

Northumbria Research Link

Citation: Best, Andrew, James, Katherine, Dalgliesh, Caroline, Hong, Elaine, Kheirolah-Kouhestani, Mahsa, Curk, Tomaz, Xu, Yaobo, Danilenko, Marina, Hussain, Rafiq, Keavney, Bernard, Wipat, Anil, Klinck, Roscoe, Cowell, Ian G., Cheong Lee, Ka, Austin, Caroline A., Venables, Julian P., Chabot, Benoit, Santibanez Koref, Mauro, Tyson-Capper, Alison and Elliott, David J. (2014) Human Tra2 proteins jointly control a CHEK1 splicing switch among alternative and constitutive target exons. *Nature Communications*, 5 (1). ISSN 2041-1723

Published by: Nature Publishing

URL: <https://doi.org/10.1038/ncomms5760> <<https://doi.org/10.1038/ncomms5760>>

This version was downloaded from Northumbria Research Link: <http://nrl.northumbria.ac.uk/42275/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



**Northumbria
University**
NEWCASTLE



UniversityLibrary

ARTICLE

Received 17 Apr 2014 | Accepted 22 Jul 2014 | Published 11 Sep 2014

DOI: 10.1038/ncomms5760

OPEN

Human Tra2 proteins jointly control a *CHEK1* splicing switch among alternative and constitutive target exons

Andrew Best¹, Katherine James², Caroline Dalgliesh¹, Elaine Hong³, Mahsa Kheirolah-Kouhestani¹, Tomaz Curk⁴, Yaobo Xu¹, Marina Danilenko¹, Rafiq Hussain¹, Bernard Keavney^{1,5}, Anil Wipat², Roscoe Klinck⁶, Ian G. Cowell⁷, Ka Cheong Lee⁷, Caroline A. Austin⁷, Julian P. Venables¹, Benoit Chabot⁶, Mauro Santibanez Koref¹, Alison Tyson-Capper³ & David J. Elliott¹

Alternative splicing—the production of multiple messenger RNA isoforms from a single gene—is regulated in part by RNA binding proteins. While the RBPs transformer2 alpha (Tra2 α) and Tra2 β have both been implicated in the regulation of alternative splicing, their relative contributions to this process are not well understood. Here we find simultaneous—but not individual—depletion of Tra2 α and Tra2 β induces substantial shifts in splicing of endogenous Tra2 β target exons, and that both constitutive and alternative target exons are under dual Tra2 α -Tra2 β control. Target exons are enriched in genes associated with chromosome biology including *CHEK1*, which encodes a key DNA damage response protein. Dual Tra2 protein depletion reduces expression of full-length CHK1 protein, results in the accumulation of the DNA damage marker γ H2AX and decreased cell viability. We conclude Tra2 proteins jointly control constitutive and alternative splicing patterns via paralog compensation to control pathways essential to the maintenance of cell viability.

¹Institute of Genetic Medicine, Newcastle University, Central Parkway, Newcastle NE1 3BZ, UK. ²School of Computing Science, Claremont Tower, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. ³Institute for Cellular Medicine, Newcastle University, Framlington Place, Newcastle NE2 4HH, UK. ⁴Faculty of Computer and Information Science, University of Ljubljana, Trzaska cesta 25, SI-1000, Ljubljana, Slovenia. ⁵Institute of Cardiovascular Sciences, The University of Manchester, Manchester M13 9NT, UK. ⁶Department of Microbiology and Infectious Diseases, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, Québec, Canada J1E 4K8. ⁷Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle NE2 4HH, UK. Correspondence and requests for materials should be addressed to D.J.E. (email: David.Elliott@ncl.ac.uk)

Human genes encode long precursor messenger RNAs (mRNAs) that are extensively processed before nuclear export. This maturation includes the splicing of exons, which normally occurs with high fidelity to create functional mRNAs¹. Constitutive exons splice into all mRNAs transcribed from a gene, while alternative exons are sometimes included and sometimes skipped². Human protein-coding genes each produce an average of three mRNA isoforms through alternative splicing, many of which are differentially regulated³. RNA binding proteins play a key role in transforming precursor RNAs into mRNAs. Although RNA-binding proteins can regulate many transcripts in parallel, some splicing regulatory proteins preferentially engage with transcripts belonging to specific functional classes, including Nova proteins (synapse functions), Fox proteins (neuromuscular, cytoskeleton and EMT functions), PTB proteins (cytoskeleton functions) and T-STAR (synapse functions)^{4–8}.

Transformer2 (Tra2) proteins are involved in splicing control^{9,10}. First discovered in insects, Tra2 proteins form an essential component of the alternative splicing complex that controls fly sexual differentiation^{11,12}. Tra2 proteins are conserved across the animal kingdom, but separate gene paralogs encoding Tra2 α and Tra2 β proteins evolved early in vertebrate evolution^{13,14}. Knockout experiments in mice show that Tra2 β is essential for embryonic and brain development^{15–18}. In humans, Tra2 β expression levels change in several cancers (reviewed by Best *et al.*¹⁹), and Tra2 β is implicated in the pathology of other diseases including spinal muscular atrophy²⁰, Alzheimer's disease²¹ and frontotemporal dementia and Parkinsonism linked to chromosome 17 (ref. 22).

Tra2 proteins have amino- and carboxy-terminal domains enriched in arginine and serine residues (RS domains) flanking a single central RNA recognition motif (RRM) and so resemble the relatively well characterized core group of 12 SR proteins that control both constitutive and alternative splicing as well as other aspects of RNA metabolism^{23–25}. Each core SR protein contains N-terminal RRMs and single C-terminal RS domains. However, unlike the core SR proteins all current data implicate Tra2 proteins solely in alternative splicing rather than constitutive splicing^{10,26}, and only SR proteins and not Tra2 proteins can provide splicing activity to S100 extracts²⁶.

To regulate splicing inclusion Tra2 β binds to AGAA-rich and CAA-rich target RNA sequences. These RNA protein interactions have been resolved at the atomic level^{9,27}. Endogenous Tra2 β target RNAs have been identified using HITS-CLIP¹⁸, RIP-seq²⁸, shRNA depletion²⁹ and microarrays¹⁷, but important fundamental questions still remain as to the identity of the biological targets and the functions of vertebrate Tra2 proteins. These include whether endogenous Tra2 α and Tra2 β proteins jointly control the same splicing targets, and if so what these shared targets are? Although Tra2 α and Tra2 β both activate splicing of the same model exons when overexpressed in transfected HEK-293 cells (suggesting redundant functions)¹⁸, the *Tra2a* gene alone is not sufficient to maintain viability in *Tra2b* knockout mice (suggesting specific functions)¹⁵. Another question relates to how Tra2 α and Tra2 β interact with each other? We previously found that Tra2 β protein binds to a poison exon in the *TRA2A* gene to activate poison exon inclusion¹⁸. Poison exons introduce premature translation termination codons into mRNAs so as to inhibit translation of full-length proteins and are often regulatory^{18,30–32}, but whether Tra2 α might reciprocally control Tra2 β expression is not known.

Here we address these questions in human MDA-MB-231 cells that model invasive breast cancer. We find asymmetric splicing feedback control pathways between Tra2 α and Tra2 β that buffer splicing defects caused by depletion of either Tra2 α or Tra2 β

protein alone. Overriding these feedback control pathways by joint depletion of both Tra2 α and Tra2 β globally identifies Tra2-dependent target exons, and reveals critical roles for these proteins in DNA damage control and cell viability.

Results

Tra2 β efficiently suppresses Tra2 α protein expression. To test for *in vivo* interactions between Tra2 α and Tra2 β proteins, we monitored their expression levels using western blots. Consistent with predictions from our previous study¹⁸, Tra2 α protein levels were normally very low but significantly increased after small interfering RNA (siRNA)-mediated depletion of Tra2 β (Fig. 1a top panel, compare lanes 1 and 3, and Fig. 1b). Although weak, the Tra2 α western blot signal was of the predicted size and was almost completely eliminated following transfection with a *TRA2A*-specific siRNA (Fig. 1a top panel, compare lanes 1 and 2). Tra2 α protein depletion had less effect on Tra2 β protein levels (Fig. 1a, middle panel and Fig. 1b). Western blot analysis confirmed this effect for two independent sets of siRNAs targeted against different parts of the respective mRNAs (Supplementary Fig. 1).

Consistent with Tra2 β protein repressing Tra2 α expression via poison exon activation, siRNA-mediated depletion of Tra2 β led to strongly reduced splicing inclusion of the *TRA2A* poison exon (Fig. 1d, upper panel). siRNA-mediated depletion of Tra2 α protein led to a smaller but detectable effect on splicing inclusion of the *TRA2B* poison exon (Fig. 1d, lower panel). Analysis of *TRA2A* and *TRA2B* steady state mRNA expression levels by quantitative PCR confirmed that each protein also negatively regulates the expression of the other at the RNA level (Fig. 1c).

The *TRA2A* and *TRA2B* genes are differentially expressed.

RNA-seq of MDA-MB-231 cells indicated that the *TRA2B* gene is expressed at much higher levels than the *TRA2A* gene (Fig. 1e shows one of three biological replicate RNA-seq analyses, with the height of the y axis showing read depth and so indicating relative gene expression levels). This provides a potential mechanism for why Tra2 β represses Tra2 α protein expression more than vice versa, since lower cellular concentrations of Tra2 α would be less able to activate splicing of the *TRA2B* poison exon.

We used iCLIP³³ to systematically map the transcriptome-wide binding sites of human Tra2 β in MDA-MB-231 cells. Endogenous Tra2 β protein was efficiently immunoprecipitated along with radiolabelled crosslinked RNA. A single radiolabelled RNA protein adduct of ~40 kDa was identified at high RNase concentrations, just above the known molecular weight of uncrosslinked endogenous Tra2 β protein (37 kDa) (arrowed in Supplementary Fig. 2a). Lower RNase concentrations enabled endogenous Tra2 β binding sites to be mapped across the MDA-MB-231 cell transcriptome in biological triplicate iCLIP experiments. Following deep sequencing, 7,443,903 reads were successfully mapped back to the human genome, of which 3,338,710 were unique cDNA reads used for downstream analysis (Supplementary Data 1). These individual sequencing reads are subsequently referred to as iCLIP tags. The only clusters of Tra2 β iCLIP tags, which mapped to the human *TRA2B* and *TRA2A* genes from all three biological replicates were within their respective poison exons (Fig. 1e). Despite much lower levels of overall *TRA2A* gene expression, the *TRA2A* poison exon had a similar number of Tra2 β iCLIP tags as the *TRA2B* poison exon. This suggests the *TRA2A* poison exon is a stronger physiological target for Tra2 β binding than the *TRA2B* poison exon (the *TRA2A* poison exon also has a much higher density of AGAA Tra2 β binding sites than the *TRA2B* poison exon¹⁸).

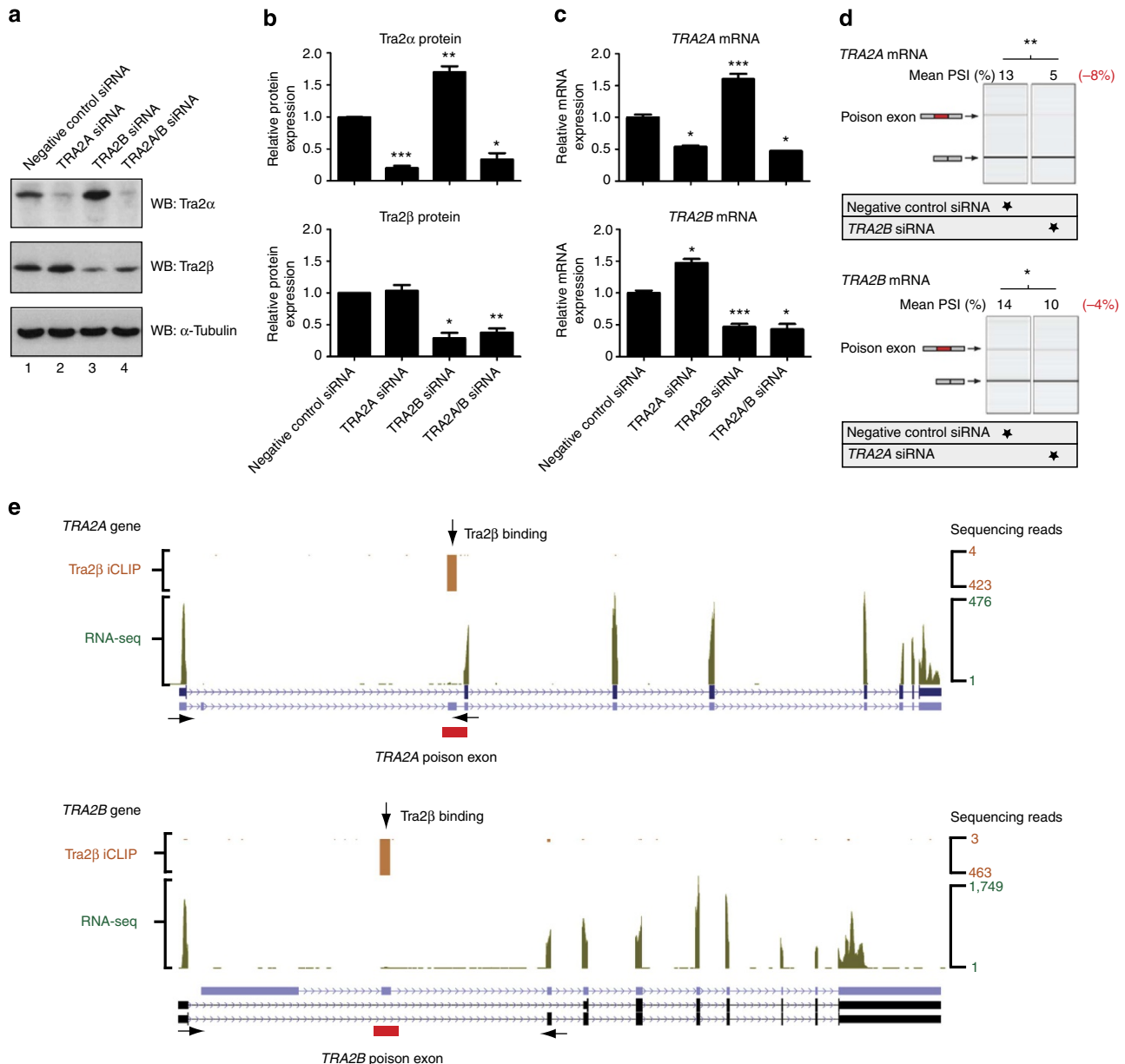


Figure 1 | Tra2 β regulates Tra2 α protein expression. (a) Western blot analysis showing depletion of *TRA2B* induces reciprocal upregulation of Tra2 α protein expression, whereas depletion of *TRA2A* had minimal effect on Tra2 β protein expression. (b) Quantitation of cross-regulation between Tra2 α and Tra2 β at the protein level (Tra2 α and Tra2 β protein expression were quantified relative to α -Tubulin from three western blots using independent biological replicates). (c) Quantitation of cross-regulation between Tra2 α and Tra2 β at the RNA level from quantitative PCR analysis of three independent biological replicates in MDA-MB-231 cells. (d) Splicing inclusion of the *TRA2A* poison exon is strongly reduced by depletion of endogenous Tra2 β protein, whereas splicing inclusion of the *TRA2B* poison exon is less affected by depletion of Tra2 α protein. Splicing patterns were monitored by RT-PCR between flanking exons (arrowed) followed by capillary electrophoresis. (e) Screenshot from the UCSC genome browser³⁵ showing the *TRA2B* and *TRA2A* genes, and the positions of aligned RNA-seq reads (green peaks) and Tra2 β binding (orange clusters of significant cross-linking by Tra2 β protein identified by biological triplicate iCLIP experiments) in MDA-MB-231 cells. Probability (*P*) values were calculated using an independent two-sample *t*-test between negative control siRNA-treated cells and the gene-specific siRNA-treated cells (statistical significance shown as: **P* < 0.05, ***P* < 0.01, ****P* < 0.0001). All data represented by bar charts was generated from three biological replicates and error bars represent the s.e.m.

Endogenous Tra2 α functionally compensates for loss of Tra2 β .

The most frequently enriched pentamers recovered in the iCLIP tags were highly enriched in AGAA nucleotide sequences (Supplementary Fig. 2b), which is the Tra2 β binding site predicted by HITS-CLIP for endogenous mouse Tra2 β , RIP-seq and from SELEX experiments using purified Tra2 β proteins^{18,26,28}. However, our combined human iCLIP data in MDA-MB-231 cells provided substantially more coverage than previously

obtained in mouse testis¹⁸ (in which just 177,457 reads were mapped back to the mouse genome). In total, 1,546,290 (44.8%) of unique cDNAs mapped to intronic regions, suggesting the Tra2 β iCLIP experiment largely captured Tra2 β interactions with pre-mRNAs. However, a further 1,169,374 (33.8%) of unique cDNAs mapped to exons (5'UTR, 3'UTR or ORF), despite exons comprising only ~1% of the genome. After correcting for the relative size of each genomic region (by dividing the

number of unique cDNAs mapping to each genomic region by the relative size of that region within the genome), we find that Tra2 β binding is highly enriched within exons: 76.8% of Tra2 β iCLIP tags mapped to exons (5'UTR, 3'UTR or ORF), while a further 20.4% mapped to non-coding RNAs (Supplementary Fig. 2c).

We used our iCLIP data to screen for endogenous exons controlled by human Tra2 β . Alternative exon junctions (\pm 300 bp of either splice site) were stratified to find those with the highest number of iCLIP tags relative to overall iCLIP coverage within the same gene. To test whether upregulation of Tra2 α protein expression could functionally compensate in splicing regulation for depletion of Tra2 β , we then monitored percentage splicing inclusion (PSI) of associated exons after single depletion of either Tra2 α or Tra2 β , or after combined depletion of both Tra2 α and Tra2 β proteins.

Clusters of Tra2 β iCLIP tags mapped to alternative exons in the *ATRX*, *GLYR1* and *CEP95* genes (Fig. 2a). Single depletion of either Tra2 α or Tra2 β had only a small effect on the endogenous splicing pattern of these three exons, but joint depletion of both Tra2 α and Tra2 β substantially decreased their splicing inclusion

(Fig. 2b). We obtained similar results with 14/14 Tra2 β target exons identified by iCLIP analysis (Fig. 2c. Individual data for each of these tested exons are shown in Supplementary Fig. 3). In fact, of these 14 tested exons, some less-responsive exons including cassette exons within *PAM* and *BDP1* only responded to depletion of both Tra2 proteins, and not to single depletion of Tra2 β at all (Supplementary Fig. 3).

These data are consistent with maintenance of splicing patterns via paralogue compensation, that is, following depletion of Tra2 β , upregulated Tra2 α is able to functionally substitute for Tra2 β and largely maintain Tra2 target exon inclusion. The Tra2 β target exons inhibited more substantially by joint Tra2 protein depletion compared with single depletion of either Tra2 α or Tra2 β included *SMN2* exon 7 (Supplementary Fig. 3), which is a candidate target for gene therapy in spinal muscular atrophy²⁰.

Tra2 α and Tra2 β control constitutive exon splicing patterns.

Splicing profiles of candidate Tra2 β target exons (containing Tra2 β iCLIP tag coverage) were next analysed using RNA-seq after joint depletion of Tra2 α and Tra2 β proteins, and changes

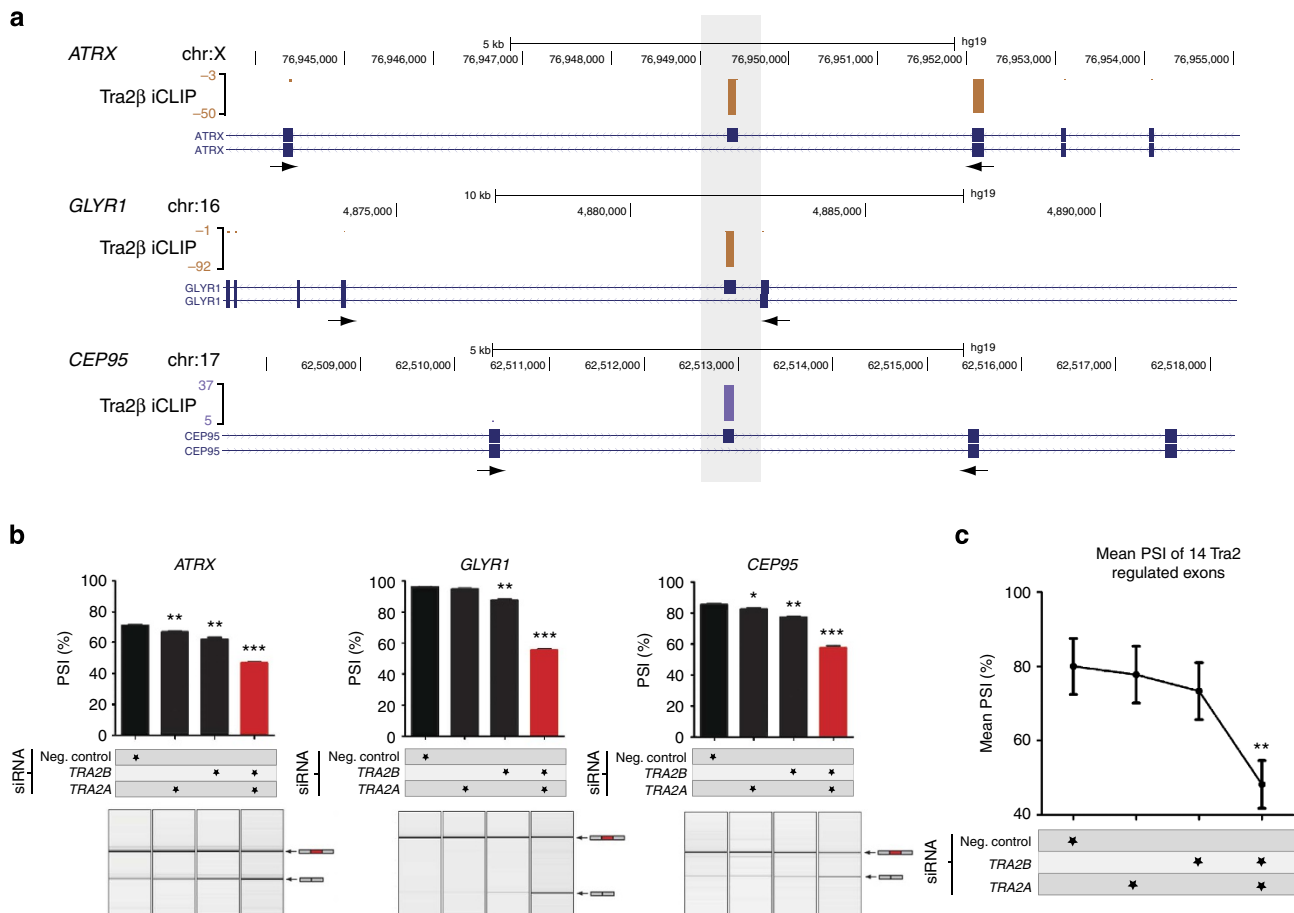


Figure 2 | Endogenous Tra2 α functionally compensates for loss of Tra2 β . (a) UCSC genome browser screenshot³⁵ showing significant clusters of iCLIP tags mapping directly to alternatively spliced exons within the *CEP95*, *GLYR1* and *ATRX* genes (position of target alternative exons highlighted in grey). (b) Splicing inclusion of novel Tra2 β target exons within *ATRX*, *GLYR1* and *CEP95* were only slightly affected by depletion of either endogenous Tra2 α or Tra2 β proteins, but were strongly affected by joint depletion of both Tra2 α and Tra2 β (red). PSI levels were measured by RT-PCR and capillary gel electrophoresis (lower panels) in three biological replicates (upper panels). (c) Splicing inclusion of 14 novel Tra2 β target exons showed minimal splicing response to single depletion of either Tra2 α or Tra2 β , but showed highly significant splicing changes after joint depletion of both Tra2 proteins (complete data for all 14 exons is provided in Supplementary Fig. 4). Probability (*P*) values were calculated using an independent two-sample *t*-test between PSI levels of negative control siRNA-treated cells and TRA2A/TRA2B siRNA-treated cells (statistical significance: **P* < 0.05, ***P* < 0.01, ****P* < 0.0001). All data represented by bar charts was generated from three biological replicates where error bars represent the s.e.m.

different levels after Tra2 protein depletion (Fig. 3b). The constitutively spliced exon within *ANKRD1* showed the largest change (Fig. 3c), a -73% point switch in PSI after joint depletion of Tra2 α and Tra2 β . To further dissect splicing control, we cloned the *ANKRD1* exon and its flanking intronic sequences between β -globin exons in a minigene construct. Transfection experiments showed the *ANKRD1* exon is included at 72% PSI when expressed from this minigene, indicating it lacks some important sequences for splicing (the endogenous *ANKRD1* exon was 100% included in these transfected HEK-293 cells, Fig. 3d). Two clusters of GAA-rich Tra2 β binding sites were present in the *ANKRD1* exon (Tra2 β binding site clusters 1 and 2). Mutation of either Tra2 β binding site cluster negatively impacted splicing inclusion, showing an essential role for these binding sites in modulating the inclusion of the *ANKRD1* exon. In particular, mutation of Tra2 β binding site cluster 1 (to create Mutant M1) completely abolished splicing inclusion of the *ANKRD1* constitutive exon (Fig. 3d, right panel).

In total, we identified and validated 53 human splicing targets, which were both directly bound and jointly controlled by Tra2 proteins, including both alternative and constitutive exons (Fig. 4a; Supplementary Fig. 10; Supplementary Data 2). As well as *SMN2* exon 7 (ref. 20), these included the *NASP-T* exon and the *TRA2A* poison exon, orthologs of which have both been previously identified as functional Tra2 β splicing targets in the mouse testis by HITS-CLIP¹⁸. The *NASP-T* exon was also in the data set of Tra2 β targets identified by RIP-seq²⁸. However, the vast majority of the dual Tra2 α /Tra2 β target exons identified here are novel. The PSI changes for individual genes in response to joint Tra2 α and Tra2 β protein depletion ranged between -4 and -92% points (measured by RT-PCR, Fig. 4a; Supplementary Data 2), indicating that individual Tra2-dependent exons have different intrinsic requirements for Tra2 proteins. The length of Tra2-dependent exons ranged from 64 nucleotides at the smallest (a cassette exon in the *SMYD2* gene, which showed a -5% point switch in PSI in response to Tra2 protein depletion), to 5916 nucleotides at the largest (an unusually large internal cassette exon in the *SON* gene, which showed a -8 point PSI change in response to endogenous Tra2 protein depletion) (Supplementary Data 2). A further 38 exons had Tra2 β iCLIP tag coverage, but did not detectably respond to Tra2 protein depletion in MDA-MB-231 cells: possibly this latter class of exons either might need a relatively small amount of Tra2 protein to be included, or alternatively they might not be controlled by Tra2 proteins in MDA-MB-231 cells (Supplementary Data 2). In a comparison between Tra2 responsive and non-responsive exons, the only statistically significant difference was the density of Tra2 β binding sites in the more highly responsive exons (exons showing >15 PSI change following joint Tra2 α /Tra2 β depletion, Fig. 4b). High resolution iCLIP maps of the individual exons are shown in Supplementary Fig. 10.

Tra2 proteins are highly specific splicing regulators. To establish the relative role of Tra2 proteins in controlling the identified panel of target exons, we probed a custom plate containing cDNAs, where we had systematically knocked down 53 known splicing regulators in MDA-MB-231 cells⁵ (Fig. 4c). Strikingly, of all the knockdowns tested, only double knockdown of Tra2 proteins shifted *ANKRD1* splicing (constitutive exon) and joint Tra2 depletion also had the largest effect on splicing for *GLYR1* (alternative exon). An intermediate situation was observed for *SMC4* (constitutive exon) in which knockdown of *SNRP70* (encoding U170K) also reduced splicing inclusion, as did knockdown of *SRPK1*. The splicing inclusion pattern of *CHEK1* was strongly shifted (-78 point PSI switch) by joint depletion of Tra2 proteins, but consistent with

broader mechanisms of combinatorial control, significant shifts were also seen after depletion of three core U2 snRNP components, which are thought to be important for splice site commitment for all exons. Depletion of other constitutive splicing factors such as *SFRS2*, *hnRNPK*, *hnRNPC2*, *KHSRP* and *CDC5L* also affected *CHEK1* exon 3 splicing.

Since our panel of splicing factor knockdowns was not exhaustive, we cannot exclude all combinations of combinatorial control. However, our data are at least consistent with Tra2 proteins being among the most quantitatively important splicing regulators for their individual target exons.

Tra2 splicing targets associate with chromosome biology. Gene ontology (GO) enrichment analysis of the 53 human genes containing Tra2-dependent exons revealed an enrichment of five functionally similar biological processes (Fig. 5a). Eleven of the 53 genes were annotated to one or more of these processes, with ten of the eleven genes being annotated to the term 'chromosome organization'. There was significant overlap in annotation between this process and annotation to the conceptually related terms 'histone modification' and 'chromatin modification' (Fig. 5b).

The BioGRID database³⁶ was used to retrieve a network of functional interaction data involving genes containing Tra2-dependent exons. In addition to the eleven genes directly annotated to the five enriched GO processes in Fig. 5a, a further 23 of the 53 genes (43.4%) that contain Tra2-dependent exons also have functional interactions with genes annotated to these terms (Supplementary Fig. 4, and summarized in Fig. 5c). Interestingly, although indirectly connected within the network via these annotated genes, none of the 53 genes containing validated Tra2-dependent exons directly interact with one another in the BioGRID database (Supplementary Fig. 4).

Tra2 proteins control splicing of a key checkpoint protein.

Among the Tra2 target exons involved in chromosome biology was exon 3 of the *CHEK1* gene, which encodes the serine/threonine protein kinase CHK1 that is involved in checkpoint control in response to DNA damage. iCLIP analysis identified significant Tra2 β binding over *CHEK1* exon 3 (Fig. 6a), and we observed a -55 point PSI switch for this exon in MDA-MB-231 cells after joint Tra2 protein depletion (Fig. 6b). Joint depletions of Tra2 α and Tra2 β proteins also indicate *CHEK1* exon 3 splicing is under similar control in multiple cell types including MCF7, PC3 and HeLa (Fig. 6b).

Tra2 β iCLIP tags mapped throughout *CHEK1* exon 3, but were particularly enriched towards the 3' splice site (Fig. 7a). We confirmed *CHEK1* exon 3 is a direct target for Tra2 β binding *in vitro* by electrophoretic mobility shift assays (EMSA) using radiolabelled RNA probes corresponding to portions of the exon sequence (Fig. 7b). RNA probe A corresponds to the part of *CHEK1* exon 3 with the most Tra2 β iCLIP tags, and also contains the most predicted binding sites for Tra2 β (shaded green in Fig. 7B, right hand side). RNA probe A was very efficiently shifted by even the lowest tested concentrations (25 ng) of Tra2 β protein. RNA probe B did not bind Tra2 β protein as tightly (around 200 ng Tra2 β protein was needed to see a comparable shift) and also contained fewer Tra2 β binding sites and mapped iCLIP tags. A control RNA probe corresponding to the flanking intron sequence did not shift even at the highest concentrations of Tra2 β protein (this intron sequence contained no predicted Tra2 β binding sites).

We confirmed that *CHEK1* exon 3 is a direct target for Tra2 β splicing regulation using a minigene construct in which *CHEK1* exon 3 is flanked by β -globin exons (Fig. 7c). After transfection of

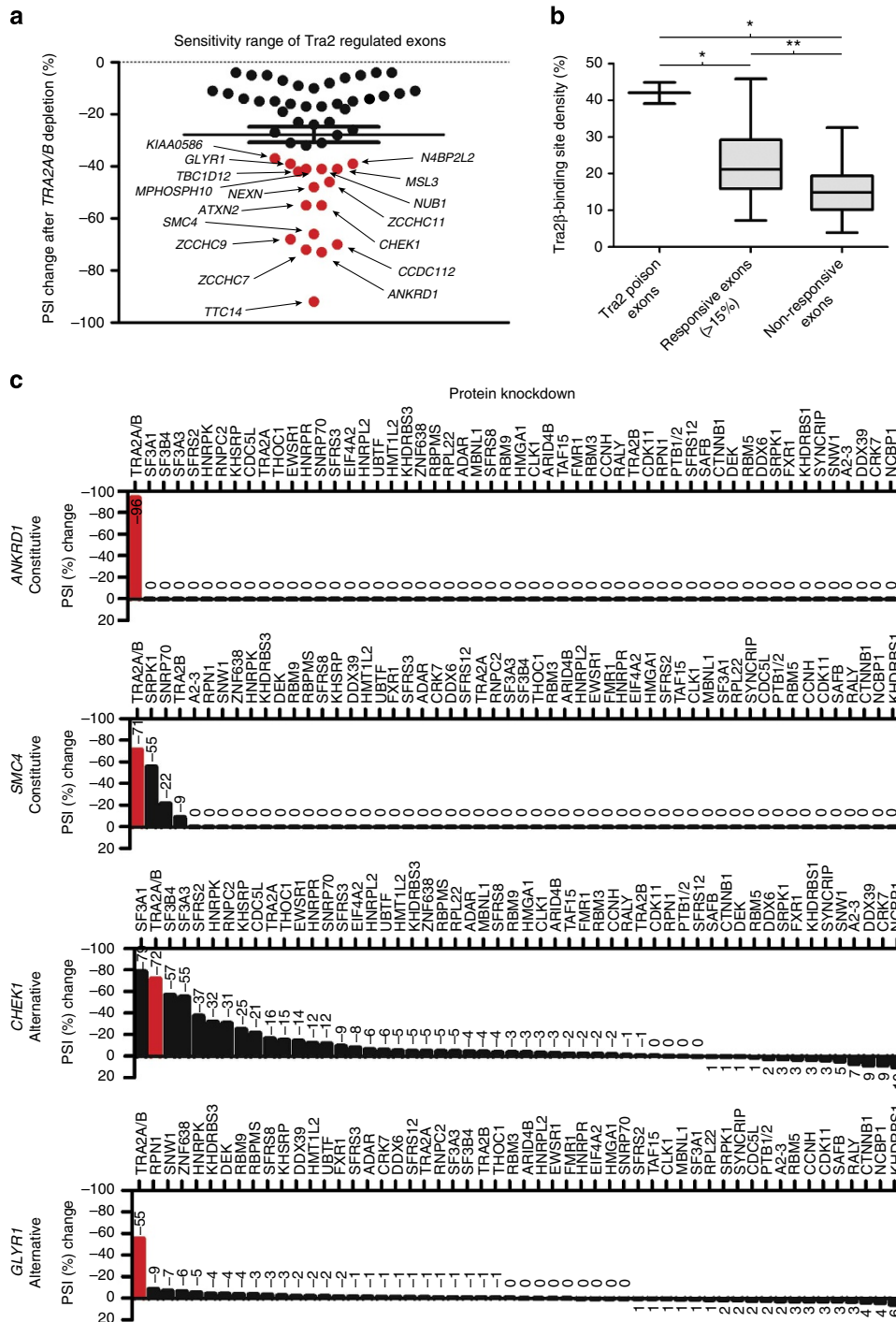


Figure 4 | Identification of exons jointly controlled by Tra2 α and Tra2 β proteins. (a) Scatterplot showing amplitude of splicing response of 53 exons to joint depletion of endogenous Tra2 α and Tra2 β in MDA-MB-231 cells. The genes corresponding to the highest amplitude PSI changes after joint Tra2 protein depletion are labelled and highlighted in red. **(b)** Analysis of Tra2 β binding site density (measured as a percentage of exon content) within groups of Tra2 β target exons identified by iCLIP. Tra2 β binding site density comparisons are shown between the Tra2 α and Tra2 β poison exons; all exons that showed a greater than 15% point PSI change following joint Tra2 α and Tra2 β depletion; and in the exons that bound Tra2 β based on iCLIP tag coverage but did not respond to Tra2 α and Tra2 β depletion. Probability (*P*) values were calculated using an independent two-sample *t*-test (statistical significance: **P* < 0.05, ***P* < 0.01, ****P* < 0.0001). **(c)** Regulation of ANKRD1, SMC4, GLYR1 and CHEK1 splice variants following knockdown of a panel of RNA binding proteins (RBPs) in MDA-MB-231 cells⁵. The y axis shows PSI change after joint Tra2 α and Tra2 β depletion, with a negative number indicating splicing repression in the absence of these proteins.

this minigene into HEK-293 cells, CHEK1 exon 3 was skipped, but its splicing inclusion was strongly induced in response to co-transfection with either Tra2 β -GFP or Tra2 α -GFP. No CHEK1 exon 3 splicing activation was observed after co-

transfection of either Tra2 β Δ RRM-GFP (lacking the RRM) or GFP alone. Furthermore, point mutations of the Tra2 β binding sites within the exon (wild-type binding sites shaded green, mutations shaded red in Fig. 7c right hand side) completely

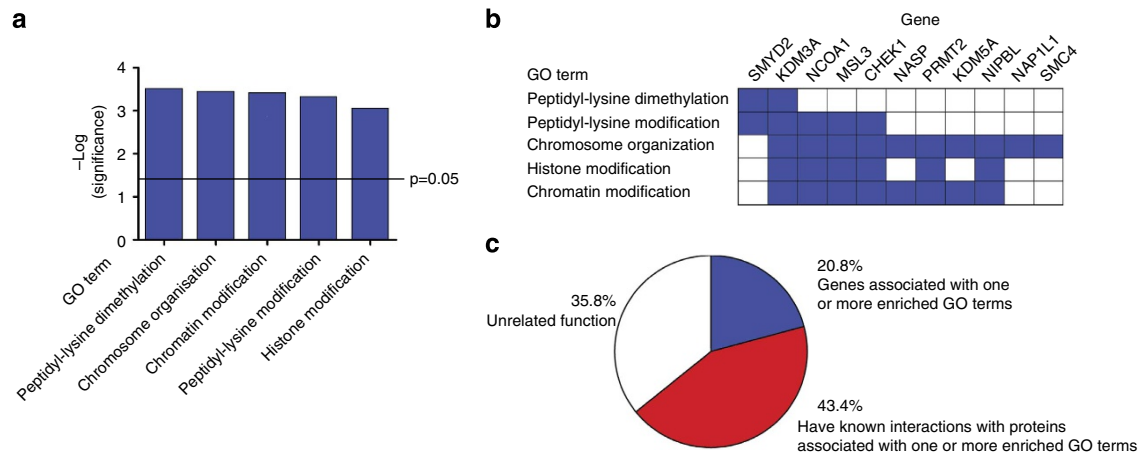


Figure 5 | Tra2 splicing targets are enriched in GO terms associated with chromosome biology. (a) GO enrichment analysis reveals splicing targets responsive to endogenous Tra2 α /Tra2 β protein concentrations are enriched in particular biological processes associated with chromosome biology. (b) GO enrichment analysis showed some joint Tra2 α /Tra2 β -responsive exons were annotated to multiple overlapping biological processes. (c) Summary of GO and network analyses of joint Tra2 α and Tra2 β -dependent splicing targets. Individual segments of the pie chart show the percentage of Tra2 α /Tra2 β target genes directly annotated to the GO biological processes shown in part a; the percentage of Tra2 α /Tra2 β target genes that interact within the BioGrid database with partners known to be involved in the biological processes shown in part a; and the percentage of Tra2 α /Tra2 β target genes, which have unknown or unrelated functions. Full details of the BioGrid analysis are given in Supplementary Fig. 4.

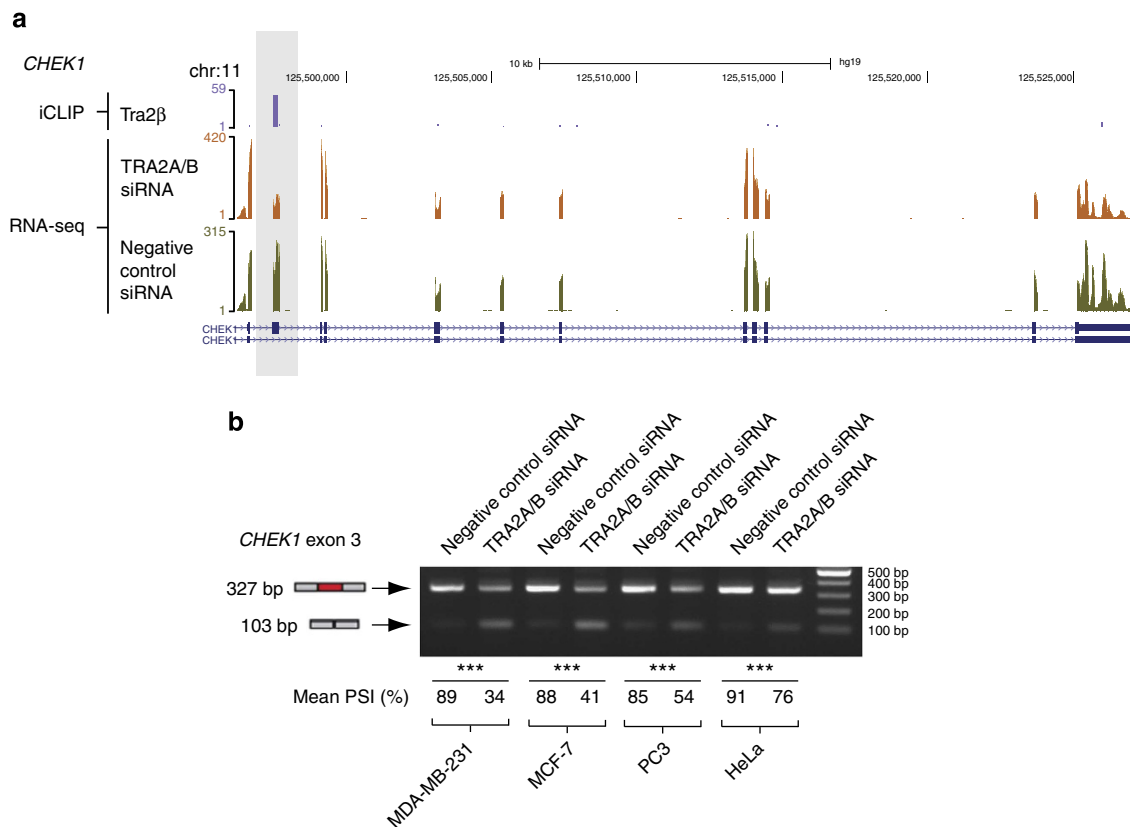


Figure 6 | Splicing inclusion of CHEK1 exon 3 decreases following joint depletion of Tra2 α and Tra2 β . (a) Screenshot³⁵ from the UCSC genome browser showing the CHEK1 gene. Tra2 β iCLIP tags are shown at the top in purple, and map predominantly to a known alternative exon (CHEK1 exon 3) that is skipped in one mRNA isoform (shaded grey). Example RNA-seq tracks from control MDA-MB-231 cells and after joint Tra2 α and Tra2 β protein depletion are shown in green and orange respectively. (b) Splicing inclusion of CHEK1 exon 3 is inhibited by siRNA-mediated Tra2 α and Tra2 β protein depletion in multiple cell types. Splicing inclusion was monitored by RT-PCR and agarose gel electrophoresis, and shown as a mean PSI value after analysis of three independent samples. Probability (P) values were calculated using an independent two-sample t -test between PSI levels of negative control siRNA-treated cells and TRA2A/B siRNA-treated cells (statistical significance: * P < 0.05, ** P < 0.01, *** P < 0.0001). An uncropped gel is shown in Supplementary Fig. 13.

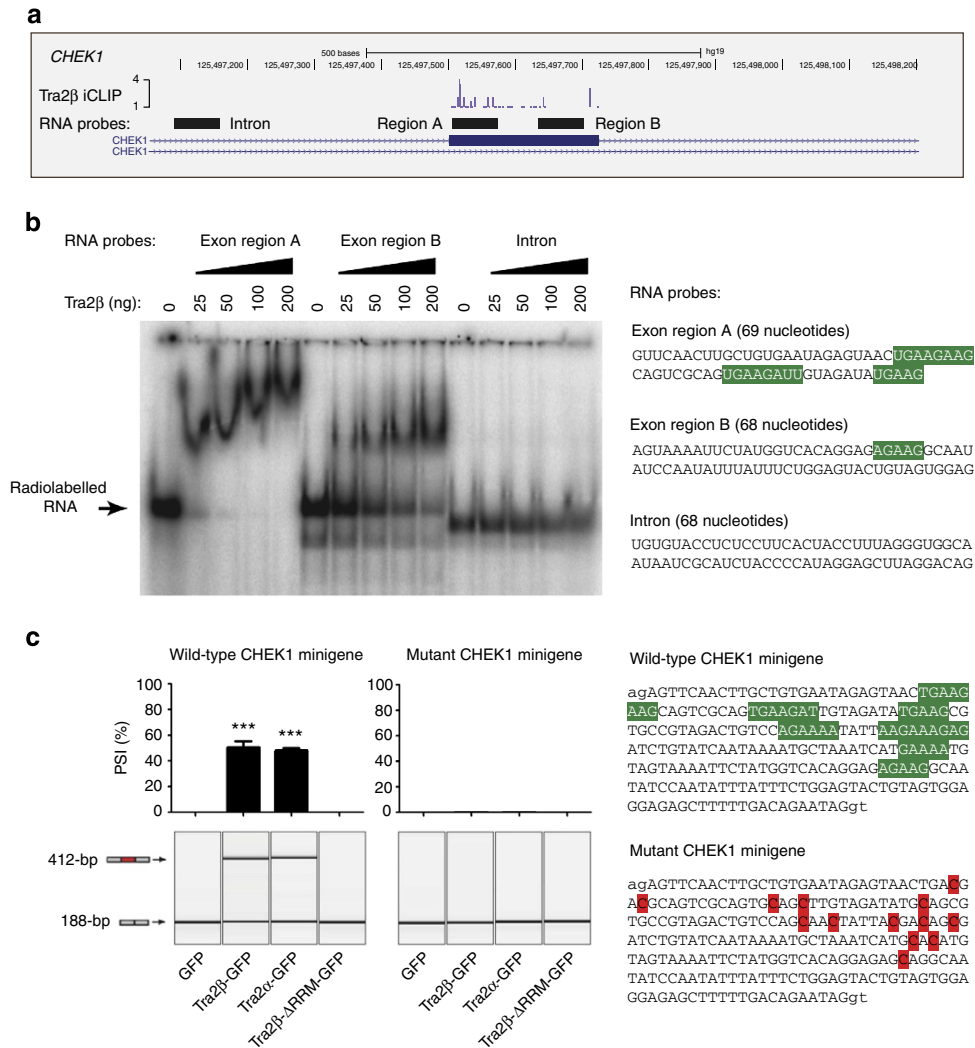


Figure 7 | Human Tra2 α and Tra2 β regulate *CHEK1* exon 3 through direct RNA protein interactions. (a) High resolution map *CHEK1* exon 3 showing mapped iCLIP tags (purple bars) and the three subregions of the pre-mRNA, which were used to generate RNA probes for EMSAs. This screenshot was downloaded from the UCSC genome browser³⁵. **(b)** Molecular interactions between purified Tra2 β protein and RNA probes in and around *CHEK1* exon 3 (the location of these probes is shown in part a). The sequences of the probes are shown to the right, with predicted Tra2 β binding sites shaded green. **(c)** Splicing patterns of mRNAs made from a minigene containing *CHEK1* exon 3 in response to coexpressed fusion proteins, expressed either as a PSI (upper bar chart, $n = 3$ independent experiments) or shown as one of the original capillary electrophoresis gel-like images from a single experiment (lower image). The sequence of *CHEK1* exon 3 is shown to the right, with the predicted Tra2 β binding sites shaded green (above) and the altered sequence after these sites were mutated (below, the altered nucleotides are shaded red). Probability (P) values were calculated using an independent two-sample t -test between PSI levels of the minigene-derived *CHEK1* exon 3 in cells cotransfected with GFP and each of the different Tra2 constructs (statistical significance: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$). All data represented by bar charts was generated from three biological replicates and error bars represent s.e.m.

abolished splicing activation in response to coexpressed Tra2 proteins.

Tra2 proteins are required for CHK1 protein expression. We carried out further experiments to test if Tra2 proteins are also required for expression of full-length CHK1 protein. On western blots, we could detect expression of a single major protein CHK1 isoform in MDA-MB-231 cells, corresponding to the expected size of full-length CHK1 protein (54 kDa). This band was substantially reduced following treatment with an siRNA directed against *CHEK1* mRNA (Fig. 8a). Consistent with joint control of *CHEK1* expression by Tra2 α and Tra2 β , levels of full-length CHK1 protein were also substantially reduced after joint

depletion of Tra2 α and Tra2 β . Expression of full-length CHK1 protein was also reduced after joint Tra2 α and Tra2 β protein depletion in MCF7, PC3 and to a lesser extent HeLa cells (Supplementary Fig. 7a).

A shorter isoform of the CHK1 protein (termed CHK1-S) has previously been reported to be translated from an alternative downstream translational initiation site in exon 3-skipped *CHEK1* mRNA³⁷. In our experiments, although depletion of Tra2 proteins switched splicing of *CHEK1* exon 3, they did not lead to an observable increase in any shorter isoform of the CHK1 protein. We detected much lower expression levels of possible shorter CHK1 protein isoforms on western blots compared with full-length CHK1 protein (the ~43 kDa protein that would correspond in size to CHK1-S could only be seen on long

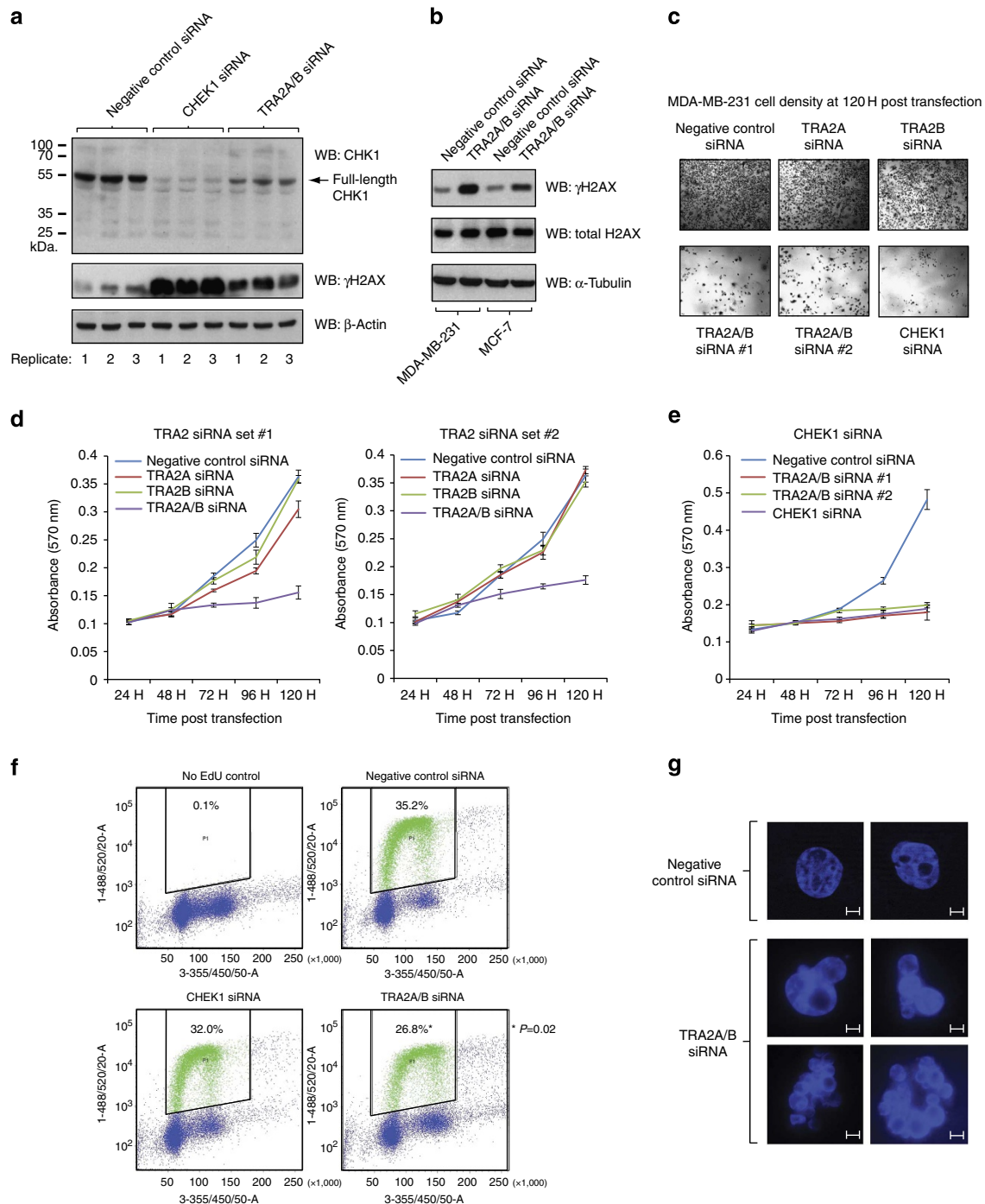


Figure 8 | Human Tra2 proteins are essential for expression of full-length CHK1 protein and to maintain cell viability. (a) Full-length CHK1 protein expression is depleted by siRNAs specific to *CHEK1* mRNA and also by joint siRNAs specific to the *TRA2A* and *TRA2B* mRNAs. In each case, samples from three independent replicate experiments were analysed in parallel. Also detected in these samples are levels of γ H2AX and α -tubulin. (b) Expression of total H2AX and γ H2AX after joint *Tra2 α* and *Tra2 β* depletion, or depletion with a control siRNA in MDA-MB-231 cells and MCF7 cells. (c) Measurement of cell density 120 h after transfection of siRNAs targeting different regions of the *TRA2A* and *TRA2B* mRNAs or *CHEK1* mRNA. (d) Joint (but not single) depletion of *Tra2 α* and *Tra2 β* proteins reduced MDA-MB-231 cell viability measured by MTT assays after siRNA transfection. (e) Depletion of *CHEK1* protein alone was sufficient to reduce viability of MDA-MB-231 cells measured by MTT assay at different time points after siRNA transfection. (f) Joint depletion of *Tra2 α* and *Tra2 β* reduced the proportion of EdU-positive MDA-MB-231 cells 96 h after siRNA transfection. Separate panels, shown clockwise from top left, show fluorescence-activated cell sorting analysis of control MDA-MB-231 cells incubated without EdU; cells transfected with a negative control siRNA and incubated with EdU; cells transfected with siRNAs specific for *TRA2A* and *TRA2B* and incubated with EdU; and cells transfected with a single siRNA specific to *CHEK1* and incubated with EdU. Probability (*P*) values were calculated using an independent two-sample *t*-test comparing the percentage of EdU-positive cells of negative control siRNA-treated cells and the gene-specific siRNA-treated cells (statistical significance: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$). Data were generated from three biological replicates. (g) Examples of abnormal nuclear morphology observed within cells transfected with siRNAs specific for *TRA2A* and *TRA2B* (lower panel) compared with the normal morphology seen in negative control siRNA-treated cells (upper panel). Cells were stained with 4',6-diamidino-2-phenylindole, and these images were taken 96 h after siRNA transfection. The scale bar shows 5 μ m. Uncropped western blots are shown in Supplementary Figs 11–15.

exposure, and decreased on siRNA treatment, Supplementary Fig. 7b). CHK1-S protein is reported to be regulated over the cell cycle and in tumours³⁷. To test for cell cycle regulated splicing inclusion of *CHEK1* exon 3, we prepared RNAs from KGI cell populations enriched in different cell cycle stages prepared using elutriation (Supplementary Fig. 5). When analysed by RT-PCR, very similar patterns of *CHEK1* exon 3 splicing inclusion were observed in each of the cell populations even though they contain different cell cycle profiles. We could detect high levels of both *CHEK1* splice isoforms in RNA purified from a small panel of breast cancer tissues, although we did not see an enrichment of either isoform in any particular tumour type at the RNA level (Supplementary Fig. 6). Overall, the above data are most consistent with Tra2 α and Tra2 β activity being essential for expression of full-length CHK1 protein rather than inducing expression of a shorter protein isoform of CHK1.

Tra2 protein depletion affects DNA damage and cell viability.

Although it is not as a direct target of CHK1 phosphorylation, γ H2AX has been used as a marker for the replication stress that can be induced by depleted CHK1 levels^{38,39}. Similar to previous observations³⁸, we observed greatly increased levels of the DNA damage marker γ H2AX following depletion of CHK1 protein by siRNA, compared with cells treated with a negative control siRNA (Fig. 8a). Increased γ H2AX levels were also observed after joint depletion of Tra2 α and Tra2 β proteins in MDA-MB-231 cells and in MCF7 cells (Fig. 8a,b). The relative increased levels of γ H2AX following *CHEK1* or TRA2A/B siRNA treatment appear proportional to the reduction in full-length CHK1 protein expression observed by western blot.

Microscopy and MTT assays also indicated reduced cell viability 120 hours after joint Tra2 α and Tra2 β depletion (Fig. 8c,d). In contrast, single depletion of either Tra2 α or Tra2 β had negligible effect on cell viability compared with mock depleted cells. Similar results were obtained using two independent sets of siRNAs targeted at different regions of the mRNAs. This reduction in cell viability from joint removal of Tra2 α and Tra2 β , compared with the negligible effects of removing either protein alone, suggest that Tra2 α and Tra2 β are functionally interchangeable for maintaining cell viability in MDA-MB-231 cells, as well as in splicing control.

Depletion of CHK1 also reduced cell viability in MDA-MB-231 cells (Fig. 8e). This suggests that depletion of full-length CHK1 protein would likely be sufficient by itself to contribute to the loss of cell viability observed after joint Tra2 α and Tra2 β depletion. To test if re-introduction of full-length CHK1 protein would be sufficient to restore viability of joint Tra2 α and Tra2 β protein-depleted cells, we made a stable cell line in the FLP-in HEK-293 cell background in which a full-length FLAG-tagged CHK1 protein was expressed under control of a tetracycline promoter. Similar to the result obtained in MDA-MB-231 cells, joint depletion of Tra2 α and Tra2 β reduced cell viability in this stable HEK-293 cell line. However, although the full-length FLAG-tagged CHK1 protein was efficiently induced by tetracycline, it was not sufficient to rescue cell viability after joint Tra2 protein depletion (Supplementary Figs 8 and 9). While we cannot rule out that the tagged full-length CHK1 protein failed to rescue viability of this cell line for another reason, this result is consistent with multiple exons controlled by Tra2 proteins (including *CHEK1* exon 3) being important for cell viability.

Finally, we monitored incorporation of the thymidine analogue EdU using flow cytometry to determine whether joint Tra2 protein depletion affected cell proliferation of MDA-MB-231 cells (Fig. 8f). After joint Tra2 protein depletion, we observed a significant reduction in the proportion of EdU-positive cells 96 h

after siRNA transfection (an 8.4% reduction, $P = 0.02$), indicating fewer cells had initiated DNA replication after joint Tra2 protein depletion. A slight reduction in the proportion of EdU-positive cells was observed after single CHK1 protein depletion, but this was not statistically significant when compared with negative control siRNA-treated cells. Joint Tra2 protein depletion also caused an increase in the proportion of cells containing abnormally shaped nuclei 96 h after siRNA transfection, consistent with major biological defects (Fig. 8g).

Discussion

Here we find that only joint depletion of both Tra2 α and Tra2 β proteins (and not single depletion of either protein alone) could induce substantial splicing switches in endogenous Tra2 β target exons in MDA-MB-231 cells. This joint depletion strategy has enabled us to derive the most comprehensive map of dual Tra2-dependent target exons in any organism to date. Among the jointly regulated exons identified here was a key exon in the *CHEK1* gene, which encodes a protein essential for monitoring DNA damage and controlling cell cycle progression^{37,39}. Exon 3 of the *CHEK1* gene is 224 nucleotides long; hence, skipping of this exon in the absence of Tra2 proteins would frameshift the reading frame of the *CHEK1* mRNA if a downstream translational initiation site is not selected³⁷. Joint depletion of both Tra2 proteins quantitatively switched *CHEK1* pre-mRNA splicing, reduced expression of full-length CHK1 protein, and led to an increase in DNA damage as monitored by accumulation of γ H2AX. We also confirmed *SMN2* exon 7 as a joint Tra2 α /Tra2 β target exon. Joint control by Tra2 α provides an explanation why *SMN2* exon 7 is a target for Tra2 β in transfected cells, but not appreciably affected in *Tra2b* single knockout mice¹⁵.

This strategy also reveals that Tra2 proteins are required for splicing inclusion of some constitutively spliced exons. To the best of our knowledge, Tra2 proteins have only previously been described as alternative splicing factors²⁶. Low levels of apparent alternative splicing of constitutive exons might be ascribed to error prone exon recognition by the spliceosome⁴⁰. However, the *ANKRD1* and *SMC4* exons are not annotated as alternatively spliced in any tissue consistent with them being true constitutive exons, yet also show high-amplitude splicing changes upon Tra2 protein depletion. A role in constitutive splicing brings the Tra2 proteins closer to the core SR group in described molecular functions²³. Consistent with this newly discovered role, Tra2 β protein is fairly evenly expressed across mouse tissues, so would be available in most cells for splicing inclusion of constitutive exons^{10,18}. Although we only detected exons activated by Tra2 proteins in this study, exons have previously been described that are repressed by Tra2 proteins^{29,41}. Such repressed exons might have either eluded our search criteria or occur less frequently.

Genes containing Tra2-dependent exons are enriched in processes associated with chromosome biology. Although GO annotations are known to be incomplete and can differ in accuracy⁴², this pattern of functional enrichment observed in our data is notably coherent. A potential association between the Tra2-dependent exons and chromosome biology is additionally supported by the connectivity of the BioGRID functional interaction network. In addition to the regulated *CHEK1* alternative splice, 6/17 of the strongest Tra2 protein-responsive exons (showing > 40 point PSI change after Tra2 depletion) were also in genes involved in chromosome structure and epigenetic regulation, including *MSL3* (– 41 point PSI exon switch after Tra2 depletion). *MSL3* is the human ortholog⁴³ of the *Drosophila melanogaster* *MSL3* gene, which regulates chromatin remodelling during sex determination, and is under control of Sex Lethal (a protein just upstream of Tra2 in the *Drosophila* sex determination

pathway) in flies. The *SMC4* gene (–66 point PSI exon switch after Tra2 depletion) encodes a protein important for DNA repair and chromosome condensation, and also interacts with the CTCF transcription factor that modifies chromatin structure⁴⁴. The *ANKRD1* (–73 point PSI exon switch after Tra2 depletion) encodes a transcription factor, which is a negative regulator of cardiac genes⁴⁵. Also among the genes with highly Tra2 protein-responsive exons was *GLYR1* (–41 point PSI exon switch after Tra2 depletion), which is a cofactor for histone demethylation⁴⁶, and the zinc-finger protein *ZCCHC9* (–68 point PSI exon switch after Tra2 depletion), which targets histone mRNAs for degradation⁴⁷. *ZCCHC7* (–72 point PSI exon switch after Tra2 depletion) and *ZCCHC11* (–46 point PSI exon switch after Tra2 depletion) encode zinc-finger proteins homologous to *ZCCHC9*, but with roles in non-coding RNA metabolism^{47–49}. The *MPHOSPH10* gene (–42 point PSI exon switch after Tra2 depletion) encodes a protein involved in ribosomal RNA processing in interphase, and is associated with chromosomes during mitosis⁵⁰. Other genes involved in chromatin modification that are controlled by Tra2 proteins but did not fit into the most responsive group include the *NASP* (*NASP-T*, –12 point PSI exon switch after Tra2 depletion) and *ATRX* genes (–23 point PSI exon switch after Tra2 depletion)^{18,51}. Interestingly, the two known *Drosophila* Tra2 splicing targets *Doublesex* and *Fruitless* are both transcription factors¹², and one of the major functions of the *Drosophila* sex determination pathway is dosage compensation via chromatin modification. Our data lend further support to the association of particular splicing regulators with the regulation of coherent cellular functions, also described for NOVA, RBFox2, PTB and T-STAR^{4–7}.

Our data indicate a high degree of functional redundancy between Tra2 α and Tra2 β , and a powerful homeostatic repressive feedback activity of Tra2 β over Tra2 α that buffers splicing changes when just one of these proteins is missing. Although splicing defects after single Tra2 α and Tra2 β depletion were small, they were often individually statistically significant (for example, in the *ATRX* gene). Such fine tuning of splicing profiles by joint Tra2 protein concentration by splicing feedback control might be important in whole organisms and at particular points of development (for example, in brain or testis development). Even individually, small splicing defects over many Tra2 β -target exons might cumulatively cause physiological defects. This might explain why *Tra2b*-null mice are embryonic lethal despite containing a *Tra2a* gene^{15,17,18}. Similar asymmetric expression patterns, in which a dominantly expressed splicing factor cross-regulates other family member proteins, have been found in the PTB family, where PTBP1 cross-regulates PTBP2 and PTBP3 (ref. 53). Comprehensive identification of PTBP1 targets similarly required joint depletion of PTBP1 and PTBP2 (ref. 53) Future studies of Tra2 β -regulated splicing may also benefit by considering expression levels of both Tra2 proteins, rather than in the context of Tra2 β expression alone.

Joint depletion of both Tra2 protein levels reduced cell viability in MDA-MB-231 cells, likely at least in part because of the requirement for productive splicing of the *CHEK1* mRNA. CHK1 protein expression is critical to reduce replication stress in cancer cells undergoing rapid proliferation driven by oncogenes including *RAS* and *MYC*^{54,55}. Our data thus suggest the Tra2 proteins may represent novel targets to inhibit cancer cell growth.

Methods

Cell culture. MDA-MB-231 cells and MCF7 cells were maintained in DMEM (no phenol red) plus 10% fetal bovine serum and 1% Penicillin Streptomycin. HEK-293, HeLa and PC3 cells were maintained in DMEM plus 10% fetal bovine serum. Cells lines were originally purchased from the American Type Culture Collection and IGC Standards, Europe.

iCLIP. Triplicate iCLIP experiments were performed following the iCLIP protocol³³. Briefly, MDA-MB-231 cells were irradiated with 400 mJ cm^{–2} ultraviolet-C light on ice, lysed and subject to partial RNase digestion. The crosslinked Tra2 β -RNA complexes were then immunoprecipitated using Protein A Dynabeads (Invitrogen) and a rabbit polyclonal anti-Tra2 β antibody (Abcam, ab31353). cDNA libraries were prepared according to the published iCLIP protocol. High throughput sequencing of cDNA libraries was performed using an Illumina GAIIX.

RNA-seq. RNA was extracted from cells using RNeasy Plus Mini Kit (Qiagen) following manufacturer's instructions and re-suspended in nuclease-free water. All RNA samples were DNase treated using DNA-free kit (Invitrogen) and stored at –80 °C prior to RNA quality control check using 2100 Agilent Bioanalyser and mRNA library prep using TruSeq mRNA library kit (Illumina). Pair-end sequencing was done in total for six samples (three biological replicates of negative control siRNA-treated cells and three biological replicates from *TRA2A* and *TRA2B* siRNA-treated cells) using an Illumina HiSeq 2000.

Bioinformatics (iCLIP and RNA-seq analysis). iCLIP data analysis, crosslink site identification and quantification, randomization of iCLIP positions and pentamer enrichment analysis were performed according to published procedures³³. Briefly, we used the human genome annotation version hg19, and gene annotations from Ensembl 59. Experiment barcode and random barcodes were registered and removed from iCLIP reads. After trimming, we ignored reads shorter than 11 nucleotides. Remaining trimmed reads were then mapped using Bowtie⁵⁶, allowing two mismatches and accepting only reads with single hits. Crosslink sites were initially identified as the first nucleotide upstream of the iCLIP tag, and then filtered to determine statistically significant crosslink sites and those which occurred in clusters within 15 nucleotides windows and with a significant iCLIP tag count, compared with randomized positions, as described in König *et al.*⁵⁶. For RNA-seq analysis, the base quality of raw sequencing reads were checked with FastQC (ref. 57) and refined with Seqtk (ref. 58) and Trim-galore (ref. 59). Reads were mapped to the hg19 reference with Tophat2 (ref. 60) and matches analyzed with Bedtools (ref. 61). Differentially expressed genes and exon usage were determined with DESeq (ref. 62) and DEXSeq (ref. 63) respectively.

siRNA transfection. Efficient knockdown of endogenous Tra2 α , Tra2 β and CHK1 proteins were achieved by transfecting cells with Silencer Select Pre-designed siRNAs (Ambion), targeting *TRA2A* mRNA (Ambion IDs: s26664 and s26665), *TRA2B* mRNA (Ambion IDs: s12749 and s12751) or *CHEK1* mRNA (Ambion ID: s503) respectively, with siPORT NeoFX Transfection Agent (Ambion). Control cells were transfected with a negative control siRNA (Ambion Cat#: 4390843). MDA-MB-231 cells grown in 100 mm tissue culture dishes were transfected with either 24 μ l of 10 μ M negative control siRNA (control), 12 μ l of 10 μ M siRNA targeting *TRA2A*, *TRA2B* or *CHEK1* (single Tra2 α , Tra2 β or CHK1 knockdown) or 12 μ l of 10 μ M siRNA targeting *TRA2A* and 12 μ l of 10 μ M siRNA targeting *TRA2B* (joint Tra2 α and Tra2 β knockdown). Cells were incubated for 72 h post siRNA transfection, before RNA extraction or western blotting.

Splicing assays: RNA extraction, RT-PCR and PCR. RNA was extracted using standard Trizol RNA extraction. cDNA was synthesized from 500 ng total RNA in a 10 μ l reaction, using Superscript VILO cDNA synthesis kit (Invitrogen) following manufacturer's instructions. Splicing profiles were monitored by PCR using primers in flanking exons. For each PCR, 1 μ l diluted cDNA (1/8) was used as template in a 10 μ l PCR reaction using Phusion High-Fidelity PCR Kit (NEB, UK) following manufacturer's instructions. Splicing profiles were monitored and quantified using the Qiagxel capillary electrophoresis system (Qiagen) and PSI was calculated as described previously¹⁸. All primers used for splicing assays are provided in Supplementary Data 3.

Quantitative PCR. Relative gene expression was determined by quantitative real-time PCR using the SYBR Green PCR Master Mix kit (Applied Biosystems) and an Applied Biosystems 7900HT Fast Real-Time PCR Machine. cDNA was generated from equal quantities of total RNA for each sample using Superscript VILO cDNA synthesis kit (Invitrogen) following manufacturer's instructions. Gene expression was calculated relative to three housekeeping genes *ACTB*, *GAPDH* and *TUBB*. C_t values for each sample were calculated using SDS 2.4 software (Applied Biosystems) and relative mRNA expression was calculated using the $2^{-\Delta\Delta C_t}$ method.

Calculation of Tra2 β binding site density. Tra2 β binding site density was calculated as the percentage of nucleotides within an exon that correspond to the top 10 kmers identified from the Tra2 β iCLIP experiments (Supplementary Fig. 2b).

Detection of proteins using western blotting. Endogenous proteins were detected by western blot analysis using the following primary antibodies and dilutions: Tra2 α (Novus Biologicals, H00029896-B01P;1:500 dilution), Tra2 β (Abcam, ab31353;1:2,000 dilution), CHEK1 (Proteintech, 10362-1-AP;1:250 dilution), Histone H2AX (Santa Cruz Biotechnology, sc54-606;1:500 dilution),

γ H2AX (Ser 139) (Santa Cruz Biotechnology, sc-101696;1:500 dilution), FLAG (Sigma-Aldrich, F3040;1:2,000 dilution), β -Actin (Sigma-Aldrich, A5441;1:2,000 dilution) and α -Tubulin (Sigma-Aldrich, T5168;1:2,000 dilution).

ANKRD1 and CHEK1 minigene construction and mutagenesis. The *ANKRD1* constitutive exon and ~200 nucleotides of flanking intronic region was amplified from human genomic DNA using the cloning primers ANKRD1 F (5'-AAAA AAAAAGAATTCAAATCTAAGACTTGTCTTATGGCATT-3') and ANKRD1 R (5'-AAAAAAAAGAATTCAAGATGAGAGTTACCGTGAGC-3'). The PCR products were digested with *Bam*HI restriction enzyme and cloned into the pXJ41 vector⁶⁴ using the *Mfe*I site midway through the 757 nucleotide rabbit β -globin intron 2. Tra2 β binding site mutations were made using site directed mutagenesis with the following primers: ANKRD1 M1F (5'- AGAACACATATCAAAGCTT GCACATTTATACGACCTTGA AAA-3'), ANKRD1 M1R (5'-CAAGTTCGTATAA ATGTGCAAGCTTTGATATGTGTTCTAG-3'), ANKRD1 M2F (5'-ATCATT CAACTGCAGCAACGGCAACAATACAGGCACACTAAAG-3') and ANKRD1 M2R (5'-GAACCTTTAGTGTGCTGTATTGTTGCCGTTGCTGCAGTTGAA TG-3'). The *CHEK1* alternative exon and approximately 250 nucleotides of flanking intronic region was synthesised *in vitro* and similarly cloned into the pXJ41 vector. A mutated version that disrupted Tra2 β binding sites was also synthesised (sequence provided in Fig. 7c) and cloned into the pXJ41 vector. Analysis of splicing patterns of mRNAs transcribed from minigenes was carried out in HEK-293 cells as previously described^{18,24}, using primers within the β -globin exons of pXJ41; PXJRTF (5'-GCTCCGGATCGATCCTGAGA ACT-3') and PXJB (5'-GCTGCAATAACAAGTTCTGCT-3').

EMSA. Gel shift experiments^{18,65} were performed using full-length Tra2 β protein and *in vitro*-translated RNA probes made from constructs containing amplified regions of the human *CHEK1* gene, cloned into the pBluescript vector. Three regions of the human *CHEK1* gene were amplified using the following primers: CHEK1 intronic F (5'-AAAAAAAAGGTACCTGTGTACCTCTCCTTCACTA CC-3'), CHEK1 intronic R (5'-AAAAAAAAGAATTCTCTCTCAAGCTCCT ATGGGG-3'), CHEK1 exon region A F (5'-AAAAAAAAGGTACCgttcaactgc ttgtaataagagt-3'), CHEK1 exon region A R (5'-AAAAAAAAGAATTCggcagc CTTCAatctacaATCT-3'), CHEK1 exon region B F (5'-AAAAAAAAGGTACC agtaaaattcatgttcacagga-3') and CHEK1 exon region B R (5'-AAAAAAAAGAA TTCctcactacagtactcagaat-3').

GO and functional network analysis. GO⁶⁶ enrichment analysis was carried out using the Bioconductor GOSTats package version 2.24.0refs^{67,68}. Enrichments of GO biological process terms were calculated using the conditional hypergeometric test with a significance cut-off of 0.001 and using a background of genes that are normally expressed in MDA-MB-231 cells. Annotations were taken from the Bioconductor *Homo sapiens* annotation package org.Hs.eg.db version 2.8.0ref.⁶⁹. The analysis was run in the open source statistical package R version 3.0.1ref.⁷⁰.

Interaction data for *Homo sapiens* was retrieved from the BioGRID database (version 110). These data were integrated into a network in which nodes represented genes or gene products, and edges represented any type of BioGRID interaction between the nodes. The network was visualised using the Cytoscape visualization platform⁷¹, and was coloured based on annotations to top five enriched GO biological processes (as downloaded from QuickGO⁷¹). Where a protein was annotated to more than one term, the most specific annotation was chosen.

MTT assay. MTT assays were performed using MTT Cell Proliferation Assay Kit (Cayman Chemical), following manufacturer's instructions. An siRNA transfection mix was added to a suspension of $\sim 2 \times 10^5$ MDA-MB-231 cells in 10 ml media. The siRNA/cell suspension was gently mixed and a 100 μ l aliquot was added per well to a 96-well plate. Absorbance from the MTT assay was measured at 24, 48, 72, 96 and 120 h after siRNA transfection/seeding of cells. Relative density of cells was also compared 120 h after seeding cells by microscopy.

Fluorescence-activated cell sorting analysis of EdU-positive cells. MDA-MB-231 cells were incubated with 10 μ M EdU for 4 h, 96 h after siRNA transfection. Cell fixation, permeabilization and EdU detection was performed using the Click-iT EdU Flow Cytometry Assay Kit (Life Technologies) following the manufacturer's instructions. Data were collected and analysed using a BD LSR II flow cytometer using 488 nm excitation and a 520/20 band-pass for detection of EdU Alexa Fluor488 azide and 355 nm excitation and a 450/50 band-pass for detection of 4',6-diamidino-2-phenylindole. Experiments were performed with biological triplicate samples and 30,000 cells were analysed per sample. A no-EdU control sample was used to inform our gating strategy to calculate the proportion of EdU-positive cells.

Analysis of nuclear morphology. To investigate nuclear morphology, cells were fixed with 4% paraformaldehyde followed by nuclear staining with 4',6-diamidino-2-phenylindole, 96 h after siRNA transfection (siRNA transfection as described above).

Elutriation and cell cycle evaluation. Elutriation: Cells were size fractionated by centrifugal elutriation, using flow rates of 10, 13, 17, 20, 24 and 28 ml min⁻¹ (ref. 72). Cell cycle evaluation: Cell cycle phase enrichment of cells was assessed using immunofluorescence staining for CENPF (late S, G₂, G₂/M)⁷³ and phospho-histone H3S10 (G₂/M, M)⁷⁴. Asynchronous and elutriated KG1 cells were suspended in PBS, spotted onto poly-lysine-coated slides and processed for immunofluorescence. Images were captured and cells were scored for CENPF and phospho-H3S10 staining.

Generation of tetracycline-inducible HEK-293 cells. Full-length CHK1-FLAG cDNA was amplified from the pcDNA4-Chk1-Flag plasmid (Addgene plasmid #22894) using the primers CHEK1 FLAG F (5'-AAAAAAAAGCGGCCGC atggcagtcctctgtggaagac-3') and CHEK1 FLAG R (5'-AAAAAAAAGTTCGAC tcatgtggcaggaagcacaatcttc-3') and cloned into the Flp-In expression vector (pCDNA5). To generate an inducible cell line, the CHK1-FLAG-pCDNA5 vector was cotransfected with the Flp recombinase plasmid (pOG44) into Flp-In HEK-293 cells and selected for using treatment with Hygromycin B. Following Hygromycin B selection, CHK1-FLAG expression was induced by the addition of tetracycline to promote expression CHK1-FLAG expression via a tetracycline-inducible promoter.

References

1. Fox-Walsh, K. L. & Hertel, K. J. Splice-site pairing is an intrinsically high fidelity process. *Proc. Natl Acad. Sci. USA* **106**, 1766–1771 (2009).
2. Kelemen, O. *et al.* Function of alternative splicing. *Gene* **514**, 1–30 (2013).
3. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
4. Ehrmann, I. *et al.* The tissue-specific RNA binding protein T-STAR controls regional splicing patterns of neurexin pre-mRNAs in the brain. *PLoS Genet* **9**, e1003474 (2013).
5. Venables, J. P. *et al.* RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. *Mol. Cell Biol.* **33**, 396–405 (2013).
6. Ule, J. *et al.* Nova regulates brain-specific splicing to shape the synapse. *Nat. Genet.* **37**, 844–852 (2005).
7. Zhang, C. *et al.* Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* **22**, 2550–2563 (2008).
8. Llorian, M. *et al.* Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat. Struct. Mol. Biol.* **17**, 1114–1123 (2010).
9. Tsuda, K. *et al.* Structural basis for the dual RNA-recognition modes of human Tra2-beta RRM. *Nucleic Acids Res.* **39**, 1538–1553 (2011).
10. Elliott, D. J., Best, A., Dalglish, C., Ehrmann, I. & Grellscheid, S. How does Tra2beta protein regulate tissue-specific RNA splicing? *Biochem Soc Trans* **40**, 784–8 (2012).
11. Baker, B. S. Sex in flies: the splice of life. *Nature* **340**, 521–524 (1989).
12. Forch, P. & Valcarcel, J. Splicing regulation in *Drosophila* sex determination. *Prog. Mol. Subcell. Biol.* **31**, 127–151 (2003).
13. Dauwalder, B., Amaya-Manzanares, F. & Mattox, W. A human homologue of the *Drosophila* sex determination factor transformer-2 has conserved splicing regulatory functions. *Proc. Natl Acad. Sci. USA* **93**, 9004–9009 (1996).
14. Beil, B., Scream, G. & Stamm, S. Molecular cloning of htra2-beta-1 and htra2-beta-2, two human homologs of tra-2 generated by alternative splicing. *DNA Cell Biol.* **16**, 679–690 (1997).
15. Mende, Y. *et al.* Deficiency of the splicing factor Sfrs10 results in early embryonic lethality in mice and has no impact on full-length SMN/Smn splicing. *Hum. Mol. Genet.* **19**, 2154–2167 (2010).
16. Roberts, J. M. *et al.* Splicing factor TRA2B is required for neural progenitor survival. *J. Comp. Neurol.* **522**, 372–392 (2014).
17. Storbeck, M. *et al.* Neuronal-specific deficiency of the splicing factor tra2b causes apoptosis in neurogenic areas of the developing mouse brain. *PLoS ONE* **9**, e89020 (2014).
18. Grellscheid, S. *et al.* Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2beta in development. *PLoS Genet.* **7**, e1002390 (2011).
19. Best, A. *et al.* Expression of Tra2 beta in cancer cells as a potential contributory factor to neoplasia and metastasis. *Int. J. Cell Biol.* **2013**, 843781 (2013).
20. Hofmann, Y., Lorson, C. L., Stamm, S., Androphy, E. J. & Wirth, B. Htra2-beta 1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2). *Proc. Natl Acad. Sci. USA* **97**, 9618–9623 (2000).
21. Glatz, D. C. *et al.* The alternative splicing of tau exon 10 and its regulatory proteins CLK2 and TRA2-BETA1 changes in sporadic Alzheimer's disease. *J. Neurochem.* **96**, 635–644 (2006).
22. Long, J. C. & Caceres, J. F. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* **417**, 15–27 (2009).
23. Zhou, Z. & Fu, X. D. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* **122**, 191–207 (2013).

24. Venables, J. P. *et al.* Up-regulation of the ubiquitous alternative splicing factor Tra2beta causes inclusion of a germ cell-specific exon. *Hum. Mol. Genet.* **14**, 2289–2303 (2005).
25. Busch, A. & Hertel, K. J. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip. Rev. RNA* **3**, 1–12 (2012).
26. Tacke, R., Tohyama, M., Ogawa, S. & Manley, J. L. Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* **93**, 139–148 (1998).
27. Clery, A. *et al.* Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-beta1. *Nat. Struct. Mol. Biol.* **18**, 443–450 (2011).
28. Uren, P. J. *et al.* Site identification in high-throughput RNA-protein interaction data. *Bioinformatics* **28**, 3013–3020 (2012).
29. Anderson, E. S. *et al.* The cardiotonic steroid digitoxin regulates alternative splicing through depletion of the splicing factors SRSF3 and TRA2B. *RNA* **18**, 1041–1049 (2012).
30. Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**, 926–929 (2007).
31. Stoilov, P., Daoud, R., Nayler, O. & Stamm, S. Human tra2-beta1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. *Hum. Mol. Genet.* **13**, 509–524 (2004).
32. Saltzman, A.L., Pan, Q. & Blencowe, B. J. Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev* **25**, 373–84 (2011).
33. Konig, J. *et al.* iCLIP—transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J. Vis. Exp.* **50**: pii 2638 (2011).
34. Dreszer, T. R. *et al.* The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res.* **40**, D918–D923 (2012).
35. Meyer, L. R. *et al.* The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.* **41**, D64–D69 (2013).
36. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539 (2006).
37. Pabla, N., Bhatt, K. & Dong, Z. Checkpoint kinase 1 (Chk1)-short is a splice variant and endogenous inhibitor of Chk1 that regulates cell cycle and DNA damage checkpoints. *Proc. Natl Acad. Sci. USA* **109**, 197–202 (2012).
38. Gagou, M. E., Zuazua-Villar, P. & Meuth, M. Enhanced H2AX phosphorylation, DNA replication fork arrest, and cell death in the absence of Chk1. *Mol. Biol. Cell* **21**, 739–752 (2010).
39. Syljuasen, R. G. *et al.* Inhibition of human Chk1 causes increased initiation of DNA replication, phosphorylation of ATR targets, and DNA breakage. *Mol. Cell Biol.* **25**, 3553–3562 (2005).
40. Melamud, E. & Moul, J. Stochastic noise in splicing machinery. *Nucleic Acids Res.* **37**, 4873–4886 (2009).
41. Chandler, D. S., Qi, J. L. & Mattox, W. Direct repression of splicing by transformer-2. *Mol. Cell Biol.* **23**, 5174–5185 (2003).
42. Skunca, N., Altenhoff, A. & Dessimoz, C. Quality of computationally inferred gene ontology annotations. *PLoS Comput. Biol.* **8**, e1002533 (2012).
43. Smith, E.R. *et al.* A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol Cell Biol* **25**, 9175–88 (2005).
44. Huang, K. *et al.* Ribosomal RNA gene transcription mediated by the master genome regulator protein CCCTC-binding factor (CTCF) is negatively regulated by the condensin complex. *J Biol Chem* **288**, 26067–77 (2013).
45. Zolk, O. *et al.* Cardiac ankyrin repeat protein, a negative regulator of cardiac gene expression, is augmented in human heart failure. *Biochem Biophys Res Commun* **293**, 1377–1382 (2002).
46. Fang, R. *et al.* LSD2/KDM1B and its cofactor NPAC/GLYR1 endow a structural and molecular model for regulation of H3K4 demethylation. *Mol. Cell* **49**, 558–570 (2013).
47. Schmidt, M. J., West, S. & Norbury, C. J. The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. *RNA* **17**, 39–44 (2011).
48. Fasken, M. B. *et al.* Air1 zinc knuckles 4 and 5 and a conserved IWRXY motif are critical for the function and integrity of the Trf4/5-Air1/2-Mtr4 polyadenylation (TRAMP) RNA quality control complex. *J. Biol. Chem.* **286**, 37429–37445 (2011).
49. Hagan, J. P., Piskounova, E. & Gregory, R. I. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **16**, 1021–1025 (2009).
50. Westendorf, J.M. *et al.* M phase phosphoprotein 10 is a human U3 small nucleolar ribonucleoprotein component. *Mol Biol Cell* **9**, 437–49 (1998).
51. Eustermann, S. *et al.* Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. *Nat. Struct. Mol. Biol.* **18**, 777–782 (2011).
52. Ule, J. *et al.* An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**, 580–586 (2006).
53. Spellman, R., Llorian, M. & Smith, C. W. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol. Cell* **27**, 420–434 (2007).
54. Bartek, J., Mistrik, M. & Bartkova, J. Thresholds of replication stress signaling in cancer development and treatment. *Nat. Struct. Mol. Biol.* **19**, 5–7 (2012).
55. Murga, M. *et al.* Exploiting oncogene-induced replicative stress for the selective killing of Myc-driven tumors. *Nat. Struct. Mol. Biol.* **18**, 1331–1335 (2011).
56. Konig, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**, 909–15 (2010).
57. Andrews, S. FastQC A quality control tool for high throughput sequence data. Available at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
58. Li, H. seqtk. Available at <https://github.com/lh3/seqtk>.
59. Krueger, F. Trim Galore! Available at http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
60. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36 (2013).
61. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
62. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biology* **11**, R106 (2010).
63. Anders, S. *et al.* Detecting differential usage of exons from RNA-seq data. *Genome Research* **22**, 2008–2017 (2012).
64. Bourgeois, C. F., Popielarz, M., Hildwein, G. & Stevenin, J. Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5' or 3' splice site activation. *Mol. Cell. Biol.* **19**, 7347–7356 (1999).
65. Greltscheid, S. N. *et al.* Molecular design of a splicing switch responsive to the RNA binding protein Tra2beta. *Nucleic Acids Res.* **39**, 8092–8104 (2011).
66. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
67. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
68. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
69. Carlson, M. org.Hs.eg.db: Genome wide annotation for Human. R package version 2.8.0. (2012).
70. Rcoreteam. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
71. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
72. Ly, T. *et al.* A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *Elife* **3**, e01630 (2014).
73. Liao, H., Winkfein, R. J., Mack, G., Rattner, J. B. & Yen, T. J. CENP-F is a protein of the nuclear matrix that assembles onto kinetochores at late G2 and is rapidly degraded after mitosis. *J. Cell Biol.* **130**, 507–518 (1995).
74. Hendzel, M. J. *et al.* Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* **106**, 348–360 (1997).

Acknowledgements

We thank Professor Jernej Ule and Dr Julian Konig for help with the iCLIP analysis, and Professor Ian Eperon, Dr Ingrid Ehrmann and Dr Jennifer Munkley for comments on the manuscript. Andrew Best was a Breast Cancer Campaign funded PhD student. This project was supported by the Breast Cancer Campaign (Grant 2009NovPhd21), the Wellcome Trust (Grant numbers WT080368MA and WT089225/Z/09/Z), the BBSRC (grant numbers BB/D013917/1 and BB/I006923/1), a Canadian Institute of Health Research grant to B.C. (MOP93917) and the Leukaemia and Lymphoma Research Specialist Programme Grant 12031 (to C.A.A. and I.G.C.). B.K. was a BHF Professor of Cardiology.

Author contributions

A.B. performed iCLIP, RNA-seq, splicing and cellular analyses, and performed experiments; K.J. performed GO and BioGrid analyses; C.D. performed EMSAs; E.H. helped with iCLIP experiments; M.K.-K. made stable HEK-293 cell lines; T.C. performed iCLIP sequence analysis; Y.X. and M.S.K. performed RNA-seq bioinformatic analysis; M.D. made minigenes; R.H. and B.K. performed iCLIP sequencing; A.W. helped with analysis; J.P.V., R.K. and B.C. analysed the role of multiple RNA binding proteins for CHEK1, SMC4, GLYR1 and ANKRD1; I.G.C., K.C.L. and C.A.A. purified and characterized KG1 cell fractions; A.B., A.T.-C. and D.J.E. were involved in study design and data analysis. A.B. and D.J.E. prepared the manuscript. All authors discussed the results and commented on the manuscript.

Additional information

Accession codes: iCLIP and RNA-seq data available via the Gene Expression Omnibus (GEO) using GEO accession number GSE59335.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Best, A. *et al.* Human Tra2 proteins jointly control a *CHEK1* splicing switch amongst alternative and constitutive target exons. *Nat. Commun.* 5:4760 doi: 10.1038/ncomms5760 (2014).



This work is licensed under a Creative Commons Attribution 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/4.0/>