Published online 20 April 2015

Nucleic Acids Research, 2015, Vol. 43, No. 14 ***e95***
doi: 10.1093/nar/gkv390

Dissecting the nascent human transcriptome by analysing the RNA content of transcription factories

Maïwen Caudron-Herger[1], Peter R. Cook[2], Karsten Rippe[1] and Argyris Papantonis[2,3,*]

[1]Deutsches Krebsforschungszentrum (DKFZ) & BioQuant, D-69120 Heidelberg, Germany, [2]Sir William Dunn School of Pathology, University of Oxford, OX1 3RE Oxford, UK and [3]Center for Molecular Medicine, University of Cologne, D-50931 Cologne, Germany

Received February 16, 2015; Revised March 24, 2015; Accepted April 13, 2015

ABSTRACT

While mapping total and poly-adenylated human transcriptomes has now become routine, characterizing nascent transcripts remains challenging, largely because nascent RNAs have such short half-lives. Here, we describe a simple, fast and cost-effective method to isolate RNA associated with transcription factories, the sites responsible for the majority of nuclear transcription. Following stimulation of human endothelial cells with the pro-inflammatory cytokine TNFα, we isolate and analyse the RNA content of factories by sequencing. Comparison with total, poly(A)$^+$ and chromatin RNA fractions reveals that sequencing of purified factory RNA maps the complete nascent transcriptome; it is rich in intronic unprocessed transcript, as well as long intergenic non-coding (lincRNAs) and enhancer-associated RNAs (eRNAs), micro-RNA precursors and repeat-derived RNAs. Hence, we verify that transcription factories produce most nascent RNA and confer a regulatory role via their association with a set of specifically-retained non-coding transcripts.

INTRODUCTION

Over the last years, the human transcriptome has been repeatedly revisited and found to be more complex than expected, mainly due to the large fraction of non-coding RNA (ncRNA) (1). Besides the classical ribosomal, transfer and small nucleolar RNAs, many other non-protein-coding RNAs have been uncovered (2,3) including micro-RNAs (miRNAs)(4), long inter-genic non-coding RNAs (lincRNAs)(5), copies of repeats (6,7), as well as short transcripts in and around transcription start sites (TSSs)(8,9), active enhancers (eRNAs)(10–12) and introns (13). In order to identify changes of this complex mixture of RNAs in a cell population, it is highly informative to map the dynamics of the nascent transcriptome. Most nascent transcripts have short half-lives, are variably processed and studies have shown that profiles change rapidly. Of the various approaches described for assessing nascent transcriptomes, including 'GRO-seq' (8), 'NET-seq' (14), 'chromatin RNA-seq' (15), 'poly(A)-depleted RNA-seq' (16,17) and the metabolic tagging of newly-made RNA using 4-thiouridine (18,19) each has particular shortcomings, and—in general—all are laborious and/or require a high 'sequencing depth'; most also focus on just parts of the transcriptome.

Here we introduce a simple, fast and cost-effective method to isolate and sequence nascent RNAs. It is based on the biochemical isolation of 'transcription factories', the megadalton large complexes that harbour >95% of all nuclear transcription (20–22). Earlier controversy on the extent to which factories contribute to the act of transcription (23) has been relieved by recent studies confirming that (i) RNA synthesis in eukaryotic nuclei occurs in discrete foci that contain several active transcription units (24–26), (ii) factories can be isolated biochemically and their protein constituents catalogued by mass-spectrometry (27) and (iii) co-regulated genes and their enhancers that lie distant in genomic space often come together in 3D nuclear space when transcribed (28–33).

Here, by analysing the RNAs differentially-produced by (and associated with) factories, upon stimulation of human umbilical vein endothelial cells (HUVECs) with the pro-inflammatory cytokine TNFα, we validate that they represent the transcriptional hotspots of the nucleoplasm. Moreover, by comparison to different sub-cellular RNA fractions, we identify particular long and short non-coding transcripts that can confer much of the processing and regulatory activities attributed to transcription factories.

MATERIALS AND METHODS

Cell culture

HUVECs from pooled donors (Lonza), grown to 90% confluence in Endothelial Basal Medium 2-MV with supplements (EBM; Lonza) and 5% fetal bovine serum (FBS), were 'starved' for 16 h in EBM + 0.5% FBS, treated with

*To whom correspondence should be addressed: Tel: +49 221 478 96987; Fax: +49 221 478 4833; Email: argyris.papantonis@uni-koeln.de

TNFα (10 ng/ml; Peprotech) and harvested 0 or 30 min post-stimulation.

Isolation of sub-cellular fractions and RNA extraction

The protocol was modified from Melnik *et al.* (27) to result in isolation of RNA from the various fractions along its course; the general workflow is outlined in Figure 1A. All buffers were prepared using water treated with diethylpyrocarbonate (DEPC) and using nuclease- and protease-free reagents; they were used ice-cold unless stated otherwise. All washes/spins were at 600 × *g* for 5 min at 4°C. Approximately 1.5 – 2 × 10^7 cells (stimulated with TNFα for 0 or 30 min) were harvested in complete physiological buffer (PB*) using a rubber scraper (StarLabs). [PB* is based upon 100 mM potassium acetate, 30 mM KCl, 10 mM Na_2HPO_4 and 1 mM $MgCl_2$. Immediately before each experiment, the following components were added: 1 mM Na_2ATP, 1 mM dithiothreitol, 25 units/ml RNase inhibitor (RiboLock; Fermentas), 10 mM β-glycerophosphate, 10 mM NaF, 0.2 mM Na_3VO_4 and a 1/1000 dilution of complete protease inhibitor cocktail (PIC; Roche). As the acidity of adenosine triphosphate (ATP) batches varies, 100 mM KH_2PO_4 is used to adjust the pH to 7.4.] Nuclei were isolated by washing 3× (10 min) in 5 ml PB* plus 0.4% NP-40 (Igepal; Sigma-Aldrich), pelleted by centrifugation, gently resuspended in PB* + 0.4% NP-40 (100 μl per 10^7 cells), and treated (30 min, 33°C) with DNase I (Worthington; 10 units/10^7 cells plus 0.5 mM $CaCl_2$) or *Hae*III (1000 units/10^7 cells) without shaking in round-bottomed 2-ml tubes (Greiner). Reactions were stopped by adding ethylenediaminetetraacetic acid to 2.5 mM and cooling on ice and digested nuclei collected by centrifugation. This nucleolytic treatment releases most chromatin into the supernatant, from which 'chromatin-associated' RNA was isolated by immediately mixing with 750 μl Trizol LS (Invitrogen) and subsequent purification using the manufacturer's instructions. The pellet was now resuspended in Native Lysis Buffer (NLB; 50 μl per 10^7 cells). NLB contains 40 mM Tris-acetate pH 7.4, 2 M 6-aminocaproic acid, 7% sucrose, plus 1/1000 PIC and 50 units/ml RiboLock. After vortexing vigorously, the mixture was incubated (20 min; ice). Next, a mixture of caspases (2 units/μl of caspases 6, 8, 9 and 10 in PB* + NP-40; BioVision) was added to the resuspended pellet (2 μl of mix/10^7 nuclei) and the mixture shaken (900 rpm) and incubated (30 min, 33°C) in a thermomixer (Eppendorf). [None of the core subunits of RNA polymerase I, II or III, contain sites for these caspases, except RPB9 (27).] The reaction was stopped by adding Caspase Inhibitor III (Calbiochem) to 0.2 mM, incubating shortly on ice and spinning samples; the resulting supernatant contains large fragments of factories. 'Factory' RNA was now purified from this fraction using 750 μl of Trizol LS (Invitrogen) using the manufacturer's instructions.

For the analysis in Supplementary Figure S1, the workflow was slightly modified to include a nuclear 'run-on'. Intact HUVECs were permeabilized in PB* with 0.025% saponin and RNA polymerases were allowed to extend their transcripts by <40 nt in the presence of a 'run-on' mix (complete PB* plus 100 μM of ATP, CTP and GTP, 0.1 μM UTP, 50 μCi/ml [^{32}P]UTP and $MgCl_2$ to a concentration equimolar to that of all the triphosphates) that allowed subsequent detection of labelled RNA by autoradiography. Then, following treatment with DNase I and caspases as above, released 'factory' complexes were resolved in two-dimensional acrylamide-agarose composite gels using the Mini-Protean 3 system (BioRad); distributions of nascent transcripts was visualized by autoradiography, while those of RNA polymerase II by western blotting using the iBlot transfer system (Invitrogen) and a mouse monoclonal antibody against its RPB1 subunit (7C2; ref. 34)—all as described in ref. (27).

Finally, total RNA was isolated by directly lyzing ∼10^6 HUVECs in 1 ml of Trizol LS (Invitrogen) as described by the manufacturer and poly(A)$^+$ fractions selected using the TrueSeq protocol (Illumina). In all cases, long (>200 nt) and short (<200 nt) RNAs were separated and DNase-treated using the miRNeasy mini kit (Qiagen) as per the manufacturer's instructions.

RNA sequencing (RNA-seq) and data analysis

Long (>200 nt) and short (<200 nt) RNAs, isolated from the different fractions of HUVECs stimulated with TNFα for 0 or 30 min, were depleted of rRNA species using the RiboMinus kit (Epicentre) according to the manufacturer's instructions, reverse-transcribed using random hexamers in a protocol that allows generation of strand-specific libraries (35) and each of two biological replicates sequenced to ≥40 million (50-bp) reads on the Illumina platform. Resulting 'fastq' files were processed using the Bioconductor package for R programming (http://www.bioconductor.org) to assess read quality and produce a read-coverage file after alignment of reads with Bowtie on the GRCh37/hg19 assembly of the human genome (reporting unique hits and allowing no mismatches). The Integrative Genomics Viewer (http://www.broadinstitute.org/igv) was then used to visualize coverage, while rRNAs (representing <10% of mapped reads) were excluded from downstream analysis. Transcripts were assembled using cufflinks (http://cufflinks.cbcb.umd.edu) and annotated using the Genomatix software suite (Genomatix). After annotation, the transcripts were compared to the Genomatix 'primary transcript' database and regrouped into a single primary transcript (when both exonic and intronic transcripts would correspond to the same transcript model) or into spliced transcripts (when only exonic signal was seen).

Pre-miRNA RT-qPCR analysis

Short factory RNA (<200 nt), isolated 0 or 30 min post-stimulation with TNFα as described above, was treated with RQ1 DNase (1 unit of DNase/μg of RNA; 37°C for 45 min; Promega) and precursor miRNAs amplified from nascent RNA using miScript assays (Qiagen) as per manufacturer's instructions on a Rotor-Gene 3000 cycler (Corbett). The presence of single amplimers was confirmed by melting-curve analysis; reactions in which the reverse-transcription step was omitted were performed to ensure amplimers did not result from residual genomic DNA and pre-miRNA levels were normalized relative to those of *RNU6* snRNA.

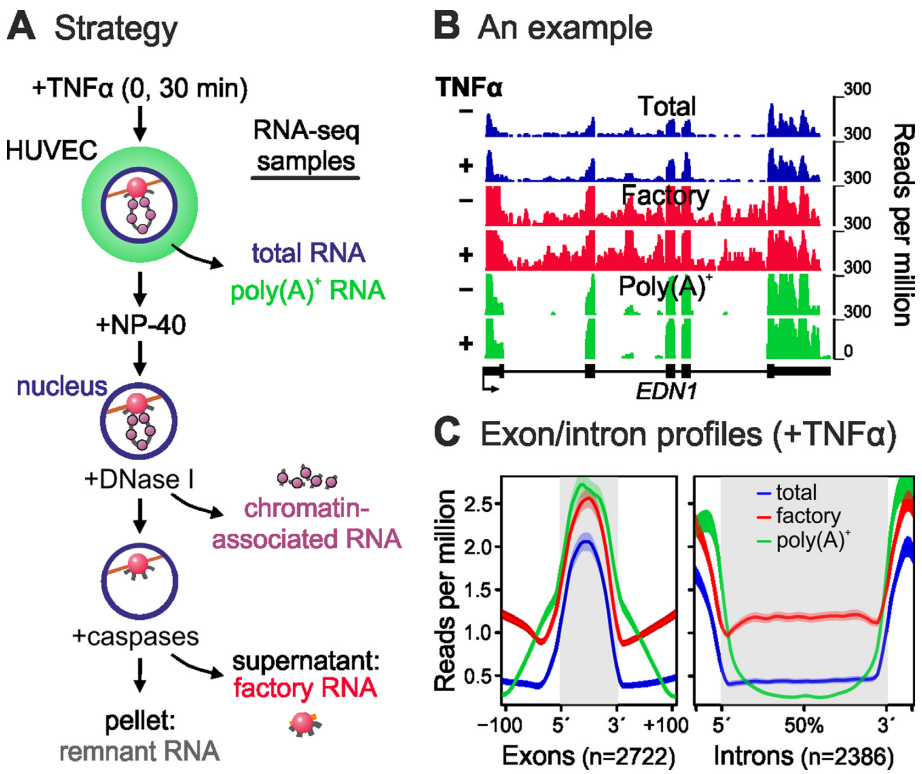


Figure 1. Overview of the approach. (**A**) Strategy. HUVECs were stimulated ± TNFα for 0 or 30 min, and harvested; intact nuclei were isolated in a physiological buffer and digested with DNase I. After spinning, most chromatin (and 'chromatin-associated' RNA) is released into the supernatant. The resulting pellet was treated with a mixture of caspases to detach factories (red) from the underlying sub-structure (brown line). Following centrifugation, the supernatant yields 'factory' RNA, whilst the pellet contains factory remnants and residual chromatin. Total (ribodepleted), and poly(A)$^+$-enriched RNA fractions were also collected. (**B**) Genome browser view of a typical, constitutively-expressed, gene (*EDN1*) illustrating read densities obtained by sequencing total, factory and poly(A)$^+$ RNA. (**C**) Average coverage profiles from two biological replicates along 2722 and 2386 concatenated exons and introns, respectively, belonging to 309 TNFα-responsive genes. Plots also include 100 bp up/downstream of exons/introns.

RESULTS

Sequencing the RNA content of transcription factories

We previously devised a method to biochemically isolate discrete factory (>8 MDa) complexes harbouring the active forms of RNA polymerases I, II or III, and analysed their protein content by mass-spectrometry (27). Here, we have modified this approach and isolated a single fraction containing all three active polymerizing activities (Figure 1A). In brief, HUVECs were harvested, 0 or 30 min after stimulation with TNFα in a 'physiological buffer' (PB*; ref. 25), nuclei isolated after mild treatment with NP-40, washed (to remove most cytoplasmic RNAs) and DNase I-treated to detach most chromatin, but leave essentially all transcriptional activity bound to the nuclear sub-structure. The active polymerases were then released from the sub-structure by digestion with a mixture of caspases. After spinning, the resulting supernatant contained large fragments of factories that still harbour much of the original transcriptional activity (shown using a nuclear 'run-on' assay) and chromatin rich in active marks, like histone 3 trimethylation at lysines 4 and 36, and acetylation at lysine 27 (Supplementary Figure S1). RNA purified from this fraction (hereafter called 'factory' RNA) was compared with that prepared from whole cells both before ('total') and after poly(A)$^+$-selection ('poly(A)$^+$'), as well as that from the detached chromatin ('chromatin'; Figure 1A).

Total, poly(A)$^+$, and factory RNAs were sequenced, to yield ∼40 million reads per sample (replicates were highly reproducible; Supplementary Figure S2A). Following data mapping, factory RNA was found to be rich in intronic unprocessed—and so nascent—transcripts (a typical example is shown in Figure 1B); this verifies that factories are the active sites of transcription. Reassuringly, read profiles closely matched those obtained via ChIP-seq (33) using an antibody targeting active RNA polymerase II (Supplementary Figure S2B). While reads mapping to exons and 5′/3′-UTRs are the prevalent species in rRNA-depleted total and poly(A)$^+$ RNA, factory profiles are dominated by intronic signal (Figure 1C) and by reads mapping to intergenic (non-coding) regions (Supplementary Figure S2C).

Factory RNA-seq allows sensitive detection of changes in gene expression profiles

The transcripts responsible for determining a particular tissue or developmental state are often identified by comparing poly(A)$^+$ RNA profiles. However, TNFα up/downregulates many genes within minutes—well before any changes are seen in poly(A)$^+$ RNA. Therefore, comparing nascent RNA profiles should significantly improve detection of such responsive genes and we exemplify this

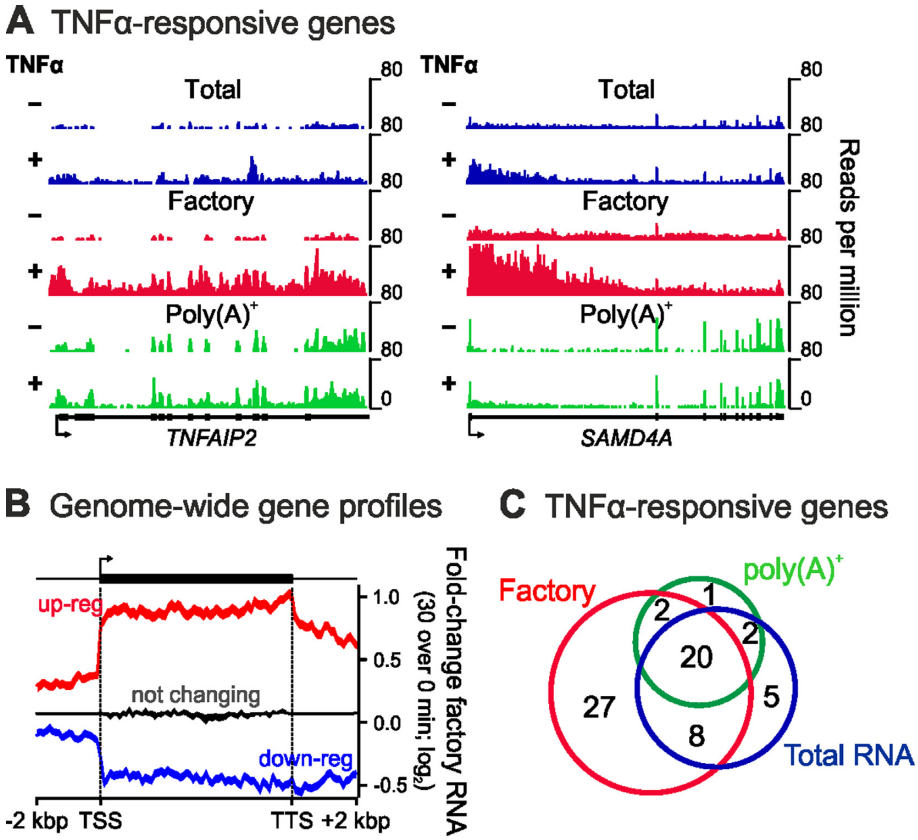


Figure 2. Changing factory RNA levels accurately portray transcriptional changes. (**A**) Examples of RNA-seq coverage along typical TNFα-responsive genes. Genome-browser views illustrating read densities along 11-kbp *TNFAIP2* and 221-kbp *SAMD4A* obtained using total, factory and poly(A)$^+$ RNA at 0 (−TNFα) or 30 min (+TNFα) post-stimulation. (**B**) Genome-wide profiles of factory RNA. Up/downregulated and non-responsive genes were aligned at transcription start (TSS) and termination (TTS) sites, and the $\log_2$ fold changes in factory RNA-seq signal (30- compared to 0-min data ± SD) from two biological replicates was plotted; changes within 2 kbp up/downstream of the TSS/TTS, are also shown. (**C**) Total and poly(A)$^+$ RNA-seq detect ∼50% less genes as significantly upregulated by TNFα compared to those identified using factory RNA.

using two 'early' genes analysed previously—one short (i.e. 11-kbp *TNFAIP2*) and one long (i.e. 221-kbp *SAMD4A*; ref. 36). Poly(A)$^+$ profiles along both genes before and after stimulation were essentially the same (Figure 2A); consequently, both would be classified as non-responsive. However, new reads in factory RNA are seen after 30 min throughout *TNFAIP2* and in the first half of *SAMD4A* (Figure 2A), indicating that both are responsive (there are few new reads in the second half of *SAMD4A* as pioneering polymerases have not yet transcribed that far into the gene) (36). Use of whole-genome factory RNA data enables up/downregulated genes to be readily detected (Figure 2B) and with higher sensitivity than when using poly(A)$^+$ and total RNA (Figure 2C; Supplementary Figure S2D, E and Table S1).

A genome-wide variant of 'nuclear run-on', GRO-seq, allows high-resolution analysis of the positioning of engaged RNA polymerases on genes (8). When applied to AC16 human cardiomyocytes stimulated with TNFα (37), it yields similar profiles to those obtained by factory RNA-seq in HUVECs, albeit with lower read coverage (Supplementary Figure S3A). Results from the two cell types correlate to an extent that allows comparison (Supplementary Figure S3B), as the inflammatory cascade is ubiquitous. Again, factory RNA-seq allows up/downregulated genes to be detected more sensitively (Supplementary Figure S3C–E and Table S1). As seen in various cell types, including AC16 (8,37), divergent transcripts are found at TSSs. These are less obvious in factory RNA profiles, largely because they are masked by the high intronic levels detected in gene bodies; even so, low-level anti-sense species are still detectable (Supplementary Figure S3F and G) and this matches the model of transcriptional initiation on the surface of factories (21).

Factory and chromatin RNAs represent different stages in the processing pathway

Factory RNA from HUVECs is also especially rich in intergenic transcripts (compared to total and poly(A)$^+$ RNA; Figure 3A); therefore, we examined whether it contained many long inter-genic non-coding RNAs (lincRNAs; ref. 5). It did (Figure 3B illustrates two examples), with ∼2000 lincRNAs at each time point being enriched >2-fold relative to levels in total RNA (Figure 3C).

It has been proposed that a significant fraction of nascent RNA associates with chromatin (15) and thus chromatin isolation coupled to RNA sequencing could provide a simple means for uncovering nascent RNAs (38). Therefore, we extracted RNA associated with chromatin as we iso-

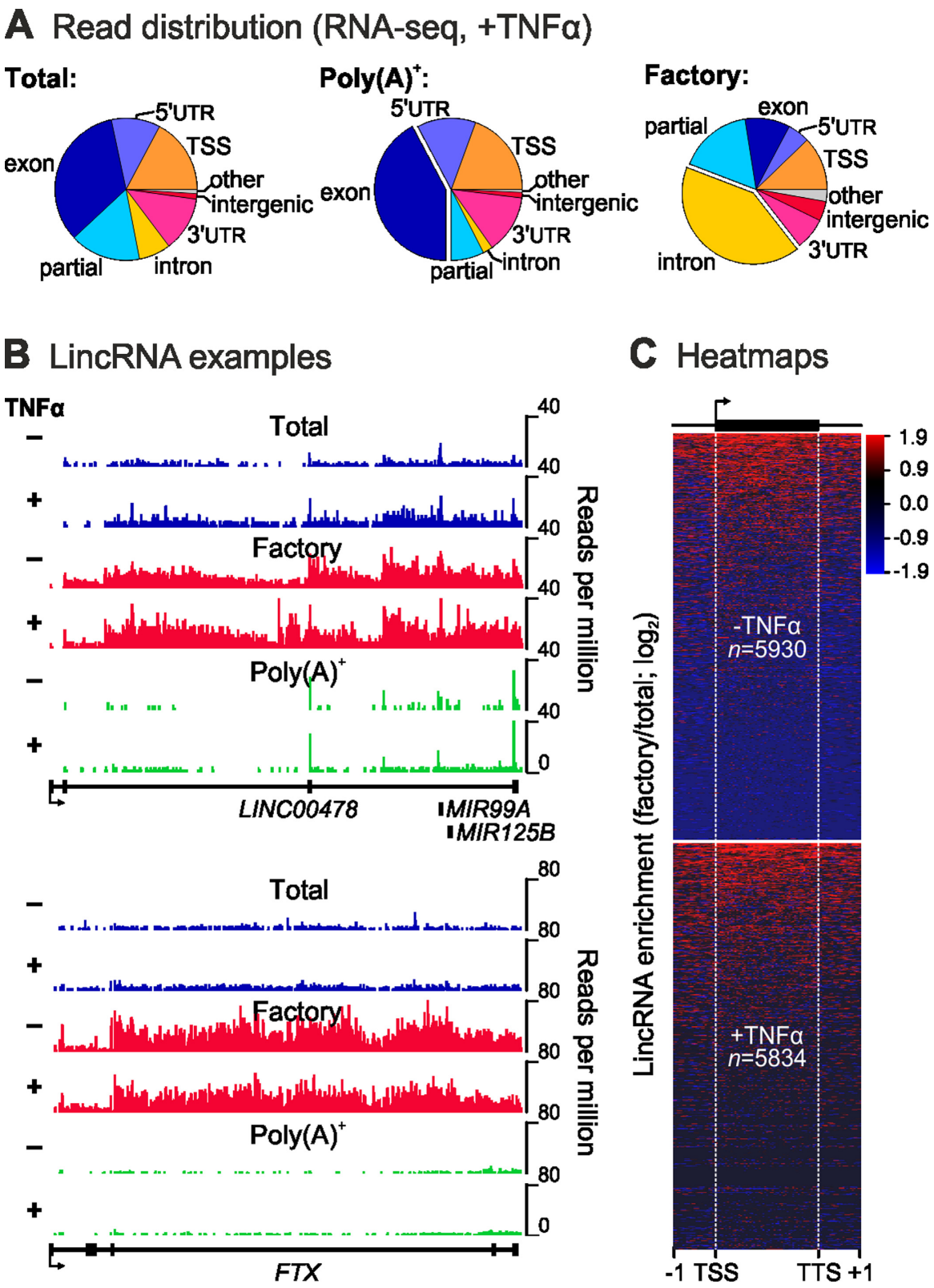


Figure 3. Factory RNA is rich in nascent and non-coding RNA. (**A**) Pie charts depict distributions of reads obtained with the different RNA fractions that map uniquely to genomic features. Reads mapping to exons are the prevalent species in poly(A)$^+$ RNA, while those mapping to introns or across exon-intron boundaries ('partial') predominate in factory RNA. (**B**) Genome-browser views of regions encoding two long inter-genic non-coding RNAs (lincRNAs) enriched in factories, 470-kbp *LINC00478* (which also hosts TNFα-responsive *MIR99A* and non-expressed *MIR125B*) and 265-kbp *FTX*. Tracks show read density (reads per million) of total, factory and poly(A)$^+$ RNA, before (–) or after (+) TNFα stimulation. (**C**) Heatmaps covering all DNA segments encoding lincRNAs detected 0 or 30 min after stimulation (from TSS to TTS, plus 2 kbp up/downstream). Each row in a heatmap represents one locus amongst the 5930 and 5834 detected before (–) and after (+) TNFα stimulation; loci are ranked from most (top) to least (bottom) enriched (compared to total RNA; $\log_2$ scale).

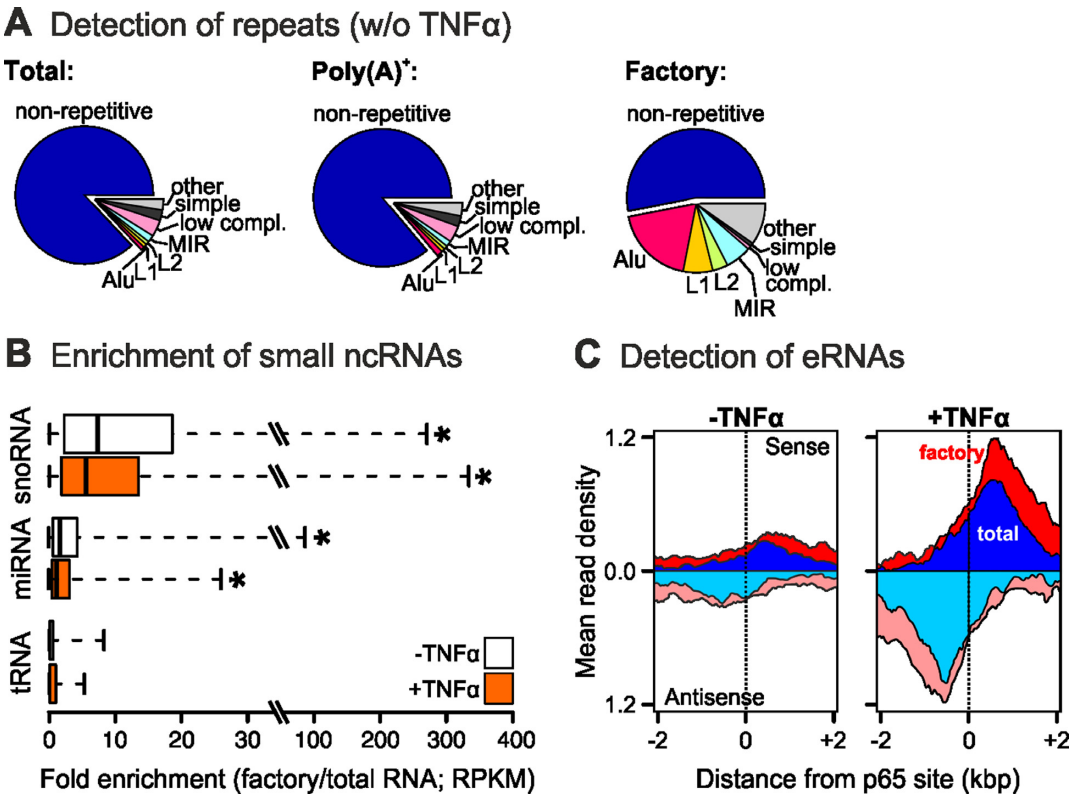


Figure 4. Factory RNA is rich in short non-coding transcripts. (**A**) Pie charts depict distributions of reads mapping to particular genomic features; reads mapping to the non-repetitive genome are the prevalent species in all fractions, but more of those mapping to repeats (colour-coded) are found enriched in factory RNA (data from unstimulated HUVECs). (**B**) Box plots present the fold enrichment of sno- and micro-RNAs (RPKM; factory compared to total RNA levels); tRNAs serve as a negative control. *: *P*-value < 0.05; unpaired Student's *t*-test. (**C**) eRNAs derived from active enhancers are robustly detected using factory RNA. Graphs show mean read densities of sense (red) and anti-sense (pink) factory eRNAs around 1221 active inter-genic enhancers (defined as genomic segments carrying H3K4me1 and H3K27ac marks, and binding NF-κB 30 min post-stimulation; see Supplementary Figure S6A). Coverage using total and chromatin RNA is also shown for both sense (blue/purple) and anti-sense (light blue/light purple) strands.

lated factories (Figure 1A) and sequenced it to obtain 63–65 million reads—roughly the coverage obtained with factory RNA. Chromatin-associated RNA was rich in nascent transcripts, although not as much as factory RNA (Supplementary Figure S4A and B); nevertheless, expression levels in factory and chromatin RNA-seq data correlated well (Supplementary Figure S4C). Notably, chromatin and factories were associated with a distinct subset of lincRNAs and repeat transcripts (Supplementary Figure S4D and E), and the coverage profiles of highly-active genes probably reflect transcripts in discrete processing states (also seen in ref. 15; Supplementary Figure S4F). By analysing the top 5000 introns enriched in either fraction we can attribute this to the differential processing of long introns on chromatin; introns in each group also belong to genes involved in distinct functional pathways (Supplementary Figure S4G).

Short non-coding transcripts are abundant constituents at active sites of transcription

To investigate the contribution of short non-coding RNAs to factories, we selected those of <200 nt from total and factory RNA (again, 0 and 30 min after stimulation) and deep-sequenced them (to ∼50 million reads per sample). Factory RNA was rich in repeat-derived transcripts, especially *Alus* and *L1s* (Figure 4A; Supplementary Figure S5A); this confirms early results obtained by low-throughput sequencing of the residual RNA that remained tightly associated with the nuclear sub-structure of HeLa cells (39). As expected, factory RNA is depleted of tRNAs (Figure 4B), with the exception of tRNA-Asp-GAY, -Lys-AAG and -Glu-GAG (encoded by chromosomes 9, 7 and 13, respectively). It also contains few mature miRNAs (which are also mostly cytoplasmic; Figure 4B). However, there were again exceptions—including miR-1291, -145, -1248 or -452 which were enriched >6-fold (Supplementary Table S2)—and we note that some of these are known to be regulated by TNFα (40). In contrast (but again as predicted), miRNA precursors were produced in factories (Supplementary Figure S5B). Small nucleolar RNAs (snoRNAs) are also most enriched in factories (Figure 4B). This is in agreement with their role in co-transcriptional splicing also during inflammatory signalling (1,41), as well as with the idea of maintaining an 'open' chromatin state, perhaps necessary for efficient association with factories (42). Finally, many transcripts made by RNA polymerase III are also found in factory RNA (details in Supplementary Figure S5C).

Enhancers are now defined as active transcription units that encode short divergent transcripts of ∼500 nt (eRNAs; ref. 12) and we expected to find such eRNAs in factories (21). Of >5500 putative enhancers active 30 min after stim-

ulation (defined by the presence of H3K4me1, H3K27ac and NF-κB; ref. 33), we focused on the subset of 1221 that lay in intergenic regions (to eliminate possible 'contamination' from intronic transcripts, and simplify analysis; Supplementary Figure S6A and Table S3). Factory RNA contained many sense and anti-sense reads from these 1221 sites and these were more enriched than those in total or chromatin RNA; this also holds true when looking at intragenic enhancers or HUVEC-specific 'super-enhancers' (43) (Figure 4C; Supplementary Figure S6C–E). Thus, factories are also the sites where eRNAs are produced and factory RNA-seq efficiently detects them.

DISCUSSION

The human transcriptome is a complex multi-layer catalogue of protein-coding and non-coding RNAs (1), but the relative abundance, biogenesis, sub-cellular localization and functional roles of the non-coding constituents remain controversial (44). Here we describe a fast and cost-effective method for purifying nascent transcripts from their sites of production—transcription factories (Figure 1A; ref. 19). 'Factory RNA-seq' provides excellent read coverage of nascent RNA whether that unit is protein-coding or non-coding, even at the relatively-low sequencing depth used here, largely due to the biochemical enrichment achieved by factory purification. This readout is depleted of exonic signal and allows for a redefinition of the 'active' subset of genes as we detect ∼350 genes characterized by obvious signal in poly(A)$^+$ but barely any reads in factory RNA. We conclude that the corresponding set of genes most probably reflects those that rarely engage in productive transcription but yield long-lived mRNAs (Supplementary Figure S7). Overall, our approach complements others that look at specific fractions of the transcriptome, such as GRO-seq (37) and chromatin-associated RNA-seq (15,38). We anticipate that a combined use of such methods, plus novel *in silico* analysis tools that focus on intronic RNA levels (45), will facilitate increasingly accurate definition of transcriptomes in different cell types.

Transcription factories are multi-MDa, multi-protein complexes (27) and the idea of specific RNAs acting as structural scaffolds therein has long been entertained. Here, we have identified a number of non-coding RNAs that might play such a structural role. About 250 lncRNAs are uniquely enriched in factories compared to the other fractions (Supplementary Figure S4D), some of which are known to be both nuclear and to regulate transcription (46)—including *MALAT1*, *NEAT1* and *GAS5* (47). Factories are also rich in short ncRNAs, including snoRNAs (Figure 4)—which is expected as splicing occurs co-transcriptionally (1,21). They also contain many RNAs copied from repeats (Figure 4A), including those from *Alus* (which are known to engage with the key pro-inflammatory driver, transcription factor NF-κB; ref. 48) and *L1s* (which are abundant, euchromatin-associated transcripts; ref. 49).

In conclusion, this work introduces a general method for analysing RNAs associated with active sites of transcription. Much of this RNA is intronic, but a significant fraction is extra-genic. Faithful cataloguing of when these ncRNAs are both produced upon gene induction and which ones are associated with transcription factories provides a novel approach for dissecting their role in regulating gene expression.

DATA AVAILABILITY

RNA-seq data generated here can be accessed via the EBI Array Express archive under accession number E-MTAB-3020 (summarized in Supplementary Table S4).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We thank Binwei Deng and Lilija Brant for help with isolating factories and members of our labs for helpful discussions.

FUNDING

Biotechnology and Biological Sciences Research Council (to P.R.C.) and Federal Ministry for Education and Research (BMBF) (to K.R.) for the Epigensys consortium via the ERASysBio+/FP7 initiative; DKFZ scholarship (to M.C.-H.); CMMC; UoC Advanced Researcher Grant via the DFG Excellence Initiative and CMMC intramural funding (to A.P.). Funding for open access charge: CMMC (to A.P.).
Conflict of interest statement. None declared.

REFERENCES

1. Tilgner,H., Knowles,D.G., Johnson,R., Davis,C.A., Chakrabortty,S., Djebali,S., Curado,J., Snyder,M., Gingeras,T.R. and Guigó,R. (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.*, **22**, 1616–1625.
2. van Bakel,H., Nislow,C., Blencowe,B.J. and Hughes,T.R. (2010) Most "dark matter" transcripts are associated with known genes. *PLoS Biol.*, **8**, e1000371.
3. Kapranov,P., St Laurent,G., Raz,T., Ozsolak,F., Reynolds,C.P., Sorensen,P.H., Reaman,G., Milos,P., Arceci,R.J., Thompson,J.F. *et al.* (2010) The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. *BMC Biol.*, **8**, 149.
4. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
5. Rinn,J.L. (2014) lncRNAs: linking RNA to chromatin. *Cold Spring Harb. Perspect. Biol.*, **6**, a018614.
6. Dewannieux,M. and Heidmann,T. (2005) LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modeling. *Cytogenet. Genome Res.*, **110**, 35–48.
7. Ichiyanagi,K. (2013) Epigenetic regulation of transcription and possible functions of mammalian short interspersed elements, SINEs. *Genes Genet. Syst.*, **88**, 19–29.
8. 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.
9. Seila,A.C., Calabrese,J.M., Levine,S.S., Yeo,G.W., Rahl,P.B., Flynn,R.A., Young,R.A. and Sharp,P.A. (2008) Divergent transcription from active promoters. *Science*, **322**, 1849–1851.
10. 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. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature*, **465**, 182–187.

11. 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. *et al.* (2013) Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell*, **51**, 310–325.
12. Andersson,R., Gebhard,C., Miguel-Escalada,I., Hoof,I., Bornholdt,J., Boyd,M., Chen,Y., Zhao,X., Schmidl,C., Suzuki,T. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455–461.
13. St Laurent,G., Shtokalo,D., Tackett,M.R., Yang,Z., Eremina,T., Wahlestedt,C., Urcuqui-Inchima,S., Seilheimer,B., McCaffrey,T.A. and Kapranov,P. (2012) Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. *BMC Genomics*, **13**, 504.
14. Churchman,L.S. and Weissman,J.S. (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, **469**, 368–373.
15. Bhatt,D.M., Pandya-Jones,A., Tong,A.J., Barozzi,I., Lissner,M.M., Natoli,G., Black,D.L. and Smale,S.T. (2012) Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell*, **150**, 279–290.
16. Ameur,A., Zaghlool,A., Halvardson,J., Wetterbom,A., Gyllensten,U., Cavelier,L. and Feuk,L. (2011) Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol.*, **18**, 1435–1440.
17. Khodor,Y.L., Rodriguez,J., Abruzzi,K.C., Tang,C.H., Marr,M.T. 2nd and Rosbash,M. (2011) Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev.*, **25**, 2502–2512.
18. Rabani,M., Levin,J.Z., Fan,L., Adiconis,X., Raychowdhury,R., Garber,M., Gnirke,A., Nusbaum,C., Hacohen,N., Friedman,N. *et al.* (2011) Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.*, **29**, 436–442.
19. Windhager,L., Bonfert,T., Burger,K., Ruzsics,Z., Krebs,S., Kaufmann,S., Malterer,G., L'Hernault,A., Schilhabel,M., Schreiber,S. *et al.* (2012) Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Res.*, **22**, 2031–2042.
20. Edelman,L.B. and Fraser,P. (2012) Transcription factories: genetic programming in three dimensions. *Curr. Opin. Genet. Dev.*, **22**, 110–114.
21. Papantonis,A. and Cook,P.R. (2013) Transcription factories: genome organization and gene regulation. *Chem. Rev.*, **113**, 8683–86705.
22. Buckley,M.S. and Lis,J.T. (2014) Imaging RNA Polymerase II transcription sites in living cells. *Curr. Opin. Genet. Dev.*, **25**, 126–130.
23. Sutherland,H. and Bickmore,W.A. (2009) Transcription factories: gene expression in unions? *Nat. Rev. Genet.*, **10**, 457–466.
24. Eskiw,C.H. and Fraser,P. (2011) Ultrastructural study of transcription factories in mouse erythroblasts. *J. Cell Sci.*, **124**, 3676–3683.
25. Cisse,I.I., Izeddin,I., Causse,S.Z., Boudarene,L., Senecal,A., Muresan,L., Dugast-Darzacq,C., Hajj,B., Dahan,M. and Darzacq,X. (2013) Real-time dynamics of RNA polymerase II clustering in live human cells. *Science*, **341**, 664–667.
26. Ghamari,A., van de Corput,M.P., Thongjuea,S., van Cappellen,W.A., van Ijcken,W., van Haren,J., Soler,E., Eick,D., Lenhard,B. and Grosveld,F.G. (2013) In vivo live imaging of RNA polymerase II transcription factories in primary cells. *Genes Dev.*, **27**, 767–777.
27. Melnik,S., Deng,B., Papantonis,A., Baboo,S., Carr,I.M. and Cook,P.R. (2010) The proteomes of transcription factories containing RNA polymerases I, II or III. *Nat. Methods*, **8**, 963–968.
28. Schoenfelder,S., Sexton,T., Chakalova,L., Cope,N.F., Horton,A., Andrews,S., Kurukuti,S., Mitchell,J.A., Umlauf,D., Dimitrova,D.S. *et al.* (2010) Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.*, **42**, 53–61.
29. Li,G., Ruan,X., Auerbach,R.K., Sandhu,K.S., Zheng,M., Wang,P., Poh,H.M., Goh,Y., Lim,J., Zhang,J. *et al.* (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*, **148**, 84–98.
30. Denholtz,M., Bonora,G., Chronis,C., Splinter,E., de Laat,W., Ernst,J., Pellegrini,M. and Plath,K. (2013) Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell*, **13**, 602–616.
31. Li,H.B., Ohno,K., Gui,H. and Pirrotta,V. (2013) Insulators target active genes to transcription factories and polycomb-repressed genes to polycomb bodies. *PLoS Genet.*, **9**, e1003436.
32. Park,S.K., Xiang,Y., Feng,X. and Garrard,W.T. (2014) Pronounced cohabitation of active immunoglobulin genes from three different chromosomes in transcription factories during maximal antibody synthesis. *Genes Dev.*, **28**, 1159–1164.
33. Papantonis,A., Kohro,T., Baboo,S., Larkin,J.D., Deng,B., Short,P., Tsutsumi,S., Taylor,S., Kanki,Y., Kobayashi,M. *et al.* (2012) TNFα signals through specialized factories where responsive coding and miRNA genes are transcribed. *EMBO J.*, **31**, 4404–4414.
34. Besse,S., Vigneron,M., Pichard,E. and Puvion-Dutilleul,F. (1995) 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.
35. Sultan,M., Dökel,S., Amstislavskiy,V., Wuttig,D., Sültmann,H., Lehrach,H. and Yaspo,M.L. (2012) A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and dUTP methods. *Biochem. Biophys. Res. Commun.*, **422**, 643–646.
36. Wada,Y., Ohta,Y., Xu,M., Tsutsumi,S., Minami,T., Inoue,K., Komura,D., Kitakami,J., Oshida,N., Papantonis,A. *et al.* (2009) A wave of nascent transcription on activated human genes. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 18357–18361.
37. Danko,C.G., Hah,N., Luo,X., Martins,A.L., Core,L., Lis,J.T., Siepel,A. and Kraus,W.L. (2012) Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol. Cell*, **50**, 212–222.
38. Weber,C.M., Ramachandran,S. and Henikoff,S. (2014) Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol. Cell*, **53**, 819–830.
39. Jackson,D.A., Bartlett,J. and Cook,P.R. (1996) Sequences attaching loops of nuclear and mitochondrial DNA to underlying structures in human cells: the role of transcription units. *Nucleic Acids Res.*, **24**, 1212–1219.
40. Suárez,Y., Wang,C., Manes,T.D. and Pober,J.S. (2010) TNF-induced microRNAs regulate TNF-induced expression of E-selectin and intercellular adhesion molecule-1 on human endothelial cells: feedback control of inflammation. *J. Immunol.*, **184**, 21–25.
41. Hao,S. and Baltimore,D. (2013) RNA splicing regulates the temporal order of TNF-induced gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 11934–11939.
42. Schubert,T., Pusch,M.C., Diermeier,S., Benes,V., Kremmer,E., Imhof,A. and Längst,G. (2012) Df31 protein and snoRNAs maintain accessible higher-order structures of chromatin. *Mol. Cell*, **48**, 434–444.
43. Hnisz,D., Abraham,B.J., Lee,T.I., Lau,A., Saint-André,V., Sigova,A.A., Hoke,H.A. and Young,R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell*, **155**, 934–947.
44. Clark,M.B., Choudhary,A., Smith,M.A., Taft,R.J. and Mattick,J.S. (2013) The dark matter rises: the expanding world of regulatory RNAs. *Essays Biochem.*, **54**, 1–16.
45. Madsen,J.G., Schmidt,S.F., Larsen,B.D., Loft,A., Nielsen,R. and Mandrup,S. (2015) iRNA-seq: computational method for genome-wide assessment of acute transcriptional regulation from total RNA-seq data. *Nucleic Acids Res.*, **43** ,e40.
46. Vance,K.W. and Ponting,C.P. (2014) Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet.*, **30**, 348–355.
47. West,J.A., Davis,C.P., Sunwoo,H., Simon,M.D., Sadreyev,R.I., Wang,P.I., Tolstorukov,M.Y. and Kingston,R.E. (2014) The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell*, **55**, 791–802.
48. Antonaki,A., Demetriades,C., Polyzos,A., Banos,A., Vatsellas,G., Lavigne,M.D., Apostolou,E., Mantouvalou,E., Papadopoulou,D., Mosialos,G. *et al.* (2011) Genomic analysis reveals a novel nuclear factor-κB (NF-κB)-binding site in Alu-repetitive elements. *J. Biol. Chem.*, **286**, 38768–38782.
49. Hall,L.L., Carone,D.M., Gomez,A.V., Kolpa,H.J., Byron,M., Mehta,N., Fackelmayer,F.O. and Lawrence,J.B. (2014) Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell*, **156**, 907–919.

SUPPLEMENTARY DATA

Dissecting the nascent human transcriptome by analyzing the RNA content of transcription factories

Maïwen Caudron-Herger[1], Peter R. Cook[2], Karsten Rippe[1], and Argyris Papantonis[2,3,*]

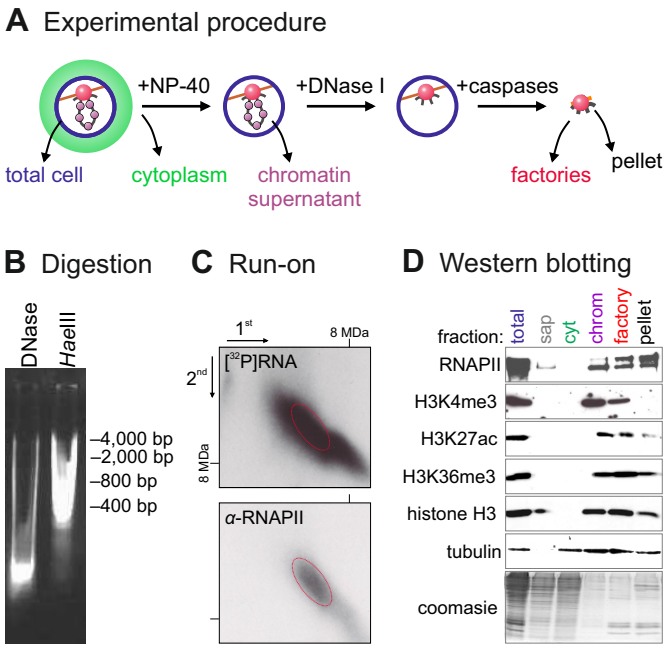
[1]Deutsches Krebsforschungszentrum & BioQuant, D-69120 Heidelberg, Germany

[2]Sir William Dunn School of Pathology, University of Oxford, OX1 3RE Oxford, UK

[3]Center for Molecular Medicine, University of Cologne, D-50931 Cologne, Germany

*To whom correspondence should be addressed: Tel: +49-221-478-96987; Fax: +49-221-478-4833; E-mail: argyris.papantonis@uni-koeln.de

Supplementary Figure S1



Supplementary Figure S1. **Verifying transcriptional activity of "factory" fractions.**

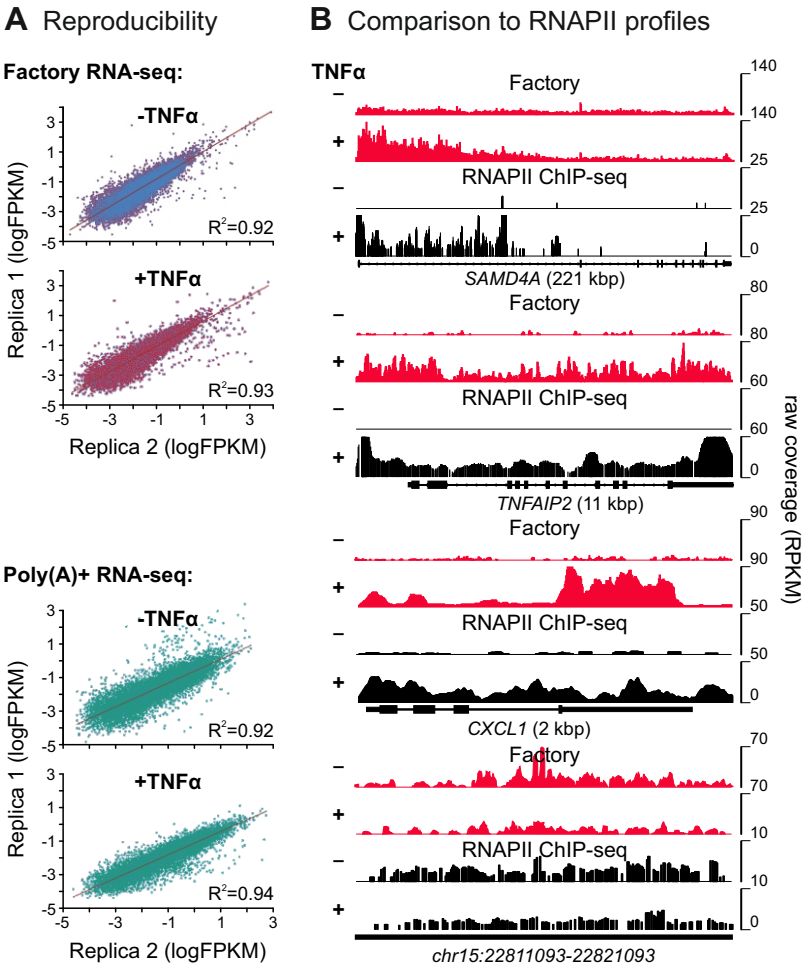
(***A***) Overview of the experimental procedure. HUVECs were stimulated with TNFα for 0 or 30 min, and harvested; now, intact nuclei were isolated in a physiological buffer, digested with DNase I, and spun. The resulting pellet was treated with caspases -6, -8, -9, and -10, which detach factories (*red*) from the underlying sub-structure (*brown line*). Following centrifugation, "factory" RNA is purified from the supernatant; the pellet contains remnants (black) of factories and residual chromatin.

(***B***) Comparison of chromatin fragments remaining associated with factories after digesting nuclei with DNase I (10 units/10^7 cells) or *Hae*III (800 units/10^7 cells), purifying DNA, and electrophoretic separation on a 1.5% agarose gel. An image of a stained gel is shown (sizes of markers indicated).

(***C***) 2D electrophoresis of isolated factory fractions. Half the sample is used in a nuclear run-on in the presence of [^{32}P]UTP (*top*), and the other half for immuno-blotting using an antibody against the largest subunit, RPB1, of RNA polymerase II (*bottom*). The dotted ovals denote regions of the gel containing large fragments of factories rich in RNA polymerase II (the position of the largest size-marker of 8 MDa that is available is indicated for each dimension).

(***D***) Histone marks in the different fractions. The fractions indicated in panel A – plus a saponin-extracted fraction ("sap"; derived after lysing cells in the saponin used prior to the run-on shown in panel *C*) – were analyzed by SDS-PAGE gel electrophoresis and immuno-blotting using antibodies against RNA polymerase II (as in panel *C*) or the "active" histone marks indicated; antibodies targeting histone H3 and tubulin – and the Coomassie-stained gel – provide loading controls.

Supplementary Figure S2



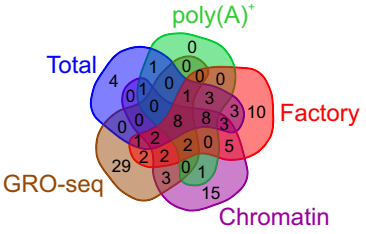
C Comparison of RNA-seq datasets

	Total		Poly(A)⁺		Factory	
TNFα	−	+	−	+	−	+
% reads mapping to:						
TSS	32	28	32	34	16	17
5' UTR	21	18	22	23	7	7
Exon	**64**	**50**	**61**	**65**	13	15
Intron	8	12	5	4	**59**	**60**
Partial	26	26	13	20	24	23
3' UTR	25	20	24	23	8	8
Intergenic	2	2	2	2	**4**	**5**
Other	4	5	4	3	4	4

D Detection sensitivity

	30- compared to 0-min TNFα	
	↑up-reg	↓down-reg
Total	44 / 89	14 / 3
Poly(A)⁺	53 / 45	3 / 6
Factory	135 / 146	17 / 37
GRO-seq	68 / 42	0 / 2
Chromatin	79	18

E Top TNFα-responsive genes



Supplementary Figure S2. **Reproducibility and features of RNA-seq.**

(***A***) Reproducibility of (**i**) factory and (**ii**) poly(A)$^+$ RNA-seq biological replicates obtained from HUVECs treated ±TNFα. Levels of individual transcripts (logFPKM; fragments per kilobase per million) in two replicas are plotted, and Spearman's correlation coefficients (R^2) were calculated.

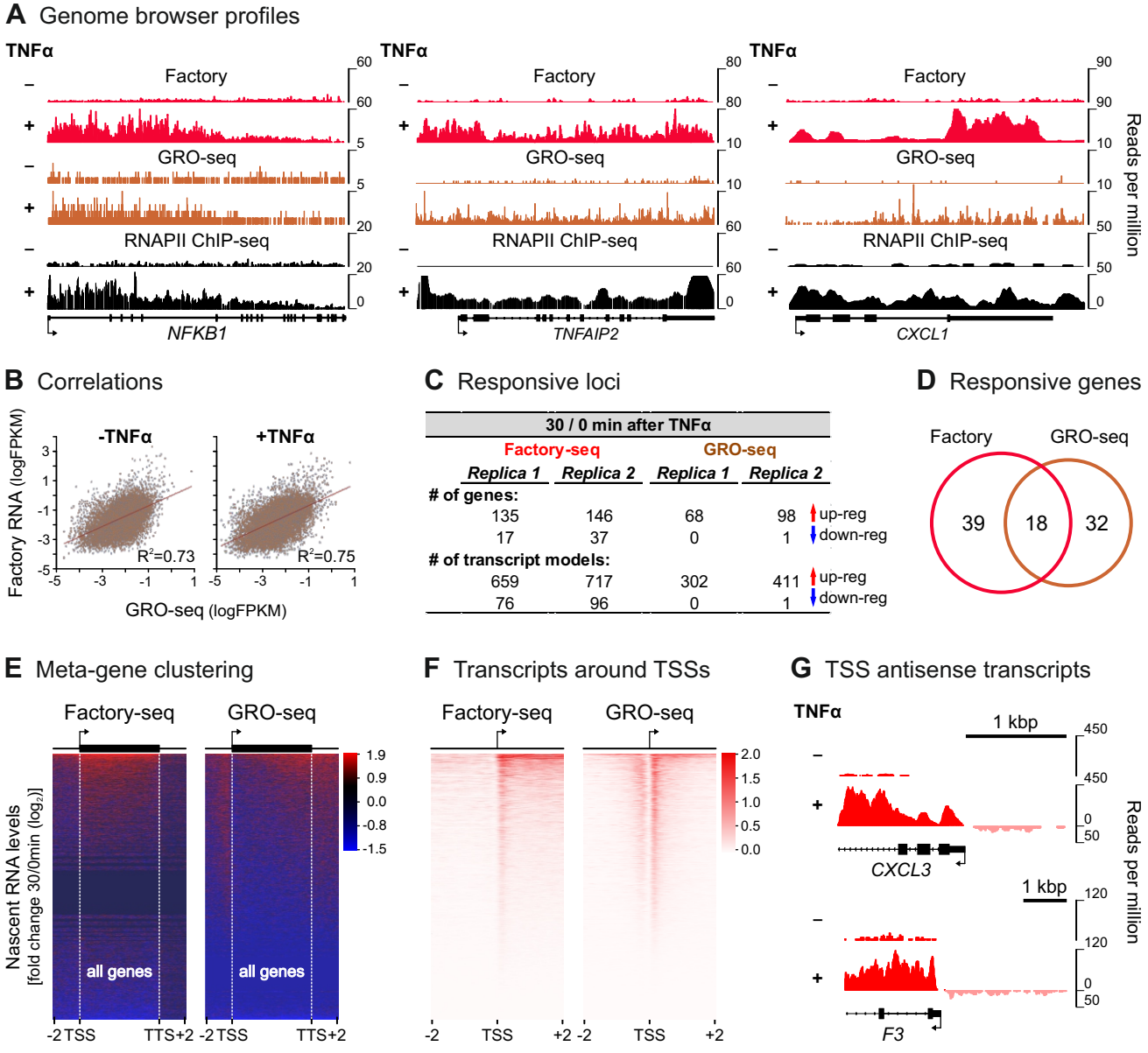
(***B***) Comparison of factory RNA and RNA polymerase II binding-profiles (from ChIP-seq data of ref. 33) obtained before (-) or after (+) TNFα stimulation. Browser views (*y*-axes show RPKM values), include TNFα-responsive *SAMD4A*, *TNFAIP2*, and *CXCL1*, and a locus on chromosome 15 encoding a novel lincRNA.

(***C***) Percentages of reads obtained using the different RNA fractions that map to selected genomic features. Values particularly enriched in a fraction are highlighted; e.g., factory RNA is rich in long inter-genic non-coding RNAs (lincRNAs) and transcripts copied from repeats.

(***D***) Comparison of the number of up-/down-regulated genes detected 30 min post-stimulation, using the different RNA-seq approaches. Note that "chromatin" RNA-seq data come from a single biological replicate.

(***E***) A Venn diagram illustrating the overlap in detection of the top TNFα-responsive genes using each of the different RNA-seq approaches (see also **Supplementary Table S1**).

Supplementary Figure S3



Supplementary Figure S3. Comparison of factory RNA-seq to GRO-seq. Factory RNA-seq data was obtained here using HUVECs treated ±TNFα (10 ng/μl) for 0 or 30 min; the GRO-seq data is from ref. 37 who used AC16 cells stimulated ±TNFα (25 ng/μl) for the same times.

(***A***) Genome-browser views illustrating read profiles (RPKM) of factory RNA-seq to global nuclear run-on (GRO-seq) along TNFα-responsive *SAMD4A*, *TNFAIP2*, and *CXCL1* genes; RNA polymerase II binding-profiles (using ChIP-seq data from ref. 33) are included for comparison.

(***B***) Correlation of GRO-seq and factory RNA-seq data. Expression levels of individual transcripts (logFPKM) from two replicas are plotted and Spearman's correlation coefficients (R^2) were calculated.

(***C***) Numbers of up-/down-regulated genes (*top*) or transcript models (*bottom*) detected 30 min post-stimulation using factory RNA-seq data from two biological replicates is compared to those returned using GRO-seq.

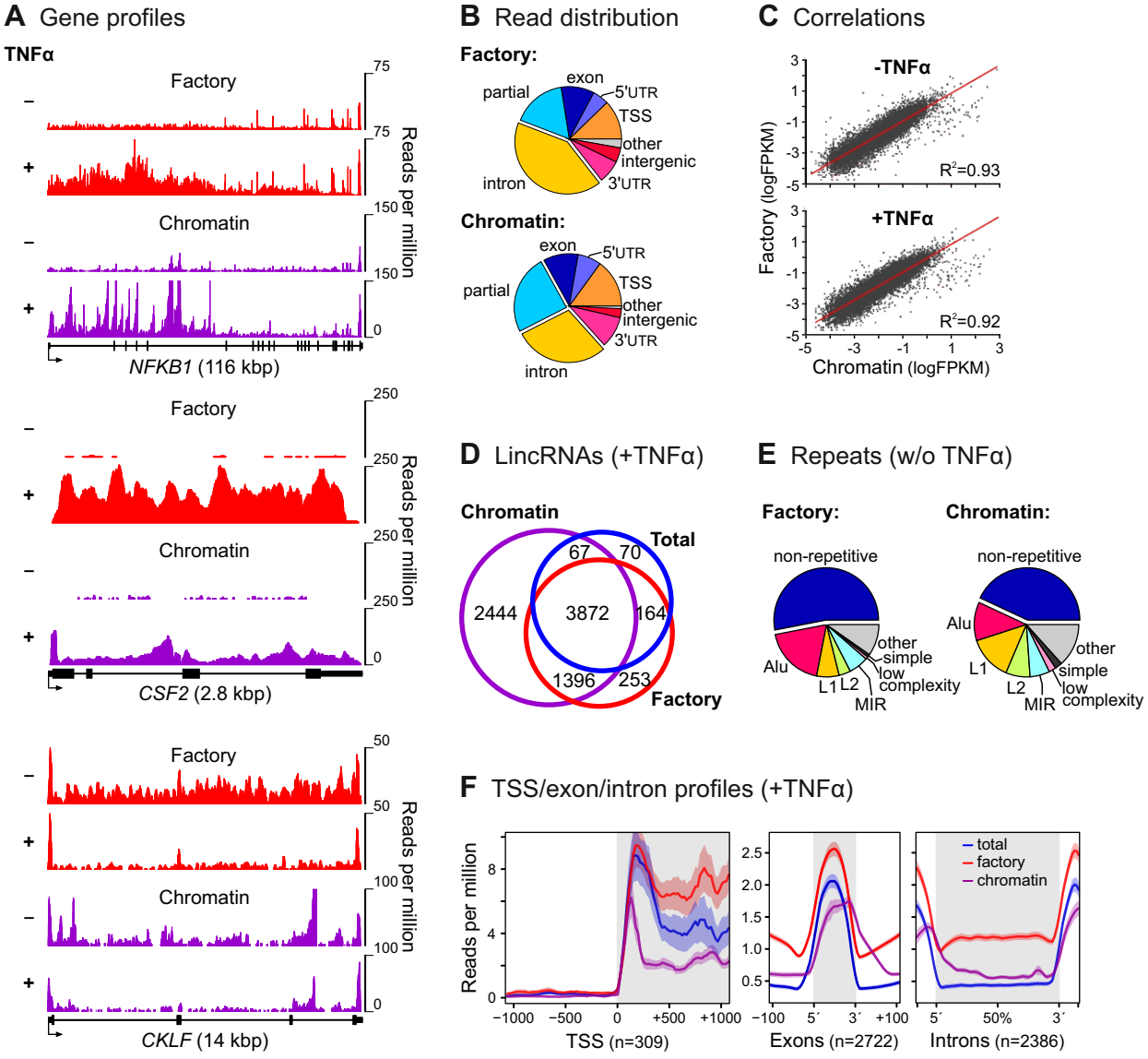
(***D***) Overlap of detected TNFα-responsive gene sets. Genes found to be significantly up-regulated (>2-fold) using factory RNA-seq replicates (*red*) include 36% of those detected in GRO-seq (*brown*) ones.

(***E***) Heatmaps covering gene bodies (from TSS to TTS plus 2 kbp up-/down-stream), of all up-regulated genes (30 compared to 0 min) for factory RNA-seq and GRO-seq are shown. Each row in a heatmap represents a single gene of 20,193 genes, and genes are ranked from most (*top*) to least (*bottom*) responsive.

(***F***) Identification of TSS-associated transcripts. All expressed genes were aligned at the TSSs ± 2 kbp, and signal from factory RNA-seq and GRO-seq plotted.

(***G***) Examples of TSS-associated transcripts in factory RNA. Browser views of sense (*red*) and anti-sense (*pink*) transcripts around the TSS of TNFα-responsive *CXCL3* and *F3*.

Supplementary Figure S4



G Longer introns are processed on chromatin

Protein Class	# in DB	# in sample	expected	Fold Enriched	P value
RNA binding protein	983	186	103.6	1.8	8.98E-12
nucleic acid binding	2524	367	266.01	1.38	4.15E-08
RNA helicase	89	30	9.38	3.2	1.23E-05
helicase	146	40	15.39	2.6	2.21E-05
translation factor	113	31	11.91	2.6	5.46E-04
translation initiation factor	78	24	8.22	2.92	1.12E-03
transferase	1313	178	123.63	1.44	2.37E-04
enzyme modulator	1439	191	135.49	1.41	3.57E-04
extracellular matrix protein	466	76	43.88	1.73	1.08E-03
protein kinase	404	67	38.04	1.76	2.27E-03
kinase	545	84	51.32	1.64	2.73E-03
hydrolase	1654	208	155.74	1.34	3.63E-03

Supplementary Figure S4. Comparison of factory and chromatin-associated RNA.

(***A***) Genome-browser views illustrating read profiles (reads per million) obtained using factory or chromatin RNA-seq before (-) or after (+) TNFα stimulation along up-regulated *NFKB1* and *CSF2*, and down-regulated *CKLF*.

(***B***) Pie charts of read distribution uniquely mapped to particular genomic features (using data from unstimulated HUVECs). Reads mapping to introns and exon/intron boundaries ("partial") predominate in both libraries, but the former were more enriched in factory RNA.

(***C***) Correlation between RNA-seq data using factory and chromatin RNA. Expression levels of individual transcripts (logFPKM), obtained before (-) or after (+) TNFα stimulation, are plotted, and Spearman's correlation coefficients (R^2) were calculated.

(***D***) Detection of enriched lincRNAs. RNA-seq data obtained using total, factory, and chromatin RNA were mined for lincRNAs and the Venn diagram depicts unique and shared ones.

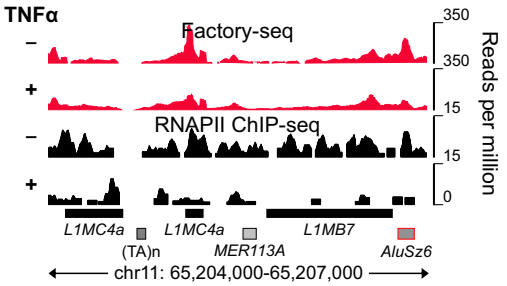
(***E***) Pie charts depict reads mapping to the non-repetitive genome compared to those mapping to repeats (*colour-coded*); both factory and chromatin RNA fractions were rich in repeat-derived transcripts.

(***F***) Average coverage profiles, from the different RNA-seq approaches, ±1 kbp around the TSS and along the 2722 and 2386 concatenated exons and introns, respectively, that belong to 309 TNFα-responsive genes. Plots also include 100 bp up-/down-stream of exons/introns.

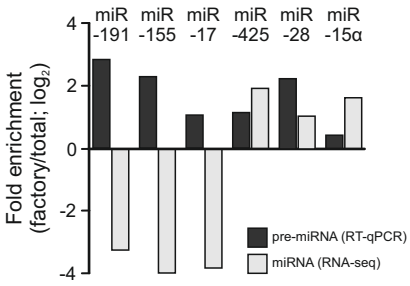
(***G***) Longer introns, associated with particular GO terms, are processed on chromatin. *Left*: The number of introns exhibiting different lengths of the top 5,000 most enriched (≥5.5-fold) in factory (*red*) or chromatin RNA (*purple*) is shown; chromatin RNA is especially rich for longer introns (average and median lengths: 15.8 and 1.6 kbp, respectively, compared to 2.37 and 0.87 kbp for factory-retained ones). *Right*: Protein class decomposition, using Gene Ontology (GO) terms from the Panther database (http://www.pantherdb.org/), of the genes carrying the introns most enriched in factory (*red*) or chromatin RNA (*purple*).

Supplementary Figure S5

A Example of identified repeats



B pre-miRNA factory enrichment



C Detection of RNAPIII-transcribed sequences

	total 0 min	total 30 min	poly(A) 0 min	poly(A) 30 min	chrom 0 min	chrom 30 min	factory 0 min	factory 30 min	GRO 0 min	GRO 30 min
5S rRNA	4	6	5	5	4	4	7	7	11	10
7SK	4	3	1	1	**22**	**14**	4	2	2	1
7SL	10	11	4	4	9	4	14	10	7	5
Alu	1278	1397	954	836	1499	1471	**1893**	**2049**	607	460
SINE	32	35	15	18	34	44	37	33	6	7
tRNA	71	90	11	9	29	30	33	36	**190**	**219**
RNU6	4	5	2	0	14	10	6	9	2	2

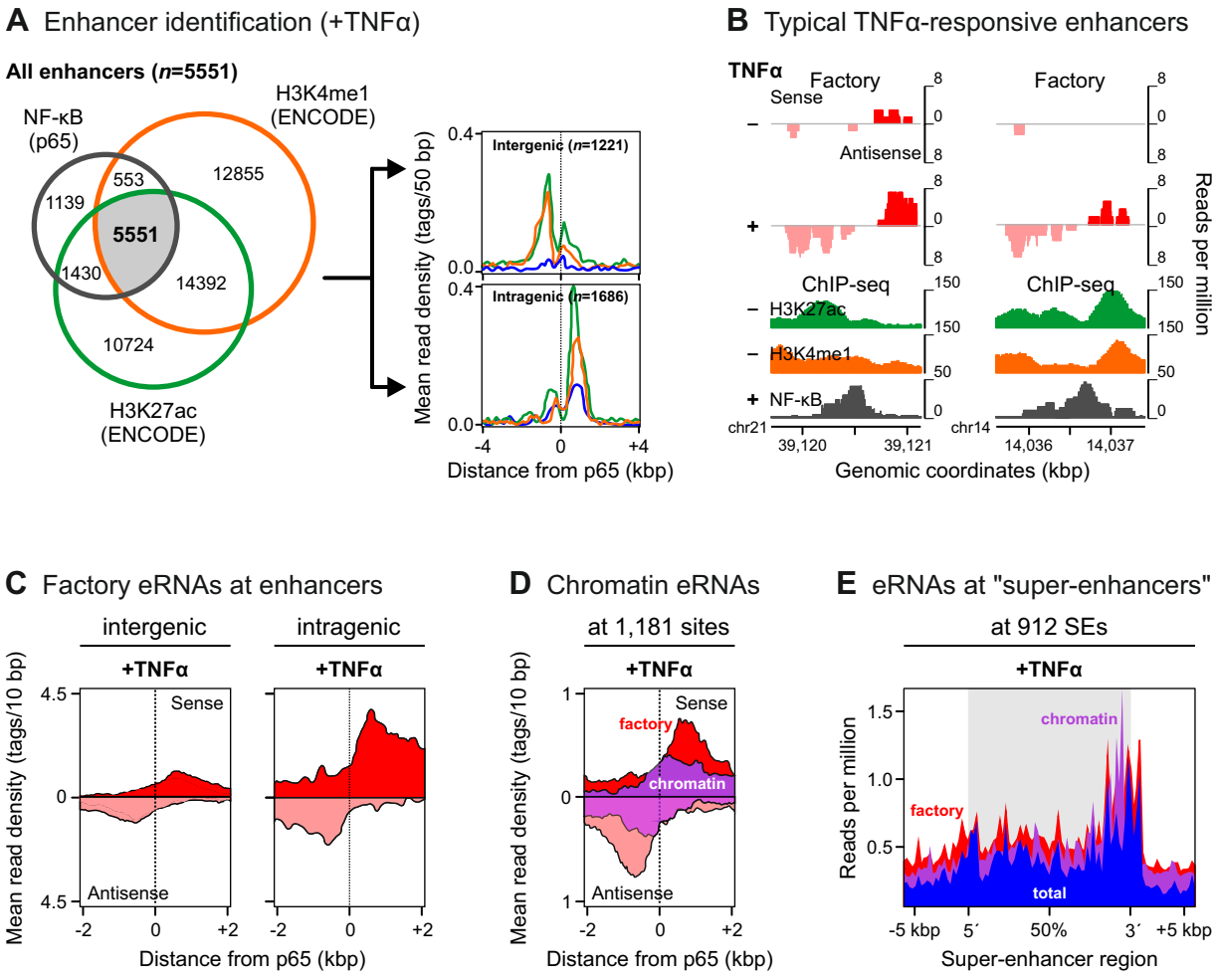
Supplementary Figure S5. Detection of repeats and pre-miRNAs in factory RNA.

(***A***) Examples. Typical genome-browser views illustrating read profiles (RPKM) obtained using factory RNA-seq and ChIP-seq (with an antibody targeting RNA polymerase II; from ref. 33) along a repeat-rich region on chromosome 11.

(***B***) Verification of the enrichment of pre-miRNAs in factory RNA. Enrichments (assessed by RT-qPCR) of six nascent pre-miRNAs (*black*) in factory RNA (comparison with total; $\log_2$ fold-change), 30 min after TNFα stimulation; enrichments are presented alongside those seen for the respective mature miRNAs using RNA-seq (*light grey*).

(***B***) Detection of sequences transcribed by RNA polymerase III. The number of genes and repeat elements known to be transcribed by RNAPIII as mined from the mapped reads of each RNA-seq approach (also using the Repeat Masker annotation; www.repeatmasker.org/). Values deviating significantly from the average detection levels are shown in bold.

Supplementary Figure S6



Supplementary Figure S6. Detection of eRNAs produced in response to TNFα signaling.

(***A***) Identifying active enhancers. Enhancer genomic segments were required to be marked by mono-methylation at lysine 4 and acetylation of lysine 27 of histone H3 (ENCODE ChIP-seq; www.encodeproject.org), and to bind NF-κB (p65) 30 min post-TNFα (ChIP-seq data from ref. 33); this gave 5,551 putative active enhancer regions (Venn diagram; *left*). These positions were further filtered for location in two ways. First, they had to be ≥1 kbp from an annotated TSS and not to overlap known genes; this gave 1,221 putatively-active intergenic enhancers (see **Supplementary Table S3**). Second, they had to lie within an annotated gene body and ≥1 kbp from annotated TSSs; this gave 1,686 putatively-active intragenic enhancers. Histone modification and NF-κB binding profiles for ±4 kbp around each of these sites are illustrated using mean read-density profiles. Both groups were de-enriched tri-methylation at lysine 4 of histone H3 (H3K4me3; *blue line*).

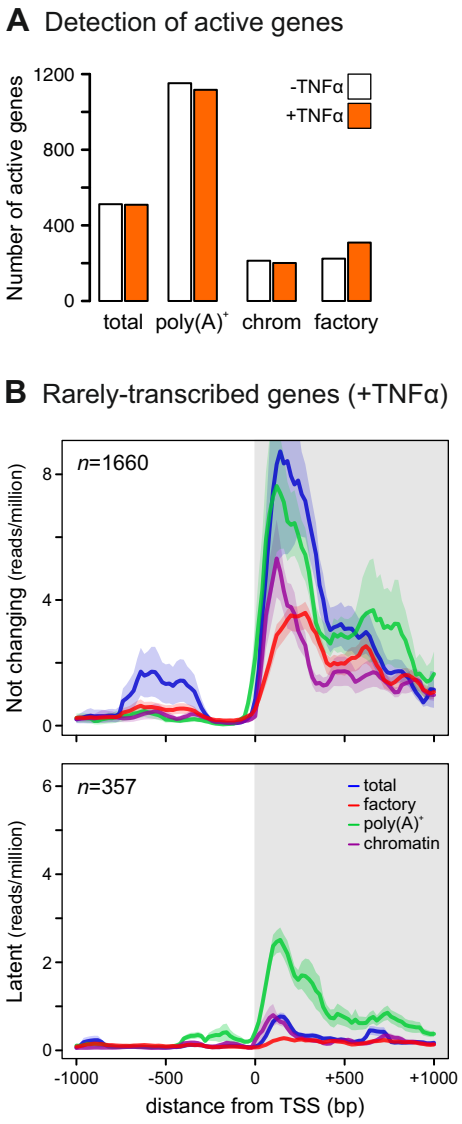
(***B***) Genome-browser views around two typical intergenic enhancers (on chromosomes 21 and 14) illustrating factory RNA (sense: *red*; anti-sense: *pink*), and ChIP-seq profiles for H3K27ac (*green*), H3K4me1 (*orange*), and NF-κB (*grey*).

(***C***) Comparison of eRNA profiles using strand-specific factory RNA-seq obtained 30 min after TNFα stimulation. Mean read densities of sense (*red*) and anti-sense (*pink*) eRNAs seen in factory RNA around 1,221 intergenic and 1,686 intragenic enhancers (defined as in panel *A*) are plotted for ±2 kbp around NF-κB (p65) binding sites.

(***D***) Comparison of eRNA profiles using strand-specific factory and chromatin RNA-seq data obtained 30 min after TNFα stimulation. Mean read densities of sense and anti-sense eRNAs seen at 1,181 enhancers detected in both datasets are plotted for ±2 kbp around NF-κB (p65) binding sites.

(***E***) Comparison of eRNA profiles at "super-enhancers" (SEs) using total (ribodepleted), factory, and chromatin RNA-seq data obtained 30 min after TNFα stimulation. Mean read densities of eRNA signal seen at 912 HUVEC-specific SEs is plotted after aligning them at their 5' and 3' ends; signal 5 kbp up-/down-stream SEs is also shown. SE genomic coordinates were obtained via the "dbSUPER" database (http://bioinfo.au.tsinghua.edu.cn/dbsuper/index.php).

Supplementary Figure S7



Supplementary Figure S7. Detection of active genes in HUVECs.

(***A***) The number of genes with normalized expression levels of >1 (i.e., >0.1 FPKM) in the different RNA-seq libraries before (*white*) or after TNFα stimulation (*orange*).

(***B***) Average coverage profiles, from the different RNA-seq approaches, ±1 kbp around the TSS of 1660 and 357 non-TNFα-responsive genes that do ("not-changing") or do not have ("latent") factory RNA-seq signal.

Supplementary Table S1. **The top 50 TNFα-responsive gene per dataset.** The top (up to 50) TNFα-responsive genes identified at 30 min post-stimulation using each of four RNA-seq approaches, and shared between biological replicates, are listed alongside the mean (log_2) fold change and average 0- and 30-min normalized expression levels (reads per kilobase per million). [Provided as a .xls file.]

Supplementary Table S2. Enrichment of short non-coding RNAs. Full lists of sno- and micro-RNAs detected using total (ribodepleted) and factory "short" (<200 nt) RNA-seq are shown. In each the genomic location (chromosome, start, end, strand), gene name, raw and normalized expression levels and fold-change rations are shown. [Provided as a .xls file.]

Supplementary Table S3. A catalogue of NFκB-bound intergenic enhancers. Chromosomal locations of 1,221 TNFα-induced intergenic enhancers in HUVECs. [Provided as a .BED file.]

Supplementary Table S4. A summary of all RNA-seq generated. All datasets generated here are listed with details on "long" (>200 nt) or "short" RNA (<200 nt) sub-selection, number of generated reads and their length, strand-specificity, mapping efficiency, and stimulation times.

RNA type	Sample type	generated reads	read length (bp)	strand-specific	uniquely mapped reads	min after TNFα
Total	Long / 1st replica	38785582	36	N	14156051	0
	Long / 2nd replica	39229692	100	Y	18192957	0
	Long / 1st replica	36461434	36	N	17792420	30
	Long / 2nd replica	62519832	100	Y	28854334	30
	Short / 1st replica	120305889	50	Y	85307536	0
	Short / 1st replica	109027012	50	Y	65651611	30
Poly(A)+	Long / 1st replica	33397101	36	N	22079154	0
	Long / 2nd replica	74133132	50	Y	40704898	0
	Long / 1st replica	34870984	36	N	22875836	30
	Long / 2nd replica	50965297	50	Y	33215626	30
Factory	Long / 1st replica	34158749	36	N	25892711	0
	Long / 2nd replica	49655048	100	Y	28496017	0
	Long / 1st replica	35703199	36	N	24461627	30
	Long / 2nd replica	57382481	100	Y	30839433	30
	Short / 1st replica	52984430	50	Y	48215918	0
	Short / 1st replica	48044506	50	Y	40579351	30
Chromatin	Long / 1st replica	82488420	50	Y	64722271	0
	Long / 1st replica	76065770	50	Y	62925618	30