Somatic MED12 Nonsense Mutation Escapes mRNA Decay and Reveals a Motif Required for Nuclear Entry

Heikkinen, Tuomas

2017-03

Heikkinen , T , Kämpjärvi , K , Keskitalo , S , von Nandelstadh , P , Liu , X , Rantanen , V , Pitkänen , E , Kinnunen , M , Kuusanmäki , H , Kontro , M , Turunen , M , Mäkinen , N , Taipale , J , Heckman , C , Lehti , K , Mustjoki , S , Varjosalo , M & Vahteristo , P 2017 , ' Somatic MED12 Nonsense Mutation Escapes mRNA Decay and Reveals a Motif Required for Nuclear Entry ' , Human Mutation , vol. 38 , no. 3 , pp. 269-274 . https://doi.org/10.1002/humu.23157

http://hdl.handle.net/10138/348416
https://doi.org/10.1002/humu.23157

acceptedVersion

Somatic *MED12* nonsense mutation escapes mRNA decay and reveals a motif required for nuclear entry

Tuomas Heikkinen[1,2,†], Kati Kämpjärvi[1,2,†], Salla Keskitalo[3,†], Pernilla von Nandelstadh[1,4], Xiaonan Liu[3], Ville Rantanen[1,5], Esa Pitkänen[1, 2], Matias Kinnunen[3], Heikki Kuusanmäki[6,7], Mika Kontro[6], Mikko Turunen[1,4], Netta Mäkinen[1,2], Jussi Taipale[1,4], Caroline Heckman[7], Kaisa Lehti[1,8,9], Satu Mustjoki[6,10], Markku Varjosalo[3], Pia Vahteristo[1,2,*]

Author affiliation

[1]Research Programs Unit, Genome-Scale Biology and [2]Medicum, Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland; [3]Institute of Biotechnology, University of Helsinki, Helsinki, Finland; [4]Department of Pathology, University of Helsinki, Helsinki, Finland; [5]Institute of Biomedicine, University of Helsinki, Helsinki, Finland; [6]Hematology Research Unit Helsinki, Department of Hematology, University of Helsinki and Helsinki University Hospital Comprehensive Cancer Center, Helsinki, Finland; [7]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; [8]Finnish Cancer Institute, Helsinki, Finland; [9]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden; [10]Department of Clinical Chemistry, University of Helsinki, Helsinki, Finland

† These authors contributed equally to the work.

* To whom the correspondence may be addressed.

Correspondence to: Pia Vahteristo, Ph.D., Adjunct Professor
Tel: +358-2941-25600
Fax: +358-2941-25105
E-mail: pia.vahteristo@helsinki.fi

Contract Grant Sponsors: Academy of Finland (Academy Research Fellow grants 260370 and 292769 for PV, 288475 and 294173 for MV; Postdoctoral Researcher grant 295640 for TH), the Sigrid Jusélius Foundation, the Cancer Society of Finland, and the Finnish Cultural Foundation.

ABSTRACT

MED12 is a key component of the transcription regulating Mediator complex. Specific missense and in-frame insertion/deletion mutations in exons 1 and 2 have been identified in uterine leiomyomas, breast tumors, and chronic lymphocytic leukemias. Here, we characterize the first *MED12* 5' end nonsense mutation (c.97G>T, p.E33X) identified in acute lymphoblastic leukemia and show that it escapes nonsense mediated mRNA decay (NMD) by using an alternative translation initiation site. The resulting N-terminally truncated protein is unable to enter the nucleus due to the lack of identified nuclear localization signal (NLS). Absence of NLS prevents the mutant MED12 protein to be recognized by importin-α and subsequent loading into the nuclear pore complex. Due to this mislocalization, all interactions between the MED12 mutant and other Mediator components are lost. Our findings provide new mechanistic insights into the MED12 functions and indicate that somatic nonsense mutations in early exons may avoid NMD.

Key words: MED12, nonsense mutation, nuclear localization signal (NLS), nonsense mediated mRNA decay (NMD), acute lymphoblastic leukemia (ALL), BioID, affinity purification mass spectrometry

Mediator complex is a large protein assembly, which participates in RNA-polymerase II-dependent transcription regulation (Allen and Taatjes, 2015). Mediator complex subunit 12 (MED12; MIM# 300188, NM_005120.2) comprises the Mediator kinase module together with MED13, CDK8/19, and Cyclin C (Clark et al., 2015). *MED12* mutations were first associated with human tumorigenesis when highly specific mutations affecting exon 2 were identified in 70% of uterine leiomyomas (Mäkinen et al., 2011). Mutations have subsequently been identified also in exon 1, and both exon 1 and 2 mutations (henceforth 5' *MED12* mutations) have been shown to result in a unique global gene expression profile, loss of interactions with other kinase module components, and diminished Mediator-associated CDK kinase activity (Mehine et al., 2013; Kämpjärvi et al., 2014; Turunen et al., 2014). Following the original observation, identical mutations have been reported in uterine leiomyosarcomas and in breast fibroadenomas and phyllodes tumors (Pérot et al., 2012; Kämpjärvi et al., 2012; Lim et al., 2014; Cani et al., 2015). In addition to hormone-dependent female solid tumors, similar 5' *MED12* mutations were recently observed in ~5% of chronic lymphocytic leukemias (CLL) (Wang et al., 2011; Kämpjärvi et al., 2015; Guieze et al., 2015). In CLL mutations were shown to associate with poor prognosis markers (Kämpjärvi et al., 2015). All mutations in both solid tumors and CLL have been missense changes or small insertions, deletions or splice-site mutations retaining the reading frame and suggesting an oncogenic nature of the mutations. This is supported by a recent mouse-model study, which reported the most common *MED12* exon 2 mutation, G44D, to promote tumor formation via gain-of-function mechanism (Mittal et al., 2015).

Recently, exome sequencing revealed the first *MED12* mutation in T-cell acute lymphoblastic leukemia (T-ALL) (Kontro et al., 2014). Mutation (c.97G>T, p.E33X) was present in both diagnostic and relapse samples, and it affected the last codon of exon 1. The site has been previously identified as a mutation hotspot in CLL (Kämpjärvi et al., 2015) but not in leiomyomas (Mäkinen et al., 2011; McGuire et al., 2012) or other solid tumors (The Catalogue of Somatic Mutations in Cancer,

COSMIC). In contrast to CLL, where all mutations affecting codon E33 have been missense changes, the mutation in T-ALL is a nonsense mutation and thus predicted to result in a truncated protein product.

The female T-ALL patient with the c.97G>T (p.E33X) mutation has received chemotherapy and hematopoietic stem cell transplantation and is in complete remission. Small amount of blood-derived DNA and RNA were, however, available for small-scale molecular analyses. The *MED12* mutation observed in exome sequencing was validated by Sanger sequencing in genomic DNA (Figure 1A). *MED12* resides on X chromosome and displays monoallelic expression in females due to random inactivation of one of the two X chromosomes. Interestingly, cDNA sequencing showed that the mutant allele of *MED12* is expressed while the wild type is not (Figure 1A, Supp. Figure S1). Alternative transcript lacking the last four nucleotides of exon 1 (r.96_99del) and leading to a premature stop codon at codon 35 was also observed. Transcript is likely caused by an alternative splice site introduced by the c.97G>T mutation. This frameshift mutation may lead to the production of a short, likely unstable protein product of 34 amino acids in length, or it may result in bypassing the NMD. RNA sequencing data available from earlier studies (Kontro et al., 2014) showed that the r.96_99del was present in one sequence read whereas the mutant r.97T was present in the remaining 46 reads. Neither cDNA nor RNA sequencing showed the wild-type allele confirming that it was not expressed in the patient sample.

To further analyze the expression and possible functional effects of the E33X mutation, site-directed mutagenesis was utilized to introduce the *MED12* c.97G>T (p.E33X) and c.97G>A (p.E33K; CLL hotspot mutation) mutations into a pTO_HA_StrepIII_c_GW_FRT expression vector containing a wild type *MED12* cDNA (Turunen et al., 2014). Stable Flp-In 293 T-REx cell lines expressing MED12 E33X, E33K, G44D (uterine leiomyoma hotspot mutation) (Turunen et al., 2014) and wild

type (WT) MED12 were created. Tagged MED12 derivatives have previously been shown to attain induced expression level comparable to the level of endogenous MED12 in Flp-In 293 T-REx cells (Turunen et al., 2014). All cell lines expressed C-terminally HA-tagged MED12 in Western blot, but the E33X mutant protein showed clearly smaller molecular weight than the WT or missense mutant derivatives (Figure 1B). With an N-terminal anti-MED12 antibody, no signal was observed for MED12 E33X while other derivatives showed similar expression as the WT MED12 protein (Figure 1B). Peptide identification using mass spectrometry confirmed that while the WT construct produces a full-length MED12 protein, the N-terminal region is missing in the E33X mutant (Figure 1C, Table S1). Together these results indicate that the c.97G>T (p.E33X) mutation escapes nonsense-mediated mRNA decay, a conserved surveillance mechanism normally preventing the production of truncated proteins, and is translated into a truncated MED12 protein without the N-terminal region. The closest potential downstream translation initiation sites in human *MED12* are codons M154 and M162, both showing strong evolutionary conservation in vertebrates (Supp. Figure S2). To analyze which of these is used in the translation of E33X mutant, we disrupted these sites using site-directed mutagenesis and created stable Flp-In 293 T-REx cell lines with MED12 E33X_M154A, E33X_M162A, and E33X_M154A_M162A. Only the cells with intact M154 codon (E33X_M162A) produced HA-tagged protein (Supp. Figure S3) presenting M154 as the probable translation start site for the E33X mutant.

The intriguing mechanism, an escape from NMD through the use of an alternative translation initiation site, has only rarely been reported in the literature. Specifically, it has been described in the context of hereditary conditions such as β-thalassemia (*β-globin*) (Neu-Yilik et al., 2011), attenuated adenomatous polyposis coli (*APC*) (Heppner Goss et al., 2002), and hereditary predisposition to retinoblastoma (*RB1*) (Sanchez-Sanchez et al., 2007), where individuals with such mutations have shown milder or even healthy phenotype. In a recent study, NMD escape was analyzed in cancer

using public sequencing and expression data of almost 10 000 tumors from The Cancer Genome Atlas (Lindeboom et al., 2016). NMD escape through the use of an alternative translation start site was observed as a relatively common mechanism in the context of somatic 5' terminal nonsense mutations which, in our study, is experimentally shown for *MED12* c.97G>T (p.E33X). In general, somatic nonsense and frameshift mutations are very rare in *MED12*, none being identified in tumor types with recurrent *MED12* mutations. In support of the critical role of MED12 expression, most cells do not tolerate almost complete suppression of MED12 (Huang et al., 2012), and *Med12*-null mouse embryos are not able to develop beyond embryonic day 7.5 (Rocha et al., 2010).

The physical effects of the E33X mutation were further analyzed by proteome-wide affinity purification mass spectrometry (AP-MS). These analyses showed that all interactions between the truncated MED12 and other Mediator components were abolished (Figure 1D, Table S2). This is strikingly different compared to the missense mutations G44D and E33K, which only affected interactions between MED12 and kinase module components and interactions with other Mediator components remained intact (this study and Turunen et al., 2014). These results indicate that the E33X mutation may, in addition to the effects on the kinase module, impact also other functional properties of MED12 and/or the Mediator complex.

Immunofluorescence staining was then used to study whether the extensive changes observed in protein-protein interactions result from the mislocalization of the MED12 E33X mutant protein. In contrast to the full-length MED12 derivatives, which showed predominant nuclear localization, N-terminally truncated protein was mainly found in the cytoplasm (Figure 2A). Nucleus/cytoplasm signal intensity ratios of MED12 clearly differed between MED12 WT and E33X (Figure 2B, Table S3).

Four *in silico* prediction tools (SeqNLS, PSORT II, cNLS Mapper, and NLStradamus) were utilized to analyze the 154 amino acid region lacking in E33X derivative for the presence of a nuclear localization signal (NLS). All algorithms predicted an NLS to exist between amino acids 13 and 19, a region with strong evolutionary conservation (Figure 2C). Two constructs disrupting the predicted NLS (NLS1, amino acids 13-16 PLKR replaced with AAAA and NLS2, amino acids 13-19 PLKRPRL replaced with AAAAAAA) were generated and stable Flp-In 293 T-REx cell lines created. In Western blot, both NLS constructs displayed similar sized proteins as the WT MED12 and immunofluorescence staining showed that the nuclear localization of the NLS mutant proteins is prevented similarly to the E33X mutation (Supp. Figure S4). When the MED12 NLS domain (amino acids 11-20; HRPLKRPRLG) was cloned into a GFP containing vector and transfected into HeLa cells, all GFP with MED12-NLS construct was localized in the nucleus while the GFP without this domain was distributed evenly across the cell (in 100/100 cells analyzed for each construct, Supp. Figure S5). AP-MS analysis further showed that the interactions between NLS mutant MED12 and all other Mediator complex components were severely disrupted (Supp. Figure S6). These data indicate that the predicted NLS is functional and required for nuclear entry.

To obtain more insight into the molecular mechanisms of MED12 nuclear entry, MED12 constructs (wild type and mutants) were cloned into BioID-vector (pTO_BioID_c_myc_GW_FRT). BioID approach allows subsequent biotinylation of proteins in the bait proximity and thus detection of even transient molecular interactions (Roux et al., 2012) (Supp. Figure S7). MS analysis of the purified biotinylated interactors showed lost or decreased binding of E33X and NLS mutants to all Mediator components (Figure 4A, Table S4), in concordance with the results from AP-MS. Interestingly, BioID analysis also revealed interactions of MED12 with importins and several nuclear pore complex (NPC) components. When comparing the E33X mutant with the wild type, lost or severely decreased interactions between the E33X and the main components of NPC nuclear basket NUP50, NUP153

and TPR were observed (Figure 4B). Diminished interactions between the E33X and importin-α-proteins (IMA1, IMA4 and IMA7) were also observed, whereas the interaction with importin-β (IMB1) was retained (Figure 4A, B). Similarly decreased interactions were observed between the NLS mutants and the NPC nuclear basket components (NUP50, NUP153 and TPR) as well as importin-α-proteins when compared to the wild type. In addition, increased interactions of NLS mutants with importin-β and cytoplasmic components of the NPC (RBP2/NUP358, NUP214, NUP88), i.e. cytoplasmic ring (Raices and D'Angelo, 2012) were detected. These results suggest that due to the decreased importin-α binding, MED12 mutants fail to enter the NPC central channel governed by the cargo assembly platform RBP2 (Tran et al., 2014) and receptor-mediated import docking site NUP214 (Forler et al., 2004) (Figure 4C). In accordance with AP-MS results, BioID analysis verified that missense mutations (E33K and G44D) affect only the interactions between MED12 and the Mediator kinase module components CDK8/19 and Cyclin C (CCNC) (Supp. Figure S8).

While Mediator has a key role in regulating transcription, its structure and functions have been difficult to study due to large size and complexity. Mediator components have also only a few known functional domains, making them challenging to approach. Thus the identification of specific *MED12* mutations in human tumors has provided a valuable tool to characterize the molecular basis of tumor development as well as for bringing new knowledge on the normal function of the protein (Turunen et al., 2014). In the absence of clear functional domains, enrichment of particular amino acid residues has been used to divide the MED12 protein sequence into L domain (leucine-rich; residues 1-500), LS domain (leucine-, and serine rich; residues 501-1650), PQL domain (proline-, glutamine-, and leucine-rich; residues 1651-2086), and opposite paired (OPA) domain (glutamine-rich; residues 2087-2212) (Zhou et al., 2002). Thus far, the only characterized functional sites in MED12 have been the N-terminal binding site for Cyclin C (Turunen et al., 2014), and SOX9 (Zhou et al., 2002), β-

catenin (Kim et al., 2006), and Gli3 (Zhou et al., 2006) binding regions on PQL. Here, we describe another functional site as a region required for nuclear localization (residues 11-20; HRPLKRPRLG) was identified. This finding opens novel opportunities to study the cytoplasmic functions of MED12. Thus far, the only reported cytoplasmic functions of MED12 are related to drug response through physical interaction with TGF-βR2 (Huang et al., 2012). Even though TGF-βR2 was not listed as a significant interaction partner of MED12 in our MS analyses, it cannot be excluded that these mutations do not affect TGF-βR2 binding. There may also be other, thus far unidentified, cytoplasmic MED12 functions which are necessary for the cells.

It is challenging to evaluate the impact of *MED12* mutations on T-ALL tumorigenesis based on a single case. In addition to the *MED12* mutation, the patient had nonsynonymous mutations in known leukemia genes *NOTCH1*, *KRAS*, *SUZ12*, *KDM6A*, and *STAT5B14* (Kontro et al., 2014). Interestingly, however, the *MED12* mutation affects the same amino acid which has been previously identified as a mutational hotspot in CLL (Kämpjärvi et al., 2015). *MED12* mutation has also been suggested as a single somatic founder mutation in CLL (Schuh et al., 2012) and also here it was present in both the diagnostic and relapse samples. As impaired interactions between MED12 and Mediator kinase module components CDK8/19 and Cyclin C have been shown to diminish the RNA polymerase II C-terminal directed kinase activity (Turunen et al., 2014), it is evident that this kinase activity is similarly abolished for the truncated MED12 E33X derivative, either due to its inability to enter the nucleus or the missing region required for the interactions. Aberrant CDK8 kinase activity may well play a role in tumorigenesis, and in addition to Pol II CTD, products of known leukemia genes *NOTCH1* and *STAT* family members, among others, have been indicated as plausible biological substrates for the CDK8 phosphorylation (reviewed in Clark et al., 2015). Interestingly, MED12 was recently reported as an essential regulator of normal hematopoiesis further supporting *MED12* mutations as putative drivers of tumorigenesis (Aranda-Orgilles et al., 2016).

Taken together, our results indicate that somatic nonsense mutations in early exons of *MED12* may escape NMD by using an alternative translation start site. A novel functional site at the N-terminus of MED12 was identified, as the nuclear localization signal was discovered and experimentally validated. This information can be utilized to further study the cytoplasmic functions of MED12. In addition, new information on the mechanism and the molecules required for MED12 nuclear entry were revealed. These results emphasize the value of individual mutations in identifying and understanding the protein functions and the underlying molecular mechanisms in tumorigenesis.

ACKNOWLEDGEMENTS

We would like to thank Annukka Ruokolainen, Inga-Lill Svedberg, Alison Ollikainen, and Sini Miettinen for technical assistance and the Biomedicum Imaging Unit for imaging facilities. The authors have no conflicts of interest to declare.

FIGURE LEGENDS

Figure 1: *MED12* c.97G>T (p.E33X) mutation is expressed and leads to the production of an N-terminally truncated protein product. (A) Sequence chromatogram of *MED12* c.97G>T (p.E33X) mutation on genomic DNA of the T-ALL sample (left upper panel). cDNA sequencing of the mutant sample shows that the mutant T-allele is expressed while the WT G-allele is not (right upper panel). Expression of an alternative transcript (r.96_99del) was also observed (see also Supp. Figure S1). Sequence chromatograms of the WT reference samples are presented in the lower panel. **(B)** MED12 constructs (WT, E33K, E33X, G44D) with C-terminal HA-tag were transfected into Flp-In 293 T-REx cells. Western-blot analysis with anti-HA-antibody shows that all MED12 derivatives are expressed, but the protein product of E33X construct is smaller than other derivatives (upper panel).

An N-terminal anti-MED12-antibody (sc-5374; epitope between aa25 and aa75) identifies other derivatives but does not bind to the E33X mutant product (lower panel). Vinculin was used as a loading control. **(C)** Peptide identification with mass spectrometry reveals that the WT MED12 (green) peptides are identified from the beginning of the protein, whereas the first peptide recognized from the E33X derivative (red) is at amino acid positions 163-174. Alternative translation start site used in E33X derivative, M154, is marked with bold red.

Figure 2: MED12 E33X protein loses interactions with all Mediator components and is unable to enter the nucleus. (A) Affinity purification mass spectrometry shows that all interactions between the MED12 E33X derivative and other Mediator complex components are lost. Missense mutations E33K and G44D result in reduced interactions between the mutant MED12 and Mediator kinase module components CDK8, CDK19, and Cyclin C (CCNC), while interactions with other Mediator components remain intact. Similar results have been previously shown for G44D (Turunen et al., 2014). Error bars represent ± standard deviation (SD). **(B)** Immunofluorescence staining of HA-tagged MED12 derivatives shows that while the WT, G44D, and E33K products localize in the nucleus, the E33X product remains in the cytoplasm. Scale bar, 20 μm. **(C)** Density plot of MED12 immunofluorescence signal quantification (R function density with parameter bw = 0.03). **(D)** Four independent *in silico* tools predict nuclear localization signal between MED12 amino acids 13 and 19. The domain predicted by the algorithms shows strong evolutionary conservation. **(E)** BioID analysis identifies interactions between MED12 and proteins involved in nuclear import. Compared to the WT MED12, E33X mutant shows decreased interactions with importin-α proteins (IMA1, IMA4, and IMA7), and diminished interactions with NPC nuclear ring proteins (NUP50, NUP153, and TPR). NLS mutants show similar lowered interactions as the E33X, and in addition display increased interactions with NPC cytoplasmic ring proteins (RBP2, NUP214, and NUP88). This suggests that the mutant proteins fail to enter the NPC central channel and progress into the nucleus.

Error bars represent ± SD. Asterisks denote statistical difference of individual mutant versus WT (*p < 0.05, **p < 0.01, ***p < 0.001). **(F)** Schematic presentation of the normal MED12 nuclear import (green) and the import failure of E33X and NLS mutants (red).

REFERENCES

Allen BL, Taatjes DJ. 2015. The mediator complex: A central integrator of transcription. Nat Rev Mol Cell Biol 16:155-166.

Aranda-Orgilles B, Saldana-Meyer R, Wang E, Trompouki E, Fassl A, Lau S, Mullenders J, Rocha PP, Raviram R, Guillamot M, Sanchez-Diaz M, Wang K et al. 2016. MED12 regulates HSC-specific enhancers independently of mediator kinase activity to control hematopoiesis. Cell Stem Cell In Press.

Cani AK, Hovelson DH, McDaniel AS, Sadis S, Haller MJ, Yadati V, Amin AM, Bratley J, Bandla S, Williams PD, Rhodes K, Liu CJ et al. 2015. Next-gen sequencing exposes frequent MED12 mutations and actionable therapeutic targets in phyllodes tumors. Mol Cancer Res 13:613-619.

Clark AD, Oldenbroek M, Boyer TG. 2015. Mediator kinase module and human tumorigenesis. Crit Rev Biochem Mol Biol 50:393-426.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797.

Forler D, Rabut G, Ciccarelli FD, Herold A, Kocher T, Niggeweg R, Bork P, Ellenberg J, Izaurralde E. 2004. RanBP2/Nup358 provides a major binding site for NXF1-p15 dimers at the nuclear pore complex and functions in nuclear mRNA export. Mol Cell Biol 24:1155-1167.

Guieze R, Robbe P, Clifford R, de Guibert S, Pereira B, Timbs A, Dilhuydy MS, Cabes M, Ysebaert L, Burns A, Nguyen-Khac F, Davi F et al. 2015. Presence of multiple recurrent mutations confers poor trial outcome of relapsed/refractory CLL. Blood 126:2110-2117.

Heppner Goss K, Trzepacz C, Tuohy TM, Groden J. 2002. Attenuated APC alleles produce functional protein from internal translation initiation. Proc Natl Acad Sci U S A 99:8161-8166.

Huang S, Holzel M, Knijnenburg T, Schlicker A, Roepman P, McDermott U, Garnett M, Grernrum W, Sun C, Prahallad A, Groenendijk FH, Mittempergher L et al. 2012. MED12 controls the response to multiple cancer drugs through regulation of TGF-beta receptor signaling. Cell 151:937-950.

Kämpjärvi K, Järvinen TM, Heikkinen T, Ruppert AS, Senter L, Hoag KW, Dufva O, Kontro M, Rassenti L, Hertlein E, Kipps TJ, Porkka K et al. 2015. Somatic MED12 mutations are associated with poor prognosis markers in chronic lymphocytic leukemia. Oncotarget 6:1884-1888.

Kämpjärvi K, Mäkinen N, Kilpivaara O, Arola J, Heinonen HR, Böhm J, Abdel-Wahab O, Lehtonen HJ, Pelttari LM, Mehine M, Schrewe H, Nevanlinna H et al. 2012. Somatic MED12 mutations in uterine leiomyosarcoma and colorectal cancer. Br J Cancer 107:1761-1765.

Kämpjärvi K, Park MJ, Mehine M, Kim NH, Clark AD, Bützow R, Böhling T, Böhm J, Mecklin JP, Järvinen H, Tomlinson IP, van der Spuy ZM et al. 2014. Mutations in exon 1 highlight the role of MED12 in uterine leiomyomas. Hum Mutat 35:1136-1141.

Kim S, Xu X, Hecht A, Boyer TG. 2006. Mediator is a transducer of Wnt/beta-catenin signaling. J Biol Chem 281:14066-14075.

Kontro M, Kuusanmäki H, Eldfors S, Burmeister T, Andersson EI, Bruserud O, Brummendorf TH, Edgren H, Gjertsen BT, Itala-Remes M, Lagstrom S, Lohi O et al. 2014. Novel activating STAT5B mutations as putative drivers of T-cell acute lymphoblastic leukemia. Leukemia 28:1738-1742.

Kosugi S, Hasebe M, Tomita M, Yanagawa H. 2009. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc Natl Acad Sci U S A 106:10171-10176.

Lim WK, Ong CK, Tan J, Thike AA, Ng CC, Rajasegaran V, Myint SS, Nagarajan S, Nasir ND, McPherson JR, Cutcutache I, Poore G et al. 2014. Exome sequencing identifies highly recurrent MED12 somatic mutations in breast fibroadenoma. Nat Genet 46:877-880.

Lin JR, Hu J. 2013. SeqNLS: Nuclear localization signal prediction based on frequent pattern mining and linear motif scoring. PLoS One 8:e76864.

Lindeboom RG, Supek F, Lehner B. 2016. The rules and impact of nonsense-mediated mRNA decay in human cancers. Nat Genet 48:1112-1118.

Mäkinen N, Mehine M, Tolvanen J, Kaasinen E, Li Y, Lehtonen HJ, Gentile M, Yan J, Enge M, Taipale M, Aavikko M, Katainen R et al. 2011. MED12, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. Science 334:252-255.

Mattila OS, Strbian D, Saksi J, Pikkarainen TO, Rantanen V, Tatlisumak T, Lindsberg PJ. 2011. Cerebral mast cells mediate blood-brain barrier disruption in acute experimental ischemic stroke through perivascular gelatinase activation. Stroke 42:3600-3605.

McGuire MM, Yatsenko A, Hoffner L, Jones M, Surti U, Rajkovic A. 2012. Whole exome sequencing in a random sample of north american women with leiomyomas identifies MED12 mutations in majority of uterine leiomyomas. PLoS One 7:e33251.

Mehine M, Kaasinen E, Mäkinen N, Katainen R, Kämpjärvi K, Pitkänen E, Heinonen HR, Bützow R, Kilpivaara O, Kuosmanen A, Ristolainen H, Gentile M et al. 2013. Characterization of uterine leiomyomas by whole-genome sequencing. N Engl J Med 369:43-53.

Mittal P, Shin YH, Yatsenko SA, Castro CA, Surti U, Rajkovic A. 2015. Med12 gain-of-function mutation causes leiomyomas and genomic instability. J Clin Invest 125:3280-3284.

Nakai K, Horton P. 1999. PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci 24:34-36.

Neu-Yilik G, Amthor B, Gehring NH, Bahri S, Paidassi H, Hentze MW, Kulozik AE. 2011. Mechanism of escape from nonsense-mediated mRNA decay of human beta-globin transcripts with nonsense mutations in the first exon. RNA 17:843-854.

Nguyen Ba AN, Pogoutse A, Provart N, Moses AM. 2009. NLStradamus: A simple hidden markov model for nuclear localization signal prediction. BMC Bioinformatics 10:202.

Pérot G, Croce S, Ribeiro A, Lagarde P, Velasco V, Neuville A, Coindre JM, Stoeckle E, Floquet A, MacGrogan G, Chibon F. 2012. MED12 alterations in both human benign and malignant uterine soft tissue tumors. PLoS One 7:e40015.

Raices M, D'Angelo MA. 2012. Nuclear pore complex composition: A new regulator of tissue-specific and developmental functions. Nat Rev Mol Cell Biol 13:687-699.

Rantanen V, Valori M, Hautaniemi S. 2014. Anima: Modular workflow system for comprehensive image data analysis. Front Bioeng Biotechnol 2:25.

Rocha PP, Scholze M, Bleiss W, Schrewe H. 2010. Med12 is essential for early mouse development and for canonical wnt and Wnt/PCP signaling. Development 137:2723-2731.

Roux KJ, Kim DI, Raida M, Burke B. 2012. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell Biol 196:801-810.

Sanchez-Sanchez F, Ramirez-Castillejo C, Weekes DB, Beneyto M, Prieto F, Najera C, Mittnacht S. 2007. Attenuation of disease phenotype through alternative translation initiation in low-penetrance retinoblastoma. Hum Mutat 28:159-167.

Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R, Feller SM, Grocock R, Henderson S, Khrebtukova I, Kingsbury Z, Luo S et al. 2012. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. Blood 120:4191-4196.

Tran EJ, King MC, Corbett AH. 2014. Macromolecular transport between the nucleus and the cytoplasm: Advances in mechanism and emerging links to disease. Biochim Biophys Acta 1843:2784-2795.

Turunen M, Spaeth JM, Keskitalo S, Park MJ, Kivioja T, Clark AD, Mäkinen N, Gao F, Palin K, Nurkkala H, Vähärautio A, Aavikko M et al. 2014. Uterine leiomyoma-linked MED12 mutations disrupt mediator-associated CDK activity. Cell Rep 7:654-660.

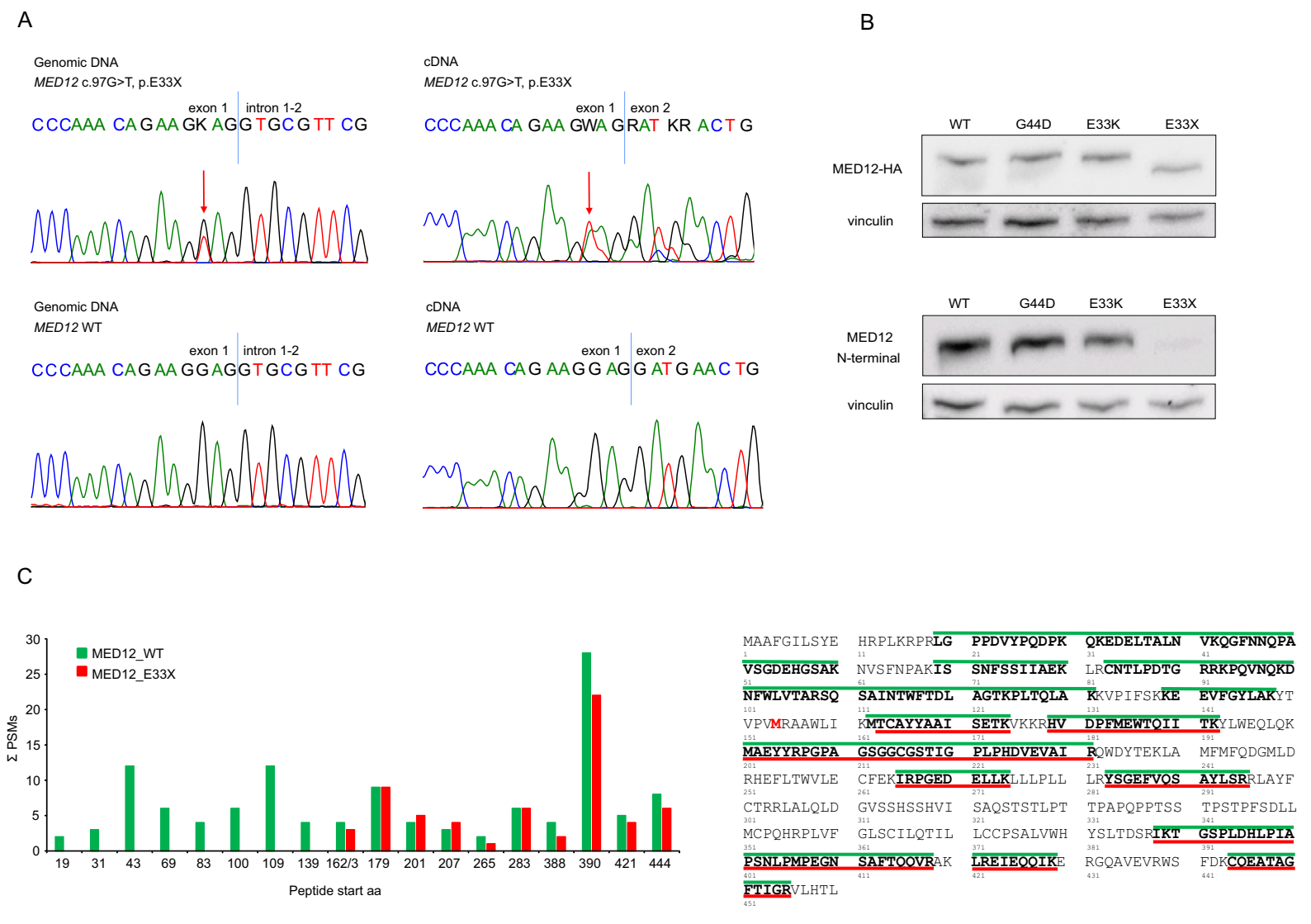
Varjosalo M, Keskitalo S, Van Drogen A, Nurkkala H, Vichalkovski A, Aebersold R, Gstaiger M. 2013a. The protein interaction landscape of the human CMGC kinase group. Cell Rep 3:1306-1320.

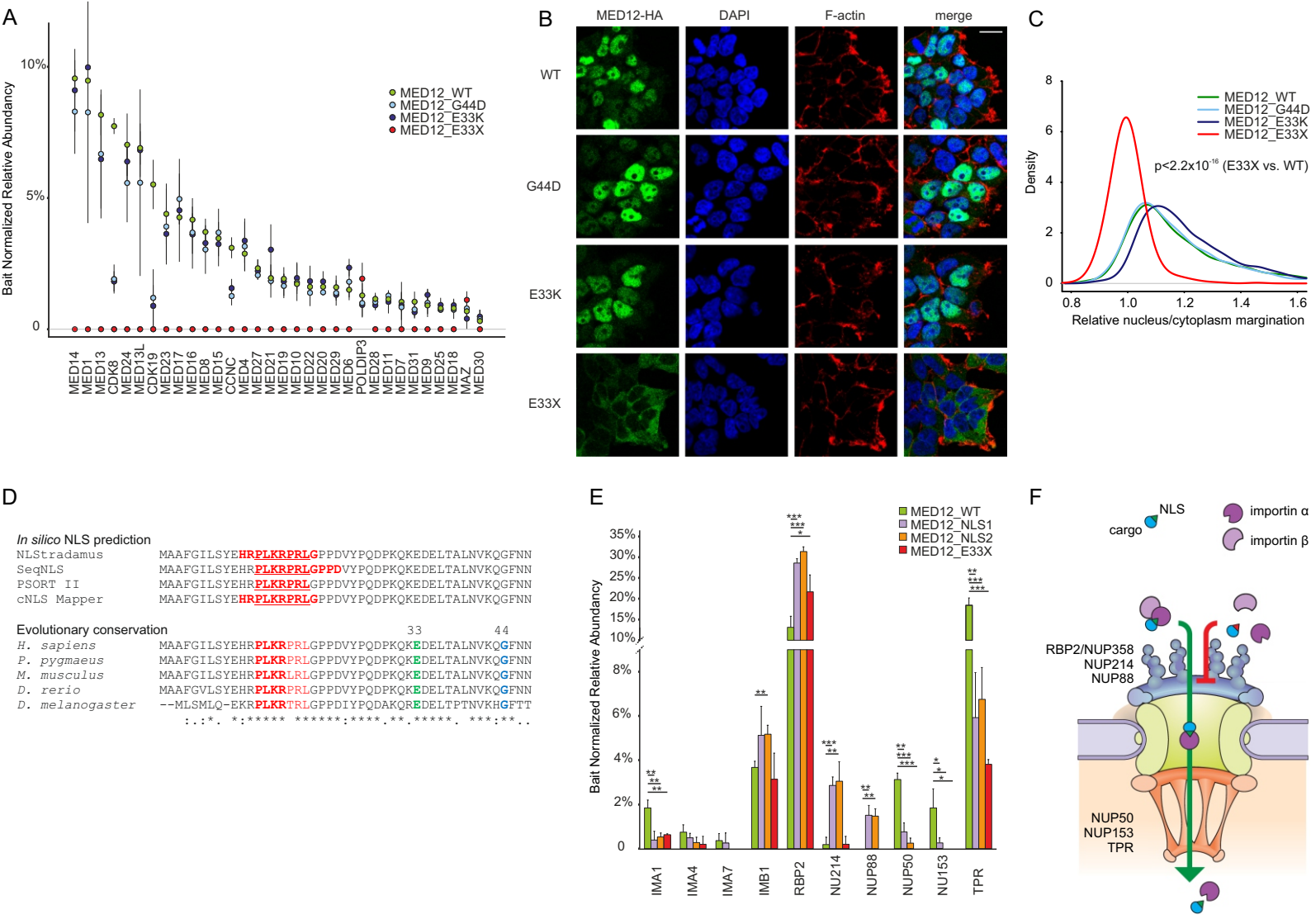
Varjosalo M, Sacco R, Stukalov A, van Drogen A, Planyavsky M, Hauri S, Aebersold R, Bennett KL, Colinge J, Gstaiger M, Superti-Furga G. 2013b. Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. Nat Methods 10:307-314.

Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L, Zhang W, Vartanov AR et al. 2011. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med 365:2497-2506.

Zhou H, Kim S, Ishii S, Boyer TG. 2006. Mediator modulates Gli3-dependent sonic hedgehog signaling. Mol Cell Biol 26:8667-8682.

Zhou R, Bonneaud N, Yuan CX, de Santa Barbara P, Boizet B, Schomber T, Scherer G, Roeder RG, Poulat F, Berta P. 2002. SOX9 interacts with a component of the human thyroid hormone receptor-associated protein complex. Nucleic Acids Res 30:3245-3252.


A
Genomic DNA
MED12 c.97G>T, p.E33X
exon 1
intron 1-2
CCCAAA CA G AA GK AGG TGC G TT CG
cDNA
MED12 c.97G>T, p.E33X
exon 1
exon 2
CCCAAA CA GAA GWA GRAT KR ACT G
Genomic DNA
MED12 WT
exon 1
intron 1-2
CCCAAA CA G AA GG AGG TGC G TT CG
cDNA
MED12 WT
exon 1
exon 2
CCCAAA CA G AA G G AG G AT G AA C TG
B
WT
G44D
E33K
E33X
MED12-HA
vinculin
WT
G44D
E33K
E33X
MED12
N-terminal
vinculin
C
MED12_WT
MED12_E33X
Σ PSMs
0
5
10
15
20
25
30
19
31
43
69
83
100
109
139
162/3
179
201
207
265
283
388
390
421
444
Peptide start aa
MAAFGILSYE HRPLKRPRLG PPDVYPQDPK QKEDELTALN VKQGFNNQPA
VSGDEHGSAK NVSFNPAKIS SNFSSIIAEK LRCNTLPDTG RRKPQVNQKD
NFWLVTARSQ SAINTWFTDL AGTKPLTQLA KKVPIFSKKE EVFGYLAKYT
VPVMRAAWLI KMTCAYYAAI SETKVKKRHV DPFMEWTQII TKYLWEQLQK
MAEYYRPGPA GSGGCGSTIG PLPHDVEVAI RQWDYTEKLA MFMFQDGMLD
RHEFLTWVLE CFEKIRPGED ELLKLLLPLL LRYSGEFVQS AYLSRRLAYF
CTRRLALQLD GVSSHSSHVI SAQSTSTLPT TPAPQPPTSS TPSTPFSDLL
MCPQHRPLVF GLSCILQTIL LCCPSALVWH YSLTDSRIKT GSPLDHLPIA
PSNLPMPEGN SAFTQQVRAK LREIEQQIKE RGQAVEVRWS FDKCQEATAG
FTIGRVLHTL


A
Bait Normalized Relative Abundancy
10%
5%
0
MED12_WT
MED12_G44D
MED12_E33K
MED12_E33X
MED14
MED1
MED13
CDK8
MED24
MED13L
CDK19
MED23
MED17
MED16
MED8
MED15
CCNC
MED4
MED27
MED21
MED19
MED10
MED22
MED20
MED29
MED6
POLDIP3
MED28
MED11
MED7
MED31
MED9
MED25
MED18
MAZ
MED30
B
MED12-HA
DAPI
F-actin
merge
WT
G44D
E33K
E33X
C
Density
8
6
4
2
0
MED12_WT
MED12_G44D
MED12_E33K
MED12_E33X
p<2.2x10-16 (E33X vs. WT)
0.8
1.0
1.2
1.4
1.6
Relative nucleus/cytoplasm margination
D
In silico NLS prediction
NLStradamus MAAFGILSYEHRPLKRPRLGPPDVYPQDPKQKEDELTALNVKQGFNN
SeqNLS MAAFGILSYEHRPLKRPRLGPPDVYPQDPKQKEDELTALNVKQGFNN
PSORT II MAAFGILSYEHRPLKRPRLGPPDVYPQDPKQKEDELTALNVKQGFNN
cNLS Mapper MAAFGILSYEHRPLKRPRLGPPDVYPQDPKQKEDELTALNVKQGFNN
Evolutionary conservation
33
44
H. sapiens MAAFGILSYEHRPLKRPRLGPPDVYPQDPKQKEDELTALNVKQGFNN
P. pygmaeus MAAFGILSYEHRPLKRPRLGPPDVYPQDPKQKEDELTALNVKQGFNN
M. musculus MAAFGILSYEHRPLKRLRLGPPDVYPQDPKQKEDELTALNVKQGFNN
D. rerio MAAFGVLSYEHRPLKRPRLGPPDVYPQDPKQKEDELTALNVKQGFNN
D. melanogaster --MLSMLQ-EKRPLKRTRLGPPDIYPQDAKQREDELTPTNVKHGFTT
E
Bait Normalized Relative Abundancy
35%
30%
25%
20%
15%
10%
8%
6%
4%
2%
0
MED12_WT
MED12_NLS1
MED12_NLS2
MED12_E33X
IMA1
IMA4
IMA7
IMB1
RBP2
NU214
NUP88
NUP50
NU153
TPR
F
cargo
NLS
importin α
importin β
RBP2/NUP358
NUP214
NUP88
NUP50
NUP153
TPR

Supporting Information

Somatic *MED12* nonsense mutation escapes mRNA decay and reveals a motif required for nuclear entry

Tuomas Heikkinen[†], Kati Kämpjärvi[†], Salla Keskitalo[†], Pernilla von Nandelstadh, Xiaonan Liu, Ville Rantanen, Esa Pitkänen, Matias Kinnunen, Heikki Kuusanmäki, Mika Kontro, Mikko Turunen, Netta Mäkinen, Jussi Taipale, Caroline Heckman, Kaisa Lehti, Satu Mustjoki, Markku Varjosalo, Pia Vahteristo[*]

[†] Equal contribution

[*] Corresponding author: Pia Vahteristo, Ph.D., Adjunct Professor; Research Programs Unit, Genome-Scale Biology and Medicum, Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland; Tel: +358-2941-25600; E-mail: pia.vahteristo@helsinki.fi

1. Supp. Materials and Methods
2. Supp. Figures S1-9
3. Supp. Tables S1-5 (Supp. Tables S2 and S4 as separate files)

Supp. Materials and methods

Sequencing of the patient samples

Genomic DNA isolated from a bone marrow sample of a T-ALL patient (Kontro et al., 2014) was PCR amplified and bi-directionally sequenced to validate the presence of the *MED12* c.97G>T (p.E33X) mutation. An RNA sample of the T-ALL with *MED12* c.97G>T (p.E33X) was converted to cDNA, PCR amplified and bi-directionally sequenced. The sequence traces were analyzed using Mutation Surveyor software (SoftGenetics) and by visual inspection. *MED12* NM_005120.2 was used as a cDNA reference sequence, where the A of the ATG translation initiation codon was used as +1 in nucleotide numbering. The primer sequences are presented in Supp. Table S5. Study has been approved by the ethics committee of the Hospital District of Helsinki and Uusimaa, Helsinki, Finland.

Creating MED12 expressing Flp-In 293 T-REx cell lines

The Flp-In 293 T-REx cells (Invitrogen, Life Technologies) were cultured prior to transfection in Dulbecco's modified Eagle medium (Biowhittaker ® DMEM 4.5 g/L glucose; Lonza) supplemented with fetal bovine serum (10%), L-alanyl-L-glutamine (2 mmol/L), penicillin (50 mg/mL), and streptomycin (50 mg/mL). Cell line was authenticated with Promega GenePrint10 System (Genomics Unit of Technology Centre, Institute for Molecular Medicine Finland, FIMM) and tested negative for mycoplasma contamination using Venor®GeM Mycoplasma PCR Detection Kit (Minerva Biolabs). *MED12* mutant plasmids were created using QuickChange Site-Directed Mutagenesis kit (Agilent Technologies) to pTO_HA_StrepIII_c_GW_FRT expression vector (Varjosalo et al., 2013a) and to Gateway pDONR221 entry clone containing wild type *MED12* cDNA construct (Turunen et al., 2014). The mutagenesis primers are listed in Table S5. The *MED12* constructs were further subcloned from Gateway pDONR221 vector into pTO_MYC_BirA_c vector. All the generated constructs were confirmed with direct sequencing. The pTO_HA_StrepIII_c_GW_FRT_MED12 and pTO_MYC_BirA_c_MED12 constructs were transfected into Flp-In 293 T-REx cells with FuGENE® HD Transfection Reagent (Promega) together with pOG44 Flp recombinase expression vector (Thermo Fisher Scientific). The cells were grown under selection with Hygromycin B (Thermo Fisher Scientific) and Blasticidin S HCL (Thermo Fisher Scientific) DMEM to create stable cell line.

Western Blot

The Flp-In 293 T-REx cells were induced with 1 µg/ml doxycycline (MP Biomedicals) for 24 hr and lysed using RIPA buffer (Sigma-Aldrich) with HALT^{tm} Protease inhibitors (Thermo Fisher Scientific) and the protein concentration was determined using a Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). Proteins were separated on pre-casted 10% Tris–HCl polyacrylamide gels (Mini-PROTEAN® TGXTM Gels) and transferred to PVDF membranes using Trans-Blot® Turbo^{TM} Transfer System (all from Bio-Rad Laboratories). The proteins were detected by overnight incubation with mouse anti-HA.11 monoclonal antibody (16B12; Covance), or goat anti-MED12 (anti-TRAP230, sc-5374, Santa Cruz Biotechnology), and rabbit anti-Vinculin (sc-5573, Santa Cruz Biotechnology) was used as a loading control. Secondary antibodies (Anti-mouse IgG A4416, Anti-goat IgG A5420, and Anti-rabbit IgG A6154; Sigma–Aldrich) were incubated at room temperature for 1 hr. The Western blots were detected using ChemiDoc^{TM} MP System (Bio-Rad Laboratories).

Affinity Purification and BioID

For each pull-down approximately 5×10^7 cells (5 x 15 cm dishes) in three biological replicates were induced with 1 µg/ml doxycycline (MP Biomedicals) for 24 hr (for BioID additional 50 µM biotin was added). After induction, cells were washed with 0.1 mM $MgCl_2$, 0.1 mM $CaCl_2$ in PBS and harvested in 1 mM EDTA-PBS. Cells were pelleted by centrifugation 400 g, 5 min, 4 °C. Strep-tag affinity purification was performed as described in (Turunen et al., 2014). For BioID experiments harsher lysis buffer and conditions were used. Cell pellets were lysed on ice for 10 min in HNN-buffer supplemented with 0.5% NP-40, 1.5 mM Na_3VO_4, 1.0 mM PMSF (phenylmethanesulfonylfluoride), 10 µl/ml protease inhibitor cocktail (Sigma-Aldrich), 0.1% SDS and 80 U/ml Benzonase Nuclease (Santa Cruz Biotechnology). Incubation was followed by three cycles of 3 min sonication and 5 min break. Lysate was centrifuged twice at 16 000 g for 20 min at 4 °C to remove the insoluble material. During centrifugation Bio-Spin chromatography columns (Bio-Rad Laboratories) were loaded with Strep-Tactin Sepharose beads (400 µl 50% Slurry; IBA GmbH), and washed once with 1 ml HNN-buffer containing 0.5% NP-40 and inhibitors (HNN-wash buffer). Cleared lysate was loaded on spin columns followed by three washes with 1 ml ice-cold HNN-wash buffer and four washes with 1 ml ice-cold HNN-

buffer without supplements. Bound proteins were eluted with 2 x 300 µl freshly prepared 0.5 mM D-biotin (Thermo Fisher Scientific) in HNN-buffer into a fresh 2 ml tube.

Mass spectrometry

For liquid chromatography-mass spectrometry (LC-MS/MS) samples were prepared as follows: cysteine bonds were reduced with 5 mM Tris (2-carboxyethyl)phosphine (TCEP) (Sigma-Aldrich) for 20 min at 37 °C and alkylated with 10 mM iodoacetamide (Fluka, Sigma-Aldrich) for 20 min at room temperature in the dark. A total of 1 µg trypsin (Promega) was added and samples digested overnight at 37 °C. Samples were quenched with 10% trifluoroacetic acid (TFA) and purified with C-18 Micro SpinColumns (The Nest Group) eluting the samples to 0.1% TFA in 50% acetonitrile (ACN). Samples were dried by vacuum concentration and peptides were reconstituted in 30 µl buffer A (0.1% TFA and 1% ACN in LC-MS grade water) and vortexed thoroughly.

LC-MS/MS analysis was performed on an Orbitrap Elite ETD hybrid mass spectrometer using the Xcalibur version 2.2 SP 1.48 coupled to EASY-nLCII-system (all from Thermo Fisher Scientific) via a nanoelectrospray ion source. 6 µl and 5 µl of peptides were loaded from Strep-Tag and BioID-samples, respectively. Samples were separated using a two-column setup consisting of a C18 trap column (EASY-Column™ 2cm x 100 µm, 5 µm, 120 Å, Thermo Fisher Scientific), followed by C18 analytical column (EASY-Column™ 10 cm x 75 µm, 3 µm, 120 Å, Thermo Fisher Scientific). Peptides were eluted from the analytical column with a 60 min linear gradient from 5 to 35% buffer B (buffer A: 0.1% FA, 0.01% TFA in 1% acetonitrile; buffer B: 0.1% FA, 0.01% TFA in 98% acetonitrile). This was followed by 5 min 80% buffer B, 1 min 100% buffer B followed by 9 min column wash with 100% buffer B at a constant flow rate of 300 nl/min. Analysis was performed in data-dependent acquisition mode where a high resolution (60 000) FTMS full scan (m/z 300-1 700) was followed by top20 CID-MS2 scans (energy 35) in ion trap. Maximum fill time allowed for the FTMS was 200 ms (Full AGC target 1 000 000) and 200 ms for the ion trap (MSn AGC target 50 000). Precursor ions with more than 500 ion counts were allowed for MSn. To enable the high resolution in FTMS scan preview mode was used.

SEQUEST search algorithm in Proteome Discoverer software (Thermo Fisher Scientific) was used for peak extraction and protein identification with the human reference proteome of UniProtKB/SwissProt database (www.uniprot.org). Allowed error tolerances were 15 ppm and 0.8 Da for the precursor and fragment ions, respectively. Database searches were limited to fully tryptic peptides allowing one missed cleavage (in peptide mapping semi-tryptic with one missed cleavage allowed), and carbamidomethyl +57 021 Da (C) of cysteine residue was set as fixed, and oxidation of methionine +15 995 Da (M) as dynamic modifications. For peptide identification FDR was set to <0.05.

The high confidence protein-protein interactions were identified using stringent filtering against control contaminant database. The bait normalized relative protein abundances (% to the bait) were calculated from the spectral counts. Each average and SD was calculated from 3 biological replicates. Statistical difference of the nuclear pore protein abundance of individual mutant versus WT were calculated with Student's t-test.

Immunofluorescence

Flp-In 293 T-REx cells containing HA-tagged wild type or mutant *MED12* construct were plated on Poly-L-lysine hydrobromide (Sigma-Aldrich) coated coverslips. The expression of the construct gene was induced with doxycycline (MP Biomedicals) and after 24 hr the cells were fixed with 4% paraformaldehyde in PBS. HA.11 antibody and Alexa-488 goat anti-mouse IgG (Thermo Fisher Scientific) were used for staining of the cells. F-actin was visualized with Alexa Fluor®568 Phalloidin (Thermo Fisher Scientific). Vectashield® Mounting Media (Vector Laboratories) containing DAPI (4', 6-diamidino-2-phenylindole) to visualize the nuclei was used to mount the coverslips. The stained cells were visualized with Axioplan 2 upright epifluorescence microscope and imaged with LSM 780 confocal microscope with Plan-neofluar 40x, 1.3 NA oil objective (Carl Zeiss). Photoshop CS5.1 (Adobe) was applied for adjustment of the brightness and contrast with a combination of two serial optical sections used for image display.

The image analysis pipeline was developed on Anima platform (Rantanen et al., 2014). Cell nuclei were detected from a DAPI staining using a graph-cuts-based segmentation. The cell cytoplasm was defined as a 20 pixel wide ring around the nucleus. Intensities of HA-tagged MED12 staining within the nuclei and the

cytoplasms were measured using the masks. Margination, or the difference in nuclear and cytoplasmic intensity, was calculated as explained by Mattila et al. (Mattila et al., 2011). R software, version 3.0.2 (R Foundation for Statistical Computing), was utilized for statistical analysis. Margination differences between cells expressing MED12 WT and mutants were calculated using Wilcoxon Rank Sum test.

NLS Predictions and Sequence Alignments

To evaluate the presence of Nuclear Localization Signals on the N-terminal region of 154 amino acids (lost through c.97G>T, p.E33X mutation), four independent *in silico* prediction tools were applied. The amino acid sequence was analyzed using SeqNLS (Lin and Hu, 2013), PSORT II (Nakai and Horton, 1999), cNLS Mapper (Kosugi et al., 2009) with default settings and NLStradamus (Nguyen Ba et al., 2009) with prediction cutoff value of 0.1. The N-terminal region of MED12 was aligned to evaluate the evolutionary conservation of potential alternative translation start sites (Supp. Figure S2) and predicted NLS (Figure 2D) over different species. The protein sequences were retrieved from Uniprot database (www.uniprot.org) for *Homo sapiens* (Q93074), *Pongo pygmaeus* (Q5RCU2), Mus musculus (A2AGH6), *Danio rerio* (Q2QCI8), and *Drosophila melanogaster* (Q9VW47) and aligned using MUSCLE multiple sequence alignment tool (Edgar, 2004).

NLS Validation

To validate the ability of the identified NLS to localize a protein into the nucleus the motif was cloned into a GFP containing vector and transfected into HeLa cells. The vector backbones were generated by synthesizing MED12-NLS sequence (amino acids 11-20; HRPLKRPRLG), which was digested with restriction enzymes and inserted before coding part of eGFP into pcDNA5/FRT/TO/GW backbone (Varjosalo et al., 2013b). For transfection, HeLa cells were plated on coverslips in a 4-well plate (Thermo Fisher Scientific), grown to 70% density and transfected using HelaFect (OZbiosciences) reagent according to the manufacturer's protocol. For immunofluorescence staining, HeLa cells were fixed 24 hours after transfection with 4% PFA. F-actin was visualized with Alexa Fluor®568 Phalloidin (Thermo Fisher Scientific). Vectashield® Mounting Media (Vector Laboratories) containing DAPI (4', 6-diamidino-2-phenylindole) to visualize the nuclei was used to mount the coverslips. Images were acquired with an Orca-Flash4.0 V2 sCMOS camera (Hamamatsu) on an upright Leica DM6000B fluorescent microscope (Leica). All images were taken with a 63x objective.

References for Supp. Materials and Methods

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797.

Kontro M, Kuusanmaki H, Eldfors S, Burmeister T, Andersson EI, Bruserud O, Brummendorf TH, Edgren H, Gjertsen BT, Itala-Remes M, Lagstrom S, Lohi O et al. 2014. Novel activating STAT5B mutations as putative drivers of T-cell acute lymphoblastic leukemia. Leukemia 28:1738-1742.

Kosugi S, Hasebe M, Tomita M, Yanagawa H. 2009. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc Natl Acad Sci U S A 106:10171-10176.

Lin JR, Hu J. 2013. SeqNLS: Nuclear localization signal prediction based on frequent pattern mining and linear motif scoring. PLoS One 8:e76864.

Mattila OS, Strbian D, Saksi J, Pikkarainen TO, Rantanen V, Tatlisumak T, Lindsberg PJ. 2011. Cerebral mast cells mediate blood-brain barrier disruption in acute experimental ischemic stroke through perivascular gelatinase activation. Stroke 42:3600-3605.

Nakai K, Horton P. 1999. PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci 24:34-36.

Nguyen Ba AN, Pogoutse A, Provart N, Moses AM. 2009. NLStradamus: A simple hidden markov model for nuclear localization signal prediction. BMC Bioinformatics 10:202.

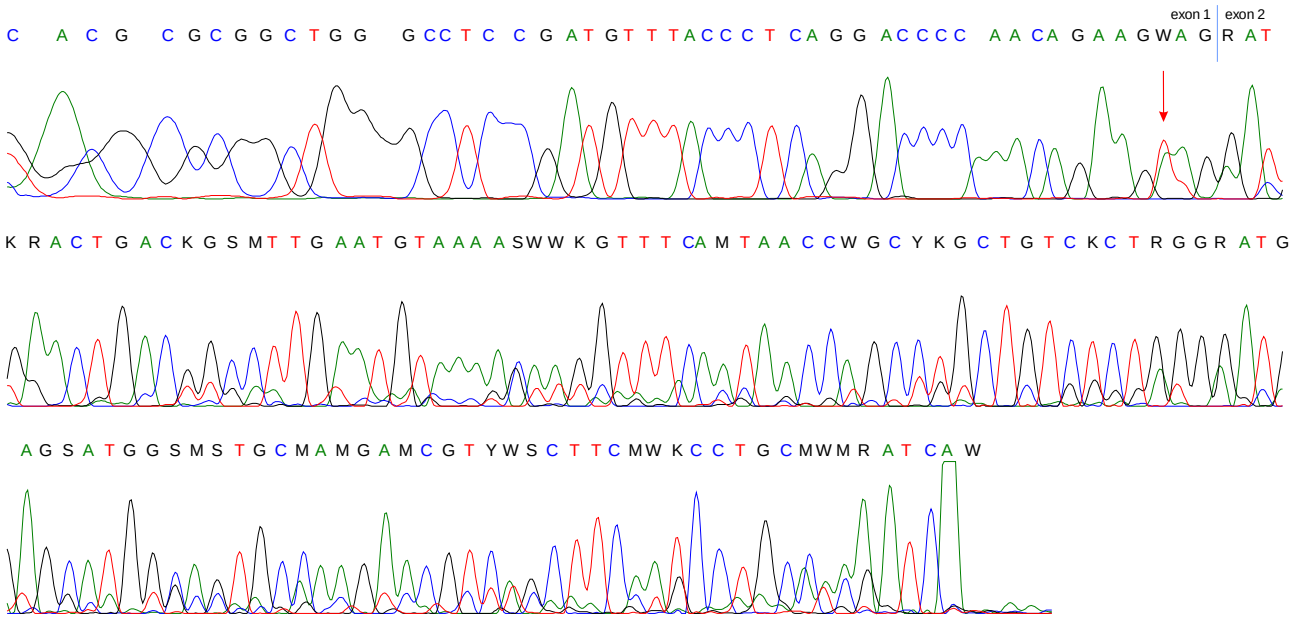
Rantanen V, Valori M, Hautaniemi S. 2014. Anima: Modular workflow system for comprehensive image data analysis. Front Bioeng Biotechnol 2:25.

Turunen M, Spaeth JM, Keskitalo S, Park MJ, Kivioja T, Clark AD, Makinen N, Gao F, Palin K, Nurkkala H, Vaharautio A, Aavikko M et al. 2014. Uterine leiomyoma-linked MED12 mutations disrupt mediator-associated CDK activity. Cell Rep 7:654-660.

Varjosalo M, Keskitalo S, Van Drogen A, Nurkkala H, Vichalkovski A, Aebersold R, Gstaiger M. 2013a. The protein interaction landscape of the human CMGC kinase group. Cell Rep 3:1306-1320.

Varjosalo M, Sacco R, Stukalov A, van Drogen A, Planyavsky M, Hauri S, Aebersold R, Bennett KL, Colinge J, Gstaiger M, Superti-Furga G. 2013b. Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. Nat Methods 10:307-314.

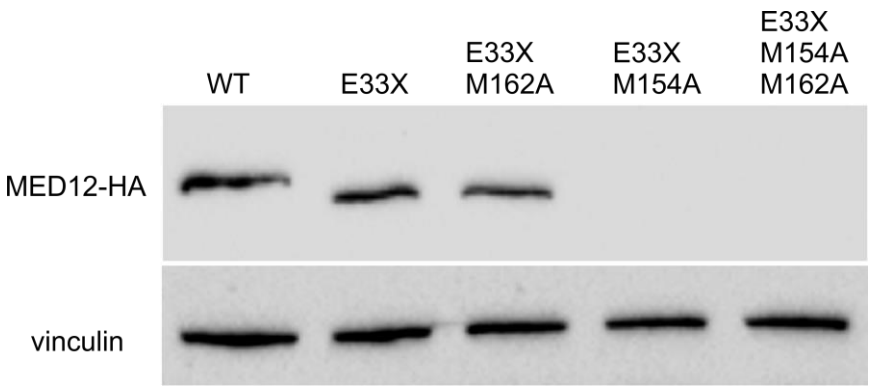
Supp. Figures



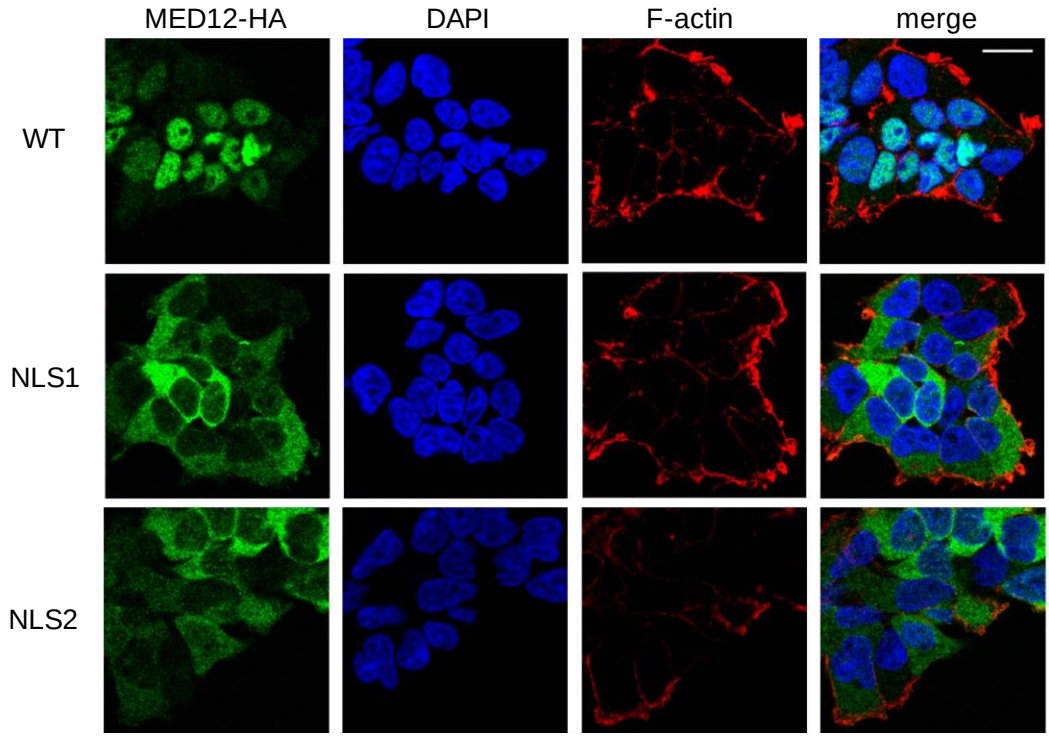
Supp. Figure S1. Full length sequence chromatogram of cDNA sequencing of the T-ALL patient sample with *MED12* c.97G>T (p.E33X) mutation. The mutation site is indicated with a red arrow.

```
                                                    154     162
H. sapiens          AGTKPLTQLAKKVPIFSKKEEVFGYLAKYTVPVMRAAWLIKMTCAYYAAISETKVKKR
P. pygmaeus         AGTKPLTQLAKKVPIFSKKEEVFGYLAKYTVPVMRAAWLIKMTCAYYAAISETKVKKR
M. musculus         AGTKPLTHLAKKVPIFSKKEEVFGYLAKYTVPVMRAAWLIKMTCAYYAAMSETKVKKK
D. rerio            AGTKPLTQLAKKVPIFSKKEEVFGYLAKYSVPVMRSAWMIKMTCAYHAAITENKMKKR
D. melanogaster     AGNKPLLSLAKRAPSFNKKEEIFITLCENQVNMQRATWFIKLSAAYTLSFTESKNKKR
                    **.***  ***..* *.****:*  *.:  * : *::*:**::.**  :::*.* **.
```

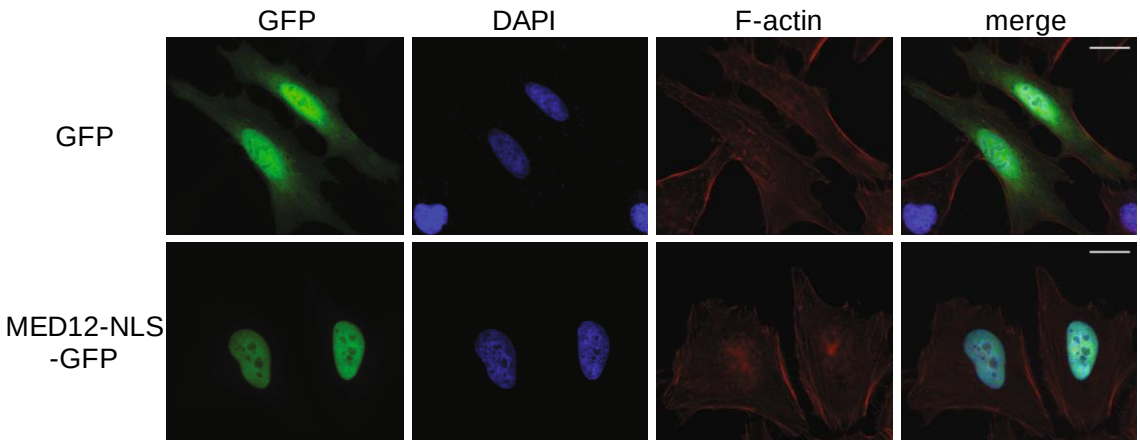
Supp. Figure S2. Species conservation of the region with potential alternative translation start sites (M154 and M162). Both sites are highly conserved among vertebrates.



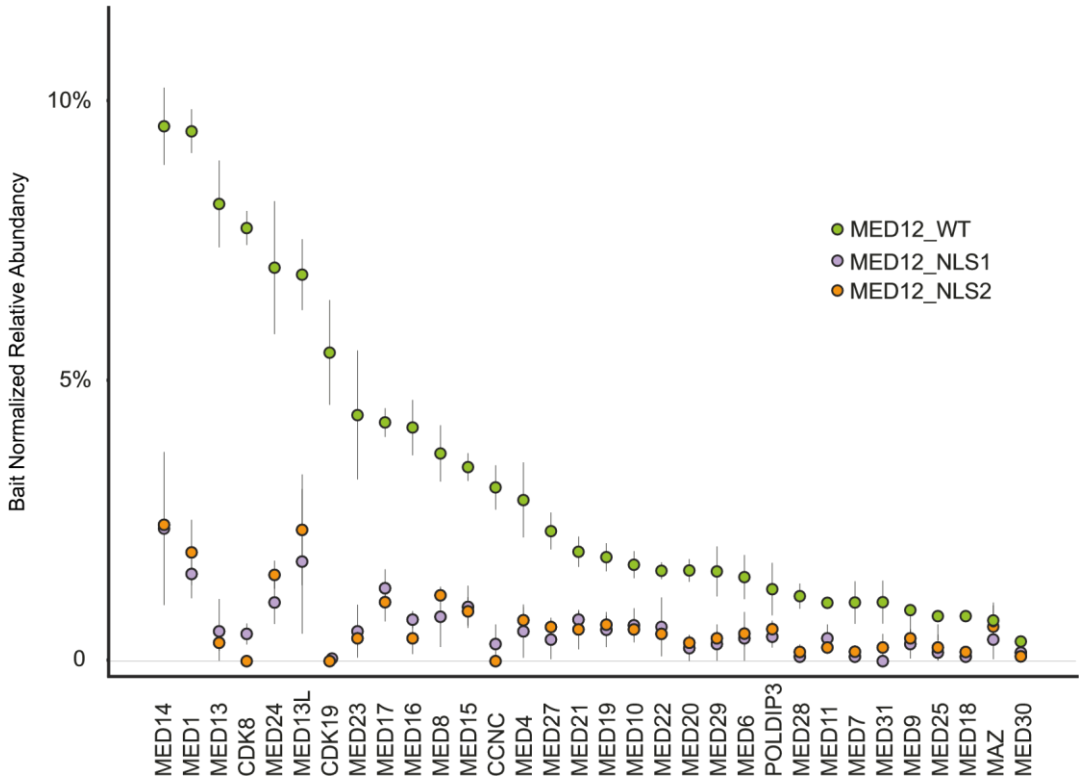
Supp. Figure S3. Western blot image of the MED12 derivatives with alternative translation start sites mutated together with E33X. Only M154 is able to function as a translation start site.



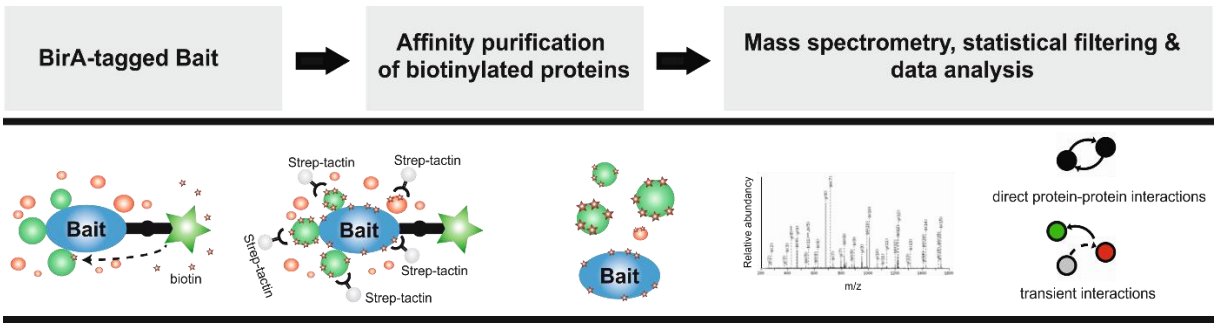
Supp. Figure S4. Immunofluorescence staining of MED12 NLS1 and NLS2 alanine substitution mutant derivatives. Similarly to E33X mutation the NLS mutations also prevent the protein from entering the nucleus in FlpIn 293 T-REx cells. Scale bar, 20 µm.



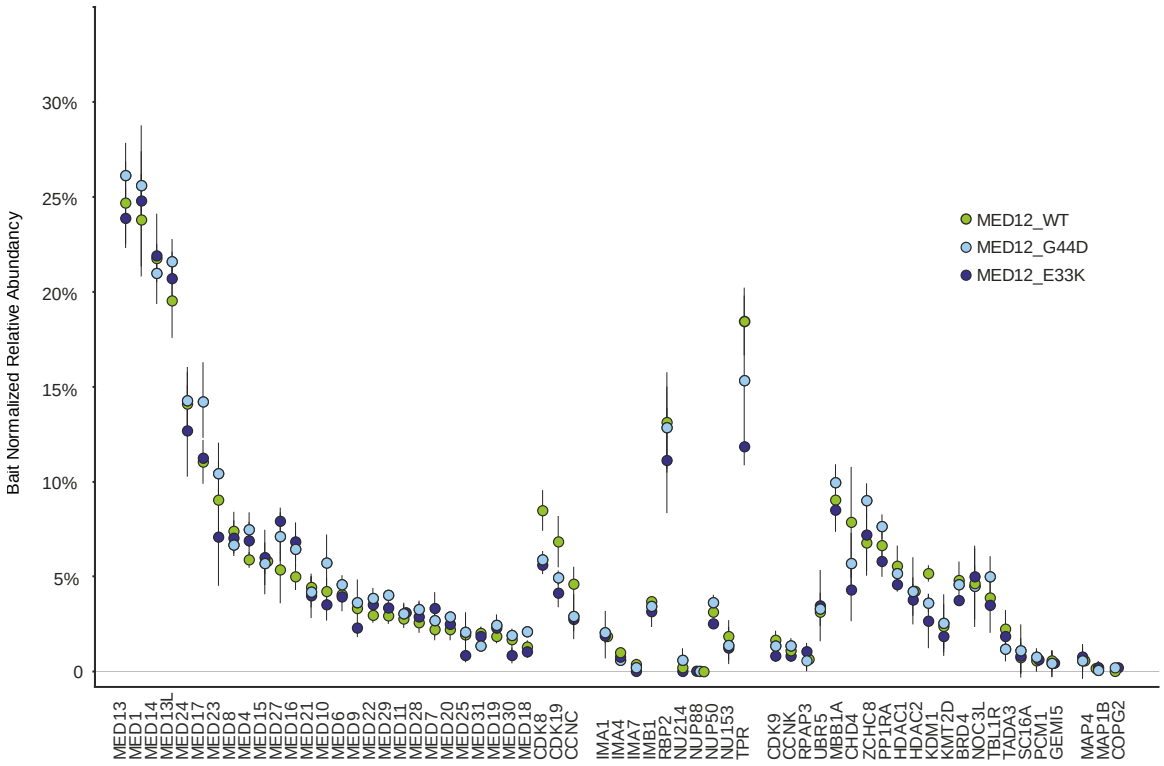
Supp. Figure S5. Immunofluorescence imaging of GFP and MED12-NLS-GFP. The native GFP protein localizes both to the nucleus and to the cytoplasm. In contrast, the MED12-NLS -GFP fusion protein, when expressed in HeLa, localizes predominantly to the nucleus. Scale bar, 20 µm.



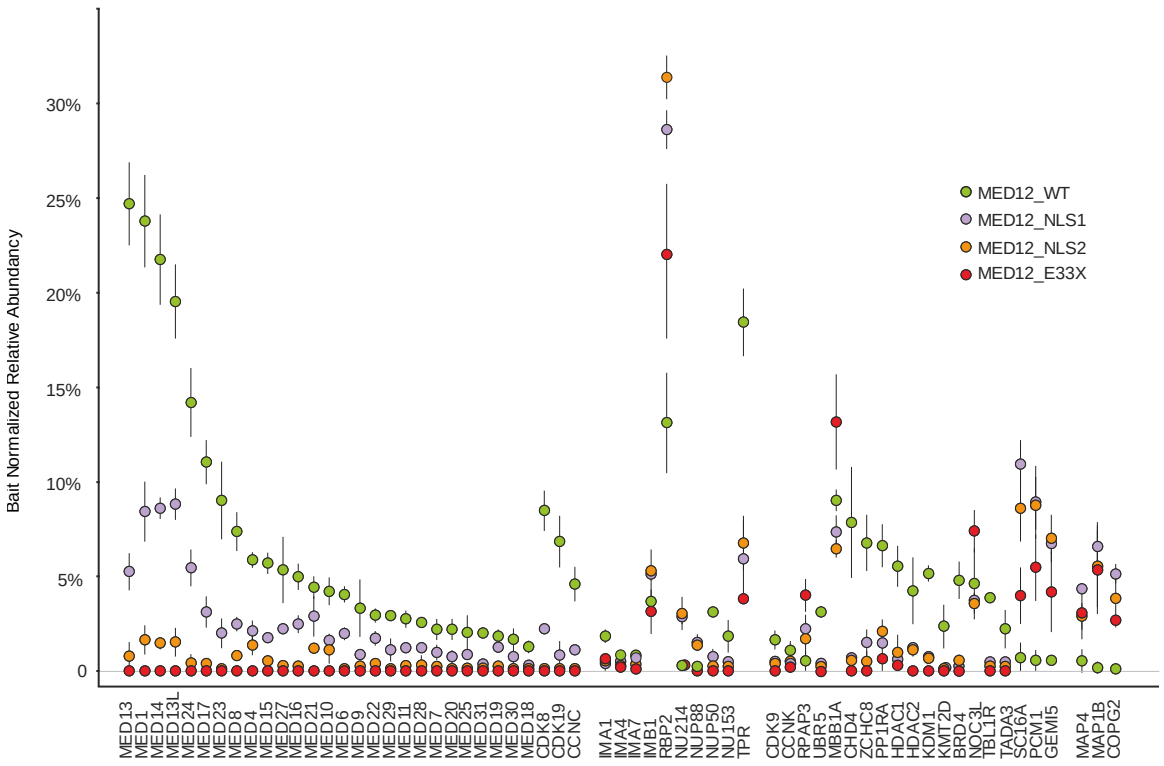
Supp. Figure S6. AP-MS analysis of MED12 WT and NLS-mutant protein derivatives. The interactions between the NLS mutants and Mediator complex components are significantly reduced, but not completely lost as with E33X. Error bars represent ± SD.



Supp. Figure S7. Schematic illustration of the BioID pipeline. The minimal biotin ligase BirA is activated by addition of biotin to the cell culture media and the proximity biotinylated interacting proteins are purified using affinity purification with Strep-tactin. The biotinylated proteins are then analyzed using LC-MS and the high-confidence protein-protein interactions are resolved using statistical filtering.



Supp. Figure S8. BioID MS analysis of MED12 WT, G44D, and E33K. In concordance with previous findings, the MED12 exon 1 and 2 missense mutations affect the interactions between MED12 and the Mediator kinase module components CyclinC (CCNC) and CDK8/19. Error bars represent ± SD.



Supp. Figure S9. BioID MS analysis of MED12 WT, NLS1, NLS2, and E33X. The analysis confirms the results from AP-MS and show that interactions between MED12 mutant derivatives and other components of the Mediator are lost (E33X) or significantly reduced (NLS1, NLS2). Error bars represent ± SD.

Supp. Table S1. Semi-tryptic peptides recognized in MED12-E33X and MED12-WT

peptides STOP			WT peptides		
POSITION	SUM PSM	Sequence	POSITION	SUM PSM	Sequence
			19	2	LGPPDVYPQDPK
			31	3	QKEDELTALNVK
			43	12	QGFNNQPAVSGDEHGSAK
			69	6	ISSNFSSIIAEK
			83	4	cNTLPDTGR
			100	6	DNFWLVTAR
			109	12	SQSAINTWFTDLAGTKPLTQLAK
			139	4	KEEVFGYLAK
163	3	TcAYYAAISETK	162	4	MTcAYYAAISETK
179	9	HVDPFMEWTQIITK	179	9	HVDPFMEWTQIITK
201	5	MAEYYRPGPAGSGGcGSTIGPLPHDVEVAIR	201	4	MAEYYRPGPAGSGGcGSTIGPLPHDVEVAIR
207	4	PGPAGSGGcGSTIGPLPHDVEVAIR	207	3	PGPAGSGGcGSTIGPLPHDVEVAIR
265	1	IRPGEDELLK	265	2	IRPGEDELLK
283	6	YSGEFVQSAYLSR	275	2	LLLPLLLR
388	2	IKTGSPLDHLPIAPSNLPMPEGNSAFTQQVR	283	6	YSGEFVQSAYLSR
390	22	TGSPLDHLPIAPSNLPMPEGN	388	4	IKTGSPLDHLPIAPSNLPMPEGNSAFTQQVR
407	2	PEGNSAFTQQVR	390	28	TGSPLDHLPIAPSNLPMPEGN
421	4	LREIEQQIK	421	5	LREIEQQIK
444	6	cQEATAGFTIGR	430	1	ERGQAVEVR
456	12	VLHTLEVLDSHSFER	444	8	cQEATAGFTIGR
471	7	SDFSNSLDSLcNR	456	14	VLHTLEVLDSHSFER
484	5	IFGLGPSKDGHEISSDDDAVVSLLcEWAVScK	471	6	SDFSNSLDSLcNR
492	18	DGHEISSDDDAVVSLLcEWAVScK	484	8	IFGLGPSK
532	7	RQAEIEAER	492	33	DGHEISSDDDAVVSLLcEWAVScK
541	1	cGESEAADEKGSIASGSLSAPSAPIFQDVLLQFLDTQAPMLTDPR	532	3	RQAEIEAER
551	11	GSIASGSLSAPSAPIFQDVLLQFLDTQAPMLTDPR	541	11	cGESEAADEKGSIASGSLSAPSAPIFQDVLLQFLDTQAPmLTDPR
606	9	HDVFSHNMYTcTLISR	551	16	GSIASGSLSAPSAPIFQDVLLQFLDTQAPMLTDPR
622	5	GDLAFGAPGPRPPSPFDDPADDPEHK	606	10	HDVFSHNMYTcTLISR
657	1	LEDPGLSESMDIDPSSSVLFEDMEKPDFSLFSPTMPcEGK	622	2	GDLAFGAPGPRPPSPFDDPADDPEHK
697	3	GSPSPEKPDVEK	657	4	LEDPGLSESMDIDPSSSVLFEDMEKPDFSLFSPTMPcEGK
716	5	EKIEGTLGVLYDQPR	697	3	GSPSPEKPDVEK
718	10	IEGTLGVLYDQPR	716	4	EKIEGTLGVLYDQPR
731	11	HVQYATHFPIPQEEScSHEcNQR	718	12	IEGTLGVLYDQPR
754	6	LVVLFGVGK	731	14	HVQYATHFPIPQEEScSHEcNQR
785	8	KGTAETDQLAPIVPLNPGDLTFLGGEDGQK	754	7	LVVLFGVGK
786	9	GTAETDQLAPIVPLNPGDLTFLGGEDGQK	785	9	KGTAETDQLAPIVPLNPGDLTF
818	6	NRPEAFPTAEDIFAK	786	15	GTAETDQLAPIVPLNPGDLTFLGGEDGQK
833	15	FQHLSHYDQHQVTAQVSR	818	7	NRPEAFPTAEDIFAK
840	1	DQHQVTAQVSR	833	15	FQHLSHYDQHQVTAQVSR
932	4	HYHAcLILNQDQMAQVFEGLcGVVK	840	2	DQHQVTAQVSR
970	7	cILAYLYDLYTScSHLK	932	8	HYHAcLILNQDQMAQVFEGLcGVVK
987	6	NKFGELFSDFcSK	970	9	cILAYLYDLYTScSHLK
989	4	FGELFSDFcSK	975	4	LYDLYTScSHLK
1000	1	VKNTIYcNVEPSESNMR	987	6	NKFGELFSDFcSK
1002	13	NTIYcNVEPSESNmR	989	4	FGELFSDFcSK
1017	11	WAPEFMIDTLENPAAHTF	1000	3	VKNTIYcNVEPSESNMR
1042	1	SLSENPANR	1002	12	NTIYcNVEPSESNMR
1051	3	YSFVcNALMHVcVGHHDPDR	1017	14	WAPEFMIDTLENPAAH
1061	1	VcVGHHDPDR	1042	3	SLSENPANR
1071	6	VNDIAILcAELTGYcK	1051	6	YSFVcNALMHVcVGHHDPDR
1087	6	SLSAEWLGVLK	1071	3	VNDIAILcAELTGYcK
1139	6	QcLLLEDLIR	1087	6	SLSAEWLGVLK
1149	8	cAAIPSLLNAAcSEQDSEPGAR	1139	5	QcLLLEDLIR
1157	3	NAAcSEQDSEPGAR	1149	5	cAAIPSLLNAAcSEQDSEPGAR
1175	4	ILLHLFK	1157	3	NAAcSEQDSEPGAR
1182	18	TPQLNPcQSDGNKPTVGIR	1175	4	ILLHLFK
1206	1	HLLAASQNR	1182	15	TPQLNPcQSDGNKPTVGIR
1215	9	IVDGAVFAVLK	1215	9	IVDGAVFAVLK
1226	6	AVFVLGDAELK	1226	8	AVFVLGDAELK
1237	12	GSGFTVTGGTEELPEEEGGGGSGGR	1237	9	GSGFTVTGGTEELPEEEGGGGSGGR
1267	5	NISVETASLDVYAK	1241	2	TVTGGTEELPEEEGGGGSGGR
1285	6	SIcQQEWVGER	1267	4	NISVETASLDVYAK
1299	14	SLcEDSNDLQDPVLSSAQAQR	1285	3	SIcQQEWVGER
1310	1	PVLSSAQAQR	1299	16	SLcEDSNDLQDPVLSSAQAQR
1320	7	LMQLIcYPHR	1320	6	LMQLIcYPHR
1330	9	LLDNEDGENPQR	1330	6	LLDNEDGENPQR
1347	5	ILQNLDQWTMR	1347	6	ILQNLDQWTMR
1358	2	QSSLELQLMIK	1358	2	QSSLELQLMIK
1369	7	QTPNNEMNSLLENIAK	1369	7	QTPNNEMNSLLENIAK
1385	12	ATIEVFQQSAETGSSSGSTASNMPSSSK	1385	18	ATIEVFQQSAETGSSSGSTASNmPSSSK
1413	4	TKPVLSSLER	1413	4	TKPVLSSLER
1423	9	SGVWLVAPLIAK	1423	6	SGVWLVAPLIAK
1435	6	LPTSVQGHVLK	1435	8	LPTSVQGHVLK
1446	8	AAGEELEKGQHLGSSSR	1436	1	PTSVQGHVLK
1472	14	SMSLLSQQPFLSLVLTcLK	1446	4	AAGEELEKGQHLGSSSR
1482	1	LSLVLTcLK	1472	16	SmSLLSQQPFLSLVLTcLK
1497	1	EGLLTSLYSQVHQIVNNWR	1497	4	EGLLTSLYSQVH
1516	5	DDQYLDDcKPK	1516	3	DDQYLDDcKPK
1537	7	LNLVGGMFDTVQR	1537	10	LNLVGGMFDTVQR
1637	1	QLLPLPK	1637	1	QLLPLPK
1647	9	DVITcEPQGSLIDTK	1647	10	DVITcEPQGSLIDTK
1665	4	IAGFDSIFKK	1665	5	IAGFDSIFKK
1685	9	ISPWDLFEGLKPSAPL	1685	10	ISPWDLFEGLK
1696	1	PSAPLSWGWFGTVR	1696	2	PSAPLSWGWFGTVR
1736	33	AYYLEPLPLPPEDEEPPAPTLLEPEKK	1736	33	AYYLEPLPLPPEDEEPPAPTLLEPEK
1769	1	TDKPGAAPPSTEER	1769	4	TDKPGAAPPSTEER
1793	1	SQPATKTEDYGMGPGR	1793	1	SQPATKTEDYGMGPGR
1799	6	TEDYGmGPGR	1799	4	TEDYGmGPGR
1809	3	SGPYGVTVPPDLLHHPNPGSITHLNYR	1809	9	SGPYGVTVPPDLLHHPNPGSITHLNYR
1817	2	PPDLLHHPNPGSITHLNYR	1817	3	PPDLLHHPNPGSITHLNYR
1836	11	QGSIGLYTQNQPLPAGGPR	1836	10	QGSIGLYTQNQPLPAGGPR
1868	13	LPTRPTYPGVLPTTMTGVMGLEPSSYK	1868	20	LPTRPTYPGVLPTTMTGVMGLEPSSYK
1900	5	QQQPAVPQGQR	1900	4	QQQPAVPQGQR
1948	3	VSHVGLQQHTGPAGTMVPPSYSSQPYQSTHPSTNPTLVDPTR	1913	1	QQLQQSQGMLGQSSVHQMTPSSSYGLQTSQGYTPY
1957	3	TGPAGTMVPPSYSSQPYQSTHPSTNPTLVDPTR	1948	6	VSHVGLQQHTGPAGTMVPPSYSSQPYQSTHPSTNPTLVDPTR
1990	15	HLQQRPSGYVHQQAPTYGHGLTSTQR	1957	4	TGPAGTMVPPSYSSQPYQSTHPSTNPTLVDPTR
1995	9	PSGYVHQQAPTYGHGLTSTQR	1963	1	MVPPSYSSQPYQSTHPSTNPTLVDPTR
2016	17	FSHQTLQQTPMISTMTPM	1974	2	QSTHPSTNPTLVDPTR
2027	8	ISTMTPMSAQGVQAGVR	1990	18	HLQQRPSGYVHQQAPTYGH
2034	7	SAQGVQAGVR	1995	9	PSGYVHQQAPTYGHGLTSTQR
2044	6	STAILPEQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQYHIR	1999	5	VHQQAPTYGHGLTSTQR
2081	7	QQQQQQILR	2016	27	FSHQTLQQTPM
2140	15	QGLQQTQQQQQTAALVR	2027	13	ISTMTPMSAQGVQAGVR
2157	14	QLQQQLSNTQPQPSTNIFGR	2034	4	SAQGVQAGVR
2163	1	SNTQPQPSTNIFGR	2044	10	STAILPEQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQY
			2081	6	QQQQQQILR
			2140	14	QGLQQTQQQQQTAALVR
			2157	12	QLQQQLSNTQPQPSTNIFGR
			2163	1	SNTQPQPSTNIFGR

Supp. Table S2. High-confidence protein-protein interactions identified using AP-MS. The relative abundancy of each interacting protein is bait normalized (prey spectral count value divided by the bait spectral counts). Each of the MED12 constructs were analyzed in three biological replicates (separate file SuppTableS2.xlsx).

Supp. Table S3. Number of Flp-In 293 T-REx cells expressing different MED12 derivatives analyzed on Anima platform after immunofluorescence staining.

	WT	G44D	E33K	E33X	NLS1	NLS2
Cell number	3044	1720	1333	1830	1683	1828
Relative nucleus/cytoplasm margination						
Average	1.21	1.18	1.22	1.00	0.99	1.01
StDev	0.24	0.20	0.19	0.07	0.14	0.13
Median	1.14	1.13	1.18	1.00	0.98	0.99
p-value vs. WT	-	0.03476	3.419×10^{-12}	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$

Supp. Table S4. High-confidence protein-protein interactions identified using BioID. The relative abundancy of each interacting protein is bait normalized (prey spectral count value divided by the bait spectral counts). Each of the MED12 constructs were analyzed in three biological replicates (separate file SuppTableS4.xlsx).

Supp. Table S5. Primers used for sequencing and for site-directed mutagenesis in creating the *MED12* cDNA constructs.

PCR Amplicon	Forward Primer Sequence	Reverse Primer Sequence
MED12 exon 1	ccccttttcggctccctc	gtcagtgcctcctcctagg
MED12 cDNA exon 1-2	cttcgggatcttgagctacg	gatcttggcaggattgaagc
Mutagenesis c.97G>A, p.E33K	gccgtcagttcatccttcttctgtttgggg tcc	ggaccccaaacagaagaaggatgaactgac ggc
Mutagenesis c.97G>T, p.E33X	gccgtcagttcatcctacttctgtttgggg tcc	ggaccccaaacagaagtaggatgaactgac ggc
Mutagenesis NLS1 c.37-48 CCCCTGAAGCGG>GCCGCGGCGGCG, p.13-16 PLKR>AAAA	aggccccagccgcggcgccgccgcggcccg gtgttcgtagctc	gagctacgaacaccgggccgcggcggcgcc gcggctggggcct
Mutagenesis NLS2 c.37-57 CCCCTGAAGCGGCCGCGGCTG>GCCGCGGCGGC GGCGGCGGCG, p.13-19 PLKRPRL>[A]7	agggtaaacatcgggaggccccgccgccgc cgccgccgcggcccggtgttcgtagctcaa gatc	gatcttgagctacgaacaccgggccgcggc ggcggcggcggcggggcctcccgatgttta ccct
Mutagenesis c.560-561 AT>GC, p.M154A	ccaggcagcccgcgccacaggcactgtg	cacagtgcctgtggcgcgggctgcctgg
Mutagenesis c.584-585 AT>GC, p.M162A	agtaggcacaggtcgccttaatgagccagg cagccc	gggctgcctggctcattaaggcgacctgtg cctact