eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Facultad de Medicina y Odontología

Departamento de Genética, Antropología Física y Fisiología Animal

Tesis doctoral

Genetic variants involved in Childhood Acute Lymphoblastic Leukemia Susceptibility

Ángela Gutiérrez Camino

Leioa, 2016

Este trabajo ha sido posible gracias a la concesión de una beca predoctoral para Personal Investigador en Formación otorgada por el Departamento de educación, universidades e investigación del Gobierno Vasco y una beca de Investigación de la Fundación Jesús de Gangoiti Barrera.

A mis padres

A Daniel

Agradecimientos

Me gustaría aprovechar estas líneas para dar las gracias a todas las personas que han estado a mi lado y han hecho posible este proyecto.

En primer lugar, a África, porque hemos conseguido llegar enteras a la meta, aunque no ha sido un camino fácil. Gracias por haberme abierto el camino de la investigación y haber confiado en mí para este proyecto. Idoia, gracias por todo lo que me has enseñado, pero sobre todo por transmitirme tu pasión por este trabajo y por hacerme confiar más en mi misma.

A los doctores y doctoras Aurora Navajas, Aizpea Echebarria, Itziar Astigarraga, Mª Angeles Piñán, Ana Carboné, Purificación García-Miguel, Ana Sastre, Nagore Garcia de Andoin y Javier Uriz. A las enfermeras Silvia, Tere, Marta, Mayte y Begoña. Sin vuestra ayuda no habría sido posible este trabajo.

A Javier Ballesteros y a Borja por vuestra ayuda con la estadística, porque me habéis enseñado a ver su lado bueno.

Thanks to Dr. Marry van den Heuvel-Eibrink and Dr. Robert de Jonge, for opening for me the door to Rotterdam and to all the people in the group for turning my 4 months there in an unforgettable experience, specially Cherina, Pooja, Maja, Pamela, Samira, Gigi, Marianne and all the people in the AKC Department. Also thanks to Marissa for her help and support.

A mis compañeras y amigas. Nerea, mil gracias por haber decidido empezar este camino a mi lado, no sabes lo que significa para mí que me hayas apoyado siempre, porque compartir las cargas las hace menos pesadas. Idoia, gracias por ayudarme a quitarle importancia a las cosas cuando no veía la forma de seguir y por ser un apoyo en los momentos más difíciles. No se cómo hemos conseguido llegar al punto de reírnos cuando las cosas van mal, pero funciona. Eli, muchísimas gracias por tu infinita paciencia, por tu ayuda en los comienzos, y por esa gran boda que pasará a la historia por lo mal que nos lo pasamos ☺. Maria y Leire, aunque ha sido relativamente poco el tiempo que hemos trabajado juntas, muchas gracias por vuestra ayuda incondicional, por vuestros ánimos y por haber estado conmigo en algún momento de este camino. A todos los demás compañeros de laboratorio, Ander, Sarai, Julen, Ainara, Nora, Cris, Nerea, Alba, Bea, Ziortza, Jaione y Maitane.

Gracias a nuestro grupo vecino, Buli, Izortze, Amaia, Irati, Leti, Nora, Ainara, Koldo y Teresa, por animar el laboratorio, por los cafés, por las comidas… Gracias porque habéis hecho que me sienta menos sola.

Me gustaría agradecer de manera especial a mis amigas porque son el espejo en el que me miro para saber que vamos por el buen camino (o por uno muy malo, depende de cómo se mire ☺). Marta, Nerea, Elena, Vicky y Silvia, gracias a vosotras todo merece la pena, gracias por ser mi vía de escape y ser el lugar en el que me curo cuando estoy herida.

A César, que llegaste a mi vida para hacerla más bonita en un momento en el que todo era negro. Gracias por entenderme y ayudarme a superar los malos momentos, por enseñarme que para tener días buenos, también hay que tenerlos malos.

A mi familia, que son el pilar fundamental de mi vida. Escribiría páginas y páginas agradeciendo todo el apoyo que recibo, todos los ánimos y todo el cariño. Nada sería posible sin vosotros. Sobre todo gracias a mi abuela, que me cuida tanto, a mis tías, que creen más en mí que yo misma y especialmente a mis padres, a quienes les debo todo. Gracias por enseñarme a luchar y a levantarme cuando me he caído. Nos ha costado, pero lo hemos conseguido.

A Daniel, a quien jamás recuerdo porque nunca olvido.

Y a todos los que me han ayudado de alguna manera pero que, por culpa de mi mala memoria, no están aquí mencionados.

Publications

The work of this thesis is reflected in the following publications:

Gutierrez-Camino A, Martín-Guerrero I, Santos B, García de Andoin N, Sastre A, Carboné Bañeres A, Astigarraga I, Navajas A García-Orad A. Involvement of polymorphisms in *IKZF1* and *CEBPE* in childhood Acute Lymphoblastic Leukemia susceptibility (*Under preparation*)

Gutierrez-Camino A, Martín-Guerrero I,García de Andoin N, Sastre A Carboné Bañeres A Astigarraga I, Navajas A, García-Orad A. Involvement of SNPs in CDKN2A/B locus in Acute Lymphoblastic Leukemia susceptibility (*Under review in Leukemia and Lymphoma*)

Gutierrez-Camino A, Martín-Guerrero I, Dolzan V, Jazbec J, Carbone Bañeres A, Garcia de Andoin N, Sastre A, Astigarraga I, Navajas A, Garcia-Orad A. MicroRNAs SNPS involved in acute lymphoblastic leukemia susceptibility (*Under review in Haematologica*)

Gutiérrez-Camino A, López-López E, Martin-Guerrero I, Piñan MA, Garcia-Miguel P, Sánchez-Toledo J, Carboné Bañeres, A. Uriz J, Navajas A, García-Orad, A. Non-coding RNAs-related polymorphisms in pediatric acute lymphoblastic leukemia susceptibility. *Pediatr Res. 2014 Jun;75(6):767-73.* (Pediatrics Q1 19/122 IF: 2.67)

Gutiérrez-Camino A, López-López E, Uriz J, Sánchez-Toledo J, Carboné Bañeres A, García-Miguel P, Navajas A, García-Orad, A. Intron 3 of ARID5B gene: a hotspot for Acute Lymphoblastic leukemia susceptibility. *Journal of Cancer Research and Clinical Oncology. 2013. Nov;139(11):1879-86.* (Oncology Q2 84/196 IF: 2.914).

Lopez-Lopez E, **Gutierrez-Camino A**, Martin-Guerrero I, Garcia-Orad A. Re: Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations. *JNCI-Journal of the National Cancer Institute 2013. Oct 2;105(19):1512.* (Oncology Q1 6/196 IF: 14.33)

Furthermore, during the development of this thesis, I have collaborated in the following projects:

Iparraguirre L, **Gutierrez-Camino A**, Umerez M, Martin-Guerrero I, Astigarraga I, Navajas A, Sastre A, Garcia de Andoin N, Garcia-Ora A. MiR-pharmacogenetics of methotrexate in childhood Acute Lymphoblastic Leukemia *Pharmacogenet Genomics. 2016 Nov;26(11):517-525.* (Pharmacology and Pharmacy Q1 61/255 IF: 3.4)

Lopez-Lopez E, **Gutierrez-Camino A,** Astigarraga I, Navajas A, Echebarria-Barona A, Garcia-Miguel P, Garcia de Andoin N, Lobo C, Guerra-Merino I, Martin-Guerrero I, Garcia-Orad A. Vincristine pharmacokinetics pathway and neurotoxicity during early phases of treatment in pediatric acute lymphoblastic leukemia *Pharmacogenomics. 2016 May;17(7):731-41* (Pharmacology & Pharmacy Q1 47/261 IF: 3.857)

Gutierrez-Camino A, Martin-Guerrero I, Lopez-Lopez E, Echebarria-Barona A, Zabalza I, Ruiz I, Guerra-Merino I, Garcia-Orad A. Lack of association of rs924607 TT genotype in CEP72 with vincristine peripheral neuropathy in early phase of pediatric ALL treatment in Spanish

population. *Pharmacogenet Genomics. 2016 Feb;26(2):100-2.* (Pharmacology and Pharmacy Q1 61/255 IF: 3.4)

Martin-Guerrero I, **Gutiérrez-Camino A**, Lopez-Lopez E, Bilbao-Aldaiturriaga N, Pombar-Gómez M, Ardanaz M, Garcia-Orad A. Genetic variants in microRNA processing genes and pre-miRNAs are associated with the risk of Chronic Lymphocytic Leukemia *PLoS One. 2015 Mar 20;10(3):e0118905.* (Multidisciplinary Sciences Q1 8/55 IF: 3.534)

Bilbao-Aldaiturriaga N, **Gutiérrez-Camino A**, Martin-Guerrero I, Pombar-Gómez M, Zalacain-Díez M, Patiño-García A, López-López E, Garcia-Orad A. Polymorphisms in miRNAprocessing genes and their role in osteosarcoma risk. *Pediatr Blood Cancer. 2015 Feb 7.* (Pediatrics Q1 27/122 IF: 2.353)

M.A.H. den Hoed, Lopez-Lopez E, te Winkel M.L., Tissing W, de Rooij J.D,E, **Gutierrez-Camino A**, Garcia-Orad A, den Boer E, Pieters R, Pluijm S.M.F, de Jonge R, and van den Heuvel-Eibrink M.M. Genetic and metabolic determinants of methotrexate induced mucositis in pediatric acute lymphoblastic leukemia. *Pharmacogenomics J. 2014 Nov 4* (Pharmacology & Pharmacy Q1 14/254 IF: 5.513)

Gutierrez-Camino A, Lopez-Lopez E, Garcia-Orad A. SLC19A1 hot spot for MTX plasma concentration. *Med Oncol. 2014 Oct;31(10):204* (Oncology Q3 136/202 IF: 2.058)

Lopez-Lopez E, **Gutierrez-Camino A**, Bilbao-Aldaiturriaga N, Pombar-Gómez M, Martin-Guerrero I, Garcia-Orad A. Pharmacogenetics of childhood acute lymphoblastic leukemia. *Pharmacogenomics. 2014 Jul;15(10):1383-98* (Pharmacology & Pharmacy Q1 47/261 IF: 3.857)

Lopez-Lopez E, **Gutiérrez-Camino A**, Piñan MA, Sanchez-Toledo J, Uriz JJ, Ballesteros J, García-Miguel P, Navajas A, Garcia-Orad A Pharmacogenetics of microRNAs and microRNAs biogenesis machinery in pediatric Acute Lymphoblastic Leukemia. *PLoS One. 2014 Mar 10;9(3):e91261.*(Multidisciplinary Sciences Q1 7/56 IF: 3.7)

Abbreviations

3UTR	3'UTR regulation
5UTR	5'UTR regulation
ABL1	Abelson murine leukemia viral oncogene homolog 1
AF4	AF4/FMR2 family, member 1
AF9	Myeloid/lymphoid or mixed-lineage leukemia; translocated to, 3
AGO	Argonaute
ALL	Acute lymphoblastic leukemia
RUNX1	Runt-related transcription factor 1
ANRIL	Antisense RNA in the INK4 locus
ARID5B	AT rich interactive domain 5B
ARMS	Amplification-refractory mutation system
ASO	Allele specific oligos
ATXN7L3B	Ataxin 7 like 3B
B- ALL	B-cell lineage acute lymphoblastic leukemia.
BCR	B-cell receptor
BIB	Bibliographic
BIM	BCL2-like 11 (apoptosis facilitator)
BRMS1L	Breast cancer metastasis-suppressor 1-like
BTK	Bruton tyrosine kinase
CACNA1D	Calcium voltage-gated channel subunit alpha1 D
CACNG1	Calcium voltage-gated channel auxiliary subunit gamma 1
CACNG8	Calcium voltage-gated channel auxiliary subunit gamma 8
CD20	Membrane-spanning 4-domains, subfamily A, member 1
CD22	Sialic acid binding Ig-like lectin 2
CDK	Cyclin- dependent kinase
CDKN2A/B	Cyclin-dependent kinase inhibitor 2A/B
CEBPB	CCAAT/enhancer binding protein beta
CEBPE	CCAAT/enhancer binding protein epsilon
CeGen	Spanish national genotyping center.
CEP68	Centrosomal protein 68kda
CG	Cpg site
CHAS	Chromosome analysis suite.
Chrom	Chromosome
CI	Confidence interval
CIMA	Centre for Applied Medical Research
CN State	Copy Number state
CNOT	CCR4-NOT transcription complex
CNS	Central nervous sistem
CNVs	Copy number variations.
CPdB	Consensuspath database
cSNPs	Coding snps
CYP1A1	Cytochrome P450 family 1 subfamily A member 1
CYP2E1	Cytochrome P450 family 2 subfamily E member 1
del	Deletion

DGCR8	Digeorge syndrome critical region gene 8
DICER	Ribonuclease type III
DNA	Deoxyribonucleic acid
DROSHA	Double-stranded RNA-specific endoribonuclease
dsRNA	Double-stranded RNA
E2A	Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
EBF1	Early B-cell factor 1
EIF2C1	Eukaryotic translation initiation factor 2C, 1
EIF2C2	Eukaryotic translation initiation factor 2C
ENL	Myeloid/lymphoid or mixed-lineage leukemia; translocated to, 1
eQTL	Expression quantitative trait locus
ERCC2	Excision repair-complementing group 2
ERG	V-ets erythroblastosis virus E26 oncogene homolog
ERK	EPH receptor B2
ETV6	Ets variant 6
FDR	False discovery rate
FOXO1A	Forkhead homolog in rhabdomyosarcoma
FOXP1	Fork head-related protein like B
GEMIN3	DEAD (Asp-Glu-Ala-Asp) box polypeptide 20
GEMIN4	Gem (nuclear organelle) associated protein 4
GEMIN5	Gem (nuclear organelle) associated protein 5
GRB2	Growth factor receptor bound protein 2
GSTM1	Glutathione S-transferase M1
GSTT1	Glutathione S-transferase theta 1
GWAS	Genome wide association study
HGP	Human Genome Project
HIWI	Piwi-like 1
HLA	Human leukocyte antigen
HWE	Hardy-Weinberg equilibrium
IGL@	Immunoglobulin lambda locus
IKZF1	IKAROS gene
IL3	Interleukin 3
ISCIII	Instituto de Salud Carlos III
KDM5A	Lysine demethylase 5A
KIF24	Kinesin family member 24
LD	Linkage disequilibrium
lncRNA	Long non coding RNA
LRT	Log-likelihood ratio test
LSO	Locus specific oligos
MAF	Minor allele frequency
MAP2K4	Mitogen-activated protein kinase kinase 4
MAPK	Mitogen-activated protein kinase
MDM2	Mouse double minute 2
MDR1	ATP-binding cassette, sub-family B (MDR/TAP), member 1
MFE	Minimum free energy
MHC	Major histocompatibility complex

miRNA	Microrna
MIRTS	Mirna target site
MLL	Myeloid/lymphoid or mixed-lineage leukemia
mRNA	Messenger RNA
MS4A1	Cd20
MTHFR	5,10-methylenetetrahydrofolate reductase
MTR	Methionine synthase
MTRR	Methionine synthase reductase
N.E.	Not estimable.
N.S.	Non-significant
NA	Not available
ncRNA	Non coding rnas
NFQ	Nonfluorescent quencher
NQO1	NAD (P) H quinone oxidoreductase
NS	Non-synonimous
OR	Odds ratio.
PAX5	Paired box 5 gene.
PBX1	Pre-B-cell leukemia homeobox 1
PCR	Polymerase chain reaction
PCR-RFLP	Polymerase chain reaction - restriction fragment length polymorphism
PDGFR	Platelet derived growth factor receptor beta
PI3K	Phosphatidylinositol 3-kinase
PIP2	Phosphatidylinositol-5,4-bisphosphate
PIP4K2A	Phosphatidylinositol-5-phosphate 4-kinase, type II, alpha
PLC	Phospholipase C
pre-miRNA	Precursor mirna
pri-miRNA	Primary mirna
PTDR	Post-traductional regulation
PTEN	Tumor suppressor and cell cycle regulatory genes.
Pter	P terminus
PTPN11	Protein tyrosine phosphatase, non-receptor type 11
Qter	Q terminus
RAN	Member RAS oncogene family
RAP1B	Member of RAS oncogene family
RB1	Retinoblastoma 1
RFC1	Reduced folate carrier
RFC5	Replication factor C subunit 5
RFLP	Restriction fragment length polymorphism
RISC	RNA-inducing silencing complex
RNA	Ribonucleic Acid
RUNX1	Runt related transcription factor 1
SD	Standard deviation.
SH2B3	SH2B adaptor protein 3
SHMT1	Serine hydromethyl transferase.
SND1	Staphylococcal nuclease and tudor domain containing 1
SNP	Single nucleotide polymorphisms

SOS1	SOS Ras/Rac guanine nucleotide exchange factor 1
SR	Splicing regulation
SRF	Serum response factor
ssRNA	Single stranded RNA
TADA2A	Transcriptional adaptor 2A
TAG	Tagsnp
TARBP2	TAR (HIV-1) RNA binding protein 2
TBP	Tatabox binding protein
TCR	Transcriptional regulation
TEL	Ets variant 6
TFBS	Transcription factor binding site
TNRC6A/B	Trinucleotide repeat containing 6A/B
TR	Transcriptional regulation
TS	Thymidylate synthetase
UR	Upstream regulation.
WGA	Whole genome amplification
WHO	World health organization.
XPO5	Exportin 5
XRCC1	X-Ray repair-cross complementing group 1

Index

INTRODUCTION

ACUTE LYMPHOBLASTIC LEUKEMIA

INTRODUCTION

Acute Lymphoblastic Leukemia (ALL) is the most common type of cancer in children, being the leading cause of death by disease in developed countries. Incidence rates for ALL vary several-fold, internationally, with the highes rates occurring in Spain, among Hispanics in Los Angeles, and in Caucasians in Canada and New Zealand (Wantenberg et al., 2008). In fact, in Spain, ALL corresponds to about 36% of all cases of cancer. However, the causation pathways of this disease are poorly understood.

DEFINITION

ALL is a neoplasm of precursor cells (lymphoblasts), committed to the B- or T-cell lineage. Acquisition by the precursor of a series of genetic abnormalities disturbs its normal maturation process (Figure 1), leading to differentiation arrest and proliferation of the transformed cell. As a consequence, there is accumulation of an immature B- or T-cell clone, typically composed of small to medium-sized blast cells with scant cytoplasm, moderately condensed to dispersed chromatin and inconspicuous nucleoli. By definition, bone marrow is involved in all cases and peripheral blood is usually affected. Extramedullary involvement is frequent, with particular predilection for the central nervous system, lymph nodes, spleen, liver and testis in males (Swerdlov et al. 2008).

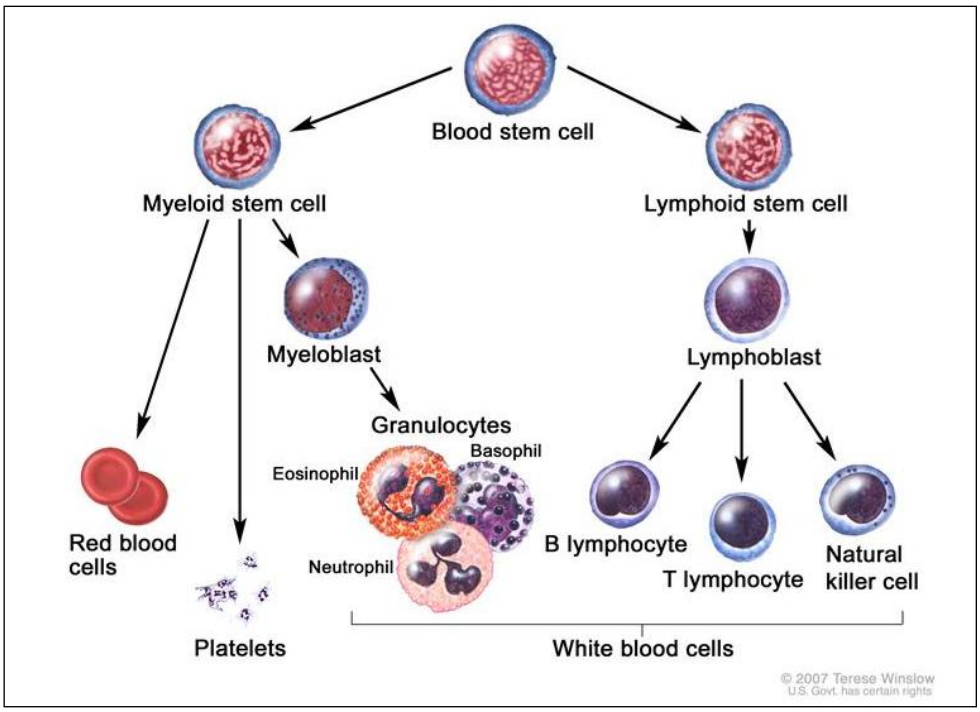

Figure 1: Normal blood cell development.

(http://www.cancer.gov/cancertopics/pdq/treatment/childALL/Patient/page1 NCI).

EPIDEMIOLOGY

Acute Lymphoblastic Leukemia represents the third part of pediatric cancer cases, accounting about 75-80% of all cases of acute leukemia in children. Approximately, 75% of cases occur in children under six years of age and there is a frequency peak between 2 and 5 years (Swerdlov et al. 2008). The worldwide incidence is estimated at 1-4.75/100,000 cases per year. ALL is slightly more common in males than females. Additionally, racial and ethnic differences have been described. Hispanics are more likely to develop acute leukemia than Caucasians and these show higher incidence than African-Americans (Lim et al. 2014).

In Spain, according to the National Child Tumor Registry (RNTI-SEHOP, 1980-2015), the incidence of ALL is 36.6% (CI95%: 34.7-38.6) of all pediatric malignancies in children <14 years old (Figure 2).

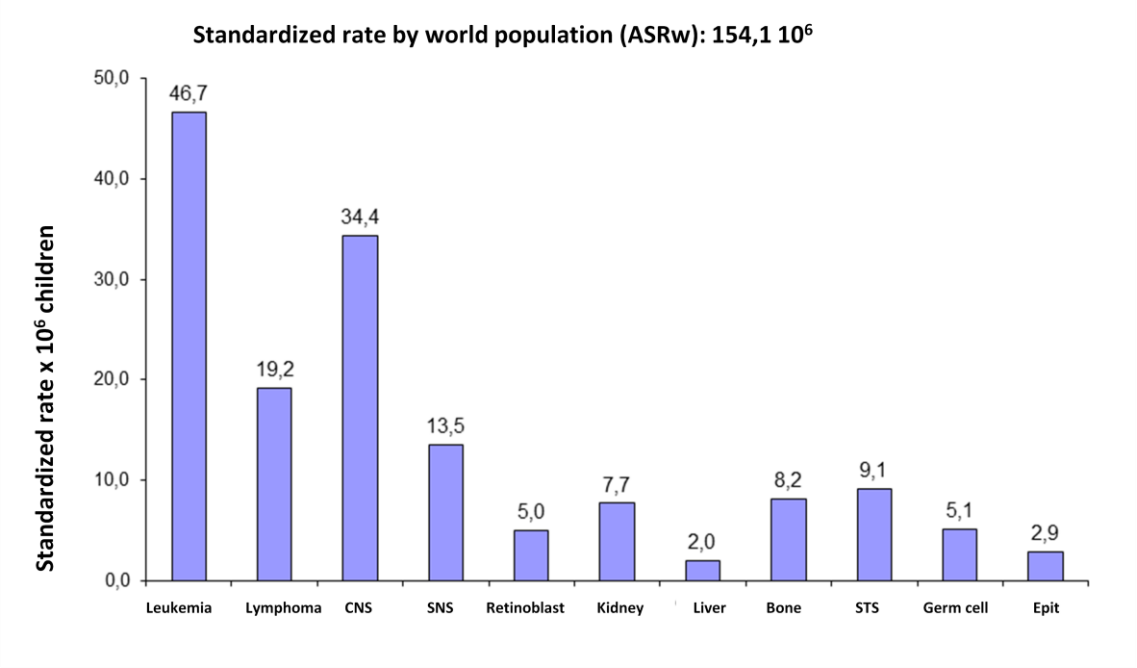

Figure 2: ALL incidence in Spain (RNTI-SEHOP).

The 80–85% of all ALL cases are of B-cell lineage (B-ALL) (Silverman et al. 2000). From now on, we are going to focus on this majority subtype.

DISEASE HETEROGENEITY

One of the main characteristics of B-ALL is its great heterogeneity, with marked differences between individuals at diagnosis, clinical behavior and response to chemotherapeutic agents. The identification of these differences has led to define significant outcome predictors that are currently used to determine a patient's "risk group" and to stratify the intensity of delivered therapy.

Nowadays, the prognostic factors most used in B-ALL risk stratification include age, white blood cell count at diagnosis, extramedullar affection, minimal residual disease and cytogenetics.

Infancy, increasing age (>10 years) and higher white blood cell count are all associated with adverse prognosis. The presence of central nervous system (CNS) disease at diagnosis is also associated with adverse outcome, and requires specific therapy. Additionally, certain cytogenetic abnormalities have been identified as relevant prognostic factors. Based on this knowledge, in 2008 the World Health Organization revised the nomenclature from precursor

B-ALL and established a classification based on 7 specific recurring genetic lesions associated with unique clinical, immunophenotypic, and/or prognostic features (Vardiman et al. 2009) (Table 1) (Figure 3):

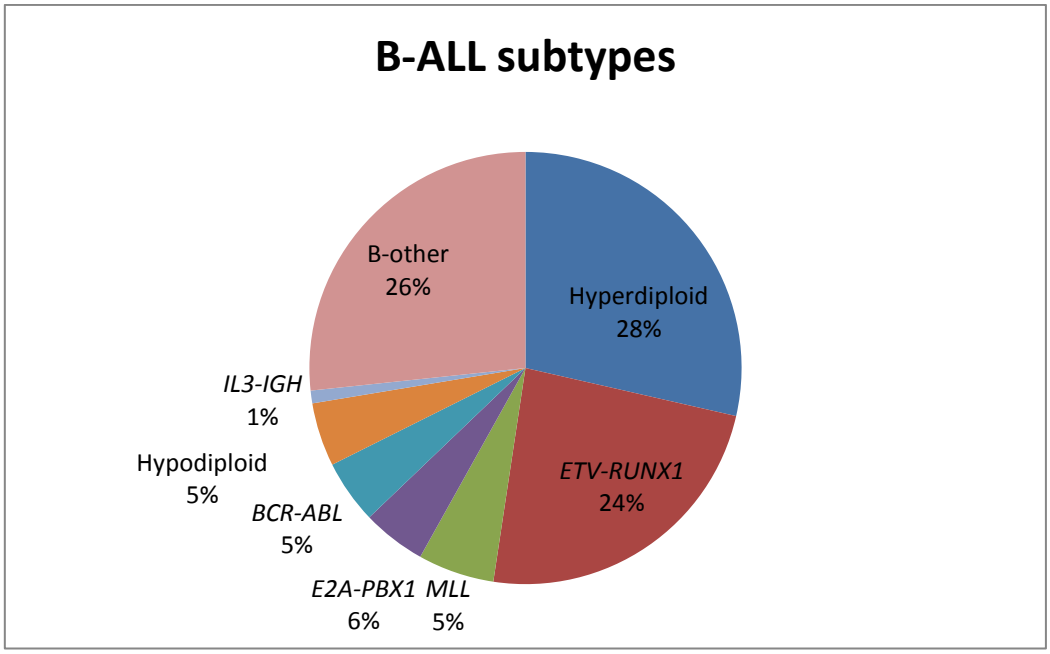

Figure 3: B-ALL subtypes by World Health Organization 2008.

1. Hyperdiploidy ALL: It is the most common subtype accounting for about 30% of B-ALL cases (Paulsson 2016). It is not seen in infants, and decreases in frequency in older children. Hyperdiploid B-ALL contains a numerical increase in chromosomes, usually without structural abnormalities. Extra copies of chromosomes 21, X, 14 and 4 are the most common and chromosomes 1, 2 and 3 are the least often seen (Heerema et al. 2007). This subtype is associated with favorable outcome.

2. *ETV6-RUNX1* ALL: This subtype accounts for about 20-25% of B-ALL cases. It is not seen in infants and decreases in frequency in older children. *ETV6-RUNX1* leukemia is characterized by the translocation t(12;21), also known as *TEL-AML1*, and results in the production of a fusion protein that interferes with normal function of the transcription factor RUNX1, involved in normal hematopoiesis. This leukemia appears to derive from a B-cell progenitor rather than from a hematopoietic stem cell. It is also associated with good prognosis (Swerdlov et al. 2008).

3. *E2A-PBX1* ALL: This leukemia accounts for about 6% of all childhood ALL. This ALL is characterized by the translocation t(1;19)(q23;p13.3). The *E2A-PBX* translocation

results in the production of a fusion protein that has an oncogenic role as a transcriptional activator and also likely interferes with the normal function of the transcription factors involved in lymphocyte development. It is associated with intermediated prognosis (Swerdlov et al. 2008).

4. *MLL* ALL: ALL with rearrangements in the gene *MLL* (11q23) is the most common leukemia in infants <1 year of age. It is less common in older children and increases with age into adulthood. Over 50 fusion partner genes have been identified so far. The most common partner genes are *AF4* transcription factor on chromosome 4q21, *ENL* on chromosome 19p13 and *AF9* on chromosome 9p22. In contrast to the previous subtypes, this subtype has a poor prognosis (Swerdlov et al. 2008).

5. *BCR-ABL* ALL: This subtype accounts for 2-4% of childhood ALL. It is formed by t(9;22)(q34;q11), resulting from fusion of *BCR* at 22q11.2 and the cytoplasmic tyrosine kinase gene *ABL1* at 9q34, with production of a BCR-ABL1 fusion protein. It is believed that the cell of origin of t(9;22) ALL is more immature than that of other B-ALL cases (Cobaleda et al. 2000). This subtype is also a prognostically unfavorable group (Swerdlov et al. 2008).

6. Hypodiploidy ALL: It accounts for about 5% of ALL overall. All patients by definition show loss of one or more chromosomes, having from 45 chromosomes to near haploid (23-29 chromosomes). Structural abnormalities may be seen in the remaining chromosomes though there are no specific abnormalities that are characteristically associated. It has a poor prognosis (Swerdlov et al. 2008).

7. *IL3-IGH* ALL: This is a rare subtype, accounting for 1% of B-ALL cases. This subtype is characterized by the translocation t(5;14)(q31;q32). The unique characteristic of this neoplasm derives from a functional rearrangement between the *IL3* gene on chromosome 5 and the *IGH* gene on chromosome 14, resulting in overexpression of the *IL3* gene. The prognosis of this subtype is not considered to be different from other cases of ALL (Swerdlov et al. 2008).

Table 1: Most common cytogenetic abnormalities in ALL and their prognostic value.

Prognosis	Cytogenetic abnormalities
Favorable or no unfavorable	Hyperdiploidy 51-81 chromosomes
	t(12;21) ETV6-RUNX1+
	Normal Karyotype
Unfavorable	Hypodiploidy 30-45 chromosomes
	Almost tetraploidy 82-94 chromosomes
	Other structural changes not included in the other groups
Very unfavorable	Almost haploidy 24-29 chromosomes
	t(9;22) BCR/ABL+
	t(4;11) MLL+

Interestingly, the existence of biologically different subtypes suggests different etiologies (Inaba et al. 2013, Greaves 2006). For example, in infants ALL is usually associated with *MLL* rearrangement, and the remarkably high concordance rate in monozygotic twins suggests that leukemogenesis is largely complete at birth. By contrast, incidence of other subtypes of B-ALL peaks between 2 and 5 years and has a concordance rate of 10–15%, suggesting that, although initiation in utero usually occurs, other factors are probably necessary for disease emergence (Greaves et al. 2003). However, specific etiology of each subtype is unknown.

ETIOLOGY

Most of pediatric patients usually exhibit the genetic characteristics aforementioned at time of birth, suggesting a prenatally origin of ALL. These prenatal chromosomal aberrations are likely initiating genetic events, occurring during fetal hematopoiesis and operating within a minimal two-hit disease model. Current evidence indicates that transition to overt disease will occur in only a small proportion of children carrying this pre-leukemic clone, after a sufficient second genetic event which likely occurs postnatally. These initiating events, as well as secondary events, are originated due to unknown causes, but like cancer in general, probably arise from interactions between environmental exposures, genetic (inherited) susceptibility, and chance (Figure 4). The challenge is to identify the relevant exposures and inherited genetic variants contributing to the multistep natural history of acute lymphoblastic leukemia (Greaves 2006, Inaba et al. 2013, Pui et al. 2008).

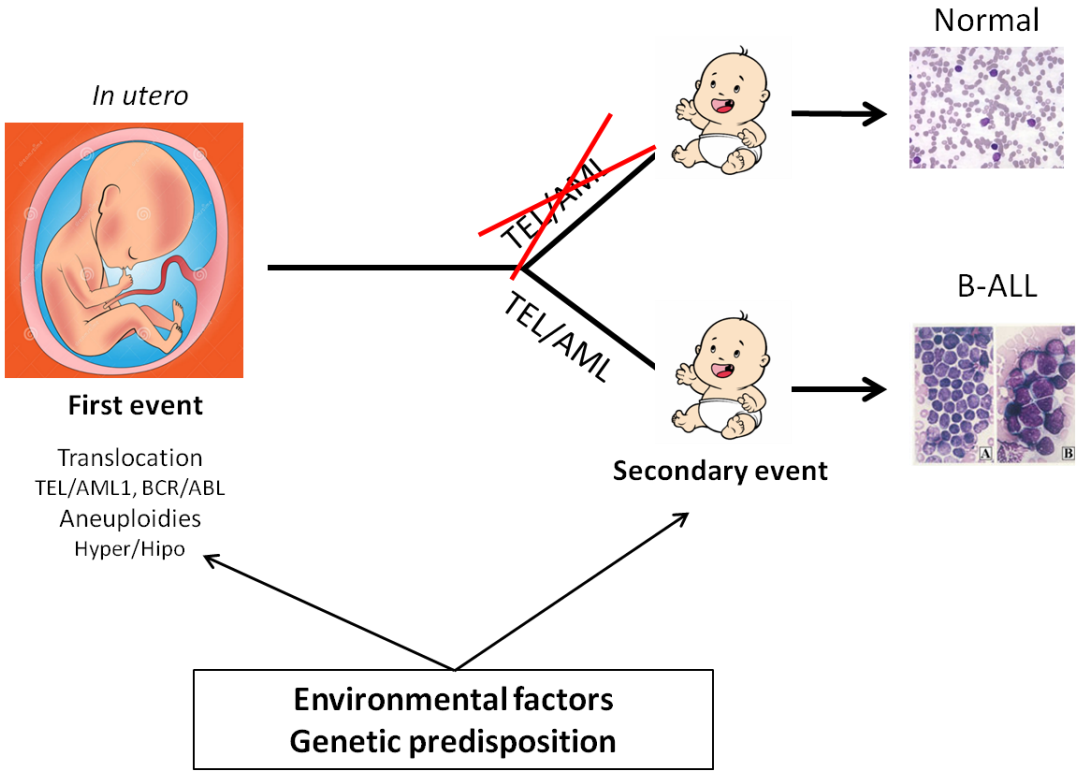

Figure 4: Representation of the development of B-ALL.

A) Environmental factors

Epidemiological and case-control studies have found more than twenty candidate exposures that contribute to childhood ALL but very few are based on reproducibly significant data or are biologically plausible. Some are of public concern, such us ionising radiation, which is an established causal exposure for childhood ALL, as evidenced by the impact in 1945 of atomic bombs in Japan and by the modestly but significantly elevated risk caused by X-ray pelvimetry during pregnancy (Greaves 1997). Others, such as exposures to electromagnetic fields have been particularly controversial. A meta-analysis suggests that high levels of electromagnetic field radiation are associated with slightly increased risk, but the reliability of this finding is uncertain (Schüz 2011). To prove that exposure to electromagnetic fields never causes ALL is impossible, but at most such radiation might be implicated in only a few cases (Inaba et al. 2013).

Apart from this, infection was the first suggested causal exposure for childhood ALL and remains the strongest candidate. Two specific hypotheses have been proposed and both are supported by epidemiological data (Figure 5):

1. Population mixing hypothesis: childhood leukemia may result from an abnormal immune response to specific, although unidentified, infections commonly seen with the influx of infected people into an area previously populated with non-immune, and thus, susceptible individuals (Kinlen 1988).

2. Delayed infection hypothesis: childhood leukemia, particularly common B-cell precursor ALL, may be caused by a proliferative stress-induced effect of common infections on the developing immune system of the child. Implicit in this explanation is that an adverse immune response to infections is a result of insufficient priming of the immune system usually influenced by a delay in exposure to common infectious agents during early childhood (Greaves 1988).

Both, the population mixing and delayed infection hypotheses, are compatible with the available evidence, and in some populations, it is possible that both mechanisms may be operative (Urayama et al. 2008).

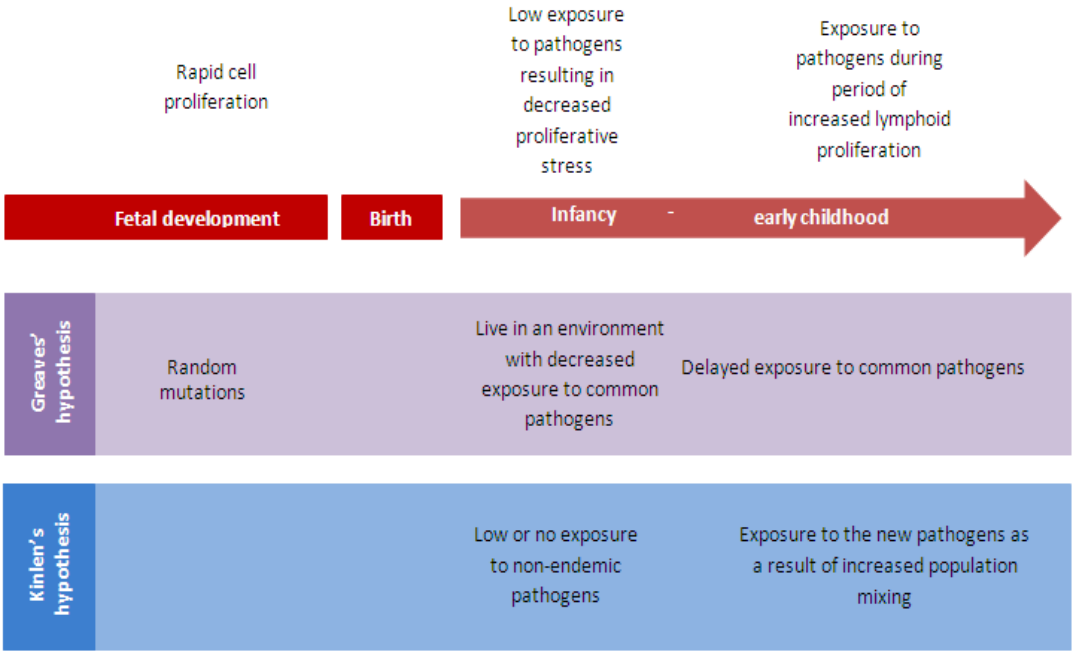

Figure 5: Infection-base models of leukemia development (Pui et al. 2008).

Nevertheless, causation pathways are likely to be multifactorial and it is probable that the risk of ALL from environmental exposure is influenced by genetic variation (Sherborne et al. 2010)

B) Genetic factors

Since years, there are a lot of evidences of the genetic influence in pediatric ALL, such as the early disease onset of the disease; which suggests a strong inherited genetic basis for ALL susceptibility.

In fact, this genetic component is supported by the high risk of ALL associated with Bloom's syndrome, Li-Fraumeni, neurofibromatosis, ataxia telangiectasia and constitutional trisomy 21 (Hsu et al. 2015), most of them caused by rare mutation of high penetrance in genes that predisposes to cancer risk, like *TP53*, *ATM*, *NBN*. In addition, in studies of familial ALL, other rare germline mutations with high penetrance have been identified in *PAX5*, *SH2B3* and *ETV6* (Moriyama et al. 2015a, Shah et al. 2013, Perez-Garcia et al. 2013).

However, the vast majority of childhood ALL is not familial. In contrast, ALL susceptibility is likely influenced by the co-inheritance of multiple low penetrant variants associated with a modestly increased risk of ALL (Papaemmanuil et al. 2009, Moriyama et al. 2015b).

.

LOW PENETRANCE VARIANTS IN B-ALL

The low penetrant variants are common genetic variants such as copy number variations (CNV) and single nucleotide polymorphisms (SNP); being these latter the most often used in the association studies. SNPs are single base substitutions of one nucleotide with another, observed in the general population at a frequency greater than 1%. SNPs are the simplest form of DNA variation among individuals occurring throughout the genome at a frequency of about one in 200-300 bp. Recent large-scale studies have identified approximately 15 million SNPs in the human genome (Consortium 2010).

SNPs can be found across human genome in genes as well as in non-genic regions. Within a gene, its location suggests its function. These potentially functional effects include changes in the exonic sequences, alternative splicing, regulation in the promoter region, changes in transcription factor binding sites, or disruption/creation of CpG sites, that could carry changes in the methylation pattern, or microRNA (miRNA) target sites, involved in the downregulation of gene expression at the post-transcriptional level (Figure 6).

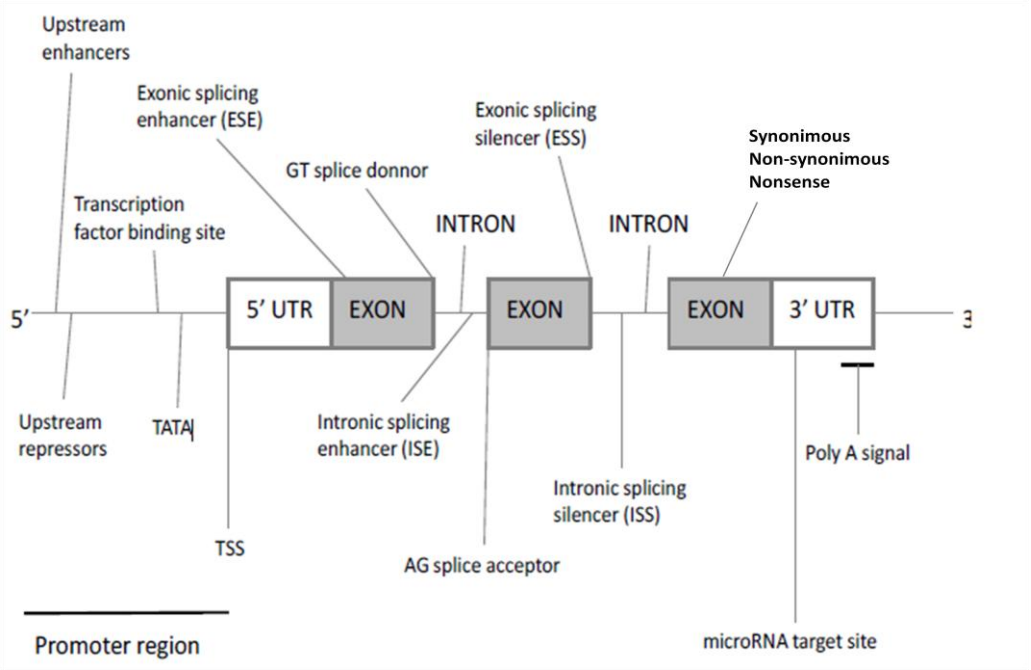

Figure 6: Different functions of SNPs on the gene.

Given the high number of SNPs, it is impractical genotyping all existing common variants. SNPs in the same region of DNA form haplotypes that are typically inherited together. The human genome is composed of stretches of high linkage disequilibrium (LD) (regions with a high level of concomitant inheritance), punctuated by recombination hotspots or points of extremely low LD (Gabriel et al. 2002, Goldstein and Weale 2001, Reich et al. 2001). This means that many SNPs located in the same haplotype block are not inherited independently and show correlated genotypes due to linkage disequilibrium (Sachidanandam et al. 2001, Risch and Merikangas 1996), which results in redundancy of information. The knowledge of the haplotype structure of the genomic region of interest allows the selection of a reduced number of SNPs which 'tag' the common haplotypes of a region, resulting in a great reduction of cost and time (Goldstein and Cavalleri 2005, Howie et al. 2006).

Therefore, a tagSNP is a representative SNP in a region of the genome with high LD. Nowadays, the selection of tagSNPs is facilitated by the existence of The International HapMap Project, a multi-country effort to identify and catalogue genetic similarities and differences in human beings (Figure 7).

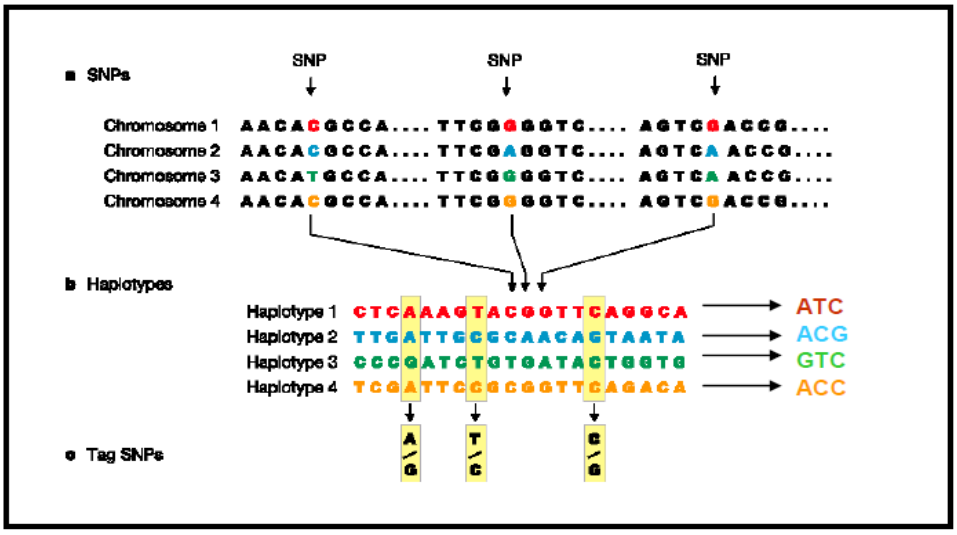

Figure 7: Example of tagSNP a) SNPs identified in DNA samples from multiple individuals. b) Adjacent SNPs that are inherited together are compiled into "haplotypes." c) Detection of "Tag" SNPs within haplotypes that identify uniquely those compiled haplotypes. By genotyping the three tag SNPs shown in this figure, it can be identified which of the four haplotypes shown here are present in each individual.

In B-ALL, SNPs involved in the risk of developing the disease have been studied following two strategies: candidate genes studies and Genome Wide Association Studies or GWAS.

In the candidate gene approach, groups of genes are selected based on their hypothetical function on the disease, while the GWAS approach lacks of hypothesis because the whole genome is analyzed, and sometimes, the results are of difficult interpretation.

CANDIDATE GENE STUDIES

Up to date, there is an extensive body of work that examines the contribution of a number of "candidate" pathways to B-ALL susceptibility, and the number of studies continues to increase. In summary, the genetic candidate studies that have been evaluated in B-ALL risk can be delineated in four main pathways: immune system, carcinogen metabolism, folate metabolism and DNA repair.

A) Immune response

Exposure to common infections and the role of immune-related processes have emerged as strong candidate risk factors for ALL. Thus, the human leukocyte antigen (HLA) genes in the major histocompatibility complex (MHC) are among the most studied genes in relation to the risk of developing B-ALL. For instance, germline SNPs at the classical HLA genes of the class II region HLA-DP and HLA-DOA were associated with ALL susceptibility in admixed populations (Chang et al. 2010, Urayama et al. 2012), although a comprehensive analysis of the MHC region in 824 B-ALL cases and 4737 controls of European genetic ancestry did not find statistically significant association signals (Hosking et al. 2011). These results suggest that caution needs to be exercised when examining HLA variants, especially in diverse populations, because of the complex LD and excessive diversity at these loci in different races and ethnic groups (Moriyama et al. 2015b).

B) Carcinogen metabolism genes

Since children are particularly vulnerable to environmental toxins because of their greater relative exposure, immature metabolism and higher rate of cell division and growth. Detoxifications enzymes, such as cytochrome P4501A1 (CYP1A1), glutathione S-transferases GSTM1 or GSTT1 have been extensively studied. Other metabolic gene variants investigated included polymorphisms in *NQO1* (NADPH:quinone oxidoreductase), a cytosolic enzyme catalyzing reduction of quinones and prevention of their participation in redox cycling and thus in oxidative stress (Brisson et al. 2015, Han et al. 2013). Variation in *MDR1*, which encodes the

P-glycoprotein, was also studied as a possible risk factor for childhood ALL on the basis that it provides a cellular defense against toxic xenobiotic compounds (Urayama et al. 2013, Zhang et al. 2015). However, a recent meta-analysis summarizing 25 studies with a total of 13 polymorphisms in 8 genes of this pathway observed only modest significant associations with ALL susceptibility for 4 variants in *GSTM1, CYP1A1, CYP2E1* and *NQO1* (Vijayakrishnan and Houlston 2010) (Table 2).

C) Folate metabolism genes

Folate and its bioactive metabolic substrates are essential to numerous bodily functions, particularly for their role in DNA methylation and synthesis that aid the rapid cell division and growth requirements associated with pregnancy and early infancy (Lautner-Csorba et al. 2013). Folate deficiency may contribute to carcinogenesis via hypomethylation of important regulatory genes as well as induction of DNA damage through uracil misincorporation during DNA replication (Urayama et al. 2013), thus potentially increasing the risk of chromosomal aberration. Dysfunctional folate metabolism is, therefore, an attractive candidate in the etiology of ALL. Important for folate metabolism are the enzymes 5,10-methylenetetrahydrofolate reductase (MTHFR), methionine synthase (MTR), methionine synthase reductase (MTRR) serine hydroxymethyltransferase (SHMT), thymidylate synthetase (TS) and reduced folate carrier 1 (RFC1 or SLC19A1) which have been investigated as risk factors for ALL (Li et al. 2015, Fang et al. 2014). Vijayakrishnan et al., in their meta-analysis included 18 studies in which 7 SNP were analyzed in 6 genes related to folate metabolism. SNPs in *MTRR*, *SHMT1* and *RFC1* showed significant results (Vijayakrishnan and Houlston 2010) (Table 2).

D) DNA repair genes

Childhood ALL results from chromosomal alterations and somatic mutations that disrupt the normal process by which lymphoid progenitor cells differentiate and senesce. These are the result of unrepaired DNA damage. Since repair of DNA damage is critical, alterations in innate DNA repair pathways may play a role in leukemia development (Urayama et al. 2013, Brisson et al. 2015). *XRCC1* (X-Ray repair-cross complementing group 1) plays a role in DNA single strand repair by forming protein complexes with DNA repair associated proteins and polymorphisms in it have been associated with ALL susceptibility (Wang et al. 2015). Variants in other DNA repair genes, including *ERCC2* (excision repair-complementing group 2), have been also evaluated as risk factors for ALL for similar reasons (Liu et al. 2014). *XRCC1* C26304T

was the only SNP that showed statistically significant association in the meta-analysis performed by Vijayakrishnan et al. (Vijayakrishnan and Houlston 2010) (Table 2).

Table 2: SNPs significantly associated with ALL risk in the meta-analysis of Vijayakrishnan et al.

Gene	SNP	P
Carcinogen matabolism		
CYP1A1	rs4646903	0.003
CYP2E1	rs3813867	0.001
GSTM1	Deletion	0.008
NQO1	rs1800566	0.03
Folate metabolism		
MTRR	rs1805087	0.005
RFC1	rs1051266	0.003
SHMT1	rs1979277	0.028
DNA repair pathway		
XRCC1	rs25487	0.001

Although some polymorphic variants of these pathways were only examined once, most were evaluated as risk factors in several studies, but often with discordant findings. The 8 SNPs in *GSTM1*, *MTRR*, *SHMT1*, *RFC1*, *CYP1A1*, *CYP2E1*, *NQO1* and *XRCC1* associated with ALL susceptibility had a false-positive probability of 20% (Moriyama et al. 2015b, Vijayakrishnan and Houlston 2010). In addition, in the GWAS performed by Papaemmanuil et al. in 2009, several of these candidate genes were interrogated in their analysis, and despite the limitations of the genotyping platforms, as well as differences in the study cohorts used, they did not find that variants in previously reported genes were associated with the risk of ALL in their study (Papaemmanuil et al. 2009).

GENOME WIDE ASSOCIATION STUDIES (GWAS)

This approach enables screens of genetic variation across the entire human genome by using a high-throughput genotyping technology with up to a few million genetic markers tested per patient. Therefore, it avoids the possibility of missing the identification of important variants in hitherto unstudied genes. This is possible due to the availability of comprehensive sets of tagging SNPs that capture much of the common sequence variation in the genome and the availability of high-resolution LD maps that allow GWAS for disease associations to be conducted efficiently (Moriyama et al. 2015b, Houlston 2010, Sherborne et al. 2011). Because of the large number of variants tested in GWASs, the required level of significance for association between a variant and a phenotype is generally set very high ($p>10^{-7}$) rather than

the typical level of .05 used for most power calculations. They also require validation in a second, independent series of patients and independent confirmation by another group of researchers. It should be also noted that commercial genotyping platforms that have been used in GWASs predominantly focus on tagSNPs relatively common. Most of these variants are intronic and may not be directly functional; instead, they are in at least partial linkage with other variants that are likely biologically active. As a result, findings from GWASs often require extensive follow-up studies to discover the true causal genetic variants underlying the GWAS signal (Moriyama et al. 2015b, Sherborne and Houlston 2010).

In ALL, there are up to six GWAS that have identified five loci associated with ALL risk with p values much more significant than the values obtained in the candidate gene approach (Table 3).

Table 3: The most significant SNPs in each loci reported by GWAS studies.

Gene	SNP	P	Study
ARID5B	rs7073837	$p=1,03 \times 10^{-15}$	Papaemmanuil et al., 2009, Treviño et al., 2009, Migliorini et al., 2013, Xu et al., 2013, Orsi et al., 2012
	rs10740055	$p=1,61x\ 10^{-14}$	
	rs7089424	$p=1,41 \times 10^{-19}$	
	rs10821936	$p=1,40 \times 10^{-15}$	
	rs10994982	$p=5,7 \times 10^{-9}$	
IKZF1	rs6964823	$p=1,8 \times 10^{-13}$	Papaemmanuil et al., 2009, Treviño et al., 2009, Migliorini et al., 2013, Xu et al., 2013, Orsi et al., 2012
	rs4132601	$p=9,3 \times 10^{-20}$	
	rsrs6944602	$p=1,5 \times 10^{-15}$	
	rs11978267	$p=8,8 \times 10^{-11}$	
CEBPE	rs2239633	$p=5,6 \times 10^{-8}$	Papaemmanuil et al., 2009, Migliorini et al., 2013, Xu et al., 2013, Orsi et al., 2012
	rs10143875	$p=1 \times 10^{-3}$	
	rs4982731	$p=1 \times 10^{-12}$	
CDKNA2A/B	rs3731217	$p=1.13 \times 10^{-8}$	Sherborne et al., 2010 Orsi et al., 2012
	rs2811709	$p=1 \times 10^{-3}$	
	rs17756311	$p=1 \times 10^{-5}$	
PIP4K2A	rs7901152	$p=1.89\times10^{-8}$	Migliorini et al., 2013 Xu et al., 2013
	rs11013046	$p=2.92\times10^{-9}$	
	rs7088318	$p=1.13\times10^{-11}$	
	rs7075634	$p=2.06\times10^{-10}$	

The first two GWAS performed in ALL were carried out by Treviño et al., and Papaemmanuil et al., in 2009. Both GWAS independently identified two significant loci at 10q21.2 containing AT-

rich interactive domain 5B (*ARID5B)* gene, and at 7p12.2 including Ikaros family zinc finger protein 1 (*IKZF1*) gene (Treviño et al. 2009, Papaemmanuil et al. 2009).

The *ARID5B* gene is a member of the ARID family of transcription factors with important role in embryogenesis and growth regulation. The specific role of *ARID5B* in childhood ALL remains unknown but accumulating evidence appears to indicate that *ARID5B* has a role in ALL development. For instance, Arid5b knockout mice exhibit abnormalities in B-lymphocyte development (Lahoud et al. 2001, Paulsson et al. 2010), and ARID5B mRNA expression is upregulated in hematologic malignancies such as acute promyelocytic leukemia (Chang et al. 2008) and acute megakaryioblastic leukemia (Bourquin et al. 2006).

A total of 5 SNPs in *ARID5B* were associated with childhood B-ALL in both studies. The highest association signal was found for rs7089424 in the GWAS performed by Papaemmanuil et al., and it was in high LD with rs10821936, reported by Treviño et al. In addition, *ARID5B* SNPs were found to be more significantly associated with childhood hyperdiploid B-ALL subtype. The association of these SNPs with ALL risk was a novel finding and was confirmed in subsequent studies and different ethnic groups (Han et al. 2010, Healy et al. 2010, Vijayakrishnan et al. 2010, Yang et al. 2010), which supports the hypothesis that *ARID5B* is involved in a general mechanism that contributes to the etiology of childhood ALL. Of note was the fact that all the significant SNPs in *ARID5B* are located in intron 3 or exhibit high LD with intron 3, without a known function. Then, the mechanism(s) by which these SNPs affect the risk of ALL remain to be elucidated and can be diverse. On the one hand, these SNPs may be markers in LD with copy number variation (CNVs) in the region, which have been previously described (Perry et al. 2008, Gusev et al. 2009). On the other hand, SNPs in *ARID5B* might also have a role in transcriptional regulation, thereby affecting the expression of ARID5B or the splicing, generating different isoforms.

The *IKZF1* gene encodes the early lymphoid transcription factor Ikaros, which is a DNA-binding zinc finger transcription factor involved in the development of all lymphoid lineages (Dai et al. 2014). Germline mutant mice expressing only non–DNA binding dominant-negative leukemogenic Ikaros isoforms develop an aggressive form of lymphoblastic leukemia (Georgopoulos et al. 1994). Chromosomal deletions involving *IKZF1* are common (30%) in high-risk/poor prognosis B-cell precursor ALL and are highly prevalent (95%) in ALL with *BCR-ABL1* fusions (Mullighan et al. 2008, Mullighan et al. 2009).

In *IKZF1*, a total of 4 SNPs were associated with childhood B-ALL in both GWAS. The highest association signal was found for rs4132601 in the GWAS performed by Papemmanuil et al., which was in high LD with rs11978267, reported by Treviño et al. A relatively large number of studies have evaluated the association between *IKZF1* rs4132601 polymorphism and ALL risk, but the results have been inconsistent due to limited sample sizes and different study populations (Li et al. 2015). In order to clarify the possible association between rs4132601 and risk of ALL, two meta-analyses were carried out, all of them confirming the existence of association (Li et al. 2015, Dai et al. 2014). In the last year, new studies showing controversial results have been performed (Kreile et al. 2016, Gharbi et al. 2016, Bahari et al. 2016), so it could be interesting to include these data in a new meta-analysis. The SNP rs4132601 maps in the 3´untranslate region (UTR) of *IKZF1*. Another SNP of the four significant ones, rs6944602, was also localized in this 3´UTR region, suggesting an important role of this region in mRNA stability. In fact, Papaemmanuil et al. found that IKZF1 mRNA expression was significantly associated with rs4132601 genotype in a dose-dependent fashion, with lower expression being associated with risk alleles (Papaemmanuil et al. 2009). However, up to date, this SNP has an unknown function.

The third locus identified was *CEBPE* (CCAAT/enhancer binding protein epsilon), reported by Papaemmanuil et al. at 14q11.2. *CEBPE* is a member of CEBPs family of transcription factors and is involved in terminal differentiation and functional maturation of myeloid cells, especially neutrophils and macrophages (Wang et al. 2015). In childhood ALL, intrachromosomal translocations involving *IGH* and *CEBPE* have been described, resulting in the upregulation of CEBPE expression (Akasaka et al. 2007).

At *CEBPE*, the highest association signal was found for rs2239633. This SNP also showed a strong association with B-hyperdiploid subtype (Wiemels et al. 2016). This finding was replicated in some populations (Prasad et al. 2010, Orsi et al. 2012, Hungate et al. 2016), but not in others (Vijayakrishnan et al. 2010, Healy et al. 2010, Emerenciano et al. 2014). A recent meta-analysis evaluating the association between this polymorphism and the risk of ALL concluded that rs2239633 was associated with the disease (Wang et al. 2015). Nevertheless, some inaccuracies were detected in the study, such as the lack of some important studies (Healy et al. 2010, Ellinghaus et al. 2012). Additionally, the meta-analysis did not include analyses by subtypes and since its publication in 2015 new studies have been published (Kreile et al. 2016, Gharbi et al. 2016) that could help to elucidate if rs2239633 is really associated with the risk of B-ALL. The SNP rs2239633 maps within a 25.7-kb region of LD that

encompasses the gene *CEBPE* and is located in 5´UTR region of the gene. Two other SNPs associated with ALL risk at $P=10^{-5}$ (rs7157021 and rs10143875) map within this region of LD, providing additional support for 14q11.2 as a susceptibility locus. However, the rs2239633 has an unknown function, suggesting that additional polymorphisms underlie the association peak near *CEBPE*.

Recently, Wiemels et al. performed an imputation-based fine-mapping and functional validation analyses of the chromosome 14q11.2 locus in a multi-ethnic case-control population and identified a SNP, rs2239635 located in the promoter region, more significantly associated with B-ALL risk than the previously reported GWAS hit (Wiemels et al. 2016). The SNP rs2239635 is a cis-eQTL for CEBPE, with an increased gene expression associated with risk allele. However, *CEBPE* is not required for B-cell maturation or function, opening a question as to why a polymorphism affecting a subtype of pre-B-ALL may be located proximal to the gene. Interestingly, rs2239635 is located within an Ikaros transcription factor binding site, and the risk allele disrupts Ikaros binding near CEBPE. One of Ikaros functions in the normal lymphoid development is to silence CEBPE. Therefore, incomplete suppression of CEBPE by Ikaros due to rs2239635 may lead to lineage confusion, a common feature of leukemogenesis (Yamanaka et al. 1997), and in turn, promote B-ALL. In addition, Wielmens et al. in their study tested the interaction between rs2239635 and rs4132601, the SNP in *IKZF1* aforementioned, and found that the combined effect of rs2239635 and rs4132601 risk alleles was greater than it would be expected if they operated independently, in contrast of the independent effect of each locus which had been suggested.

The fourth locus identified was at 9p21.3 associated with B-ALL risk (Sherborne et al. 2010). The locus 9p21.3 is particularly noteworthy since it is deleted in around 30% of childhood ALL patients (Walsh et al. 2015). This region comprises *CDKN2A* and *CDKN2B* genes and a long noncoding RNA (lncRNA) known as *ANRIL* (or *CDKN2B-AS*). *CDKN2A* codifies for INK4-class cyclin dependent kinase (CDK) inhibitors $p16^{INK4A}$ and $p14^{ARF}$ (Iacobucci et al. 2011). These proteins are tumor suppressors that block cell cycle division during the G1/S phase and inhibit MDM2, respectively. The second gene *CDKN2B* encodes for the tumour suppressor $p15^{INK4B}$, which is also a ciclyn kinase inhibitor. Finally, *ANRIL* has widespread influences on gene expression, impacting the cell cycle by regulating the expression of tumour suppressors $p14^{ARF}$, $p15^{INK4B}$ and $p16^{INK4A}$ (Congrains et al. 2013).

The most significant variant found by Sherborne et al. was rs3731217 in intron 1 of *CDKN2A*, identified in children from the United Kingdom (Sherborne et al. 2010). This association was replicated in several populations such as Germany, Canada (Sherborne et al. 2010) and France (Orsi et al. 2012), but not in others, like Poland (Pastorczak et al. 2011), Hispanic (Chokkalingam et al. 2013) or Thai population (Vijayakrishnan et al. 2010). In 2012, Orsi et al. (Orsi et al. 2012) in another GWAS also associated one variant located in intron 1 of *CDKN2A*, rs2811709, with B-ALL in French children, a variant in low LD with rs3731217 (r^2<0.8). In a posterior GWAS in 2013, rs17756311 located in *ANRIL*, was identified as the highest associated variant with B-ALL in European Americans, but not in African or Hispanic Americans (Xu et al. 2013). In 2015, three independent studies using genotyping and imputation-based fine-mapping, pointed to rs3731249 in exon 2 of *CDKN2A* as the hit associated variant that conferred high risk for B-ALL in European and Hispanic children (Xu et al. 2015, Walsh et al. 2015, Vijayakrishnan et al. 2015).

Therefore, although there is an obvious implication of CDKN2A/B locus in B-ALL susceptibility, the variants annotated by the different studies are different and are in low LD among them. This may be due to the fact that different variants in each population could alter CDKN2A/B locus function through diverse mechanisms. In fact, it has been suggested that the alleles of rs3731217 create two overlapping cis-acting intronic splice enhancer motifs (CCCAG**G** and CAG**T**AC) that may regulate alternative splicing of *CDKN2A* (Hungate et al. 2016). Regarding rs17756311, Hungate et al. found that a SNP in high LD with it (r^2>0.8), rs662463 in *ANRIL,* regulates *CDKN2B* expression by disrupting a transcription factor binding site (TFBS) for *CEBPB* (Hungate et al. 2016). Finally, rs3731249 is a missense SNP in *CDKN2A* which produces an alanine-to-threonine change in amino-acid-sequence, resulting in reduced tumour suppressor function of $p16^{INK4A}$ (Xu et al. 2015). Interestingly, this SNP is located in the 3´UTR region of $p14^{ARF}$, where it creates binding site for mir-132-5p and mir-4642 (Gong et al. 2012). Therefore, this SNP could cause the downregulation of CDKN2A/B locus. More than other 40 SNPs in 3´UTR region of *CDKN2A* and *CDKN2B* that disrupt or create miRNA binding sites have been described, suggesting their importance in CDKN2A/B regulation. However, studies focused on SNPs in miRNA binding sites are almost absent.

Finally, in 2013, the last locus identified was at 10p12.2, where *PIP4K2A* gene is located (Xu et al. 2013, Migliorini et al. 2013). PIP4K2A is a member of the family of enzymes that catalyze phosphorylation of phosphatidylinositol-5-phosphate to form phosphatidylinositol-5,4-bisphosphate (PIP2), a precursor of the important second messenger molecule, PIP3. Upon B-

cell receptor activation, *PIP4K2A* is directly recruited by BTK to the plasma membrane as a means of stimulating local PIP2 synthesis. Similarly, PIP5K enzymes also interact with the Rho-family small GTP-binding proteins (eg, Rac1) to regulate membrane PIP2 synthesis and PI3K and PLC signaling in B cells (Xu et al. 2013, Migliorini et al. 2013).

In *PIP4K2A*, a total of five SNPs were reported to be associated with B-ALL risk in high LD. The most significant SNP at this locus was rs7088318. In addition, the SNP rs7088318 also showed association with B-hyperdiploid ALL (Xu et al. 2013). However, the only study which tried to confirm this finding in a Hispanic population found no association with B-ALL, although when analyses were limited to hyperdiploid B-ALL, the association approached significance (Walsh et al. 2013).

Another interesting result can be extraced from the previous GWAS. When we examined in detail the exact location of all the SNPs significantly associated with ALL in these GWAS, some of them were located in intergenic regions and not in the genes pointed out in the GWAS. In fact, 37.5% of SNPs found in these GWASs corresponds to intergenic regions (Table 4 and Annex I Table I). In addition, most of the significant SNPs described in GWAS are located in introns or regulatory regions. These data suggest that non-coding regions could play an important role in the risk of ALL. Similar results have been described in GWAs of other cancers.

Table 4: The 20 most significant SNPs found in all GWAS.

SNP	Location	P	Exact Location	Reference
rs7090445	*ARID5B*	4.98×10^{-54}	Intron in *ARID5B*	Migliorini et al., 2013
rs7089424	*ARID5B*	8.41×10^{-51}	Intron in *ARID5B*	Migliorini et al., 2013
rs4506592	*ARID5B*	2.68×10^{-46}	Intron in *ARID5B*	Migliorini et al., 2013
rs10821936	*ARID5B*	$5{,}88 \times 10^{-46}$	Intron in *ARID5B*	Xu et al., 2013
rs7896246	*ARID5B*	$7{,}4 \times 10^{-43}$	Intron in *ARID5B*	Xu et al., 2013
rs10821938	*ARID5B*	6.44×10^{-36}	Intron in *ARID5B*	Migliorini et al., 2013
rs7073837	*ARID5B*	4.12×10^{-34}	Intron in *ARID5B*	Migliorini et al., 2013
rs11980379	*IKZF1*	2.91×10^{-33}	3´UTR of *IKZF1*	Migliorini et al., 2013
rs4132601	*IKZF1*	3.74×10^{-33}	3´UTR of *IKZF1*	Migliorini et al., 2013
rs7923074	*ARID5B*	$1{,}13 \times 10^{-32}$	Intron in *ARID5B*	Xu et al., 2013
rs6964969	*IKZF1*	$1{,}59 \times 10^{-29}$	-	Xu et al., 2013
rs11978267	*IKZF1*	$5{,}29 \times 10^{-24}$	Intron in *IKZF1*	Xu et al., 2013
rs11770117	*IKZF1*	$3{,}98 \times 10^{-23}$	-	Xu et al., 2013
rs12719019	*IKZF1*	$8{,}5 \times 10^{-23}$	-	Xu et al., 2013
rs6952409	*IKZF1*	1.39×10^{-22}	Intron in *IKZF1*	Migliorini et al., 2013
rs10994982	*ARID5B*	$2{,}15 \times 10^{-20}$	Intron in *ARID5B*	Xu et al., 2013
rs6944602	*IKZF1*	$1{,}28 \times 10^{-19}$	-	Xu et al., 2013
rs2239633	*CEBPE*	1.29×10^{-16}	-	Migliorini et al., 2013
rs12434881	*CEBPE*	$1{,}21 \times 10^{-15}$	5´UTR of *CEBPE*	Xu et al., 2013
rs10740055	*ARID5B*	$1{,}61 \times 10^{-14}$	Intron in *ARID5B*	Papaemmanuil et al. 2009

NON CODING REGIONS

The Human Genome Project (HGP) revealed in 2001 that the number of genes codifying for proteins are about 20.000-25.000, which corresponds for only 1.5% of the genome approximately, a proportion that increases to 2% if untranslated regions are considered (Esteller 2011).

By the end of the human genome sequencing, new challenges were proposed in order to interpret the data generated by the HGP, and a subsequent Project called ENCODE (Encyclopedia Of DNA Elements) was launched. The Project aimed to prepare a complete catalog which contained all functional elements codified in the human genome, and one of the main conclusions was that about 80% of our genome is transcribed as elements that do not codify for proteins. Early considered "junk DNA", it was determined that a large part of the non-protein coding regions were functional. These elements are non-coding RNAs (ncRNA) (Consortium 2012).

NcRNAs are classified by its length. Among the small non coding RNAs, microRNAs (miRNAs) are the most studied and it is kwnon their involvement in cancer (Sana et al. 2012) (Figure 8).

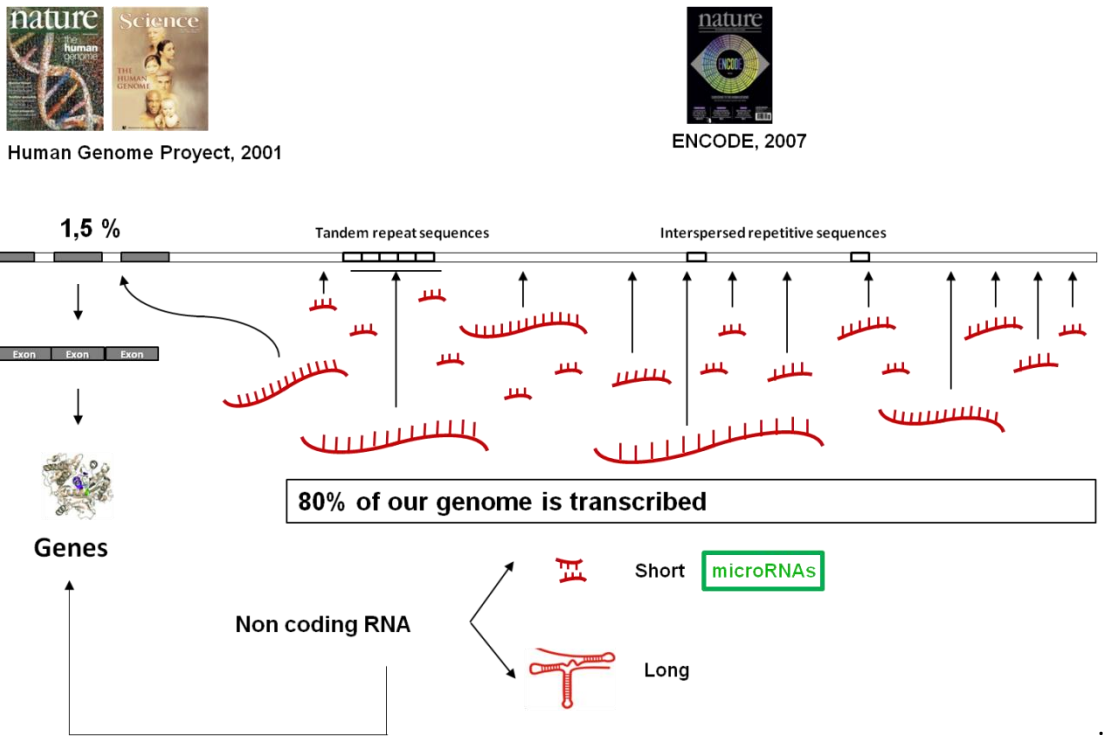

Figure 8: Transcription of our genome.

MIRNAS

MiRNAs comprise a large family of 18-22 nucleotide-long RNAs that have emerged as key regulators of genes at the post-transcriptional level. MiRNAs are transcribed from different locations in the genome by RNA polymerase II into long primary transcripts called pri-miRNAs (dsDNA, 300-5000pb). The pri-miRNAs are characterized by a central region of double-stranded RNA (dsRNA) of about 30-40 nucleotides, a terminal loop and two single stranded RNA (ssRNA) opposite among each other. These pri-miRNAs are processed in the nucleus by the complex formed by DROSHA RNase and DGCR8 containing dsRNA binding domains. DsRNA sequence determines its secondary structure and its binding with processing proteins.

After processing of pri-miRNAs, these smaller molecules (about 70 nucleotides) are known as pre-miRNAs. The pre-miRNAs are exported from the nucleus to the cytoplasm through Exportine5 (XPO5) and RAN GTPase (Bohnsack et al. 2004, Kim 2004) protein. In the cytoplasm, the pre-miRNAs are processed by Dicer (Hutvágner et al. 2001, Merritt et al. 2010) and TARBP2 enzyme, which eliminate the loop, generating a dsRNA molecule known as miRNA duplex (Song et al. 2003). The miRNA duplex is separated to form the mature miRNA as a single strand. The strand selected of the miRNA duplex is incorporated into multiprotein complex known as RISC (RNA-inducing silencing complex), composed of the *EIF2C1* (*AGO1*), *EIF2C2* (*AGO2*), *SND1*, *GEMIN3*, *GEMIN4* and CCR-NOT complex (Inada and Makino 2014) genes. The mature miRNA is transported by the RISC complex to messenger RNAs (mRNAs), target of the regulation (Li et al. 2014). The miRNA binds to messenger RNA complementary bases at the 3´ UTR region (Figure 9).

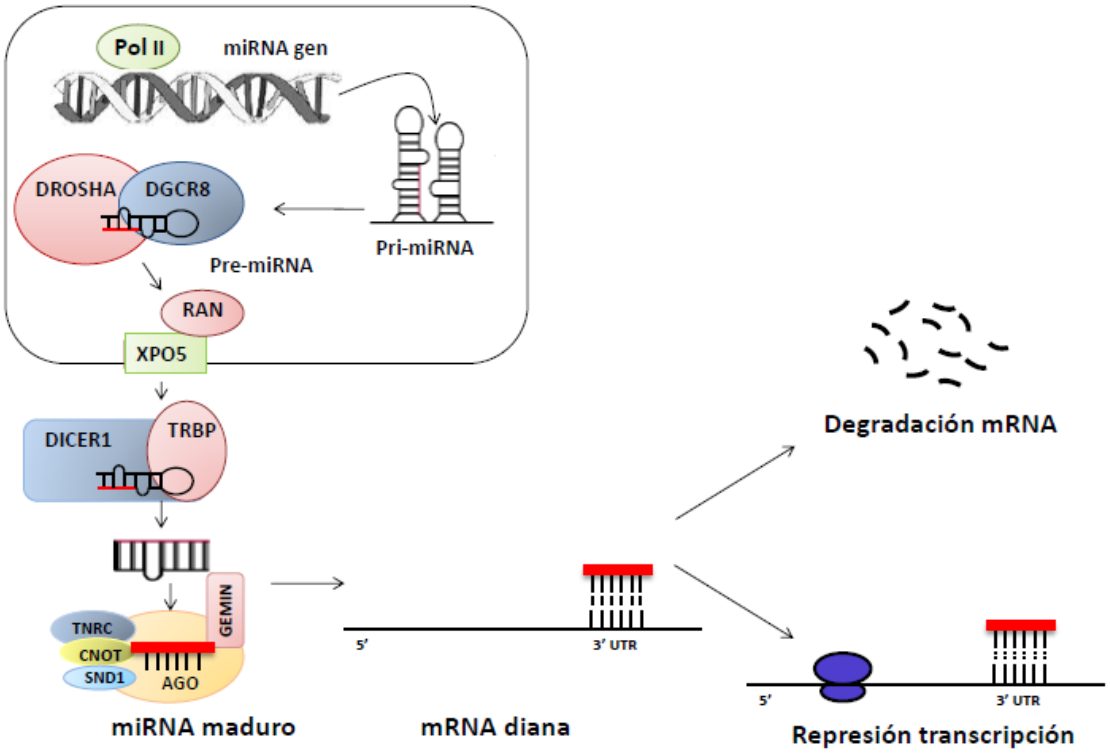

Figure 9: miRNA biogenesis and mechanism of action (adapted from (Ryan et al. 2010)).

MiRNAs have a characteristic sequence for target recognition of approximately 7 bp, known as *seed* region. The miRNA acts by specific binding of the *seed* sequence to a complementary target sequence. The regulation mechanism depends on the degree of miRNA-mRNA complementarity: direct cleavage and degradation when the complementarity is perfect; protein translation blocking/inhibition in the case of imperfect base pairing (Gregory et al. 2006).

Since the sequence complementary to a miRNA *seed* is short, a miRNA may degrade or repress translation of many targets mRNA containing complementary sequences to the *seed* region. But there is also the possibility that a gene can be regulated by multiple miRNAs. MiRNAs sharing all or part of the nucleotide sequence of the *seed* region can be grouped into families. Members of the same family of miRNAs are potential regulators of the same set of mRNAs (Lewis et al. 2005, Friedman et al. 2009)

For each miRNA, there are many putative targets predicted by databases such as Mirwalk (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/) and TARGETSCAN (http://www.targetscan.org/vert_71/), among others. However, nowadays miRNA targets are not completely defined and few interactions are validated experimentally.

Through this mechanism of regulation, miRNAs could regulate more than 50% of human genes, having an enormous impact on the function of any cell (Johanson et al. 2014), including B-lymphocytes.

Gene regulation mediated by miRNAs can be affected by both, changes in the levels of miRNAs or changes in the binding sequence. MiRNA levels could be altered due to changes in processing genes and/or changes in pre-miRNAs. Therefore, SNPs affecting the proteins involved in miRNA biogenesis may have deleterious effects on the general miRNAome, while SNPs in pre-miRNAs which determine the miRNAs secondary structure, may affect their stability or processing efficiency, affecting their own levels. On the other hand, the miRNA *seed* sequence determines the miRNA binding. Therefore, SNPs in the *seed* region could affect the accurate recognition of its targets (Figure 10) (Ryan et al. 2010).

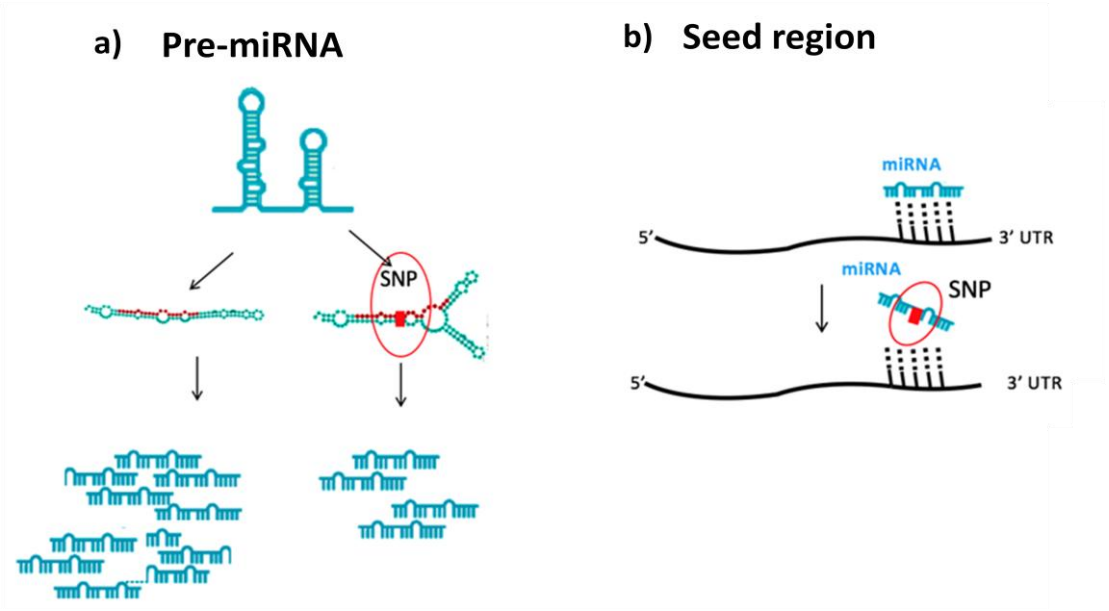

Figure 10: SNPs in pre-miRNA could a) alter miRNA secondary structure affecting its own levels or b) miRNA binding with its targets.

MIRNA PROCESSING GENES AND ACUTE LYMPHOBLASTIC LEUKEMIA

Any alteration in miRNA processing genes could affect the levels of miRNAs. When the expression of members of the Ago family or Dicer are removed, the synthesis of mature miRNAs in mouse models is impaired and B-cell differentiation is affected, highlighting the importance of miRNAs in the formation of B-cells (Marques et al. 2015). For instance, deletion of *EIF2C1* in bone marrow cells was shown to compromise B-lymphoid lineage development

due to impaired miRNA-mediated gene silencing (O'Carroll et al. 2007). Similarly, Dicer ablation in early B-cell progenitors resulted in an almost complete developmental block at the pro- to pre-B-cell transition due to a massive increase in apoptosis (Koralov et al. 2008, Belver et al. 2010).

Several genes of this pathway have been shown to be deregulated in cancer, either because they are over-expressed, as *EIF2C2* (AGO complex) and *TARBP2* (Dicer complex) in prostate cancer, or because they are under-expressed, as DROSHA and Dicer in breast cancer (Huang et al. 2014). Any alteration in these genes could affect miRNA levels, and in turn, the levels of their targets. Therefore, SNPs affecting the proteins involved in miRNA biogenesis may affect the miRNAome. In fact, a SNP in *DROSHA* (rs640831) causes deregulation of 56 miRNAs (Rotunno et al. 2010) in lung cancer. Several SNPs in miRNA processing genes have been described in association with different cancer types (Liu et al. 2014, Kim et al. 2010, Martin-Guerrero et al. 2015, Bilbao-Aldaiturriaga et al. 2015) (Table 5). However, despite all these evidences, there are no studies analyzing SNPs in processing genes associated with childhood B-ALL.

Table 5: SNPs in processing genes associated with the risk ofdifferent cancer types.

Gene	SNP	Disease	Reference
XPO5	rs11077	Hepatocellular carcinoma	Liu et al., 2014
RAN	rs14035	Oral cancer	Roy R et al., 2014
AGO1	rs636832	Lung cancer	Kim et al., 2010
DGCR8	rs417309	Breast cancer	Jiang Y et al., 2013
GEMIN3	rs197412	Oral cancer	Roy R et al., 2014
	rs197414	Bladder cancer	Yang H et al., 2008
GEMIN4	rs2740348	Prostate cancer	Liu J et al., 2012
	rs7813		
	rs2740351	Ovarian cancer	Liang D et al., 2010
CNOT1	rs11866002	Osteosarcoma	Bilbao-Aldaiturriaga et al., 2015
DROSHA	rs3805500	Chronic Lymphocytic Leukemia	Martín-Guerrero et al., 2015
	rs6877842		

MIRNA AND ACUTE LYMPHOBLASTIC LEUKEMIA

The role of individual miRNAs expressed in the B-cell lineage was also examined. For instance, the miR-17-92 cluster was shown to have an essential role in B-cell development as its deletion led to the pro- to pre-B-cell transition block (Ventura et al. 2008). The functionally important targets of miR-17-92 comprise pro-apoptotic proteins PTEN and BIM. The pro- to pre-B-cell transition also appears to be regulated by another miRNA, miR-34a (Rao et al. 2010). MiR-34a

is expressed at the highest levels in pro-B cells, and its constitutive expression in the bone marrow leads to a developmental block at the pro- to pre-B-cell stage resulting in a decrease of mature B cells. In contrast, miR-34a knockdown resulted in increased amounts of mature B cells. The key mediator of the miR-34a effect is the transcription factor FOXP1, which was found to be its direct target required for B-cell development (Rao et al. 2010). Therefore, because various miRNAs were shown to act as direct regulators of B-cell development and differentiation, it is not surprising that the deregulation of miRNAs is a common event in B-cell malignancies.

In ALL, several studies have shown the deregulation of miRNAs compared to normal donors (Schotte et al. 2009, Schotte et al. 2011a, Schotte et al. 2011b, Zhang et al. 2009, Ju et al. 2009, Duyu et al. 2014, de Oliveira et al. 2012). A report examining 40 newly diagnosed pre-B-ALL samples showed that in childhood ALL, miR-222, miR-339, and miR-142-3p were overexpressed along with the downregulation of miR-451, miR-373 (Ju et al. 2009). Additionally, a report by Schotte et al. (Schotte et al. 2009) examined miRNAs expression levels in pediatric ALL samples in comparison to normal CD34 positive cells, providing evidence for the upregulation of miR-128a, miR-142, miR-150, miR-181, miR-30e-5p, miR-193, miR-34b, miR-365, miR-582, and miR-708 and the downregulation of miR-100, miR-125b, miR-99a, miR-196b, and miR-let-7e. In a posterior study, using high-throughput sequencing technology, Schotte et al. discovered 153 known miRNAs, 16 novel and 170 candidate novel mature miRNAs and miRNA-star strands only expressed in ALL, whereas 140 known, 2 novel and 82 candidate novel mature miRNAs and miRNA-star strands were unique to normal hematopoietic cells (Schotte et al. 2011a). More recently, additional studies have identified miR-143, miR-145, and miR-223 high expression in normal bone marrow and miR-127, miR-299-5p, and miR-411 high expression in normal CD34 positive cells, differentiating these cells from ALL samples (Schotte et al. 2011b).

In addition, unique miRNA signatures were identified for various ALL subtypes including ETV6-RUNX1, MLL-rearranged, hyperdiploidy, E2A-PBX1 and BCR-ABL (Akbari Moqadam et al. 2014, Schotte et al. 2009, Schotte et al. 2011a, Schotte et al. 2011b). A comparative analysis of MLL-rearranged leukemia with precursor B-ALL negative for common cytogenetic aberrations (B-other) in children, demonstrated low expression of miR-193, miR-151-5p, let-7e, miR-30e-5p, miR-34b, miR-582, and miR-708 and high expression of miR-196b. Additionally, MLL-rearranged cells displayed lower miR-708 and higher expression of miR-196b when compared to other ALL subtypes (*ETV6-RUNX1*, *BCR-ABL*, *E2A-PBX1*, hyperdiploid, and B-other) (Schotte

et al. 2009). Schotte et al. identified unique miRNA expression profiles for each pediatric ALL subtype measuring the expression level of 397 miRNAs in 81 cases of ALL (Schotte et al. 2011b). The authors were able to differentiate many of the major subtypes of ALL, such as *MLL*, *TEL-AML1*, *E2A-PBX1*, and hyperdiploid cells.

Alterations in miRNAs can alter their function affecting their targets genes expression. Genetic variants can modify the miRNA expression levels if they are located in the pre-miRNA or the mRNA-miRNA binding if they are located in the seed region. Therefore, SNPs in miRNA could be involved in B-ALL susceptibility.

Despite all these evidences, only three studies analyzing the involvement of SNPs in miRNAs in the risk of B-ALL have been performed (Hasani et al. 2013, Tong et al. 2014, Tong et al. 2015). Hasani and colleagues studied rs3746444 in miR-499 and rs2910164 in mir-146a and found rs2910164 in mir-146a associated with ALL susceptibility in a Iranian population of 75 children diagnosed with ALL (Hasani et al. 2013). Tong and colleagues in 2014 found association between rs11614913 in mir196a-2 and ALL risk in a Chinese population of 574 pediatric ALL patients, and with the same population, in 2015 found association between rs4938723 in pri-mir-34b. Remarkably, although a relatively low number of SNPs were analyzed in miRNAs and B-ALL susceptibility, significant results were found.

ANNEX I

Annex table 1: All the SNPs reported by GWAS performed by Papaemmanuil et al., Treviño et al., Sherborne et al., Xu et al. and Migliorini et al. In red, SNPs in intergenic regions or in genes that not codify for proteins.

SNP	Gene	Function	Location	Reference
rs2784140	*SLC6A17*	Synonymous codon	1p13.3	Migliorini et al., 2013
rs17115122			1p21.3	Sherborne et al., 2010
rs10873876	*SIAT7C* *ST6GALNAC3*	Intron variant	1p31.1	Treviño et al., 2009
rs11211481	*TAL1*	Intron variant	1p33	Orsi et al., 2012
rs2248907	*PDZK1I* *PDZK1IP1*	upstream variant 2KB flanking_5UTR	1p33	Orsi et al., 2012
rs11799849			1p36.12	Sherborne et al., 2010
rs6428370	*LOC100996886*	lncRNA	1q31.3	Treviño et al., 2009
rs2405523	*CNIH3*	Intron variant	1q42.12	Orsi et al., 2012
rs7554607	*RYR2*	Intron variant	1q43	Treviño et al., 2009
rs1881797	*GCSALM* *GCSALM-AS1*	Intron variant	1q44	Treviño et al., 2009
rs4853946	*MYT1L*	Intron variant	2p25.3	Sherborne et al., 2010
rs896232			2p25.3	Sherborne et al., 2010
rs930372	*ANKRD44*	Intron variant	2q33.1	Orsi et al., 2012
rs12621643	*KCNE4*	Missense	2q36.1	Treviño et al., 2009
rs12162384	*SLC16A14* (20K)	flanking_5UTR	2q36.3	Orsi et al., 2012
rs774588	*ROBO2*	Intron variant	3p12.3	Orsi et al., 2012
rs267103	*ROBO2*	Intron variant	3p12.3	Orsi et al., 2012
rs11708509	*OR5K4* (2K)	flanking_3UTR	3q11.2	Orsi et al., 2012
rs9290663	*KCNMB2* *KCNMB2-AS1*	Intron variant	3q26.32	Treviño et al., 2009
rs2130904	LOC102723846	Intron variant	4p15.1	Sherborne et al., 2010
rs10002424			4q13.1	Sherborne et al., 2010
rs4916794	*CETN3* (110K)	flanking_3UTR	5q14.3	Orsi et al., 2012
	LINC01339	Intron variant	5q14.3	
rs33584			5q23.1	Migliorini et al., 2013
rs405510			5q23.1	Sherborne et al., 2010
rs10061417			5q23.1	Sherborne et al., 2010
rs1800197	*PROP1*	Missense	5q35.3	Sherborne et al., 2010
rs7448421	*ZNF354B* *ZFP2* (19K)	Intron variant	5q35.3	Sherborne et al., 2010
rs2935505	*CETN3* (150K)	flanking_3UTR	5q41.3	Orsi et al., 2012
	LINC01339	Intron variant	5q41.3	
rs1870262	*IMPG1* (550K)	flanking_5UTR	6q14.1	Orsi et al., 2012
rs1336767	*NKAIN2; TCBA1*	Intron variant	6q22.31	Sherborne et al., 2010
rs11155133	*LOC102723724*	Intron variant	6q24.1	Treviño et al., 2009
rs4716398	*C6orf7* *ERMARD*	Intron variant	6q27	Orsi et al., 2012
rs7809758	*IKZF1*	Intron variant	7p12.1	Migliorini et al., 2013; Papaemmanuil et al., 2009
rs6592961	*IKZF1* (100K) *DDC*	Intron variant	7p12.1	Migliorini et al., 2013
rs12718572	*IKZF1* (100K) *DDC*	Intron variant	7p12.1	Migliorini et al., 2013

Annex table 1: All the SNPs reported by GWAS performed by Papaemmanuil et al., Treviño et al., Sherborne et al., Xu et al. and Migliorini et al. In red, SNPs in intergenic regions or in genes that not codify for proteins (continue).

SNP	Gene	Function	Location	Reference
rs17133853	*DDC*	Intron variant	7p12.1	Xu et al., 2013
rs7791875	*DDC*	Intron variant	7p12.1	Xu et al., 2013
rs3823674	*DDC*	Intron variant	7p12.1	Xu et al., 2013
rs4947584	*DDC*	Intron variant	7p12.1	Xu et al., 2013
rs6944090	*DDC*	Intron variant	7p12.1	Xu et al., 2013
rs7808025	*DDC*	Intron variant	7p12.1	Xu et al., 2013
rs6592963	*DDC*	Intron variant	7p12.1	Xu et al., 2013
rs12535064	*DDC*	Intron variant	7p12.1	Xu et al., 2013
rs12538830	*DDC*	Intron variant	7p12.1	Xu et al., 2013
rs9918702	*DDC*	Intron variant	7p12.1	Xu et al., 2013
rs9791817	*GRB10*	Intron variant	7p12.1	Xu et al., 2013
rs4245555	*GRB10*	Intron variant	7p12.1	Xu et al., 2013
rs4245556	*GRB10*	Intron variant	7p12.1	Xu et al., 2013
rs12540874	*GRB10*	Intron variant	7p12.1	Xu et al., 2013
rs4947709	*GRB10*	Intron variant	7p12.1	Xu et al., 2013
rs2074778	*GRB10*	Intron variant	7p12.1	Xu et al., 2013
rs4947737	*GRB10*	Intron variant	7p12.1	Xu et al., 2013
rs2237450	*GRB10*	Intron variant	7p12.1	Xu et al., 2013
rs2244372	*GRB10*	Intron variant	7p12.1	Xu et al., 2013
rs2244353	*GRB10*	Intron variant	7p12.1	Xu et al., 2013
rs3779084	*DDC*	Intron variant	7p12.1	Papaemmanuil et al., 2009
rs880028	*DDC*	Intron variant	7p12.1	Papaemmanuil et al., 2009
rs11980379	*IKZF1*	UTR variant 3 prime	7p12.2	Migliorini et al., 2013
rs4132601	*IKZF1*	UTR variant 3 prime	7p12.2	Migliorini et al., 2013; Orsi et al., 2012; Papaemmanuil et al., 2009
rs6952409	*IKZF1* *LOC105375275*	Intron variant nc transcript variant	7p12.2	Migliorini et al., 2013; Orsi et al., 2012; Papaemmanuil et al., 2009
rs4490786	*IKZF1*	Intron variant	7p12.2	Migliorini et al., 2013
rs10499691	*IKZF1* (20K)	flanking_3UTR	7p12.2	Migliorini et al., 2013
rs11976368	*IKZF1* (20K)	flanking_3UTR	7p12.2	Migliorini et al., 2013
rs11575553	*IKZF1* (50K)	flanking_3UTR	7p12.2	Migliorini et al., 2013; Xu et al., 2013
	DDC	utr variant 3 prime	7p12.2	
rs7809377	*IKZF1*	Intron variant	7p12.2	Migliorini et al., 2013
rs7797772	*IKZF1* *LOC105375275*	Intron variant nc transcript variant	7p12.2	Migliorini et al., 2013
rs7781977	*IKZF1*	Intron variant	7p12.2	Migliorini et al., 2013
rs11575548	*IKZF1* (60K) *DDC*	Intron variant	7p12.2	Migliorini et al., 2013
rs4917017	*IKZF1* (5K)	flanking_5UTR	7p12.2	Migliorini et al., 2013
rs10235226	*IKZF1* (100K)	flanking_5UTR	7p12.2	Xu et al., 2013
rs921909	*IKZF1* (100K)	flanking_5UTR	7p12.2	Xu et al., 2013
rs921910	*IKZF1* (100K)	flanking_5UTR	7p12.2	Xu et al., 2013
rs9886239	*IKZF1* (6K)	flanking_5UTR	7p12.2	Xu et al., 2013
rs11765988	*IKZF1*	Intron variant	7p12.2	Xu et al., 2013
rs7800411	*IKZF1*	Intron variant	7p12.2	Xu et al., 2013

Annex table 1: All the SNPs reported by GWAS performed by Papaemmanuil et al., Treviño et al., Sherborne et al., Xu et al. and Migliorini et al. In red, SNPs in intergenic regions or in genes that not codify for proteins (continue).

SNP	Gene	Function	Location	Reference
rs12719039	*IKZF1*	Intron variant	7p12.2	Xu et al., 2013
rs7806674	*IKZF1*	Intron variant	7p12.2	Xu et al., 2013
rs7790846	*IKZF1*	Intron variant	7p12.2	Xu et al., 2013
rs12669559	*IKZF1*	Intron variant Synonymous codon	7p12.2	Xu et al., 2013
rs11978267	*IKZF1* *LOC105375275*	Intron variant nc transcript variant	7p12.2	Xu et al., 2013; Treviño et al., 2009
rs6964969	*IKZF1* (450bp)	downstream variant 500B flanking_3UTR	7p12.2	Xu et al., 2013
rs6944602	*IKZF1* (950bp)	flanking_3UTR	7p12.2	Xu et al., 2013; Papaemmanuil et al., 2009
rs11770117	*IKZF1* (950bp)	flanking_3UTR	7p12.2	Xu et al., 2013
rs12719019	*IKZF1* (3K)	flanking_3UTR	7p12.2	Xu et al., 2013
rs10261922	*DDC* (55K)	flanking_3UTR	7p12.2	Xu et al., 2013
rs11575575	*DDC* (2K)	flanking_3UTR	7p12.2	Xu et al., 2013
rs7803247	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs3887825	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs12718527	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs12718528	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs4580999	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs10899734	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs10899735	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs10899736	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs11575457	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs4948196	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs17152020	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs1037351	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs2167364	*DDC*	Intron variant	7p12.2	Xu et al., 2013; Treviño et al., 2009
rs4947582	*DDC*	Intron variant	7p12.2	
rs11575387	*DDC*	Intron variant	7p12.2	Xu et al., 2013
rs6964823	*IKZF1* *LOC105375275*	Intron variant Nc transcript variant	7p12.2	Papaemmanuil et al., 2009
rs6945253			7p12.2	Papaemmanuil et al., 2009
rs1349492	*DDC*	Intron variant	7p12.2	
rs6592952	*DDC*	downstream variant 500B intron variant	7p12.2	Papaemmanuil et al., 2009
rs2242041	*DDC*	Intron variant	7p12.2	Treviño et al., 2009
rs6951648	*DDC*	Intron variant	7p12.2-12.1	Xu et al., 2013
rs2107449	*SP4*	Intron variant	7p15.3	Migliorini et al., 2013
rs10249014	*SP4* (2K)		7p15.3	Migliorini et al., 2013
rs2390538	*SP4*	Intron variant	7p15.3	Migliorini et al., 2013
rs11764793	*TES* (19K)	flanking_5UTR	7q31.2	Sherborne et al., 2010
rs11136067	*DUSP4* (100K)	flanking_5UTR	8p12	Orsi et al., 2012

Annex table 1: All the SNPs reported by GWAS performed by Papaemmanuil et al., Treviño et al., Sherborne et al., Xu et al. and Migliorini et al. In red, SNPs in intergenic regions or in genes that not codify for proteins (continue).

SNP	Gene	Function	Location	Reference
rs667656	*DUSP4* (70K)	flanking_5UTR	8p12	Orsi et al., 2012
rs7835507	*MSRA*	Intron variant	8p23.1	Sherborne et al., 2010
rs6997224	*MSRA*	Intron variant	8p23.1	Sherborne et al., 2010
rs1427050	*RALYL*	Intron variant	8q21.2	Migliorini et al., 2013
rs3935421	*CCDC26* (160K)	flanking_3UTR	8q24.21	Orsi et al., 2012
rs7843653	*KCNQ3*	Intron variant	8q24.22	Orsi et al., 2012
rs10813050	*LRRN6C* (100K) *LINGO2*	flanking_5UTR	9p21.1	Orsi et al., 2012
rs1331876	*LINGO2*	Intron variant	9p21.1	Sherborne et al., 2010
rs3731217	*CDKN2A*	Intron variant	9p21.3	Migliorini et al., 2013; Sherborne et al., 2010
rs7049105	*CDKN2B-AS1*	Intron variant	9p21.3	
rs10120688	*CDKN2B-AS1*	Intron variant	9p21.3	Migliorini et al., 2013
rs2811712	*CDKN2B-AS1*	Intron variant	9p21.3	Migliorini et al., 2013
rs1333034	*CDKN2B-AS1*	Intron variant	9p21.3	Migliorini et al., 2013
rs10965212	*CDKN2B-AS1*	Intron variant	9p21.3	Xu et al., 2013
rs10965215	*CDKN2B-AS1*	intron variant nc transcript variant	9p21.3	Xu et al., 2013
rs662463	*CDKN2B-AS1*	Intron variant	9p21.3	Xu et al., 2013
rs10965219	*CDKN2B-AS1*	Intron variant	9p21.3	Xu et al., 2013
rs17756311	*CDKN2B-AS1*	Intron variant	9p21.3	Xu et al., 2013
rs17694572	*CDKN2B-AS1*	Intron variant	9p21.3	Xu et al., 2013
rs10511573			9p23	Sherborne et al., 2010
rs1001919			9q31.2	Sherborne et al., 2010
rs872863	*DENND1A*	Intron variant	9q33.3	Sherborne et al., 2010
	CRB2 (10K)	flanking_3UTR	9q33.3	
	OR2C3 (4K)	flanking_3UTR		
rs563507	*PARD3*	Intron variant	10p11.21	Treviño et al., 2009
rs10828317				
has merged into rs2230469	*PIP4K2A*	Missense	10p12.2	Migliorini et al., 2013
rs3824662	*GATA3*	Intron variant	10p14	Migliorini et al., 2013
rs569421	*GATA3*	Intron variant	10p14	Migliorini et al., 2013
rs9746	*GATA3*	UTR variant 3 prime	10p14	Migliorini et al., 2013
rs7090445	*ARID5B*	Intron variant	10q21.2	Migliorini et al., 2013
rs7089424	*ARID5B*	Intron variant	10q21.2	Migliorini et al., 2013; Orsi et al., 2012; Papaemmanuil et al., 2009
rs4506592	*ARID5B*	Intron variant	10q21.2	Migliorini et al., 2013
rs10821938	*ARID5B*	Intron variant	10q21.2	Migliorini et al., 2013; Xu et al., 2013
rs7073837	*ARID5B*	Intron variant	10q21.2	Migliorini et al., 2013; Xu et al., 2013; Orsi 2012
rs7087125	*ARID5B*	Intron variant	10q21.2	Migliorini et al., 2013; Xu et al., 2013
rs4948491	*ARID5B*	Intron variant	10q21.2	Migliorini et al., 2013
rs7894504	*ARID5B*	Intron variant	10q21.2	Migliorini et al., 2013

Annex table 1: All the SNPs reported by GWAS performed by Papaemmanuil et al., Treviño et al., Sherborne et al., Xu et al. and Migliorini et al. In red, SNPs in intergenic regions or in genes that not codify for proteins (continue).

SNP	Gene	Function	Location	Reference
rs10740055	*ARID5B*	Intron variant	10q21.2	Orsi et al., 2012; Papaemmanuil et al., 2009
rs4948488	*ARID5B*	Intron variant	10q21.2	Xu et al., 2013
rs2893881	*ARID5B*	Intron variant	10q21.2	Xu et al., 2013
rs6479778	*ARID5B*	Intron variant	10q21.2	Xu et al., 2013
rs6479779	*ARID5B*	Intron variant	10q21.2	Xu et al., 2013
rs10994982	*ARID5B*	Intron variant	10q21.2	Xu et al, 2013; Treviño et al., 2009
rs10994983	*ARID5B*	Intron variant	10q21.2	Xu et al., 2013
rs7923074	*ARID5B*	Intron variant	10q21.2	Xu et al., 2013
rs10821936	*ARID5B*	Intron variant	10q21.2	Xu et al., 2013; Treviño et al., 2009
rs7896246	*ARID5B*	Intron variant	10q21.2	Xu et al., 2013
rs9415636	*ARID5B*	Intron variant	10q21.2	Xu et al., 2013
rs11188661	*BLNK*	Intron variant	10q24.1	Sherborne et al., 2010
	ZNF518A (38K)	flanking_3UTR	10q24.1	
rs11188664	*BLNK*	Intron variant	10q24.1	Sherborne et al., 2010
	ZNF518A (40K)	flanking_3UTR	10q24.1	
rs7084370			10q25.2	Sherborne et al., 2010
rs6571245			10q26.3	Migliorini et al., 2013
rs3802765	*MAML2*	Intron variant	11q21	Orsi et al., 2012
rs12582396	*DUSP16* (20K)	flanking_5UTR	12p13.2	Sherborne et al., 2010
	CREBL2 (34K)	flanking_5UTR	12p13.2	
rs10849033	*C12orf5* (5K)	flanking_3UTR	12p13.32	Treviño et al., 2009
rs10877094			12q14.1	Sherborne et al., 2010
rs7971479	*CPM*	UTR variant 3 prime	12q15	Sherborne et al., 2010
	MDM2 (8K)	flanking_3UTR	12q15	
rs2089222	*MAP1LC3B2*	Intron variant	12q24.22	Treviño et al., 2009
	KRTHB5		12q24.22	
rs7984659	*PCDH20* (650K)	flanking_5UTR	13q21.31	Orsi et al., 2012
rs7317221			13q21.33	Sherborne et al., 2010
rs1832050	*LOC105370342*	Intron variant	13q33.2	Sherborne et al., 2010
rs2239633	*CEBPE*	Missense Upstream variant 2KB	14q11.2	Migliorini et al., 2013; Papaemmanuil et al., 2009
rs12434881	*CEBPE*	Intron variant Utr variant 5 prime	14q11.2	Migliorini et al., 2013
rs761874	*CEBPE* (5K) *SLC7A8*	nc transcript variant Utr variant 3 prime	14q11.2	Migliorini et al., 2013
rs7157021	*SLC7A8*	nc transcript variant Synonymous codon	14q11.2	Migliorini et al., 2013
	CEBPE (10K)	flanking_5UTR	14q11.2	
rs10143875	*CEBPE* (2K)	flanking_3UTR	14q11.2	Migliorini et al., 2013
rs4982731	*CEBPE* (1K)	flanking_3UTR	14q11.2	Migliorini et al., 2013; Xu et al., 2013
rs4982729	*CEBPE* (10K)	flanking_3UTR	14q11.2	Xu et al., 2013
	LOC100128908	upstream variant 2KB flanking_5UTR	14q11.2	
rs12887958	*CEBPE* (4K)	flanking_3UTR	14q11.2	Xu et al., 2013

Annex table 1: All the SNPs reported by GWAS performed by Papaemmanuil et al., Treviño et al., Sherborne et al., Xu et al. and Migliorini et al. In red, SNPs in intergenic regions or in genes that not codify for proteins (continue).

SNP	Gene	Function	Location	Reference
rs8015478	*CEBPE* (500bp)	downstream variant 500B flanking_3UTR	14q11.2	Xu et al., 2013
rs17794251	*CEBPE* (4K)	flanking_5UTR	14q11.2	Xu et al., 2013
rs1394759	*FLJ395 FLJ39531* (*C15orf54*) (10K)	flanking_5UTR	15q14	Orsi et al., 2012
rs6151562	*BNIP2*	Intron variant	15q22.2	Orsi et al., 2012
rs5021303	*LINGO1*	Intron variant	15q24.3	Sherborne et al., 2010
rs984999	*LOC105371010*	Intron variant	15q26.3	Migliorini et al., 2013
rs11647078	*IRX3* (175)	flanking_5UTR	16q12.2	Orsi et al., 2012
rs13331075	*CDYL2*	Intron variant	16q23.2	Orsi et al., 2012
rs1366754	*COTL1* (5K)	flanking_3UTR	16q24.1	Orsi et al., 2012
rs7224669	*RGS9* (150K)	flanking_3UTR	17q24.1	Orsi et al., 2012
rs1879352			18p11.32	Treviño et a.l, 2009
rs8088707	*PMAIP1* (35K)	flanking_3UTR	18q21.32	Orsi et al., 2012
	LOC105372151	Intron variant	18q21.32	
rs567379			18q22.2	Sherborne et al., 2010
rs2191566	*ZNF230*	Intron variant	19q13.31	Treviño et al., 2009
rs6509133	*ZNF230* (4K)	flanking_3UTR	19q13.31	Treviño et al., 2009
rs2284378	*RALY*	Intron variant	20q11.22	Sherborne et al., 2010
	EIF2S2 (90K)	flanking_3UTR	20q11.22	
rs4911414	*RALY* (50K)	flanking_3UTR	20q11.22	Sherborne et al., 2010
	EIF2S2 (30K)	flanking_5UTR	20q11.22	
rs2903908	*NCOA5*	Intron variant	20q13.12	Sherborne et al., 2010
	SLC12A5 (5K)	flanking_3UTR	20q13.12	
rs6027571			20q13.33	Sherborne et al., 2010
rs2822553	*ABCC13*	Intron variant	21q11.2	Orsi et al., 2012
rs9613221	*CRYBB1* (8K)	flanking_3UTR	22q12.1	Sherborne et al., 2010
	CRYBA4 (17K)	flanking_5UTR	22q12.1	
	TPST2 (700bp)	upstream variant 2KB flanking_5UTR	22q12.1	

HYPOTHESIS AND OBJECTIVES

HYPOTHESIS

The early disease onset of pediatric B-ALL and the high risk of the disease associated with some congenital genetic disorders, suggest a strong genetic component in its origin. Supporting this idea, several GWAS studies have identified genetic variants at five genes associated with B-ALL susceptibility. When we analyzed in deep these GWAS, we also found genetic variants in non coding region associated with B-ALL risk. Interestingly, miRNAs, non-coding RNAs, are deregulated in B-ALL, and some genetic variants in miRNA genes have been associated with the susceptibility to cancer.

Considering all these evidences, we propose that common genetic variants in coding genes and in non coding genes are involved in B-ALL susceptibility in our Spanish population. The detection of these variants could help to improve the knowledge of this disease.

OBJECTIVES

The main goal of the present study was to prove the strong genetic component in the etiology of childhood B-ALL by identifying genetic susceptibility markers in coding genes and in non coding genes in our Spanish population.

In order to achieve this goal, we proposed the following specific objectives:

A. To validate the association of genetic variants in B-ALL risk previously proposed in our Spanish population.
    a) We performed a case-control study in 264 children with B-ALL analyzing the most significant SNPs at *ARID5B*, *IKZF1*, *CEBPE*, *CDKN2A/B* and *PIP4K2A*, previously proposed.

B. To determine the association of genetic variants in B-ALL risk in miRNA related genes.
    a) We performed a case-control study in 264 children with B-ALL analyzing 72 SNPs in 21 miRNA processing genes in our Spanish population.
    b) We performed a case-control study in 343 children with B-ALL of two cohorts of Spanish and Slovenian patients analyzing 213 SNPs in 203 miRNAs.

MATERIAL & METHODS

1. POPULATION OF THE STUDY

This study included a total of 343 pediatric patients all diagnosed with B-ALL and 815 healthy controls. The samples of B-ALL patients were collected from 2000 to 2011 at the Pediatric Oncology Units of 5 Spanish reference hospitals (University Hospital Cruces, University Hospital Donostia, University Hospital Miguel Servet, University Hospital Vall D´Hebrón and University Hospital La Paz) (n=264) and the Unit of Pediatric Oncology of Hospital "University Children´s" of Liubliana, Slovenia (n=79). The control group consisted of Spanish unrelated healthy individuals of C.0001171 collection registred in Instituto de Salud Carlos III (ISCIII) (n=719) and Slovenian cancer non-related controls (n=96).

The study was approved by medical ethical committees (PI2014039 and 62/07/03) and informed consent was obtained by parents or guardians and patients (in case they were <12 years) according to the Declaration of Helsinki.

Clinical data were collected objectively by two independent researchers, blinded to genotypes, from the patients' medical files. Data collected included: immunophenotype, molecular alterations, cytogenetics alterations, age at diagnosis and sex (Table 6).

Table 6: Characteristics of the B-ALL patients and controls examined in this study in the Spanish and Slovenian cohort.

	Spanish Cohort		Slovenian Cohort	
	Patients	**Controls**	**Patients**	**Controls**
No. of individuals	264	719	79	96
Mean age ± SE, y	4.1 ± 3.6	52.1 ± 20.5	4.7 ± 5.4	44.5 ± 9.4
Sex*				
Males, n (%)	144 (54.8)	356 (49.5)	41(51.9)	58
Females, n (%)	119 (45.2)	363 (50.5)	38 (48.1)	38
Immunophenotype				
B-type	260 (98.5)	-	78 (98.7)	-
Biphenotipic	4 (1.5)	-	1 (1.3)	-
Genetic alterations[#]				
Hyperdiploid	66 (25)	-	9 (11.4)	-
ETV6-RUNX1	42 (16)	-	12 (15.2)	-
MLL	13 (5)	-	4 (5.1)	-
BCR-ABL	6 (2.3)	-	1 (1.3)	-
E2A-PBX1	6 (2.3)	-	-	-
Hypodiploid	3 (1.1)	-	1 (1.3)	-
Other	2 (0.7)	-	6 (7.6)	-
No alteration	106 (40.1)	-	48 (60.8)	-
No available	26 (9.8)	-	0	-

SE: standard error. *There is no datum of one patient in the Spanish cohort. [#]three of the Spanish patients are both B-hyperdiploid and *ETV6-RUNX1*, one is *BCR-ABL* and B-hyperdiploid, one is *MLL* and B-hyperdiploid and another one is *ETV6-RUNX1* and *MLL*; three Slovenian patients are both B-hyperdiploid and *ETV6-RUNX1*, one is *BCR-ABL* and B-hyperdiploid and another one is *MLL* and B-hyperdiploid.

2. STUDIES OF GENES IDENTIFIED IN GWAS

2.1 *ARID5B* gene

Polymorphisms selection

With the objective of mapping the whole intron 3 of *ARID5B*, the SNPs were selected based on the following criteria: (1) TagSNPs that were defined using Haploview software version 4.2 (http:// www.broadinstitute.org/haploview/haploview) with an r^2 threshold value of 0.8; (2) SNPs predicted to have functional effects (e.g., putative transcription factor-binding sites and CpG sites) according to the bioinformatic analyses (F-SNP(http://compbio.cs.queensu.ca/F-SNP/), Ensembl (http://www.ensembl.org/index.html), and Genome Browser (http://genome.ucsc.edu/)); (3) SNPs previously reported to be associated with ALL susceptibility. Of these, only SNPs with a reported minor allele frequency (MAF)>10% were analyzed, and this included a total of 10 polymorphisms.

Genotype analysis

Genomic DNA was extracted from remission peripheral blood or bone marrow (with<5 % blast cells) using the phenol–chloroform method previously described (Sambrook and Russell 2001) or from saliva samples using Oragene DNA kit (DNA Genotek, Ottawa, Ontario, Canada), according to the manufacturer's instructions.

Genotyping was performed at the General Research Services (SGIker) of the University of the Basque Country using TaqMan Open Array technology (Applied Biosystems), according to the published Applied Biosystems protocol.

TaqMan OpenArray Genotyping Plates contain the selected TaqMan SNP Genotyping Assays pre-loaded and dried down in the through-holes. Each assay contains: a specific fluorescent-dye labeled probe for each allele of the target SNP (the probes contain different fluorescent reporter dyes in 5' to differentiate each allele), a forward primer, a reverse primer and a nonfluorescent quencher (NFQ) at the 3'end of each probe.

During PCR, each probe anneals specifically to its complementary sequence between the forward and reverse primer sites. The DNA polymerase can cleave only probes that hybridize to their specific SNP allele (match). Cleavage separates the reporter dye from the quencher

dye, substantially increasing fluorescence of the reporter dye. Thus, the fluorescence signals generated during PCR amplification indicate the alleles that are present in the sample. A substantial increase in VIC dye fluorescence indicates homozygosity for allele 1, an increase in FAM dye fluorescence indicates homozygosity for allele 2 and both fluorescence signals indicates heterozygosity (Figure 11). A total of 300 ng of DNA were required from each sample to carry out the analysis.

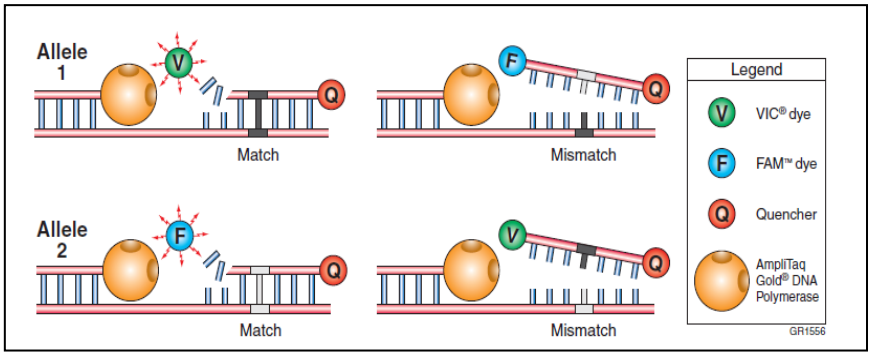

Figure 11: Results from matches and mismatches between target and probe sequences in TaqMan SNP Genotyping Assays (Livak et al. 1995).

Data were analyzed with Taqman Genotyper software for genotype clustering and calling (Figure 12). Duplicate samples were genotyped across the plates. SNPs showing discordant genotypes were excluded from the analysis.

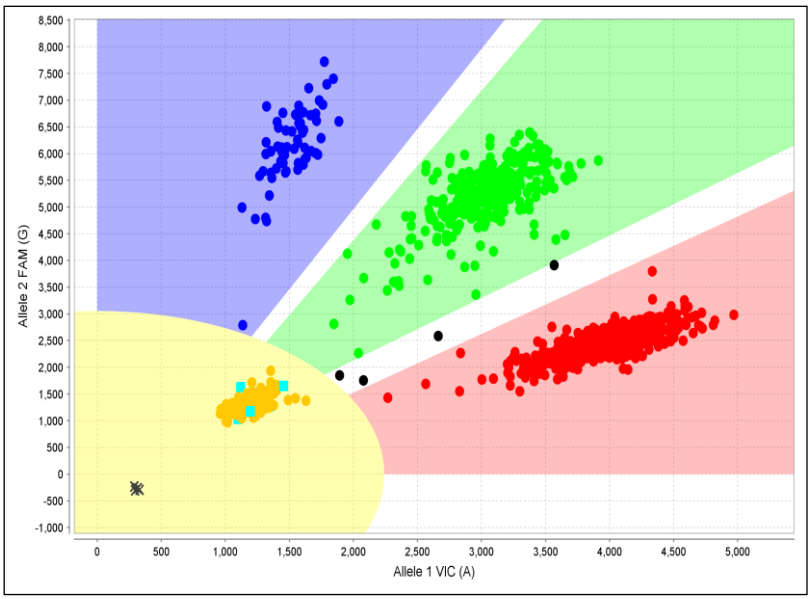

Figure 12: Genotype clustering and calling with Taqman Genotyper software. AA individuals are marked in red, AG in green and GG in dark blue.

Copy number variation (CNV) analysis

Copy number detection was carried out at the Center for Applied Medical Research (CIMA) with the Cytogenetics Whole-Genome 2.7M platform (Affymetrix). This array contains a total number of 2,761,979 copy number probes that enable a high-resolution genome-wide DNA copy number analysis.

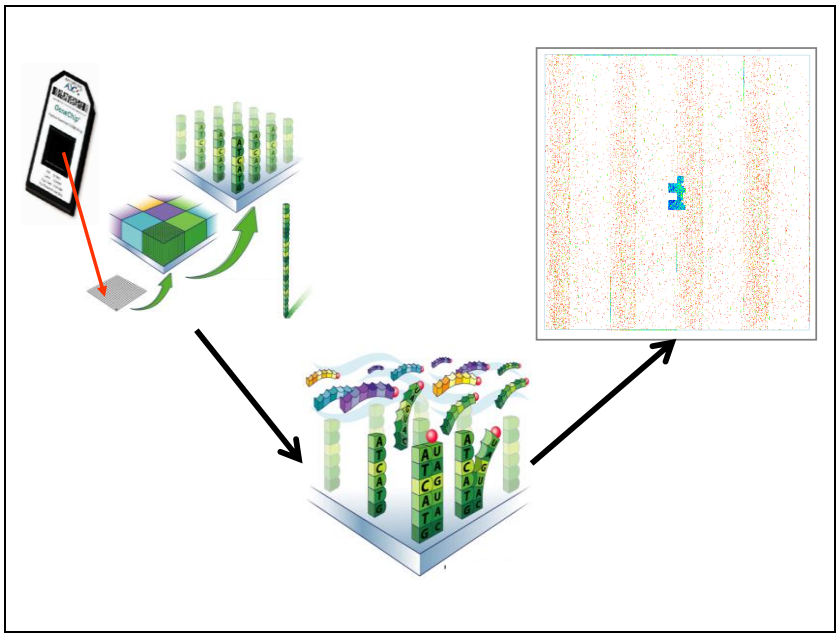

Figure 13: Amplified, fragmented and marked DNA is hybridized into the array that contains probes representing the whole genome. The intensity of each spot is indicative of the number of copies of each region of the genome.

We started with 100ng of genomic DNA adjusted to a concentration of approximately 33ng/µl, which was denatured and amplified. Amplified DNA was purified using magnetic beads and its purity and concentration (above 0.55µg/µl) was validated with the Nanodrop spectrophotometer. Subsequently, the DNA was enzymatically digested to obtain fragments of 50-100bp. We validated the correct fragmentation by electrophoresis on agarose gel and proceeded to the hybridization to the microarray. Washing and staining of the Cyto-array was carried out with the GeneChip Fluidics Station 450 (Affymetrix) and subsequent scanning using the GeneChip Scanner 3000 (Affymetrix), which generated the raw data of the Cyto-array. The intensity of each spot is indicative of the number of copies of each region of the genome (Figure 13). During the protocol, specific reagents provided in the commercial Cyto-array kit (Cytogenetics Reagent Kit, Affymetrix) were used in each stage.

Gene expression

Quantitative real-time PCR (qPCR) was performed using an ABI PRISM 7900HT Fast Real-Time PCR System (Applied Biosystems, Life Technologies, Carlsbad, USA) to detect ARID5B mRNA expression in seven different cell lines, according to the manufacturer's protocol. Briefly, RNA was extracted using Ultraspec (Biotecx, Houston, TX, USA) according to the manufacturer's instructions. Primers and probes for the *ARID5B* gene (Hs01381961_m1) and the TATAbox binding protein gene (*TBP*) [used as a housekeeping control (Hs00427620_m1)] were obtained from Applied Biosystems. Of the two transcript variants previously characterized for *ARID5B*, the primers and probes selected for this study were specific for transcript variant 1. This variant contains exons 2 and 3, which limits the region analyzed (Figure 14b) in order to avoid errors due to the quantification of gene expression for both transcripts.

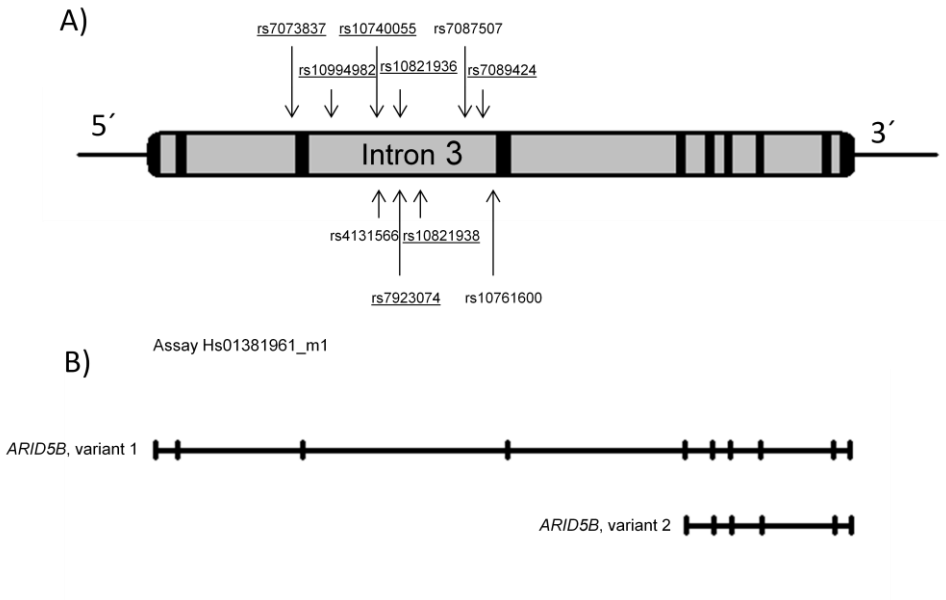

Figure 14: Distribution of the ten selected SNPs analyzed within the *ARID5B* gene. A) SNPs analyzed in other studies are underlined. B) Transcript variants of ARID5B and Hs01381961_m1 were specifically selected to quantify levels of ARID5B variant 1. The vertical markers along the gene indicate exons.

2.2 *IKZF1* GENE

Selection of polymorphisms

The polymorphism rs4132601 in *IKZF1,* previously reported to be highly associated with ALL susceptibility in the literature, was selected.

Genotype analyses

Genomic DNA was extracted from remission peripheral blood or bone marrow using the phenol-chloroform method as previously described (Sambrook and Russell 2001). Genotyping analyses were performed by using PCR followed by restriction analysis. Duplicates were included in each assay. The PCR products were visualized after electrophoresis on 2% agarose gels (Figure 15). Primer sequences and PCR conditions are described in detail in Table 7.

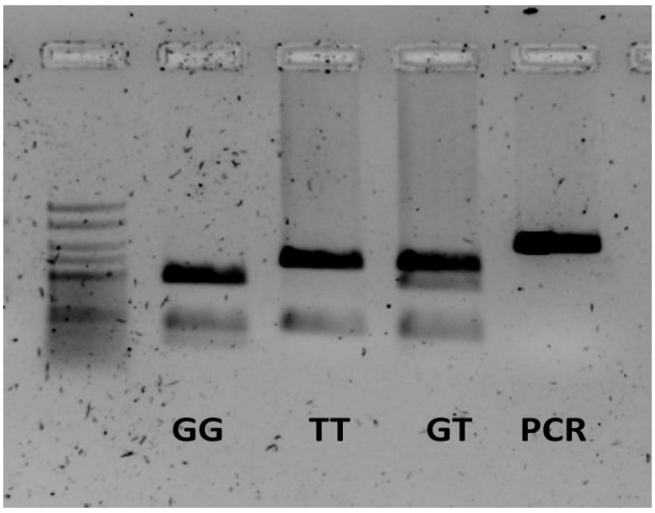

Figure 15: Agarosa gel showing results of rs4132601at *IKZF1*.

Table 7: Primers and PCR conditions for the amplification of rs4132601 in *IKZF1*.

SNP	Primer Sequences (5'- 3')	Genotype method	Restriction Enzyme	Fragment length (bp) according to the genotype
rs4132601	F1: TCTGCTCACAGAAGGGTGTG R1: AGGAAAGGGCAAAGCAGTTT	PCR-RFLP	*MboI*	**GG**:203, 86, 57 **GT**:260, 203, 86, 57 **TT**:260, 86

Systematic review and meta-analysis

Search strategy. We performed an exhaustive search to identify studies that examined the association between the rs4132601 at IKZF1 and ALL susceptibility. We used the keywords and subject terms "(IKZF1 or rs4132601) and acute lymphoblastic leukemia" for Pubmed (www.ncbi.nlm.nih.gov/pubmed) searches for articles published until July 2016. All references cited in the studies were then reviewed to possibly identify additional publications.

Inclusion and exclusion criteria. Original studies that investigated the association between the rs4132601 polymorphism and ALL risk with sufficient data to calculate crude OR values were

included. Reviews, meta-analyses and studies analyzing other regions or variants were excluded.

Data extraction. For each article, we gathered year of publication, first author, country of origin, sample size and genotype frequencies. When it was not possible to extract the genotype data from the article, we contacted the authors to obtain them.

Quality Assessment. The quality of included studies was assessed by scoring according to a "methodological quality assessment scale" (Table 8) (Bilbao-Aldaiturriaga et al. 2016). In the scale, five items, including the representativeness of cases, source of controls, sample size, quality control of genotyping methods and Hardy–Weinberg equilibrium (HWE) were carefully checked. Quality scores ranged from 0 to 10 and a higher score indicated better quality of the study. Scores > 5 were considered acceptable.

Table 8. Scale for methodological quality assessment.

Criteria	**Score**
1.Representativeness of cases	
ALL diagnosed according to acknowledged criteria	2
Mentioned the diagnosed criteria but not specifically described	1
Not described	0
2.Source of controls	
Population or community based	3
Hospital-based ALL-free controls	2
Healthy volunteers without total description	1
ALL-free controls with related diseases	0.5
Not described	0
3.Sample size	
>100	2
25-100	1
<25	0
4.Quality control of genotyping methods	
Repetition of partial/total tested samples with a different method	2
Repetition of partial/total tested samples with the same method	1
Not described	0
5.Hardy-Weinberg equilibrium (HWE)	
Hardy-Weinberg equilibrium in control subjects	1
Hardy-Weinberg disequilibrium in control subjects	0

2.3 *CEBPE* GENE

Selection of polymorphisms

The polymorphisms rs2239633 and rs22396635 in *CEBPE,* previously reported to be highly associated with ALL susceptibility in the literature, were selected.

Genotype analyses

Genomic DNA was extracted from remission peripheral blood or bone marrow using the phenol-chloroform method as previously described (Joseph Sambrook and David W Russell 2001). Genotyping analyses were performed by using amplification-refractory mutation system polymerase chain reaction (ARMS-PCR). Duplicates were included in each assay. The PCR products were visualized after electrophoresis on 2% agarose gels (Figure 16). Primer sequences and PCR conditions are described in detail in Table 9.

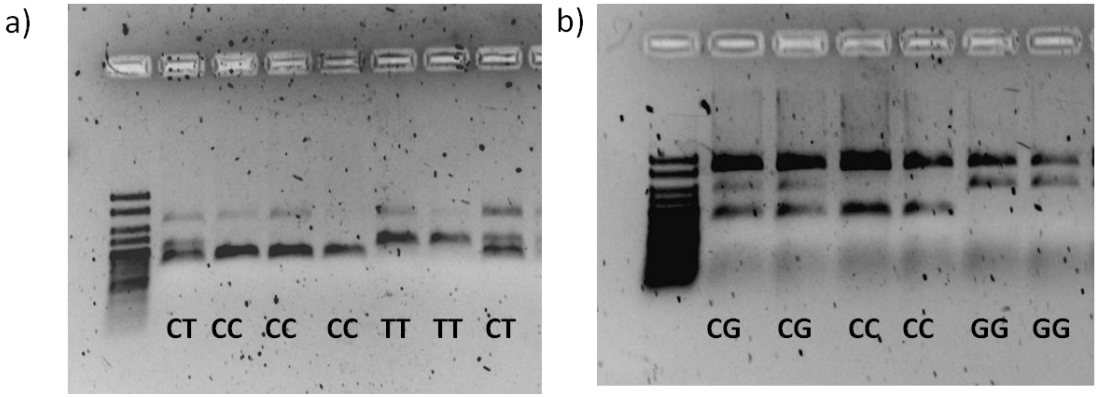

Figure 16: Agarosa gel showing results of a) rs2239633 and b) rs2239635.

Table 9: Primers and PCR conditions for the amplification of rs2239633 and rs22396635 in *CEBPE.*

SNP	Primer Sequences (5'- 3')	Genotype method	Fragment length (bp) according to the genotype
rs2239633	F1: CACCACGCAGGCTCGT R1: CAGTTGGGTCTCACCTCCTC F2: AAGCCCAGGGAGTTAGGAAG R2: GTCCTAGGAACAAGCTCTACACG	ARMS-PCR	**CC**: 159, 325 **CT**: 159, 204, 325 **TT**: 204, 325
rs2239635	F1: GATTTGGAGTCCCCTGGC R1: CCCCAGAGGGAGAGATGTAAG F2: CCCTGGCCTACACTCAGAGA R2: CACTCCCTGCTGGGAGC	ARMS-PCR	**CC**: 167, 419 **CG**: 167, 286, 419 **GG**: 286, 419

Systematic review and meta-analysis

Search strategy. We performed an exhaustive search to identify studies that examined the association between rs2239633 polymorphism of *CEBPE* and ALL susceptibility. We used the keywords and subject terms "(CEBPE or rs2239633 or 14q11.2) and leukemia" for Pubmed (www.ncbi.nlm.nih.gov/pubmed) searches for articles published until July 2016. All references cited in the studies were then reviewed to possibly identify additional publications.

Inclusion and exclusion criteria, Data extraction and *Quality Assessment* were done as previously described in section 2.2.

2.4 *CDKN2A/B* locus

Selection of polymorphisms

A total of six SNPs at the locus 9p21.3 were selected (Table 10). Selection was done based on the following criteria: (i) four SNPs previously reported to be highly associated with ALL susceptibility in the literature or in high LD defined using the International HapMap Project (release #24; http://hapmap.ncbi.nlm.nih.gov/) (The HapMap Data Coordination Center (DCC), Bethesda, MD) and Haploview software v.4.2 (http://www.broad.mit.edu/mpg/haploview/) (Broad Institute, Cambridge, USA) with an r^2 threshold of 0.8 and a MAF of 10%, (ii) SNPs in miRNA binding sites at 3´UTR of *CDKN2A* and *CDKN2B* with a MAF>10% identified using bioinformatics tools: Ensembl (http://www.ensembl.org/) (Welcome Trust Genome Campus, Cambridge, UK), and miRNASNP (http://bioinfo.life.hust.edu.cn/miRNASNP2/index.php) (College of Life Science and Technology, HUST). Out of 47 SNPs identified in the 3´UTR that disrupt or create a miRNA binding site (Table 11), only two had a MAF>10%.

Table 10: SNPs selected and selection criteria.

SNP	Gene	Alleles	Reason for selection
rs3731222	*CDKN2A*	A>G	In LD with rs3731217 (Sherborne et al. 2010)
rs2811709	*CDKN2A*	G>A	Bibliography (Orsi et al. 2012)
rs2811712	*ANRIL*	A>G	In LD with rs662463 and rs17756311 (Xu et al. 2013, Hungate et al. 2016)
rs3731249	*CDKN2A*	C>T	Bibliography (Xu et al. 2015, Walsh et al. 2015, Vijayakrishnan et al. 2015)
rs1063192	*CDKN2B, ANRIL*	T>C	3´UTR miRNA binding site
rs3217992	*CDKN2B, ANRIL*	G>A	3´UTR miRNA binding site

Table 11: SNPs identified in 3´UTR of *CDKN2A/B CDKN2B.*

CDKN2A			***CDKN2B***		
SNP	**Alleles**	**MAF**	**SNP**	**Alleles**	**MAF**
rs111532782	G>C	0.004	**rs1063192***	T>C	0.45
rs113798404	-	-	rs140726127	A>G	-
rs121913388	C>G	0.000008	rs148421170	C>T	0.002
rs142371511	G>C	-	rs148786939	T>C	-
rs145697272	-	-	rs150924737	T>A	0.001
rs182558871	C>A	0.001	rs181736450	C>A	-
rs190538376	G>A	-	rs185130567	T>C	0.005
rs200429615	T>G	0.001	rs187514719	A>G	0.001
rs201314211	T>G	-	rs200344272	G>A	-
rs34886500	C>T	0.0001	rs2285329	T>C	-
rs34968276	C>A	0.0001	rs3217988	G>A	-
rs36204273	G>A	0.0001	rs62637622	G>A	-
rs3731249	G>A	0.032	rs111751296	-	-
rs3731253	C>G	-	rs140430251	-	-
rs3731255	C>G	-	rs142570894	C>T	0.0003
rs4987127	G>A	-	rs144131923	-	-
rs6413463	T>A	-	rs183610933	T>C	-
rs11552822	-	-	rs187657501	T>C	-
rs121913383	-	-	rs3217983	C>T	-
rs121913385	-	-	rs3217990	C>A	0.002
rs137854597	G>A	0.000008	**rs3217992***	G>A	0.4
rs137854599	-	-			
rs181022755	-	-			
rs200863613	C>A	-			
rs3088440	G>A	0.078			
rs45476696	-	-			

SNPs with a MAF>10% are in bold and marked with *.

Genotype analyses

Genomic DNA was extracted from remission peripheral blood or bone marrow using the phenol-chloroform method as previously described (Sambrook and Russell 2001).

Genotyping was performed at the Spanish National Genotyping Center (CeGen) using the GoldenGate Genotyping Assay with Illumina Bead Array System (Illumina Inc., San Diego; USA). In this approach during the liquid phase, allele specific oligos (ASO) are hybridized to genomic DNA, extended and ligated to a locus specific oligo (LSO). PCR is performed using universal primers. The multiplexed products are hybridized to a universal Sentrix Array for detection and analysis. A schematic view of the principle of the assay is shown in Figure 17.

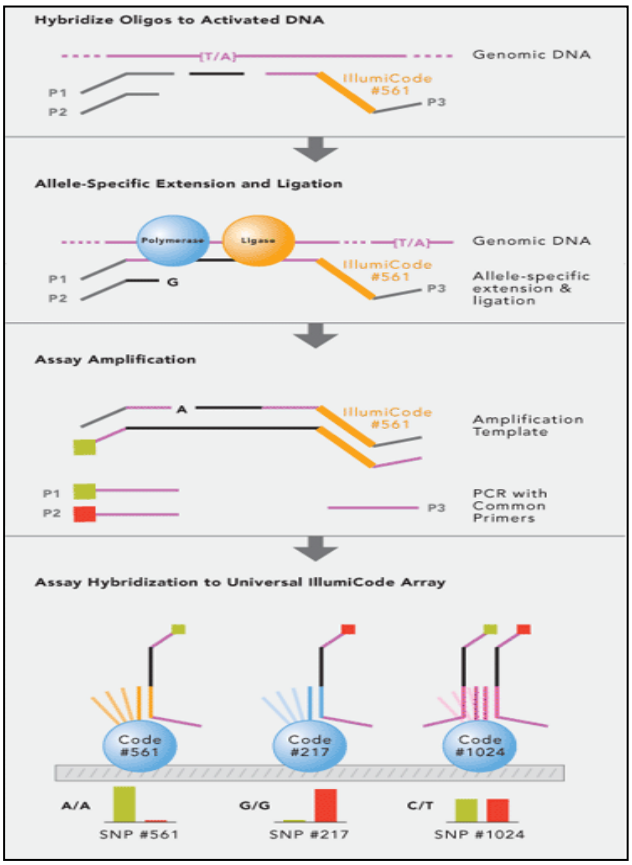

Figure 17: Goldengate assay overview.

Each reaction required a total of 400 ng of DNA. The DNA was re-quantified at the Spanish Genotyping Center using PicoGreen technique (Invitrogen Corp., Carlsbad, CA) and diluted to a final concentration of 50 ng/µl. With this technique, the concentration of DNA is determined by means of a fluorescent dye that binds to double stranded DNA (PicoGreen®, Molecular Probes), which is then quantified with a fluorometer.

Data were analyzed with GenomeStudio software for genotype clustering and calling. Duplicate samples and CEPH trios (Coriell Cell Repository, Camden, NJ) were genotyped across the plates. SNPs showing Mendelian allele-transmission errors or showing discordant genotypes were excluded from the analysis.

For rs3731249, the genotyping analyses were performed by using PCR followed by restriction analysis with BstUI enzyme. Duplicates were included in each assay. The PCR products were visualized after electrophoresis on 3% agarose gels (Figure 18). Primer sequences and PCR conditions are described in detail in Table 12.

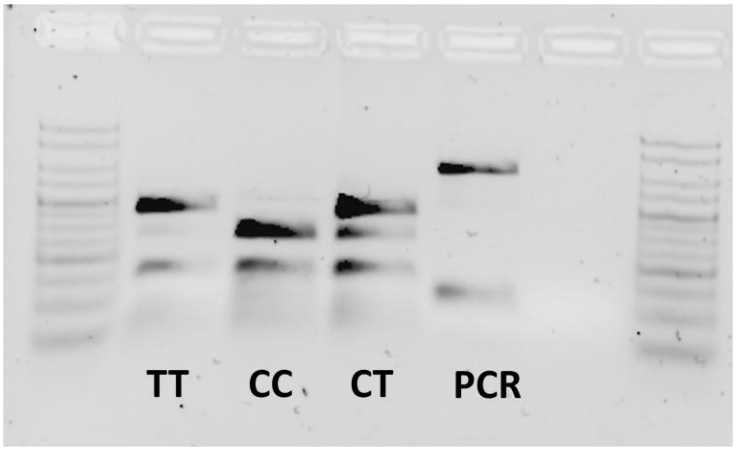

Figure 18: Agarose gel showing results for rs3731249 at *CDKN2A.*

.

Table 12: Primers and PCR conditions for the amplification of rs3731249 in CDKN2A.

SNP	Primer Sequences (5'- 3')	Genotype method	Restriction Enzyme	Fragment length (bp) according to the genotype
rs3731249	F1: TGGACCTGGCTGAGGAG R1:TCGGGATTATTTCCCATTTG	PCR-RFLP	*BstUI*	**CC**:133, 73, 48, 29,21,2 **CT**:181, 133, 73, 48, 29, 21, 2 **TT**:181, 73, 48, 29, 21, 2

2.5 *PIPK42A* GENE

Polymorphism selection

We decided to select rs7088318 as a tagSNP of the association hotspot at *PIP4K2A* because the four SNPs reported in the article of Xu et al. (Xu et al. 2013) were in linkage disequilibrium with an r^2 greater than 0.8. We made a linkage disequilibrium analysis using Hapmap database and Haploview version 4.2 software (Daly lab at the Broad Institute, Cambridge, MA, USA).

Genotyping

Genotyping analyses were performed by ARMS-PCR. Duplicates were included in each assay. The PCR products were visualized after electrophoresis on 2% agarose gels (Figure 19). Primer sequences and PCR conditions are described in detail in Table 13.

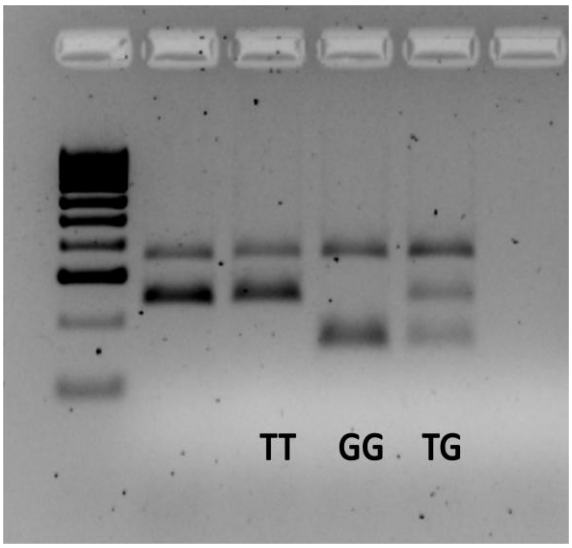

Figura 19: Agarose gel showing results for rs7088318 at *PIP4K2A*.

Table 13: Primers used to genotype rs7088318 at *PIP4K2A*.

SNP	Primer Sequences (5'- 3')	Genotype method	Fragment length (bp) according to the genotype
rs7088318	F1:CCTATGAAAAGAGAATAATAGAATTTGTTTG R1: CAACCAGTGCCAGTTGTGAC F2: ACAGAGAGGAAAGCCACACG R2: CCGAGAAAGATGAGTTCTCGGTA	ARMS-PCR	**TT**: 248, 348 **TG**: 248, 163, 348 **GG**: 163, 348

3. STUDIES OF NEW GENETIC VARIATIONS

3.1 SNPS IN GENES OF PROCESSING MACHINERY

Selection of Genes and Polymorphisms

Twenty-one genes involved in miRNA biogenesis and processing, as determined based on available literature and the Patrocles database (http:// www.patrocles.org/; University of Liege, Liège, Belgium), were selected for the analysis (Table 14). For each gene, all of the SNPs with potential functional effects were examined using F-SNP (http:// compbio.cs.queensu.ca/F-SNP/; Queen's University, Kingston, Canada), Fast-SNP (http://fastsnp.ibms.sinica.edu.tw; Academia Sinica, Taipei, Taiwan), polymirTS (http://compbio.uthsc.edu/ miRSNP/; University of Tennessee Health Science Center, Memphis, TN), and Patrocles databases. Functional effects were considered to be those that resulted in amino acid changes and/or alternate splicing, those that were located in the

promoter region of putative TFB sites, or those that disrupted or created miRNAs binding interactions. SNPs previously included in association studies were also examined. The final selection of SNPs was made based on those having a MAF reater than 5% (i.e., ≥0.05) in European/Caucasian populations (Table 15). A total of 72 SNPs in 21 processing genes were included in the analysis.

Table 14: Genes involved in processing and biogénesis of miRNAs.

RISC complex	GENIM complex	*GENIM3*
		GENIM4
		GENIM5
	AGO	*EIF2C1*
		EIF2C2
		HIWI
	CCR-NOT complex	*CNOT1*
		CNOT2
		CNOT3
		CNOT4
		CNOT5
		CNOT6
	GW182	*TNRC6A*
		TNRC6B
	SND1	*SND1*
DROSHA/DGR8	DGCR8	*DGCR8*
	DROSHA	*DROSHA*
DICER complex	XPO5	*XPO5*
	RAN	*RAN*
	DICER	*DICER*
	TRBP	*TRBP*

Table 15: SNPs identified in microRNA processing genes and their selection criteria.

Gene	SNP	Alleles	Chr	Location	Function	Reason for selection
CNOT1	rs11644694	G > A	16	58557342	Non-synonymous	NS, SR
	rs11866002	C > T	16	58587737	Synonymous	SR
	rs37060	C > T	16	58566304	Intronic	SR
CNOT2	rs10506586	C > A	12	70715490	Non-synonymous	NS, SR
CNOT3	rs42318	G > A	19	54657069	Non-synonymous	NS
CNOT4	rs1003226	T > C	7	135046552	3'UTR	SR
	rs3763425	C > T	7	135195320	Upstream	UR
	rs3812265	C > T	7	135048804	Non-synonymous	NS, SR
CNOT6	rs11738060	T > A	5	180004154	3'UTR	MIRTS
	rs6877400	T > C	5	179996111	Synonymous	SR
DGCR8	rs1640299	T > G	22	20098359	3'UTR	BIB
	rs35987994	T > C	22	20074006	Non-synonymous	NS
	rs3757	G > A	22	20099331	3'UTR	MIRTS, BIB
	rs417309	G > A	22	20098544	3'UTR	3UTR, BIB
	rs9606248	A > G	22	20087539	Intronic	BIB
DICER1	rs1057035	T > C	14	95554142	3'UTR	MIRTS
	rs1209904	C > T	14	95563712	Intronic	BIB
	rs13078	T > A	14	95556747	3'UTR	3UTR, BIB
	rs3742330	A > G	14	95553362	3'UTR	BIB
DROSHA	rs10035440	T > C	5	31539463	Upstream	BIB
	rs10719	C > T	5	31401447	3'UTR	3UTR, BIB
	rs17408716	A > G	5	31467952	Intronic	BIB
	rs2287584	T > C	5	31423007	Synonymous	SR, BIB
	rs3792830	T > C	5	31416248	Intronic	BIB
	rs3805500	T > C	5	31462977	Intronic	BIB
	rs4867329	A > C	5	31435627	Intronic	BIB
	rs493760	T > C	5	31437040	Intronic	BIB
	rs55656741	G > A	5	31515657	Non-synonymous	NS, SR
	rs639174	C > T	5	31433647	Intronic	BIB
	rs6877842	G > C	5	31532638	Upstream	BIB
	rs6884823	G > A	5	31491121	Intronic	BIB
	rs7719666	C > T	5	31520778	Intronic	BIB
	rs7735863	G > A	5	31486540	Intronic	BIB
EIF2C1	rs595961	A > G	1	36367780	Intronic	BIB
	rs636832	G > A	1	36363475	Intronic	BIB
EIF2C2	rs2292778	C > T	1	141568622	Synonymous	SR
	rs2293939	G > A	1	141551407	Synonymous	SR
	rs4961280	C > A	1	141647414	Upstream	UR, BIB
DDX20	rs197388	T > A	1	112297482	Upstream	UR, BIB
	rs197412	T > C	1	112308953	Non-synonymous	NS, BIB
	rs197414	C > A	1	112309123	Non-synonymous	NS, BIB
	rs563002	T > C	1	112317135	Downstream	BIB
GEMIN4	rs1062923	T > C	17	649067	Non-synonymous	NS, BIB
	rs2740348	G > C	17	649935	Non-synonymous	NS, BIB
	rs34610323	C > T	17	648546	Non-synonymous	NS
	rs3744741	C > T	17	649232	Non-synonymous	NS, BIB
	rs7813	C > T	17	648186	Non-synonymous	NS, BIB
	rs910924	C > T	17	655920	5'UTR	5UTR, BIB
GEMIN5	rs1974777	A > G	5	154291409	Non-synonymous	NS
	rs6865950	G > A	5	154275786	Non-synonymous	NS
	rs816736	T > C	5	154271948	Synonymous	SR
PIWIL1	rs1106042	G > A	12	130841638	Non-synonymous	NS, SR, BIB
RAN	rs11061209	G > A	12	131364988	Downstream	BIB
	rs14035	C > T	12	131361241	3'UTR	MIRTS, BIB
SMAD5	rs3764941	A > C	5	135469527	Non-synonymous	NS, SR
	rs3764942	G > A	5	135469500	Intronic	SR
SND1	rs17151639	A > G	7	127637816	Non-synonymous	NS
	rs17676986	C > T	7	127636958	Intronic	TR
	rs322825	C > T	7	127721507	Synonymous	SR
	rs3823994	T > A	7	127669857	Intronic	SR
TNRC6A	rs6497759	G > A	16	24801737	Non-synonymous	NS
TNRC6B	rs139919	T > C	22	40726183	3'UTR	MIRTS
	rs2413621	T > C	22	40673999	Intronic	SR
	rs470113	A > G	22	40729614	3'UTR	MIRTS
	rs4821943	A > G	22	40722745	3'UTR	MIRTS
	rs9611280	G > A	22	40552119	Non-synonymous	NS, SR
TARBP2P	rs784567	C > T	12	53894465	Upstream	BIB
XPO5	rs1106841	A > C	6	43496662	Synonymous	SR
	rs2227301	G > A	6	43485283	Downstream	BIB

Table 15: SNPs identified in microRNA processing genes and their selection criteria (continuation).

Gene	SNP	Alleles	Chr	Location	Function	Reason for selection
	rs2257082	C > T	6	43492578	Synonymous	SR, BIB
	rs34324334	C > T	6	43535018	Non-synonymous	NS, SR
	Rs7755135	C > T	6	43490809	3'UTR	MIRTS

Chr: chromosome; 3UTR: 3'UTR regulation; 5UTR: 5'UTR regulation; BIB: Bibliographic; MIRTS: miRNA target site; NS: Non-synonymous; SR: Splicing regulation; UR: Upstream regulation.

3.2 SNPS IN MICRORNA GENES

Selection of Genes and Polymorphisms

We decided to include all miRNAs with SNPs, due to the fact that they can regulate a wide range of genes that are not completely defined. Therefore, any miRNA could be implicated in the regulation of genes affecting ALL risk. To this purpose, we performed two studies.

In the first study, we selected all the known SNPs in miRNAs at the moment of the selection (Jun 2011) with a MAF>0.01 in European/Caucasoid populations, using Patrocles and Ensembl (http://www.ensembl.org/; Welcome Trust Genome Campus, Cambridge, UK), miRNA SNiPer (http://www.integratomics-time.com/miRNA-SNiPer/) databases and literature review. A total of 46 SNPs present in 42 pre-miRNA genes were selected (Table 16).

Table 16: Characteristics of the SNPs located in microRNAs that were analyzed.

Gene	SNP	Alleles	Chr	Location	Position
mir-106b	rs72631827	G > T	7	99691652	Premature
mir-1178	rs7311975	T > C	12	120151493	Premature
mir-1206	rs2114358	T > C	8	129021179	Premature
mir-1255b-1	rs6841938	G > A	4	36428048	Mature
mir-1265	rs11259096	T > C	10	14478618	Premature
mir-1269	rs73239138	G > A	4	67142620	Mature
mir-1274a	rs318039	C > T	5	41475766	Premature
mir-1282	rs11269	G > T	15	44085909	Premature
mir-1294	rs13186787	A > G	5	153726769	Premature
mir-1302-4	rs10173558	T > C	2	208133995	PremiRNA Flanking region
mir-1307	rs7911488	A > G	10	105154089	Premature
mir-146a	rs2910164	G > C	5	159912418	Premature
mir-149	rs2292832	C > T	2	241395503	Premature
mir-154	rs41286570	G > A	14	101526127	Mature
mir-16-1	rs72631826	T > C	13	50623143	Premature
mir-1908	rs174561	T > C	11	61582708	Premature
mir-196a-2	rs11614913	C > T	12	54385599	Premature
mir-2053	rs10505168	A > G	8	113655752	Premature
mir-2110	rs17091403	C > T	10	115933905	Premature
mir-216a	rs41291179	A > T	2	56216090	Premature
mir-220a	rs72631817	T > C	X	122696014	Premature
mir-222	rs72631825	G > A	X	45606471	Premature
mir-27a	rs895819	T > C	19	13947292	Premature
mir-300	rs12894467	C > T	14	101507727	Premature
mir-423	rs6505162	A > C	17	28444183	Premature
mir-449b	rs10061133	A > G	5	54466544	Mature
mir-453	rs56103835	T > C	14	101522556	Premature
mir-492	rs2289030	C > G	12	95228286	Premature
mir-499	rs3746444	T > C	20	33578251	Seed
mir-548a-1	rs12197631	T > G	6	18572056	Premature
mir-548h-3	rs9913045	G > A	12	13446924	Mature
mir-548h-4	rs73235382	A > T	8	26906437	Premature
mir-577	rs34115976	C > G	4	115577997	Premature
mir-585	rs62376934	G > A	5	168690612	Premature
mir-595	rs4909237	C > T	7	158325503	Premature
mir-603	rs11014002	C > T	10	24564653	Premature
mir-604	rs2368392	C > T	10	29834003	Premature
mir-604	rs2368393	T > C	10	29833998	Premature
mir-605	rs2043556	A > G	10	53059406	Premature
mir-608	rs4919510	C > G	10	102734778	Mature
mir-612	rs12803915	G > A	11	65211979	Premature
mir-612	rs550894	G > T	11	65211940	Premature
mir-618	rs2682818	C > A	12	81329536	Premature
mir-624	rs11156654	T > A	14	31483955	Premature Flanking region
mir-656	rs58834075	C > T	14	101533093	Premature
mir-943	rs1077020	T > C	4	1988193	Premature

Chr: chromosome; Premature: precursor miRNA molecule of 70-100 nucleotides (pre-miRNA); Mature: mature miRNA; Seed: Nucleotides in the position 2-8 of the miRNA that bind mRNA target; Flanking region: Sequence 1 kb upstream and 1kb downstream.

Taking into account that after our previous study the number of annotated miRNAs increased substantially (Kozomara and Griffiths-Jones 2014), we performed a second study in which following the same criteria, a total of 213 SNPs in 2016 miRNAs were selected (Table 17) at the moment of the selection (May 2014).

Table 17: SNPS selected in miRNA genes.

	Gene	SNP	Alelles	Chromosome	Location
1	hsa-mir-449b	rs10061133	A>G	5	54466544
2	hsa-mir-1302-4	rs10173558	T>C	2	208133995
3	hsa-mir-5196	rs10406069	G>A	19	35836530
4	hsa-mir-4745	rs10422347	C>T	19	804959
5	hsa-mir-548ae-2	rs10461441	T>T	5	57825920
6	hsa-mir-2053	rs10505168	A>G	8	113655752
7	hsa-mir-4700	rs1055070	T>G	12	121161048
8	hsa-mir-943	rs1077020	T>T	4	1988193
9	hsa-mir-6074	rs10878362	T>T	12	66417493
10	hsa-mir-544b	rs10934682	T>G	3	124451312
11	hsa-mir-603	rs11014002	T>T	10	24564653
12	hsa-mir-1343	rs11032942	T>T	11	34963459
13	hsa-mir-624	rs11156654	T>A	14	31483955
14	hsa-mir-5579	rs11237828	T>T	11	79133220
15	hsa-mir-1265	rs11259096	T>C	10	14478618
16	hsa-mir-196a-2	rs11614913	C>T	12	54385599
17	hsa-mir-548at	rs11651671	T>T	17	40646803
18	hsa-mir-5092	rs11713052	C>G	3	124870376
19	hsa-mir-4792	rs11714172	T>G	3	24562877
20	hsa-mir-3192	rs11907020	T>C	20	18451325
21	hsa-mir-4653	rs11983381	A>G	7	100802786
22	hsa-mir-548a-1	rs12197631	T>T	6	18572056
23	hsa-mir-202	rs12355840	T>C	10	135061112
24	hsa-mir-3117	rs12402181	G>A	1	67094171
25	hsa-mir-1269b	rs12451747	T>T	17	12820632
26	hsa-mir-4744	rs12456845	T>C	18	46576058
27	hsa-mir-4433	rs12473206	T>T	2	64567916
28	hsa-mir-4274	rs12512664	A>G	4	7461769
29	hsa-mir-4277	rs12523324	T>T	5	1708983
30	hsa-mir-4293	rs12780876	T>A	10	14425204
31	hsa-mir-612	rs12803915	G>A	11	65211979
32	hsa-mir-4309	rs12879262	G>C	14	103006047
33	hsa-mir-300	rs12894467	C>T	14	101507727
34	hsa-mir-1294	rs13186787	T>T	5	153726769
35	hsa-mir-3152	rs13299349	G>A	9	18573360
36	hsa-mir-548ac	rs1414273	T>T	1	117102649
37	hsa-mir-3175	rs1439619	A>C	15	93447631
38	hsa-mir-5007	rs1572687	C>T	13	55748673
39	hsa-mir-3612	rs1683709	C>T	12	128778703
40	hsa-mir-5700	rs17022749	T>T	12	94955603
41	hsa-mir-2110	rs17091403	C>T	10	115933905
42	hsa-mir-4422	rs17111728	T>C	1	55691384
43	hsa-mir-1908	rs174561	T>C	11	61582708
44	hsa-mir-3143	rs17737028	A>G	6	27115467

Table 17: SNPS selected in miRNA genes (continuation).

	Gene	SNP	Alelles	Chromosome	Location
45	hsa-mir-633	rs17759989	A>G	17	61021611
46	hsa-mir-3652	rs17797090	G>A	12	104324266
47	hsa-mir-4733	rs17885221	C>T	17	29421443
48	hsa-mir-5197	rs2042253	A>G	5	143059433
49	hsa-mir-605	rs2043556	A>G	10	53059406
50	hsa-mir-4511	rs2060455	T>T	15	66011630
51	hsa-mir-3620	rs2070960	C>T	1	228284991
52	hsa-mir-1206	rs2114358	T>C	8	129021179
53	hsa-mir-4494	rs215383	G>A	12	47758032
54	hsa-mir-3130-1	rs2241347	T>T	2	207647981
55	hsa-mir-4707	rs2273626	C>A	14	23426182
56	hsa-mir-492	rs2289030	C>G	12	95228286
57	hsa-mir-1229	rs2291418	C>T	5	179225324
58	hsa-mir-564	rs2292181	G>C	3	44903434
59	hsa-mir-149	rs2292832	T>T	2	241395503
60	hsa-mir-604	rs2368392	C>T	10	29834003
61	hsa-mir-4432	rs243080	C>T	2	60614572
62	hsa-mir-4636	rs257095	A>G	5	9053945
63	hsa-mir-1208	rs2648841	C>A	8	129162433
64	hsa-mir-3183	rs2663345	T>T	17	925764
65	hsa-mir-4804	rs266435	C>G	5	72174432
66	hsa-mir-6128	rs2682818	C>A	12	81329536
67	hsa-mir-4308	rs28477407	C>T	14	55344901
68	hsa-mir-378d-1	rs28645567	G>A	4	5925054
69	hsa-mir-4472-1	rs28655823	G>C	8	143257760
70	hsa-mir-1255a	rs28664200	T>C	4	102251501
71	hsa-mir-146a	rs2910164	G>C	5	159912418
72	hsa-mir-5695	rs2967897	G>G	19	13031210
73	hsa-mir-4803	rs3112399	T>A	5	71465361
74	hsa-mir-577	rs34115976	C>G	4	115577997
75	hsa-mir-4669	rs35196866	T>T	9	137271318
76	hsa-mir-2278	rs356125	G>A	9	97572244
77	hsa-mir-5189	rs35613341	C>G	16	88535407
78	hsa-mir-6076	rs35650931	G>C	14	50433227
79	hsa-mir-449c	rs35770269	A>T	5	54468124
80	hsa-mir-3166	rs35854553	A>T	11	87909673
81	hsa-mir-3936	rs367805	G>A	5	131701279
82	hsa-mir-6499	rs3734050	C>T	5	150901699
83	hsa-mir-499a	rs3746444	T>C	20	33578251
84	hsa-mir-5090	rs3823658	G>A	7	102106201
85	hsa-mir-4751	rs4112253	C>G	19	54786022
86	hsa-mir-96	rs41274239	A>G	7	129414574
87	hsa-mir-187	rs41274312	G>A	18	33484792
88	hsa-mir-154	rs41286570	G>G	14	101526127

Table 17: SNPS selected in miRNA genes (continuation).

	Gene	SNP	Alelles	Chromosome	Location
89	hsa-mir-216a	rs41291179	A>T	2	56216090
90	hsa-mir-122	rs41292412	C>T	18	56118358
91	hsa-mir-3135b	rs4285314	T>T	6	32717702
92	hsa-mir-548ap	rs4414449	T>C	15	86368898
93	hsa-mir-6084	rs45530340	C>C	1	20960230
94	hsa-mir-548ap	rs4577031	A>T	15	86368959
95	hsa-mir-4268	rs4674470	T>C	2	220771223
96	hsa-mir-941-1	rs4809383	C>T	20	62550780
97	hsa-mir-548j	rs4822739	C>G	22	26951185
98	hsa-mir-5680	rs487571	T>T	8	103137693
99	hsa-mir-595	rs4909237	C>T	7	158325503
100	hsa-mir-608	rs4919510	C>G	10	102734778
101	hsa-mir-548al	rs515924	A>G	11	74110353
102	hsa-mir-3671	rs521188	A>G	1	65523519
103	hsa-mir-4424	rs56088671	T>T	1	178646884
104	hsa-mir-323b	rs56103835	T>C	14	101522556
105	hsa-mir-548aw	rs56195815	T>T	9	135821099
106	hsa-mir-5189	rs56292801	G>A	16	88535341
107	hsa-mir-1283-1	rs57111412	T>T	19	54191743
108	hsa-mir-559	rs58450758	T>T	2	47604866
109	hsa-mir-656	rs58834075	C>T	14	101533093
110	hsa-mir-888	rs5965660	T>G	X	145076302
111	hsa-mir-3928	rs5997893	G>A	22	31556103
112	hsa-mir-4762	rs60308683	T>T	22	46156446
113	hsa-mir-4326	rs6062431	G>C	20	61918164
114	hsa-mir-4467	rs60871950	G>A	7	102111936
115	hsa-mir-596	rs61388742	T>C	8	1765425
116	hsa-mir-3922	rs61938575	G>A	12	104985443
117	hsa-mir-412	rs61992671	G>A	14	101531854
118	hsa-mir-4772	rs62154973	C>T	2	103048780
119	hsa-mir-585	rs62376935	C>T	5	168690635
120	hsa-mir-4482	rs641071	T>T	10	106028157
121	hsa-mir-3679	rs6430498	G>A	2	134884700
122	hsa-mir-423	rs6505162	T>T	17	28444183
123	hsa-mir-646	rs6513496	T>C	20	58883534
124	hsa-mir-4731	rs66507245	T>T	17	15154966
125	hsa-mir-3622a	rs66683138	T>T	8	27559214
126	hsa-mir-6128	rs67042258	G>A	11	56511354
127	hsa-mir-3167	rs670637	T>T	11	126858392
128	hsa-mir-4642	rs67182313	A>G	6	44403438
129	hsa-mir-4431	rs6726779	T>C	2	52929680
130	hsa-mir-3910-1, hsa-mir-3910-2	rs67339585	T>T	9	94398581
131	hsa-mir-3135a	rs6787734	T>T	3	20179097

Table 17: SNPS selected in miRNA genes (continuation).

	Gene	SNP	Alelles	Chromosome	Location
132	hsa-mir-4305	rs67976778	T>T	13	40238175
133	hsa-mir-3144	rs68035463	C>A	6	120336327
134	hsa-mir-1255b-1	rs6841938	T>T	4	36428048
135	hsa-mir-3683	rs6977967	A>G	7	7106636
136	hsa-mir-3686	rs6997249	T>T	8	130496365
137	hsa-mir-4427	rs701213	T>T	1	233759918
138	hsa-mir-378h	rs702742	A>G	5	154209024
139	hsa-mir-548aj-2	rs7070684	T>T	10	12172775
140	hsa-mir-1283-2	rs71363366	C>G	19	54261549
141	hsa-mir-140	rs7205289	C>C	16	69967005
142	hsa-mir-2117	rs7207008	T>A	17	41522213
143	hsa-mir-4741	rs7227168	C>T	18	20513374
144	hsa-mir-3188	rs7247237	C>T	19	18392894
145	hsa-mir-3689f	rs72502717	T>T	9	137742597
146	hsa-mir-105-2	rs72631816	T>A	X	151562938
147	hsa-mir-222	rs72631825	G>A	X	45606471
148	hsa-mir-16-1	rs72631826	T>T	13	50623143
149	hsa-mir-106b	rs72631827	G>G	7	99691652
150	hsa-mir-323b	rs72631831	G>G	7	1062656
151	hsa-mir-183	rs72631833	G>G	7	129414804
152	hsa-mir-3972	rs72646786	C>T	1	17604437
153	hsa-mir-3976	rs72855836	G>A	18	5840810
154	hsa-mir-4999	rs72996752	A>G	19	8454236
155	hsa-mir-4459	rs73112689	T>T	5	53371399
156	hsa-mir-1178	rs7311975	T>C	12	120151493
157	hsa-mir-647	rs73147065	T>T	20	62574006
158	hsa-mir-4532	rs73177830	T>T	20	56470471
159	hsa-mir-548h-4	rs73235381	T>T	8	26906402
160	hsa-mir-1269a	rs73239138	G>A	4	67142620
161	hsa-mir-4739	rs73410309	T>T	17	77681036
162	hsa-mir-4474	rs74428911	G>T	9	20502274
163	hsa-mir-6504	rs74469188	T>C	16	81644970
164	hsa-mir-3615	rs745666	C>G	17	72744798
165	hsa-mir-518d	rs74704964	C>T	19	54238208
166	hsa-mir-2682	rs74904371	C>T	1	98510847
167	hsa-mir-5702	rs74949342	C>G	2	227523436
168	hsa-mir-4719	rs7500280	T>T	16	76902847
169	hsa-mir-4477a	rs75019967	A>A	9	68415338
170	hsa-mir-4742	rs7522956	A>C	1	224585958
171	hsa-mir-520f	rs75598818	G>A	19	54185492
172	hsa-mir-944	rs75715827	T>C	3	189547735
173	hsa-mir-4298	rs75966923	C>A	11	1880730
174	hsa-mir-182	rs76481776	C>T	7	129410227
175	hsa-mir-4521	rs76800617	A>G	17	8090294

Table 17: SNPS selected in miRNA genes (continuation).

	Gene	SNP	Alelles	Chromosome	Location
176	hsa-mir-1303	rs77055126	T>T	5	154065348
177	hsa-mir-4634	rs7709117	A>G	5	174178774
178	hsa-mir-576	rs77639117	A>T	4	110409933
179	hsa-mir-4743	rs78396863	G>C	18	46196971
180	hsa-mir-6075	rs78541299	G>A	5	1510904
181	hsa-mir-6083	rs78790512	G>A	3	124093220
182	hsa-mir-4789	rs78831152	C>T	3	175087408
183	hsa-mir-4786	rs78832554	G>A	2	240882476
184	hsa-mir-4481	rs7896283	A>G	10	12695177
185	hsa-mir-1307	rs7911488	A>G	10	105154089
186	hsa-mir-597	rs79397096	G>A	8	9599276
187	hsa-mir-3976	rs79512808	T>G	5	82136024
188	hsa-mir-5707	rs80128580	G>A	7	158384368
189	hsa-mir-3176	rs8054514	T>G	16	593277
190	hsa-mir-4520a	rs8078913	C>T	17	6558768
191	hsa-mir-4698	rs832733	T>T	12	47581629
192	hsa-mir-550a-3	rs850108	T>T	7	29720404
193	hsa-mir-4751	rs8667	G>A	19	50436371
194	hsa-mir-4671	rs877722	A>T	1	234442257
195	mir-27a	rs895819	T>C	19	13947292
196	hsa-mir-4519	rs897984	T>T	16	30886643
197	hsa-mir-5689	rs9295535	T>T	6	10439968
198	hsa-mir-3141	rs936581	G>A	5	153975576
199	hsa-mir-5186	rs9842591	C>A	3	151283691
200	hsa-mir-5680	rs9877402	A>G	3	120768492
201	hsa-mir-548h-3	rs9913045	T>T	17	13446924
202	hsa-mir-4302	rs11048315	G>A	12	26026988
203	hsa-mir-3908	rs111803974	T>T	12	124021017
204	hsa-mir-299, hsa-mir-380	rs111906529	T>C	14	101489703
205	hsa-mir-520G	rs112328520	C>T	19	54225501
206	hsa-mir--1282	rs11269	G>G	15	44085909
207	hsa-mir-4532	rs113808830	C>T	20	56470456
208	hsa-mir-4479	rs116932476	G>A	9	139781193
209	hsa-mir-296	rs117258475	G>A	20	57392686
210	hsa-mir-6717	rs117650137	G>A	14	21491532
211	hsa-mir-3649	rs117723462	T>G	12	1769533
212	hsa-mir-4436B2	rs163642	T>T	2	111042483
213	hsa-mir-3689	rs62571442	A>G	9	137742124

Genotype analyses

Genomic DNA was extracted from remission peripheral blood or bone marrow (with <5 % blast cells) using the phenol-chloroform method as previously described (Sambrook and Russell

2001). DNA was quantified using PicoGreen (Invitrogen Corp., Carlsbad, CA). For the first study, SNP genotyping was performed using TaqMan OpenArray Genotyping technology (Applied Biosystems, Carlsbad, CA) according to the published Applied Biosystems protocol. For the second one, the GoldenGate Genotyping Assay with Veracode technology according to the published Illumina protocol was used.

4. DATA ANALYSIS

Association study

To identify any deviation in HWE for the healthy controls, a $\chi 2$ test was used. The association between genetic polymorphisms in cases and controls, as well as between ALL subtypes and controls, was also evaluated using the $\chi 2$ or Fisher's exact test. The effect sizes of the associations were estimated by the odds ratio (OR) from univariate logistic regression. The most significant test among codominant, dominant, recessive, and additive genetic models was selected. The results were adjusted for multiple comparisons by the False Discovery Rate (FDR) (Benjamini and Hochberg 1995). In all cases the significance level was set at 5%. Analyses were performed by using R v2.11 software.

Meta-analysis study

For the meta-analysis, we used a recessive model. The overall pooled OR and corresponding 95%CI were estimated using Mantel-Haenszel´s method, with random effects model. The heterogeneity was quantified using the I^2 statistic (0-25% no heterogeneity, 25-50% moderate heterogeneity, 50-75% large heterogeneity and 75-100% extreme heterogeneity). Begg's funnel plot and Egger's test (Egger et al. 1997) were performed to access the publication bias of literatures in this meta-analysis.

Gene-gene interactions

Gene-gene interactions were calculated performing log-likelihood ratio test (LRT) under three genetic models: log-additive (doses dependent effect: major allele homozygotes vs. heterozygotes vs. minor allele homozygotes), dominant (major allele homozygotes vs. heterozygotes + minor allele homozygotes) and recessive (major allele homozygotes + heterozygotes vs. minor allele homozygotes).

mRNA expression analysis

The qPCR data were analyzed using the comparative CT method (Livak and Schmittgen 2001). In addition, the nonparametric Mann–Whitney test, applied by SPSSTM for Windows19.0 (SPSS, Chicago, IL), was used to determine significance.

CNV study

The interpretation of images obtained by scanning the arrays was performed using Chromosome Analysis Suite software (ChAS, Affimetrix, Santa Clara, CA, USA), Affymetrix annotations, and NetAffx versionbuild-3.1.0, and was based on the version NCBIv37 genome (hg19). Filters were applied for ChAS to report only the gains or losses that affected at least 50 markers within 100 kb. When the quality parameters were not optimal, the restriction filters were increased (e.g., 200 markers altered within 200 kb) to avoid false positives.

Bioinformatical analysis

miRNAs secondary structures prediction

The RNAfold web tool (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) was used to calculate the minimum free energy (MFE) secondary structures and to predict the most stable secondary structures of the miRNAs showing significant SNPs.

Gene targets selection and pathways analysis

MirWalk (Dweep and Gretz 2015) (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/) database was used to select miRNA targets. Targets predicted by at least 6 different algorithms provided by miRWalk were selected. Enriched pathway analyses of putative target genes were determined with ConsensusPath database (CPdB) (http://consensuspathdb.org/) using the over-representation analysis module. Gene list was analyzed against the default collection of KEGG (http://www.genome.jp/kegg/), Reactome (http://www.reactome.org/) and BioCarta (http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways) pathways databases. A conservative *p*-value cutoff (0.0001) was used.

RESULTS AND DISCUSSION

J Cancer Res Clin Oncol
DOI 10.1007/s00432-013-1512-3

ORIGINAL PAPER

Intron 3 of the *ARID5B* gene: a hot spot for acute lymphoblastic leukemia susceptibility

Gutiérrez-Camino A [1]; López-López E[1]; Martín-Guerrero I[1]; Sanchez-Toledo J[2]; García-de Andoin N[3]; Carboné A[4]; García-Miguel P[5]; Navajas A[6]; Garcia-Orad A [1].

[1]Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Leioa, Spain; [2] Service of Pediatric Oncology and Hematology, University Hospital Vall d' Hebron, VHIR, Barcelona, Spain; [3] Unit of Pediatric Oncohematology, University Hospital Donostia, San Sebastian, Spain; [4]Department of Paediatrics, Hospital Miguel Servet, Zaragoza, Spain; [5]Service of Pediatric Oncohematology, University Hospital La Paz, Madrid, Spain; [6]Unit of Pediatric Hematology/Oncology, University Hospital Cruces, Bilbao, Spain

Abstract

Single-nucleotide polymorphisms (SNPs) in AT-rich interactive domain 5B (*ARID5B*) have been associated with risk for pediatric acute lymphoblastic leukemia (ALL). After reviewing previous studies, we realized that the most significant associations were restricted to intron 3, but the mechanism(s) by which those SNPs affect ALL risk remain to be elucidated. Therefore, the aim of this study was to analyze the association between genetic variants of the intron 3 region of *ARID5B* and the incidence of B-ALL in a Spanish population. We also aimed to find a functional explanation for the association, searching for copy number variations (CNVs), and changes in ARID5B expression associated with the genotypes of the SNPs. We analyzed 10 SNPs in intron 3 of *ARID5B* in a Spanish population of 219 B-ALL patients and 397 unrelated controls with the Taqman Open Array platform. CNVs were analyzed in 23 patients and 17 controls using the Cytogenetics Whole-genome 2.7 M platform. Expression of ARID5B transcript 1 was quantified by qPCR and related to SNPs genotype in seven ALL cell lines. Association between intron 3 and B-ALL risk was confirmed for all of the SNPs evaluated in our Spanish population. We could not explain this association by the presence of CNVs. We neither detected changes in the expression of ARID5B isoform associated with the genotype of the SNPs. The intron 3 of *ARID5B* gene was found to be strongly associated with B-ALL risk in the Spanish population examined. However, neither CNVs nor changes in mRNA expression were found to be responsible for this association.

Keywords Acute lymphoblastic leukemia, Childhood, *ARID5B*, Susceptibility

INTRODUCTION

Despite acute lymphoblastic leukemia (ALL) being the most common malignancy in children (Johnston et al. 2010), its etiology remains poorly understood. It is hypothesized that both environmental and genetic factors contribute to the initiation of leukemogenesis (Healy et al. 2010). Recently, single nucleotide polymorphisms (SNPs) in two genetic loci, AT-rich interactive domain 5B (*ARID5B*) and Ikaros family zinc finger protein 1 (*IKZF1*), were found to be strongly associated with ALL risk (Treviño et al. 2009, Papaemmanuil et al. 2009). The association of *ARID5B* with ALL risk was a novel finding, and was confirmed in subsequent studies (Yang et al. 2010, Han et al. 2010, Healy et al. 2010, Prasad et al. 2010, Xu et al. 2012, Vijayakrishnan et al. 2010, Ellinghaus et al. 2012, Orsi et al. 2012, Pastorczak et al. 2011). In some studies, *ARID5B* SNPs were found to be more significantly associated with childhood hyperdiploid B-ALL (Treviño et al. 2009, Papaemmanuil et al. 2009). When gender-specific effects were considered, contradictory results for these associations have been reported (Orsi et al. 2012, Healy et al. 2010). The association between *ARID5B* and ALL risk has also been confirmed for different ethnic groups (Healy et al. 2010, Yang et al. 2010, Han et al. 2010, Xu et al. 2012, Vijayakrishnan et al. 2010), which supports the hypothesis that *ARID5B* is involved in a general mechanism that contributes to the etiology of childhood ALL.

ARID5B is a member of the ARID family of transcription factors and has important roles in embryogenesis and growth regulation. Accumulating evidence also appears to indicate that *ARID5B* has a role in ALL development. For example, *Arid5b* knockout mice exhibit abnormalities in B-lymphocyte development (Paulsson et al. 2010, Lahoud et al. 2001), and ARID5B mRNA expression is upregulated in hematologic malignancies such as acute promielocytic leukemia (Chang et al. 2008) and acute megakaryioblastic leukemia (Bourquin et al. 2006). However, the role of *ARID5B* in childhood ALL remains unknown, although alterations in gene function may contribute to an increased risk for this disease (Treviño et al. 2009).

Of note was the fact that all the significant SNPs in *ARID5B* are located in intron 3, or exhibit high linkage disequilibrium ($r^2 > 0.8$) with intron 3, without a known function. In fact, the mechanism(s) by which these SNPs affect the risk of ALL remain to be elucidated.

The mechanism by which these SNPs are associated with an increased risk of ALL can be diverse. On the one hand, it is known that copy number variations (CNVs) have the potential to influence a healthy individual's susceptibility to cancer, possibly by varying the gene dosage of

tumor suppressors or oncogenes (Huang et al. 2012). CNVs have previously been described for this region (Perry et al. 2008, Gusev et al. 2009). So, these SNPs may be markers in linkage disequilibrium with CNVs in the region. Furthermore, alterations in the frequency of SNPs present in the intron 3 of *ARID5B* could be caused by an increase in the frequency of homozygosity due to the localization of CNVs to intron 3. On the other hand, SNPs in *ARID5B* might also have a role in transcriptional regulation, thereby affecting expression of *ARID5B* or the splicing, generating different isoforms. Accordingly, changes in the concentration of different RNA isoforms of certain genes have been described in cancer cells (Pal et al. 2012).

The aim of this study was to analyze the association between genetic variants of the intron 3 region of *ARID5B* and the incidence of B-ALL in a Spanish population. We selected other SNPs with a putative effect and searched for possible functional explanations.

METHODS

Study subjects and cell lines

In this study, 219 children diagnosed with B-cell precursor ALL (BCP-ALL) between 1995 and 2010 in the Pediatric Oncology Units of five Spanish hospitals (e.g., Hospital Cruces; Hospital Donostia; Hospital La Paz; Hospital Miguel Servet; Hospital Vall d´Hebrón), were enrolled. Controls included 397 unrelated healthy individuals (Table 18). Informed consent was obtained from all participants, or their parents, before samples were collected. Patients were classified with B-hyperdiploid ALL if their DNA index value was > 1.16, and/or their karyotype included more than 50 chromosomes. For 95/219 (43.4%) B-ALL patients, cytogenetic data were not available.

For functional studies, seven B-ALL-derived cell lines were studied: MY, 697, TOM1, TANOUE, REH, NALM20, and SEM.

Table 18. Characteristics of the B-ALL patients and controls examined in this study.

	Patients (n = 219)	**Controls (n = 397)**
Mean age ± SE (y)	5.48 ± 3.48	51.16 ± 7.81
Males, n (%)	124 (56.22)	202 (51.92)
Females, n (%)	95 (43.37)	187 (48.07)
Hyperdiploid, n (%)	53 (24.89)	-
Absence of hyperdiploid, n (%)	85 (39.91)	-
Chromosomal Translocations, n(%)		
ETV6-RUNX1	27 (12.32)	-
MLL	7 (3.19)	-
BCR-ABL	4 (1.82)	-
E2A-PBX1	1 (0.450)	-
Overall 5-year survival, n (%)	214 (93.15)	-
5-year event-free survival, n (%)	188 (85.84)	-

SE: standard error.

Polymorphism selection

To map the whole intron 3 of *ARID5B*, the SNPs were selected based on the following criteria: (i) TagSNPs that were defined using Haploview software v.4.2 (http://www.broadinstitute.org/haploview/haploview) with an r^2 threshold value of 0.8; (ii) SNPs predicted to have functional effects (e.g., putative transcription factor binding sites and CpG sites) according to bioinformatic analyses (F-SNP, Fast-SNP, Ensembl, and Genome Browser); (iii) SNPs previously reported to be associated with ALL susceptibility. Of these, only SNPs with a reported minor allele frequency (MAF) > 10% or 0.10 were analyzed, and this included a total of 10 polymorphisms (Figure 13A).

Genotype analysis

Genomic DNA was extracted from remission peripheral blood or bone marrow (with < 5% blast cells) using the phenol-chloroform method previously described (Lopez-Lopez et al. 2011), or from saliva samples using Oragene DNA kit (DNA Genotek, Ottawa, Ontario, Canada), according to the manufacturer´s instructions. Genotyping was performed using TaqMan Open Array technology (Applied Biosystems, Life Technologies, Carlsbad, CA, USA) according to the published Applied Biosystems protocol. Data were analyzed using Taqman Genotyper software for genotype clustering and genotype calling. Duplicate samples were genotyped.

Copy number variation (CNV) analysis

Genomic DNA was extracted from lymphocytes isolated with Ficoll-Paque™ PLUS (GE Healthcare, Uppsala, Sweden) and from remission bone marrow or peripheral blood

(containing < 5% blast cells) from 23 patients diagnosed at Cruces Hospital, and from 17 controls, using QIAamp DNA Blood Mini Kits (Qiagen, Hilden, Germany). Copy number detection was performed using the Cytogenetics Whole-Genome 2.7M platform (Affymetrix, Santa Clara, CA, USA), containing 2,761,979 copy number probes with 735 bp being the median distance between markers. Analyses were performed according to the manufacturer's instructions.

Gene expression

Quantitative real-time PCR (qPCR) was performed using an ABI PRISM 7900HT Fast Real-Time PCR System (Applied Biosystems, Life Technologies, Carlsbad, USA) to detect *ARID5B* mRNA expression in seven different cell lines, according to the manufacturer's protocol. Briefly, RNA was extracted using Ultraspec (Biotecx, Houston, TX, USA) according to the manufacturer's instructions. Primers and probes for the *ARID5B* gene (Hs01381961_m1) and the TATAbox binding protein gene (*TBP*) [used as a housekeeping control (Hs00427620_m1)], were obtained from Applied Biosystems. Of the two transcript variants previously characterized for *ARID5B*, the primers and probes selected for this study were specific for transcript variant 1. This variant contains exons 2 and 3, which limits the region analyzed (Figure 14B) in order to avoid errors due to the quantifycation of gene expression for both transcripts.

Data analysis

Statistical analyses were performed using R software (version v2.14.1). $\chi 2$ was used to search for any deviation of Hardy-Weinberg equilibrium (HWE) for the healthy controls. The association between genetic polymorphisms in cases and controls, and between B-hyperdiploid ALL and non B-hyperdiploid ALL cases, were evaluated using $\chi 2$ or Fisher's exact tests. Gender-specific associations were examined using a stratified analysis of male cases versus male controls and female cases versus female controls. The effect sizes of the associations were estimated using odds ratio (OR) values obtained from univariate logistic regression. A co-dominant test was used to determine the statistical significance of each SNP. The results were adjusted for multiple comparisons using the False Discovery Rate (FDR) (Klipper-Aurbach et al. 1995). In all cases the significance level was set at 5%.

In the CNV study, the interpretation of images obtained by scanning the arrays was performed using Chromosome Analysis Suite software (ChAS, Affimetrix, Santa Clara, CA, USA), Affymetrix annotations, and NetAffx version-build-3.1.0, and was based on the version NCBIv37 genome (hg19). Filters were applied for ChAS to report only the gains or losses that affected at least 50

markers within 100 kb. When the quality parameters were not optimal, the restriction filters were increased (e.g., 200 markers altered within 200 kb) to avoid false positives.

The qPCR data were analyzed using the comparative CT method (Livak and Schmittgen 2001). In addition, the nonparametric Mann-Whitney test, applied by SPSS™ for Windows 19.0 (SPSS, Chicago, IL), was used to determine significance.

RESULTS

ARID5B SNPs and B-ALL Susceptibility

A total of 576 DNA samples (93.51%) were successfully genotyped. All ten of the SNPs present in intron 3 of *ARID5B* were genotyped satisfactorily. The average genotyping rate for all SNPs was 95.39%. Of these SNPs, only one (rs10740055) was not in HWE in the control population, and therefore, only the other nine SNPs were examined in association studies.

To identify genetic variations associated with B-ALL risk, the genotype frequencies of the nine *ARID5B* SNPs identified were calculated for 219 pediatric B-ALL cases and 397 controls. All of the SNPs were found to be significantly associated with B-ALL risk. Furthermore, after FDR correction, all of the SNPs remained significantly associated with B-ALL susceptibility (Table 19).

Table 19: Distribution of the *ARID5B* genotypes among B-ALL cases and controls.

SNP	Genotype	Controls n (%)	Cases n (%)	OR (CI 95%)	P	Adjusted P value*
rs10821936	TT	169 (45.9)	55 (27.0)	Reference	4.5×10^{-7}	4.5×10^{-6}
	CT	150 (40.8)	90 (44.1)	1.84(1.23–2.75)		
	CC	49 (13.3)	59 (28.9)	3.70 (2.28–6.01)		
rs7089424	TT	163 (44.1)	55 (27.4)	Reference	2.19×10^{-5}	1.2×10^{-4}
	GT	155 (41.9)	92 (45.8)	1.76(1.18–2.62)		
	GG	52 (14.1)	54 (26.9)	3.08(1.89–5.02)		
rs7073837	CC	119 (34.1)	32 (17.2)	Reference	5.93×10^{-5}	1.09×10^{-4}
	AC	157 (45.0)	96 (51.6)	2.27 (1.43–3.62)		
	AA	73 (20.9)	58 (31.2)	2.95(1.76–4.97)		
rs7087507	AA	157 (43.0)	55 (28.4)	Reference	1.54×10^{-4}	3.8×10^{-4}
	AG	154 (42.2)	86 (44.3)	1.59 (1.06–2.39)		
	GG	54 (14.8)	53 (27.3)	2.8(1.72–4.56)		
rs10821938	CC	127 (34.4)	47 (22.9)	Reference	2.5×10^{-4}	5.1×10^{-4}
	AC	170 (46.1)	89 (43.4)	1.41 (0.93–2.16)		
	AA	72 (19.5)	69 (33.7)	2.59(1.62–4.14)		
rs7923074	CC	122 (33.5)	44 (21.7)	Reference	6.2×10^{-4}	1.04×10^{-3}
	AC	171 (47.0)	94 (46.3)	1.52 (0.99–2.34)		
	AA	71 (19.5)	65 (32.0)	2.54 (1.57–4.11)		
rs10761600	AA	133 (36.2)	102 (50.0)	Reference	0.005	0.006
	AT	181 (49.3)	82 (40.2)	0.59(0.41–0.85)		
	TT	53 (14.4)	20 (9.80)	0.49(0.28–0.87)		
rs4131566	TT	57 (21.8)	56 (32.9)	Reference	0.0124	0.0138
	CT	129 (49.2)	81 (47.6)	0.64 (0.40–1.01)		
	CC	76 (29.0)	33 (19.4)	0.44 (0.25–0.77)		
rs10994982	AA	94 (25.6)	67 (34.4)	Reference	0.018	0.018
	AG	178 (48.5)	95 (48.7)	0.75 (0.5–1.12)		
	GG	95 (25.9)	33 (16.9)	0.49 (0.29–0.81)		

OR: Odds ratio; CI: Confidence Interval. *Adjusted for multiple comparisons using the False Discovery Rate ($p \leq 0.05$).

Association between ARID5B SNP genotype and B-ALL susceptibility according to B-hyperdiploid genetic subtype and gender

Polymorphisms in intron 3 of *ARID5B* have been shown to confer an increased risk for developing B-hyperdiploid ALL. Therefore, SNP genotype frequencies between B-hyperdiploid ALL and non B-hyperdiploid patients were compared. However, no significant difference was found between these two sets of patients ($p > 0.05$). When B-hyperdiploid ALL patients and controls were compared, a statistically significant association was observed for SNPs rs10821936, rs7089424, rs7087507, and rs4131566 ($p < 0.05$) (Table 20).

Table 20: Association analysis between B-hyperdiploid ALL patients and controls.

SNP	Risk allele	Genotype	Controls n (%)	Cases n (%)	OR (CI 95%)	P
		TT	169 (45.9)	14(28.6)	Reference	
rs10821936	C	CT	150 (40.8)	21(42.9)	1.69(0.83-3.44)	0.012
		CC	49 (13.3)	14(28.6)	3.45 (1.54-7.72)	
		TT	163 (44.1)	13(26.5)	Reference	
rs7089424	G	GT	155 (41.9)	22(44.9)	1.78(0.87-3.66)	0.014
		GG	52 (14.1)	14(28.6)	3.38(1.49-7.64)	
		AA	157 (43)	13(27.7)	Reference	
rs7087507	G	AG	154 (42.2)	21(44.7)	1.65 (0.8-3.41)	0.042
		GG	54 (14.8)	13(27.7)	2.91(1.27-6.66)	
		CC	119 (34.1)	11(23.9)	Reference	
rs7073837	A	AC	157 (45)	20(43.5)	1.38 (0.64-2.99)	0.16
		AA	73 (20.9)	15(32.6)	2.22(0.97-5.1)	
		CC	127 (34.4)	10(20.4)	Reference	
rs10821938	A	AC	170 (46.1)	25(51)	1.87 (0.87-4.03)	0.09
		AA	72 (19.5)	14(28.6)	2.47(1.04-5.84)	
		CC	122 (33.5)	11(22.9)	Reference	
rs7923074	A	AC	171 (47)	23(47.9)	1.49 (0.7-3.17)	0.18
		AA	71 (19.5)	14(29.2)	2.19 (0.94-5.08)	
		CC	101 (29)	17(35.4)	Reference	
rs10740055	C	AC	154 (44.3)	24(50)	0.93(0.47-1.81)	0.15
		AA	93 (26.7)	7(14.6)	0.45 (0.18-1.13)	
		AA	133 (36.2)	23(47.9)	Reference	
rs10761600	A	AT	181 (49.3)	21(43.8)	0.67(0.36-1.26)	0.21
		TT	53 (14.4)	4(8.3)	0.44 (0.14-1.32)	
		CC	76 (29)	6(14.6)	Reference	
rs4131566	T	CT	129 (49.2)	20(48.8)	1.96 (0.76-5.1)	0.04
		TT	57 (21.8)	15(36.6)	3.33 (1.22-9.13)	
		AA	94 (25.6)	15(32.6)	Reference	
rs10994982	A	AG	178 (48.5)	25(54.3)	0.88 (0.44-1.75)	0.11
		GG	95 (25.9)	6(13)	0.4 (0.15-1.06)	

OR: Odds ratio; CI: Confidence Interval.

Due to conflicting reports regarding the possibility of a gender-specific association for B-ALL, the association between *ARID5B* polymorphisms and B-ALL susceptibility in relation to gender was also investigated. In this analysis, *ARID5B* SNPs were more often associated with females than males (Table 21).

Table 21: Stratified analysis of ALL patients between males and females.

		Males				Females			
		Controls	Cases			Controls	Cases		
SNP	Genotype	n (%)	n (%)	OR (CI 95%)	P	n (%)	N (%)n	OR (CI 95%)	P
rs10821936	TT	79(27.4)	33(29.5)	Reference	0.008	85(48.3)	22(23.9)	Reference	3.75 e-5
	CT	82(46.9)	50(44.6)	1.46(0.85-2.5)		67(38.1)	40(43.5)	2.31(1.25-4.25)	
	CC	45(25.7)	29(25.9)	2.89(1.47-5.69)		24(13.6)	30(32.6)	4.83(2.37-9.85)	
rs7089424	TT	76(40.9)	31(27.9)	Reference	0.0208	82(46.3)	24(26.7)	Reference	0.00118
	GT	87(46.8)	55(49.5)	1.55(0.91-2.65)		67(37.9)	37(41.1)	1.89(1.03-3.46)	
	GG	23(12.4)	25(22.5)	2.66(1.32-5.39)		28(15.8)	29(32.2)	3.54(1.77-7.06)	
rs7073837	CC	58(33.1)	19(18.1)	Reference	0.0174	60(35.5)	13(16)	Reference	0.0023
	AC	82(46.9)	57(54.3)	2.12(1.14-3.94)		72(42.6)	39(48.1)	2.5(1.22-5.11)	
	AA	35(20)	29(27.6)	2.53(1.24-5.17)		37(21.9)	29(35.8)	3.62(1.67-7.83)	
rs7087507	AA	75(41)	33(30.3)	Reference	0.0342	78(44.6)	22(25.9)	Reference	0.0034
	AG	83(45.4)	49(45)	1.34(0.78-2.3)		69(39.4)	37(43.5)	1.9(1.02-3.53)	
	GG	25(13.7)	27(24.8)	2.45(1.24-4.85)		28(16)	26(30.6)	3.29(1.61-6.72)	
rs10821938	CC	59(31.9)	27(23.9)	Reference	0.1	65(36.7)	20(21.7)	Reference	0.0017
	AC	89(48.1)	52(46)	1.28(0.72-2.26)		78(44.1)	37(40.2)	1.54(0.82-2.91)	
	AA	37(20)	34(30.1)	2.01(1.05-3.85)		34(19.2)	35(38)	3.35(1.68-6.66)	
rs7923074	CC	58(31.5)	26(23.4)	Reference	0.15	61(35.3)	18(19.6)	Reference	0.0024
	AC	90(48.9)	54(48.6)	1.34(0.76-2.37)		78(45.1)	40(43.5)	1.74(0.91-3.33)	
	AA	36(19.6)	31(27.9)	1.92(0.99-3.74)		34(19.7)	34(37)	3.39(1.67-6.88)	
rs10740055	CC	48(27.4)	39(34.5)	Reference	0.15	51(30.7)	43(47.3)	Reference	0.0058
	AC	82(46.9)	55(48.7)	0.83(0.48-1.42)		69(41.6)	36(39.6)	0.62(0.35-1.1)	
	AA	45(25.7)	19(16.8)	0.52(0.26-1.03)		46(27.7)	12(13.2)	0.31(0.15-0.66)	
rs10761600	AA	66(36.1)	54(48.2)	Reference	0.058	65(36.7)	48(52.2)	Reference	0.00605
	AT	96(52.5)	43(38.4)	0.55(0.033-0.91)		83(46.9)	39(42.4)	0.64(0.37-1.08)	
	TT	21(11.5)	15(13.4)	0.87(0.41-1.86)		29(16.4)	5(5.4)	0.23(0.08-0.65)	
rs4131566	TT	36(27.7)	21(22.8)	Reference	0.5	28(22)	31(39.7)	Reference	0.0094
	CT	65(50)	46(50)	1.21(0.63-2.34)		62(48.8)	35(44.9)	0.51(0.26-0.98)	
	CC	29(22.3)	25(27.2)	1.48(0.69-3.16)		37(29.1)	12(15.4)	0.29(0.13-0.67)	
rs10994982	AA	47(25.5)	30(27.8)	Reference	0.3	45(25.6)	37(42.5)	Reference	0.0119
	AG	92(50)	59(54.6)	1(0.57-1.76)		83(47.2)	36(41.4)	0.53(0.29-0.95)	
	GG	45(14.5)	19(17.6)	0.66(0.33-1.34)		48(27.3)	14(16.1)	0.35(0.17-0.74)	

OR: Odds ratio; CI: Confidence Interval.

CNV analysis

No CNV or loss-of-heterozygosity (LOH) was detected for intron 3 of *ARID5B* (Figure 20) in 23 B-ALL patients or 17 healthy controls. Furthermore, no CNVs were detected for the entire gene. With regard to the genotyping study, there was also no excess of homozygosis events.

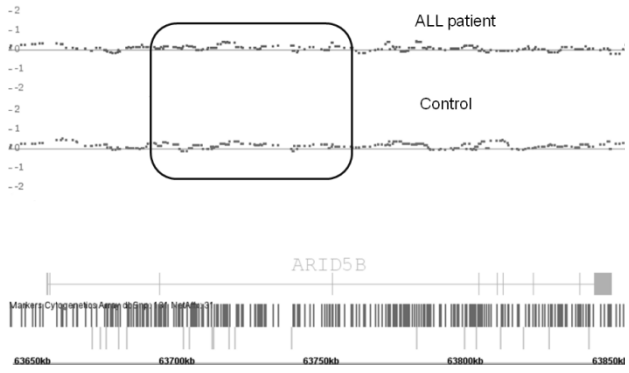

Figure 20: Representative genomic profiles of *ARID5B* from one B-ALL patient and one control. The boxed sequence represents the intron 3 region. The weighted log 2 ratio of the copy number state is indicated, and values > 1 represent gains and values < -1 represent losses. The *ARID5B* gene and the probes specific for this gene that were included in the array are shown at the bottom of the image.

Expression analysis

To investigate whether different genotypes of the SNPs selected have a functional effect on *ARID5B* expression, *ARID5B* mRNA levels of transcript variant 1 were detected in seven different B-ALL cell lines using qPCR (Table 22). The cell lines were grouped according to the recessive or dominant genotype model they represented, in order to compare the risk and protection genotypes for 7/9 SNPs (e.g., CC vs. CT+TT in rs10821936; GG vs. GT+TT in rs7089424; AA vs. AC+CC in rs7073837; GG vs. GA+AA in rs7087507; AA vs. AC+CC in rs10821938; AA vs. AC+CC in rs7923074; and AA vs. AG+GG in rs10994982). For SNPs rs10761600 and rs4131566, this was not possible due to a lack of sufficiently different genotypes in the cell lines analyzed. No significant correlation was found between the SNPs studied and the corresponding mRNA levels detected, even for the rs10821936 and rs7089424 genotypes, which are the most significant SNPs previously associated with B-ALL risk (Figure 21).

Table 22: Genotypes of the cell lines analyzed.

	697	SEM	TOM1	REH	MY	NALM20	TANOUE
rs10821936	CC	CC	TT	CT	CT	CT	CC
rs7089424	GG	GG	TT	GT	GT	GG	GG
rs7073837	-	AA	CC	AC	AC	AC	AA
rs7087507	GG	GG	AA	AG	AG	GG	GG
rs10821938	AA	AA	AA	AC	AC	AC	AA
rs7923074	AA	AA	AA	AC	AC	AC	AA
rs10761600	AA	AA	AA	AA	AA	AA	AA
rs4131566	-	-	-	CT	TT	-	TT
rs10994982	AA	AA	AA	AG	AG	AG	AA

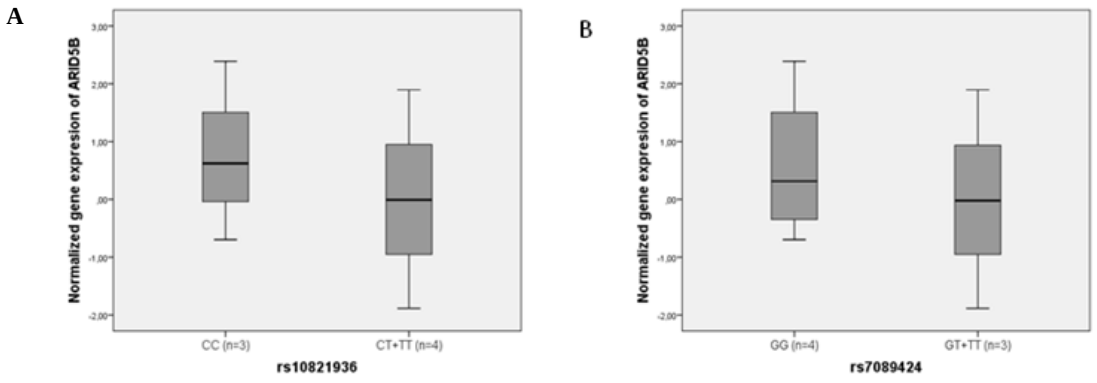

Figure 21: The correlation between *ARID5B* mRNA levels and A) rs10821936 and B) rs7089424. The horizontal line within each box represents the median value, the length of the box corresponds to the interquartile range, and the whiskers represent 1.5 interquartile ranges.

DISCUSSION

After a deep revision of *ARID5B* SNPs associated with ALL susceptibility in two GWAS studies, we observed that these SNPs were all located in intron 3, or exhibited high linkage disequilibrium with intron 3. Moreover, functions for these SNPs had not been identified. Thus, the goal of this study was to evaluate SNPs present in intron 3 of *ARID5B* to cover completely this region previously associated with ALL susceptibility, in order to identify whether other SNPs were more significant, or if putative functional effects of these SNPs could be detected. We performed an association study in a large Spanish population and searched for a functional explanation for the association.

The nine SNPs that were in HWE and were subsequently analyzed, showed significant association with pediatric B-ALL, even after correction by FDR test. Four of these nine SNPs (rs10821936, rs10994982, rs7073837 and rs7089424) had previously been associated with ALL susceptibility in two GWAS (Treviño et al. 2009, Papaemmanuil et al. 2009), and this association was confirmed in additional studies (Prasad et al. 2010, Healy et al. 2010, Yang et al. 2010, Pastorczak et al. 2011, Xu et al. 2012). An association between ALL susceptibility and rs10821938 (Vijayakrishnan et al. 2010) and rs7923074 (Xu et al. 2012) had also previously been demonstrated. Thus, to our knowledge, this is the first study to demonstrate that the SNPs, rs4131566, rs7087507, and rs10761600, are associated with ALL risk. So, our results in a Spanish population confirm the association shown by other authors in other populations.

SNPs present in intron 3 of *ARID5B* have been associated with the risk of developing high hyperdiploid childhood ALL (Treviño et al. 2009, Papaemmanuil et al. 2009, Healy et al. 2010, Xu et al. 2012). However, in the present study, when B-hyperdiploid patients were compared to controls, only four SNPs remain significant with lower p-values, and there was no significant difference in the incidence of B-hyperdiploid ALL and non B-hyperdiploid ALL. These results are consistent with those of Lautner-Csorba et al. (Lautner-Csorba et al. 2012). Taken together, these contradictory results suggest that hyperdiploidy is not an essential factor in *ARID5B*-mediated B-ALL susceptibility.

In this study, *ARID5B* SNPs were more often associated with females than males, and this is in agreement with previous data reported by Orsi et al. (Orsi et al. 2012). However, the role of gender remains controversial. For example, Xu et al. (Xu et al. 2012) and Lautner-Csorba et al. (Lautner-Csorba et al. 2012) found no differences between the incidence of childhood ALL in males versus females, while Healy et al. (Healy et al. 2010) reported a male bias. Therefore,

these contradictory results do not point to a very relevant role of gender in *ARID5B* mediated susceptibility to pediatric B-ALL.

Significant *ARID5B* SNPs have been found in intron regions but their function remains unknown. In the present study, it was investigated whether CNVs could affect this association in intron 3 of *ARID5B*. If CNV was associated with intron 3, the deletion could increase the frequency of homozygous alleles detected, thereby explaining the statistical association. Another possibility is that CNVs, in linkage disequilibrium with SNPs, might influence a healthy individual's susceptibility to cancer, by varying protein size or gene dosage of tumor suppressors or oncogenes (Huang et al. 2012). However, CNV was not observed in either the cases or controls, in intron 3, or within the entire *ARID5B* gene. In addition, no excess homozygosis was detected with statistical analyses. Therefore, it does not appear that CNVs are involved in the B-ALL risk associated with this region.

Treviño et al. proposed that *ARID5B* germline variations could affect susceptibility to B-ALL by altering *ARID5B* function during B-lineage development (Treviño et al. 2009). Hence, to identify a putative function for the nine SNPs identified for *ARID5B*, *in silico* analysis was employed. For SNPs rs10761600 and rs10994982, no putative function was found. In contrast, putative roles in transcriptional regulation were identified for SNPs rs10821936, rs7073837, rs7089424, rs7087507, and rs1082193. For example, risk alleles of rs10821936 and rs10821938 were reported to eliminate the binding site of different transcriptional factors (NIT2 and CCAAT, respectively) (Lee and Shatkay 2008), which could affect gene expression. In addition, SNPs rs7923074 and rs4131566 were found to localize to CpG sites, and the presence of risk alleles in both cases removed those CpG sites (C→A and C→T, respectively), avoiding the possibility of methylation (Samuelsson et al. 2011). These changes could potentially affect *ARID5B* regulation and expression.

Paulsson et al. suggested that involvement of the *ARID5B* gene in ALL susceptibility may involve gene transcription (Paulsson et al. 2010). Consistent with the latter, aberrant *ARID5B* expression in the developing fetus was shown to halt B-lymphocyte maturation and contribute to leukemogenesis (Healy et al. 2010). To further investigate whether mRNA levels could be altered by the different genotypes of the SNPs examined, mRNA levels of *ARID5B* were assayed in seven B-ALL cell lines. As levels of specific RNA isoforms have been shown to be affected in cancer (Pal et al. 2012), we analyzed specifically an only isoform, the *ARID5B* transcript 1, which contains the region limiting with intron 3. No differences in the mRNA levels of *ARID5B*

transcript 1 were detected in the seven cell lines assayed, according to the genotypes of the SNPs in intron 3 of *ARID5B*. In addition, an *in silico* analysis using miRBase and USCS Genome Browser databases did not identify any validated noncoding RNA genes present in the *ARID5B* intron 3. Thus, the mechanism(s) by which SNPs in intron 3 of *ARID5B* affect B-ALL susceptibility remains unclear.

In conclusion, the intron 3 of *ARID5B* gene was found to be strongly associated with B-ALL risk in the Spanish population examined. However, neither CNV nor changes in mRNA expression were found to be responsible for this association. Therefore, additional functional studies are needed to determine the role of intronic SNPs in *ARID5B*.

ACKNOWLEDGMENTS

The authors would like to thank Xabier Agirre for providing us the RNA of the seven cell lines of ALL and Nerea Bilbao for her contribution to this study. Support by SGIker (UPV/EHU) is gratefully acknowledged. This project was supported by RTICS (RD/06/0020/0048), Basque Government (GIC10/71) and UPV/EHU (UFI11/35). AGC was supported by a predoctoral grant from the Gangoiti Barrera Foundation. ELL was supported by a "Fellowship for recent Doctors until their integration in postdoctoral programs" by the Investigation Vice-rector's office of the UPV/EHU.

CONFLICT OF INTEREST STATEMENT

The authors reported no potential conflicts of interest.

REFERENCES

Bourquin, J. P., Subramanian, A., Langebrake, C., Reinhardt, D., Bernard, O., Ballerini, P., Baruchel, A., Cavé, H., Dastugue, N., Hasle, H., Kaspers, G. L., Lessard, M., Michaux, L., Vyas, P., van Wering, E., Zwaan, C. M., Golub, T. R. and Orkin, S. H. (2006) 'Identification of distinct molecular phenotypes in acute megakaryoblastic leukemia by gene expression profiling', *Proc Natl Acad Sci U S A,* 103(9), 3339-44.

Chang, L. W., Payton, J. E., Yuan, W., Ley, T. J., Nagarajan, R. and Stormo, G. D. (2008) 'Computational identification of the normal and perturbed genetic networks involved in myeloid differentiation and acute promyelocytic leukemia', *Genome Biol,* 9(2), R38.

Ellinghaus, E., Stanulla, M., Richter, G., Ellinghaus, D., te Kronnie, G., Cario, G., Cazzaniga, G., Horstmann, M., Panzer Grümayer, R., Cavé, H., Trka, J., Cinek, O., Teigler-Schlegel, A., ElSharawy, A., Häsler, R., Nebel, A., Meissner, B., Bartram, T., Lescai, F., Franceschi, C., Giordan, M., Nürnberg, P., Heinzow, B., Zimmermann, M., Schreiber, S., Schrappe, M. and Franke, A. (2012) 'Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia', *Leukemia,* 26(5), 902-9.

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L., Friedman, J. M. and Pe'er, I. (2009) 'Whole population, genome-wide mapping of hidden relatedness', *Genome Res,* 19(2), 318-26.

Han, S., Lee, K. M., Park, S. K., Lee, J. E., Ahn, H. S., Shin, H. Y., Kang, H. J., Koo, H. H., Seo, J. J., Choi, J. E., Ahn, Y. O. and Kang, D. (2010) 'Genome-wide association study of childhood acute lymphoblastic leukemia in Korea', *Leuk Res,* 34(10), 1271-4.

Healy, J., Richer, C., Bourgey, M., Kritikou, E. A. and Sinnett, D. (2010) 'Replication analysis confirms the association of ARID5B with childhood B-cell acute lymphoblastic leukemia', *Haematologica,* 95(9), 1608-11.

Huang, L., Yu, D., Wu, C., Zhai, K., Jiang, G., Cao, G., Wang, C., Liu, Y., Sun, M., Li, Z., Tan, W. and Lin, D. (2012) 'Copy number variation at 6q13 functions as a long-range regulator and is associated with pancreatic cancer risk', *Carcinogenesis,* 33(1), 94-100.

Johnston, W. T., Lightfoot, T. J., Simpson, J. and Roman, E. (2010) 'Childhood cancer survival: a report from the United Kingdom Childhood Cancer Study', *Cancer Epidemiol,* 34(6), 659-66.

Klipper-Aurbach, Y., Wasserman, M., Braunspiegel-Weintrob, N., Borstein, D., Peleg, S., Assa, S., Karp, M., Benjamini, Y., Hochberg, Y. and Laron, Z. (1995) 'Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus', *Med Hypotheses,* 45(5), 486-90.

Lahoud, M. H., Ristevski, S., Venter, D. J., Jermiin, L. S., Bertoncello, I., Zavarsek, S., Hasthorpe, S., Drago, J., de Kretser, D., Hertzog, P. J. and Kola, I. (2001) 'Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs', *Genome Res,* 11(8), 1327-34.

Lautner-Csorba, O., Gézsi, A., Semsei, A. F., Antal, P., Erdélyi, D. J., Schermann, G., Kutszegi, N., Csordás, K., Hegyi, M., Kovács, G., Falus, A. and Szalai, C. (2012) 'Candidate gene association study in pediatric acute lymphoblastic leukemia evaluated by Bayesian network based Bayesian multilevel analysis of relevance', *BMC Med Genomics,* 5, 42.

Lee, P. H. and Shatkay, H. (2008) 'F-SNP: computationally predicted functional SNPs for disease association studies', *Nucleic Acids Res,* 36(Database issue), D820-4.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', *Methods,* 25(4), 402-8.

Lopez-Lopez, E., Martin-Guerrero, I., Ballesteros, J., Piñan, M. A., Garcia-Miguel, P., Navajas, A. and Garcia-Orad, A. (2011) 'Polymorphisms of the SLCO1B1 gene predict methotrexate-related toxicity in childhood acute lymphoblastic leukemia', *Pediatr Blood Cancer,* 57(4), 612-9.

Orsi, L., Rudant, J., Bonaventure, A., Goujon-Bellec, S., Corda, E., Evans, T. J., Petit, A., Bertrand, Y., Nelken, B., Robert, A., Michel, G., Sirvent, N., Chastagner, P., Ducassou, S., Rialland, X., Hémon, D., Milne, E., Scott, R. J., Baruchel, A. and Clavel, J. (2012) 'Genetic polymorphisms and childhood acute lymphoblastic leukemia: GWAS of the ESCALE study (SFCE)', *Leukemia,* 26(12), 2561-4.

Pal, S., Gupta, R. and Davuluri, R. V. (2012) 'Alternative transcription and alternative splicing in cancer', *Pharmacol Ther,* 136(3), 283-94.

Papaemmanuil, E., Hosking, F. J., Vijayakrishnan, J., Price, A., Olver, B., Sheridan, E., Kinsey, S. E., Lightfoot, T., Roman, E., Irving, J. A., Allan, J. M., Tomlinson, I. P., Taylor, M., Greaves, M. and

Houlston, R. S. (2009) 'Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia', *Nat Genet,* 41(9), 1006-10.

Pastorczak, A., Górniak, P., Sherborne, A., Hosking, F., Trelińska, J., Lejman, M., Szczepański, T., Borowiec, M., Fendler, W., Kowalczyk, J., Houlston, R. S. and Młynarski, W. (2011) 'Role of 657del5 NBN mutation and 7p12.2 (IKZF1), 9p21 (CDKN2A), 10q21.2 (ARID5B) and 14q11.2 (CEBPE) variation and risk of childhood ALL in the Polish population', *Leuk Res,* 35(11), 1534-6.

Paulsson, K., Forestier, E., Lilljebjörn, H., Heldrup, J., Behrendtz, M., Young, B. D. and Johansson, B. (2010) 'Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia', *Proc Natl Acad Sci U S A,* 107(50), 21719-24.

Perry, G. H., Ben-Dor, A., Tsalenko, A., Sampas, N., Rodriguez-Revenga, L., Tran, C. W., Scheffer, A., Steinfeld, I., Tsang, P., Yamada, N. A., Park, H. S., Kim, J. I., Seo, J. S., Yakhini, Z., Laderman, S., Bruhn, L. and Lee, C. (2008) 'The fine-scale and complex architecture of human copy-number variation', *Am J Hum Genet,* 82(3), 685-95.

Prasad, R. B., Hosking, F. J., Vijayakrishnan, J., Papaemmanuil, E., Koehler, R., Greaves, M., Sheridan, E., Gast, A., Kinsey, S. E., Lightfoot, T., Roman, E., Taylor, M., Pritchard-Jones, K., Stanulla, M., Schrappe, M., Bartram, C. R., Houlston, R. S., Kumar, R. and Hemminki, K. (2010) 'Verification of the susceptibility loci on 7p12.2, 10q21.2, and 14q11.2 in precursor B-cell acute lymphoblastic leukemia of childhood', *Blood,* 115(9), 1765-7.

Samuelsson, J., Alonso, S., Ruiz-Larroya, T., Cheung, T. H., Wong, Y. F. and Perucho, M. (2011) 'Frequent somatic demethylation of RAPGEF1/C3G intronic sequences in gastrointestinal and gynecological cancer', *Int J Oncol,* 38(6), 1575-7.

Treviño, L. R., Yang, W., French, D., Hunger, S. P., Carroll, W. L., Devidas, M., Willman, C., Neale, G., Downing, J., Raimondi, S. C., Pui, C. H., Evans, W. E. and Relling, M. V. (2009) 'Germline genomic variants associated with childhood acute lymphoblastic leukemia', *Nat Genet,* 41(9), 1001-5.

Vijayakrishnan, J., Sherborne, A. L., Sawangpanich, R., Hongeng, S., Houlston, R. S. and Pakakasama, S. (2010) 'Variation at 7p12.2 and 10q21.2 influences childhood acute lymphoblastic leukemia risk in the Thai population and may contribute to racial differences in leukemia incidence', *Leuk Lymphoma,* 51(10), 1870-4.

Xu, H., Cheng, C., Devidas, M., Pei, D., Fan, Y., Yang, W., Neale, G., Scheet, P., Burchard, E. G., Torgerson, D. G., Eng, C., Dean, M., Antillon, F., Winick, N. J., Martin, P. L., Willman, C. L., Camitta, B. M., Reaman, G. H., Carroll, W. L., Loh, M., Evans, W. E., Pui, C. H., Hunger, S. P., Relling, M. V. and Yang, J. J. (2012) 'ARID5B genetic polymorphisms contribute to racial disparities in the incidence and treatment outcome of childhood acute lymphoblastic leukemia', *J Clin Oncol,* 30(7), 751-7.

Yang, W., Treviño, L. R., Yang, J. J., Scheet, P., Pui, C. H., Evans, W. E. and Relling, M. V. (2010) 'ARID5B SNP rs10821936 is associated with risk of childhood acute lymphoblastic leukemia in blacks and contributes to racial differences in leukemia incidence', *Leukemia,* 24(4), 894-6.

Under preparation

Involvement of polymorphisms in *IKZF1* and *CEBPE* in childhood Acute Lymphoblastic Leukemia susceptibility

Gutierrez-Camino, A[1]; Martín-Guerrero, I[1]; Santos, B[2]; García de Andoin, N[3,4]; Sastre, A[5]; Carboné Bañeres, A[6]; Astigarraga, I[7,8]; Navajas, A[8]; García-Orad, Á[1, 8].

[1]Department of Genetics, Physic Anthropology and Animal Physiology, University of the Basque Country, UPV/EHU, Leioa, Spain; [2]Department of Neurosciences, Psychiatry, University of the Basque Country, UPV/EHU, Leioa, Spain; [3]Department of Paediatrics, University Hospital Donostia, San Sebastian, Spain; [4]BioDonostia Health Research Institute, San Sebastian, Spain; [5]Department of Oncohematology, University Hospital La Paz, Madrid, Spain; [6]Department of Paediatrics, University Hospital Miguel Servet, Zaragoza, Spain; [7]Department of Paediatrics, University Hospital Cruces, Barakaldo, Spain; [8]BioCruces Health Research Institute, Barakaldo, Spain.

Abstract

The genetic basis of acute lymphoblastic leukemia (ALL) susceptibility is supported by genome-wide association studies (GWAS), which have reported up to five loci. Although it has been suggested that each locus has an independent role, recently, it has been shown an interaction between *IKZF1* and *CEBPE*. The highest association signal found at *IKZF1* was for rs4132601, which have not been replicated in all studies. Several meta-analyses have been done, but new studies with different results have been performed. At *CEBPE* locus, the highest association signal was found for rs2239633, replicated in some populations, but not in others. In the last meta-analysis of this SNP, some inaccuracies have been detected. In addition, rs2239633 has an unknown function, suggesting that additional polymorphism underlie the association signal. Recently, a cis-eQTL SNP, rs2239635, more significantly associated with B-ALL risk, has been reported. Therefore, the aim of this study was to determine the involvement of rs4132601 in *IKZF1* and rs2239633 and rs2239635 in *CEBPE* in the susceptibility of B-ALL in a Spanish population of 155 children and 170 controls, compare all relevant studies analyzing rs4132601 and rs2239633 and test the interaction between both SNPs. In this study, all the SNPs were associated with B-ALL susceptibility. In the meta-analysis for rs4132601 (5953 ALL childhood patients and 9807 controls) and rs2239633 (6520 ALL childhood patients and 10748 controls) we found association between both SNPs and ALL susceptibility. Finally, the gene-gene interaction analysis showed association between both genes. These results confirm the involvement of these genes in B-ALL development.

Keywords: IKZF1, CEBPE, ALL, susceptibility

INTRODUCTION

Acute lymphoblastic leukemia (ALL) is the most common childhood malignancy (Pui and Evans 2006, Greaves 2006). The genetic basis of ALL susceptibility is broadly supported by its association with certain congenital abnormalities (Xu et al. 2013) and, by genome-wide association studies (GWAS), which have reported up to five loci associated with ALL (*ARID5B*, *IKZF1*, *CEBPE*, *CDKN2A/B* and *PIP4K2A*) (Papaemmanuil et al. 2009, Treviño et al. 2009, Sherborne et al. 2010, Xu et al. 2013, Migliorini et al. 2013). Although it has been suggested that each locus has an independent role in B-ALL development (Houlston 2010, Pastorczak et al. 2011), Wiemels et al. recently showed an interaction between *IKZF1* and *CEBPE* (Wiemels et al. 2016).

IKZF1 encodes the early lymphoid transcription factor IKAROS, a transcription factor involved in the development of all lymphoid lineages (Dai et al. 2014). Moreover, alterations in Ikaros directly contribute to the pathogenesis of *BCR–ABL1* ALL (Treviño et al. 2009). The highest association signal found at this locus by diverse GWAS was for rs4132601 (Papaemmanuil et al. 2009). Although several studies have replicated the association between *IKZF1* rs4132601 polymorphism and ALL risk (Prasad et al. 2010, Lautner-Csorba et al. 2012), some studies did not (Healy et al. 2010, Emerenciano et al. 2014). In order to clarify the possible association between rs4132601 and risk of ALL, two meta-analyses were carried out, all of them confirming the existence of association (Li et al. 2015, Dai et al. 2014). In the last year, new studies showing controversial results have been performed (Kreile et al. 2016, Gharbi et al. 2016, Bahari et al. 2016), so it could be interesting to include these data in a new meta-analysis. The SNP rs4132601 maps in the 3´untrasnlate region (UTR) of *IKZF1* and is involved in IKZF1 mRNA expression, with lower expression for the risk allele. However, up to date, this SNP has an unknown function.

CEBPE is a member of CEBPs family of transcription factors and is involved in terminal differentiation and functional maturation of myeloid cells, especially neutrophils and macrophages (Wang et al. 2015). In childhood ALL, intrachromosomal translocations involving *IGH* and *CEBPE* have been described, resulting in the upregulation of CEBPE expression (Akasaka et al. 2007). At this locus, the highest association signal was found for rs2239633 and B-ALL susceptibility (Papaemmanuil et al. 2009), which, moreover, showed a strong association with B-hyperdiploid subtype (Chokkalingam et al. 2013). This finding was replicated in some studies (Prasad et al. 2010, Orsi et al. 2012, Hungate et al. 2016), but not in others

(Vijayakrishnan et al. 2010, Healy et al. 2010, Emerenciano et al. 2014). A recent meta-analysis evaluating the association between this polymorphism and the risk of ALL concluded that rs2239633 was associated with the disease (Wang et al. 2015), nevertheless, some inaccuracies were detected in the study, such as the lack of the association studies performed by Healy et al., and Ellinghaus et al.,(Healy et al. 2010, Ellinghaus et al. 2012). Additionally, the meta-analysis did not include analyses by subtypes and since its publication in 2015 new studies have been published (Kreile et al. 2016, Gharbi et al. 2016) that could help to elucidate if rs2239633 is really associated with the risk of B-ALL. The SNP rs2239633 is located in upstream region of *CEBPE* (14q11.2), mapping within a 25.7-kb region of high LD. However, rs2239633 has an unknown function, suggesting that additional polymorphisms underlie the association peak near *CEBPE*. Recently, Wiemels et al. performing imputation-based fine-mapping and functional validation analyses of this locus identified another polymorphism more significantly associated with B-ALL risk (rs2239635 at the promoter region of *CEBPE*) (Wiemels et al. 2016). The SNP rs2239635 is a cis-eQTL for *CEBPE*, showing an increased gene expression for the risk allele. Remarkably, the risk allele of rs2239635 was shown to disrupt the binding of Ikaros, avoiding CEBPE repression. In addition, an interaction between rs2239635 and rs4132601 was suggested (Wiemels et al. 2016).

Therefore, the aim of this study was to determine the involvement of rs4132601 at *IKZF1* and rs2239633 and rs2239635 at *CEBPE* in the susceptibility of B-ALL in Spanish population. Moreover, we performed an updated meta-analysis including all the studies analyzing rs4132601 and rs2239633 published so far. Finally, we tested the interaction between both SNPs.

MATERIALS AND METHODS

Study Participants

A total of 155 Caucasian children diagnosed with B-ALL between 2000 and 2011 in the Paediatric Oncology Units of four Spanish hospitals (University Hospital Cruces, University Hospital Donostia, University Hospital La Paz and University Hospital Miguel Servet) and 170 unrelated healthy controls were included in this study (Table 23).

Table 23: Characteristics of study population.

	Patients	Controls
No. of individuals	155	170
Mean age ± SE, y	3.9 ± 3.4	37.8 ± 12.8
Sex		
Males, n (%)	83 (53.5)	97 (57)
Females, n (%)	72 (46.5)	73 (43)
Genetic alterations[#]		
Hyperdiploid	39 (25.7)	-
ETV6-RUNX1	28 (18.1)	-
MLL	11 (7.1)	-
BCR-ABL	2 (1.3)	-
E2A-PBX1	3 (1.9)	-
Hypodiploid	3 (1.9)	-
Other	0	-
No alteration	62 (40.8)	-
No available	13 (8.4)	-

SE: standard error, y: years. [#]Six patients have more than one alteration.

Data were collected objectively, blinded to genotypes, from the patients' medical files. High hyperdiploid (>50 chromosomes) and *ETV6-RUNX1* ALL genetic subtypes were considered also for the analyses. The other subtypes were not included due to the low number of patients in our cohort. Sex and age data were systematically recorded. Informed consent was obtained from all participants, or from their parents prior to sample collection. The study was approved by the local ethics committees (PI2014039) and was carried out according to the Declaration of Helsinki.

Genotype analyses

Genomic DNA was extracted from remission peripheral blood or bone marrow using the phenol-chloroform method as previously described (Sambrook and Russell 2001). Genotyping analyses were performed by using amplification-refractory mutation system polymerase chain reaction (ARMS)-PCR for rs2239633 and rs2239635 and PCR followed by restriction analysis for rs4132601. Duplicates were included in each assay. The PCR products were visualized after electrophoresis on 2% agarose gels (Figure 15 and 16). Primer sequences and PCR conditions are described in detail in Table 7 and 9.

Systematic review and meta-analysis

Search strategy. We performed an exhaustive search to identify studies that examined the association between the rs4132601 and *IKZF1* and ALL susceptibility and rs2239633

polymorphism of *CEBPE* and ALL susceptibility. We used the keywords and subject terms "(IKZF1 or rs4132601) and acute lymphoblastic leukemia" and "(CEBPE or rs2239633 or 14q11.2) and leukemia" for Pubmed (www.ncbi.nlm.nih.gov/pubmed) searches for articles published until July 2016. All references cited in the studies were then reviewed to possibly identify additional publications.

Inclusion and exclusion criteria. Original studies that investigated the association between the rs4132601 and rs2239633 polymorphisms and ALL risk with sufficient data to calculate crude Odd Ratio (OR) values were included. Reviews, meta-analyses and studies analyzing other regions or variants were excluded.

Data extraction. For each article, we gathered year of publication, first author, country of origin, sample size and genotype frequencies. When it was not possible to extract the genotype data from the article, we contacted the authors to obtain them.

Quality Assessment. The quality of included studies was assessed by scoring according to a "methodological quality assessment scale" (Table 8) (Bilbao-Aldaiturriaga et al. 2016). In the scale, five items, including the representativeness of cases, source of controls, sample size, quality control of genotyping methods and HWE were carefully checked. Quality scores ranged from 0 to 10 and a higher score indicated better quality of the study. Scores > 5 were considered acceptable.

Statistical analysis

The data were statistically processed by R v2.15 software (http://www.R-project.org). Genotype frequencies in cases and controls were compared using a $\chi 2$ test. The deviation from HWE was also calculated by a $\chi 2$ test (in the healthy population). The effect sizes of the associations were estimated by the OR from univariate logistic regression using different genetic models. In all cases the significance level was set at 5%. The results were adjusted for multiple comparisons using the FDR (Benjamini and Hochberg 1995). For the meta-analysis, we used a recessive model. The overall pooled OR and corresponding 95%CI were estimated using Mantel-Haenszel´s method, with random effects model. The heterogeneity was quantified using the I^2 statistic (0-25% no heterogeneity, 25-50% moderate heterogeneity, 50-75% large heterogeneity and 75-100% extreme heterogeneity). Begg's funnel plot and Egger's test (Egger et al. 1997) were performed to access the publication bias of literatures in this meta-analysis. Gene-gene interactions were calculated performing log-likelihood ratio test (LRT).

RESULTS

Genotyping Results

A total of 155 patients with B-ALL and 170 unrelated healthy controls were included in the study. The genotyping success rate was 95.4%. All SNPs were in HWE in the control cohort.

Genotype association study of B-ALL

We found the three SNPs analyzed in *IKZF1* and *CEBPE* significantly associated with B-ALL risk (Table 24). The SNP rs4132601 at *IKZF1* was significantly associated with B-ALL risk under the additive model, displaying a 1.47-fold increased risk of B-ALL for GG genotype (95% CI: 1.04-2.06; P=0.026). In *CEBPE* gene, rs2239633 displayed the most significant value under the dominant model (CC vs CT+TT). The CT+TT genotypes showed a 0.49-fold decreased risk of B-ALL (95% CI: 0.30-0.79; P=0.003). The CC genotype of rs2239635 produced a 1.54 –fold increased risk of B-ALL (95% CI: 1.09-2.17; P=0.012). The three SNPs remained statistically associated with B-ALL risk after FDR correction.

Table 24: Association results of SNPs in *IKZF1* and *CEBPE* and B-ALL.

SNP	Genotype	N (controls) (N=170)	N(cases) (N=151)	OR (CI 95%)	P
rs4132601	TT	83 (53.9)	63 (41.7)	Additive 1.47 (1.04-2.06)	0.026*
	GT	58 (37.7)	68 (45.0)		
	GG	13 (8.4)	20 (13.2)		
	T	224 (72.7)	194 (64.2)	1.48 (1.05-2.09)	0.024*
	G	84 (27.3)	108 (35.8)		
rs2239633	CC	41 (24.6)	59 (40.1)	Dominant 0.49 (0.30-0.79)	0.003*
	CT	96 (57.5)	67 (45.6)		
	TT	30 (18)	21 (14.3)		
	C	178 (53.3)	185 (62.9)	0.67 (0.48-0.92)	0.014*
	T	156 (46.7)	109 (37.1)		
rs2239635	GG	92 (55.8)	65 (43.3)	Additive 1.54 (1.09-2.17)	0.012*
	GC	62 (37.6)	66 (44.0)		
	CC	11 (6.7)	19 (12.7)		
	G	246 (74.5)	196 (65.3)	1.55 (1.10-2.19)	0.011*
	C	84 (25.5)	104 (34.7)		

Abbreviations: CI, confidence interval; OR, odds ratio; *Significant after FDR correction.

Genotype association study of B-ALL subtypes

When we performed the analysis by subtype, we found that GG genotype increased the risk of developing B-hyperdiploid ALL (OR: 1.97; 95% CI: 1.17-3.32; P=0.01) (Table 25). When we analyzed rs2239633 and rs22396635 at *CEBPE* we found that CT+TT genotypes of rs2239633 decreased the risk of B-hyperdiploid ALL (OR: 0.31; 95% CI: 0.15-0.64; P=0.001) while CC genotype of rs2239635 increased the risk of both B-hyperdiploid ALL (OR: 1.81; 95% CI: 1.08-3.06; P=0.026) and *ETV6-RUNX1* ALL (OR: 2.22; 95% CI: 1.21-4.09; P=0.01) (Table 25). After FDR correction, all the associations remained statistically significant.

Table 25: Association results of SNPs IKZF1 and CEBPE and B-hyperdiploid ALL and ETV6-RUNX1 ALL.

			B-hyperdiploid ALL			ETV6-RUNX1 ALL		
SNP	**Genotype**	**N (controls) (N=170)**	**N (cases) (N=39)**	**OR (CI 95%)**	**P**	**N (cases) (N=26)**	**OR (CI 95%)**	**P**
rs4132601	TT GT GG	83 (53.9) 58 (37.7) 13 (8.4)	12 (32.4) 18 (48.6) 7 (18.9)	Additive 1.97 (1.17-3.32)	0.010*	14 (53.8) 8 (30.8) 4 (15.4)	Additive 0.74 (0.30-1.8)	0.49
	T G	224 (72.7) 84 (27.3)	42 (56.8) 32 (43.2)	2.03 (1.20-3.43)	0.008*	36 (69.2) 16 (30.8)	1.18 (0.62-2.24)	0.60
rs2239633	CC CT TT	41 (24.6) 96 (57.5) 30 (18)	19 (51.4) 15 (40.5) 3 (8.19	Dominant 0.31 (0.15-0.64)	0.001*	11 (44) 11 (44) 3 (12)	Dominant 0.41 (0.17-0.98)	0.050
	C T	178 (53.3) 156 (46.7)	53 (71.6) 21 (28.4)	0.45 (0.26-0.78)	0.004*	33 (66) 17 (34)	0.58 (0.31-1.09)	0.09
rs2239635	GG GC CC	92 (55.8) 62 (37.6) 11 (6.7)	15 (38.5) 18 (46.2) 6 (15.4)	Additive 1.81 (1.08-3.06)	0.026*	9 (36) 10 (40) 6 (24)	Additive 2.22 (1.21-4.09)	0.010*
	G C	246 (74.5) 84 (25.5)	48 (61.5) 30 (38.5)	1.83 (1.08-3.07)	0.022*	28 (56) 22 (44)	2.30 (1.24-4.23)	0.007*

Abbreviations: CI, confidence interval; OR, odds ratio; *Significant after FDR correction.

Meta-analysis

The search for rs4132601 *IKZF1* provided 234 records. Of these, 210 were discarded after reviewing the abstracts because they did not meet the required criteria for inclusion. The full texts of the remaining 24 studies were examined in detail. Of these, we identified a total of 9 studies that investigated the association between *IKZF1* SNP rs4132601 and ALL risk. After revision of references cited, 3 studies were included (Figure 22). The characteristics of the studies are presented in Table 26. The distribution of genotypes in the controls of each study was in agreement with HWE ($p>0.05$).

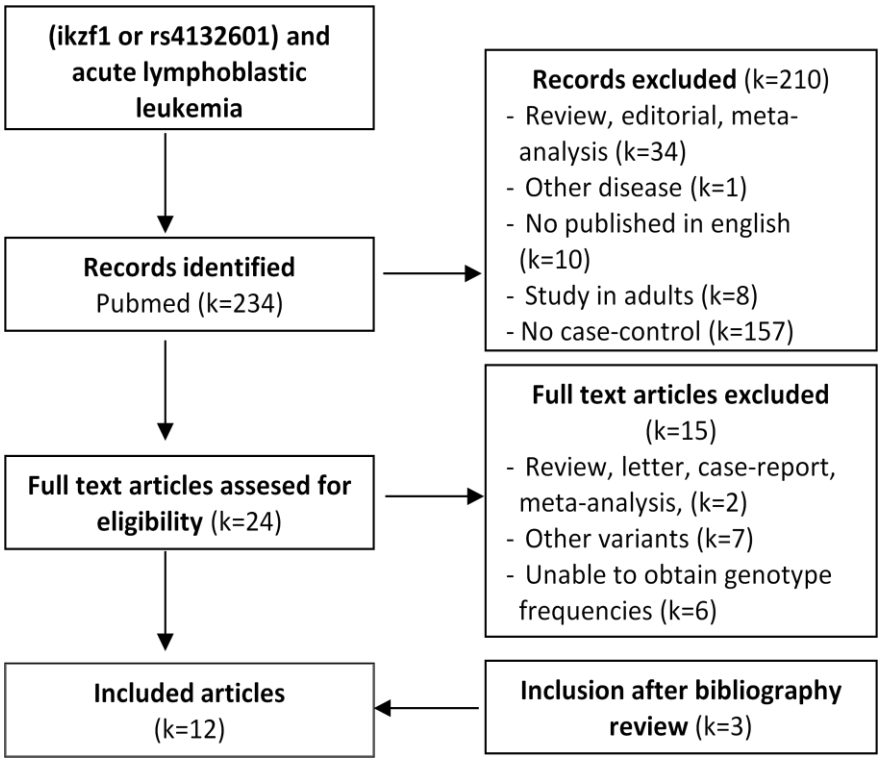

Figure 22: Flow-chart of study selection.

Table 26: Characteristic of the studies included for rs4132601 at *IKZF1.*

Study	ALL patients	B-ALL patients	Controls	Country
Gutierrez-Camino 2016	151	151	154	Spain
Papaemmanuil et al. 2009 (1)	503	459	1438	UK
Papaemmanuil et al. 2009 (2)	404	365	960	UK
Orsi et al. 2012	365	365	415	France
Pastorczak et al. 2011	389		715	Poland
Prasad et al. 2010(1)	1189	1189	1501	German
Prasad et al. 2010 (2)	188	188	360	UK
Kreile et al. 2016	82	82	121	Latvia
Healy et al. 2010	273	273	265	Canada
Ellinghaus et al. 2012	1404	1404	2674	German/Italy
Gharbi et al. 2016	58		150	Tunisia
Bahari et al. 2016	110		120	Iran
Vijayakrishnan et al. 2010	190	172	182	Thailand
Lin et al. 2014	79	45	80	Taiwan
Wang et al. 2013	568		672	China
Total	5953	4693	9807	

The original search for rs2239633 *CEBPE* provided 81 records. Of these, 65 were discarded after reviewing the abstracts because they did not meet the required criteria for inclusion. The full texts of the remaining 16 studies were examined in detail. Of these, we identified a total of 9 studies that investigated the association between *CEBPE* SNP rs2239633 and ALL risk. After

revision of references cited, 4 studies were included (Figure 23). The characteristics of the studies are presented in Table 27. The distribution of genotypes in the controls of each study was in agreement with HWE (p>0.05).

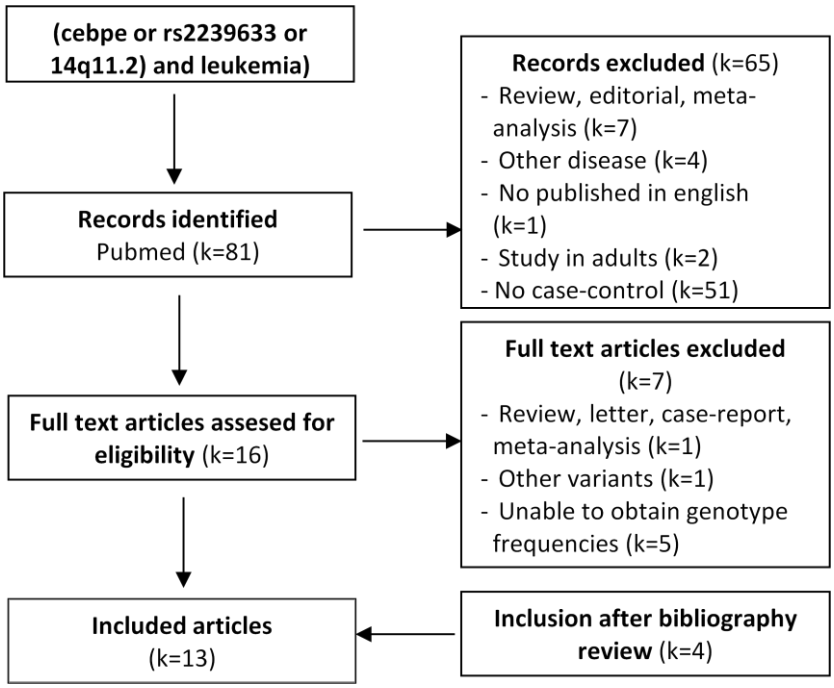

Figure 23: Flow-chart of study selection.

Table 27: Characteristic of the studies included for rs2239633 at *CEBPE.*

Study	ALL patients	B-ALL patients	Controls	Country
Gutierrez-Camino 2016	147	147	167	Spain
Papaemmanuil et al. 2009 (1)	503	459	1435	UK
Papaemmanuil et al. 2009 (2)	404	365	960	UK
Orsi et al. 2012	364	364	442	France
Pastorczak et al. 2011	388		711	Poland
Lautner-Csorba et al. 2012	541		529	Hungary
Prasad et al. 2012 (1)	1193	1193	1510	German
Prasad et al. 2012 (2)	183	183	352	UK
Kreile et al. 2016	76	76	121	Latvia
Healy et al. 2010	278	278	266	Canada
Ellinghaus et al. 2012	1382	1382	2632	German/Italy
Ross et al. 2013	85	85	363	US/Canada
Gharbi et al. 2016	58		150	Tunisia
Emerenciano et al. 2014	160	160	483	Brazil
Vijayakrishnan et al. 2010	190	172	182	Thailand
Wang et al. 2013	568		445	China
Total	6520	4864	10748	

IKZF1 rs4132601 polymorphism

A total of 13 populations including 5953 patients with ALL, 4693 of them with B-ALL, and 9807 controls were analyzed in the meta-analysis. Overall, this SNP was found to be significantly associated with ALL (p<0.0001; OR=1.8; CI 95%=1.56-2.08) (Figure 24A) and B-ALL (p<0.0001; OR=1.9; CI 95%=1.68-2.14) (Figure 24B) under the recessive model. The heterogeneity of studies on this polymorphism was 24.8% in ALL patients, while when analysis was restricted to B-ALL subtype there was no heterogeneity. The individual study's influence on the pooled results was also analyzed showing that no study affected the pooled OR significantly.

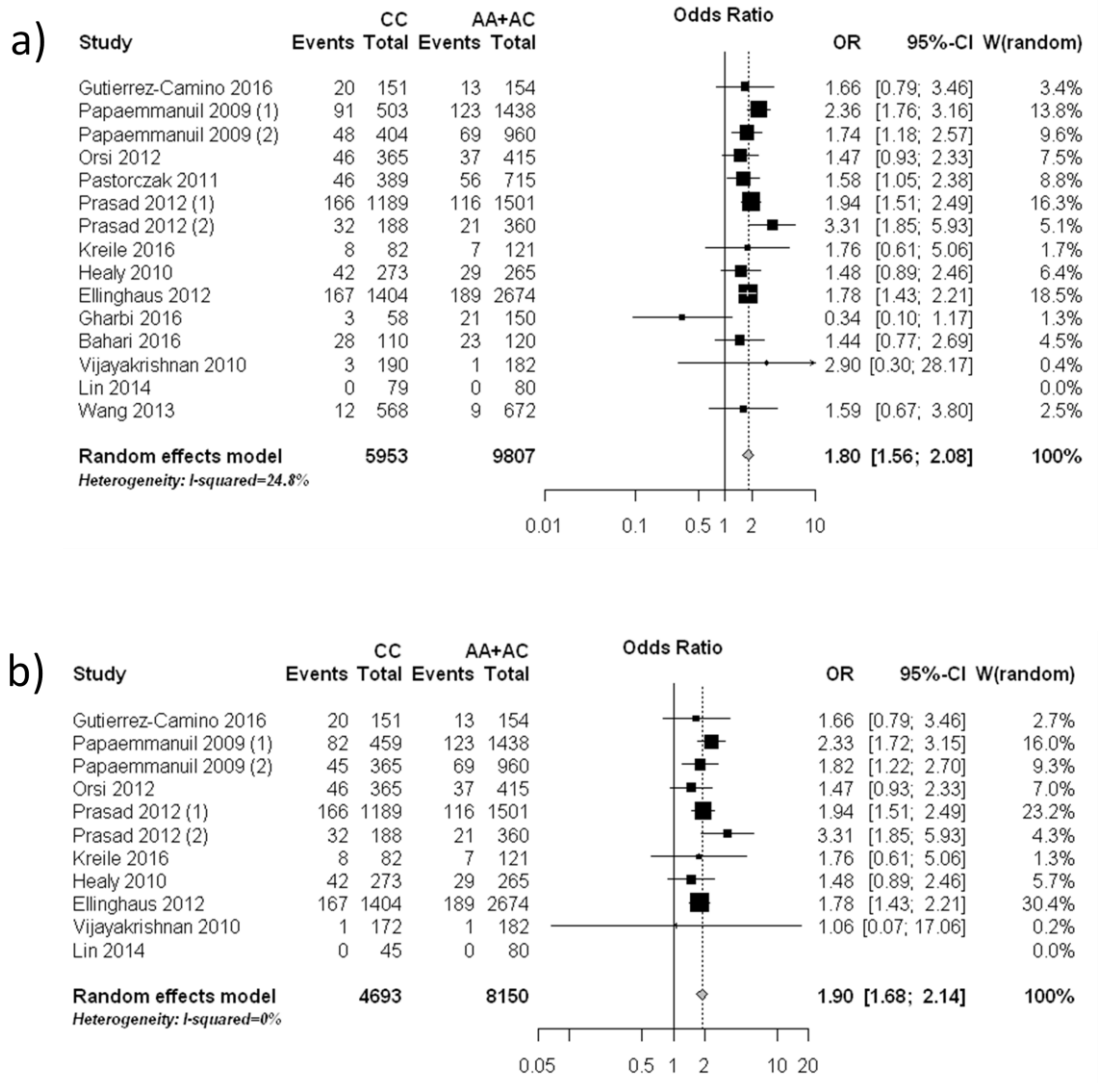

Figure 24: Forest plot for meta-analysis of the association between rs4132601 polymorphism and a) ALL risk and b)B-ALL risk.

CEBPE rs2239633 polymorphism

The meta-analysis on rs2239633 included a total of 14 studies populations with 6520 ALL patients, 4864 of them with B-ALL, and 10748 controls. Overall, this SNP was found to be significantly associated with ALL (p<0.0001; OR=0.77; CI 95%=0.69-0.86) (Figure 25A) and B-ALL

(p<0.0001; OR=0,75; CI 95%=0.68-0.82) (Figure 25B) under the recessive model. The heterogeneity of studies on this polymorphism was 34% in ALL patients, while when analysis was restricted to B-ALL subtype there was no heterogeneity. The individual study's influence on the pooled results was also analyzed showing that no study affected the pooled OR significantly.

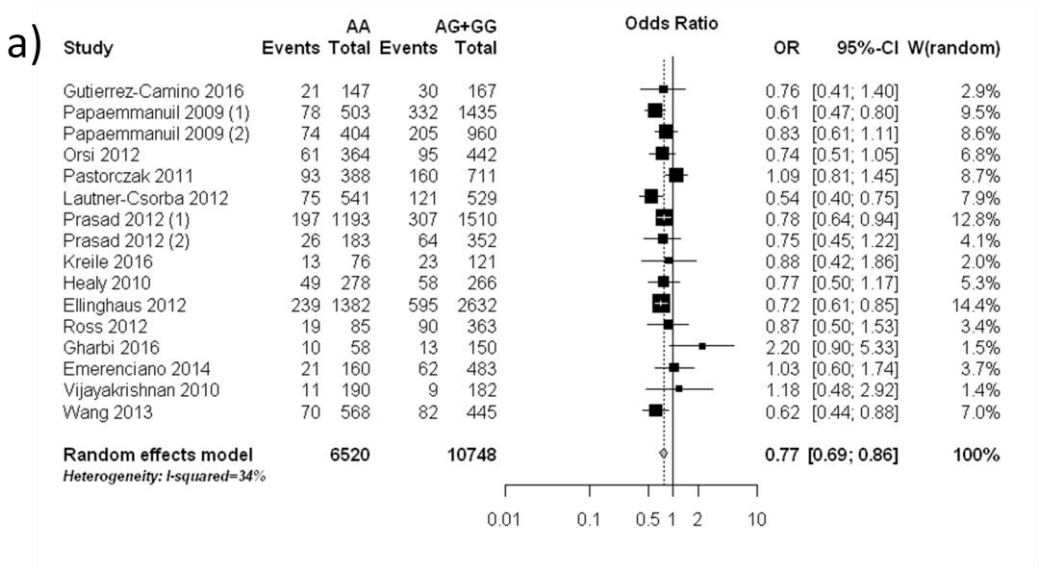

Figure 25: Meta-analysis of the association between rs2239633 polymorphism and a) ALL risk and b) B-ALL risk.

Publication bias

The shapes of funnel plot did not reveal obvious evidence of asymmetry (Figure 26) for both SNPs, and all the p values of Egger's tests were more than 0.05, providing statistical evidence of the funnel plots' symmetry. This indicates that biases from publication may not have influence on the results.

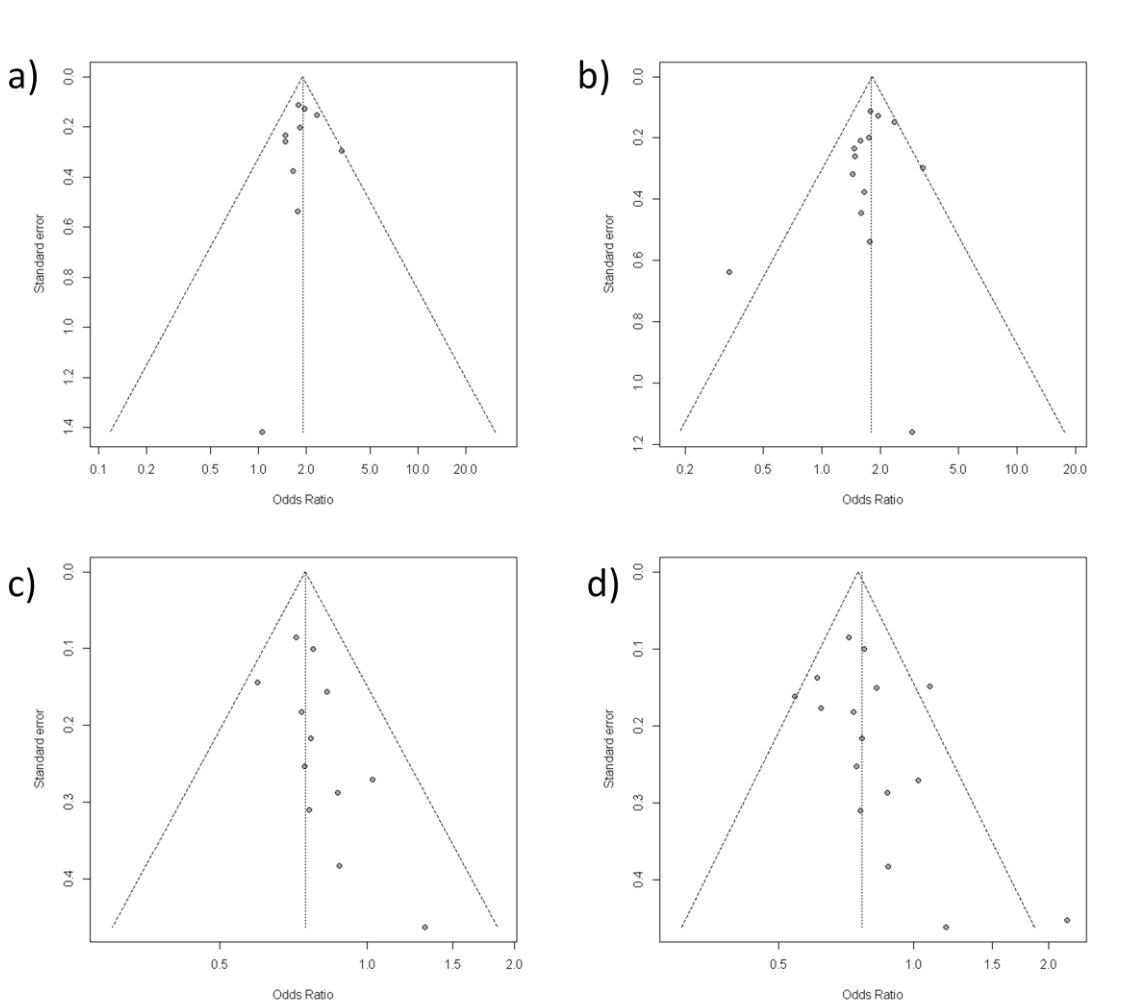

Figure 26: Funnel plots of the Egger´s test of allele comparison for publication bias. a) rs4132601 in ALL, b) rs4132601 in B-ALL, c) rs2239633 in ALL and d) rs2239633 in B-ALL.

Gene-gene interactions

Epistasis or gene-gene interactions were evaluated under three genetic models (dominant, recessive and log-additive). In case-control analyses of B-ALL, rs4132601 and rs2239633 displayed significant interaction (p= 0.02) under the recessive model. In B-hyperdiploid patients, we found interaction between rs4132601 and rs2239633 (p=0.03) under the dominant model and between rs4132601 and rs2239635 (p=0.02) under the log-additive model. Finally, in patients carrying *ETV6-RUNX1* translocation we found interaction between rs4132601 and rs2239633 (p=0.03) under the recessive model.

DISCUSSION

In the current study, we analyzed rs4132601 at *IKZF1* and rs2239633 and rs2239635 at *CEBPE* in 155 children with B-ALL and 170 controls in a Spanish cohort. All the SNPs were associated

with B-ALL susceptibility and B-hyperdiploid subtype. In addition, rs2239635 was associated with *ETV6-RUNX1* ALL. We also performed a meta-analysis with all available data for rs4132601 (5953 ALL childhood patients and 9807 controls) and rs2239633 (6520 ALL childhood patients and 10748 controls) including 13 and 14 studies respectively, and found association between both SNPs and ALL susceptibility. Finally, the gene-gene interaction analysis showed association between both genes.

We found the GG genotype of rs4132601 at *IKZF1* associated with an increased risk of B-ALL. A relatively large number of studies have evaluated the association between *IKZF1* rs4132601 polymorphism and ALL susceptibility, but the results have been contradictory. Therefore, we performed a meta-analysis including the results of 13 studies with a total of 5953 children with ALL and 9807 controls. The results showed that GG genotype of rs4132601 at *IKZF1* increased the risk of ALL. Furthermore, when the analyses were restricted to the B-ALL group (4693 children with B-ALL), we found a higher risk for B-ALL, indicating a higher effect in this subgroup. In addition, in our study rs4132601 was associated with B-hyperdiploid subtype. This is in line with results reported by Chokkalingam et al (Chokkalingam et al. 2013), which found association in a Hispanic population. *IKZF1* is a transcriptional factor involved in lymphoid differentiation, and this SNP, located in 3´UTR region of *IKZF1*, was associated with a lower expression of the gene (Papaemmanuil et al. 2009), however the functional significance of the *IKZF1* polymorphism was not fully elucidated. Recently, thanks to the publication of ENCyclopedia of DNA Elements (ENCODE) (Consortium 2012), which aims to identify all functional elements in the human genome sequence, we have found that rs4132601 is located in a miRNA binding site, in which the G allele creates a binding for mir-4772 and mir-3937 (Gong et al. 2015). The creation of these miRNA binding sites could decrease IKZF1 expression previously described. Therefore, the risk allele G of *IKZF1* creates a miRNA binding site for mir-4772 and mir-3937, and the miRNAs could downregulate the expression of IKZF1.

At *CEBPE* gene, TT genotype of rs2239633 showed the highest association signal, decreasing the risk of B-ALL. This association has been replicated in several studies, but not in others. Our meta-analysis provides a robust evidence for association of the TT genotype with a decrease of ALL risk, as well as with B-ALL risk when the analysis was restricted to this subgroup. In addition, in our study rs2239633 was associated with B-hyperdiploid subtype, which is in agreement with the results reported by Chokkalingam et al., Hsu et al., and Walsh et al. (Chokkalingam et al. 2013, Hsu et al. 2015, Walsh et al. 2013), which found association in the Hispanic population. Regarding rs2239635 polymorphism, CC genotype was associated with an

increased risk of B-ALL, and a higher risk for B-hyperdiploid, which confirm the results of Wiemels et al. (Wiemels et al. 2016). Moreover, the CC genotype of rs2239635 was also associated with *ETV6-RUNX1* subtype, although the results have to be taken with caution due to the low number of patients in this group. Remarkably, Wiemels et al. in their study showed for the first time an interaction between *CEBPE* and *IKZF1*, results that have been replicated in the present study. The risk allele of rs2239635 was shown to disrupt the binding of Ikaros, answering the question as to why *CEBPE*, a critical modulator of mielopoiesis not required for B-cell maturation or function, is involved in B-ALL (Wiemels et al. 2016). Wiemels et al. suggested that lineage commitment for pre-B cells involves the suppression of *CEBPE* by Ikaros, and a polymorphism disrupting this repression promotes B-ALL risk. Incomplete suppression of *CEBPE* by Ikaros may lead to lineage confusion, a common feature of leukemogenesis (Yamanaka et al. 1997).

In conclusion, we have validated the association of rs4132601 at *IKZF1* and rs2239633 and rs2239635 at *CEBPE* and B-ALL susceptibility in the Spanish population and the present meta-analyses indicates that rs4132601 and rs2239633 SNPs are genetic risk factors for B-ALL susceptibility in the different populations. In addition we have confirmed the interaction of *IKZF1* and *CEBPE* genes. These results support the involvement of these genes in B-ALL development.

ACKNOWLEDGMENTS

This study was funded by the Basque Government (IT661-13), UPV/EHU (UFI11/35) and RTICC (RD12/0036/0060, RD12/0036/0036). AGC was supported by a pre-doctoral grant from the Basque Government.

CONFLICT OF INTEREST STATEMENT

None declared

REFERENCES

Akasaka, T., Balasas, T., Russell, L. J., Sugimoto, K. J., Majid, A., Walewska, R., Karran, E. L., Brown, D. G., Cain, K., Harder, L., Gesk, S., Martin-Subero, J. I., Atherton, M. G., Brüggemann, M., Calasanz, M. J., Davies, T., Haas, O. A., Hagemeijer, A., Kempski, H., Lessard, M., Lillington, D. M., Moore, S., Nguyen-Khac, F., Radford-Weiss, I., Schoch, C., Struski, S., Talley, P., Welham, M. J., Worley, H., Strefford, J. C., Harrison, C. J., Siebert, R. and Dyer, M. J. (2007) 'Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL)', *Blood,* 109(8), 3451-61.

Bahari, G., Hashemi, M., Naderi, M. and Taheri, M. (2016) 'IKZF1 gene polymorphisms increased the risk of childhood acute lymphoblastic leukemia in an Iranian population', *Tumour Biol,* 37(7), 9579-86.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the false discovery rate: A practical and powerful approach to multiple testing', Soc Series B(

Bilbao-Aldaiturriaga, N., Askaiturrieta, Z., Granado-Tajada, I., Goričar, K., Dolžan, V., For The Slovenian Osteosarcoma Study Group, Garcia-Miguel, P., Garcia de Andoin, N., Martin-Guerrero, I. and Garcia-Orad, A. (2016) 'A systematic review and meta-analysis of MDM2 polymorphisms in osteosarcoma susceptibility', *Pediatr Res,* 80(4), 472-9.

Chokkalingam, A. P., Hsu, L. I., Metayer, C., Hansen, H. M., Month, S. R., Barcellos, L. F., Wiemels, J. L. and Buffler, P. A. (2013) 'Genetic variants in ARID5B and CEBPE are childhood ALL susceptibility loci in Hispanics', *Cancer Causes Control,* 24(10), 1789-95.

Consortium, E. P. (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature,* 489(7414), 57-74.

Dai, Y. E., Tang, L., Healy, J. and Sinnett, D. (2014) 'Contribution of polymorphisms in IKZF1 gene to childhood acute leukemia: a meta-analysis of 33 case-control studies', *PLoS One,* 9(11), e113748.

Egger, M., Davey Smith, G., Schneider, M. and Minder, C. (1997) 'Bias in meta-analysis detected by a simple, graphical test', *BMJ,* 315(7109), 629-34.

Ellinghaus, E., Stanulla, M., Richter, G., Ellinghaus, D., te Kronnie, G., Cario, G., Cazzaniga, G., Horstmann, M., Panzer Grümayer, R., Cavé, H., Trka, J., Cinek, O., Teigler-Schlegel, A., ElSharawy, A., Häsler, R., Nebel, A., Meissner, B., Bartram, T., Lescai, F., Franceschi, C., Giordan, M., Nürnberg, P., Heinzow, B., Zimmermann, M., Schreiber, S., Schrappe, M. and Franke, A. (2012) 'Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia', *Leukemia,* 26(5), 902-9.

Emerenciano, M., Barbosa, T. C., Lopes, B. A., Blunck, C. B., Faro, A., Andrade, C., Meyer, C., Marschalek, R., Pombo-de-Oliveira, M. S. and Leukemia, B. C. S. G. o. I. A. (2014) 'ARID5B polymorphism confers an increased risk to acquire specific MLL rearrangements in early childhood leukemia', *BMC Cancer,* 14, 127.

Gharbi, H., Ben Hassine, I., Soltani, I., Safra, I., Ouerhani, S., Bel Haj Othmen, H., Teber, M., Farah, A., Amouri, H., Toumi, N. H., Abdennebi, S., Abbes, S. and Menif, S. (2016) 'Association of genetic variation in IKZF1, ARID5B, CDKN2A, and CEBPE with the risk of acute lymphoblastic leukemia in Tunisian children and their contribution to racial differences in leukemia incidence', *Pediatr Hematol Oncol,* 33(3), 157-67.

Gong, J., Liu, C., Liu, W., Wu, Y., Ma, Z., Chen, H. and Guo, A. Y. (2015) 'An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools', *Database (Oxford),* 2015, bav029.

Greaves, M. (2006) 'Infection, immune responses and the aetiology of childhood leukaemia', *Nat Rev Cancer,* 6(3), 193-203.

Healy, J., Richer, C., Bourgey, M., Kritikou, E. A. and Sinnett, D. (2010) 'Replication analysis confirms the association of ARID5B with childhood B-cell acute lymphoblastic leukemia', *Haematologica,* 95(9), 1608-11.

Houlston, R. S. (2010) 'Low-penetrance susceptibility to hematological malignancy', *Curr Opin Genet Dev,* 20(3), 245-50.

Hsu, L. I., Chokkalingam, A. P., Briggs, F. B., Walsh, K., Crouse, V., Fu, C., Metayer, C., Wiemels, J. L., Barcellos, L. F. and Buffler, P. A. (2015) 'Association of genetic variation in IKZF1, ARID5B, and CEBPE and surrogates for early-life infections with the risk of acute lymphoblastic leukemia in Hispanic children', *Cancer Causes Control,* 26(4), 609-19.

Hungate, E. A., Vora, S. R., Gamazon, E. R., Moriyama, T., Best, T., Hulur, I., Lee, Y., Evans, T. J., Ellinghaus, E., Stanulla, M., Rudant, J., Orsi, L., Clavel, J., Milne, E., Scott, R. J., Pui, C. H., Cox, N. J., Loh, M. L., Yang, J. J., Skol, A. D. and Onel, K. (2016) 'A variant at 9p21.3 functionally implicates CDKN2B in paediatric B-cell precursor acute lymphoblastic leukaemia aetiology', *Nat Commun,* 7, 10635.

Kreile, M., Piekuse, L., Rots, D., Dobele, Z., Kovalova, Z. and Lace, B. (2016) 'Analysis of possible genetic risk factors contributing to development of childhood acute lymphoblastic leukaemia in the Latvian population', *Arch Med Sci,* 12(3), 479-85.

Lautner-Csorba, O., Gézsi, A., Semsei, A. F., Antal, P., Erdélyi, D. J., Schermann, G., Kutszegi, N., Csordás, K., Hegyi, M., Kovács, G., Falus, A. and Szalai, C. (2012) 'Candidate gene association study in pediatric acute lymphoblastic leukemia evaluated by Bayesian network based Bayesian multilevel analysis of relevance', *BMC Med Genomics,* 5, 42.

Li, S., Ren, L., Fan, L. and Wang, G. (2015) 'IKZF1 rs4132601 polymorphism and acute lymphoblastic leukemia susceptibility: a meta-analysis', *Leuk Lymphoma,* 56(4), 978-82.

Lin, C. Y., Li, M. J., Chang, J. G., Liu, S. C., Weng, T., Wu, K. H., Yang, S. F., Huang, F. K., Lo, W. Y. and Peng, C. T. (2014) 'High-resolution melting analyses for genetic variants in ARID5B and IKZF1 with childhood acute lymphoblastic leukemia susceptibility loci in Taiwan', *Blood Cells Mol Dis,* 52(2-3), 140-5.

Migliorini, G., Fiege, B., Hosking, F. J., Ma, Y., Kumar, R., Sherborne, A. L., da Silva Filho, M. I., Vijayakrishnan, J., Koehler, R., Thomsen, H., Irving, J. A., Allan, J. M., Lightfoot, T., Roman, E., Kinsey, S. E., Sheridan, E., Thompson, P., Hoffmann, P., Nöthen, M. M., Mühleisen, T. W., Eisele, L., Zimmermann, M., Bartram, C. R., Schrappe, M., Greaves, M., Stanulla, M., Hemminki, K. and Houlston, R. S. (2013) 'Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype', *Blood,* 122(19), 3298-307.

Orsi, L., Rudant, J., Bonaventure, A., Goujon-Bellec, S., Corda, E., Evans, T. J., Petit, A., Bertrand, Y., Nelken, B., Robert, A., Michel, G., Sirvent, N., Chastagner, P., Ducassou, S., Rialland, X., Hémon, D., Milne, E., Scott, R. J., Baruchel, A. and Clavel, J. (2012) 'Genetic polymorphisms and childhood acute lymphoblastic leukemia: GWAS of the ESCALE study (SFCE)', *Leukemia,* 26(12), 2561-4.

Papaemmanuil, E., Hosking, F. J., Vijayakrishnan, J., Price, A., Olver, B., Sheridan, E., Kinsey, S. E., Lightfoot, T., Roman, E., Irving, J. A., Allan, J. M., Tomlinson, I. P., Taylor, M., Greaves, M. and Houlston, R. S. (2009) 'Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia', *Nat Genet,* 41(9), 1006-10.

Pastorczak, A., Górniak, P., Sherborne, A., Hosking, F., Trelińska, J., Lejman, M., Szczepański, T., Borowiec, M., Fendler, W., Kowalczyk, J., Houlston, R. S. and Młynarski, W. (2011) 'Role of 657del5 NBN mutation and 7p12.2 (IKZF1), 9p21 (CDKN2A), 10q21.2 (ARID5B) and 14q11.2 (CEBPE) variation and risk of childhood ALL in the Polish population', *Leuk Res,* 35(11), 1534-6.

Prasad, R. B., Hosking, F. J., Vijayakrishnan, J., Papaemmanuil, E., Koehler, R., Greaves, M., Sheridan, E., Gast, A., Kinsey, S. E., Lightfoot, T., Roman, E., Taylor, M., Pritchard-Jones, K., Stanulla, M., Schrappe, M., Bartram, C. R., Houlston, R. S., Kumar, R. and Hemminki, K. (2010) 'Verification of the susceptibility loci on 7p12.2, 10q21.2, and 14q11.2 in precursor B-cell acute lymphoblastic leukemia of childhood', *Blood,* 115(9), 1765-7.

Pui, C. H. and Evans, W. E. (2006) 'Treatment of acute lymphoblastic leukemia', *N Engl J Med,* 354(2), 166-78.

Ross, J. A., Linabery, A. M., Blommer, C. N., Langer, E. K., Spector, L. G., Hilden, J. M., Heerema, N. A., Radloff, G. A., Tower, R. L. and Davies, S. M. (2013) 'Genetic variants modify susceptibility to leukemia in infants: a Children's Oncology Group report', *Pediatr Blood Cancer,* 60(1), 31-4.

Sambrook, J. and Russell, D. W. (2001) 'Preparation and Analysis of Eukaryotic Genomic DNA' in *Molecular cloning: A laboratory manual*, 3rd edition ed., Cold Spring Harbor, USA: CSHL Press, 4-12.

Sherborne, A. L., Hosking, F. J., Prasad, R. B., Kumar, R., Koehler, R., Vijayakrishnan, J., Papaemmanuil, E., Bartram, C. R., Stanulla, M., Schrappe, M., Gast, A., Dobbins, S. E., Ma, Y., Sheridan, E., Taylor, M., Kinsey, S. E., Lightfoot, T., Roman, E., Irving, J. A., Allan, J. M., Moorman, A. V., Harrison, C. J., Tomlinson, I. P., Richards, S., Zimmermann, M., Szalai, C., Semsei, A. F., Erdelyi, D. J., Krajinovic, M., Sinnett, D., Healy, J., Gonzalez Neira, A., Kawamata, N., Ogawa, S., Koeffler, H. P., Hemminki, K., Greaves, M. and Houlston, R. S. (2010) 'Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk', *Nat Genet,* 42(6), 492-4.

Treviño, L. R., Yang, W., French, D., Hunger, S. P., Carroll, W. L., Devidas, M., Willman, C., Neale, G., Downing, J., Raimondi, S. C., Pui, C. H., Evans, W. E. and Relling, M. V. (2009) 'Germline genomic variants associated with childhood acute lymphoblastic leukemia', *Nat Genet,* 41(9), 1001-5.

Vijayakrishnan, J., Sherborne, A. L., Sawangpanich, R., Hongeng, S., Houlston, R. S. and Pakakasama, S. (2010) 'Variation at 7p12.2 and 10q21.2 influences childhood acute lymphoblastic leukemia risk in the Thai population and may contribute to racial differences in leukemia incidence', *Leuk Lymphoma,* 51(10), 1870-4.

Walsh, K. M., Chokkalingam, A. P., Hsu, L. I., Metayer, C., de Smith, A. J., Jacobs, D. I., Dahl, G. V., Loh, M. L., Smirnov, I. V., Bartley, K., Ma, X., Wiencke, J. K., Barcellos, L. F., Wiemels, J. L. and Buffler, P. A. (2013) 'Associations between genome-wide Native American ancestry, known risk alleles and B-cell ALL risk in Hispanic children', *Leukemia,* 27(12), 2416-9.

Wang, C., Chen, J., Sun, H., Sun, L. and Liu, Y. (2015) 'CEBPE polymorphism confers an increased risk of childhood acute lymphoblastic leukemia: a meta-analysis of 11 case-control studies with 5,639 cases and 10,036 controls', *Ann Hematol,* 94(2), 181-5.

Wang, Y., Chen, J., Li, J., Deng, J., Rui, Y., Lu, Q., Wang, M., Tong, N., Zhang, Z. and Fang, Y. (2013) 'Association of three polymorphisms in ARID5B, IKZF1 and CEBPE with the risk of childhood acute lymphoblastic leukemia in a Chinese population', *Gene,* 524(2), 203-7.

Wiemels, J. L., de Smith, A. J., Xiao, J., Lee, S. T., Muench, M. O., Fomin, M. E., Zhou, M., Hansen, H. M., Termuhlen, A., Metayer, C. and Walsh, K. M. (2016) 'A functional polymorphism in the CEBPE gene promoter influences acute lymphoblastic leukemia risk through interaction with the hematopoietic transcription factor Ikaros', *Leukemia,* 30(5), 1194-7.

Xu, H., Yang, W., Perez-Andreu, V., Devidas, M., Fan, Y., Cheng, C., Pei, D., Scheet, P., Burchard, E. G., Eng, C., Huntsman, S., Torgerson, D. G., Dean, M., Winick, N. J., Martin, P. L., Camitta, B. M., Bowman, W. P., Willman, C. L., Carroll, W. L., Mullighan, C. G., Bhojwani, D., Hunger, S. P., Pui, C. H., Evans, W. E., Relling, M. V., Loh, M. L. and Yang, J. J. (2013) 'Novel Susceptibility Variants at 10p12.31-12.2 for Childhood Acute Lymphoblastic Leukemia in Ethnically Diverse Populations', *J Natl Cancer Inst*.

Yamanaka, R., Barlow, C., Lekstrom-Himes, J., Castilla, L. H., Liu, P. P., Eckhaus, M., Decker, T., Wynshaw-Boris, A. and Xanthopoulos, K. G. (1997) 'Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein epsilon-deficient mice', *Proc Natl Acad Sci U S A,* 94(24), 13187-92.

Submitted to Leukemia and Lymphoma

Involvement of SNPs in CDKN2A/B locus in Acute Lymphoblastic Leukemia susceptibility

Gutierrez-Camino, A[1]; Martín-Guerrero, I [1]; García de Andoin, N[2,3]; Sastre, A[4]; Carboné Bañeres, A[5]; Astigarraga, I[6,7]; Navajas, A[7]; García-Orad, Á[1, 7].

[1]Department of Genetics, Physic Anthropology and Animal Physiology, University of the Basque Country, UPV/EHU, Leioa, Spain; [2]Department of Paediatrics, University Hospital Donostia, San Sebastian, Spain; [3]BioDonostia Health Research Institute, San Sebastian, Spain; [4]Department of Oncohematology, University Hospital La Paz, Madrid, Spain; [5]Department of Paediatrics, University Hospital Miguel Servet, Zaragoza, Spain; [6]Department of Paediatrics, University Hospital Cruces, Barakaldo, Spain; [7]BioCruces Health Research Institute, Barakaldo, Spain.

Abstract

CDKN2A/B (9p21.3) locus has been repeatedly associated with childhood acute lymphoblastic leukemia (ALL) susceptibility in several genome wide association studies (GWAS). Despite the clear and evident association of this locus with B-ALL susceptibility, the variants associated in the diverse studies are different and in low linkage disequilibrium (LD). This may be due to the fact that different variants in each population could alter CDKN2A/B locus function through diverse mechanisms. Therefore, the aim of this study was to determine the involvement of SNPs in the CDKN2A/B locus in the susceptibility of B-ALL in a Spanish population. We analysed 6 SNPs in CDKN2A/B locus in blood samples of 217 paediatric patients with B-cell ALL in complete remission and 330 healthy controls. The SNPs rs2811712, rs3731249, rs3217992 and rs2811709 were associated with B-ALL susceptibility**.** In addition, rs2811712 was associated with B-hyperdiploid ALL. All of them remained statistically significant after FDR correction. These results provide evidence for the influence of genetic variants at CDKN2A/B locus with the risk of developing B-ALL.

Keywords: SNP, CDKN2A/B, acute lymphoblastic leukaemia, susceptibility

INTRODUCTION

Acute lymphoblastic leukemia (ALL) is the most common childhood malignancy (Pui and Evans 2006, Greaves 2006). The genetic basis of ALL susceptibility is broadly supported by, on the one hand, its association with certain congenital abnormalities (Xu et al. 2013) and, on the other hand, by genome-wide association studies (GWAS). The two first GWAS identified independently three loci associated with childhood ALL susceptibility: 10q21.2 (*ARID5B*), 7p12.2 (*IKZF1*) (Treviño et al. 2009, Papaemmanuil et al. 2009) and 14q11.2 (*CEBPE*) (Papaemmanuil et al. 2009), results widely validated (Gutiérrez-Camino et al. 2013, Prasad et al. 2010, Vijayakrishnan et al. 2010, Xu et al. 2013). Some of these loci were associated with specific genetic subtypes of ALL, such as locus 10q21.2 (*ARID5B*) and B-hyperdiploid ALL (Treviño et al. 2009, Papaemmanuil et al. 2009). Subsequent GWASs discovered additional susceptibility loci at 10p12.2 (*BMI1-PIP4K2A*) (Xu et al. 2013), validated in some populations (Migliorini et al. 2013), but no in others (Lopez-Lopez et al. 2013), and 9p21.3 (*CDKN2A/B*) (Sherborne et al. 2010).

The locus 9p21.3 (*CDKN2A/B*) is particularly noteworthy since it is deleted in around 30% of childhood ALL patients (Walsh et al. 2015), suggesting the involvement of the genes of this region in leukemogenesis. This region comprises *CDKN2A* and *CDKN2B* genes and a long noncoding RNA (lncRNA) known as *ANRIL* (or *CDKN2B-AS*). *CDKN2A* codifies for INK4-class cyclin dependent kinase (CDK) inhibitors $p16^{INK4A}$ and $p14^{ARF}$ (Iacobucci et al. 2011). These proteins are tumour suppressors that block cell cycle division during the G1/S phase and inhibit *MDM2*, respectively. The second gene *CDKN2B* encodes for the tumour suppressor $p15^{INK4B}$, which is also a cyclin kinase inhibitor. Finally, *ANRIL* has widespread influences on gene expression, impacting the cell cycle by regulating the expression of tumour suppressors $p14^{ARF}$, $p15^{INK4B}$ and $p16^{INK4A}$ (Congrains et al. 2013).

Despite the clear and evident association of this locus CDKN2A/B with B-ALL susceptibility, the variants associated in the diverse GWAS are different. The first variant identified in children from the United Kingdom in 2010 was rs3731217 (Sherborne et al. 2010), which is located in intron 1 of *CDKN2A.*This association was replicated in several populations such as Germany, Canada (Sherborne et al. 2010) and France (Orsi et al. 2012), but not in others like Poland (Pastorczak et al. 2011), Hispanic (Chokkalingam et al. 2013) or Thai population (Vijayakrishnan et al. 2010). In 2012, Orsi et al. (Orsi et al. 2012) also associated one variant located in intron 1 of *CDKN2A,* rs2811709, with B-ALL in French children, a variant in low linkage disequilibrium (LD) with rs3731217 (r^2<0.8). In a posterior GWAS in 2013, rs17756311 located in *ANRIL*, was

identified as the highest associated variant with B-ALL in European Americans, but not in African or Hispanic Americans (Xu et al. 2013). In 2015, three independent studies using genotyping and imputation-based fine-mapping, pointed to rs3731249 in exon 2 of *CDKN2A* as the hit associated variant that conferred high risk for B-ALL in European and Hispanic children (Xu et al. 2015, Walsh et al. 2015, Vijayakrishnan et al. 2015).

Therefore, although there is an obvious implication of CDKN2A/B locus in B-ALL susceptibility, the variants annotated by these studies are different and are in low LD among them. This may be due to the fact that different variants in each population could alter CDKN2A/B locus function through diverse mechanisms. In fact, it has been suggested that the alleles of rs3731217 create two overlapping cis-acting intronic splice enhancer motifs (CCCAG**G** and CAG**T**AC) that may regulate alternative splicing of *CDKN2A* (Hungate et al. 2016). Regarding rs17756311, Hungate et al. found that a SNP in high LD with it (r^2>0.8), rs662463 in *ANRIL,* regulates *CDKN2B* expression by disrupting a transcription factor binding site (TFBS) for CEBPB (Hungate et al. 2016). Finally, rs3731249 is a missense SNP in *CDKN2A* which produces an alanine-to-threonine change in amino-acid-sequence, resulting in reduced tumour suppressor function of $p16^{INK4A}$ (Xu et al. 2015). Interestingly, this SNP is also located in the 3´UTR region of $p14^{ARF}$, where it creates a binding site for mir-132-5p and mir-4642 (Gong et al. 2012). Therefore, this SNP could cause the downregulation of CDKN2A/B locus. More than other 40 SNPs in 3´UTR region of *CDKN2A* and *CDKN2B* that disrupt or create miRNA binding sites have been described, suggesting their importance in CDKN2A/B regulation. However, studies focused on SNPs in miRNA binding sites are almost absent.

Therefore, the aim of this study was to determine the involvement of variants in the CDKN2A/B locus in the susceptibility of B-ALL. For this aim we have analyzed four SNPs previously proposed by the literature and SNPs in miRNA binding sites in a large cohort of Spanish children diagnosed with B-ALL.

MATERIALS AND METHODS

Study Participants

A total of 231 Caucasian children diagnosed with B-ALL between 2000 and 2011 in the Paediatric Oncology Units of four Spanish hospitals (University Hospital Cruces, University Hospital Donostia, University Hospital La Paz and University Hospital Miguel Servet) and 338 unrelated healthy controls were included in this study (Table 28).

Data were collected objectively, blinded to genotypes, from the patients' medical files. The two most common ALL subtypes, B-lineage hyperdiploid ALL with more than 50 chromosomes (B-hyperdiploid) and B-lineage ALL bearing the t(12;21)(p13;q22) translocation leading to an *ETV6-RUNX1* gene fusion, were also analyzed. The other subtypes were not considered due to the low number of patients in our cohort. Sex and age data were systematically recorded (Table 27). Informed consent was obtained from all participants, or from their parents prior to sample collection. The study was approved by the local ethics committees (PI2014039) and was carried out according to the Declaration of Helsinki.

Table 28: Characteristics of study population.

	Patients	**Controls**
No. of individuals	231	338
Mean age ± SE, y	4.04 ± 3.61	57.8 ± 28.1
Sex*		
Males, n (%)	128 (55.7)	157 (46.4)
Females, n (%)	102 (44.3)	181 (53.6)
Genetic alterations[#]		
Hyperdiploid	56 (24.2)	-
ETV6-RUNX1	37 (16.0)	-
MLL	13 (5.6)	-
BCR-ABL	6 (2.6)	-
E2A-PBX1	6 (2.6)	-
Hipodiploid	2 (0.9)	-
Other	1 (0.4)	-
No alteration	95 (41.1)	-
No available	21 (9.1)	-

SE: standard error, y: years. *There is no data for one patient. [#]Six patients have more than one alteration.

Selection of polymorphisms

A total of six SNPs at the locus 9p21.3 were selected (Table 10). Selection was done based on the following criteria: (i) four SNPs previously reported to be highly associated with ALL susceptibility in the literature or in high LD defined using the International HapMap Project (release #24; http://hapmap.ncbi.nlm.nih.gov/) (The HapMap Data Coordination Center (DCC), Bethesda, MD) and Haploview software v.4.2 (http://www.broad.mit.edu/mpg/haploview/) (Broad Institute, Cambridge, USA) with an r^2 threshold of 0.8 and a minimum minor allele frequency (MAF) of 0.10, (ii) SNPs in miRNA binding sites of 3´UTR region of *CDKN2A* and *CDKN2B* with a MAF>10% identified using bioinformatics tools: Ensembl (http://www.ensembl.org/) (Welcome Trust Genome Campus, Cambridge, UK), and miRNASNP

(http://bioinfo.life.hust.edu.cn/miRNASNP2/index.php) (College of Life Science and Technology, HUST). Among 47 SNPs identified in the 3´UTR region that disrupt or create a miRNA binding site (Table 11), only two had a MAF>10%.

Genotype analyses

Genomic DNA was extracted from remission peripheral blood or bone marrow using the phenol-chloroform method as previously described (Sambrook and Russell 2001). DNA was quantified using PicoGreen (Invitrogen Corp., Carlsbad, CA).

For each sample, 400 ng of DNA were genotyped using the GoldenGate Genotyping Assay with Veracode technology according to the published Illumina protocol. Data were analyzed with GenomeStudio software for genotype clustering and calling. Duplicate samples and CEPH trios (Coriell Cell Repository, Camden, NJ) were genotyped across the plates. For rs3731249, the genotyping analyses were performed by using PCR followed by restriction analysis with *BstUI* enzyme. Duplicates were included in each assay. The PCR products were visualized after electrophoresis on 3% agarose gels (Figure 18). Primer sequences and PCR conditions are described in detail in Table 12.

Statistical analysis

To identify any deviation in Hardy-Weinberg equilibrium (HWE) for the healthy controls, a $\chi 2$ test was used. The association between genetic polymorphisms in cases and controls, as well as ALL subtypes and controls, was also evaluated using the $\chi 2$ or Fisher's exact test. The effect sizes of the associations were estimated by the odds ratio from univariate logistic regression. The most significant test among codominant, dominant, recessive, and additive genetic models was selected. The results were adjusted for multiple comparisons using the false discovery rate (FDR) (Benjamini and Hochberg 1995). In all cases, the significance level was set at 5%.

RESULTS

Genotyping Results

A total of 231 patients with B-ALL and 338 unrelated healthy controls were available for genotyping. Successful genotyping was achieved for 217 B-ALL patients and 330 controls (96.1%). Of the SNPs, 6/6 (100%) were genotyped satisfactorily. All of them were in HWE in the control cohort.

Genotype association study of B-ALL

Of the 6 SNPs analyzed, we found 4 significantly associated with B-ALL risk (Table 29 and Figure 27). From them, rs2811712 at *CDKN2B* displayed the most significant value under the log-additive genetic model (AA vs AG vs GG). The GG genotype showed a 1.98-fold increased risk of B-ALL (CI 95%: 1.39-2.82; p=0.0001). The second most significant association signal was found for rs3731249 at *CDKN2A.* In this case, the TT genotype produced a 2.61-fold increased risk of B-ALL (CI 95%: 1.38-4.92; p=0.002). We also found AA genotype of rs3217992 associated with a decreased risk of B-ALL (OR: 0.56; CI 95%: 0.36-0.88; p=0.009). Finally, rs2811709 AA genotype was associated with a 1.7-fold increased risk of B-ALL. All the SNPs remained statistically associated with B-ALL risk after FDR correction. The SNPs rs3731222 and rs1063192 were not associated with B-ALL risk in our population.

Table 29: Association results of SNPs in CDKN2A/B and B-ALL.

Gene SNP	Genotype	N (controls) (N=330)	N (cases) (N=217)	OR (CI 95%)	P
ANRIL **rs2811712**	AA AG GG	264 (80.2) 62 (18.8) 3 (0.9)	143 (66.5) 64 (29.8) 8 (3.7)	Aditive 1.98 (1.39-2.82)	0.0001*
	A G	590 (89.7) 68 (10.3)	350 (81.4) 80 (18.6)	1.98 (1.39-2.81)	0.0001*
CDKN2A **rs3731249**	CC CT TT	217 (92.7) 16 (6.8) 1 (0.4)	142 (83) 28 (16.4) 1 (0.6)	Dominant 2.61 (1.38-4.92)	0.002*
	C T	450 (96.2) 18 (3.8)	312 (91.2) 30 (8.8)	2.4 (1.31-4.38)	0.004*
CDKN2B, ANRIL **rs3217992**	GG AG AA	95 (28.9) 153 (46.5) 81 (24.6)	72 (33.8) 108 (50.7) 33 (15.5)	Recessive 0.56 (0.36-0.88)	0.009*
	G A	343 (52.1) 315 (47.9)	252 (59.2) 174 (40.8)	0.75 (0.58-0.96)	0.023*
CDKN2A **rs2811709**	GG AG AA	203 (79.6) 49 (19.2) 3 (1.2)	145 (69.7) 59 (28.4) 4 (1.9)	Dominant 1.7 (1.11-2.59)	0.014*
	G A	455 (89.2) 55 (10.8)	349 (83.9) 67 (16.1)	1.58 (1.08-2.32)	0.017*
CDKN2A **rs3731222**	AA AG GG	246 (74.8) 78 (23.7) 5 (1.5)	165 (77.5) 44 (20.7) 4 (1.9)	Dominant 0.86 (0.57-1.29)	0.47
	A G	570 (86.6) 88 (13.4)	374 (87.8) 52 (12.2)	0.9 (0.62-1.29)	0.57
CDKN2B, ANRIL **rs1063192**	TT CT CC	125 (38.1) 162 (49.4) 41 (12.5)	86 (39.8) 98 (45.4) 32 (14.8)	Dominant 0.93 (0.65-1.32)	0.68
	T C	412 (62.8) 244 (37.2)	270 (62.5) 162 (37.5)	1.01 (0.78-1.3)	0.91

Abbreviations: CI, confidence interval; OR, odds ratio; *Significant after FDR correction.

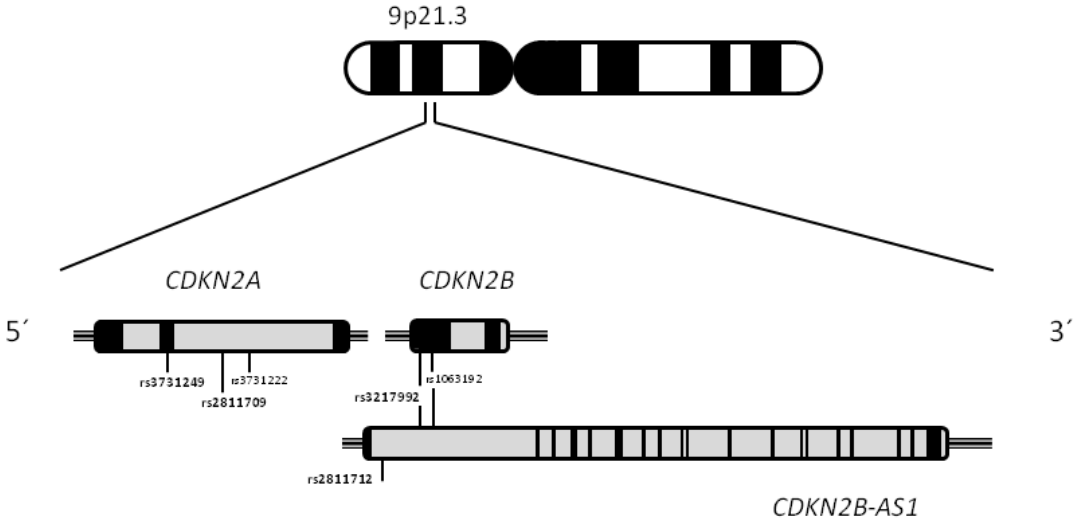

Figure 27: Diagram of CDKN2A/B locus. In bold, the SNPs significantly associated with B-ALL risk in our study.

Genotype association study of B-ALL subtypes

When we analyzed the 6 SNPs considering B-hyperdiploid ALL and *ETV-RUNX1* ALL subtype, we found association between TT genotype of rs3731249 and B-hyperdiploid ALL (OR:2.62; CI 95%:1.06-6.48; p=0.048), AA genotype of rs3217992 with *ETV6-RUNX1* ALL (OR:0.58; CI 95%:0.34-0.96; p=0.03) and GG genotype of rs2811712 with both B-hyperdiploid (OR:8.69; CI 95%:1.89-40.0; p=0.007) and *ETV6-RUNX1* ALL (OR:2.4; CI 95%:1.15-5.02; p=0.024) (Table 30). After FDR correction, the association between rs2811712 and B-hyperdiploid ALL remained statistically significant (p=0.042).

Table 30: Association results of SNPs in CDKN2A/B and B-hyperdiploid ALL and *ETV6-RUNX1* ALL.

			B-hyperdiploid ALL			ETV6-RUNX1 ALL		
Gene SNP	**Genotype**	**N (controls) (N=330)**	**N (cases) (N=54)**	**OR (CI 95%)**	**P**	**N (cases) (N=54)**	**OR (CI 95%)**	**P**
***ANRIL* rs2811712**	AA AG GG	264 (80.2) 62 (18.8) 3 (0.9)	40 (74.1) 10 (18.5) 4 (7.4)	Recessive 8.69 (1.89-40.0)	0.007*	22 (62.9) 12 (34.3) 1 (2.9)	Dominant 2.4 (1.15-5.02)	0.024
	A G	590 (91.2) 68 (8.8)	90 (90.7) 18 (9.3)	1.73 (0.98-3.05)	0.055	56 (80) 14 (20)	2.16 (1.14-4.1)	0.017
***CDKN2A* rs3731249**	CC CT TT	217 (92.7) 16 (6.8) 1 (0.4)	39 (83) 8 (17) 0	Dominant 2.62 (1.06-6.48)	0.048	28 (90.3) 3 (9.7) 0	Dominant 1.37 (0.38-4.96)	0.64
	C T	450 (96.2) 18 (3.8)	86 (91.5) 8 (8.5)	2.32 (0.98-5.51)	0.55	59 (95.2) 3 (4.8)	1.27 (0.36-4.44)	0.70
***CDKN2A* rs2811709**	GG AG AA	203 (79.6) 49 (19.2) 3 (1.2)	38 (70.4) 13 (24.1) 3 (5.6)	Additive 1.72 (0.98-3.01)	0.06	23 (67.6) 11 (32.4) 0	Dominant 1.87 (0.86-4.07)	0.12
	G A	455 (89.2) 55 (10.8)	89 (82.4) 19 (17.6)	1.76 (0.99-3.11)	0.05	57 (83.8) 11 (16.2)	1.59 (0.79-3.22)	0.19
***CDKN2B, ANRIL* rs1063192**	TT CT CC	125 (38.1) 162 (49.4) 41 (12.5)	23 (42.6) 24 (44.4) 7 (13)	Dominant 0.83 (0.46-1.49)	0.53	12 (34.3) 17 (48.6) 6 (17.1)	Dominant 1.18 (0.57-2.46)	0.65
	T C	412 (62.8) 244 (37.2)	70 (64.8) 38 (35.2)	0.91 (0.59-1.4)	0.68	41 (58.8) 29 (41.4)	1.19 (0.72-1.97)	0.48
***CDKN2A* rs3731222**	AA AG GG	246 (74.8) 78 (23.7) 5 (1.5)	41 (77.4) 12 (22.6) 0	Dominant 0.87 (0.44-1.73)	0.68	27 (77.1) 7 (20) 1 (2.9)	Dominant 0.88 (0.38-2.01)	0.75
	A G	570 (86.6) 88 (13.4)	94 (88.7) 12 (11.3)	0.82 (0.43-1.57)	0.56	61 (87.1) 9 (12.9)	0.95 (0.45-1.99)	0.90
***CDKN2B, ANRIL* rs3217992**	GG AG AA	95 (28.9) 153 (46.5) 81 (24.6)	17 (31.5) 24 (44.4) 13 (24.1)	Recessive 0.97 (0.5-1.9)	0.93	15 (44.1) 15 (44.1) 4 (11.8)	Aditive 0.58 (0.34-0.96)	0.030
	G A	343 (52.1) 315 (47.9)	58 (53.7) 50 (46.3)	0.93 (0.62-1.41)	0.76	45 (66.2) 23 (33.8)	0.55 (0.32-0.94)	0.028

Abbreviations: CI, confidence interval; OR, odds ratio; *Significant after FDR correction.

DISCUSSION

In the current study, we analyzed 6 SNPs at the CDKN2A/B locus in 217 children with B-ALL and 330 controls in a Spanish cohort. SNPs rs2811712, rs3731249, rs3217992 and rs2811709 were associated with B-ALL susceptibility. In the subtype analysis, rs2811712 was associated with the risk of developing B-hyperdiploid.

The most significant finding was the association between G allele of rs2811712 and the increased risk of developing B-ALL. This result is in line with most of the studies (Sherborne et al. 2010, Chokkalingam et al. 2013, Migliorini et al. 2013, Hungate et al. 2016). The SNP rs2811712 is in strong LD with two of the previously reported SNPs, rs17756311 (r^2=0.83) and rs662463 (r^2=1), both associated with B-ALL risk in European Americans (Xu et al. 2013, Hungate et al. 2016) and African Americans (Hungate et al. 2016), respectively. In the subtype analysis, GG genotype of rs2811712 was also associated with B-hyperdiploid ALL after FDR correction, result that is in line with Chokkalingam et al., study performed in Hispanics (Chokkalingam et al. 2013). These associations could be explained considering that rs2811712

is located in intron 1 of the lncRNA *ANRIL.* SNPs in lncRNA may affect its expression or its structure. In this case, this SNP could interfere with lncRNA folding or by modulating its protein-lncRNA interactions (Boon et al. 2016). *ANRIL* has been shown to regulate *CDKN2A* and *CDKN2B* genes. Specifically, acting in cis, ANRIL binds various Polycomb proteins resulting in histone modification of the *CDKN2A/CDKN2B* locus, and in turn, silencing the cluster (Meseure et al. 2016). In fact, the G allele of rs2811712 was shown to decrease CDKN2B mRNA levels (Consortium 2013). Therefore, the G allele of rs2811712 in *ANRIL* could be involved in the downregulation of the locus, contributing to increased susceptibility to B-ALL.

The second most significant association was found for the T allele of rs3731249, which produced a 2.6-fold increased risk of B-ALL. This association was also described recently by 3 independent studies, all of them pointing out the high impact of this variant, since it confers in all studies between two and three-fold increased risk of B-ALL susceptibility in children of European and Hispanic origin (Xu et al. 2015, Walsh et al. 2015, Vijayakrishnan et al. 2015). rs3731249 localizes to exon 2 of *CDKN2A*, being shared by both $p16^{INK4A}$ and $p14^{ARF}$. For the $p16^{INK4A}$, the C-to-T nucleotide substitution resulted in an alanine-to-threonine change (p.A148T). There is evidence that the variant $p16^{INK4A}$ (p.148T) is preferentially retained in the nucleus, compromising its ability to inhibit CDK4 and CDK6 in the cytoplasm (Xu et al. 2015) and favouring proliferation, and therefore contributing to the association of ALL risk. In $p14^{ARF}$, rs3731249 is in the 3´UTR region, where the risk allele creates a miRNA binding site for mir132-5p and mir4642 (Gong et al. 2012). These miRNAs could downregulate $p14^{ARF}$ expression, and then, attenuate its function as cyclin inhibitor. Therefore, T allele of rs3731249 in *CDKN2A* could be involved in B-ALL through its effect on the function of both $p16^{INK4A}$ and $p14^{ARF}$.

The third finding was the association between the A allele of rs3217992 and a decreased B-ALL risk. This SNP is located in a miRNA binding site in *CDKN2B* in which the A allele disrupts the binding for mir-138 and mir-205 (Gong et al. 2015). The loss of binding of these miRNAs could increase *CDKN2B* expression, explaining its protective role. As far as we know, this is the first time that this SNP is associated with B-ALL risk.

Finally, regarding rs2811709, the A allele was associated with an increased in the risk of B-ALL susceptibility in our population. This SNP was also associated with B-ALL risk in two previous studies of children of European origin (Sherborne et al. 2010, Orsi et al. 2012). rs2811709 is a cis-eQTL for *CDKN2B*, with a decreased expression of CDKN2B mRNA for the risk allele (Consortium 2013), which could describe the involvement of rs2811709 in B-ALL.

On the other hand, we found no association between rs3731222 and rs1063192 and B-ALL susceptibility. One of them, rs3731222, is in high LD with rs3731217 (r^2=1), the SNP identified in the work performed by Sherborne et al. (Sherborne et al. 2010) and replicated in several studies (Migliorini et al. 2013, Orsi et al. 2012). However, we and others could not replicate this association (Pastorczak et al. 2011, Vijayakrishnan et al. 2010). This lack of replication could be due to differences in the variants that are involved in the disease in different populations.

In conclusion, three of the variants previously proposed by the literature, rs2811712, rs3731249 and rs2811709, and a new variant, rs3217992, are associated with B-ALL susceptibility in the Spanish cohort. These results confirmed the implication of CDKN2A/B locus in the development of B-ALL.

ACKNOWLEDGMENTS

This study was funded by the Basque Government (IT661-13), UPV/EHU (UFI11/35) and RTICC (RD12/0036/0060, RD12/0036/0036). AGC was supported by a pre-doctoral grant from the Basque Government.

CONFLICT OF INTEREST STATEMENT

None declared

REFERENCES

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the false discovery rate: A practical and powerful approach to multiple testing', Soc Series B(

Boon, R. A., Jaé, N., Holdt, L. and Dimmeler, S. (2016) 'Long Noncoding RNAs: From Clinical Genetics to Therapeutic Targets?', *J Am Coll Cardiol,* 67(10), 1214-26.

Chokkalingam, A. P., Hsu, L. I., Metayer, C., Hansen, H. M., Month, S. R., Barcellos, L. F., Wiemels, J. L. and Buffler, P. A. (2013) 'Genetic variants in ARID5B and CEBPE are childhood ALL susceptibility loci in Hispanics', *Cancer Causes Control,* 24(10), 1789-95.

Congrains, A., Kamide, K., Ohishi, M. and Rakugi, H. (2013) 'ANRIL: molecular mechanisms and implications in human health', *Int J Mol Sci,* 14(1), 1278-92.

Consortium, G. (2013) 'The Genotype-Tissue Expression (GTEx) project', *Nat Genet,* 45(6), 580-5.

Gong, J., Liu, C., Liu, W., Wu, Y., Ma, Z., Chen, H. and Guo, A. Y. (2015) 'An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools', *Database (Oxford),* 2015, bav029.

Gong, J., Tong, Y., Zhang, H. M., Wang, K., Hu, T., Shan, G., Sun, J. and Guo, A. Y. (2012) 'Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis', *Hum Mutat,* 33(1), 254-63.

Greaves, M. (2006) 'Infection, immune responses and the aetiology of childhood leukaemia', *Nat Rev Cancer,* 6(3), 193-203.

Gutiérrez-Camino, Á., López-López, E., Martín-Guerrero, I., Sánchez-Toledo, J., García de Andoin, N., Carboné Bañeres, A., García-Miguel, P., Navajas, A. and García-Orad, Á. (2013) 'Intron 3 of the ARID5B gene: a hot spot for acute lymphoblastic leukemia susceptibility', *J Cancer Res Clin Oncol,* 139(11), 1879-86.

Hungate, E. A., Vora, S. R., Gamazon, E. R., Moriyama, T., Best, T., Hulur, I., Lee, Y., Evans, T. J., Ellinghaus, E., Stanulla, M., Rudant, J., Orsi, L., Clavel, J., Milne, E., Scott, R. J., Pui, C. H., Cox, N. J., Loh, M. L., Yang, J. J., Skol, A. D. and Onel, K. (2016) 'A variant at 9p21.3 functionally implicates CDKN2B in paediatric B-cell precursor acute lymphoblastic leukaemia aetiology', *Nat Commun,* 7, 10635.

Iacobucci, I., Sazzini, M., Garagnani, P., Ferrari, A., Boattini, A., Lonetti, A., Papayannidis, C., Mantovani, V., Marasco, E., Ottaviani, E., Soverini, S., Girelli, D., Luiselli, D., Vignetti, M., Baccarani, M. and Martinelli, G. (2011) 'A polymorphism in the chromosome 9p21 ANRIL locus is associated to Philadelphia positive acute lymphoblastic leukemia', *Leuk Res,* 35(8), 1052-9.

Lopez-Lopez, E., Gutierrez-Camino, A., Martin-Guerrero, I. and Garcia-Orad, A. (2013) 'Re: novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations', *J Natl Cancer Inst,* 105(19), 1512.

Meseure, D., Vacher, S., Drak Alsibai, K., Nicolas, A., Chemlali, W., Caly, M., Lidereau, R., Pasmant, E., Callens, C. and Bieche, I. (2016) 'Expression of ANRIL - Polycomb Complexes - CDKN2A/B/ARF Genes in Breast Tumors: Identification of a Two-gene (EZH2/CBX7) Signature with Independent Prognostic Value', *Mol Cancer Res*.

Migliorini, G., Fiege, B., Hosking, F. J., Ma, Y., Kumar, R., Sherborne, A. L., da Silva Filho, M. I., Vijayakrishnan, J., Koehler, R., Thomsen, H., Irving, J. A., Allan, J. M., Lightfoot, T., Roman, E., Kinsey, S. E., Sheridan, E., Thompson, P., Hoffmann, P., Nöthen, M. M., Mühleisen, T. W., Eisele, L., Zimmermann, M., Bartram, C. R., Schrappe, M., Greaves, M., Stanulla, M., Hemminki, K. and Houlston, R. S. (2013) 'Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype', *Blood,* 122(19), 3298-307.

Orsi, L., Rudant, J., Bonaventure, A., Goujon-Bellec, S., Corda, E., Evans, T. J., Petit, A., Bertrand, Y., Nelken, B., Robert, A., Michel, G., Sirvent, N., Chastagner, P., Ducassou, S., Rialland, X., Hémon, D., Milne, E., Scott, R. J., Baruchel, A. and Clavel, J. (2012) 'Genetic polymorphisms and childhood acute lymphoblastic leukemia: GWAS of the ESCALE study (SFCE)', *Leukemia,* 26(12), 2561-4.

Papaemmanuil, E., Hosking, F. J., Vijayakrishnan, J., Price, A., Olver, B., Sheridan, E., Kinsey, S. E., Lightfoot, T., Roman, E., Irving, J. A., Allan, J. M., Tomlinson, I. P., Taylor, M., Greaves, M. and Houlston, R. S. (2009) 'Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia', *Nat Genet,* 41(9), 1006-10.

Pastorczak, A., Górniak, P., Sherborne, A., Hosking, F., Trelińska, J., Lejman, M., Szczepański, T., Borowiec, M., Fendler, W., Kowalczyk, J., Houlston, R. S. and Młynarski, W. (2011) 'Role of 657del5 NBN mutation and 7p12.2 (IKZF1), 9p21 (CDKN2A), 10q21.2 (ARID5B) and 14q11.2 (CEBPE) variation and risk of childhood ALL in the Polish population', *Leuk Res,* 35(11), 1534-6.

Prasad, R. B., Hosking, F. J., Vijayakrishnan, J., Papaemmanuil, E., Koehler, R., Greaves, M., Sheridan, E., Gast, A., Kinsey, S. E., Lightfoot, T., Roman, E., Taylor, M., Pritchard-Jones, K., Stanulla, M.,

Schrappe, M., Bartram, C. R., Houlston, R. S., Kumar, R. and Hemminki, K. (2010) 'Verification of the susceptibility loci on 7p12.2, 10q21.2, and 14q11.2 in precursor B-cell acute lymphoblastic leukemia of childhood', *Blood,* 115(9), 1765-7.

Pui, C. H. and Evans, W. E. (2006) 'Treatment of acute lymphoblastic leukemia', *N Engl J Med,* 354(2), 166-78.

Sambrook, J. and Russell, D. W. (2001) 'Preparation and Analysis of Eukaryotic Genomic DNA' in *Molecular cloning: A laboratory manual*, 3rd edition ed., Cold Spring Harbor, USA: CSHL Press, 4-12.

Sherborne, A. L., Hosking, F. J., Prasad, R. B., Kumar, R., Koehler, R., Vijayakrishnan, J., Papaemmanuil, E., Bartram, C. R., Stanulla, M., Schrappe, M., Gast, A., Dobbins, S. E., Ma, Y., Sheridan, E., Taylor, M., Kinsey, S. E., Lightfoot, T., Roman, E., Irving, J. A., Allan, J. M., Moorman, A. V., Harrison, C. J., Tomlinson, I. P., Richards, S., Zimmermann, M., Szalai, C., Semsei, A. F., Erdelyi, D. J., Krajinovic, M., Sinnett, D., Healy, J., Gonzalez Neira, A., Kawamata, N., Ogawa, S., Koeffler, H. P., Hemminki, K., Greaves, M. and Houlston, R. S. (2010) 'Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk', *Nat Genet,* 42(6), 492-4.

Treviño, L. R., Yang, W., French, D., Hunger, S. P., Carroll, W. L., Devidas, M., Willman, C., Neale, G., Downing, J., Raimondi, S. C., Pui, C. H., Evans, W. E. and Relling, M. V. (2009) 'Germline genomic variants associated with childhood acute lymphoblastic leukemia', *Nat Genet,* 41(9), 1001-5.

Vijayakrishnan, J., Henrion, M., Moorman, A. V., Fiege, B., Kumar, R., da Silva Filho, M. I., Holroyd, A., Koehler, R., Thomsen, H., Irving, J. A., Allan, J. M., Lightfoot, T., Roman, E., Kinsey, S. E., Sheridan, E., Thompson, P. D., Hoffmann, P., Nöthen, M. M., Mühleisen, T. W., Eisele, L., Bartram, C. R., Schrappe, M., Greaves, M., Hemminki, K., Harrison, C. J., Stanulla, M. and Houlston, R. S. (2015) 'The 9p21.3 risk of childhood acute lymphoblastic leukaemia is explained by a rare high-impact variant in CDKN2A', *Sci Rep,* 5, 15065.

Vijayakrishnan, J., Sherborne, A. L., Sawangpanich, R., Hongeng, S., Houlston, R. S. and Pakakasama, S. (2010) 'Variation at 7p12.2 and 10q21.2 influences childhood acute lymphoblastic leukemia risk in the Thai population and may contribute to racial differences in leukemia incidence', *Leuk Lymphoma,* 51(10), 1870-4.

Walsh, K. M., de Smith, A. J., Hansen, H. M., Smirnov, I. V., Gonseth, S., Endicott, A. A., Xiao, J., Rice, T., Fu, C. H., McCoy, L. S., Lachance, D. H., Eckel-Passow, J. E., Wiencke, J. K., Jenkins, R. B., Wrensch, M. R., Ma, X., Metayer, C. and Wiemels, J. L. (2015) 'A Heritable Missense Polymorphism in CDKN2A Confers Strong Risk of Childhood Acute Lymphoblastic Leukemia and Is Preferentially Selected during Clonal Evolution', *Cancer Res,* 75(22), 4884-94.

Xu, H., Yang, W., Perez-Andreu, V., Devidas, M., Fan, Y., Cheng, C., Pei, D., Scheet, P., Burchard, E. G., Eng, C., Huntsman, S., Torgerson, D. G., Dean, M., Winick, N. J., Martin, P. L., Camitta, B. M., Bowman, W. P., Willman, C. L., Carroll, W. L., Mullighan, C. G., Bhojwani, D., Hunger, S. P., Pui, C. H., Evans, W. E., Relling, M. V., Loh, M. L. and Yang, J. J. (2013) 'Novel Susceptibility Variants at 10p12.31-12.2 for Childhood Acute Lymphoblastic Leukemia in Ethnically Diverse Populations', *J Natl Cancer Inst*.

Xu, H., Zhang, H., Yang, W., Yadav, R., Morrison, A. C., Qian, M., Devidas, M., Liu, Y., Perez-Andreu, V., Zhao, X., Gastier-Foster, J. M., Lupo, P. J., Neale, G., Raetz, E., Larsen, E., Bowman, W. P., Carroll, W. L., Winick, N., Williams, R., Hansen, T., Holm, J. C., Mardis, E., Fulton, R., Pui, C. H., Zhang, J., Mullighan, C. G., Evans, W. E., Hunger, S. P., Gupta, R., Schmiegelow, K., Loh, M. L., Relling, M. V. and Yang, J. J. (2015) 'Inherited coding variants at the CDKN2A locus influence susceptibility to acute lymphoblastic leukaemia in children', *Nat Commun,* 6, 7553.

JNCI Journal of the National Cancer Institute Advance Access published September 6, 2013

CORRESPONDENCE

Re: Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations

Lopez-Lopez E[1]; Gutierrez-Camino A[1]; Martin-Guerrero I[1]; Garcia-Orad A[1]

[1]Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Leioa, Spain

We have read with interest the contribution by Xu and collaborators (Xu et al. 2013) regarding the study entitled "Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations".

In this article, the authors carried out the first multiethnic GWAS with pediatric acute lymphoblastic leukemia (ALL) and, besides replicating previously described loci, they have found a new association locus in *PIP4K2A* gene. This association was confirmed in three replication cohorts of different ethnicities (European/American, African/American and Asian/American).

In our group, we tried to replicate this association in a Spanish population of 191 B-ALL patients and 342 unrelated healthy controls in order to verify if this is a general mechanism involved in ALL risk.

We made a linkage disequilibrium (LD) analysis of the 4 SNPs in *PIP4K2A* that were associated with ALL risk in the article by Xu et al, using Hapmap database and Haploview v4.2 software. As the 4 SNPs were in LD with an $r^2 > 0.8$, we decided to analyze rs7088318, as a tagSNP of this association hotspot. We carried out an allele-specific PCR and carried out the genotyping with a 100% success. Both cases and controls were in Hardy-Weinberg equilibrium. We carried out an association study under the codominant, dominant, recessive and log-additive models. However, we did not replicate the statistically significant association between rs7088318 genotype and ALL risk under any of the inheritance models analyzed (Table 31).

Table 31: Association study of *PIP4K2A* rs7088318 and ALL risk.

Genotype	Controls N=342 n (%)	N cases N=191 n (%)	OR (CI 95%) Codominant (P)	OR (CI 95%) Dominant (P)	OR (CI 95%) Recessive (P)	OR (CI 95%) Additive (P)
TT	139 (40.6)	94 (49.2)	Reference	Reference	Reference	Reference
GT	164 (48)	75 (39.6)	0.68 (0.46-0.99)	0.71 (0.49-1.01)	1.01 (0.58-1.76)	0.83 (0.63-1.08)
GG	39 (11.4)	22 (11.5)	0.83 (0.46-1.5) (P=0.1)	(P=0.056)	(P=0.9)	(P=0.1)

Abbreviations: CI, confidence interval; OR, odds ratio.

Differences in results might be due to differences in genetic composition among populations. In fact, if we have a look at the risk allele frequency (RAF), in our control population it is higher (RAF=0.646) than in the control European/American population analyzed by Xu et al. (RAF=0.59), and similar to what they observed in the case European/American population (RAF=0.65). This peculiarity makes it difficult to find significant differences in frequency between cases and controls.

Therefore, we conclude that SNPs in *PIP4K2A* may be associated with pediatric ALL risk in some populations but it does not seem to be a good susceptibility marker in the Spanish population. Consequently, it does not seem to be a susceptibility marker as general as other previously proposed as SNPs in *ARID5B* or *IKZF1*, which have been replicated in multiple populations.

FUNDING

This work was supported by Basque Government (IT663-13, SAI10/03 and 2006111015), and UPV/EHU (UFI11/35). ELL was supported by a predoctoral grant from the Basque Government and “Fellowship for recent Doctors until their integration in postdoctoral programs” by the Investigation Vice-rector’s office of the UPV/EHU.

REFERENCES

Xu, H., Yang, W., Perez-Andreu, V., Devidas, M., Fan, Y., Cheng, C., Pei, D., Scheet, P., Burchard, E. G., Eng, C., Huntsman, S., Torgerson, D. G., Dean, M., Winick, N. J., Martin, P. L., Camitta, B. M., Bowman, W. P., Willman, C. L., Carroll, W. L., Mullighan, C. G., Bhojwani, D., Hunger, S. P., Pui, C. H., Evans, W. E., Relling, M. V., Loh, M. L. and Yang, J. J. (2013) 'Novel Susceptibility Variants at 10p12.31-12.2 for Childhood Acute Lymphoblastic Leukemia in Ethnically Diverse Populations', *J Natl Cancer Inst*.

nature publishing group

Clinical Investigation | Articles

Received 5 August 2013; accepted 1 January 2014; advance online publication 2 April 2014. doi:10.1038/pr.2014.43

Non-coding RNAs-related polymorphisms in pediatric acute lymphoblastic leukemia susceptibility

Gutierrez-Camino A[1]; Lopez-Lopez E[1]; Martin-Guerrero I[1]; Piñan MA[2]; Garcia-Miguel P[3]; Sanchez-Toledo J[4]; Carbone Bañeres A[5]; Uriz J[6]; Navajas A[7]; Garcia-Orad A[1].

[1]Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Leioa, Spain [2]Department of Hematology and Hemotherapy, University Hospital Cruces, Bilbao, Spain [3]Service of Pediatric Oncohematology, University Hospital La Paz, Madrid, Spain [4]Service of Pediatric Oncology and Hematology, University Hospital Vall d' Hebron, VHIR, Barcelona, Spain [5]Unit of Pediatric Oncohematology, University Hospital Miguel Servet, Zaragoza, Spain [6]Unit of Pediatric Oncohematology, University Hospital Donostia, San Sebastian, Spain [7]Unit of Pediatric Hematology/Oncology, University Hospital Cruces, Bilbao, Spain

Abstract

Evidence for an inherited genetic risk for pediatric acute lymphoblastic leukemia has been provided in several studies. Most of them focused on coding regions. However, those regions represent only 1.5% of the entire genome. In acute lymphoblastic leukemia (ALL), it has been suggested that the expression of microRNAs (miRNAs) is dysregulated, which suggests that they may have a role in ALL risk. Changes in miRNA function may occur through single-nucleotide polymorphisms (SNPs). Therefore, the aim of this study was to evaluate whether polymorphisms in premiRNAs, and/or miRNA-processing genes, contribute to a predisposition for childhood ALL. In this study, we analyzed 118 SNPs in pre-miRNAs and miRNA-processing genes in 213 B-cell ALL patients and 387 controls. We found 11 SNPs significantly associated with ALL susceptibility. These included three SNPs present in miRNA genes (*miR-612*, *miR-499*, and *miR-449b*) and eight SNPs present in six miRNA biogenesis pathway genes (*TNRC6B, DROSHA*, *DGCR8*, *EIF2C1, CNOT1*, and *CNOT6*). Among the 118 SNPs analyzed, rs12803915 in *mir-612* and rs3746444 in mir-499 exhibited a more significant association, with a *P* value <0.01. The results of this study indicate that SNP rs12803915 located in pre-mir-612, and SNP rs3746444 located in pre-mir-499, may represent novel markers of B-cell ALL susceptibility.

Keywords: miRNAs, biogenesis pathway, SNPs, ALL, susceptibility

INTRODUCTION

Acute lymphoblastic leukemia (ALL) is the most common pediatric hematological malignancy in developed countries. Its etiology is believed to be multifactorial, with both environmental and genetic risk factors being relevant (Ellinghaus et al. 2012). Recently, several studies have provided evidence for an inherited genetic risk for pediatric ALL (Treviño et al. 2009, Papaemmanuil et al. 2009). Most of these studies focused on the coding regions of these genetic components. However, this represents only ~1.5% of the entire genome, and noncoding regions of the genome have also been shown to mediate regulatory functions. For example, microRNAs (miRNAs) are a class of small noncoding RNA molecules that regulate gene expression at the post-transcriptional level by binding to the 3´untranslated region of a target gene (Ryan et al. 2010). This can lead to an inhibition of translation or enhanced degradation of a target mRNA (Figure 28). Primary double-stranded miRNA transcripts (pri-miRNA) are processed in the nucleus by microprocessor machinery, which includes DROSHA RNase and the double-stranded RNA-binding protein, DGCR8. A hairpin precursor miRNA molecule of 70–100 nucleotides (pre-miRNA) is then produced, and its translocation into the cytoplasm is facilitated by RAN GTPase and Exportin 5 (*XPO5*). In the cytoplasm, pre-miRNAs are further processed by a protein complex that includes DICER1, TRBP, EIF2C1, EIF2C2, GEMIN3, and GEMIN4, resulting in the production of mature miRNAs (Ryan et al. 2010). It has been predicted that there are more than 1,000 miRNA genes in the human genome (Vinci et al. 2013), and ~30% of human genes are regulated by miRNAs.

In the past few years, it was suggested that miRNAs in ALL are dysregulated. For example, in the study of Zhang et al., (Zhang et al. 2009) up to 171 miRNAs have been found to be differentially expressed between ALL patients and normal donors. These results suggest that dysregulation of these miRNAs may be associated with an increased risk for ALL. Changes in miRNAs function have the potential to affect the expression of a large number of genes, including genes involved in the origin and evolution of pediatric ALL (H. Zhang et al. 2009, Schotte et al. 2011). Changes in miRNA function may occur through genetic variations (Ryan et al. 2010). For example, single-nucleotide polymorphisms (SNPs) present in genes involved in miRNA processing can affect levels of miRNA expression, whereas SNPs in miRNA genes can affect miRNA biogenesis and function. There have been several polymorphisms found to be associated with other malignancies, and a recent pilot study has found an association between rs2910164 in mir-146a and ALL risk (Hasani et al. 2014).

However, despite accumulating evidence that inherited genetic variation can contribute to a predisposition for pediatric ALL and the suggested role of miRNAs in the development of this disease, as well as the role of miRNA-related polymorphisms in cancer risk, the SNPs in miRNA genes and miRNA-processing genes have not been extensively studied in association with pediatric ALL risk. Therefore, the aim of this study was to evaluate whether polymorphisms in pre-miRNAs, and/or miRNA-processing genes, contribute to a predisposition for childhood ALL.

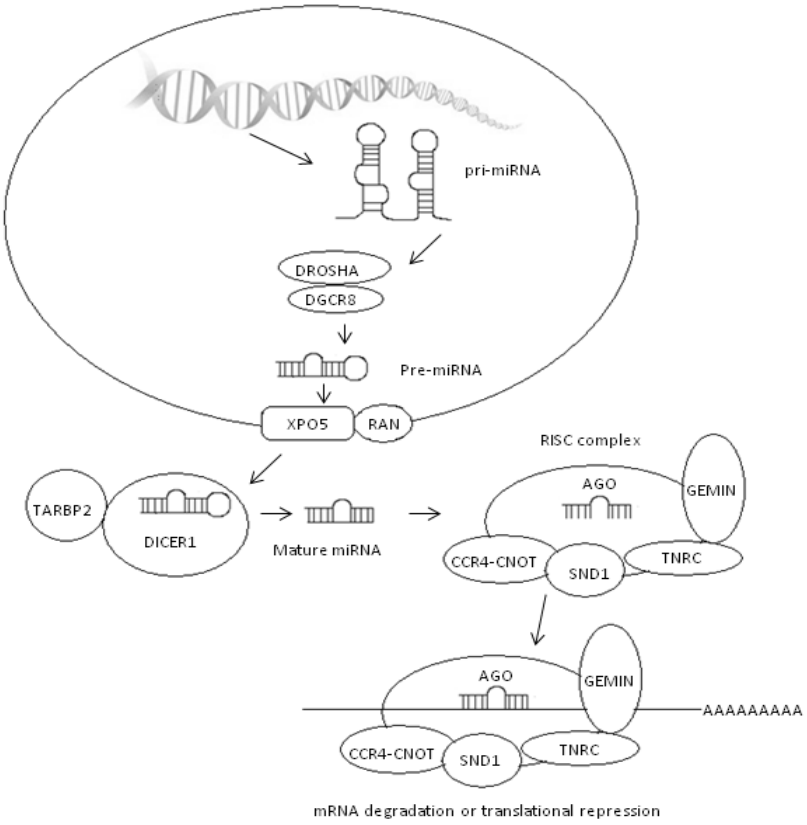

Figure 28: MicroRNA biogenesis pathway. Pre-miRNA, precursor microRNA; Pri-miRNA, primary double-stranded microRNA; RISC, RNA-induced silencing complex.

METHODS

Study Participants

A total of 213 children (1–15 years) of European origin all diagnosed with precursor B-ALL in the Pediatric Oncology Units of five Spanish hospitals (Hospital Cruces, Hospital Donostia, Hospital Vall d'Hebrón, Hospital La Paz, and Hospital Miguel Servet) were enrolled in this study. These patients were the entire incident population diagnosed and treated in the participating centers between 1995 and 2011. In addition, 387 Spanish healthy individuals of European origin with no previous history of cancer from the collection C.0001171 registered in the Institute of Health Carlos III were enrolled as controls (Table 32). Patients were classified

with B-hyperdiploid ALL if the DNA index was > 1.16 and/or the karyotype had more than 50 chromosomes. For 51 cases out of 213 B-ALL patients, cytogenetic data were not available. Informed consent was obtained from all participants, or from their parents, prior to sample collection. The study was approved by the Ethics Committee for Clinical Research and conducted in accordance with the Declaration of Helsinki.

Table 32: Characteristics of the B-ALL patients and controls examined in this study.

	Patients	Controls
No of individuals (n)	213	387
Male (n, %)	124 (56)	199 (51)
Female (n, %)	95 (43)	187 (48)
Mean age ± SE, years	5.7± 3.5	51.2±7.7
Age at diagnosis		
1 to 9 years	173 (81)	-
≥10 years	38 (18)	-
Leukocytes at diagnosis (WBC)		
<20 (x10^9/l)	94 (30)	-
20 to 200 (x10^9/l)	54 (25)	-
>200 (x10^9/l)	6 (3)	-
Hyperdiploid (n, %)	55 (26)	-
No-hyperdiploid (n, %)	107 (50)	-
Chromosomal Translocations		
ETV6-RUNX1	28(13)	-
MLL	10 (5)	-
BCR-ABL	5 (2)	-
E2A-PBX1	3 (1)	-

SE: standard error.

Selection of Genes and Polymorphisms

Twenty-one genes involved in miRNA biogenesis and processing, as determined based on available literature and the Patrocles (http:// www.patrocles.org/; University of Liege, Liège, Belgium) database, were selected for the analysis. For each gene, all of the SNPs with potential functional effects were examined using F-SNP (http:// compbio.cs.queensu.ca/F-SNP/; Queen's University, Kingston, Canada), Fast-SNP (http://fastsnp.ibms.sinica.edu.tw; Academia Sinica, Taipei, Taiwan), polymirTS (http://compbio.uthsc.edu/ miRSNP/; University of Tennessee Health Science Center, Memphis, TN), and Patrocles databases. Functional effects were considered to be those that resulted in amino acid changes and/or alternate splicing, those that were located in the promoter region of putative transcription factor–binding sites, or those that disrupted or created miRNAs binding interactions. SNPs previously included in association studies were also examined. The final selection of SNPs was made based on those having a minor allele frequency greater than 5% (i.e., ≥0.05) in European/Caucasian populations.

Considering that miRNAs can regulate a wide range of genes and that the number of polymorphisms in miRNAs was affordable, we selected all the known SNPs at the moment of the selection with a minor allele frequency > 0.01 in European/Caucasoid populations, using Patrocles and Ensembl (http://www.ensembl.org/; Welcome Trust Genome Campus, Cambridge, UK) databases and literature review.

Genotyping

Genomic DNA was extracted from remission (containing less than 5% blast cells) peripheral blood, bone marrow slides or granulocytes isolated with Ficoll-Plaque PLUS (GE Healthcare Life Sciences, Piscataway, NJ), using the phenol–chloroform method (Sambrook and Russell 2001) or from saliva samples using Oragene DNA kit (DNA Genotek, Ottawa, Canada) according to the manufacturer's instructions.

SNP genotyping was performed using TaqMan OpenArray Genotyping technology (Applied Biosystems, Carlsbad, CA) according to the published Applied Biosystems protocol. Initially, 131 SNPs were considered for analysis. After considering compatibility with the Taqman OpenArray platform, 118 SNPs were included in a Taqman OpenArray Plate (Applied Biosystems), and these included 72 SNPs present in 21 genes involved in miRNA biogenesis and 46 SNPs present in 42 pre-miRNA genes (Tables 15 and 16).

Data were analyzed using Taqman Genotyper software (Applied Biosystems) for genotype clustering and genotype calling. Duplicate samples were genotyped across the plates.

Statistical Analyses

Statistical analyses were performed using R software (version v2.14.1; Institute for Statistics and Mathematics, Wien, Austria). To identify any deviation in Hardy–Weinberg equilibrium for the healthy controls (n = 387), a $\chi 2$ test was used. The association between genetic polymorphisms in B-ALL patients and controls was also evaluated using the $\chi 2$ or Fisher's exact test. Fisher's exact test was used if a genotype class had less than five individuals. We also tested the association considering genetic characteristics (hyperdiploid subtype and chromosomal translocations). The effect sizes of the associations were estimated by the odds ratio from univariate logistic regression and multivariate logistic regression to account for the possible confounding effect of sex.

The most significant test among codominant (major allele homozygotes vs. heterozygotes and major allele homozygotes vs. minor allele homozygotes), dominant (major allele homozygotes vs. heterozygotes + minor allele homozygotes), recessive (major allele homozygotes + heterozygotes vs. minor allele homozygotes), and additive (doses dependent effect: major allele homozygotes vs. heterozygotes vs. minor allele homozygotes) genetic models was selected. In all cases, the significance level was set at 5%. The results were adjusted for multiple comparisons using the Bonferroni correction.

RESULTS

Genotyping Results

A total of 213 patients with B-cell ALL (B-ALL) and 387 unrelated healthy controls were available for genotyping. Successful genotyping was achieved for 550 DNA samples (91.67%). Among the SNPs, 106/118 (89.83%) were genotyped satisfactorily. Failed genotyping was due to an absence of PCR amplification, insufficient intensity for cluster separation, poor cluster definition, or an inability to define clusters. The average genotyping rate for all SNPs was 98.12%. Furthermore, of the 106 SNPs genotyped, 14 were not in Hardy–Weinberg equilibrium in the population of 387 healthy controls, and therefore, were not considered for further analysis. In total, 26 SNPs were excluded from the association study (Table 33), leaving 92 SNPs available for association studies.

Table 33: SNP excluded from the association study of microRNA processing genes.

SNP	Gene	Alleles	Reason for exclusion
rs1003226	*CNOT4*	T > C	Genotyping failure
rs11738060	*CNOT6*	T > A	Genotyping failure
rs34610323	*GEMIN4*	C > T	Genotyping failure
rs73239138	mir-1269	G > A	Genotyping failure
rs318039	mir-1274a	C > T	Genotyping failure
rs72631826	mir-16-1	T > C	Genotyping failure
rs72631825	mir-222	G > A	Genotyping failure
rs12197631	mir-548a-1	T > G	Genotyping failure
rs11014002	mir-603	C > T	Genotyping failure
rs2368392	mir-604	C > T	Genotyping failure
rs11061209	*RAN*	G > A	Genotyping failure
rs493760	*DROSHA*	T > C	Genotyping failure
rs42318	*CNOT3*	G > A	Absence of HWE
rs3757	*DGCR8*	G > A	Absence of HWE
rs3742330	*DICER1*	A > G	Absence of HWE
rs7813	*GEMIN4*	C > T	Absence of HWE
rs910924	*GEMIN4*	C > T	Absence of HWE
rs816736	*GEMIN5*	T > C	Absence of HWE
rs2292832	mir-149	C > T	Absence of HWE
rs174561	mir-1908	T > C	Absence of HWE
rs4919510	*mir-608*	C > G	Absence of HWE
rs11156654	mir-624	T > A	Absence of HWE
rs55656741	*DROSHA*	G > A	Absence of HWE
rs7719666	*DROSHA*	C > T	Absence of HWE
rs2413621	*TNRC6B*	T > C	Absence of HWE
rs470113	*TNRC6B*	A > G	Absence of HWE

HWE: Hardy-Weinberg equilibrium.

Analysis of Association

To investigate if genetic variation influences the risk of ALL, the 92 polymorphisms successfully genotyped were compared between cases and controls. As shown in Tables 34 and 35, statistically significant associations ($p < 0.05$) were observed for 11 polymorphisms present in miRNA-related genes. Of these, three were located in pre-miRNAs (Table 34), and eight were located in miRNA-processing genes (Table 35).

Table 34: Genotype frequencies of selected SNPs present in miRNA genes.

miRNA	SNP	Best fitting model	Genotype	Controls n (%)	Cases n (%)	OR (CI 95%)	P
mir-612	**rs12803915**	Additive	CC CT TT	232 (67.6) 100 (29.2) 11 (3.2)	152 (77.6) 42 (21.4) 2 (1.0)	0.61 (0.42–0.88)	0.007
mir-499	**rs3746444**	Additive	AA AG GG	206 (59.4) 117 (40.6) 24 (6.9)	138 (69.0) 56 (28.0) 6 (3.0)	0.67 (0.49–0.91)	0.009
mir-449b	**rs10061133**	Dominant	AA GA/GG	283 (81.8) 63 (18.2)	180 (89.6) 21 (10.4)	Reference 0.52 (0.31 –0.89)	0.012

Abbreviations: CI, confidence interval; OR, odds ratio.

Table 35: Genotype frequencies of selected SNPs present in miRNA processing genes.

Gene	SNP	Best fitting model	Genotype	Control n (%)	Case n (%)	OR (CI 95%)	P
TNRC6B	**rs139919**	Recessive	TT/CT CC	340 (99.1) 3 (0.90)	181 (95.8) 8 (4.2)	Reference 5.01 (1.31–19.11)	0.011
CNOT6	**rs6877400**	Additive	TT CT CC	274 (78.5) 68 (19.8) 7 (2.5)	173 (86.5) 26 (13.0) 1 (0.50)	0.58 (0.37–0.9)	0.011
DGCR8	**rs9606248**	Additive	AA AG GG	205 (58.9) 126 (36.3) 17 (4.9)	135 (67.8) 61 (30.7) 3 (1.5)	0.67 (0.48–0.92)	0.012
	rs1640299	Recessive	GG/GT TT	267 (76.5) 82 (23.5)	168 (84.4) 31 (15.6)	Reference 0.6 (0.38–0.95)	0.025
CNOT1	**rs11866002**	Dominant	CC CT/TT	134 (38.7) 212 (61.3)	97 (49.2) 100 (50.8)	Reference 0.65 (0.46–0.93)	0.017
DROSHA	**rs10035440**	Additive	TT CT CC	213 (62.8) 109 (32.2) 17 (5.0)	108 (54.5) 72 (36.4) 18 (9.1)	1.38 (1.04–1.83)	0.025
EIF2C1	**rs636832**	Recessive	GG/AG AA	344 (99.1) 3 (0.90)	194 (96.5) 7 (3.5)	Reference 4.14 (1.06–16.2)	0.031
	rs595961	Recessive	AA/AG GG	329 (97.1) 10 (2.9)	182 (93.3) 13 (6.7)	Reference 2.35 (1.01–5.47)	0.046

Abbreviations: CI, confidence interval; OR, odds ratio.

Among the SNPs located in pre-miRNA genes, the SNP that was the most significantly associated with ALL risk was SNP rs12803915 in premature mir-612 (Figure 29a). The A allele for this SNP was found to be protective (OR: 0.61; CI 95%: 0.42–0.88; p = 0.007) in the log-additive (GG vs. GA vs. AA) genetic model. The second most significant association involved SNP rs3746444 in mir-499 (Figure 29b). Moreover, this SNP is located in the seed region of mature miR-499-3p. The G allele of this SNP was found to be protective (OR: 0.67; CI 95%: 0.49–0.91; p = 0.009) in the log-additive (AA vs. AG vs. GG) genetic model. An association with rs10061133 in mir-449b was also identified (Table 34).

These two top SNPs, rs12803915 and rs3746444, were also studied in association with B-hyperdiploid ALL subtype (Tables 36 and 37) and chromosomal translocations, and no significant differences were found.

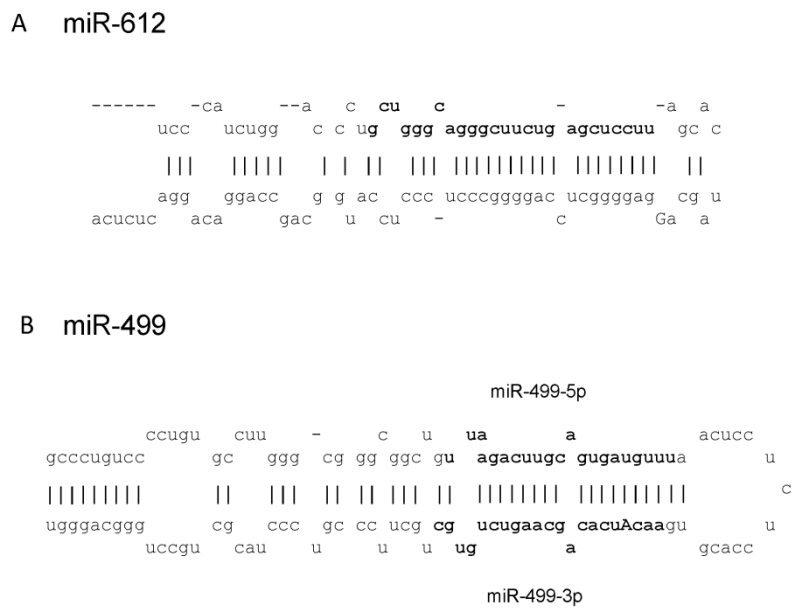

Figure 29: Schematic diagram of the hairpin loop structure of the mir-612 and mir-499. The sequence for mature microRNA is in bold. The polymorphisms sites are indicated in capital letters. (**a**) The A>G polymorphism is located in the loop of mir-612; this variation might alter the secondary structure. (**b**) The G>A polymorphism is located in the stem region opposite to the mature mir-499 sequence, which results in a change from A:U pair to G:U mismatch in the stem structure of mir-499 precursor.

Table 36: Association analysis between hyperdiploid ALL patients and non-hyperdiploid ALL patients.

miRNA	SNP	Best fitting model	Genotype	Hyperdiploid n (%)	No-Hyperdiploid n (%)	OR (CI 95%)	P
mir-612	**rs12803915**	Additive	GG GA AA	42 (77.8) 11 (20.4) 1 (1.9)	77 (77.8) 21 (21.2) 1 (1)	1.04 (0.5-2.16)	0.91
mir-499	**rs3746444**	Additive	AA AG GG	36 (66.7) 14 (25.9) 4 (7.4)	74 (74) 26 (26) 0 (0)	1.71 (0.91-3.20)	0.03

Abbreviation: PM: pre-miRNA, OR: Odd ratio, CI: Confidence interval

Table 37: Association analysis between hyperdiploid ALL patients and controls.

miRNA	SNP	Best fitting model	Genotype	Hyperdiploid n (%)	Controls n (%)	OR (CI 95%)	P
mir-612	**rs12803915**	Additive	GG GA AA	42 (77.8) 11 (20.4) 1 (1.9)	232 (67.6) 100 (29.2) 11 (3.2)	0.63 (0.34-1.17)	0.12
mir-499	**rs3746444**	Additive	AA AG GG	36 (66.7) 14 (25.9) 4 (7.4)	206 (59.4) 117 (33.7) 24 (6.9)	0.83 (0.51-1.35)	0.44

Abbreviation: OR: Odd ratio, CI: Confidence interval

In the genes of the miRNA-processing components, the most significant SNP was rs139919, a SNP located in the *TNRC6B* gene. In the recessive genetic model, the variant homozygous genotype, CC, of this SNP was associated with a 5.1-fold increase in ALL risk (CI 95%: 1.31–19.11; p = 0.011). Associations involving rs9606248 and rs1640299 in *DGCR8*, rs11866002 in *CNOT1*, rs6877400 in *CNOT6*, rs10035440 in *DROSHA*, and rs636832 and rs595961 in *EIF2C1* were also identified (Table 35).

All the SNPs remained significantly associated with ALL risk after multivariate logistic regression to account for the possible confounding effect of sex. These SNPs did not reach the significant value when Bonferroni correction was applied.

DISCUSSION

In this study, 11 SNPs were found to be significantly associated with ALL susceptibility. These included three SNPs present in miRNA genes (miR-612, miR-499, and miR-449b) and eight SNPs present in six miRNA biogenesis pathway genes (*TNRC6B*, *DROSHA*, *DGCR8*, *EIF2C1*, *CNOT1*, and *CNOT6*). Among them, rs12803915 in mir-612 and rs3746444 in mir-499 exhibited a more significant association, with a *P* value <0.01. In spite of not reaching a significant *P* value after the restrictive Bonferroni correction, our results point to a putative role of these SNPs in ALL susceptibility, which could be of lower penetrance.

SNP rs12803915 located in the premature region of mir-612 showed the strongest association with ALL risk, with the A allele being protective (p = 0.007). To our knowledge, this is the first report of this SNP being associated with cancer risk. It has been suggested that a SNP in the premature region of a miRNA could alter its secondary structure and inhibit or enhance pri-miRNA processing (Salzman and Weidhaas 2013). Therefore, SNPs in miRNAs could lead to dysregulation of miRNA expression (Ryan et al. 2010). Accordingly, it has been shown that SNP rs12803915 significantly decreases mature miR-612 levels, and this may represent a mechanism by which cancer risk is increased (HK et al. 2012). One of the potential targets of mir-612 is *IKZF2* (Friedman et al. 2009), a member of the Ikaros family of zinc-finger proteins. This protein is a hematopoietic-specific transcription factor involved in the regulation of lymphocyte development, and other members of this family have been associated with ALL susceptibility (Treviño et al. 2009, Papaemmanuil et al. 2009). Therefore, the presence of SNP rs12803915 in pre-mir-612 may contribute to an increased susceptibility to ALL based on its capacity to affect the expression of mir-612, as well as its downstream targets, such as *IKZF2*.

The second most significant association identified in this study involved SNP rs3746444 (p = 0.009) located in pre-mir-499. In this case, the G allele was associated with a lower risk for ALL. This SNP has recently been implicated in the etiology of several types of cancer with controversial results. The results of this study are consistent with those of three previous studies, in which the rs3746444 GG genotype was shown to be associated with a decreased risk of cancer (Kim et al. 2012, Liu et al. 2010, DH et al. 2013). By contrast, in six other studies

(Vinci et al. 2013, Xiang et al. 2012, Zhou et al. 2011, Hu et al. 2008, Alshatwi et al. 2012, George et al. 2011), the GG genotype was found to be associated with a higher risk of cancer development. There have also been studies that have found no association between this SNP and cancer risk (Hasani et al. 2014, Akkiz et al. 2011, Catucci et al. 2010, KT et al. 2012, Okubo et al. 2010, Tian et al. 2009). The inconsistency of these results may be due to differences in the carcinogenic mechanisms of different cancers, as well as differences in genetic backgrounds (Yang et al. 2008). Correspondingly, a meta-analysis observed that a tendency for reduced cancer risk was associated with the mir-499 rs3746444 GG genotype in Caucasian populations (Qiu et al. 2012).

This SNP, rs3746444, is located in pre-mir-499, in the premature sequences of mir-499-5p and mir-499-3p and in the seed sequence of mir-499-3p. Therefore, it could have a double effect. On one hand, a SNP located in a pre-miRNA region has the potential to impact the processing of pre-miRNAs into mature miRNAs (DH et al. 2013). Correspondingly, different genotypes of rs3746444 in pre-mir-499 have been associated with dysregulated expression of mir-499-5p in a colorectal cancer model (Vinci et al. 2013). These data are of particular interest considering that mir-499-5p seems to be upregulated in ALL (H. Zhang et al. 2009). On the other hand, SNP rs3746444 is also present in the seed region of a mature mir-499-3p and could be essential for the accurate recognition of target mRNA sequences. Potential targets of mir-499-3p (Friedman et al. 2009) include *FOXO1A* (a transcription factor that is dysregulated in B-ALL (Andersson et al. 2007)), *MS4A1* or *CD20* (a B-lymphocyte surface molecule that plays a role in the development and differentiation of B cells into plasma cells), and *PBX1* (pre–B-cell leukemia homeobox 1 that is a dysregulated transcription factor in ALL) (Li et al. 2009). Therefore, alterations in the sequence of mir-499-5p and mir-499-3p, which affect expression of these miRNAs and/or the binding of mir-499-3p to target mRNAs, may have functional consequences for ALL.

Among the SNPs located in pre-miRNA genes, other interesting result was that the rs10061133 G allele in mir-449b was associated with a decreased risk of ALL ($p = 0.012$). To our knowledge, this is the first time that this polymorphism has been associated with cancer risk. Based on its location, this polymorphism is also present in the mature sequence of the miRNA; therefore, it could affect the strength of miRNA–mRNA binding, as well as miRNA levels. This miRNA has also been found to be upregulated in endometrial cancer and bladder cancer (Chung et al. 2012, Catto et al. 2009). However, in this larger population, we have not replicated the

association previously found in a small pilot study between rs2910164 in mir-146a and ALL risk (Hasani et al. 2014).

When we analyzed 72 polymorphisms present in miRNA biogenesis pathway, 8 of these were found to be significantly associated with ALL risk. Polymorphism rs139919 in *TNRC6B* was the SNP most highly associated with ALL susceptibility among the miRNA-processing genes, with the CC genotype associated with an increase in ALL risk ($p = 0.011$). *TNRC6B* encodes an RNA interference machinery component, which contributes to the RNA-induced silencing complex, and is crucial for miRNA-dependent translational repression or degradation of target mRNAs (Tao et al. 2012). Although this SNP has not been analyzed previously, other genetic variants in *TNRC6B* have been associated with prostate cancer risk (Tao et al. 2012, Sun et al. 2009). It has also been suggested that alterations in the expression of *TNRC6B* are due to genetic variations that may affect mRNA levels that are normally regulated by *TNRC6B*, thereby affecting carcinogenesis (Sun et al. 2009).

We have also found other SNPs located in genes which produce proteins that contribute to the RNA-induced silencing complex associated with ALL susceptibility. These included two SNPs in *EIF2C1* (rs595961 and rs636832), one SNP in *CNOT1* (rs11866002), and one SNP in *CNOT6* (rs6877400). The SNPs rs595961 and rs636832 in *EIF2C1* have been previously shown to be associated with renal cell carcinoma risk in males (Horikawa et al. 2008) and lung cancer (Kim et al. 2010), respectively. However, to our knowledge, this is the first study to identify an association between the SNPs, rs11866002 in *CNOT 1* and rs6877400 in *CNOT6*, and cancer risk. In addition, it is hypothesized that rs595961, rs11866002, and rs6877400 have putative roles in transcription and/or regulation of splicing events, thereby affecting gene expression. SNPs that affect expression levels of these proteins may have deleterious effects on miRNA–mRNA interactions and may affect cancer development and progression. Correspondingly, *EIF2C1* is frequently lost in human cancers such as Wilms tumor, neuroblastoma, and carcinomas of the breast, liver, and colon (Koesters et al. 1999).

Among the genes of biogenesis machinery, when we studied genes that contribute to the processing of pri-miRNAs to pre-miRNAs, we identified an association between SNPs and ALL risk. These SNPs included rs10035440 in *DROSHA*, and rs9606248 and rs1640299 in *DGCR8*. Interestingly, these three SNPs have putative roles in transcriptional regulation and may affect the expression levels of *DROSHA* or *DGCR8*. As a result, levels of miRNAs could be affected, leading to an increased ALL risk. Consistent with this hypothesis, a differentially expressed

miRNA profile and *DROSHA* gene expression have been observed in relation to another SNP, rs640831, being present in the lung tissue (Rotunno et al. 2010). Furthermore, increased expression of *DROSHA* and *DGCR8* has been shown to dysregulate miRNAs present in the pleomorphic adenomas of the salivary gland (X. Zhang et al. 2009). In contrast, decreased expression of *DROSHA* and *DGCR8* have been shown to accelerate cellular transformation and tumorigenesis (Kumar et al. 2007). Surprisingly, the association of these polymorphisms with the risk to develop other tumors has previously been analyzed (Yang et al. 2008, Horikawa et al. 2008, Sung et al. 2011), and no significant association was found. However, they had not been analyzed in ALL patients until now.

In conclusion, the results of this study indicate that SNP rs12803915 located in pre-mir-612 and SNP rs3746444 located in pre-mir-499 may represent novel markers of B-ALL susceptibility. It would be of great interest to confirm these results in different cohorts of patients. To our knowledge, this is the first extensive study to report miRNA-related SNPs associated with ALL risk.

STATEMENT OF FINANCIAL SUPPORT

A.G.-C. was supported by a predoctoral grant from the Gangoiti Barrera Foundation, Bilbao, Spain. E.L.-L. was supported by a predoctoral grant of the Basque Government and “Fellowship for recent doctors until their integration in postdoctoral programs” by the Investigation Vice-rector’s office of the University of Basque Country (UPV/EHU). This project was supported by Spanish Thematic Network of Cooperative Research in Cancer (RD/06/0020/0048), the Basque Government (IT661-13, S-PE12UN060), and UPV/EHU (UFI11/35). Support by SGIker (UPV/EHU) is gratefully acknowledged.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to disclose.

REFERENCES

Akkiz, H., Bayram, S., Bekar, A., Akgöllü, E. and Üsküdar, O. (2011) 'Genetic variation in the microRNA-499 gene and hepatocellular carcinoma risk in a Turkish population: lack of any association in a case-control study', *Asian Pac J Cancer Prev,* 12(11), 3107-12.

Alshatwi, A. A., Shafi, G., Hasan, T. N., Syed, N. A., Al-Hazzani, A. A., Alsaif, M. A. and Alsaif, A. A. (2012) 'Differential expression profile and genetic variants of microRNAs sequences in breast cancer patients', *PLoS One,* 7(2), e30049.

Andersson, A., Ritz, C., Lindgren, D., Edén, P., Lassen, C., Heldrup, J., Olofsson, T., Råde, J., Fontes, M., Porwit-Macdonald, A., Behrendtz, M., Höglund, M., Johansson, B. and Fioretos, T. (2007) 'Microarray-based classification of a consecutive series of 121 childhood acute leukemias: prediction of leukemic and genetic subtype as well as of minimal residual disease status', *Leukemia,* 21(6), 1198-203.

Catto, J. W., Miah, S., Owen, H. C., Bryant, H., Myers, K., Dudziec, E., Larré, S., Milo, M., Rehman, I., Rosario, D. J., Di Martino, E., Knowles, M. A., Meuth, M., Harris, A. L. and Hamdy, F. C. (2009) 'Distinct microRNA alterations characterize high- and low-grade bladder cancer', *Cancer Res,* 69(21), 8472-81.

Catucci, I., Yang, R., Verderio, P., Pizzamiglio, S., Heesen, L., Hemminki, K., Sutter, C., Wappenschmidt, B., Dick, M., Arnold, N., Bugert, P., Niederacher, D., Meindl, A., Schmutzler, R. K., Bartram, C. C., Ficarazzi, F., Tizzoni, L., Zaffaroni, D., Manoukian, S., Barile, M., Pierotti, M. A., Radice, P., Burwinkel, B. and Peterlongo, P. (2010) 'Evaluation of SNPs in miR-146a, miR196a2 and miR-499 as low-penetrance alleles in German and Italian familial breast cancer cases', *Hum Mutat,* 31(1), E1052-7.

Chung, T. K., Lau, T. S., Cheung, T. H., Yim, S. F., Lo, K. W., Siu, N. S., Chan, L. K., Yu, M. Y., Kwong, J., Doran, G., Barroilhet, L. M., Ng, A. S., Wong, R. R., Wang, V. W., Mok, S. C., Smith, D. I., Berkowitz, R. S. and Wong, Y. F. (2012) 'Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXC1', *Int J Cancer,* 130(5), 1036-45.

DH, A., H, R., YK, C., YJ, J., KT, M., K, K., SP, H., SG, H. and NK, K. (2013) 'Association of the miR-146aC>G , miR-149T>C , miR-196a2T>C , and miR-499A>G polymorphisms with gastric cancer risk and survival in the Korean population', 52(Suppl 1:E39-51),

Ellinghaus, E., Stanulla, M., Richter, G., Ellinghaus, D., te Kronnie, G., Cario, G., Cazzaniga, G., Horstmann, M., Panzer Grümayer, R., Cavé, H., Trka, J., Cinek, O., Teigler-Schlegel, A., ElSharawy, A., Häsler, R., Nebel, A., Meissner, B., Bartram, T., Lescai, F., Franceschi, C., Giordan, M., Nürnberg, P., Heinzow, B., Zimmermann, M., Schreiber, S., Schrappe, M. and Franke, A. (2012) 'Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia', *Leukemia,* 26(5), 902-9.

Friedman, R. C., Farh, K. K., Burge, C. B. and Bartel, D. P. (2009) 'Most mammalian mRNAs are conserved targets of microRNAs', *Genome Res,* 19(1), 92-105.

George, G. P., Gangwar, R., Mandal, R. K., Sankhwar, S. N. and Mittal, R. D. (2011) 'Genetic variation in microRNA genes and prostate cancer risk in North Indian population', *Mol Biol Rep,* 38(3), 1609-15.

Hasani, S. S., Hashemi, M., Eskandari-Nasab, E., Naderi, M., Omrani, M. and Sheybani-Nasab, M. (2014) 'A functional polymorphism in the miR-146a gene is associated with the risk of childhood acute lymphoblastic leukemia: a preliminary report', *Tumour Biol,* 35(1), 219-25.

HK, K., L, P.-O. and SJ., C. (2012) 'Common genetic variants in miR-1206 (8q24.2) and miR-612 (11q13.3) affect biogenesis of mature miRNA forms', 7(10), e47454.

Horikawa, Y., Wood, C. G., Yang, H., Zhao, H., Ye, Y., Gu, J., Lin, J., Habuchi, T. and Wu, X. (2008) 'Single nucleotide polymorphisms of microRNA machinery genes modify the risk of renal cell carcinoma', *Clin Cancer Res,* 14(23), 7956-62.

Hu, Z., Chen, J., Tian, T., Zhou, X., Gu, H., Xu, L., Zeng, Y., Miao, R., Jin, G., Ma, H., Chen, Y. and Shen, H. (2008) 'Genetic variants of miRNA sequences and non-small cell lung cancer survival', *J Clin Invest,* 118(7), 2600-8.

Kim, J. S., Choi, Y. Y., Jin, G., Kang, H. G., Choi, J. E., Jeon, H. S., Lee, W. K., Kim, D. S., Kim, C. H., Kim, Y. J., Son, J. W., Jung, T. H. and Park, J. Y. (2010) 'Association of a common AGO1 variant with lung cancer risk: a two-stage case-control study', *Mol Carcinog,* 49(10), 913-21.

Kim, W. H., Min, K. T., Jeon, Y. J., Kwon, C. I., Ko, K. H., Park, P. W., Hong, S. P., Rim, K. S., Kwon, S. W., Hwang, S. G. and Kim, N. K. (2012) 'Association study of microRNA polymorphisms with hepatocellular carcinoma in Korean population', *Gene,* 504(1), 92-7.

Koesters, R., Adams, V., Betts, D., Moos, R., Schmid, M., Siermann, A., Hassam, S., Weitz, S., Lichter, P., Heitz, P. U., von Knebel Doeberitz, M. and Briner, J. (1999) 'Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence, genomic organization, localization to chromosomal bands 1p34-p35, and expression', *Genomics,* 61(2), 210-8.

KT, M., JW, K., YJ, J., MJ, J., SY, C., D, O. and NK, K. (2012) 'Association of the miR-146aC > G , 149C > T , 196a2C > T , and 499A > G polymorphisms with colorectal cancer in the Korean population', 51(Suppl 1:E65-73),

Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. and Jacks, T. (2007) 'Impaired microRNA processing enhances cellular transformation and tumorigenesis', *Nat Genet,* 39(5), 673-7.

Li, Z., Zhang, W., Wu, M., Zhu, S., Gao, C., Sun, L., Zhang, R., Qiao, N., Xue, H., Hu, Y., Bao, S., Zheng, H. and Han, J. D. (2009) 'Gene expression-based classification and regulatory networks of pediatric acute lymphoblastic leukemia', *Blood,* 114(20), 4486-93.

Liu, Z., Li, G., Wei, S., Niu, J., El-Naggar, A. K., Sturgis, E. M. and Wei, Q. (2010) 'Genetic variants in selected pre-microRNA genes and the risk of squamous cell carcinoma of the head and neck', *Cancer,* 116(20), 4753-60.

Okubo, M., Tahara, T., Shibata, T., Yamashita, H., Nakamura, M., Yoshioka, D., Yonemura, J., Ishizuka, T., Arisawa, T. and Hirata, I. (2010) 'Association between common genetic variants in pre-microRNAs and gastric cancer risk in Japanese population', *Helicobacter,* 15(6), 524-31.

Papaemmanuil, E., Hosking, F. J., Vijayakrishnan, J., Price, A., Olver, B., Sheridan, E., Kinsey, S. E., Lightfoot, T., Roman, E., Irving, J. A., Allan, J. M., Tomlinson, I. P., Taylor, M., Greaves, M. and Houlston, R. S. (2009) 'Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia', *Nat Genet,* 41(9), 1006-10.

Qiu, M. T., Hu, J. W., Ding, X. X., Yang, X., Zhang, Z., Yin, R. and Xu, L. (2012) 'Hsa-miR-499 rs3746444 polymorphism contributes to cancer risk: a meta-analysis of 12 studies', *PLoS One,* 7(12), e50887.

Rotunno, M., Zhao, Y., Bergen, A. W., Koshiol, J., Burdette, L., Rubagotti, M., Linnoila, R. I., Marincola, F. M., Bertazzi, P. A., Pesatori, A. C., Caporaso, N. E., McShane, L. M., Wang, E. and Landi, M. T. (2010) 'Inherited polymorphisms in the RNA-mediated interference machinery affect microRNA expression and lung cancer survival', *Br J Cancer,* 103(12), 1870-4.

Ryan, B. M., Robles, A. I. and Harris, C. C. (2010) 'Genetic variation in microRNA networks: the implications for cancer research', *Nat Rev Cancer,* 10(6), 389-402.

Salzman, D. W. and Weidhaas, J. B. (2013) 'SNPing cancer in the bud: microRNA and microRNA-target site polymorphisms as diagnostic and prognostic biomarkers in cancer', *Pharmacol Ther,* 137(1), 55-63.

Sambrook, J. and Russell, D. W. (2001) 'Preparation and Analysis of Eukaryotic Genomic DNA' in *Molecular cloning: A laboratory manual*, 3rd edition ed., Cold Spring Harbor, USA: CSHL Press, 4-12.

Schotte, D., Akbari Moqadam, F., Lange-Turenhout, E. A., Chen, C., van Ijcken, W. F., Pieters, R. and den Boer, M. L. (2011) 'Discovery of new microRNAs by small RNAome deep sequencing in childhood acute lymphoblastic leukemia', *Leukemia,* 25(9), 1389-99.

Sun, J., Zheng, S. L., Wiklund, F., Isaacs, S. D., Li, G., Wiley, K. E., Kim, S. T., Zhu, Y., Zhang, Z., Hsu, F. C., Turner, A. R., Stattin, P., Liu, W., Kim, J. W., Duggan, D., Carpten, J., Isaacs, W., Grönberg, H., Xu, J. and Chang, B. L. (2009) 'Sequence variants at 22q13 are associated with prostate cancer risk', *Cancer Res,* 69(1), 10-5.

Sung, H., Lee, K. M., Choi, J. Y., Han, S., Lee, J. Y., Li, L., Park, S. K., Yoo, K. Y., Noh, D. Y., Ahn, S. H. and Kang, D. (2011) 'Common genetic polymorphisms of microRNA biogenesis pathway genes and risk of breast cancer: a case-control study in Korea', *Breast Cancer Res Treat,* 130(3), 939-51.

Tao, S., Wang, Z., Feng, J., Hsu, F. C., Jin, G., Kim, S. T., Zhang, Z., Gronberg, H., Zheng, L. S., Isaacs, W. B., Xu, J. and Sun, J. (2012) 'A genome-wide search for loci interacting with known prostate cancer risk-associated genetic variants', *Carcinogenesis,* 33(3), 598-603.

Tian, T., Shu, Y., Chen, J., Hu, Z., Xu, L., Jin, G., Liang, J., Liu, P., Zhou, X., Miao, R., Ma, H., Chen, Y. and Shen, H. (2009) 'A functional genetic variant in microRNA-196a2 is associated with increased susceptibility of lung cancer in Chinese', *Cancer Epidemiol Biomarkers Prev,* 18(4), 1183-7.

Treviño, L. R., Yang, W., French, D., Hunger, S. P., Carroll, W. L., Devidas, M., Willman, C., Neale, G., Downing, J., Raimondi, S. C., Pui, C. H., Evans, W. E. and Relling, M. V. (2009) 'Germline genomic variants associated with childhood acute lymphoblastic leukemia', *Nat Genet,* 41(9), 1001-5.

Vinci, S., Gelmini, S., Mancini, I., Malentacchi, F., Pazzagli, M., Beltrami, C., Pinzani, P. and Orlando, C. (2013) 'Genetic and epigenetic factors in regulation of microRNA in colorectal cancers', *Methods,* 59(1), 138-46.

Xiang, Y., Fan, S., Cao, J., Huang, S. and Zhang, L. P. (2012) 'Association of the microRNA-499 variants with susceptibility to hepatocellular carcinoma in a Chinese population', *Mol Biol Rep,* 39(6), 7019-23.

Yang, H., Dinney, C. P., Ye, Y., Zhu, Y., Grossman, H. B. and Wu, X. (2008) 'Evaluation of genetic variants in microRNA-related genes and risk of bladder cancer', *Cancer Res,* 68(7), 2530-7.

Zhang, H., Yang, J. H., Zheng, Y. S., Zhang, P., Chen, X., Wu, J., Xu, L., Luo, X. Q., Ke, Z. Y., Zhou, H., Qu, L. H. and Chen, Y. Q. (2009) 'Genome-wide analysis of small RNA and novel MicroRNA discovery in human acute lymphoblastic leukemia based on extensive sequencing approach', *PLoS One,* 4(9), e6849.

Zhang, X., Cairns, M., Rose, B., O'Brien, C., Shannon, K., Clark, J., Gamble, J. and Tran, N. (2009) 'Alterations in miRNA processing and expression in pleomorphic adenomas of the salivary gland', *Int J Cancer,* 124(12), 2855-63.

Zhou, B., Wang, K., Wang, Y., Xi, M., Zhang, Z., Song, Y. and Zhang, L. (2011) 'Common genetic polymorphisms in pre-microRNAs and risk of cervical squamous cell carcinoma', *Mol Carcinog,* 50(7), 499-505.

Submitted to Haematologica

MicroRNAs SNPS involved in acute lymphoblastic leukemia susceptibility

Gutierrez-Camino A[1], Martín-Guerrero I[1], Dolzan V[2], Jazbec J[3], Carbone Bañeres A[4], Garcia de Andoin N[5,6], Sastre A[7], Astigarraga I[8,9], Navajas A[9], Garcia-Orad A[1,9]

[1]Department of Genetics, Physic Anthropology and Animal Physiology, University of the Basque Country, UPV/EHU, Leioa, Spain; [2]Institute of Biochemistry, Faculty of Medicine, Ljubljana, Slovenia; [3]Department of oncology and haematology, University Children's Hospital, University Medical Centre Ljubljana, Ljubljana, Slovenia; [4]Department of Paediatrics, University Hospital Miguel Servet, Zaragoza, Spain; [5]Department of Paediatrics, University Hospital Donostia, San Sebastian, Spain; [6]BioDonostia Health Research Institute, San Sebastian, Spain; [7]Department of Oncohematology, University Hospital La Paz, Madrid, Spain; [8]Department of Paediatrics, University Hospital Cruces, Barakaldo, Spain; [9]BioCruces Health Research Institute, Barakaldo, Spain.

Abstract

The genetic basis of acute lymphoblastic leukemia (ALL) susceptibility has been supported by genome-wide association studies. Interestingly, 37.5% of SNPs found in these studies corresponds to intergenic regions, suggesting that non-coding regions, such as microRNAs (miRNAs), could play an important role in ALL risk. MiRNAs regulated more than 50% of human genes, including those involved in B-cell maturation, differentiation and proliferation. SNPs in miRNAs can alter their own levels or function, affecting its target gene expression. In ALL, only three studies of miRNA SNPs have been performed and nowadays the number of annotated miRNAs has increased substantially, therefore, the aim of this study was to determine the role of the currently described miRNAs SNPs in B-ALL susceptibility. We analyzed all variants in pre-miRNAs (MAF>1%) in two independent cohorts from Spain and Slovenia and evaluated the putative functional implication by *in silico* analysis. SNPs rs12402181 in mir3117 and rs62571442 in mir3689d2 were associated with B-ALL risk in both cohorts, possibly through its effect on MAPK signaling pathway. SNP rs10406069 in mir5196 was associated with the risk of developing B-hyperdiploid ALL, possibly by its own levels changes or a direct effect on *CD22* gene. These SNPs could be novel markers for B-ALL susceptibility.

Keywords: SNP, miRNAs, acute lymphoblastic leukaemia, susceptibility

INTRODUCTION

Acute lymphoblastic leukemia (ALL) is the most common childhood malignancy and a leading cause of death due to disease in children (Pui and Evans 2006, Greaves 2006). The genetic basis of ALL susceptibility has been supported, on one hand, by its association with certain congenital abnormalities (Xu et al. 2013) and, more recently, by several genome-wide association studies (GWAS). These GWAS identified common variants at *ARID5B*, *IKZF1*, *CEBPE* and *CDKN2A* influencing ALL risk in children of European descent (Treviño et al. 2009, Papaemmanuil et al. 2009, Xu et al. 2013, Sherborne et al. 2010, Migliorini et al. 2013, Orsi et al. 2012), results that have been repeatedly validated (Gutiérrez-Camino et al. 2013, Healy et al. 2010, Prasad et al. 2010, Ross et al. 2013). Interestingly, 37.5% of SNPs found in these GWASs corresponds to intergenic regions. These data suggest that non-coding regions could play an important role in the risk of ALL.

MicroRNAs (miRNAs) are non-coding RNA that regulate gene expression at the post-transcriptional level by binding to the 3´ untranslated region (UTR) of a target mRNA, leading to its translation inhibition or degradation (Ryan et al. 2010). Through this mechanism, miRNAs regulate more than 50% of human genes, having an enormous impact on the function of any cell (Johanson et al. 2014), including B-lymphocytes.

It has been widely shown that miRNAs regulate B-cell maturation and function, controlling B-cell receptor (BCR) signaling, B-cell migration/adhesion, cell–cell interactions in immune niches, and the production and class-switching of immunoglobulins (Marques et al. 2015, Musilova and Mraz 2015). They also contribute to the regulation of important signaling pathways such as tyrosine kinase and Ras signaling (Musilova and Mraz 2015). The deregulation of these pathways has been demonstrated in ALL (Layton Tovar and Mendieta Zerón 2016). In fact, recent studies have found more than 200 miRNAs deregulated in pediatric B-ALL patients (Schotte et al. 2011, Schotte et al. 2009, Zhang et al. 2009, Duyu et al. 2014). All these data show the role of miRNAs in pediatric B-ALL evolution.

Genetic variations in miRNAs can alter their function affecting their targets genes expression. These variants can modify the miRNA expression levels if they are located in the pre-miRNA or the mRNA-miRNA binding if they are located in the seed region. Nowadays, several works have already described polymorphisms in miRNAs associated with the susceptibility to different types of cancer (Xia et al. 2014, Srivastava and Srivastava 2012). Despite all these evidences, only three studies analyzing the involvement of SNPs in miRNAs in the risk of ALL have been

performed (Hasani et al. 2013, Tong et al. 2014, Gutierrez-Camino et al. 2014). Hasani and colleagues found rs2910164 in mir-146a associated with ALL susceptibility in a Iranian population of 75 children diagnosed with ALL (Hasani et al. 2013). Tong and colleagues found association between rs11614913 in mir196a-2 and ALL risk in a Chinese population of 574 pediatric ALL patients (Tong et al. 2014). Recently, our group found association between rs12803915 in miR-612, rs3746444 in miR-499 and rs10061133 in miR-449b and B-ALL risk in a Spanish cohort of 213 children (Gutierrez-Camino et al. 2014). Of note is the fact that although a relatively low number of SNPs were analyzed in miRNAs and B-ALL susceptibility, significant results were found.

Considering all these data and that nowadays the number of annotated miRNAs has increased substantially up to 2500 miRNAs approximately (Kozomara and Griffiths-Jones 2014), the aim of this study was to determine the role of the currently described SNPs in miRNAs in the risk of B-ALL. For this aim, we analyzed all variants in pre-miRNAs genes with a minor allele frequency higher than 1% in two independent cohorts of Spanish and Slovenian origin. The putative functional implication of significant variants was evaluated by *in silico* analysis.

MATERIALS AND METHODS

Study Participants

A total of 310 Caucasian children diagnosed with B-ALL and 434 unrelated healthy controls were included in this study (Table 37). The Spanish cohort consisted of 231 children diagnosed with B-ALL between 2000 and 2011 in the Pediatric Oncology Units of four Spanish hospitals (University Hospital Cruces, University Hospital Donostia, University Hospital La Paz and University Hospital Miguel Servet) and 338 unrelated healthy individuals. The Slovenian cohort consisted of 79 Caucasian children diagnosed with B-ALL between 1993 and 2009, at the Department of Hematology and Oncology of the University Children's Hospital Ljubljana and 96 unrelated healthy individuals.

Data were collected objectively, blinded to genotypes, from the patients' medical files. The two most common ALL subtypes (B-lineage hyperdiploid ALL with more than 50 chromosomes (B-hyperdiploid) and B-lineage ALL bearing the t(12;21)/ETV6-RUNX1 fusion) were also analyzed. Sex and age data were systematically recorded from the clinical records (Table 38). Informed consent was obtained from all participants, or from their parents prior to sample collection. The study was approved by the ethics committees (PI2014039 and 62/07/03) and was carried out according to the Declaration of Helsinki.

Table 38: Characteristic of study population.

	Spanish cohort		Slovenian cohort	
	Patients	**Controls**	**Patients**	**Controls**
No. of individuals	231	338	79	96
Mean age ± SE, y	4.04 ± 3.61	57.8 ± 28.1	4.65 ± 5.41	44.5 ± 9.4
Sex*				
Males, n (%)	128 (55.7)	157 (46.4)	41 (51.9)	58 (60.4)
Females, n (%)	102 (44.3)	181 (53.6)	38 (48.1)	38 (39.6)
Genetic alterations[#]				
Hyperdiploid	56 (24.2)	-	9 (11.4)	-
ETV6-RUNX1	37 (16.0)	-	12 (15.2)	-
MLL	13 (5.6)	-	4 (5.1)	-
BCR-ABL	6 (2.6)	-	1 (1.3)	-
E2A-PBX1	6 (2.6)	-	-	-
Hipodiploid	2 (0.9)	-	1 (1.3)	-
Other	1 (0.4)	-	6 (7.6)	-
No alteration	95 (41.1)	-	48 (60.8)	-
No available	21 (9.1)	-	0	-

SE: standard error, y: years. *There is no datum for one patient of the Spanish cohort. [#]Six patients have more than one alteration in the Spanish cohort and two of the patients have more than one alteration in the Slovenian cohort.

Selection of genes and polymorphisms

We selected all the SNPs in pre-miRNAs with a MAF>0.01 in European/Caucasian populations described in the databases until May 2014. We decided to include all miRNAs due to the fact that they can regulate a wide range of genes that are not completely defined. Therefore, any miRNA could be implicated in the regulation of genes affecting ALL risk. Of a total of 1910 SNPs in 969 miRNAs found at the moment of the study, we included all the SNPs with a MAF>0.01, a total of 213 SNPs in 206 pre-miRNAs (Table 16). The SNP selection was performed using miRNA SNiPer (www.integratomics-time.com/miRNA-SNiPer/), NCBI (http://www.ncbi.nlm.nih.gov/snp/) and literature review.

Genotype analyses

Genomic DNA was extracted from remission peripheral blood or bone marrow (with <5 % blast cells) as previously described (Sambrook and Russell 2001). DNA was quantified using PicoGreen (Invitrogen Corp., Carlsbad, CA).

For each sample, 400 ng of DNA were genotyped using the GoldenGate Genotyping Assay with Veracode technology according to the published Illumina protocol. Data were analyzed

with GenomeStudio software for genotype clustering and calling. Duplicate samples and CEPH trios (Coriell Cell Repository, Camden, NJ) were genotyped across the plates. SNPs showing Mendelian allele-transmission errors or showing discordant genotypes were excluded from the analysis.

Statistical analysis

To identify any deviation in Hardy-Weinberg equilibrium (HWE) for the healthy controls, a $\chi 2$ test was used. The association between genetic polymorphisms in cases and controls, as well as ALL subtypes and controls, was also evaluated using the $\chi 2$ or Fisher's exact test. The effect sizes of the associations were estimated by the odds ratio from univariate logistic regression. The most significant test among codominant, dominant, recessive, and additive genetic models was selected. The results were adjusted for multiple comparisons by the False Discovery Rate (FDR) (Benjamini and Hochberg 1995). In all cases the significance level was set at 5%. Analyses were performed by using R v2.11 software.

Bioinformatic analysis

miRNAs secondary structures prediction

The RNAfold web tool (http://rna.tbi.univie.ac.at) was used to calculate the minimum free energy (MFE) secondary structures and to predict the most stable secondary structures of the miRNAs showing significant SNPs.

Gene targets selection and pathways analysis

MirWalk (Dweep and Gretz 2015) (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/) database was used to select miRNA targets. Targets predicted by at least 6 different algorithms provided by miRWalk were selected. Enriched pathway analyses of putative target genes were determined with ConsensusPath database (CPdB) (http://consensuspathdb.org/) (Kamburov et al. 2013) using the over-representation analysis module. Gene list were analyzed against the default collection of KEGG (Kanehisa et al. 2016), Reactome (Fabregat et al. 2016) and BioCarta (http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways) pathways databases. A conservative *p*-value cutoff (0.0001) was used.

RESULTS

Genotyping Results

Genotyping analyses were performed in 310 patients with B-ALL (231 from Spain and 79 from Slovenia) and 434 unrelated healthy controls (338 from Spain and 96 from Slovenia). Successful genotyping was achieved for 718 of 744 DNA samples (96.5%), 217 children with B-ALL and 330 controls from the Spanish cohort and 75 children with B-ALL and 96 controls from the Slovenian cohort. From the total of 213 SNPs, after eliminating SNPs with genotyping failures (<80%), monomorphic in our population or with deviations from HWE in controls, a total of 135 SNPs (63.4%) were included in the association analysis (Table 39).

Table 39: SNPs selected for the study.

N	SNP	Gene	% Genotyping	MAF	Alleles	Exclusion criteria
1	**rs10061133**	hsa-mir-449b	100.0	0.079	A:G	
2	**rs10173558**	hsa-mir-1302-4	100.0	0.123	T:C	
3	**rs10406069**	hsa-mir-5196	99.3	0.209	G:A	
4	**rs10422347**	hsa-mir-4745	98.2	0.082	C:T	
5	**rs10461441**	hsa-mir-548ae-2	0.0	0.0	T:T	Genotyping failure
6	**rs10505168**	hsa-mir-2053	99.4	0.307	A:G	
7	**rs1055070**	hsa-mir-4700	99.7	0.059	T:G	
8	**rs1077020**	hsa-mir-943	0.0	0.0	T:T	Genotyping failure
9	**rs10878362**	hsa-mir-6074	0.0	0.0	T:T	Genotyping failure
10	**rs10934682**	hsa-mir-544b	100.0	0.159	T:G	
11	**rs11014002**	hsa-mir-603	0.0	0.0	T:T	Genotyping failure
12	**rs11032942**	hsa-mir-1343	0.0	0.0	T:T	Genotyping failure
13	**rs11156654**	hsa-mir-624	99.3	0.248	T:A	
14	**rs11237828**	hsa-mir-5579	0.0	0.0	T:T	Genotyping failure
15	**rs11259096**	hsa-mir-1265	99.9	0.056	T:C	
16	**rs11614913**	hsa-mir-196a-2	99.7	0.349	C:T	
17	**rs11651671**	hsa-mir-548at	0.0	0.0	T:T	Genotyping failure
18	**rs11713052**	hsa-mir-5092	100.0	0.038	C:G	
19	**rs11714172**	hsa-mir-4792	99.2	0.363	T:G	
20	**rs11907020**	hsa-mir-3192	99.9	0.02	T:C	
21	**rs11983381**	hsa-mir-4653	97.1	0.188	A:G	
22	**rs12197631**	hsa-mir-548a-1	0.0	0.0	T:T	Genotyping failure
23	**rs12355840**	hsa-mir-202	85.7	0.185	T:C	No HWE Slovenia No HWE total
24	**rs12402181**	hsa-mir-3117	100.0	0.135	G:A	
25	**rs12451747**	hsa-mir-1269b	0.0	0.0	T:T	Genotyping failure
26	**rs12456845**	hsa-mir-4744	99.9	0.036	T:C	
27	**rs12473206**	hsa-mir-4433	0.0	0.0	T:T	Genotyping failure
28	**rs12512664**	hsa-mir-4274	99.2	0.471	A:G	
29	**rs12523324**	hsa-mir-4277	0.0	0.0	T:T	Genotyping failure
30	**rs12780876**	hsa-mir-4293	99.4	0.289	T:A	

Table 39: SNPs selected for the study (continuation).

N	SNP	Gene	% Genotyping	MAF	Alleles	Exclusion criteria
31	**rs12803915**	hsa-mir-612	99.3	0.161	G:A	
32	**rs12879262**	hsa-mir-4309	99.7	0.145	G:C	
33	**rs12894467**	hsa-mir-300	99.7	0.385	C:T	
34	**rs13186787**	hsa-mir-1294	0.0	0.0	T:T	Genotyping failure
35	**rs13299349**	hsa-mir-3152	85.5	0.348	G:A	
36	**rs1414273**	hsa-mir-548ac	0.0	0.0	T:T	Genotyping failure
37	**rs1439619**	hsa-mir-3175	86.1	0.498	C:A	
38	**rs1572687**	hsa-mir-5007	99.4	0.45	C:T	
39	**rs1683709**	hsa-mir-3612	100.0	0.199	C:T	
40	**rs17022749**	hsa-mir-5700	0.0	0.0	T:T	Genotyping failure
41	**rs17091403**	hsa-mir-2110	99.9	0.089	C:T	
42	**rs17111728**	hsa-mir-4422	100.0	0.068	T:C	
43	**rs174561**	hsa-mir-1908	57.0	0.254	T:C	Genotyping failure
44	**rs17737028**	hsa-mir-3143	100.0	0.008	A:G	
45	**rs17759989**	hsa-mir-633	99.9	0.017	A:G	
46	**rs17797090**	hsa-mir-3652	99.9	0.109	G:A	
47	**rs17885221**	hsa-mir-4733	99.7	0.057	C:T	
48	**rs2042253**	hsa-mir-5197	99.6	0.22	A:G	
49	**rs2043556**	hsa-mir-605	99.0	0.218	A:G	
50	**rs2060455**	hsa-mir-4511	0.0	0.0	T:T	Genotyping failure
51	**rs2070960**	hsa-mir-3620	99.2	0.061	C:T	
52	**rs2114358**	hsa-mir-1206	99.2	0.433	T:C	
53	**rs215383**	hsa-mir-4494	99.2	0.178	G:A	
54	**rs2241347**	hsa-mir-3130-1	0.0	0.0	T:T	Genotyping failure
55	**rs2273626**	hsa-mir-4707	89.0	0.471	A:C	
56	**rs2289030**	hsa-mir-492	99.7	0.068	C:G	
57	**rs2291418**	hsa-mir-1229	100.0	0.03	C:T	
58	**rs2292181**	hsa-mir-564	99.7	0.042	G:C	
59	**rs2292832**	hsa-mir-149	0.0	0.0	T:T	Genotyping failure
60	**rs2368392**	hsa-mir-604	99.4	0.248	C:T	No HWE Spain
61	**rs243080**	hsa-mir-4432	98.3	0.43	C:T	
62	**rs257095**	hsa-mir-4636	99.7	0.154	A:G	
63	**rs2648841**	hsa-mir-1208	98.9	0.123	C:A	
64	**rs2663345**	hsa-mir-3183	0.0	0.0	T:T	Genotyping failure
65	**rs266435**	hsa-mir-4804	99.3	0.136	C:G	
66	**rs2682818**	hsa-mir-618	99.3	0.131	C:A	
67	**rs28477407**	hsa-mir-4308	100.0	0.088	C:T	
68	**rs28645567**	hsa-mir-378d-1	99.7	0.015	G:A	
69	**rs28655823**	hsa-mir-4472-1	89.6	0.114	G:C	
70	**rs28664200**	hsa-mir-1255a	82.3	0.247	T:C	
71	**rs2910164**	hsa-mir-146a	99.9	0.249	G:C	
72	**rs2967897**	hsa-mir-5695	100.0	0.0	G:G	Monomorphic
73	**rs3112399**	hsa-mir-4803	99.3	0.447	T:A	No HWE Spain No HWE total
74	**rs34115976**	hsa-mir-577	98.9	0.196	C:G	No HWE Spain

Table 39: SNPs selected for the study (continuation).

N	SNP	Gene	% Genotyping	MAF	Alleles	Exclusion criteria
75	**rs35196866**	hsa-mir-4669	0.0	0.0	T:T	Genotyping failure
76	**rs356125**	hsa-mir-2278	99.9	0.052	G:A	
77	**rs35613341**	hsa-mir-5189	99.6	0.339	C:G	No HWE Spain No HWE total
78	**rs35650931**	hsa-mir-6076	99.9	0.094	G:C	
79	**rs35770269**	hsa-mir-449c	99.6	0.345	A:T	
80	**rs35854553**	hsa-mir-3166	86.5	0.068	A:T	
81	**rs367805**	hsa-mir-3936	99.0	0.309	G:A	
82	**rs3734050**	hsa-mir-6499	99.9	0.053	C:T	
83	**rs3746444**	hsa-mir-499a	99.3	0.205	T:C	
84	**rs3823658**	hsa-mir-5090	99.7	0.141	G:A	
85	**rs4112253**	hsa-mir-4751	99.9	0.365	C:G	
86	**rs41274239**	hsa-mir-96	99.7	0.001	A:G	
87	**rs41274312**	hsa-mir-187	99.7	0.006	G:A	
88	**rs41286570**	hsa-mir-154	100.0	0.0	G:G	Monomorphic
89	**rs41291179**	hsa-mir-216a	100.0	0.056	A:T	
90	**rs41292412**	hsa-mir-122	99.9	0.002	C:T	
91	**rs4285314**	hsa-mir-3135b	0.0	0.0	T:T	Genotyping failure
92	**rs4414449**	hsa-mir-548ap	81.3	0.376	T:C	
93	**rs45530340**	hsa-mir-6084	99.9	0.0	C:C	Monomorphic
94	**rs4577031**	hsa-mir-548ap	99.6	0.366	A:T	
95	**rs4674470**	hsa-mir-4268	99.7	0.214	T:C	
96	**rs4809383**	hsa-mir-941-1	86.4	0.117	C:T	No HWE Spain
97	**rs4822739**	hsa-mir-548j	99.9	0.063	C:G	
98	**rs487571**	hsa-mir-5680	0.0	0.0	T:T	Genotyping failure
99	**rs4909237**	hsa-mir-595	99.6	0.164	C:T	
100	**rs4919510**	hsa-mir-608	99.4	0.198	C:G	
101	**rs515924**	hsa-mir-548al	99.3	0.1	A:G	
102	**rs521188**	hsa-mir-3671	100.0	0.035	A:G	
103	**rs56088671**	hsa-mir-4424	0.0	0.0	T:T	Genotyping failure
104	**rs56103835**	hsa-mir-323b	100.0	0.179	T:C	
105	**rs56195815**	hsa-mir-548aw	0.0	0.0	T:T	Genotyping failure
106	**rs56292801**	hsa-mir-5189	89.1	0.299	G:A	Ho HWE Spain No HWE total
107	**rs57111412**	hsa-mir-1283-1	0.0	0.0	T:T	Genotyping failure
108	**rs58450758**	hsa-mir-559	0.0	0.0	T:T	Genotyping failure
109	**rs58834075**	hsa-mir-656	99.9	0.02	C:T	
110	**Rs5965660**	hsa-mir-888	99.9	0.157	T:G	No HWE Spain No HWE Slovenia
111	**rs5997893**	hsa-mir-3928	99.3	0.319	G:A	
112	**rs60308683**	hsa-mir-4762	0.0	0.0	T:T	Genotyping failure
113	**rs6062431**	hsa-mir-4326	98.6	0.328	G:C	
114	**rs60871950**	hsa-mir-4467	98.7	0.475	G:A	
115	**rs61388742**	hsa-mir-596	99.6	0.094	T:C	
116	**rs61938575**	hsa-mir-3922	85.9	0.285	G:A	

Table 39: SNPs selected for the study (continuation).

N	SNP	Gene	% Genotyping	MAF	Alleles	Exclusion criteria
117	**rs61992671**	hsa-mir-412	99.6	0.493	G:A	
118	**rs62154973**	hsa-mir-4772	99.2	0.101	C:T	
119	**rs62376935**	hsa-mir-585	99.4	0.066	C:T	No HWE Slovenia No HWE total
120	**rs641071**	hsa-mir-4482	0.0	0.0	T:T	Genotyping failure
121	**rs6430498**	hsa-mir-3679	99.0	0.326	G:A	
122	**rs6505162**	hsa-mir-423	0.0	0.0	T:T	Genotyping failure
123	**rs6513496**	hsa-mir-646	99.6	0.197	T:C	
124	**rs66507245**	hsa-mir-4731	0.0	0.0	T:T	Genotyping failure
125	**rs66683138**	hsa-mir-3622a	0.0	0.0	T:T	Genotyping failure
126	**rs67042258**	hsa-mir-6128	99.4	0.26	G:A	
127	**rs670637**	hsa-mir-3167	94.7	0.0	T:T	Monomorphic
128	**rs67182313**	hsa-mir-4642	99.4	0.182	A:G	
129	**rs6726779**	hsa-mir-4431	99.4	0.377	T:C	
130	**rs67339585**	hsa-mir-3910-1, hsa-mir-3910-2	0.0	0.0	T:T	Genotyping failure
131	**rs6787734**	hsa-mir-3135a	0.0	0.0	T:T	Genotyping failure
132	**rs67976778**	hsa-mir-4305	0.0	0.0	T:T	Genotyping failure
133	**rs68035463**	hsa-mir-3144	99.4	0.223	C:A	
134	**rs6841938**	hsa-mir-1255b-1	0.0	0.0	T:T	Genotyping failure
135	**rs6977967**	hsa-mir-3683	99.9	0.193	A:G	
136	**rs6997249**	hsa-mir-3686	0.0	0.0	T:T	Genotyping failure
137	**rs701213**	hsa-mir-4427	0.0	0.0	T:T	Genotyping failure
138	**rs702742**	hsa-mir-378h	99.9	0.101	A:G	
139	**rs7070684**	hsa-mir-548aj-2	0.0	0.0	T:T	Genotyping failure
140	**rs71363366**	hsa-mir-1283-2	99.2	0.038	C:G	
141	**rs7205289**	hsa-mir-140	87.5	0.0	C:C	Monomorphic
142	**rs7207008**	hsa-mir-2117	99.3	0.468	T:A	
143	**rs7227168**	hsa-mir-4741	99.3	0.116	C:T	
144	**rs7247237**	hsa-mir-3188	99.2	0.291	C:T	
145	**rs72502717**	hsa-mir-3689f	0.0	0.0	T:T	Genotyping failure
146	**rs72631816**	hsa-mir-105-2	100.0	0.0	T:T	Monomorphic
147	**rs72631825**	hsa-mir-222	100.0	0.0	G:G	Monomorphic
148	**rs72631826**	hsa-mir-16-1	99.9	0.0	T:T	Monomorphic
149	**rs72631827**	hsa-mir-106b	99.9	0.0	G:G	Monomorphic
150	**rs72631831**	hsa-mir-323b	100.0	0.0	G:G	Monomorphic
151	**rs72631833**	hsa-mir-183	100.0	0.0	G:G	Monomorphic
152	**rs72646786**	hsa-mir-3972	99.3	0.114	C:T	
153	**rs72855836**	hsa-mir-3976	99.6	0.052	G:A	No HWE Slovenia No HWE total
154	**rs72996752**	hsa-mir-4999	95.0	0.249	A:G	
155	**rs73112689**	hsa-mir-4459	0.0	0.0	T:T	Genotyping failure
156	**rs7311975**	hsa-mir-1178	99.4	0.041	T:C	
157	**rs73147065**	hsa-mir-647	0.0	0.0	T:T	Genotyping failure
158	**rs73177830**	hsa-mir-4532	0.0	0.0	T:T	Genotyping failure
159	**rs73235381**	hsa-mir-548h-4	0.0	0.0	T:T	Genotyping failure

Table 39: SNPs selected for the study (continuation).

N	SNP	Gene	% Genotyping	MAF	Alleles	Exclusion criteria
160	**rs73239138**	hsa-mir-1269a	99.4	0.247	G:A	
161	**rs73410309**	hsa-mir-4739	0.0	0.0	T:T	Genotyping failure
162	**rs74428911**	hsa-mir-4474	99.9	0.01	G:T	No HWE Spain No HWE total
163	**rs74469188**	hsa-mir-6504	85.7	0.122	T:C	
164	**rs745666**	hsa-mir-3615	99.6	0.364	C:G	
165	**rs74704964**	hsa-mir-518d	85.8	0.035	C:T	
166	**rs74904371**	hsa-mir-2682	99.6	0.029	C:T	
167	**rs74949342**	hsa-mir-5702	100.0	0.006	C:G	
168	**rs7500280**	hsa-mir-4719	0.0	0.0	T:T	Genotyping failure
169	**rs75019967**	hsa-mir-4477a	99.9	0.0	A:A	Monomorphic
170	**rs7522956**	hsa-mir-4742	99.7	0.233	A:C	
171	**rs75598818**	hsa-mir-520f	99.7	0.027	G:A	
172	**rs75715827**	hsa-mir-944	99.6	0.071	T:C	
173	**rs75966923**	hsa-mir-4298	100.0	0.029	C:A	
174	**rs76481776**	hsa-mir-182	99.3	0.086	C:T	
175	**rs76800617**	hsa-mir-4521	100.0	0.023	A:G	
176	**rs77055126**	hsa-mir-1303	0.0	0.0	T:T	Genotyping failure
177	**rs7709117**	hsa-mir-4634	98.6	0.446	A:G	
178	**rs77639117**	hsa-mir-576	99.4	0.017	A:T	
179	**rs78396863**	hsa-mir-4743	99.4	0.011	G:C	
180	**rs78541299**	hsa-mir-6075	100.0	0.003	G:A	
181	**rs78790512**	hsa-mir-6083	100.0	0.175	G:A	
182	**rs78831152**	hsa-mir-4789	99.7	0.091	C:T	
183	**rs78832554**	hsa-mir-4786	99.9	0.024	G:A	
184	**rs7896283**	hsa-mir-4481	67.4	0.404	A:G	Genotyping failure
185	**rs7911488**	hsa-mir-1307	2.2	0.406	A:G	Genotyping failure
186	**rs79397096**	hsa-mir-597	100.0	0.015	G:A	
187	**rs79512808**	hsa-mir-3976	100.0	0.013	T:G	
188	**rs80128580**	hsa-mir-5707	100.0	0.026	G:A	
189	**rs8054514**	hsa-mir-3176	99.9	0.146	T:G	
190	**rs8078913**	hsa-mir-4520a	94.4	0.447	C:T	
191	**rs832733**	hsa-mir-4698	0.0	0.0	T:T	Genotyping failure
192	**rs850108**	hsa-mir-550a-3	0.0	0.0	T:T	Genotyping failure
193	**rs8667**	hsa-mir-4751	93.7	0.374	G:A	
194	**rs877722**	hsa-mir-4671	100.0	0.125	A:T	
195	**rs895819**	hsa-mir-27a	2.5	0.167	T:C	Genotyping failure
196	**rs897984**	hsa-mir-4519	0.0	0.0	T:T	Genotyping failure
197	**rs9295535**	hsa-mir-5689	0.0	0.0	T:T	Genotyping failure
198	**rs936581**	hsa-mir-3141	99.6	0.169	G:A	
199	**rs9842591**	hsa-mir-5186	85.9	0.457	C:A	
200	**rs9877402**	hsa-mir-5680	85.7	0.052	A:G	
201	**rs9913045**	hsa-mir-548h-3	0.0	0.0	T:T	Genotyping failure
202	**rs11048315**	hsa-mir-4302	99.4	0.127	G:A	No HWE Spain No HWE total

Table 39: SNPs selected for the study (continuation).

N	SNP	Gene	% Genotyping	MAF	Alleles	Exclusion criteria
203	**rs111803974**	hsa-mir-3908	0.0	0.0	T:T	Genotyping failure
204	**rs111906529**	hsa-mir-299, hsa-mir-380	99.9	0.013	T:C	
205	**rs112328520**	hsa-mir-520G	99.2	0.064	C:T	
206	**rs11269**	hsa-mir-1282	100.0	0.0	G:G	Monomorphic
207	**rs113808830**	hsa-mir-4532	99.4	0.102	C:T	
208	**rs116932476**	hsa-mir-4479	99.4	0.008	G:A	
209	**rs117258475**	hsa-mir-296	99.9	0.017	G:A	
210	**rs117650137**	hsa-mir-6717	100.0	0.033	G:A	
211	**rs117723462**	hsa-mir-3649	100.0	0.008	T:G	
212	**rs163642**	hsa-mir-4436B2	0.0	0.0	T:T	Genotyping failure
213	**rs62571442**	hsa-mir-3689	98.9	0.43	A:G	

Genotype association study of B-ALL

From the total of 135 SNPs, we found two SNPs in two miRNAs, rs12402181 at mir3117-3p and rs62571442 at mir3689d2, significantly associated with B-ALL risk in the Spanish population, and validated in the Slovenian cohort (Table 40). In the Spanish cohort, the AA genotype of rs12402181 at mir3117-3p displayed a 1.44-fold increased risk of B-ALL (CI 95%: 1.01-2.08; p = 0.047) under the log-additive genetic model (GG vs AG vs AA). The same effect was observed in the Slovenian cohort (OR: 2.01; CI 95%: 1.02-3.95; p=0.041). When both populations were analyses together, they showed the same trend increasing the p value (OR: 1.53; CI 95%: 1.12-2.09; P = 0.006), showing the A allele as the risk allele (p=0.007).

The second SNP was rs62571442 at mir3689d2. In the Spanish cohort the CT/CC genotype showed a 1.48-fold increased risk of B-ALL (CI 95%: 1.02-2.15; p=0.039). In the Slovenian cohort, a higher p value was observed (OR: 3.57; CI 95%: 1.57-8.12; p=0.001). When both populations were analyzed together, they showed the same tendency (OR: 1.31; CI 95%: 1.06-1.60; p=0.011), being the C allele the risk allele in the total population (p=0.012).

When we analyzed the Spanish cohort independently, other 13 SNPs in 13 miRNAs were significantly associated with B-ALL risk. Among them, TT genotype of rs35854553 in mir3166 displayed the most significant association (OR: 0.35; CI 95%: 0.18-0.67; p=0.0006). None of these 13 SNPs were replicated in the Slovenian population (Table 41).

Table 40: Polymorphisms in miRNAs associated with B-ALL risk in the Spanish and Slovenian population.

			Spanish and Slovenian cohort				Spanish cohort				Slovenian cohort			
Gene (Location)	**SNP (Position)**	**Genotype**	**N (controls) N=426**	**N (cases) N=292**	**OR(CI 95%)**	**P**	**N (controls) (N=330)**	**N (cases) (N=217)**	**OR(CI 95%)**	**P**	**N (controls) (N=96)**	**N (cases) (N=75)**	**OR(CI 95%)**	**P**
mir3117-3p (1p31.3)	rs12402181 (seed)	GG AG AA	332 (77.9) 90 (21.1) 4 (0.9)	203 (69.5) 82 (28.1) 7 (2.4)	Additive 1.53 (1.12-2.09)	0.006*	257 (77.9) 71 (21.5) 2 (0.6)	155 (71.4) 57 (26.3) 5 (2.3)	Additive 1.44 (1.01-2.08)	0.047*	75 (78.1) 19 (19.8) 2 (2.1)	48 (64) 25 (33.3) 2 (2.7)	Dominant 2.01 (1.02-3.95)	0.041*
		G A	754 (88.5) 98 (11.5)	488 (83.6) 96(16.4)	1.51 (1.11-2.05)	0.007*	585 (88.6) 75 (11.4)	367 (84.6) 67 (15.4)	1.42 (0.99-2.02)	0.050*	169 (88) 23 (12)	121 (80.7) 29 (19.3)	1.76 (0.97-3.19)	0.06
mir3689d2 (9q34.3)	rs62571442 (PM)	TT CT CC	147 (35.1) 207 (49.4) 65 (15.5)	82 (28.2) 145 (49.8) 64 (22)	Additive 1.32 (1.06-1.64)	0.011*	117 (36.2) 151 (46.7) 55 (17.0)	60 (27.8) 114 (52.8) 42 (19.4)	Dominant 1.48 (1.02-2.15)	0.039*	30 (31.2) 56 (58.3) 10 (10.4)	22 (29.3) 31 (41.3) 22 (29.3)	Recessive 3.57 (1.57-8.12)	0.001*
		T C	501 (59.8) 337 (40.2)	309 (53.1) 273 (46.9)	1.31 (1.06-1.6)	0.012*	385 (59.6) 261 (40.4)	234 (54.2) 198 (45.8)	1.24 (0.97-1.59)	0.07	116 (60.4) 76 (39.6)	75 (50) 75 (50)	1.52 (0.99-2.34)	0.054

Abbreviation: PM: pre-miRNA, OR: Odd ratio, CI: Confidence interval *Significant SNPs.

Table 41: Polymorphisms in miRNAs associated with B ALL risk in the Spanish cohort.

Gene (Location)	SNP (Position)	Genotype	N (controls) (N=330)	N (cases) (N=217)	OR(CI 95%)	P
mir3166 (11q14.2)	rs35854553 (PM)	AA AT TT	232 (83.2) 44 (15.8) 3 (1.1)	183 (93.4) 13 (6.6) 0	Dominant 0.35 (0.18-0.67)	0.0006
mir3144 (6q22.31)	rs68035463 (PM)	CC AC AA	209 (63.7) 107 (32.6) 12 (3.7)	115 (53.5) 82 (38.1) 18 (8.4)	Additive 1.51(1.13-2.01)	0.004
mir4745 (19p13.3)	rs10422347 (M)	CC CT TT	283 (87.3) 40 (12.3) 1 (0.3)	166 (78.3) 44 (20.8) 2 (0.9)	Recessive 1.91 (1.2-3.04)	0.005
mir5196 (19q13.12)	Rs10406069 (PM)	GG AG AA	208 (63.8) 103 (31.6) 15 (4.6)	123 (56.9) 90 (41.7) 3 (1.4)	Codominant 1.48 (1.03-2.12) 0.34 ((0.1-1.19)	0.010
mir612 (11q13.1)	rs12803915 (PM)	GG AG AA	213 (65.1) 104 (31.8) 10 (3.1)	161 (74.2) 54 (24.9) 2 (0.9)	Additive 0.65(0.44-0.95)	0.012
mir300 (14q32.31)	rs12894467 (PM)	CC CT TT	148 (45.0) 141 (42.9) 40 (12.2)	75 (34.7) 114 (52.8) 27 (12.5)	Dominant 1.54 (1.08-2.19)	0.016
mir595 (7q36.3)	rs4909237 (PM)	CC CT TT	233 (70.8) 90 (27.4) 6 (1.8)	145 (67.1) 59 (27.3) 12 (5.6)	Recessive 3.17 (1.17-8.57)	0.018
mir4653 (7q22.1)	rs11983381 (PM)	AA AG GG	227 (69) 96 (29.2) 6 (1.8)	121 (61.1) 68 (34.3) 9 (4.5)	Additive 1.43 (1.04-1.98)	0.029
mir-2278 (9q22.32)	rs356125 (PM)	GG AG AA	291 (88.2) 36 (10.9) 3 (0.9)	202 (93.5) 14 (6.5) 0	Dominant 0.52 (0.27-0.98)	0.034
mir4308 (14q22.3)	Rs28477407 (PM)	CC CT TT	281 (85.2) 46 (13.9) 3 (0.9	172 (79.3) 45 (20.7) 0	Codominant 1.6(1.02-2.5) 0	0.037
mir4432	rs243080 (PM)	CC CT TT	98 (29.8) 175 (53.2) 56 (17)	69 (32.9) 90 (42.9) 51 (24.3)	Recessive 1.56 (1.02-2.4)	0.040
mir3683 (7p22.1)	rs6977967 (PM)	AA AG GG	201 (60.9) 113 (34.2) 16 (4.8)	150 (69.4) 57 (26.4) 9 (4.2)	Dominant 0.69(0.48-0.99)	0.040
mir4634 (5q35.2)	rs7709117 (PM)	AA AG GG	104 (31.7) 155 (47.3) 69 (219	50 (23.7) 123 (58.3) 38 (18)	Dominant 1.49 (1.01-2.22)	0.042

Abbreviation: PM: pre-miRNA, : mature, OR: Odd ratio, CI: Confidence interval

In the Slovenian cohort, other 11 SNPs in 10 miRNAs showed association with B-ALL risk, being TT genotype of rs72646786 in mir3972 the most significant (OR: 0.24; CI 95%: 0.09-0.61; p=0.001). None of these 11 SNPs were replicated in the Spanish population (Table 42).

Table 42: Polymorphisms in miRNAs associated with B ALL risk in the Slovenian cohort.

Gene (Location)	SNP (Position)	Genotype	N (controls) (N=96)	N (cases) (N=79)	OR(CI 95%)	P
mir3972 (1p36.13)	rs72646786 (PM)	CC CT TT	70 (72.9) 25 (26) 1 (1)	68 (91.9) 4 (5.4) 2 (2.7)	Dominant 0.24 (0.09-0.61)	0.0011
mir5189 (16q24.2)	rs56292801 (PM)	GG AG AA	51 (53.1) 41 (42.7) 4 (4.2)	33 (44.6) 27 (36.5) 14 (18.9)	Recessive 5.37 (1.69-17.08)	0.0017
mir4293 (10p13)	rs12780876 (PM)	TT AT AA	28 (29.2) 54 (56.2) 14 (14.6)	36 (48.6) 35 (47.3) 3 (4.1)	Additive 0.45 (0.27-0.76)	0.0018
mir5189 (16q24.2)	rs35613341 (PM)	CC CG GG	49 (51) 41 (42.7) 6 (6.2)	29 (39.2) 30 (40.5) 15 (20.3)	Recessive 3.81 (1.4-10.39)	0.005
mir3175 (15q26.1)	rs1439619 (PM)	AA AC CC	34 (35.4) 41 (42.7) 21 (21.9)	8 (16) 23 (46) 19 (38)	Additive 1.93 (1.19-3.11)	0.006
mir5682 (3q13.33)	rs9877402 (PM)	AA AG	88 (93.6) 6 (6.4)	39 (78) 11 (22)	Codominant 4.14 (1.43-11.9)	0.007
mir3615 (17q25.1)	rs745666 (PM)	CC CG GG	40 (41.7) 39 (40.6) 17 (17.7)	30 (40) 41 (54.7) 4 (5.3)	Recessive 0.27 (0.08-0.81)	0.010
mir4772 (2q12.1)	rs62154973 (M)	CC CT TT	69 (71.9) 27 (28.1) 0	60 (83.3) 10 (13.9) 2 (2.8)	Codominant 0.43 (0.19-0.95) 0	0.018
mir4520-1 (17p13.1)	rs8078913 (PM)	CC CT TT	28 (32.2) 41 (47.1) 18 (20.7)	10 (16.7) 31 (51.7) 19 (31.7)	Additive 1.7 (1.06-2.74)	0.026
mir3166 (11q14.2)	rs35854553 (PM)	AA AT	86 (89.6) 10 (10.4)	38 (76) 12 (24)	Codominant 2.72 (1.08-6.83)	0.033
mir548AL (11q13.4)	rs515924 (seed)	AA AG GG	74 (77.1) 21 (21.9) 1 (1.0)	46 (62.2) 27 (36.5) 1 (1.4)	Dominant 2.05 (1.05-4)	0.034

Abbreviation: PM: pre-miRNA, M: mature, OR: Odd ratio, CI: Confidence interval

Genotype association study considering B-ALL subtype

In the Spanish cohort we performed the subtype analysis in B-hyperdiploid group (n=56) and in the group of patients who carried ETV-RUNX1 fusion gene (n=37). In the Slovenian population the analyses were not performed due to the low number of patients in B-hyperdiploid (n=9) and ETV-RUNX1 (n=12) groups.

In the Spanish cohort rs10406069 at mir5196 was the SNP most significantly associated with B-hyperdiploid ALL (OR: 1.78; CI 95%: 1.11-2.85; p=0.0001), with the G allele being the protection allele. This SNP remained statistically significant after FDR correction (p=0.017). Other 10 SNPs showed associations, but they did not reach a significant value after correction (Table 43). Regarding patients carrying ETV6-RUNX1 fusion gene, none of the SNPs reached significant results after correction (Table 44).

Table 43: Polymorphism associated with B-hyperdiploid ALL risk in the Spanish population.

Gene (Location)	SNP (Position)	Genotype	N (controls) N=330	N (cases) N=54	OR (CI 95%)	P
mir5196 (19q13.12)	rs10406069 (PM)	GG AG AA	208 (63.8) 103 (31.6) 15 (4.6)	21 (38.9) 33 (61.1) 0	Additive 1.78 (1.11-2.85)	0.0001*
mir6128 (12q21.31)	rs2682818 (PM)	CC AC AA	241 (73.3) 82 (24.9) 6 (1.8)	48 (88.9) 6 (11.1) 0	Dominant 0.34 (0.14-0.83)	0.007
mir5090 (7q22.1)	rs3823658 (seed)	GG AG AA	238 (72.1) 89 (27) 3 (0.9)	40 (74.1) 10 (18.5) 4 (7.4)	Recessive 8.72 (1.9-40.12)	0.007
mir4751 (19q13.33)	rs8667 (PM)	GG AG AA	109 (38) 133 (46.3) 45 (15.7)	24 (44.4) 28 (51.9) 2 (3.7)	Recessive 0.21 (0.05-0.88)	0.007
mir1302-4 (2q33.3)	rs10173558 (PM)	TT CT CC	251 (76.1) 77 (23.3) 2 (0.6)	49 (90.7) 5 (9.3) 0 (0.0)	Dominant 0.32 (0.12-0.84)	0.008
mir3683 (7p22.1)	rs6977967 (PM)	AA AG GG	201 (60.9) 113 (34.2) 16 (4.8)	42 (77.8) 9 (16.7) 3 (5.6)	Dominant 0.45 (0.23-0.88)	0.013
mir4742 (1q42.11)	rs7522956 (PM)	AA AC CC	204 (62) 110 (33.4) 15 (4.6)	26 (48.1) 21 (38.9) 7 (13)	Additive 1.75 (1.12-2.73)	0.014
mir4521 (17p13.1)	rs76800617 (PM)	AA AG	315 (95.5) 15 (4.5)	47 (87) 7 (13)	Codominant 3.13 (1.21-8.07)	0.027
mir3144 (6q22.31)	rs68035463 (PM)	CC AC AA	209 (63.7) 107 (32.6) 12 (3.7)	29 (53.7) 19 (35.2) 6 (11.1)	Recessive 3.29 (1.18-9.18)	0.033
mir4653 (7q22.1)	rs11983381 (PM)	AA AG GG	227 (69) 96 (29.2) 6 (1.8)	27 (54) 22 (44) 1 (2)	Dominant 1.9 (1.04-3.47)	0.039
mir300 (14q32.31)	rs12894467 (PM)	CC CT TT	148 (45) 141 (42.9) 40 (12.2)	16 (30.2) 28 (52.8) 9 (17)	Dominant 1.89 (1.01-3.53)	0.040

Abbreviation: PM: pre-miRNA, OR: Odd ratio, CI: Confidence interval. *Significant after FDR correction.

Table 44: Polymorphisms in miRNAs associated with *ETV6-RUNX1* ALL risk in the Spanish population.

Gene (Location)	SNP (Position)	Genotype	N (controls) (N=330)	N (cases) (N=35)	OR(CI 95%)	P
mir595 (7q36.3)	rs4909237 (PM)	CC CT TT	233 (70.8) 90 (27.4) 6 (1.8)	20 (57.1) 11 (31.4) 4 (11.4)	Recessive 6.95 (1.86-25.9)	0.009
mir4293 (10p13)	rs12780876 (PM)	TT AT AA	186 (56.9) 117 (35.8) 24 (7.3)	14 (40) 14 (40) 7 (20)	Additive 1.89 (1.15-3.1)	0.013
mir300 (14q32.31)	rs12894467 (PM)	CC CT TT	148 (45) 141 (42.9) 40 (12.2)	16 (45.7) 19 (54.3) 0	-	0.021
mir5186 (3q25.1)	rs9842591 (PM)	CC AC AA	74 (26.5) 147 (52.7) 58 (20.8)	14 (46.7) 10 (33.3) 6 (20)	Dominant 0.41 (0.19-0.89)	0.025
mir4751 (19q13.33)	rs8667 (PM)	GG AG AA	109 (38) 133 (46.3) 45 (15.7)	7 (20) 21 (60) 7 (20)	Dominant 2.45 (1.03-5.8)	0.029
mir1269a (4q13.2)	rs73239138 (M)	GG AG AA	193 (59) 115 (35.2) 19 (5.8)	27 (77.1) 8 (22.9) 0	Dominant 1.00 0.43 (0.19-0.97)	0.030
mir4467 (7q22.1)	rs60871950 (M)	AA AG GG	88 (26.9) 163 (49.8) 76 (23.2)	4 (11.8) 18 (52.9) 12 (35.3)	Additive 1.75 (1.04-2.93)	0.031
mir585 (5q35.1)	rs62376935 (seed)	CC CT TT	297 (90) 31 (9.4) 2 (0.6)	26 (76.5) 8 (23.5) 0	Dominant 2.77 (1.16-6.61)	0.032
mir4789 (3q26.31)	rs78831152 (PM)	CC CT TT	279 (84.5) 49 (14.8) 2 (0.6)	30 (85.7) 3 (8.6) 2 (5.7)	Recessive 9.94 (1.36-72.8)	0.037
mir4432 (2p16.1)	rs243080 (PM	CC CT TT	98 (29.8) 175 (53.2) 56 (17)	10 (29.4) 13 (38.2) 11 (32.4)	Recessive 2.33 (1.08-5.06)	0.040

Abbreviation: PM: pre-miRNA, M: mature, OR: Odd ratio, CI: Confidence interval.

Bioinformatic analysis

miRNAs secondary structures prediction

We analyzed *in silico* the energy change (|ΔΔG|) and the secondary structures modifications of the SNPs associated with B-ALL risk (rs12402181 at mir-3117-3p and rs62571442 at mir-3689d2) and the SNP associated with B-hyperdiploid ALL risk (rs10406069 at mir-5196).

The SNP rs12402181 is located in the seed region of mir-3117-3p. The change from G to A allele did not show either and energy change or change in the secondary structure. In the case of rs62571442, located in the pre-miRNA of mir-3689d2, showed an energy change from -31.2 kcal/mol for the T allele to -30.0 kcal/mol for the risk allele C (1.2 kcal/mol). This SNP also produced an evident change in the secondary structure (Figure 30A). Finally, rs10406069 at mir-5196, located in the pre-miRNA, showed an energy change from -53.50 kcal/mol for the G allele to -53.9 kcal/mol for the A allele (-0.4 kcal/mol) and a slight change in the secondary structure (Figure 30B).

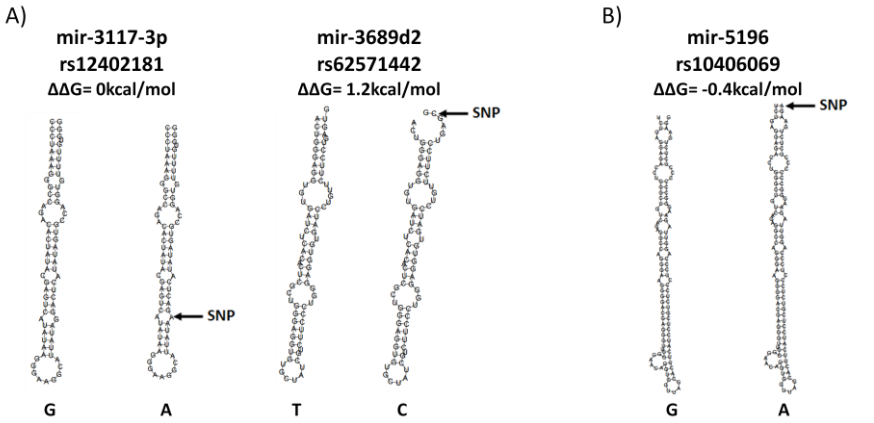

Figure 30: Secondary structures of the miRNAs showing significant SNPs associated with B-ALL risk (A) and B-hyperdiploid ALL (B), predicted by the miRNA SNP tool.

Pathway analysis

In order to evaluate the pathways that could be affected by the miRNAs with the most significant SNPs, we performed a pathway enrichment analysis by using miRWalk and the ConsensusPathDB web tools.

For mir-3117-3p, among the ten most significant pathways, we found the mitogen-activated protein kinase (MAPK) signaling pathway over-represented (p-value of 4.9×10^{-7}) (Table 45). In this pathway, mir-3117-3p targeted up to 24 genes (Table 46). Moreover, 7 out of top 10 enriched pathways were related with Ras signaling cascade, which is one of the MAPK pathways (Table 45). For mir-3689d2, among the ten most significant pathways, 6 were also related with Ras signaling, and it targeted up to 32 genes involved in MAPK signaling (Table 47 and 48). When we analyzed the putative target genes of both miRNAs together, the association for MAPK signaling pathway increased up to $p=5.75 \times 10^{-13}$, with both miRNAs targeting up to 55 genes of the pathway. Moreover, 8 out of the top 10 pathways are related with Ras cascade (Figure 31 and Table 49).

Table 45: The enriched patways for mir-3117-3p.

Pathway name	Set size	Candidates	P	q-value	Pathway source
MAPK signaling pathway - Homo sapiens (human)	257	24 (9.3%)	$4.94x10^{-07}$	0.000322	**KEGG**
Ras signaling pathway - Homo sapiens (human)	228	21 (9.2%)	$3.24x10^{-06}$	0.00106	**KEGG**
Choline metabolism in cancer - Homo sapiens (human)	101	13 (12.9%)	$7.18x10^{-06}$	0.00149	**KEGG**
PDGF signaling pathway	27	7 (25.9%)	$9.14x10^{-06}$	0.00149	**BioCarta**
Renal cell carcinoma - Homo sapiens (human)	66	10 (15.2%)	$1.95x10^{-05}$	0.00254	**KEGG**
FoxO signaling pathway - Homo sapiens (human)	134	14 (10.4%)	$3.61x10^{-05}$	0.00317	**KEGG**
ErbB signaling pathway - Homo sapiens (human)	87	11 (12.6%)	$4.32x10^{-05}$	0.00317	**KEGG**
Signaling by EGFR in Cancer	15	5 (33.3%)	$4.87x10^{-05}$	0.00317	**Reactome**
Constitutive Signaling by Ligand-Responsive EGFR Cancer Variants	15	5 (33.3%)	$4.87x10^{-05}$	0.00317	**Reactome**
Signaling by Ligand-Responsive EGFR Variants in Cancer	15	5 (33.3%)	$4.87x10^{-05}$	0.00317	**Reactome**
Diseases of signal transduction	180	16 (8.9%)	$7.45x10^{-05}$	0.00442	**Reactome**

Table 46: Genes of MAPK signalling pathway targeted by mir-3117-3p.

Entrez-gene ID	*Entrez-gene name*
3845	*KRAS* : Kirsten rat sarcoma viral oncogene homolog
5156	*PDGFRA* : platelet-derived growth factor receptor, alpha polypeptide
6197	*RPS6KA*3 : ribosomal protein S6 kinase, 90kDa, polypeptide 3
51776	*ZAK* : sterile alpha motif and leucine zipper containing kinase AZK
6722	*SRF* : serum response factor (c-fos serum response element-binding transcription factor)
2317	*FLNB* : filamin B, beta
23162	*MAPK8IP3* : mitogen-activated protein kinase 8 interacting protein 3
6789	*STK4* : serine/threonine kinase 4
4763	*NF1* : neurofibromin 1
6416	*MAP2K4* : mitogen-activated protein kinase kinase 4
8913	*CACNA1G* : calcium channel, voltage-dependent, T type, alpha 1G subunit
786	*CACNG1* : calcium channel, voltage-dependent, gamma subunit 1
5908	*RAP1B* : RAP1B, member of RAS oncogene family
10000	*AKT3* : v-akt murine thymoma viral oncogene homolog 3
5922	*RASA2* : RAS p21 protein activator 2
5923	*RASGRF1* : Ras protein-specific guanine nucleotide-releasing factor 1
5924	*RASGRF2* : Ras protein-specific guanine nucleotide-releasing factor 2
2885	*GRB2* : growth factor receptor-bound protein 2
1386	*ATF2* : activating transcription factor 2
7046	*TGFBR1* : transforming growth factor, beta receptor 1
59283	*CACNG8* : calcium channel, voltage-dependent, gamma subunit 8
9693	*RAPGEF2* : Rap guanine nucleotide exchange factor (GEF) 2
5599	*MAPK8* : mitogen-activated protein kinase 8
6654	*SOS1* : son of sevenless homolog 1 (Drosophila)

Table 47: The top ten enriched patways for mir-3689d2.

Pathway name	Set size	Candidates	P	q-value	Pathway source
Axon guidance	459	59 (12.9%)	$4.69x10^{-13}$	$4.24e^{-10}$	**Reactome**
Developmental Biology	586	68 (11.6%)	$9.47x10^{-13}$	$4.28e^{-10}$	**Reactome**
NGF signalling via TRKA from the plasma membrane	310	43 (13.9%)	$7.24x10^{-11}$	$1.84e^{-08}$	**Reactome**
Signalling by NGF	386	49 (12.7%)	$8.14x10^{-11}$	1.84e-08	**Reactome**
Signaling by PDGF	301	41 (13.6%)	$3.61x10^{-10}$	6.52e-08	**Reactome**
Signaling by EGFR	292	40 (13.7%)	$5.03x10^{-10}$	7.57e-08	**Reactome**
DAP12 interactions	298	40 (13.4%)	$9.29x10^{-10}$	1.2e-07	**Reactome**
Downstream signal transduction	279	38 (13.6%)	$1.64x10^{-09}$	1.25e-07	**Reactome**
Downstream signaling of activated FGFR2	267	37 (13.9%)	$1.66x10^{-09}$	1.25e-07	**Reactome**
Downstream signaling of activated FGFR1	267	37 (13.9%)	$1.66x10^{-09}$	1.25e-07	**Reactome**

Table 48: Genes of MAPK signalling pathway targeted by mir-3689d.

Entrez-gene ID	***Entrez-gene name***
5154	*PDGFA* : platelet-derived growth factor alpha polypeptide
5155	*PDGFB* : platelet-derived growth factor beta polypeptide
5159	*PDGFRB* : platelet-derived growth factor receptor, beta polypeptide
4137	*MAPT* : microtubule-associated protein tau
9479	*MAPK8IP1* : mitogen-activated protein kinase 8 interacting protein 1
6195	*RPS6KA1* : ribosomal protein S6 kinase, 90kDa, polypeptide 1
6722	*SRF* : serum response factor (c-fos serum response element-binding transcription factor)
208	*AKT2* : v-akt murine thymoma viral oncogene homolog 2
627	*BDNF* : brain-derived neurotrophic factor
5778	*PTPN7* : protein tyrosine phosphatase, non-receptor type 7
51347	*TAOK3* : TAO kinase 3
5494	*PPM1A* : protein phosphatase, Mg2+/Mn2+ dependent, 1A
2768	*GNA12* : guanine nucleotide binding protein (G protein) alpha 12
2250	*FGF5* : fibroblast growth factor 5
10454	*TAB1* : TGF-beta activated kinase 1/MAP3K7 binding protein 1
782	*CACNB1* : calcium channel, voltage-dependent, beta 1 subunit
10000	*AKT3* : v-akt murine thymoma viral oncogene homolog 3
1850	*DUSP8* : dual specificity phosphatase 8
1852	*DUSP9* : dual specificity phosphatase 9
57551	*TAOK1* : TAO kinase 1
5609	*MAP2K7* : mitogen-activated protein kinase kinase 7
5532	*PPP3CB* : protein phosphatase 3, catalytic subunit, beta isozyme
1956	*EGFR* : epidermal growth factor receptor
5058	*PAK1* : p21 protein (Cdc42/Rac)-activated kinase 1
5578	*PRKCA* : protein kinase C, alpha
5579	*PRKCB* : protein kinase C, beta
2002	*ELK1* : ELK1, member of ETS oncogene family
27092	*CACNG4* : calcium channel, voltage-dependent, gamma subunit 4
2005	*ELK4* : ELK4, ETS-domain protein (SRF accessory protein 1)
5594	*MAPK1* : mitogen-activated protein kinase 1
5595	*MAPK3* : mitogen-activated protein kinase 3
3554	*IL1R1* : interleukin 1 receptor, type I

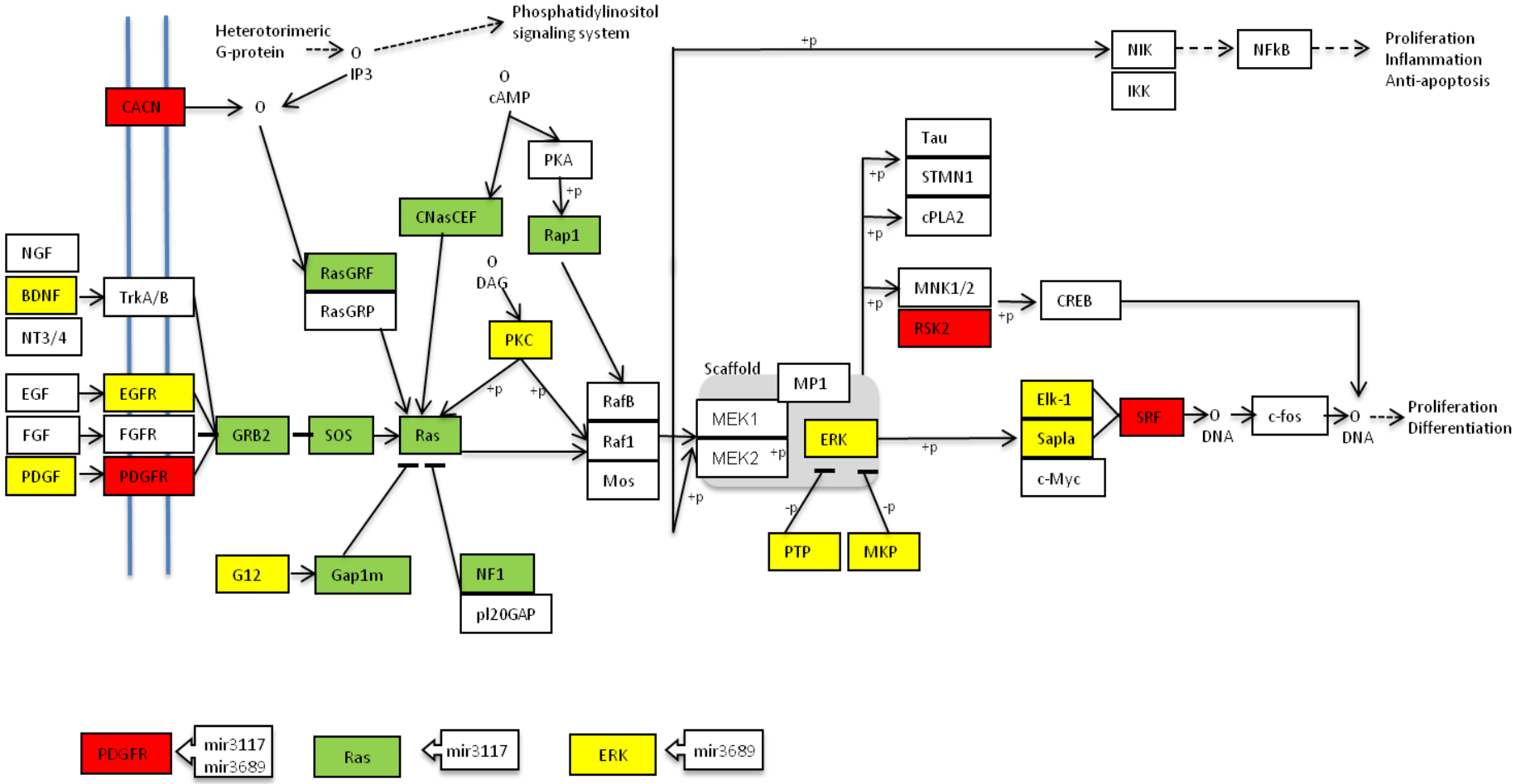

Figure 31: Genes of the MAPK signaling pathway targeted by mir-3117-3p and mir-3689d2 (adapted from KEGG database).

Table 49: Top ten enriched patways for mir-3117-3p and mir-3689d2.

Pathway name	Set size	Candidates contained	P	q-value	Pathway source
Axon guidance	459	83 (18.1%)	$2.77x10^{-15}$	$3.13e^{-12}$	**Reactome**
Developmental Biology	586	96 (16.4%)	$1.09x10^{-14}$	$6.16e^{-12}$	**Reactome**
NGF signalling via TRKA from the plasma membrane	310	62 (20.0%)	$1.1x10^{-13}$	$3.86e^{-11}$	**Reactome**
Signalling by NGF	386	71 (18.4%)	$1.37x10^{-13}$	$3.86e^{-11}$	**Reactome**
MAPK signaling pathway - Homo sapiens (human)	257	54 (21.0%)	$5.75x10^{-13}$	$1.3e^{-10}$	**KEGG**
Signaling by EGFR	292	58 (19.9%)	$9.62x10^{-13}$	$1.65e^{-10}$	**Reactome**
Signaling by PDGF	301	59 (19.6%)	$1.1x10^{-12}$	$1.65e^{-10}$	**Reactome**
Signaling by FGFR3	270	55 (20.4%)	$1.32x10^{-12}$	$1.65e^{-10}$	**Reactome**
Signaling by FGFR4	270	55 (20.4%)	$1.32x10^{-12}$	$1.65e^{-10}$	**Reactome**
Signaling by FGFR1	271	55 (20.3%)	$1.54x10^{-12}$	$1.74e^{-10}$	**Reactome**

Finally, for mir-5196, we did not find any pathway related with B-hyperdiploid ALL susceptibility (Table 50). However, we found up to 8 genes involved in histone modifications or centromeric chromatid formation (*PTPN11, CEP68, KDM5A, RFC5, TADA2A, BRMS1L, KIF24, ATXN7L3B*) that could be involved in aneuploidy.

Table 50: Top ten enriched patways for mir-5196.

Pathway name	Set size	Candidates	P	q-value	Pathway source
Neurotrophin signaling pathway - Homo sapiens (human)	120	15 (12.5%)	$1.05x10^{-05}$	0.00701	**KEGG**
AMPK signaling pathway - Homo sapiens (human)	124	14 (11.3%)	$6.5x10^{-05}$	0.0184	**KEGG**

DISCUSSION

In the current study, rs12402181 in mir-3117 and rs62571442 in mir-3689d2 showed statistically significant association with B-ALL risk in the Spanish cohort, and were validated in the Slovenian population. In the subtype analysis, the most interesting result was the association between rs10406069 in mir-5196 and the risk of developing B-hyperdiploid ALL, which remains statistically significant after FDR correction. Our results point to a putative role of SNPs in B-ALL susceptibility.

The SNP rs12402181 in mir-3117-3p was associated with an increased risk of developing B-ALL under the additive model, being the A allele the risk allele. This result was observed in the Spanish cohort (p=0.047) and validated in the Slovenian population (p=0.041). The association increased when both populations were analyzed together (p=0.006). These results support the

idea that the A allele could represent a risk allele of low penetrance for B-ALL independently of the population studied. rs12402181 is located in the seed region of mir-3117-3p, therefore, the change of the G allele for the A allele could affect the accurate recognition of its target mRNA sequences. Among the target genes of mir-3117-3p, *in silico* analysis determined that genes of MAPK signaling pathway are over-represented, mainly those of the MAPK/ERK family or classical pathway (Kamburov et al. 2013, Dweep and Gretz 2015, Kanehisa et al. 2016). Interestingly, the genes predicted to be targeted for mir-3117-3p are in the first steps of the cascade (*CACNA1D, CACNG1, CACNG8*, *PDGFR*, *GRB2*, *SOS1* and *RAS),* which in turn could produce the deregulation of the following steps. In addition, it was predicted that the risk allele of this SNP causes the loss of up to 7 target genes in this pathway (*CACNA1D*, *CACNG1, CACNG8*, *SRF*, *MAP2K4*, *RAP1B* and *GRB2*) (Gong et al. 2015), which could lead to an increased in their expression. Aberrant expression of this pathway is a major and highly prevalent oncogenic event in many human cancers (Masliah-Planchon et al. 2015), including childhood ALL (Barbosa et al. 2014, Case et al. 2008). Of note, miRNAs have been suggested to be involved in this process (Masliah-Planchon et al. 2015). In summary, failed recognition between mir-3117-3p and its targets due to the change of the G allele for the A allele of rs12402181 in the seed region could contribute to leukemogenesis by leading to an aberrant activation of the RAS-MAPK pathway.

The second SNP, rs62571442 in mir-3689d2, was associated with an increased risk of B-ALL under the additive model, being the C allele the allele which increased the risk of B-ALL. This result was observed in the Spanish cohort (p=0.039), as well as in the Slovenian population (p=0.001), indicating that this SNP could be a general marker for B-ALL. The change C>T in this SNP, located in the pre-miRNA sequence, modified the secondary structure and induced a positive energy change of 1.2 kcal/mol for the C risk allele. The hairpin structure changes from stable (U:A) to unstable status (C:A). When the SNP decreases the stability, the product of mature miRNA is reduced, which in turn may increase the target gene expression (Gong et al. 2012). In the pathway analysis of mir-3689d2, 6 out of the top 10 enriched pathways were again related to Ras signaling, aforementioned. Therefore, a decreased expression of this miRNA could increase the expression of Ras-related genes.

When pathways analysis for mir-3117-3p and mir-3689d2 were performed together, 8 out of the 10 first pathways were related with Ras signaling, and the association obtained is higher, as expected. Therefore, both miRNAs could contribute to the activation of this pathway: on the one hand, because of the loss of target recognition due to the SNP in the seed region of mir-

3117-3p, and on the other hand, because of the reduction in mature levels of mir3689d2 due to the risk allele in the hairpin structure. The fact that these SNPs showed the same trend in both populations, indicate that they could be considered as B-ALL markers. It would be interesting to study these associations in other populations.

In the subtype analysis we have found a strong association between rs10406069 AA genotype in mir-5196 and the risk to develop B-hyperdiploid ALL (p=0.0001), remaining significant after FDR correction (p=0.017). The rs10406069 SNP is located in the pre-miRNA, where the A risk allele increased slightly the stability of the hairpin structure (ΔΔG=-0.4kcal/mol), which in turn may increase mir5196 expression (Gong et al. 2012). In fact, it has already been described that mir-5196 is overexpressed in B-hyperdiploid ALL (Schotte et al. 2011). This overexpression could repress its target genes. Among them, we found, for instance, *PTPN11*, a gene that encodes for Shp2, a tyrosine phosphatase whose depletion causes chromosome instability (Liu et al. 2012), which could be involved in hyperdiploid ALL. In addition, mir-5196 is hosted in CD22 gene, and it has been suggested that intronic miRNAs may affect the levels of their own host genes (Gao et al. 2012, Hinske et al. 2010). Moreover, this SNP also produces a missense variation (p.Gly745Asp) in exon 12 of *CD22*, which could affect its function. It has been described that alterations in CD22 could affect the transmission of apoptotic signals in primary leukemic cells from infants with B-ALL (Ma et al. 2012, Uckun et al. 2010). Therefore, rs10406069 A allele could increase the risk of B-hyperdiploid ALL either by repression of its target genes (*PTPN11* and *CD22*) or by its direct effect in *CD22*.

This study has some limitations that might be addressed, such as the relatively high failure rate in genotyping technique. However, this high chance of failure was accepted from the beginning of the study, because despite the predicted low score for genotyping, no other design option to amplify the polymorphisms in question was possible. Another possible weakness is the inaccuracy of the prediction algorithms of the databases used (Lee et al. 2015, Akhtar et al. 2016), but nowadays this limitation has to be assumed.

In conclusion, rs12402181 in mir-3117-3p and rs62571442 in mir-3689d2 could be involved in B-ALL susceptibility through its effect on the regulation of MAPK signaling-related pathways. The SNP rs10406069 in mir-5196 could be implicated in B-hyperdiplod ALL susceptibility by changes in expression levels of its own host gene or a direct effect on *CD22* gene.

ACKNOWLEDGMENTS

This study was funded by the Basque Government (IT661-13), UPV/EHU (UFI11/35) and RTICC (RD12/0036/0060, RD12/0036/0036). AGC was supported by a pre-doctoral grant from the Basque Government.

AUTHOR CONTRIBUTIONS

AGC, AGO: performed research, analysed data, and wrote the paper. IMG: performed research and analysed data. AS, AGA, IA, AN, ACB: provided clinical information and patient samples, analysed and interpreted data and wrote the paper.

CONFLICT OF INTEREST

The authors declare no competing financial interests

REFERENCES

Akhtar, M. M., Micolucci, L., Islam, M. S., Olivieri, F. and Procopio, A. D. (2016) 'Bioinformatic tools for microRNA dissection', *Nucleic Acids Res,* 44(1), 24-44.

Barbosa, T. C., Andrade, F. G., Lopes, B. A., de Andrade, C. F., Mansur, M. B., Emerenciano, M. and Pombo-de-Oliveira, M. S. (2014) 'Impact of mutations in FLT3, PTPN11 and RAS genes on the overall survival of pediatric B cell precursor acute lymphoblastic leukemia in Brazil', *Leuk Lymphoma,* 55(7), 1501-9.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the false discovery rate: A practical and powerful approach to multiple testing', Soc Series B(

Case, M., Matheson, E., Minto, L., Hassan, R., Harrison, C. J., Bown, N., Bailey, S., Vormoor, J., Hall, A. G. and Irving, J. A. (2008) 'Mutation of genes affecting the RAS pathway is common in childhood acute lymphoblastic leukemia', *Cancer Res,* 68(16), 6803-9.

Duyu, M., Durmaz, B., Gunduz, C., Vergin, C., Yilmaz Karapinar, D., Aksoylar, S., Kavakli, K., Cetingul, N., Irken, G., Yaman, Y., Ozkinay, F. and Cogulu, O. (2014) 'Prospective evaluation of whole genome microRNA expression profiling in childhood acute lymphoblastic leukemia', *Biomed Res Int,* 2014, 967585.

Dweep, H. and Gretz, N. (2015) 'miRWalk2.0: a comprehensive atlas of microRNA-target interactions', *Nat Methods,* 12(8), 697.

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., Matthews, L., May, B., Milacic, M., Rothfels, K., Shamovsky, V., Webber, M., Weiser, J., Williams, M., Wu, G., Stein, L., Hermjakob, H. and D'Eustachio, P. (2016) 'The Reactome pathway Knowledgebase', *Nucleic Acids Res,* 44(D1), D481-7.

Gao, X., Qiao, Y., Han, D., Zhang, Y. and Ma, N. (2012) 'Enemy or partner: relationship between intronic micrornas and their host genes', *IUBMB Life,* 64(10), 835-40.

Gong, J., Liu, C., Liu, W., Wu, Y., Ma, Z., Chen, H. and Guo, A. Y. (2015) 'An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools', *Database (Oxford),* 2015, bav029.

Gong, J., Tong, Y., Zhang, H. M., Wang, K., Hu, T., Shan, G., Sun, J. and Guo, A. Y. (2012) 'Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis', *Hum Mutat,* 33(1), 254-63.

Greaves, M. (2006) 'Infection, immune responses and the aetiology of childhood leukaemia', *Nat Rev Cancer,* 6(3), 193-203.

Gutierrez-Camino, A., Lopez-Lopez, E., Martin-Guerrero, I., Piñan, M. A., Garcia-Miguel, P., Sanchez-Toledo, J., Carbone Bañeres, A., Uriz, J., Navajas, A. and Garcia-Orad, A. (2014) 'Noncoding RNA-related polymorphisms in pediatric acute lymphoblastic leukemia susceptibility', *Pediatr Res,* 75(6), 767-73.

Gutiérrez-Camino, Á., López-López, E., Martín-Guerrero, I., Sánchez-Toledo, J., García de Andoin, N., Carboné Bañeres, A., García-Miguel, P., Navajas, A. and García-Orad, Á. (2013) 'Intron 3 of the ARID5B gene: a hot spot for acute lymphoblastic leukemia susceptibility', *J Cancer Res Clin Oncol,* 139(11), 1879-86.

Hasani, S. S., Hashemi, M., Eskandari-Nasab, E., Naderi, M., Omrani, M. and Sheybani-Nasab, M. (2013) 'A functional polymorphism in the miR-146a gene is associated with the risk of childhood acute lymphoblastic leukemia: a preliminary report', *Tumour Biol*.

Healy, J., Richer, C., Bourgey, M., Kritikou, E. A. and Sinnett, D. (2010) 'Replication analysis confirms the association of ARID5B with childhood B-cell acute lymphoblastic leukemia', *Haematologica,* 95(9), 1608-11.

Hinske, L. C., Galante, P. A., Kuo, W. P. and Ohno-Machado, L. (2010) 'A potential role for intragenic miRNAs on their hosts' interactome', *BMC Genomics,* 11, 533.

Johanson, T. M., Skinner, J. P., Kumar, A., Zhan, Y., Lew, A. M. and Chong, M. M. (2014) 'The role of microRNAs in lymphopoiesis', *Int J Hematol,* 100(3), 246-53.

Kamburov, A., Stelzl, U., Lehrach, H. and Herwig, R. (2013) 'The ConsensusPathDB interaction database: 2013 update', *Nucleic Acids Res,* 41(Database issue), D793-800.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2016) 'KEGG as a reference resource for gene and protein annotation', *Nucleic Acids Res,* 44(D1), D457-62.

Kozomara, A. and Griffiths-Jones, S. (2014) 'miRBase: annotating high confidence microRNAs using deep sequencing data', *Nucleic Acids Res,* 42(Database issue), D68-73.

Layton Tovar, C. F. and Mendieta Zerón, H. (2016) 'Intracellular Signaling Pathways Involved in Childhood Acute Lymphoblastic Leukemia; Molecular Targets', *Indian J Hematol Blood Transfus,* 32(2), 141-53.

Lee, Y. J., Kim, V., Muth, D. C. and Witwer, K. W. (2015) 'Validated MicroRNA Target Databases: An Evaluation', *Drug Dev Res,* 76(7), 389-96.

Liu, X., Zheng, H. and Qu, C. K. (2012) 'Protein tyrosine phosphatase Shp2 (Ptpn11) plays an important role in maintenance of chromosome stability', *Cancer Res,* 72(20), 5296-306.

Ma, H., Qazi, S., Ozer, Z., Gaynon, P., Reaman, G. H. and Uckun, F. M. (2012) 'CD22 Exon 12 deletion is a characteristic genetic defect of therapy-refractory clones in paediatric acute lymphoblastic leukaemia', *Br J Haematol,* 156(1), 89-98.

Marques, S. C., Laursen, M. B., Bødker, J. S., Kjeldsen, M. K., Falgreen, S., Schmitz, A., Bøgsted, M., Johnsen, H. E. and Dybkaer, K. (2015) 'MicroRNAs in B-cells: from normal differentiation to treatment of malignancies', *Oncotarget,* 6(1), 7-25.

Masliah-Planchon, J., Garinet, S. and Pasmant, E. (2015) 'RAS-MAPK pathway epigenetic activation in cancer: miRNAs in action', *Oncotarget*.

Migliorini, G., Fiege, B., Hosking, F. J., Ma, Y., Kumar, R., Sherborne, A. L., da Silva Filho, M. I., Vijayakrishnan, J., Koehler, R., Thomsen, H., Irving, J. A., Allan, J. M., Lightfoot, T., Roman, E., Kinsey, S. E., Sheridan, E., Thompson, P., Hoffmann, P., Nöthen, M. M., Mühleisen, T. W., Eisele, L., Zimmermann, M., Bartram, C. R., Schrappe, M., Greaves, M., Stanulla, M., Hemminki, K. and Houlston, R. S. (2013) 'Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype', *Blood,* 122(19), 3298-307.

Musilova, K. and Mraz, M. (2015) 'MicroRNAs in B-cell lymphomas: how a complex biology gets more complex', *Leukemia,* 29(5), 1004-17.

Orsi, L., Rudant, J., Bonaventure, A., Goujon-Bellec, S., Corda, E., Evans, T. J., Petit, A., Bertrand, Y., Nelken, B., Robert, A., Michel, G., Sirvent, N., Chastagner, P., Ducassou, S., Rialland, X., Hémon, D., Milne, E., Scott, R. J., Baruchel, A. and Clavel, J. (2012) 'Genetic polymorphisms and childhood acute lymphoblastic leukemia: GWAS of the ESCALE study (SFCE)', *Leukemia,* 26(12), 2561-4.

Papaemmanuil, E., Hosking, F. J., Vijayakrishnan, J., Price, A., Olver, B., Sheridan, E., Kinsey, S. E., Lightfoot, T., Roman, E., Irving, J. A., Allan, J. M., Tomlinson, I. P., Taylor, M., Greaves, M. and Houlston, R. S. (2009) 'Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia', *Nat Genet,* 41(9), 1006-10.

Prasad, R. B., Hosking, F. J., Vijayakrishnan, J., Papaemmanuil, E., Koehler, R., Greaves, M., Sheridan, E., Gast, A., Kinsey, S. E., Lightfoot, T., Roman, E., Taylor, M., Pritchard-Jones, K., Stanulla, M., Schrappe, M., Bartram, C. R., Houlston, R. S., Kumar, R. and Hemminki, K. (2010) 'Verification of the susceptibility loci on 7p12.2, 10q21.2, and 14q11.2 in precursor B-cell acute lymphoblastic leukemia of childhood', *Blood,* 115(9), 1765-7.

Pui, C. H. and Evans, W. E. (2006) 'Treatment of acute lymphoblastic leukemia', *N Engl J Med,* 354(2), 166-78.

Ross, J. A., Linabery, A. M., Blommer, C. N., Langer, E. K., Spector, L. G., Hilden, J. M., Heerema, N. A., Radloff, G. A., Tower, R. L. and Davies, S. M. (2013) 'Genetic variants modify susceptibility to leukemia in infants: a Children's Oncology Group report', *Pediatr Blood Cancer,* 60(1), 31-4.

Ryan, B. M., Robles, A. I. and Harris, C. C. (2010) 'Genetic variation in microRNA networks: the implications for cancer research.', *Nat Rev Cancer,* 10(6), 389-402.

Sambrook, J. and Russell, D. (2001) *Molecular cloning: A laboratory manual,* 3rd ed.*,* Cold Spring Harbor, USA: CSHL Press.

Schotte, D., Akbari Moqadam, F., Lange-Turenhout, E. A., Chen, C., van Ijcken, W. F., Pieters, R. and den Boer, M. L. (2011) 'Discovery of new microRNAs by small RNAome deep sequencing in childhood acute lymphoblastic leukemia', *Leukemia,* 25(9), 1389-99.

Schotte, D., Chau, J. C., Sylvester, G., Liu, G., Chen, C., van der Velden, V. H., Broekhuis, M. J., Peters, T. C., Pieters, R. and den Boer, M. L. (2009) 'Identification of new microRNA genes and aberrant microRNA profiles in childhood acute lymphoblastic leukemia', *Leukemia,* 23(2), 313-22.

Sherborne, A. L., Hosking, F. J., Prasad, R. B., Kumar, R., Koehler, R., Vijayakrishnan, J., Papaemmanuil, E., Bartram, C. R., Stanulla, M., Schrappe, M., Gast, A., Dobbins, S. E., Ma, Y., Sheridan, E., Taylor, M., Kinsey, S. E., Lightfoot, T., Roman, E., Irving, J. A., Allan, J. M., Moorman, A. V., Harrison, C. J., Tomlinson, I. P., Richards, S., Zimmermann, M., Szalai, C., Semsei, A. F., Erdelyi, D. J., Krajinovic, M., Sinnett, D., Healy, J., Gonzalez Neira, A., Kawamata, N., Ogawa, S., Koeffler, H. P., Hemminki, K., Greaves, M. and Houlston, R. S. (2010) 'Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk', *Nat Genet,* 42(6), 492-4.

Srivastava, K. and Srivastava, A. (2012) 'Comprehensive review of genetic association studies and meta-analyses on miRNA polymorphisms and cancer risk', *PLoS One,* 7(11), e50966.

Tong, N., Xu, B., Shi, D., Du, M., Li, X., Sheng, X., Wang, M., Chu, H., Fang, Y., Li, J., Wu, D. and Zhang, Z. (2014) 'Hsa-miR-196a2 polymorphism increases the risk of acute lymphoblastic leukemia in Chinese children', *Mutat Res,* 759, 16-21.

Treviño, L. R., Yang, W., French, D., Hunger, S. P., Carroll, W. L., Devidas, M., Willman, C., Neale, G., Downing, J., Raimondi, S. C., Pui, C. H., Evans, W. E. and Relling, M. V. (2009) 'Germline genomic variants associated with childhood acute lymphoblastic leukemia', *Nat Genet,* 41(9), 1001-5.

Uckun, F. M., Goodman, P., Ma, H., Dibirdik, I. and Qazi, S. (2010) 'CD22 EXON 12 deletion as a pathogenic mechanism of human B-precursor leukemia', *Proc Natl Acad Sci U S A,* 107(39), 16852-7.

Xia, L., Ren, Y., Fang, X., Yin, Z., Li, X., Wu, W., Guan, P. and Zhou, B. (2014) 'Prognostic role of common microRNA polymorphisms in cancers: evidence from a meta-analysis', *PLoS One,* 9(10), e106799.

Xu, H., Yang, W., Perez-Andreu, V., Devidas, M., Fan, Y., Cheng, C., Pei, D., Scheet, P., Burchard, E. G., Eng, C., Huntsman, S., Torgerson, D. G., Dean, M., Winick, N. J., Martin, P. L., Camitta, B. M., Bowman, W. P., Willman, C. L., Carroll, W. L., Mullighan, C. G., Bhojwani, D., Hunger, S. P., Pui, C. H., Evans, W. E., Relling, M. V., Loh, M. L. and Yang, J. J. (2013) 'Novel Susceptibility Variants at 10p12.31-12.2 for Childhood Acute Lymphoblastic Leukemia in Ethnically Diverse Populations', *J Natl Cancer Inst*.

Zhang, H., Yang, J. H., Zheng, Y. S., Zhang, P., Chen, X., Wu, J., Xu, L., Luo, X. Q., Ke, Z. Y., Zhou, H., Qu, L. H. and Chen, Y. Q. (2009) 'Genome-wide analysis of small RNA and novel MicroRNA discovery in human acute lymphoblastic leukemia based on extensive sequencing approach', *PLoS One,* 4(9), e6849.

ANNEX II

Annex table 1: All the results of SNPs in miRNA processing genes

SNP	Gene	Genotype	N (controls) (N=387)	N(cases) (N=213)	OR(CI 95%)	P dom
		TT	247 (72.0)	138 (73.0)	Dominant	
rs139919	TNRC6B	CT	93 (27.1)	43 (22.8)	1	0.80400
		CC	3 (0.9)	8 (4.2)	0.95 (0.64-1.42)	
		TT	139 (40.4)	83 (42.1)	Dominant	
rs197412	DDX20	CT	152 (44.2)	75 (38.1)	1	0.6949
		CC	53 (15.4)	39 (19.8)	0.93 (0.65-1.33)	
		GG	283 (81.6)	159 (79.1)	Dominant	
rs636832	AGO1	AG	61 (17.6)	35 (17.4)	1	0.48550
		AA	3 (0.9)	7 (3.5)	1.17 (0.76-1.80)	
		GG	231 (67.7)	128 (65.0)	Dominant	
rs6877842	DROSHA	CG	99 (29.0)	58 (29.4)	1	0.5124
		CC	11 (3.2)	11 (5.6)	1.13 (0.78 -1.64)	
		GG	188 (55.6)	108 (54.3)	Dominant	
rs2293939	AGO2	AG	132 (39.1)	77 (38.7)	1	0.7614
		AA	18 (5.3)	14 (7.0)	1.06 (0.74-1.50)	
		CC	194 (56.1)	110 (55.0)	Dominant	
rs1209904	DICER1	CT	127 (36.7)	77 (38.5)	1	0.8086
		TT	25 (7.2)	13 (6.5)	1.04 (0.74-1.48)	
		GG	201 (59.1)	125 (62.8)	Dominant	
rs6497759	TNRC6A	AG	117 (34.4)	64 (32.2)	1	0.3962
		AA	22 (6.5)	10 5.0	0.86 (0.60-1.23)	
		GG	302 (87.3)	175 (89.3)	Dominant	
rs417309	DGCR8	AG	41 (11.8)	20 (10.2)	1	0.4874
		AA	3 (0.9)	1 (0.5)	0.82 (0.47-1.43)	
		CC	301 (87.2)	173 (86.1)	Dominant	
rs3764942	SMAD5	TC	40 (11.6)	26 (12.9)	1	0.6960
		TT	4 (1.2)	2 (1.0)	1.11 (0.67-1.84)	
		GG	307 (88.2)	179 (89.5)	Codominante	
rs11644694	CNOT1	AG	41 (11.8)	21 (10.5)	1	0.6469
					0.88 (0.5-1.53)	
		GG	208 (59.8)	105 (52.2)	Dominant	
rs37060	CNOT1	AG	123 (35.3)	81 (40.3)	1	0.08636
		AA	17 (4.9)	15 (7.5)	1.36 (0.96-1.93)	
		CC	134 (38.7)	97 (49.2)	Dominant	
rs11866002	CNOT1	CT	174 (50.3)	80 (40.6)	1	0.01742
		TT	38 (11.0)	20 (10.2)	0.65 (0.46-0.93)	
		CC	270 (78.3)	143 (74.5)	Dominant	
rs2740348	GEMIN4	CG	69 (20.0)	47 (24.5)	1	0.3213
		GG	6 (1.7)	2 (1.0)	1.23 (0.82-1.86)	
		AA	255 (73.3)	144 (73.1)	Dominant	
rs1062923	GEMIN4	AG	87 (25.0)	45 (22.8)	1	0.9638
		GG	6 (1.7)	8 (4.1)	1.01 (0.68-1.50)	

Annex table 1: All the results of SNPs in miRNA processing genes (continuation)

SNP	Gene	Genotype	N (controls) (N=387)	N(cases) (N=213)	OR(CI 95%)	P dom
		AA	113 (33.5)	66 (34.6)	Dominant	
rs1106841	XPO5	AC	166 (49.3)	98 (51.3)	1	0.8114
		CC	58 (17.2)	27 (14.1)	0.96 (0.66-1.39)	
		AA	125 (37.0)	74 (38.9)	Dominant	
rs3805500	DROSHA	AG	158 (46.7)	87 (45.8)	1	0.6550
		GG	55 (16.3)	29 (15.3)	0.92 (0.64-1.33)	
		TT	156 (45.1)	76 (39.8)	Dominant	
rs1057035	DICER1	CT	152 (43.9)	94 (49.2)	1	0.2348
		CC	38 (11.0)	21 (11.0)	1.24 (0.87-1.78)	
		TT	220 (63.8)	132 (66.7)	Dominant	
rs13078	DICER1	AT	111 (32.2)	58 (29.3)	1	0.4952
		AA	14 (4.1)	8 (4.0)	0.88 (0.61-1.27)	
		GG	85 (24.4)	44 (22.1)	Dominant	
rs1640299	DGCR8	TG	182 (52.1)	124 (62.3)	1	0.55013
		TT	82 (23.5)	31 (15.6)	1.13 (0.75-1.72)	
		GG	193 (55.9)	113 (58.9)	Dominant	
rs10719	DROSHA	AG	134 (38.8)	64 (33.3)	1	0.5132
		AA	18 (5.2)	15 (7.8)	0.89 (0.62-1.27)	
		GG	304 (88.6)	177 (91.7)	Dominant	
rs1106042	PIWIL1	AG	36 (10.5)	15 (7.8)	1	0.2524
		AA	3 (0.9)	1 (0.5)	0.70 (0.38-1.30)	
		GG	77 (22.8)	53 (27.0)	Dominant	
rs784567	TARBP2	AG	181 (53.6)	95 (48.5)	1	0.2712
		AA	80 (23.7)	48 (24.5)	0.80 (0.53-1.19)	
		TT	273 (78.4)	160 (80.0)	Dominant	
rs1974777	GEMIN5	TC	72 (20.7)	38 (19.0)	1	
		CC	3 (0.9)	2 (1.0)	0.91 (0.59-1.40)	
		CC	138 (40.4)	77 (39.3)	Dominant	
rs14035	RAN	CT	164 (48.0)	88 (44.9)	1	0.8081
		TT	40 (11.7)	31 (15.8)	1.05 (0.73-1.50)	
		CC	212 (63.1)	136 (70.5)	Dominant	
rs3812265	CNOT4	CT	109 (32.4)	49 (25.4)	1	0.08367
		TT	15 (4.5)	8 (4.1)	0.72 (0.49-1.05)	
		AA	90 (25.9)	57 (28.8)	Dominant	
rs2292778	AGO2	AG	175 (50.3)	90 (45.5)	1	0.4600
		GG	83 (23.9)	51 (25.8)	0.86 (0.58-1.27)	
		GG	202 (58.4)	107 (54.3)	Dominant	
rs2257082	XPO5	GA	123 (35.5)	76 (38.6)	1	0.3579
		AA	21 (6.1)	14 (7.1)	1.18 (0.83-1.68)	
		TT	186 (53.4)	108 (54.3)	Dominant	
rs2287584	DROSHA	CT	143 (41.1)	76 (38.2)	1	0.8526
		CC	19 (5.5)	15 (7.5)	0.97 (0.68-1.37)	

Annex table 1: All the results of SNPs in miRNA processing genes (continuation)

SNP	Gene	Genotype	N (controls) (N=387)	N(cases) (N=213)	OR(CI 95%)	P dom
		CC	260 (74.9)	150 (75.0)	Dominant	
rs3744741	GEMIN4	CT	85 (24.5)	45 (22.5)	1	0.98506
		TT	2 (0.6)	5 (2.5)	1.00 (0.67-1.49)	
		CC	228 (65.5)	139 (69.2)	Dominant	
rs2227301	XPO5	CT	106 (30.5)	55 (27.4)	1	0.3819
		TT	14 (4.0)	7 (3.5)	0.85 (0.58-1.23)	
		AA	189 (54.3)	103 (52.0)	Dominant	
rs17151639	SND1	AG	128 (36.8)	83 (41.9)	1	0.6061
		GG	31 (8.9)	12 (6.1)	1.10 (0.77-1.56)	
		GG	276 (79.5)	164 (82.8)	Dominant	
rs6865950	GEMIN5	AG	69 (19.9)	32 (16.2)	1	0.3459
		AA	2 (0.6)	2 (1.0)	0.81 (0.51-1.27)	
		AA	330 (95.1)	188 (94.0)	Dominant	
rs35987994	DGCR8	AG	16 (4.6)	10 (5.0)	1	0.5830
		GG	1 (0.3)	2 (1.0)	1.24 (0.58-2.65)	
		CC	298 (86.9)	178 (90.8)	Codominant	
rs34324334	XPO5	CT	45 (13.1)	18 (9.2)	1	0.1645
					0.67 (0.38-1.19)	
		TT	274 (78.5)	173 (86.5)	Dominant	
rs6877400	CNOT6	CT	68 (19.5)	26 (13.0)	1	0.01822
		CC	7 (2.0)	1 (0.5)	0.57 (0.35-0.92)	
		GG	263 (76.0)	143 (72.2)	Dominant	
rs7735863	DROSHA	AG	73 (21.1)	52 (26.3)	1	0.3304
		AA	10 (2.9)	3 (1.5)	1.22 (0.82-1.81)	
		CC	166 (47.6)	91 (47.6)	Dominant	
rs639174	DROSHA	CT	154 (44.1)	81 (42.4)	1	0.9859
		TT	29 (8.3)	19 (9.9)	1.00 (0.70-1.42)	
		TT	158 (46.2)	86 (43.9)	Dominant	
rs3764941	SMAD5	GT	150 (43.9)	88 (44.9)	1	0.6026
		GG	34 (9.9)	22 (11.2)	1.10 (0.77-1.56)	
		AA	163 (46.8)	98 (48.8)	Dominant	
rs3823994	SND1	AT	157 (45.1)	87 (43.3)	1	0.6648
		TT	28 (8.0)	16 (8.0)	0.93 (0.65-1.31)	
		CC	259 (74.6)	155 (77.5)	Dominant	
rs3763425	CNOT4	CT	79 (22.8)	43 (21.5)	1	0.4509
		TT	9 (2.6)	2 (1.0)	0.85 (0.57-1.29)	
		AA	312 (89.9)	176 (88.0)	Dominant	
rs3792830	DROSHA	AG	34 (9.8)	23 (11.5)	1	0.4899
		GG	1 (0.3)	1 (0.5)	1.22 (0.70-2.11)	
		AA	106 (30.4)	69 (34.3)	Dominant	
rs4867329	DROSHA	AC	166 (47.6)	91 (45.3)	1	0.3388
		CC	77 (22.1)	41 (20.4)	0.83 (0.58-1.21)	

Annex table 1: All the results of SNPs in miRNA processing genes (continuation)

SNP	Gene	Genotype	N (controls) (N=387)	N(cases) (N=213)	OR(CI 95%)	P dom
		AA	205 (58.9)	135 (67.8)	Dominant	
rs9606248	DGCR8	AG	126 (36.2)	61 (30.7)	1	0.03725
		GG	17 (4.9)	3 (1.5)	0.68 (0.47-0.98)	
		GG	273 (78.4)	167 (83.1)	Dominant	
rs9611280	TNRC6B	AG	70 (20.1)	34 (16.9)	1	0.1856
		AA	5 (1.4)	0 (0.0)	0.74 (0.47-1.16)	
		AA	170 (49.4)	92 (48.7)	Dominant	
rs4821943	TNRC6B	AG	143 (41.6)	76 (40.2)	1	0.8699
		GG	31 (9.0)	21 (11.1)	1.03 (0.72-1.47)	
		AA	242 (71.2)	137 (71.4)	Dominant	
rs197388	DDX20	AT	89 (26.2)	48 (25.0)	1	0.9653
		TT	9 (2.6)	7 (3.6)	0.99 (0.67-1.47)	
		AA	302 (88.3)	173 (93.5)	Codominant	
rs17408716	DROSHA	AG	40 (11.7)	12 (6.5)	1	0.04847
					0.52 (0.27-1.03)	
		CC	246 (71.3)	148 (75.5)	Dominant	
rs17676986	SND1	CT	95 (27.5)	44 (22.4)	1	0.2881
		TT	4 (1.2)	4 (2.0)	0.81 (0.54-1.20)	
		GG	281 (82.2)	169 (86.7)	Dominant	
rs6884823	DROSHA	AG	54 (15.8)	26 (13.3)	1	0.16814
		AA	7 (2.0)	0 (0.0)	0.71 (0.43-1.17)	
		CC	241 (69.1)	137 (68.5)	Dominant	
rs4961280	AGO2	CA	101 (28.9)	53 (26.5)	1	0.89266
		AA	7 (2.0)	10 (5.0)	1.03 (0.71-1.49)	
		AA	242 (71.4)	132 (67.7)	Dominant	
rs595961	AGO1	AG	87 (25.7)	50 (25.6)	1	0.37098
		GG	10 (2.9)	13 (6.7)	1.19 (0.81-1.74)	
		GG	248 (71.7)	139 (69.2)	Dominant	
rs7755135	XPO5	AG	92 (26.6)	57 (28.4)	1	0.5327
		AA	6 (1.7)	5 (2.5)	1.13 (0.77-1.65)	
		TT	217 (62.7)	130 (66.0)	Dominant	
rs563002	DDX20	CT	113 (32.7)	58 (29.4)	1	0.4442
		CC	16 (4.6)	9 (4.6)	0.87 (0.60-1.25)	
		TT	213 (62.8)	108 (54.5)	Dominant	
rs10035440	DDX20	CT	109 (32.2)	72 (36.4)	1	0.05939
		CC	17 (5.0)	18 (9.1)	1.41 (0.99-2.01)	

Annex table 2: All the results of SNPs in miRNA genes of the first study

SNP	Gene	Genotype	N (controls) (N=387)	N(cases) (N=213)	OR(CI 95%)	P dom
		CC	235 (68.7)	134 (67.3)	Dominant	
rs34115976	hsa-mir-577	CG	93 (27.2)	58 (29.1)	1	0.7404
		GG	14 (4.1)	7 (3.5)	1.07 (0.73-1.55)	
		CC	280 (81.4)	153 (76.1)	Dominant	
rs2682818	hsa-mir-618	AC	61 (17.7)	45 (22.4)	1	0.1442
		AA	3 (0.9)	3 (1.5)	1.37 (0.90-2.10)	
		CC	244 (74.4)	144 (76.2)	Dominant	
rs550894	hsa-mir-612	AC	76 (23.2)	39 (20.6)	1	0.6480
		AA	8 (2.4)	6 (3.2)	0.91 (0.60-1.38)	
		CC	162 (47.1)	86 (43.2)	Dominant	
rs322825	hsa-mir-593	CT	152 (44.2)	98 (49.2)	1	0.3818
		TT	30 (8.7)	15 (7.5)	1.17 (0.82-1.66)	
		CC	125 (36.1)	67 (34.0)	Dominant	
rs12894467	hsa-mir-300	CT	165 (47.7)	96 (48.7)	1	0.6193
		TT	56 (16.2)	34 (17.3)	1.10 (0.76-1.58)	
		CC	214 (62.8)	122 (61.3)	Dominant	
rs4919510	hsa-mir-608	CG	108 (31.7)	69 (34.7)	1	0.7376
		GG	19 (5.6)	8 (4.0)	1.06 (0.74-1.52)	
		AA	110 (31.6)	64 (32.2)	Dominant	
rs2114358	hsa-mir-1206	AG	181 (52.0)	100 (50.3)	1	0.8940
		GG	57 (16.4)	35 (17.6)	0.97 (0.67-1.42)	
		TT	170 (49.0)	83 (41.9)	Dominant	
rs895819	hsa-mir-27a	CT	137 (39.5)	96 (48.5)	1	0.11082
		CC	40 (11.5)	19 (9.6)	1.33 (0.94-1.89)	
		TT	213 (62.8)	108 (54.5)	Dominant	
rs10035440	C5orf22	CT	109 (32.2)	72 (36.4)	1	0.05939
		CC	17 (5.0)	18 (9.1)	1.41 (0.99-2.01)	
		CC	84 (24.7)	55 (28.2)	Dominant	
rs6505162	hsa-mir-423	AC	166 (48.8)	98 (50.3)	1	0.3761
		AA	90 (26.5)	42 (21.5)	0.84 (0.56-1.24)	
		TT	202 (61.0)	105 (58.0)	Dominant	
rs2043556	hsa-mir-605	CT	112 (33.8)	63 (34.8)	1	0.5060
		CC	17 (5.1)	13 (7.2)	1.13 (0.78-1.64)	
		GG	305 (87.9)	178 (88.6)	Dominant	
rs2289030	hsa-mir-492	CG	39 (11.2)	22 (10.9)	1	0.8172
		CC	3 (0.9)	1 (0.5)	0.94 (0.55-1.61)	
		GG	191 (54.7)	113 (56.2)	Dominant	
rs2910164	hsa-mir-146a	CG	128 (36.7)	66 (32.8)	1	0.7348
		CC	30 (8.6)	22 (10.9)	0.94 (0.66-1.34)	
		AA	193 (55.5)	122 (60.7)	Dominant	
rs2368393	hsa-mir-604	AG	129 (37.1)	72 (35.8)	1	0.23117
		GG	26 (7.5)	7 (3.5)	0.81 (0.57-1.15)	

Annex table 2: All the results of SNPs in miRNA genes of the first study (continuation)

SNP	Gene	Genotype	N (controls) (N=387)	N(cases) (N=213)	OR(CI 95%)	P dom
		CC	294 (85.7)	173 (86.1)	Dominant	
rs17091403	hsa-mir-2110	CT	46 (13.4)	26 (12.9)	1	0.9085
		TT	3 (0.9)	2 (1.0)	0.97 (0.59-1.60)	
		AA	161 (48.3)	80 (41.2)	Dominant	
rs7911488	hsa-mir-1307	AG	136 (40.8)	90 (46.4)	1	0.1134
		GG	36 (10.8)	24 (12.4)	1.33 (0.93-1.91)	
		TT	170 (49.4)	90 (45.9)	Dominant	
rs10505168	hsa-mir-2053	CT	141 (41.0)	85 (43.4)	1	0.4336
		CC	33 (9.6)	21 (10.7)	1.15 (0.81-1.64)	
		TT	302 (86.8)	174 (87.4)	Codominant	
rs11259096	hsa-mir-1265	CT	46 (13.2)	25 (12.6)	1	0.8259
					0.94 (0.56-1.59)	
		CC	137 (39.7)	90 (45.5)	Dominant	
rs11614913	hsa-mir-196a-2	CT	159 (46.1)	79 (39.9)	1	0.1921
		TT	49 (14.2)	29 (14.6)	0.79 (0.56-1.13)	
		AA	334 (96.3)	192 (96.5)	Codominant	
rs13186787	hsa-mir-1294	AG	13 (3.7)	7 (3.5)	1	0.8907
					0.94 (0.37-2.39)	
		AA	283 (81.8)	180 (89.6)	Dominant	
rs10061133	hsa-mir-449b	GA	61 (17.6)	20 (10.0)	1	0.01290
		GG	2 (0.6)	1 (0.5)	0.52 (0.31-0.89)	
		TT	264 (76.1)	156 (80.8)	Dominant	
rs10173558	hsa-mir-1302-4	TC	81 (23.3)	35 (18.1)	1	0.1997
		CC	2 (0.6)	2 1.0	0.75 (0.49-1.17)	
		AA	198 (56.9)	124 (62.3)	Dominant	
rs1077020	hsa-mir-943	AG	125 (35.9)	62 (31.2)	1	0.2147
		GG	25 (7.2)	13 (6.5)	0.80 (0.56-1.14)	
		GG	343 (99.1)	193 (97.5)	Codominante	
rs11269	hsa-mir-1282	TG	3 (0.9)	5 (2.5)	1	0.1317
					2.96 (0.7-12.53)	
		CC	232 (67.6)	152 (77.6)	Dominant	
rs12803915	hsa-mir-612	TC	100 (29.2)	42 (21.4)	1	0.013334
		TT	11 (3.2)	2 (1.0)	0.61 (0.40-0.91)	
		AA	206 (59.4)	138 (69.0)	Dominant	
rs3746444	hsa-mir-499a	AG	117 (33.7)	56 (28.0)	1	0.023822
		GG	24 (6.9)	6 (3.0)	0.66 (0.45-0.95)	
		GG	348 (100.0)	199 (99.5)	Codominant	
rs41286570	hsa-mir-154	AG	0 (0.0)	1 (0.5)	1	0.365
					0 (0.0)	
		TT	307 (88.7)	181 (91.0)	Dominant	
rs41291179	hsa-mir-216a	CT	39 (11.3)	16 (8.0)	1	0.40906
		CC	0 (0.0)	2 (1.0)	0.78 (0.43-1.41)	

Annex table 2: All the results of SNPs in miRNA genes of the first study (continuation)

SNP	Gene	Genotype	N (controls) (N=387)	N(cases) (N=213)	OR(CI 95%)	P dom
		GG	247 (71.0)	134 (67.3)	Dominant	
rs4909237	hsa-mir-595	AG	90 (25.9)	53 (26.6)	1	0.3743
		AA	11 (3.2)	12 (6.0)	1.19 (0.81-1.73)	
		AA	230 (66.3)	124 (62.9)	Dominant	
rs56103835	hsa-mir-323b	AG	104 (30.0)	65 (33.0)	1	0.4333
		GG	13 (3.7)	8 (4.1)	1.16 (0.80-1.67)	
		GG	335 (96.0)	189 (95.5)	Dominant	
rs58834075	hsa-mir-656	GA	13 (3.7)	8 (4.0)	1	0.7661
		AA	1 (0.3)	1 (0.5)	1.14 (0.48-2.68)	
		GG	146 (43.3)	88 (45.6)	Dominant	
rs62376934	hsa-mir-585	GA	140 (41.5)	82 (42.5)	1	0.6123
		AA	51 (15.1)	23 (11.9)	0.91 (0.64-1.30)	
		AA	93 (26.9)	45 (22.5)	Dominant	
rs6505162	hsa-mir-423	AC	169 (48.8)	101 (50.5)	1	0.2540
		CC	84 (24.3)	54 (27.0)	1.27 (0.84-1.90)	
		GG	301 (86.2)	164 (81.6)	Dominant	
rs6841938	hsa-mir-1255b-1	AG	44 (12.6)	35 (17.4)	1	0.1497
		AA	4 (1.1)	2 (1.0)	1.41 (0.89-2.26)	
		AA	343 (99.7)	195 (99.0)	Dominant	
rs72631817	hsa-mir-220a	GA	1 (0.3)	0 (0.0)	1	0.2869
		GG	0 (0.0)	2 (1.0)	3.52 (0.32-39.05)	
		GG	343 (98.8)	199 (99.5)	Codominant	
rs72631827	hsa-mir-106b	GT	4 (1.2)	1 (0.5)	1	0.4184
					0.43 (0.05-3.88)	
		TT	342 (98.8)	194 (98.5)	Dominant	
rs7311975	hsa-mir-1178	TC	4 (1.2)	3 1.5	1	0.7186
					1.32 (0.29-5.97)	
		TT	301 (97.7)	153 (93.3)	Dominant	
rs73235382	hsa-mir-548h-4	TC	7 (2.3)	10 (6.1)	1	0.02019
		CC	0 (0.0)	1 (0.6)	3.09 (1.17-8.13)	
		CC	119 (36.5)	66 (35.7)	Dominant	
rs9913045	hsa-mir-548h-3	CT	142 (43.6)	69 (37.3)	1	0.85157
		TT	65 (19.9)	50 (27.0)	1.04 (0.71-1.51)	

Annex table 3: All the results for SNPs in miRNA genes in the second study

			Spanish and Slovenian population				Spanish population				Slovenian population			
SNP	**Gene**	**Genotype**	**N (controls) (N=426)**	**N(cases) (N=296)**	**OR(CI 95%)**	**P dom**	**N (controls) (N=330)**	**N(cases) (N=217)**	**OR(CI 95%)**	**P dom**	**N (controls) (N=96)**	**N(cases) (N=79)**	**OR(CI95%)**	**P dom**
rs10061133	hsa-mir-449b	AA	353 (82.9)	256 (87.7)	Dominant	0.07486	276 (83.6)	190 (87.6)	Dominant	0.2025	77 (80.2)	66 (88.0)	CoDominant	0.1666
		AG	71 (16.7)	34 (11.6)	1		52 (15.8)	25 (11.5)	1		19 (19.8)	9 (12.0)	1	
		GG	2 (0.5)	2 (0.7)	0.68 (0.44-1.05)		2 (0.6)	2 (0.9)	0.73 (0.44-1.19)				0.55 (0.23-1.3)	
rs10173558	mir-1302-4	TT	325 (76.3)	227 (77.7)	Dominant	0.65061	251 (76.1)	173 (79.7)	Dominant	0.3133	74 (77.1)	54 (72.0)	Dominant	0.44803
		CT	98 (23.0)	58 (19.9)	1		77 (23.3)	41 (18.9)	1		21 (21.9)	17 (22.7)	1	
		CC	3 (0.7)	7 (2.4)	0.92 (0.65-1.31)		2 (0.6)	3 (1.4)	0.81 (0.53-1.23)		1 (1.0)	4 (5.3)	1.31 (0.65-2.62)	
rs10461441	hsa-mir-548ae-2					ERROR				ERROR				ERROR
rs10505168	hsa-mir-2053	AA	212 (49.9)	136 (47.1)	Dominant	0.4587	167 (50.8)	97 (45.1)	Dominant	0.1976	45 (46.9)	39 (52.7)	Dominant	0.4510
		AG	172 (40.5)	122 (42.2)	1		130 (39.5)	95 (44.2)	1		42 (43.8)	27 (36.5)	1	
		GG	41 (9.6)	31 (10.7)	1.12 (0.83-1.51)		32 (9.7)	23 10.7	1.25 (0.89-1.77)		9 (9.4)	8 (10.8)	0.79 (0.43-1.45)	
rs1055070	hsa-mir-4700	TT	370 (87.1)	264 (90.7)	Dominant	0.12648	281 (85.4)	196 (90.3)	Dominant	0.01857	89 (92.7)	68 (91.9)	CoDominant	0.8429
		GT	55 (12.9)	25 (8.6)	1		48 (14.6)	19 (8.8)	1		7 (7.3)	6 (8.1)	1	
		GG	0 (0.0)	2 (0.7)	0.69 (0.42-1.12)		0 (0.0)	2 (0.9)	(0.00)				1.12 (0.36-3.49)	
rs1077020	hsa-mir-943					ERROR				ERROR				ERROR
rs10878362	hsa-mir-6074					ERROR				ERROR				ERROR
rs10934682	hsa-mir-544b	TT	307 (72.1)	200 (68.5)	Dominant	0.3029	240 (72.7)	155 (71.4)	Dominant	0.7403	67 (69.8)	45 (60.0)	Dominant	0.1820
		GT	110 (25.8)	83 (28.4)	1		85 (25.8)	58 (26.7)	1		25 (26.0)	25 (33.3)	1	
		GG	9 (2.1)	9 (3.1)	1.19 (0.86-1.64)		5 (1.5)	4 (1.8)	1.07 (0.73-1.56)		4 (4.2)	5 (6.7)	1.54 (0.82-2.91)	
rs11014002	hsa-mir-603					ERROR				ERROR				ERROR
rs11032942	hsa-mir-1343					ERROR				ERROR				ERROR
rs11156654	mir-624	TT	247 (58.5)	167 (57.4)	Dominant	0.7613	200 (61.2)	127 (58.8)	Dominant	0.5817	47 (49.5)	40 (53.3)	Dominant	0.6171
		AT	146 (34.6)	98 (33.7)	1		109 (33.3)	70 (32.4)	1		37 (38.9)	28 (37.3)	1	
		AA	29 (6.9)	26 (8.9)	1.05 (0.77-1.42)		18 (5.5)	19 (8.8)	1.10 (0.78-1.57)		11 (11.6)	7 (9.3)	0.86 (0.47-1.57)	
rs11237828	hsa-mir-5579					ERROR				ERROR				ERROR

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs11259096	hsa-mir-1265	TT	383 (90.1)	256 (87.7)	Dominant	0.3038	298 (90.3)	186 (85.7)	Dominant	0.10309	85 (89.5)	70 (93.3)	CoDominant	0.3727
		CT	41 (9.6)	34 (11.6)	1		31 (9.4)	29 (13.4)	1		10 (10.5)	5 (6.7)	1	
		CC	1 (0.2)	2 (0.7)	1.28 (0.80-2.06)		1 (0.3)	2 (0.9)	1.55 (0.92-2.63)				0.61 (0.2-1.86)	
rs11614913	hsa-mir-196a-2	CC	174 (40.9)	131 (45.0)	Dominant	0.2790	130 (39.5)	99 (45.6)	Dominant	0.1573	44 (45.8)	32 (43.2)	Dominant	0.7362
		CT	198 (46.6)	124 (42.6)	1		156 (47.4)	92 (42.4)	1		42 (43.8)	32 (43.2)	1	
		TT	53 (12.5)	36 (12.4)	0.85 (0.63-1.14)		43 (13.1)	26 (12.0)	0.78 (0.55-1.10)		10 (10.4)	10 (13.5)	1.11 (0.60-2.04)	
rs11651671	hsa-mir-548at					ERROR				ERROR				ERROR
rs11713052	hsa-mir-5092	CC	396 (93.0)	269 (92.1)	Dominant	0.6754	305 (92.4)	199 (91.7)	CoDominant	0.7604	91 (94.8)	70 (93.3)	Dominant	0.6878
		CG	30 (7.0)	22 (7.5)	1		25 (7.6)	18 (8.3)	1		5 (5.2)	4 (5.3)	1	
		GG	0 (0.0)	1 (0.3)	1.13 (0.64-1.99)				1.1 (0.59-2.08)		0 (0.0)	1 (1.3)	1.30 (0.36-4.67)	
rs11714172	hsa-mir-4792	TT	176 (41.6)	115 (39.8)	Dominant	0.6283	126 (38.5)	84 (39.3)	Dominant	0.8665	50 (52.1)	31 (41.3)	Dominant	0.1618
		GT	189 (44.7)	136 (47.1)	1		152 (46.5)	100 (46.7)	1		37 (38.5)	36 (48.0)	1	
		GG	58 (13.7)	38 (13.1)	1.08 (0.79-1.46)		49 (15.0)	30 (14.0)	0.97 (0.68-1.38)		9 (9.4)	8 (10.7)	1.54 (0.84-2.84)	
rs11907020	hsa-mir-3192	TT	405 (95.1)	284 (97.6)	CoDominant	0.07758	309 (93.6)	210 (97.2)	CoDominant	0.0498	96 (100.0)	74 (98.7)	CoDominant	0.4386
		CT	21 (4.9)	7 (2.4)	1		21 (6.4)	6 (2.8)	1		0 (0.0)	1 (1.3)	1	
					0.48 (0.2-1.13)				0.42 (0.17-1.06)				0 (0.0)	
rs12197631	hsa-mir-548a-1					ERROR				ERROR				ERROR
rs12355840	hsa-mir-202	TT				H-W	183 (66.1)	120 (61.9)	Dominant	0.3486				H-W
		CT					78 (28.29	65 (33.5)	1					
		CC					16 (5.8)	9 (4.6)	1.20 (0.82-1.76)					
rs12451747	hsa-mir-1269b					ERROR				ERROR				ERROR
rs12456845	hsa-mir-4744	TT	400 (94.1)	266 (91.1)	CoDominant	0.1252	311 (94.5)	200 (92.2)	CoDominant	0.2743	89 (92.7)	66 (88.0)	CoDominant	0.2963
		CT	25 (5.9)	26 (8.9)	1		18 (5.5)	17 (7.8)	1		7 (7.3)	9 (12.0)	1	
					1.56 (0.88-2.77)				1.47 (0.74-2.92)				1.73 (0.61-4.89)	

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs12473206	hsa-mir-4433					ERROR				ERROR				ERROR
rs12512664	hsa-mir-4274	AA	118 (28.0)	84 (29.0)	Dominant	0.7706	89 (27.3)	57 (26.4)	Dominant	0.8147	29 (30.2)	27 (36.5)	Dominant	0.3886
		AG	217 (51.4)	132 (45.5)	1		170 (52.1)	102 (47.2)	1		47 (49.0)	30 (40.5)	1	
		GG	87 (20.6)	74 (25.5)	0.95 (0.68-1.33)		67 (20.6)	57 (26.4)	1.05 (0.71-1.54)		20 (20.8)	17 (23.0)	0.75 (0.40-1.43)	
rs12523324	hsa-mir-4277					ERROR				ERROR				ERROR
rs12780876	hsa-mir-4293	TT	214 (50.6)	150 (51.5)	Dominant	0.8019	186 (56.9)	114 (52.5)	Dominant	0.3185	28 (29.2)	36 (48.6)	Dominant	0.009347
		AT	171 (40.4)	117 (40.2)	1		117 (35.8)	82 (37.8)	1		54 (56.2)	35 (47.3)	1	
		AA	38 (9.0)	24 (8.2)	0.96 (0.71-1.30)		24 (7.3)	21 (9.7)	1.19 (0.84-1.68)		14 (14.6)	3 (4.1)	0.43 (0.23-0.82)	
rs12879262	hsa-mir-4309	GG	305 (71.9)	218 (74.7)	Dominant	0.4186	228 (69.5)	155 (71.4)	Dominant	0.63136	77 (80.2)	63 (84.0)	Dominant	0.5213
		CG	108 (25.5)	70 (24.0)	1		90 (27.4)	60 (27.6)	1		18 (18.8)	10 (13.3)	1	
		CC	11 (2.6)	4 (1.4)	0.87 (0.62-1.22)		10 (3.0)	2 (0.9)	0.91 (0.63-1.33)		1 (1.0)	2 (2.7)	0.77 (0.35-1.71)	
rs13186787	hsa-mir-1294					ERROR				ERROR				ERROR
rs13299349	hsa-mir-3152	GG	158 (42.4)	114 (47.3)	Dominant	0.2288	120 (43.3)	91 (47.2)	Dominant	0.4118	38 (39.6)	23 (47.9)	Dominant	0.3411
		AG	163 (43.7)	94 (39.0)	1		113 (40.8)	74 (38.3)	1		50 (52.1)	20 (41.7)	1	
		AA	52 (13.9)	33 (13.7)	0.82 (0.59-1.13)		44 (15.9)	28 (14.5)	0.86 (0.59-1.24)		8 (8.3)	5 (10.4)	0.71 (0.35-1.43)	
rs1414273	hsa-mir-548ac					ERROR				ERROR				ERROR
rs1439619	hsa-mir-3175	CC	92 (24.7)	71 (29.0)	Dominant	0.23527	71 (25.6)	52 (26.7)	Dominant	0.8010	21 (21.9)	19 (38.0)	Dominant	0.011090
		AC	171 (45.8)	123 (50.2)	1		130 (46.9)	100 (51.3)	1		41 (42.7)	23 (46.0)	1	
		AA	110 (29.5)	51 (20.8)	0.80 (0.56-1.15)		76 (27.4)	43 (22.1)	0.95 (0.62-1.44)		34 (35.4)	8 (16.0)	2.88 (81.21-6.83)	
rs1572687	hsa-mir-5007	CC	128 (30.3)	86 (29.6)	Dominant	0.8394	106 (32.4)	64 (29.5)	Dominant	0.4706	22 (22.9)	22 (29.7)	Dominant	0.3159
		CT	208 (49.2)	150 (51.5)	1		152 (46.5)	113 (52.1)	1		56 (58.3)	37 (50.0)	1	
		TT	87 (20.6)	55 (18.9)	1.03 (0.75-1.43)		69 (21.1)	40 (18.4)	1.15 (0.79-1.66)		18 (18.8)	15 (20.3)	0.70 (0.35-1.40)	
rs1683709	hsa-mir-3612	CC	276 (64.8)	183 (62.7)	Dominant	0.5619	212 (64.2)	136 (62.7)	Dominant	0.7091	64 (66.7)	47 (62.7)	Dominant	0.5869
		CT	133 (31.2)	99 (33.9)	1		107 (32.4)	74 (34.1)	1		26 (27.1)	25 (33.3)	1	
		TT	17 (4.0)	10 (3.4)	1.10 (0.80-1.49)		11 (3.3)	7 (3.2)	1.07 (0.75-1.53)		6 (6.2)	3 (4.0)	1.19 (0.63-2.24)	

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs17022749	hsa-mir-5700					ERROR				ERROR				ERROR
rs17091403	hsa-mir-2110	CC	350 (82.2)	248 (85.2)	Dominant	0.2764	268 (81.2)	190 (87.6)	Dominant	0.04607	82 (85.4)	58 (78.4)	Dominant	0.2346
		CT	71 (16.7)	40 (13.7)	1		58 (17.6)	24 (11.1)	1		13 (13.5)	16 (21.6)	1	
		TT	5 (1.2)	3 (1.0)	0.80 (0.53-1.20)		4 (1.2)	3 (1.4)	0.61 (0.38-1.00)		1 (1.0)	0 (0.0)	1.62 (0.73-3.57)	
rs17111728	hsa-mir-4422	TT	367 (86.2)	255 (87.3)	Dominant	0.6477	290 (87.9)	192 (88.5)	Dominant	0.8316	77 (80.2)	63 (84.0)	Dominant	0.5213
		CT	58 (13.6)	36 (12.3)	1		39 (11.8)	25 (11.5)	1		19 (19.8)	11 (14.7)	1	
		CC	1 (0.2)	1 (0.3)	0.90 (0.58-1.40)		1 (0.3)	0 (0.0)	0.94 (0.55-1.6)		0 (0.0)	1 (1.3)	0.77 (0.35-1.71)	
rs17737028	hsa-mir-3143	AA	419 (98.4)	287 (98.3)	CoDominant	0.9435	325 (98.5)	213 (98.2)	CoDominant	0.7692	94 (97.9)	74 (98.7)	CoDominant	0.7074
		AG	7 (1.6)	5 (1.7)	1		5 (1.5)	4 (1.8)	1		2 (2.1)	1 (1.3)	1	
					1.04 (0.33-3.32)				1.22 (0.32-4.6)				0.64 (0.06-7.14)	
rs17759989	hsa-mir-633	AA	408 (96.0)	284 (97.3)	CoDominant	0.3595	318 (96.7)	213 (98.2)	CoDominant	0.2817	90 (93.8)	71 (94.7)	Dominant	0.7992
		AG	17 (4.0)	8 (2.7)	1		11 (3.3)	4 (1.8)	1		6 (6.2)	4 (5.3)	1	
					0.68 (0.29-1.59)				0.54 (0.17-1.73)				0.85 (0.23-3.11)	
rs17797090	hsa-mir-3652	GG	334 (78.4)	231 (79.4)	Dominant	0.7529	251 (76.1)	164 (75.9)	Dominant	0.9713	83 (86.5)	67 (89.3)	Dominant	0.5678
		AG	88 (20.7)	59 (20.3)	1		76 (23.0)	51 (23.6)	1		12 (12.5)	8 (10.7)	1	
		AA	4 (0.9)	1 (0.3)	0.94 (0.65-1.36)		3 (0.9)	1 (0.5)	1.01 (0.67-1.51)		1 (1.0)	0 (0.0)	0.76 (0.30-1.95)	
rs17885221	hsa-mir-4733	CC	377 (88.7)	260 (89.3)	Dominant	0.7876	283 (86.0)	188 (87.0)	Dominant	0.7336	94 (97.9)	72 (96.0)	CoDominant	0.4624
		CT	47 (11.1)	30 (10.3)	1		45 (13.7)	27 (12.5)	1		2 (2.1)	3 (4.0)	1	
		TT	1 (0.2)	1 (0.3)	0.94 (0.58-1.51)		1 (0.3)	1 (0.5)	0.92 (0.55-1.52)				1.96 (0.32-12.03)	
rs2042253	hsa-mir-5197	AA	248 (58.5)	180 (61.9)	Dominant	0.3667	196 (59.6)	136 (63.0)	Dominant	0.4272	52 (54.7)	44 (58.7)	Dominant	0.6077
		AG	157 (37.0)	102 (35.1)	1		118 (35.9)	72 (33.3)	1		39 (41.1)	30 (40.0)	1	
		GG	19 (4.5)	9 (3.1)	0.87 (0.64-1.18)		15 (4.6)	8 (3.7)	0.87 (0.61-1.23)		4 (4.2)	1 (1.3)	0.85 (0.46-1.57)	
rs2043556	hsa-mir-605	AA	262 (61.6)	172 (60.1)	Dominant	0.6863	203 (61.7)	131 (61.5)	Dominant	0.9627	59 (61.5)	41 (56.2)	Dominant	0.4882
		AG	146 (34.4)	98 (34.3)	1		113 (34.3)	70 (32.9)	1		33 (34.4)	28 (38.4)	1	
		GG	17 (4.0)	16 (5.6)	1.07 (0.78-1.45)		13 (4.0)	12 (5.6)	1.01 (0.71-1.44)		4 (4.2)	4 (5.5)	1.24 (0.67-2.31)	

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs2060455	hsa-mir-4511					ERROR				ERROR				ERROR
rs2070960	hsa-mir-3620	CC	370 (87.3)	255 (88.5)	Dominant	0.6084	295 (89.7)	192 (90.1)	CoDominant	0.8578	75 (78.9)	63 (84.0)	CoDominant	0.4002
		CT	54 (12.7)	33 (11.5)	1		34 (10.3)	21 (9.9)	1		20 (21.1)	12 (16.0)	1	
					0.89 (0.56-1.41)				0.95 (0.53-1.68)				0.71 (0.32-1.57)	
rs2114358	hsa-mir-1206	TT	140 (33.0)	92 (31.9)	Dominant	0.7639	107 (32.5)	69 (32.1)	Dominant	0.9165	33 (34.7)	23 (31.5)	Dominant	0.6594
		CT	200 (47.2)	143 (49.7)	1		153 (46.5)	111 (51.6)	1		47 (49.5)	32 (43.8)	1	
		CC	84 (19.8)	53 (18.4)	1.05 (0.76-1.45)		69 (21.0)	35 (16.3)	1.02 (0.71-1.47)		15 (15.8)	18 (24.7)	1.16 (0.60-2.22)	
rs215383	hsa-mir-4494	GG	292 (68.9)	196 (68.1)	Dominant	0.8189	221 (67.4)	143 (66.8)	Dominant	0.8929	71 (74.0)	53 (71.6)	Dominant	0.7341
		AG	114 (26.9)	80 (27.8)	1		90 (27.4)	59 (27.6)	1		24 (25.0)	21 (28.4)	1	
		AA	18 (4.2)	12 (4.2)	1.04 (0.75-1.43)		17 (5.2)	12 (5.6)	1.03 (0.71-1.48)		1 (1.0)	0 (0.0)	1.13 (0.57-2.22)	
rs2241347	hsa-mir-3130-1					ERROR				ERROR				ERROR
rs2273626	hsa-mir-4707	AA	100 (28.7)	79 (27.1)	Dominant	0.6560	69 (27.4)	58 (26.9)	Dominant	0.8979	31 (32.3)	21 (28.0)	Dominant	0.5442
		AC	175 (50.3)	143 (49.1)	1		130 (51.6)	104 (48.1)	1		45 (46.9)	39 (52.0)	1	
		CC	73 (21.0)	69 (23.7)	1.08 (0.76-1.53)		53 (21.0)	54 (25.0)	1.03 (0.68-1.55)		20 (20.8)	15 (20.0)	1.23 (0.63-2.38)	
rs2289030	hsa-mir-492	CC	364 (85.4)	257 (88.6)	Dominant	0.2155	283 (85.8)	191 (88.8)	Dominant	0.2926	81 (84.4)	66 (88.0)	Dominant	0.4958
		CG	61 (14.3)	32 (11.0)	1		47 (14.2)	23 (10.7)	1		14 (14.6)	9 (12.0)	1	
		GG	1 (0.2)	1 (0.3)	0.75 (0.48-1.18)		0 (0.0)	1 (0.5)	0.76 (0.45-1.28)		1 (1.0)	0 (0.0)	0.74 (0.30-1.79)	
rs2291418	hsa-mir-1229	CC	397 (93.2)	279 (95.5)	Dominant	0.1797	307 (93.0)	208 (95.9)	Dominant	0.1600	90 (93.8)	71 (94.7)	CoDominant	0.7992
		CT	28 (6.6)	13 (4.5)	1		22 (6.7)	9 (4.1)	1		6 (6.2)	4 (5.3)	1	
		TT	1 (0.2)	0 (0.0)	0.64 (0.33-1.25)		1 (0.3)	0 (0.0)	0.58 (0.26 - 1.27)				0.85 (0.23-3.11)	
rs2292181	hsa-mir-564	GG	388 (91.3)	270 (92.8)	Dominant	0.4705	303 (92.1)	204 (94.4)	Dominant	0.2862	85 (88.5)	66 (88.0)	CoDominant	0.913
		CG	36 (8.5)	20 (6.9)	1		25 (7.6)	11 (5.1)	1		11 (11.5)	9 (12.0)	1	
		CC	1 (0.2)	1 (0.3)	0.82 (0.47-1.42)		1 (0.3)	1 (0.5)	0.69 (0.34-1.39)				1.05 (0.41-2.69)	
rs2292832	hsa-mir-149					ERROR				ERROR				ERROR

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs2368392	hsa-mir-604	CC	233 (54.8)	176 (60.9)	Dominant	0.10663				H-W	47 (49.0)	43 (59.7)	Dominant	0.1655
		CT	155 (36.5)	101 (34.9)	1						42 (43.8)	24 (33.3)	1	
		TT	37 (8.7)	12 (4.2)	0.78 (0.57-1.06)						7 (7.3)	5 (6.9)	0.65 (0.35-1.20)	
rs243080	hsa-mir-4432	CC	128 (30.2)	101 (35.8)	Dominant	0.11869	98 (29.8)	69 (32.9)	Dominant	0.45311	30 (31.6)	32 (44.4)	Dominant	0.08876
		CT	224 (52.8)	123 (43.6)	1		175 (53.2)	90 (42.9)	1		49 (51.6)	33 (45.8)	1	
		TT	72 (17.0)	58 (20.6)	0.77 (0.56-1.07)		56 (17.0)	51 (24.3)	0.87 (0.60-1.26)		16 (16.8)	7 (9.7)	0.58 (0.31-1.09)	
rs257095	hsa-mir-4636	AA	303 (71.3)	213 (73.2)	Dominant	0.5770	236 (71.7)	163 (75.1)	Dominant	0.3818	67 (69.8)	50 (67.6)	Dominant	0.7564
		AG	107 (25.2)	73 (25.1)	1		81 (24.6)	52 (24.0)	1		26 (27.1)	21 (28.4)	1	
		GG	15 (3.5)	5 (1.7)	0.91 (0.65-1.27)		12 (3.6)	2 (0.9)	0.84 (0.57-1.24)		3 (3.1)	3 (4.1)	1.11 (0.58-2.13)	
rs2648841	hsa-mir-1208	CC	330 (78.0)	222 (77.4)	Dominant	0.83518	251 (76.5)	159 (74.6)	Dominant	0.6192	79 (83.2)	63 (85.1)	Dominant	0.7272
		AC	87 (20.6)	55 (19.2)	1		71 (21.6)	45 (21.1)	1		16 (16.8)	10 (13.5)	1	
		AA	6 (1.4)	10 (3.5)	1.04 (0.73-1.49)		6 (1.8)	9 (4.2)	1.11 (0.74-1.65)		0 (0.0)	1 (1.4)	0.86 (0.37-1.99)	
rs2663345	hsa-mir-3183					ERROR				ERROR				ERROR
rs266435	hsa-mir-4804	CC	324 (76.2)	206 (71.5)	Dominant	0.15929	250 (76.0)	154 (71.6)	Dominant	0.2571	74 (77.1)	52 (71.2)	Dominant	0.3883
		CG	97 (22.8)	75 (26.0)	1		75 (22.8)	55 (25.6)	1		22 (22.9)	20 (27.4)	1	
		GG	4 (0.9)	7 (2.4)	1.28 (0.91-1.79)		4 (1.2)	6 (2.8)	1.25 (0.85-1.85)		0 (0.0)	1 (1.4)	1.36 (0.68-2.72)	
rs2682818	hsa-mir-6128	CC	315 (74.3)	221 (76.5)	Dominant	0.5078	241 (73.3)	165 (76.7)	Dominant	0.3584	74 (77.9)	56 (75.7)	Dominant	0.7344
		AC	103 (24.3)	64 (22.1)	1		82 (24.9)	46 (21.4)	1		21 (22.1)	18 (24.3)	1	
		AA	6 (1.4)	4 (1.4)	0.89 (0.63-1.26)		6 (1.8)	4 (1.9)	0.83 (0.56-1.24)				1.13 (0.55-2.32)	
rs28645567	hsa-mir-378d-1	GG	411 (96.9)	284 (97.3)	CoDominant	0.7986	319 (97.3)	211 (97.2)	CoDominant	0.9883	92 (95.8)	73 (97.3)	CoDominant	0.5922
		AG	13 (3.1)	8 (2.7)	1		9 (2.7)	6 (2.8)	1		4 (4.2)	2 (2.7)	1	
					0.89 (0.36-2.18)				1.01 (0.35-2.87)				0.63 (0.11-3.54)	
rs28655823	hsa-mir-4472-1	GG	273 (77.1)	235 (81.3)	Dominant	0.1923	215 (79.3)	178 (82.4)	Dominant	0.3922	58 (69.9)	57 (78.1)	Dominant	0.2437
		CG	75 (21.2)	49 (17.0)	1		52 (19.2)	33 (15.3)	1		23 (27.7)	16 (21.9)	1	
		CC	6 (1.7)	5 (1.7)	0.77 (0.53-1.14)		4 (1.5)	5 (2.3)	0.82 (0.52-1.29)		2 (2.4)	0 (0.0)	0.65 (0.32-1.35)	

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs28664200	hsa-mir-1255a	TT	202 (55.0)	128 (57.1)	Dominant	0.6175	151 (55.7)	98 (56.0)	Dominant	0.9536	51 (53.1)	30 (61.2)	Dominant	0.3515
		CT	143 (39.0)	87 (38.8)	1		103 (38.0)	71 (40.6)	1		40 (41.7)	16 (32.7)	1	
		CC	22 (6.0)	9 (4.0)	0.92 (0.66-1.28)		17 (6.3)	6 (3.4)	0.99 (0.67-1.45)		5 (5.2)	3 (6.1)	0.72 (0.36-1.45)	
rs2910164	hsa-mir-146a	GG	240 (56.3)	172 (59.1)	Dominant	0.4613	180 (54.5)	125 (57.6)	Dominant	0.4809	60 (62.5)	47 (63.5)	Dominant	0.8921
		CG	158 (37.1)	95 (32.6)	1		128 (38.8)	73 (33.6)	1		30 (31.2)	22 (29.7)	1	
		CC	28 (6.6)	24 (8.2)	0.89 (0.66-1.21)		22 (6.7)	19 (8.8)	0.88 (0.63-1.25)		6 (6.28)	5 (6.8)	0.96 (0.51-1.79)	
rs2967897	hsa-mir-5695					M				M				M
rs3112399	hsa-mir-4803	TT				H-W				H-W	32 (33.7)	29 (38.7)	Dominant	0.5017
		AT									46 (48.4)	29 (38.7)	1	
		AA									17 (17.9)	17 (22.7)	0.81 (0.43-1.51)	
rs34115976	hsa-mir-577	CC	278 (65.6)	187 (65.4)	Dominant	0.9602				H-W	55 (57.3)	46 (64.8)	Dominant	0.3261
		CG	125 (29.5)	87 (30.4)	1						38 (39.6)	20 (28.2)	1	
		GG	21 (5.0)	12 (4.2)	1.01 (0.74-1.38)						3 (3.1)	5 (7.0)	0.73 (0.39-1.37)	
rs35196866	hsa-mir-4669					ERROR				ERROR				ERROR
rs35613341	hsa-mir-5189	CC				H-W				H-W	49 (51.0)	29 (39.2)	Dominant	0.123343
		CG									41 (42.7)	30 (40.5)	1	
		GG									6 (6.2)	15 (20.3)	1.62 (0.87-2.99)	
rs35650931	hsa-mir-6076	GG	349 (82.1)	240 (82.2)	Dominant	0.9797	278 (84.2)	183 (84.3)	Dominant	0.9776	71 (74.7)	57 (76.0)	Dominant	0.8495
		CG	72 (16.9)	49 (16.8)	1		49 (14.8)	32 (14.7)	1		23 (24.2)	17 (22.7)	1	
		CC	4 (0.9)	3 (1.0)	0.99 (0.67-1.47)		3 (0.9)	2 (0.9)	0.99 (0.62-1.59)		1 (1.1)	1 (1.3)	0.93 (0.46-1.89)	
rs35770269	hsa-mir-449c	AA	179 (42.2)	132 (45.4)	Dominant	0.4050	137 (41.8)	97 (44.9)	Dominant	0.4696	42 (43.8)	35 (46.7)	Dominant	0.7037
		AT	191 (45.0)	123 (42.3)	1		146 (44.5)	91 (42.1)	1		45 (46.9)	32 (42.7)	1	
		TT	54 (12.7)	36 (12.4)	0.88 (0.65-1.19)		45 (13.7)	28 (13.0)	0.88 (0.62-1.24)		9 (9.4)	8 (10.7)	0.89 (0.48-1.63)	

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs367805	hsa-mir-3936	GG	206 (48.6)	139 (48.4)	Dominant	0.9681	164 (50.0)	108 (50.2)	Dominant	0.9577	42 (43.8)	31 (43.1)	Dominant	0.9284
		AG	171 (40.3)	121 (42.2)	1		131 (39.9)	88 (40.9)	1		40 (41.7)	33 (45.8)	1	
		AA	47 (11.1)	27 (9.4)	1.01 (0.75-1.36)		33 (10.1)	19 (8.8)	0.99 (0.70-1.40)		14 (14.6)	8 (11.1)	1.03 (0.56-1.91)	
rs3734050	hsa-mir-6499	CC	382 (89.9)	259 (88.7)	Dominant	0.6139	294 (89.4)	193 (88.9)	CoDominant	0.8767	88 (91.7)	66 (88.0)	CoDominant	0.4284
		CT	43 (10.1)	33 (11.3)	1		35 (10.6)	24 (11.1)	1		8 (8.3)	9 (12.0)	1	
					1.13 (0.7-1.83)				1.04 (0.6-1.81)				1.5 (0.55-4.1)	
rs3746444	hsa-mir-499a	TT	265 (62.5)	187 (64.7)	Dominant	0.5480	201 (61.3)	144 (67.0)	Dominant	0.1763	64 (66.7)	43 (58.1)	Dominant	0.25254
		CT	143 (33.7)	87 (30.1)	1		113 (34.5)	62 (28.8)	1		30 (31.2)	25 (33.8)	1	
		CC	16 (3.8)	15 (5.2)	0.91 (0.67-1.24)		14 (4.3)	9 (4.2)	0.78 (0.54-1.12)		2 (2.1)	6 (8.1)	1.44 (0.77-2.70)	
rs3823658	hsa-mir-5090	GG	312 (73.2)	218 (75.2)	Dominant	0.56194	238 (72.1)	168 (77.8)	Dominant	0.13650	74 (77.1)	50 (67.6)	Dominant	0.1673
		AG	108 (25.4)	62 (21.4)	1		89 (27.0)	43 (19.9)	1		19 (19.8)	19 (25.7)	1	
		AA	6 (1.4)	10 (3.4)	0.90 (0.64-1.27)		3 (0.9)	5 (2.3)	0.74 (0.49-1.10)		3 (3.1)	5 (6.8)	1.61 (0.82-3.19)	
rs4112253	hsa-mir-4751	CC	170 (39.9)	122 (41.9)	Dominant	0.5893	127 (38.5)	95 (44.0)	Dominant	0.20156	43 (44.8)	27 (36.0)	Dominant	0.2450
		CG	203 (47.7)	124 (42.6)	1		162 (49.1)	88 (40.7)	1		41 (42.7)	36 (48.0)	1	
		GG	53 (12.4)	45 (15.5)	0.92 (0.68-1.25)		41 (12.4)	33 (15.3)	0.80 (0.56-1.13)		12 (12.5)	12 (16.0)	1.44 (0.78-2.68)	
rs41274239	hsa-mir-96	AA	425 (99.8)	290 (100.0)	CoDominant	1	329 (99.7)	216 (100.0)	CoDominant	1				M
		AG	1 (0.2)	0 (0.0)	1		1 (0.3)	0 (0.0)	1					
					0 (0.0)				(0.0)					
rs41274312	hsa-mir-187	GG	420 (98.6)	288 (99.3)	CoDominant	0.3544	325 (98.5)	215 (99.5)	CoDominant	0.221	95 (99.0)	73 (98.6)	CoDominant	0.8533
		AG	6 (1.4)	2 (0.7)	1		5 (1.5)	1 (0.5)	1		1 (1.0)	1 (1.4)	1	
					0.49 (0.1-2.43)				0.3 (0.04-2.61)				1.3 (0.08-21.16)	
rs41286570	hsa-mir-154					M				M				M
rs41291179	hsa-mir-216a	AA	377 (88.5)	265 (90.8)	Dominant	0.3312	287 (87.0)	197 (90.8)	Dominant	0.16643	90 (93.8)	68 (90.7)	CoDominant	0.4523
		AT	48 (11.3)	24 (8.2)	1		42 (12.7)	17 (7.8)	1		6 (6.2)	7 (9.3)	1	
		TT	1 (0.2)	3 (1.0)	0.78 (0.48-1.29)		1 (0.3)	3 (1.4)	0.68 (0.39-1.19)				1.54 (0.5-4.8)	

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs41292412	hsa-mir-122	CC	423 (99.5)	291 (99.7)	CoDominant	0.7916	328 (99.4)	217 (100.0)	CoDominant	0.5205	95 (100.0)	74 (98.7)	CoDominant	0.4412
		CT	2 (0.5)	1 (0.3)	1		2 (0.6)	0 (0.0)	1		0 (0.0)	1 (1.3)	1	
					0.73 (0.07-8.05)				(0.0)				0 (0.0)	
rs4285314	hsa-mir-3135b					ERROR				ERROR				ERROR
rs4414449	hsa-mir-548ap	TT	131 (39.6)	100 (39.5)	Dominant	0.9900	107 (43.9)	81 (41.3)	Dominant	0.5944	24 (27.6)	19 (33.3)	Dominant	0.4626
		CT	149 (45.0)	118 (46.6)	1		106 (43.4)	90 (45.9)	1		43 (49.4)	28 (49.1)	1	
		CC	51 (15.4)	35 (13.8)	1.00 (0.72-1.40)		31 (12.7)	25 (12.8)	1.11 (0.76-1.62)		20 (23.0)	10 (17.5)	0.76 (0.37-1.57)	
rs45530340	hsa-mir-6084					M				M				M
rs4577031	hsa-mir-548ap	AA	161 (38.0)	117 (40.2)	Dominant	0.5473	135 (41.2)	92 (42.4)	Dominant	0.7742	26 (27.1)	25 (33.8)	Dominant	0.3455
		AT	206 (48.6)	145 (49.8)	1		154 (47.0)	104 (47.9)	1		52 (54.2)	41 (55.4)	1	
		TT	57 (13.4)	29 (10.0)	0.91 (0.67-1.24)		39 (11.9)	21 (9.7)	0.95 (0.67-1.35)		18 (18.8)	8 (10.8)	0.73 (0.38-1.41)	
rs4674470	hsa-mir-4268	TT	262 (61.6)	185 (63.6)	Dominant	0.6008	212 (64.4)	144 (66.4)	Dominant	0.6442	50 (52.1)	41 (55.4)	Dominant	0.6667
		CT	142 (33.4)	90 (30.9)	1		102 (31.0)	59 (27.2)	1		40 (41.7)	31 (41.9)	1	
		CC	21 (4.9)	16 (5.5)	0.92 (0.68-1.25)		15 (4.6)	14 (6.5)	0.92 (0.64-1.32)		6 (6.2)	2 (2.7)	0.87 (0.48-1.61)	
rs4809383	hsa-mir-941-1	CC	285 (76.4)	197 (79.8)	Dominant	0.3243				H-W	73 (76.8)	36 (72.0)	Dominant	0.5237
		CT	85 (22.8)	46 (18.6)	1						19 (20.0)	13 (26.0)	1	
		TT	3 (0.8)	4 (1.6)	0.82 (0.56-1.22)						3 (3.2)	1 (2.0)	1.29 (0.59-2.81)	
rs4822739	hsa-mir-548j	CC	377 (88.5)	250 (85.9)	CoDominant	0.3068	293 (88.8)	190 (88.0)	CoDominant	0.7684	84 (87.5)	60 (80.0)	CoDominant	0.1837
		CG	49 (11.5)	41 (14.1)	1		37 (11.2)	26 (12.0)	1		12 (12.5)	15 (20.0)	1	
					1.26 (0.81-1.97)				1.08 (0.64-1.85)				1.75 (0.76-4.01)	
rs487571	hsa-mir-5680					ERROR				ERROR				ERROR
rs4919510	hsa-mir-608	CC	269 (63.1)	186 (64.6)	Dominant	0.6949	201 (60.9)	130 (60.5)	Dominant	0.9174	68 (70.8)	56 (76.7)	Dominant	0.3898
		CG	142 (33.3)	93 (32.3)	1		115 (34.8)	77 (35.8)	1		27 (28.1)	16 (21.9)	1	
		GG	15 (3.5)	9 (3.1)	0.94 (0.69-1.28)		14 (4.2)	8 (3.7)	1.02 (0.72-1.45)		1 (1.0)	1 (1.4)	0.74 (0.37-1.48)	

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs515924	hsa-mir-548al	AA	349 (82.3)	226 (78.2)	Dominant	0.1744	275 (83.8)	180 (83.7)	Dominant	0.9703	74 (77.1)	46 (62.2)	Dominant	0.03468
		AG	73 (17.2)	61 (21.1)	1		52 (15.9)	34 (15.8)	1		21 (21.9)	27 (36.5)	1	
		GG	2 (0.5)	2 (0.7)	1.30 (0.89-1.89)		1 (0.3)	1 (0.5)	1.01 (0.63-1.61)		1 (1.0)	1 (1.4)	2.05 (1.05-4.00)	
rs521188	hsa-mir-3671	AA	391 (91.8)	277 (94.9)	Dominant	0.1053	303 (91.8)	204 (94.0)	CoDominant	0.33	88 (91.7)	73 (97.3)	CoDominant	0.1023
		AG	35 (8.2)	15 (5.1)	1		27 (8.2)	13 (6.0)	1		8 (8.3)	2 (2.7)	1	
					0.6 (0.32-1.13)				0.72 (0.36-1.42)				0.3 (0.06-1.46)	
rs56088671	hsa-mir-4424					ERROR				ERROR				ERROR
rs56103835	hsa-mir-323b	TT	288 (67.6)	195 (66.8)	Dominant	0.8171	233 (70.6)	147 (67.7)	Dominant	0.4775	55 (57.3)	48 (64.0)	Dominant	0.3730
		CT	127 (29.8)	86 (29.5)	1		88 (26.7)	63 (29.0)	1		39 (40.6)	23 (30.7)	1	
		CC	11 (2.6)	11 (3.8)	1.04 (0.76-1.43)		9 (2.7)	7 (3.2)	1.14 (0.79-1.66)		2 (2.1)	4 (5.3)	0.75 (0.41-1.40)	
rs56195815	hsa-mir-548aw					ERROR				ERROR				ERROR
rs56292801	hsa-mir-5189	GG				H-W				H-W	51 (53.1)	33 (44.6)	Dominant	0.269700
		AG									41 (42.7)	27 (36.5)	1	
		AA									4 (4.2)	14 (18.9)	1.41 (0.77-2.59)	
rs57111412	hsa-mir-1283-1					ERROR				ERROR				ERROR
rs58450758	hsa-mir-559					ERROR				ERROR				ERROR
rs58834075	hsa-mir-656	CC	410 (96.2)	279 (95.9)	CoDominant	0.8033	322 (97.6)	208 (96.3)	CoDominant	0.3912	88 (91.7)	71 (94.7)	Dominant	0.4406
		CT	16 (3.8)	12 (4.1)	1		8 (2.4)	8 (3.7)	1		8 (8.3)	4 (5.3)	1	
					1.1 (0.51-2.37)				1.55 (0.57-4.19)				0.62 (0.18-2.14)	
rs5965660	hsa-mir-888					H-W				H-W				H-W
rs5997893	hsa-mir-3928	GG	210 (49.5)	128 (44.3)	Dominant	0.1688	174 (53.0)	107 (50.0)	Dominant	0.4875	36 (37.5)	21 (28.0)	Dominant	0.18909
		AG	168 (39.6)	127 (43.9)	1		121 (36.9)	90 (42.1)	1		47 (49.0)	37 (49.3)	1	
		AA	46 (10.8)	34 (11.8)	1.23 (0.91-1.67)		33 (10.1)	17 (7.9)	1.13 (0.80-1.59)		13 (13.5)	17 (22.7)	1.54 (0.80-2.96)	
rs60308683	hsa-mir-4762					ERROR				ERROR				ERROR

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs6062431	hsa-mir-4326	GG	191 (45.4)	148 (51.6)	Dominant	0.1049	155 (47.5)	110 (51.4)	Dominant	0.3806	36 (37.9)	38 (52.1)	Dominant	0.06676
		CG	172 (40.9)	102 (35.5)	1		131 (40.2)	77 (36.0)	1		41 (43.2)	25 (34.2)	1	
		CC	58 (13.8)	37 (12.9)	0.78 (0.58-1.05)		40 (12.3)	27 (12.6)	0.86 (0.61-1.21)		18 (18.9)	10 (13.7)	0.56 (0.30-1.04)	
rs60871950	hsa-mir-4467	AA	103 (24.3)	68 (23.8)	Dominant	0.4494	88 (26.9)	58 (27.1)	Dominant	0.9609	15 (15.6)	10 (13.9)	Dominant	0.4428
		AG	201 (47.5)	130 (45.5)	1		163 (49.8)	96 (44.9)	1		38 (39.6)	34 (47.2)	1	
		GG	119 (28.1)	88 (30.8)	0.88 (0.63-1.22)		76 (23.2)	60 (28.0)	0.99 (0.67-1.46)		43 (44.8)	28 (38.9)	1.27 (0.68-2.37)	
rs61388742	hsa-mir-596	TT	343 (80.9)	245 (84.2)	Dominant	0.2548	264 (80.5)	183 (84.3)	Dominant	0.2496	79 (82.3)	62 (83.8)	Dominant	0.7973
		CT	77 (18.2)	43 (14.8)	1		62 (18.9)	31 (14.3)	1		15 (15.6)	12 (16.2)	1	
		CC	4 (0.9)	3 (1.0)	0.80 (0.53-1.18)		2 (0.6)	3 (1.4)	0.77 (0.49-1.21)		2 (2.1)	0 (0.0)	0.90 (0.40-2.02)	
rs61938575	hsa-mir-3922	GG	186 (49.6)	128 (52.9)	Dominant	0.42435	138 (49.3)	101 (52.1)	Dominant	0.55221	48 (50.5)	27 (56.2)	Dominant	0.5171
		AG	165 (44.0)	89 (36.8)	1		124 (44.3)	71 (36.6)	1		41 (43.2)	18 (37.5)	1	
		AA	24 (6.4)	25 (10.3)	0.88 (0.63-1.21)		18 (6.4)	22 (11.3)	0.89 (0.62-1.29)		6 (6.3)	3 (6.2)	0.79 (0.40-1.60)	
rs61992671	hsa-mir-412	GG	118 (27.8)	67 (23.1)	Dominant	0.1605	99 (30.1)	53 (24.7)	Dominant	0.1648	19 (19.8)	14 (18.7)	Dominant	0.5267
		AG	210 (49.4)	145 (50.0)	1		157 (47.7)	106 (49.3)	1		53 (55.2)	39 (52.0)	1	
		AA	97 (22.8)	78 (26.9)	1.28 (0.91-1.81)		73 (22.2)	56 (26.0)	1.32 (0.89-1.94)		24 (25.0)	22 (29.3)	0.80 (0.41-1.58)	
rs62154973	hsa-mir-4772	CC	336 (79.1)	243 (84.7)	Dominant	0.05724	267 (81.2)	183 (85.1)	Dominant	0.2287	69 (71.9)	60 (83.3)	Dominant	0.07787
		CT	82 (19.3)	40 (13.9)	1		55 (16.7)	30 (14.0)	1		27 (28.1)	10 (13.9)	1	
		TT	7 (1.6)	4 (1.4)	0.68 (0.46-1.02)		7 (2.1)	2 (0.9)	0.75 (0.47-1.20)		0 (0.0)	2 (2.8)	0.51 (0.24-1.10)	
rs62376935	hsa-mir-585	CC				H-W	297 (90.0)	182 (85.0)	Dominant	0.08481				H-W
		CT					31 (9.4)	31 (14.5)	1					
		TT					2 (0.6)	1 (0.5)	1.58 (0.94-2.66)					
rs641071	hsa-mir-4482					ERROR				ERROR				ERROR
rs6430498	hsa-mir-3679	GG	192 (45.3)	127 (44.3)	Dominant	0.7860	145 (44.2)	89 (41.6)	Dominant	0.5472	47 (49.0)	38 (52.1)	Dominant	0.6900
		AG	185 (43.6)	136 (47.4)	1		144 (43.9)	105 (49.1)	1		41 (42.7)	31 (42.5)	1	
		AA	47 (11.1)	24 (8.4)	1.04 (0.77-1.41)		39 (11.9)	20 (9.3)	1.11 (0.79-1.58)		8 (8.3)	4 (5.5)	0.88 (0.48-1.62)	

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs6505162	hsa-mir-423					ERROR				ERROR				ERROR
rs6513496	hsa-mir-646	TT	278 (65.3)	185 (64.0)	Dominant	0.7326	215 (65.2)	146 (67.9)	Dominant	0.5055	63 (65.6)	39 (52.7)	Dominant	0.08835
		CT	134 (31.5)	89 (30.8)	1		104 (31.5)	59 (27.4)	1		30 (31.2)	30 (40.5)	1	
		CC	14 (3.3)	15 (5.2)	1.06 (0.77-1.44)		11 (3.3)	10 (4.7)	0.88 (0.61-1.27)		3 (3.1)	5 (6.8)	1.71 (0.92-3.19)	
rs66507245	hsa-mir-4731					ERROR				ERROR				ERROR
rs66683138	hsa-mir-3622a					ERROR				ERROR				ERROR
rs67042258	hsa-mir-6128	GG	240 (56.7)	144 (49.5)	Dominant		186 (56.9)	112 (51.9)	Dominant	0.2493	54 (56.2)	32 (42.7)	Dominant	0.07747
		AG	160 (37.8)	129 (44.3)	1		125 (38.2)	93 (43.1)	1		35 (36.5)	36 (48.0)	1	
		AA	23 (5.4)	18 (6.2)			16 (4.9)	11 (5.1)	1.22 (0.87-1.73)		7 (7.3)	7 (9.3)	1.73 (0.94-3.18)	
rs670637	hsa-mir-3167					M				M				M
rs67182313	hsa-mir-4642	AA	277 (65.0)	200 (69.4)	Dominant	0.2173	212 (64.2)	146 (67.9)	Dominant	0.3775	65 (67.7)	54 (74.0)	Dominant	0.3751
		AG	134 (31.5)	80 (27.8)	1		106 (32.1)	64 (29.8)	1		28 (29.2)	16 (21.9)	1	
		GG	15 (3.5)	8 (2.8)	0.82 (0.59-1.13)		12 (3.6)	5 (2.3)	0.85 (0.59-1.22)		3 (3.1)	3 (4.1)	0.74 (0.38-1.45)	
rs6726779	hsa-mir-4431	TT	160 (37.6)	110 (38.1)	Dominant	0.9106	126 (38.2)	79 (36.7)	Dominant	0.7348	34 (35.8)	31 (41.9)	Dominant	0.4189
		CT	215 (50.6)	134 (46.4)	1		164 (49.7)	103 (47.9)	1		51 (53.7)	31 (41.9)	1	
		CC	50 (11.8)	45 (15.6)	0.98 (0.72-1.34)		40 (12.1)	33 (15.3)	1.06 (0.75-1.52)		10 (10.5)	12 (16.2)	0.77 (0.41-1.44)	
rs67339585	MIR3910-1, MIR3910-2					ERROR				ERROR				ERROR
rs6787734	hsa-mir-3135a					ERROR				ERROR				ERROR
rs67976778	hsa-mir-4305					ERROR				ERROR				ERROR
rs6841938	hsa-mir-1255b-1					ERROR				ERROR				ERROR
rs6997249	hsa-mir-3686					ERROR				ERROR				ERROR
rs701213	hsa-mir-4427					ERROR				ERROR				ERROR

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs702742	hsa-mir-378h	AA	342 (80.5)	236 (80.8)	Dominant	0.9069	255 (77.5)	175 (80.6)	Dominant	0.3785	87 (90.6)	61 (81.3)	Dominant	0.07822
		AG	79 (18.6)	54 (18.5)	1		71 (21.6)	41 (18.9)	1		8 (8.3)	13 (17.3)	1	
		GG	4 (0.9)	2 (0.7)	0.98 (0.67-1.43)		3 (0.9)	1 (0.5)	0.83 (0.54-1.26)		1 (1.0)	1 (1.3)	2.22 (0.90-5.45)	
rs7070684	hsa-mir-548aj-2					ERROR				ERROR			Dominant	ERROR
rs71363366	hsa-mir-1283-2	CC	391 (92.9)	267 (91.8)	CoDominant	0.5798	307 (93.9)	201 (92.6)	CoDominant	0.5656	84 (89.4)	66 (89.2)	Dominant	0.9714
		CG	30 (7.1)	24 (8.2)	1		20 (6.1)	16 (7.4)	1		10 (10.6)	8 (10.8)	1	
					1.17 (0.67-2.05)				1.22 (0.62-2.41)				1.02 (0.38-2.72)	
rs7205289	hsa-mir-140				Dominant	M				M				M
rs7207008	hsa-mir-2117	TT	128 (30.1)	75 (26.0)	Dominant	0.2352	99 (30.1)	51 (23.9)	Dominant	0.1161	29 (30.2)	24 (32.0)	Dominant	0.8016
		AT	205 (48.2)	147 (51.0)	1		162 (49.2)	114 (53.5)	1		43 (44.8)	33 (44.0)	1	
		AA	92 (21.6)	66 (22.9)	1.22 (0.88-1.71)		68 (20.7)	48 (22.5)	1.37 (0.92-2.03)		24 (25.0)	18 (24.0)	0.92 (0.48-1.77)	
rs7227168	hsa-mir-4741	CC	333 (78.7)	224 (77.2)	Dominant	0.6387	265 (81.0)	171 (79.5)	Dominant	0.6663	68 (70.8)	53 (70.7)	Dominant	0.9810
		CT	86 (20.3)	61 (21.0)	1		58 (17.7)	40 (18.6)	1		28 (29.2)	21 (28.0)	1	
		TT	4 (0.9)	5 (1.7)	1.09 (0.76-1.56)		4 (1.2)	4 (1.9)	1.10 (0.71-1.69)		0 (0.0)	1 (1.3)	1.01 (0.52-1.96)	
rs7247237	hsa-mir-3188	CC	201 (47.6)	149 (51.4)	Dominant	0.3255	158 (48.3)	109 (50.2)	Dominant	0.6622	43 (45.3)	40 (54.8)	Dominant	0.2203
		CT	192 (45.5)	117 (40.3)	1		148 (45.3)	90 (41.5)	1		44 (46.3)	27 (37.0)	1	
		TT	29 (6.9)	24 (8.3)	0.86 (0.64-1.16)		21 (6.4)	18 (8.3)	0.93 (0.66-1.31)		8 (8.4)	6 (8.2)	0.68 (0.37-1.26)	
rs72502717	hsa-mir-3689f					ERROR				ERROR				ERROR
rs72631816	hsa-mir-105-2					M				M				M
rs72631825	hsa-mir-222					M				M				M
rs72631826	hsa-mir-16-1					M				M				M
rs72631827	hsa-mir-106b					M				M				M
rs72631831	hsa-mir-323b					M				M				M
rs72631833	hsa-mir-183					M				M				M

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs72646786	hsa-mir-3972	CC	324 (76.1)	236 (82.2)	Dominant	0.04706	254 (77.0)	168 (78.9)	Dominant	0.6019	70 (72.9)	68 (91.9)	Dominant	0.0011005
		CT	96 (22.5)	47 (16.4)	1		71 (21.5)	43 (20.2)	1		25 (26.0)	4 (5.4)	1	
		TT	6 (1.4)	4 (1.4)	0.69 (0.47-1.00)		5 (1.5)	2 (0.9)	0.90 (0.59-1.36)		1 (1.0)	2 (2.7)	0.24 (0.09-0.61)	
rs72855836	hsa-mir-3976	GG				H-W	297 (90.5)	200 (92.6)	Dominant	0.4024				H-W
		AG					29 (8.8)	15 (6.9)	1					
		AA					2 (0.6)	1 (0.5)	0.77 (0.41-1.44)					
rs72996752	hsa-mir-4999	AA	240 (58.0)	149 (55.6)	Dominant	0.5409	197 (60.1)	115 (55.3)	Dominant	0.2754	43 (50.0)	34 (56.7)	Dominant	0.4269
		AG	145 (35.0)	101 (37.7)	1		115 (35.1)	81 (38.9)	1		30 (34.9)	20 (33.3)	1	
		GG	29 (7.0)	18 (6.7)	1.10 (0.81-1.50)		16 (4.9)	12 (5.8)	1.22 (0.86-1.73)		13 (15.1)	6 (10.0)	0.76 (0.39-1.48)	
rs73112689	hsa-mir-4459					ERROR				ERROR				ERROR
rs7311975	hsa-mir-1178	TT	394 (92.5)	263 (91.3)	Dominant	0.5734	309 (93.6)	197 (92.1)	Dominant	0.4828	85 (88.5)	66 (89.2)	Dominant	0.8942
		CT	31 (7.3)	24 (8.3)	1		21 (6.4)	16 (7.5)	1		10 (10.4)	8 (10.8)	1	
		CC	1 (0.2)	1 (0.3)	1.17 (0.68-2.02)		0 (0.0)	1 (0.5)	1.27 (0.65-2.47)		1 (1.0)	0 (0.0)	0.94 (0.36-2.46)	
rs73147065	hsa-mir-647					ERROR				ERROR				ERROR
rs73177830	hsa-mir-4532					ERROR				ERROR				ERROR
rs73235381	hsa-mir-548h-4					ERROR				ERROR				ERROR
rs73239138	hsa-mir-1269a	GG	243 (57.4)	161 (55.3)	Dominant	0.5744	193 (59.0)	120 (55.6)	Dominant	0.4240	50 (52.1)	41 (54.7)	Dominant	0.7369
		AG	153 (36.2)	114 (39.2)	1		115 (35.2)	87 (40.3)	1		38 (39.6)	27 (36.0)	1	
		AA	27 (6.4)	16 (5.5)	1.09 (0.81-1.47)		19 (5.8)	9 (4.2)	1.15 (0.81-1.63)		8 (8.3)	7 (9.3)	0.90 (0.49-1.65)	
rs73410309	hsa-mir-4739					ERROR				ERROR				ERROR
rs74428911	hsa-mir-4474	GG				H-W				H-W	94 (97.9)	71 (95.9)	CoDominant	0.4531
		GT									2 (2.1)	3 (4.1)	1	
													1.99 (0.32-12.2)	

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs74469188	hsa-mir-6504	TT	286 (76.9)	188 (77.4)	Dominant	0.8888	208 (75.1)	148 (76.3)	Dominant	0.7655	78 (82.1)	40 (81.6)	Dominant	0.9444
		CT	81 (21.8)	51 (21.0)	1		65 (23.5)	42 (21.6)	1		16 (16.8)	9 (18.4)	1	
		CC	5 (1.3)	4 (1.6)	0.97 (0.66-1.43)		4 (1.4)	4 (2.1)	0.94 (0.61-1.44)		1 (1.1)	0 (0.0)	1.03 (0.42-2.52)	
rs745666	hsa-mir-3615	CC	171 (40.2)	118 (40.7)	Dominant	0.9033	131 (39.8)	88 (40.9)	Dominant	0.7959	40 (41.7)	30 (40.0)	Dominant	0.82587
		CG	191 (44.9)	140 (48.3)	1		152 (46.2)	99 (46.0)	1		39 (40.6)	41 (54.7)	1	
		GG	63 (14.8)	32 (11.0)	0.98 (0.72-1.33)		46 (14.0)	28 (13.0)	0.95 (0.67-1.36)		17 (17.7)	4 (5.3)	1.07 (0.58-1.98)	
rs74704964	hsa-mir-518d	CC	346 (93.3)	227 (92.7)	CoDominant	0.7724	262 (94.9)	183 (93.4)	CoDominant	0.4745	84 (88.4)	44 (89.8)	CoDominant	0.8024
		CT	25 (6.7)	18 (7.3)	1		14 (5.1)	13 (6.6)	1		11 (11.6)	5 (10.2)	1	
					1.1 (0.59-2.06)				1.33 (0.61-2.89)				0.87 (0.28-2.66)	
rs74904371	hsa-mir-2682	CC	400 (93.9)	275 (95.2)	Dominant	0.4687	310 (93.9)	205 (95.8)	Dominant	0.3398	90 (93.8)	70 (93.3)	CoDominant	0.9123
		CT	25 (5.9)	14 (4.8)	1		19 (5.8)	9 (4.2)	1		6 (6.2)	5 (6.7)	1	
		TT	1 (0.2)	0 (0.0)	0.78 (0.40-1.53)		1 (0.3)	0 (0.0)	0.68 (0.30-1.52)				1.07 (0.31-3.66)	
rs74949342	hsa-mir-5702	CC	420 (98.6)	289 (99.0)	CoDominant	0.6484	324 (98.2)	216 (99.5)	CoDominant	0.1384	96 (100.0)	73 (97.3)	Dominant	0.1909
		CG	6 (1.4)	3 (1.0)	1		6 (1.8)	1 (0.5)	1		0 (0.0)	2 (2.7)	1	
					0.73 (0.18-2.93)				0.25 (0.03-2.09)				0 (0.0)	
rs7500280	hsa-mir-4719					ERROR				ERROR				ERROR
rs75019967	hsa-mir-4477a					M				M				M
rs7522956	hsa-mir-4742	AA	257 (60.5)	162 (55.7)	Dominant	0.2007	204 (62.0)	129 (59.7)	Dominant	0.5929	53 (55.2)	33 (44.0)	Dominant	0.1454
		AC	150 (35.3)	110 (37.8)	1		110 (33.4)	72 (33.3)	1		40 (41.7)	38 (50.7)	1	
		CC	18 (4.2)	19 (6.5)	1.22 (0.90-1.65)		15 (4.6)	15 (6.9)	1.10 (0.77-1.56)		3 (3.1)	4 (5.3)	1.57 (0.85-2.88)	
rs75598818	hsa-mir-520f	GG	404 (95.1)	274 (94.2)	Dominant	0.5991	316 (96.0)	209 (96.3)	Dominant	0.8746	88 (91.7)	65 (87.8)	CoDominant	0.4115
		AG	20 (4.7)	17 (5.8)	1		12 (3.6)	8 (3.7)	1		8 (8.3)	9 (12.2)	1	
		AA	1 (0.2)	0 (0.0)	1.19 (0.62-2.30)		1 (0.3)	0 (0.0)	0.93 (0.38-2.28)				1.52 (0.56-4.16)	

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs75715827	hsa-mir-944	TT	363 (85.2)	254 (87.9)	Dominant	0.3040	278 (84.2)	189 (88.3)	Dominant	0.1784	85 (88.5)	65 (86.7)	Dominant	0.7115
		CT	59 (13.8)	35 (12.1)	1		50 (15.2)	25 (11.7)	1		9 (9.4)	10 (13.3)	1	
		CC	4 (0.9)	0 (0.0)	0.79 (0.51-1.24)		2 (0.6)	0 (0.0)	0.71 (0.42-1.18)		2 (2.1)	0 (0.0)	1.19 (0.48-2.97)	
rs75966923	hsa-mir-4298	CC	401 (94.1)	278 (95.2)	Dominant	0.5301	312 (94.5)	207 (95.4)	Dominant	0.6585	89 (92.7)	71 (94.7)	CoDominant	0.6017
		AC	24 (5.6)	13 (4.5)	1		17 (5.2)	9 (4.1)	1		7 (7.3)	4 (5.3)	1	
		AA	1 (0.2)	1 (0.3)	0.81 (0.41-1.58)		1 (0.3)	1 (0.5)	0.84 (0.38-1.85)				0.72 (0.2-2.54)	
rs76481776	hsa-mir-182	CC	359 (84.7)	236 (81.7)	Dominant	0.29043	275 (83.8)	171 (79.9)	Dominant	0.24327	84 (87.5)	65 (86.7)	Dominant	0.8718
		CT	60 (14.2)	53 (18.3)	1		49 (14.9)	43 (20.1)	1		11 (11.5)	10 (13.3)	1	
		TT	5 (1.2)	0 (0.0)	1.24 (0.83-1.85)		4 (1.2)	0 (0.0)	1.30 (0.84-2.04)		1 (1.0)	0 (0.0)	1.08 (0.44-2.65)	
rs76800617	hsa-mir-4521	AA	409 (96.0)	276 (94.5)	Dominant	0.353	315 (95.5)	202 (93.1)	CoDominant	0.2391	94 (97.9)	74 (98.7)	CoDominant	0.7074
		AG	17 (4.0)	16 (5.5)	1		15 (4.5)	15 (6.9)	1		2 (2.1)	1 (1.3)	1	
					1.39 (0.69-2.81)				1.56 (0.75-3.26)				0.64 (0.06-7.14)	
rs77055126	hsa-mir-1303					ERROR				ERROR				ERROR
rs77639117	hsa-mir-576	AA	410 (96.7)	280 (96.6)	CoDominant	0.9152	317 (96.6)	209 (96.8)	CoDominant	0.9425	93 (96.9)	71 (95.9)	CoDominant	0.7458
		AT	14 (3.3)	10 (3.4)	1		11 (3.4)	7 (3.2)	1		3 (3.1)	3 (4.1)	1	
					1.05 (0.46-2.39)				0.97 (0.37-2.53)				1.31 (0.26-6.68)	
rs78396863	hsa-mir-4743	GG	413 (97.6)	286 (98.3)	CoDominant	0.5498	319 (97.6)	213 (98.2)	CoDominant	0.6354	94 (97.9)	73 (98.6)	Dominant	0.7159
		CG	10 (2.4)	5 (1.7)	1		8 (2.4)	4 (1.8)	1		2 (2.1)	1 (1.4)	1	
					0.72 (0.24-2.13)				0.75 (0.22-2.52)				0.64 (0.06-7.24)	
rs78541299	hsa-mir-6075	GG	423 (99.3)	291 (99.7)	CoDominant	0.5094	327 (99.1)	216 (99.5)	CoDominant	0.5348				M
		AG	3 (0.7)	1 (0.3)	1		3 (0.9)	1 (0.5)	1					
					0.48 (0.05-4.68)				0.5 (0.05-4.88)					
rs78790512	hsa-mir-6083	GG	291 (68.3)	199 (68.2)	Dominant	0.9641	215 (65.2)	144 (66.4)	Dominant	0.7709	76 (79.2)	55 (73.3)	Dominant	0.3724
		AG	121 (28.4)	84 (28.8)	1		103 (31.2)	65 (30.0)	1		18 (18.8)	19 (25.3)	1	
		AA	14 (3.3)	9 (3.1)	1.01 (0.73-1.39)		12 (3.6)	8 (3.7)	0.95 (0.66-1.36)		2 (2.1)	1 (1.3)	1.38 (0.68-2.81)	

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs78831152	hsa-mir-4789	CC	354 (83.1)	237 (81.7)	Dominant	0.6350	279 (84.5)	177 (82.3)	Dominant	0.4947	75 (78.1)	60 (80.0)	Dominant	0.7650
		CT	68 (16.0)	51 (17.6)	1		49 (14.8)	36 (16.7)	1		19 (19.8)	15 (20.0)	1	
		TT	4 (0.9)	2 (0.7)	1.10 (0.74-1.63)		2 (0.6)	2 (0.9)	1.17 (0.74-1.86)		2 (2.1)	0 (0.0)	0.89 (0.42-1.88)	
rs78832554	hsa-mir-4786	GG	402 (94.6)	281 (96.2)	CoDominant	0.3024	309 (93.6)	209 (96.3)	CoDominant	0.1621	93 (97.9)	72 (96.0)	CoDominant	0.4696
		AG	23 (5.4)	11 (3.8)	1		21 (6.4)	8 (3.7)	1		2 (2.1)	3 (4.0)	1	
					0.68 (0.33-1.43)				0.56 (0.24-1.3)				1.94 (0.32-11.9)	
rs7896283	hsa-mir-4481	AA	85 (35.1)	93 (38.4)	Dominant	0.4507	57 (38.8)	78 (40.2)	Dominant	0.7890	28 (29.5)	15 (31.2)	Dominant	0.8272
		AG	113 (46.7)	108 (44.6)	1		68 (46.3)	85 (43.8)	1		45 (47.4)	23 (47.9)	1	
		GG	44 (18.2)	41 (16.9)	0.87 (0.60-1.26)		22 (15.0)	31 (16.0)	0.94 (0.61-1.46)		22 (23.2)	10 (20.8)	0.92 (0.43-1.95)	
rs7911488	hsa-mir-1307					ERROR				ERROR				ERROR
rs79397096	hsa-mir-597	GG	410 (96.2)	286 (97.9)	CoDominant	0.1832	317 (96.1)	212 (97.7)	CoDominant	0.2834	93 (96.9)	74 (98.7)	CoDominant	0.4283
		AG	16 (3.8)	6 (2.1)	1		13 (3.9)	5 (2.3)	1		3 (3.1)	1 (1.3)	1	
					0.54 (0.21-1.39)				0.58 (0.2-1.64)				0.42 (0.04-4.11)	
rs79512808	hsa-mir-3976	TT	413 (96.9)	287 (98.3)	Dominant	0.2485	322 (97.6)	213 (98.2)	Dominant	0.6465	91 (94.8)	74 (98.7)	CoDominant	0.1492
		GT	13 (3.1)	4 (1.4)	1		8 (2.4)	3 (1.4)	1		5 (5.2)	1 (1.3)	1	
		GG	0 (0.0)	1 (0.3)	0.55 (0.20-1.57)		0 (0.0)	1 (0.5)	0.76 (0.22-2.54)				0.25 (0.03-2.15)	
rs80128580	hsa-mir-5707	GG	405 (95.1)	276 (94.5)	Dominant	0.7441	316 (95.8)	205 (94.5)	CoDominant	0.4919	89 (92.7)	71 (94.7)	Dominant	0.6017
		AG	21 (4.9)	15 (5.1)	1		14 (4.2)	12 (5.5)	1		7 (7.3)	3 (4.0)	1	
		AA	0 (0.0)	1 (0.3)	1.12 (0.57-2.18)				1.32 (0.6-2.91)		0 (0.0)	1 (1.3)	0.72 (0.20-2.54)	
rs8054514	hsa-mir-3176	TT	301 (70.8)	217 (74.3)	Dominant	0.3037	227 (68.8)	156 (71.9)	Dominant	0.4376	74 (77.9)	61 (81.3)	Dominant	0.5808
		GT	118 (27.8)	70 (24.0)	1		97 (29.4)	57 (26.3)	1		21 (22.1)	13 (17.3)	1	
		GG	6 (1.4)	5 (1.7)	0.84 (0.60-1.17)		6 (1.8)	4 (1.8)	0.86 (0.59-1.26)		0 (0.0)	1 (1.3)	0.81 (0.38-1.72)	
rs8078913	hsa-mir-4520a	CC	130 (31.2)	75 (28.6)	Dominant	0.4679	102 (31.0)	65 (32.2)	Dominant	0.7772	28 (32.2)	10 (16.7)	Dominant	0.03134
		CT	206 (49.5)	134 (51.1)	1		165 (50.2)	103 (51.0)	1		41 (47.1)	31 (51.7)	1	
		TT	80 (19.2)	53 (20.2)	1.13 (0.81-1.59)		62 (18.8)	34 (16.8)	0.95 (0.65-1.38)		18 (20.7)	19 (31.7)	2.37 (1.05-5.36)	

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs832733	hsa-mir-4698					ERROR				ERROR				ERROR
rs850108	hsa-mir-550a-3					ERROR				ERROR				ERROR
rs8667	hsa-mir-4751	GG	142 (37.2)	117 (40.2)	Dominant	0.4232	109 (38.0)	84 (38.9)	Dominant	0.8355	33 (34.7)	33 (44.0)	Dominant	0.2189
		AG	183 (47.9)	141 (48.5)	1		133 (46.3)	109 (50.5)	1		50 (52.6)	32 (42.7)	1	
		AA	57 (14.9)	33 (11.3)	0.88 (0.64-1.20)		45 (15.7)	23 (10.6)	0.96 (0.67-1.38)		12 (12.6)	10 (13.3)	0.68 (0.36-1.26)	
rs877722	hsa-mir-4671	AA	318 (74.6)	230 (78.8)	Dominant	0.2002	242 (73.3)	167 (77.0)	Dominant	0.3378	76 (79.2)	63 (84.0)	CoDominant	0.4189
		AT	102 (23.9)	58 (19.9)	1		82 (24.8)	46 (21.2)	1		20 (20.8)	12 (16.0)	1	
		TT	6 (1.4)	4 (1.4)	0.79 (0.56-1.13)		6 (1.8)	4 (1.8)	0.82 (0.55-1.23)				0.72 (0.33-1.59)	
rs895819	mir-27a					ERROR				ERROR				ERROR
rs897984	hsa-mir-4519					ERROR				ERROR				ERROR
rs9295535	hsa-mir-5689					ERROR				ERROR				ERROR
rs936581	hsa-mir-3141	GG	299 (70.5)	195 (67.0)	Dominant	0.31943	243 (74.1)	156 (71.9)	Dominant	0.5716	56 (58.3)	39 (52.7)	Dominant	0.46365
		AG	109 (25.7)	92 (31.6)	1		79 (24.1)	59 (27.2)	1		30 (31.2)	33 (44.6)	1	
		AA	16 (3.8)	4 (1.4)	1.18 (0.85-1.62)		6 (1.8)	2 (0.9)	1.12 (0.76-1.64)		10 (10.4)	2 (2.7)	1.26 (0.68-2.31)	
rs9842591	hsa-mir-5186	CC	102 (27.2)	81 (33.5)	Dominant	0.09711	74 (26.5)	64 (33.3)	Dominant	0.11173	28 (29.2)	17 (34.0)	Dominant	0.5500
		AC	198 (52.8)	106 (43.8)	1		147 (52.7)	84 (43.8)	1		51 (53.1)	22 (44.0)	1	
		AA	75 (20.0)	55 (22.7)	0.74 (0.52-1.05)		58 (20.8)	44 (22.9)	0.72 (0.48-1.08)		17 (17.7)	11 (22.0)	0.80 (0.38-1.66)	
rs9877402	hsa-mir-5680	AA	341 (91.7)	210 (86.4)	CoDominant	0.03931	253 (91.0)	171 (88.6)	CoDominant	0.3941	88 (93.6)	39 (78.0)	CoDominant	0.007147
		AG	31 (8.3)	33 (13.6)	1		25 (9.0)	22 (11.4)	1		6 (6.4)	11 (22.0)	1	
					1.73 (1.03-2.91)				1.3 (0.71-2.38)				4.14 (1.43-11.99)	
rs9913045	hsa-mir-548h-3					ERROR				ERROR				ERROR
rs11048315	MIR4302	GG				H-W				H-W	71 (74.7)	58 (77.3)	Dominant	0.6939
		AG									21 (22.1)	15 (20.0)	1	
		AA									3 (3.2)	2 (2.7)	0.87 (0.43-1.77)	
rs111803974	MIR3908					ERROR				ERROR				ERROR

Annex table 3: All the results for SNPs in miRNA genes in the second study (continuation)

SNP	Gene	Genotype	N (controls) (N=426)	N(cases) (N=296)	OR(CI 95%)	P dom	N (controls) (N=330)	N(cases) (N=217)	OR(CI 95%)	P dom	N (controls) (N=96)	N(cases) (N=79)	OR(CI95%)	P dom
rs111906529	MIR299, MIR380	TT	417 (98.1)	282 (96.6)	CoDominant	0.1995	323 (98.2)	211 897.2)	CoDominant	0.4676	94 (97.9)	71 (94.7)	CoDominant	0.2523
		CT	8 (1.9)	10 (3.4)	1		6 (1.8)	6 (2.8)	1		2 (2.1)	4 (5.3)	1	
					1.85 (0.72-4.74)				1.53 (0.49-4.81)				2.65 (0.47-14.86)	
rs112328520	MIR520G	CC	364 (86.1)	257 (88.9)	CoDominant	0.2558	282 (86.2)	194 (89.8)	CoDominant	0.2102	82 (85.4)	63 (86.3)	CoDominant	0.8702
		CT	59 (13.9)	32 (11.1)	1		45 (13.8)	22 (10.2)	1		14 (14.6)	10 (13.7)	1	
					0.77 (0.49-1.22)				0.71 (0.41-1.22)				0.93 (0.39-2.23)	
rs11269	mir-1282					M				M				M
rs113808830	MIR4532	CC	340 (80.2)	233 (80.3)	Dominant	0.9589	270 (82.1)	180 (83.7)	Dominant	0.6169	70 (73.7)	53 (70.7)	Dominant	0.6626
		CT	80 (18.9)	57 (19.7)	1		56 (17.0)	35 (16.3)	1		24 (25.3)	22 (29.3)	1	
		TT	4 (0.9)	0 (0.0)	0.99 (0.68-1.44)		3 (0.9)	0 (0.0)	0.89 (0.56-1.41)		1 (1.1)	0 (0.0)	1.16 (0.59-2.28)	
rs116932476	hsa-mir-4479	GG	418 (98.8)	284 (97.6)	CoDominant	0.2166	326 (99.1)	211 (97.7)	CoDominant	0.1891	92 97.9	73 97.3	Dominant	0.8194
		AG	5 (1.2)	7 (2.4)	1		3 (0.9)	5 (2.3)	1		2 (2.1)	2 (2.7)	1	
					2.06 (0.65-6.56)				2.58 (0.61-10.89)				1.26 (0.17-9.16)	
rs117258475	MIR296	GG	414 (97.4)	281 (96.2)	Dominant	0.3725	320 (97.3)	209 (96.3)	Dominant	0.5346	94 (97.9)	72 (96.0)	Dominant	0.4624
		AG	11 (2.6)	9 (3.1)	1		9 (2.7)	7 (3.2)	1		2 (2.1)	2 (2.7)	1	
		AA	0 (0.0)	2 (0.7)	1.47 (0.63-3.44)		0 (0.0)	1 (0.5)	1.36 (0.52-3.58)		0 (0.0)	1 (1.3)	1.96 (0.32-12.03)	
rs117650137	hsa-mir-6717	GG	399 (93.7)	271 (92.8)	CoDominant	0.6539	310 (93.9)	202 (93.1)	CoDominant	0.6916	89 (92.7)	69 (92.0)	CoDominant	0.8625
		AG	27 (6.3)	21 (7.2)	1		20 (6.1)	15 (6.9)	1		7 (7.3)	6 (8.0)	1	
					1.15 (0.63-2.07)				1.15 (0.58-2.3)				1.11 (0.36-3.44)	
rs117723462	MIR3649	TT			CoDominant		326 (98.8)	212 (97.7)	Dominant	0.3326	96 (100.0)	73 (97.3)	CoDominant	0.1909
		GT			1		4 (1.2)	5 (2.3)	1		0 (0.0)	2 (2.7)	1	
									1.92 (0.51-7.24)				0 (0.0)	
rs163642	MIR4436B2					ERROR				ERROR				ERROR

Abbreviations: OR Odd Ratio, CI Confidence Interval, M monomorphic, H-W No Hardy-Weimber equilibrium

DISCUSSION

The main goal of the present study was to prove the strong genetic component in the etiology of childhood B-ALL by identifying genetic susceptibility markers in coding regions as well as non coding regions. To that end, on the one hand we have determined whether the loci previously proposed by the GWAS were associated with B-ALL risk in the Spanish population, and on the other hand, we have determined the involvement of genetic variants in miRNA related genes in the susceptibility of B-ALL.

Considering the obtained results in the GWAS in relation with B-ALL susceptibility, up to date, five loci have been proposed to be involved in the disease: *ARID5B*, *IKZF1*, *CEBPE*, *CDKN2A/B* and *PIP4K2A*.

In *ARID5B*, we confirmed the association previously proposed by the all the GWAS (Treviño et al. 2009, Papaemmanuil et al. 2009, Orsi et al. 2012, Migliorini et al. 2013, Xu et al. 2013). Interestingly, this association between *ARID5B* and B-ALL risk was also confirmed for different ethnic groups (Healy et al. 2010, Prasad et al., 2010, Pastorczack et al., 2011, Ross el al, 2012, Lautner-Csorba et al., 2012, Chokkalingam et al., 2013, Yang et al. 2010, Han et al. 2010, Xu et al. 2012, Vijayakrishnan et al. 2010, Lin et al. 2014, Wang et al. 2013, Emerenciano et al. 2014, Kreile et al. 2016, Gharbi et al. 2016, Bahari et al. 2016), which supports the hypothesis that *ARID5B* is involved in a general mechanism that contributes to the etiology of childhood ALL. Of note is that the specific role of *ARID5B* in B-ALL remains to be elucidated, as well as the function of the SNPs associated. By *in silico* analysis, we could determine that rs10821936, rs7073837, rs7089424, rs7087507, and rs10821938 disrupt the binding site of transcriptional factors (Lee and Shatkay 2008) and rs7923074 and rs4131566 removed CpG sites (Samuelsson et al. 2011). These changes could potentially affect *ARID5B* regulation but further functional studies are needed to explain the implication of these SNPs and the *ARID5B* gene in B-ALL susceptibility.

In *IKZF1*, our results have confirmed the association of rs4132601, one of the highest association signals found by GWAS (Treviño et al. 2009, Papaemmanuil et al. 2009, Orsi et al. 2012, Migliorini et al. 2013, Xu et al. 2013), with B-ALL susceptibility in our cohort. Although several studies replicated the association between *IKZF1* rs4132601 polymorphism and ALL risk (Prasad et al. 2010, Vijayakrishnan et al., 2010 Pastorczack et al., 2011, Ross el al, 2012, Ellinghaus et al, 2012, Lautner-Csorba et al., 2012, Linabery et al., 2013, Chokkalingam et al., 2013 Bhandari et al., 2016 Gharbi et al., 2016), some studies did not (Healy et al. 2010, Wang et al., 2013, Emerenciano et al. 2014, Lin et al., 2014, Kennedy et al., 2015, Kreile et al., 2016).

In order to clarify the possible association between rs4132601 and risk of ALL we performed a meta-analysis. Our results confirmed that GG genotype of rs4132601 at *IKZF1* increased the risk of B-ALL. This SNP is located in 3´UTR region of *IKZF1*, a region involved in gene regulation. In this line, Papaemmanuil et al. found a significantly decreased expression associated with the G risk allele (Papaemmanuil et al. 2009); however, the functional explanation of this SNP was not fully elucidated. Recently, thanks to the publication of ENCyclopedia of DNA Elements (ENCODE) (Consortium 2012), we have found that rs4132601 is located in a miRNA binding site, in which the G allele creates a binding for mir-4772 and mir-3937 (Gong et al. 2015). The creation of these miRNA binding sites could decrease IKZF1 expression previously described. Therefore, the risk allele G of *IKZF1* creates a miRNA binding site, and the binding of miRNAs could downregulate the expression of IKZF1.

At *CEBPE* locus, our results confirmed that TT genotype of rs2239633 decreased the risk of B-ALL, previously described by GWAS (Papaemmanuil et al. 2009, Orsi et al. 2012, Migliorini et al. 2013, Xu et al. 2013). Again, this finding were replicated in some populations (Prasad et al. 2010, Han et al., 2010, Lautner-Csorba et al. 2012, Chokkalingam et al., 2013, Gharbi et al., 2016), but not in others (Vijayakrishnan et al., 2010, Healy et al. 2010, Pastorczack et al., 2011, Ross et al. 2013, Wang et al., 2013, Emerenciano et al. 2014, Kennedy et al., 2015, Bhandari et al., 2016, Kreile et al., 2016), and therefore, we conducted a meta-analysis. Our meta-analysis provides a robust evidence for association of the TT genotype and a decreased risk of B-ALL. However, in spite of this clear association, the SNP rs2239633, located 25.7-kb upstream of *CEBPE,* has an unknown function. Recently, Wiemels et al. identified the SNP rs2239635 located in the promoter region highly associated with B-ALL risk. In line with their results, we also found rs2239635 associated with an increased risk of B-ALL. Interestingly, the risk allele C of rs2239635 disrupts the binding of Ikaros (Wiemels et al. 2016), indicating an interaction between both genes, which has been confirmed in our study. This disruption avoids CEBPE repression and it has been suggested that incomplete suppression of *CEBPE* by Ikaros may lead to lineage confusion, a common feature of leukemogenesis (Wiemels et al. 2016).

Regarding *CDKN2A/B* locus, in our study, we confirmed the previos association described by the GWAS (Sherborne et al. 2010, Xu et al. 2013, Migliorini et al. 2013). We analyzed 4 SNPs previosly identified and 2 new functional SNPs. The most significant finding was the association between rs2811712, which was confirmed by different studies (Sherborne et al. 2010, Chokkalingam et al. 2013, Migliorini et al. 2013, Hungate et al. 2016). rs2811712 is located in intron 1 of *ANRIL,* which may affect its expression or its structure. *ANRIL* silences *CDKN2A* and

CDKN2B genes (Meseure et al. 2016), and therefore, alterations in *ANRIL* function could affect the expression of both genes. The second most significant association was found for the T allele of rs3731249, association that was also described recently by 3 independent studies (Walsh et al. 2015, Xu et al. 2015, Vijayakrishnan et al. 2015). rs3731249 localizes to exon 2 of *CDKN2A*, being shared by both $p16^{INK4A}$ and $p14^{ARF}$. The variant $p16^{INK4A}$ is preferentially retained in the nucleus, compromising its ability to inhibit CDK4 and CDK6 in the cytoplasm (Xu et al. 2015) and favouring proliferation. In the $p14^{ARF}$, rs3731249 is in the 3´UTR region, where the risk allele creates a miRNA binding site (Gong et al. 2012). It is the same for the third finding at rs3217992, which is also located in a miRNA binding site of *CDKN2B.* MiRNAs could downregulate $p14^{ARF}$ and $p15^{INK4B}$ expression, and then, attenuate its function as cyclin inhibitor. Finally, we found rs2811709 associated with risk of B-ALL, confirming the results of two previous studies (Sherborne et al. 2010, Orsi et al. 2012). rs2811709 is a cis-eQTL for *CDKN2B*, with a decreased expression of CDKN2B mRNA for the risk allele (Consortium 2013).

Contrary as it was expected, we found no association between rs3731222, in high LD with the first SNP found by Sherborne et al., rs3731217, and B-ALL susceptibility. However, we and others could not replicate this association (Pastorczak et al. 2011, Vijayakrishnan et al. 2010). This may be due to the fact that in each population CDKN2A/B function could be altered by diverse mechanisms, supporting the hypothesis that there are different association signals at *CDKN2A/B* associated with B-ALL risk.

Finally, at *PIP4K2A* locus, we did not validate the association between rs7088318 and B-ALL risk in our population. Although this locus was identified as the fith loci associated with B-ALL in two GWAS independently, the only study that tried to replicate this association (Chokkalingam et al. 2013), could not confirme it. Differences in results might be due to differences in genetic composition among populations. In fact, if we have a look at the risk allele frequency (RAF), in our control population this is higher than in the control European/American population analyzed by Xu et al., and similar to what they observed in the case European/American population (Xu et al. 2013). This peculiarity makes it difficult to find significant differences in frequency between cases and controls. Therefore, SNPs in *PIP4K2A* may be associated with pediatric ALL risk in some populations but it does not seem to be a good susceptibility marker in the Spanish cohort.

However, all the loci reported in ALL GWASs thus far cumulatively accounted for only 8% of genetic variation in ALL risk, suggesting additional susceptibility variants yet to be discovered (Enciso-Mora et al. 2012).

In this line, when we analyzed in deep all the significant SNPs found by GWAS, we observed that almost 40% of significant signals reported were located in non-coding region, suggesting that these regions, such as miRNAs, could play an important role B-ALL risk. MiRNAs regulated more than 50% of human genes, including those involved in B-cell maturation, differentiation and proliferation. Changes in miRNA function may occur through SNPs in miRNA-related genes. For instance, SNPs in genes involved in miRNA processing can affect levels of miRNA expression, whereas SNPs in miRNA genes can affect miRNA biogenesis and function. Therefore, alterations in miRNA function could be involved in the origin of B-ALL.

Regarding miRNA processing genes, we found 8 SNPs in 6 genes associated with B-ALL risk in our population. Among them, the most significant finding was the SNP rs139919 in *TNRC6B*. This gene encodes a RNA interference machinery component, which contributes to the RISC complex, and is crucial for miRNA-dependent translational repression or degradation of target mRNAs (Tao et al. 2012). It has been suggested that alterations in the expression of *TNRC6B* due to genetic variations may affect mRNA levels that are normally regulated by *TNRC6B*, thereby affecting carcinogenesis (Sun et al. 2009). We also found other SNPs located in genes of the RISC complex associated with B-ALL susceptibility. These included two SNPs in *EIF2C1* (rs595961 and rs636832), one SNP in *CNOT1* (rs11866002), and one SNP in *CNOT6* (rs6877400). SNPs that affect expression levels of these proteins may have deleterious effects on miRNA–mRNA interactions and may affect cancer development and progression. Finally, we also found rs10035440 in *DROSHA*, and rs9606248 and rs1640299 in *DGCR8* associated with B-ALL. These three SNPs have putative roles in transcriptional regulation and may affect the expression levels of *DROSHA* or *DGCR8*. As a result, levels of miRNAs could be affected, leading to an increased ALL risk.

Regarding miRNA genes, the SNPs rs12402181 in mir-3117 and rs62571442 in mir-3689d2 showed statistically significant association with B-ALL risk. rs12402181 is located in the seed region of mir-3117-3p, therefore, it could affect the accurate recognition of its target mRNA sequences. Among the target genes of mir-3117-3p, *in silico* analysis determined that genes of MAPK signaling pathway are over-represented, mainly those of the MAPK/ERK family or classical pathway (Kamburov et al. 2013, Dweep and Gretz 2015, Kanehisa et al. 2016).

Remarkably, the genes predicted to be targeted for mir-3117-3p are in the first steps of the cascade, which in turn could produce the deregulation of the following steps. The second SNP, rs62571442 in mir-3689d2 is located in the pre-miRNA sequence and the hairpin structure of the miRNA changes from stable to unstable status with the presence of the risk allele. When the SNP decreases the stability, the product of mature miRNA is reduced, which in turn may increase the target gene expression (Gong et al. 2012). In the pathway analysis of mir-3689d2, enriched pathways were again related to Ras signaling. Therefore, an alteration in the function of both miRNAs could affect the expression of Ras-related genes. Aberrant expression of this pathway is a major and highly prevalent oncogenic event in childhood ALL (Barbosa et al. 2014, Case et al. 2008). Therefore, miRNAs could contribute in this process (Masliah-Planchon et al. 2015).

In summary, we have validated the association of four of the loci proposed by literature at *ARID5B*, *IKZF1*, *CEBPE*, *CDKN2A/B* and B-ALL risk. Searching for new variants associated with B-ALL susceptibility, we found 8 significant SNPs in 6 genes of miRNA processing pathway, and 2 SNPs in mir-3117-3p and mir-3689d2, associated with B-ALL susceptibility.

Therefore, our results support the strong genetic component of B-ALL, which could be due, at least in part, to genetic variants in coding genes as well as in non coding genes.

CONCLUSIONS

CONCLUSIONS

A. After the analysis of genetic variants previously proposed by the literature as possible candidate risk variants for LLA-B, we have validated four out of five loci in the Spanish population. With these results:
    a) We confirm the involvement of *ARID5B* as a general susceptibility marker for B-ALL risk.
    b) We replicate the association of *IKZF1* with B-ALL risk and we propose that rs4132601 could do its effect through the creation of a miRNA binding site.
    c) We validate the association of *CEBPE* with B-ALL risk, supporting that the alteration of Ikaros binding could be the causal effect.
    d) We replicate the results in *CDKN2A/B* confirming the genetic heterogeneity among different populations in the association at this locus.

B. Our results in miRNA processing genes identify eight SNPs in six genes of miRNA processing pathway. We propose that these SNPs alter miRNA level and, in turn, affect genes involved in B-ALL susceptibility

C. In miRNA genes, we identify two SNPs in mir-3117-3p and mir-3689d2 associated with B-ALL risk. These SNPs could alter the miRNA function, and then, affect their target genes in the MAPK signaling pathway, which is involved in ALL.

In summary, our results support a strong genetic component of B-ALL in our Spanish cohort. This genetic component is based on genetic variants in coding genes as well as in non coding genes. We open a new field in B-ALL susceptibility based on the study of non coding regions.

REFERENCES

A

Akhtar, M. M., Micolucci, L., Islam, M. S., Olivieri, F. and Procopio, A. D. (2016) 'Bioinformatic tools for microRNA dissection', *Nucleic Acids Res,* 44(1), 24-44.

Akasaka, T., Balasas, T., Russell, L. J., Sugimoto, K. J., Majid, A., Walewska, R., Karran, E. L., Brown, D. G., Cain, K., Harder, L., Gesk, S., Martin-Subero, J. I., Atherton, M. G., Brüggemann, M., Calasanz, M. J., Davies, T., Haas, O. A., Hagemeijer, A., Kempski, H., Lessard, M., Lillington, D. M., Moore, S., Nguyen-Khac, F., Radford-Weiss, I., Schoch, C., Struski, S., Talley, P., Welham, M. J., Worley, H., Strefford, J. C., Harrison, C. J., Siebert, R. and Dyer, M. J. (2007) 'Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL)', *Blood,* 109(8), 3451-61.

Akkiz, H., Bayram, S., Bekar, A., Akgöllü, E. and Üsküdar, O. (2011) 'Genetic variation in the microRNA-499 gene and hepatocellular carcinoma risk in a Turkish population: lack of any association in a case-control study', *Asian Pac J Cancer Prev,* 12(11), 3107-12.

Akbari Moqadam, F., Lange-Turenhout, E. A., van der Veer, A., Marchante, J. R., Boer, J. M., Pieters, R. and den Boer, M. (2014) 'MicroRNA signature in BCR-ABL1-like and BCR-ABL1-positive childhood acute lymphoblastic leukemia: similarities and dissimilarities', *Leuk Lymphoma,* 55(8), 1942-5.

Alshatwi, A. A., Shafi, G., Hasan, T. N., Syed, N. A., Al-Hazzani, A. A., Alsaif, M. A. and Alsaif, A. A. (2012) 'Differential expression profile and genetic variants of microRNAs sequences in breast cancer patients', *PLoS One,* 7(2), e30049.

Andersson, A., Ritz, C., Lindgren, D., Edén, P., Lassen, C., Heldrup, J., Olofsson, T., Råde, J., Fontes, M., Porwit-Macdonald, A., Behrendtz, M., Höglund, M., Johansson, B. and Fioretos, T. (2007) 'Microarray-based classification of a consecutive series of 121 childhood acute leukemias: prediction of leukemic and genetic subtype as well as of minimal residual disease status', *Leukemia,* 21(6), 1198-203.

B

Bahari, G., Hashemi, M., Naderi, M. and Taheri, M. (2016) 'IKZF1 gene polymorphisms increased the risk of childhood acute lymphoblastic leukemia in an Iranian population', *Tumour Biol,* 37(7), 9579-86.

Barbosa, T. C., Andrade, F. G., Lopes, B. A., de Andrade, C. F., Mansur, M. B., Emerenciano, M. and Pombo-de-Oliveira, M. S. (2014) 'Impact of mutations in FLT3, PTPN11 and RAS genes on the overall survival of pediatric B cell precursor acute lymphoblastic leukemia in Brazil', *Leuk Lymphoma,* 55(7), 1501-9.

Belver, L., de Yébenes, V. G. and Ramiro, A. R. (2010) 'MicroRNAs prevent the generation of autoreactive antibodies', *Immunity,* 33(5), 713-22.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the false discovery rate: A practical and powerful approach to multiple testing', Soc Series B(

Bilbao-Aldaiturriaga, N., Askaiturrieta, Z., Granado-Tajada, I., Goričar, K., Dolžan, V., For The Slovenian Osteosarcoma Study Group, Garcia-Miguel, P., Garcia de Andoin, N., Martin-Guerrero, I. and Garcia-Orad, A. (2016) 'A systematic review and meta-analysis of MDM2 polymorphisms in osteosarcoma susceptibility', *Pediatr Res,* 80(4), 472-9.

Bilbao-Aldaiturriaga, N., Gutierrez-Camino, A., Martin-Guerrero, I., Pombar-Gomez, M., Zalacain-Diez, M., Patiño-Garcia, A., Lopez-Lopez, E. and Garcia-Orad, A. (2015) 'Polymorphisms in miRNA processing genes and their role in osteosarcoma risk', *Pediatr Blood Cancer,* 62(5), 766-9.

Bohnsack, M. T., Czaplinski, K. and Gorlich, D. (2004) 'Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs', *RNA,* 10(2), 185-91.

Boon, R. A., Jaé, N., Holdt, L. and Dimmeler, S. (2016) 'Long Noncoding RNAs: From Clinical Genetics to Therapeutic Targets?', *J Am Coll Cardiol,* 67(10), 1214-26.

Bourquin, J. P., Subramanian, A., Langebrake, C., Reinhardt, D., Bernard, O., Ballerini, P., Baruchel, A., Cavé, H., Dastugue, N., Hasle, H., Kaspers, G. L., Lessard, M., Michaux, L., Vyas, P., van Wering, E., Zwaan, C. M., Golub, T. R. and Orkin, S. H. (2006) 'Identification of distinct molecular

phenotypes in acute megakaryoblastic leukemia by gene expression profiling', *Proc Natl Acad Sci U S A,* 103(9), 3339-44.

Brisson, G. D., Alves, L. R. and Pombo-de-Oliveira, M. S. (2015) 'Genetic susceptibility in childhood acute leukaemias: a systematic review', *Ecancermedicalscience,* 9, 539.

C

Case, M., Matheson, E., Minto, L., Hassan, R., Harrison, C. J., Bown, N., Bailey, S., Vormoor, J., Hall, A. G. and Irving, J. A. (2008) 'Mutation of genes affecting the RAS pathway is common in childhood acute lymphoblastic leukemia', *Cancer Res,* 68(16), 6803-9.

Catto, J. W., Miah, S., Owen, H. C., Bryant, H., Myers, K., Dudziec, E., Larré, S., Milo, M., Rehman, I., Rosario, D. J., Di Martino, E., Knowles, M. A., Meuth, M., Harris, A. L. and Hamdy, F. C. (2009) 'Distinct microRNA alterations characterize high- and low-grade bladder cancer', *Cancer Res,* 69(21), 8472-81.

Catucci, I., Yang, R., Verderio, P., Pizzamiglio, S., Heesen, L., Hemminki, K., Sutter, C., Wappenschmidt, B., Dick, M., Arnold, N., Bugert, P., Niederacher, D., Meindl, A., Schmutzler, R. K., Bartram, C. C., Ficarazzi, F., Tizzoni, L., Zaffaroni, D., Manoukian, S., Barile, M., Pierotti, M. A., Radice, P., Burwinkel, B. and Peterlongo, P. (2010) 'Evaluation of SNPs in miR-146a, miR196a2 and miR-499 as low-penetrance alleles in German and Italian familial breast cancer cases', *Hum Mutat,* 31(1), E1052-7.

Chang, J. S., Wiemels, J. L., Chokkalingam, A. P., Metayer, C., Barcellos, L. F., Hansen, H. M., Aldrich, M. C., Guha, N., Urayama, K. Y., Scélo, G., Green, J., May, S. L., Kiley, V. A., Wiencke, J. K. and Buffler, P. A. (2010) 'Genetic polymorphisms in adaptive immunity genes and childhood acute lymphoblastic leukemia', *Cancer Epidemiol Biomarkers Prev,* 19(9), 2152-63.

Chang, L. W., Payton, J. E., Yuan, W., Ley, T. J., Nagarajan, R. and Stormo, G. D. (2008) 'Computational identification of the normal and perturbed genetic networks involved in myeloid differentiation and acute promyelocytic leukemia', *Genome Biol,* 9(2), R38.

Chokkalingam, A. P., Hsu, L. I., Metayer, C., Hansen, H. M., Month, S. R., Barcellos, L. F., Wiemels, J. L. and Buffler, P. A. (2013) 'Genetic variants in ARID5B and CEBPE are childhood ALL susceptibility loci in Hispanics', *Cancer Causes Control,* 24(10), 1789-95.

Chung, T. K., Lau, T. S., Cheung, T. H., Yim, S. F., Lo, K. W., Siu, N. S., Chan, L. K., Yu, M. Y., Kwong, J., Doran, G., Barroilhet, L. M., Ng, A. S., Wong, R. R., Wang, V. W., Mok, S. C., Smith, D. I., Berkowitz, R. S. and Wong, Y. F. (2012) 'Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXC1', *Int J Cancer,* 130(5), 1036-45.

Cobaleda, C., Gutiérrez-Cianca, N., Pérez-Losada, J., Flores, T., García-Sanz, R., González, M. and Sánchez-García, I. (2000) 'A primitive hematopoietic cell is the target for the leukemic transformation in human philadelphia-positive acute lymphoblastic leukemia.', *Blood,* 95(3), 1007-13.

Congrains, A., Kamide, K., Ohishi, M. and Rakugi, H. (2013) 'ANRIL: molecular mechanisms and implications in human health', *Int J Mol Sci,* 14(1), 1278-92.

Consortium, E. P. (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature,* 489(7414), 57-74.

Consortium, G. (2010) 'A map of human genome variation from population-scale sequencing.', 467(7319)(1061-1073.

Consortium, G. (2013) 'The Genotype-Tissue Expression (GTEx) project', *Nat Genet,* 45(6), 580-5.

D

Dai, Y. E., Tang, L., Healy, J. and Sinnett, D. (2014) 'Contribution of polymorphisms in IKZF1 gene to childhood acute leukemia: a meta-analysis of 33 case-control studies', *PLoS One,* 9(11), e113748.

de Oliveira, J. C., Scrideli, C. A., Brassesco, M. S., Morales, A. G., Pezuk, J. A., Queiroz, R. e. P., Yunes, J. A., Brandalise, S. R. and Tone, L. G. (2012) 'Differential miRNA expression in childhood acute lymphoblastic leukemia and association with clinical and biological features', *Leuk Res,* 36(3), 293-8.

DH, A., H, R., YK, C., YJ, J., KT, M., K, K., SP, H., SG, H. and NK, K. (2013) 'Association of the miR-146aC>G , miR-149T>C , miR-196a2T>C , and miR-499A>G polymorphisms with gastric cancer risk and survival in the Korean population', 52(Suppl 1:E39-51),

Duyu, M., Durmaz, B., Gunduz, C., Vergin, C., Yilmaz Karapinar, D., Aksoylar, S., Kavakli, K., Cetingul, N., Irken, G., Yaman, Y., Ozkinay, F. and Cogulu, O. (2014) 'Prospective evaluation of whole genome microRNA expression profiling in childhood acute lymphoblastic leukemia', *Biomed Res Int,* 2014, 967585.

Dweep, H. and Gretz, N. (2015) 'miRWalk2.0: a comprehensive atlas of microRNA-target interactions', *Nat Methods,* 12(8), 697.

E

Egger, M., Davey Smith, G., Schneider, M. and Minder, C. (1997) 'Bias in meta-analysis detected by a simple, graphical test', *BMJ,* 315(7109), 629-34.

Ellinghaus, E., Stanulla, M., Richter, G., Ellinghaus, D., te Kronnie, G., Cario, G., Cazzaniga, G., Horstmann, M., Panzer Grümayer, R., Cavé, H., Trka, J., Cinek, O., Teigler-Schlegel, A., ElSharawy, A., Häsler, R., Nebel, A., Meissner, B., Bartram, T., Lescai, F., Franceschi, C., Giordan, M., Nürnberg, P., Heinzow, B., Zimmermann, M., Schreiber, S., Schrappe, M. and Franke, A. (2012) 'Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia', *Leukemia,* 26(5), 902-9.

Emerenciano, M., Barbosa, T. C., Lopes, B. A., Blunck, C. B., Faro, A., Andrade, C., Meyer, C., Marschalek, R., Pombo-de-Oliveira, M. S. and Leukemia, B. C. S. G. o. I. A. (2014) 'ARID5B polymorphism confers an increased risk to acquire specific MLL rearrangements in early childhood leukemia', *BMC Cancer,* 14, 127.

Esteller, M. (2011) 'Non-coding RNAs in human disease', *Nat Rev Genet,* 12(12), 861-74.

F

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., Matthews, L., May, B., Milacic, M., Rothfels, K., Shamovsky, V., Webber, M., Weiser, J., Williams, M., Wu, G., Stein, L., Hermjakob, H. and D'Eustachio, P. (2016) 'The Reactome pathway Knowledgebase', *Nucleic Acids Res,* 44(D1), D481-7.

Fang, D. H., Ji, Q., Fan, C. H., An, Q. and Li, J. (2014) 'Methionine synthase reductase A66G polymorphism and leukemia risk: evidence from published studies', *Leuk Lymphoma,* 55(8), 1910-4.

Friedman, R. C., Farh, K. K., Burge, C. B. and Bartel, D. P. (2009) 'Most mammalian mRNAs are conserved targets of microRNAs', *Genome Res,* 19(1), 92-105.

G

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J. and Altshuler, D. (2002) 'The structure of haplotype blocks in the human genome', *Science,* 296(5576), 2225-9.

Gao, X., Qiao, Y., Han, D., Zhang, Y. and Ma, N. (2012) 'Enemy or partner: relationship between intronic micrornas and their host genes', *IUBMB Life,* 64(10), 835-40.

George, G. P., Gangwar, R., Mandal, R. K., Sankhwar, S. N. and Mittal, R. D. (2011) 'Genetic variation in microRNA genes and prostate cancer risk in North Indian population', *Mol Biol Rep,* 38(3), 1609-15.

Georgopoulos, K., Bigby, M., Wang, J. H., Molnar, A., Wu, P., Winandy, S. and Sharpe, A. (1994) 'The Ikaros gene is required for the development of all lymphoid lineages', *Cell,* 79(1), 143-56.

Gharbi, H., Ben Hassine, I., Soltani, I., Safra, I., Ouerhani, S., Bel Haj Othmen, H., Teber, M., Farah, A., Amouri, H., Toumi, N. H., Abdennebi, S., Abbes, S. and Menif, S. (2016) 'Association of genetic variation in IKZF1, ARID5B, CDKN2A, and CEBPE with the risk of acute lymphoblastic leukemia in

Tunisian children and their contribution to racial differences in leukemia incidence', *Pediatr Hematol Oncol,* 33(3), 157-67.

Goldstein, D. B. and Cavalleri, G. L. (2005) 'Genomics: understanding human diversity', *Nature,* 437(7063), 1241-2.

Goldstein, D. B. and Weale, M. E. (2001) 'Population genomics: linkage disequilibrium holds the key', *Curr Biol,* 11(14), R576-9.

Gong, J., Liu, C., Liu, W., Wu, Y., Ma, Z., Chen, H. and Guo, A. Y. (2015) 'An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools', *Database (Oxford),* 2015, bav029.

Gong, J., Tong, Y., Zhang, H. M., Wang, K., Hu, T., Shan, G., Sun, J. and Guo, A. Y. (2012) 'Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis', *Hum Mutat,* 33(1), 254-63.

Greaves, M. (2006) 'Infection, immune responses and the aetiology of childhood leukaemia', *Nat Rev Cancer,* 6(3), 193-203.

Greaves, M. F. (1988) 'Speculations on the cause of childhood acute lymphoblastic leukemia', *Leukemia,* 2(2), 120-5.

Greaves, M. F. (1997) 'Aetiology of acute leukaemia', *Lancet,* 349(9048), 344-9.

Greaves, M. F., Maia, A. T., Wiemels, J. L. and Ford, A. M. (2003) 'Leukemia in twins: lessons in natural history', *Blood,* 102(7), 2321-33.

Gregory, R. I., Chendrimada, T. P. and Shiekhattar, R. (2006) 'MicroRNA biogenesis: isolation and characterization of the microprocessor complex', *Methods Mol Biol,* 342, 33-47.

Guo, L., Zhao, Y., Zhang, H., Yang, S. and Chen, F. (2014) 'Integrated evolutionary analysis of human miRNA gene clusters and families implicates evolutionary relationships', *Gene,* 534(1), 24-32.

Gutierrez-Camino, A., Lopez-Lopez, E., Martin-Guerrero, I., Piñan, M. A., Garcia-Miguel, P., Sanchez-Toledo, J., Carbone Bañeres, A., Uriz, J., Navajas, A. and Garcia-Orad, A. (2014) 'Noncoding RNA-related polymorphisms in pediatric acute lymphoblastic leukemia susceptibility', *Pediatr Res,* 75(6), 767-73.

Gutiérrez-Camino, Á., López-López, E., Martín-Guerrero, I., Sánchez-Toledo, J., García de Andoin, N., Carboné Bañeres, A., García-Miguel, P., Navajas, A. and García-Orad, Á. (2013) 'Intron 3 of the ARID5B gene: a hot spot for acute lymphoblastic leukemia susceptibility', *J Cancer Res Clin Oncol,* 139(11), 1879-86.

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L., Friedman, J. M. and Pe'er, I. (2009) 'Whole population, genome-wide mapping of hidden relatedness', *Genome Res,* 19(2), 318-26.

H

Han, F., Tan, Y., Cui, W., Dong, L. and Li, W. (2013) 'Novel insights into etiologies of leukemia: a HuGE review and meta-analysis of CYP1A1 polymorphisms and leukemia risk', *Am J Epidemiol,* 178(4), 493-507.

Han, S., Lee, K. M., Park, S. K., Lee, J. E., Ahn, H. S., Shin, H. Y., Kang, H. J., Koo, H. H., Seo, J. J., Choi, J. E., Ahn, Y. O. and Kang, D. (2010) 'Genome-wide association study of childhood acute lymphoblastic leukemia in Korea', *Leuk Res,* 34(10), 1271-4.

Hasani, S. S., Hashemi, M., Eskandari-Nasab, E., Naderi, M., Omrani, M. and Sheybani-Nasab, M. (2013) 'A functional polymorphism in the miR-146a gene is associated with the risk of childhood acute lymphoblastic leukemia: a preliminary report', *Tumour Biol*.

Healy, J., Richer, C., Bourgey, M., Kritikou, E. A. and Sinnett, D. (2010) 'Replication analysis confirms the association of ARID5B with childhood B-cell acute lymphoblastic leukemia', *Haematologica,* 95(9), 1608-11.

Heerema, N. A., Raimondi, S. C., Anderson, J. R., Biegel, J., Camitta, B. M., Cooley, L. D., Gaynon, P. S., Hirsch, B., Magenis, R. E., McGavran, L., Patil, S., Pettenati, M. J., Pullen, J., Rao, K., Roulston, D., Schneider, N. R., Shuster, J. J., Sanger, W., Sutcliffe, M. J., van Tuinen, P., Watson, M. S. and Carroll, A. J. (2007) 'Specific extra chromosomes occur in a modal number dependent pattern in pediatric acute lymphoblastic leukemia.', *Genes Chromosomes Cancer,* 46(7), 684-93.

Hinske, L. C., Galante, P. A., Kuo, W. P. and Ohno-Machado, L. (2010) 'A potential role for intragenic miRNAs on their hosts' interactome', *BMC Genomics,* 11, 533.

HK, K., L, P.-O. and SJ., C. (2012) 'Common genetic variants in miR-1206 (8q24.2) and miR-612 (11q13.3) affect biogenesis of mature miRNA forms', 7(10), e47454.

Hosking, F. J., Leslie, S., Dilthey, A., Moutsianas, L., Wang, Y., Dobbins, S. E., Papaemmanuil, E., Sheridan, E., Kinsey, S. E., Lightfoot, T., Roman, E., Irving, J. A., Allan, J. M., Taylor, M., Greaves, M., McVean, G. and Houlston, R. S. (2011) 'MHC variation and risk of childhood B-cell precursor acute lymphoblastic leukemia', *Blood,* 117(5), 1633-40.

Horikawa, Y., Wood, C. G., Yang, H., Zhao, H., Ye, Y., Gu, J., Lin, J., Habuchi, T. and Wu, X. (2008) 'Single nucleotide polymorphisms of microRNA machinery genes modify the risk of renal cell carcinoma', *Clin Cancer Res,* 14(23), 7956-62.

Houlston, R. S. (2010) 'Low-penetrance susceptibility to hematological malignancy', *Curr Opin Genet Dev,* 20(3), 245-50.

Howie, B. N., Carlson, C. S., Rieder, M. J. and Nickerson, D. A. (2006) 'Efficient selection of tagging single-nucleotide polymorphisms in multiple populations', *Hum Genet,* 120(1), 58-68.

Hsu, L. I., Chokkalingam, A. P., Briggs, F. B., Walsh, K., Crouse, V., Fu, C., Metayer, C., Wiemels, J. L., Barcellos, L. F. and Buffler, P. A. (2015) 'Association of genetic variation in IKZF1, ARID5B, and CEBPE and surrogates for early-life infections with the risk of acute lymphoblastic leukemia in Hispanic children', *Cancer Causes Control,* 26(4), 609-19.

Hu, Z., Chen, J., Tian, T., Zhou, X., Gu, H., Xu, L., Zeng, Y., Miao, R., Jin, G., Ma, H., Chen, Y. and Shen, H. (2008) 'Genetic variants of miRNA sequences and non-small cell lung cancer survival', *J Clin Invest,* 118(7), 2600-8.

Huang, L., Yu, D., Wu, C., Zhai, K., Jiang, G., Cao, G., Wang, C., Liu, Y., Sun, M., Li, Z., Tan, W. and Lin, D. (2012) 'Copy number variation at 6q13 functions as a long-range regulator and is associated with pancreatic cancer risk', *Carcinogenesis,* 33(1), 94-100.

Huang, J. T., Wang, J., Srivastava, V., Sen, S. and Liu, S. M. (2014) 'MicroRNA Machinery Genes as Novel Biomarkers for Cancer', *Front Oncol,* 4, 113.

Hungate, E. A., Vora, S. R., Gamazon, E. R., Moriyama, T., Best, T., Hulur, I., Lee, Y., Evans, T. J., Ellinghaus, E., Stanulla, M., Rudant, J., Orsi, L., Clavel, J., Milne, E., Scott, R. J., Pui, C. H., Cox, N. J., Loh, M. L., Yang, J. J., Skol, A. D. and Onel, K. (2016) 'A variant at 9p21.3 functionally implicates CDKN2B in paediatric B-cell precursor acute lymphoblastic leukaemia aetiology', *Nat Commun,* 7, 10635.

Hutvágner, G., McLachlan, J., Pasquinelli, A. E., Bálint, E., Tuschl, T. and Zamore, P. D. (2001) 'A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA', *Science,* 293(5531), 834-8.

I

Iacobucci, I., Sazzini, M., Garagnani, P., Ferrari, A., Boattini, A., Lonetti, A., Papayannidis, C., Mantovani, V., Marasco, E., Ottaviani, E., Soverini, S., Girelli, D., Luiselli, D., Vignetti, M., Baccarani, M. and Martinelli, G. (2011) 'A polymorphism in the chromosome 9p21 ANRIL locus is associated to Philadelphia positive acute lymphoblastic leukemia', *Leuk Res,* 35(8), 1052-9.

Inaba, H., Greaves, M. and Mullighan, C. G. (2013) 'Acute lymphoblastic leukaemia', *Lancet,* 381(9881), 1943-55.

Inada, T. and Makino, S. (2014) 'Novel roles of the multi-functional CCR4-NOT complex in post-transcriptional regulation', *Front Genet,* 5, 135.

J

Jiang Y, Chen J, Wu J, Hu Z, Qin Z, Liu X, et al. Evaluation of genetic variants in microRNA biosynthesis genes and risk of breast cancer in Chinese women. Int J Cancer. 2013;133(9):2216-24.

Johanson, T. M., Skinner, J. P., Kumar, A., Zhan, Y., Lew, A. M. and Chong, M. M. (2014) 'The role of microRNAs in lymphopoiesis', *Int J Hematol,* 100(3), 246-53.

Johnston, W. T., Lightfoot, T. J., Simpson, J. and Roman, E. (2010) 'Childhood cancer survival: a report from the United Kingdom Childhood Cancer Study', *Cancer Epidemiol,* 34(6), 659-66.

Ju, X., Li, D., Shi, Q., Hou, H., Sun, N. and Shen, B. (2009) 'Differential microRNA expression in childhood B-cell precursor acute lymphoblastic leukemia', *Pediatr Hematol Oncol,* 26(1), 1-10.

K

Kamburov, A., Stelzl, U., Lehrach, H. and Herwig, R. (2013) 'The ConsensusPathDB interaction database: 2013 update', *Nucleic Acids Res,* 41(Database issue), D793-800.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2016) 'KEGG as a reference resource for gene and protein annotation', *Nucleic Acids Res,* 44(D1), D457-62.

Kim, J. S., Choi, Y. Y., Jin, G., Kang, H. G., Choi, J. E., Jeon, H. S., Lee, W. K., Kim, D. S., Kim, C. H., Kim, Y. J., Son, J. W., Jung, T. H. and Park, J. Y. (2010) 'Association of a common AGO1 variant with lung cancer risk: a two-stage case-control study', *Mol Carcinog,* 49(10), 913-21.

Kim, V. N. (2004) 'MicroRNA precursors in motion: exportin-5 mediates their nuclear export', *Trends Cell Biol,* 14(4), 156-9.

Kim, W. H., Min, K. T., Jeon, Y. J., Kwon, C. I., Ko, K. H., Park, P. W., Hong, S. P., Rim, K. S., Kwon, S. W., Hwang, S. G. and Kim, N. K. (2012) 'Association study of microRNA polymorphisms with hepatocellular carcinoma in Korean population', *Gene,* 504(1), 92-7.

Kinlen, L. (1988) 'Evidence for an infective cause of childhood leukaemia: comparison of a Scottish new town with nuclear reprocessing sites in Britain', *Lancet,* 2(8624), 1323-7.

Klipper-Aurbach, Y., Wasserman, M., Braunspiegel-Weintrob, N., Borstein, D., Peleg, S., Assa, S., Karp, M., Benjamini, Y., Hochberg, Y. and Laron, Z. (1995) 'Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus', *Med Hypotheses,* 45(5), 486-90.

Koesters, R., Adams, V., Betts, D., Moos, R., Schmid, M., Siermann, A., Hassam, S., Weitz, S., Lichter, P., Heitz, P. U., von Knebel Doeberitz, M. and Briner, J. (1999) 'Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence, genomic organization, localization to chromosomal bands 1p34-p35, and expression', *Genomics,* 61(2), 210-8.

Koralov, S. B., Muljo, S. A., Galler, G. R., Krek, A., Chakraborty, T., Kanellopoulou, C., Jensen, K., Cobb, B. S., Merkenschlager, M., Rajewsky, N. and Rajewsky, K. (2008) 'Dicer ablation affects antibody diversity and cell survival in the B lymphocyte lineage', *Cell,* 132(5), 860-74.

Kozomara, A. and Griffiths-Jones, S. (2014) 'miRBase: annotating high confidence microRNAs using deep sequencing data', *Nucleic Acids Res,* 42(Database issue), D68-73.

Kreile, M., Piekuse, L., Rots, D., Dobele, Z., Kovalova, Z. and Lace, B. (2016) 'Analysis of possible genetic risk factors contributing to development of childhood acute lymphoblastic leukaemia in the Latvian population', *Arch Med Sci,* 12(3), 479-85.

KT, M., JW, K., YJ, J., MJ, J., SY, C., D, O. and NK, K. (2012) 'Association of the miR-146aC > G , 149C > T , 196a2C > T , and 499A > G polymorphisms with colorectal cancer in the Korean population', 51(Suppl 1:E65-73),

Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. and Jacks, T. (2007) 'Impaired microRNA processing enhances cellular transformation and tumorigenesis', *Nat Genet,* 39(5), 673-7.

L

Lahoud, M. H., Ristevski, S., Venter, D. J., Jermiin, L. S., Bertoncello, I., Zavarsek, S., Hasthorpe, S., Drago, J., de Kretser, D., Hertzog, P. J. and Kola, I. (2001) 'Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs', *Genome Res,* 11(8), 1327-34.

Lautner-Csorba, O., Gézsi, A., Erdélyi, D. J., Hullám, G., Antal, P., Semsei, Á., Kutszegi, N., Kovács, G., Falus, A. and Szalai, C. (2013) 'Roles of genetic polymorphisms in the folate pathway in childhood acute lymphoblastic leukemia evaluated by Bayesian relevance and effect size analysis', *PLoS One,* 8(8), e69843.

Lautner-Csorba, O., Gézsi, A., Semsei, A. F., Antal, P., Erdélyi, D. J., Schermann, G., Kutszegi, N., Csordás, K., Hegyi, M., Kovács, G., Falus, A. and Szalai, C. (2012) 'Candidate gene association study in pediatric acute lymphoblastic leukemia evaluated by Bayesian network based Bayesian multilevel analysis of relevance', *BMC Med Genomics,* 5, 42.

Layton Tovar, C. F. and Mendieta Zerón, H. (2016) 'Intracellular Signaling Pathways Involved in Childhood Acute Lymphoblastic Leukemia; Molecular Targets', *Indian J Hematol Blood Transfus,* 32(2), 141-53.

Lewis, B. P., Burge, C. B. and Bartel, D. P. (2005) 'Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets', *Cell,* 120(1), 15-20.

Lee, P. H. and Shatkay, H. (2008) 'F-SNP: computationally predicted functional SNPs for disease association studies', *Nucleic Acids Res,* 36(Database issue), D820-4.

Lee, Y. J., Kim, V., Muth, D. C. and Witwer, K. W. (2015) 'Validated MicroRNA Target Databases: An Evaluation', *Drug Dev Res,* 76(7), 389-96.

Li, S., Ren, L., Fan, L. and Wang, G. (2015) 'IKZF1 rs4132601 polymorphism and acute lymphoblastic leukemia susceptibility: a meta-analysis', *Leuk Lymphoma,* 56(4), 978-82.

Li, S., Wang, L., Fu, B., Berman, M. A., Diallo, A. and Dorf, M. E. (2014) 'TRIM65 regulates microRNA activity by ubiquitination of TNRC6', *Proc Natl Acad Sci U S A,* 111(19), 6970-5.

Li, S. Y., Ye, J. Y., Liang, E. Y., Zhou, L. X. and Yang, M. (2015) 'Association between MTHFR C677T polymorphism and risk of acute lymphoblastic leukemia: a meta-analysis based on 51 case-control studies', *Med Sci Monit,* 21, 740-8.

Li, Z., Zhang, W., Wu, M., Zhu, S., Gao, C., Sun, L., Zhang, R., Qiao, N., Xue, H., Hu, Y., Bao, S., Zheng, H. and Han, J. D. (2009) 'Gene expression-based classification and regulatory networks of pediatric acute lymphoblastic leukemia', *Blood,* 114(20), 4486-93.

Liang D, Meyer L, Chang DW, Lin J, Pu X, Ye Y, Gu J, Wu X, Lu K. (2010) Genetic variants in MicroRNA biosynthesis pathways and binding sites modify ovarian cancer risk, survival, and treatment response. Cancer Res. 2010 Dec 1;70(23):9765-76.

Liu J, Wei M, He Y, Liao B, Liao G, Li H (2012) Genetic variants in the microRNA machinery gene GEMIN4 are associated with risk of prostate cancer: a case-control study of the Chinese Han population. DNA Cell Biol.;31(7):1296-302.

Liu, Z., Li, G., Wei, S., Niu, J., El-Naggar, A. K., Sturgis, E. M. and Wei, Q. (2010) 'Genetic variants in selected pre-microRNA genes and the risk of squamous cell carcinoma of the head and neck', *Cancer,* 116(20), 4753-60.

Liu, X., Zheng, H. and Qu, C. K. (2012) 'Protein tyrosine phosphatase Shp2 (Ptpn11) plays an important role in maintenance of chromosome stability', *Cancer Res,* 72(20), 5296-306.

Lim, J. Y., Bhatia, S., Robison, L. L. and Yang, J. J. (2014) 'Genomics of racial and ethnic disparities in childhood acute lymphoblastic leukemia', *Cancer,* 120(7), 955-62.

Lin, C. Y., Li, M. J., Chang, J. G., Liu, S. C., Weng, T., Wu, K. H., Yang, S. F., Huang, F. K., Lo, W. Y. and Peng, C. T. (2014) 'High-resolution melting analyses for genetic variants in ARID5B and IKZF1 with childhood acute lymphoblastic leukemia susceptibility loci in Taiwan', *Blood Cells Mol Dis,* 52(2-3), 140-5.

Liu, D., Wu, D., Li, H. and Dong, M. (2014) 'The effect of XPD/ERCC2 Lys751Gln polymorphism on acute leukemia risk: a systematic review and meta-analysis', *Gene,* 538(2), 209-16.

Liu, S., An, J., Lin, J., Liu, Y., Bao, L., Zhang, W. and Zhao, J. J. (2014) 'Single nucleotide polymorphisms of microRNA processing machinery genes and outcome of hepatocellular carcinoma', *PLoS One,* 9(3), e92791.

Livak, K. J., Flood, S. J., Marmaro, J., Giusti, W. and Deetz, K. (1995) 'Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization.', *PCR Methods Appl,* 4(6), 357-62.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', *Methods,* 25(4), 402-8.

Lopez-Lopez, E., Gutierrez-Camino, A., Martin-Guerrero, I. and Garcia-Orad, A. (2013) 'Re: novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations', *J Natl Cancer Inst,* 105(19), 1512.

Lopez-Lopez, E., Martin-Guerrero, I., Ballesteros, J., Piñan, M. A., Garcia-Miguel, P., Navajas, A. and Garcia-Orad, A. (2011) 'Polymorphisms of the SLCO1B1 gene predict methotrexate-related toxicity in childhood acute lymphoblastic leukemia', *Pediatr Blood Cancer,* 57(4), 612-9.

M

Ma, H., Qazi, S., Ozer, Z., Gaynon, P., Reaman, G. H. and Uckun, F. M. (2012) 'CD22 Exon 12 deletion is a characteristic genetic defect of therapy-refractory clones in paediatric acute lymphoblastic leukaemia', *Br J Haematol,* 156(1), 89-98.

Marques, S. C., Laursen, M. B., Bødker, J. S., Kjeldsen, M. K., Falgreen, S., Schmitz, A., Bøgsted, M., Johnsen, H. E. and Dybkaer, K. (2015) 'MicroRNAs in B-cells: from normal differentiation to treatment of malignancies', *Oncotarget,* 6(1), 7-25.

Martin-Guerrero, I., Gutierrez-Camino, A., Lopez-Lopez, E., Bilbao-Aldaiturriaga, N., Pombar-Gomez, M., Ardanaz, M. and Garcia-Orad, A. (2015) 'Genetic variants in miRNA processing genes and pre-miRNAs are associated with the risk of chronic lymphocytic leukemia', *PLoS One,* 10(3), e0118905.

Masliah-Planchon, J., Garinet, S. and Pasmant, E. (2015) 'RAS-MAPK pathway epigenetic activation in cancer: miRNAs in action', *Oncotarget*.

Merritt, W. M., Bar-Eli, M. and Sood, A. K. (2010) 'The dicey role of Dicer: implications for RNAi therapy', *Cancer Res,* 70(7), 2571-4.

Meseure, D., Vacher, S., Drak Alsibai, K., Nicolas, A., Chemlali, W., Caly, M., Lidereau, R., Pasmant, E., Callens, C. and Bieche, I. (2016) 'Expression of ANRIL - Polycomb Complexes - CDKN2A/B/ARF Genes in Breast Tumors: Identification of a Two-gene (EZH2/CBX7) Signature with Independent Prognostic Value', *Mol Cancer Res*.

Migliorini, G., Fiege, B., Hosking, F. J., Ma, Y., Kumar, R., Sherborne, A. L., da Silva Filho, M. I., Vijayakrishnan, J., Koehler, R., Thomsen, H., Irving, J. A., Allan, J. M., Lightfoot, T., Roman, E., Kinsey, S. E., Sheridan, E., Thompson, P., Hoffmann, P., Nöthen, M. M., Mühleisen, T. W., Eisele, L., Zimmermann, M., Bartram, C. R., Schrappe, M., Greaves, M., Stanulla, M., Hemminki, K. and Houlston, R. S. (2013) 'Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype', *Blood,* 122(19), 3298-307.

Moriyama, T., Relling, M. V. and Yang, J. J. (2015) 'Inherited genetic variation in childhood acute lymphoblastic leukemia', *Blood,* 125(26), 3988-95.

Mullighan, C. G., Miller, C. B., Radtke, I., Phillips, L. A., Dalton, J., Ma, J., White, D., Hughes, T. P., Le Beau, M. M., Pui, C. H., Relling, M. V., Shurtleff, S. A. and Downing, J. R. (2008) 'BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros', *Nature,* 453(7191), 110-4.

Mullighan, C. G., Su, X., Zhang, J., Radtke, I., Phillips, L. A., Miller, C. B., Ma, J., Liu, W., Cheng, C., Schulman, B. A., Harvey, R. C., Chen, I. M., Clifford, R. J., Carroll, W. L., Reaman, G., Bowman, W. P., Devidas, M., Gerhard, D. S., Yang, W., Relling, M. V., Shurtleff, S. A., Campana, D., Borowitz, M. J., Pui, C. H., Smith, M., Hunger, S. P., Willman, C. L., Downing, J. R. and Group, C. s. O. (2009) 'Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia', *N Engl J Med,* 360(5), 470-80.

Musilova, K. and Mraz, M. (2015) 'MicroRNAs in B-cell lymphomas: how a complex biology gets more complex', *Leukemia,* 29(5), 1004-17.

O

O'Carroll, D., Mecklenbrauker, I., Das, P. P., Santana, A., Koenig, U., Enright, A. J., Miska, E. A. and Tarakhovsky, A. (2007) 'A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway', *Genes Dev,* 21(16), 1999-2004.

Orsi, L., Rudant, J., Bonaventure, A., Goujon-Bellec, S., Corda, E., Evans, T. J., Petit, A., Bertrand, Y., Nelken, B., Robert, A., Michel, G., Sirvent, N., Chastagner, P., Ducassou, S., Rialland, X., Hémon, D., Milne, E., Scott, R. J., Baruchel, A. and Clavel, J. (2012) 'Genetic polymorphisms and childhood acute lymphoblastic leukemia: GWAS of the ESCALE study (SFCE)', *Leukemia,* 26(12), 2561-4.

Okubo, M., Tahara, T., Shibata, T., Yamashita, H., Nakamura, M., Yoshioka, D., Yonemura, J., Ishizuka, T., Arisawa, T. and Hirata, I. (2010) 'Association between common genetic variants in pre-microRNAs and gastric cancer risk in Japanese population', *Helicobacter,* 15(6), 524-31.

P

Pal, S., Gupta, R. and Davuluri, R. V. (2012) 'Alternative transcription and alternative splicing in cancer', *Pharmacol Ther,* 136(3), 283-94.

Papaemmanuil, E., Hosking, F. J., Vijayakrishnan, J., Price, A., Olver, B., Sheridan, E., Kinsey, S. E., Lightfoot, T., Roman, E., Irving, J. A., Allan, J. M., Tomlinson, I. P., Taylor, M., Greaves, M. and

Houlston, R. S. (2009) 'Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia', *Nat Genet,* 41(9), 1006-10.

Pastorczak, A., Górniak, P., Sherborne, A., Hosking, F., Trelińska, J., Lejman, M., Szczepański, T., Borowiec, M., Fendler, W., Kowalczyk, J., Houlston, R. S. and Młynarski, W. (2011) 'Role of 657del5 NBN mutation and 7p12.2 (IKZF1), 9p21 (CDKN2A), 10q21.2 (ARID5B) and 14q11.2 (CEBPE) variation and risk of childhood ALL in the Polish population', *Leuk Res,* 35(11), 1534-6.

Paulsson, K. (2016) 'High hyperdiploid childhood acute lymphoblastic leukemia: Chromosomal gains as the main driver event', *Mol Cell Oncol,* 3(1), e1064555.

Paulsson, K., Forestier, E., Lilljebjörn, H., Heldrup, J., Behrendtz, M., Young, B. D. and Johansson, B. (2010) 'Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia', *Proc Natl Acad Sci U S A,* 107(50), 21719-24.

Perry, G. H., Ben-Dor, A., Tsalenko, A., Sampas, N., Rodriguez-Revenga, L., Tran, C. W., Scheffer, A., Steinfeld, I., Tsang, P., Yamada, N. A., Park, H. S., Kim, J. I., Seo, J. S., Yakhini, Z., Laderman, S., Bruhn, L. and Lee, C. (2008) 'The fine-scale and complex architecture of human copy-number variation', *Am J Hum Genet,* 82(3), 685-95.

Prasad, R. B., Hosking, F. J., Vijayakrishnan, J., Papaemmanuil, E., Koehler, R., Greaves, M., Sheridan, E., Gast, A., Kinsey, S. E., Lightfoot, T., Roman, E., Taylor, M., Pritchard-Jones, K., Stanulla, M., Schrappe, M., Bartram, C. R., Houlston, R. S., Kumar, R. and Hemminki, K. (2010) 'Verification of the susceptibility loci on 7p12.2, 10q21.2, and 14q11.2 in precursor B-cell acute lymphoblastic leukemia of childhood', *Blood,* 115(9), 1765-7.

Pui, C. H. and Evans, W. E. (2006) 'Treatment of acute lymphoblastic leukemia', *N Engl J Med,* 354(2), 166-78.

Pui, C. H., Robison, L. L. and Look, A. T. (2008) 'Acute lymphoblastic leukaemia', *Lancet,* 371(9617), 1030-43.

Pui, C. H., Sandlund, J. T., Pei, D., Campana, D., Rivera, G. K., Ribeiro, R. C., Rubnitz, J. E., Razzouk, B. I., Howard, S. C., Hudson, M. M., Cheng, C., Kun, L. E., Raimondi, S. C., Behm, F. G., Downing, J. R., Relling, M. V., Evans, W. E. and Hospital, T. T. S. X. a. S. J. C. s. R. (2004) 'Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIB at St Jude Children's Research Hospital', *Blood,* 104(9), 2690-6.

Q

Qiu, M. T., Hu, J. W., Ding, X. X., Yang, X., Zhang, Z., Yin, R. and Xu, L. (2012) 'Hsa-miR-499 rs3746444 polymorphism contributes to cancer risk: a meta-analysis of 12 studies', *PLoS One,* 7(12), e50887.

R

Rao, D. S., O'Connell, R. M., Chaudhuri, A. A., Garcia-Flores, Y., Geiger, T. L. and Baltimore, D. (2010) 'MicroRNA-34a perturbs B lymphocyte development by repressing the forkhead box transcription factor Foxp1', *Immunity,* 33(1), 48-59.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R. and Lander, E. S. (2001) 'Linkage disequilibrium in the human genome', *Nature,* 411(6834), 199-204.

Risch, N. and Merikangas, K. (1996) 'The future of genetic studies of complex human diseases', *Science,* 273(5281), 1516-7.

Ross, J. A., Linabery, A. M., Blommer, C. N., Langer, E. K., Spector, L. G., Hilden, J. M., Heerema, N. A., Radloff, G. A., Tower, R. L. and Davies, S. M. (2013) 'Genetic variants modify susceptibility to leukemia in infants: a Children's Oncology Group report', *Pediatr Blood Cancer,* 60(1), 31-4.

Rotunno, M., Zhao, Y., Bergen, A. W., Koshiol, J., Burdette, L., Rubagotti, M., Linnoila, R. I., Marincola, F. M., Bertazzi, P. A., Pesatori, A. C., Caporaso, N. E., McShane, L. M., Wang, E. and Landi, M. T. (2010) 'Inherited polymorphisms in the RNA-mediated interference machinery affect microRNA expression and lung cancer survival', *Br J Cancer,* 103(12), 1870-4.

Roy R, De Sarkar N, Ghose S, Paul RR, Pal M, Bhattacharya C, Chowdhury SK, Ghosh S, Roy B. (2014) Genetic variations at microRNA and processing genes and risk of oral cancer. Tumour Biol. 35(4):3409-14

Ryan, B. M., Robles, A. I. and Harris, C. C. (2010) 'Genetic variation in microRNA networks: the implications for cancer research', *Nat Rev Cancer,* 10(6), 389-402.

S

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., Hunt, S. E., Cole, C. G., Coggill, P. C., Rice, C. M., Ning, Z., Rogers, J., Bentley, D. R., Kwok, P. Y., Mardis, E. R., Yeh, R. T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R. H., McPherson, J. D., Gilman, B., Schaffner, S., Van Etten, W. J., Reich, D., Higgins, J., Daly, M. J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M. C., Linton, L., Lander, E. S., Altshuler, D. and Group, I. S. M. W. (2001) 'A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms', *Nature,* 409(6822), 928-33.

Salzman, D. W. and Weidhaas, J. B. (2013) 'SNPing cancer in the bud: microRNA and microRNA-target site polymorphisms as diagnostic and prognostic biomarkers in cancer', *Pharmacol Ther,* 137(1), 55-63.

Sambrook, J. and Russell, D. W. (2001) 'Preparation and Analysis of Eukaryotic Genomic DNA' in *Molecular cloning: A laboratory manual*, 3rd edition ed., Cold Spring Harbor, USA: CSHL Press, 4-12.

Sana, J., Faltejskova, P., Svoboda, M. and Slaby, O. (2012) 'Novel classes of non-coding RNAs and cancer', *J Transl Med,* 10, 103.

Samuelsson, J., Alonso, S., Ruiz-Larroya, T., Cheung, T. H., Wong, Y. F. and Perucho, M. (2011) 'Frequent somatic demethylation of RAPGEF1/C3G intronic sequences in gastrointestinal and gynecological cancer', *Int J Oncol,* 38(6), 1575-7.

Schotte, D., Akbari Moqadam, F., Lange-Turenhout, E. A., Chen, C., van Ijcken, W. F., Pieters, R. and den Boer, M. L. (2011a) 'Discovery of new microRNAs by small RNAome deep sequencing in childhood acute lymphoblastic leukemia', *Leukemia,* 25(9), 1389-99.

Schotte, D., Chau, J. C., Sylvester, G., Liu, G., Chen, C., van der Velden, V. H., Broekhuis, M. J., Peters, T. C., Pieters, R. and den Boer, M. L. (2009) 'Identification of new microRNA genes and aberrant microRNA profiles in childhood acute lymphoblastic leukemia', *Leukemia,* 23(2), 313-22.

Schotte, D., De Menezes, R. X., Akbari Moqadam, F., Khankahdani, L. M., Lange-Turenhout, E., Chen, C., Pieters, R. and Den Boer, M. L. (2011b) 'MicroRNA characterize genetic diversity and drug resistance in pediatric acute lymphoblastic leukemia', *Haematologica,* 96(5), 703-11.

Schüz, J. (2011) 'Exposure to extremely low-frequency magnetic fields and the risk of childhood cancer: update of the epidemiological evidence', *Prog Biophys Mol Biol,* 107(3), 339-42.

Sherborne, A. L., Hemminki, K., Kumar, R., Bartram, C. R., Stanulla, M., Schrappe, M., Petridou, E., Semsei, A. F., Szalai, C., Sinnett, D., Krajinovic, M., Healy, J., Lanciotti, M., Dufour, C., Indaco, S., El-Ghouroury, E. A., Sawangpanich, R., Hongeng, S., Pakakasama, S., Gonzalez-Neira, A., Ugarte, E. L., Leal, V. P., Espinoza, J. P., Kamel, A. M., Ebid, G. T., Radwan, E. R., Yalin, S., Yalin, E., Berkoz, M., Simpson, J., Roman, E., Lightfoot, T., Hosking, F. J., Vijayakrishnan, J., Greaves, M. and Houlston, R. S. (2011) 'Rationale for an international consortium to study inherited genetic susceptibility to childhood acute lymphoblastic leukemia', *Haematologica,* 96(7), 1049-54.

Sherborne, A. L., Hosking, F. J., Prasad, R. B., Kumar, R., Koehler, R., Vijayakrishnan, J., Papaemmanuil, E., Bartram, C. R., Stanulla, M., Schrappe, M., Gast, A., Dobbins, S. E., Ma, Y., Sheridan, E., Taylor, M., Kinsey, S. E., Lightfoot, T., Roman, E., Irving, J. A., Allan, J. M., Moorman, A. V., Harrison, C. J., Tomlinson, I. P., Richards, S., Zimmermann, M., Szalai, C., Semsei, A. F., Erdelyi, D. J., Krajinovic, M., Sinnett, D., Healy, J., Gonzalez Neira, A., Kawamata, N., Ogawa, S., Koeffler, H. P., Hemminki, K., Greaves, M. and Houlston, R. S. (2010) 'Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk', *Nat Genet,* 42(6), 492-4.

Sherborne, A. L. and Houlston, R. S. (2010) 'What are genome-wide association studies telling us about B-cell tumor development?', *Oncotarget,* 1(5), 367-72.

Silverman, L. B., Declerck, L., Gelber, R. D., Dalton, V. K., Asselin, B. L., Barr, R. D., Clavell, L. A., Hurwitz, C. A., Moghrabi, A., Samson, Y., Schorin, M. A., Lipton, J. M., Cohen, H. J. and Sallan, S. E. (2000)

'Results of Dana-Farber Cancer Institute Consortium protocols for children with newly diagnosed acute lymphoblastic leukemia (1981-1995)', *Leukemia,* 14(12), 2247-56.

Song, J. J., Liu, J., Tolia, N. H., Schneiderman, J., Smith, S. K., Martienssen, R. A., Hannon, G. J. and Joshua-Tor, L. (2003) 'The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes', *Nat Struct Biol,* 10(12), 1026-32.

Sun, J., Zheng, S. L., Wiklund, F., Isaacs, S. D., Li, G., Wiley, K. E., Kim, S. T., Zhu, Y., Zhang, Z., Hsu, F. C., Turner, A. R., Stattin, P., Liu, W., Kim, J. W., Duggan, D., Carpten, J., Isaacs, W., Grönberg, H., Xu, J. and Chang, B. L. (2009) 'Sequence variants at 22q13 are associated with prostate cancer risk', *Cancer Res,* 69(1), 10-5.

Sung, H., Lee, K. M., Choi, J. Y., Han, S., Lee, J. Y., Li, L., Park, S. K., Yoo, K. Y., Noh, D. Y., Ahn, S. H. and Kang, D. (2011) 'Common genetic polymorphisms of microRNA biogenesis pathway genes and risk of breast cancer: a case-control study in Korea', *Breast Cancer Res Treat,* 130(3), 939-51.

Srivastava, K. and Srivastava, A. (2012) 'Comprehensive review of genetic association studies and meta-analyses on miRNA polymorphisms and cancer risk', *PLoS One,* 7(11), e50966.

Swerdlov, S., Campo, E. and Harris, N. (2008) 'WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues',

T

Tao, S., Wang, Z., Feng, J., Hsu, F. C., Jin, G., Kim, S. T., Zhang, Z., Gronberg, H., Zheng, L. S., Isaacs, W. B., Xu, J. and Sun, J. (2012) 'A genome-wide search for loci interacting with known prostate cancer risk-associated genetic variants', *Carcinogenesis,* 33(3), 598-603.

Tian, T., Shu, Y., Chen, J., Hu, Z., Xu, L., Jin, G., Liang, J., Liu, P., Zhou, X., Miao, R., Ma, H., Chen, Y. and Shen, H. (2009) 'A functional genetic variant in microRNA-196a2 is associated with increased susceptibility of lung cancer in Chinese', *Cancer Epidemiol Biomarkers Prev,* 18(4), 1183-7.

Tong, N., Xu, B., Shi, D., Du, M., Li, X., Sheng, X., Wang, M., Chu, H., Fang, Y., Li, J., Wu, D. and Zhang, Z. (2014) 'Hsa-miR-196a2 polymorphism increases the risk of acute lymphoblastic leukemia in Chinese children', *Mutat Res,* 759, 16-21.

Tong N, Chu H, Wang M, Xue Y, Du M, Lu L, Zhang H, Wang F, Fang Y, Li J, Wu D, Zhang Z, Sheng X. (2016). Pri-miR-34b/c rs4938723 polymorphism contributes to acute lymphoblastic leukemia susceptibility in Chinese children. Leuk Lymphoma. 2016;57(6):1436-41.

Treviño, L. R., Yang, W., French, D., Hunger, S. P., Carroll, W. L., Devidas, M., Willman, C., Neale, G., Downing, J., Raimondi, S. C., Pui, C. H., Evans, W. E. and Relling, M. V. (2009) 'Germline genomic variants associated with childhood acute lymphoblastic leukemia', *Nat Genet,* 41(9), 1001-5.

U

Uckun, F. M., Goodman, P., Ma, H., Dibirdik, I. and Qazi, S. (2010) 'CD22 EXON 12 deletion as a pathogenic mechanism of human B-precursor leukemia', *Proc Natl Acad Sci U S A,* 107(39), 16852-7.

Urayama, K. Y., Chokkalingam, A. P., Manabe, A. and Mizutani, S. (2013) 'Current evidence for an inherited genetic basis of childhood acute lymphoblastic leukemia', *Int J Hematol,* 97(1), 3-19.

Urayama, K. Y., Chokkalingam, A. P., Metayer, C., Ma, X., Selvin, S., Barcellos, L. F., Wiemels, J. L., Wiencke, J. K., Taylor, M., Brennan, P., Dahl, G. V., Moonsamy, P., Erlich, H. A., Trachtenberg, E. and Buffler, P. A. (2012) 'HLA-DP genetic variation, proxies for early life immune modulation and childhood acute lymphoblastic leukemia risk', *Blood,* 120(15), 3039-47.

Urayama, K. Y., Ma, X. and Buffler, P. A. (2008) 'Exposure to infections through day-care attendance and risk of childhood leukaemia', *Radiat Prot Dosimetry,* 132(2), 259-66.

V

Vardiman, J. W., Thiele, J., Arber, D. A., Brunning, R. D., Borowitz, M. J., Porwit, A., Harris, N. L., Le Beau, M. M., Hellström-Lindberg, E., Tefferi, A. and Bloomfield, C. D. (2009) 'The 2008 revision of the

World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes', *Blood,* 114(5), 937-51.

Ventura, A., Young, A. G., Winslow, M. M., Lintault, L., Meissner, A., Erkeland, S. J., Newman, J., Bronson, R. T., Crowley, D., Stone, J. R., Jaenisch, R., Sharp, P. A. and Jacks, T. (2008) 'Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters', *Cell,* 132(5), 875-86.

Vijayakrishnan, J., Henrion, M., Moorman, A. V., Fiege, B., Kumar, R., da Silva Filho, M. I., Holroyd, A., Koehler, R., Thomsen, H., Irving, J. A., Allan, J. M., Lightfoot, T., Roman, E., Kinsey, S. E., Sheridan, E., Thompson, P. D., Hoffmann, P., Nöthen, M. M., Mühleisen, T. W., Eisele, L., Bartram, C. R., Schrappe, M., Greaves, M., Hemminki, K., Harrison, C. J., Stanulla, M. and Houlston, R. S. (2015) 'The 9p21.3 risk of childhood acute lymphoblastic leukaemia is explained by a rare high-impact variant in CDKN2A', *Sci Rep,* 5, 15065.

Vijayakrishnan, J. and Houlston, R. S. (2010) 'Candidate gene association studies and risk of childhood acute lymphoblastic leukemia: a systematic review and meta-analysis', *Haematologica,* 95(8), 1405-14.

Vijayakrishnan, J., Sherborne, A. L., Sawangpanich, R., Hongeng, S., Houlston, R. S. and Pakakasama, S. (2010) 'Variation at 7p12.2 and 10q21.2 influences childhood acute lymphoblastic leukemia risk in the Thai population and may contribute to racial differences in leukemia incidence', *Leuk Lymphoma,* 51(10), 1870-4.

Vinci, S., Gelmini, S., Mancini, I., Malentacchi, F., Pazzagli, M., Beltrami, C., Pinzani, P. and Orlando, C. (2013) 'Genetic and epigenetic factors in regulation of microRNA in colorectal cancers', *Methods,* 59(1), 138-46.

W

Walsh, K. M., Chokkalingam, A. P., Hsu, L. I., Metayer, C., de Smith, A. J., Jacobs, D. I., Dahl, G. V., Loh, M. L., Smirnov, I. V., Bartley, K., Ma, X., Wiencke, J. K., Barcellos, L. F., Wiemels, J. L. and Buffler, P. A. (2013) 'Associations between genome-wide Native American ancestry, known risk alleles and B-cell ALL risk in Hispanic children', *Leukemia,* 27(12), 2416-9.

Walsh, K. M., de Smith, A. J., Chokkalingam, A. P., Metayer, C., Dahl, G. V., Hsu, L. I., Barcellos, L. F., Wiemels, J. L. and Buffler, P. A. (2013) 'Novel childhood ALL susceptibility locus BMI1-PIP4K2A is specifically associated with the hyperdiploid subtype', *Blood,* 121(23), 4808-9.

Walsh, K. M., de Smith, A. J., Hansen, H. M., Smirnov, I. V., Gonseth, S., Endicott, A. A., Xiao, J., Rice, T., Fu, C. H., McCoy, L. S., Lachance, D. H., Eckel-Passow, J. E., Wiencke, J. K., Jenkins, R. B., Wrensch, M. R., Ma, X., Metayer, C. and Wiemels, J. L. (2015) 'A Heritable Missense Polymorphism in CDKN2A Confers Strong Risk of Childhood Acute Lymphoblastic Leukemia and Is Preferentially Selected during Clonal Evolution', *Cancer Res,* 75(22), 4884-94.

Wang, C., Chen, J., Sun, H., Sun, L. and Liu, Y. (2015) 'CEBPE polymorphism confers an increased risk of childhood acute lymphoblastic leukemia: a meta-analysis of 11 case-control studies with 5,639 cases and 10,036 controls', *Ann Hematol,* 94(2), 181-5.

Wang, F., Zhao, Q., He, H. R., Zhai, Y. J., Lu, J., Hu, H. B., Zhou, J. S., Yang, Y. H. and Li, Y. J. (2015) 'The association between XRCC1 Arg399Gln polymorphism and risk of leukemia in different populations: a meta-analysis of case-control studies', *Onco Targets Ther,* 8, 3277-87.

Wang, Y., Chen, J., Li, J., Deng, J., Rui, Y., Lu, Q., Wang, M., Tong, N., Zhang, Z. and Fang, Y. (2013) 'Association of three polymorphisms in ARID5B, IKZF1 and CEBPE with the risk of childhood acute lymphoblastic leukemia in a Chinese population', *Gene,* 524(2), 203-7.

Wartenberg , D., Groves, FD. and Adelmand, AS. (2008) 'Acute Lymphoblastic leukemia:Epidemiology and Etiology' in *Hematologic Malignancies: Acute Leukemias.* Springer, Berlin.

Wiemels, J. L., de Smith, A. J., Xiao, J., Lee, S. T., Muench, M. O., Fomin, M. E., Zhou, M., Hansen, H. M., Termuhlen, A., Metayer, C. and Walsh, K. M. (2016) 'A functional polymorphism in the CEBPE gene promoter influences acute lymphoblastic leukemia risk through interaction with the hematopoietic transcription factor Ikaros', *Leukemia,* 30(5), 1194-7.

X

Xia, L., Ren, Y., Fang, X., Yin, Z., Li, X., Wu, W., Guan, P. and Zhou, B. (2014) 'Prognostic role of common microRNA polymorphisms in cancers: evidence from a meta-analysis', *PLoS One,* 9(10), e106799.

Xiang, Y., Fan, S., Cao, J., Huang, S. and Zhang, L. P. (2012) 'Association of the microRNA-499 variants with susceptibility to hepatocellular carcinoma in a Chinese population', *Mol Biol Rep,* 39(6), 7019-23.

Xu, H., Cheng, C., Devidas, M., Pei, D., Fan, Y., Yang, W., Neale, G., Scheet, P., Burchard, E. G., Torgerson, D. G., Eng, C., Dean, M., Antillon, F., Winick, N. J., Martin, P. L., Willman, C. L., Camitta, B. M., Reaman, G. H., Carroll, W. L., Loh, M., Evans, W. E., Pui, C. H., Hunger, S. P., Relling, M. V. and Yang, J. J. (2012) 'ARID5B genetic polymorphisms contribute to racial disparities in the incidence and treatment outcome of childhood acute lymphoblastic leukemia', *J Clin Oncol,* 30(7), 751-7.

Xu, H., Yang, W., Perez-Andreu, V., Devidas, M., Fan, Y., Cheng, C., Pei, D., Scheet, P., Burchard, E. G., Eng, C., Huntsman, S., Torgerson, D. G., Dean, M., Winick, N. J., Martin, P. L., Camitta, B. M., Bowman, W. P., Willman, C. L., Carroll, W. L., Mullighan, C. G., Bhojwani, D., Hunger, S. P., Pui, C. H., Evans, W. E., Relling, M. V., Loh, M. L. and Yang, J. J. (2013) 'Novel Susceptibility Variants at 10p12.31-12.2 for Childhood Acute Lymphoblastic Leukemia in Ethnically Diverse Populations', *J Natl Cancer Inst*.

Xu, H., Zhang, H., Yang, W., Yadav, R., Morrison, A. C., Qian, M., Devidas, M., Liu, Y., Perez-Andreu, V., Zhao, X., Gastier-Foster, J. M., Lupo, P. J., Neale, G., Raetz, E., Larsen, E., Bowman, W. P., Carroll, W. L., Winick, N., Williams, R., Hansen, T., Holm, J. C., Mardis, E., Fulton, R., Pui, C. H., Zhang, J., Mullighan, C. G., Evans, W. E., Hunger, S. P., Gupta, R., Schmiegelow, K., Loh, M. L., Relling, M. V. and Yang, J. J. (2015) 'Inherited coding variants at the CDKN2A locus influence susceptibility to acute lymphoblastic leukaemia in children', *Nat Commun,* 6, 7553.

Y

Yamanaka, R., Barlow, C., Lekstrom-Himes, J., Castilla, L. H., Liu, P. P., Eckhaus, M., Decker, T., Wynshaw-Boris, A. and Xanthopoulos, K. G. (1997) 'Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein epsilon-deficient mice', *Proc Natl Acad Sci U S A,* 94(24), 13187-92.

Yang, H., Dinney, C. P., Ye, Y., Zhu, Y., Grossman, H. B. and Wu, X. (2008) 'Evaluation of genetic variants in microRNA-related genes and risk of bladder cancer', *Cancer Res,* 68(7), 2530-7.

Yang, W., Treviño, L. R., Yang, J. J., Scheet, P., Pui, C. H., Evans, W. E. and Relling, M. V. (2010) 'ARID5B SNP rs10821936 is associated with risk of childhood acute lymphoblastic leukemia in blacks and contributes to racial differences in leukemia incidence', *Leukemia,* 24(4), 894-6.

Z

Zhang, H., Yang, J. H., Zheng, Y. S., Zhang, P., Chen, X., Wu, J., Xu, L., Luo, X. Q., Ke, Z. Y., Zhou, H., Qu, L. H. and Chen, Y. Q. (2009) 'Genome-wide analysis of small RNA and novel MicroRNA discovery in human acute lymphoblastic leukemia based on extensive sequencing approach', *PLoS One,* 4(9), e6849.

Zhang, H., Zhang, Z. and Li, G. (2015) 'ABCB1 polymorphism and susceptibility to acute lymphoblastic leukemia: a meta analysis', *Int J Clin Exp Med,* 8(5), 7585-91.

Zhang, X., Cairns, M., Rose, B., O'Brien, C., Shannon, K., Clark, J., Gamble, J. and Tran, N. (2009) 'Alterations in miRNA processing and expression in pleomorphic adenomas of the salivary gland', *Int J Cancer,* 124(12), 2855-63.

Zhou, B., Wang, K., Wang, Y., Xi, M., Zhang, Z., Song, Y. and Zhang, L. (2011) 'Common genetic polymorphisms in pre-microRNAs and risk of cervical squamous cell carcinoma', *Mol Carcinog,* 50(7), 499-505.