In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading

Julian C. Knight[1,2], Brendan J. Keating[1], Kirk A. Rockett[1] & Dominic P. Kwiatkowski[1,3]

In vivo characterization of regulatory polymorphisms is a key requirement for next-generation human genetic analysis. Here we describe haploChIP, a method that uses chromatin immunoprecipitation (ChIP) and mass spectrometry to identify differential protein–DNA binding *in vivo* associated with allelic variants of a gene. We demonstrate this approach with the imprinted gene *SNRPN*. HaploChIP showed close correlation between the level of bound phosphorylated RNA polymerase II at the *SNRPN* locus and allele-specific expression. Application of the approach to the *TNF/LTA* locus identified functionally important haplotypes that correlate with allele-specific transcription of *LTA*. The haploChIP method may be useful in high-throughput screening for common DNA polymorphisms that affect gene regulation *in vivo*.

Introduction

Single-nucleotide polymorphisms (SNPs) comprise a large part of human diversity, and their inheritance may alter susceptibility to disease. For SNPs in coding regions, it is often possible to predict consequences for protein structure and function, but most SNPs are found in non-coding DNA. These SNPs may affect gene regulation or they may have no biological consequence. Screening for SNPs that affect transcriptional regulation is commonly done by *in vitro* assays of protein–DNA interaction and plasmid reporter-gene expression. The biological relevance of this approach is limited by the absence of a natural chromatin structure and regulation. Furthermore, SNPs are typically analyzed in isolation, whereas it may be the precise combination of SNPs on a given allele (the haplotype) that determines regulatory significance. We therefore need ways of evaluating the effects of SNPs on transcriptional regulation *in vivo*, that is, in natural chromosomes in a normally functioning nucleus. The simplest approach is to compare expression levels in cells from people of different genotype; this may, however, be confounded by many genetic, environmental and technical factors that differ between individuals. These factors can be controlled if the alleles are compared in cells from an individual who is heterozygous with respect to a given SNP, for example, by comparing the amount of transcript by allele-specific RT–PCR. Unfortunately, this approach has limited applicability because it requires that the polymorphism or a linked marker appear in the RNA transcript.

This study aimed to find a sensitive method for detecting SNPs that affect gene regulation *in vivo*, even if there are no suitable genetic markers on the RNA transcript. We reasoned that in a cell that is heterozygous with respect to one or more SNPs in the location of a gene of interest, these SNPs could permit allele-specific discrimination of protein–DNA interactions occurring *in vivo* and thus identify differential effects on specific molecular events occurring in transcription. We chose to focus here on protein–DNA interactions that could provide a surrogate measure of transcriptional activity, as these have the broadest applicability. The central event of the transcriptional process begins when RNA polymerase II (Pol II) is released from the initiation complex to start synthesis of the nascent transcript. This is associated with phosphorylation of specific serine residues in the C-terminal domain (CTD) of Pol II[1–4]. Phosphorylation of Ser5 correlates with the transition from initiation to elongation and mRNA capping[4–6] and is essential for Pol II to be in the processive elongation form[7], whereas phosphorylation at Ser2 occurs later in elongation[8]. Previous studies have confirmed that the amount of phosphorylated Pol II associated with a segment of chromatin is related to the transcriptional activity of the corresponding gene[3,9–11]. Here we examine whether it is possible to detect differences in the amount of phosphorylated Pol II bound to two different alleles of a gene in cells using haplotype-specific chromatin immunoprecipitation (haploChIP) and whether this is related to allelic differences in gene expression *in vivo*.

Results

Distinct allele-specific loading in an imprinted gene

To test this concept, we studied expression of *SNRPN* (encoding small nuclear ribonucleoprotein polypeptide N), an imprinted gene in which the two alleles have markedly different levels of expression according to their parental origin[12]. To distinguish between alleles, we typed a panel of B-cell lines transformed with

[1]*Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.* [2]*Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA.* [3]*University Department of Pediatrics, John Radcliffe Hospital, Oxford OX3 9DU, UK. Correspondence should be addressed to J.C.K. (e-mail: julian@well.ox.ac.uk).*

Epstein–Barr virus (EBV) for a SNP at nucleotide position 377 (numbered according to cDNA) in exon 4 of *SNRPN*[13]. Comparison of genomic DNA and cDNA confirmed that, in cell lines heterozygous with respect to this SNP, only one of the two alleles present in the genomic DNA produced an mRNA transcript (Fig. 1*a*). We then carried out a ChIP assay for phosphorylated Pol II using antibodies highly specific for certain phosphorylated serine residues of the CTD[14]. When chromatin from heterozygous cell lines was analyzed for the SNP at nt 377 in *SNRPN*, the starting material (total input chromatin) contained similar amounts of each allele, whereas the immunoprecipitated material (chromatin to which phosphorylated Pol II was bound) contained predominantly one allele, corresponding to the single allele that produced an mRNA transcript (Fig. 1*b*). This shows that one

copy of *SNRPN* is transcriptionally suppressed before phosphorylation of Pol II by TFIIH, that is, before the synthesis of mRNA commences. This is expected, as genomic imprinting is known to depend on DNA methylation[15] and to involve chromatin compaction[16] and histone acetylation[17,18].

Accurate quantification of haploChIP

Because the effects of common regulatory polymorphisms may be subtle, the haploChIP approach requires a sensitive quantitative assay of the relative abundance of two different alleles in a sample of immunoprecipitated chromatin that has the potential for high-throughput analysis. To achieve this, we used primer extension with detection by matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI–TOF or PE/MS; refs. 19,20). Mass
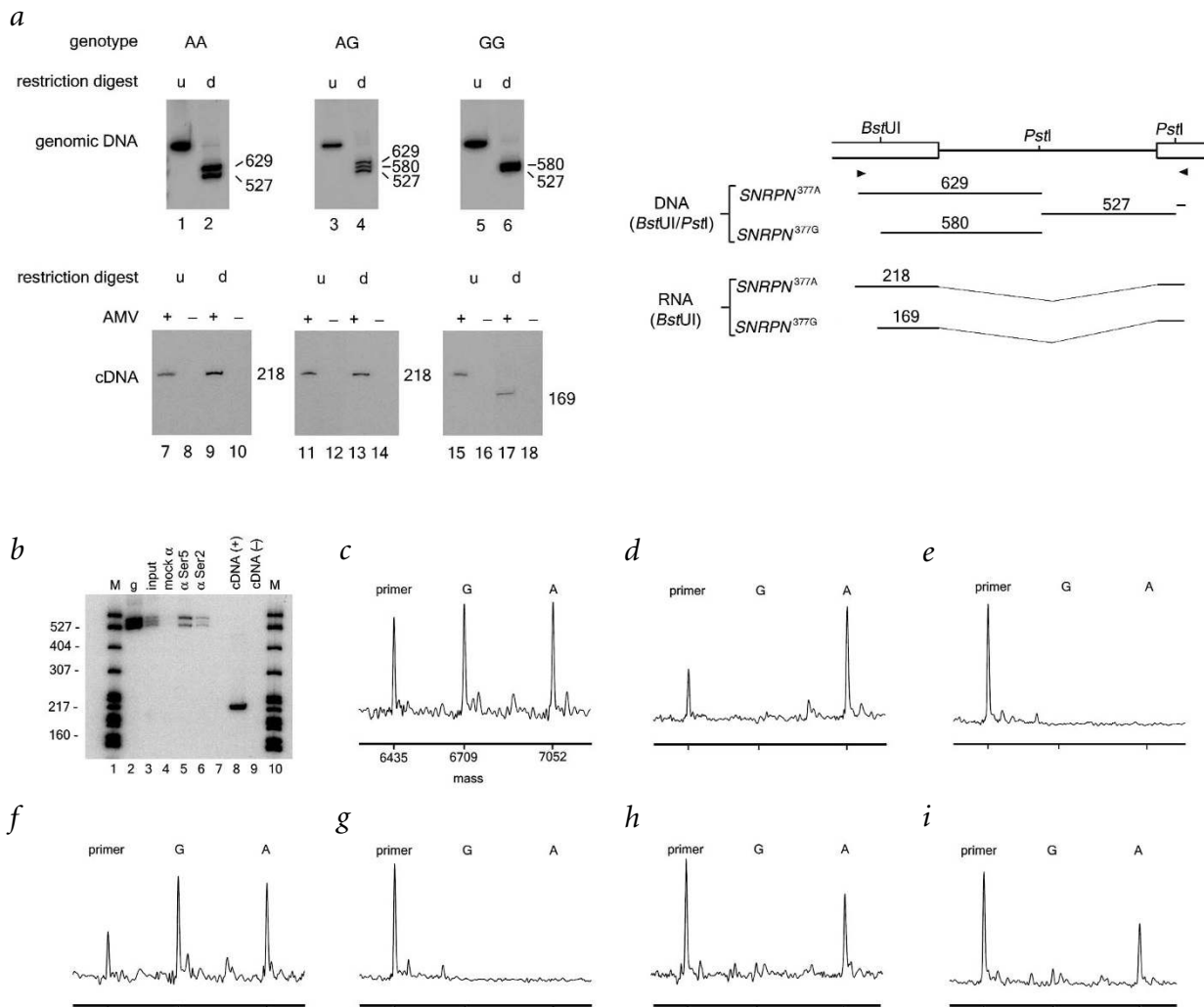
Fig. 1 Allele-specific loading of phosphorylated Pol II at imprinted gene *SNRPN*. *a*, Examples of different genotypes for the SNP at nt 377 of *SNRPN* identified by restriction-enzyme digestion of genomic DNA using *Bst*UI and *Pst*I (lanes 1–6, either before (u) or after (d) addition of restriction enzymes). Restriction-enzyme digestion using *Bst*UI of cDNA (lanes 7–18) synthesized from mRNA in the presence (+) or absence (–) of AMV reverse transcriptase showed that only one allele was expressed in the heterozygous cell line. *b–i*, Analysis of patterns of occupancy of Pol II at *SNRPN* exon 4 using chromatin immunoprecipitation for an unstimulated cell line derived from an individual who was heterozygous with respect to the SNP at nt 377 in *SNRPN*. *b*, We used restriction-enzyme digestion with *Bst*UI and *Pst*I to differentiate relative abundance of the two alleles in genomic DNA (lane 2); input chromatin used in chromatin immunoprecipitation reactions (lane 3); products of ChIP using antibodies to SV40 large T antigen (lane 4) as a mock antibody control or to phosphorylated Ser5 (lane 5) or Ser2 (lane 6) residues of Pol II CTD; and cDNA synthesized from mRNA in the presence (lane 8) or absence (lane 9) of AMV reverse transcriptase. Lanes 1 and 10 show a marker (M) derived from pBR322 DNA *Msp*I digestion. Sizes of marker bands are given in bp. *c–i*, To quantify relative levels of abundance of allele-specific fragments, we used PE/MS to analyze genomic DNA (*c*); cDNA synthesized from mRNA in the presence (*d*) or absence (*e*) of AMV reverse transcriptase; input chromatin used in chromatin immunoprecipitation reactions (*f*); products of ChIP using specific antibodies to SV40 large T antigen as a mock antibody control (*g*) or to phosphorylated residues of Pol II CTD at Ser5 (*h*) or Ser2 (*i*). Graphs show mass (Da) along the *x* axis and intensity of signal along the *y* axis. Primer peak (at 6435 Da) and peaks corresponding to G (6709 Da) and A (7052 Da) alleles are shown.

Fig. 2 Quantification of differential protein–DNA binding *in vivo* between two alleles of a gene in a cell that is heterozygous with respect to a marker SNP using haploChIP. Experimentally, we begin by crosslinking protein that is bound to DNA using formaldehyde. Chromosomes are then broken into chromatin fragments by sonication. After immunoprecipitation, we use PCR and primer extension to determine in an allele-specific manner the relative abundance of those gene-specific DNA fragments to which the DNA-binding protein had bound over the genome.
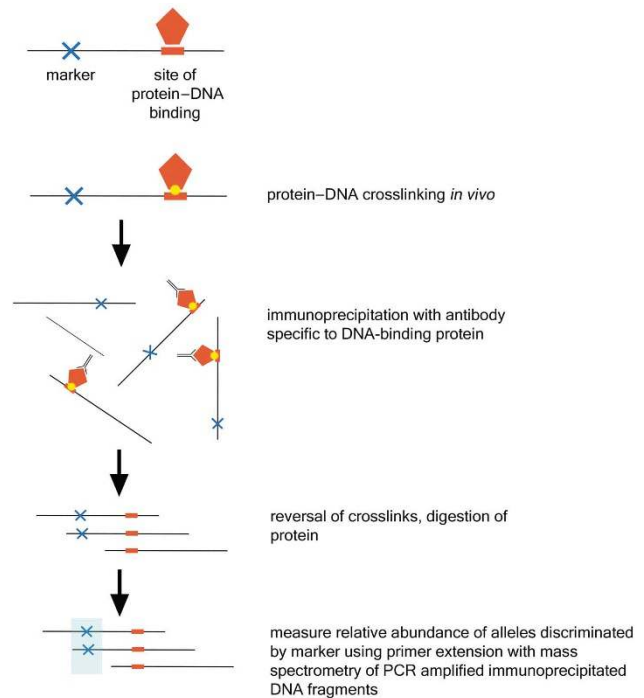
spectrometry allows unequivocal detection of a specific DNA product by its mass-to-charge ratio and avoids problems of secondary-structure formation that can affect detection by hybridization methods[21]. We amplified specific DNA sequences from a sample of immunoprecipitated chromatin by PCR and designed primers to anneal immediately adjacent to the SNP of interest. Primer-extension reactions containing the appropriate dideoxy-terminated nucleotides result in different sized products for each allele, typically different by 300 Da. The PE/MS method confirmed that in a cell line heterozygous with respect to the SNP at nt 377 in *SNRPN*, the two alleles had similar abundance in the total chromatin but differed markedly in abundance in the chromatin bound to phosphorylated Pol II and in the mRNA (Fig. 1*c–i*). We detected no allele-specific difference in chromatin immunoprecipitated with mock antibodies or in the products of RT–PCR reactions without reverse transcriptase.

To assess the accuracy and sensitivity of this method, we measured the area under the mass-spectrometry peak corresponding to each allele for different ratios of the two alleles for the SNP at nt 377 in *SNRPN* and SNPs in two other genes: at nt −308 in *TNF* and nt 252 in *LTA* (see Supplementary Fig. 1 online). We observed a close linear relationship between the true ratio of the alleles and the observed ratio of peak areas ($r^2 = 0.985$, 0.96 and 0.99, respectively, for the three SNPs). Analysis of genomic DNA and of total chromatin from heterozygous cell lines showed that the ratio between peak areas for the two alleles was close to the 1:1 ratio expected (see Supplementary Fig. 1 online). In later experiments, a SNP at nucleotide 723 in *LTA* showed a systematic allelic difference in PE/MS signal intensity: this was calibrated to obtain an accurate correction factor for the data below (see Supplementary Fig. 2 online).

Application of haploChIP to a candidate polymorphism
We applied the haploChIP method (Fig. 2) to a SNP whose functional significance has been widely debated. The cytokine tumor necrosis factor (encoded by *TNF*) has a pivotal role in inflammation, immunity and apoptosis. A SNP located 308 nt 5′ of the transcriptional start site has been associated with susceptibility to malaria, leishmaniasis, leprosy, meningococcal disease, asthma and other diseases[22–25]. But reporter-gene and other functional studies have provided conflicting information as to whether this polymorphism acts to modulate the level of *TNF* transcription[26,27].

We began by evaluating phosphorylated Pol II loading and *TNF* mRNA expression in EBV-transformed human B cells. After stimulation with 4-β-phorbol-12-myristate-13-acetate (PMA) and ionomycin, *TNF* mRNA levels increased after 30 minutes and peaked after 1 hour, causing substantial amounts of TNF to appear in the cell supernatant after 2 hours (Fig. 3*a,b*). Pol II recruitment at *TNF* was investigated by ChIP, analyzed using primer pairs as shown in Figure 3*c*. Loading of total Pol II and Pol II containing a CTD phosphorylated at Ser5 onto the region between −308 and +2 nt relative to the *TNF* transcriptional start site increased after 15 minutes and returned to baseline after 3 hours (Fig. 3*d*). For Pol II containing a CTD phosphorylated at Ser2, binding to the promoter region stayed relatively low,

whereas levels at downstream transcribed regions increased during the first hour after stimulation (Fig. 3*e*). This is consistent with other evidence that Ser5 and Ser2 are phosphorylated at different stages in the transcriptional process[4]. In subsequent experiments, we focused on phosphorylation of Ser5 as a marker of phosphorylated Pol II loading because it was more reliably immunoprecipitated and easier to interpret.

To explore the relationship between the SNP at nt −308 of *TNF* and loading of phosphorylated Pol II, we identified B-cell lines heterozygous with respect to the SNP from the CEPH collection[28] and repeated the above experiment, quantifying relative phosphorylated Pol II loading on the two alleles by haploChIP using PE/MS. Over a detailed time course in five different heterozygous cell lines, we found no significant difference between the two alleles discriminated by this SNP in loading of phosphorylated Pol II to the 5′ region of *TNF* (Fig. 4 and Supplementary Fig. 3 online).

Haplotype structure at *TNF/LTA*
The haploChIP analysis of the polymorphism at position −308 in *TNF* suggested that alleles resolved using this SNP did not have different levels of phosphorylated Pol II loading at *TNF* in the cell types and conditions of stimulation assayed. We next explored the haplotype(s) bearing this polymorphism more widely across the locus. This included a neighboring gene, *LTA*, that shares many biological and structural characteristics with *TNF*[29,30]. We determined the precise haplotypic structure of the five cell lines heterozygous with respect to the SNP at nt −308 in *TNF* by cloning and sequencing 10 kb of DNA encompassing *TNF* and *LTA* and flanking regions. This identified a total of 14 SNPs (Fig. 5). Of the 7 located in exons or introns, all were in *LTA* and none in *TNF*. In these cell lines, the 14 SNPs comprised only 3 haplotypes. We found that the allele *TNF*[−308A] was always accompanied by three polymorphisms of *LTA*: *LTA*[10A], *LTA*[252G] and *LTA*[723A]. We refer to this as haplotype A and to the other two haplotypes as B and C. Of the five cell lines described here, four were A/B heterozygotes and one was A/C.

Haplotype-specific polymerase loading at *LTA* correlates with transcript expression

We investigated the three haplotypes at *LTA* using heterozygous B-cell lines stimulated with PMA and ionomycin. Within 15 minutes of stimulation, the amount of total Pol II and of phosphorylated Pol II loaded onto *LTA* intron 1 increased, and *LTA* mRNA expression increased from its constitutive level within 60 minutes (Fig. 6*a*). To examine haplotype-specific loading of phosphorylated Pol II, we used the SNP at nt 252 of *LTA* as a marker. In all four A/B heterozygous cell lines, loading of phosphorylated Pol II was greater onto haplotype A versus B (Fig. 6*b–e*). This was most marked 30 minutes after stimulation (mean A:B ratio = 1.31; 95% confidence interval (c.i.) = 1.19–1.44). In contrast, loading of phosphorylated Pol II was not different between haplotypes A and C in the A/C heterozygous cell line (Fig. 6*f*).

In contrast to *TNF* in which we found no transcribed marker on the three haplotypes, haplotype A includes a marker in an *LTA* exon, allowing us to pursue this result in terms of *LTA* mRNA expression. Allele-specific RT–PCR analysis using the SNP at nt 723 in *LTA* showed that haplotypes A and B differed in mRNA expression in a similar way to that observed for loading

of phosphorylated Pol II (Fig. 6*b–e*). Thirty minutes after stimulation, the mean ratio of *LTA* mRNA expression from haplotype A versus B was 1.71 (95% c.i. = 1.42–2.01). Over the time course of stimulation, the A:B ratios for loading of phosphorylated Pol II and for mRNA abundance were significantly correlated (Spearman $r = 0.77$; $P = 0.05$). As with loading of phosphorylated Pol II, no difference in *LTA* mRNA synthesis was observed between haplotypes A and C (Fig. 6*f*).

Discussion

Analysis of TNF/LTA haplotypes. It is generally recognized that many genetic associations may derive from linked markers, necessitating exploration of haplotype structure at a locus. We report here a clear haplotype-specific difference in levels of mRNA transcript for *LTA*, but we found no transcribed marker with which to differentiate these haplotypes for *TNF*. The haplotype-specific difference in transcript at *LTA* was exactly mirrored by differential loading of phosphorylated Pol II, whereas we found no such difference at *TNF*. This raises several important issues. First, the absence of a transcribed marker at *TNF* would previously have precluded analysis of allele-specific
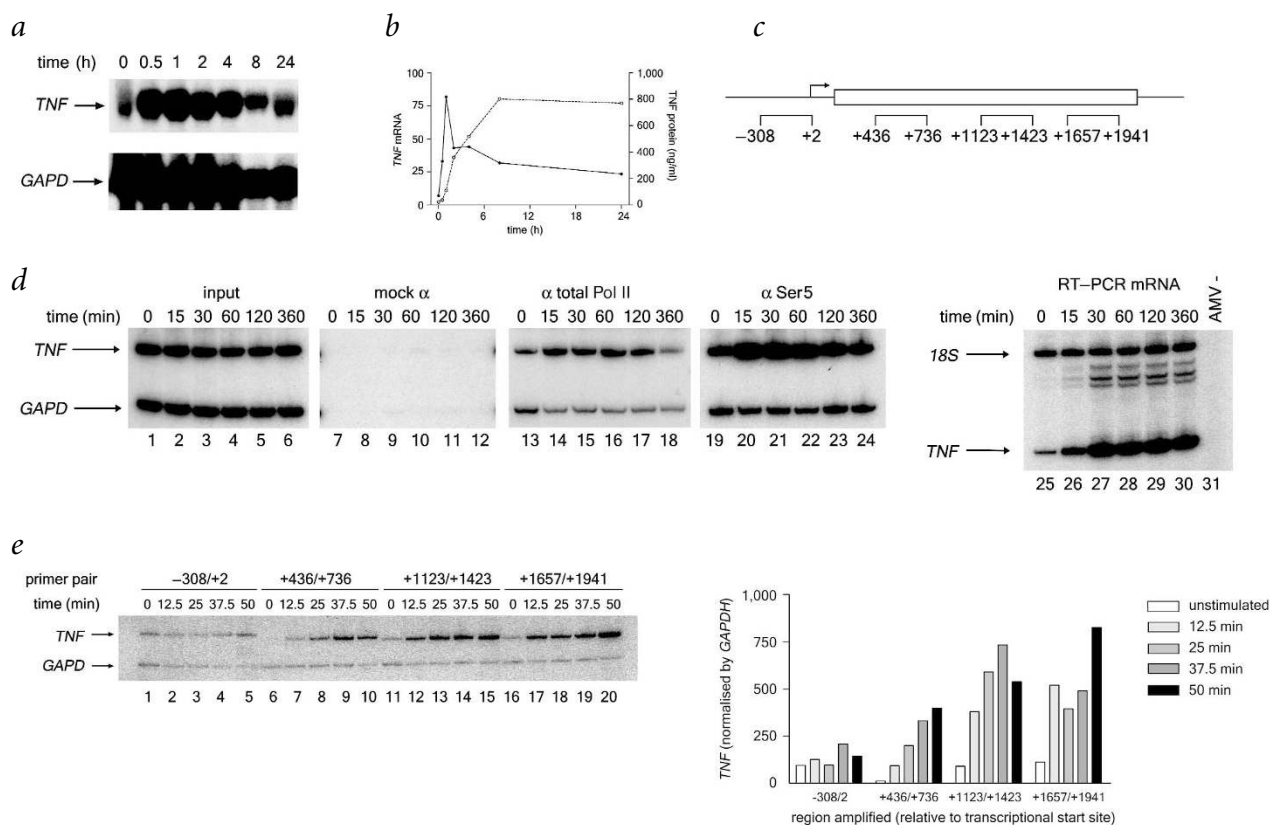
Fig. 3 *In vivo* loading of phosphorylated Pol II at *TNF*. **a**, Northern blot demonstrating time course of *TNF* and *GAPD* mRNA expression after stimulation with ionomycin and PMA in EBV-transformed B cells. **b**, Results from northern blot for *TNF* mRNA (normalized by *GAPD*; filled circles with solid connecting lines) derived using phosphoimager analysis are shown together with TNF protein concentration in supernatant after stimulation measured by ELISA (open circles with dashed connecting lines). **c**, Schematic of *TNF* showing location of primers used in chromatin immunoprecipitation analysis. **d**, Chromatin immunoprecipitation experiment using B cells stimulated with ionomycin and PMA (lanes 1–24) for the indicated times. PCR amplification (lanes 1–24) was done using primers spanning the *TNF* promoter region (-308/+2) multiplexed with *GAPD* promoter primer pair. Input (lanes 1–6) shows products of PCR reactions containing 0.5% of total amount of chromatin used in immunoprecipitations reactions. Mock immunoprecipitations using irrelevant control antibodies (against T Antigen pAb101) are shown (lanes 7–12); immunoprecipitations were done using antibodies specific for total Pol II (lanes 13–18) or for phosphorylated Pol II containing CTD phosphorylated at Ser5 (lanes 19–24). mRNA taken concomitantly with chromatin crosslinking at the indicated times after stimulation with ionomycin and PMA was analyzed by relative RT–PCR (lanes 25–31) using primers specific for *TNF* multiplexed with primers for *18S*; cDNA was generated with (lanes 25–30) or without AMV reverse transcriptase (lane 31). **e**, ChIP assay for phosphorylated Pol II containing CTD phosphorylated at Ser2 showed progressive accumulation along the *TNF* coding region as compared with the promoter after stimulation with ionomycin and PMA (lanes 1–20). ChIP assay was analyzed by multiplex PCR combining *TNF*-specific primers with *GAPD* promoter primer pair. Graph shows results of ChIP assay plotting *TNF* band intensity (normalized by *GAPD*) across *TNF* promoter/coding regions derived from phosphoimager analysis.

expression *in vivo*. In contrast, the haploChIP approach does not require a transcribed marker and thus allows interrogation of allele-specific effects that is broadly applicable. Second, the correlation of phosphorylated Pol II loading with transcript synthesis at *LTA* and *SNRPN* shows the potential of assaying for phosphorylated Pol II loading as a measure of allele-specific gene expression. Further studies are needed to examine the exact nature of this relationship, including, for example, investigation of other specific phosphorylated residues of the Pol II CTD, and whether the approach holds for all genes.

These data add to a growing body of information concerning the potential clinical and biological importance of *LTA*; studies in transgenic mice, for example, have recently implicated LTA in the pathogenesis of bacterial and parasitic infections[31,32], and a recent publication shows an *LTA* association with susceptibility to myocardial infarction in a large genomic association study[33]. Our data showing greater loading of phosphorylated Pol II and transcript synthesis for haplotype A versus B are consistent with previous reports that the *LTA*[252G] allele is associated with higher levels of LTA production[34,35]. The functional basis of this haplotypic difference is unresolved; prime candidates are six SNPs, located in or immediately upstream of *LTA*, that differentiate haplotypes A and B (Fig. 5), though more distant polymorphisms may be responsible. Our finding of no apparent functional difference between haplotypes A and C illustrates the utility of further haplotypic analysis to identify the functional polymorphism, as these two haplotypes are identical for three of the six *LTA* SNPs that differentiate haplotypes A and B, including that at nt 252.

Potential applications of haploChIP. HaploChIP has the advantage of analysis *in vivo* in a chromosomal environment of naturally occurring haplotypes, subject to all the regulatory processes that operate in the specific cell type investigated. The method uses a genetic marker to differentiate the binding of a protein between the chromosomal alleles in a cell. Protein–DNA complexes are crosslinked *in vivo* and then chromatin is fragmented by sonication. The size of the fragments defines the boundaries of the unit of analysis that the method can resolve; in this study, the average size was approximately 800 nt. We then use PCR and primer extension to detect sequences containing a genetic marker (Fig. 2). The sensitivity of the method in detecting allele-specific differences in protein loading is probably highest where this occurs close to the genetic marker. Any allele-specific difference observed ascribes functional significance to the haplotype defined by the genetic marker[36] but not necessarily to the marker itself.

It would be useful to apply this approach to screening on a genomic level for regulatory polymorphisms. Mass spectrometry is becoming the modality of choice for high-throughput genotyping[36], and its use for the quantification of haploChIP allows for the interrogation of large numbers of SNPs using panels of tissue-specific cell lines from different individuals and of primary cells. We have investigated allele-specific loading of phosphorylated Pol II in this paper, but the same general approach could be applied to any naturally occurring DNA–protein interaction. This would allow allele-specific modulation of binding by specific candidate factors to be interrogated *in vivo* at different stages of the transcriptional process.
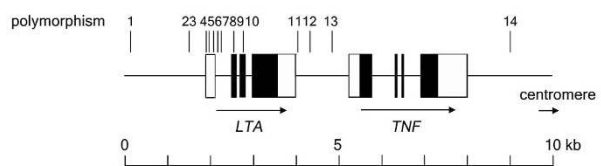
Methods

Cell culture, stimulation and cytokine assays. We obtained established EBV-transformed B cell lines from the Centre d'Etude du Polymorphisme Humain (CEPH) collection (Coriell Institute for Medical Research) and maintained them at 2×10^6 cells per ml in RPMI 1640 medium (Sigma) with 2 mM glutamine (Sigma) and 10% fetal calf serum (Sigma) at 37 °C
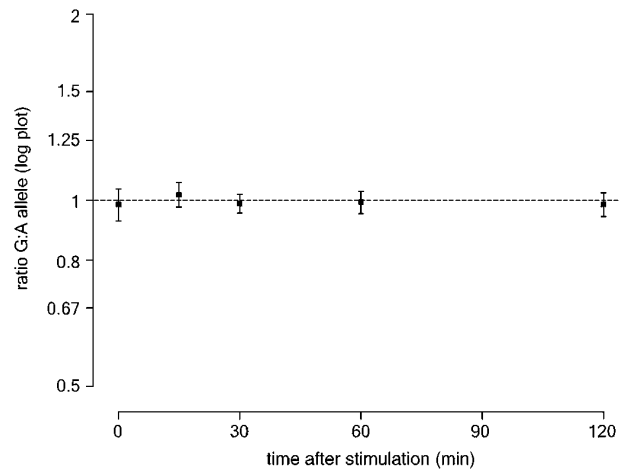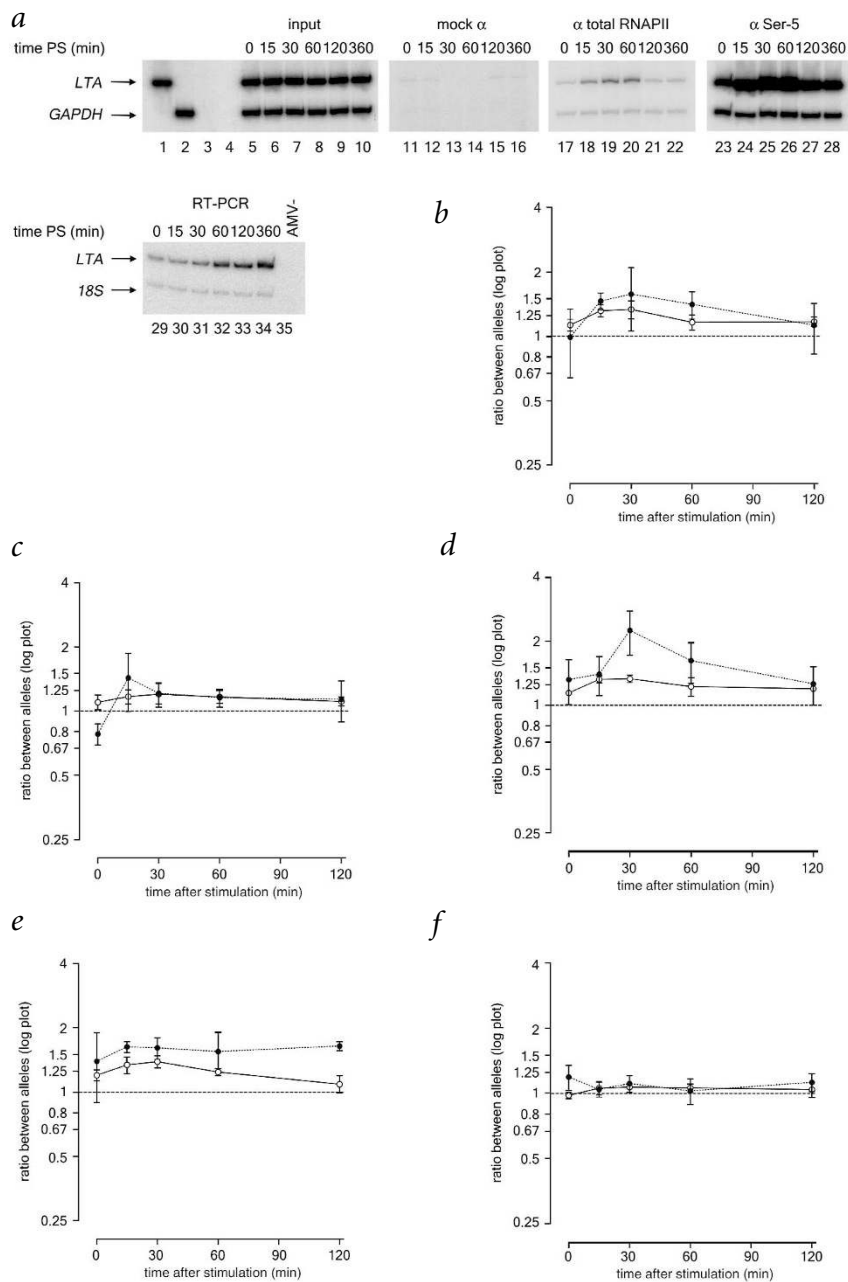

Fig. 4 Allele-specific loading of phosphorylated Pol II analyzed for the SNP at nt −308 of *TNF* by PE/MS. We stimulated five cell lines heterozygous for this SNP with ionomycin and PMA. The overall mean ratios (± s.d.) of phosphorylated Pol II loading between the A and G alleles were assayed using antibodies against Pol II CTD phosphorylated at Ser5. For each cell line, we determined the mean of three immunoprecipitation reactions for a given time point, with each immunoprecipitation analyzed by three independent PCR amplification reactions, which were in turn each spotted as four replicates on the chip before PE/MS.

and 5% CO_2. We stimulated cells with 200 nM PMA (Sigma) and 1 μM ionomycin (Sigma) unless otherwise stated. For northern-blot analysis of TNF, we isolated total RNA using Tri reagent (Sigma). We separated 15 μg RNA on a 1.4% agarose–1.8 M formaldehyde gel. We transferred products to Hybond-N[+] (Amersham Biosciences) then hybridized them with ^{32}P-labeled *TNF* cDNA corresponding to exon 4 (a gift from I. Udalova) and with a β-actin control (Qiagen). We measured amount of TNF by enzyme-linked immunosorbent assay as previously described[37]. For RT–PCR analysis of *TNF* and *LTA*, we isolated total RNA using an Absolutely RNA RT–PCR miniprep kit (Stratagene). We prepared cDNA using random decamers and AMV reverse transcriptase and then carried out relative quantitative RT–PCR (Ambion) according to manufacturer's instructions, multiplexing gene-specific primers with appropriate 18S rRNA primers and competimers. Cycling parameters were 94 °C for 5 min followed by 30

Fig. 5 Haplotypes at the *TNF/LTA* locus. Cloning and sequencing of a 10-kb region of the locus spanning *TNF* and *LTA* in five cell lines heterozygous for the SNP at nt −308 of *TNF* identified three major haplotypes. Designations of polymorphisms refer to the nucleotide position relative to the transcriptional start site.

	polymorphism	designation	contig location	haplotype A	B	C
1	TAA(TnAn)CAT		48892	8T5A	8T5A	7T6A
2	AAC[C]/TGCC	LTA−294	47622	C	C	T
3	ACC[G]/ACCT	LTA−293	47621	G	A	G
4	TCC[G]/ACAC	LTA+10	47318	A	G	G
5	CCC[C]/AGCA	LTA+80	47248	C	A	C
6	ATG[A]/GTTC	LTA+252	47076	G	A	A
7	GCT[G]/CTCT	LTA+368	46960	G	C	G
8	CTG(TCn)	LTA+438	46890	(TC)9	(TC)9	(TC)10
9	GTG[T]/CGTG	LTA+495	46833	T	T	C
10	GCA[C]/ACCT	LTA+723	46605	A	C	C
11	AGA[T]/CGAA	TNF−1031	45081	T	T	C
12	CCC[C]/ACTT	TNF−863	44913	C	C	A
13	ATG[G]/AGGA	TNF−308	44358	A	G	G
14	AGA[G]/ACAA		40549	A	G	G

Fig. 6 Allele-specific loading of phosphorylated Pol II *in vivo* at *LTA*. *a*, ChIP assay at the indicated times after stimulation with ionomycin and PMA. Input chromatin and immunoprecipitated products amplified by primers spanning *LTA* intron 1 (+160/+475) and *GAPD* promoter primers either singly (lanes 1,2) or as a multiplex (lanes 5–28) are shown. mRNA was taken concomitantly with chromatin crosslinking at the indicated times after stimulation with ionomycin and PMA and was analyzed by relative RT–PCR using primers specific for *LTA* multiplexed with primers for *18S* (lanes 29–35). *b–f*, Allele-specific transcription at *LTA*. The ratio of phosphorylated Pol II loading between the G and A alleles at the SNP at nt 252 of *LTA* (open circles with solid connecting lines) and the ratio of mRNA from the A and C alleles of the SNP at nt 723 of *LTA* (filled circles with dashed connecting lines) for each of five heterozygous B-cell lines stimulated with ionomycin and PMA are shown. For each cell line, the mean ± s.d. of three immunoprecipitation reactions for phosphorylated Pol II loading using antibodies against Pol II CTD phosphorylated at Ser5 are shown for a given time point. We analyzed each immunoprecipitation by three independent PCR amplification reactions, which were in turn each spotted as four replicates on the chip before PE/MS. For mRNA analysis, we amplified cDNA in four independent PCR amplification reactions, which were in turn each spotted as four replicates on the chip before PE/MS. Haplotypes were A/B for four of the cell lines (panels *b–e*) and A/C for one cell line (*f*).

We lysed cells, collected nuclei, sonicated and purified chromatin fragments on a cesium-chloride gradient and carried out dialysis as described previously[40] with the following modifications. To inhibit any potential phosphatase activity, we included 10 mM sodium pyrophosphate (Sigma) in all buffers. We sonicated nuclei at 4 °C using a microtip attached to a Branson 450 Sonifier in 30 s bursts (six times at setting 4 then six times at setting 5), cooling samples on ice for 1 min between pulses. For immunoprecipitation of chromatin, we used magnetic beads as described[41]. We incubated Dynabeads M-280 precoated with sheep antibody against mouse IgG (Dynal) overnight with mouse monoclonal antibodies at 4 °C in phosphate-buffered saline containing 5 mg ml⁻¹ bovine serum albumin (Sigma). We used specific antibodies against phosphorylated serine residues of the CTD of Pol II (Ser5, MMS-134R clone H14; Ser2, MMS-129R clone H5; Covance); we also did mock antibody controls using anti-T Antigen pAb101 (sc-147; Santa Cruz Biotechnology). For total Pol II chromatin immunoprecipitations, we incubated rabbit polyclonal antibody against Pol II (N-20; sc-899; Santa Cruz Biotechnology) against the N terminus of the large subunit of Pol II with Dynabeads M-280 precoated with sheep antibody against rabbit IgG (Dynal) and then added them to chromatin as described. After washing to remove unbound antibody, we incubated antibody beads overnight at 4 °C with 50 µg chromatin in 1× RIPA buffer[11] containing 10 mM sodium pyrophosphate, 1× complete protease inhibitors (Roche) and 5 µg ml⁻¹ pepstatin (Roche) on a nutator. We washed bead–chromatin immunoprecipitations as described[11] with minor modifications. We washed immunoprecipitates twice with 1× RIPA buffer, twice with 1× RIPA buffer containing 100 µg ml⁻¹ salmon sperm DNA (Promega) for 5 min with rotation, twice with 1× RIPA buffer containing 300 mM NaCl final plus 100 µg ml⁻¹ salmon sperm DNA for 5 min with rotation and once with 1× RIPA buffer containing 250 mM LiCl (Sigma). For total Pol II, we washed immunoprecipitates seven times in 1× RIPA buffer. We resuspended beads in 100 µl of TE digestion buffer containing 0.5% SDS with 1 µg RNaseA (Roche) and 2 µg proteinase K (Roche) and incubated them at 55 °C for 3 h and then for 12 h at 65 °C to

cycles (for *TNF*) or 24 cycles (for *LTA*) of 94 °C for 10 s, 57 °C for 30 s and 72 °C for 60 s. Primer sequences are available on request. We resolved RT–PCR products by 6% PAGE and quantified them using the Cyclone storage phosphor system (PerkinElmer).

Genotyping and sequencing. We carried out long-range PCR from genomic DNA stocks using the 'Expand' (BI) Long-Range system (Roche) to amplify a 10-kb region spanning *LTA* and *TNF* (from 39231 to 49111 on accession number Y14678). We cloned PCR products using TOPO-XL (Invitrogen) and then sequenced them using the Big-Dye terminator kit on an ABI3700 (Applied Biosystems). We used polyphred alignment to create haplotypes from each hemizygous clone. We typed the SNP at nt 377 of *SNRPN* using *Bst*UI and *Pst*I as previously described[38] and the SNP at nt –308 of *TNF* by amplification refractory mutation system as previously described[39].

Chromatin immunoprecipitation. We crosslinked 5×10^8 cells using formaldehyde (1% final concentration) for 45 min at room temperature and quenched reactions by adding glycine (final concentration 0.125 M).

reverse crosslinks. We extracted DNA using phenol–chloroform and precipitated it with ethanol in the presence of 20 μg glycogen. We resuspended pellets in 100 μl TE buffer and assayed contents by semiquantitative PCR as described[40], multiplexing with *GAPD* promoter primer pair (primer sequences available on request).

Primer extension and mass spectrometry. We carried out first-round PCR using 5 ng genomic DNA or 5 μl ChIP DNA in a 25 μl reaction volume using 0.5 U BioTaq (Bioline) with 0.8 mM dNTPs, 1.9 mM $MgCl_2$ and 0.2 μM each primer (see Supplementary Table 1 online). Thermal cycling using an MJ Tetrad was 96 °C for 1 min; six cycles of 94 °C for 45 s, 56 °C for 45 s, 72 °C for 30 s; 30 cycles of 94 °C for 45 s, 65 °C for 45 s, 72 °C for 30 s; and a final extension at 72 °C for 10 min. We divided each first-round product into 4 wells on a 384-well plate and then removed unincorporated dNTPs using shrimp alkaline phosphatase by incubating at 37 °C for 20 min and then at 85 °C for 5 min. We carried out primer extension using a homogeneous MassEXTEND (Sequenom) reaction comprising a cocktail of 100 μM extension primer (see Supplementary Table 1 online), 0.576 U MassEXTEND enzyme, buffer and a termination mix of dTTP and ddATP, ddCTP and ddGTP (except for the SNP at nt 723 of *LTA*, for which we used dGTP, ddATP, ddCTP and ddTTP). We carried out primer extension at 94 °C for 2 min and then 40 cycles of 94 °C for 5 s, 52 °C for 5 s and 72 °C for 5 s. We desalted products of primer extension using SpectroCLEAN (Sequenom) resin and transferred them onto a SpectroCHIP (Sequenom) microarray by SpectroPOINT (Sequenom) nanoliter dispenser. We carried out MALDI–TOF analysis using a SpectroREADER (Sequenom) mass spectrometer.

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

Competing interests statement

The authors declare that they have no competing financial interests.

1. Bartholomew, B., Dahmus, M.E. & Meares, C.F. RNA contacts subunits IIo and IIc in HeLa RNA polymerase II transcription complexes. *J. Biol. Chem.* **261**, 14226–14231 (1986).
2. Laybourn, P.J. & Dahmus, M.E. Transcription-dependent structural changes in the C-terminal domain of mammalian RNA polymerase subunit IIa/o. *J. Biol. Chem.* **264**, 6693–6698 (1989).
3. O'Brien, T., Hardin, S., Greenleaf, A. & Lis, J.T. Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation. *Nature* **370**, 75–77 (1994).
4. Komarnitsky, P., Cho, E.J. & Buratowski, S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14**, 2452–2460 (2000).
5. Dahmus, M.E. Reversible phosphorylation of the C-terminal domain of RNA polymerase II. *J. Biol. Chem.* **271**, 19009–19012 (1996).
6. Schroeder, S.C., Schwer, B., Shuman, S. & Bentley, D. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev.* **14**, 2435–2440 (2000).
7. Yamamoto, S. *et al.* Studies of nematode TFIIE function reveal a link between Ser5 phosphorylation of RNA polymerase II and the transition from transcription initiation to elongation. *Mol. Cell. Biol.* **21**, 1–15 (2001).
8. Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J. & Buratowski, S. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev.* **15**, 3319–3329 (2001).
9. Cadena, D.L. & Dahmus, M.E. Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. *J. Biol. Chem.* **262**, 12468–12474 (1987).
10. Weeks, J.R., Hardin, S.E., Shen, J., Lee, J.M. & Greenleaf, A.L. Locus-specific variation in phosphorylation state of RNA polymerase II *in vivo*: correlations with gene activity and transcript processing. *Genes Dev.* **7**, 2329–2344 (1993).
11. Nissen, R.M. & Yamamoto, K.R. The glucocorticoid receptor inhibits NFκB by interfering with serine-2 phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **14**, 2314–2329 (2000).
12. Ozcelik, T. *et al.* Small nuclear ribonucleoprotein polypeptide N (*SNRPN*), an expressed gene in the Prader–Willi syndrome critical region. *Nat. Genet.* **2**, 265–269 (1992).
13. Reed, M.L. & Leff, S.E. Maternal imprinting of human *SNRPN*, a gene deleted in Prader–Willi syndrome. *Nat. Genet.* **6**, 163–167 (1994).
14. Patturajan, M. *et al.* Growth-related changes in phosphorylation of yeast RNA polymerase II. *J. Biol. Chem.* **273**, 4689-4694 (1998).
15. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).
16. Feil, R. & Khosla, S. Genomic imprinting in mammals: an interplay between chromatin and DNA methylation? *Trends Genet.* **15**, 431–435 (1999).
17. Hu, J.F., Oruganti, H., Vu, T.H. & Hoffman, A.R. The role of histone acetylation in the allelic expression of the imprinted human insulin-like growth factor II gene. *Biochem. Biophys. Res. Commun.* **251**, 403–408 (1998).
18. Saitoh, S. & Wada, T. Parent-of-origin specific histone acetylation and reactivation of a key imprinted gene locus in Prader–Willi syndrome. *Am. J. Hum. Genet.* **66**, 1958–1962 (2000).
19. Braun, A., Little, D.P. & Koster, H. Detecting *CFTR* gene mutations by using primer oligo base extension and mass spectrometry. *Clin. Chem.* **43**, 1151–1158 (1997).
20. Jurinke, C., van den Boom, D., Cantor, C.R. & Koster, H. Automated genotyping using the DNA MassArray technology. *Methods Mol. Biol.* **170**, 103–116 (2001).
21. Griffin, T.J. & Smith, L.M. Single-nucleotide polymorphism analysis by MALDI–TOF mass spectrometry. *Trends Biotechnol.* **18**, 77–84 (2000).
22. McGuire, W., Hill, A.V.S., Allsopp, C.E.M., Greenwood, B.M. & Kwiatkowski, D. Variation in the TNF-α promoter region associated with susceptibility to cerebral malaria. *Nature* **371**, 508–511 (1994).
23. Cabrera, M. *et al.* Polymorphism in tumor necrosis factor genes associated with mucocutaneous leishmaniasis. *J. Exp. Med.* **182**, 1259–1264 (1995).
24. Moffatt, M.F. & Cookson, W.O. Tumour necrosis factor haplotypes and asthma. *Hum. Mol. Genet.* **6**, 551–554 (1997).
25. Shaw, M.A. *et al.* Association and linkage of leprosy phenotypes with HLA class II and tumour necrosis factor genes. *Genes Immun.* **2**, 196–204 (2001).
26. Wilson, A.G., Symons, J.A., McDowell, T.L., McDevitt, H.O. & Duff, G.W. Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation. *Proc. Natl. Acad. Sci. USA* **94**, 3195–3199 (1997).
27. Brinkman, B.M.N., Zuijdgeest, D., Kaigzel, E.L., Breedveld, F.C. & Verweij, C.L. Relevance of the tumor necrosis factor α-308 promoter polymorphism in TNFα gene regulation. *J. Inflamm.* **46**, 32–41 (1996).
28. Dausset, J. *et al.* Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* **6**, 575–577 (1990).
29. Nedwin, G.E. *et al.* Human lymphotoxin and tumor necrosis factor genes: structure, homology and chromosomal location. *Nucleic Acids Res.* **13**, 6361–6373 (1985).
30. Paul, N.L. & Ruddle, N.H. Lymphotoxin. *Annu. Rev. Immunol.* **6**, 407–438 (1988).
31. Roach, D.R. *et al.* Secreted lymphotoxin-α is essential for the control of an intracellular bacterial infection. *J. Exp. Med.* **193**, 239–246 (2001).
32. Engwerda, C.R. *et al.* Locally up-regulated lymphotoxin-α, not systemic tumor necrosis factor α, is the principle mediator of murine cerebral malaria. *J. Exp. Med.* **195**, 1371–1377 (2002).
33. Ozaki, K. *et al.* Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650–654 (2002).
34. Messer, G. *et al.* Polymorphic structure of the tumor necrosis factor (TNF) locus: an *Nco*I polymorphism in the first intron of the human TNF-β gene correlates with a variant amino acid in position 26 and a reduced level of TNF-β production. *J. Exp. Med.* **173**, 209–219 (1991).
35. Whichelow, C.E., Hitman, G.A., Raafat, I., Bottazzo, G.F. & Sachs, J.A. The effect of *TNFB* gene polymorphism on TNF-α and -β secretion levels in patients with insulin-dependent diabetes mellitus and healthy controls. *Eur. J. Immunogenet.* **23**, 425–435 (1996).
36. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
37. Kwiatkowski, D. *et al.* TNF concentration in fatal cerebral, non-fatal cerebral, and uncomplicated *Plasmodium falciparum* malaria. *Lancet* **336**, 1201–1204 (1990).
38. Sakatani, T. *et al.* Epigenetic heterogeneity at imprinted loci in normal populations. *Biochem. Biophys. Res. Commun.* **283**, 1124–1130 (2001).
39. Richardson, A. *et al.* Nucleotide diversity of the TNF gene region in an African village. *Genes Immun.* **2**, 343–348 (2001).
40. Takahashi, Y., Rayman, J.B. & Dynlacht, B.D. Analysis of promoter binding by the E2F and pRB families *in vivo*: distinct E2F proteins mediate activation and repression. *Genes Dev.* **14**, 804–816 (2000).
41. Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).