UNIVERSITY OF LEEDS

This is a repository copy of *Isolation of the protein and RNA content of active sites of transcription from mammalian cells.*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/96057/

Version: Accepted Version

Article:
Melnik, S, Caudron-Herger, M, Brant, L et al. (4 more authors) (2016) Isolation of the protein and RNA content of active sites of transcription from mammalian cells. Nature Protocols, 11 (3). pp. 553-565. ISSN 1754-2189

https://doi.org/10.1038/nprot.2016.032

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Isolation of the protein and RNA content of active sites of transcription in mammalian cells

Svitlana Melnik[1,5,§], Maïwen Caudron-Herger[2,5], Lilija Brant[3], Ian M. Carr[4], Karsten Rippe[2], Peter R. Cook[1] & Argyris Papantonis[3]

[1]Sir William Dunn School of Pathology, University of Oxford, OX1 3RE Oxford, United Kingdom.
[2]German Cancer Research Center (DKFZ) & BioQuant, D-69120 Heidelberg, Germany.
[3]Center for Molecular Medicine, University of Cologne, D-50931 Cologne, Germany.
[4]Institute of Biomedical and Clinical Sciences, University of Leeds, LS9 7TF Leeds, United Kingdom.
[5]These authors contributed equally to this work.
[§]Present Address: DNA Vectors, German Cancer Research Center (DKFZ), D-69120 Heidelberg, Germany.

Correspondence should be addressed to A.P. (argyris.papantonis@uni-koeln.de). Tel.: +49 221 478 96987; Fax: +49 221 478 4855.

Subject terms: biochemistry; nucleic acids-based molecular biology; transcriptomics; mass spectrometry; next-generation sequencing

Keywords: transcription factory; proteome; nascent RNA; RNA polymerase; eRNAs

---

Associated publications:

Melnik S, Deng B, Papantonis A, Baboo S, Carr IM & Cook PR. (2011) The proteomes of transcription factories containing polymerases I, II, or III. *Nat Methods* **8**: 963-968.

Caudron-Herger M, Cook PR, Rippe K & Papantonis A. (2015) Dissecting the nascent human transcriptome by analysing the RNA content of transcription factories. *Nucleic Acids Res* **43**: e95.

ABSTRACT

Mammalian nuclei contain three RNA polymerases (I, II, and III) transcribing different gene subsets, the active forms of which are contained in supramolecular complexes known as "transcription factories". These are difficult to isolate because they are embedded in the 3D structure of the nucleus. Factories exchange components with the soluble nucleoplasmic pool over time as gene expression programs change during development or disease. Analyzing their content can provide information on the nascent transcriptome and its regulators. Here, we describe a protocol for the isolation of large factory fragments under isotonic salt concentrations that requires <72 hours. It relies on chromatin detachment by DNase I from the nuclear substructure of freshly-isolated, unfixed, cells, followed by caspase treatment to release multi-megadalton factory complexes. These retain transcriptional activity, and their contents can be analyzed by mass-spectrometry or RNA sequencing to catalogue the proteins and RNA associated with sites of active transcription.

INTRODUCTION

Eukaryotic nuclei contain three RNA polymerases (I, II, III)[1,2], which are part of larger complexes; for example, RNA polymerase II is found with enzymatic activities involved in initiation, capping, splicing, and poly-adenylation[3,4]. These complexes may be organized into even larger structures that harbor most of the machinery required for transcript production known as "transcription factories"[5,6]. These are typically defined as sites containing at least two different active transcription units (simply to distinguish them from the case where two polymerizing complexes are active on one template), and harbor more than 95% of total transcriptional activity in a mammalian cell[5]. The biochemical isolation of such factories would allow dissection of the protein content of the active sites of transcription, and most nascent transcripts being produced at any given moment. However, their isolation has proved challenging due to their large mass and tight association with nuclear substructures. Furthermore, the failure to isolate transcription factories had raised concerns as to whether factories actually exist[7]. A few years ago we addressed these issues by introducing an approach for purifying large fragments of factories with an apparent mass of >8 MDa (the size of the largest protein marker commercially available), and went on to analyze the proteomes of different complexes containing either polymerase I, or II, or III[8]. More recently we adapted this method to catalogue their RNA content revealing a strong enrichment for nascent transcripts[9]. Here, we present a protocol for a unified workflow to purify protein and/or RNA from transcription factories based on the combination of both procedures (**Fig. 1**).

Overview of the method and its applications

There is a long-standing research interest in complexes involved in the regulation of gene expression, and a growing volume of literature points to transcription factories as important contributors in this (reviewed in ref. 5). Until recently, neither the protein composition of factories nor their primary

products – the nascent, factory-associated, transcripts – could be directly interrogated.

Two factors make purification of mammalian polymerases engaged on endogenous templates difficult. First, active enzymes represent only a quarter of the total; the remaining majority form a rapidly-diffusing soluble pool that tends to aggregate in non-isotonic buffers[8,10,11]. Therefore, we use isotonic conditions whilst removing the inactive fraction. Second, engaged polymerases plus their templates and transcripts are housed in factories bound to the underlying nuclear scaffold[8,10-12]. As caspases deconstruct nuclei during apoptosis, we reasoned they might be used to also release factories. In brief, the method relies on the isolation of nuclei from living (unfixed) cells in close-to-physiological salt concentrations. Then, the most chromatin is detached by DNase I, and factories are disengaged from the nuclear substructure using a mixture of group-III caspases. After pelleting nuclear debris, large fragments of the factories (>8 MDa) are found in the supernatant. Once the RNA and/or the proteins in this supernatant have been purified (see **Fig. 1**), high-throughput sequencing of RNA and/or mass-spectrometry can be applied to comprehensively catalogue factory content[8,9].

Overall, the current protocol is reproducible, straightforward to implement, and requires limited hands-on time. The whole process concludes in less than 72 hours (not including the time needed for mass-spectrometry or next-generation sequencing) and is particularly suited to dissect cellular responses to extra-cellular cues. For example, one could monitor the changes in the protein and RNA content of the active transcription compartment when transcription is reprogrammed by different cytokines.

Limitations of the method and comparison to existing approaches

Although the content of mature and nascent proteins in large subcellular complexes like nucleoli have been analyzed[13,14], factory-associated proteins have been difficult to interrogate. While our method facilitates proteomic analysis of isolated factories (**Fig. 1**), it still requires large numbers of cells (in the order of ~$5x10^7$), especially if the three polymerizing activities are to be separated by 2D electrophoresis[8]. In addition, caspase treatment inevitably hinder the detection of those proteins that are most targeted by the enzymes used.

Similarly, there exist many methods for isolating and characterizing newly-synthesized RNA, for example "GRO-seq"[15], "mNET-seq"[16,17], "chromatin RNA-seq"[18], "poly(A)-depleted RNA-seq"[19], "nascent-seq"[20], and the metabolic tagging of newly-made RNA using 4-thiouridine[21,22]. Each of these has its own particular shortcomings, and all are relatively laborious and/or require a high sequencing depth; most also focus on particular parts of the transcriptome (for a comparison of RNA-seq methods, see **Table 1**). Hence, cataloguing nascent mammalian transcriptomes is challenging (in contrast to the now-routine mapping of mRNAs) due to the short half-lives, low abundance, and variety of processing pathways. Nonetheless, such cataloguing can be highly informative, especially when analyzing short-term responses to stimuli, and the associated changes in messenger, non-

coding, and enhancer RNAs (eRNAs[23,24]). Analysis of factory-associated transcripts offers a simple, low-cost, way for assessing transcriptional changes at the level of nascent RNA. However, the current approach uses freshly-isolated cells, and analogous methods for fixed cells have yet to be developed. Nevertheless, eRNAs (and other short-lived transcripts) can be detected with high sensitivity (see ref. 9 for an example) and most chromatin-associated RNA can be studied as a separate fraction, despite factory- and chromatin-associated fractions inevitably sharing components.

Experimental design

As already noted, the isolation of transcription factories is performed on living cells collected in a "physiological" buffer. To date, we have successfully implemented this protocol on human, mouse, and Chinese hamster cells, either pluri-/multi-potent or differentiated; the list includes HeLa, HUVECs, IMR-90s, human induced pluripotent stem cells and cardiac progenitors, mouse E14 stem cells and embryonic fibroblasts, and Chinese hamster ovary cells. The general workflow for the isolation of factories relies on efficient isolation of nuclei that should be optimized for each cell type. Although, labeling nascent RNA is not required, this may be included (see **Fig. 2**), especially when setting up the technique to quantitate retention of transcriptional activity and nascent RNA (**Box 1**). Then, cells are permeabilized using saponin, polymerases allowed to extend their transcripts (by "running-on") by <40 nucleotides in a tagged precursor (e.g., Br-UTP or [^{32}P]UTP), and the amount of label incorporated into RNA measured (e.g., using anti-BrU antisera and immunofluorescence, or autoradiography/scintillation counting[8]). After the run-on, nuclei are isolated and chromatin is digested with DNase I to release most chromatin into the supernatant. Nuclei are still largely intact at this stage. They are now resuspended in a "native lysis buffer" and digested with caspases 6, 8, 9, and 10 (core subunits of the three RNA polymerases lack sites recognized by these caspases, with the exception of RPB9). After stopping the reaction, the sample can be retreated with DNase I (to trim off as much chromatin as possible). After repelleting, the resulting supernatant contains large (>8-MDa) fragments of factories, and its contents can be analyzed in various ways to fit the needs of each end-user.

For proteomic studies, released complexes can be resolved electrophoretically using a "native" (preparative) two-dimensional acrylamide-agarose composite gel; the three polymerizing activities are separated into partially-overlapping regions of the gel. To trace the positions of the complexes containing the different RNA polymerases and verify separation, additional small aliquots of the same sample from the same preparation are separated in a set of identical (analytical) mini-gels. These can be used for immuno-detection of one or other RNA polymerase, of other proteins of interest, and controls. For example, Western blots can be used to assess cross-contamination between the three complexes, or from other cellular compartments using proteins typically not found in factories like MacroH2A.2 and Hsp70. If nascent RNA has been labeled, autoradiography can

be used to detect the associated [^{32}P]RNA (**Box 2**). Next, preparative gels are stained with SYBR Green to reveal the boundaries of regions rich in RNA, and wanted regions in the preparative gel excised using the corresponding regions in the mini-gels as a guide. Finally, proteins in the excised regions are analyzed by mass-spectrometry. If separation of the factory fraction into the three different RNA polymerizing complexes is not required, the user can omit this step, and proceed directly to proteome analysis. This largely alleviates a major limitation of this method: the need for large numbers of cells (typically >5x10^7). Consequently, we envisage the use of quantitative SILAC-based[25,26] approaches becoming far more cost-effective, while also offering a means for additional controls (quantitatively assessing contaminants between replicas/conditions, etc.).

For transcriptome analysis, RNA is isolated from the factory-containing supernatant using Trizol, and cDNA libraries prepared for next-generation sequencing (in our case, on an Illumina platform; **Box 3**). For such an analysis, significantly lower cell counts can be used, and robust coverage of mRNAs and eRNAs can be achieved using as few as 0.5x10^6 cells (M.C.-H., K.R. and A.P., *unpublished data*). A sequencing depth of ~35 million reads suffices to catalogue comprehensively factory-associated RNAs, including long non-coding (lncRNAs) and eRNAs from ~10 million cells[9]. It is noteworthy that in the course of isolating factories, additional fractions can easily be retained at the appropriate stages to provide “total cell”, “cytoplasmic”, “nuclear”, and “chromatin-associated” ones (**Fig. 1**); subsequent comparison of their contents can provide additional insight into the organization of mammalian transcription along the genome.

MATERIALS

General reagents/chemicals

- [^{32}P]UTP (3,000 Ci/mmol; Perkin Elmer, cat no. BLU507H250UC)
- [methyl-^{3}H]thymidine (50 Ci/mmol; Perkin Elmer, cat. no. NET027W)
- 30% acrylamide/bis-acrylamide solution (37.5:1; Bio-Rad, cat. no. 1610158)
- 5-Bromouridine 5'-triphosphate sodium salt (5-BrUTP; Sigma-Aldrich, cat. no. B7166)
- 6-aminocaproic acid (Sigma-Aldrich, cat. no. A7824)
- Acetic acid (Sigma-Aldrich, cat. no. A6283)
- Ammonium persulphate (APS; Sigma-Aldrich, cat. no. A3678)
- ATP disodium salt (Sigma-Aldrich, cat. no. A2383)
- Blotto non-fat milk (Santa Cruz Biotech., cat. no. sc2325)
- Blue carrier immunogenic protein (Thermo Scientific, cat. no. PI77661)
- Bromophenol blue (Sigma-Aldrich, cat. no. 32712)
- $CaCl_2 \cdot 2H_2O$ (Sigma-Aldrich, cat. no. C3306)
- Coomassie Blue G-250 (Serva, cat. no.)
- Dithiothreitol (DTT; Sigma-Aldrich, cat. no. D9779)
- EDTA disodium salt (Sigma-Aldrich, cat. no. 4884)
- Glycine (Sigma-Aldrich, cat. no. G8898)
- Hyperfilm MP (GE Healthcare, cat. no. 28-9068-42)
- iBlot™ nitrocellulose transfer stacks, regular size (Life Technologies, cat. no. IB23001)
- Igepal (NP-40; Sigma-Aldrich, cat. no. I8896)
- Imperial protein stain (Thermo Scientific, cat. no. PI24615)
- KCl (Sigma-Aldrich, cat. no. P9541)
- $MgCl_2 \cdot 6H_2O$ (Sigma-Aldrich, cat. no. M2670)
- Na_2HPO_4 (Sigma-Aldrich, cat. no. S3264)
- Na_3VO_4 (Sigma-Aldrich, cat. no. S6508)
- NaF (Sigma-Aldrich, cat. no. 30105)
- NTPs (100 mM each; Roche, cat. no. 11277057001)
- Ponceau S (Sigma-Aldrich, cat. no. P3504)
- Potassium acetate (Sigma-Aldrich, cat. no. P1190)
- Protein-Free (TBS) blocking buffer (Thermo Scientific, cat. no. PI37570)
- RestorePlus™ Western Blot Stripping buffer (Thermo Scientific, cat. no. PI46428)
- Saponin (Sigma-Aldrich, cat. no. 47036)
- SeaKem® Gold Agarose (Lonza, cat. no. 50152)
- Sodium dodecyl-sulphate (SDS; Sigma-Aldrich, cat. no. L3771)
- Sucrose (Sigma-Aldrich, cat. no. S0389)
- SuperSignal™ West Pico ECL detection kit (Pierce, cat. no. PI34081)
- SYBR Green II nucleic acids gel stain (Life Technologies, cat. no. S33102)
- Tetra-methyl-ethylenediamine (TEMED; Sigma-Aldrich, cat. no. T22500)
- Tris (Sigma-Aldrich, cat. no. T1503)
- Triton X-100 (Sigma-Aldrich, cat. no. T8787)
- Trizol® LS (Life Technologies, cat. no. 10296-010)
  ! CAUTION: Trizol-based solutions are toxic and must be handled and disposed of according to established national and institutional regulations.
- Tween-20 (Sigma-Aldrich, cat. no. P9416)
- Whatman® filter paper, Grade 1 (Sigma-Aldrich, cat. no. Z240087)
- Xylene cyanol (Sigma-Aldrich, cat. no. X4126)
- β-glycerophosphate (Sigma-Aldrich, cat. no. G9422)
- β-mercaptoethanol (Sigma-Aldrich, cat. no. M3148)

Enzymes/inhibitors:

- Active human caspases, Group III (6-,8-,9-,10) mix (BioCat, cat. no. K243-4-25-BV)
- Complete protease inhibitor cocktail, EDTA-free (PIC; Roche, cat. no. 6538282001)
- DNase I (Worthington, cat. no. LS006331)
  Δ CRITICAL: If another brand is used, check for enzyme activity in PB.
- Group-III caspase inhibitor (Calbiochem, cat. no. 2188745)
- Proteinase K (Roche, cat. no. 3115828001)
- RNase inhibitor (RiboLock; Thermo Scientific, cat. no. E00384)

Antibodies:

- Chicken polyclonal anti-RPC62 (Abcam, cat. no. ab26185; used at 1/1000 dilution)
- Mouse monoclonal anti-RPA194 (Santa-Cruz Biotech., cat. no. sc-48385; used at 1/100 dilution)
- Mouse monoclonal anti-RPB1 (7C2[27], used at 1/10,000 dilution).

EQUIPMENT

- Cooling microcentrifuge (Hermel, cat. no. Z216MK)
- iBlot™ dry blotting system (Life Technologies, cat. no. IB1001)
- Mini-Protean3® electrophoresis cell (Bio-Rad, cat.no. 1658000)
- Nanodrop spectrophotometer (LabTech, ND-1000)
- Protean II xi® electrophoresis cell (Bio-Rad, cat. no. 1651801)
- Thermomixer (HLC, cat. no. DITA15001)

REAGENT SETUP

Δ CRITICAL STEP: Prepare all solutions using Milli-Q H_2O (Millipore) or DEPC-treated water in glassware decontaminated by baking (>180°C, overnight) or in nuclease-free plasticware.

Complete protease inhibitor, EDTA-free (PIC). Dissolve one tablet in 1 ml of PBS to create a 50× working solution. Store the solution at −20°C for up to 3 months; avoid repeated freeze-thaw cycles.

Complete physiological buffer (PB; adapted from Jackson *et al.*[10]). Basal PB (100 mM potassium acetate, 30 mM KCl, 10 mM Na_2HPO_4, 1 mM $MgCl_2$) is supplemented just before each experiment with 1 mM Na_2ATP, 1 mM dithiothreitol, 25 units/ml RNaseOUT, 10 mM β-glycerophosphate, 10 mM NaF, 0.2 mM Na_3VO_4, and a 1/1000 dilution of PIC. Always use PB ice-cold. Basal PB (which does not contain ATP) can be stored at 4°C for up to 6 months. **Δ CRITICAL STEP:** As the acidity of each ATP batch varies, 100 mM KH_2PO_4 is used to adjust the pH to 7.4 with the highest possible precision.

Native lysis buffer, pH 7.4 (NLB; modified from Novakova *et al.*[28]). 40 mM Tris-acetate, 2 M 6-aminocaproic acid, 7% sucrose, 1/1000 dilution of PIC, and 50 units/ml RNaseOUT. Can be stored at −20°C for up to 6 months; PIC and RNaseOUT are always added immediately before use.

"Run-on" mix: PB plus 100 µM ATP, 100 µM CTP, 100 µM GTP, 0.1 µM UTP, 50 µCi/ml [^{32}P]UTP or 80 µM 5-bromo-UTP plus $MgCl_2$ added to give a concentration of Mg^{2+} ions equimolar to that of all triphosphates. Always prepare this mix immediately before use.

5x Composite gel buffer (modified from Nadano *et al.*[29]): 200 mM Tris-acetate (pH 7.4), 35% sucrose, 0.05% Triton X-100. Stored at 4°C for a maximum of 6 months.

10x Composite gel running buffer: 400 mM Tris-acetate (pH 7.4).

Coomassie solution: 500 mM 6-aminocaproic acid, 2% (w/v) Coomassie Blue G-250. Stored at 4°C for a maximum of 6 months.

TBS-T buffer: 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.05% Tween-20.

10x Tris-glycine running buffer: 250 mM Tris (pH 8.3), 1.92 M glycine, 1% SDS. Glycine stocks should be stored at 4°C for a maximum of 6 months.

Starting material: The protocol is implemented on living cells, either grown in suspension or as a monolayer. Cells grown in suspension ensure fewer losses as they do not need to be scraped or trypsinized off the surface. We typically start with $5x10^7$ cells for proteomic analyses, and 4-$8x10^6$ cells for nascent RNA sequencing (see **Fig. 3** for some data on use of lower cell counts).

PROCEDURE

Preparation of cells ● TIMING: 1 h

1| The protocol is implemented on living cells grown in suspension or as a monolayer. If DNA recovery needs to be monitored, scintillation counting can be used; add 0.25 µCi/ml [methyl-^{3}H]thymidine to the culture medium a day before to achieve uniform labeling by incorporation into cellular DNA[8]. Harvest cells according to either option A or B.

A For cells grown as a monolayer on 15-cm culture dishes (e.g., HUVECs[9]):

i Split cells the day before the experiment. Harvest them when they reach 75-85% confluence.

ii Aspirate medium from plates, and immediately place them on ice.

iii Add 2.5 ml of ice-cold PB per plate. Gently scrape cells and transfer to a 50-ml falcon tube. **Δ CRITICAL STEP:** Soft rubber scrapers are strongly recommended.

iv Collect cells by centrifugation at 400 x *g* for 5 min at 4°C; keep pellets on ice.

B For cells grown in suspension (e.g., HeLa in spinner culture[8]):

i Collect cells by centrifugation at 400 x *g* for 5 min at 4°C.

ii Remove medium, immediately place cells on ice.

Note: We have verified that the immediate transfer of cells to ice "freezes" polymerizing activity during the response to the inflammatory cytokine, TNFα, at the desired time-point (see ref. 9).

Isolation of nuclei and chromatin digestion ● TIMING: 1-1.5 h

2| Resuspend cells in 5 ml of ice-cold PB+0.4% NP-40, incubate 20 min on ice.

Δ CRITICAL STEP: Different cell types respond variably to this treatment. When releasing HUVEC nuclei, for example, 2x 10-min washes with PB+0.4% NP-40 are optimal. Assess the fraction of lysed cells by placing a drop of the cell suspension after each wash on a haemocytometer and examining under a light microscope.

3| Pellet nuclei at 400 x *g* for 5 min at 4°C, and remove the supernatant.

4| Wash briefly once more in PB+0.4% NP-40, and pellet as above.

5| Gently resuspend pellet in PB+0.4% NP-40 supplemented with 0.5 mM $CaCl_2$ (100 µl/10^7 cells), and transfer to a 2-ml tube.

Δ CRITICAL STEP: From this step onwards use 2-ml round-bottom (nuclease-free) tubes.

6| Add DNase I (10 units/10^7 cells) and incubate for 30 min at 33°C.

Δ CRITICAL STEP: No shaking at this step.

? TROUBLESHOOTING (see **Table 2**)

7| Stop the reaction by adding EDTA to 2.5 mM and cooling on ice. A small aliquot (depending on the starting amount of cells) can be kept here and used to isolate DNA, and verify the extent of DNA fragmentation by DNase I (a smear around ~200 bp is expected; see **Fig. 4a**).

? TROUBLESHOOTING

8| Spin digested nuclei at 600 x *g* for 5 min at 4°C, keep both the supernatant and pellet separately. Pelleted nuclei are used in subsequent steps; the supernatant contains "chromatin-associated" RNA and can be directly mixed with 0.75-1 ml Trizol and purified (see **Box 3**) or stored at -80°C.

Caspase digestion and factory release ● TIMING: 1-1.5 h

9| Resuspend nuclei in NLB (50 µl/10^7 cells, but larger volumes may be used to allow efficient resuspension). Incubate on ice for 20 min.

Δ CRITICAL STEP: Vortex pellet vigorously at this step to disaggregate clusters of nuclei. Always use the smallest possible volume of NLB.

10| Prepare the mixture of caspases as follows. Allow the vials containing lyophilized caspase 6, 8, 9, and 10 to thaw on ice and resuspend in PB+0.4% NP40 to achieve a final concentration of 2 unit of each caspase/µl of the buffer. Snap-freeze small aliquots of the caspases mix in liquid nitrogen and store at -80°C. For each experiment always use a freshly-thawed aliquot.

11| Add the caspases to the sample (2 units of caspases mix/10^7 nuclei) and incubate at 33°C for 30 min in a thermomixer with vigorous (~900 rpm) shaking.

12| Transfer on ice and stop the reaction by adding group-III caspase inhibitor to a final concentration of 0.2 mM (we typically use a 20 mM stock which is stored at -20°C).

13| Following a brief incubation on ice, spin at 600 x *g* for 5 min at 4°C.

14| Collect supernatant and treat it with DNase I as before (steps **7-8**); this is the fraction that contains large fragments of factories. The quality of the large fragments can be assessed by performing 2D native electrophoresis in (small) analytical gels (**Box 2**). From this point onwards proceed according to either option **A** or **B**.

A For analysis of factory-associated RNA:

i Directly mix the supernatant with 0.75-1 ml Trizol and allow the sample to sit at room temperature (22°C) for 5 min.

ii Now RNA is purified directly or the Trizol mix is stored at -80°C. Once RNA is purified, strand-specific cDNA libraries may be prepared as described (**Box 3**).

B For proteomic analysis proceed with steps **16-26**.

PAUSE POINT: Samples can be aliquoted, snap-frozen in liquid nitrogen, and stored at -80°C.

? TROUBLESHOOTING

Native 2D electrophoresis, preparative ● TIMING: 36-48 h

Conditions for electrophoresis in a native 2D gel are modified from those used previously[28,29].

15| Cast 1st dimension gel (0.7% agarose, 1.5% acrylamide in 40 mM Tris-acetate pH 7.4, 7% sucrose, 0.01% Triton X-100) for example using the Protean II® xi system (gel dimensions of 16 x 20 x 0.15 cm) using 1.5 mm a 5-well wide-tooth comb (Bio-Rad). Assemble the casting plates and measure the exact amount of water needed for one such gel, and prepare as follows (example for a 10-ml gel is given; scale amounts up proportionately for larger gel volumes):

Component	Amount	Final concentration
Agarose (SeaKem® Gold)	0.07 g (in 7.425 ml water)	0.7%
Acrylamide (30%; 37.5:1)	0.5 ml	1.5%
Composite gel buffer (5x)	2.0 ml	1x
TEMED (100% w/v)	15 µl	0.15%
APS (25% w/v)	60 µl	0.15%
Total	10 ml	

Δ CRITICAL STEP: Weigh the 0.07 g of agarose in a glass beaker, add 7.425 ml of Milli-Q water and weigh the beaker before melting the agarose in a microwave oven. After slowly bringing the gel to boiling, weigh the beaker again and add back the amount of water that evaporated. Also, preheat the casting plates to 37°C before pouring the gel. Pour the gel and allow it to polymerize at room temperature for ≥1 hour before the run.

16| Load the preparative gels with a sample equivalent to that isolated from 4-5x10^7 cells (not radioactively labeled). Bromophenol blue (0.004% w/v) and xylene cyanol (0.004% w/v) are added to the sample to monitor migration. Then, for separating/isolating complexes containing RNA polymerases II and III, add Coomassie blue solution to the sample (1/1000 dilution) to be loaded onto the gel, incubate sample for 30 min on ice, and add Coomassie solution into the 1x cathode buffer at a 1/100 dilution. For the separation/isolation of complexes containing RNA polymerases I (nucleolar factories), do not add Coomassie (as it affects migration of the RNAPI complex); hence, proceed directly to sample loading.

Note: It is also possible to proceed directly to mass-spectrometry of the supernatant isolated in step **14**, without separating the different RNA polymerase-containing complexes by 2D electrophoresis. Such an approach can reduce the starting amount of cells required for analysis, and it is the end-user's decision which variation of the protocol best fits the question in hand.

17| Load the sample onto the 1st dimension gel. Use Blue carrier protein (200 ng in NLB with 1/1000 Coomassie solution) as a protein marker (8 MDa).

18| Run the gel at a constant voltage of 100 V in 1x composite running buffer until the sample has fully entered the gel. Then reduce to 40 V and electrophorese overnight in a cold room. At the end of the run, xylene cyanol has migrated ~¾ of the length of the gel and bromophenol blue is lost.

19| Following electrophoresis, cut a sample-containing strip out of the gel using a clean scalpel.

20|Cast 2nd dimension gel using 2 mm spacers; no comb is used for the 2nd dimension. The composition and running conditions of the 2nd dimension gel are the same as for the 1st dimension.

21| Polymerize the sample-containing strip to the 2nd dimension. For this, prepare a small amount of composite gel and pour it over the sample strip as it placed on top of the 2nd-dimension gel. This takes place with the gel placed horizontally; make sure all gaps between the strip and the already-polymerized gel are filled.
Δ CRITICAL STEP: Avoid air bubbles between the sample strip and the 2nd-dimension gel. To run Blue carrier protein as a marker, use a short 2-mm thick comb next to the sample strip to cast a single well. Cover the top of the gel with Milli-Q water and let it cool for ≥1 hour.

22| Run the 2nd-dimension gel at 40 V for 8-12 hours in a cold room.

23| Following electrophoresis, locate the separated complexes by staining the gel with SYBR Green II nucleic acids gel stain for 20-30 min at room temperature while rocking gently. The relative positions of the different RNA polymerase complexes can be determined by reference to those in the analytical gels (**Box 2**).
? TROUBLESHOOTING

Preparation of samples for mass-spectrometry ● TIMING: 1-2 h
24| Cast a standard 7% SDS-polyacrylamide gel (gel dimensions of 8.6 x 6.7 x 0.2 cm) and wide-tooth comb with 5 wells (Bio-Rad).

25| Equilibrate each excised gel piece from the 2nd dimension run in 1x Tris-glycine running buffer for 10 min at room temperature. Repeat this step once more.

26| Each equilibrated gel piece is loaded into separate well of the SDS-polyacrylamide gel together with a small volume of 1x SDS loading buffer to allow monitoring of migration. Electrophoresis is allowed to run for 10-15 min after the dye has entered the separating part of the gel at a constant voltage of 100 V to ensure all samples have completely migrated from the stacking into the separating gel.

27| Once electrophoresis concludes, excise the whole band with the dye from the gel that now contains the concentrated proteins in resolving gel (10-15 mm-long gel slice). The excised gel slice can now be cut in small pieces with a clean scalpel and subjected to trypsin digestion and extraction for mass spectrometry according to the specifications of the instrument used (for an example using HeLa cells, see ref. 8).

Δ CRITICAL STEP: In order to avoid excessive contamination by skin keratin it is advisable to handle all related equipment and consumables while wearing gloves.

TIMING

The two legs of this protocol are variably timed, but the core is common to both (steps **1-14**) and concludes within less than a day and requires little hands-on time. The transcriptomics leg is also not laborious; it requires standard RNA isolation, followed by the preparation of cDNA libraries for high-throughput sequencing (**Box 3**). Then, depending on the sequencing facility of choice, the turnaround times may vary. As regards the proteomics leg, analysis by native 2D electrophoresis is quite laborious, and may prove even more so for inexperienced users. However, if the separation of complexes belonging to each of the three RNA polymerases is not required, then the factory fraction obtained by step **14** can be essentially directly analyzed by mass-spectrometry. Nonetheless, for both legs it is important to devote some time to quality controls using either RT-qPCR (**Box 3**) or Western blots (**Box 2**); these should take place after obtaining the factory fraction and before proceeding to downstream high-throughput analyses of either kind. Of course, addition of a "run-on" (**Box 1**) may prove helpful at the beginning, without dramatically increasing the total time required.

ANTICIPATED RESULTS

This protocol results in the cataloguing of RNA and proteins associated with transcription factories. Depending on the desired end-point analysis, anticipated results and necessary controls differ.

For high-throughput sequencing of HUVEC factory RNA we previously performed an analysis using standard bioinformatic tools[9]. In brief, we mapped the resulting 100-bp reads, reported as a

".fastq" file, to the human genome (reference build GRCh37/hg19) using Bowtie[30] and default parameters. Reassuringly <10% of mapped reads mapped to rDNA regions, and these reads were excluded from downstream analyses. The Integrative Genomics Viewer (www.broadinstitute.org/igv) was used to visualize genomic coverage, transcripts were assembled using Cufflinks[31] and annotated via the Genomatix software suite (Genomatix). It should be noted here that there now exist dedicated *in silico* pipelines[32,33] for assessing changes in gene expression based on intronic RNA levels (instead of exonic levels that standard algorithms use), and we anticipate that their implementation will further enhance the sensitivity of factory RNA-seq.

It is also recommended to perform side-by-side analysis of the factory and total or poly(A)-enriched RNA-seq datasets from the same cell type and set of conditions. Typically factory RNA-seq returns read profiles that are rich in intronic signal (**Fig. 4b-d**). The expected read distribution, compared to poly(A)-enriched RNA, for example, should show marked depletion in exonic reads with a concomitant enrichment of reads mapping to introns, intron-exon junctions, and intergenic space (**Fig. 4c**). A key advantage of our method is the detection of eRNAs produced at active enhancers, despite them being so unstable. Here, detection sensitivity (as exemplified using a set of ~1300 active intergenic enhancers[9] in **Fig. 4e**) is superior to that achieved by total (ribodepleted) or chromatin RNA-seq.

For proteomics analyses, depending on the instrument used and the implementation of metabolic protein labeling[25], results will vary. Here, we present an exemplary analysis on isolated HeLa factory complexes for all three polymerase as a guide (from ref. 8). After resolving complexes on native 2D gels (**Fig. 5a,b**), each sample was excised as a single gel piece, proteins analyzed by SDS-PAGE and trypsin-treated, and the now-separated peptides extracted, vacuum dried, and injected (typically three injections/sample; 120 min/injection) into a Dionex U3000 nanoHPLC system coupled to an LTQ Orbitrap mass spectrometer (Thermo Scientific). The three resulting data files were merged, converted to ".mzXML" format and submitted to the Central Proteomics Facilities pipeline[34]. This pipeline uses a combination of search engines and scores, and sorts peptides hits using the integrated iProphet toolkit[35]. Then, predictions are validated by comparison to a concatenated target-decoy database (International Protein Index human v3.64) to return the final protein lists with <1% false discovery rate (FDR; empirical). As a further test, we also generate protein lists using a more stringent 0.5% FDR cutoff; these should typically contain ≥90% of proteins predicted using the 1% FDR cutoff. Biological replicates of the above are expected to display ≥75% overlap in the proteins identified (full data from ref. 8 in: http://users.path.ox.ac.uk/~pcook/data/ContentOfFactories.html). Note that we chose to apply the "normalized spectral index" approach[36] to our data to estimate relative protein abundancies, in the absence of metabolic labeling.

Additional analysis of the protein lists retrieved per complex may be performed as follows. First, proteins identified for the complexes of RNA polymerase I should display significant (≥80% in data from ref. 8) overlap to the proteome of isolated nucleoli[13]. Then, each complex should contain proteins uniquely associated with the respective polymerase (e.g., respective RNA polymerase subunits, general transcription factors like TFIIH for polymerase II, or the Lupus La antigen for polymerase III), as well as a considerable number of shared protein components (for an overview see **Fig. 5c**). Finally, in order to determine the Gene Ontology (GO) terms that concisely describe the contents of each factory complex, we produced the "MS-prot" algorithm (freely-available at http://www.ms-prot.co.uk/) that connects a given UniProt accession number to all associated GO terms, and allows user-defined GO term groups to be used as a filter. Then, when factory-associated GO terms are compared to those associated with terms like "cytoplasm" and "nucleus", or with S100 cytosolic extracts, the particular enrichments of factory fractions are revealed (**Fig. 5d**).

ACKNOWLEDGEMENTS

We thank Binwei Deng and Jon Bartlett for their help, and the SWDSOP proteomics and BioQuant/ DKFZ sequencing facilities for the high-throughput analyses. This work was supported by a Medical Research Council grant (to P.R.C.), by the ERASysBio+/FP7 initiative (to K.R. and P.R.C.), by a DKFZ intramural grant (to M.C.-H.), and by CMMC core funding (to A.P.).

AUTHOR CONTRIBUTIONS

S.M. and P.R.C. conceived and developed the factory isolation procedure. S.M., A.P., and P.R.C. implemented and validated the procedure. I.M.C. produced the "MS-prot" script. M.C.-H., L.B., K.R., and A.P. adapted and implemented the protocol for nascent RNA-seq. All authors analyzed data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

REFERENCES

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. Papantonis, A. & Cook, P.R. Transcription factories: genome organization and gene regulation.

Chem. Rev. **113**. 8683-8705 (2013).

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. Melnik, S., Deng, B., Papantonis, A., Baboo, S., Carr, I.M. & Cook, P.R. The proteomes of transcription factories containing polymerases I, II, or III. *Nat. Methods* **28**, 963-968 (2011).
9. Caudron-Herger, M., Cook, P.R., Rippe, K. & Papantonis, A. Dissecting the nascent human transcriptome by analyzing the RNA content of transcription factories. *Nucleic Acids Res*. **pii**, gkv390 (2015).
10. 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).
11. 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).
12. Jackson, D.A. & Cook, P.R. Transcription occurs at a nucleoskeleton. *EMBO J.* **4**, 919-925 (1985).
13. Ahmad, Y., Boisvert, F.M., Gregor, P., Cobley, A. & Lamond, A.I. NOPdb: Nucleolar Proteome Database--2008 update. *Nucleic Acids Res.* **37**, D181-184 (2009).
14. Aviner, R., Geiger, T. & Elroy-Stein, O. Novel proteomic approach (PUNCH-P) reveals cell cycle-specific fluctuations in mRNA translation. *Genes Dev*. **27**, 1834-1844 (2013).
15. Core, L.J., Waterfall, J.J. and Lis, J.T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845-1848.
16. Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A. & Churchman, L.S. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**, 541-554 (2015).
17. Nojima, T., Gomes, T., Grosso, A.R., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M. & Proudfoot, N.J. Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell*. **161**, 526-40 (2015).
18. Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L. & Smale, S.T. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279-290 (2012).
19. Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L. & Feuk, L. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol*. **18**, 1435-1440 (2011).

20. Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H., Marr, M.T. 2nd & Rosbash, M. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev*. **25**, 2502-2512 (2011).
21. Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., Amit, I. & Regev, A. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol*. **29**, 436-442 (2011).
22. Windhager, L., Bonfert, T., Burger, K., Ruzsics, Z., Krebs, S., Kaufmann, S., Malterer, G., L'Hernault, A., Schilhabel, M., Schreiber, S., Rosenstiel, P., Zimmer, R., Eick, D., Friedel, C.C. & Dölken, L. Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Res*. **22**, 2031-2042 (2012).
23. Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H,, Worley, P.F., Kreiman, G. and Greenberg, M.E. (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-187.
24. Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun H.B., Tough, D.F., Prinjha, R.K., Benner, C. and Glass, C.K. (2013) Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* **51**, 310-325.
25. Mann, M. Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell. Biol*. **7**, 952-958 (2006).
26. Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J.V. & Mann, M. A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat Protoc*. **4**, 698-705 (2009).
27. 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).
28. 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).
29. Nadano, D., Aoki, C., Yoshinaka, T., Irie, S. & Sato, T.A. Electrophoretic characterization of ribosomal subunits and proteins in apoptosis: specific downregulation of S11 in staurosporine-treated human breast carcinoma cells. *Biochemistry* **40**, 15184-15193 (2001).
30. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359 (2012).
31. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L.,

Rinn, J.L. & Pachter, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc*. **7**, 562-578 (2012).

32. Madsen, J.G., Schmidt, S.F., Larsen, B.D., Loft, A., Nielsen, R. & Mandrup, S. iRNA-seq: computational method for genome-wide assessment of acute transcriptional regulation from total RNA-seq data. *Nucleic Acids Res*. **43**, e40 (2015).
33. Gaidatzis, D., Burger, L., Florescu, M. & Stadler, M.B. Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nat Biotechnol*. **33**, 722-729 (2015).
34. Trudgian, D.C., Thomas, B., McGowan, S.J., Kessler, B.M., Salek, M. & Acuto, O. CPFP: a central proteomics facilities pipeline. *Bioinformatics* **26**, 1131-1132 (2010).
35. Shteynberg, D., Deutsch, E.W., Lam, H., Eng, J.K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R.L., Aebersold, R. & Nesvizhskii, A.I. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol. Cell. Proteomics* **10**, M111.007690 (2011).
36. Griffin, N.M., Yu, J., Long, F., Oh, P., Shore, S., Li, Y., Koziol, J.A. & Schnitzer, J.E. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat. Biotechnol*. **28**, 83-89 (2010).
37. Lovén, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., Burge, C.B., Levens, D.L., Lee, T.I. & Young, R.A. Revisiting global gene expression analysis. *Cell* **151**, 476-482 (2012).

FIGURE LEGENDS

Figure 1 | Overview of the factory-isolation strategy. Live (not cross-linked) cells are collected in a physiological buffer (PB), and nuclei are isolated after treating cells with NP-40. Next, whole nuclei (*black oval*) are treated with DNase I in a step where most chromatin (*blue lines*) is detached and can be separated from the rest of the nucleus by centrifugation. Finally, transcription factories (*orange spheres*), and any attached DNA (*blue*) and RNA (*red*) are released into the supernatant by digestion with a mixture of active caspases. As a result, different subcellular fractions (each shown boxed) may be isolated along the course of this procedure.

Figure 2 | Workflow for the isolation of transcription factories. This procedure can be applied to perform either an analysis of RNA ("transcriptomics"; *left leg*) or of proteins ("proteomics"; *right leg*) associated with transcription factories. Here, the general workflow is presented with an annotation of possible pause points (*green*), steps at which quality checks are performed (*blue*), and timing.

Figure 3 | Factory RNA-sequencing performed on low-cell counts. (**a**) Typical genome-browser profiles (in reads per million; rpm) along two neighboring genes on human chromosome 14, *CGRRF1* and *SAMD4A*, obtained by deep-sequencing factory RNA prepared from 5 (*black*), 1 (*grey*), and 0.5 million (*light grey*) osteosarcoma U2OS cells. (**b**) Correlation plots of gene expression levels obtained using factory RNA from 0.5-5 million cells. Calculated Spearman's correlation coefficients (ρ) indicate the reproducibility of the results despite using decreasing cell counts.

Figure 4 | Anticipated results following factory RNA-seq analysis. (**a**) Analysis of factory-associated chromatin isolated from cells untreated (-) or treated (+) with DNase I; the middle lane represents a partially-undigested sample, with the optimal profile shown on the right. (**b**) Typical genome-browser profiles (in reads per million; rpm) along a non-coding, *LINC00478*, and a protein-coding gene, *SAMD4A*, obtained by deep-sequencing poly(A)-enriched (*orange*), factory (*magenta*), and total ribodepleted RNA (*blue*) prepared from primary endothelial cells (HUVECs). (**c**) Expected transcript distributions of mapped reads over genomic features as seen in total ribodepleted, poly(A)-enriched, factory, and chromatin-associated RNA. The percentage of reads belonging to the prevailing feature per experiment is shown. (**d**) Average read density for 1 kbp up- and down-stream of the TSS of ~8,000 active genes plotted using RNA-seq data from total ribodepleted (*blue*), poly(A)-enriched (*orange*), factory (*magenta*), and chromatin-associated (*black*) RNA. (**e**) Average eRNA read density for 2 kbp up- and down-stream of ~1,300 active HUVEC enhancers plotted using RNA-seq data from total ribodepleted (*blue*), factory (*magenta*), and chromatin-associated (*black*) RNA. All data for panels **a**-**e** were generated using HUVECs, and have been adapted by permission from ref. 9.

Figure 5 | Anticipated results following factory proteomic analysis. (**a**) Resolving RNA polymerase (RNAP) complexes II and III by native 2D electrophoresis using Coomassie blue in the first dimension. The cartoon (*top left*) shows regions excised for mass spectrometry. An autoradiograph (*top right*) of the gel shows overlapping spots of (nascent) [^{32}P]RNA along the diagonal. The region with prominent signal contained ~0.03%, ~0.8%, and ~5% of the protein, DNA, and [^{32}P]RNA initially present, respectively. The protein sample was next transferred onto nitrocellulose and immuno-blotted successively for RNAPII (*bottom left*) and III (*bottom right*). (**b**) Resolving RNAPI complexes by native 2D electrophoresis in the absence of Coomassie blue in the first dimension. All details are as in panel a, except for immuno-blotting for RNAPI (*bottom left*) and II (*bottom right*). (**c**) The Venn diagram shows the number of proteins detected per RNAP complex and their overlap. (**d**) The bar plot shows a comparison of Gene Ontology (GO) terms associated with the proteins identified in each RNAP fraction, as well as the cytoplasmic, nuclear, and nucleolar fractions, plus those in S100 cytosolic extracts. All data for panels **a**-**d** were generated using HeLa cells, and have been adapted by permission from ref. 8.

TABLES

Table 1 | Comparison of factory RNA-seq to other methods that capture nascent RNA.

RNA-seq method	Applied in	Advantages	Limitations
metabolic RNA 4sU-tagging[21,22]	human, mouse	Straightforward 4sU incorporation; robust intronic RNA coverage.	4sU pulse labeling might perturb transcript biogenesis and distribution; difficult to adjust pulse length so as to avoid labeling mRNAs.
factory RNA-seq[9]	human, mouse	Straightfoward and fast protocol; robust intronic and eRNA coverage; low sequencing depth required; separates chromatin- from factory-associated RNA.	Side-effects of caspase treatment are not clear; not yet adapted to fixed cells.
GRO-seq[15]	human, mouse	Useful for identifying paused RNAPs; robust coverage of enhancer and promoter-associated RNAs.	Laborious; relying on antibody affinity for the pull-down; artificial induction of RNAP elongation; potential overestimation of transcript levels.
nascent-/chromatin-seq[18,20]	human, mouse, fruitfly	Good coverage of intronic RNA; strong mRNA depletion.	Extraction under non-physiological conditions; high sequencing depth required; long-lived chromatin-associated and nascent RNAs mixed.
(m)NET-seq[16,17]	human, yeast	Single-nucleotide resolution; strong mRNA depletion.	Extraction under non-physiological conditions; chromatin-associated and nascent RNAs mixed; eRNA coverage unclear; the "pull-down" variant[13] relies on antibody affinity.
poly(A)-minus RNA-seq[19]	human, mouse, fruitfly, yeast	Straightforward, based on total RNA isolation.	mRNA depletion is not quantitative; high sequencing depth required; chromatin-associated and nascent RNAs mixed.

Table 2 | Point-by-point troubleshooting.

Protocol step	Problem	Possible source	Proposed solution
5-6	Aggregation of isolated nuclei.	This is a cell type-specific effect and can be aggravated by DNA leakage.	Optimize detergent concentration and treatment times for each cell type. Dilute nuclei into larger volumes; add PB to the nuclei in a stepwise manner.
7	Poor DNase I digestion.	The coating of cell culture containers, or the brand of DNAse I used.	We have observed that DNases I from different manufacturers display variable cutting efficiencies when treating cells grown on coated cell-culture containers (e.g., mouse ES cells on gelatin-coated plates); here, trypsinization is preferred to scraping and cells should then be washed twice in PB. The recommended DNase I shows robust digestion across conditions. DNase I performance is also pH-dependent, and can be inefficient if the PB buffer has a pH even slightly higher than 7.4.
14	Very low yield in factory fraction.	Poor DNase I and/or caspase digestion.	Isolate RNA from the respective pellet and quantify relative to the released factory fraction. If not applied to mammalian cells, investigate the efficiency of caspase digestion in your system. It is also critical to get complete solubization of the pellet in NLB buffer; no pellet clumps should be observed. For DNase I see above.
23	Large overlap between the three different RNA polymerase complexes.	Poor DNase I and/or caspase digestion, or poor electrophoretic separation.	See above. Also check that 2D gel concentration and electrophoresis parameters are exactly as described, and that gels are well polymerized.
Box 1	Poor “run-on” signal.	Cells not in exponential growth phase.	It is advised to split cells the day before the experiment.
Box 3	Highly-degraded RNA profiles.	High titers of free bivalent cations in PB, or RNase contamination.	Do not exceed the proposed 0.5 mM of $CaCl_2$ for DNase digestion. Always use buffers prepared with DEPC water, RNase inhibitors and RNase-free tubes and glassware.

Box 1 | Labeling nascent RNA by "run-on" ● TIMING: 2 h

Every time a new cell type is used for isolating factories, it is recommended that "run-on" is performed to check the extent to which transcriptional activity is retained in PB. The steps described below also enable tracking of labeled nascent RNA during the course of the procedure.

1| Wash cells twice in PB; pellet cells pelleted by centrifugation at 400 x g, 5 min, at 4°C.

2| Resuspend cells in PB and permeabilize for 5 min on ice by the adding saponin to a final concentration of 170 µg/ml.

Δ CRITICAL STEP: The saponin concentration required for permeabilization needs to be established during preliminary experiments for each cell type and batch of saponin. For this, cells are incubated in increasing concentrations of saponin (50-300 µg/ml) for 5 min on ice, and the optimal concentration is determined as that giving ~98% permeabilized cells without extensive cell lysis (assessed using Trypan blue).

3| Wash permeabilized cells 4 times in PB (as above, making sure cells stay in PB for at least 5 min/wash), while the "run-on" mix is prepared.

4| Prewarm permeabilized cells in PB at 33°C for 3-5 min.

5| Add the "run-on" mix to the prewarmed cells so that it just covers cells (e.g., 2.5 ml for a 15-cm culture dish, or 3-5 ml for a cell pellet). Incubate at 33°C for 10 min.

6| Wash twice in PB as above to remove the mix.

Box 2 | Native 2D electrophoresis, analytical ● TIMING: 3-4 h

Samples are separated in 2D analytical gels before the preparative one to assess the quality of the sample, and to determine the relative positions of the complexes containing each RNA polymerase and of the associated RNA (which can be traced if labelled during the "run-on") as follows.

1| Cast three identical composite gels are cast side-by-side (e.g., using the Mini-Protean3® system with gel dimensions of 8.6 x 6.7 x 0.15 cm). One gel is used for protein staining to assess separation of the complexes, another to trace nascent [^{32}P]RNA (by autoradiography), and a third to locate the different RNA polymerase complexes (by Western blot analysis after transferring protein complexes on a nitrocellulose membrane). Gel casting is exactly as described for preparative 2D gels, using 1 mm-thick spacers for the 1st dimension gel, and 1.5-mm ones for the 2nd dimension gel.

2| Load a sample equivalent to 2x10^{6} cells per each gel.

3| Run gels at a constant voltage of 100 V until the xylene cyanol dye added to the loading buffer has migrated ¾ of the gel's length, and the bromophenol blue dye is lost.

4| Once the run concludes, stain one of the gels using Imperial stain as per manufacturer's instructions to assess the efficiency of complexes separation. Place another gel on a Whatman paper, dry it, and visualize nascent [^{32}P]RNA by autoradiography using Hyperfilm MP (GE Healthcare). In both cases three major stained spots with good separation should be observed (see **Fig. 5**).

5| Transfer the complexes on the third gel onto a nitrocellulose membrane using the iBlot system (Program #3 for 7 min; Life Technologies). Use the membrane in Western blots with antibodies targeting major subunits of the three mammalian RNA polymerases, or other proteins of interest.

Δ CRITICAL STEP: We recommend confirming efficient protein transfer by staining with Ponceau S. If the antibody used for the Western blot produces a high background (as we experienced with the best available antibody to detect RNA pol III, chicken anti-RPC62), blocking membranes using protein-free blocking buffer (BB; Thermo Scientific) is recommended, rather than 5% non-fat milk in TBS-T (for 30 min at room temperature). Imaging membranes, after performing chemiluminscence detection, on a Fuji LAS-4000 scanner (rather than exposing a film) is also recommended. The membrane can be successively probed with antibodies recognizing subunits of RNA polymerase I, II, or III (see MATERIALS); membranes can be stripped up to four times using the RestorePlus™ buffer (Thermo Scientific) for successive blotting with another antibody.

Box 3 | Preparation of cDNA libraries for RNA sequencing ● TIMING: 5 h

Factory RNA and—if desired—chromatin-associated RNA from step **9** are purified by the Trizol reagent as per manufacturer's instructions or using a kit. We typically use the Direct-zol™ RNA MiniPrep kit (Zymo Research, cat. no. R2050) including a DNase I-treatment step. Following this procedure, preparations mostly contain long (>150-200 nt) RNAs. But it is possible to also isolate short (20-200 nt) RNAs in a separate fraction that can be used for high-throughput sequencing when using a dedicated cDNA library construction kit (e.g., TruSeq Small RNA library preparation kit; Illumina, cat. no. RS-200-0012).

The resulting RNA can also be used in standard RT-qPCR tests to verify its enrichment for nascent (intronic) RNA in comparison to total cell RNA preparations; for this we typically use primers spanning exon-intron junctions of actively-transcribed genes in the cell type of interest (and exon-exon junctions as controls, although these should be expected to be depleted from factory RNA preparations). Following these test, factory RNA is depleted of rRNA species using the RiboZero™ rRNA removal kit (Illumina, cat. no. MRZH116) and used directly with the NEBNext® Ultra™ Directional RNA Library Prep kit (New England Biolabs) to generate cDNA libraries for sequencing on an Illumina platform as per manufacturer's instructions. This way, the strand-specific information of the RNA pool is preserved. It is also advisable to "spike" preparations with exogenous RNA of known concentrations to facilitate comparison of RNA sequencing data across cell types and conditions[37].

Finally, cDNA libraries are multiplexed and (typically paired-end) sequenced on a HiSeq2000 platform (Illumina). For our own work we generated between 30 and 50 million 100-bp reads per sample, in two biological replicates, and such sequencing depth proved sufficient for both the robust genome-wide coverage of intronic RNA and the detection of low-level, unstable, species such as eRNAs[19,20]. Note that in the meantime we have successfully managed to obtain reproducible factory RNA-seq data from U2OS cells using increasingly lower cell counts (down to $5x10^5$ cells; see **Fig. 3**).

Δ CRITICAL STEP: The RNA isolated via this protocol may not retain the typical profile of whole cell RNA with its prominent rRNA bands. As a result, the profiles might appear to fail the RIN criteria for pre-sequencing quality control. This should be ignored, unless excessive degradation is observed as apparent from fragment sizes below 50-100 nt average length.

Figure 1

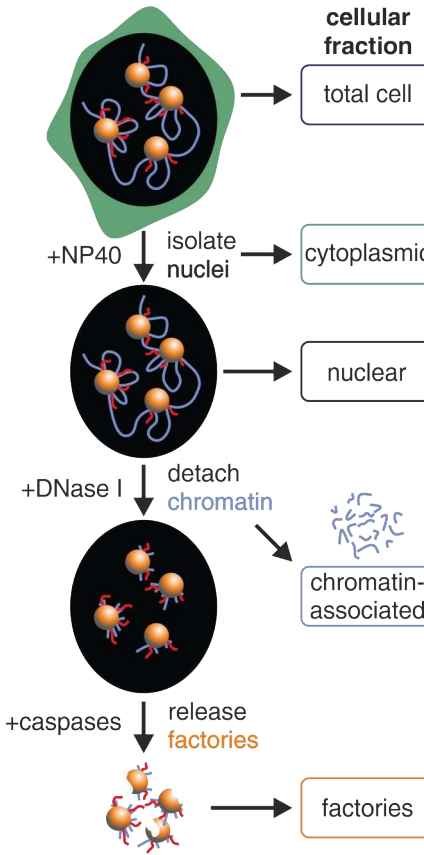

cellular fraction
total cell
+NP40
isolate nuclei
cytoplasmic
nuclear
+DNase I
detach chromatin
chromatin-associated
+caspases
release factories
factories

Figure 2

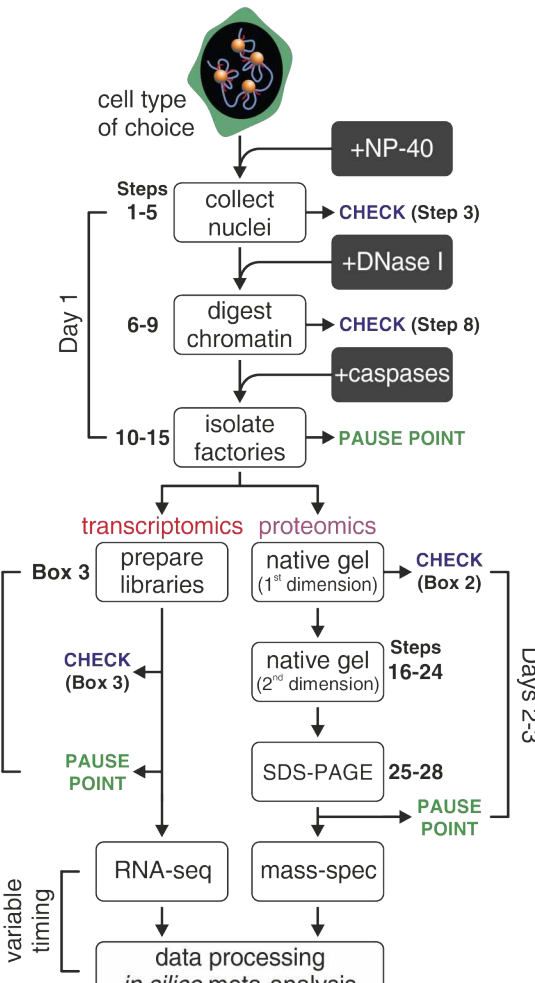

cell type of choice
+NP-40
Steps 1-5
collect nuclei
CHECK (Step 3)
+DNase I
Day 1
6-9
digest chromatin
CHECK (Step 8)
+caspases
10-15
isolate factories
PAUSE POINT
transcriptomics
proteomics
Box 3
prepare libraries
native gel (1st dimension)
CHECK (Box 2)
Day 2
CHECK (Box 3)
native gel (2nd dimension)
Steps 16-24
Days 2-3
PAUSE POINT
SDS-PAGE
25-28
PAUSE POINT
RNA-seq
mass-spec
variable timing
data processing
in silico meta-analysis

Figure 3

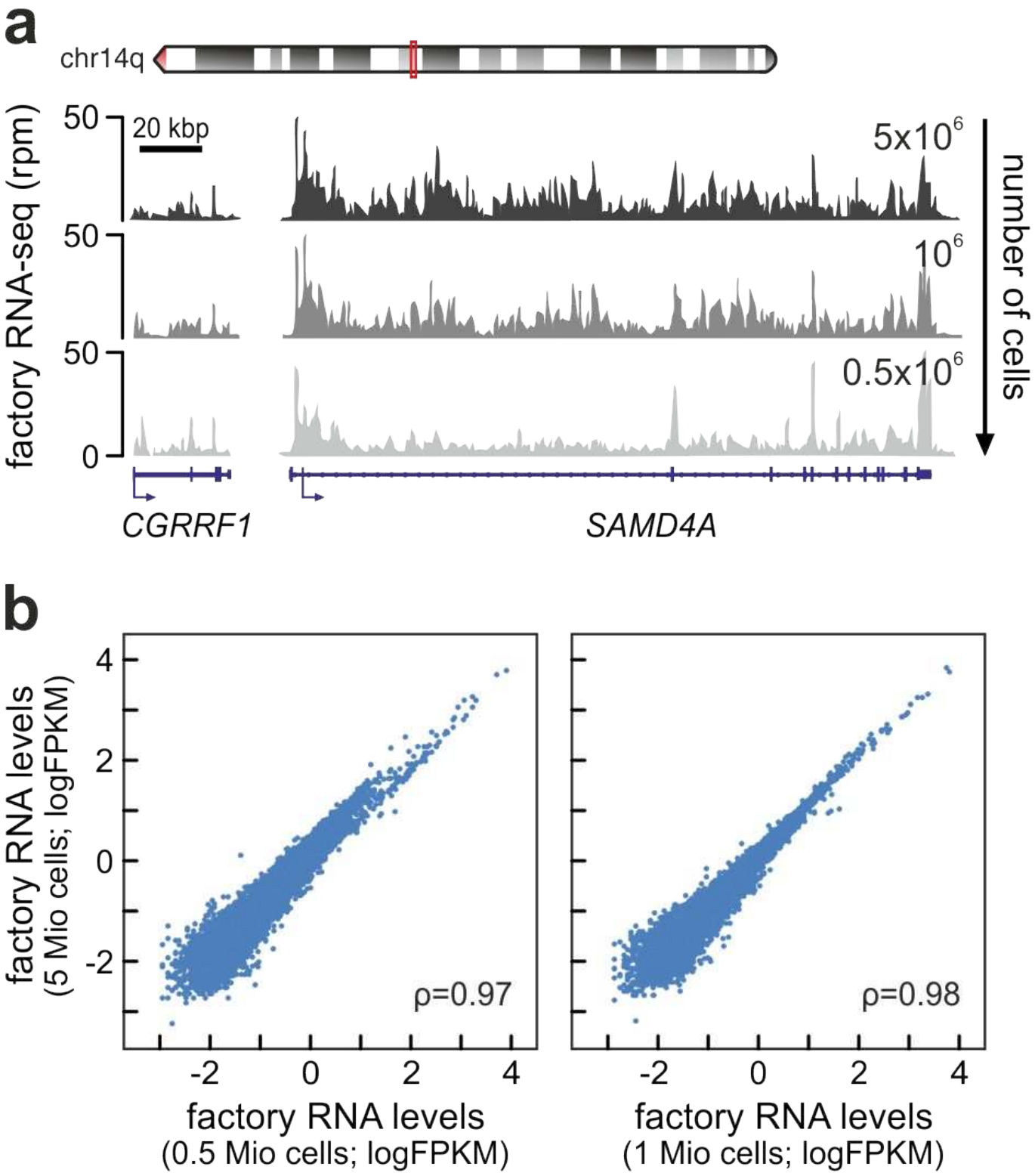

a
chr14q
factory RNA-seq (rpm)
50
20 kbp
50
50
0
5x10^6
10^6
0.5x10^6
number of cells
CGRRF1
SAMD4A
b
factory RNA levels
(5 Mio cells; logFPKM)
4
2
0
-2
ρ=0.97
ρ=0.98
-2
0
2
4
factory RNA levels
(0.5 Mio cells; logFPKM)
factory RNA levels
(1 Mio cells; logFPKM)

Figure 4

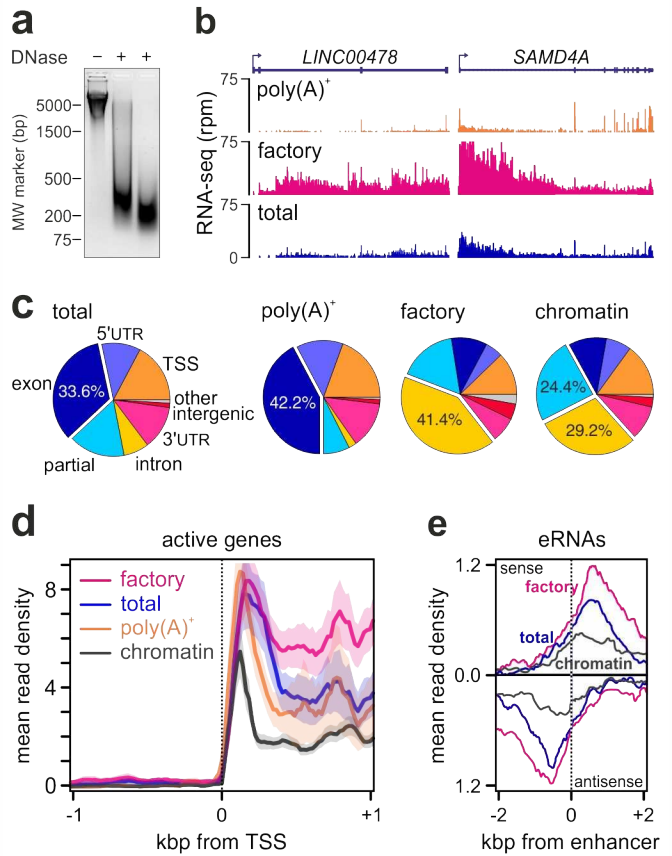
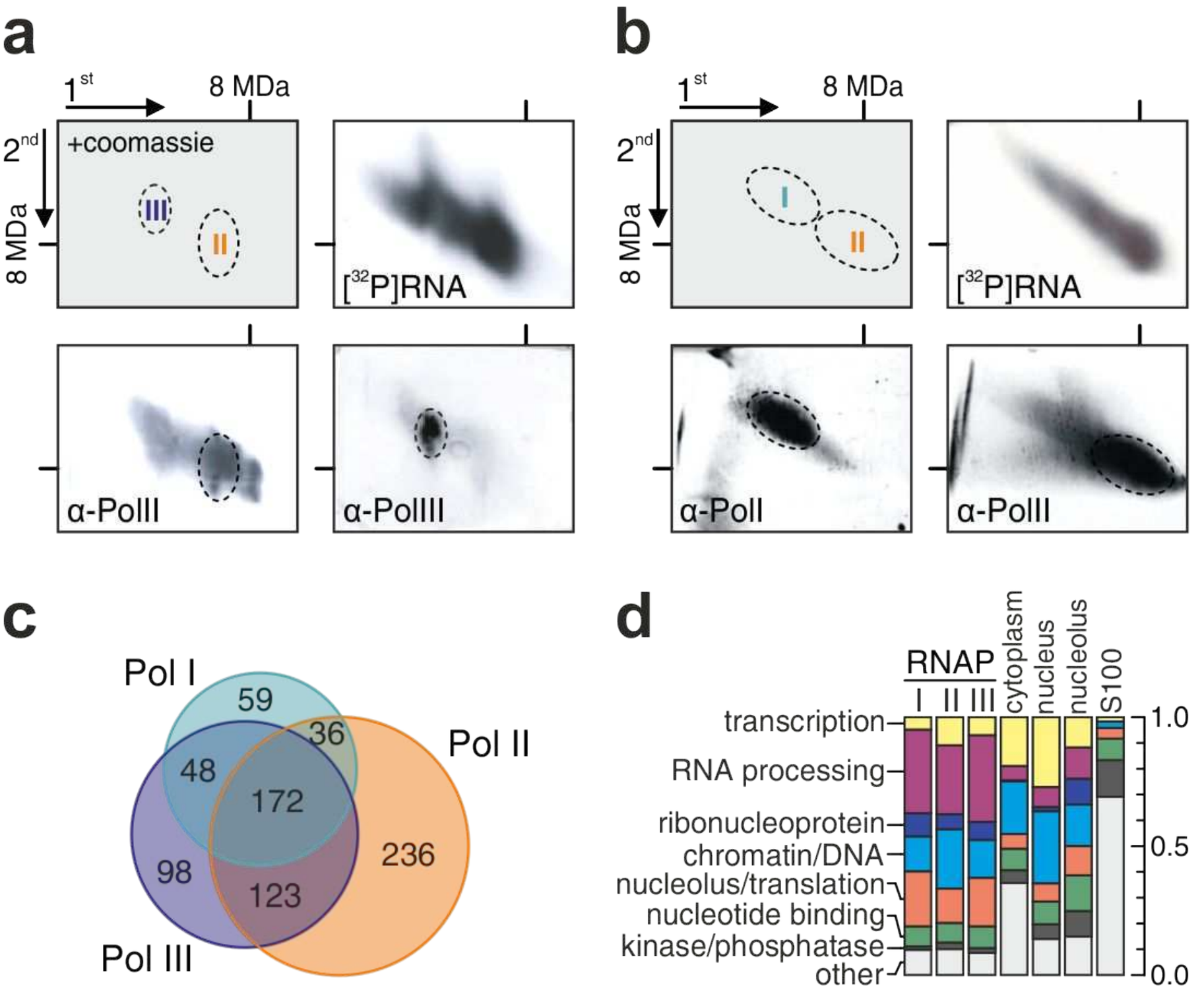

a
DNase
– + +
MW marker (bp)
5000
1500
500
200
75
b
LINC00478
SAMD4A
RNA-seq (rpm)
poly(A)+
factory
total
75
75
75
0
c
total
poly(A)+
factory
chromatin
5'UTR
TSS
exon
33.6%
other intergenic
3'UTR
partial
intron
42.2%
41.4%
24.4%
29.2%
d
active genes
factory
total
poly(A)+
chromatin
mean read density
8
4
0
-1
0
+1
kbp from TSS
e
eRNAs
sense
factory
total
chromatin
antisense
mean read density
1.2
0.0
1.2
-2
0
+2
kbp from enhancer

Figure 5


a
1st
8 MDa
2nd
8 MDa
+coomassie
III
II
[32P]RNA
α-PolIII
α-PolIII
b
1st
8 MDa
2nd
8 MDa
I
II
[32P]RNA
α-PolI
α-PolII
c
Pol I
59
36
Pol II
48
172
236
98
123
Pol III
d
RNAP
I
II
III
cytoplasm
nucleus
nucleolus
S100
transcription
RNA processing
ribonucleoprotein
chromatin/DNA
nucleolus/translation
nucleotide binding
kinase/phosphatase
other
1.0
0.5
0.0