Large common deletions associate with mortality at old age

Maris Kuningas1,*, Karol Estrada2, Yi-Hsiang Hsu4,5, Kannabiran Nandakumar4, André G. Uitterlinden1,2,6, Kathryn L. Lunetta7, Cornelia M. van Duijn1,6, David Karasik4, Albert Hofman1,6, Joanne Murabito8,9, Fernando Rivadeneira2,6, Douglas P. Kiel4 and Henning Tiemeier1,3,6

1Department of Epidemiology, 2Department of Internal Medicine and 3Department Psychiatry, Erasmus Medical Center, Rotterdam, CA 3000, The Netherlands 4Hebrew Senior Life Institute for Aging Research and Harvard Medical School, Boston, MA 02131, USA, 5Molecular and Integrative Physiological Sciences Program, Harvard School of Public Health, Boston, MA 02115, USA, 6Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), Leiden, RC 2300, The Netherlands, 7Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA, 8Sections of General Internal Medicine, Preventive Medicine, Cardiology and Neurology, Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA and 9The National Heart Lung and Blood Institute’s Framingham Heart Study, Framingham, MA 01702-5827, USA

Received February 15, 2011; Revised July 1, 2011; Accepted August 1, 2011

Copy-number variants (CNVs) are a source of genetic variation that increasingly are associated with human disease. However, the role of CNVs in human lifespan is to date unknown. To identify CNVs that influence mortality at old age, we analyzed genome-wide CNV data in 5178 participants of Rotterdam Study (RS1) and positive findings were evaluated in 1714 participants of the second cohort of the Rotterdam Study (RS2) and in 4550 participants of Framingham Heart Study (FHS). First, we assessed the total burden of rare (frequency <1%) and common (frequency >1%) CNVs for association with mortality during follow-up. These analyses were repeated by stratifying CNVs by type and size. Secondly, we assessed individual common CNV regions (CNVR) for association with mortality. We observed that the burden of common but not of rare CNVs influences mortality. A higher burden of large (≥500 kb) common deletions associated with 4% higher mortality [hazard ratio (HR) per CNV 1.04, 95% confidence interval (CI) 1.02–1.07, \( P = 5.82 \times 10^{-5} \)] in the 11 442 participants of RS1, RS2 and FHS. In the analysis of 312 individual common CNVRs, we identified two regions (11p15.5; 14q21.3) that associated with higher mortality in these cohorts. The 11p15.5 region (combined HR 1.59, 95% CI 1.31–1.93, \( P = 2.87 \times 10^{-6} \)) encompasses 41 genes, of which some have previously been related to longevity, whereas the 14q21.3 region (combined HR 1.57, 95% CI 1.19–2.07, \( P = 1.53 \times 10^{-3} \)) does not encompass any genes. In conclusion, the burden of large common deletions, as well as common CNVs in 11p15.5 and 14q21.3 region, associate with higher mortality.

INTRODUCTION

It has long been observed that human lifespan has a genetic component (heritability 25–30%) (1–3), but to date only few genes have been identified to influence it. The recent genome-wide association studies (GWASs) relating single nucleotide polymorphisms (SNPs) to longevity have provided only modest additional information (4,5). These studies were well powered to detect the association of common SNPs with longevity, suggesting that either rare variants or other types of genetic variation might be at play. Recently, the genome-wide study of copy number variations (CNVs) has become feasible and thereby also the study of CNVs in relation to human lifespan.

*To whom correspondence should be addressed. Tel: +31 107043488; Fax: +31 107044657; Email: m.kuningas@erasmusmc.nl

© The Author 2011. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
CNVs are caused by chromosomal rearrangements that result in the loss (deletion) or gain (duplication) of stretches of DNA sequence, and they represent a significant source of genetic variation in the human genome (6). The mechanisms through which CNVs can lead to phenotypic effects include among others gene interruption, gene fusion and changes in gene expression. Differences in gene expression can result from a dose effect for genes that are within a deletion or duplication, or from altered expression through disruption of regulatory elements (7–9). Gene interruption, on the other hand, can lead to gene inactivation, whereas gene fusion caused by genomic rearrangements between different genes or their regulatory sequences can generate a gene with a new and unexpected function (7–9). It has been noticed that CNVs can be detrimental, beneficial or have no phenotypic consequences at all (10).

In humans, CNVs have been shown to cause sporadic and Mendelian diseases (10), but recent evidence implicates CNVs also in a variety of complex traits (11–15). In these studies, individual CNVs and the total burden of rare CNVs have been analyzed. Analyzing the burden of rare CNVs (frequency <1%) enables to estimate the combined effect of rare variants. Through this approach, the burden of rare CNVs has been identified as a possible cause of schizophrenia (14), autism (11), bipolar disorder (16), amyotrophic lateral sclerosis (17,18) and obesity (19). Likewise, lifespan could be affected by the combined effect of rare CNVs, but the same could hold true for the combined effect of common CNVs. Common CNVs are individually likely to have small or moderate phenotypic effects, but in combination they might result in considerable phenotypic effects. Such effects can reliably be assessed with moderate sample sizes. To date, however, common CNVs have typically been analyzed separately for their association with various traits. It has been found that individual common CNVs (frequency >5%) associate with Crohn’s disease, rheumatoid arthritis, type 1 diabetes and type 2 diabetes (15). Most of these common CNVs are tagged by SNPs, and have therefore been indirectly studied with GWASs (6,15). Nevertheless, as not all CNVs are tagged by SNPs, the study of individual common CNVs remains important. The same hold true for the evaluation of their combined effect, i.e. CNV burden.

In this study, we assessed the impact of CNVs on mortality in two ways: first as individuals’ total burden of rare and common CNVs, and second, by searching for common individual CNV regions. We hypothesized that a higher burden of CNVs, either rare or common, is associated with higher mortality. In addition, we expected to identify novel genomic locations that influence lifespan by analyzing individual common CNV regions (CNVRs). The study was carried out in the first cohort (RS1) of the prospective population-based Rotterdam study (RS), and positive findings were replicated in the second cohort of RS (RS2) and in the Framingham Heart Study.

RESULTS

We analyzed the total burden of CNVs that were present in <1% (rare CNVs) or in >1% (common CNVs) of the RS1 study population separately. Altogether, 95% of these participants carried at least one rare CNV, whereas 99% carried at least one common CNV (Supplementary Material, Table S1). Per individual we observed a median number of three rare and six common CNVs (Supplementary Material, Table S1). Among common CNVs, no very large (≥1 Mb) deletions or duplications were observed.

First, we investigated the effect of the total burden of rare and common CNVs on mortality in RS1. Of the 5178 participants, 2716 (53%) died during a median follow-up period of 15 years (Table 1). For the total burden of rare CNVs, no associations with mortality were observed (Table 2). This result did not change when the rare CNV burden was analyzed separately for deletions and duplications, or CNV size (Table 2). In contrast, for the burden of common CNVs, we observed that
a higher burden of deletions which are >500 kb is associated with 17% higher mortality in RS1 (HR 1.17, 95% CI 1.05–1.30, \( P = 5.0 \times 10^{-3} \)) (Table 2). Similar associations were also observed in RS2 (HR 1.58, 95% CI 0.58–4.25, \( P = 0.37 \)) and FHS (HR 1.04, 95% CI 1.02–1.06, \( P = 4.3 \times 10^{-3} \)). In the overall meta-analysis of the 11442 participants of RS1, RS2 and FHS, a 4% increase in mortality (HR 1.04, 95% CI 1.02–1.07, \( P = 5.82 \times 10^{-5} \)) was observed.

Next, we examined the association of individual CNVRs that have a frequency >1% with mortality. In defining CNVRs, overlapping CNVs within a region were merged, and the outermost boundaries of those CNVs were taken. This approach yielded 312 CNVRs with an average size of 847 kb (range 1.70 kb–14.5 Mb), of which nine were associated with mortality in RS1 (Table3). From these, the association with mortality was most consistently replicated for the 11p15.5 and 14q21.3 CNVRs, within which we repeated the initial analysis comparing with non-carriers, as observed in 11 442 participants and it contributes to higher mortality as observed in 11 442 participants of RS1, RS2 and FHS (combined HR 1.59, 95% CI 1.31–1.93, \( P = 2.87 \times 10^{-5} \)) (Table 3).

Within the RS1 discovery cohort, we repeated all analyses for CNVRs with frequency >5%. This did not reveal any significant associations (data not shown). In addition, to test whether the association between higher mortality and the burden of deletions >500 kb in RS1 is accounted for by the 11 442 participants of RS1, RS2 and FHS (combined HR 1.59, 95% CI 1.31–1.93, \( P = 2.87 \times 10^{-5} \)) (Table 3).

Table 3. Association between mortality and common individual CNVRs

<table>
<thead>
<tr>
<th>CNVR</th>
<th>Locus</th>
<th>Size (KB)</th>
<th>Type</th>
<th>Freq.</th>
<th>HR (95% CI)</th>
<th>Freq.</th>
<th>HR (95% CI)</th>
<th>Freq.</th>
<th>HR (95% CI)</th>
<th>Overall HR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr4:16092457–162509675</td>
<td>4q32.1</td>
<td>1229.1</td>
<td>DEL/DUP</td>
<td>0.12</td>
<td>0.99 (0.89–1.11)</td>
<td>0.01</td>
<td>0.98 (0.78–1.24)</td>
<td>0.04</td>
<td>1.11 (0.80–1.53)</td>
<td>1.52 (1.23–1.88)</td>
<td>1.53E–03</td>
</tr>
<tr>
<td>chr14:104197399–106066496</td>
<td>14q32.33</td>
<td>1869.1</td>
<td>DEL/DUP</td>
<td>0.04</td>
<td>1.09 (0.89–1.33)</td>
<td>0.01</td>
<td>0.87 (0.64–1.18)</td>
<td>0.01</td>
<td>0.88 (0.65–1.20)</td>
<td>0.52 (0.40–0.70)</td>
<td>8</td>
</tr>
<tr>
<td>chr11:280233–1074362</td>
<td>11p15.5</td>
<td>794.1</td>
<td>DEL/DUP</td>
<td>0.02</td>
<td>1.33 (1.05–1.68)</td>
<td>0.02</td>
<td>2.23 (1.02–4.83)</td>
<td>0.01</td>
<td>2.38 (1.59–3.54)</td>
<td>1.59 (1.31–1.93)</td>
<td>2.87E–06</td>
</tr>
<tr>
<td>chr9:5257043–5354592</td>
<td>9p24.1</td>
<td>97.6</td>
<td>DEL/DUP</td>
<td>0.01</td>
<td>1.46 (1.06–2.01)</td>
<td>0.01</td>
<td>0.41 (0.06–2.90)</td>
<td>NA</td>
<td>NA</td>
<td>1.46 (1.06–2.01)</td>
<td>0.027</td>
</tr>
<tr>
<td>chr14:44254784–44320168</td>
<td>14q21.3</td>
<td>65.4</td>
<td>DEL</td>
<td>0.01</td>
<td>1.49 (1.04–2.12)</td>
<td>0.01</td>
<td>2.49 (1.05–6.05)</td>
<td>0.01</td>
<td>1.47 (0.83–2.54)</td>
<td>1.57 (1.19–2.07)</td>
<td>1.53E–03</td>
</tr>
</tbody>
</table>

DISCUSSION

In this genome-wide CNV analysis we observed in RS1 and replicated in RS2 and FHS that a higher burden of deletions larger than 500 kb associates with 4% higher mortality. In addition, we identified two common CNV regions (CNVRs) that associate with higher mortality as well. One of these CNVRs (11p15.5) is located in a gene-rich region, whereas the other (14q21.3) does not encompass any genes.

In this study we expected that the higher burden of CNVs, either rare or common would be associated with higher mortality. For the burden of rare CNVs, we observed no significant associations. It is likely that rare CNVs are highly detrimental and that carriers of these do not reach old age. Therefore, we speculate that the rare CNVs observed in this study of older individuals have negligible functional consequences. Alternatively, they could have moderate phenotypic effects, but due to power limitations, were unable to detect these associations,
Figure 1. The 11p15.5 CNV region. The chromosome (chr)11:280233–1074362 region (794 kb) contains overlapping deletions (red) and duplications (blue), which were merged into a CNVR. The lines for duplications and deletions represent RS1 individuals who carry these CNVs. Within this region, 41 RefSeq genes are located. The breakdown of linkage disequilibrium between SNPs within this region and Database of Genomic Variants entries from other studies are shown at the bottom part of the figure.
leading to a potential false-negative result. Common CNVs are likely to contribute to moderate phenotypic effects, but due to their higher frequency their combined effect is more easily detectable. Here, we observed that individuals with a higher burden of common large deletions have 4% higher mortality. This association could not be explained by the two individual common CNVRs (11p15.5 and 14q21.3) that also associated with higher mortality. This suggests that not only the number of CNVs, but also the combination of CNVs that an individual carries may have an influence on the risk of mortality.

In this study, we tested 312 common individual CNVRs, of which two (11p15.5 and 14q21.3) were consistently associated with higher mortality across the three populations. Considering the low number of CNVRs tested and the level of significance these associations reached, we are confident that these loci represent true findings. Furthermore, both of these CNVRs have been reported by the Database of Genomic Variants (DGV, http://projects.tcag.ca/variation/). The 11p15.5 CNV is located in a gene-rich region and it encompasses 41 genes. Some of these genes have been associated with cardiovascular disease (CVD; PKP3) (20), cancer (HRAS, RNHI1, TALDO1) (21–23), insulin resistance (PNPLA2) (24) and longevity (HRAS) (25,26). In the HRAS gene, a 3’ variable number of tandem repeat allele was observed to be less frequent in centenarians compared with young controls (25), and in an independent study, a haplotype was shown to enhance the effect of APOE on exceptional survival (26). However, even though there is prior evidence for the influence of HRAS gene on human lifespan, other genes in this region or in close proximity might also be of relevance. Especially when taking into account that the 11p15.5 region is located nearby genes (SIRT3, TH, INS and IGF2) that are considered to be prominent candidate genes for human longevity (27–29). Based on our results, we cannot conclude the independence or uniform effect of the CNV events spanning this region, where specific evaluation of the CNV breaking points is needed. The other, 14q21.3, CNVR encompasses no genes, making functional implications less clear cut. The closest gene to that region is c14orf28, which is located 116 kb downstream. Recently, it was demonstrated that the expression of c14orf28 is altered in schizophrenia and bipolar disorder samples relative to controls (30), implicating this gene in dopamine-signaling pathway. Although the 14q21.3 deletion is located far upstream, it could influence the transcription of c14orf28 gene. Another alternative functional consequence of the 14q21.3 deletion includes a disruption of highly evolutionarily conserved DNA region with functions either in chromosome assembly or in DNA replication (31).

This in study, we called CNVs common, when they were present in >1% of the study population (i.e. polymorphism). Under this definition, we were able to observe significant associations for CNVs that have a relatively low frequency (11p15.5 2%; 14q21.3 1%). In several other studies, different cut-off criteria for determining common CNVs have been used. One of these criteria includes frequency >5% (15). When we applied this criterion, no significant associations with CNVRs and mortality were observed. Previously it has been demonstrated that most of the CNVs with a frequency >5% are well tagged by SNPs and have indirectly been studied in GWASs (6,15). Hence, the lack of associations with CNVRs with frequency >5% in this study is in line with the lack of genome-wide significant associations between SNPs and longevity. On the other hand, in this study, we were powered enough to study CNVs that are rare but not extremely rare, representing frequencies for which very large samples sizes would be needed in GWASs.

The current study has several strengths. One of these is the availability of genome-wide CNV data within a large population-based cohort, implying unbiased phenotyping. In addition, the participants of this study have been followed for mortality for a relatively long period of time, yielding a high number of events to study. Also, by making use of RS2 and FHS cohorts, we were able to validate the CNV events and their frequency in different cohorts and across different genotyping array types. Furthermore, we were able to replicate the associations observed in RS1. A strength but also a drawback of the study is the use of high-density SNP arrays for detecting CNVs. These arrays enable high genomic resolution, but they have been designed to target unique non-duplicated regions of the human genome. Therefore, not all regions in the genome are equally well covered, especially those which are prone to genomic rearrangements (32). Consequently, some genomic regions were not assessed and remain to be studied. In addition, we observed CNVs in the same genomic regions with different frequencies in different study populations. This is likely to stem from the genotype data used, since if SNPs for some reason do not pass quality control, it affects CNV inference (33). Despite these limitations, we identified CNVRs that associate with higher mortality across three study populations. The use of specific CNV arrays in the future would enable a more comprehensive assessment of the relationship between CNVs and lifespan.

In conclusion, in this study, we found that the burden of large common deletions, as well as common CNVs in the 11p15.5 and 14q21.3 regions, contribute to higher mortality at old age.

**SUBJECTS AND METHODS**

**Study populations**

This study is embedded in the RS and the FHS. The RS is an ongoing population-based study of risk factors for chronic diseases in the elderly. Detailed information on design, objectives and methods has been presented elsewhere (34). The current study included 5178 participants of the RS1 and 1714 participants of RS2 for whom genome-wide CNV data were available. All participants of the RS were followed for mortality until 1 January 2009. Altogether 2716 (53%) of the 5178 participants in RS1 and 223 (13%) of the 1714 participants RS2 died during the follow-up period (Table 1). The Medical Ethics Committee of the Erasmus Medical Center approved the RS and written informed consent was obtained from all participants. The FHS was initiated to study the determinants of CVD and other major illnesses. The original cohort was recruited in 1948 and the offspring of the original cohort participants and offspring spouses were enrolled in 1971 (35–38). DNA was obtained for genetic studies in the 1990s from surviving original cohort and offspring participants. The time at
which DNA was obtained is considered the baseline examination for these analyses. For the current study, genetic information and data on mortality were available for 4550 participants.

CNV detection

In RS, CNVs were inferred from the genome-wide genotype data of the RS participants. The genome-wide genotyping was performed with Illumina 550K array (Illumina, San Diego, CA, USA) in individuals of European descent (sample call rate ≥ 97.5%). Individuals with excess of autosomal heterozygosity mismatch between genotypic and phenotypic gender, and outliers identified by the identity-by-state clustering analysis, were excluded. For CNV detection, log R ratio (LRR) signal intensity and B allele frequency (BAF) were extracted using BeadStudio 3.1.3 (Illumina). A hidden-Markov model, implemented in the software QuantiSNP (32), was used to make CNV calls. For each CNV call, a log Bayes factor was computed as a measure of confidence. A correction for local difference in GC content is implemented in the algorithm, which enables to adjust for irregularities in signal intensity. Samples with a mean autosomal LRR SD > 0.35 or a BAF SD > 0.15 were excluded, as well as CNVs calls that spanned the centromere. In addition, all CNV calls with a log Bayes factor value < 10, a CNV size < 1 kb or CNVs with only one consecutive SNP in the CNV event, were excluded. For samples with an LRR SD between 0.3 and 0.35, a stricter confidence threshold was applied (BAF = 15). Finally, samples with an excess of CNV calls [expressed as upper quartile + 1.5 × (inter-quartile range)] were also excluded. In FHS, genotyping was performed using the Affymetrix genome wide array 5.0 for 8735 individuals. The raw CEL files were read, normalized and CNVs called using Affymetrix power tools and the PennCNV software. A total of 112 746 CNVs were obtained from the analysis. Extensive quality control of data was done using several measures. CNVs that did not adhere to a one or more of these criteria were excluded.

Generation of CNV-burden and CNVR datasets

The data on total CNV burden and on specific CNVRs per individual was estimated by using PLINK v1.07 software (39). In estimating the total CNV burden, CNVs that overlapped by at least 50% of their length with another CNV were considered to co-localize and counted once. For CNVRs, completely non-redundant CNVRs were constructed by merging overlapping CNVs and taking the outermost boundaries of those CNVs. These datasets were generated for deletions and duplications combined and separated.

Statistical analysis

The association between mortality and the burden of CNVs and individual CNVRs was calculated with Cox proportional hazard regression, adjusted for age at baseline and sex. All association analyses were performed with SPSS version 17 (SPSS Inc., Chicago, IL, USA) statistical software. Results were combined in a meta-analysis using the inverse variance method to combine effect size estimates from each sample (40).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters for their help in creating the GWAS database. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Research Institute for Diseases in the Elderly (grant number 014-93-015; RIDE2); and the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) project (grant number 050-060-810; NCHA); Erasmus Medical Center and Erasmus University, Rotterdam; Netherlands Organization for the Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); the Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam. The generation and management of GWAS genotype data for the Rotterdam Study are supported by the Netherlands Organization of Scientific Research NWO Investments (grant numbers 175.010.2005.011, 911-03-012 and 017-106-370 VIDI to H.T.). The FHS was funded by grants from the US National Institute for Arthritis, Musculoskeletal and Skin Diseases and National Institute on Aging (R01 AR/AG 41398 D.P.K.; R01 AR 050066 D.K.; R01 AG 29451 J.M. and K.L.L.) and Hebrew SeniorLife Men’s Associate. The Framingham Heart Study of the National Heart, Lung, and Blood Institute of the National Institutes of Health, National Heart, Lung, and Blood Institute’s Framingham Heart Study of the National Heart, Lung, and Blood Institute of the National Institutes of Health and Boston University School of Medicine were supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health and Boston University School of Medicine. Analyses reflect intellectual and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARE) project. A portion of this research was conducted using the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

REFERENCES


