Exploring the world of non-coding genes in stem cells and autoimmunity.

Messemaker, T.C.

Chapter 5

A novel long non-coding RNA in the rheumatoid arthritis risk locus *TRAF1-C5* influences C5 mRNA levels

Genes Immun. 2016 Mar;17(2):85-92. doi: 10.1038/gene.2015.54.

Messemaker TC, Frank-Bertoncelj M, Marques RB, Adriaans A, Bakker AM, Daha N, Gay S, Huizinga TW, Toes RE, Mikkers HM, Kurreeman F

Abstract

Long non-coding RNAs (lncRNAs) can regulate the transcript levels of genes in the same genomic region. These locally acting lncRNAs have been found deregulated in human disease and some have been shown to harbour quantitative trait loci (eQTLs) in autoimmune diseases. However, lncRNAs linked to the transcription of candidate risk genes in loci associated to rheumatoid arthritis (RA) have not yet been identified. The *TRAF1* and *C5* risk locus shows evidence of multiple eQTLs and transcription of intergenic non-coding sequences. Here, we identified a non-coding transcript (C5T1lncRNA) starting in the 3' untranslated region (UTR) of *C5*. RA-relevant cell types express *C5T1lncRNA* and RNA levels are further enhanced by specific immune stimuli. *C5T1lncRNA* is expressed predominantly in the nucleus and its expression correlates positively with C5 mRNA in various tissues (*P*=0.001) and in peripheral blood mononuclear cells (*P*=0.02) indicating transcriptional co-regulation. Knockdown results in a concurrent decrease in C5 mRNA levels but not of other neighbouring genes. Overall, our data show the identification of a novel lncRNA *C5T1lncRNA* that is fully located in the associated region and influences transcript levels of *C5*, a gene previously linked to RA pathogenesis.

Introduction

A large proportion of the mammalian genome is transcribed into non-coding RNAs (ncRNAs), which are thought to be equally important for normal development and physiology as coding genes and have been found deregulated in human disease.[1] ncRNAs with a length of >200 nucleotides are termed long non-coding RNA genes (lncRNAs).[2] The identification of lncRNA genes in the human genome has increased over the years, and was recently estimated at, minimally, 50 000 genes.[3] A variety of mechanisms by which lncRNAs positively or negatively regulate coding genes are known although our understanding is likely going to increase in the coming years.[4,5] These include miRNA sponges,[6] recruitment of proteins that directly enhance or interfere with transcription,[7] and recruitment of chromatin modifiers like polycomb repressor complexes 1 (PRC1)[8] and 2 (PRC2),[9] histone demethylases (LSD1),[10] and methyl transferases (G9a).[11,12]

As certain locally acting lncRNAs have been shown to control transcript levels of neighbouring genes,[13] deregulation of the regulatory lncRNAs may contribute to

disease by affecting the expression of adjacent coding genes. However, in spite of the vast amount of lncRNA genes in the human genome, the percentage of lncRNAs directly linked to disease is extraordinarily low in comparison to coding genes.[1] Genome wide association studies (GWAS) have identified tens of thousands of genomic loci that are associated to complex, multifactorial disorders.[14] One example of a ncRNA disease gene is the lncRNA *ANRIL*. *ANRIL* is located in a 50-kb genetic risk region of atherosclerosis and regulates transcription *in trans* by recruiting PRC2 to genome-wide ALU elements affecting various atherogenic cell functions as cell proliferation, cell adhesion and apoptosis.[15]

We and others have previously identified the *TRAF1-C5* region associated with rheumatoid arthritis (RA), a common disease of autoimmune origin.[16,17] The *TRAF1-C5* region of association is relatively large containing multiple immune-related candidate genes. *TRAF1* and *C5* are excellent, evidence based, candidate RA disease genes. TRAF1 negatively regulates TNFa signalling,[18] which is a successful therapeutic target for RA.[19] Increased levels of C5 have been found in inflamed joints of RA patients,[20] and C5-deficient mice are resistant to the development of collagen-induced arthritis.[21] In addition, both genes have a role in innate immune responses.[22,23] Deciphering disease-associated risk loci to reveal causal genes and mechanisms has remained a challenge with the majority of disease-associated single-nucleotide polymorphisms located in non-coding regions.[24,25] However, recent studies have reported that such regions may contain regulatory RNAs.[26,27] In this study, we performed expression analysis of the intergenic region between *TRAF1* and *C5*. We identified a lncRNA in this region and studied its effect on neighbouring genes in RA relevant cell types.

Results

The intergenic region of the RA risk locus TRAF1 and C5 is transcribed

The number of RA-associated loci have reached over 100, with few causal genes identified.[27,28] Recent insight suggests the importance of regulatory RNAs in GWAS loci.[27] Regulatory lncRNAs can influence the expression of multiple genes in a locus.[11,12] As we previously identified the *TRAF1-C5* locus as a susceptibility locus for RA, we decided to further pursue characterisation of this locus. Analysis of the *TRAF1-C5* risk region in the University of California Santa Cruz (UCSC),

genome browser did not show the presence of non-coding genes and neither did exploration of various lncRNA databases including LNCipedia[29] and LncRBase.[30] As most commonly used RNA-sequencing approaches lack sensitivity,[31,32] we continued expression analysis of the intergenic region of *C5* and *TRAF1* in a panel of 22 tissues using intergenic primer sets (Figure 1a). We observed expression of the intergenic region in all tissues analysed (Figure 1b). Highest expression was seen in liver although in general, intergenic expression was rather low in comparison with coding genes. Before a full characterisation of the transcript, we wished to confirm that the identified transcript of >200 bp is the result of legitimate transcription. As the transcript levels were the highest in liver, we inhibited transcription in a hepatocyte cell line (Huh7) with α-amanitin. Concentrations of α-amanitin (50 µM) that specifically inhibit RNA polymerase II (RNAPII)[33] resulted in a decrease of the intergenic transcript indicating that transcription of the intergenic region is mediated by RNAPII (Figure 1c).

The intergenic region of TRAF1 and C5 contains a lncRNA gene

We next wished to further characterise the nature of the intergenic transcript. Intergenic non-coding genes can be uni- or bidirectionally transcribed. For example, non-coding RNAs transcribed from enhancer elements (eRNA) are most of the time generated by bidirectional transcription.[34,35] Therefore, we first analysed the direction of transcription by strand-specific cDNA synthesis. Here, we could only detect RNA originating from the lagging strand (data not shown), aligning transcriptional directionality in the same orientation as that of the neighbouring genes *TRAF1* and *C5*. Consequently, we reasoned that the identified intergenic transcribed sequence could still be a part of the upstream-located *C5* gene or downstream-located *TRAF1* gene. Identifying the 5′ and 3′ ends of the transcript would allow discrimination between an independent transcript or an alternative TRAF1 or C5 transcript. Like coding genes, RNAPII-derived non-coding transcripts possess 5′-methylguanosine caps.[36] To identify the 5′ end of the intergenic transcript, we conducted RNA linker-mediated (RLM) RACE (Supplementary Figure S1).[37] We detected multiple transcription start sites (TSS) located within the 3′ untranslated region (UTR) of *C5* in Huh7 cells (Figure 2a, Supplementary Figure S1). Interestingly, transcriptional start sites in the 3′ UTR of *C5* were confirmed by Fantom5 5′cap analysis gene expression in primary hepatocytes (Supplementary Figure S2).[38] The large majority of RNAPII-derived non-coding transcripts are polyadenylated at their 3′ ends.[36]

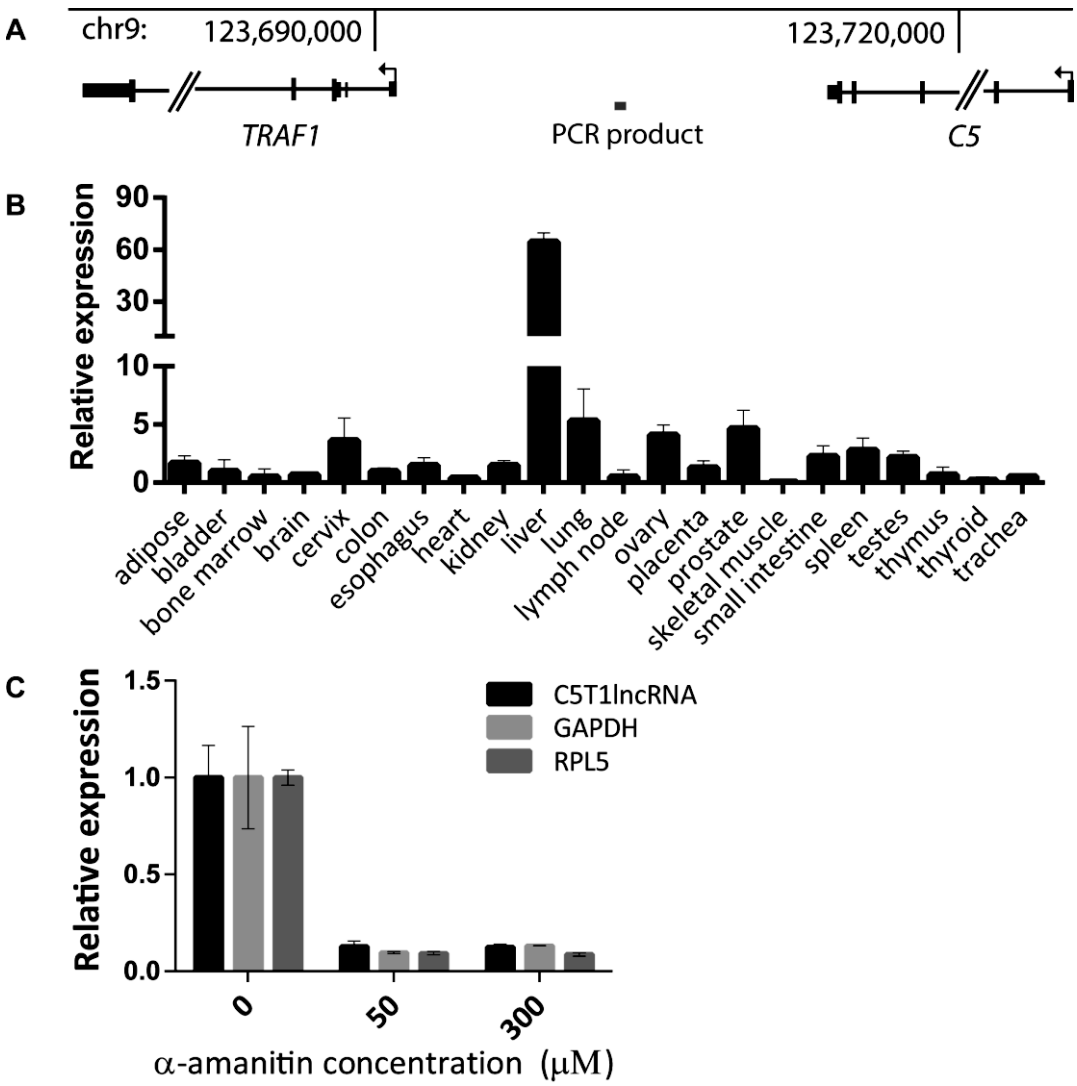

Figure 1. Identification of a novel transcript intergenic of *C5* and *TRAF1*. (a) Schematic overview of the intergenic *TRAF1-C5* region including PCR product used for RT-qPCR. Chromosome positions are according to the UCSC genome browser build GRCh37/hg19. (b) RNA levels of the intergenic amplicon (C5T1lncRNA) were measured in 22 different human tissues. Expression was normalised to GAPDH. The data are representative of two independent experiments ±s.d. (c) Inhibition of RNAPII reduced levels of the intergenic amplicon (C5T1lncRNA). A total of 50 μM α-amanitin inhibited expression of the intergenic amplicon in Huh7 cells. Expression of the intergenic transcript (C5T1lncRNA), GAPDH, RPL5 and 18S RNA was measured 36 h after addition of α-amanitin. GAPDH and RPL5 were used as positive controls, known to be transcribed by RNAPII. 18S RNA is transcribed by RNAPI and was used for normalisation. The data represents the mean±s.e.m. of two independent experiments.

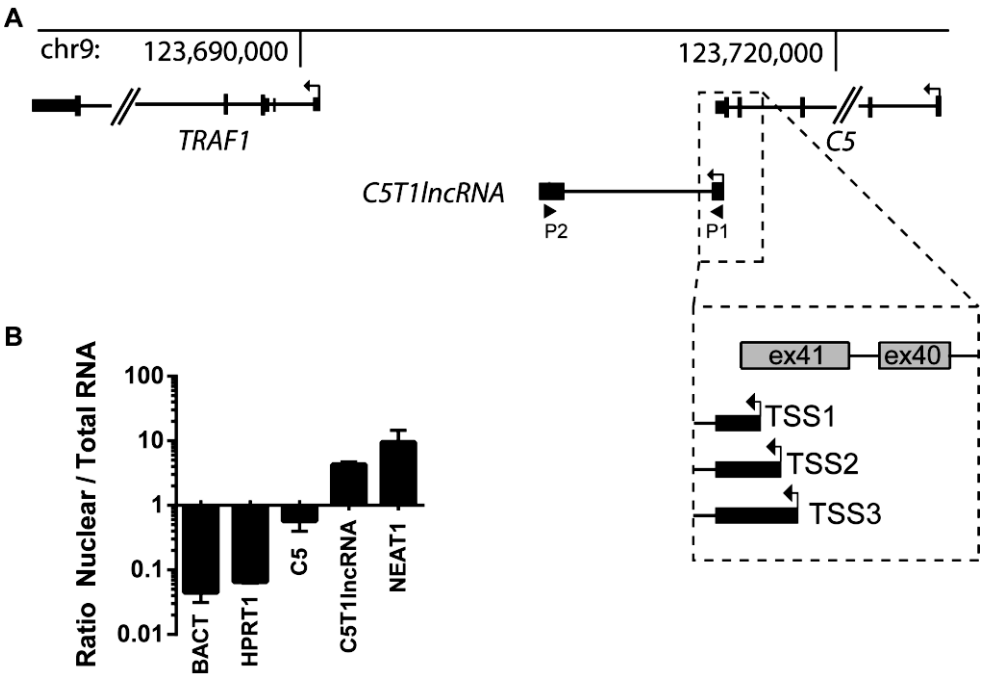

Figure 2. Characterisation of the intergenic transcript named *C5T1lncRNA*. (a) Schematic overview of an identified splice variant and the transcriptional start sites in the region harbouring the 3' UTR of *C5*. Enlargement of *C5* exon 40 and exon 41 shows the location of multiple transcriptional start sites (black blocks (TSS1, TSS2 and TSS3)). cDNA walking experiments yielded the identification of a splice variant. The spliced product was obtained by PCR on Huh7 cDNA using two primers: P1 (C5T1lncRNA5) and P2 (C5T1lncRNA2) indicated by the arrowheads. (b) Analysis of the subcellular localisation of C5T1lncRNA. Huh7 cells were separated and used for a nuclear RNA isolation and total RNA isolation, and RT-qPCR was performed to compute nuclear/total RNA ratio. qRT-PCR results are depicted as nuclear/total RNA ratio. NEAT1 is a known nuclear-localised lncRNA, whereas β-actin, HPRT1 and C5 are protein-coding mRNAs expected to localise in the cytoplasm.

3' RACE analysis in Huh7 cells revealed a 3' polyadenlation signal that does not overlap with the start of *TRAF1* (Figure 2a, Supplementary Figure S3) further confirming the presence of an independent transcript in the *TRAF1-C5* intergenic region. Further characterisation using Sanger sequencing identified two splice sites (Supplementary Figure S4).[39] The full sequence of the obtained transcript can be found in Supplementary Figure S5. Finally, we investigated whether the identified transcript is part of *C5* by a cDNA walking experiment. Primers on various exons of *C5* and the identified RNA failed to amplify a hybrid product

containing RNA from both transcripts, suggesting that the identified RNA is an independent RNA transcript (Supplementary Figure S6). We next searched for protein-coding potential of the novel transcript using NCBI open reading frame finder. All predicted open reading frames are small (<70 amino acids), do not contain known protein motifs, and more importantly have poor Kozak consensus sequences (Supplementary Figure S7a).[40,41] Moreover, the protein-coding potential score of the transcript sequence predicted by coding potential calculator (CPC) was similar to other non-coding RNAs, suggesting that it is likely of non-coding nature (Supplementary Figure S7b).[42] To find further evidence for the lncRNA nature of the identified transcript, we investigated its cellular localisation. Many lncRNAs, for example, NEAT1[43] are enriched in the nucleus.[13] We also found the identified transcripts (C5T1lncRNA) to be highly enriched in the nucleus in contrast to their neighbouring coding *C5* gene (Figure 2b). Given the presented evidence, we believe that we identified a novel lncRNA transcript in the *TRAF1-C5* region. Therefore, we refer to the transcript as *C5T1lncRNA* (*C5-TRAF1*-long non-coding RNA) from here onwards.

Correlated expression of C5T1lncRNA, TRAF1 and C5

As C5T1lncRNA is enriched in the nucleus, we hypothesised a transcriptional regulatory role for C5T1lncRNA as many regulatory lncRNAs are nuclear enriched, for example, *Pint*,[44] *Kcnq1ot1*[45] and *PANDAR*.[46] Evidence supporting a regulatory role would be either positively or negatively correlated expression of C5T1lncRNA and adjacent genes. To determine whether TRAF1 and/or C5 transcript levels correlate with C5T1lncRNA, we measured C5 and TRAF1 mRNA expression in the same tissues that were analysed for C5T1lncRNA expression (Figures 3a and b).

We only observed a strong significant positive correlation between C5 and C5T1lncRNA throughout the tissues analysed (r=0.87, P-value<0.001; Figure 3c). When separated in expression clusters, tissues belonging to cluster A predominantly coexpress *C5T1lncRNA* with *C5*. However, *C5T1lncRNA* and *TRAF1* are both expressed in spleen suggesting that perhaps *C5T1lncRNA* has a regulatory role on *TRAF1* in the cells residing there (Figure 3c, cluster D high mRNA levels of C5T1 and TRAF1). Recent data has shown that lipopolysaccharide (LPS)-induced inflammatory response in human monocytes depends on the transcription of a large number of lncRNA genes.[47]

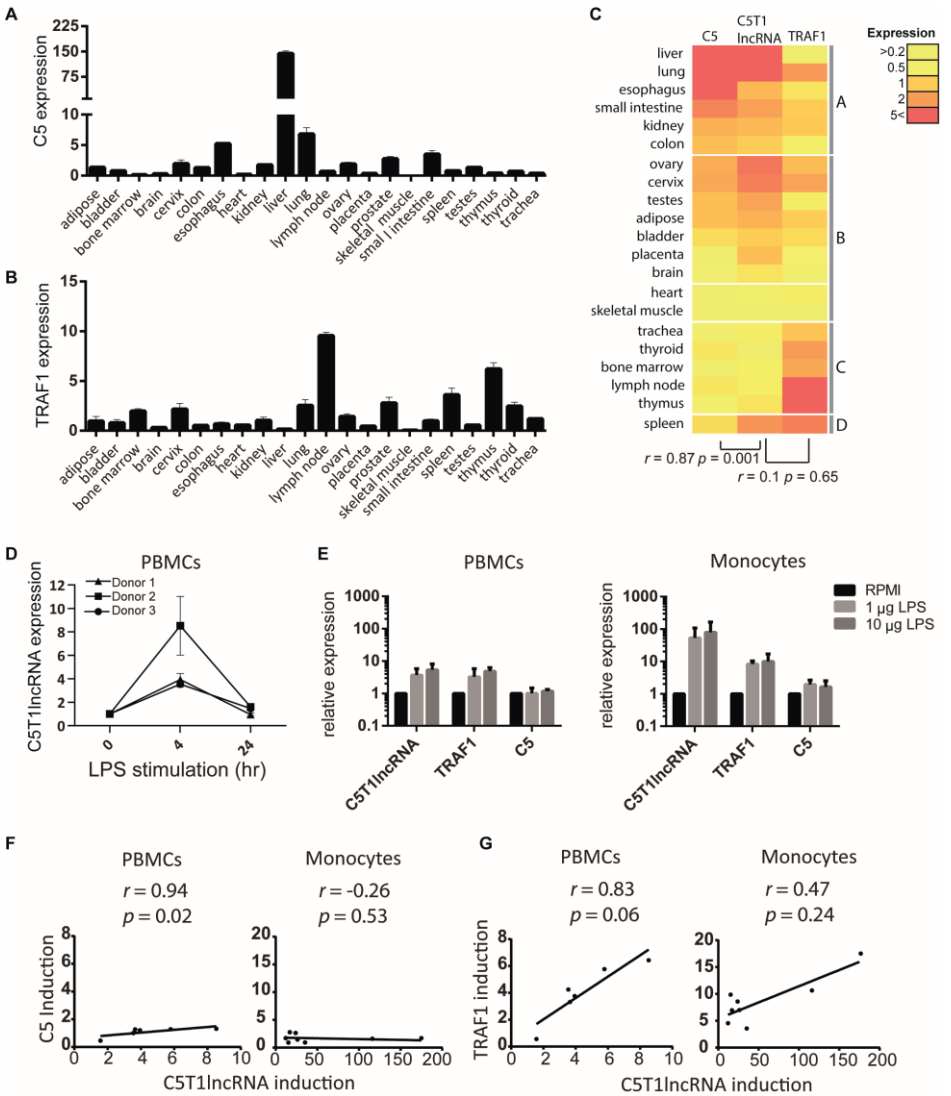

Figure 3. *C5T1lncRNA* is induced by LPS and correlates with C5 expression in various tissues and peripheral blood mononuclear cells (PBMCs). (a, b) Distribution of C5 and TRAF1 mRNA in 22 different human tissues. Expression was normalised using GAPDH. (c) Heat map of C5T1lncRNA, C5 and TRAF1 expression. Expression of each gene was compared with the mean of the 22 tissues and is indicated by colour (red=high expression/yellow=low expression). Spearman *r*-values using the expression data of all tissues are depicted underneath the heat map. Expression cluster analysis result in four clusters and are depicted as grey bars right of the heat map named A–D. (d) Time course of C5T1lncRNA expression in PBMCs after LPS stimulation (10 µg ml^{-1}). (e) Dose response curve of LPS-induced expression of C5T1lncRNA, C5 and TRAF1 in PBMCs (left) and monocytes (right). Total RNA was isolated and measured after 4 h of stimulation of either PBMCs or monocytes from three individual donors. Expression of PBMCs and monocytes were normalised for GAPDH and RPL5. C5 (f) and TRAF1 (g) RNA levels were measured within the same stimulated donors and correlated to the induction of C5T1lncRNA. Spearman *r*-values are depicted above the graphs.

As TRAF1 is abundantly expressed in monocytes[38] we analysed the expression of *C5T1lncRNA* in relation to C5 and TRAF1 transcript levels in LPS stimulated peripheral blood mononuclear cells (PBMCs) and primary monocytes. An induction of C5T1lncRNA was observed in stimulated PBMCs that was part of an early response as the highest induction was observed after 4 h (Figure 3d). A dose-dependent induction of C5T1lncRNA was observed for all tested donors and was greatest in primary monocytes (up to 200-fold for one of the donors; Figure 3e). This indicates that expression of *C5T1lncRNA* is not spurious but under specific regulation, here, being part of the innate immune response. *C5T1lncRNA* expression correlates strongly with C5 in LPS-stimulated PBMCs (r=0.94, P-value=0.02; Figures 3e and f). Interestingly, a suggestive correlation was also found between C5T1lncRNA and TRAF1 (r=0.83, P-value=0.06; Figures 3e and g). Analysis of LPS-stimulated monocytes did not show correlation between any of the genes (Figures 3f and g).

C5T1lncRNA knockdown decreases C5 levels in hepatocytes

Our data indicate co-expression of C5 with C5T1lncRNA in various tissues and PBMCs. To demonstrate that C5T1lncRNA can affect C5 mRNA levels we aimed to interfere with C5T1lncRNA levels in cells abundantly expressing the lncRNA. To this end we created C5T1lncRNA knockdowns in Huh7 cells using lentiviral vector-mediated expression of shRNAs targeting exon 2 of C5T1lncRNA. Two independent C5T1lncRNA-specific shRNAs resulted in ~60% reduction of C5T1lncRNA transcript in comparison with the shRNA control (Figures 4a and b). Knockdown of C5T1lncRNA also yielded reduced levels of C5 mRNA (Figure 4c). In contrast, C5T1lncRNA knockdown did not affect TRAF1 (Figure 4d). As the abundance of the C5T1lncRNA transcripts is an estimated 50 times lower than C5 mRNA (Supplementary Figure S8), we exclude the possibility that these results are caused by an alternative C5 transcript. Expression analysis of the generated shRNA lines demonstrated a significant correlation between C5T1lncRNA and C5 expression (r=0.85, P<0.05) (Supplementary Figure S9) corroborating our initial tissue-wide correlation data, and suggesting that C5T1lncRNA influences C5 mRNA levels.

C5T1lncRNA influences C5 RNA levels in synovial fibroblasts from RA patients

As C5T1lncRNA knockdown yielded lower C5 mRNA levels in Huh7 cells, we wished to investigate whether the observed correlated C5T1lncRNA and C5

expression can be observed in RA-relevant cell types as well. siRNA studies in primary PBMCs are not feasible, because transfection or transduction of primary blood cells is notoriously difficult. We therefore performed the knockdown experiment in synovial fibroblasts. Analysis of the mRNA in synovial fibroblasts derived from RA patients showed that *TRAF1*, *C5* and *C5T1lncRNA* are expressed (Figure 5a). The expression levels are independent of RA as synovial fibroblasts from osteoarthritis patients have similar levels of TRAF1, C5 and C5T1lncRNA (Figure 5b). As primary synovial fibroblasts grow relatively slow in culture we performed transient transfections using GapmeR sequences instead of shRNA transduction.

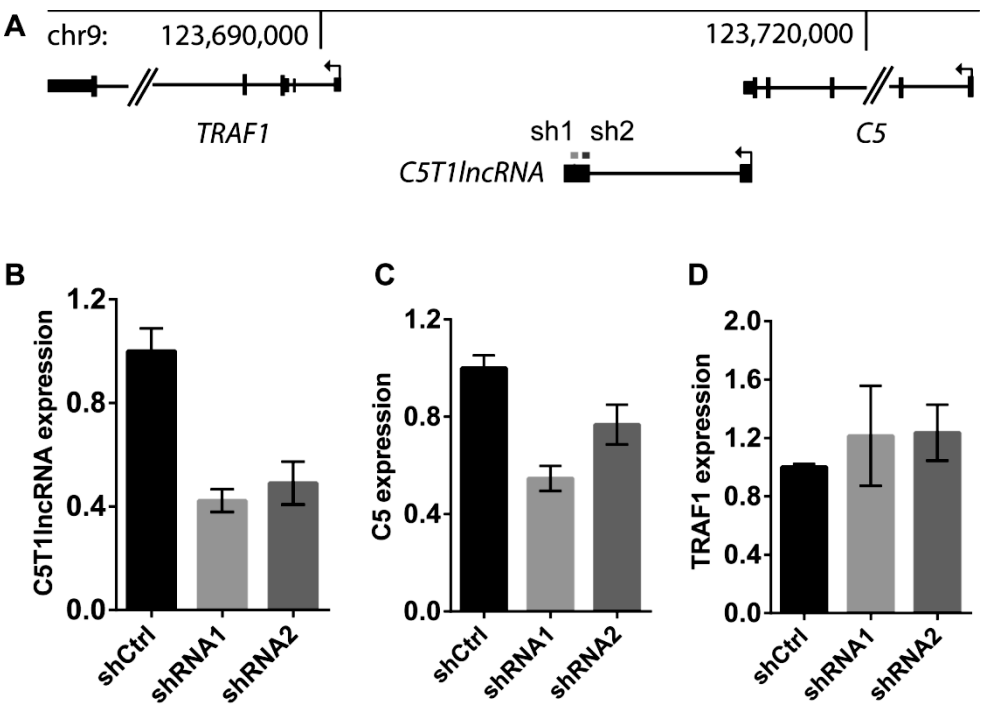

Figure 4. Knockdown of C5T1lncRNA using lentiviral shRNA. (a) Schematic representation of the location of shRNA1 (sh1) and shRNA2 (sh2) sequences that were used to knockdown C5T1lncRNA with a lentiviral vector. As a control Huh7 cells were transduced with empty viral vector and a nonspecific shRNA (shCtrl). (b-d) C5T1lncRNA (b), C5 (c) and TRAF1 (d) RNA levels were quantified using RT-qPCR. RNA levels were normalised for GAPDH. Data represent the mean±s.e.m. of two independent experiments.

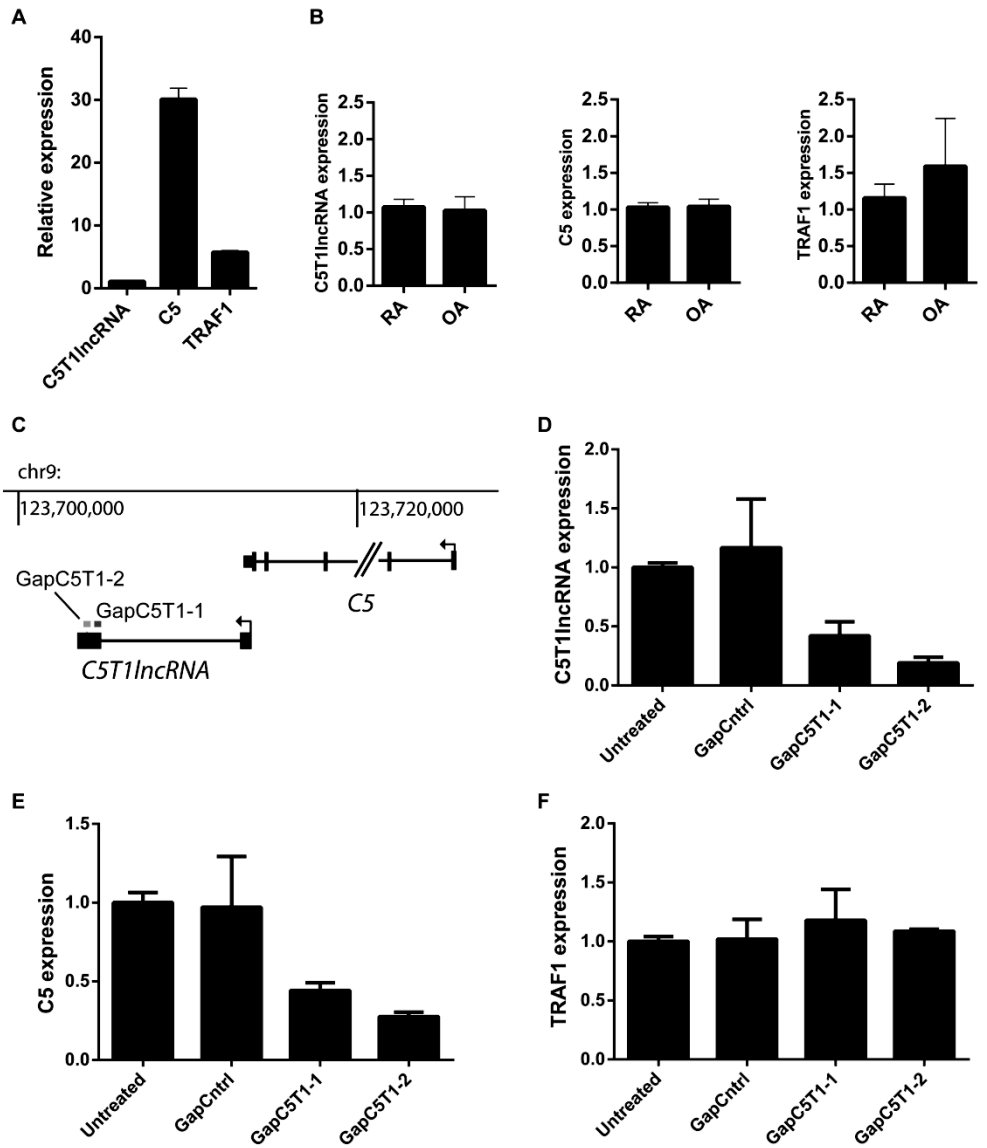

Figure 5. Expression of C5T1lncRNA, C5 and TRAF1 in synovial fibroblasts. (a) C5T1lncRNA, C5 and TRAF1 expression was measured in synovial fibroblasts obtained from two RA patients. Expression of C5T1lncRNA was set to 1. (b) C5T1lncRNA, C5 and TRAF1 RNA levels measured in synovial fibroblasts obtained from 18 RA and 10 osteoarthritis patients. Mean expression in RA patients was set to 1. RNA levels were normalised for housekeeping genes HPRT1 and B2M. (c–f) Knockdown of C5T1lncRNA using GapmeRs in synovial fibroblasts. Synovial fibroblasts were cultured untreated for 48 h or treated with a nonspecific GapmeR (GapCntrl) or two specific GapmeR sequences (GapC5T1-1 and GapC5T1-2). (c) Schematic representation of the target location of the GapmeR sequences (GapC5T1-1 and GapC5T1-2) used to knockdown C5T1lncRNA. C5T1lncRNA (d), C5 (e) and TRAF1 (f) RNA levels were quantified using RT-qPCR. RNA levels were normalised to B2M. The data represent the mean±s.e.m. of two independent donors.

GapmeR-mediated knockdown of C5T1lncRNA in synovial fibroblasts resulted in strongly reduced C5 mRNA levels reiterating the results we observed in hepatocytes (Figures 5c–e). In contrast, mRNA levels of TRAF1 were not affected by the C5T1lncRNA knockdown (Figure 5f). Taken together, we identified a novel lncRNA in a locus genetically associated to RA that influences C5 RNA levels in both liver-derived cells and synovial fibroblast.

Discussion

lncRNAs that act locally by regulating the levels of neighbouring genes are thought to be widely present in the genome. A few locally acting lncRNAs have shown to be responsible for important developmental processes and physiology, and have been linked to human disease.[48] However, locally acting lncRNAs are highly understudied as possible disease candidate genes in loci associated to complex genetic diseases. Challenges include low expression levels, higher tissue specificity compared with protein-coding genes, and transcriptional overlap (on both strands). A more prevalent role for locally acting lncRNAs in complex genetic diseases has been suggested by a study correlating lncRNA expression to disease-associated polymorphisms using PBMCs as 29 eQTLs affected expression of the lncRNAs and neighbouring protein-coding genes.[49] The outcome was however hampered by low expression levels of lncRNAs. Also more advanced RNA-sequencing methods may be hindered in the identification of lncRNA disease genes by aforementioned low lncRNA expression and complex sequence characteristics. The success of 'forward' approaches hinges on the correct choice of tissues, in which transcript levels are abundant enough to be quantitatively assessed. The aetiology of many complex genetic diseases is largely enigmatic confounding the cell type choice. We took a reverse approach based on the presence of multiple candidate genes, eQTLs, and preliminary unpublished evidence of transcription in the GWAS region. Interestingly, analysis of expression of the intergenic region between *TRAF1* and *C5* revealed the presence of a novel, spliced non-coding gene that starts in the region containing the 3' UTR of *C5*, and ends half-way between *C5* and *TRAF1*. The overlap with the region encompassing the 3' UTR of *C5*, and the low expression of C5T1lncRNA imposes that microarray or regular RNA sequencing transcriptome analyses will fail to identify this non-coding transcript. Knocking down the expression of C5T1lncRNA in cells, in which C5T1lncRNA expression correlates with C5, yielded reduced C5 mRNA levels. As

mRNA levels affected, the protein C5 is likely to be similarly decreased as well. Thus our data suggest that C5T1lncRNA influences the mRNA levels of the RA candidate gene C5, albeit we do not yet grasp the mechanism. Possible paths for the observed decreased C5 levels include miRNA sponging and transcriptional enhancement. The fact that C5T1lncRNA and C5 expression are positively correlated could point to a miRNA sponge function.[6] Sponging miRNAs is a common mechanism by which lncRNAs enhance mRNA levels of protein coding genes when the lncRNA gene overlaps with that of protein-coding genes.[6,50] To our knowledge, miRNAs targeting the 3' UTR of C5 mRNA have been predicted but not yet validated.[51] Another plausible functional mechanism would be based on the presence of a Tigger4B repeat in C5T1lncRNA. This repeat sequence is 93% identical to a stretch of 130 bp in intron 28 of *C5*. As these sequences are in the reverse orientation, the RNA could potentially hybridise to either C5 pre-mRNA or its genomic DNA. Recently, lncRNA/pre-mRNA interaction was hypothesised as a strategy used by many ncRNAs as part of their regulatory function.[52] Moreover, repeats can be found in the majority of lncRNAs and have been associated to various functional mechanisms.[53] Alu repeat sequences in lncRNAs have been linked to Staufen-mediated decay of coding transcripts,[54] Staufen-mediated mRNA stabilisation[55] and transcriptional activation.[15] Moreover, *BACE1-AS* a non-coding RNA that is transcribed from the opposite DNA strand as *BACE1* was shown to stabilise the mRNA of *BACE1* by hybridisation between both transcripts.[56] A similar regulatory mechanism might apply to *C5T1lncRNA* and *C5*, where C5 RNA is stabilised by hybridisation to C5T1lncRNA. However, the precise mechanism and its potential role in RA require further investigation. Finally, eQTL data have revealed that lower levels of C5 and TRAF1 are linked to the risk alleles. It will be of interest to see whether C5T1lncRNA expression or function is related to the disease haplotype. Remarkably, two highly RA-associated single-nucleotide polymorphisms (rs10818488 and rs35517037) are present in the exonic region of *C5T1lncRNA*, of which rs35517037 is located within the Tigger4B repeat. Such single-nucleotide polymorphisms could alter RNA function by either influencing RNA stability, RNA structure or RNA-binding ability and therefore need further investigation.[57]

Taken together, we identified a novel RA candidate gene that is of non-coding nature using an alternative approach based on the presence of protein coding candidate disease genes in genetic risk locus. This lncRNA could be considered as

a novel candidate gene as it is fully located within the associated region, is expressed in RA-relevant cells and influences the mRNA levels of at least one of the postulated candidate disease genes. The strategy employed in this study constitutes a practical alternative for the identification of novel regulatory lncRNAs as disease candidate genes of complex genetic immune-related diseases. It would be interesting to further dissect the role for this non-coding RNA in disease over the next years. It is expected that discovery of these lncRNAs will take a massive leap in near future and may represent pivotal players in the pathogenesis of complex immune diseases.

Materials and methods

Cell culture

Huh7 cells obtained from ATCC were cultured in Dulbecco's Modified Eagles Medium (Gibco, Life Technologies, Paisley, UK) supplemented with 15% fetal bovine serum (Gibco, Life Technologies), 2 mM l-glutamine (Gibco, Life Technologies), 10 U ml^{-1} penicillin–streptomycin (Gibco, Life Technologies) at 37 °C and 10% CO2. Huh7 cells were tested negative for mycoplasma multiple times during the various experiments.

Primary PBMCs isolated on a Ficoll gradient (pharmacy LUMC, Leiden, The Netherlands) were cultured in RPMI 1640 Medium (Gibco, Life Technologies) supplemented with 10% heat-inactivated fetal bovine serum, 2 mM l-glutamine and 10 U ml^{-1} penicillin–streptomycin at 37 °C and 5% CO2. Monocytes (CD14+) were isolated from PBMCs using magnetic beads (Miltenyi Biotec, Bergisch Gladbach, Germany) according to the manufacturer's instructions. The purity of isolated monocytes was determined by flow cytometry using a PerCPCy5.5 conjugated antibody against human CD14 (catalogue no. 45-0149-41: eBioscience, San Diego, CA, USA).

LPS stimulation experiments were performed with either 1 × 10^6 PBMCs or monocytes per ml with 1 or 10 µg LPS obtained from *Salmonella typhosa* (Sigma-Aldrich, Saint Louis, MO, USA). Primary cultures of synovial fibroblasts were established by collagenolytic digestion of synovial tissue specimens obtained from RA and osteoarthritis patients during joint replacement surgery. RA patients fulfilled the American College of Rheumatology 1987 criteria for RA.[58] Synovial

fibroblasts cultures were maintained in Dulbecco's Modified Eagles Medium/F-12 (Gibco, Life Technologies) supplemented with 10% fetal bovine serum, 2 mM l-glutamine, 10 mM HEPES (Gibco, Life Technologies), 50 U ml^{-1} penicillin–streptomycin, 0.5 mg ml^{-1} amphotericin B (Gibco, Life Technologies) and used for experiments between passages 5 and 8. The study was approved by the local ethic committee and all patients provided informed consent.

RNA isolation, first-strand cDNA synthesis and real-time quantitative PCR
Total RNA was extracted using nucleospin RNA purification columns (Macherey Nagel, Duren, Germany) according to the manufacturer's instructions. Purified RNA was subjected to additional DNase I (1 u μg^{-1}) (Invitrogen, Carlsbad, CA, USA) treatment. First-strand cDNA was synthesised using reverse transcriptase (Fermentas, Vilnius, Lithuania) according to the manufacturer's instructions. Real-time quantitative PCR (RT-qPCR) was conducted using sensifast no-rox SYBR green mix (Bioline, Luckenwalde, Germany) on a CFX96 or CFX384 real-time PCR system (Bio-Rad, Temse, Belgium).

The total RNA tissue panel was obtained from Ambion (RNA survey pool, Leusden, The Netherlands), whereas total RNA from human bone marrow and lymph nodes was obtained from Clontech (Saint-Germain-en-Laye, France). For these tissues 1 µg of total RNA was used for first-strand cDNA synthesis using SuperScript III reverse transcriptase (Invitrogen). For synovial fibroblast experiments, RNA was isolated using the miRNEAsy kit (Qiagen, Hilden, Germany) including on column DNAse digestion. cDNA was synthesised using MultiScribe Reverse Transcriptase (Invitrogen). FastStart Universal SYBR Green Master (Rox) (Roche, Mannheim, Germany) was used for qRT-PCR. Expression in synovial fibroblast was measured on a 7900HT Fast Real-Time PCR System (Applied Biosciences, Foster City, CA, USA). Multiple housekeeping genes were used for normalisation: glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (RNA tissue panel), GAPDH and ribosomal protein L5 (RPL5) (PBMCs and monocytes), and beta-2-microglobulin (β2M) and hypoxanthine phosphoribosyltransferase (HPRT1) (synovial fibroblasts). Relative expression was calculated using the ΔΔCT method.[59]

Expression of untreated samples was set to 1 unless otherwise stated. Sequences of the used primer pairs are listed in Supplementary Table S1. Expression of

C5T1lncRNA was measured using C5T1lncRNA primer1 and C5T1lncRNA primer2 (Supplementary Table S1). All qPCR measurements were performed in triplicate including minus RT samples.

C5T1lncRNA characterisation

5' RLM-RACE

A schematic overview of the protocol is depicted in Supplementary Figure S1. A total of 30 µg RNA from Huh7 cells was treated with DNase I (Invitrogen). A phenol/chloroform extraction was performed after each enzymatic step. RNA was treated with calf intestine alkaline phosphatase (New England Biolabs (NEB), Frankfurt, Germany). Next, RNA was treated with tobacco acid pyrophosphatase (Epicentre, Warsaw, Poland) and the RNA adapter (for sequence see Supplementary Table S1) was ligated to the liberated 5' end with RNA ligase (Thermo Scientific, Waltham, MA, USA). Strand-specific cDNA was produced using C5T1lncRNA-primer3 and amplified with an adapter-specific primer (for sequence see Supplementary Table S1) and C5T1lncRNA primer4.

PCR products were Sanger sequenced. cDNA walking experiments were performed to identify the possible splice transcripts. A splice variant was identified using primers C5T1lncRNA primer2 and C5T1lncRNA primer5. PCR products were separated on a 1% agarose gel, purified using GeneJET gel extraction kit (Thermo Scientific) and Sanger sequenced.

3' RACE

cDNA synthesis of liver RNA was performed using a 3' RACE primer (for sequence see Supplementary Table S1). C5T1lncRNA was amplified (Phusion, NEB) using the 3'RACE adapter-primer and C5T1lncRNA primer1. A second round of PCR was performed using the nested primer: C5T1lncRNA primer6. PCR products were purified from a 1% agarose gel and Sanger sequenced. Obtained sequences were aligned to the human genome (UCSC genome browser build hg19).[60]

cDNA walking experiment

A PCR was performed on cDNA from Huh7 cells using Phusion polymerase according to the manufacturer's protocol using various primers located in *C5* and *C5T1lncRNA* (C5 primer 2 to 5, C5T1lncRNA primer 2 and 5). The C5T1lncRNA

sequence is available at the DNA Data Bank of Japan (DDBJ) with accession number LC094347.

Inhibition of transcription

3×10^5 Huh7 cells were first cultured for 24 h in one well of a 24-well plate. Cells were then treated with 0, 50 and 300 µM α-amanitin (Sigma-Aldrich) for 36 h. After treatment, RNA was collected and analysed using RT-qPCR. RNA levels of GAPDH, and RPL5 (both RNAPOLII transcribed), and 18S (RNAPOLI) were analysed and used for normalisation.

Nuclear isolation, relative abundance of transcript levels and RNA localisation

Huh7 cells were isolated by incubating 5×10^6 cells in 1 ml lysis buffer I (50 mM HEPES–KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1 × cOmplete protease inhibitors (Roche) for 10 min on 4 °C roller bank. Nuclei were pelleted (1350 *g* at 4 °C for 5 min), washed carefully in lysis buffer II (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1 × cOmplete protease inhibitors (Roche) by rocking for 10 min at RT, and repelleted. Nuclei were lysed using RA1 RNA lysis buffer (Macherey-Nagel) and RNA was isolated using nucleospin columns (Macherey-Nagel). Primers representing NEAT1, a nuclear located RNA were used as positive control. To calculate the difference between relative abundance of RNA transcripts of C5 and C5T1lncRNA, cDNA from Huh7 cells was used for qPCR. To amplify C5T1lncRNA, primer4 and primer7 were used. C5 RNA was amplified using primer7 and C5-primer1 (Supplementary Table S1). Ct-values were normalised for primer efficiencies by taking along genomic DNA.

Lentiviral knockdown of C5T1lncRNA

shRNA sequences targeting C5T1lncRNA were designed, synthesised by ShineGene (Shanghai, China), and Sanger sequenced. Sequences are depicted in Supplementary Table S2. shRNA sequences were introduced into a pRRL.Super plasmid, which was generated by introducing the H1-mcs-PGKPuro part of pRetroSuper into the pRRL vector. Lentiviral vectors were generated in 293T cells using the three plasmid lentiviral production system.[61] Lentiviral titres were estimated by ELISA on basis of P24 levels (ZeptoMetrix, Buffalo, NY, USA). As controls either an empty lentiviral vector or a vector containing a shRNA sequence against mouse ATF3 were used.

Huh7 cells were transduced at a multiplicity of infection of five based on the levels of P24 (1 ng ml^{-1} of P24 equals the transduction of 2500 cells). shRNA-expressing cells were selected using 1 μg ml^{-1} Puromycin. Two weeks after puromycin selection cells were collected for RNA analysis.

GapmeR transfection

Custom designed LNA longRNA GapmeRs targeting C5T1lncRNA; 5'-GGCCTCTTCACGTAGT-3' (GapC5T1-1) and 5'-CGGGATCTGGAACATT-3' (GapC5T1-2) were purchased from Exiqon (Vedbaek, Denmark) along with the negative control LNA longRNA GapmeR - Negative Control A (catalogue no. 300610). RASF were transfected with 10 nM GapmeRs using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. After 24 h, medium was replaced and 48 h after transfection cells were lysed for RNA isolation.

Correlation and statistical analysis

Unless otherwise stated, expression experiments were performed at least three independent times or with three independent donors and is presented as mean±s.d. Statistical analysis was performed using GraphPad (GraphPad, San Diego, CA, USA). Correlation analysis was performed using spearman rank *r* correlation test. Spearman rank analysis was applied using relative expression values. Differences with *P*-values <0.05 are considered significant.

Supplementary material

Supplementary information is available online on the Genes and Immunity website (http://www.nature.com/gene): Supplementary Figure S1-S9 and Supplementary Table S1 and S2.

Acknowledgements

We wish to thank SJ Cramer and M Rabelink for the generation of shRNA-containing lentiviral particles and H van Dam for providing reagents. We also thank the volunteers for donating blood and AWG Berendsen for experimental support. Finally we thank RC Hoeben and A Zaldumbide for comments and critical reading of the manuscript. This work was supported by the Dutch Arthritis

Foundation, The Netherlands. F Kurreeman is supported via UNESCO-L'Oreal for Women in Science Fellowship, Marie Curie FP7 Outgoing Fellowship and an LUMC Research Fellowship.

References

1. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet 2011; 12: 861–874.
2. Churchill GA, Holland MJ, Draghici S, Khatri P, Eklund AC, Szallasi Z *et al*. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 2007; 316: 1484–1488.
3. Hangauer MJ, Vaughn IW, McManus MT. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. PLoS Genet 2013; 9: e1003569.
4. Van Leeuwen S, Mikkers H. Long non-coding RNAs: guardians of development. Differentiation 2010; 80: 175–183.
5. Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol 2013; 14: 699–712.
6. Xia T, Liao Q, Jiang X, Shao Y, Xiao B, Xi Y *et al*. Long noncoding RNA associated-competing endogenous RNAs in gastric cancer. Sci Rep 2014; 4: 6088.
7. Wang X, Arai S, Song X, Reichart D, Du K, Pascual G *et al*. Induced ncRNAs allosterically modify RNA-binding proteins *in cis* to inhibit transcription. Nature 2008; 454: 126–130.
8. Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S *et al*. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol Cell 2010; 38: 662–674.
9. Terranova R, Yokobayashi S, Stadler MB, Otte AP, van Lohuizen M, Orkin SH *et al*. Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev Cell 2008; 15: 668–679.
10. Tsai M, Manor O, Wan Y, Mosammaparast N, Wang JK, Shi Y *et al*. Long noncoding RNA as modular scaffold of histone modification complexes. Science 2010; 329: 689–693.
11. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J *et al*. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 2008; 32: 232–246.
12. Nagano T, Mitchell JA, Sanz LA, Pauler FM. The air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science 2008; 322: 1717–1720.
13. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H *et al*. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 2012; 22: 1775–1789.
14. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H *et al*. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 2014; 42: D1001–D1006.
15. Holdt LM, Hoffmann S, Sass K, Langenberger D, Scholz M, Krohn K *et al*. Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks. PLoS Genet 2013; 9: e1003588.
16. Kurreeman FaS, Padyukov L, Marques RB, Schrodi SJ, Seddighzadeh M, Stoeken-Rijsbergen G *et al*. A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. PLoS Med 2007; 4: e278.
17. Plenge RM, Seielstad M, Padyukov L, Lee T, Remmers EF, Ding B *et al*. TRAF1–C5 as a risk locus for rheumatoid arthritis — a genomewide study. NEJM 2007; 357: 1199–1209.
18. Lavorgna A, De Filippi R, Formisano S, Leonardi A. TNF receptor-associated factor 1 is a positive regulator of the NF-kappaB alternative pathway. Mol Immunol 2009; 46: 3278–3282.
19. Ma X, Xu S. TNF inhibitor therapy for rheumatoid arthritis. Biomed Rep 2013; 1: 177–184.

20. Cooke TD, Hurd ER, Jasin HE, Bienenstock J, Ziff M. Identification of immunoglobulins and complement in rheumatoid articular collagenous tissues. Arthritis Rheum 1975; 18: 541–551.
21. Wang Y, Kristan J, Hao L, Lenkoski CS, Shen Y, Matis LA. A role for complement in antibody-mediated inflammation: C5-deficient DBA/1 mice are resistant to collagen-induced arthritis. J Immunol 2000; 164: 4340–4347.
22. Dunkelberger JR, Song W-C. Complement and its role in innate and adaptive immune responses. Cell Res 2010; 20: 34–50.
23. Su X, Li S, Meng M, Qian W, Xie W, Chen D *et al*. TNF receptor-associated factor-1 (TRAF1) negatively regulates Toll/IL-1 receptor domain-containing adaptor inducing IFN-beta (TRIF)-mediated signaling. Eur J Immunol 2006; 36: 199–206.
24. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS *et al*. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci USA 2009; 106: 9362–9367.
25. Messemaker T, Toes REM, Kurreeman F, Mikkers HMM. Comment on "functional analysis of a complement polymorphism (rs17611) associated with rheumatoid arthritis. J Immunol 2015; 195: 3–4.
26. Ding J, Eyre S, Worthington J. Genetics of RA susceptibility, what comes next? RMD Open 2015; 1: e000028–e000028.
27. Messemaker TC, Huizinga TW, Kurreeman F. Immunogenetics of rheumatoid arthritis: understanding functional implications. J Autoimmun 2015; 64: 74–81.
28. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K *et al*. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2013; 506: 376–381.
29. Volders P-J, Helsens K, Wang X, Menten B, Martens L, Gevaert K *et al*. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res 2013; 41: D246–D251.
30. Chakraborty S, Deb A, Maji RK, Saha S, Ghosh Z. LncRBase: an enriched resource for lncRNA information. PLoS One 2014; 9: e108010.
31. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-vega B, Regev A *et al*. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 2011; 25: 1915–1927.
32. Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh JA *et al*. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat Biotechnol 2012; 30: 99–104.
33. Warfel AH. Specific Inhibition of nuclear RNA polymerase II by a-Amanitin. Science 1970; 170: 447–449.
34. Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J *et al*. Widespread transcription at neuronal activity-regulated enhancers. Nature 2010; 465: 182–187.
35. Ren B. Transcription: enhancers make non-coding RNA. Nature 2010; 465: 173–174.
36. Tuck AC, Tollervey D. A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. Cell 2013; 154: 996–1009.
37. Liu X, Gorovsky MA. Mapping the 5' and 3' ends of Tetrahymena therrmophila mRNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE). Nucleic Acids Res 1993; 21: 4954–4960.
38. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S *et al*. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol 2015; 16: 22.
39. Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human splicing finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009; 37: e67.
40. Kozak M, Limited IRLP. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 1987; 15: 8125–8148.

41. Nishikawa T, Ota T, Isogai T. Prediction whether a human cDNA sequence contains initiation codon by combining statistical information and similarity with protein sequences. Bioinformatics 2000; 16: 960–967.
42. Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L *et al*. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 2007; 35: W345–W349.
43. Imamura K, Imamachi N, Akizuki G, Kumakura M, Kawaguchi A, Nagata K *et al*. Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. Mol Cell 2014; 53: 393–406.
44. Marín-Béjar O, Marchese FP, Athie A, Sánchez Y, González J, Segura V *et al*. Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. Genome Biol 2013; 14: R104.
45. Lee JT, Bartolomei MS. X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell 2013; 152: 1308–1323.
46. Han L, Zhang E, Yin D, Kong R, Xu T, Chen W *et al*. Low expression of long noncoding RNA PANDAR predicts a poor prognosis of non-small cell lung cancer and affects cell apoptosis by regulating Bcl-2. Cell Death Dis 2015; 6: e1665.
47. IIott NE, Heward JA, Roux B, Tsitsiou E, Fenwick PS, Lenzi L *et al*. Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. Nat Commun 2014; 5: 3979.
48. Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X *et al*. LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Res 2013; 41: D983–D986.
49. Kumar V, Westra H-J, Karjalainen J, Zhernakova DV, Esko T, Hrdlickova B *et al*. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet 2013; 9: e1003201.
50. Kallen AN, Zhou X-B, Xu J, Qiao C, Ma J, Yan L *et al*. The imprinted H19 LncRNA antagonizes Let-7 microRNAs. Mol Cell 2013; 52: 101–112.
51. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005; 120: 15–20.
52. Engreitz JM, Sirokman K, McDonel P, Shishkin AA, Surka C, Russell P *et al*. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. Cell 2014; 159: 188–199.
53. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G *et al*. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet 2013; 9: e1003470.
54. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature 2011; 470: 284–288.
55. Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K *et al*. Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature 2013; 493: 231–235.
56. Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE *et al*. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of B-secretase expression. Nat Med 2008; 14: 723–730.
57. Naslavsky MS, Crovella S, Lima Filho JL, Rocha CRC. The sound of silence: human beta-defensin-1 gene untranslated SNPs change the predicted mRNA secondary structure in a length-dependent manner. Immunol Lett 2010; 129: 53–55.
58. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988; 31: 315–324.
59. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001; 25: 402–408.

60. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M *et al*. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res 2015; 43 (Database issue): D670–D681.

61. Bai Y, Soda Y, Izawa K, Tanabe T, Kang X, Tojo A *et al*. Effective transduction and stable transgene expression in human blood cells by a third-generation lentiviral vector. Gene Ther 2003; 10: 1446–1457.