

## Accepted Manuscript

An Intronic Polymorphism of IRF4 Gene Influences Gene Transcription in vitro and Shows a Risk Association with Childhood Acute Lymphoblastic Leukemia in Males

Thuy N. Do, Esma Ucisik-Akkaya, Charronne F. Davis, Brittany A. Morrison, M. Tevfik Dorak

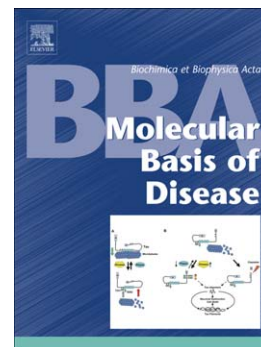
PII: S0925-4439(09)00251-8  
DOI: doi:[10.1016/j.bbadis.2009.10.015](https://doi.org/10.1016/j.bbadis.2009.10.015)  
Reference: BBADIS 63031

To appear in: *BBA - Molecular Basis of Disease*

Received date: 12 August 2009  
Revised date: 27 October 2009  
Accepted date: 28 October 2009

Please cite this article as: Thuy N. Do, Esma Ucisik-Akkaya, Charronne F. Davis, Brittany A. Morrison, M. Tevfik Dorak, An Intronic Polymorphism of IRF4 Gene Influences Gene Transcription in vitro and Shows a Risk Association with Childhood Acute Lymphoblastic Leukemia in Males, *BBA - Molecular Basis of Disease* (2009), doi:[10.1016/j.bbadis.2009.10.015](https://doi.org/10.1016/j.bbadis.2009.10.015)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



**An Intronic Polymorphism of IRF4 Gene Influences Gene  
Transcription in vitro and Shows a Risk Association  
with Childhood Acute Lymphoblastic Leukemia in Males**

**Thuy N Do, Esma Ucisik-Akkaya, Charronne F Davis, Brittany A Morrison,  
M Tefvik Dorak**

Genomic Immunoepidemiology Laboratory  
HUMIGEN LLC, The Institute for Genetic Immunology  
2439 Kuser Road  
Hamilton, NJ 08690-3303  
USA

**Corresponding author**

M. Tefvik DORAK, MD PhD  
Genomic Immunoepidemiology Laboratory  
HUMIGEN LLC, The Institute for Genetic Immunology  
2439 Kuser Road  
Hamilton, NJ 08690-3303  
USA

Phone: 1-609-570-1032

Fax: 1-609-570-1039

E-mail: [m.t.dorak@humigen.org](mailto:m.t.dorak@humigen.org)

**Abstract**

The interferon regulatory factor (IRF) family of DNA-binding proteins regulates expression of interferon-inducible genes with roles in the immune response and carcinogenesis. IRF4 is involved in the differentiation of B and T cells and is overexpressed in B-cell malignancies as a result of c-REL (NF- $\kappa$ B) hyperactivation. IRF4 polymorphisms are associated with susceptibility to chronic lymphoid leukemia (CLL) and non-Hodgkin lymphoma (NHL). We examined 13 IRF4 SNPs in 114 cases of childhood acute lymphoblastic leukemia (ALL) and 388 newborn controls from Wales (U.K.) using TaqMan assays. IRF4 intron 4 SNP rs12203592 showed a male-specific risk association (OR = 4.4, 95% CI = 1.5 to 12.6,  $P = 0.007$ ). Functional consequences of the C>T substitution at this SNP were assessed by cell-based reporter assays using three different cell lines. We found a repressive effect of the rs12203592 wildtype allele C on IRF4 promoter activity ( $P < 0.001$ ) but no repression by the variant allele in any cell line tested. Thus, homozygosity for the rs12203592 variant allele would result in increased IRF4 expression. This increase would be compounded by high levels of NF- $\kappa$ B activity in males due to the absence of estrogen. IRF4 differs from other IRFs in its anti-interferon activity which interferes with immune surveillance. We propose that a detailed study of IRF4 can provide information on the mechanism of the sex effect and the role of immune surveillance in childhood ALL development.

**Key words**

Genetic Predisposition to Disease

Lymphoblastic Leukemia, Acute, Childhood

Interferon Regulatory Factor 4

Polymorphism, Single Nucleotide

Luciferase Reporter Assay

Sex Effect

## Introduction

The interferon regulatory protein (IRF) family consists of at least nine proteins [1]. These proteins have roles in B- and T-lymphocyte differentiation [2] and maturation as well as in regulation of interferon-induced gene expression. One member of this family, IRF4 (also known as multiple myeloma 1 (MUM1) and lymphocyte-specific interferon regulatory factor (LSIRF)) functions as a transcription factor in lymphocyte differentiation [3, 4]. IRF4 expression is restricted to lymphocytes of the B-cell type (pre-B, B and plasma cells) and mature T cells [5]. IRF4 is highly expressed in multiple myeloma due to t(6;14)(p25;q32) translocation [6] but it is overexpressed in most B-cell malignancies [7-9] even without a chromosomal rearrangement. IRF4 differs from other IRFs in several ways: its expression is induced by c-REL [10]; HTLV1 Tax protein induces leukemic transformation via induction of IRF4 [5]; it represses IFN-induced gene expression [5], acts as a negative regulator of TLR signaling [11] and increases telomerase activity in B-lymphocytes [12]. Thus, while IRF family may be considered as a major player in cancer immune surveillance, IRF4 appears to have a pro-carcinogenic activity [13].

The IRF4 gene is located at the tip of the short arm of chromosome 6 (6p25) and recent genome-wide association studies found associations between IRF4 polymorphisms and several conditions. Most importantly, 3' UTR SNPs rs872071 and rs12211228 are associated with chronic lymphoid leukemia (CLL) [14] and non-Hodgkin lymphoma risk [15], respectively. An intron 7 SNP (rs2001508) has been found to correlate with height in subjects with type 2 diabetes [16]. Another finding is that the intron 4 SNP rs12203592 is a marker for population substructure in Europe and associated with eye color and particularly with hair color [17, 18]. A different SNP located in 3' UTR of the IRF4 gene (rs9378805) is a marker for UK population substructure [19].

Given the possible role played by IRF4 in B-cell development and associations between its polymorphisms and hematologic malignancies, we included the IRF4 gene in our

childhood acute lymphoblastic leukemia (ALL) candidate gene association studies. Here, we provide a comprehensive analysis of IRF4 polymorphisms in a case-control study and functional assessment of an intronic SNP association.

## **Materials and Methods**

### ***Genetic Association Studies***

#### *Case and control groups*

The cases analyzed were 114 childhood ( $\leq 16$  years) ALL samples consecutively diagnosed from 1988 to 1999 in South Wales (U.K.) and originally used to describe the male-specific HLA-DRB4 and C282Y associations [20, 21]. Controls were 388 anonymously collected cord blood samples also from South Wales. The control samples were collected in two hospitals in Cardiff in 1997 and 1998 without any selection other than they were all naturally born full-term healthy newborns. Further characteristics of these samples were described elsewhere [22]. We only had information on age-at-onset, sex and ALL subtype in cases. Thus, we could only examine primary susceptibility with age and sex effects, but not associations with clinical outcome. The samples were used with the approval of the South Wales (U.K.) Research Ethics Committee.

#### *SNP selection and genotypings*

We used the HapMap data and NCBI ENTREZ SNP to select SNPs that will provide maximum information about the polymorphism content about the IRF4 gene and 10kb flanking region on each side. HapMap lists 11 haplotype tagging SNPs (htSNP) for approximately 40kb region that includes IRF4. We genotyped all but two htSNPs (rs4959234 and rs9391997) in the IRF4 region. We included all IRF4 SNPs that have shown associations with any condition as reported by Johnson & O'Donnell [23], National Human Genome Research Institute (NHGRI) Catalog of Published Genome-Wide Association Studies [24] and HuGE Navigator [25]. The resulting 18 SNPs covered a genomic region of 35kb including at

least 6kb on each side of the IRF4 gene. Thus, the selected SNP list contains all SNPs in the IRF4 gene region with known associations with a disease or trait.

Although NCBI ENTREZ SNP (build 129) lists seven coding region SNPs (three missense ones) for IRF4, but none has double hits (i.e., they are not confirmed polymorphisms) or frequency information (i.e., not observed in any population sample). We still went ahead and genotyped one of them, rs9378788 in exon 2 encoding L70L in our control group but it did not show any polymorphism. We did not proceed with genotyping of the other six unconfirmed coding region SNPs.

We also screened other publicly available resources for selection of functional SNPs of IRF4. The list of SNPs that show strong correlation with gene expression levels does not contain any from the IRF4 genomic region [26]. Examination of bioinformatics resources like SNP500Cancer, SNPs3D, FastSNP, SNPSeek and TFSEARCH did not result in addition of any other potentially functional SNP.

Two IRF4 SNPs have been identified as geographic differentiation markers in Europe (rs12203592) [18, 27] and in the UK (rs9378805) [19]. Since this feature of IRF4 SNPs may cause spurious associations in case of cryptic population substructure, we included the top ranking marker (lactase gene SNP rs1042712) for population substructure in the UK identified by the Wellcome Trust Case-Control Consortium Study [19].

All genotypings were achieved by TaqMan allelic discrimination assays purchased from Applied Biosystems (ABI, Foster City, CA) on Stratagene Mx3000 real-time PCR instruments using standard allelic discrimination assay conditions. TaqMan assay numbers are provided in Table 1. Stratagene Mx3000P software was used to assign genotypes. Safeguards against genotyping errors included inclusion of blind duplicates (intra-plate, inter-plate), negative controls in each assay, randomly repeated samples and examination of the data for unexpected break of haplotypic homozygosity in homozygous samples. We also repeated genotypings when haplotype construction yielded rare haplotypes due to one SNP

in that haplotype. At the analytical phase, genotype distributions were examined for their compliance with Hardy-Weinberg equilibrium proportions. The SNP that showed an association (rs12203592) was genotyped for a second time by a different operator.

#### *Data analysis for genetic associations*

Statistical analysis was performed on Stata/IC v.10 (StataCorp, College Station, TX). We used the Stata command “genhwcci” which tests the Hardy-Weinberg equilibrium (HWE). Statistical significance of associations with minor allele-positivity (dominant model) was assessed by logistic regression (command “logistic”) and odds ratios (OR) and their 95% confidence intervals (CI) were obtained. We also used gene-dosage effect analysis by the genotype-based trend test (additive model) which reveals associations that depend additively upon the minor allele. This test compares the heterozygote and then variant homozygote frequencies with wildtype homozygote frequency, and yields genotype-specific odds ratios as well as an odds ratio per allele for stepwise changes with the number of minor allele in the genotype (OR per allele). Using the wild type genotype as reference, independence of associations was assessed by multivariable logistic regression. All *P* values presented are two-sided. Linkage disequilibrium (LD) analysis, LD plot generation and haplotype construction were achieved on Haploview v4.0 (<http://www.broad.mit.edu/haploview/haploview>). Haploview also provided statistical assessment of allele frequency comparisons between cases and controls (corresponding to multiplicative model).

#### ***Functional Assessment of the Intronic Risk Marker***

For further details of the methods described below, see the Supplementary Material.

#### *Cells and reagents*

The EBV-positive human Burkitt lymphoma B cells (Raji) (#CCL-86), the human adrenal NCI-H295R cells (#CRL-2128), and the human embryonic kidney HEK-293T (#CRL-1573) were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA).

#### *Amplification of IRF4 promoter and IRF4 intron 4 fragments from genomic DNA*

The PCR was used to amplify IRF4 2.4kb promoter fragment and its full 1.2kb intron 4 from two different cell lines (PGF (IHWG #9318) and LWAGS (IHWG #79)). Cell line PGF is wildtype homozygous (CC) for our SNP of interest (rs12203592) in intron 4 and cell line LWAGS is homozygous for the variant allele (TT).

Four IRF4 intron 4 fragments from the two cell lines were obtained by PCR. All four fragments from each cell line were identical except at their terminal sequences where unique endonuclease restriction sites were introduced. These unique restriction sites facilitated the sub-cloning of intron 4 in the sense and antisense direction both proximal and distal to the luciferase reporter gene.

#### *Construction of expression vectors*

All of *Taq* polymerase-amplified PCR products (IRF4 promoter and intron 4) were first subcloned into the pCR<sup>®</sup>2.1-TOPO vector (Invitrogen) and transformed to TOP10 competent cells (Invitrogen). At this step, all PCR amplified products were fully sequenced in both directions and all sequences are identical to the sequences published in GenBank.

The IRF4 promoters were liberated from TOPO vector using *XhoI* and *HindIII* enzymes. Intron 4 fragments were released from TOPO vectors using the restriction enzymes. *KpnI* and *XhoI* for intron 4 fragments to be cloned 5' immediately upstream of IRF4 promoter and *Sall* and *BamHI* for those to be downstream of the luciferase gene.

The IRF4 promoters were subcloned into the *XhoI*-and-*HindIII*-digested pGL4.10[luc2] luciferase vector (Promega). Since the sequences of both IRF4 promoters from two different cell lines were identical and their transcriptional activities were confirmed to be similar in all three cell lines (see Results), IRF4 promoter from the PGF cell line was used for further experiments with different IRF4 intron 4 fragments.

#### *Site-directed mutagenesis*

Construct C\_D(+) was used to change the wild type allele C in intron 4 to the variant allele T (C\_D(+)CtoT) using a Change-IT multiple-mutation site-directed mutagenesis kit (USB,



Cleveland, Ohio, USA) with phosphorylated primers ordered from Integrated DNA Technology (IDT, Coralville, IA, USA). The fidelity of all constructions was verified by sequencing in both directions, and after that, they were purified using Qiagen columns.

#### *Sequencing*

Sequencing analysis was performed with the CEQ dye terminator cycle sequencing kit on the CEQ8000 Genetic Analysis System (Beckman Coulter, Fullerton, CA, USA).

#### *Transient transfection and luciferase assay*

For Raji cells, transfection was performed with Nucleofector™ Technology (Lonza, Walkersville, MD, USA), with  $2 \times 10^6$  logarithmically growing cells nucleofected using a Cell Line Nucleofector™ Kit V, program M-13 and 2 µg of plasmid in each well of a 12-well plate. NCI-H295R and HEK-293T cells were plated at a density of  $10^6$  or  $10^5$  cells/well in each well of a 12-well plate, respectively, and incubated for 24 h before transfection using the Lipofectamine™2000 Transfection Reagent (Invitrogen, Carlsbad, CA, USA). The relative light units at 48 hours after incubation were similar to that of the 24 hours; therefore, 24 hours was chosen for all experiments. Cells were transfected with either luciferase vector driven by IRF4 promoter or control plasmids (pGL4.10[luc2] and pGL4.13[luc2/SV40]). Co-transfection of the *Renilla* luciferase reporter pGL4.73[hRenilla/SV40] (Promega) was used as a control for transfection efficiency in a 25:1 molar ratio (firefly:renilla). Where necessary, pcDNA3.1[cmvP] was used to make up the total amount of DNA in the transfection. Cells were harvested and the luciferase activity was measured in cellular extracts using the Dual-Glo Luciferase Reporter Assay System (Promega, Madison, WI, USA). The firefly luciferase activity was normalized by the *Renilla* luciferase, and the result was expressed as relative luciferase activity. All constructs were transfected in triplicates, and the results represent the mean of three independent experiments  $\pm$  SE. All statistical analyses were done using non-parametric Mann-Whitney test.

#### *Oligonucleotide Synthesis for DNA-protein binding assay*

In silico analyses suggested the binding of the transcription factor AP-2 $\alpha$  to the SNP that was found to be the risk marker (see Results), we designed a DNA-protein binding assay to experimentally test this prediction.

#### *DNA-Protein Binding Assay*

HEK-293T cells at 80-90% confluent were harvested, and cytoplasmic- and nuclear-extracts were prepared using the ProteoJet™ Cytoplasmic and Nuclear Protein Extraction Kit (Fermentas, Germany). NanoDropND-1000 Spectrophotometer (Nano-Drop Technologies, Wilmington, DE, USA) was used to estimate protein concentration. DNA-protein binding assays were performed using a streptavidin-agarose bead pulldown assay [28]. Relative intensity of Western blot bands was quantified using the GS-800 calibrated densitometer (BioRad, CA, USA).

#### *Western Blotting*

The eluted proteins were subjected to a 4-12% gradient Bis-Tris gel (BioRad, Hercules, CA, USA) by electrophoresis and transferred electrophoretically to a polyvinylidene difluoride membrane (Millipore Corp., Bedford, MA, USA). The membrane was then blocked in a blocking solution (5% nonfat dry milk/Tris-buffered solution) and then incubated with primary antibody overnight. For both AP-2 $\alpha$  and Tst-1 (Oct-6) detection, a 1:1000 dilution of an anti-AP-2 $\alpha$  polyclonal antibody (sc-8975) or an anti-Oct-6 (sc-28593) (Santa Cruz Biotechnology Inc., Santa Cruz, CA) were used. The unbound primary antibody was removed by washing the membrane with 0.1 % Tween/Tris-buffered solution, followed by incubation with horseradish peroxidase-conjugated secondary antibody diluted 1:3000 in 0.5% nonfat dry milk/Tween/Tris-buffered solution. Proteins were visualized using enhanced chemiluminescence reagent (Amersham Pharmacia Biotech) and X-ray film.

## **Results**

### *Cases and controls*

The case-control group consisted of 114 childhood ( $\leq 16$  yr) ALL cases (62 males and 52 females) and 388 newborn controls (201 girls and 187 boys) collected at overlapping time periods from the same geographic region. This case-control set has been used in association studies before [20, 21], and more recently, TP53 [29] and HSPA1B [22] association were also reported. Of the 114 cases, 93 (81.6%) had B-cell precursor ALL, 10 (8.8%) T-cell ALL, 6 (5.3%) B-cell ALL, 4 (3.5%) null and 1 case had unclassified ALL. While the overall male-to-female ratio was 1.19, it was 2.33 in the T-ALL group and the average age was also higher in the T-ALL group (95 vs 59 months). The overall mean age of onset was 62.4 months (median = 51, range = 1 to 180). In the most common subtype B-cell precursor ALL ( $n=93$ ), the mean age was 57.8 months (median = 48, range = 1 to 180) and in the T-ALL subgroup ( $n=10$ ), the mean age was 95 months (median = 96, range = 46 to 156).

#### *Genotyping results*

Four of the SNPs (rs2797301, rs4985288, rs7768807, rs9503644) listed in Table 1 could not be genotyped because of TaqMan assay failure. As mentioned in the Materials section, one coding region SNP (rs9378788) listed in the NCBI ENTREZ SNP was not polymorphic and not genotyped. Thus, out of the 18 SNPs selected for genotyping, 13 of them were genotyped in cases and controls. Distributions of genotypes at all these SNPs were in Hardy-Weinberg equilibrium with the exception of rs9378805 that showed slight deviation in the whole of control group ( $P = 0.01$ ). In the analysis after stratification by sex, the deviation was exclusive to male newborns ( $P = 0.008$  in males and  $P = 0.18$  in females) suggesting possible involvement in prenatal selection rather than an explanation concerning whole of the control group. The genotype distribution of the lactase gene SNP rs1042712 genotyped as a marker for geographic differentiation in the UK did not violate HWE, and did not show any statistically significant association in case-control comparison, thus ruling out detectable population substructure.

#### *Linkage disequilibrium analysis and haplotype construction*

Using the default options of Haploview, the only strong ( $r^2 > 0.80$ ) LD was between the two 3' end SNPs rs1877175 and rs11242865 ( $r^2 > 0.91$ ;  $D' = 0.96$ ). Sex-specific LD patterns did not show any difference.

#### *Univariable data analysis*

In allele frequency comparisons, there was no statistically significant change between cases and controls in the overall group. The largest difference was noted for the intron 4 SNP rs12203592 (0.255 vs 0.197,  $P = 0.07$ ). The difference was larger in males-only analysis (0.277 vs 0.169,  $P = 0.01$ ) with no difference in female cases and controls (0.228 vs 0.222,  $P = 0.90$ ).

In Table 2, we present the results of additive gene-dosage model statistics for the overall group. Again, the largest difference was observed for intron 4 SNP with the homozygosity rates between cases and controls showing a significant difference (in reference to wildtype genotype,  $P = 0.009$ ). These findings prompted us to perform a detailed analysis of the association of rs12203592 (alternative name: NT\_034880.3:g.336321C>T) with childhood ALL. Sex-specific analysis confirmed that the association was exclusive to males. Homozygosity for the variant allele T conferred an increased risk for childhood ALL in males (in reference to the wildtype genotype, OR = 4.86, 95% CI = 1.64 to 14.4;  $P = 0.004$ ). In females, the same association did not reach statistical significance ( $P = 0.42$ ). As the data presented in Table 2 suggested, the recessive model was the best fit to the male-specific risk association of rs12203592 with childhood ALL. The homozygous genotype increased the risk in males very strongly (in reference to the other genotypes, OR = 4.36, 95% CI = 1.51 to 12.6,  $P = 0.007$ ) but not in females, although there was a non-significant increase in odds ratio (OR = 1.79, 95% CI = 0.60 to 5.28,  $P = 0.29$ ). All  $r^2$  values for LD between this SNPs and the other 12 IRF4 SNPs were  $<0.10$ . Thus, this association was not confounded by any of the IRF4 SNPs examined in the present study.

The male-specific risk marker was present in 3 of 7 male T-ALL cases and 5 of 48 male B-cell precursor ALL (42.9% vs 10.4%) but the numbers were too small for assessment of the statistical significance. Likewise the average age of onset in males with the risk genotype group was 77 months compared with 70 months in the remainder of male cases.

We also examined the CLL associated SNP rs872071 [14], NHL associated SNP rs12211228 [15] and the 3'UTR SNP rs9378805 [16] associated with height in subjects with T2D in detail. None of these SNPs showed a statistically significant association with childhood ALL in either sex or in the overall sample. There was no LD between any of these SNPs and the intron 4 SNP rs12203592.

#### *Multivariable analysis*

In the presence of other male-specific genetic associations with immunologic functions in the same case-control group [20, 22], we considered the independence of all these associations. In a multivariable model, the DRB4, HSPA1B rs1061581 and IRF4 rs12203592 homozygosity all retained their statistical significance for their associations in males. In this model, the adjusted OR for the IRF4 SNP was 4.26 (95% CI = 1.33 to 13.6,  $P = 0.01$ ) using the wildtype genotype as reference.

#### *Promoter reporter assays*

In order to study the functional role of IRF4 intron 4 SNP in the regulation of IRF4 expression level, two IRF4 promoter (IRF4P) fragments (2.4kb) were amplified from two different cell lines. The sequences of both homozygous (C)IRF4P and (T)IRF4P were fully sequenced in both directions to check for any promoter SNP that might be in LD with the studied SNP rs12203592 C>T in intron 4. There are seven known SNPs in this 2.4kb IRF4 promoter according to the published sequence in the Ensembl database; however, both of our IRF4 promoters possess all the wild type alleles for these seven SNPs. We did not detect any new polymorphism. Functionally, the luciferase activities from these two IRF4 promoters were

determined to be very similar in all three cell lines used in this study: Raji (Figure 1), HEK-293T, and NCI-H295R (not shown).

Raji is a suspension cell line and therefore was transfected using Cell Line Nucleofector™ Kit V from Amaxa. The transfection efficiency determined by FACS analysis and EGFP plasmid was around 70% (data not shown). An included positive control pGL4.13[luc2/SV40] in every experiment always yielded a very high luciferase reading (~ 600 units), whereas that for IRF4 promoter is only in the range of 5 to 9 units. The consistent data of the positive control suggested that the low IRF4 promoter's transcriptional activity in Raji cells was not due to low transfection efficiency. The same amount of plasmid DNA was used to transfect the other two adherent cell lines (HEK-293T and NCI-H295R); however, in these two cell lines the luciferase reading of IRF4 promoters were at least as high as that of the positive control pGL4.13[luc2/SV40]. For all three cell lines, data was collected at both 24 and 48 hours, the luciferase readings of both firefly and *Renilla* were higher in 48 hours; however the normalized relative units (firefly/*renilla*) were similar for both 24 and 48 data points. Therefore, 24 hour was chosen for analysis in all experiments.

Next, to determine whether the IRF4 intron 4 has any transcriptional regulatory function, firefly luciferase reporter gene vectors were generated with IRF4 promoter (2.4kb) in the presence of IRF4 intron 4. Recombinant plasmids with this IRF4 intron 4 fragment inserted in both orientations (+ and -), both upstream (C\_U(+)) and C\_U(-)) and downstream (C\_D(+)) and C\_D(-)) of the IRF4 promoter, as well as the parent plasmids with just IRF4 promoter (IRF4P), were independently nucleofected into a human Burkitt lymphoma B lymphocyte cell line called Raji and assayed for luciferase activity at 5 and 24 hours later. The data from both time points were similar in which this intronic 4 fragment repressed expression by the IRF4 promoter more than 2 folds (Figure 1). Intron 4 fragment downstream of IRF4 promoter had the ability to repress the transcriptional activity more than when it was placed

upstream of the promoter. The results demonstrated that the IRF4 intron 4 contains a repressor element that functions in an orientation- and position-independent fashion.

Our goal was to determine if a change from C to T at SNP rs12203592 in this IRF4 intron 4 will change the repressive effect of IRF4 intron 4 upon IRF4 promoter activity. In order to test this, we generated four additional plasmids with IRF4 promoter and IRF4 intron 4 with homozygous genotype for the variant allele T : T\_U(+), T\_U(-), T\_D(+) and T\_D(-).

Intron 4 with the variant allele T had less repressive effect than the wild type allele C especially when it was placed at the downstream position (Figure 1). To determine if this effect was due specifically from a change of C to T at this SNP in intron 4, the construct containing intron 4 with the wild type allele C downstream in sense direction (C\_D(+)) was mutated to T using PCR. The data showed that the mutated plasmid CD(+)\_CtoT also had less repressive effect than that of its partner CD(+). These results suggested that the repressive effect of intron 4 might result mostly due to a transcription factor binding to the wild type sequence with allele C.

We postulated that the level of IRF4 would be induced in most of the malignant cells. Using qPCR with TaqMan expression assay for IRF4 and 18sRNA, we confirmed the presence of IRF4 mRNA at a similar level of that in Raji cells in the two adherent cell lines available to us: the human embryonic kidney 293T (HEK-293T) transformed with adenovirus 5 DNA and the carcinoma human adrenocortical cell line (NCI-H295R) (data not shown). The transcriptional activity of the 2.4kb IRF4 promoter driving the luciferase reporter gene was very strong in both HEK-293 and NCI-H295R as compared to that of a positive luciferase vector driven by a constitutively expressed promoter SV40 (pGL4.13[luc2/SV40]). Since the IRF4 promoter functions strongly in these two cell lines, we were also interested to confirm the differential effect of IRF4 intron 4 SNP rs12203592.

In both HEK-293T and NCI-H295R cells, only four key plasmids were tested along with the negative and positive controls (pGL4.10[luc2] and pGL4.13[luc2/SV40]) (Figure 2).

The data consistently showed the same pattern as that seen in Raji cells. Intron 4 with the wild type allele C downstream of IRF4 promoter C\_D(+) repressed the transcriptional activity of IRF4 promoter about 2 fold and the effect is greatly alleviated with intron 4 having the variant T allele (T\_D(+)) or when the allele C was mutated to the variant allele T (C\_D(+))CtoT).

#### *Streptavidin-agarose pulldown assay*

Two bioinformatics web tools were used to predict transcription factor binding sites in IRF4 intron 4 sequences spanning the SNP rs12203592 C>T. FastSNP predicted that the SNP lies in a binding site for NF- $\kappa$ B or Tst-1 but allelic variation does not change binding affinity. The transcription element search system (TESS) predicted that transcription factor AP-2 $\alpha$  binds to the DNA sequence when the SNP has allele C but not to the variant allele containing sequence. Out of the three cell lines used for reporter gene assay in this study, HEK-293T cell line was used as a representative candidate for the examination of differential binding of IRF4 intron 4 SNP oligonucleotides to AP-2 $\alpha$ . Using western blotting, we were able to detect a high level of Tst-1 and AP-2 $\alpha$  in both the cytosol and nuclear extract of HEK-293T cells; however, both p65 and p50 subunits of NF- $\kappa$ B were detected only in the cytosol and not in the nuclear extract (data not shown). The original reporter gene assay was performed in an unstimulated cell system and since NF- $\kappa$ B binding is not predicted to vary by sequence variation, we did not repeat western blotting using stimulated cells.

To examine the prediction that allelic variation at the SNP site alters AP-2a binding affinity, we chose a streptavidin-agarose pulldown assay over the electrophoretic mobility shift assay (EMSA) as it does not require radiolabeled probes. This method involves incubation of nuclear extract protein with 5'-biotinylated double-stranded DNA probes and streptavidin-agarose resins [28]. The protein-DNA complex is pulled down by centrifugation, and proteins in the complex are dissociated and analyzed by Western blotting.



We used two 25 bp biotinylated-double-stranded DNA sequences spanning the IRF4 SNP rs12203592 with either the wild type allele C or the variant allele T as the probes to assess their relative binding affinities to transcription factor AP-2 $\alpha$  in the nuclear extract of HEK-293T by the streptavidin-agarose resins pulldown assay. Positive control oligonucleotides known to bind to AP-2 $\alpha$  were also included in every experiment. Results from four independent experiments consistently showed that oligonucleotides with the wild type allele C were able to bind to AP-2 $\alpha$  about 60% stronger than that with the variant allele (Figure 3). A negative control including everything except oligonucleotides did not pull down any AP-2 $\alpha$  protein (data not shown).

We then stripped the same blot to and reprobed with Tst-1(Oct-6) antibody since FastSNP predicted binding of Tst-1 to the same sequence without specificity to the allelic variation. We did not try the same with NF-kB because it was not detectable in the nuclear extract of the unstimulated cell line used. Both oligonucleotides spanning the SNP (with alleles C or T) were able to pull down similar amounts of Tst-1 protein. The positive control oligonucleotide for AP-2 $\alpha$  did not bind the Tst-1 protein (Figure 3). The similar amounts of Tst-1 protein detected in both lanes confirmed that the differences seen in AP-2 $\alpha$  bindings to C or T carrying oligonucleotides were due to differential binding affinities and not to technical reasons.

## Discussion

This comprehensive study of IRF4 polymorphisms identified a male-specific risk association between childhood ALL and an intronic SNP and demonstrated its effect on gene transcription in vitro. While rs12203592 was associated with childhood ALL, other IRF4 SNPs associated with CLL (rs872071) or NHL (rs12211228) did not show any association and there was no LD between the intron 4 SNP and the other SNPs associated with other hematopoietic cancers. Bioinformatics analysis on SNPSeek shows that the ALL- and CLL-associated SNPs are located in a regulatory-potential region from 7-species alignment and

the NHL-associated SNP is in a human-mouse-rat conserved region. Of these three SNPs, only rs12203592 has been examined for its functional effect experimentally (the present study).

The intronic SNP rs12203592 has an interesting feature. The examination of the HapMap data showed that the variant allele only occurs in the European population sample with no polymorphism detected in African and Asian samples. The HapMap European frequencies are 0.167 for the variant allele and 0.033 for variant allele homozygosity. These are similar to the frequencies in the control group of the present study (0.197 and 0.050, respectively). Even within Europe, the allele frequencies show regional variation between the Irish population and Northern/Eastern/Central European populations [18, 27]. This follows the finding that the variant allele of this SNP is enriched in people with black hair, fair skin and blue eyes with poor tanning ability. We did not have hair color data to exclude a confounding of the observed association by differential representation of individuals with different hair colors in cases and controls but this is an unlikely explanation for the observed association.

Besides rs12203592 being a geographic differentiation marker in Europe, another IRF4 SNP rs9378805 has been identified as a one of the 13 similar markers specifically for the UK population [19]. The 3'UTR SNP rs9378805 ranks fourth as an informative marker among the UK-specific markers [19]. To rule out a spurious association possibility due to cryptic population substructure in our case-control sample, we also examined the top ranking UK-specific marker (LCT rs1042712). Neither SNP showed a frequency difference between cases and controls or between sexes to suggest population stratification that may account for the observed association. Statistical adjustments for these two SNPs did not cause an appreciable change in the male-specific association of the intron 4 SNP either. Given the fact that the detailed principal component analysis of the UK population sample by the Wellcome Trust Case-Control Consortium concluded that within the UK, the population structure is very weak, we are confident that the association is not a spurious finding. The sex effect and

association with a homozygous genotype rather than variant allele positivity also make population stratification an unlikely alternative explanation.

Like any statistical association, the rs12203592 association required an independent replication or functional confirmation. We first examined bioinformatics resources for predicted functions. FastSNP/TFSEARCH analysis suggested an intronic enhancer function for rs12203592 due to its location in transcription factors NF- $\kappa$ B and Tst-I binding sites. However, the SNP does not change the binding characteristics of these transcription factors in silico. Having eliminated LD and identified the rs12203592 association as the primary association, we went ahead and examined the functional consequences of nucleotide substitution on gene transcription by in vitro assays. Using three cell lines and natural sequence variants as well as site-directed mutagenesis variants, we showed a repressive effect of intron 4 with the wildtype allele C on IRF4 promoter activity ( $P < 0.001$ ) but no repression by the variant allele T. Thus, homozygosity for the rs12203592 variant allele would result in increased IRF4 expression if IRF4 expression is induced.

The results from our in vitro experiments confirmed the predictions of in silico analyses and suggested a role of AP-2 $\alpha$  in regulating the transcriptional activity of IRF4 gene by differentially binding to IRF4 rs12203592 alleles. We used the streptavidin-agarose pulldown assay and determined that the wild type sequence of IRF4 intron 4 binds more than 2 folds stronger to the transcription factor AP-2 $\alpha$  than that with the variant allele T (Figure 3). The reporter gene assay showed that the wild type intron 4 has a repressive effect on the transcriptional activity of IRF4 promoter and the repressive effect was significantly alleviated when the nucleotide C at the rs12203592 position is replaced by the variant allele T. Based on the observations from this study, we propose that AP-2 $\alpha$  may be the mediator of the repressive effect on the IRF4 gene as it does with other genes in other cancers [30, 31] Although AP-2 family is generally considered to be an activator of transcription, recent studies

have also shown that AP-2 $\alpha$  may negatively regulate the transcription of certain genes in cancer cells [30, 31].

IRF4 is a distinct member of the IRF family of transcription factors. It differs from other IRFs in several ways including its repressive effect on IFN-induced gene expression [5, 32] and induction by proliferative stimuli [5, 33]. These and other unique effects [5, 10-12, 34] culminate into an oncogenic activity [13]. Most B-cell malignancies have increased IRF4 expression [7-9] presumably required to maintain their malignant phenotype. Our finding of a risk association with an intronic SNP that would facilitate increased IRF4 expression in childhood ALL, which is predominantly a B-cell origin malignancy, is therefore biologically plausible.

The sex effect we observed in the IRF4 association is not unexpected given the previous observations of sex-specific associations in childhood ALL [20-22, 29, 35-37]. The identification of the mechanism of the male-specificity of this risk association with an IRF4 polymorphism requires detailed studies but a working hypothesis can be put forward as follows: IRF4 overexpression in malignancies is due to hyperactivation of c-REL which is a common occurrence [10, 38] including hematopoietic cancers [39, 40], specifically in childhood ALL [41]. Experimentally, induction of IRF4 expression by v-Rel facilitates transformation of fibroblasts [34]. Overexpression of REL, which is accompanied by increased expression of IRF4, increases the oncogenic properties of the human B-cell lymphoma BJAB cell line [39]. The REL gene encodes c-Rel, a transcription factor that is a member of the Rel/NF- $\kappa$ B family. Since the effects of NF- $\kappa$ B on its downstream target genes are suppressed by estrogens [42], overexpression of IRF4 in cancer is expected to be stronger in males. This may be the basis of the male-specificity of the rs12203592 association.

There are only limited examples of mechanisms of sex effects in cancer and no functional mechanism is shown for the sex-specificity of any genetic association. Our data may hint the first such mechanism if followed up and confirmed by signal transduction studies

of NF- $\kappa$ B pathway and its interaction with IRF4. Two other known mechanisms of the sex effect in cancer predisposition in thyroid cancer [43] and liver cancer [44, 45] both involve sex hormones. Thyroid follicular cancer is one of those cancers that are more frequent in females. It has been suggested that males may be protected due to androgen receptor expression in thyroid follicular cells through which androgens reduce proliferation of follicular cells. Lacking this effect, females are at higher risk of developing thyroid follicular cancer. In the case of liver cancer, it has been shown that estrogens inhibit IL-6 secretion which is a key mediator of liver cancer development and females are protected owing to estrogen effect.

We recently reported another sex-specific association of MDM2 SNP (rs2279744, SNP309) with age-at-onset in the same case-control group [29]. The association with earlier age-at-onset was observed only in females, as has been noted also in adult cancers. This effect is attributed to the effect of estrogen [29]. The MDM2 association with age-at-onset and the findings of the present study implicate estrogen action in childhood ALL. However, there is only negligible estrogen activity in childhood, and raises the possibility of mediation of fetal programming of childhood ALL susceptibility by sex hormones. Fetal programming of susceptibility to adult diseases has been shown in several diseases [46] but it has not been considered for childhood leukemia.

The present study provided yet another example of a functional intronic polymorphism. It has already been recognized that intronic SNPs should not be dismissed as non-functional polymorphisms in genetic association studies [47-49]. Besides, the present study also exemplified the importance of taking sex into account in analysis. Unfortunately, most study results are reported with adjustment for sex but that analysis only adjusts for confounding. To uncover a sex effect or effect modification by sex, either stratified analysis or examination of interaction is required. Finally, we chose to examine the functionality of the associated SNP to validate the statistical association. This way, we avoided further statistical manipulations (such as the conservative Bonferroni correction) to increase the confidence

levels in our statistical association. No statistical manipulation can substitute experimental data for the functionality of a SNP that shows an association. Overall our results suggested a mechanism for the greater predilection of males to childhood ALL and perhaps to other cancers. The hypothesis that we generated in the present study which proposes that REL activation induces the oncogenic IRF4 expression more strongly in males and especially if they are homozygous for the intron 4 SNP variant allele is testable in any cancer. If confirmed, this mechanism can be targeted pharmacologically as already suggested [38].

### **Acknowledgments**

This work was funded intramurally by HUMIGEN LLC, The Institute for Genetic Immunology (Hamilton, New Jersey, USA). All authors are employees of HUMIGEN LLC.

### **URLs of Online Resources Used**

NCBI ENTREZ SNP: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp>

HapMap project: <http://www.hapmap.org>

NHGRI Catalog of Published Genome-Wide Association Studies:

<http://www.genome.gov/gwastudies>

HuGE Navigator: <http://www.hugenavigator.net>

SNP500Cancer: <http://snp500cancer.nci.nih.gov>

SNPs3D: <http://www.snps3d.org>

FastSNP: <http://fastsnp.ibms.sinica.edu.tw>

SNPSeek: <http://snp.wustl.edu/cgi-bin/SNPseek/index.cgi>

TFSEARCH: <http://www.cbrc.jp/research/db/TFSEARCH.html>

Database of Genomic Variants: <http://projects.tcag.ca/variation>

**References**

- [1] T. Tamura, H. Yanai, D. Savitsky and T. Taniguchi, The IRF family transcription factors in immunity and oncogenesis, *Annu Rev Immunol*, 26 (2008) 535-84.
- [2] R. Lu, Interferon regulatory factor 4 and 8 in B-cell development, *Trends Immunol*, 29 (2008) 487-92.
- [3] R. Lu, K.L. Medina, D.W. Lancki and H. Singh, IRF-4,8 orchestrate the pre-B-to-B transition in lymphocyte development, *Genes Dev*, 17 (2003) 1703-8.
- [4] H.W. Mittrucker, T. Matsuyama, A. Grossman, T.M. Kundig, J. Potter, A. Shahinian, A. Wakeham, B. Patterson, P.S. Ohashi and T.W. Mak, Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function, *Science*, 275 (1997) 540-3.
- [5] T. Yamagata, J. Nishida, S. Tanaka, R. Sakai, K. Mitani, M. Yoshida, T. Taniguchi, Y. Yazaki and H. Hirai, A novel interferon regulatory factor family transcription factor, ICSAT/Pip/LSIRF, that negatively regulates the activity of interferon-regulated genes, *Mol Cell Biol*, 16 (1996) 1283-94.
- [6] S. Iida, P.H. Rao, M. Butler, P. Corradini, M. Boccadoro, B. Klein, R.S. Chaganti and R. Dalla-Favera, Deregulation of MUM1/IRF4 by chromosomal translocation in multiple myeloma, *Nat Genet*, 17 (1997) 226-30.
- [7] A. Carbone, A. Gloghini, M.R. Cozzi, D. Capello, A. Steffan, P. Monini, L. De Marco and G. Gaidano, Expression of MUM1/IRF4 selectively clusters with primary effusion lymphoma among lymphomatous effusions: implications for disease histogenesis and pathogenesis, *Br J Haematol*, 111 (2000) 247-57.
- [8] A. Carbone, A. Gloghini, L.M. Larocca, D. Capello, F. Pierconti, V. Canzonieri, U. Tirelli, R. Dalla-Favera and G. Gaidano, Expression profile of MUM1/IRF4, BCL-6, and CD138/syndecan-1 defines novel histogenetic subsets of human immunodeficiency virus-related lymphomas, *Blood*, 97 (2001) 744-51.

- [9] A. Chadburn, E.M. Hyjek, W. Tam, Y. Liu, T. Rengifo, E. Cesarman and D.M. Knowles, Immunophenotypic analysis of the Kaposi sarcoma herpesvirus (KSHV; HHV-8)-infected B cells in HIV+ multicentric Castleman disease (MCD), *Histopathology*, 53 (2008) 513-24.
- [10] R.J. Grumont and S. Gerondakis, Rel induces interferon regulatory factor 4 (IRF-4) expression in lymphocytes: modulation of interferon-regulated gene expression by rel/nuclear factor kappaB, *J Exp Med*, 191 (2000) 1281-92.
- [11] H. Negishi, Y. Ohba, H. Yanai, A. Takaoka, K. Honma, K. Yui, T. Matsuyama, T. Taniguchi and K. Honda, Negative regulation of Toll-like-receptor signaling by IRF-4, *Proc Natl Acad Sci USA*, 102 (2005) 15989-94.
- [12] R. Hrdlickova, J. Nehyba and H.R. Bose, Jr., Regulation of telomerase activity by interferon regulatory factors 4 and 8 in immune cells, *Mol Cell Biol*, 29 (2009) 929-41.
- [13] N. Tanaka and T. Taniguchi, The interferon regulatory factors and oncogenesis, *Semin Cancer Biol*, 10 (2000) 73-81.
- [14] M.C. Di Bernardo, D. Crowther-Swanepoel, P. Broderick, E. Webb, G. Sellick, R. Wild, K. Sullivan, J. Vijayakrishnan, Y. Wang, A.M. Pittman, N.J. Sunter, A.G. Hall, M.J. Dyer, E. Matutes, C. Dearden, T. Mainou-Fowler, G.H. Jackson, G. Summerfield, R.J. Harris, A.R. Pettitt, P. Hillmen, D.J. Allsup, J.R. Bailey, G. Pratt, C. Pepper, C. Fegan, J.M. Allan, D. Catovsky and R.S. Houlston, A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia, *Nat Genet*, 40 (2008) 1204-10.
- [15] S.S. Wang, M.P. Purdue, J.R. Cerhan, T. Zheng, I. Menashe, B.K. Armstrong, Q. Lan, P. Hartge, A. Krickler, Y. Zhang, L.M. Morton, C.M. Vajdic, T.R. Holford, R.K. Severson, A. Grulich, B.P. Leaderer, S. Davis, W. Cozen, M. Yeager, S.J. Chanock, N. Chatterjee and N. Rothman, Common gene variants in the tumor necrosis factor



- (TNF) and TNF receptor superfamilies and NF- $\kappa$ B transcription factors and non-Hodgkin lymphoma risk, *PLoS One*, 4 (2009) e5360.
- [16] R. Saxena, B.F. Voight, V. Lyssenko, N.P. Burtt, P.I. de Bakker, H. Chen, J.J. Roix, S. Kathiresan, J.N. Hirschhorn, M.J. Daly, T.E. Hughes, L. Groop, D. Altshuler, P. Almgren, J.C. Florez, J. Meyer, K. Ardlie, K. Bengtsson Bostrom, B. Isomaa, G. Lettre, U. Lindblad, H.N. Lyon, O. Melander, C. Newton-Cheh, P. Nilsson, M. Orho-Melander, L. Rastam, E.K. Speliotes, M.R. Taskinen, T. Tuomi, C. Guiducci, A. Berglund, J. Carlson, L. Gianniny, R. Hackett, L. Hall, J. Holmkvist, E. Laurila, M. Sjogren, M. Sterner, A. Surti, M. Svensson, M. Svensson, R. Tewhey, B. Blumenstiel, M. Parkin, M. Defelice, R. Barry, W. Brodeur, J. Camarata, N. Chia, M. Fava, J. Gibbons, B. Handsaker, C. Healy, K. Nguyen, C. Gates, C. Sougnez, D. Gage, M. Nizzari, S.B. Gabriel, G.W. Chirn, Q. Ma, H. Parikh, D. Richardson, D. Ricke and S. Purcell, Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels, *Science*, 316 (2007) 1331-6.
- [17] A.H. Gathany, P. Hartge, S. Davis, J.R. Cerhan, R.K. Severson, W. Cozen, N. Rothman, S.J. Chanock and S.S. Wang, Relationship between interferon regulatory factor 4 genetic polymorphisms, measures of sun sensitivity and risk for non-Hodgkin lymphoma, *Cancer Causes Control* (2009), in press.
- [18] J. Han, P. Kraft, H. Nan, Q. Guo, C. Chen, A. Qureshi, S.E. Hankinson, F.B. Hu, D.L. Duffy, Z.Z. Zhao, N.G. Martin, G.W. Montgomery, N.K. Hayward, G. Thomas, R.N. Hoover, S. Chanock and D.J. Hunter, A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation, *PLoS Genet*, 4 (2008) e1000074.
- [19] Wellcome Trust Case Control Consortium, Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls, *Nature*, 447 (2007) 661-78.

- [20] M.T. Dorak, T. Lawson, H.K. Machulla, C. Darke, K.I. Mills and A.K. Burnett, Unravelling an HLA-DR association in childhood acute lymphoblastic leukemia, *Blood*, 94 (1999) 694-700.
- [21] M.T. Dorak, A.K. Burnett, M. Worwood, A.M. Sproul and B.E. Gibson, The C282Y mutation of HFE is another male-specific risk factor for childhood acute lymphoblastic leukemia, *Blood*, 94 (1999) 3957.
- [22] E. Ucisik-Akkaya, C. Davis, C. Gorodezky, C. Alaez and M.T. Dorak, HLA complex-linked heat shock protein genes and childhood acute lymphoblastic leukemia susceptibility., *Cell Stress Chaperones* (2009), in press.
- [23] A.D. Johnson and C.J. O'Donnell, An open access database of genome-wide association results, *BMC Med Genet*, 10 (2009) 6.
- [24] L.A. Hindorff, P. Sethupathy, H.A. Junkins, E.M. Ramos, J.P. Mehta, F.S. Collins and T.A. Manolio, Potential etiologic and functional implications of genome-wide association loci for human diseases and traits, *Proc Natl Acad Sci USA*, 106 (2009) 9362-7.
- [25] W. Yu, M. Gwinn, M. Clyne, A. Yesupriya and M.J. Khoury, A navigator for human genome epidemiology, *Nat Genet*, 40 (2008) 124-5.
- [26] A.L. Dixon, L. Liang, M.F. Moffatt, W. Chen, S. Heath, K.C. Wong, J. Taylor, E. Burnett, I. Gut, M. Farrall, G.M. Lathrop, G.R. Abecasis and W.O. Cookson, A genome-wide association study of global gene expression, *Nat Genet*, 39 (2007) 1202-7.
- [27] C. Tian, R.M. Plenge, M. Ransom, A. Lee, P. Villoslada, C. Selmi, L. Klareskog, A.E. Pulver, L. Qi, P.K. Gregersen and M.F. Seldin, Analysis and application of European genetic substructure using 300 K SNP information, *PLoS Genet*, 4 (2008) e4.

- [28] W.G. Deng, Y. Zhu, A. Montero and K.K. Wu, Quantitative analysis of binding of transcription factor complex to biotinylated DNA probe by a streptavidin-agarose pulldown assay, *Anal Biochem*, 323 (2003) 12-8.
- [29] T.N. Do, E. Ucisik-Akkaya, C.F. Davis, B.A. Morrison and M.T. Dorak, TP53 R72P and MDM2 SNP309 polymorphisms in modification of childhood acute lymphoblastic leukemia susceptibility, *Cancer Genet Cytogenet*, 195 (2009) 31-6.
- [30] M. Ruiz, C. Pettaway, R. Song, O. Stoeltzing, L. Ellis and M. Bar-Eli, Activator protein 2alpha inhibits tumorigenicity and represses vascular endothelial growth factor transcription in prostate cancer cells, *Cancer Res*, 64 (2004) 631-8.
- [31] C. Tellez, M. McCarty, M. Ruiz and M. Bar-Eli, Loss of activator protein-2alpha results in overexpression of protease-activated receptor-1 and correlates with the malignant phenotype of human melanoma, *J Biol Chem*, 278 (2003) 46632-42.
- [32] S.J. Gao, C. Boshoff, S. Jayachandra, R.A. Weiss, Y. Chang and P.S. Moore, KSHV ORF K9 (vIRF) is an oncogene which inhibits the interferon signaling pathway, *Oncogene*, 15 (1997) 1979-85.
- [33] T. Matsuyama, A. Grossman, H.W. Mittrucker, D.P. Siderovski, F. Kiefer, T. Kawakami, C.D. Richardson, T. Taniguchi, S.K. Yoshinaga and T.W. Mak, Molecular cloning of LSIRF, a lymphoid-specific member of the interferon regulatory factor family that binds the interferon-stimulated response element (ISRE), *Nucleic Acids Res*, 23 (1995) 2127-36.
- [34] R. Hrdlickova, J. Nehyba and H.R. Bose, Jr., Interferon regulatory factor 4 contributes to transformation of v-Rel-expressing fibroblasts, *Mol Cell Biol*, 21 (2001) 6369-86.
- [35] M. Krajcinovic, D. Labuda, C. Richer, S. Karimi and D. Sinnett, Susceptibility to childhood acute lymphoblastic leukemia: influence of CYP1A1, CYP2D6, GSTM1, and GSTT1 genetic polymorphisms, *Blood*, 93 (1999) 1496-501.

- [36] M.H. Ng, K.M. Lau, B.R. Hawkins, K.W. Chik, N.P. Chan, W.S. Wong, K.S. Tsang, M.M. Shing and C.K. Li, HLA-B67 may be a male-specific HLA marker of susceptibility to relapsed childhood ALL in Hong Kong Chinese and HLA-A33 or HLA-B17 signifies a higher presentation leukocytosis: A retrospective analysis on 53 transplant candidates (1989-2003), *Ann Hematol*, 85 (2006) 535-41.
- [37] H. Reddy and K. Jamil, Polymorphisms in the MTHFR gene and their possible association with susceptibility to childhood acute lymphocytic leukemia in an Indian population, *Leuk Lymphoma*, 47 (2006) 1333-9.
- [38] T.D. Gilmore, D. Kalaitzidis, M.C. Liang and D.T. Starczynowski, The c-Rel transcription factor and B-cell proliferation: a deal with the devil, *Oncogene*, 23 (2004) 2275-86.
- [39] M. Chin, M. Herscovitch, N. Zhang, D.J. Waxman and T.D. Gilmore, Overexpression of an activated REL mutant enhances the transformed state of the human B-lymphoma BJAB cell line and alters its gene expression profile, *Oncogene*, 28 (2009) 2100-11.
- [40] B. Rayet and C. Gelinas, Aberrant rel/nfkb genes and activity in human cancer, *Oncogene*, 18 (1999) 6938-47.
- [41] U. Kordes, D. Krappmann, V. Heissmeyer, W.D. Ludwig and C. Scheidereit, Transcription factor NF-kappaB is constitutively activated in acute lymphoblastic leukemia cells, *Leukemia*, 14 (2000) 399-402.
- [42] D. Kalaitzidis and T.D. Gilmore, Transcription factor cross-talk: the estrogen receptor and NF-kappaB, *Trends Endocrinol Metab*, 16 (2005) 46-52.
- [43] R. Rossi, M.C. Zatelli, P. Franceschetti, I. Maestri, E. Magri, G. Aguiari, P. Cavazzini, E.C. degli Uberti and L. del Senno, Inhibitory effect of dihydrotestosterone on human thyroid cell growth, *J Endocrinol*, 151 (1996) 185-94.

- [44] W.E. Naugler, T. Sakurai, S. Kim, S. Maeda, K. Kim, A.M. Elsharkawy and M. Karin, Gender disparity in liver cancer due to sex differences in MyD88-dependent IL-6 production, *Science*, 317 (2007) 121-4.
- [45] T. Lawrence, T. Hageman and F. Balkwill, Cancer. Sex, cytokines, and cancer, *Science*, 317 (2007) 51-2.
- [46] J.G. Hall, The importance of the fetal origins of adult disease for geneticists, *Clin Genet*, 72 (2007) 67-73.
- [47] K. Ozaki, Y. Ohnishi, A. Iida, A. Sekine, R. Yamada, T. Tsunoda, H. Sato, H. Sato, M. Hori, Y. Nakamura and T. Tanaka, Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction, *Nat Genet*, 32 (2002) 650-4.
- [48] S. Tokuhiro, R. Yamada, X. Chang, A. Suzuki, Y. Kochi, T. Sawada, M. Suzuki, M. Nagasaki, M. Ohtsuki, M. Ono, H. Furukawa, M. Nagashima, S. Yoshino, A. Mabuchi, A. Sekine, S. Saito, A. Takahashi, T. Tsunoda, Y. Nakamura and K. Yamamoto, An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis, *Nat Genet*, 35 (2003) 341-8.
- [49] D. Bellizzi, G. Rose, P. Cavalcante, G. Covello, S. Dato, F. De Rango, V. Greco, M. Maggolini, E. Feraco, V. Mari, C. Franceschi, G. Passarino and G. De Benedictis, A novel VNTR enhancer within the SIRT3 gene, a human homologue of SIR2, is associated with survival at oldest ages, *Genomics*, 85 (2005) 258-63.

## Figure Legends

**Figure 1.** IRF4 intron 4 with wild type allele C at SNP rs12203592 represses IRF4 promoter activity while IRF4 intron 4 with variant allele T significantly alleviates this repressive effect. Both work in an orientation- and position-independent manner. The full 1.2kb fragment of intron 4 of the human IRF4 gene (contains either a wild type C or variant allele T at SNP rs12203592) was subcloned into the luciferase-reporter plasmid driven by a 2.4kb IRF4 promoter (the big black arrow right before luciferase gene (LUC)). In all of the constructs, the LUC is used as a reporter gene whose mRNA is stabilized by a polyadenylation/splice signal from the simian virus 40 (Poly A). Raji cells were co-nucleofected with these constructs and with the internal control plasmid pGL4.13[hRenilla/SV40] and then assayed for both firefly and *Renilla* luciferases after 24h. To adjust for differences in transfection efficiencies, firefly luciferase values were standardized to *Renilla* luciferase values. The results are from three independent experiments. The error bars represent standard errors. \* Comparison between intron 4 with the variant allele T and intron 4 with the wild type allele C; # comparison between CD(+)CtoT with TD(+).

**Figure 2.** IRF4 wild type intron 4 represses IRF4 promoter activity while intron 4 with the variant allele T alleviates the repressive effect also in two more cell lines: **(A)** HEK-293T and **(B)** NCI-H295R. Cells were co-transfected with the constructs described in Figure 1 and the internal control plasmid pGL413[hRenilla/SV40]; then assayed for both firefly and *Renilla* luciferases after 24h. To adjust for differences in transfection efficiencies, firefly luciferase values were standardized to *Renilla* luciferase values to give the relative light unit (RLU). The results are from three to four independent experiments. The error bars represent standard errors. \*  $P < 0.0001$ .

**Figure 3:** Representative immunoblots showing AP-2 $\alpha$  and Tst-1 detection in (column 1) HEK-293T nuclear extract, (column 2) nuclear extract incubated with 5 pmol of biotin-labeled DNA oligonucleotides known to bind to AP-2 $\alpha$ , (column 3 and 4) 50pmol of the biotin-labeled DNA oligonucleotides spanning the IRF4 intron 4 SNP alleles C and T. Nuclear extract prepared from HEK-293T cells were incubated with the biotin-labeled DNA oligonucleotides and streptavidin-agarose resins for 1 h and proteins in the complex analyzed by Western blots. The western blot was first probed with anti-AP-2 $\alpha$ , then stripped and reprobed with anti-Tst-1 (Oct-6) antibody. The bar graph shows mean  $\pm$  SE of four experiments.

Table 1. IRF4 SNPs included in the study

SNP ID	Chromosome 6 position <sup>a</sup>	Location	Nucleotide change <sup>b</sup>	ABI TaqMan kit no
rs2797301 <sup>c</sup>	327 111	5' flanking	C>G	C__27071526_10
rs4985288 <sup>c</sup>	327 246	5' flanking	G>T	C__27957646_10
rs9405192	327 537	5' flanking	G>A	C__27071545_10
rs1033180	328 546	5' flanking	C>T	C__1576195_10
rs9378788 <sup>c</sup>	338 362	exon 2	not polymorphic	(PCR-RFLP)
rs12203592	341 321	intron 4	C>T	C__31918199_10
rs3778607	348 799	intron 7	A>G	C__27499861_10
rs2001508	349 632	intron 7	A>C	C__1576163_20
rs1131442	352 656	exon 9-3'UTR	G>A	C__1576159_20
rs7768807 <sup>c</sup>	353 246	exon 9-3'UTR	T>C	C__1576158_10
rs12211228	353 833	exon 9-3'UTR	G>C	C__31918237_10
rs1877175	355 493	exon 9-3'UTR	G>A	C__1576157_10
rs9392502	355 608	exon 9-3'UTR	C>G	C__1576155_10
rs872071	356 064	exon 9-3'UTR	A>G	C__8770093_10
rs11242865	356 954	3' flanking	C>T	C__1576153_10
rs7757906	357 741	3' flanking	G>A	C__29334722_20
rs9503644 <sup>c,d</sup>	360 406	3' flanking	G>A	C__30106657_20
rs9378805	362 727	3' flanking	A>C	C__30340659_10

<sup>a</sup> NCBI Reference assembly (build 129) nucleotide positions (from telomere to centromere).

The IRF4 gene lies between 336 760 and 356 193. <sup>b</sup> The wildtype and variant alleles in the



HapMap European population as well as in the current study control group. <sup>c</sup> These SNPs could not be typed (see Results). <sup>d</sup> Formerly rs11242867.

ACCEPTED MANUSCRIPT

**Table 2. Univariable analysis of associations between 13 IRF4 SNPs  
and childhood ALL<sup>a,b</sup>**

SNP	Wildtype	Heterozygote	Variant homozygote	<i>P</i> (additive model) <sup>c</sup>
<b>rs9405192</b>	1.0 (referent) (0.531 vs 0.521) <sup>b</sup>	1.04 (0.66 to 1.66) <sup>a</sup> (0.418 vs 0.393) <sup>b</sup>	0.58 (0.22 to 1.57) <sup>a</sup> (0.051 vs 0.086) <sup>b</sup>	0.89 (0.63 to 1.28) <i>P</i> = 0.54
<b>rs1033180</b>	1.0 (referent) (0.906 vs 0.875)	0.69 (0.32 to 1.45) (0.085 vs 0.012)	1.83 (0.16 to 20.4) (0.005 vs 0.009)	0.80 (0.41 to 1.54) <i>P</i> = 0.50
<b>rs12203592</b>	1.0 (referent) (0.606 vs 0.654)	0.98 (0.60 to 1.61) (0.269 vs 0.296)	<b>2.71 (1.28 to 5.75)</b> <b>(0.125 vs 0.050)</b>	1.37 (0.98 to 1.91) <i>P</i> = 0.07
<b>rs3778607</b>	1.0 (referent) (0.353 vs 0.326)	0.81 (0.49 to 1.34) (0.422 vs 0.480)	1.07 (0.59 to 1.96) (0.225 vs 0.194)	1.01 (0.74 to 1.37) <i>P</i> = 0.96

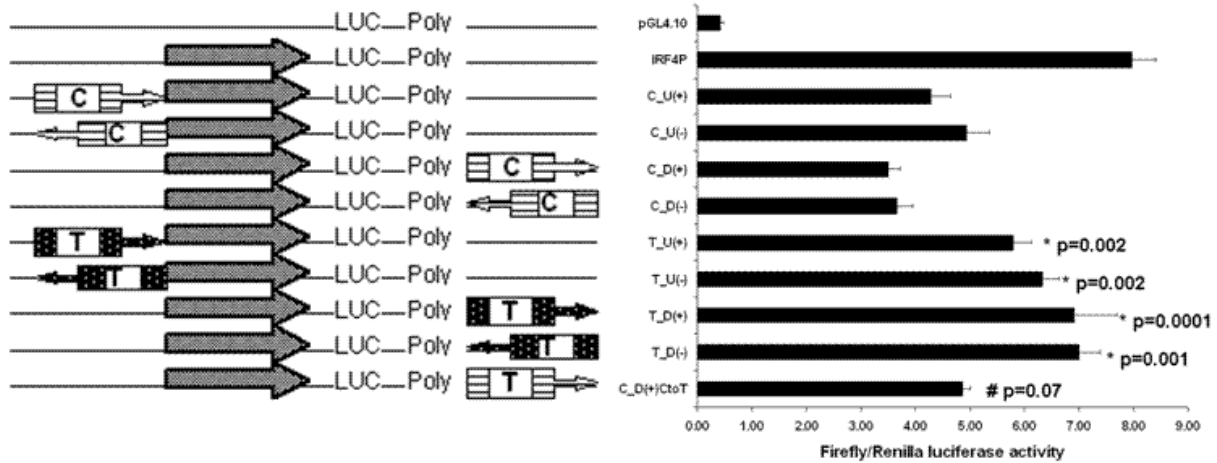
<b>rs2001508</b>	1.0 (referent) (0.472 vs 0.503)	1.12 (0.71 to 1.77) (0.426 vs 0.402)	1.12 (0.53 to 2.36) (0.102 vs 0.096)	1.08 (0.78 to 1.49) $P = 0.63$
<b>rs1131442</b>	1.0 (referent) (0.480 vs 0.513)	1.11 (0.70 to 1.75) (0.430 vs 0.415)	1.32 (0.59 to 2.97) (0.090 vs 0.073)	1.13 (0.80 to 1.59) $P = 0.48$
<b>rs12211228</b>	1.0 (referent) (0.750 vs 0.742)	0.93 (0.55 to 1.74) (0.220 vs 0.235)	1.31 (0.35 to 4.95) (0.030 vs 0.023)	1.00 (0.64 to 1.55) $P = 1.00$
<b>rs1877175</b>	1.0 (referent) (0.686 vs 0.645)	0.88 (0.54 to 1.42) (0.286 vs 0.306)	0.55 (0.16 to 1.92) (0.029 vs 0.049)	0.82 (0.55 to 1.22) $P = 0.33$
<b>rs9392502</b>	1.0 (referent) (0.694 vs 0.707)	0.99 (0.60 to 1.61) (0.259 vs 0.268)	1.93 (0.63 to 5.94) (0.046 vs 0.024)	1.13 (0.76 to 1.68) $P = 0.55$
<b>rs872071</b>	1.0 (referent) (0.340 vs 0.293)	0.79 (0.48 to 1.30) (0.453 vs 0.493)	0.84 (0.46 to 1.53) (0.208 vs 0.214)	0.90 (0.66 to 1.22) $P = 0.50$
<b>rs11242865</b>	1.0 (referent) (0.705 vs 0.666)	0.95 (0.59 to 1.54) (0.295 vs 0.293)	0.11 (0.01 to 1.79) (0.000 vs 0.041)	0.75 (0.49 to 1.15) $P = 0.19$
<b>rs7757906</b>	1.0 (referent) (0.621 vs 0.559)	0.77 (0.49 to 1.24) (0.330 vs 0.384)	0.77 (0.28 to 2.12) (0.049 vs 0.057)	0.82 (0.56 to 1.19) $P = 0.30$

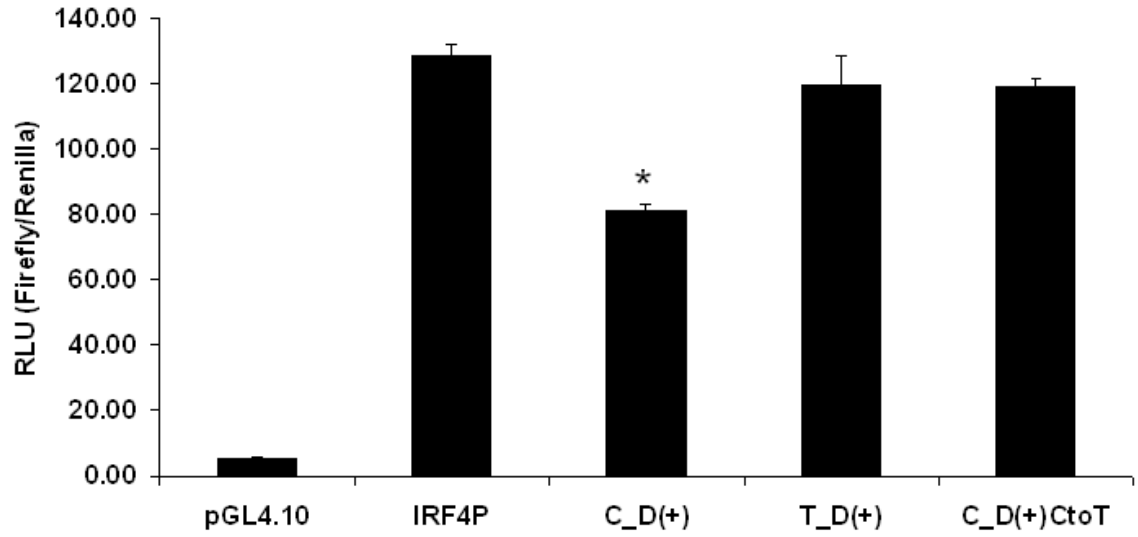
---

<b>rs9378805</b>	1.0 (referent)	1.16 (0.69 to 1.94)	1.29 (0.73 to 2.28)	1.14 (0.85 to 1.51)
	(0.273 vs 0.312)	(0.431 vs 0.431)	(0.291 vs 0.258)	<i>P</i> = 0.38

---

<sup>a</sup> Odds ratios and 95% confidence intervals using the wildtype homozygous genotype as reference and <sup>b</sup> frequencies in cases (n=114) and controls (n=388); <sup>c</sup> Odds ratio (additive model) given is per allele.



**A.****B.**