

## SHORT COMMUNICATION

## Comprehensive translational control of tyrosine kinase expression by upstream open reading frames

K Wethmar<sup>1,2,6</sup>, J Schulz<sup>1,6</sup>, EM Muro<sup>3,4</sup>, S Talyan<sup>4</sup>, MA Andrade-Navarro<sup>3,4</sup> and A Leutz<sup>1,5</sup>

Post-transcriptional control has emerged as a major regulatory event in gene expression and often occurs at the level of translation initiation. Although overexpression or constitutive activation of tyrosine kinases (TKs) through gene amplification, translocation or mutation are well-characterized oncogenic events, current knowledge about translational mechanisms of TK activation is scarce. Here, we report the presence of translational *cis*-regulatory upstream open reading frames (uORFs) in the majority of transcript leader sequences of human TK mRNAs. Genetic ablation of uORF initiation codons in TK transcripts resulted in enhanced translation of the associated downstream main protein-coding sequences (CDSs) in all cases studied. Similarly, experimental removal of uORF start codons in additional non-TK proto-oncogenes, and naturally occurring loss-of-uORF alleles of the *c-met* proto-oncogene (*MET*) and the kinase insert domain receptor (*KDR*), was associated with increased CDS translation. Based on genome-wide sequence analyses we identified polymorphisms in 15.9% of all human genes affecting uORF initiation codons, associated Kozak consensus sequences or uORF-related termination codons. Together, these data suggest a comprehensive role of uORF-mediated translational control and delineate how aberrant induction of proto-oncogenes through loss-of-function mutations at uORF initiation codons may be involved in the etiology of cancer. We provide a detailed map of uORFs across the human genome to stimulate future research on the pathogenic role of uORFs.

*Oncogene* advance online publication, 22 June 2015; doi:10.1038/onc.2015.233

## INTRODUCTION

Systems biology data have demonstrated the importance of many layers of protein expression control beyond transcriptional gene regulation, including regulation at the translational and post-translational levels. Specifically, ribosome profiling has uncovered abundant non-canonical translational initiation at upstream open reading frames (uORFs) located in the transcript leader sequence (TLS) of eukaryotic mRNA.<sup>1,2</sup> A potential uORF is defined by a translational initiation site that precedes the initiation codon of the main protein-coding sequence (CDS) and by a subsequent in-frame termination codon within a given transcript. Translation of a uORF is considered to affect the translation rates of a subsequent CDS by interfering with unrestrained progression of the scanning ribosome or by consuming a functional preinitiation complex.<sup>3–6</sup> In uORF-containing transcripts, leaky ribosomal scanning across the uORF initiation codon or reinitiation of ribosomes after termination at the uORF stop codon, permit translation of the CDS.<sup>7</sup> Multiple modes of translational control through uORFs have been reported for ~100 eukaryotic genes,<sup>4</sup> yet systematic analyses on the relevance of uORF-regulation in specific groups of functionally important genes are lacking. Although there is evidence that uORF-activating or -creating point mutations in the tumor-suppressor genes *CDKN1B* and *CDKN2A* are involved in the development of multiple endocrine neoplasia syndrome type 4 and hereditary melanoma,

respectively,<sup>8,9</sup> few data exist on a potential role of loss-of-uORF mutations in the development of human disease.<sup>10</sup>

Tyrosine kinases (TKs) represent a prototype family of oncogenic proteins that are key enzymatic regulators of cellular signaling cascades involved in proliferation and differentiation control. TKs are frequently overexpressed or constitutively activated in malignant cells<sup>11,12</sup> and oncogenic functions of TKs have been attributed to (I) genetic translocations (for example, *BCR/ABL* in chronic myeloid leukemia), (II) gene amplifications (for example, *ERBB2* in breast cancer) or (III) mutations within their kinase domains causing enhanced or constitutive signaling activity (for example, *c-Kit* in acute myeloid leukemia). Here, we investigated the prevalence and regulatory impact of uORF-mediated translational control in TKs and other proto-oncogenes by sequence analysis and by targeted mutational ablation of uORF initiation and termination codons. Our data suggest frequent uORF-mediated translational repression of human TKs and imply a potential driver function for loss-of-uORF mutations in tumorigenesis.

## RESULTS AND DISCUSSION

Characterization of prevalence and properties of TK uORFs

The set of TKs analyzed in this study includes all human proteins with assigned 'tyrosine kinase activity' according to the amigo.geneontology.org website (GO:0004713) accessed in April 2012.

<sup>1</sup>Department of Cell Differentiation and Tumorigenesis, Max-Delbrueck-Center for Molecular Medicine, Berlin, Germany; <sup>2</sup>Department of Hematology, Oncology and Pneumology, University Hospital Muenster, Muenster, Germany; <sup>3</sup>Department of Computational Biology and Data Mining, Max-Delbrueck-Center for Molecular Medicine, Berlin, Germany; <sup>4</sup>Department of Biology, Institute of Molecular Biology, Johannes-Gutenberg University, Mainz, Germany and <sup>5</sup>Department of Biology, Humboldt-University, Berlin, Germany. Correspondence: Dr K Wethmar, Department of Hematology, Oncology and Pneumology, University Hospital Muenster, Albert-Schweitzer-Campus 1, D-48149 Muenster, Germany. E-mail: klaus.wethmar@ukmuenster.de

<sup>6</sup>These authors contributed equally to this work.

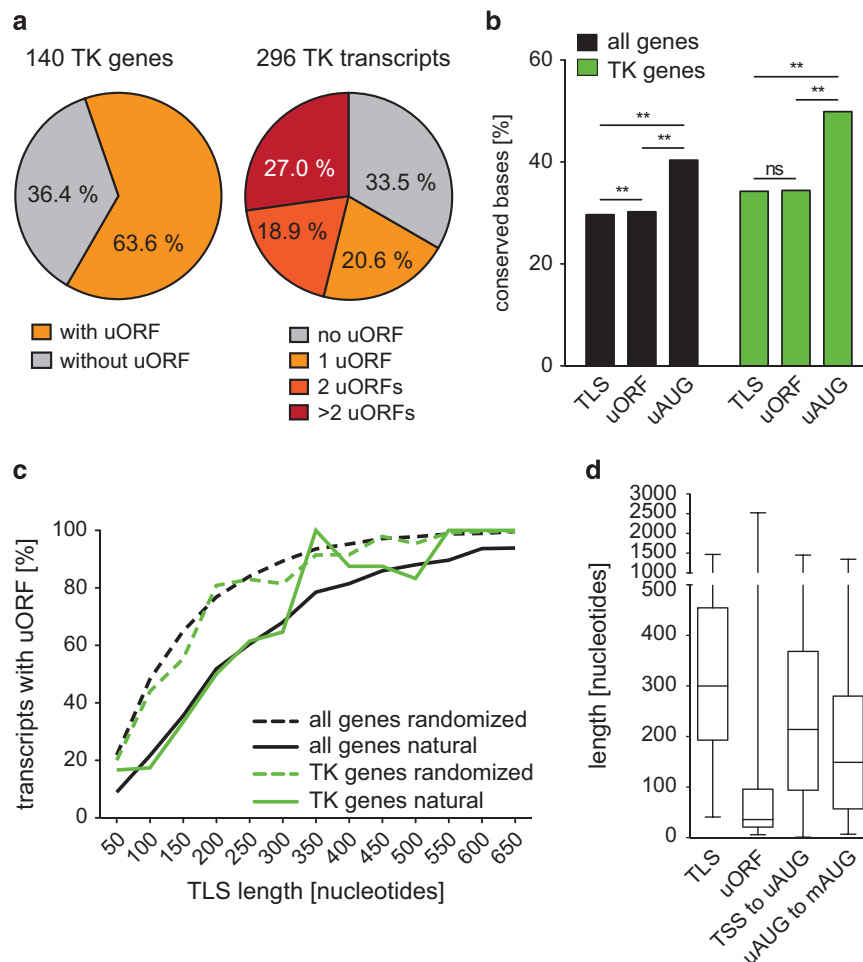
Received 29 November 2014; revised 2 April 2015; accepted 10 May 2015

A total of 140 individual proteins were identified that were encoded by 296 transcript variants annotated in the RefSeq (hg19) database.<sup>13</sup> Within this set, a total of 409 distinct AUG-initiated uORFs were detected. For simplicity, and because translational initiation at uAUG codons appeared to be most relevant in downstream translational control,<sup>2</sup> the analysis of near cognate non-AUG upstream initiation codons was excluded from this study, despite a potential role in serving as alternative uORF start sites.<sup>1,2,14</sup>

Of the 140 human TK genes, 89 (63.6%) encoded at least one transcript variant that contained one or more uORF(s) (Figure 1a and Supplementary Table 1). The observed proportion of uORF-bearing genes in the TK subset was higher as compared with the average uORF prevalence of 55.5% in the whole genome (Supplementary Table 2). Most likely, longer TLSs of TK transcripts accounted for the higher proportion of uORF-bearing genes in the

TK subset, as the ratio of uORF initiation codons per base of TLS was not increased above genome-wide frequencies (data not shown).

Within the 296 individual TK transcripts, one, two or more uORFs were found in 20.6%, 18.9% and 27.0%, respectively, showing that the occurrence of multiple uORFs within the same transcript is a rather frequent event (Figure 1a). In all, 18.6% of the uORFs overlapped with the CDS initiation codon (Supplementary Table 1), meaning that translation of these uORFs would prevent ribosomal reinitiation at the downstream CDS. By mapping uORF genomic positions to the '100 vertebrates conserved elements' provided by the UCSC,<sup>15</sup> we found that uAUGs were significantly higher conserved as compared with the remaining uORF or TLS sequences in both, the TK subset and genome wide (Figure 1b). Looking at all human genes, also uORF sequences excluding the uAUGs were significantly higher conserved as compared with



**Figure 1.** Presence, conservation and position of uORFs in human tyrosine kinases and all human genes. **(a)** Pie charts indicating the prevalence of uORFs in TK genes and the number of uORFs per transcript for alternative TK mRNAs. **(b)** Bar graph displaying the frequency of conserved bases within TLSs, uORFs and uAUGs of TKs and all human genes. Note that columns labeled 'TLS' do not include uORF bases and columns labeled 'uORF' do not include uAUGs, respectively. For genome-wide transcript analyses, Hg19 annotations were downloaded from the UCSC genome browser database.<sup>15</sup> Of 44 113 entries, we excluded 7794 non-coding RNAs, 1861 sequences not aligning to autosomes or sexual chromosomes, and 1237 transcripts lacking a TLS. Finally, we positively selected the 32 332 mRNAs with unique genome alignments corresponding to 17 893 distinct genes ([ftp://ftp.ncbi.nih.gov/refseq/H\\_sapiens/mRNA\\_Prot/](ftp://ftp.ncbi.nih.gov/refseq/H_sapiens/mRNA_Prot/)). Conservation was defined by mapping the respective genomic positions to the '100 vertebrates conserved elements' provided by UCSC (<http://genome.ucsc.edu/cgi-bin/hgTrackUi?g=cons100way>). Statistical significance was analyzed using the  $\chi^2$  test and indicated as  $***P < 0.01$  and NS (not significant). **(c)** Graph indicating the prevalence of uORFs in respect to the length of the TLS for TKs and all human genes (solid green and black lines, respectively). Dashed lines indicate expected uORF frequencies for TKs and all human genes following a 10 000-times randomization of TLS nucleotides (green and black, respectively). **(d)** Box plots summarizing the variable length of TLSs and uORFs in TK transcripts. The graph indicates the range of nucleotide distances from the transcriptional start site (TSS) to the uAUG and from the uAUG to the main AUG (mAUG).

non-uORF TLS sequences; yet this observation did not hold significance for the smaller gene set of TKs. Extending a previous report that identified AUG as the triplet most frequently conserved among human and mouse,<sup>16</sup> our data based on conserved elements among 100 vertebrate species suggest functional importance of uAUGs and occasionally also of the subsequent uORF sequences.

In all human transcripts and in the TK subset, uORFs were less prevalent than expected, as shown by quantification of uAUG frequencies in natural and randomized TLSs (Figure 1c). These observations are in line with previous findings implying negative selection of uAUGs.<sup>17</sup> Moreover, we observed a high variability of uORF length and position within all transcripts (data not shown) and individual TK-TLSs (Figure 1d and Supplementary Table 1), suggesting that the regulatory potential of a uORF may be highly dependent on the structural vicinity within individual transcripts.

#### Translational repression of TKs by uORFs

Having characterized the high prevalence of uORFs in TK transcripts, we set out to monitor a potential *cis*-regulatory function on downstream translation. A genetically modified luciferase reporter plasmid was generated to facilitate the insertion of complete TLSs, including the gene-specific CDS initiation codon and the core-Kozak nucleotide at position +4 of the main AUG, to retain the endogenous initiation context of individual transcripts (Figure 2a and Supplementary Figure 1). The inserted TLSs either contained the wild-type uORF initiation codon (wt uORF) or a point mutation at the uORF start site (UUG), known to abolish ribosomal initiation ( $\Delta$ uORF).<sup>18</sup> The functionality of the TLS luciferase reporter system was tested by inserting wt or  $\Delta$ uORF versions of the TLSs of *CEBPA*, *CEBPB* and *ERBB2* (Figures 2b–d, positive controls), representing paradigm transcripts with conserved uORFs and previously documented regulatory potential.<sup>19,20</sup> As expected, elevated luciferase activity was detected for the  $\Delta$ uORF-containing reporter constructs, as compared with the wt uORF controls, whereas mRNA abundance was not significantly affected. Hence, the TLS reporter system proved to be suited to determine the regulatory impact of uORFs on downstream translation.

Following the genome-wide identification of potential uORFs in human TK transcripts, 10 TLSs harboring a single uORF were selected to exclude competing impacts of subsequent uORFs within the same TLS during functional analyses. Selection was further based on uORF properties that implied functional importance,<sup>3</sup> including the conservation among vertebrate species and the quality of the Kozak initiation context. Transcripts of *HCK*, *MAP2K3*, *ERBB3* (short and long transcript variants), *LCK*, *MAP2K2*, *TYRO3* and *RET* displayed considerably high conservation of the uORF initiation codon, whereas in the transcripts of *YES1* and *ZAP70* interspecies conservation was lower (Figure 2b). Introduction of wt and  $\Delta$ uORF variants of the 10 selected TK TLSs into the luciferase reporter construct revealed that deletion of the uORF initiation codon enhanced downstream translation in all cases (Figure 2c). Luciferase activity increased >3-fold after removal of conserved uORFs in *HCK*, *MAP2K3*, *ERBB3*-short and *LCK* transcripts. Here, the  $\Delta$ uORF-induced de-repression of translation was similar as observed with the positive controls *CEBPA*, *CEBPB* and *ERBB2*. Of note, deletion of the less conserved uORFs (*ZAP70* and *YES1*) resulted in lower, yet still significant de-repression of reporter activity, implying translational activity also at less conserved uORFs. Quantification of mRNA excluded the possibility that an increase in mRNA abundance predominantly accounted for the higher luciferase activity observed (Figure 2d), although mildly elevated transcript levels of  $\Delta$ uORF TLS versions were seen on some occasions. Similar results of consistently enhanced luciferase translation in response to

functional ablation of the TK uORFs were obtained by using HEK293 instead of HeLa cells (Supplementary Figure 2), suggesting cell type independence.

Next, the analysis was extended to additional non-TK proto-oncogenic transcripts, known to be overexpressed or amplified in human cancer (Figures 2b–d).<sup>21</sup> Two independent ribosomal profiling studies in human and mouse had demonstrated translational activity at specific uORFs of *SKP2*, *MDM2*, *CDK4* and *SHC1*.<sup>2,14</sup> Similar to the observations for TKs, deletion of uORF initiation codons of aforementioned proto-oncogenes resulted in increased downstream translation of their CDS, whereas occasional stabilizing effects on the related mRNAs were much smaller.

Taken together, the results imply that translation of many TKs and other proto-oncogenic proteins is constitutively repressed by uORFs. Beyond previous observations in individual transcripts,<sup>20,22</sup> the universal regulatory effect of uORF deletions in all cases analyzed here support the existence of a widespread translational control mechanism, where proto-oncogenes may gain enhanced expression and tumorigenic potential through loss-of-uORF mutations within their TLSs.

#### Mutation-induced elongation of uORFs sustains translational repression

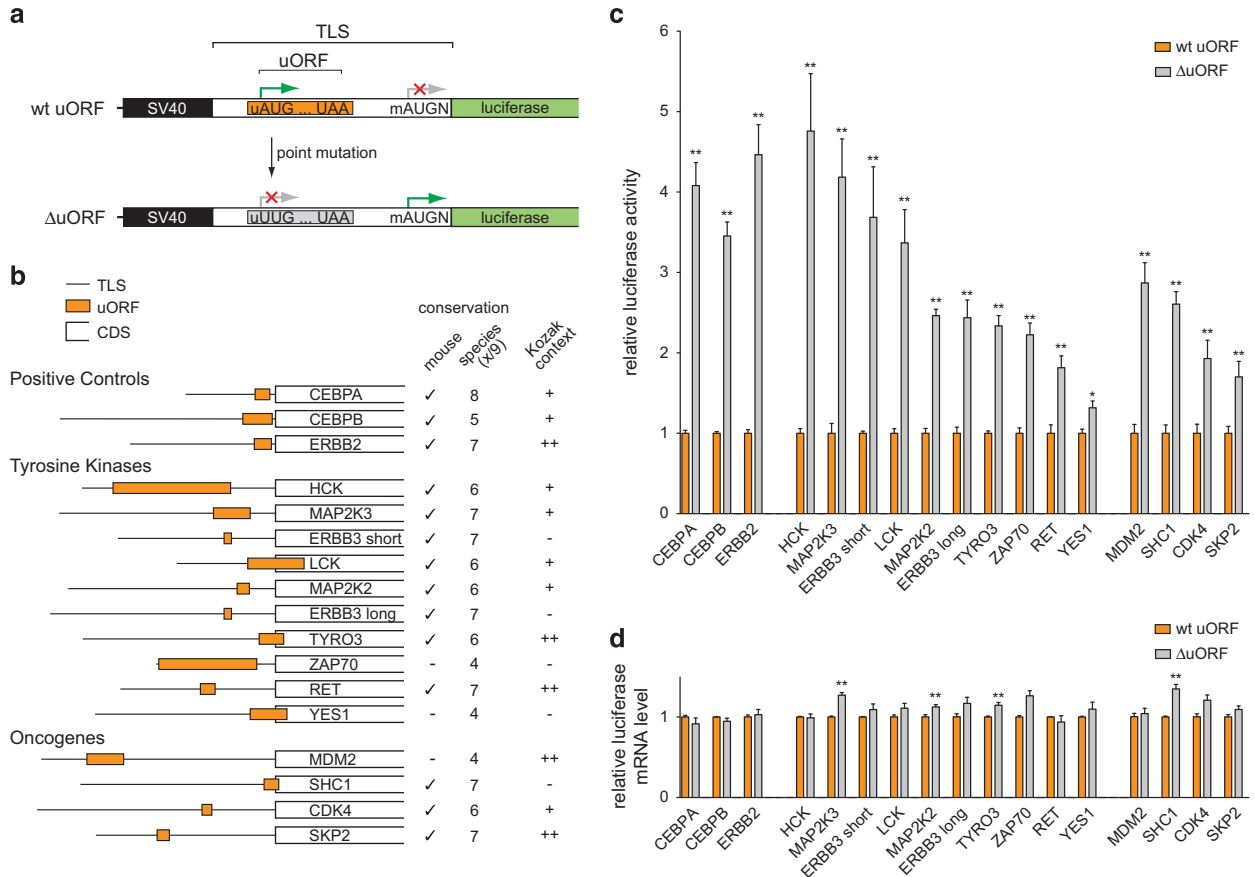
Within the sample set of TKs and proto-oncogenes described above no obvious difference in the regulatory impact of overlapping vs non-overlapping uORFs could be observed. This is in line with a previous report, where naturally occurring overlapping- and non-overlapping uORFs, on average, showed similar inhibitory effects on downstream translation.<sup>3</sup>

To investigate the consequences of de-novo mutations that may convert a non-overlapping into an overlapping uORF, stop codons of the uORFs (uStop) in *CEBPA*, *CEBPB*, *ERBB2*, *HCK*, *MAP2K3* and *RET* were replaced by non-stop CUU codons ( $\Delta$ uStop) that created overlapping uORFs in luciferase constructs and prevented the possibility of ribosomal reinitiation (Figure 3a). In all transcripts analyzed, luciferase translation efficacy was markedly reduced in response to the loss-of-uStop mutations (Figure 3b), suggesting that in the wt TLSs reinitiating ribosomes contribute substantially to the translation of the CDS. Consistently, re-introduction of alternative uStop codons downstream of the natural uStop and upstream of the CDS in *HCK*, *MAP2K3* and *RET* reverted this inhibitory effect to various extents (Figures 3a and b). Of note, for most of the  $\Delta$ uStop versions of the TLS, the luciferase mRNA levels mildly increased as compared with the wt TLS (Figure 3c). As previously reported, reduced nonsense-mediated mRNA decay in the absence of uORF termination codons may account for the observed elevation of respective mRNAs levels.<sup>23</sup>

The observation of reduced CDS translation after deletion of uStop codons in all four TK transcripts suggested that the translation of these and other uORF-bearing TKs may substantially depend on reinitiating ribosomes that previously translated a uORF and terminated at a uStop codon. As translational reinitiation is dependent on the reconstitution of a functional preinitiation complex,<sup>24</sup> TK translation may be highly sensitive to environmental signals and global translational conditions of a cell.

#### Widespread prevalence of natural polymorphic uORFs

To investigate whether naturally occurring sequence variations in TK uORF initiation codons or the surrounding Kozak consensus sequences may alter translational control, we screened TK TLSs for single-nucleotide polymorphisms (SNPs) using the UCSC genome browser<sup>15</sup> and dbSNP.<sup>25</sup> In the TK gene set, we found eight SNPs and one deletion depleting a uAUG ( $\Delta$ puORF). Accordingly, ~5.8% of TKs may be subject to translational variability due to polymorphic uORF initiation codons (puORF; Figure 4a and Supplementary Table 5). In addition, 26 SNPs and 1 deletion in

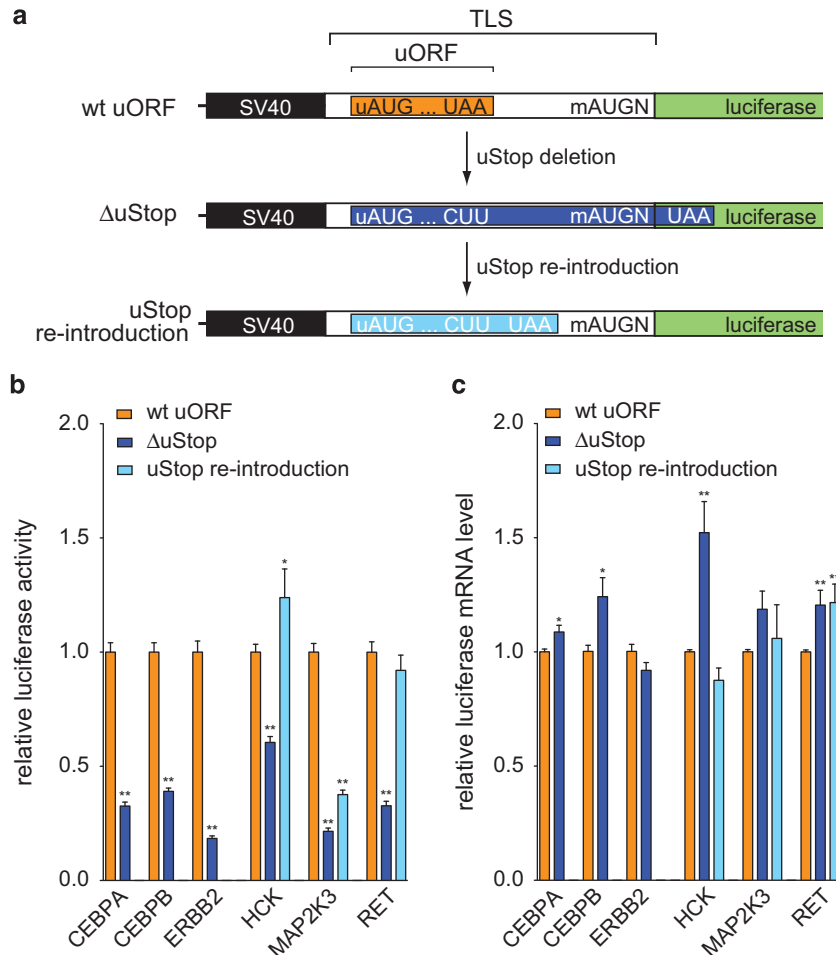


**Figure 2.** uORFs repress downstream translation of TKs and other proto-oncogenes. **(a)** Graphic representation of the luciferase reporter construct. The plasmid allows the insertion of a complete TLS including the endogenous mAUG initiation codon plus the surrounding Kozak sequence including base +4 (N) to retain the gene-specific initiation context. For functional experiments, the wild-type (wt) uORF initiation codon was deleted by the insertion of a point mutation that creates a UUG codon instead ( $\Delta$ uORF). To generate the reporter plasmid, an SV40 promoter originating from the pSG5 vector (Agilent Technologies) was cloned into the pGL3-Basic vector (Promega). Then, the SV40 promoter plus the remainder of the multiple cloning (MCS) site of the pGL3-Basic vector were PCR-amplified and cloned back into the *KpnI* and a newly generated *SacI* restriction site, obtained by mutating the Firefly luciferase initiation codon of pSG5 to GCG. This cloning strategy resulted in a reporter plasmid with the required structure: 5'-SV40 promoter-MCS-luciferase gene-3'. TLSs containing either the wt uORF start site or a respective uUUG mutation ( $\Delta$ uORF) were synthesized by GeneArt Gene Synthesis (Life Technologies) and introduced in-frame with the luciferase coding sequence. (For details on restriction sites and inserted sequences, see Supplementary Figure 1 and Supplementary Table 3 and 4). **(b)** Schematic representation of the position and length of uORFs within the TLSs of three positive control transcripts, 10 selected TKs and 4 other proto-oncogenes. Conservation of the uAUG among human and mouse (check mark) and among a total of nine vertebrate species (human, rhesus, mouse, rat, cow, dog, elephant, chicken, zebrafish) is depicted. The Kozak context is indicated as strong (++) a purine base at -3 and a guanine base at +4, intermediate (+, one of the core Kozak bases matches) or weak (-, no core Kozak base matches). Note that *SKP2*, *MDM2* and *CDK4* contain additional potential uORFs that were excluded from display and functional analysis as they were not consistently translated in ribosome profiling studies. **(c)** Bar graph showing the relative reporter activity in the presence of wt uORF and  $\Delta$ uORF containing TLSs of indicated proteins. For Luciferase assays, HeLa cells were seeded in duplicates and cultured under standard conditions. After 6 h, cells were transfected with 1  $\mu$ g/12-well of the TLS luciferase reporter construct and 30 ng of Renilla luciferase reporter construct (pRL-CMV, Promega) using Metafectene transfection reagent (Biontix). After 42 h, Firefly and Renilla luciferase activities were determined following a published protocol.<sup>32</sup> For each construct, Firefly luciferase signals were normalized to Renilla luciferase internal control signals. **(d)** Bar graph indicating the relative luciferase mRNA levels of wt uORF and  $\Delta$ uORF reporter constructs for indicated TLSs. Quantitative real-time PCRs were performed after RNA extraction and cDNA synthesis following standard protocols and using custom primers that targeted Firefly and Renilla luciferase as an internal control (Supplementary Table 4). Error bars represent the s.e.m. of three independent experiments. Data were analyzed by using GraphPad Prism (Version 5.0a) and statistically evaluated by the two-tailed, nonparametric Mann-Whitney test. Asterisks indicate statistical significance (\*\* $P < 0.01$  and \* $P < 0.05$ ).

14.3% of TK genes were found to affect uORF-related Kozak consensus sequences (pKozak).

For experimental exploration of these SNPs, we focused on *KDR*, where the uORF-deleting SNP (rs7667298) affected one single uORF, and on *MET*, where the second of two uORF initiation codons was altered by two independent SNPs (rs13235174 and rs13222452, Figure 4b). Allele frequencies annotated in the dbSNP were 44.5% vs 55.5% for the puORF vs  $\Delta$ puORF allele of *KDR* and unknown for the *MET* SNPs. The  $\Delta$ puORF allele of *KDR* was associated with mildly enhanced translation of the luciferase

reporter gene (Figure 4c). In the *MET* transcript, the two alternative polymorphic ablations of the second *MET* uORF (UUG and AAG), or a combination of both (UAG), resulted in mild de-repression of downstream translation. Additional deletion of the first *MET* uORF, located upstream and in-frame to the polymorphic uAUG, resulted in further enhancement of reporter expression, but did not alter the regulatory potential of the  $\Delta$ puORF variants observed at the second *MET* uORF (Supplementary Figure 3), suggesting more complex functions of translation initiation control. As we observed slightly higher mRNA levels for some of the polymorphic  $\Delta$ puORF



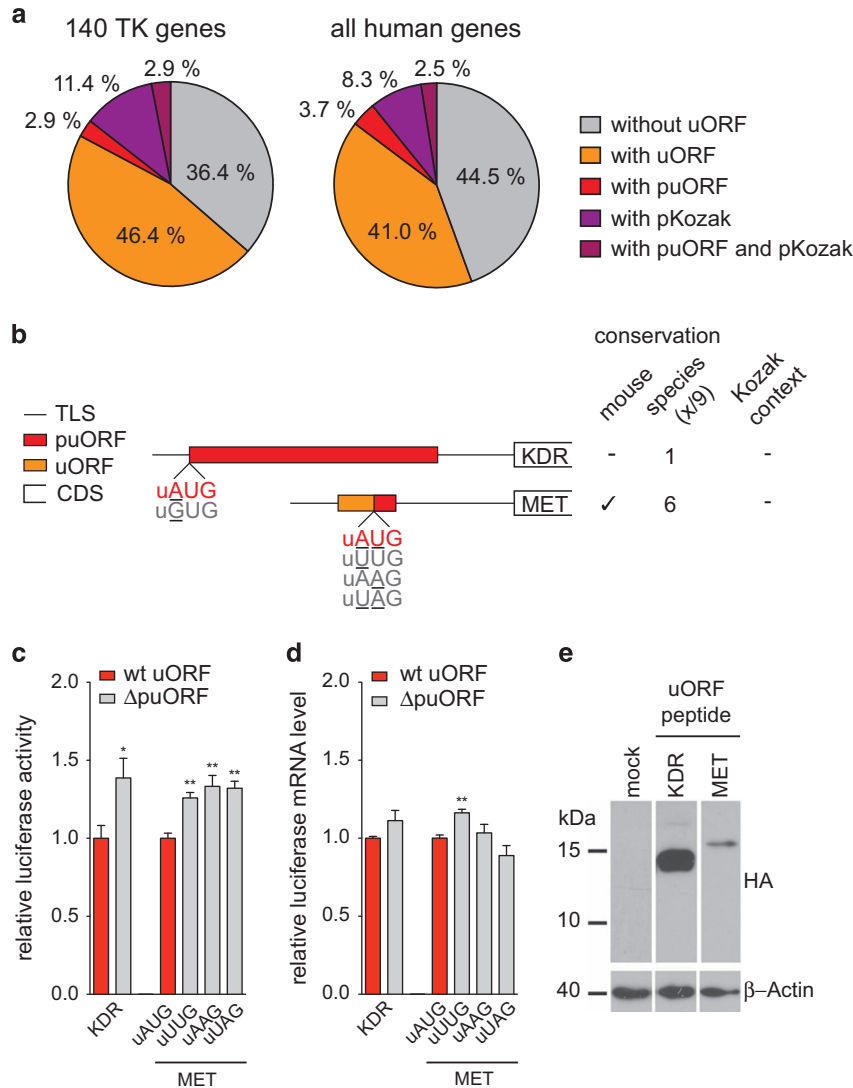
**Figure 3.** Reinitiating ribosomes contribute to TK translation. **(a)** Schematic representation of mutations introduced into the TLS reporter construct to determine the contribution of reinitiating ribosomes to TK translation. The deletion of uORF-related upstream stop codons (uStops to CUU, ΔuStop) created extended uORFs that overlapped the mAUG start site. Second, an alternative uStop codon was re-introduced between the natural uStop and the mAUG codon (uStop re-introduction). For *CEBPA*, *CEBPB* and *ERBB2*, the uStop re-introduction was not performed as the distances between the uStop and the CDS were as short as 7, 4 and 5 nt, respectively. All mutations were introduced by site-directed mutagenesis of the wt uORF version of the respective TLS using the PfuPlusI DNA Polymerase (Roboklon) and customized primers (Supplementary Tables 3 and 4). **(b)** Bar graph representing the relative luciferase activity detected in the presence of wt uORF-, ΔuStop- and re-introduced uStop-containing TLSs. **(c)** Bar graph indicating the relative luciferase mRNA levels of wt uORF, ΔuStop and re-introduced uStop reporter constructs for indicated TLSs. Error bars represent the s.e.m. of at least three independent experiments. Asterisks indicate statistical significance (\*\* $P < 0.01$  and \* $P < 0.05$ ).

versions of the *KDR* and *MET* TLSs, we cannot exclude some contribution of mRNA stabilization to the increased reporter function (Figure 4d and Supplementary Figure 3). Nevertheless, active translation of the *KDR* and *MET* uORFs could be confirmed independently by the detection of C-terminally HA-tagged uORF peptides in immunoblot analyses (Figure 4e).

Irrespective of the mode of translational induction, our data suggest that naturally occurring uORF polymorphisms may alter the expression of the respective downstream proteins. In support of the hypothesis of loss-of-uORF-mediated proto-oncogene activation, the *KDR* ΔpuORF allele was independently found to be associated with increased *KDR* protein levels in lung cancer samples<sup>26</sup> and, together with two additional *KDR* SNPs, with increased risk of glioma development.<sup>27</sup> Furthermore, the ΔpuORF allele alone was associated with an acute course of sarcoidosis<sup>28</sup> and a trend toward shorter overall survival of patients suffering from pancreatic carcinoma.<sup>29</sup> For the *MET* SNP, no clinical association data are available to date.

Given the highly reproducible translational activity of uORFs in TKs and other proto-oncogenes described above, we determined

the genome-wide prevalence of SNPs within uORF start codons and the surrounding Kozak consensus sequences. First, computational sequence analyses generated a current map of all human uORFs and provided information on the transcripts affected, the genomic position and the length of each individual uORF (Supplementary Table 2). Second, we analyzed how many of the 56 248 699 human SNPs listed in dbSNP 137 mapped to specific uORF sequences. We identified 1375 SNPs affecting uAUGs and 2724 SNPs affecting uAUG-related Kozak sequences in 2610 individual genes (Supplementary Table 6). These observations reveal that the translation rates of up to 14.6% of annotated human genes may be dependent on SNPs affecting the uORF initiation context (Figure 4a). In addition, we detected 697 SNPs at uStop codons in 3.4% of the genes (Supplementary Table 6), further increasing the number of proteins whose expression may be subject to inter-individual variability in response to uORF-related SNPs. For eight of the uORF-related SNPs described here, clinical association data have been documented according to the current dbSNP annotations in the UCSC database (Supplementary Table 7). The low number may in part reflect the fact that



**Figure 4.** Polymorphic uORFs affect TK translation. **(a)** Pie charts indicating the fraction of TK genes and all human genes with SNPs in uORF start codons (puORF) and/or corresponding Kozak sequence contexts (pKozak). SNPs were analyzed for all RefSeq-annotated transcript variants of TKs and the whole human genome using dbSNP 137.<sup>25</sup> **(b)** Schematic representation of the position and length of a uORF and puORFs within the TLSs of *KDR* and *MET*. SNPs that disrupt the uORF initiation codon are underlined and the resulting alternative codons are displayed in gray. Conservation of the uAUG among human and mouse (check mark) and among a total of nine vertebrate species is depicted. The weak quality of the Kozak context (–, no core Kozak base matches) is indicated. **(c)** Bar graph representing the relative luciferase activity in the presence of the wt puORF and the ΔpuORF containing TLSs of indicated TKs. For the two *MET* SNPs, all three possible SNP combinations were analyzed. **(d)** Bar graph indicating the relative luciferase mRNA levels of wt puORF and ΔpuORF reporter constructs for indicated TLSs. Error bars represent the s.e.m. of at least three independent experiments. Asterisks indicate statistical significance (\*\* $P < 0.01$  and \* $P < 0.05$ ). **(e)** Immunoblot shows HA-tagged peptides translated from *KDR* and *MET* uORFs. The TLSs of *KDR* and *MET* including the cap-to-uORF sequence, the uORF-coding sequences (with a disrupted uStop codon) and a C-terminally added triple HA-tag (Supplementary Table 4) were cloned into the pcDNA3 vector (Invitrogen). HEK293 cells were transfected following standard protocols and protein lysates were separated 42 h later on 18% SDS-polyacrylamide gels. After transfer, polyvinylidene fluoride membranes (Roth) were probed with specific antibodies: HA (HA.11, Covance) and β-Actin (Clone AC-15, Sigma-Aldrich) and horseradish peroxidase-linked secondary antibody (GE Healthcare Life Sciences).

genome-wide sequence data from clinical samples largely focused on coding exons and excluded uORFs in the TLSs. In addition, dbSNP annotations may be incomplete, as recent publications reported higher numbers of genotype-phenotype associations and potential clinical associations of altered uORF-mediated translational control in various types of diseases.<sup>3,10</sup>

The data described above demonstrate a wide range of regulatory potential of uORF-mutations at initiation and termination codons. Our results demand for systematic searches for loss-of-uORF mutations in proto-oncogenes, and gain-of-uORF or loss-of-uStop mutations in tumor-suppressor genes, as

both types of translational deregulation may cause mis-expression of the respective proteins and may result in dominant phenotypes.<sup>30</sup> Ultimately, ongoing re-sequencing of whole genomes, together with resources implemented to survey functional data from ribosome profiling<sup>31</sup> and individual studies,<sup>4</sup> will help to precisely characterize the contribution of uORF-related genetic variants to the etiology of disease and, more general, to phenotypic divergence. Our data suggest that the translational activation of proto-oncogenic proteins through loss-of-uORF mutations may contribute to increased tumor susceptibility.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**ACKNOWLEDGEMENTS**

This work was supported by the Deutsche Krebshilfe e.V., Bonn, Germany (grant 110525 to KW and AL).

**REFERENCES**

- Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 2009; **324**: 218–223.
- Lee S, Liu B, Huang SX, Shen B, Qian SB. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci USA* 2012; **109**: E2424–E2432.
- Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci USA* 2009; **106**: 7507–7512.
- Wethmar K, Barbosa-Silva A, Andrade-Navarro MA, Leutz A. uORFdb—a comprehensive literature database on eukaryotic uORF biology. *Nucleic Acids Res* 2014; **42** (Database issue): D60–D67.
- Somers J, Poyry T, Willis AE. A perspective on mammalian upstream open reading frame function. *Int J Biochem Cell Biol* 2013; **45**: 1690–1700.
- Wethmar K. The regulatory potential of upstream open reading frames (uORFs) in eukaryotic gene expression. *WIREs RNA* 2014; **5**: 765–778.
- Kozak M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 2002; **299**: 1.
- Occhi G, Regazzo D, Trivellin G, Boaretto F, Ciato D, Bobisse S *et al*. A novel mutation in the upstream open reading frame of the CDKN1B gene causes a MEN4 phenotype. *PLoS Genet* 2013; **9**: e1003350.
- Liu L, Dilworth D, Gao L, Monzon J, Summers A, Lassam N *et al*. Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. *Nat Genet* 1999; **21**: 128–132.
- Barbosa C, Peixeiro I, Romao L. Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet* 2013; **9**: e1003529.
- Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R *et al*. A census of human cancer genes. *Nat Rev Cancer* 2004; **4**: 177–183.
- Krause DS, Van Etten RA. Tyrosine kinases as targets for cancer therapy. *N Engl J Med* 2005; **353**: 172–187.
- Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O *et al*. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* 2014; **42** (D1): D756–D763.
- Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 2010; **466**: 835–840.
- Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M *et al*. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* 2015; **43** (D1): D670–D681.
- Churbanov A, Rogozin IB, Babenko VN, Ali H, Koonin EV. Evolutionary conservation suggests a regulatory function of AUG triplets in 5'-UTRs of eukaryotic genes. *Nucleic Acids Res* 2005; **33**: 5512–5520.
- Iacono M, Mignone F, Pesole G. uAUG and uORFs in human and rodent 5'untranslated mRNAs. *Gene* 2005; **349**: 97–105.
- Ivanov IP, Loughran G, Sachs MS, Atkins JF. Initiation context modulates auto-regulation of eukaryotic translation initiation factor 1 (eIF1). *Proc Natl Acad Sci USA* 2010; **107**: 18056–18060.
- Calkhoven CF, Muller C, Leutz A. Translational control of C/EBPalpha and C/EBP-beta isoform expression. *Genes Dev* 2000; **14**: 1920–1932.
- Spevak CC, Park EH, Geballe AP, Pelletier J, Sachs MS. her-2 upstream open reading frame effects on the use of downstream initiation codons. *Biochem Biophys Res Commun* 2006; **350**: 834–841.
- Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS. A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* 2010; **10**: 59–64.
- Jin X, Turcott E, Englehardt S, Mize GJ, Morris DR. The two upstream open reading frames of oncogene mdm2 have different translational regulatory properties. *J Biol Chem* 2003; **278**: 25716.
- Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* 2004; **36**: 1073–1078.
- Jackson RJ, Hellen CU, Pestova TV. Termination and post-termination events in eukaryotic translation. *Adv Protein Chem Struct Biol* 2012; **86**: 45–93.
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM *et al*. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001; **29**: 308–311.
- Glubb DM, Cerri E, Giese A, Zhang W, Mirza O, Thompson EE *et al*. Novel functional germline variants in the VEGF receptor 2 gene and their effect on gene expression and microvessel density in lung cancer. *Clin Cancer Res* 2011; **17**: 5257–5267.
- Chen H, Wang W, Xingjie Z, Song X, Fan W, Keke Z *et al*. Association between genetic variations of vascular endothelial growth factor receptor 2 and glioma in the Chinese Han population. *J Mol Neurosci* 2012; **47**: 448–457.
- Pabst S, Karpushova A, Diaz-Lacava A, Herms S, Walier M, Zimmer S *et al*. VEGF gene haplotypes are associated with sarcoidosis. *Chest* 2010; **137**: 156–163.
- Uzunoglu FG, Kolbe J, Wikman H, Gungor C, Bohn BA, Nentwich MF *et al*. VEGFR-2, CXCR-2 and PAR-1 germline polymorphisms as predictors of survival in pancreatic carcinoma. *Ann Oncol* 2013; **24**: 1282–1290.
- Wethmar K, Smink JJ, Leutz A. Upstream open reading frames: Molecular switches in (patho)physiology. *BioEssays* 2010; **32**: 885–893.
- Michel AM, Fox G, M Kiran A, De Bo C, O'Connor PB, Heaphy SM *et al*. GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res* 2013; **42**(Database issue): D859–D864.
- Hampf M, Gossen M. A protocol for combined Photinus and Renilla luciferase quantification compatible with protein assays. *Anal Biochem* 2006; **356**: 94–99.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Supplementary Information accompanies this paper on the Oncogene website (<http://www.nature.com/onc>)