

The proteomes of transcription factories containing RNA polymerases I, II or III

Svitlana Melnik^{1,3}, Binwei Deng¹, Argyris Papantonis¹, Sabyasachi Baboo¹, Ian M Carr² & Peter R Cook¹

Human nuclei contain three RNA polymerases (I, II and III) that transcribe different groups of genes; the active forms of all three are difficult to isolate because they are bound to the substructure. Here we describe a purification approach for isolating active RNA polymerase complexes from mammalian cells. After isolation, we analyzed their protein content by mass spectrometry. Each complex represents part of the core of a transcription factory. For example, the RNA polymerase II complex contains subunits unique to RNA polymerase II plus various transcription factors but shares a number of ribonucleoproteins with the other polymerase complexes; it is also rich in polymerase II transcripts. We also describe a native chromosome conformation capture method to confirm that the complexes remain attached to the same pairs of DNA templates found *in vivo*.

Eukaryotic nuclei contain three RNA polymerases (I, II and III) that are currently defined by the sets of genes they transcribe¹. Polymerase I produces 45S ribosomal RNA (rRNA) (a precursor of 18S and 28S rRNA), polymerase II transcribes most genes that encode proteins, and polymerase III makes various small RNAs (including 7SK small nuclear RNA and tRNAs). The core of each polymerase has been purified and the structure determined, and we now have detailed knowledge of the way each works *in vitro*². The RNA polymerases also form parts of larger complexes; for example, the polymerase II complex is also involved in capping, splicing and polyadenylation^{3,4}. These megacomplexes may, in turn, be organized into larger 'factories' that contain high concentrations of most machinery required for transcript production^{5,6}. Transcription factories are defined as nuclear sites containing at least two different active transcription units⁵. However, the existence of such factories remains controversial, and one reason for this is that they have not been isolated⁷.

Much of our knowledge about transcription was obtained using isolated polymerase cores assayed on exogenous templates. Two factors make purification of mammalian polymerases engaged on endogenous templates difficult. First, active enzymes represent a quarter of the total enzyme population; most are part of a rapidly diffusing soluble pool that aggregates in nonisotonic

buffers^{8,9}. Therefore, we used isotonic conditions when removing the inactive fraction. Second, engaged polymerases plus their templates and transcripts are housed in factories that are bound to the underlying nuclear substructure^{9,10}. Thus, a typical polymerase I factory in HeLa cells contains about four ribosomal cistrons transcribed on the surface of a 'fibrillar center', which is embedded with others in a nucleolus⁸. Whole nucleoli can be freed from the substructure and purified, and mass spectrometry has yielded a detailed inventory of their contents¹¹. Active polymerases II and III are found in dedicated nucleoplasmic factories, and polymerase II factories have been characterized in detail; high-resolution imaging¹² and quantitative analyses⁸ have shown that one polymerase II factory typically contains about eight polymerizing complexes on the surface of a polymorphic protein-rich core (average diameter ~90 nm, mass ~10 MDa). As caspases deconstruct nuclei during apoptosis, we reasoned that they might be used to release factories from the substructure. (Core subunits of the three polymerases lack sites recognized by the caspases used, except RPB9.)

Here we describe an approach for partial purification and characterization of the three transcription factory complexes from mammalian cells. All have apparent molecular masses of >8 MDa, the size of the largest protein marker available. Each contains a characteristic proteome, as well as shared components. We suggest that these complexes represent large fragments of factory cores that are still bound to the substructure. We anticipate that individual complexes in the pool that we call complex II will be heterogeneous, as different types of nucleoplasmic factories are being uncovered^{5,6}. We have also developed a method, referred to as native 3C (chromosome conformation capture), to validate that these complexes are not aggregation artifacts. With native 3C we show that isolated complexes remain associated with the same templates as found *in vivo* by conventional 3C.

RESULTS

Purification approach

To develop a method to purify transcription factories (Fig. 1a), we began by permeabilizing HeLa cells in a 'physiological buffer' (PB); essentially all transcriptional activity is retained⁸ as the inactive

¹Sir William Dunn School of Pathology, University of Oxford, Oxford, UK. ²Leeds Institute of Molecular Medicine, University of Leeds, St. James's Hospital, Leeds, UK.

³Present address: Division of Molecular Biology of the Cell II, German Cancer Research Center, Heidelberg, Germany. Correspondence should be addressed to P.R.C. (peter.cook@path.ox.ac.uk).

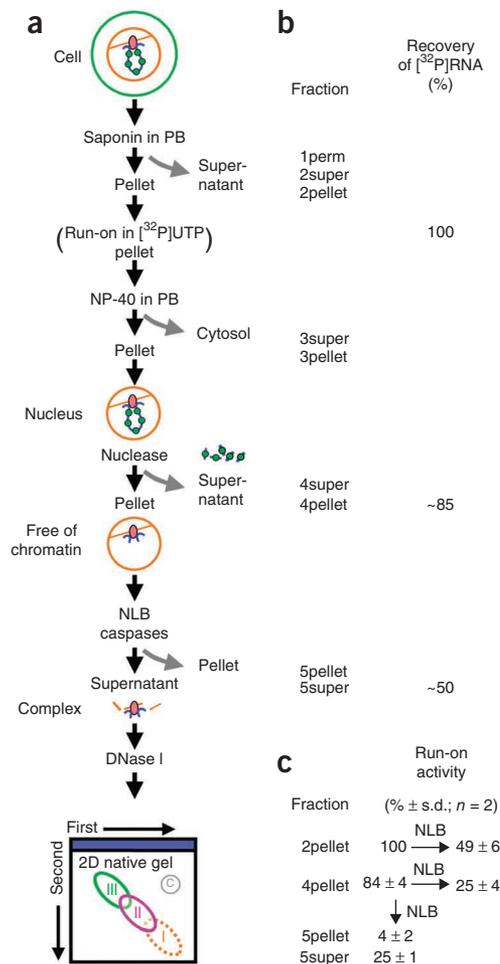


Figure 1 | Purification procedure. **(a)** Strategy. Cartoon shows a chromatin loop with nucleosomes (green circle) tethered to a polymerizing complex (oval) attached to the substructure (brown). The cells are permeabilized and in some cases a run-on is performed in [³²P]UTP so that nascent RNA can be tracked. The nuclei are then washed with NP-40, most of the chromatin is detached with a nuclease (here, DNase I), the chromatin-depleted nuclei are resuspended in NLB and polymerizing complexes are released from the substructure with caspases. After pelleting, chromatin associated with polymerizing complexes in the supernatant is degraded with DNase I, and the complexes are partially resolved in two-dimensional (2D) gels (using blue native and native gels in the first and second dimensions, respectively); rough positions of complexes (and a control region, labeled 'C') are shown. Finally, different regions are excised, and their content is analyzed by mass spectrometry. **(b)** Recovery of [³²P]RNA, after including a run-on. Fractions correspond to those at the same level in **a**. **(c)** Run-on activity assayed later during fractionation (as in **a**, but without run-on at beginning). Different fractions, with names as in **a**, were allowed to extend transcripts by <40 nucleotides in [³²P]UTP, and the amount of [³²P]RNA per cell was determined by scintillation counting. Fractions '2pellet' and '4pellet' were also resuspended in NLB before run-ons were performed; results indicate that NLB reduces incorporation to half or less. Despite this, '5super' has 25% of the run-on activity of permeabilized cells ('2pellet'), which is equivalent to half of the original (after correction for the effects of NLB).

pool is lost⁹ (**Supplementary Note**). Next we isolated nuclei using NP-40, treated them with DNase I and centrifuged the sample to leave most of the inactive chromatin in the supernatant. We then resuspended the pellet in 'native lysis buffer' (NLB), treated the sample with caspases to release large fragments of transcription factories and respun the pellet (**Supplementary Fig. 1** shows the experiments used to optimize release). The supernatant was then retreated with DNase to degrade residual chromatin.

As polymerase II activity is associated with an ~10-MDa core¹², we tested various techniques for purifying large complexes. Free-flow electrophoresis (both zone and isotachopheresis) failed to resolve different complexes. Sedimentation through sucrose or glycerol gradients allowed purification of a minority of polymerase I in polymorphic, ~100-nm complexes (**Supplementary Fig. 2**), without resolving polymerase II and III complexes (which sediment less rapidly). Electrophoresis in 'blue native gels'¹³ was more successful. After running a second dimension without Coomassie blue, we resolved three partially overlapping complexes; all ran slower than the largest (8 MDa) protein marker available.

We monitored the recovery of nascent RNA during purification by allowing polymerases in permeabilized cells to extend their transcripts by 'running on' in [³²P]UTP by <40 nucleotides⁸. Then ~85% of the resulting [³²P]RNA is spun down to pellet after treatment with DNase I (in fraction '4pellet'; **Fig. 1b**). About half this (nascent) [³²P]RNA can be released by a set of caspases (into fraction '5super'; **Fig. 1b**). Substantial amounts of run-on activity

are also released, but determining exactly how much is released is complicated by the truncation of endogenous templates by DNase I and transfer of the pellet to the NLB, which halves run-on activity (**Fig. 1c**). Nevertheless, 25% of the original activity remains in the 5super fraction (**Fig. 1c**), which is equivalent to ~50% after correction for losses due to the buffer. Immunoblotting confirmed that much of polymerases I and II was retained in 5super, whereas more polymerase III was lost (**Supplementary Fig. 1d**).

Polymerizing complexes of >8 MDa

After two-dimensional gel electrophoresis, we found complexes containing nascent [³²P]RNA and protein along the diagonal; immunoblots revealed that the three polymerases were partially resolved and ran as overlapping complexes of >8 MDa (**Fig. 2a**). We named these complexes I, II and III after the polymerases they contain. Complex I ran the fastest, even though it also sedimented the fastest in sucrose gradients (**Supplementary Fig. 2**). We traced this discrepancy to a destabilization induced by the Coomassie blue in the first dimension. In the absence of the stain, complex I runs the slowest (**Fig. 2b**), so we used Coomassie-free gels when purifying complex I. Excised regions of two-dimensional gels enriched in the different complexes contained different proteins (**Fig. 2c**).

Proteomes of the complexes

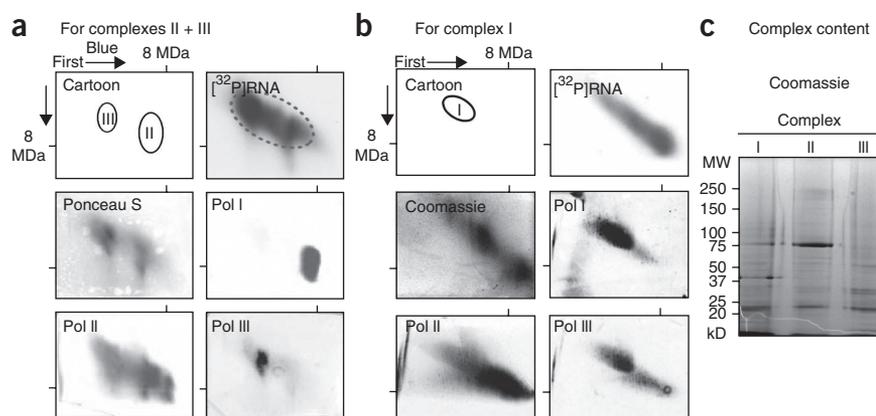
We analyzed the protein content of the transcription factory complexes by liquid chromatography followed by tandem mass spectrometry. We identified peptides using a pipeline¹⁴ that combines three search engines to provide a lower false discovery rate (FDR) compared to the use of only one engine; even so, we selected a conservative FDR of <1%. We detected several hundred proteins in each complex: some unique, others shared (**Fig. 3a**, **Table 1** and **Supplementary Table 1**).

Complexes I and II contained three and five subunits that are unique to RNA polymerases I and II, respectively (**Table 1**). Complex III contained one subunit shared by polymerases I and III

Figure 2 | Resolving different polymerases in native two-dimensional gels (run-ons in [32 P]UTP are included). (a) Resolving complexes II and III with Coomassie blue in the first dimension. The cartoon shows regions selected for mass spectrometry analysis. First, an autoradiograph of the gel was prepared; overlapping spots of (nascent) [32 P]RNA are present along the diagonal. The region indicated (dotted outline) contained ~0.03% of the protein, ~0.8% of the DNA and ~5% of the nascent [32 P]RNA initially present.

After blotting, the membrane was stained with Ponceau S; most protein is present on the diagonal. Next, the membrane was immunoprobed successively for three polymerases

(using antibodies against RPA194, RPB1 and RPC62); the three are partially resolved. Note that complex I is destabilized by the Coomassie blue in the first dimension, and so it migrates rapidly. (b) Resolving complex I (no Coomassie in either dimension). The cartoon shows regions selected for mass spectrometry analysis. First, an autoradiograph was prepared; overlapping spots of (nascent) [32 P]RNA are again present along the diagonal. After staining with Coomassie, spots are seen to overlap regions rich in [32 P]RNA. After blotting, the membrane was probed for the polymerases (as above); complex I now runs the slowest. (c) Proteins in regions indicated in a and b were resolved on a 4–15% SDS-acrylamide gel and stained with Coomassie.



(RPAC1), but none that was unique to polymerase III, consistent with the losses seen in fraction '3super' (Supplementary Fig. 1d). Each complex possessed a characteristic set of proteins (Table 1 and Supplementary Table 1). Reassuringly, 83% of the proteins identified in complex I are also present in the proteome of isolated nucleoli¹¹. Complex II contained general transcription factors such as AP-2, CEBPB and TFIIH (represented by ERCC3), specific regulators such as CTCF and SAFB (B2), and histone methyltransferases (EZH2, SUV39H1 and SUV39H2). Complex III contained Lupus La antigen (a polymerase III factor).

All three complexes share proteins involved in DNA or RNA metabolism including helicases, nucleic acid-binding and nucleotide-binding proteins, ribonucleoproteins (RNPs) and structural proteins such as spectrin and actin (Table 1 and Supplementary Table 1). Many are probably essential constituents of all complexes, whereas others are likely to be cross-contaminants

(for example, polymerase I-specific or polymerase III-specific subunits RPA2, RPA12 and RPAC1 in complex II) resulting from incomplete resolution in the gel.

As determining absolute amounts of proteins by mass spectrometry remains challenging, we used the normalized spectral index method to estimate relative abundances¹⁵. Structural proteins were among the most abundant proteins (Supplementary Table 2), including RNA-binding proteins (the small nucleolar ribonucleoprotein (snoRNP) dyskerin, and heterogeneous nuclear ribonucleoproteins (hnRNPs) H and K), spectrins and lamins in complex I, nucleophosmin in complex II and α -actinin-1 in complex III.

Analysis of GO terms

More than half the proteins in each complex are associated with the gene ontology (GO) term 'gene expression' (Fig. 3a,b), and each complex contained many proteins with expected terms. For example, complex II contained more proteins with 'transcription from RNA polymerase II promoter' (GO: 0006366) than did complexes I and III (Fig. 3b). To place analysis on a more systematic basis, we compared GO terms associated with our

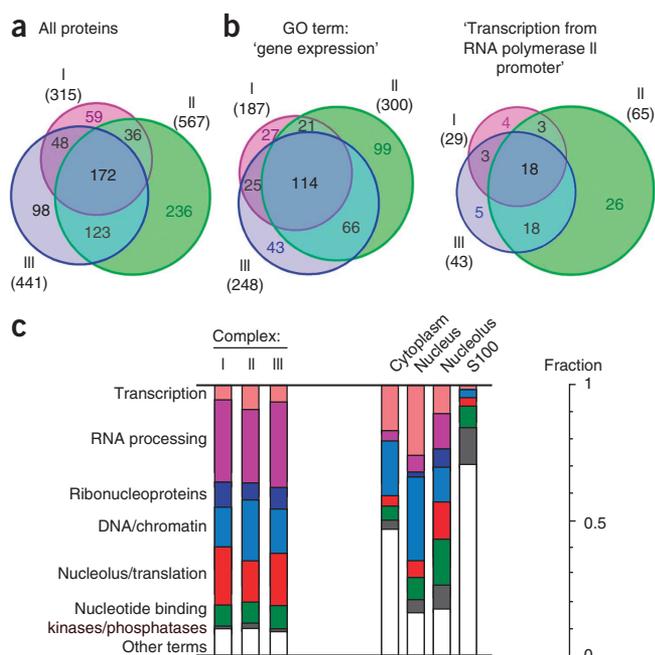


Figure 3 | The content of complexes I, II and III as determined by mass spectrometry. (a) Numbers of proteins in the different complexes and their overlap. (b) Many proteins in each complex are associated with the GO term 'gene expression' (GO: 0010467), and complex II contains more with 'transcription from RNA polymerase II promoter' (GO: 0006366) than do complexes I and III. (c) Most proteins in each complex possess GO terms related to transcript production. Selected GO terms were incorporated into eight groups; for example, 'transcription' includes terms 'RNA polymerase', 'transcription factor' and 'transcription regulation', and 'other terms' includes those not in the other seven groups. Four additional sets of proteins are included for comparison on the right. Some proteins possess terms in more than one group, and terms in each group are expressed as a fraction of the total in all groups. In each complex, 2% of proteins lacked any GO term, and many proteins in the complexes associated with 'other terms' nevertheless turn out to have a role in transcript production (for example, actin²¹ proteasomal constituents¹⁷ and nucleoporins²²). Each complex has a characteristic pattern, which is distinct from those given by proteins with the terms 'cytoplasm' and 'S100'.

Table 1 | Selection of proteins detected by MS in the three complexes

Complex (protein group)	Protein
Complex I	
RNA polymerase	RPA2; RPA34; RPA49; RPABC1.
Transcription regulators	LYRIC; ILF2; SMARCA4.
Complex II	
RNA polymerase	RPB2; RPB3; RPB4; RPB7; RPB9; RPABC3; RPA2 ^a ; RPA12 ^a ; RPAC1 ^a .
Transcription factors	Activator of basal transcription 1; TFII-I; TFIIH subunit 1; XPB helicase; TF20; TF AP-2 α ; TF AP-4; TF Sp3; CCAAT/enhancer-binding protein- β ; CTCF; ATRX; USF1.
Transcription regulators	Scaffold attachment factors B1 and B2; SAFB-like transcription modulator; sex comb on midleg-like protein; splicing factor 1; SWI/SNF-related matrix-associated actin-dependent regulator; major centromere autoantigen B; far upstream element-binding protein 1; HMG20A; chromatin assembly factor 1 subunit B.
Histone modification enzymes	Histone-lysine N-methyltransferases EZH2, SUV39H1 and SUV39H2.
Complex III	
RNA polymerase	RPAC1.
Transcription regulators	Nuclear receptor coactivator 5; SWI/SNF complex subunit 2.
tRNA modification	Lupus La.
Ribosome biogenesis	60S ribosomal protein L35a; probable ribosome biogenesis protein RPL24.
RNA processing	Exosome complex exonuclease MTR3; RNase P protein subunit p14; U6 snRNA-associated Sm-like protein LSm8.
Complexes I + II + III	
RNA helicases	Helicases A, DDX1, DDX18, DDX24, DDX3X, DDX10, DDX47, DDX49, DDX5, DHX15.
Ribonucleoproteins	HnRNPs—A0, A2/B1, A3, C1/C2, F, H, H2, H3, K, L, M, Q, R, U, U-like protein 2. snRNPs—E, Sm D1, Sm D2, Sm D3, U1 RNP A and A', U5 200 kDa helicase, U1 70 kDa, U4/U6 RNP Prp31, 116 kDa U5 component, H/ACA RNP subunit 2 and 4.
Processing factors	Spliceosomal protein SAP 155; SF-3 subunit 1 and 2; SF-3B subunit 3 and 4; U2AF 65 kDa subunit; SF-arg/ser rich 7; SF-13A; CSTF 77 kDa subunit; CPSF subunit 6 and 7.

^aSuggested contaminants.

proteins with the 87,130 terms in a database of all human proteins, or with the 9,682 that are associated just with the GO term 'nucleus' (**Supplementary Fig. 3**). We found that, for example, the five most over-represented terms for the transcription factory proteins compared with all human proteins had obvious connections with transcription, with terms 'RNA binding', 'RNP complex' and 'RNA processing' heading the lists in the GO domains 'molecular function', 'cellular components' and 'biological processes', respectively (**Supplementary Fig. 3a**). Compared to all human proteins, complex II also contained more terms associated with 'gene expression' (GO: 0010467, 300 proteins, $P < 10^{-109}$; see Online Methods for the statistical test used), 'transcription' (GO: 0006251, 149 proteins, $P < 10^{-54}$), 'splicing' (GO: 0008380, 114 proteins, $P < 10^{-65}$) and 'polyadenylation' (GO: 0043631, three proteins, $P < 10^{-3}$). Complex II also contained terms associated with processes closely coupled to (polymerase II) transcription such as 'DNA replication' (GO: 0006260, 58 proteins, $P < 10^{-19}$) and 'DNA repair' (GO: 0006281, 76 proteins, $P < 10^{-24}$). Complex I was enriched in proteins with the terms 'ribosome biogenesis' (GO: 0042254, 88 proteins, $P < 10^{-98}$) and 'rRNA processing' (GO: 0006364, 61 proteins, $P < 10^{-64}$).

To determine which GO terms concisely describe all proteins in the complexes, we developed a software tool, 'MS-prot', which links UniProt accession numbers to associated GO terms. We combined selected terms (for example, 'mRNA cleavage' and 'splicing') into one user-defined group ('RNA processing'); almost all terms associated with our complexes can then be contained in only seven groups related to transcript production (the group 'other terms' contains the remainder). Last, we expressed the number of terms in each group as a fraction of terms in all groups (**Fig. 3c**); proteins in the database associated with terms such as 'cytoplasm' and 'nucleus' serve as controls (**Fig. 3c**). Our complexes yielded different patterns from those of controls; there

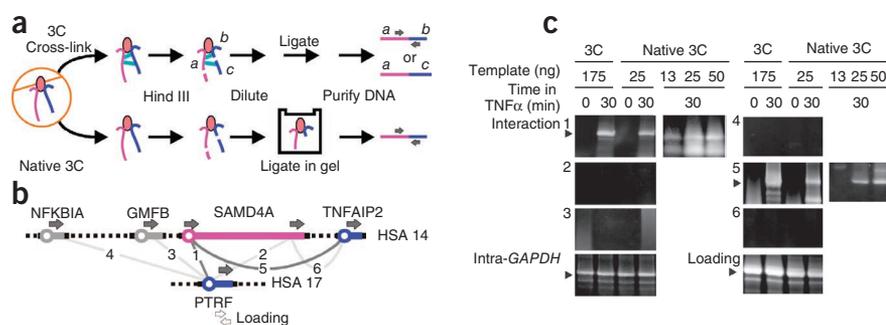
appear to be few contaminants (as 'other terms' is small), and 'RNA processing' is the largest. The 'nucleolus/translation' group is also large; this was expected as active polymerases I and III are found in or on nucleoli where ribosomes are assembled¹⁶, and nascent RNA made by polymerase II colocalizes with >20 ribosomal proteins¹⁷. Taken together, this analysis suggests that each complex has a distinct set of proteins (relevant transcription and processing factors), a large pool of shared ones (RNPs) and few external contaminants.

Confirming selected associations

We next confirmed that some proteins seen by mass spectrometry co-immunoprecipitated with nascent RNA; polymerase II (a positive control), ribosomal protein RPS6, nonsense-mediated decay factor RENT1 and a protein found in many nuclear complexes (PCNA) all co-immunoprecipitated with nascent RNA (**Supplementary Fig. 4a**). We used immunofluorescence (applied conventionally, and coupled to proximity ligation and antibody blocking) to confirm that proteins found only in complex II (for example, CTCF, Sp3 and ATRX) were found in close proximity to active RNA polymerase II, others only in complex III (for example, Lupus La and EXOSC6) lay close to polymerase III (although some Lupus La was found near polymerase II) and still others in all three complexes (for example, DDX1, hnRNPs A2 and B1, and U2AF65) lay close to both polymerases II and III (**Supplementary Fig. 4b,c**).

We also examined whether each complex contained the expected nascent RNAs using quantitative reverse-transcriptase PCR and intronic probes; for example, complex I contained ≥ 33 -fold more nascent 45S rRNA than did the other complexes (**Supplementary Fig. 5a**). The different complexes were also still associated with expected DNA fragments (inevitably some DNA survives DNase I treatment). Complex I contained relatively more DNA encoding 45S rRNA than did the other two, complex II

Figure 4 | Isolated complexes remain associated with DNA sequences found *in vivo*. (a) Strategies for 3C and native 3C. Magenta and blue genes on different chromosomes are co-transcribed by one complex (oval) attached to the substructure (brown). Conventional 3C involves covalently cross-linking (turquoise lines) DNA, cutting (here, with HindIII), dilution, ligation and detection of ligated products by PCR. Note that *a* is joined to *c*, even though there was no stable molecular bridge between the two before cross-linking; such products yield an inevitable background. Native 3C omits cross-linking and relies on pre-existing (native) contacts. As most (inactive) cellular DNA is lost during isolation (including fragment *c*), unwanted background is lower, and wanted 3C products are present in higher concentrations. (b) Targets of primers (gray arrows) used to monitor interactions 1–6; only the contacts that are due to interactions 1 and 6 (purple lines) are detected by both 3C and native 3C. White arrows show primers used for loading controls. (c) 3C and native 3C yield similar bands or contacts (although less template is needed with native 3C). HUVECs were treated with TNF α (0, 30 min), and interactions 1–6 were monitored by 3C and native 3C. Arrowheads indicate relevant 3C bands (all verified by sequencing; additional, nonspecific bands are amplified during the 36 PCR cycles used). 'Intra-GAPDH' 3C and 'loading' controls apply to all panels. Controls (with 13–50 ng template) show that PCR is conducted in the linear amplification range.



was richest in three genes transcribed by polymerase II (*RPS6*, *ARHGAP5* and *MIR191*), and complex III contained the highest amounts of two polymerase III genes (*RN7SK* and *tRNA-leuCAA*; **Supplementary Fig. 5b**).

Native 3C: structure in complex II is similar to that *in vivo*

Our purification strategy (**Fig. 1a**) yields largely template-free complexes. However, treatment with HindIII (instead of DNase I) enables complexes containing more DNA to be isolated, albeit at the cost that the three complexes can no longer be resolved (**Supplementary Fig. 6a**). We therefore developed a new method to show that complexes are associated with the same active templates found *in vivo*.

Chromosome conformation capture (3C) is a powerful tool for detecting the proximity of two DNA sequences in three-dimensional space¹⁸ and involves fixation, which cross-links DNA sequences lying together (**Fig. 4a**). In native 3C (**Fig. 4a**), we omit fixation, and rely on the natural interactions that hold sequences together¹⁹. Here we treated the nuclei with HindIII to remove most of the DNA, released the complexes with caspases, ran the gel (which separates inactive DNA fragments from transcribed fragments attached to complexes), excised the relevant region (which now contains a diluted solution of factories and associated DNA embedded in agarose), added ligase to the gel, recovered the DNA and detected new ligation products by PCR.

For this experiment we used human umbilical vein endothelial cells (HUVECs) because we previously analyzed (by 3C) the changing contacts between a number of their genes induced by tumor necrosis factor- α (TNF α)²⁰. *NFKB1A*, *SAMD4A*, *TNFAIP2* and *PTRF* are normally silent in HUVECs, but 30 min after adding TNF α they become active. Then, the 5' end of *SAMD4A* comes to lie near *TNFAIP2* (on the same chromosome) and *PTRF* (on a different chromosome)²⁰. We first confirmed these 3C results. Before adding TNF α , interactions 1–6 shown in **Figure 4b** did not yield bands on a gel (**Fig. 4c**). But after 30 min, interactions 1 and 5, in which both partners are responsive genes, yielded bands that were indicative of contacts (**Fig. 4c**). Interaction 2 remained undetected; we previously showed that this is because 221-kb *SAMD4A* is so long that the first polymerase to begin transcribing it after stimulation does not reach the region involved in interaction 2 until

~85 min after stimulation, and only then are contacts with *PTRF* or *TNFAIP2* detected²⁰. Interaction 3 (involving a constitutively active gene lying immediately next to responsive *SAMD4A*), interaction 4 (involving two responsive genes lying 20 Mb apart on the same chromosome) and interaction 6 (involving an as-yet untranscribed part of *SAMD4A* and another responsive gene) also remain undetected (**Fig. 4c**). These results confirm those obtained earlier²⁰, and are consistent with some TNF α -responsive genes (but not others), and some parts of responsive genes (but not others), coming together to be transcribed in the same dedicated factory²⁰.

Native 3C yields exactly the same pattern as conventional 3C (**Fig. 4c**). Therefore, we conclude that the contacts we detected in isolated complexes are the same as those *in vivo* and are unlikely to result from artifactual aggregation. Moreover, these interactions are specific, as both 3C and native 3C yield no bands using primers targeting (i) two responding but nonassociating genes (**Fig. 4c**, interaction 4), so contacts do not result simply from an aggregation of active genes, (ii) a polymerase II gene (*PTRF*) and either the (repeated) polymerase I rDNA gene or a polymerase III gene (*RN7SK*), so contacts do not result simply from the effects of high copy number or hyperactivity, and (iii) the polymerase I gene (rDNA) and a polymerase III gene (*RN7SK*; **Supplementary Fig. 6b**), so contacts again do not result from the effects of high copy number or hyperactivity. Notably, less DNA prepared by native 3C gives bands of equivalent intensity (**Fig. 4c**, compare loadings for interactions 1 and 5), which is consistent with fragments still attached to factories being purified away from unattached ones (**Fig. 4a**).

These results also show that our general purification strategy can be extended to a different cell type (that is, HUVECs). Finally, we have used our ability to switch on transcription of selected genes in HUVECs to confirm that (residual) relevant templates are found only in complex II when transcribed. Thus, when uninduced, *SAMD4A*, *EXT1* and *MIR17* are inactive²⁰ and not found in complex II; however, when induced by TNF α , they are enriched in complex II (but not complex III; **Supplementary Fig. 6c**).

DISCUSSION

The existence of transcription factories has been controversial, and one reason given for this is that they have not been isolated⁷. Here we reported a method to isolate large fragments of transcription

factories, and then we characterized their proteomes. We hope that this will encourage a re-evaluation of whether transcription occurs in local sites, the factories, in the nucleus.

In vitro systems for transcribing mammalian genes remain inefficient; the efficiency of our system could be increased by adding purified factors and endogenous templates to our complexes. However, two major difficulties remain. First, we have been unable to recover complexes from two-dimensional gels without aggregation. Second, added templates will also have to displace tightly bound endogenous ones. As a result, recovered 'complexes' have only the usual low transcriptional activity on added templates.

Native 3C may prove to be a useful alternative to 3C for various applications (Fig. 4a). It mainly detects contacts between active alleles in the population, which may be the minority^{6,20}, as most inactive alleles are lost during isolation. Background in native 3C may also be lower, as chemical fixation can stabilize adventitious contacts (Fig. 4a), much of the DNA distant from (contact-rich) nodes is discarded during isolation and less template is required for detection (Fig. 4c).

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturemethods/>.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank J. Bartlett for technical assistance, M. Vigneron (Institut de Génétique et de Biologie Moléculaire et Cellulaire, Strasbourg) for the 7C2 antibody, B. Thomas, D. Trudgian, G. Ridlova and M. Dreger for help with proteomics, M. Shaw for help with electron microscopy, and the Medical Research Council (S.M. and B.D.), EP Abraham Research Fund (B.D.), Biotechnology and Biological Sciences Research Council (A.P.), Wellcome Trust (A.P.) and Felix Scholarship Trust of Oxford University (S.B.) for support.

AUTHOR CONTRIBUTIONS

Experiments were designed by S.M., B.D., A.P., S.B. and P.R.C. S.M. developed the isolation procedure and carried out many of the validation experiments, S.M. and B.D. performed gel electrophoreses and mass spectrometry, A.P. developed native 3C and carried out RT-PCR, S.B. did the light microscopy, and I.M.C. developed software. All authors wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturemethods/>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Roeder, R.G. The eukaryotic transcriptional machinery: complexities and mechanisms unforeseen. *Nat. Med.* **9**, 1239–1244 (2003).
2. Cramer, P. *et al.* Structure of eukaryotic RNA polymerases. *Annu. Rev. Biophys.* **37**, 337–352 (2008).
3. Das, R. *et al.* SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* **26**, 867–881 (2007).
4. Shi, Y. *et al.* Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell* **33**, 365–376 (2009).
5. Cook, P.R. A model for all genomes; the role of transcription factories. *J. Mol. Biol.* **395**, 1–10 (2010).
6. Chakalova, L. & Fraser, P. Organization of transcription. *Cold Spring Harb. Perspect. Biol.* **2**, a000729 (2010).
7. Sutherland, H. & Bickmore, W.A. Transcription factories: gene expression in unions? *Nat. Rev. Genet.* **10**, 457–466 (2009).
8. Jackson, D.A., Iborra, F.J., Manders, E.M.M. & Cook, P.R. Numbers and organization of RNA polymerases, nascent transcripts and transcription units in HeLa nuclei. *Mol. Biol. Cell* **9**, 1523–1536 (1998).
9. Kimura, H., Tao, Y., Roeder, R.G. & Cook, P.R. Quantitation of RNA polymerase II and its transcription factors in an HeLa cell: little soluble holoenzyme but significant amounts of polymerases attached to the nuclear substructure. *Mol. Cell. Biol.* **19**, 5383–5392 (1999).
10. Jackson, D.A. & Cook, P.R. Transcription occurs at a nucleoskeleton. *EMBO J.* **4**, 919–925 (1985).
11. Ahmad, Y., Boisvert, F.M., Gregor, P., Cobley, A. & Lamond, A.I. NOPdb: Nucleolar Proteome Database–2008 update. *Nucleic Acids Res.* **37**, D181–D184 (2009).
12. Eskiw, C.H., Rapp, A., Carter, D.R.F. & Cook, P.R. RNA polymerase II activity is located on the surface of ~87 nm protein-rich transcription factories. *J. Cell Sci.* **121**, 1999–2007 (2008).
13. Novakova, Z., Man, P., Novak, P., Hozak, P. & Hodny, Z. Separation of nuclear protein complexes by blue native polyacrylamide gel electrophoresis. *Electrophoresis* **2**, 1277–1287 (2006).
14. Trudgian, D.C. *et al.* CFP – The Oxford Central Proteomics Facility Pipeline. *Clin. Proteomics* **5** (suppl. 1), 94 (2009).
15. Griffin, N.M. *et al.* Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat. Biotechnol.* **28**, 83–89 (2010).
16. Hopper, A.K., Pai, D.A. & Engelke, D.R. Cellular dynamics of tRNAs and their genes. *FEBS Lett.* **584**, 310–317 (2010).
17. Iborra, F.J., Escargueil, A.E., Kwek, K.Y., Akoulitchev, A. & Cook, P.R. Molecular cross-talk between the transcription, translation, and nonsense-mediated decay machineries. *J. Cell Sci.* **117**, 899–906 (2004).
18. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
19. Cullen, K.E., Kladdé, M.P. & Seyfred, M.A. Interaction between transcription regulatory regions of prolactin chromatin. *Science* **261**, 203–206 (1993).
20. Papantonis, A. *et al.* Active RNA polymerases: mobile or immobile molecular machines? *PLoS Biol.* **8**, e1000419 (2010).
21. Zheng, B., Han, M., Bernier, M. & Wen, J.K. Nuclear actin and actin-binding proteins in the regulation of transcription and gene expression. *FEBS J.* **276**, 2669–2685 (2009).
22. Hou, C. & Corces, V.G. Nups take leave of the nuclear envelope to regulate transcription. *Cell* **140**, 306–308 (2010).

ONLINE METHODS

Cells, general procedures. Monolayer cells were grown in DMEM (Invitrogen) with 5% (vol/vol) FCS (FCS; Biosera); suspension HeLa cells were grown in S-MEM (Invitrogen), 5% (vol/vol) FCS, nonessential amino acids, 2 mM L-glutamine and 11 mg ml⁻¹ sodium pyruvate (all from PAA Laboratories). HUVECs from pooled donors (Lonza) were grown to 80–90% confluency in endothelial basal medium 2-MV with supplements (EBM; Lonza). Recoveries of DNA were measured by scintillation counting after growing cells in [methyl-³H]thymidine (0.25 μCi ml⁻¹; ~50 Ci mmol⁻¹) overnight¹⁰. Unless stated otherwise, all buffers used with permeabilized cells were treated with diethylpyrocarbonate (DEPC) or prepared with DEPC-treated water and kept ice cold, and all washes and spins were done at 400g for 5 min at 4 °C. The amount of protein in the area of a gel containing three complexes (**Fig. 2a**) was measured by densitometry using abstract interfaces for data analysis (AIDA) software and blue carrier immunogenic protein (8 MDa; Pierce) as a standard. Recoveries of [³H]DNA and [³²P]RNA in the same areas were measured by scintillation counting. Protein concentrations were monitored using a Nanodrop ND-1000 spectrophotometer (LabTech). Sequences of some PCR primers are available in **Supplementary Table 3**, and those of others are available on request (see **Supplementary Note**).

Permeabilization and run-on in [³²P]UTP. Run-on transcription was performed using triphosphate concentrations limiting elongation to <40 nucleotides⁸. In brief, HeLa cells were permeabilized with saponin (170 μg ml⁻¹, 5 min; Sigma) in PB. PB (pH 7.4) contains 100 mM potassium acetate, 30 mM KCl, 10 mM Na₂HPO₄, 1 mM MgCl₂, 1 mM Na₂ATP, 1 mM dithiothreitol, 25 units ml⁻¹ RNaseOUT (Invitrogen), 10 mM β-glycerophosphate, 10 mM NaF, 0.2 mM Na₃VO₄ and a 1:1,000 dilution of protease inhibitor cocktail (PIC; Sigma). As the acidity of ATP batches varies, 100 mM KH₂PO₄ was used to adjust the pH. After pelleting, the supernatant is called fraction '2super'. Permeabilized cells in the pellet were then resuspended in PB, incubated (5 min on ice) and pelleted; this process was repeated three times. After resuspension again in PB, permeabilized cells were pre-incubated (33 °C, 3 min), and a run-on was performed (10 min, 33 °C) in 100 μM ATP, 100 μM CTP, 100 μM GTP, 0.1 μM UTP, 50 μCi ml⁻¹, [³²P]UTP (3,000 Ci mmol⁻¹; PerkinElmer) and MgCl₂ giving a concentration of Mg²⁺ ions that was equimolar to that of triphosphates. Reactions were stopped by transfer to ice and immediate addition of EDTA to 2.5 mM. Incorporation of ³²P into acid-insoluble material, and subsequent recoveries of [³²P]RNA (as in **Fig. 1b**) were measured by scintillation counting¹⁰. Permeabilized cells were washed twice with PB to remove unincorporated label before factories were isolated.

Isolating factories. Caspases release polymerases bound to the nuclear substructure more efficiently from HeLa cells growing in suspension as compared to monolayers, so suspension HeLa were used unless stated otherwise. Cells were permeabilized with saponin and washed four times in PB; in some cases, a run-on in [³²P]UTP was performed and the cells were washed twice to remove free label (as above). After resuspension, permeabilized cells were lysed (5 min) in PB plus 0.4% (vol/vol) NP-40, and spun; the supernatant is called fraction '3super'. Nuclei in the

pellet were washed twice in PB + NP-40 (with a 5-min incubation on ice after each resuspension, as above) to give '3pellet'. Resuspended nuclei were digested (30 min, 33 °C) with either (i) DNase I (10 units per 10⁷ cells in 100 μl PB plus 0.5 mM CaCl₂, protease- and RNase-free; Worthington), or (ii) HaeIII (1,000 units per 10⁷ cells, Invitrogen) or (iii) HindIII (1,000 units per 10⁷ cells; New England Biolabs) in PB. Reactions were stopped by adding EDTA to 2.5 mM and cooling in iced water. Chromatin-depleted nuclei were spun (600g, 5 min), and the supernatant ('4super') was collected. The pellet ('4pellet') was resuspended (10⁷ cells per 100 μl) in NLB (pH 7.4). NLB was modified from ref. 13 and contained 40 mM Tris-acetate, 2 M 6-aminocaproic acid (Fluka), 7% (wt/vol) sucrose, 1:1,000 dilution of PIC and 50 units ml⁻¹ RNaseOUT. After 20 min, recombinant caspases 6, 8, 9 and 10 (Calbiochem or Biovision; a total of 2 units in NLB per 10⁷ nuclei) were added; after 30 min at 33 °C, the reaction was stopped with caspase inhibitor III (0.2 mM; Calbiochem), the solution was spun (600g, 5 min) and the supernatant ('5super') and pellet ('5pellet') were collected. '5super' was then treated with DNase I (as above), EDTA (to 2.5 mM), and the sample was split into aliquots, frozen rapidly in dry ice and stored at -80 °C. Conditions for electrophoresis in a native two-dimensional gel were modified from those used previously^{13,23} by increasing the pore size of the gel, modifying the running buffer (to retain run-on activity) and reducing the concentration of Coomassie blue used to provide charge to the hydrophobic complexes analyzed originally. Composite (analytical) gels contained 1.5% acrylamide and 0.7% agarose (SeaKem Gold, Lonza) in 40 mM Tris-acetate (pH 7.4), 7% (wt/vol) sucrose, and 0.01% (vol/vol) Triton X-100, and were run (~1 h, 100 V, constant voltage) in 40 mM Tris-acetate (pH 7.4). A sample with bromophenol blue and xylene cyanol (both added to 0.04% (wt/vol)) was run until the xylene cyanol reached three-quarters of the length (and bromophenol blue is lost). For the 'blue' version, 0.02% and 0.002% (wt/vol) Coomassie blue G-250 were added to samples, and cathode buffers were used in the first dimension, respectively. After running the first dimension, the lane containing the sample was cut out of the gel and polymerized with the second dimension using the same gel and buffers as in the first. For preparative gels used for mass spectrometry, '5super' (from 5 × 10⁷ cells unlabeled with ³²P) was applied to a gel lacking Triton X-100; runs (overnight, 4 °C) began at 100 V (until the sample entered the gel) and then continued at 40 V. Blue carrier immunogenic protein (8 MDa; Pierce) was used as a marker. Gels were stained with Coomassie blue (Imperial protein stain, Pierce).

Mass spectrometry. After fractionation on two-dimensional gels, regions corresponding to those rich in [³²P]RNA and one of the polymerases (detected by autoradiography and immunoblotting using analytical gels run in parallel) were excised, equilibrated (10 min) in 2 changes of 1× Tris-glycine running buffer, loaded on a SDS-acrylamide gel, and subjected briefly to electrophoresis so that all denatured proteins just entered the resolving gel. The whole sample was excised as one gel piece and treated with trypsin, and the resulting peptides were extracted, vacuum dried and injected (usually three injections per sample, 120 min per injection) into a Dionex U3000 nanoHPLC system coupled to a Thermo LTQ Orbitrap mass spectrometer. The three resulting raw data files were merged, converted to .mzXML format

using ReAdW v4.2.1 (<http://tools.proteomecenter.org/wiki/index.php?title=Software:ReAdW>) and submitted to the Central Proteomics Facilities Pipeline¹⁴ (CPFP). Mass spectrometry data are typically analyzed using a single search engine such as ‘Mascot’ (Matrix Science). CPFP uses multiple search engines, modeling tools and target-decoy validation to provide peptide and protein identifications with substantially higher confidence; this provides a stringent test, and proteins in complexes I, II and III were identified with FDRs below 1%. Briefly, .mzXML files are submitted to Mascot, X! Tandem (used with the k-score plugin²⁴), and the Open Search Algorithm²⁵; resulting peptide identifications were then validated with PeptideProphet²⁶. iProphet was used to combine peptide ‘hits’ from the three search engines and to refine identification probabilities according to additional criteria²⁷. All searches were performed against a concatenated target-decoy database (International Protein Index human v3.64; precursor mass tolerance ± 10 p.p.m.; fragment mass tolerance ± 0.5 Da; fixed modification—carbamidomethyl for C; variable modifications—acetylated protein for N-terminal, deamidation for N and Q, and oxidation for M), providing empirical FDRs²⁸ that were compared with estimated ones from the Prophet tools to validate results. By default, results are reported at a 1% target-decoy FDR for both peptides and proteins. For results shown in experiment 1, 90%, 95% and 97% proteins in complexes I, II and III, respectively, were retained when the FDR filter was set more stringently at 0.5%. Two additional experiments (experiments 2 and 3) were also conducted; in both, blotting showed that polymerases were less well resolved, and in experiment 3 complex I was not analyzed. Of the proteins seen in the first experiment in complexes I, II and III, 73%, 60% and 81%, respectively, were also seen in the second. Of the proteins seen in the first experiment in complexes II and III, 39% and 53%, respectively, were also seen in the third (in which fewer proteins were seen). Details of the contents of each complex can be found in **Supplementary Tables 4–7**, and complete proteomic datasets are available at <http://users.path.ox.ac.uk/~pcook/data/ContentOfFactories.html> and <https://proteomecommons.org/tranche/> using the following hash codes: (i) ‘read me’ file, lysDE6I7cXJA140DP5-FCpSYtJKPBWgUUNmOgyTBb04HNd7DKVVzzbzWcUCgho9lrypjaIQWMnN0Zfg0Z+WN0fjk1mc8AAAAAAAABYw==, (ii) experiment 1, lqeHRUGUiEPR4v7WLY0epG4aSLRYid4aCBkJ6ZHYPxzoXB89gRcrX+RQ/98alnP7VT4DVAQLnRLvMW902MsqHyzn5fYAAAAAAAAZpg==, (iii) experiment 2, v3Wi7PA3krKsjlA241eRfMWMcyu8pYnqlimft82ZnZLm39F0BfrmYc/Aguo8jYMR6u1sU8z+rDGx4adsF4BjgqblDYAAAAAAAAM0w==, (iv) experiment 3, pAF+fdNbP/2tkcWx1huqyHh oUejqQTera1UfRnDSHIIpHFPrjDn8V7eu7+fA8PGJ3F1GZXSylU7RYyOjLplwJRvTEAAAAAAAARuA== and (v) comparing complexes I, II and III (seen in all three experiments), I7Cdw8venrUMm8VW0sg5H0sKzCd58MdiJ+n3+Hn3PM1BS6It5NypoQKFNIgIiRSjNr4xNc32woycFb4Q8TNPB99+HgAAAAAAAAC+w==.

GO term analysis. To analyze complex content, protein identifications were exported from CPFP into ProteinCenter (Proxeon); FDR filters of 0.82%, 0.8% and 0.84% (average FDRs of each data set) were maintained throughout analysis in ProteinCenter. Over- and under-represented GO categories (**Supplementary Fig. 3**) were obtained by comparison of frequencies seen with those

obtained with either a standard set of all human proteins (that is, the >87,000 entries in the human International Protein Index; <http://www.ebi.ac.uk/IPI/IPIhelp.html>) or the 9,682 (nuclear) proteins obtained by filtering this database with the GO term ‘nucleus’ (GO: 0005634). *P*-values relating to the significance of any differences seen were evaluated using the statistical test incorporated into ProteinCenter²⁹. To compare GO terms associated with complexes (**Fig. 3c**), we developed software (MS-prot; <http://www.ms-prot.co.uk/>; freely available) that connects an UniProt accession number in a protein database to associated GO terms, and allows the user to define a group of GO terms and filter out proteins linked to terms in the group. The group ‘Transcription’ contained the GO terms ‘RNA polymerase’, ‘transcription factor’ and ‘transcription regulation’; group ‘RNA processing’ contained terms ‘exosome’, ‘mRNA cleavage’, ‘mRNA polyadenylation’, ‘nonsense-mediated decay’, ‘RNA binding’, ‘RNA helicase’, ‘RNA metabolism’, ‘RNA modification’ and ‘splicing’. Group ‘RNPs’ contained the term ‘ribonucleoprotein’. Group ‘DNA/chromatin’ contained the terms ‘DNA binding’, ‘DNA topology’, ‘DNA helicase’, ‘DNA replication’, ‘DNA damage’ and ‘DNA repair’. Group ‘nucleolus/translation’ contained the terms ‘nucleolus’, ‘ribosome’, ‘ribosome biogenesis’ and ‘translation’. Group ‘nucleotide binding’ contained the terms ‘nucleotide binding’ and ‘nucleoside binding’. Group ‘kinases/phosphatases’ contained the terms ‘kinase’ and ‘phosphatase’. Group ‘other terms’ contained all those not included above. Four other sets of proteins are included for comparison: (i) 18,679 proteins associated with the term ‘cytoplasm’ (GO: 0005737), and 9,682 proteins with the term ‘nucleus’ (GO: 0005634) from the International Protein Index (above), (ii) 4,666 proteins from the nucleolus database¹¹ (<http://www.lamondlab.com/NOPdb3.0/>) and (iii) 67 ‘S100’ proteins obtained by filtering entries in the UniProt database (<http://www.uniprot.org/>) with the key word ‘S100’.

Protein quantification. Label-free relative quantification of proteins in samples was performed using the normalized spectral index (SI) method¹⁵, which combines three abundance features (peptide count, spectral count and fragment-ion intensity). SIs were calculated using the output from one search engine. Mascot, using the default significance setting of <0.05 and a script available on request. Use of a single search engine (not three as above) results in a slightly different list of proteins to that obtained with CPFP. To increase stringency, we selected proteins yielding ≥ 3 peptides; 89%, 95% and 95% of the total SI in the output was retained at this stage for complexes I, II and III, respectively. We then ranked surviving proteins according to their SI, and the top ten are listed in **Supplementary Table 2**. As these constitute 66%, 60% and 64% of the total SI seen in complexes I–III, respectively, we are confident these ten proteins are among the most abundant. The same top ten proteins were seen in complexes II and III in experiments 1 and 3 (SI analysis was not performed in experiment 2).

Native 3C. After an initial treatment with HindIII, the region of a gel containing complexes with more DNA (**Supplementary Fig. 6e**) was excised, diced and incubated (4 °C; 3 d) in ligation buffer (NEB), 1 mM ATP and T4 DNA ligase (2,000 units ml⁻¹; NEB). DNA was isolated using a MicroElute gel extraction kit (Omega Bio-Tek). We then performed 3C as described, using

sets of validated primers targeting *SAMD4A* and *PTRF*²⁰. Other primers were selected using Primer 3.0 (<http://frodo.wi.mit.edu/primer3/>) to have an optimal length of 20–22 nucleotides, a T_m of 62 °C, and to yield amplicons of 100–200 bp (**Supplementary Table 3**). PCRs (25 μ l reactions) were performed using GoTaq polymerase (Promega) with one cycle at 95 °C for 2 min, followed by 36 cycles at 95 °C for 45 s, 59 °C for 45 s, 72 °C for 20 s and a final step of 72 °C for 2 min. Amplicons were separated in 2.5% agarose gels, stained with SYBR Green, and scanned in an FLA-5000 scanner (Fuji). The hybrid nature of 3C and native 3C bands was verified by sequencing.

23. Nadano, D., Aoki, C., Yoshinaka, T., Irie, S. & Sato, T.A. Electrophoretic characterization of ribosomal subunits and protein in apoptosis: specific

downregulation of S11 in staurosporine-treated human breast carcinoma cells. *Biochemistry* **40**, 15184–15193 (2001).

24. Maclean, B., Eng, J.K., Beavis, R.C. & McIntosh, M. General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. *Bioinformatics* **22**, 2830–2832 (2006).
25. Geer, L.Y. *et al.* Open mass spectrometry search algorithm. *J. Proteome Res.* **3**, 958–964 (2004).
26. Keller, A., Nesvizhskii, A.I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392 (2002).
27. Shteynberg, D. *et al.* iProphet: Improved validation of peptide and protein IDs in the *trans*-proteomic pipeline. Poster session at: HUPO 7th Annual World Congress (August 16–20, Amsterdam 2008).
28. Elias, J.E. & Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).
29. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc., B* **57**, 289–300 (1995).

The proteomes of transcription factories containing RNA polymerases I, II or III

Svitlana Melnik, Binwei Deng, Argyris Papantonis, Sabyasachi Baboo, Ian M Carr² & Peter R Cook¹

Supplementary Figure 1	Establishing conditions for releasing polymerizing complexes from HeLa cells.
Supplementary Figure 2	Partial resolution of a minority of polymerase I complexes in a sucrose gradient.
Supplementary Figure 3	Complexes I, II, and III from HeLa are enriched in proteins with GO terms related to transcription.
Supplementary Figure 4	Confirming that selected proteins seen by mass spectrometry in HeLa complexes colocalize with nascent RNA and the relevant polymerase.
Supplementary Figure 5	Different complexes from HeLa are significantly enriched in relevant nascent RNAs and genes.
Supplementary Figure 6	Complexes from HUVECs.
Supplementary Table 1	Proteins in different complexes.
Supplementary Table 2	Relative amounts of the ten most abundant proteins in complexes I-III.
Supplementary Table 3	Sequences of primers used for native 3C.
Supplementary Note 1	Procedures used in specific figures.

Note: Supplementary Tables 4–7 are available on the Nature Methods website.

SUPPLEMENTARY NOTE 1

Procedures used in specific Figures

For **Figure 1c**, recovery of run-on activity (expressed as a percentage of radiolabel in [³²P]RNA per permeabilized cell) was assayed as follows. For permeabilized cells in NLB, cells were permeabilized with saponin, washed 4x in PB, resuspended in NLB, and a run-on in [³²P]UTP performed. For “4pellet” and “4pellet” in NLB, nuclei treated with DNase I were pelleted, resuspended in PB or NLB respectively, and run-ons performed. For “5super” and “5pellet”, run-ons were performed using caspase-treated complexes in NLB. Run-on activity was measured by scintillation counting and expressed as a fraction of that of permeabilized cells in PB.

For **Supplementary Figure 1d** (which involved run-on in [³²P]UTP to give results included in **Fig. 1b**), 2.5×10^7 cells were fractionated; prior to electrophoresis, fractions 1, “5super”, and “5pellet” were dissolved (95°C ; 10 min) in 2x SDS-loading buffer, while “2super”, “3super”, and “4super” were concentrated by precipitation with acetone and dissolved as above. Each fraction loaded on the gel had the same volume and was derived from the same number of cells.

For **Supplementary Figure 1a and b**, cells were permeabilized, run-on in [³²P]UTP performed, cells lysed and washed with PB + NP40, and divided into 10 or 6 aliquots. For **Supplementary Figure 1a**, each of the 10 aliquots was resuspended in NLB and treated ± DNase I ± caspases 6, 8, 9, and 10 for 0-30 min. After stopping reactions with caspase inhibitor, samples were spun (600 g, 5 min), and supernatants collected; each supernatant was divided into 3 aliquots for analysis of [³²P]RNA by autoradiography (after gel electrophoresis in a 1.5% acrylamide – 0.7% agarose native gel), ³²P incorporation into acid-insoluble material (by scintillation counting), and polymerase content (by immunoblotting, after electrophoresis in 5% SDS-acrylamide gels). For **Supplementary Figure 1b**, each of the 6 aliquots was treated with DNase I and spun (600 g, 5 min), pellets resuspended in NLB and treated ± caspase 6, 8, 9, or 10 or a mixture of all four. After stopping with caspase inhibitor III and spinning, the content of supernatants was analyzed as for **Supplementary Figure 1a**. For **Supplementary Figure 1c**, monolayer cells were permeabilized, a run-on in [³²P]UTP performed, cells lysed and washed with PB + NP-40, and then divided into 20 aliquots. Each aliquot was digested with DNase I, spun, pellets resuspended in NLB (all as for **Supplementary Fig. 1b**), and treated with different amounts of caspase 9 (0, 0.5, 1, or 2 units) for 0, 15, 30, 45 or 60 min. After stopping and spinning, the amount of RNA polymerase II in supernatants was analyzed by immunoblotting (as in **Supplementary Fig.**

1a).

For **Supplementary Figure 1d**, precast 4-15% Tris-HCl Ready Gels (Bio-Rad) were used. After native or denaturing gel electrophoresis, proteins were transferred onto nitrocellulose membranes (iBlot dry gel transfer system, Invitrogen); transfer was confirmed by staining with Ponceau S (Sigma). Then, membranes were blocked (30 min; 20°C) with 5% non-fat milk (Marvel, Chivers Ireland Ltd) in TBS buffer containing 0.05% Tween-20 (TBS-T) or (for anti-RPC62) protein-free blocking buffer (BB; Pierce). All incubations with primary and secondary antibodies were done in BB. Bound antibodies were visualized using SuperSignal West Pico Chemiluminescent Substrate (Pierce), and detected using Hyperfilm (Amersham) and a Fujifilm Imager (LAS-4000, Fuji). Blots were stripped using Restore Plus western blot stripping buffer (Pierce).

The following primary antibodies were used during immunoblotting to detect: RNA polymerase I – mouse monoclonal anti-RPA194 (1/100 dilution, Santa-Cruz sc-48385); polymerase II – mouse monoclonal anti-RPB1 (7C2³⁰; 1/10,000 dilution; a gift of Marc Vigneron); polymerase III – chicken anti-RPC62 (1/1000 dilution; Abcam ab26185); rabbit anti-RPS6 (1/1000 dilution, Bethyl A300-557A); goat anti-RENT1 (1/1000 dilution, Bethyl A300-036A); rabbit anti-PCNA (1/1000 dilution, Calbiochem PC474), rabbit anti-macroH2A.2 (1/1000 dilution, Abcam ab4173), and mouse monoclonal anti-Grp75 (1/1000, Abcam ab2799).

For **Figure 4** and **Supplementary Figure 6**, HUVECs were starved (18 h) in EBM+0.5% FBS, and treated with TNF α (10 ng/ml; Peprotech) for 0 or 30 min, washed with ice-cold PB, scraped off plates, washed in PB, lysed in PB + 0.4% NP40, and complexes isolated as above except that *Hind*III (1,000 units/10⁷ cells) replaced DNase I in both digestions. As complexes in “5super” remain associated with more chromatin, they are resolved less well in gels with Coomassie blue in the first dimension; therefore, we generally use only one dimension for native 3C, but results using 2D gels are included for comparison in **Supplementary Figure 6**. Analytical gels were run in parallel, and used to locate regions enriched in nascent [³²P]RNA, polymerase II (by blotting), and nucleic acids (after staining with SYBR Green nucleic acid stain I; Invitrogen).

For **Supplementary Figure 4a**, HeLa spinners (10⁸) were grown (10 min) \pm 7.5 mM BrU to label nascent RNA¹⁷, crosslinked with formaldehyde (1%; 5 min; 20°C), the reaction quenched by adding 125 mM glycine, cells collected, washed twice in PBS, and lysed with constant shaking (10 min; 4°C) in 50 mM Tris-HCl (pH 7.4), 140 mM NaCl, 1.5 mM MgCl₂, 0.3 mM sucrose, 0.4% NP-40, 1 mM phenyl-methyl-sulphonyl fluoride (PMSF), 1/1000 dilution of protein inhibitor cocktail (PIC, Sigma-Aldrich) and 20 units/ml RNaseOUT. After

spinning (600 g; 5 min), this procedure was repeated. The resulting nuclei were lysed in 7 M urea, 50 mM Tris-HCl, 5 mM EDTA (pH 7.4), 1 mM PMSF, 1/1000 PIC, 20 units/ml RNaseOUT, and sonicated (5 s pulses followed by 60 s intervals; 12 times; amplitude 10, Sanyo MSE Soniprep 150). The lysate was cleared (4,000 g; 5 min), dialyzed (2 h; 2 changes) against PBS plus 1 mM PMSF using Slide-A-lyser cassettes (Pierce), pre-cleared (1 h) with pre-blocked protein-A agarose beads (Pierce). Proteins from 2.5×10^7 cells in 1 ml supplemented with 100 units/ml RNaseOUT were used for one immunoprecipitation (16 h) with 100 μ l 50% agarose conjugated mouse monoclonal antibody to bromo-deoxyuridine (sc-32323 AC; Santa Cruz Biotechnology). 50 μ l lysate was retained as input and mixed directly with an equal volume of 2XSDS loading buffer. Protein-A agarose beads that had been pre-blocked by incubation (2 h) with normal rabbit IgG (Upstate 12-370) were used as a control. Agarose-A beads were pre-blocked (1 h) by incubation in PBS plus 100 μ g/ml BSA and 50 μ g/ml yeast tRNA (Invitrogen), 1/1000 PIC, and 20 units/ml RNaseOUT. After immunoprecipitation, beads were washed (6 times for 10 min; each spin 4,000 rpm for 1 min) in PBS, 0.4% NP-40, 1 mM PMSF, 1/1000 PIC, and 20 units/ml RNaseOUT. To elute the complex, 100 μ l SDS loading buffer were added to the beads, the sample boiled (5 min) and spun, supernatants collected and incubated (4 h; 65°C) to reverse crosslinks, and used for western blotting.

For **Supplementary Figure 4b** (*in situ* proximity ligation assay³⁵ using Duolink kits; Olink Biosciences), HeLa cells on coverslips were fixed (20 min; 20°C) in 4% paraformaldehyde (Electron Microscopy Science) plus 250 mM HEPES (pH 7.6), and antigens indirectly immunolabelled using various primary antibodies: (i) a mouse monoclonal against RPC32, a subunit of RNA polymerase III (20 μ g/ml; Santa Cruz Biotechnology), (ii) a rabbit polyclonal targeting phospho-serine 2 in the C-terminal domain of the largest subunit of RNA polymerase II (10 ng/ml; Abcam), (iii) goat polyclonal antibodies against CTCF (20 μ g/ml; Santa Cruz Biotechnology) or EXOSC6 (200 μ g/ml; Santa Cruz Biotechnology), and (iv) a normal goat IgG (2 μ g/ml; Santa Cruz Biotechnology). Next, secondary antibodies covalently attached to oligonucleotides were bound to their targets; these secondaries were either “Duolink II PLA probe anti-mouse” or “Duolink II PLA probe anti-rabbit PLUS” applied with “Duolink II PLA probe anti-goat MINUS”. Now, the tethered oligonucleotides were detected using “Duolink II Detection Reagents Orange”. After nuclei were counterstained with 4',6'-diamidino-2-phenylindole (DAPI) in the mounting medium (Vectashield; Vector Laboratories), images were acquired with a confocal laser-scanning microscope (Olympus IX81, 100x Olympus UPlanSApo oil immersion objective with

numerical aperture of 1.4; confocal aperture 117 μm , scanning at 8 $\mu\text{s}/\text{pixel}$) equipped with 405 and 559 nm diodes plus argon (488 nm) lasers, and FLUOVIEW v2.1b software. Z-projections of each Z-stack were made, and the number of fluorescent foci/nucleus counted.

For **Supplementary Figure 4c** (immunofluorescence, both conventional and with antibody blocking), HeLa cells on coverslips were fixed as above, permeabilized (20 min; 20°C) in 0.5% Triton X100 (Sigma) and 0.5% saponin (Sigma), blocked (30 min; 20°C) with 3% bovine serum albumin plus 0.2% cold water fish skin gelatin (Sigma) in PBS, and antigens indirectly immunolabelled¹⁷. For **Supplementary Figure 4ci**, cells were incubated with the rabbit antibody targeting phospho-serine 2 in the C-terminal domain of the largest subunit of polymerase II (10 ng/ml) and the goat anti-CTCF (20 $\mu\text{g}/\text{ml}$), then with secondary donkey anti-rabbit IgG tagged with Cy3 and chicken anti-goat IgG tagged with Alexa488, and nuclei counterstained with DAPI as above; images were acquired with the confocal microscope (confocal aperture of 350 μm , scanning at 10 $\mu\text{s}/\text{pixel}$). For the blocking experiment (**Supplementary Figure 4cii-iv**), primary antibodies included (i) mouse monoclonal antibodies directed against Sp3 and ATRX (both 2 $\mu\text{g}/\text{ml}$; Santa Cruz Biotechnology), (ii) goat polyclonal antibodies against CTCF, EXOSC6, DDX1, hnRNP A2/B1, Lupus La, U2AF65 (2 $\mu\text{g}/\text{ml}$; Santa Cruz Biotechnology), and (iii) normal mouse (2.5 $\mu\text{g}/\text{ml}$; Upstate Cell Signalling Solutions) or goat IgG (2 $\mu\text{g}/\text{ml}$; Santa Cruz Biotechnology). The detection antibodies were either the mouse monoclonal antibody against RPC32 (20 $\mu\text{g}/\text{ml}$) or the rabbit polyclonal antibody against phospho-serine 2 in the C-terminal domain of the largest subunit of RNA polymerase II (10 ng/ml). Fluorescently-tagged antibodies (Jackson ImmunoResearch) were a donkey polyclonal raised against either mouse IgG (tagged with Cy3; 1/200 dilution) or rabbit IgG (tagged with Cy3; 1/2,000). Cells were first incubated (16 h; 4°C) with a blocking antibody, washed twice in 0.05% Tween 20 (Sigma) in PBS at 20°C, fixed again with 4% paraformaldehyde in PBS at 20°C, and washed with PBS (10 min; at 20°C). Next, cells were incubated (1 h; 20°C) with a detection antibody, and washed 4 times with 0.05% Tween 20 in PBS at 20°C. Now, cells were incubated (30 min; 20°C) with the appropriate antibody conjugated with a fluor, washed 3 times (each for 10 min at 20°C) with 0.05% Tween 20 in PBS and once with PBS (10 min) at 20°C, and counterstained with DAPI in mounting medium. Images were collected using an Axioplan 2e microscope (Carl Zeiss MicroImaging, GmbH) fitted with a 63X Zeiss Plan-APOCHROMAT oil immersion objective (numerical aperture 1.4) and a CoolSNAP_{HQ} camera (Photometrics) running under MetaMorph software (Molecular Devices), and analysed using ImageJ³¹. The mean fluorescent intensities of 5 nuclei from each experiment

were exported to Excel, and expressed (as percentages) relative to those given by an “irrelevant” blocking antibody (i.e., goat IgG).

For **Supplementary Figure 2**, 500 μ l “5super” were loaded on sucrose gradients (2:2.5:2.5:2 ml steps of 20:35:50:65%) in PB, spun (20,000 rpm, 15 h, SW41 rotor, Beckman ultracentrifuge L8-M), and 500 μ l fractions collected from the top; aliquots of each fractions were then mixed with SDS loading-buffer, resolved on 7% SDS acrylamide gels, and stained with silver (SilverSNAP stain, Pierce) or analyzed by western blotting. For electron microscopy, fractions were pipetted on to formvar-coated copper grids, allowed to settle (5 min, room temperature), the grids washed with water, blocked (30 min; 20°C) with PBS + 1% BSA, and proteins immuno-gold labeled. Primary antibodies included a mouse monoclonal against UBF (1/10 dilution, sc-13125; Santa-Cruz Biotechnology), a mouse monoclonal (B6-1) against shared subunit RPB6³² (1/10 dilution), rabbit anti-phospho-ser5 in the C-terminal domain of the largest subunit of polymerase II (1/100 dilution; Abcam, ab18488), and a mouse monoclonal (C39-1) against RPC39³² (1/10 dilution). After incubation (1 h) with a primary antibody, grids were washed 3 times (20 min in PBS), incubated (1 h) with goat anti-mouse secondary antibody conjugated with 10-nm gold particles (1/50 dilution, EM.GMHL10; BBI International) or goat anti-rabbit secondary antibody conjugated with 10-nm gold particles (1/50 dilution, EM.GAR10; BBI International), washed 3 times (each 20 min in PBS), stained with 2% phosphotungstic acid (pH 7.0), and imaged (FEI, Technai T12).

For **Supplementary Figures 5 and 6c**, complexes (from $\sim 2 \times 10^7$ HeLa or HUVECs) were resolved in 2D gels, RNA stained using SYBR Green stain II (Invitrogen), regions (of equal weight) containing complexes I-III and a control region excised, and total nucleic acids isolated (EZNA Gel Extraction kit, Omega Bio-Tek) using the manufacturer’s instructions and the same final elution volume for all samples. Amounts of nucleic acid were determined from the optical density at 260 nm. Half of each sample was treated with RNase A (Sigma-Aldrich), and half with RQ1 DNase (Promega) using the manufacturer’s instructions. Quantitative PCR, or quantitative RT-PCR, was conducted using Platinum or Superscript III/Platinum *Taq* Polymerase SYBR Green mix (Invitrogen) on a Rotor-Gene 3000 apparatus (Qiagen) with the following cycling conditions: (qPCR) 95°C for 5 min, followed by 40 cycles of 60°C for 40 sec; (qRT-PCR) 55°C for 10 min, 95°C for 5 min, followed by 40 cycles of 60°C for 40 sec, plus a final step at 40°C for 1 min. Results were analyzed using the standard curve method. Primer pairs targeted intronic sequences; their sequences are available upon request. Single amplicon production by each primer pair was confirmed by gel

electrophoresis and/or melting curve analysis.

References for Supplementary Note 1

30. Besse, S., Vigneron, M., Pichard, E., & Puvion-Dutilleul, F. Synthesis and maturation of viral transcripts in herpes simplex virus type 1 infected HeLa cells: the role of interchromatin granules. *Gene Expr* **4**, 143-161 (1995).
31. Rasband, W.S. ImageJ, <http://rsb.info.nih.gov/ij/> U.S. National Institutes of Health, Bethesda, Maryland, USA (1997-2007).
32. Jones, E., *et al.* Isolation and characterization of monoclonal antibodies directed against subunits of human RNA polymerases I, II, and III. *Exp Cell Res* **254**, 163-172 (2000).
33. Pombo, A., & Cook, P.R. The localization of sites containing nascent RNA and splicing factors. *Exp. Cell Res.* **229**, 201-203 (1996).
34. van Steensel, B., *et al.* Partial colocalization of glucocorticoid and mineralocorticoid receptors in discrete compartments in nuclei of rat hippocampus neurons. *J Cell Sci* **109**, 787-792 (1996).
35. Söderberg, O., *et al.* Characterizing proteins and their interactions in cells and tissues using the in situ proximity ligation assay. *Methods* **45**, 227-232 (2008).
36. Andres V., & Gonzalez, J.M. Role of A-type lamins in signaling, transcription, and chromatin organization. *J. Cell Biol.* **187**, 945-957 (2009).

Supplementary Figure 1. Establishing conditions for releasing polymerizing complexes from HeLa cells.

(a) Release by DNase I and a mixture of 4 caspases.

Cells were permeabilized with saponin in PB, and washed 4x with PB. Nascent transcripts were then labeled by allowing engaged polymerases to extend transcripts in [³²P]UTP, the permeabilized cells washed 4x with PB, and once with NP40 in PB. Next, the resulting nuclei were washed 2x with NP40 in PB and resuspended in the same buffer. After division into aliquots, samples were spun, resuspended in NLB, and either treated with DNase I (10 units/10⁷ cells) or/and a mixture of caspases 6, 8, 9, and 10 (a total of 2 units/10⁷ cells) for 0, 10, 20, or 30 min. After re-pelleting, supernatants were recovered and their contents analyzed.

(i) Release of [³²P]RNA (detected by autoradiography of a “blue native gel”). The longest combined treatment releases significant amounts of [³²P]RNA in complexes of > 8 MDa.

(ii) The amount of [³²P]RNA released into the supernatant (measured by scintillation counting) is expressed as a percentage relative to the total initially in the aliquot (mean of 2 experiments; SDs <1%). The longest combined treatment releases most nascent RNA.

(iii) Release of polymerases II and III, detected by immunoblotting using 5 or 10% acrylamide/SDS gels and antibodies against RPB1 and RPC62, respectively. The longest combined treatment releases significant amounts of the polymerases; more II_O is released than II_A.

(b) Release by different caspases.

Cells were permeabilized with saponin in PB, washed, split, and nascent transcripts in two parts labelled as in **(a)**. For all parts, permeabilized cells were washed 4x in PB, and once in PB + 0.4% NP40; released nuclei were now washed 2x in PB + 0.4% NP40 and resuspended in the same buffer. After division into aliquots, samples were treated with DNase I (10 units/10⁷ cells; 30 min), spun, resuspended in NLB, and treated (30 min) with individual caspases or a mixture of four (giving a total of 2 units/10⁷ cells in each case). After re-pelleting, supernatants were recovered and their contents analyzed as in **(a)**.

(i) Release of nascent [³²P]RNA (detected by autoradiography of a “blue native gel”); all caspases release similar amounts of large complexes.

(ii) Release of nascent [³²P]RNA (measured by scintillation counting; percentages are means from 2 experiments, and all SDs were <2%); all caspases release similar amounts.

(iii) Release of polymerases I, II, and III, detected by immunoblotting using 5% (RPA194, RPB1) or 10% (RPC62) acrylamide/SDS gels and antibodies against RPA194, RPB1, and

RPC62. The three polymerases are differentially released by the four caspases; the mixture releases all.

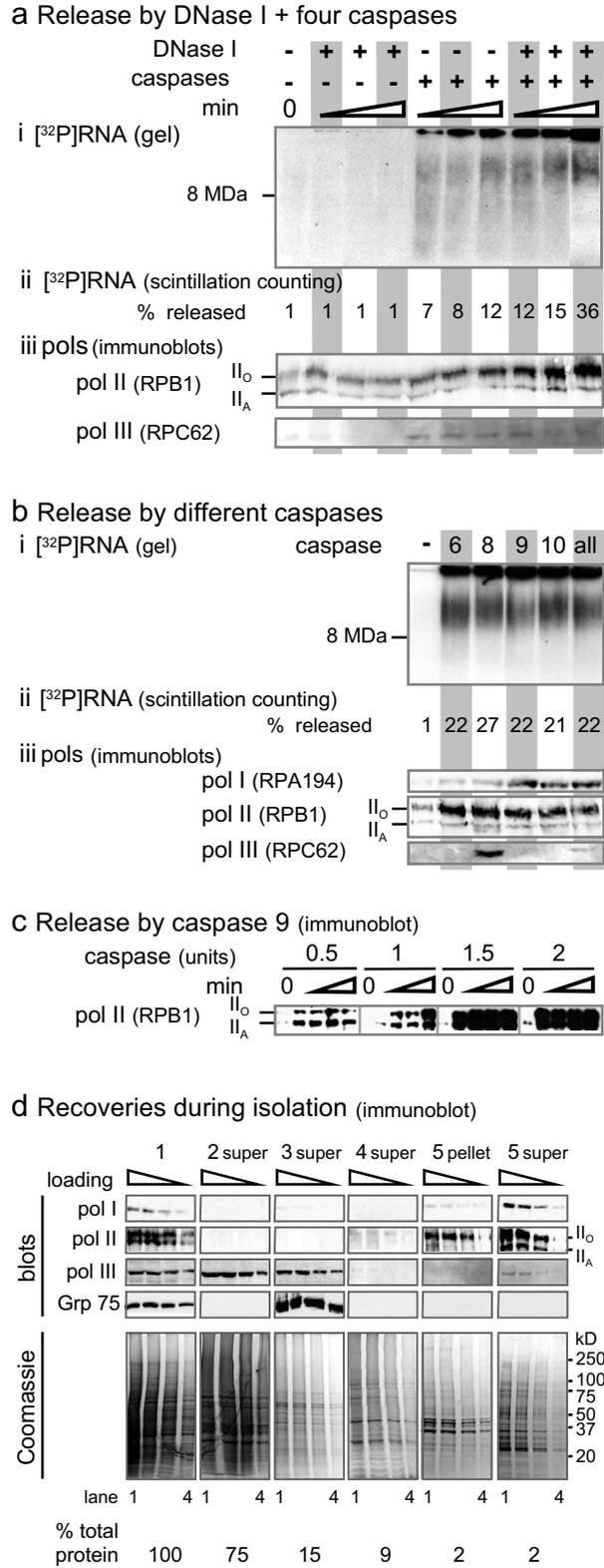
(c) Release of polymerase II (detected by immunoblotting) by different amounts of caspase 9. Treatments were as in (ciii), except that after resuspension in NLB, nuclei were incubated with 0.5, 1, 1.5 or 2 units of caspase 9 for 0, 15, 30, 45 or 60 min. Caspase 9 (2 units for 30 min) releases maximal amounts of forms II_O and II_A.

(d) Immunoblots reveal that most polymerases I and II survive fractionation.

Fractions were collected at different stages (numbered as in **Fig. 1**), proteins from known numbers of cell equivalents resolved by electrophoresis in 4-15% SDS-poly-acrylamide gradient gels, and selected ones detected by immunoblotting; images show relevant regions of blots. Photographs of gels loaded with the same samples and stained with Coomassie blue, and recoveries of total protein are also shown. For each fraction, 10^6 , 7.5×10^5 , 5×10^5 , and 2.5×10^5 cell equivalents were loaded in lanes 1-4.

Coomassie staining reveals the different fractions contain different proteins. Considerable amounts of RNA polymerases I (RPA194) and II (forms II_A and II_O of RPB1 indicated) are found in “5super” (but not in intermediate fractions); however, much of polymerase III (RPC62) is lost in fractions 2 and 3. [Note that it is impossible to determine exact recoveries as protein transfer and/or detection is better the more dilute the solution (so “5super” invariably yields more intense bands than 1, despite loadings from equal numbers of cells).] In contrast, essentially all Grp75 – a mitochondrial marker – is lost in “3super”.

Supplementary Fig. 1



Supplementary Figure 2. Partial resolution of a minority of polymerase I complexes in a sucrose gradient.

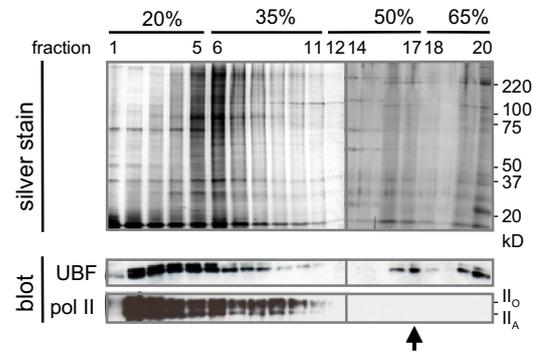
Complexes were released with caspases from HeLa, spun on a sucrose gradient (with steps of 20, 35, 50, and 65%), and fractions collected.

(a) The content of the different fractions analyzed by electrophoresis through SDS-acrylamide gels; gels were either stained with silver, or probed for UBF (or RPB1 using the 7C2 antibody) by immunoblotting. Positions of molecular weight markers, forms II_A and II_O, and fraction 16 used for further analysis, are indicated. Most UBF (in complex I) and polymerase II (in complex II) are found at the top of the gradient, complex III in the middle, and some UBF at the bottom; a minority of UBF is found in fractions 15 and 16.

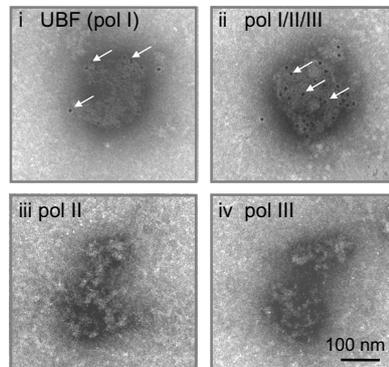
(b) Large complexes in fraction 16 analyzed by immuno-electron microscopy. (i) UBF and (ii) the shared B6 subunit (present in all three polymerases) are found in these complexes (marked by gold particles; arrows), but not (iii) polymerase II or (iv) polymerase III. [The antibodies used in (iii) and (iv) label polymorphic particles in fractions 6 and 13, respectively.] Bar: 100 nm.

Supplementary Fig. 2

a sucrose gradient



b immuno-EM (fraction 16)



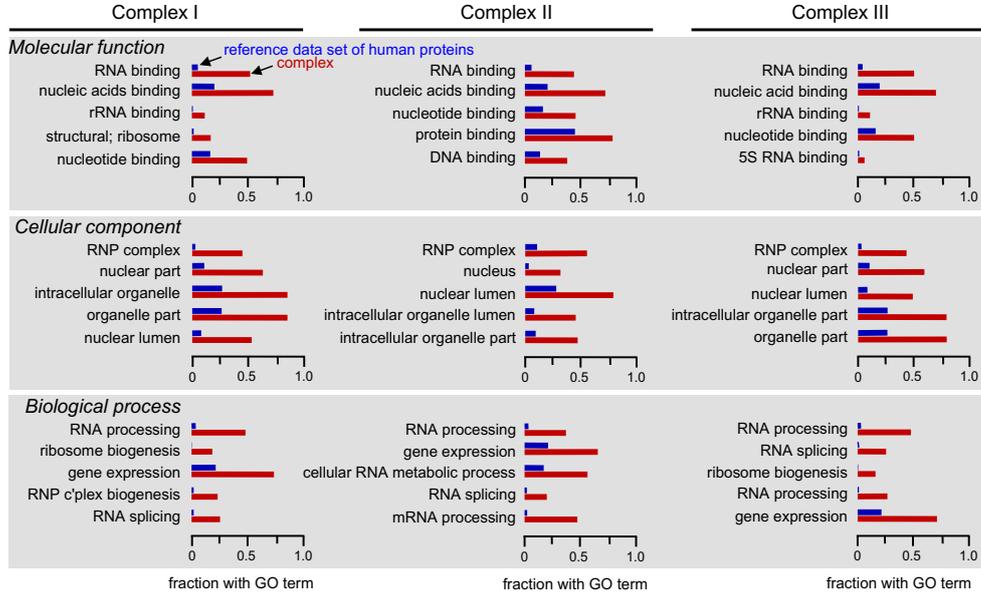
Supplementary Figure 3. Complexes I, II, and III from HeLa are enriched in proteins with GO terms related to transcription.

(a) Comparison with all human proteins. Blue bars indicate the fraction of proteins in the International Protein Index – a reference set of ~87,130 human proteins – that contain the GO terms indicated (derived from each of the three GO domains – “molecular function”, “cellular components”, and “biological processes”); brown bars give corresponding fractions for complexes I, II, and III. Comparison with the reference set reveals that each complex is highly enriched in terms associated with transcript production, and the five terms showing the most significant enrichments are listed in rank order (the difference between every pair shown is significant at a P value of $< 10^{-32}$). Note that one protein may be associated with many terms, and some terms are associated with a significant fraction of all reference proteins. Nevertheless, proteins in the complexes are associated with both highly-restricted and highly-inclusive terms (indicated by short and long blue bars, respectively). For example, the “molecular function” domain includes the highly-inclusive term “nucleic acids binding” and the highly-restricted term “rRNA binding”, and proteins associated with these terms are enriched in complex I. [The two terms most significantly depleted in the “molecular function” domain were: “serine/threonine kinase activity” and “receptor activity” (complex I), “molecular transducer binding” and signal transducer activity (complex II), and “transmembrane receptor binding” and “receptor activity” (complex III); all these have no connection with transcript production.]

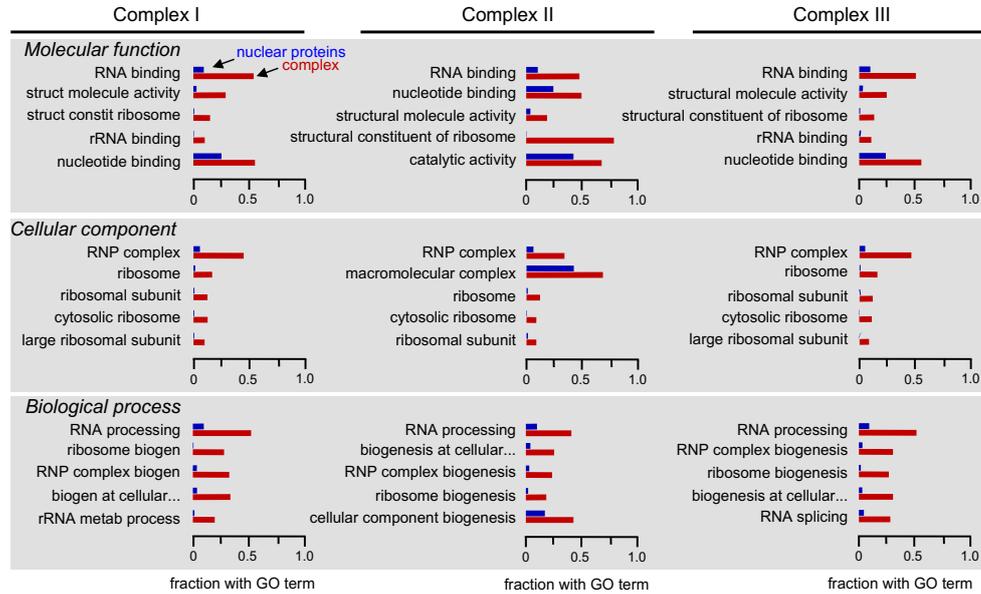
(b) Comparison with nuclear proteins. Here, blue bars refer to the 9,682 proteins in the International Protein Index that are associated with the GO term “nucleus” in the domain “cellular components” (i.e., GO: 0005634); again, the five terms showing the most significant enrichments are listed in rank order (the difference between every pair shown is significant at a P value of $< 10^{-44}$). The terms “RNA binding”, “RNP complex”, and “RNA processing” head the lists. [The two most under-represented GO terms in, for instance, the “biological process” domain were: (for complex I) “regulation transcription from RNA polymerase II promoter by nuclear hormone receptor” and “cell fate commitment”; (for complex II); “digestive system development” and “telencephalon development”; (for complex III) “cell fate commitment” and “sex differentiation”.]

Supplementary Fig. 3

a Proteins in complexes compared with random set of all proteins



b Proteins in complexes compared with nuclear proteins



Supplementary Figure 4. Confirming that selected proteins seen by mass spectrometry in HeLa complexes colocalize with nascent RNA and the relevant polymerase.

(a) Co-immunoprecipitation. Cells were grown \pm 7.5 mM BrU for 10 min to label nascent RNA, treated with 1% formaldehyde for 5 min, nuclei released and lysed, and nascent BrRNA immunoselected using agarose beads coated with anti-BrdU. [Beads coated with a control IgG, and cells grown in the absence of BrU, serve as controls; in the latter case, inputs obtained with cells grown \pm BrU were similar, and so both are not shown.] After washing, bound proteins were recovered from beads, resolved on SDS-acrylamide gels, and immunodetected by blotting using antibodies against RPB1 (7C2 antibody, which detects forms II_O and II_A), ribosomal protein S6 (RPS6), RENT1, PCNA, and macroH2A2. Proteins from 5×10^5 and 2.5×10^5 cell equivalents (for the input), and from 5×10^6 cells (for the IgG control and pull-downs), were loaded on the gel. Three proteins detected by mass spectrometry in some or all of the complexes (i.e., RPS6 in all three complexes, RENT1 in complexes II and III, and PCNA in complex II) co-immunoprecipitate with nascent BrRNA (like the positive control – RPB1). MacroH2A.2 – which is not detected in the complexes – provides a negative control. In the bottom panel, cells were grown \pm 200 μ M 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB; Sigma-Aldrich) for 30 min prior to addition of BrU, and uncoated beads serve as a control. In the absence of DRB, some polymerase II is pulled down with (nascent) BrRNA; in the presence of the transcriptional inhibitor, no polymerase II is detected; this confirms that the anti-BrRNA specifically pulls down nascent RNA, if present.

(b) *In situ* proximity ligation assay³⁵ to detect colocalization of two proteins (using an Olink Bioscience kit). It involves binding of two primary antibodies targeting different antigens, followed by binding of two secondaries, each conjugated to a different oligonucleotide. If the two targets lie within \sim 40 nm, the two oligonucleotides can hybridize to two “padlock” probes; after ligation, “rolling circle replication” amplifies the resulting circularized padlocks. Finally, amplified DNA is detected using fluorescently-tagged oligonucleotides, and cells imaged. Foci then mark sites where the two targets lay close together.

(i) Example. Antibodies targeting two components seen in the same complex (the active form of polymerase II and CTCF, which are both in complex II) yield nuclear foci, whilst antibodies targeting two proteins seen in different complexes (the polymerase III subunit RPC32 and CTCF) yield background numbers of foci. Bar: 10 μ m.

(ii) Quantitative results. Antibody pairs against polymerase II + CTCF (only in complex II), and polymerase III + EXOSC6 (only in complex III) yield significantly more nuclear foci than those given by CTCF + polymerase III (in different complexes) and polymerase III and a

control immunoglobulin ($P < 0.0001$; $n = 27-39$; two-tailed Student's t test).

(c) Immunofluorescence and antibody blocking¹⁷.

(i) A problem. Colocalization is usually demonstrated by immunolabelling one antigen with a red and another with a green fluor, and colocalization then gives yellow in the merged image. However, several inter-related reasons make the use of this approach problematic here. First, a minority of most markers studied to date are found inside factories (for example, only $1/4$ of RNA polymerase II is engaged in factories, and $\geq 90\%$ of ~ 10 transcription factors studied are in a soluble pool and not in factories⁹). Second, markers like hnRNPs and transcription/splicing factors are distributed throughout nuclei to yield immunofluorescence images in which most pixels contain signal above background (even in a single confocal section); then, two such markers inevitably overlap³³. Third, the light microscope has a resolution of ~ 200 nm at best (compared to a ~ 90 -nm factory). Here, the merged image (left; bar $5 \mu\text{m}$) illustrates a single equatorial confocal section and the complex distributions of RNA polymerase II (red; detected using an antibody recognizing phospho-serine 2 in the C-terminal repeat in the largest catalytic subunit) and CTCF (green) in a HeLa nucleus stained with DAPI (blue); arrowheads mark position of line scan. The line scan (middle) illustrates overlap between red and green signals; it is difficult to establish the degree of co-localization (if any) as so many pixels contain signals. The panel on the right illustrates one approach³⁴ that can be used to establish the cross-correlation function of the red and green components in the image (determined by shifting the red component by ± 20 pixels in the x axis, and plotting Pearson's correlation coefficient, R , against Δx). A peak at $\Delta x = 0$ would indicate colocalization between red and green foci, but the peak height is small and the peak width broad. Therefore, we use a higher-resolution approach – antibody blocking.

(ii) Principle behind antibody blocking¹⁷. This exploits the ability of one non-fluorescent antibody (“blocking Ab”) to prevent access of another fluorescent antibody (“detection Ab”) to its target (the two targets must lie within ~ 10 nm, the dimensions of an antibody). In this assay, only the detection antibody is tagged with a fluor. Note that such antibody blocking inevitably reduces the degree of colocalization obtained when using the approaches used in **Supplementary Figures 4b and 4c**.

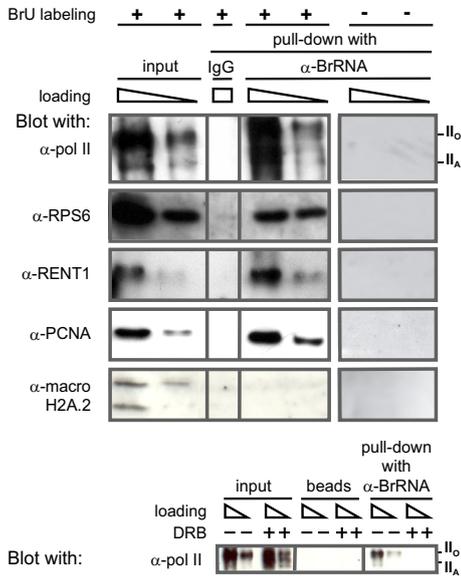
(iii) Example of antibody blocking. Cells were fixed, and incubated with two primary antibodies – one targets the active form of polymerase II (i.e., phospho-serine 2 in the C-terminal domain of the catalytic subunit) and the other was either a non-blocking control (left) or an anti-CTCF (right); after incubation with a secondary antibody tagged with Cy3 that targets only the anti-polymerase, images were collected. Active polymerase II is seen in factories throughout the nucleoplasm, but the anti-CTCF reduces signal (right); such blocking

indicates that CTCF lies within a few nm of the polymerase. Bar: 10 μ m.

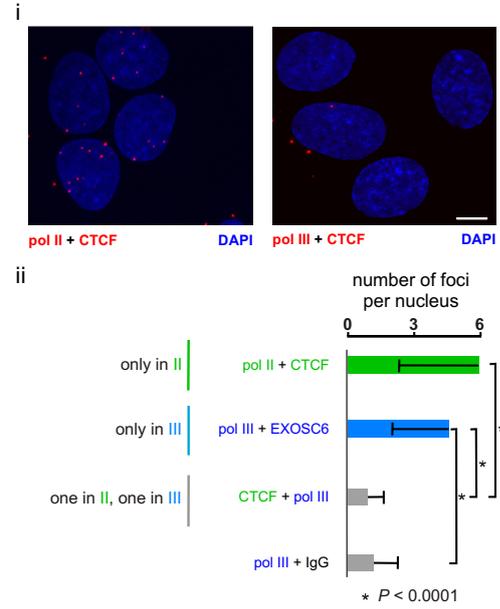
(iv) Colocalization revealed by antibody blocking. Fluorescent intensities over 5 nuclei in images like those in the right-hand panel in (iii) were expressed relative to those in the left-hand panel. Antibodies targeting three proteins seen by mass spectrometry only in complex II (CTCF, Sp3, ATRX) block access of anti-polymerase II. Similarly, antibodies targeting two proteins seen only in complex III (Lupus La antigen, EXOSC6) block access of anti-polymerase III; however, there is some blocking of access of anti-polymerase II by anti-La which ($P = 0.02$; $n = 5$; two-tailed Student's t test). Antibodies targeting proteins seen in all three complexes (DDX1, hnRNP A2/B1, U2AF65) block access of antibodies targeting both polymerases II and III. Clearly, these pairs of epitopes lie sufficiently close together (i.e., within ≤ 10 nanometers) that the unlabeled antibody can block access of the labeled one. These results confirm those obtained by mass spectrometry and in **Supplementary Figure 4b**. *: difference relative to the value given by a non-blocking control antibody was significant ($P < 0.0002$; $n = 5$; two-tailed Student's t test).

Supplementary Fig. 4

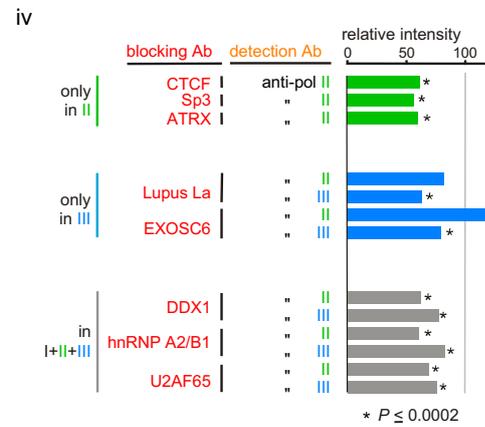
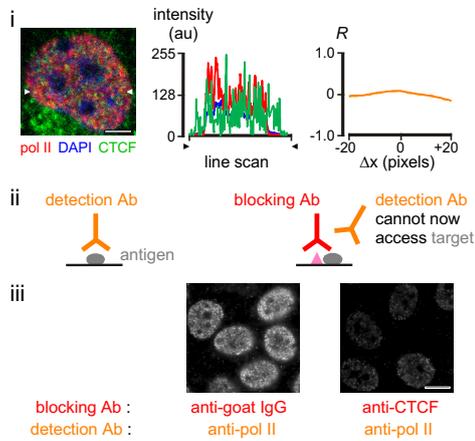
a Co-immunoprecipitation



b *In situ* proximity ligation assay



c Immunofluorescence and antibody blocking



Supplementary Figure 5. Different complexes from HeLa are significantly enriched in relevant nascent RNAs and genes.

Complexes were resolved on 2D gels, regions containing complexes I-III and control regions excised, amounts of intronic RNA (or DNA) determined using quantitative RT-PCR (or quantitative PCR) and normalized relative to amounts of nucleic acid applied to the gel.

Values are averages of triplicates obtained from two independent experiments. Pink rectangles highlight highest enrichments given by the same primer pair (i.e., in one row). *: significantly different from others in the row ($n = 6$, two-tailed Student's t test; $P < 0.01$).

(a) Complexes I, II, and III are enriched in (nascent) intronic RNA transcribed by the respective polymerases. For example, complex I contains 36-fold more nascent 45S rRNA than complex II, while complex III contains at least 34-fold more of both nascent 7SK and transfer RNA than complex II.

(b) Although DNase I was used during isolation, sufficient DNA remains for analysis and complexes I, II, and III are enriched in genes transcribed by the respective polymerases. For example, complex II contains 200-fold more *RPS6* DNA than complex I.

Supplementary Fig. 5

		enrichment relative to input (\pm SD)			
		Complex			
	control area	I	II	III	
a RNA					
Pol I transcript					
	<i>45S rRNA</i>	0.0019 \pm 0.0007	0.11 \pm 0.08*	0.003 \pm 0.003	0.0023 \pm 0.0003
Pol II transcript					
	<i>RPS6</i>	<0.0001	0.0038 \pm 0.002	0.067 \pm 0.002*	0.0005 \pm 0.0004
	<i>ARHGAP5</i>	<0.0001	0.0025 \pm 0.002	0.10 \pm 0.003*	0.0086 \pm 0.001
	<i>MIR191</i>	<0.0001	0.0089 \pm 0.008	0.044 \pm 0.004*	<0.0001
Pol III transcript					
	<i>RN7SK</i>	<0.0001	0.0005 \pm 0.0002	0.0052 \pm 0.002	0.18 \pm 0.04*
	<i>tRNA-leu(CAA)</i>	0.0099 \pm 0.006	0.0004 \pm 0.0002	0.0011 \pm 0.0003	0.017 \pm 0.01*
b DNA					
Pol I gene					
	<i>45S rRNA</i>	<0.0001	0.1 \pm 0.06*	0.011 \pm 0.004	0.0057 \pm 0.001
Pol II gene					
	<i>RPS6</i>	<0.0001	0.0017 \pm 0.002	0.36 \pm 0.02*	0.009 \pm <0.001
	<i>ARHGAP5</i>	<0.0001	0.0064 \pm <0.001	0.29 \pm 0.05*	0.087 \pm 0.02
	<i>MIR191</i>	<0.0001	<0.0001	0.36 \pm 0.3*	0.009 \pm 0.0005
Pol III gene					
	<i>RN7SK</i>	<0.0001	0.00086 \pm <0.001	0.0023 \pm 0.0005	0.2 \pm 0.02*
	<i>tRNA-leu(CAA)</i>	<0.0001	0.0013 \pm 0.001	0.001 \pm 0.001	0.15 \pm 0.13*

most highly enriched if significantly different

Supplementary Figure 6. Complexes from HUVECs.

(a) After cutting with *HindIII* and electrophoresis in 2D gels, complexes remain associated with more DNA (and so cannot be sufficiently resolved into complexes I, II, and III).

Unstimulated HUVECs were permeabilized, engaged polymerases allowed to extend their transcripts in [³²P]UTP, and nuclei isolated; after removing most chromatin with *HindIII*, complexes were released with caspases, attached chromatin trimmed with *HindIII*, and complexes run on three 2D native gels (with Coomassie blue in the first dimension).

(i) Cartoon indicating directions of migration in the two dimensions, plus the position of the 8-MDa marker.

(ii) An autoradiograph of the first gel; (nascent) [³²P]RNA is seen along the diagonal.

(iii) The first gel was stained with Coomassie blue; most protein is on the diagonal.

(iv) One gel was immunoblotted and probed using an antibody (7C2) against polymerase II. [Other immunoblots show that polymerases II and III are found in the same region.]

(v) One gel was stained with SYBR green; most DNA is on the diagonal. The area indicated was excised and used for native 3C.

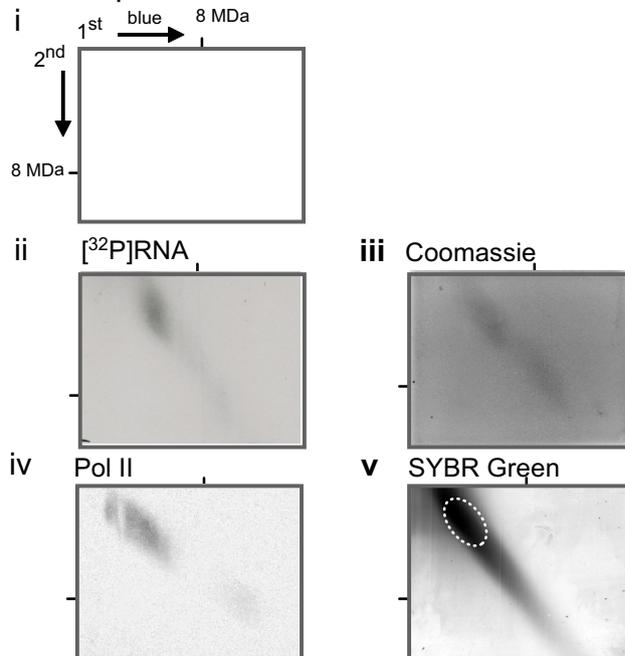
(b) Some controls demonstrating specific associations. HUVECs were stimulated with TNF α for 30 min, complexes associated with residual *HindIII* fragments isolated from a 2D gel, and native 3C conducted using 25 ng template, 36 cycles, and primer pairs indicated. Standard 3C was also conducted on the same cells. With both native 3C and 3C, only primers targeting *SAMD4A* and *PTRF* yield a band (indicating the two genes lie close together). Loading controls (as in **Fig. 4**) show equal amounts of DNA are present in each sample. Note also that although complex II is unresolved from complexes I and III in the 2D gel, native 3C shows that *PTRF* – a responsive polymerase II gene – only contacts another responsive polymerase II gene (i.e., *SAMD4A*), but not the repeated polymerase I gene (the gene encoding 45S *rRNA*; contacts assessed using two different primers, F1 and F2) or a polymerase III gene (*RN7SK*, which encodes the small nuclear RNA 7SK). This indicates that the *SAMD4A:PTRF* interaction is specific.

(c) Genes are found in the relevant complexes (isolated from HUVECs using DNase I) only when transcriptionally active. [Despite the use of the nuclease, some DNA survives to remain associated with the complexes.] HUVECs were treated \pm TNF α for 30 min, complexes prepared using DNase I and resolved on 2D gels, regions containing complexes II and III (and a control region) excised, and amounts of DNA determined using quantitative PCR; amounts were normalized first relative to amounts of nucleic acid applied to the gel, and to amounts found in the control region. Values are averages of triplicates obtained from two

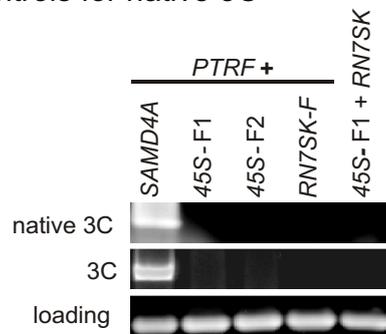
independent experiments. In untreated cells, all four polymerase II genes (*SAMD4A*, *EXT1*, *MIR17*, *AFP*) are inactive, and background levels of their DNA are found in complex II. However, after stimulation, significantly more DNA from the three responding genes (*SAMD4A*, *EXT1*, *MIR17*) – but not that from non-responding *AFP* – can be found in complex II (*: significantly different; $n = 6$, two-tailed Student's t test; $P < 0.01$). This indicates that only transcriptionally-active genes are found in complex II. Complex III is enriched in the polymerase III gene (*RN7SK*), and this does not change on stimulation. This gene is active both in the presence and absence of $\text{TNF}\alpha$, and more is always found in complex III.

Supplementary Fig. 6

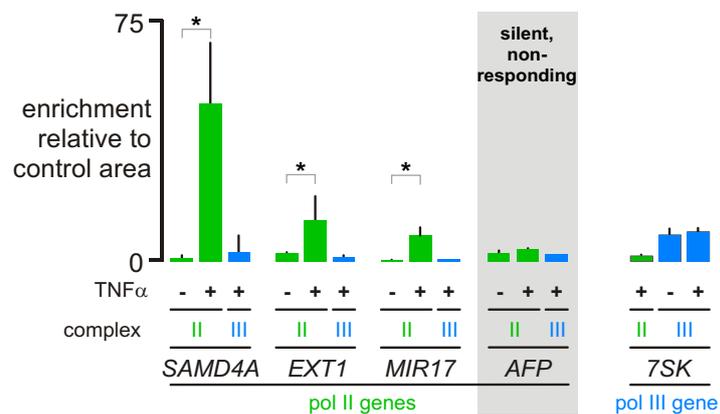
a Complexes used for native 3C



b Controls for native 3C



c Complexes are enriched in specific genes (qPCR)



Supplementary Table 1. Proteins in different complexes. X indicates protein present in the dataset analyzed, while y and z indicate the protein was also seen in two additional experiments (complex I was not analyzed in the last).

.Gene	Complex		
	I	II	III
<i>AATF</i>		Xy	Xyz
<i>ABT1</i>		X	
<i>ACIN1</i>	Xy	Xyz	Xyz
<i>ACTBL2</i>			Xz
<i>ACTL6A</i>	X	X	Xy
<i>ACTN1</i>	Xy	Xyz	Xyz
<i>ACTN4</i>	Xy	Xyz	Xyz
<i>ACTR2</i>			Xy
<i>ACTR3</i>	Xy		Xy
<i>ACTR3B</i>		Xyz	
<i>ADAR</i>		Xyz	Xyz
<i>ADARB1</i>			Xy
<i>ADD1</i>		X	
<i>ADNP</i>		X	
<i>AHCTF1</i>		Xy	
<i>AIMP1</i>	X		
<i>ALB</i>		X	
<i>ALDH18A1</i>		Xz	
<i>ALPI</i>		Xy	
<i>ANAPC1</i>		Xy	
<i>ANAPC7</i>	X		
<i>ANKFY1</i>	X		
<i>ANXA2</i>		Xy	X
<i>AP2A2</i>		X	
<i>AP2B1</i>			Xy
<i>AP2M1</i>			X
<i>APEX1</i>		Xy	
<i>API5</i>		Xyz	Xyz
<i>APOBEC3B</i>	X	Xz	Xy
<i>APOBEC3C</i>		Xz	
<i>AQR</i>			Xyz
<i>ARID4B</i>		X	
<i>ARPC1B</i>			X
<i>ARPC2</i>	Xy		Xy
<i>ATAD3A</i>	X		
<i>ATP6V1A</i>			X
<i>ATRX</i>		Xy	
<i>BAG2</i>			Xy
<i>BANF1</i>		X	
<i>BAZ1A</i>		X	
<i>BAZ1B</i>		Xy	Xyz
<i>BCAS2</i>	Xy		Xy
<i>BHLHE40</i>		X	
<i>BMS1</i>	Xy	Xy	Xyz
<i>BOP1</i>		X	Xy
<i>BRIX1</i>	Xy	Xyz	Xyz

.Gene	Complex		
	I	II	III
<i>BUB3</i>		X	
<i>BYSL</i>	X	X	
<i>C14orf166</i>	Xy	Xyz	Xy
<i>C14orf21</i>			Xz
<i>C15orf44</i>		Xy	
<i>C15orf57</i>		Xyz	Xy
<i>C17orf42</i>		X	Xy
<i>C1orf107</i>		Xy	Xy
<i>C1orf25</i>	Xy	Xy	Xy
<i>C1orf77</i>	Xy		
<i>C22orf28</i>	Xy	Xy	Xy
<i>C3orf26</i>			Xy
<i>C6orf150</i>		Xyz	
<i>CALD1</i>		X	
<i>CALR</i>	X		
<i>CAPRIN1</i>		Xy	Xy
<i>CAPZA1</i>		X	Xyz
<i>CAPZB</i>		Xyz	Xz
<i>CASP10</i>			X
<i>CASP8</i>		Xy	X
<i>CBX1</i>		Xyz	
<i>CBX5</i>		Xy	
<i>CCAR1</i>		Xy	Xy
<i>CCDC86</i>	Xy	Xy	Xy
<i>CCNL1</i>	X	X	
<i>CD3EAP</i>	X		
<i>CD44</i>		Xy	
<i>CDC16</i>		X	
<i>CDC40</i>	Xy		Xy
<i>CDC5L</i>	Xy		Xy
<i>CDC73</i>		Xy	Xy
<i>CDC48</i>		Xy	Xy
<i>CEBPB</i>		Xy	
<i>CEBPZ</i>	X	Xy	
<i>CELF1</i>	Xy	X	Xy
<i>CENPB</i>		Xy	
<i>CENPK</i>		X	
<i>CENPM</i>		X	
<i>CENPV</i>	X		
<i>CHAF1B</i>		Xz	
<i>CHERP</i>		X	
<i>CIRH1A</i>	Xy		Xy
<i>CLK2</i>		X	
<i>CLTA</i>	X		
<i>CLTC</i>	Xy		
<i>CLTCL1</i>	X		
<i>COIL</i>			Xy
<i>CORO1C</i>		X	X
<i>CPNE8</i>	Xy		X
<i>CPSF1</i>		Xyz	Xy
<i>CPSF2</i>		Xz	

.Gene	Complex		
	I	II	III
<i>CPSF6</i>	Xy	Xy	Xy
<i>CPSF7</i>	Xy	Xyz	Xyz
<i>CRNKL1</i>			Xyz
<i>CSDA</i>	X	Xyz	
<i>CSDE1</i>		X	
<i>CSNK1A1L</i>			yz
<i>CSNK2A1</i>		Xyz	Xyz
<i>CSNK2A2</i>		Xz	
<i>CSNK2B</i>		X	X
<i>CSTF1</i>			Xy
<i>CSTF3</i>	Xy	Xy	Xyz
<i>CTCF</i>		X	
<i>CTNNB1</i>		X	
<i>CTR9</i>		X	
<i>CTTN</i>		Xy	
<i>CUL4B</i>		X	
<i>CWC22</i>			Xz
<i>DAZAP1</i>		Xy	Xy
<i>DBN1</i>		Xz	
<i>DBT</i>		Xy	Xy
<i>DCAF13</i>			X
<i>DDX1</i>	Xy	Xyz	Xy
<i>DDX10</i>	Xy	Xyz	Xy
<i>DDX17</i>	X	Xz	Xz
<i>DDX18</i>	Xy	Xyz	Xyz
<i>DDX21</i>	Xy	Xyz	Xyz
<i>DDX23</i>		Xy	Xyz
<i>DDX24</i>	Xy	Xyz	Xyz
<i>DDX27</i>		Xyz	Xyz
<i>DDX3X</i>	Xy	Xyz	Xyz
<i>DDX42</i>		X	Xyz
<i>DDX46</i>	X		
<i>DDX47</i>	Xy	Xyz	Xyz
<i>DDX49</i>	Xy	Xy	Xy
<i>DDX5</i>	Xy	Xyz	Xyz
<i>DDX50</i>		Xyz	
<i>DDX51</i>	X		Xy
<i>DDX52</i>	Xy		Xyz
<i>DDX54</i>	Xy	Xyz	Xyz
<i>DDX56</i>	XY	Xy	
<i>DEK</i>		Xyz	Xy
<i>DHX15</i>	Xy	Xyz	Xyz
<i>DHX16</i>		X	
<i>DHX30</i>		Xyz	Xyz
<i>DHX36</i>		X	
<i>DHX37</i>	X	X	
<i>DHX8</i>	Xy		Xy
<i>DHX9</i>	Xy	Xyz	Xyz
<i>DIMTIL</i>	X		
<i>DKC1</i>	Xy	Xyz	Xyz
<i>DMAPI</i>	X		

.Gene	Complex		
	I	II	III
<i>DNAJA2</i>		X	
<i>DNAJA3</i>	Xy	X	Xy
<i>DNTTIP1</i>		X	
<i>DNTTIP2</i>		X	Xy
<i>DPY30</i>		Xz	
<i>EBNA1BP2</i>	Xy	Xyz	Xyz
<i>EEF1A2</i>	Xy		Xz
<i>EEF1D</i>		X	
<i>EEF1E1</i>	X		
<i>EEF1G</i>		Xz	X
<i>EFTUD2</i>	Xy	Xyz	Xyz
<i>EIF1AY</i>		X	
<i>EIF2S1</i>		Xy	
<i>EIF2S2</i>		X	
<i>EIF2S3</i>		X	
<i>EIF3A</i>		Xy	
<i>EIF3CL</i>		X	
<i>EIF3D</i>		X	
<i>EIF3F</i>		Xz	
<i>EIF3I</i>		X	
<i>EIF3L</i>		Xy	
<i>EIF4A3</i>	Xy	Xyz	Xyz
<i>EIF6</i>	Xy	Xz	Xy
<i>ELAVL1</i>	Xy	Xyz	Xyz
<i>EMG1</i>	Xy	Xyz	Xyz
<i>EPPK1</i>		Xy	Xy
<i>EPRS</i>	Xy		
<i>ERCC3</i>		X	
<i>ESF1</i>		X	
<i>ESRRA</i>		X	
<i>EXOSC2</i>	Xy	Xy	Xyz
<i>EXOSC3</i>		Xy	
<i>EXOSC4</i>	Xy	Xy	
<i>EXOSC5</i>	Xy	Xy	
<i>EXOSC6</i>			Xy
<i>EXOSC7</i>		X	
<i>EXOSC8</i>	Xy	Xy	Xy
<i>EXOSC9</i>	X	Xz	
<i>EZH2</i>		X	
<i>EZR</i>		X	
<i>FAM98A</i>		X	
<i>FAM98B</i>			Xy
<i>FARSA</i>			X
<i>FBL</i>	Xy	Xyz	Xyz
<i>FBLL1</i>			Xyz
<i>FIP1L1</i>	X		
<i>FLJ27502</i>	X		
<i>FLNA</i>	X		
<i>FLNB</i>	X		
<i>FMR1</i>		Xz	Xyz
<i>FTSJ3</i>	Xy	Xy	Xyz

.Gene	Complex		
	I	II	III
<i>FUBP1</i>		X	Xy
<i>FUS</i>		Xyz	Xyz
<i>G3BP1</i>		Xz	
<i>GAR1</i>	Xy	Xyz	Xy
<i>GLT25D1</i>		Xz	
<i>GLTSCR2</i>	Xy		Xyz
<i>GLYR1</i>		X	Xy
<i>GNB2L1</i>	Xy	Xy	Xyz
<i>GNL3</i>	Xy		Xz
<i>GPATCH4</i>	Xy		
<i>GTF2H1</i>		Xy	
<i>GTF2I</i>		Xy	
<i>GTF3C1</i>		X	
<i>GTF3C4</i>		X	
<i>GTPBP4</i>	Xy	Xy	Xyz
<i>H1F0</i>	Xy	Xyz	Xyz
<i>H1FX</i>	Xy	Xyz	Xyz
<i>H2AFV</i>		Xyz	Xyz
<i>H2AFY</i>		Xyz	Xyz
<i>HADHA</i>	Xy		Xy
<i>HDAC2</i>	Xy		Xy
<i>HDLBP</i>	X	Xy	
<i>HEATR1</i>	Xy		Xyz
<i>HIST1H1C</i>	Xy	Xyz	Xyz
<i>HIST1H1E</i>	X		X
<i>HIST1H2AA</i>			Xyz
<i>HIST1H2AB</i>	X		
<i>HIST1H2AC</i>		Xyz	
<i>HIST1H2AH</i>			Xz
<i>HIST1H2BL</i>	Xy		
<i>HIST1H3F</i>		Xz	Xyz
<i>HIST2H2AA4</i>		Xz	
<i>HIST2H2AB</i>		Xyz	
<i>HIST2H2BE</i>	Xy		
<i>HIST2H3D</i>		Xyz	
<i>HIST2H4B</i>		Xyz	Xy
<i>HMG20A</i>		Xy	
<i>HNRNPA0</i>	Xy	Xyz	Xyz
<i>HNRNPA1</i>	X		
<i>HNRNPA2B1</i>	Xy	Xyz	Xyz
<i>HNRNPA3</i>	Xy	Xyz	Xyz
<i>HNRNPAB</i>		Xyz	
<i>HNRNPC</i>	Xy	Xyz	Xyz
<i>HNRNPD</i>		Xyz	Xyz
<i>HNRNPF</i>	Xy	Xyz	Xyz
<i>HNRNPH1</i>	Xy	Xyz	Xyz
<i>HNRNPH2</i>	Xy	Xyz	Xyz
<i>HNRNPH3</i>	Xy	Xyz	Xyz
<i>HNRNPK</i>	Xy	Xyz	Xyz
<i>HNRNPL</i>	Xy	Xyz	Xyz
<i>HNRNPM</i>	Xy	Xyz	Xyz

.Gene	Complex		
	I	II	III
<i>HNRNPR</i>	Xy	Xyz	Xyz
<i>HNRNPU</i>	Xy	Xyz	Xyz
<i>HNRNPUL1</i>		Xyz	Xyz
<i>HNRNPUL2</i>	Xy	Xyz	Xyz
<i>HNRPDL</i>		Xy	
<i>HP1BP3</i>		Xyz	Xyz
<i>HSPA1A</i>		X	Xy
<i>HSPA5</i>	Xy	Xyz	Xyz
<i>HSPA8</i>	Xy	Xy	Xz
<i>HSPA9</i>	X		
<i>HSPB1</i>	X		
<i>HSPD1</i>	X	Xy	
<i>IFI16</i>		Xy	
<i>IGF2BP2</i>			X
<i>IK</i>		Xy	Xyz
<i>ILF2</i>	Xy	Xyz	Xyz
<i>ILF3</i>	Xy	Xyz	Xyz
<i>IMP3</i>		X	Xy
<i>IMP4</i>		Xy	Xy
<i>IMPDH2</i>		X	Xy
<i>IQGAP1</i>		Xy	
<i>ISG20L2</i>		X	
<i>JUP</i>		X	
<i>KARS</i>		Xz	
<i>KHDRBS1</i>		Xy	Xy
<i>KHSRP</i>		X	Xy
<i>KIAA0020</i>	Xy	Xyz	Xyz
<i>KIAA0174</i>			Xy
<i>KIAA1967</i>		Xz	
<i>KPNA2</i>		Xz	
<i>KPNA4</i>		Xyz	
<i>KPNA6</i>	X	X	
<i>KPNB1</i>		Xy	
<i>KRI1</i>		X	
<i>KRR1</i>		Xyz	Xyz
<i>KRT1</i>	Xy	Xyz	Xyz
<i>KRT10</i>	X	Xy	X
<i>KRT17</i>	Xy	Xy	Xy
<i>KRT18</i>		Xy	Xy
<i>KRT2</i>	Xy	Xyz	Xyz
<i>KRT7</i>		X	
<i>KRT8</i>	Xy		
<i>KRT9</i>		Xyz	Xy
<i>LAS1L</i>			Xy
<i>LEPRE1</i>		X	
<i>LGALS1</i>		X	
<i>LIG3</i>		X	
<i>LIMA1</i>		Xy	
<i>LIMCH1</i>		X	
<i>LLPH</i>			X
<i>LMNA</i>	Xy	Xyz	Xyz

.Gene	Complex		
	I	II	III
<i>LMNB1</i>	Xy	Xy	Xyz
<i>LMNB2</i>	Xy		Xy
<i>LMO7</i>		Xy	Xyz
<i>LOC100290337</i>		Xyz	
<i>LOC123397</i>			Xy
<i>LOC285984</i>		Xyz	Xyz
<i>LOC440926</i>		Xyz	
<i>LOC644914</i>		Xyz	
<i>LOC730732</i>	X	Xy	
<i>LRRC59</i>	Xy		Xy
<i>LRWD1</i>		X	
<i>LUC7L3</i>	X		
<i>LYAR</i>	Xy	Xy	Xyz
<i>MAGOHB</i>		Xy	Xyz
<i>MAK16</i>	Xy		Xy
<i>MATR3</i>	Xy	Xy	Xyz
<i>MCM2</i>		Xyz	Xyz
<i>MCM3</i>	X	Xyz	Xyz
<i>MCM3AP</i>		X	
<i>MCM4</i>		Xz	Xyz
<i>MCM5</i>	X	Xyz	Xyz
<i>MCM6</i>		Xyz	
<i>MCM7</i>		Xz	Xyz
<i>MDC1</i>		Xy	Xy
<i>METAP1</i>		Xy	
<i>MICALL1</i>			Xy
<i>MKI67</i>		Xz	
<i>MKI67IP</i>	Xy		Xyz
<i>MOGS</i>			Xy
<i>MORF4</i>		X	
<i>MOV10</i>		X	
<i>MPG</i>		Xy	
<i>MPHOSPH10</i>		Xy	Xy
<i>MPHOSPH8</i>		Xy	
<i>MRTO4</i>		Xy	Xyz
<i>MSI2</i>		Xz	
<i>MTDH</i>	X		
<i>MYBBP1A</i>	Xy		Xyz
<i>MYH10</i>			Xy
<i>MYH9</i>			Xy
<i>NAA38</i>			Xy
<i>NACA</i>	X	Xz	
<i>NAPIL1</i>		X	
<i>NAT10</i>	Xy	Xy	Xyz
<i>NCL</i>		Xyz	Xz
<i>NCOA5</i>			Xyz
<i>NDNL2</i>			X
<i>NEXN</i>		X	
<i>NFIC</i>		X	
<i>NGDN</i>	X	Xy	Xy
<i>NHP2</i>	Xy	Xyz	Xyz

.Gene	Complex		
	I	II	III
<i>NHP2L1</i>	X	Xyz	
<i>NIP7</i>	Xy	Xyz	Xyz
<i>NIPBL</i>	X		
<i>NLE1</i>	Xy		
<i>NMT1</i>			X
<i>NOC2L</i>	Xy		Xyz
<i>NOC3L</i>	Xy		Xy
<i>NOC4L</i>	X		
<i>NOL10</i>	Xy	Xy	Xyz
<i>NOL11</i>	Xy	Xy	Xy
<i>NOL6</i>	Xy	Xy	
<i>NOL7</i>			Xy
<i>NOL8</i>		X	
<i>NOL9</i>			Xyz
<i>NOLC1</i>	Xy	Xyz	Xyz
<i>NONO</i>		Xyz	Xyz
<i>NOP10</i>		X	Xy
<i>NOP14</i>		X	
<i>NOP16</i>			Xy
<i>NOP2</i>	Xy	Xyz	Xyz
<i>NOP56</i>	Xy	Xyz	Xyz
<i>NOP58</i>	Xy	Xyz	Xyz
<i>NPM1</i>	Xy	Xyz	Xyz
<i>NPM3</i>	X	Xyz	X
<i>NSA2</i>	Xy	X	Xy
<i>NSMCE1</i>			X
<i>NSUN5</i>	Xy	Xyz	Xyz
<i>NTHL1</i>		X	
<i>NUDT21</i>	Xy	Xz	
<i>NUP107</i>		X	
<i>NUP160</i>		X	
<i>NUP37</i>		X	
<i>NUP98</i>		Xy	
<i>NXF1</i>		Xyz	Xz
<i>OBFC2B</i>		X	
<i>OGDH</i>		Xy	
<i>PA2G4</i>			Xz
<i>PABPC1</i>		Xyz	Xyz
<i>PABPC4</i>		Xyz	
<i>PABPN1</i>	Xy	Xyz	Xyz
<i>PAK1IP1</i>	Xy	Xy	Xy
<i>PALM2-AKAP2</i>		X	
<i>PARP1</i>		Xyz	Xy
<i>PARP2</i>		Xz	
<i>PBRM1</i>		X	Xy
<i>PCBP1</i>		X	
<i>PCBP2</i>		Xyz	X
<i>PCID2</i>		X	
<i>PCNA</i>		X	
<i>PDCD11</i>	Xy	Xyz	Xyz
<i>PDCD6</i>		Xy	

.Gene	Complex		
	I	II	III
<i>PDS5B</i>		X	
<i>PELP1</i>		Xz	Xy
<i>PES1</i>	Xy		Xyz
<i>PHB2</i>	X		
<i>PHF14</i>		X	
<i>PHF2</i>		X	
<i>PHIP</i>		X	
<i>PIP</i>		X	
<i>PKP3</i>		X	
<i>PKP4</i>			X
<i>PLEC</i>		Xy	Xy
<i>PLOD3</i>		Xy	Xy
<i>PLRG1</i>	Xy		Xyz
<i>PMPCA</i>	X		
<i>PNN</i>		Xyz	Xyz
<i>PNO1</i>		Xy	Xy
<i>POGZ</i>		X	Xy
<i>POLG2</i>		X	
<i>POLR1B</i>	X	Xz	
<i>POLR1C</i>		Xz	Xyz
<i>POLR1E</i>	X		
<i>POLR2B</i>		Xyz	
<i>POLR2C</i>		Xy	
<i>POLR2D</i>		X	
<i>POLR2E</i>	X		
<i>POLR2G</i>		Xz	
<i>POLR2H</i>	Xy	Xyz	Xyz
<i>POLR2I</i>		Xy	X
<i>POP1</i>		Xy	
<i>POP5</i>		Xyz	
<i>PPAN</i>	Xy	Xyz	Xyz
<i>PPIG</i>			Xyz
<i>PPP1CB</i>		Xz	
<i>PPP1CC</i>	X	Xyz	Xyz
<i>PPP1R8</i>		X	
<i>PPP1R9A</i>		Xy	
<i>PPP1R9B</i>		Xy	
<i>PPP2R1A</i>		X	
<i>PRPF19</i>	Xy	Xyz	Xyz
<i>PRPF3</i>		Xyz	Xyz
<i>PRPF31</i>	Xy	Xy	Xz
<i>PRPF38A</i>			X
<i>PRPF38B</i>	Xy		
<i>PRPF4</i>		Xy	Xyz
<i>PRPF40A</i>		X	
<i>PRPF4B</i>		Xy	Xy
<i>PRPF8</i>	Xy	Xyz	Xyz
<i>PRPS1</i>		X	
<i>PRR3</i>			X
<i>PSAP</i>		X	
<i>PSIP1</i>		Xyz	

.Gene	Complex		
	I	II	III
<i>PSMA1</i>		X	
<i>PSMA2</i>		Xy	
<i>PSMA4</i>		X	
<i>PSMA5</i>		Xy	
<i>PSMA6</i>		Xy	
<i>PSMA7</i>	X	X	
<i>PSMB1</i>		Xy	
<i>PSMB2</i>	X	X	
<i>PSMB3</i>		Xy	
<i>PSMB4</i>	X	Xy	
<i>PSMB5</i>	X	X	
<i>PSMB7</i>		Xy	
<i>PSMC1</i>			Xy
<i>PSMC4</i>		X	
<i>PSPC1</i>			Xy
<i>PTBP1</i>	Xy	Xyz	Xyz
<i>PTBP2</i>	Xy	Xy	Xyz
<i>PUF60</i>	X	X	Xy
<i>PWP1</i>		X	
<i>PWP2</i>	Xy		Xyz
<i>PYCR2</i>	X		
<i>QARS</i>			Xz
<i>RAD21</i>			X
<i>RAD23B</i>	Xy		Xy
<i>RAI14</i>	X		
<i>RALY</i>		Xyz	Xz
<i>RALYL</i>		Xyz	X
<i>RAN</i>	Xy	Xyz	X
<i>RANBP2</i>		X	
<i>RANGAP1</i>		X	
<i>RBM10</i>		Xy	
<i>RBM12B</i>		Xyz	Xyz
<i>RBM14</i>	Xy	Xy	Xyz
<i>RBM15</i>		Xy	Xyz
<i>RBM19</i>		Xz	
<i>RBM22</i>	Xy	Xz	Xyz
<i>RBM25</i>		Xy	Xyz
<i>RBM28</i>	Xy	X	Xyz
<i>RBM3</i>		Xz	
<i>RBM34</i>	X		
<i>RBM39</i>	Xy	Xyz	Xy
<i>RBM4</i>		X	Xy
<i>RBM45</i>		X	
<i>RBM6</i>			X
<i>RBM8A</i>	Xy	Xy	Xz
<i>RBM9</i>	X		X
<i>RCC1</i>	Xy	Xyz	Xyz
<i>RCL1</i>		Xyz	Xy
<i>RIF1</i>		Xy	
<i>RNMT</i>		X	
<i>RNMTL1</i>	X		

.Gene	Complex		
	I	II	III
<i>RNPS1</i>	Xy	Xy	Xy
<i>RPA1</i>		Xy	
<i>RPF1</i>			Xy
<i>RPF2</i>	Xy	Xyz	Xyz
<i>RPL10A</i>	Xy	Xyz	Xyz
<i>RPL10L</i>			Xz
<i>RPL10P9</i>	X		
<i>RPL11</i>	Xy	Xyz	Xyz
<i>RPL12</i>	Xy	Xyz	Xyz
<i>RPL13</i>	Xy		Xz
<i>RPL13A</i>	Xy	Xyz	Xyz
<i>RPL14P1</i>	Xy	Xyz	Xyz
<i>RPL15</i>	Xy	X	
<i>RPL17</i>	Xy	Xyz	Xyz
<i>RPL18</i>	Xy	Xyz	Xyz
<i>RPL18A</i>	Xy	Xyz	Xyz
<i>RPL19</i>		Xz	Xz
<i>RPL21</i>	Xy		Xyz
<i>RPL22</i>		Xyz	Xyz
<i>RPL23</i>	Xy		Xyz
<i>RPL23A</i>	Xy	Xy	Xyz
<i>RPL27</i>		Xyz	Xyz
<i>RPL27A</i>		Xy	Xyz
<i>RPL28</i>	Xy	Xy	X
<i>RPL3</i>		Xyz	
<i>RPL30</i>	Xy		Xyz
<i>RPL31</i>		Xyz	Xy
<i>RPL32</i>		X	Xyz
<i>RPL34</i>			Xy
<i>RPL35</i>			Xyz
<i>RPL35A</i>			Xy
<i>RPL36</i>			X
<i>RPL37A</i>	Xy		Xz
<i>RPL38</i>			Xyz
<i>RPL4</i>	Xy	Xyz	Xyz
<i>RPL5</i>	Xy	Xyz	Xyz
<i>RPL6</i>	Xy	Xyz	Xyz
<i>RPL7</i>	Xy	Xyz	Xyz
<i>RPL7A</i>	Xy		Xyz
<i>RPL7L1</i>	Xy		Xyz
<i>RPL8</i>	Xy	Xyz	Xyz
<i>RPL9</i>	Xy	Xyz	Xyz
<i>RPLP0</i>	Xy	Xyz	Xyz
<i>RPLP1</i>	X	Xyz	
<i>RPLP2</i>		Xyz	
<i>RPN2</i>	Xy		
<i>RPP14</i>			Xz
<i>RPP30</i>	X	Xyz	
<i>RPP40</i>		Xy	
<i>RPRD1B</i>		Xy	
<i>RPS10</i>			Xyz

.Gene	Complex		
	I	II	III
<i>RPS11</i>	Xy	Xyz	Xyz
<i>RPS12</i>	X	Xy	Xy
<i>RPS13</i>		Xyz	Xyz
<i>RPS15</i>			Xy
<i>RPS15A</i>		Xz	
<i>RPS15AP25</i>			X
<i>RPS16</i>	Xy	Xyz	Xyz
<i>RPS17</i>	X	X	
<i>RPS18</i>			Xz
<i>RPS2</i>	X	Xyz	Xyz
<i>RPS23</i>	Xy		
<i>RPS24</i>		Xyz	
<i>RPS26P54</i>		X	X
<i>RPS27A</i>	Xy	Xyz	Xz
<i>RPS3</i>	Xy	Xyz	Xyz
<i>RPS3A</i>	Xy	Xyz	Xyz
<i>RPS4X</i>	Xy	Xyz	Xyz
<i>RPS5</i>	Xy	Xyz	Xyz
<i>RPS6</i>	Xy	Xyz	Xz
<i>RPS7</i>	Xy	Xyz	Xyz
<i>RPS8</i>	Xy	Xz	Xyz
<i>RPS9</i>	Xy	Xyz	Xyz
<i>RPSAP55</i>	Xy		X
<i>RRP1</i>		X	Xy
<i>RRP7A</i>		X	
<i>RRP9</i>	Xy	Xy	Xyz
<i>RRS1</i>		Xyz	Xyz
<i>RSF1</i>			Xyz
<i>RSL1D1</i>	Xy	Xyz	Xyz
<i>RSL24D1</i>			Xyz
<i>RUVBL1</i>	Xy	Xy	Xy
<i>RUVBL2</i>	Xy	Xy	Xy
<i>SAFB</i>		Xyz	
<i>SAFB2</i>		Xz	
<i>SAP18</i>		Xyz	Xy
<i>SARNP</i>		Xyz	
<i>SART1</i>		Xy	Xy
<i>SCIN</i>		Xy	Xy
<i>SCML2</i>		X	
<i>SEC13</i>		X	
<i>SEC61A2</i>	X		
<i>SEN3</i>		Xy	
<i>SF1</i>		X	
<i>SF3A1</i>	Xy	Xy	Xyz
<i>SF3A2</i>	Xy		Xyz
<i>SF3A3</i>	Xy	Xy	
<i>SF3B1</i>	Xy	Xz	Xz
<i>SF3B14</i>		X	
<i>SF3B2</i>	Xy	Xy	Xy
<i>SF3B3</i>	Xy	Xyz	Xyz
<i>SF3B4</i>	X	Xy	Xy

.Gene	Complex		
	I	II	III
<i>SFPQ</i>		Xyz	Xyz
<i>SFRS1</i>		X	X
<i>SFRS11</i>	X		
<i>SFRS12</i>		X	X
<i>SFRS13A</i>	X	X	X
<i>SFRS14</i>			X
<i>SFRS2IP</i>		X	
<i>SFRS3</i>		X	X
<i>SFRS4</i>		X	X
<i>SFRS5</i>		X	
<i>SFRS6</i>		X	
<i>SFRS7</i>	X	X	X
<i>SFRS9</i>		X	X
<i>SHMT2</i>		Xy	
<i>SIGMAR1</i>	X		
<i>SKIV2L2</i>		Xy	
<i>SLTM</i>	X	Xy	
<i>SMARCA1</i>		Xyz	Xz
<i>SMARCA4</i>	X		
<i>SMARCA5</i>		Xyz	
<i>SMARCB1</i>		Xz	
<i>SMARCC2</i>			Xy
<i>SMARCD1</i>	X		
<i>SMARCE1</i>	X		
<i>SMC1A</i>			Xyz
<i>SMC3</i>	Xy		Xy
<i>SMC6</i>	X		
<i>SMU1</i>	Xy	Xy	Xyz
<i>SND1</i>		Xy	
<i>SNRNP200</i>	Xy	Xyz	Xyz
<i>SNRNP35</i>			Xz
<i>SNRNP40</i>	Xy	Xy	Xyz
<i>SNRNP70</i>	Xy	Xyz	Xyz
<i>SNRPA</i>	Xy	Xy	Xy
<i>SNRPA1</i>	Xy	Xyz	Xyz
<i>SNRPB</i>		Xyz	Xyz
<i>SNRPB2</i>	Xy	X	
<i>SNRPD1</i>	Xy	Xyz	Xyz
<i>SNRPD2</i>	Xy	Xyz	Xyz
<i>SNRPD3</i>	Xy	Xyz	Xyz
<i>SNRPE</i>	Xy	Xyz	Xyz
<i>SNRPF</i>		X	
<i>SNRPG</i>			Xy
<i>SNW1</i>	Xy		Xy
<i>SON</i>		Xy	
<i>SORBS2</i>		Xy	Xyz
<i>SP3</i>		X	
<i>SPIN2A</i>		X	
<i>SPTAN1</i>	Xy	Xy	Xyz
<i>SPTBN1</i>	Xy		Xyz
<i>SRSF1</i>		Xyz	Xyz

.Gene	Complex		
	I	II	III
<i>SRSF11</i>	X		
<i>SRSF3</i>		Xyz	Xyz
<i>SRSF4</i>		Xyz	Xyz
<i>SRSF5</i>		Xyz	
<i>SRSF6</i>		Xy	
<i>SRSF7</i>	Xy	Xyz	Xyz
<i>SRSF9</i>		Xy	X
<i>SR140</i>	Xy	Xy	Xy
<i>SRP14</i>	Xy	Xyz	Xyz
<i>SRP9</i>		X	
<i>SRPK1</i>			Xy
<i>SRRM2</i>	X	X	
<i>SSB</i>			X
<i>SSR1</i>			Xy
<i>SSR4</i>		X	
<i>SSRP1</i>		Xyz	Xyz
<i>STAG2</i>		Xy	Xy
<i>STOM</i>	Xy	Xy	Xy
<i>STRAP</i>		X	
<i>STRBP</i>		Xyz	Xyz
<i>SUPT16H</i>		Xyz	Xy
<i>SUPT5H</i>		Xyz	Xy
<i>SUV39H1</i>		Xy	X
<i>SUV39H2</i>		X	
<i>SVIL</i>		X	
<i>SYNCRIP</i>	Xy	Xyz	Xyz
<i>TARDBP</i>		Xyz	
<i>TBL2</i>			X
<i>TBL3</i>	Xy		Xy
<i>TCF20</i>		X	
<i>TERF2</i>		X	
<i>TERF2IP</i>		X	
<i>TEX10</i>		Xyz	Xyz
<i>TFAM</i>		Xy	
<i>TFAP2A</i>		Xy	Xyz
<i>TFAP4</i>		X	
<i>THOC2</i>	X		X
<i>THOC4</i>	Xy	Xyz	Xyz
<i>THOC5</i>	Xy		Xyz
<i>THOC6</i>	X	X	Xyz
<i>THRAP3</i>		Xyz	Xyz
<i>TIAL1</i>		X	
<i>TJP1</i>		X	Xyz
<i>TMOD3</i>		Xy	Xy
<i>TMPO</i>		X	
<i>TOP1</i>	Xy	Xyz	Xy
<i>TOP2A</i>	Xy	Xyz	Xyz
<i>TOP2B</i>		Xyz	Xyz
<i>TPM1</i>		Xy	
<i>TPM4</i>		Xyz	
<i>TPR</i>	X		

.Gene	Complex		
	I	II	III
<i>TRA2A</i>	Xy	Xyz	Xyz
<i>TRA2B</i>		Xyz	Xyz
<i>TRIM28</i>		Xyz	
<i>TSPYL1</i>	X		
<i>TLL3</i>			X
<i>TUBB2C</i>			X
<i>U2AF1</i>	Xy		Xy
<i>U2AF2</i>	Xy	Xyz	Xyz
<i>UBR5</i>		X	
<i>UBTF</i>		Xy	
<i>UHRF1</i>		Xyz	Xy
<i>UPF1</i>		Xy	Xyz
<i>USF1</i>		Xz	
<i>USP7</i>		Xyz	
<i>UTP11L</i>			Xy
<i>UTP14A</i>	Xy		Xyz
<i>UTP15</i>	Xy	X	Xyz
<i>UTP18</i>		Xy	Xy
<i>UTP3</i>		Xy	Xy
<i>UTP6</i>			Xy
<i>VIM</i>	Xy	Xy	Xy
<i>VRK1</i>		Xy	X
<i>VTN</i>		Xyz	Xyz
<i>WBP11</i>			X
<i>WBP4</i>	X		
<i>WDR12</i>	Xy	Xy	Xyz
<i>WDR18</i>		Xy	Xy
<i>WDR3</i>	Xy	Xz	Xyz
<i>WDR33</i>			Xy
<i>WDR36</i>	Xy	Xy	Xyz
<i>WDR43</i>	Xy	Xy	Xz
<i>WDR46</i>	Xy		X
<i>WDR5</i>	Xy	Xy	X
<i>WDR61</i>		Xz	
<i>WDR74</i>		X	
<i>WDR75</i>	X		Xyz
<i>XAB2</i>			Xy
<i>XIRP2</i>			Xy
<i>XPC</i>		Xy	
<i>XRCC5</i>		Xyz	
<i>XRCC6</i>	X	Xyz	
<i>XRN2</i>	X	Xyz	Xyz
<i>YBX1</i>		Xz	Xy
<i>YLPM1</i>		X	
<i>ZC3HAV1</i>		X	
<i>ZNF326</i>		Xz	
<i>ZNF384</i>		Xy	
<i>ZNF828</i>		X	
<i>ZNRD1</i>		X	

Supplementary Table 2. Relative amounts of the ten most abundant proteins in complexes I-III.

Relative amounts were determined using the normalized spectral index (SI) method; this extracts quantitative information from peptide/spectral counts and fragment-ion intensities so that the SI roughly reflects the percentage of a protein in a complex¹⁴. Notes: (i) Different proteins are detected by mass spectrometry with different efficiencies, and this probably underlies why the core histones are not present at equal percentages. (ii) Lamin A/C plays a role in transcription³⁶, and 5%, 0.2%, and <0.01% of different lamin proteins (lamins B1, B2, A/C) were present in complexes I, II, and III, respectively.

Protein	Gene	Description	SI (%)
Complex I			
IPI00221394	<i>DKC1</i>	dyskerin, snoRNP	14
IPI00005614	<i>SPTBN1</i>	spectrin β , actin binding	13
IPI00013881	<i>HNRNPH1</i>	hnRNP H	10
IPI00844215	<i>SPTAN1</i>	spectrin α , actin binding	8
IPI00021405	<i>LMNA</i>	lamin A/C	5
IPI00031691	<i>RPL9</i>	60S ribosomal protein L9	5
IPI00216049	<i>HNRNPK</i>	hnRNP K	4
IPI00004968	<i>PRPF19</i>	pre-mRNA processing factor 19	3
IPI00217862	<i>RRP9</i>	U3 sno RNA-interacting protein 2	2
IPI00179964	<i>PTBPI</i>	polypyrimidine tract-binding protein 1	2
Complex II			
IPI00453473	<i>HIST1H4L</i>	histone H4	12
IPI00549248	<i>NPM1</i>	nucleophosmin	12
IPI00216456	<i>HIST1H2AC</i>	histone H2A type 1-C	11
IPI00302850	<i>SNRPD1</i>	snRNP Sm D1	7
IPI00396378	<i>HNRNPA2B1</i>	hnRNP A2/B1	5
IPI00465070	<i>HIST1H3G</i>	histone H3.1	4
IPI00003377	<i>SFRS7</i>	splicing factor, arg/ser-rich 7	2
IPI00025039	<i>FBL</i>	fibrillarin, snRNP	2
IPI00217465	<i>HIST1H1C</i>	histone H1.2	2
IPI00418471	<i>VIM</i>	vimentin, intermediate filament	2
Complex III			
IPI00081836	<i>HIST1H2AM</i>	histone H2A type 1	44
IPI00013508	<i>ACTN1</i>	alpha-actinin-1	4
IPI00025039	<i>FBL</i>	fibrillarin, snRNP	4
IPI00302850	<i>SNRPD1</i>	snRNP Sm D1	2
IPI00453473	<i>HIST1H4L</i>	histone H4	2
IPI00181728	<i>BRIX1</i>	Brix domain-containing protein 2	2
IPI00221089	<i>RPS13</i>	40S ribosomal protein S13	2
IPI00003918	<i>RPL4</i>	60S ribosomal protein L4	2
IPI00418471	<i>VIM</i>	vimentin, intermediate filament	2
IPI00940685	<i>SNRNP40</i>	U5 snRNP	1

Supplementary Table 3. Sequences of primers used for native 3C.

Name	Sequence	Description
<i>PTRF</i> -R	GTAGAGACAGAAAGGTGGGTCAGC	tss/intron 1
<i>SAMD4A</i> tss	TCACGTAGAGTCTGGATTTTCTGG	tss/intron 1
<i>SAMD4A</i> 3' end	AGAAACGCTCTGTCCAGTAAGTCC	intron 11
<i>GMFB</i> -F	GGCAGTTGGAAACCTTTCGAC	tss
<i>NFkBIA</i> -F	AGTAGTGGCCTCCCCATCC	intron 5
<i>45S</i> -F1	GCAATTATCCCCATGAACGAG	rDNA repeat
<i>45S</i> -F2	TATTCCCTTCCTGGAGTTGGAG	rDNA repeat
<i>7SK</i> -F	CCTCAAACAAGCTCTCAAGG	<i>7SK</i> 3' end
<i>PTRF</i> (loading-F)	AAGGATCTGAGTGGGGAGGTG	intron 1
<i>PTRF</i> (loading-R)	ATCTACCAGGTGAGCCACAG	intron 1