Published in final edited form as:
Nature. ; 484(7394): 386–389. doi:10.1038/nature10925.

DBIRD integrates alternative mRNA splicing with RNA polymerase II transcript elongation

Pierre Close[1,2], **Philip East**[3], **A. Barbara Dirac-Svejstrup**[1], **Holger Hartmann**[4], **Mark Heron**[4], **Sarah Maslen**[5], **Alain Chariot**[2], **Johannes Söding**[4], **Mark Skehel**[5], and **Jesper Q. Svejstrup**[1,‡]

[1]Mechanisms of Transcription Laboratory, Cancer Research UK London Research Institute, Clare Hall Laboratories, South Mimms, EN6 3LD, UK

[2]Unit of Medical Chemistry, GIGA-Signal Transduction, GIGA-R, University of Liege, CHU, Sart-Tilman, Liege, Belgium

[3]Bioinformatics & Biostatistics Group, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London, WC2A 3LY

[4]Gene Center and Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany

[5]Protein Analysis and Proteomics Laboratory, Clare Hall Laboratories, Cancer Research UK, London Research Institute, South Mimms, EN6 3LD, UK

Abstract

Alternative mRNA splicing is the main reason vast mammalian proteomic complexity can be achieved with a limited number of genes. Splicing is physically and functionally coupled to transcription, and is greatly affected by the rate of transcript elongation[1,2,3]. As the nascent pre-mRNA emerges from transcribing RNA polymerase II (RNAPII), it is assembled into a messenger ribonucleoprotein (mRNP) particle which is its functional form and determines the fate of the mature transcript[4]. However, factors that connect the transcribing polymerase with the mRNP particle and help integrate transcript elongation with mRNA splicing remain obscure. Here, we characterized the interactome of chromatin-associated mRNP particles. This led to the identification of Deleted in Breast Cancer 1 (DBC1) and a protein we named ZIRD as subunits of a novel protein complex, named DBIRD, which binds directly to RNAPII. DBIRD regulates alternative splicing of a large set of exons embedded in A/T-rich DNA, and is present at the affected exons. RNAi-mediated DBIRD depletion results in region-specific decreases in transcript elongation, particularly across areas encompassing affected exons. Together, these data indicate that DBIRD complex acts at the interface between mRNP particles and RNAPII, integrating transcript elongation with the regulation of alternative splicing.

The composition of mRNP particles has been the subject of a number of studies, using a variety of approaches (see, for example, ref[5] and references therein). There are likely to be different types of mRNP particles with distinct compositions and interaction partners. We

---

‡Correspondence and requests for material should be addressed to J.Q.S. (j.svejstrup@cancer.org.uk)..

Author Contributions

P.C. and A.B.D.S. performed experiments, and S.M. and M.S. did mass spec analyses. P.E., H.H., M.H., and J.S. performed bioinformatic analyses. P.C. and J.Q.S. designed the study, analyzed the experimental data, and wrote the paper. All authors discussed the results and commented on the manuscript.

Author information

Gene-expression and splicing data sets have been deposited in the GEO database under accession code GSE35480.

Europe PMC Funders Author Manuscripts

sought to specifically purify native mRNP particles and interacting proteins from the chromatin in which they are generated and active in co-transcriptional processes. As a starting point, we generated HEK293 cells expressing near-normal levels of Flag-tagged hnRNP A1 (A1), an abundant hnRNP protein in human cells[6]. A1 shuttles between the nucleus and the cytoplasm[7], but at steady state it is mainly nuclear and concentrated in chromatin (Fig. 1a), from where it can be released by RNAse A treatment (Fig. 1b, compare lanes 2 and 4). We used DNAse I digestion and mild sonication to release mRNP particles from chromatin for purification. RNAse inhibitors were present during the whole process, outlined in Fig. 1c. mRNP particles isolated by this approach are predominantly of nuclear (chromatin) origin (Suppl. Fig. S1). Native mRNP particles and their interacting partners were purified from chromatin isolated from ~10^8 nuclei (Fig. 1d). Only two major bands (namely the added, proteinaceous RNAse inhibitors) were detected upon purification from control cells (left panel), while numerous proteins were detected in A1-flag elutions (right panel). These represent a heterogeneous mixture of core mRNP particle subunits and proteins interacting with such particles. Individual protein bands were excised and identified by mass spectrometric analysis (subset indicated in Fig. 1d; see also list in Suppl. Fig S2a). Most of the known ‘core’ mRNP proteins, such as the hnRNP proteins were present in the purified fraction, confirming the biological relevance of this approach. Many other pre-mRNA processing proteins were also identified, including splicing factors, ATP-dependent RNA helicases, and a substantial number of mRNA 3′-end processing and termination factors. Co-immunoprecipitation (co-IP) experiments confirmed the RNA-dependent interaction of some of these proteins with A1-Flag (Suppl. Fig. S2b).

We next focused on two proteins that had not previously been connected to mRNP particles or mRNA processing. One of these, DBC1, is otherwise best known for its association with and regulation of the sirtuin-like deacetylase SIRT1[8,9]. We also investigated the uncharacterized zinc finger-containing protein ZNF326. Stable cell lines expressing near-normal levels of Flag-tagged versions of these proteins were established, and co-IP experiments confirmed that both DBC1 and ZNF326 interact with mRNP particles in an RNA-dependent manner (Suppl. Fig. S3a-f). Furthermore, we discovered that ZNF326 and DBC1 associate directly, in an RNA-independent manner (Fig. 2a and e). We therefore renamed ZN326 as ZIRD (ZNF-protein Interacting with nuclear RNPs and DBC1).

We previously identified DBC1 as an RNAPII-interacting protein in another proteomic screen[10], making it a particularly interesting candidate. Co-IP experiments confirmed that RNAPII associates with DBC1-Flag, in an RNA-independent manner (Fig. 2b). ZIRD was detected in RNAPII (RPB3-Flag) purifications as well, and this interaction was also RNA-independent (Fig. 2c). In further support of a ZIRD-RNAPII interaction, ZIRD-Flag also brought down RNAPII (Fig. 2d). In contrast, we failed to detect an interaction between A1 and RNAPII under the same conditions (Fig. 2c, middle panel, and data not shown), although co-IP experiments after formaldehyde cross-linking indicated that, as expected, the proteins are in close proximity *in vivo* (Suppl. Fig. S4). Together, these results indicate that DBC1 and ZIRD are not part of the core mRNP particle, but that they might work at the interface between the mRNP particle and RNAPII.

Others reported that DBC1 interacts with SIRT1[8,9]. While we confirmed that DBC1 co-precipitates SIRT1, endogenous ZIRD and ZIRD-Flag did not (Fig. 2e, and data not shown). SIRT1 is also absent from A1-containing mRNP particles (Suppl. Fig. S3g). This indicates that ZIRD and DBC1 form a complex that lacks SIRT1. To further characterize ZIRD-DBC1 interaction, ZIRD-Flag was purified. Size exclusion chromatography of highly purified material showed that ZIRD-Flag and DBC1 are part of a salt-stable ~800 kDa complex (Fig. 2f, upper two panels), which also co-purified on MonoQ (data not shown). As

expected, SIRT1 is not part of this protein complex (Fig. 2g, and data not shown). We named the complex DBIRD (DBC1/ZIRD complex).

DBC1 and ZIRD interact with RNAPII in crude extracts (Fig. 2b-d). To investigate if this interaction is direct, DBIRD complex was characterized by gel-filtration after mixing with an excess of RNAPII. In the absence of RNAPII, the DBIRD complex peaked in fractions 13-15 (Fig. 2f, upper two panels). However, when mixed with RNAPII (Fig. 2f, lower three panels), DBIRD complex elution shifted to earlier eluting fractions, peaking in fraction 10 with a sub-fraction of RNAPII, while polymerase alone peaked in fractions 17-19 (~500 kDa), as expected. DBIRD complex thus appears to form a bridging complex, which interacts with both mRNP particles and RNAPII. Interestingly, DBIRD also interacted with mRNP particles lacking hnRNP A1 (Fig. 2h), pointing to a general bridging role.

To examine the role of the DBIRD complex in transcription-associated processes *in vivo*, we analysed the transcriptome of cells that had been depleted for DBC1 or ZIRD by RNAi (Suppl. Fig. 5a). Total mRNA was hybridised to GeneChip HUMAN EXON 1.0 ST arrays, on which the abundance of individual exons can be analyzed independently. In the absence of ZIRD, a >1.5 fold increase in exon inclusion was observed in more than 2800 situations, whereas exon exclusion was observed in only 390 cases (Suppl. Table T1a). The absence of DBC1 led to increased inclusion of an exon in 796 cases (Suppl. Table T1b), and, strikingly, the majority of these events were also on the list of ZIRD-dependent exon inclusions (567 of 796 = 71%; p-value for shared exons = 6.705e-261; Suppl. Table T1c), strongly supporting the close functional relationship between the two factors, and providing a high degree of confidence in the genome-wide alternative splicing data-sets. The effect was at the level of alternative splicing, as depletion of ZIRD or DBC1 only affected the expression of a very small number of genes (Suppl. Fig. S6).

A full list of inclusion events observed in both DBC1 and ZIRD-depleted cells is in Suppl. Table T1c. Sample results were confirmed by quantitative RT-PCR (Suppl. Fig. S7). To investigate whether DBIRD was present at affected exons, we performed RNA immunoprecipitation (RIP) experiments[11]. DBC1 and ZIRD bound the relevant exon in mRNAs from seven tested genes, while other regions (or control tRNA) were either not detected, or detected to a much lower extent (Fig. 3a and b; Suppl. Fig. S8). Interestingly, some exons of the β-actin gene (whose splicing was unaffected by DBIRD depletion) had significant levels of DBIRD complex (Suppl. Fig. S8), indicating that the interaction of DBIRD with mRNA is not invariably correlated with DBIRD-dependent splicing changes.

To investigate the mechanism underlying exon inclusion, we first searched for sequence motifs in the DNA encompassing the included exons, but failed to uncover other motifs than those known to typify splice junctions. We then looked for nucleotide patterns that might be overrepresented in the sequences surrounding the included exons by counting how often each of the 1024 possible 5-mer oligonucleotides occurred. Intriguingly, A/T-rich 5-mers were significantly enriched around included exons (Fig. 3c). The frequencies of the four nucleotides in the regions around the splice sites were also analyzed. A and T were strongly overrepresented around the splice sites of DBIRD-affected exons, as well as across the exons themselves. (Fig. 3d). The observed difference in A/T content is sufficient to explain the over-representation of A/T-rich 5-mers (Suppl. Fig. S9).

The A/T-rich DNA surrounding the affected exons might influence fundamental aspects of transcription. Indeed, A- and T-tracts are difficult for RNAPII to transcribe, constituting very efficient elongation pause sites *in vitro*[12,13]. To investigate the effect of DBIRD on transcript elongation, we performed RNAPII chromatin-immunoprecipitation (ChIP) analysis after DBIRD knockdown. As control, we also knocked down SIRT1 (Suppl. Fig.

5b). Remarkably, although overall transcription of RAD50 and SLC36A4 is not affected (see Suppl. Fig. S7), depletion of DBC1 or ZIRD (but not SIRT1) dramatically affected RNAPII transcription, distinctively in regions encompassing affected exons (Fig. 4; Suppl. Fig. S11). Quantification of newly produced mRNA by Bromo-UTP incorporation/ immunoprecipitation supported the idea that elongation rates were decreased in these regions (Suppl. Fig. S10). DBIRD depletion also affected RNAPII density at other genes whose splicing was exon-specifically affected, while little or no change in RNAPII density was observed at the unaffected β-actin control gene, even at exons that had an elevated DBIRD level (Suppl. Fig. S11-12; Compare to Suppl. Fig. S8).

Our data support the idea that the DBIRD complex represents a new type of factor, which functions at the interface between 'core' mRNP particles and RNAPII, affecting local transcript elongation rates and alternative splicing at a subset of A/T-rich exon-intron junctions (Graphic model in Suppl. Fig. S13). Interestingly, several studies have shown that the rate of RNAPII elongation affects the efficiency of splicing, with slow elongation favouring exon inclusion[1,3]. One possible explanation for our data is thus that DBIRD complex acts as an elongation factor, which facilitates transcript elongation across A/T-rich regions, and thereby affects alternative splicing of exons in these regions. It has also been suggested that exons in the nascent pre-mRNA become tethered to the elongating transcription complex[14,15]. Given that DBIRD binds both mRNPs and RNAPII, it might affect such tethering as well, and thereby splicing.

Interestingly, DBC1 has been implicated in tumourigenesis as a potential tumour suppressor, regulating apoptosis and cell survival[16]. Whether DBC1's role in the DBIRD complex and alternative splicing impacts on tumourigenesis is an interesting possibility, especially in light of the recent finding that genes encoding components of the splicing machinery are often mutated in myelodysplastic syndromes and related disorders[17]. ZIRD has not previously been characterized in human cells, but its mouse homologue, ZAN75, is highly expressed in neuronal tissues[18], suggesting that regulation of DBIRD complex might contribute to tissue-specific splicing. Other proteins with homology to ZIRD and DBC1 exist in the human genome, raising the intriguing possibility that other DBIRD-like complexes are specific for other sets of genes or exons, or are involved in other transcription-related nuclear events.

METHODS SUMMARY

ORFs encoding A1, DBC1 and ZIRD were cloned into pIRESpuro (Clontech) with a C-terminal Flag tag. HEK293 cells were grown in DMEM containing 10% FBS in 5% CO_2 at 37°C. For proteomic analysis, nuclei were isolated from A1-Flag cells. These were sonicated, DNAse I treated, and the sample cleared by centrifugation and the supernatant subjected to M2 agarose (Sigma) chromatography. Proteins were eluted with 3xFLAG peptide, and mass spectrometry performed as described[10]. DBIRD was purified by M2 agarose chromatography from nuclease-treated nuclear extract from cells expressing ZIRD-Flag. DBIRD was analyzed by MonoQ, or size exclusion chromatography with or without an excess of RNAPII. Stealth siRNAs were double transfected in HEK293 cells using lipofectamine 2000 (Invitrogen). For microarray analysis, RNA was hybridized on Human Exon 1.0 ST arrays (Affymetrix) using standard techniques (bioinformatics analysis described in Full Methods). For assessment of exon abundance and transcript expression, quantitative RT-PCR was performed using primers against affected and unaffected exons. Primer details are available on request. RNA immunoprecipitation and ChIP assays were performed as described[11,19].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants from Cancer Research UK and European Research council (ERC) (to J.Q.S.), and by an EMBO long-term fellowship and by "Fonds Leon Fredericq" foundation (to P.C.) We thank the Molecular Biology Core Facility at the Paterson Institute (Manchester), and Cell Services at LRI for analysis and help. Benoit Chabot is thanked for CB3 cells. Members of the Svejstrup laboratory, Peter Verrijjzer, and Torben Heick Jensen are thanked for comments on the manuscript. P.C. is presently Senior Research Assistant at the Belgian National Foundation for Scientific Research (FNRS).

Appendix

FULL METHODS

Plasmids and antibodies

ORFs encoding human A1, DBC1 and ZIRD (ZNF326) were cloned into pIRESpuro (Clontech) with a Flag tag at the C-terminus. Antibodies used were rabbit anti-Flag, mouse anti-Flag M2 and mouse anti-hnRNP C (Sigma); and mouse anti-pCTD mAb 4H8 (Millipore); rabbit anti-lamin B2 (Acris); rabbit anti-ZNF326 and mouse anti-A1 (Santa Cruz Biotechnology); rabbit anti-DBC1 and rabbit anti-SIRT1 (Bethyl Laboratories).

Cell culture, stable cell line establishment and Stealth siRNAs transfection

HEK293 cells were grown in DMEM containing 10% FBS in 5% CO_2 at 37°C. To generate HEK293 stably expressing a Flag-tagged protein, cells were transfected with the relevant pIRESpuro construct and selected in 1μg/ml puromycin (Sigma). Cells were maintained in selecting media for 3 weeks, and surviving cells used for experiment after transgene expression was checked.

Stealth siRNAs were double transfected in HEK293 cells using lipofectamine 2000 (Invitrogen) according to the manufacturers instructions. Protein/RNA expression was checked 48 h after the second transfection. RNAi sequences were the following: Stealth siRNA anti-ZIRD: sense: 5′CGGAGGUAGUUAUGGUGGUCGAUUU3′; antisense: 5′AAAUCGACCACCAUAACUACCUCCG-3′ Stealth siRNA anti-DBC1: Sense: 5′-CCAUCUGUGACUUCCUAGAACUCCA3′; Antisense: 5′UGGAGUUCUAGGAAGUCACAGAUGG3′. Stealth siRNA anti-SIRT1: Validated stealth siRNA from Invitrogen; Oligo ID: VHS50609. Control siRNA: Stealth siRNA Negative Control Med GC (Invitrogen #12935-300).

Immunopurification of Native mRNPs

10^8 cells stably expressing A1-Flag were lyzed with cytoplasmic lysis buffer (10 mM Tris HCl pH 7.9, 340 mM sucrose, 3 mM $CaCl_2$, 2 mM $Mg(OAc)_2$, 0.1 mM EDTA, 1 mM DTT, 0.5% NP-40, protease inhibitors and 1 μl/ml RNasin Ribonuclease inhibitor (promega)) and intact nuclei were pelleted by centrifugation at 3,500g for 15 min. Nuclei were washed with cytoplasmic lysis buffer without NP-40 and then resuspended in DNAse I buffer (20 mM Hepes pH 7.9, 10% glycerol, 1.5 mM MgCl2, 1 mM DTT, protease inhibitors and 1μl/ml RNasin). After 10 strokes in a Dounce homogenizer, nuclei were sonicated using Bioruptor® (diagenode), before DNAse I (Sigma) was added to the buffer and incubated 30 min at room temperature. Buffer was then brought to 250 mM KOAc and 1% triton X-100. The sample was cleared by centrifugation at 20,000g for 30 min and the supernatant was

collected. For negative control purification, the same extracts were prepared from the same amount of untagged cells.

The sample was then applied to M2 agarose beads (Sigma) and incubated for 4h at 4°C. After binding, beads were washed extensively with washing buffer (20 mM Hepes pH7.9, 250 mM KOAc, 1% Triton X-100, 10% glycerol, 3 mM EDTA, 1 mM DTT, protease inhibitors and 1 μl/ml RNasin® Ribonuclease inhibitor). Finally, proteins were eluted by using FLAG elution buffer (20 mM Hepes pH7.9, 100 mM KOAc, 3 mM EDTA, 1 mM DTT, 200 μg/ml 3xFLAG peptide, protease inhibitors and 1μl/ml RNasin® Ribonuclease inhibitor). Eluates were resolved by 4-12% bis-Tris gradient SDS PAGE and revealed by Sypro Ruby staining (Invitrogen).

Mass Spectrometric Analysis

Protein samples were reduced, alkylated and digested with trypsin, using the Janus liquid handling system (PerkinElmer, UK). The digests were subsequently analysed by LC-MS/MS on an LTQ Orbitrap XL mass spectrometer, (ThermoScientific, San Jose, USA). LC-MS/MS data were searched against a protein database (UniProt KB) using the Mascot search engine programme (Matrix Science, UK)[21]. All data were interrogated manually.

Purification of the DBC-1/ZIRD complex

Nuclei from ~10^9 ZIRD-Flag cells were isolated, washed as above, and then sonicated (Bioruptor, 30sec ON/OFF cycles, max intensity for 15mins). Nucleic acids were digested by adding 10,000 units/ml Benzonase (Novagen) and 30 μg/ml RNAse A (Sigma), and incubating at 4°C for 1 hour. The nuclear extract was then brought to 250 mM KOAc and the insoluble fraction removed by centrifugation (20,000g for 30 mins). The supernatant was used for Flag-M2 chromatography. After extensive washes (60 column volumes 20 mM Hepes-KOH pH 7.9, 0.5% Triton, 10% glycerol, 250 mM KOAc), bound proteins were eluted in the same buffer but containing 0.5 mg/ml 3xFLAG peptide (and 100 mM KOAc). The complex was then dialyzed into buffer A (20 mM Tris-HCl pH 7.9, 10% Glycerol, 0.01% NP-40) containing 100 mM NaCl, prior to MonoQ PC 1.6/5 (GE Healthcare) or size exclusion (Superose 6 PC 3.2/30; GE Healthcare) chromatography. Proteins were eluted from MonoQ by a salt-gradient from 0.1 to 1 M NaCl in buffer A. For size exclusion chromatography, samples were loaded in buffer B (20 mM Hepes-KOH, 0.01% NP40, 10% glycerol, 250 mM KOAc) with or without pre-incubation on ice with a 5- to 10-fold molar excess of purified mammalian RNAPII, purified as described[22]. 50 μl fractions were collected, and sizes were estimated by running protein size markers (Biorad) in parallel.

Microarray Analysis

RNAs were extracted using RNeasy Kit (Qiagen), DNAse I-treated on the column, labelled and hybridized on Human Exon 1.0 ST arrays (Affymetrix) using standard techniques. Three independent experiments were performed and used as real triplicate for data analysis. We processed core probe level signals using RMA implemented in APT (apt-1.10.0, Affymetrix) to generate quantile normalised probe set and gene level signal estimates. Probe set to transcript cluster meta-grouping was obtained from Affymetrix. We removed control probe sets from further analysis. For the gene level analysis, genes displaying a coefficient of variance <0.05 were assumed uninformative and removed. We determined transcriptional effects (dbc1 vs ctr, ZIRD vs ctr and dbc1 vs ZIRD) by linear model, moderating the t-statistics by empirical Bayes shrinkage. We selected differential genes using a 0.05 p-value threshold via a nestedF method. The analysis was carried out using the limma package from Bioconductor[23].

To identify putative alternative splicing events we first filtered probe sets to reduce false positive events. We removed all probe sets not localising to unique loci in the genome (Affymetrix annotation). We removed all probe sets from transcripts identified as not expressed in a given condition since it is not possible to determine alternative splicing events against an untranscribed background. We defined a transcript as not being expressed if less than half its member probe sets had a detection p-value (dabg value) <0.05 across all replicates. We also removed probe sets displaying a dabg value >0.05 in all replicates in both the conditions being considered. Genes with less than 3 probe sets after filtering were also removed from the analysis. To identify pairwise alternative splicing events we fitted transcript cluster specific linear models to probe set signal estimates and tested for significant interaction between each probe set and gene level signal estimates across pairwise conditions using a 0.01 p-value threshold and a nestedF method. In cases where multiple probe sets mapped to a single exon, all probe sets had to be significant to be included in the results. Once again the analysis was carried out using the limma package from Bioconductor[23].

Computational analysis of sequences around affected exons

In order to investigate sequence features that might explain the differential enrichment of certain exons upon knock-down of ZIRD or DBC1, we selected a set of 505 exons that were 1.5-fold enriched in both the ZIRD- and the DBC1 knock-downs and which were neither the first nor the last exon (to avoid including 5′ and 3′ UTRs in the following analyses). As a negative control set, we further selected 3877 exons that were unaffected upon depletion of ZIRD and which were also neither the first nor the last exon in the gene. We then prepared a positive and a negative sequence set for both the 5′ and 3′ splice sites. For the 5′ splice site, the positive set contained the regions from −200nt to +50nt around the 5′ splice sites of the 505 enriched exons. The negative control set contains corresponding regions for the 3877 unaffected exons. Analogously, the positive set for the 3′ splice site contained the region from −50nt to +200nt around the 3′ splice sites of the 505 enriched exons, and the negative set contained the corresponding regions of the 3877 unaffected exons.

The P-values in Suppl. Fig. S9 were obtained in the following way. Suppose *ab* to be a dinucleotide and *i* a position around a splice site. We calculated the frequency $P_{exp}(i,ab)$ of *ab* that we would expect, given the frequency of *ab* in the negative set, $P_{neg}(i,ab)$, and given the frequencies in the positive (*pos*) and negative (*neg*) sets of mononucleotide *a* at position *i*, and *b* at position *i+1*, respectively:

$$P_{\exp}(i,ab) = P_{neg}(i,ab)\,\frac{P_{pos}(i,a)\,P_{pos}(i+1,b)}{P_{neg}(i,a)\,P_{neg}(i+1,b)}.$$

The P-values for the observed number of dinucleotides *ab* at position *i* were calculated by approximating the binomial distribution with a normal distribution:

$$P-value = \sum_{k=K}^{N} \binom{N}{k} p^k (1-p)^{N-k} = \frac{1}{2} erfc\left(\frac{K-Np}{\sqrt{Np(1-p)}}\right),$$

where *N* is the number of sequences in the set of affected exons, *K* is the number of observed dinucleotes *ab* at position *i*, *Np* is the expected number of dinucleotides at *i*,

$$p = P_{\exp}(i,ab)\,.$$

of dinucleotides at *i*, $p = P_{exp}(i, ab)$.

Quantitative RT-PCR

RNAs were extracted using RNeasy Kit (Qiagen), DNAse I-treated on the column, and 1 μg of total RNA was retro-transcribed using random primers and first strand cDNA synthesis kit (Fermentas). Quantitative RT-PCR was performed using SYBr Green detection. Specific primers against the alternatively spliced exon and unaffected exons were designed for assessment of exon abundance and transcript expression. Sequences are available upon request.

Chromatin Immunoprecipitation

ChIP assays were performed as described[19], using 4H8, antibody (Abcam), or IgG antibody as negative control, and then 1 hr with protein G/ Herring sperm DNA. The precipitated DNA fragments were analyzed by real-time PCR with SYBr Green detection. Input DNA was analyzed simultaneously and used for normalization.

RNA Immunoprecipitation

RNA immunoprecipitation were performed essentially as described[11,24]. Briefly, 10^7 HEK293 cells were crosslinked in 1% PFA for 10 mins, quenched with 0.125 M glycine for 5 mins, washed twice in PBS and finally harvested in lysis buffer (50 mM Tris-HCl pH8, 1% SDS, 10 mM EDTA, 50 U/500 μl RNAsin and protease inhibitors). Samples were sonicated 15 mins at maximum power using Bioruptor (Diagenode) to obtain RNA fragments 200-600 bases long, cleared by centrifugation, and diluted 10-fold in dilution buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA, 1% Triton, 50 U/ml RNAsin and protease inhibitors). Samples were pre-cleared by incubation with 100 μl of IgG agarose beads (Sigma) for 2 hrs at 4°C. Anti-Flag M2 Resin (Sigma) was added and samples incubated for 4 hrs at 4°C. IPs were washed 3 times in wash-buffer 150 (20 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA, 1% Triton, 0.1% SDS, RNAsin (50 U/ml), protease inhibitors), once in wash buffer 500 (same as above, but 500 mM NaCl), once in LiCl (10 mM Tris-HCl pH 8, 250 mM LiCl, 0.5% NP-40, 0.1% Deoxycholate, 1 mM EDTA, 50 U/ml RNAsin, protease inhibitors) and once in TE100 buffer (TE containing 100 mM NaCl, 50 U/ml RNAsin, protease inhibitors). Immunocomplexes were 3 times eluted in 150 μl elution buffer (TE containing 100 mM NaCl, 200 μg/ml 3xFlag-peptide, 50 U/ml RNAsin) and incubated for 30 mins at 37°C with 1 μl of proteinase K. Crosslinking was then reversed by adding 9 μl of 5 M NaCl and incubating the samples at 65°C for 1 hr. Nucleic acids were phenol/chloroform extracted, ethanol-precipitated and any remaining DNA was eliminated with Turbo DNase (Ambion) treatment. RNAs were then reverse transcribed using random primers and the cDNAs were used for subsequent PCR reaction using relevant primers.

Nuclear run-on analysis

Cells were rinsed in PBS, then in buffer A (20mM Tris-HCl pH7.4, 5mM $MgCl_2$, 0.5mM EGTA, 25% Glycerol, 1mM PMSF), and permeabilised in buffer A containing 0.02% Triton X-100 for 3 mins at room temperature. The nascent RNA labelling reaction was carried out in buffer A containing 2 mM ATP, 0.5 mM GTP, 0.5 mM CTP, 0.2 mM BrUTP (Sigma) and 25 U/ml RNAsin (Promega) for 15 mins at 37°C. In control reactions, normal UTP was used instead of BrUTP. After BrU incorporation, cells were rinsed twice in PBS and total RNA from nuclei of both labelled and control samples was isolated using TriPure Isolation Reagent (Roche). 2 μg of BrdU antibody (Sigma; also recognizes BrU) was pre-incubated with 20μl of Protein G magnetic beads (Invitrogen) per experimental condition. RNA was then heated at 80°C for 10 mins and incubated with the beads at room temperature for 1 hr with gentle shaking. The beads were washed 5 times in PBS/0.1% PVP containing RNAsin

(20U/200μl), the RNA bound to the beads was eluted and the contaminant DNA was eliminated with Turbo DNAse (Ambion). RNA was then reverse transcribed using random primers and the cDNAs were used for subsequent PCR reaction using relevant primers.

References

1. Kornblihtt AR. Coupling transcription and alternative splicing. Adv Exp Med Biol. 2007; 623:175–189. [PubMed: 18380347]
2. Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. Epigenetics in Alternative Pre-mRNA Splicing. Cell. 2011; 144:16–26. [PubMed: 21215366]
3. Perales R, Bentley D. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol Cell. 2009; 36:178–191. [PubMed: 19854129]
4. Dreyfuss G, Kim VN, Kataoka N. Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol. 2002; 3:195–205. [PubMed: 11994740]
5. Chen YI, et al. Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. Nucleic Acids Res. 2007; 35:3928–3944. [PubMed: 17537823]
6. Kiledjian, M.; Burd, CG.; Görlach, M.; Portman, DS.; Dreyfuss, G. Frontiers in Molecular Biology. Mattaj, I.; Nagai, K., editors. Oxford University Press; 1994. p. 127-149.
7. Pinol-Roma S, Dreyfuss G. Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. Nature. 1992; 355:730–732. [PubMed: 1371331]
8. Kim JE, Chen J, Lou Z. DBC1 is a negative regulator of SIRT1. Nature. 2008; 451:583–586. [PubMed: 18235501]
9. Zhao W, et al. Negative regulation of the deacetylase SIRT1 by DBC1. Nature. 2008; 451:587–590. [PubMed: 18235502]
10. Aygun O, Svejstrup J, Liu Y. A RECQ5-RNA polymerase II association identified by targeted proteomic analysis of human chromatin. Proc Natl Acad Sci U S A. 2008; 105:8580–8584. [PubMed: 18562274]
11. Selth LA, Close P, Svejstrup JQ. Studying RNA-protein interactions in vivo by RNA immunoprecipitation. Methods Mol Biol. 2011; 791:253–264. [PubMed: 21913085]
12. Sigurdsson S, Dirac-Svejstrup AB, Svejstrup JQ. Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. Mol Cell. 2010; 38:202–210. [PubMed: 20417599]
13. Saeki H, Svejstrup JQ. Stability, Flexibility, and Dynamic Interactions of Colliding RNA Polymerase II Elongation Complexes. Mol Cell. 2009; 35:191–205. [PubMed: 19647516]
14. Dye MJ, Gromak N, Proudfoot NJ. Exon tethering in transcription by RNA polymerase II. Mol Cell. 2006; 21:849–859. [PubMed: 16543153]
15. Fong N, Ohman M, Bentley DL. Fast ribozyme cleavage releases transcripts from RNA polymerase II and aborts co-transcriptional pre-mRNA processing. Nat Struct Mol Biol. 2009; 16:916–922. [PubMed: 19701200]
16. Kim JE, Chen J, Lou Z. p30 DBC is a potential regulator of tumorigenesis. Cell Cycle. 2009; 8:2932–2935. [PubMed: 19657230]
17. Yoshida K, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011; 478:64–69. [PubMed: 21909114]
18. Lee JY, et al. Characterization of a zinc finger protein ZAN75: nuclear localization signal, transcriptional activator activity, and expression during neuronal differentiation of P19 cells. DNA Cell Biol. 2000; 19:227–234. [PubMed: 10798446]
19. Close P, et al. Transcription impairment and cell migration defects in elongator-depleted cells: implication for familial dysautonomia. Mol Cell. 2006; 22:521–531. [PubMed: 16713582]
20. Ben-David Y, Bani MR, Chabot B, De Koven A, Bernstein A. Retroviral insertions downstream of the heterogeneous nuclear ribonucleoprotein A1 gene in erythroleukemia cells: evidence that A1 is not essential for cell growth. Mol Cell Biol. 1992; 12:4449–4455. [PubMed: 1406633]

21. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999; 20:3551–3567. [PubMed: 10612281]

22. Hu X, et al. A Mediator-responsive form of metazoan RNA polymerase II. Proc Natl Acad Sci U S A. 2006; 103:9506–9511. [PubMed: 16769904]

23. Gentleman RC, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004; 5:R80. [PubMed: 15461798]

24. Gilbert C, Svejstrup JQ. RNA immunoprecipitation for determining RNA-protein associations in vivo. Curr Protoc Mol Biol. 2006 Chapter 27, Unit 27 24.

Europe PMC Funders Author Manuscripts

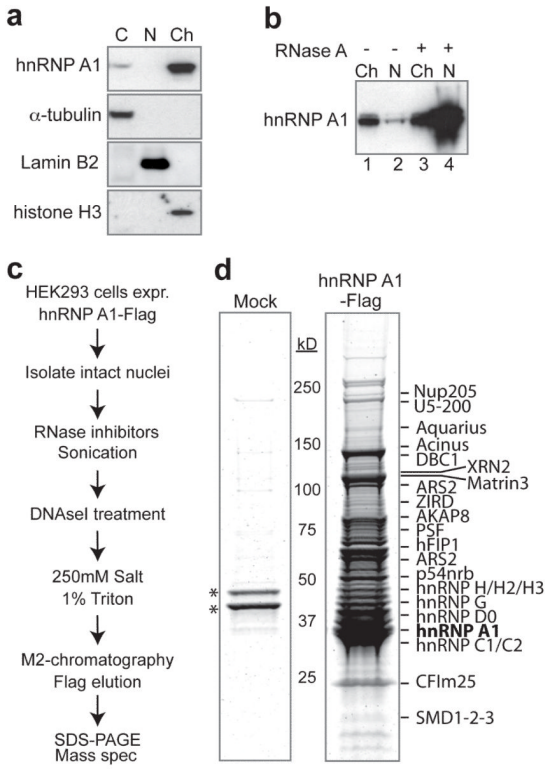

Fig. 1. Purification of nascent nuclear mRNP particles

(**a**) Western of cytoplasmn (C), nucleoplasmn (N), and chromatin (Ch), with α-tubulin, lamin B2, and histone H3 as controls for different fractions. (**b**) Fractionation as in (a), but RNAse A added to the nuclear lysis buffer where indicated. (**c**) Purification procedure outline. (**d**) Equal amount of the M2 chromatography eluates from control (Mock) and A1-Flag separated by 4-12% SDS-PAGE, stained with Sypro ruby. Arrow indicates A1-Flag, and asterisks mark RNAse inhibitor proteins. Some identified proteins indicated on the right.

Europe PMC Funders Author Manuscripts

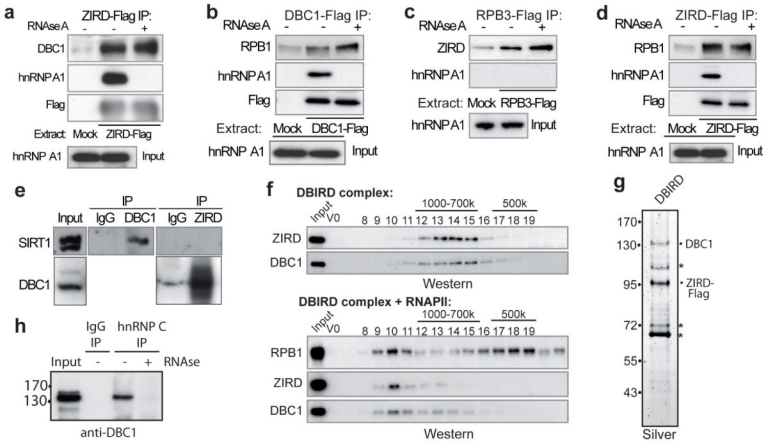

Fig. 2. DBC1 and ZIRD form a stable complex that binds RNAPII

(**a**) Western of anti-FLAG IPs from ZIRD-Flag cells. (**b**) As (a), but DBC1-Flag cells. (**c**) As (a), but RPB3-Flag. (**d**) Same as (a). (**e**) Western analysis of anti-DBC1 and -ZIRD immunoprecipitates. (**f**) DBIRD analyzed by size exclusion chromatography without (upper 2 panels), or with (lower 3 panels), RNAPII in ~5-fold molar excess. Vo, void volume fraction. (**g**) Silver stain of DBIRD. Asterisks indicate DBC1 and ZIRD degradation products. (**h**) Western of hnRNP C-containing mRNP particles from mouse CB3 cells lacking hnRNP A1[20].

Europe PMC Funders Author Manuscripts

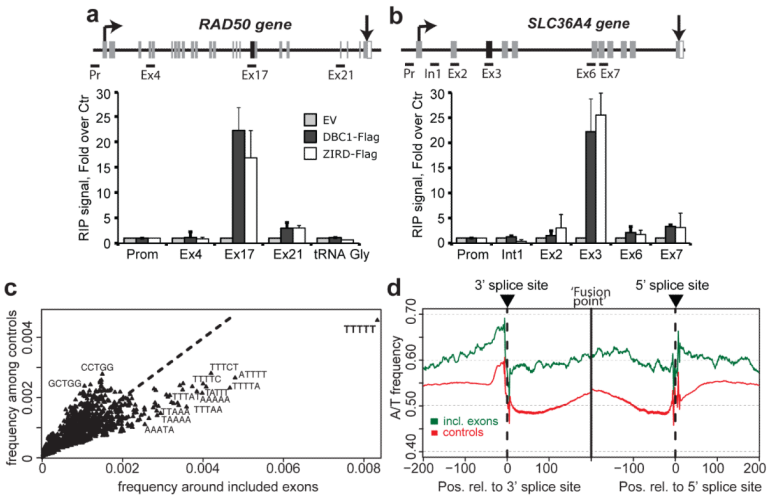

Fig. 3. DBIRD affects alternative splicing and is present at the affected exons

(**a, b**) RNA immuno-precipitated from crosslinked control, DBC1-Flag, or ZIRD-Flag cells, analyzed by qPCR. Control reactions lacking reverse transcriptase were always included (not shown). Error bars indicate standard deviations according to the Poisson statistic; n = 3. (**c**) Frequency of 5-mers in region around splice-sites, of affected (x-axis) versus unaffected exons (y-axis). Diagonal line marks equal frequencies in the positive and negative set. (**d**) Frequency of A or T around splice-sites of included exons (green) and unaffected control exons (red).

Europe PMC Funders Author Manuscripts

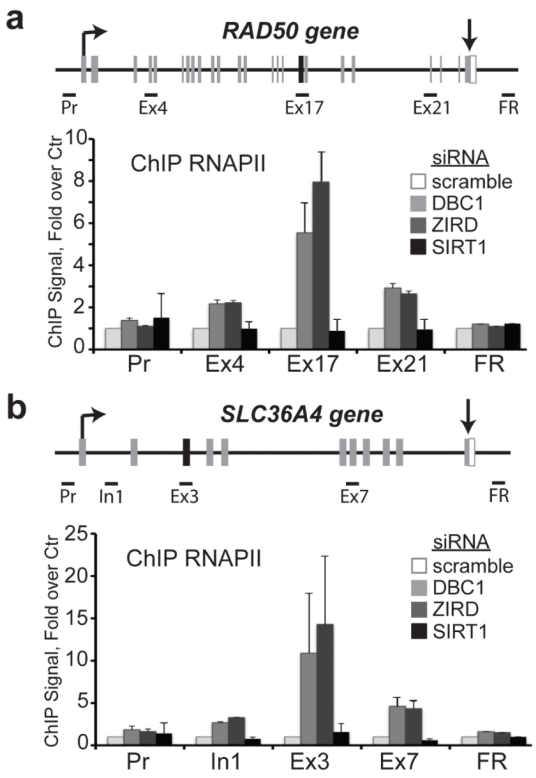

Fig. 4. DBC1 and ZIRD link exon skipping to RNAPII transcription

(**a**)(**upper**) RAD50 gene and qPCR primers (**lower**) RNAPII ChIP using cells transfected with control (scramble), DBC1-, ZIRD-, or SIRT1- stealth siRNAs. ChIP signals were normalized with inputs. Signals in control cells (scramble) were set to 1 at each position, and values obtained from factor-depleted cells expressed relative to that. Errors bars denote standard deviation; n = 3. **(b)** Same as (a), but at SLC36A4. Suppl. Fig. S12 shows the same data in a format where gene positional information is maintained.

Europe PMC Funders Author Manuscripts