53 mutant mice that display early ageing-associated phenotypes. *Nature* 2002; **415**: 45–53.
- Castro E, Edland SD, Lee L *et al*: Polymorphisms at the Werner locus: II. 1074Leu/Phe, 1367Cys/Arg, longevity and atherosclerosis. *Am J Med Genet* 2000; **95**: 374–380.
- Kim DJ, Yi SM, Lee SY *et al*: Association between the MLH1 gene and longevity. *Hum Genet* 2006; **119**: 353–354.
- Schachter F, Faure-Delanef L, Guenet F *et al*: Genetic associations with human longevity at the APOE and ACE loci. *Nat Genet* 1994; **6**: 29–32.
- Willcox BJ, Donlon TA, He Q *et al*: FOXO3A genotype is strongly associated with human longevity. *Proc Natl Acad Sci USA* 2008; **105**: 13987–13992.
- Flachsbart F, Caliebe A, Kleindorp R *et al*: Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc Natl Acad Sci USA* 2009; **106**: 2700–2705.
- Soerensen M, Dato S, Christensen K *et al*: Replication of an association of variation in the FOXO3A gene with human longevity using both case-control and longitudinal data. *Aging Cell* 2010; **9**: 1010–1017.
- Manolio TA, Collins FS, Cox NJ *et al*: Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747–753.
- Deelen J, Uh HW, Monajemi R *et al*: Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. *Age (Dordr)* 2013; **35**: 235–249.
- Wang K, Li M, Hakonarson H: Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 2010; **11**: 843–854.
- Wang L, Jia P, Wolfinger RD, Chen X, Zhao Z: Gene set analysis of genome-wide association studies: methodological issues and perspectives. *Genomics* 2011; **98**: 1–8.
- Wang K, Li M, Bucan M: Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* 2007; **81**: 1278–1283.
- Soerensen M, Dato S, Tan Q *et al*: Human longevity and variation in GH/IGF-1/insulin signaling, DNA damage signaling and repair and pro/antioxidant pathway genes: cross sectional and longitudinal studies. *Exp Gerontol* 2012; **47**: 379–387.
- Nybo H, Gaist D, Jeune B *et al*: The Danish 1905 cohort: a genetic-epidemiological nationwide survey. *J Aging Health* 2001; **13**: 32–46.
- Gaist D, Vaeth M, Tsiropoulos I *et al*: Risk of subarachnoid haemorrhage in first degree relatives of patients with subarachnoid haemorrhage: follow up study based on national registries in Denmark. *BMJ* 2000; **320**: 141–145.
- Nebel A, Kleindorp R, Caliebe A *et al*: A genome-wide association study confirms APOE at the major gene influencing survival in long-lived individuals. *Mech Ageing Dev* 2011; **132**: 324–330.
- R Development Core Team: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria 2012.
- Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 2007; **81**: 559–575.
- Goeman JJ, Buhlmann P: Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics* 2007; **23**: 980–987.
- Freidlin B, Zheng G, Li Z, Gastwirth JL: Trend tests for case-control studies of genetic markers: power, sample size and robustness. *Hum Hered* 2002; **53**: 146–152.
- de Souza-Pinto NC, Maynard S, Hashiguchi K *et al*: The recombination protein RAD52 cooperates with the excision repair protein OGG1 for the repair of oxidative lesions in mammalian cells. *Mol Cell Biol* 2009; **29**: 4441–4454.
- Harman D: The aging process: major risk factor for disease and death. *Proc Natl Acad Sci USA* 1991; **88**: 5360–5363.
- Lambert JC, Grenier-Boley B, Chouraki V *et al*: Implication of the immune system in Alzheimer's disease: evidence from genome-wide pathway analysis. *J Alzheimers Dis* 2010; **20**: 1107–1118.
- Montesanto A, Latorre V, Giordano M, Martino C, Domma F, Passarino G: The genetic component of human longevity: analysis of the survival advantage of parents and siblings of Italian nonagenarians. *Eur J Hum Genet* 2011; **19**: 882–886.
- Jacobsen R, Martinussen T, Christiansen L *et al*: Increased effect of the ApoE gene on survival at advanced age in healthy and long-lived Danes: two nationwide cohort studies. *Aging Cell* 2010; **9**: 1004–1009.
- Di Bona D, Vasto S, Capurso C *et al*: Effect of interleukin-6 polymorphisms on human longevity: a systematic review and meta-analysis. *Ageing Res. Rev* 2009; **8**: 36–42.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)