A multiprotein complex necessary for both transcription and DNA replication at the β -globin locus

Subhradip Karmakar¹, Milind C Mahajan^{1,2,*}, Vincent Schulz³, Gokul Boyapaty¹ and Sherman M Weissman^{1,*}

¹Department of Genetics, The Anlyan Center, Yale University School of Medicine, New Haven, CT, USA, ²Department of Genetics, Yale Center for Genome Analysis, Yale University, Orange, CT, USA and ³Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA

DNA replication, repair, transcription and chromatin structure are intricately associated nuclear processes, but the molecular links between these events are often obscure. In this study, we have surveyed the protein complexes that bind at β -globin locus control region, and purified and characterized the function of one such multiprotein complex from human erythroleukemic K562 cells. We further validated the existence of this complex in human CD34⁺ cell-derived normal erythroid cells. This complex contains ILF2/ILF3 transcription factors, p300 acetyltransferase and proteins associated with DNA replication, transcription and repair. RNAi knockdown of ILF2, a DNA-binding component of this complex, abrogates the recruitment of the complex to its cognate DNA sequence and inhibits transcription, histone acetylation and usage of the origin of DNA replication at the β -globin locus. These results imply a direct link between mammalian DNA replication, transcription and histone acetylation mediated by a single multiprotein complex.

The EMBO Journal (2010) 29, 3260–3271. doi:10.1038/emboj.2010.204; Published online 31 August 2010

Subject Categories: chromatin & transcription; genome stability & dynamics

Keywords: multiprotein complex; purification; replication; transcription

Introduction

Interaction between nuclear processes such as modifications of chromatin structure, DNA replication, repair and transcription have been described in several eukaryotic organisms (Gottipati and Helleday, 2009; Rampakakis *et al.*, 2009). Maintenance of genome integrity during transcription and DNA replication is accomplished by repair mechanisms such as transcription-coupled repair (Hanawalt and Spivak, 2008),

replication-associated homologous recombination and translesion synthesis (Bridges, 2005; Lehmann, 2005; Hanawalt and Spivak, 2008).

The formation of pre-replication complexes at initiation of replication (IR) sites is initiated by the origin recognizing complex (ORC)-mediated recruitment of cdc6, cdt1 and the MCM2–7 complex during G1 phase of the cell cycle and the origins are licensed to initiate a single round of DNA synthesis per cell cycle. During G1/S transition, CDC7 kinase and cyclins E/A recruit additional components, including CDC45, GINS and replicative DNA polymerase to form the pre-initiation complexes (Rampakakis *et al.*, 2009).

Sequence-specific loading of replication proteins and firing of replication origins occurs in bacteria and yeast as well as animal viruses (Kohzaki and Murakami, 2005). In contrast, ORC from metazoan cells shows very little sequence specificity other than some preference for AT-rich sequences, so that mechanisms must exist to account for the non-random distribution of origins of replication. In several animal virus infections, transcription factors have been shown to recruit the host-cell ORC to specific sites (Guo *et al.*, 1996; Ito *et al.*, 1996; Murakami *et al.*, 2007). In some metazoans, transcription is known to influence DNA replication (Danis *et al.*, 2004; Xie and Orr-Weaver, 2008). In mammalian cells, actively transcribing genes are often replicated early in S phase (Dimitrova, 2006; Falkenberg *et al.*, 2007). Many mammalian IR sites are AT rich (Gilbert, 2001) or contain a region of AT-rich sequence, but the mechanisms leading to initiation of DNA replication from specific loci are not fully understood.

The β -globin locus contains five globin genes: an embryonic gene (ϵ), two foetal globin genes (γ G and γ A) and two adult-globin genes (δ and β), and a pseudogene ($\psi\beta$), with the order of the genes corresponding to their time of expression during development. The expression of these genes is strongly dependent on a locus control region (LCR) 50 kb upstream of the β -globin gene. The LCR, in turn, contains four erythroid-specific DNase hypersensitive sites (HS1–4) that include evolutionarily conserved sequences (Figure 1A). The fifth DNase HS5 occurs in erythroid as well as several other non-erythroid haematopoietic cell systems. The β -globin locus harbours a strong IR site for DNA replication between the δ - and β -globin genes that has been used as a model system for studying mammalian DNA replication (Figure 1A) (Aladjem, 2004). Two or more independent modules exist within this β -globin IR (Wang *et al.*, 2004). In addition, initiation of DNA replication has been described at the 3' enhancer of the β -globin locus and at γ -globin genes (Kamath and Leffak, 2001; Aladjem *et al.*, 2002; Buzina *et al.*, 2005). In the chicken globin locus, the HS4 sequence of the LCR is also reported to initiate DNA replication (Prioleau *et al.*, 2003).

The globin cluster is embedded in a complex of olfactory genes that are non-functional in non-erythroid lineages.

*Corresponding authors. MC Mahajan, Yale Center for Genome Analysis, 147 Frontage Road, Yale University, Orange, CT 06477, USA. Tel.: +1 203 737 3050; Fax: +1 203 737 3104; E-mail: milind.mahajan@yale.edu or SM Weissman, Department of Genetics, The Anlyan Center, Yale University School of Medicine, New Haven, CT 06511, USA. Tel.: +1 203 737 2282; Fax: +1 203 737 2286; E-mail: sherman.weissman@yale.edu

Received: 23 February 2010; accepted: 29 July 2010; published online: 31 August 2010

Figure 1 Tiling EMSA showing the presence of various DNA–protein complexes recruited to the HS4 region of the β -globin LCR in K562 cells. (A) Architecture of human β -globin cluster showing the DNase-I hypersensitive sites HS1–5 in the LCR, five β -like-globin genes namely an embryonic ϵ gene (ϵ), two foetal-globin genes ($A\gamma$, $G\gamma$), one pseudogene ($\psi\beta$) and two adult-globin genes (β , δ) with the site of origin of DNA replication. (B) Tiling electrophoretic mobility shift assay (EMSA) screen for the detection of K562-specific DNA–protein interactions. The oligonucleotides numbered HS4–1 to HS4–10 at the top of the gel are the overlapping 5' to 3' double-stranded oligonucleotide tiles of the 350 bp core HS4 region of the β -globin LCR (see Supplementary data). Ten double-stranded 35-mer (average), ^{32}P -labelled oligonucleotide probes covering 350 bp region of the core HS4 were incubated with HeLa or K562 nuclear extracts and run on a 5% native polyacrylamide gel to display gel-retarded bands arising because of binding of K562 or HeLa nuclear proteins to each of the 10 double-stranded oligonucleotides. The EMSA bands were visualized by exposing the radioactive polyacrylamide gel to the photo-film (Kodak). Several common EMSA bands are seen with K562 and HeLa nuclear extracts. The K562-specific EMSA band obtained with HS4–9 oligonucleotide is indicated by an arrow. (C) Size determination of the HS4–9-binding protein complexes by gel exclusion chromatography. Before the size fractionation, the K562 crude nuclear extract was partially purified on a Heparin agarose column. The HS4–9 DNA-binding activity was recovered in the 0.4 M salt elute from the heparin-Sepharose column and was size fractionated by passing through a Superose-6 gel exclusion column. Seventy five 0.3 ml fractions were collected. The void volume of the Superose-6 column is indicated by an arrow. Fractions were assayed for EMSA to detect HS4–9-binding activity. This activity eluted after the void volume and peaked before the 2 MDa standard molecular marker elution peak. The elution peaks for other standard protein molecular weight markers are indicated at the top of the gel. (D) Competitive EMSA assay showing the sequence-specific binding of K562 nuclear proteins to double-stranded HS4–9 oligonucleotide sequences (sequences given in Supplementary data). In a typical competitive EMSA assay, the DNA–protein-binding reaction mixture contained K562 nuclear extract and ^{32}P -labelled HS4–9 oligonucleotide along with 10–100-fold molar excesses of non-radioactive (i) HS4–9, (ii) ϵ -12, (iii) HIF and (iv) scrambled HS4–9 oligonucleotides. The binding mix was run on a native 5% polyacrylamide gel and exposed to photo-film to visualize EMSA bands. The increasing order of 10–100-fold molar excess of non-radioactive oligonucleotides used in the EMSA lanes are indicated at the top of the figure by the expanding triangle height. The HS4–9-associated EMSA activity was efficiently competed only by non-radioactive molar excess of the same HS4–9 oligonucleotide.

Selective activation of the transcription of the globin cluster in erythroid cells involves participation of *trans*-factors that are thought to function not only at the globin promoters, but also at the LCR. The LCR was shown to be a binding region for several large protein complexes (Mahajan *et al*, 2007) and orchestrates molecular events to form an active chromatin hub (Zhou *et al*, 2006; Kooren *et al*, 2007) and transcriptional factory (Mitchell and Fraser, 2008).

The presence of intergenic deletions adjacent to globin genes of β -thalassaemia patients provides genetic evidence for

the presence of the globin IR. The Lepore deletion that removes sequences between the β - and δ -globin genes that contain an IR region results in passive replication of the locus from an outside origin (Kitsberg *et al*, 1993). In Hispanic thalassaemia in which the LCR and upstream sequences are deleted, the β -globin locus replicates late in S phase with replication originating outside of the globin locus (Forrester *et al*, 1990; Mechali, 2001). Interestingly, in Lepore patients, the direction of replication is from 5' to 3' relative to globin gene transcription, whereas it is in the opposite orientation in

Hispanic thalassemia (Kitsberg *et al*, 1993; Mechali, 2001). The different directions of DNA replication in these two types of deletions suggest that in addition to the IR sequences, upstream sequences can act as origins of DNA replication for the β -globin locus. However, in some experimental systems, deletion of the LCR neither changed the IR site nor the timing of replication (Cimbora *et al*, 2000) and the choice of IRs may be changed indirectly, for example by delaying replication firing at a potential IR site, causing replication initiated from an external region to extend across the IR before firing occurs.

In our survey of sequence-specific protein complexes in erythroid cells, we noted that antibodies against an MCM protein disrupted a sequence-specific complex that bound to an oligonucleotide from the HS4 region of the β -globin LCR (data not shown). Deletion analyses have shown that HS4 functions in maintaining normal levels of globin mRNA production (Simon *et al*, 2001; Fedosyuk and Peterson, 2007) and that deletion of this site may affect the timing of replication at the β -globin cluster (Simon *et al*, 2001), but the specifics of how HS4 functions and the function of the hypersensitive sites have not been worked out. We hypothesized that proteins recruited to the LCR might functionally co-operate with IRs in the initiation of transcription and replication. Therefore, we performed a systematic analysis for novel protein complexes from erythroid cells that bind to the globin locus HS4 region. For this purpose, we performed tiling electrophoretic mobility shift assay (EMSA) scans to isolate erythroid K562 cell line-specific DNA–protein complexes. We used a series of overlapping oligonucleotides (HS4–1 through HS4–10) tiling the core 350 bp HS4 region of the β -globin LCR and carried out EMSA assays with nuclear extracts from HeLa and K562 cells (Figure 1B). This tiling EMSA scan led to the identification of a K562-specific multiprotein complex binding to an AT-rich HS4–9 oligonucleotide. We have purified a multiprotein complex that is sequence specifically associated with this HS4–9 sequence of the HS4 and identified its protein components by mass spectrometry. This analysis uncovered a complex of at least 16 components. Follow-up experiments suggest that this DNA-associated replication and transcription complex (DARRT) affects replication, chromatin modification and transcription.

Results

Identification of a high-molecular weight protein complex associated with an AT-rich sequence in the HS4 region of the β -globin LCR

In order to identify the multiprotein complexes associated with the β -globin LCR HS1–4, we performed a tiling EMSA scan using a series of 40 overlapping 35-mer (average) double-stranded synthetic oligonucleotides, tiling four LCR DNase HS1–4 sites with 7 bp overlaps. These were used in EMSA with nuclear extracts prepared from HeLa and the human erythroleukemic cell line K562 to identify K562-specific DNA–protein interactions at the core HS1–4 sequences. In this study, we have focused on the identification and characterization of the K562-specific DNA–protein interactions formed on the core HS4 sequence (Figure 1B). In this tiling EMSA screen of the 350 bp HS4 sequence, we have used 10 oligonucleotides HS4–1 through HS4–10 that span the entire core HS4 region of the LCR. This EMSA screen of HS4 revealed three prominent K562-specific EMSA bands

with HS4–5, HS4–7 and HS4–9 oligonucleotides. The native molecular weight of the protein(s) associated with these EMSA bands were estimated by Superose-6 column-based gel exclusion chromatography. When we passed the K562 nuclear extracts through the Superose-6 size exclusion column, the protein complex associated with HS4–9 oligonucleotide was eluted with an apparent molecular weight of >2 MDa (Figure 1C). To further test whether the large protein complex associated with this HS4–9 sequence is indeed sequence specific, we carried out competitive EMSA assays. In these assays, we tested for the competitive displacement of the radioactive 32 P-labelled HS4–9-associated EMSA bands with molar excesses of non-radioactive (cold) HS4–9 oligonucleotide as well as several control oligonucleotides of the same base composition (Figure 1D). We found efficient competition with the non-radioactive HS4–9 oligonucleotide alone and not with other control oligonucleotides, indicating sequence specificity of the protein complex associated with HS4–9 oligonucleotide sequence fragment (Figure 1D).

Purification of the protein complex associated with the HS4–9 EMSA

Having established that the HS4–9 sequence-associated DNA-binding activity is sequence specific and involves a large protein complex (Figure 1), we undertook the biochemical purification of this protein complex from K562 nuclear extracts. Our purification strategy involved the enrichment of DNA-binding proteins from the K562 nuclear extracts using heparin-Sepharose chromatography followed by molecular sieving on a Superose-6 sizing column for isolation of large protein complexes. These in turn were fractionated on a DEAE-Sepharose column. The HS4–9-binding proteins eluted from the DEAE-Sepharose ion-exchange column were purified using HS4–9 oligo-affinity chromatography (Figure 2A). Extensive purification was achieved after the final HS4–9 DNA oligonucleotide-affinity column step. Active EMSA fractions eluted from the DNA-affinity column were analysed by SDS–PAGE. Intensities of various bands as seen after staining the SDS gel suggested that the proteins were present in near stoichiometric ratios (Figure 2B).

MS/MS analysis of the trypsin-digested SDS–PAGE fragments revealed that the complex contained DARRT components (Figure 2B and C). Western blot analysis of the chromatographic fractions showed co-elution of the DARRT components during the purification steps (Supplementary Figure 1A–D). Among the several protein identified, we validated the presence of ILF2, ILF3, MCM5, p300 and RAD50 in the DARRT complex by co-immunoprecipitation and immunodepletion experiments (Figure 4). Tip49a and Tip49b are involved in transcription and DNA damage response, but were not studied in further detail, as we could not obtain suitable antibodies for co-immunoprecipitation and ChIP experiments. However, the presence of Tip49 as well as other components was confirmed by an alternative procedure using immunopurification of DARRT by p300 immunoaffinity chromatography (Supplementary Figure 6). In addition, we established stable K562 cell lines expressing C-terminal 6XHis-3XFLAG-tagged ILF2. We used the nuclear extracts from these cell lines for the purification of the HS4–9-associated protein complex. We obtained the same set of proteins in this complex, when we used a modified purification procedure in which we used successively heparin agarose

Figure 2 Purification of HS4-9 oligonucleotide-binding DARRT complex. (A) Schematic flow diagram showing the strategy for the biochemical purification of the DARRT complex followed by the identification of its protein components by MS/MS analysis. (B) SDS-PAGE display of the proteins eluted from the final HS4-9 oligonucleotide-affinity column. Each band from the gel was excised, and identified by MS/MS. Mass spectrometry identification of proteins was repeated with two different batches of K562 cells with similar results. (C) Functional categorization of the protein components of DARRT identified by MS/MS analysis. Categorizations were made according to the information available in the literature describing the function of these proteins with various DNA/RNA-associated functions. (D) Gel supershift assay using antibodies against some of the HS4-9-binding proteins identified by MS/MS analysis. Note that only the antibodies against DARRT proteins disturb the EMSA bands. Antibodies against other erythroid transcription factors expressed in K562 cells, as well as the normal IgG control, do not affect the EMSA bands, suggesting the specificity of the supershifts. Although antibodies against p300 and RAD50 neutralize the EMSA bands, antibodies against MCM5 and ILF2 showed supershifts.

chromatography, superose FPLC, immunoaffinity purification with an anti-FLAG antibody, TALON column purification and finally the specific HS4-9 oligonucleotide-affinity column (Supplementary Table 1). Sucrose gradient fractionation of nuclear extracts also showed that the components of DARRT migrated as a single high-molecular weight peak (Supplementary Figure 9).

To establish that the DARRT complex is responsible for the EMSA bands with the HS4-9 oligonucleotide, we used gel supershift assays with antibodies against DARRT components (Figure 2D). Antibodies against p300 and RAD50 neutralized EMSA bands, whereas MCM5 and ILF2 antibodies resulted in band supershifts (Figure 2D). These data show *in vitro* association of DARRT with the HS4-9 oligonucleotide. ChIP-qPCR analysis using antibodies against ILF2, MCM5, RAD50 and p300 confirmed the *in vivo* association of the DARRT complex with the HS4 region (Figure 3; Supplementary Figure 2B). In addition to HS4, our ChIP qPCR results showed small but significant enrichment of ILF2, RAD50 and MCM5

at the γ -globin promoter and at the β -globin replication origin β -Rep-1 in K562 cells (Figure 3A). The p300 ChIP qPCR indicated strong binding to the HS4 region, and weaker but significant binding to other LCR and promoter sites, suggesting that it bound to multiple sites at the β globin (Figure 3A). To test whether DARRT is also associated with normal erythroid cells, we carried out ChIP analysis using antibodies against ILF2, ILF3, MCM5 and RAD50 in human CD34+ derived primary erythroid cells (Figure 3B). All antibodies showed a significant enrichment at the HS4 site. In addition, in the normal erythroid cells, ILF2, ILF3 and MCM5 showed substantial enrichment at the β -Rep-1 region of the globin IR. This suggests redundancy in the binding sites for DARRT in the normal β -globin complex and this redundancy may account for some of the variation in effects of local deletions on globin replication (Forrester *et al*, 1990; Kitsberg *et al*, 1993; Cimbora *et al*, 2000; Mechali, 2001).

Co-immunoprecipitation (co-IP) and immunodepletion experiments were carried out to test whether DARRT exists

Figure 3 ChIP qPCR of the DARRT protein components. ChIP-qPCR assays showing the *in vivo* recruitment of DARRT components at the β -globin locus in (A) K562 and (B) human primary erythroid cells differentiated from human CD34⁺ cells. Large text letters in each of the eight graphs indicate which antibody was used for that set of chromatin immunoprecipitations. The ChIP, methodology, antibodies and sequences of the primers are presented in the Supplementary data. ChIP-PCR gels showing the relative enrichments are provided as Supplementary Figure 2B. Fold enrichments were calculated relative to an IgG control. Each figure represents the average of three biological replicates and is expressed as mean $\pm$ s.e.m. The components of DARRT are all most strongly enriched over HS4 in K562 cells. In the primary erythroid cells, the components are also enriched strongly over the previously described β -globin origins of replication (β -Rep-1).

as a single homogeneous complex. Immunoprecipitating antibodies against ILF2, ILF3, MCM5 and RAD50 showed that these proteins can mutually co-immunoprecipitate each other from K562 nuclear extract, thereby suggesting that DARRT exists as a complex in the nuclear extract before fractionation (Figure 4A). Furthermore, treatment of protein extracts with DNase and RNaseA before immunoprecipitation did not have any effect on the co-IP results, suggesting that these proteins are not bound together by any RNA or DNA intermediates. Immunodepletion experiments with purified DARRT complex showed that when sufficient antibodies against the MCM5 and RAD50 proteins were added to clear them from the supernatant, there was co-clearance of each other and of ILF3, MCM3, ILF2 and DNA-PK from the IP supernatant, showing that these proteins are entirely present in a complex with MCM5 and RAD50 (Figure 4B). In addition, we show that ILF2 and RAD50 can be co-immunoprecipitated with an antibody against MCM5 in erythroid cells prepared from freshly isolated human lineage-negative CD34⁺ cells (Mahajan *et al*, 2009), indicating that DARRT also exists in primary erythroid cells (Figure 4C).

ILF2 recruits DARRT to its target DNA sequence

NF45/ILF2 associates with NF90/ILF3 as a heterodimer in the nucleus and regulates IL-2 gene transcription by binding to the antigen receptor response element/NF-AT DNA target sequence (Corthesy and Kao, 1994; Kao *et al*, 1994). To test whether these proteins are responsible for the recruitment of DARRT to its target DNA sequence on HS4, we established stable K562 lines expressing tetracycline-inducible ILF2 and MCM5 shRNA. Addition of doxycycline to these cells resulted in significant knockdown of ILF2 and MCM5 proteins (Figure 4D). Nuclear extracts from these ILF2 knockdown cells had significantly reduced DARRT DNA-binding activity to the HS4-9 oligo as revealed by EMSA (Figure 4E). siRNA-mediated double knockdown of ILF2 and ILF3 resulted in

complete loss of DARRT EMSA-binding activity (Figure 4E), indicating that these DARRT components are necessary for binding to the HS4-9 DNA sequence. Furthermore, recombinant ILF2 was found to bind *in vitro* to HS4, supporting an important function for ILF2 in DARRT recruitment to HS4 (Supplementary Figure 2A). To investigate the effect of ILF2 knockdown on the recruitment of other DARRT components to the HS4 region, we performed ChIP qPCR for MCM5 and p300 in K562 cells expressing ILF2 shRNA. The results showed a significant reduction in the ChIP signal at HS4 and other sites on the locus (Figure 4F; Supplementary Figure 2B). Collectively, these results suggest that ILF2 has an important function in recruiting the DARRT components to the HS4 sites.

DARRT components are essential for high levels of globin mRNA expression

DARRT contains transcription factors such as ILF2 and ILF3 (Corthesy and Kao, 1994) and other components such as Tip49a, TIP49b and MCM proteins that have been implicated in the regulation of transcription (Snyder *et al*, 2005, 2009; Jha and Dutta, 2009) and chromatin remodeling (Dziak *et al*, 2003). Therefore, we investigated whether DARRT regulates β -globin gene transcription by analysing the effects of ILF2, RAD50, MCM5 or ORC2 depletion on β -globin mRNA expression in K562 cells. K562 cells actively transcribe embryonic (ϵ)- and foetal (γ)-globin genes, whereas the adult globin (β) remains essentially inactive. MCM5, ORC2 and ILF2 knockdown exhibited significant inhibition of ϵ - and γ -globin expression (Figure 5A). RT-PCR experiments in K562 cells (Supplementary Figure 7) after siRNA-mediated ILF2 or MCM5 knockdown showed no significant changes in the levels of GATA1, GATA2, NFE2, EKLF (KLF1), TAL1, p18-MAF, PU.1, BACH1, LMO2, EPOR, YY1, LMO4 and RUNX1 mRNAs, suggesting that knockdown of ILF2 and MCM5 messages do not significantly alter the erythroid properties of K562 cells. Hence, the effect of ILF2 and

Figure 4 DARRT exists as a single homogeneous multiprotein complex containing the DNA-binding component ILF2. (A) Co-immunoprecipitations (co-IP) of K562 nuclear extracts using antibodies against ILF2, ILF3, MCM5 and RAD50 show that all the tested components of DARRT co-immunoprecipitate from nuclear extracts. Nuclear extracts (150 μ g) were immunoprecipitated with 10 μ g of antibodies or isotype-specific IgG and western blotting was performed. (B) Co-Immunodepletions of some of the DARRT components showing that components of DARRT are all associated in a single complex in purified DARRT preparations. For each immunodepletion, 3 μ g of the purified DARRT complex and 15 μ g of each of the antibody and control normal IgG were used. The supernatant and immunoprecipitated material was analysed by western blot using antibodies against various proteins as shown in the figure. (C) The co-IP in primary human erythroid cells showing that DARRT components co-immunoprecipitate in nuclear extracts from normal erythroid cells. (D) Western blots of doxycycline-induced shRNA-mediated knockdown of ILF2 and MCM5 in K562 cells showing >80% reduction of the specific proteins. RAD50 was knocked down with a constitutive shRNA (pKLO.1). (E) EMSA assays with K562 nuclear extracts from normal and shRNA (ILF2) as well as siRNA-based knockdown of ILF2 and ILF3 showing that depletion of ILF2 alone or in combination with ILF3 abolishes the DARRT EMSA band. Double knockdown of ILF2 and ILF3 were performed using two independent siRNA clones (targeting different regions of the template RNA) as indicated in the figure. NS, non-silencing control shRNA. (F) Chromatin immunoprecipitations with antibodies against MCM5 or p300 in ILF2 knockdown cells. The results show that knockdown of ILF2 results in a substantial reduction of MCM5 and p300 binding to HS4 and other regions of the β -globin locus. Solid bars show the ChIP results in K562 cells treated with a non-silencing shRNA and the open bars show the corresponding results in cells with ILF2 knockdown. The fold enrichment in qPCR signal is expressed relative to the signal obtained with normal IgG. Each experiment was the average of three independent biological replicates and expressed as mean $\pm$ s.e.m.

MCM5 knockdown may directly affect the ϵ - and γ -globin transcription in these cells (Supplementary Figure 7).

Knockdown of MCM5 and ILF2 proteins also resulted in a severe inhibition of K562 cell growth (Supplementary Figure 5B). Hence, we tested whether the decrease in ϵ - and γ -globin expression was a secondary effect of reduced cell proliferation. This was addressed when we slowed the growth of K562 cells by serum starvation (that is culturing in 0.5% FBS) to levels comparable with those seen with ILF2 or MCM5 depletion (Supplementary Figure 5A), we did not observe any significant change in the mRNA levels of the ϵ - and γ -globins (Supplementary Figure 5C). To further rule out the possible cell cycle arrest effect on the down regulation of ϵ - and γ -globin genes in MCM5- and ORC2-knocked down cells, we

treated the K562 cells with three different cell cycle inhibitors, namely roscovitine, nocodazole and hydroxyl urea (Supplementary Figure 11). Treatment of roscovitine and nocodazole did not have any effect on the globin gene transcription. Treatment of hydroxyurea previously reported to increase γ -globin transcription in K562 and primary erythroid cells (Tang *et al*, 2005) significantly increased the γ -globin transcription in our cells (Supplementary Figure 11). These data suggest that reduced ϵ - and γ -globin transcription may not be an effect of cell cycle arrest, but are specific effects of MCM5 and ORC2 knockdown and that ILF2 and MCM5 have an essential and specific function for globin transcription.

In addition, knockdown of ILF2 in MEL cells produced significant reduction in mRNA of adult β -major and β -minor

Figure 5 Effects of ILF2 knockdown on mRNA expression and histone modifications. Reverse transcriptase quantitative PCR of α - and β -like-globin gene expression in normal and ILF2-, MCM5-, RAD50- and ORC2-knocked down K562 cells (A) and ILF2-knocked down MEL cells (B). shRNA expression was induced in the presence of 2 μ g/ml of doxycycline for 9 days. After 6 days of doxycycline induction, the MEL cells were shifted to shRNA induction medium containing 2 μ g/ml of doxycycline and 2% of DMSO. At day 1 (D-1), day 2 (D-2), day 3 (D-3) and day 5 (D-5) of DMSO treatment, total RNA was isolated and subjected to RT-qPCR analysis. Fold change on the y axis is the change in RT-qPCR signal over the normal K562 and uninduced (UD) MEL cells. NS refers to a non-silencing shRNA control. Each RT-qPCR data set was normalized with GAPDH for K562 and $\beta 2M$ ($\beta 2$ microglobulin) for MEL cells. The values in each bar diagram are the averages of two independent biological experiments performed in duplicate. (C) Gross effect of ILF2 knockdown on MEL cells differentiated with DMSO treatment for 5 days. Note that ILF2-knocked down MEL cells failed to turn red upon DMSO-induced differentiation. (D) Significant ILF2 knockdown in MEL using a TRIPZ (Open Biosystems)-inducible shRNAmir. (E–H) Effect of ILF2 knockdown on the Histone H4 acetylation pattern across the β -globin locus. Antibodies used against each of the four acetylated lysine residues of histone H4 are shown at the top of each diagram. Solid bars represent normal K562 cells and open bars represent ILF2-knocked down cells. Data presented in each bar diagram is the average of two independent biological experiments performed in duplicate and expressed as mean $\pm$ s.e.m.

globin as well as α -globin seen by RT-PCR (Figure 5B; Supplementary Figure 3). Friend virus-transformed MEL cells provide an early erythroid precursor model system that can be used to terminally differentiate cells with chemicals such as DMSO in cell culture (Chen *et al*, 2006). Significant knockdown was achieved using a mouse-specific ILF2 shRNAmir (Figure 5D). DMSO-treated MEL cell pellets with ILF2 knockdown were pale in colour, indicating a severe reduction in globin content (Figure 5C). Whether ILF2 knockdown perturbs a broader aspect of erythroid differentiation was queried by following the mRNA expression of several other erythroid-specific genes during DMSO-induced differentiation of MEL. Although erythroid-specific ALAS-2 and spectrin $\beta 1$ mRNAs showed several fold reduction, non-erythroid CA1 and spectrin $\beta 2$ mRNAs showed an increase after ILF2 knockdown, whereas non-erythroid-specific spectrin $\alpha 1$ did not show any significant changes (Supplementary Figure 3A–K). These results suggest a specific effect of ILF2 knockdown on some aspects of erythroid differentiation in the murine system.

DARRT is required for normal histone acetylation of the β -globin locus

The presence of p300 histone acetyltransferase in the DARRT complex suggests that the complex may have a function in regulation of chromatin structure through histone acetylation. We used ChIP qPCR to study the effects of ILF2 knockdown on the acetylation of the four histone H4 residues (H4K5, H4K8, H4K12 and H4K16) for several DARRT-binding regions. Interestingly, the pattern of acetylation in control cells was different for each of these residues (Figure 5E–H). Among promoters, acetylations at K5 and K12 were present in γ - and ϵ -globin genes, respectively. K8 acetylation was present at both the γ and ϵ promoters. Among the HS sites of the LCR, K5, K12 and K16, acetylations were present mainly at the HS3 region, whereas none of the HS sites had acetylation at K8. Among the replication origin sites, we detected significant amounts of K8 acetylation at β -Rep-1 and 1.8 kb downstream of the HS4 sequence. ILF2 knockdown resulted in marked reduction of acetylation at the promoters, and

origins of DNA replication, whereas the acetylation of lysines at most of the LCR regions remained unchanged. ILF2 knockdown also resulted in reduction of acetylation at Histone H3 (K9 and K18) acetylation at the γ - and ϵ -globin promoters with no effect on the other sites including replication origins (Supplementary Figure 4A and B). These observations suggest that DARRT might directly or indirectly have a significant function in the regulation of chromatin structure through histone acetylation.

DARRT has a function in DNA replication

The presence of the MCM complex and ORC2 in DARRT suggests that DARRT may have an important function in DNA replication. This was addressed by investigating the cellular growth rate during shRNA-mediated knockdown of ILF2 and MCM5 in K562 cells (Supplementary Figure 5B). Our observations showed marked slowdown in cell growth without any appreciable cell death. FACS analysis of propidium iodide-stained ILF2 knockdown cells did not show significant changes in the relative number of cells in different stages of the cell cycle (data not shown), agreeing well with the previous reports showing a similar effect in siRNA-induced silencing of ILF2 and ILF3 in human HEK293 cells (Guan *et al*, 2008). These results suggest that the inhibition of cell growth in ILF2-knocked cells could be due to slowing multiple phases of the cell replication process rather than cell cycle arrest at a specific point in the replication cycle.

To further investigate a potential function for DARRT in DNA replication, we performed a replicating nascent-strand abundance assay in wild type and ILF2- and MCM5-knocked down K562 cells (Figure 6). This assay depends on isolation of single-stranded nascent DNA fragments that are protected from λ -exonuclease digestion because they retain an RNA primer at their 5' end. We tested known IR sites in the β -globin locus (Kamath and Leffak, 2001; Aladjem *et al*, 2002; Wang *et al*, 2004; Buzina *et al*, 2005), and several novel lesser IR sites that we have identified at HS1 and between the HS4 and HS3 regions by analysing the abundance of replicating short single-stranded RNA–DNA hybrids (Figure 6). In this assay, we found similar enrichment patterns of nascent replicating DNA molecules with both 300–600 and 600–1200 bp DNA fragments. As expected, we did not observe the enrichment of the nascent replicating strands when we size selected 2–4 kb DNA fragments (Supplementary Figure 10B), and there was no enrichment when the single-stranded DNA (ssDNA) was treated with RNase before λ -exonuclease digestion (Supplementary Figure 10A). MCM5 knockdown resulted in marked decrease of replication initiation at the IR sites in the β -globin locus as well as at the lamin-B locus, which contains a well-studied origin of replication (Figure 6E) (Lucas *et al*, 2007), whereas ILF2 depletion specifically abrogated replication initiation at the IR of the β -globin locus, but not at the lamin-B locus (Figure 6E). As ILF2 knockdown partially silenced the β -globin cluster, we wished to test whether the method

Figure 6 Nascent-strand abundance assay for mapping origins of DNA replication. Graphs show the qPCR quantization of nascent strands originating from replicating foci at the β -globin locus. Short nascent single-stranded DNA was isolated as described in Materials and methods. For each qPCR, 3.5 ng of the DNA were used. The primers used for the qPCR are listed in Supplementary data. Fold enrichment shown in the y axis is the PCR signal intensity over the signal from an equal amount of non-replicating DNA obtained from the flow through of a BND-cellulose column. The shRNA construct harbouring K562 cells were grown in the presence of doxycycline for 9 days, at the end of which they were used for the nascent-strand abundance assay. Solid lines in the graph represent ILF2 (A), MCM5 (B), RAD50 (C) and non-silencing (D) shRNA-expressing K562 cells. Non-silencing (NS) shRNA-expressing cells served as negative control, whereas MCM5-knocked down cells as positive control. Graphs with dotted lines represent normal (control) K562 cells. (E) Effect of ILF2 and MCM5 knockdown cells on the origin of DNA replication at lamin-B2 gene. The values in each point in the graphs represent the average of two independent biological experiments performed in duplicates. The results show that knockdown of ILF2 abolishes the signals for DNA origins of replication in the β -globin locus, but not in the lamin-B2 locus.

we used for detecting replication origins was still effective when the β -globin cluster was embedded in transcriptionally inactive silenced chromatin. To test this, we examined the β -globin IRs in three cell types in which globin is not expressed and the chromatin is in a relatively closed conformation. The method we used readily detected the β -globin IR in the cell lines that were not transcribing the globin mRNA (Supplementary Figure 8A). These data indicated that ILF2, which recruits DARRT to the HS4 region of the LCR, is essential for firing only from a sub-set of DNA replication origins. RAD50 knockdown, however, did not have any significant effect on origin firing (Figure 6C), suggesting that some components of DARRT are dispensable for its function in replication. The MCM proteins and ORC2 are well-known components of the initiation complex for DNA replication, and perhaps ILF2 enables origin firing at the β -globin locus at least partly by recruiting these components of DARRT to this locus.

Discussion

We have identified a protein complex DARRT that appears to have an important function in DNA replication, histone modification and transcriptional regulation at the β -globin locus, which suggests that this complex may mediate cross-talk between these processes. Several subunits of DARRT have known functions in DNA replication, transcription and repair (Snyder *et al*, 2005, 2009; Zhao *et al*, 2005; Merrill and Gromeier, 2006; Shi *et al*, 2007; Pei *et al*, 2008; Jha and Dutta, 2009; Sakamoto *et al*, 2009). Out of the several DARRT proteins, ILF2 or the ILF2/3 heterodimer are candidate DNA-binding components that function to recruit DARRT to its target DNA sequence in HS4. Interestingly, the ILF2/ILF3 heterodimer has previously been reported to be associated with DNA-PK that is part of the double-stranded break repair complex (Ting *et al*, 1998). DARRT also contains several well-known DNA repair proteins including DNA-PK, RAD50, MRE11 and APEX proteins. The significance of DNA repair proteins in DARRT remains to be determined as RAD50 knockdown did not show the same effects on transcription and origin of replication firing as did MCM5 knockdown.

The ILF2/ILF3 proteins have previously been implicated in regulating multiple processes affecting gene expression including mRNA export, stabilization and translation. The ILF2/3 heterodimer has also been implicated in transcriptional control, where it has been proposed to have a function as a transcriptional activator and in transcript elongation. Of interest, the frog homologue of this complex has been reported to control the expression of the early haematopoietic transcription factor GATA2 (Orford *et al*, 1998). However, in K562 cells, the effects of knockdown of ILF2 on erythroid gene expression does not seem to be a consequence of effects on the expression of GATA2 or the related GATA1 gene, as RT-PCR measurements have shown only a marginal reduction in GATA1 or GATA2 mRNA in the knockdown cells (Supplementary Figure 7).

Inhibition of DARRT recruitment by knockdown of ILF2 inhibits globin transcription, histone acetylation and DNA replication at downstream sequences, consistent with the previous reports of the function of the human LCR in the remote control of these processes (for review, see Mahajan *et al*, 2007). Remote control of DNA replication by distal

sequences is described in other systems such as yeast, Chinese Hamster DHFR and the mouse Th2 gene locus (Friedman *et al*, 1996; Kalejta *et al*, 1998; Hayashida *et al*, 2006). In addition to the requirement for DARRT in transcription and DNA replication, knockdown of ILF2 changes the pattern of acetylation of histones at DNA sequences involved in transcription and replication firing, indicating the involvement of the same protein complex in each of these processes. Knockdown of MCM5 had a similar effect on transcription and replication, indicating that the DARRT complex and not a free heterodimer of ILF2 and ILF3 was responsible for these effects. The significance of the discordant patterns of acetylation of different lysine residue of H3 and H4 and the site-specific effects (Figure 5) of ILF2 knockdown indicate a substantial level of complexity in the regulation of specific histone acetylations. Further, failure of ILF2-silenced MEL cells to undergo DMSO-induced differentiation and globin mRNA production indicate that ILF2 is obligate for at least some aspects of erythroid differentiation. These findings were supported by previous gene ablation experiments in mice, where deletion of ILF3, the usual binding partner of ILF2 and a component of DARRT, resulted in anaemia, cyanosis-impaired oxygen saturation and a poorly developed skeletal muscle system arising from a lack of myogenic differentiation, suggesting that ILF2/3 might function during differentiation in more than one cell lineage (Shi *et al*, 2005).

Previous observations suggest that there could be precise co-ordination of factors associated with DNA replication and transcription in mitochondria, and perhaps in nuclear DNA (Dimitrova, 2006; Falkenberg *et al*, 2007; Hyvarinen *et al*, 2007; Murakami *et al*, 2007; Rudolph *et al*, 2007). It is, however, not known how this co-ordination is achieved. The composition of DARRT suggests that formation of combinatorial complexes of DNA replication and transcription factors could bring about such co-ordination. If indeed ILF2 is involved in the recruitment of DARRT to its target DNA, and recruitment of components of DARRT is necessary for firing of the globin IR, ILF2 knockdown should inhibit the initiation of DNA replication at the globin IR as well as globin gene transcription. The shRNA knockdown studies confirm this, indicating that, apart from transcription, a sub-set of DNA replication origins are targeted by DARRT. Although we cannot absolutely exclude the alternative possibility that ILF2 knockdown acts indirectly through effects on some undiscovered factor elsewhere in the genome that is necessary for use of the globin origin of replication and transcription of the globin genes, it seems simpler to propose that the effects are a result of actions of the DARRT complex containing ILF2 directly at the globin locus where we know it binds.

Although a majority of the replication initiation sites contain AT-rich sequences (Gilbert, 2001), the mechanism for the choice of initiation sites for DNA replication is not clearly known. Frequent usage of a fairly large numbers of cryptic sites, competition between sites and flexibility of origin usage implies that multiple factors determine the final outcome of origin selection. Recruitment of DARRT-associated DNA replication initiation complexes to specific DNA sequences provides an example of transcription factor-mediated recruitment of a part of the pre-initiation complex. A few examples have been reported in animal viral systems, wherein sequence-specific transcription factors recruit the viral-encoded replication initiator/helicase to its replication

origin DNA sequence (Guo *et al*, 1996; Ito *et al*, 1996; Murakami *et al*, 2007) or stabilize the pre-initiation complex (Mul and Van der Vliet, 1992; van Leeuwen *et al*, 1997). In the case of EBV, EBNA1 loads cellular ORC into viral oriP sites (Schepers *et al*, 2001; Ritzi *et al*, 2003). In the case of human c-myc replication origin sequences, deletion of transcription factor-binding sites abolished the initiation of DNA replication (Ghosh *et al*, 2004). The MYC transcription factor was recently reported to be associated with active DNA replication origins independent of its transcriptional activity (Dominguez-Sola *et al*, 2007).

In summary, our data shows that an ILF2 containing multiprotein complex has a multiplex function in regulation of transcription, initiation of DNA replication and histone modification at the β -globin locus. This presumably is a consequence of sequence-specific recruitment of the DARRT complex to sites within the globin locus including the globin LCR HS4. This complex also contains several proteins implicated in DNA repair. Overall, the results imply an unexpected linkage between the transcription, locus-specific initiation of DNA replication and potentially DNA repair.

Materials and methods

Cell culture and media

K562 and HeLa cells were grown in RPMI 1640 (with L-glutamine) supplemented with 10% FBS and antibiotics and antimycotics (Invitrogen). The expansion and erythroid differentiation of CD34⁺ cells was as previously described (Mahajan *et al*, 2009).

EMSA and oligonucleotide probes

The oligonucleotides used in the EMSA assays are listed in Supplementary data. EMSA and gel supershift procedures were performed according to the published protocol (Mahajan and Weissman, 2002). The tiling EMSA screen for identification of K562-specific DNA-protein interactions was carried out by designing a set of short synthetic double-stranded oligonucleotides from the phylogenetically conserved core HS4 region of the LCR. The average size of the oligonucleotides was 35 bp with 7 bp overlap between the adjacent sequences.

Biochemical purification of the DARRT complex

All purification procedures were carried out in the cold room and all buffers were pre-chilled on ice and supplemented with a cocktail of protease inhibitors (Roche). The composition of buffers and other solutions used for the purification procedures are described in Supplementary data. Cells (15 billion) were grown in spinner flasks and log phase cells were harvested by spinning down at 2000 r.p.m. followed by washing twice in cold PBS supplemented with protease inhibitors. Nuclear lysate was prepared by swelling the cells for 30 min in ice cold hypotonic buffer (Buffer A) followed by disruption with a Dounce homogenizer to lyse the cells. Nuclei were collected by centrifugation at 3000 r.p.m. and the nuclear envelope was disrupted in Buffer B. This suspension was then centrifuged at 15 000 r.p.m. for 30 min to collect the nuclear lysate. Salts were dialysed out overnight against Buffer C or by passing the lysate through PD10 desalting columns followed by 10 min centrifugation (15 000 r.p.m. in a Sorvall DuPont RC5C centrifuge). Nuclear lysate was then used for the purification of the HS4-9-binding protein complex. During the purification, the HS4-9-binding proteins were tracked by EMSA assay in all elution fractions from each chromatographic step. The column elution fractions that displayed positive EMSA activity were selected for the next column fractionation. About 100 mg of the crude K562 nuclear extract prepared as described above was loaded on a 30 ml bed volume 25 cm long heparin-Sepharose column pre-equilibrated with Buffer C. The HS4-9-associated protein complex was bound to the heparin-Sepharose column as the flow through did not have any EMSA activity. The heparin-Sepharose column-bound proteins were batch eluted with 150 ml each of 0.2, 0.42 and 0.6 M NaCl in Buffer C. Seventy five 2 ml fractions were collected during each salt elution

and assayed for the HS4-9-associated EMSA activity. We found HS4-9-binding proteins only in the 0.42 M NaCl fractions, which were then concentrated and passed over a FPLC-based Superose-6 sizing column pre-equilibrated with Buffer C. Fractions that showed active HS4-9-binding eluted in the 2–5 MDa size range. These were further purified over an ion-exchange DEAE-Sepharose column that was also pre-equilibrated with Buffer C that contained 0.1 M NaCl. The high-molecular weight HS4-9 DNA-binding proteins were bound to the DEAE-Sepharose column as the flow through tested negative for the EMSA activity. The column was washed extensively with 200 ml of 0.1 M NaCl in Buffer C. The proteins bound to the DEAE-Sepharose column were batch eluted with the 0.2, 0.3 and 0.4 M NaCl salt fractions and tested for HS4-9 DNA sequence-binding activity by EMSA. We found HS4-9 DNA-binding activity only in the protein fractions eluted with 0.2 M NaCl. The EMSA-positive fractions were pooled and dialysed against Buffer C. The preparation was then passed through a scrambled HS4-9 oligonucleotide-affinity column to remove non-specific DNA-binding proteins, if any. The scrambled HS4-9 oligo-column was prepared by attaching a double-stranded scrambled HS4-9 oligonucleotide to CNBr-activated Sepharose 4B. We found the EMSA activity in the flow through of the scrambled HS4-9 oligo-affinity column. Poly-dIdC was added to a final concentration of 0.1 mg/ml to the flow through. This flow through was then passed over an HS4-9 oligo-affinity column that consisted of biotinylated concatameric HS4-9 oligonucleotides bound on Steptavidin Sepharose. The HS4-9-Sepharose column was pre-equilibrated with 0.1 M NaCl containing Buffer C. The proteins were found to bind to the HS4-9 DNA-affinity column as judged by the lack of EMSA activity in the flow through. The column was then washed serially with 50 ml of Buffer C, 15 ml of poly-dIdC in Buffer C at 0.1 mg/ml concentration and finally with 15 ml of Buffer C. The bound proteins were eluted with 0.2, 0.3 and 0.4 M NaCl in Buffer C and tested for the HS4-9-binding protein by EMSA. Fifteen 0.5 ml fractions were collected at each batch elution. We detected EMSA activity only in the 0.2 N NaCl elution fractions. These positive EMSA fractions were pooled, concentrated and displayed on an SDS-PAGE gel. The proteins run of SDS-PAGE were stained with Colloidal Coomassie Blue (Sigma-Aldrich) stain. Near equimolar staining of the protein bands was observed. The SDS-PAGE bands were excised from the gel and digested with trypsin. The tryptic peptides obtained from each protein digestion were sequenced by MS/MS for protein identification.

Mapping the origins of DNA replication

A nascent-strand abundance assay was carried out in asynchronously growing cells to determine the origins of DNA replication. The protocol was adapted from previously published procedures (Buzina *et al*, 2005; Gerbi, 2005). The compositions of the buffers are described in Supplementary data. Briefly, genomic DNA was prepared by suspending 2×10^7 cells in 100 ml of the TEN buffer containing 200 mg proteinase K at 52°C overnight. Next day, the genomic DNA was extracted with phenol: chloroform followed by precipitation with isopropanol. The DNA precipitate was dissolved in 30 ml of NET buffer and loaded onto a benzoylated naphthylated (BND)-cellulose column equilibrated with the same buffer. The flow through was collected and sonicated to an average size of 200 bp to 4 kbp and used as a control in qPCR with the purified nascent ssDNA prepared as follows. The column was washed thoroughly with NET buffer till the 260 nm OD of the wash was <0.05 . The ssDNA bound to the column was eluted with 1.8% caffeine in NET buffer. The eluted DNA was ethanol precipitated, dissolved in 1 ml H₂O and treated with T4 polynucleotide kinase followed by λ -exonuclease digestion. The λ -exonuclease-resistant DNA was run on a 1.5% alkaline agarose gel. Sonicated DNA from the BND-DEAE-cellulose column flow through loaded on the adjacent lane served as a control. DNA of 300–600 bp size range from sample and control lanes was gel extracted, digested by Gelase (Epicentre) and further purified by use of a Qiagen clean up column and subjected to qPCR analysis. For each qPCR analysis, 3.5 ng of the sample and control DNA were used with primers from various regions of the β -globin locus and from the lamin-B replication origin. To test for the specificity of the enrichment of nascent replicating strands containing RNA-DNA hybrids, we digested the ssDNA with Dnase-free RNase for 2 h at 37°C before the λ -exonuclease digestion that resulted in the lack of nascent-strand enrichment at the β -globin locus. We did not see any degradation of 2 μ g of control DNA that was treated with 7 units of RNase (Qiagen,

100 mg/ml; 7000 units/ml, solution) in 50 µl reactions for 5 h at 37°C ruling out the possibility of residual DNase contamination in the RNase.

Additional methods are described in the Supplementary data.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

We thank Dr Peter Kao (Stanford University School of Medicine) for the generous gift of ILF2 and ILF3 antibodies, the Keck DNA Facility

(Yale University) for MS/MS analysis. We thank Dr Ruby Dhar (University of Chicago) for providing reverse transfection protocol, Dr Efim Golub (Yale University School of Medicine) for preparing recombinant DNA plasmid constructs and Dr Jin Lian (Yale University School of Medicine) for maintaining laboratory chemicals and cell culture. This work was supported by funds from NIH Grant no. R01-AG23111, NIH Grant no. R01 DK54369 and in part by the Yale Center of Excellence in Molecular Hematology, NIH DK072442.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Aladjem MI (2004) The mammalian beta globin origin of DNA replication. *Front Biosci* **9**: 2540–2547
- Aladjem MI, Rodewald LW, Lin CM, Bowman S, Cimborra DM, Brody LL, Epner EM, Groudine M, Wahl GM (2002) Replication initiation patterns in the beta-globin loci of totipotent and differentiated murine cells: evidence for multiple initiation regions. *Mol Cell Biol* **22**: 442–452
- Bridges BA (2005) Error-prone DNA repair and translesion synthesis: focus on the replication fork. *DNA Repair (Amst)* **4**: 618–619, 634
- Buzina A, Aladjem MI, Kolman JL, Wahl GM, Ellis J (2005) Initiation of DNA replication at the human beta-globin 3' enhancer. *Nucleic Acids Res* **33**: 4412–4424
- Chen J, Kremer CS, Bender TP (2006) The carbonic anhydrase I locus contains a c-Myb target promoter and modulates differentiation of murine erythroleukemia cells. *Oncogene* **25**: 2758–2772
- Cimborra DM, Schubeler D, Reik A, Hamilton J, Francastel C, Epner EM, Groudine M (2000) Long-distance control of origin choice and replication timing in the human beta-globin locus are independent of the locus control region. *Mol Cell Biol* **20**: 5581–5591
- Corthesy B, Kao PN (1994) Purification by DNA affinity chromatography of two polypeptides that contact the NF-AT DNA binding site in the interleukin 2 promoter. *J Biol Chem* **269**: 20682–20690
- Danis E, Brodolin K, Menut S, Maiorano D, Girard-Reydet C, Mechali M (2004) Specification of a DNA replication origin by a transcription complex. *Nat Cell Biol* **6**: 721–730
- Dimitrova DS (2006) Nuclear transcription is essential for specification of mammalian replication origins. *Genes Cells* **11**: 829–844
- Dominguez-Sola D, Ying CY, Grandori C, Ruggiero L, Chen B, Li M, Galloway DA, Gu W, Gautier J, Dalla-Favera R (2007) Non-transcriptional control of DNA replication by c-Myc. *Nature* **448**: 445–451
- Dziak R, Leishman D, Radovic M, Tye BK, Yankulov K (2003) Evidence for a role of MCM (mini-chromosome maintenance) 5 in transcriptional repression of sub-telomeric and Ty-proximal genes in *Saccharomyces cerevisiae*. *J Biol Chem* **278**: 27372–27381
- Falkenberg M, Larsson NG, Gustafsson CM (2007) DNA replication and transcription in mammalian mitochondria. *Annu Rev Biochem* **76**: 679–699
- Fedosyuk H, Peterson KR (2007) Deletion of the human beta-globin LCR 5'HS4 or 5'HS1 differentially affects beta-like globin gene expression in beta-YAC transgenic mice. *Blood Cells Mol Dis* **39**: 44–55
- Forrester WC, Epner E, Driscoll MC, Enver T, Brice M, Papayannopoulou T, Groudine M (1990) A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. *Genes Dev* **4**: 1637–1649
- Friedman KL, Diller JD, Ferguson BM, Nyland SV, Brewer BJ, Fangman WL (1996) Multiple determinants controlling activation of yeast replication origins late in S phase. *Genes Dev* **10**: 1595–1607
- Gerbi SA (2005) Mapping origins of DNA replication in eukaryotes. *Methods Mol Biol* **296**: 167–180
- Ghosh M, Liu G, Randall G, Bevington J, Leffak M (2004) Transcription factor binding and induced transcription alter chromosomal c-myc replicator activity. *Mol Cell Biol* **24**: 10193–10207
- Gilbert DM (2001) Making sense of eukaryotic DNA replication origins. *Science* **294**: 96–100
- Gottipati P, Helleday T (2009) Transcription-associated recombination in eukaryotes: link between transcription, replication and recombination. *Mutagenesis* **24**: 203–210
- Guan D, Altan-Bonnet N, Parrott AM, Arrigo CJ, Li Q, Khaleduzzaman M, Li H, Lee CG, Pe'ery T, Mathews MB (2008) Nuclear factor 45 (NF45) is a regulatory subunit of complexes with NF90/110 involved in mitotic control. *Mol Cell Biol* **28**: 4629–4641
- Guo W, Tang WJ, Bu X, Bermudez V, Martin M, Folk WR (1996) AP1 enhances polyomavirus DNA replication by promoting T-antigen-mediated unwinding of DNA. *J Virol* **70**: 4914–4918
- Hanawalt PC, Spivak G (2008) Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* **9**: 958–970
- Hayashida T, Oda M, Ohsawa K, Yamaguchi A, Hosozawa T, Locksley RM, Giacca M, Masai H, Miyatake S (2006) Replication initiation from a novel origin identified in the Th2 cytokine cluster locus requires a distant conserved noncoding sequence. *J Immunol* **176**: 5446–5454
- Hyvarinen AK, Pohjoismaki JL, Reyes A, Wanrooij S, Yasukawa T, Karhunen PJ, Spelbrink JN, Holt IJ, Jacobs HT (2007) The mitochondrial transcription termination factor mTERF modulates replication pausing in human mitochondrial DNA. *Nucleic Acids Res* **35**: 6458–6474
- Ito K, Asano M, Hughes P, Kohzaki H, Masutani C, Hanaoka F, Kerppola T, Curran T, Murakami Y, Ito Y (1996) c-Jun stimulates origin-dependent DNA unwinding by polyomavirus large T antigen. *EMBO J* **15**: 5636–5646
- Jha S, Dutta A (2009) RVB1/RVB2: running rings around molecular biology. *Mol Cell* **34**: 521–533
- Kalejta RF, Li X, Mesner LD, Dijkwel PA, Lin HB, Hamlin JL (1998) Distal sequences, but not ori-beta/OBR-1, are essential for initiation of DNA replication in the Chinese hamster DHFR origin. *Mol Cell* **2**: 797–806
- Kamath S, Leffak M (2001) Multiple sites of replication initiation in the human beta-globin gene locus. *Nucleic Acids Res* **29**: 809–817
- Kao PN, Chen L, Brock G, Ng J, Kenny J, Smith AJ, Corthesy B (1994) Cloning and expression of cyclosporin A- and FK506-sensitive nuclear factor of activated T-cells: NF45 and NF90. *J Biol Chem* **269**: 20691–20699
- Kitsberg D, Selig S, Keshet I, Cedar H (1993) Replication structure of the human beta-globin gene domain. *Nature* **366**: 588–590
- Kohzaki H, Murakami Y (2005) Transcription factors and DNA replication origin selection. *Bioessays* **27**: 1107–1116
- Kooren J, Palstra RJ, Klous P, Splinter E, von Lindern M, Grosveld F, de Laat W (2007) Beta-globin active chromatin Hub formation in differentiating erythroid cells and in p45 NF-E2 knock-out mice. *J Biol Chem* **282**: 16544–16552
- Lehmann AR (2005) Replication of damaged DNA by translesion synthesis in human cells. *FEBS Lett* **579**: 873–876
- Lucas I, Palakodeti A, Jiang Y, Young DJ, Jiang N, Fernald AA, Le Beau MM (2007) High-throughput mapping of origins of replication in human cells. *EMBO Rep* **8**: 770–777
- Mahajan MC, Karmakar S, Krause D, Weissman SM (2009) Dynamics of alpha-globin locus chromatin structure and gene

- expression during erythroid differentiation of human CD34(+) cells in culture. *Exp Hematol* **37**: 1143–1156.e3
- Mahajan MC, Karmakar S, Weissman SM (2007) Control of beta globin genes. *J Cell Biochem* **102**: 801–810
- Mahajan MC, Weissman SM (2002) DNA-dependent adenosine triphosphatase (helicase-like transcription factor) activates beta-globin transcription in K562 cells. *Blood* **99**: 348–356
- Mechali M (2001) DNA replication origins: from sequence specificity to epigenetics. *Nat Rev Genet* **2**: 640–645
- Merrill MK, Gromeier M (2006) The double-stranded RNA binding protein 76:NF45 heterodimer inhibits translation initiation at the rhinovirus type 2 internal ribosome entry site. *J Virol* **80**: 6936–6942
- Mitchell JA, Fraser P (2008) Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev* **22**: 20–25
- Mul YM, Van der Vliet PC (1992) Nuclear factor I enhances adenovirus DNA replication by increasing the stability of a preinitiation complex. *EMBO J* **11**: 751–760
- Murakami Y, Chen LF, Sanechika N, Kohzaki H, Ito Y (2007) Transcription factor Runx1 recruits the polyomavirus replication origin to replication factories. *J Cell Biochem* **100**: 1313–1323
- Orford RL, Robinson C, Haydon JM, Patient RK, Guille MJ (1998) The maternal CCAAT box transcription factor which controls GATA-2 expression is novel and developmentally regulated and contains a double-stranded-RNA-binding subunit. *Mol Cell Biol* **18**: 5557–5566
- Pei Y, Zhu P, Dang Y, Wu J, Yang X, Wan B, Liu JO, Yi Q, Yu L (2008) Nuclear export of NF90 to stabilize IL-2 mRNA is mediated by AKT-dependent phosphorylation at Ser647 in response to CD28 costimulation. *J Immunol* **180**: 222–229
- Prioleau MN, Gendron MC, Hyrien O (2003) Replication of the chicken beta-globin locus: early-firing origins at the 5' HS4 insulator and the rho- and betaA-globin genes show opposite epigenetic modifications. *Mol Cell Biol* **23**: 3536–3549
- Rampakakis E, Arvanitis DN, Di Paola D, Zannis-Hadjopoulos M (2009) Metazoan origins of DNA replication: regulation through dynamic chromatin structure. *J Cell Biochem* **106**: 512–520
- Ritzi M, Tillack K, Gerhardt J, Ott E, Humme S, Kremmer E, Hammerschmidt W, Schepers A (2003) Complex protein-DNA dynamics at the latent origin of DNA replication of Epstein-Barr virus. *J Cell Sci* **116**: 3971–3984
- Rudolph CJ, Dhillon P, Moore T, Lloyd RG (2007) Avoiding and resolving conflicts between DNA replication and transcription. *DNA Repair (Amst)* **6**: 981–993
- Sakamoto S, Aoki K, Higuchi T, Todaka H, Morisawa K, Tamaki N, Hatano E, Fukushima A, Taniguchi T, Agata Y (2009) The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. *Mol Cell Biol* **29**: 3754–3769
- Schepers A, Ritzi M, Bousset K, Kremmer E, Yates JL, Harwood J, Diffley JF, Hammerschmidt W (2001) Human origin recognition complex binds to the region of the latent origin of DNA replication of Epstein-Barr virus. *EMBO J* **20**: 4588–4602
- Shi L, Qiu D, Zhao G, Cortes B, Lees-Miller S, Reeves WH, Kao PN (2007) Dynamic binding of Ku80, Ku70 and NF90 to the IL-2 promoter *in vivo* in activated T-cells. *Nucleic Acids Res* **35**: 2302–2310
- Shi L, Zhao G, Qiu D, Godfrey WR, Vogel H, Rando TA, Hu H, Kao PN (2005) NF90 regulates cell cycle exit and terminal myogenic differentiation by direct binding to the 3'-untranslated region of MyoD and p21WAF1/CIP1 mRNAs. *J Biol Chem* **280**: 18981–18989
- Simon I, Tenzen T, Mostoslavsky R, Fibach E, Lande L, Milot E, Gribnau J, Grosveld F, Fraser P, Cedar H (2001) Developmental regulation of DNA replication timing at the human beta globin locus. *EMBO J* **20**: 6150–6157
- Snyder M, He W, Zhang JJ (2005) The DNA replication factor MCM5 is essential for Stat1-mediated transcriptional activation. *Proc Natl Acad Sci USA* **102**: 14539–14544
- Snyder M, Huang XY, Zhang JJ (2009) The minichromosome maintenance proteins 2-7 (MCM2-7) are necessary for RNA polymerase II (Pol II)-mediated transcription. *J Biol Chem* **284**: 13466–13472
- Tang DC, Zhu J, Liu W, Chin K, Sun J, Chen L, Hanover JA, Rodgers GP (2005) The hydroxyurea-induced small GTP-binding protein SAR modulates gamma-globin gene expression in human erythroid cells. *Blood* **106**: 3256–3263
- Ting NS, Kao PN, Chan DW, Lintott LG, Lees-Miller SP (1998) DNA-dependent protein kinase interacts with antigen receptor response element binding proteins NF90 and NF45. *J Biol Chem* **273**: 2136–2145
- van Leeuwen HC, Rensen M, van der Vliet PC (1997) The Oct-1 POU homeodomain stabilizes the adenovirus preinitiation complex via a direct interaction with the priming protein and is displaced when the replication fork passes. *J Biol Chem* **272**: 3398–3405
- Wang L, Lin CM, Brooks S, Cimbora D, Groudine M, Aladjem MI (2004) The human beta-globin replication initiation region consists of two modular independent replicators. *Mol Cell Biol* **24**: 3373–3386
- Xie F, Orr-Weaver TL (2008) Isolation of a Drosophila amplification origin developmentally activated by transcription. *Proc Natl Acad Sci USA* **105**: 9651–9656
- Zhao G, Shi L, Qiu D, Hu H, Kao PN (2005) NF45/ILF2 tissue expression, promoter analysis, and interleukin-2 transactivating function. *Exp Cell Res* **305**: 312–323
- Zhou GL, Xin L, Song W, Di LJ, Liu G, Wu XS, Liu DP, Liang CC (2006) Active chromatin hub of the mouse alpha-globin locus forms in a transcription factory of clustered housekeeping genes. *Mol Cell Biol* **26**: 5096–5105