European Journal of Human Genetics (2013) 21, 240–242

www.nature.com/ejhg

SHORT REPORT

Genetic investigation of *FOXO3A* requires special attention due to sequence homology with *FOXO3B*

Friederike Flachsbart[1], Michael Möller[1], Carolin Däumer[2], Liljana Gentschew[1], Rabea Kleindorp[1], Michael Krawczak[2,3], Amke Caliebe[2], Stefan Schreiber[1,3,4,5] and Almut Nebel*[,1,5]

Our study demonstrates that the genetic investigation of forkhead box O3A gene (*FOXO3A*), a validated human longevity gene, is greatly hampered by the fact that its exonic regions have 99% sequence homology with the *FOXO3B* pseudogene. If unaccounted for, this high degree of homology can cause serious genotyping or sequencing errors. Here, we present an experimental set-up that allows reliable data generation for the highly homologous regions and that can be used for the evaluation of assay specificity. Using this design, we exemplarily showed *FOXO3A*-specific results for two single-nucleotide polymorphisms (SNPs) (rs4945816 and rs4946936) that are significantly associated with longevity in our centenarian-control sample ($P_{each} = 0.0008$). Because both SNPs are located in the 3′ untranslated region of *FOXO3A*, they could be of functional relevance for the longevity phenotype. Our experimental set-up can be used for reliable and reproducible data generation for further sequencing and genotyping studies of *FOXO3A* with the aim of discovering new SNPs of functional relevance.
European Journal of Human Genetics (2013) **21,** 240–242; doi:10.1038/ejhg.2012.83; published online 16 May 2012

Keywords: centenarians; human longevity; aging; FOXO3A; FOXO3B

INTRODUCTION

The forkhead box O3A gene (*FOXO3A*) is a key regulator of the insulin receptor/insulin-like growth factor-I signaling pathway and participates in the modulation of lifespan in *Caenorhabditis elegans* and *Drosophila*.[1–5] In humans, genetic variation in *FOXO3A* has been shown consistently in different populations to be associated with survival into old age.[6–11] The underlying molecular mechanisms, however, still remain to be elucidated because most of the longevity-associated single-nucleotide polymorphisms (SNPs) analyzed are located in intronic regions.[6–11] In order to identify the causative allelic variants that modulate gene functionality, studies are therefore needed that explore the full genetic variation of *FOXO3A* by in-depth resequencing of exons, exon–intron boundaries and the promoter.

Before embarking on such investigations, it has to be taken into consideration that a pseudogene, called *FOXO3B*, is located on chromosome 17p11. *FOXO3B* is 99% identical to *FOXO3A* over a stretch of approximately 7.1 kb of exonic sequence. The pseudogene comprises all the three exons of *FOXO3A* NM_001455 and lacks the two introns (Figure 1).[12] Because exon 2 and 3 of *FOXO3A* are comparatively large (1.4 kb and 4.9 kb, respectively), the *FOXO3A*-specific intronic regions do not provide sufficient primer targets for the analysis of these two exons. For example, to generate high-quality reads with Sanger sequencing, PCR amplicons should not exceed 600 bases. However, because exon 2 and 3 are each larger than 1400 bases, their sequencing primers need to be placed within the exonic regions, despite their strong homology with *FOXO3B*. Nevertheless, for some sequencing or genotyping assays, it may be possible to place primers particularly on *FOXO3A/FOXO3B* sequence differences (which comprise only 1% of the critical regions) (Supplementary Figure S1). For the genetic investigation of the complete exonic region (7.1 kb), however, a conventional primer design, using the homologous exonic regions for primer placement, would most likely result in unspecific genotyping or mixed *FOXO3A/FOXO3B* sequencing amplicons. Unfortunately, the outcome cannot be predicted reliably by, for instance, *in silico* test PCR prior to the actual wet lab experiment.

Here, we present an experimental set-up that allows *FOXO3A*-specific genetic investigation of exonic regions without *FOXO3B* 'contamination': with the generation of long-range products that span around 8000 bases, exon 2 and 3 can be amplified in a *FOXO3A*-specific manner using the non-homologous intronic regions for primer placement. Genotyping and sequencing experiments are described exemplarily for exon 3: first, exon 3 was amplified by long-range PCR with *FOXO3A*-specific primer placement in the nearby unique intronic regions (Supplementary Table S3) to ensure targeted amplification of the *FOXO3A* exonic region (ie, devoid of *FOXO3B* sequence). Next, the long-range amplicons were used as templates for genotyping and sequencing reactions. We tested the suitability of this experimental set-up for two SNPs located in exon 3 of *FOXO3A*, namely rs4945816 and rs4946936. These two SNPs were selected because they are the only ones in the *FOXO3A* exonic region that are listed on the dbSNP page as common SNPs in CEU-samples (http://www.ncbi.nlm.nih.gov/projects/SNP/).

RESULTS AND DISCUSSION

Sequenom test set-up: differences in *FOXO3A* specificity

Genotyping with Sequenom technology (Sequenom, Hamburg, Germany) was carried out for both SNPs on 23 individuals employing three different templates for comparison: *FOXO3A* long-range products

[1]Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany; [2]Institute of Medical Informatics and Statistics, Christian-Albrechts-University, Kiel, Germany; [3]Popgen Biobank, Christian-Albrechts-University, Kiel, Germany; [4]Clinic for Internal Medicine I, University Hospital Schleswig-Holstein, Kiel, Germany
*Correspondence: A Nebel, Institute of Clinical Molecular Biology, Christian-Albrechts-University, Schittenhelmstrasse 12, Kiel 24105, Germany. Tel: +49 431 597 1373; Fax: +49 431 597 2196; E-mail: a.nebel@mucosa.de
[5]These authors contributed equally to this work.
Received 3 November 2011; revised 6 March 2012; accepted 12 April 2012; published online 16 May 2012

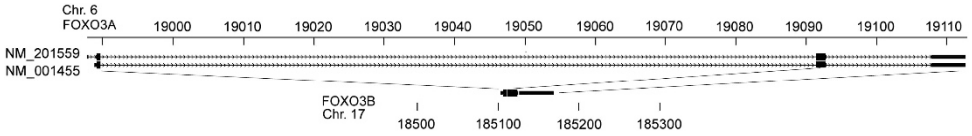

Figure 1 Illustration of the *FOXO3A* gene model showing the two isoforms NM_201559 (four exons) and NM_001455 (three exons) on chromosome 6. The exons are indicated by boxes and are presented in comparison with the *FOXO3B* pseudogene on chromosome 17. *FOXO3B* is 99% identical to *FOXO3A* over approximately 7.1 kb of related sequence comprising all the three exons of *FOXO3A* NM_001455 while lacking the two introns in-between. The physical positions (in kilobases) of the transcripts refer to the UCSC Genome Browser on the Human March 2006 Assembly.

Table 1 Genotyping and sequencing data for rs4945816 and rs4946936 on nine exemplary individuals

		(a) Test set-up			*(b)*	*(c)*
SNP	*Sample number*	*FOXO3A long-range product (Sequenom)*	*FOXO3B long-range product (Sequenom)*	*Genomic DNA (Sequenom)*	*FOXO3A long-range product (Sanger Sequencing)*	*Genomic DNA (TaqMan; AoD[a] and SDA[b])*
rs4945816						**hCV27392797[a]**
	1	**TT**	**CC**	**CT**	**TT**	**TT**
	2	CC	CC	CC	CC	CC
	3	**TT**	**CC**	**CT**	**TT**	**TT**
	4	**TT**	**CC**	**CT**	**TT**	**TT**
	5	**TT**	**CC**	**CT**	**TT**	**TT**
	6	CT	CC	NA	NA	CT
	7	CC	CC	CC	CC	CC
	8	**TT**	**CC**	**CT**	**TT**	**TT**
	9	CT	CC	CT	CT	CT
rs4946936						**rs4946936_SDA[b]**
	1	CC	CT	CC	CC	CC
	2	TT	TT	TT	TT	TT
	3	CC	NA	CC	CC	CC
	4	CC	TT	CC	CC	CC
	5	CC	TT	CC	CC	CC
	6	CT	TT	CT	CT	CT
	7	TT	TT	TT	TT	TT
	8	CC	CT	CC	CC	CC
	9	CT	TT	CT	CT	CT

Abbreviations: AoD, Assay on Demand; NA, no genotypes available; SDA, self-designed assay; SNP, single-nucleotide polymorphism.
(a) Test set-up: genotypes for both SNPs using three different templates for comparison: *FOXO3A* long-range products, *FOXO3B* long-range products and gDNA.
(b) *FOXO3A*-specific sequencing using *FOXO3A* long-range products as primer targets.
(c) Testing of TaqMan assays on gDNA.
Bold entries are those samples that demonstrate an unspecific typing by showing a mixture of *FOXO3A* and *FOXO3B* genotypes for the gDNA template.
[a]AoD = hCV27392797.
[b]SDA = rs4946936_SDA.

of exon 3 (5.523 kb), *FOXO3B* long-range products (8.540 kb) and genomic DNA (gDNA). Table 1 shows exemplary data obtained for nine individuals. Data on the remaining 14 individuals are given in Supplementary Table S1. Sequenom PCR primers for both SNPs were placed on *FOXO3A/FOXO3B* sequence differences (Supplementary Figure S1). Despite this particular primer placement, comparison of genotyping results for SNP rs4945816 obtained from the three different templates shows that the seemingly heterozygous genotypes derived from the gDNA template (samples 1, 3, 4, 5 and 8) are due to unspecific typing and represent mixtures of *FOXO3A* and *FOXO3B* genotypes (Table 1a, Supplementary Table S1a). Fictitious heterozygosity becomes evident only if *FOXO3A* and *FOXO3B* genotypes differ from each other, as it was the case for samples 1, 3, 4, 5 and 8 (Table 1a). On the other hand, the second SNP assay (rs4946936) demonstrated *FOXO3A* specificity by presenting the same genotyping results for the *FOXO3A* long-range product and the gDNA template, both of which differed from *FOXO3B* (Table 1a). The differences in *FOXO3A* specificity for the two Sequenom assays clearly indicate that *FOXO3A*-assay specificity cannot be predicted reliably prior to the actual wet lab experiment. Only the comparison of our genotyping results for the gDNA template with those derived from the *FOXO3A* and *FOXO3B* long-range products demonstrated that the assay for rs4946936 was *FOXO3A*-specific, whereas the assay for rs4945816 was not (Table 1 and Supplementary Table S1). Consequently, to obtain *FOXO3A*-specific results for rs4945816 and rs4946936 with the applied Sequenom assays, the typing of rs4945816 would need to be performed on *FOXO3A* long-range products, whereas rs4946936 could be investigated on gDNA.

FOXO3A-specific sequencing

We further showed that our experimental set-up is also suitable for *FOXO3A*-specific sequencing of the problematic homologous

exonic regions. To this end, we sequenced exon 3 of *FOXO3A* in the same 23 individuals using *FOXO3A* long-range products as targets for sequencing primers instead of gDNA. The subsequent SNP calling for rs4945816 and rs4946936 yielded the same *FOXO3A*-specific genotypes (Table 1b, Supplementary Table S1b) as before (Table 1a, Supplementary Table S1a).

FOXO3A-specific genotyping

For large-scale case–control association studies, it might be too time- and money-consuming to produce *FOXO3A*-specific long-range PCR products for thousands of samples to obtain the *FOXO3A*-specific genotyping results. Hence, in order to obviate the need for such products, our study design can also be used to test commercially available assays (eg TaqMan; Applied Biosystems, Foster City, CA, USA) for primer specificity: First, the assays of interest are typed in a small number of samples using the gDNA, *FOXO3A* and *FOXO3B* templates. The genotypes derived from the three different templates are then compared to evaluate assay specificity. Next, if the assays tested on gDNA show the same genotypes as those from *FOXO3A* long-range products, the custom assays can also be used for genotyping larger samples using gDNA as a template. We applied this set-up to evaluate *FOXO3A* specificity for one custom TaqMan assay (rs4945816; hCV27392797) and one self-designed TaqMan assay (rs4946936; rs4946936_SDA). Both TaqMan assays demonstrated *FOXO3A* specificity by yielding the same genotyping results as obtained from the *FOXO3A* long-range product (Table 1c, Supplementary Table S1c). As these two assays proved to generate reliable genotypes from normal gDNA, we decided to investigate them subsequently in our whole-study sample comprising 747 centenarians and 1102 controls. Both SNPs showed a highly significant association with longevity in an allele-based case-control analysis (rs4945816, $P = 0.0008$; rs4946936, $P = 0.0008$).

In silico investigation of functional relevance

Because both SNPs are located in the 3′ untranslated region of *FOXO3A*, they could be of functional relevance. Applying different prediction tools for *in silico* analysis, we tested the two SNPs rs4945816 and rs4946936 for potentially functional importance. There were no functional or structural effects, microRNA-binding sites or splice sites predicted (Supplementary Table S2). Only those tools that analyzed splicing regulation elements (SREs) detected differences between the reference and mutated sequences. However, these results have to be seen with caution because SREs are known to be less conserved, and consequently, this kind of tool is unlikely to generate robust and accurate results.[13] The two SNPs rs4945816 and rs4946936 have been investigated exemplarily to test our set-up. Future studies that aim to identify new SNPs of functional relevance in the *FOXO3A* exonic region should consider the aspects presented here.

CONCLUSION

Taken together, our results clearly show that, for reliable genetic investigation of *FOXO3A* exonic regions, a special study design is mandatory accounting for the *FOXO3B* sequence homology. This finding is of particular importance because some studies have already used unspecific primers for the amplification of *FOXO3A* exonic regions. For example, Mikse *et al*[14] applied various primer sets to determine *FOXO3A* deletions in tumor cells by qPCR. When we tested these pairs with the UCSC *in silico* Test-PCR (http://genome.ucsc.edu/cgi-bin/hgPcr?command=start), the output showed additional binding to *FOXO3B* for two of them (see Supplementary Materials and Methods).

The 99% *FOXO3A/FOXO3B* homology does not affect the former *FOXO3A* longevity association findings because, as has been mentioned above, most of the associated and validated SNPs are located in *FOXO3A*-specific intronic regions. Nevertheless, the pseudogene has to be taken into consideration in future *FOXO3A* genetic studies. Furthermore, the presented strategy could also be applied to other gene pairs or pseudogenes with high sequence homologies that render the design of PCR primers or reliable genotyping assays difficult, for instance the cytochrome P450 genes *CYP2D6, CYP2D7, CYP2D8* or the RH blood-group genes (*RH*).[15,16]

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all the study participants for their cooperation. For excellent technical assistance, we wish to acknowledge the laboratory personnel of the Institute of Clinical Molecular Biology and the members of the Popgen Biobank. This study was supported by the Deutsche Forschungsgemeinschaft (NE1191/1-1), the RESOLVE project (FP7-HEALTH-F4-2008-202047) and the Excellence Cluster 'Inflammation at Interfaces'.

1 Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R: A C. elegans mutant that lives twice as long as wild type. *Nature* 1993; **366**: 461–464.
2 Lin K, Dorman JB, Rodan A, Kenyon C: Daf-16: an HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. *Science* 1997; **278**: 1319–1322.
3 Kenyon C: The plasticity of aging: insights from long-lived mutants. *Cell* 2005; **120**: 449–460.
4 Ziv E, Hu D: Genetic variation in insulin/IGF-1 signaling pathways and longevity. *Ageing Res Rev* 2011; **10**: 201–204.
5 Hwangbo DS, Gershman B, Tu MP, Palmer M, Tatar M: Drosophila dFOXO controls lifespan and regulates insulin signalling in brain and fat body. *Nature* 2004; **429**: 562–566.
6 Anselmi CV, Malovini A, Roncarati R *et al*: Association of the FOXO3A locus with extreme longevity in a Southern Italian centenarian study. *Rejuvenation Res* 2009; **12**: 95–104.
7 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.
8 Li Y, Wang WJ, Cao H *et al*: Genetic association of FOXO1A and FOXO3A with longevity trait in Han Chinese populations. *Hum Mol Genet* 2009; **18**: 4897–4904.
9 Pawlikowska L, Hu D, Huntsman S *et al*: Association of common genetic variation in the insulin/IGF1 signaling pathway with human longevity. *Aging Cell* 2009; **8**: 460–472.
10 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.
11 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.
12 Anderson MJ, Viars CS, Czekay S, Cavenee WK, Arden KC: Cloning and characterization of three human forkhead genes that comprise an FKHR-like gene subfamily. *Genomics* 1998; **47**: 187–199.
13 Baralle D, Lucassen A, Buratti E: Missed threads. the impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep* 2009; **10**: 810–816.
14 Mikse OR, Blake Jr DC, Jones NR *et al*: FOXO3 encodes a carcinogen-activated transcription factor frequently deleted in early-stage lung adenocarcinoma. *Cancer Res* 2010; **70**: 6205–6215.
15 Ekman GC, Billingsly R, Hessner MJ: Rh genotyping: avoiding false-negative and false-positive results among individuals of African ancestry. *Am J Hematol* 2002; **69**: 34–40.
16 Meijerman I, Sanderson LM, Smits PHM, Beijnen JH, Schellens JHM: Pharmacogenetic screening of the gene deletion and duplications of CYP2D6*. *Drug Metab Rev* 2007; **39**: 45–60.

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