EXPRESSION AND ASSOCIATION DATA STRONGLY SUPPORT JARID2 INVOLVEMENT IN NONSINDROMIC CLEFT LIP WITH OR WITHOUT CLEFT PALATE

Luca Scapoli, Marcella Martinelli, Furio Pezzetti, Annalisa Palmieri, Ambra Girardi, Anna Savoia, Anna Monica Bianco, Francesco Carinci

Human Mutation

EXPRESSION AND ASSOCIATION DATA STRONGLY SUPPORT JARID2 INVOLVEMENT IN NONSINDROMIC CLEFT LIP WITH OR WITHOUT CLEFT PALATE

Journal:	*Human Mutation*
Manuscript ID:	humu-2010-0074.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	29-Mar-2010
Complete List of Authors:	SCAPOLI, LUCA; University of Bologna, Histology, Embryology and Applied Biology Martinelli, Marcella; University of Bologna, Department of Histology, Embryology and applied Biology Pezzetti, Furio; University of Bologna, Department of Histology, Embryology and applied Biology Palmieri, Annalisa; University of Ferrara, Department of D.M.C.C.C., Section of Maxillo-Facial Surgery Girardi, Ambra; University of Bologna, Department of Histology, Embryology and applied Biology Savoia, Anna; University of Trieste Bianco, Anna; University of Trieste Carinci, Francesco; University of Ferrara, Department of D.M.C.C.C., Section of Maxillo-Facial Surgery
Key Words:	Cleft lip, Association analysis, JARID2, jumonji

EXPRESSION AND ASSOCIATION DATA STRONGLY SUPPORT *JARID2* INVOLVEMENT IN NONSYNDROMIC CLEFT LIP WITH OR WITHOUT CLEFT PALATE

Luca Scapoli[1]*, Marcella Martinelli[1], Furio Pezzetti[1], Annalisa Palmieri[2], Ambra Girardi[1], Anna Savoia[3], Anna Monica Bianco[3], Francesco Carinci[2]

[1]Department of Histology, Embryology and Applied Biology, Centre of Molecular Genetics, University of Bologna, Via Belmeloro, 8 - 40126 Bologna, Italy.

[2] Department of D.M.C.C.C., Section of Maxillo-Facial Surgery, University of Ferrara, Corso Giovecca, 203 - 44100 Ferrara, Italy.

[3] Medical Genetics, Department of Reproductive and Developmental Sciences, IRCCS Burlo Garofolo Hospital, University of Trieste, Via dell'Istria, 65/1 - 34137 Trieste, Italy

*Corresponding author: Luca Scapoli, Ph.D.
Department of Histology, Embryology and Applied Biology
University of Bologna
Via Belmeloro, 8 - 40126 Bologna, Italy
Phone +39-051-2094100 Fax. +39-051-2094110
E-mail: luca.scapoli2@unibo.it

Manuscript information:
number of figures 3;
number of tables 2.

Abbreviations: CL/P, Nonsyndromic cleft lip with or without cleft palate.

Abstract

Nonsyndromic cleft lip with or without cleft palate (CL/P) affects approximately 1 in 1000 births. Genetic studies have provided evidence for the role of several genes and candidate loci in clefting; however conflicting results have been frequently obtained and much have to be done to unravel the complex genetic of CL/P. In the present investigation we have focused on the candidate region in 6p23, a region that have been found linked to CL/P in several investigations, in the attempt to find out the susceptibility gene provisionally named OFC1. Gene expression experiments in mice embryo of positional candidate genes revealed that *JARID2* was highly and specifically expressed in epithelial cells in merging palatal shelves. A family based linkage disequilibrium study confirmed the pivotal role of *JARID2* in orofacial development and strongly supports a role for this gene in CL/P etiology (multiallelic haplotype test $P = 6x10^{-5}$). Understanding the molecular role of *JARID2* within facial development may offer additional information to further unravel the complex genetics of CL/P.

Keywords

Cleft lip; Cleft palate; *JARID2*; Jumonji; Linkage Disequilibrium; Association; Linkage; 6p23; Complex Disease**.**

Introduction

Orofacial clefts are common congenital malformations that may occur as features in more than 300 Mendelian, chromosomal, or teratogenic syndromes. Most frequently, clefting occurs as a unique defect in the so called isolated or "nonsyndromic condition". There is evidence of a strong genetic predisposition to CL/P, with familial aggregation in about 20% of cases, elevated concordance in monozygotic twins, and a high recurrence risk in relatives. Complex segregation analyses have excluded a simple Mendelian mode of inheritance. Different models have been proposed, ranging from a major gene influence to oligogenic and multifactor models (Scapoli et al., 1999; Schliekelman and Slatkin, 2002). The idea that CL/P results from a complex interaction between genes and environmental factors is widely accepted.

Genetic studies have collected data suggesting a role for several genes and candidate loci in clefting (Carinci et al., 2007); to date, ten different genetic loci predisposing to CL/P, known as *OFC1* to *OFC10*, have been catalogued in the Online Mendelian Inheritance in Man database. *OFC1* (OMIM# 119530), which maps onto chromosome 6p, has been one of the most highly investigated loci since the publication of the pioneering linkage investigation by Eiberg and colleagues (Eiberg et al., 1987). These authors performed a low density genome wide linkage study using 42 protein polymorphisms and obtained a suggestive linkage with the blood clotting factor XIIIA (*F13A1*, OMIM# 134570). Although several studies have failed to replicate this finding, substantial data supporting the existence of a CL/P susceptibility gene in this region has been collected. Linkage to genetic markers was observed both by our group and those of others' (Scapoli et al., 1997; Prescott et al., 2000; Carreno et al., 2002; Moreno et al., 2004; Schultz et al., 2004). We observed linkage and locus heterogeneity for 60% of families associated with D6S259 (Scapoli et al., 1997). Cytogenetic abnormalities of short arm of chromosome 6 have frequently been observed in both syndromic and non syndromic cases, thus supporting the presence of a clefting gene in this region (Topping et al., 2002; Davies et

al., 2004). In this report we provide both genetic and functional data to support the role of *JARID2* (OMIM# 601594) in isolated CL/P adding a gene of major effect to the etiology of this complex trait.

Materials and Methods

Expression study

Seven genes mapping in the LOD score peak area of a previous linkage analysis were considered CL/P candidate genes (Scapoli et al., 1997). The genes from telomere to centromer were: sirtuin 5 (*SIRT5,* GenBank NM_012241.3), nucleolar protein 7 (*NOL7,* GenBank NM_016167.3), RAN binding protein 9 (*RANBP9,* GenBank NM_005493.2), coiled-coil domain containing 90A (*CCDC90A,* GenBank NM_001031713.2), ring finger protein 182 (*RNF182,* GenBank NM_001165034.1), CD83 molecule (*CD83,* GenBank NM_004233.3), and jumonji, AT rich interactive domain 2 (*JARID2,* GenBank NM_004973.2). Gene expression of mice hortologous was investigated by RT-PCR and RNA in situ hybridization.

RNA extracted from mouse palate shelves at embryonic day 14.5 was retrotranscribed and used for RT-PCR, while cDNA obtained from several different tissues was used to obtain templates for in situ hybridization probes production (primers sequences in Supp. Table S1). PCR products of candidate genes, obtained with specific primers having tails with for the T3 and T7 RNA polymerase promoter sequences, were subcloned into the pCRII-TOPO vector (TOPO TA Cloning kit; Invitrogen). The plasmid were linearized with the enzymes *Not*I or *BamH*I and transcribed with T7 and T3 RNA polymerase, respectively to obtain antisense and sense probes. Mouse embryos at E14.5 and E15.5 were harvested from C57BL/6 pregnant females and their heads were used for *in situ* RNA hybridization. Coronal sections were cryoprotected by treatment with 30% sucrose in PBS and embedded in OCT. Twenty-micrometer coronal cryosections of embryo heads were collected on superfrost plus slides and postfixed with 4% paraformaldehyde in PBS for 15 min. For *in situ* hybridization, sections were analyzed as previously described (Marigo et al., 2004). As positive controls we used adjacent sections for the hybridization of the two *Myh9* and *Tgfb3* genes because they play a critical role in palate development (Martinelli et al., 2007). Slides were coverslipped with 70%

glycerol in PBS and photographed (AxioCam digital camera, Zeiss) using a microscope with Nomarski optics (Axioplan; Zeiss).

Sets of families

A sample of 218 unrelated Italian patients with the sole clinical feature of CL/P and their parents was used in this study. 131 cases were considered sporadic or non-familial, as no other relatives manifested the malformation, whereas, 87 cases were considered familial because of a recurrence of at least one case of CL/P within each pedigree. In order to classify the CL/P as non-syndromic and exclude potential teratogenic influences, a careful clinical exam of the patient and a detailed anamnesis was carried out to evaluate the presence of any other somatic or neurological disorder in the family, and the use of known or suspected clefting substances, such as phenytoin, warfarin, and ethanol, during pregnancy. After informed consent, DNA was extracted from peripheral blood samples.

Markers

Initially, nine SNPs within the *JARID2* locus were selected among validated-assays using the Applied Biosystems SNPbrowser™ Software. Selection was done considering exon distribution, i.e. lower inter-marker distance where exons were closer to each other, and preference was accorded to SNPs with low inter-marker linkage disequilibrium and minor allele frequency higher than .2. Later, after a positive association was found, additional polymorphic sites were added to the list. Two SNPs were selected in the *DTNBP1* gene because it is situated downstream very close to *JARID2*, in addition to two insertion/deletion polymorphisms which were detected in mutation screening. Genotypes of all the SNPs were obtained using an ABI PRISM 7500 Sequence Detection System and TaqMan chemistry according to supplier protocols, while the insertions/deletions were typed as a length polymorphisms by acrylamide gel electrophoresis of PCR products. Table 1 reports a summary of marker information.

Statistical analysis

Allelic association was conducted with family-based association tests implemented in FBAT program v1.7.3 (Horvath et al., 2001). The additive genetic model was applied for each allelic association test, which examines the transmission of the marker alleles from parents to the affected offspring. In this study, for single-SNP test, FBAT multi-marker test (FBAT-MM) was used to deal with multiple comparisons. The FBAT-MM simultaneously tests H0: no linkage or association between any marker and any disease susceptibility locus underlying the trait. This method was developed to test markers in linkage disequilibrium, when a simple Bonferroni-correction for multiple comparisons may be overly conservative to adjust for the complicated multiple testing. Subsequently, each marker contribution was tested with the null hypothesis of no linkage and no association between the markers and the underlying causal locus. The Monte Carlo procedure with 100,000 permutations was performed to calculate the empirical P values for the single-SNP association.

The Haploview program was used to check for Hardy-Weinberg equilibrium and to examine linkage disequilibrium block structure of the *JARID2* gene (Barrett et al., 2005). The D' values for all pairs of SNPs were calculated and the haplotype blocks were estimated using the solid spine method which requires the first and last markers to be in a strong linkage disequilibrium (D'>.79) with all intermediate markers, but does not necessarily require the intermediate markers to be in linkage disequilibrium to each other.

The haplotype version of FBAT (HBAT) was applied to test the association between phenotype and haplotypes (Horvath et al., 2004). The mode "a" command was used to obtain both biallelic and multiallelic tests. For the haplotype analyses the Monte Carlo procedure was performed with 1,000,000 permutations.

Point estimates and asymptotic confidence intervals of genotype relative risks for heterozygous and homozygous allele carriers, as well as attributable risk were calculated with

case-parent data using the formulae proposed by Scherag and colleagues (Scherag et al., 2002). The method is based on the likelihood theory and assumes Hardy-Weinberg equilibrium and absence of population stratification.

Sequencing

Mutation searches were performed by direct sequencing of PCR products. Primer pairs were designed in the intron sequences flanking for each one of the 18 exons, in order to screen coding sequences and splicing signals (primers sequences in Supp. Table S1). In addition, 1500 bp upstream of the 5'UTR were sequenced as part of a putative promoter region. Amplimers were checked for quality in agarose gel and sent for purification and bi-directional DNA sequencing service to Macrogen (Seul, Korea). Sequences were retrieved, aligned, analyzed, and reviewed using the Sequencer 4.6 program (Gene Codes, Ann Arbor, MI, USA).

Transient expression of hybrid minigene in HeLa cells

JARID2 PCR products of 448 bp carrying the "C" or "G" allele at the rs2076056 were obtained from genomic DNA using specific primers 6F and 6R modified at their 5' end with a tail specific for restriction enzyme site NdeI (primers sequences in Supp. Table S1). The two allelic variants were inserted into the unique NdeI site of the expression vector pTBNde(min) kindly donated by Dr Franco Pagani (Pagani et al., 2000; Pagani et al., 2003). The constructs were transfected in HeLa cells and RNA was extracted after 24 ours. RT-PCR was performed with primers 2-3alpha and B2 (primers sequences in Supp. Table S1) that are expected to generate a product of 475 bp when exon 6 is correctly spliced. RT-PCR products were directly sequenced to confirm the specificity of the products.

Results

Based on our previous linkage analysis, seven positional candidate genes were picked for gene expression studies in mice. Genes - which are *SIRT5, NOL7, RANBP9, CCDC90A, RNF182, CD83*, and *JARID2* – represent all the known genes mapping on a 2 Mb area around the markers D6S259, where the higher multipoint LOD score were obtained with CL/P multiplex families (Scapoli et al., 1997). Jarid2 expression was evident after 30 cycles of RT-PCR conducted with cDNA obtained from palatal shelves of mouse embryo at E14.5, while amplimers of Sirt5, Ranbp9 and Cd83 were obtained only after 35 cycles of amplification (data not shown). Coronal sections of mouse embryos head at E14.5 and E15.5 were analyzed by RNA *in situ* hybridization to evaluate expression of the candidate genes during palate development. Jarid2 was expressed at high levels in epithelial cells surrounding the nasal cavity and the palatal processes (Fig. 1), while no signal of the other genes was detected in oral area tissues. Since palatal shelves merge and fuse starting from the primary palate and proceeding gradually to the posterior direction, palate development stages can be observed in different sections throughout the antero-posterior axis. As has been reported with *Myh9* and *Tgfb3*, *Jarid2* expression is maintained when shelves adhere to each other at their medial edge epithelia at embryonic day 14.5 giving rise to the medial epithelial seam, and in the epithelial triangles (Martinelli et al., 2007). As the fusion proceeds, the Jarid2 signal gradually decreases and becomes limited to spots of the medial epithelial seam before completely disappearing after the fusion process is completed (Fig. 1).

Since expression data supported a specific role for *JARID2* in palate development, we tested its involvement in CL/P by linkage disequilibrium using intragenic markers. A total of thirteen polymorphic loci were tested for allelic association with CL/P. First, in order to investigate a potential correlation, 9 SNPs selected within *JARID2* were investigated. Subsequently, as the association data was promising, an additional four SNPs were included in

the analysis: two insertions/deletions (markers 7 and 10) within the gene and two others (markers 12 and 13) within *DTNBP*, a gene located only 2 kb downstream of *JARID2*. In order to deal with this complex multiple testing issue, we used the FBAT multi-marker test (FBAT-MM) to test for linkage or association between any marker and any disease susceptibility locus underlying CL/P. In the FBAT-MM test the P value for association was .0019 when the whole set of 13 markers was tested, while it was .0008 when only the 11 markers in *JARID2* locus were considered. P values for the single-marker association analysis computed using 100,000 Monte-Carlo permutations, not adjusted for multiple comparisons, are displayed in Table 1. Six out of 11 SNPs within *JARID2* showed evidence of association. Among them, markers 5, 6, and 8 gave the most consistent scores for association (P values lower than .001).

Linkage disequilibrium between markers was investigated using the Haploview program (Fig. 2). Three haplotype blocks were identified: the first included markers 1 and 2, the second markers 4 to 9, and the third markers 12 and 13 (Fig. 2). This structure was consistent with data from the Hapmap Consortium showing recombination hotspots between the identified blocks. A Haplotype association analysis was then conducted for each one of the blocks using 1,000,000 Monte-Carlo permutations. P values for the whole markers strongly supported an association of the *JARID2* gene with CL/P (P = .045 for block1 and P = .000019 for block 2), while confirming the lack of association with the downstream gene *DTNBP1* (P = .71 for block 3). Association with block 2 remained highly significant even after adjusting for multiple testing using the most conservative Bonferroni correction method (for three haplotype blocks, adjusted P = .000019*3 = .000057). In particular, the third most common haplotype of block 2 resulted as highly overtransmitted to patients (Table 2). This haplotype included all the alleles that resulted overtransmitted in the pairwise analysis.

A haplotype association analysis for block 2 was then conducted separately with familial and sporadic samples. Two reasons determined this procedure. First, because part of the

familial cases (n =38) was those driven the former parametric linkage results, so sporadic cases may be considered an independent replication sample (Scapoli et al., 1997). Second, because familial and sporadic cases may have different etiology. In both groups whole marker association was statistically significant, even though it was stronger in familial than sporadic cases (P values .0007 and .02, respectively). Since higher statistical significance was obtained with a smaller sample size (n = 87 and n = 131), we can assume that association was stronger among familial cases.

FBAT statistics demonstrated a significant association between alleles/haplotypes at the *JARID2* locus and a putative CL/P susceptibility allele. Since FBAT analysis does not provide for a magnitude of genetic effects we applied the formulae described by Scherag and colleagues to calculate the genotype risk for each of the markers (Scherag et al., 2002). In line with the FBAT statistics, higher values were obtained with marker 8; relative risks were 1.83 (95% CI .99-3.41) and 3.10 (95% CI 1.54-6.27) for heterozygous and homozygous carriers, respectively. The attributable risk, that is the probability of exposure given that a person has the disease, resulted as .55 (95% CI .28-.81).

In order to identify the *JARID2* mutations responsible for CL/P we screened the 18 exons, exon-intron boundaries, and the putative promoter region by a direct sequencing of PCR products. We reasoned that CL/P may be due to rare mutations or to a common variant allele/s present in the high risk haplotype. In order to pursue both hypotheses, mutation screening was performed in 25 unrelated individuals, 15 of whom belong to the 15 pedigrees showing the highest LOD score in linkage analysis with 6p23 genetic markers (Scapoli et al., 1997), while the remaining 10 were selected from cases carrying one or two of the overtransmitted haplotypes that have been identified in the present study. No obvious mutations were detected. A number of known synonymous variations, known as rs34326651, rs7768621, rs742099, rs2235258 and the non-synonymous rs35474598 in exon 7, were detected. None of them was

likely to be involved in CL/P since each sequence variant was detected in only a single or a few samples. Two intronic variant alleles appeared relatively frequently, for this reason we included these polymorphisms in the linkage disequilibrium study (Table 1, markers 7 and 10). However, the association results did not sustain their role as a direct cause of CL/P.

Aware of the enormous difficulties in trying to determine any potential involvement of intronic variants in controlling gene expression or RNA maturation, we start to look at the effect of SNPs on splicing mechanism of *JARID2*. We select the rs2076056, marker 8 in this study, characterized by the C to G substitution of nucleotide at position +9 in intron 6 of the human *JARID2*. Genomic PCR products of the exon 6 plus boundaries containing the two variants were inserted within the NdeI restriction site of the pTBNde(min) expression vector to be a part of a minigene of three exons. Splicing process of the minigenes were tested in a transient transfection assay with Hela cells. As confirmed by direct sequence analysis, RT-PCR products obtained from transfected cells showed that the minigene is spliced as expected regardless of nucleotide change at rs2976056 (Fig. 3). This data suggest that rs2076056, even if strongly associated with the disease, does not interfere with the correct maturation processing of *JARID2* mRNA.

Discussion

CL/P is a frequent malformation with complex inheritance. Environmental factors and multiple genes have been related to clefting. Several different genes/loci have been claimed to be involved in CL/P etiology, however replication experiments have often failed to validate previous data. OFC1 locus on the short arm of chromosome 6 has been one of the most investigated regions following its initial finding by linkage analysis (Eiberg et al., 1987). Although discordant data have been reported, many authors reported positive linkage with markers in this region (Scapoli et al., 1997; Prescott et al., 2000; Carreno et al., 2002; Moreno et al., 2004; Schultz et al., 2004).

Following our linkage analyses, we decided to screen positional candidate genes with a gene expression investigation, basing on the idea that the CL/P causing gene had to be expressed during palate development. Gene expression in developing palate of seven candidate genes, was investigated by RT-PCR and RNA *in situ* hybridization experiments in mouse embryos. Among them, Jarid2 showed a high level of expression specifically localized in the medial edge epithelium of merging palatal shelves that disappear immediately after fusion. The critical time of regulated expression suggests a decisive role for Jarid2 in palate development. Public database searches revealed that *JARID2* was also expressed in different human embryo craniofacial structures relevant to palate and lip development between the 4th and the 8th week of gestation (Park et al., 2006). Although it cannot be excluded a role for other positional candidate genes, at different developmental stage or with a lower expression level, *JARID2* was prioritized for further investigations.

A family based association study was performed to determine if a common variant of *JARID2* was involved in CL/P. Different SNP alleles and haplotypes appeared significantly overtransmitted to affected individuals and thus associated with a putative CL/P susceptibility allele. Evidence of association was obtained with two independent samples, constituted by

familial and sporadic cases, respectively. Stronger association was observed among familial cases. As expected, genetic factors appear to play a relatively prominent role in this group, while environmental factors may sometimes play a determining role in sporadic CL/P. The attributable risk, that is the percentage of affecteds attributable to the risk allele, calculated for the marker 8 was .55 (95% CI .28-.81), while relative risks were 1.83 (95% CI .99-3.41) and 3.10 (95% CI 1.54-6.27) for heterozygous and homozygous carriers, respectively. The method used to calculate these risks implicitly assumes that the disease is caused by a single functional SNP which has itself been investigated. Since there is no evidence that marker 8 directly causes CL/P, the risk calculated probably represents an underestimation of the genetic effect of the real mutation/s (Franke et al., 2005).

Mutations were searched in patients carrying the risk haplotype, where a common causative variant is more likely to be found, as well as in patients of linked multiplex families where less common variants, conferring an higher risk, may be identified (Bodmer and Bonilla, 2008). No obvious mutations in *JARID2* coding regions were found among 25 selected patient and our initial attempt by functional study to correlate common variant with alteration of the splicing process was also unsuccessful. The inability to find pathogenic alleles within regions statistically associated with pathological phenotypes is a relatively common scenario in multifactor disease research, suggesting that greater efforts should be made in the attempt to identify those subtle genetic abnormalities, which lead to disease susceptibility, such as nucleotide changes in regions regulating gene expression or the splicing mechanism.

Previously, six genes which map in 6p23-p25 were considered for a linkage disequilibrium study that involved 64 candidate genes and 58 nuclear families with orofacial clefting (Park et al., 2006). A possible role for *C6orf105* was reported, while *JARID2* did not show association. These results appear in contrast with our data, but they should be considered with caution because were obtained with a small sample size. Otherwise, the discrepancy might be due to

sample composition (familial/sporadic cases ratio), to genetic heterogeneity, or to the presence of different susceptibility genes in the two samples. These argumentation may also explain why no association with 6p23, or IRF6 markers was observed in the recent published genome wide association analyses (Birnbaum et al., 2009; Grant et al., 2009; Mangold et al., 2009).

JARID2 codes for a transcription factor with an AT-Rich interaction domain (ARID) in association with two homology Jmj regions (JmjC and JmjN) of members of the jumonji family (Jung et al., 2005). Although the functional roles of Jarid2 remain largely unknown, knockout experiments have revealed its contribution to mouse embryonic development (Jung et al., 2005). Homozygous mutants are often lethal and reveal that Jarid2 plays an important role in heart and liver development, neural tube closure, and haematopoiesis. Orofacial clefting was not described as a feature of knockout mice but it should be noted that the phenotypes of mutants vary considerably, depending on genetic background. Specifically, heart defects may include double-outlet right ventricle and/or prominent interventricular septal defects. Interestingly, the ventricular septum grows and fuses to the inferior endocardial cushion, dividing the ventricle into two chambers in a process that shares some homologies with palate development (Olson, 2004). The hypothesis that heart and orofacial development share some molecular pathways is supported by the evidence that cardiac defects and CL/P, or cleft palate, are features that frequently occurs together in birth syndromes (to date, 180 entries in OMIM).

Recently published papers have shown that JARID2 interacts with the Polycomb-Repressive Complex 2 (PRC2) (Pasini et al., ; Peng et al., 2009). PRC2 regulates developmental gene expression patterns influencing chromatin state via histone H3 methylation. In the JARID2/PRC2 complex, JARID2 promotes selective DNA binding and negatively regulates histone methylation activity (Pasini et al., ; Peng et al., 2009). Genome-wide chromatin immunoprecipitation sequencing revealed that, in mouse embryonic stem cells, JARID2 can binds promoter regions of many genes essential for normal development

and differentiation. Interestingly, this group includes several genes that are known, or suspected to be involved in CL/P malformation, such as members of the fibroblast growth factor, forkhead/winged-helix, T BoX, msh homeobox, and Wnt families. It has been shown that the ARID domain of JARID2 is essential for the proper differentiation of mouse embryonic stem cells and for normal Xenopus laevis early development (Pasini et al., ; Peng et al., 2009). However, consisting with the evidence that JARID2 is able to bind different DNA sequences (Kim et al., 2003), it has been proposed that the association of PRC2 with specific target genes is not only determined by JARID2, but could depends by JARID2 in combination with other transcription factors (Pasini et al.).

In this paper compelling evidence supporting a role for JARID2 in orofacial development has been shown. The gene is specifically expressed in palatal shelves during critical time points and a family based association analysis is consistent with *JARID2* as the OFC1 gene involved in CL/P. Understanding the molecular role of JARID2 within this process, may offer additional information to detect other genetic factors and further unravel the complex genetics of CL/P.

Acknowledgements

This work is dedicated to the memory of our mentor Prof. Paolo Carinci. We are indebted to Jeff Murray for helpful discussions and critical reading of the manuscript. We thank F. Pagani whom kindly supplied the pTBNde(min) vector. This work was funded in part by the Italian Telethon Foundation (GGPO5147), University of Bologna (RFO), University of Ferrara (FAR), and Fondazione Cassa di Risparmio di Ferrara.

References

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2):263-265.

Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, Rubini M, Baluardo C, Ferrian M, Almeida de Assis N, Alblas MA, Barth S, Freudenberg J, Lauster C, Schmidt G, Scheer M, Braumann B, Berge SJ, Reich RH, Schiefke F, Hemprich A, Potzsch S, Steegers-Theunissen RP, Potzsch B, Moebus S, Horsthemke B, Kramer FJ, Wienker TF, Mossey PA, Propping P, Cichon S, Hoffmann P, Knapp M, Nothen MM, Mangold E. 2009. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet 41(4):473-477.

Bodmer W, Bonilla C. 2008. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 40(6):695-701.

Carinci F, Scapoli L, Palmieri A, Zollino I, Pezzetti F. 2007. Human genetic factors in nonsyndromic cleft lip and palate: an update. Int J Pediatr Otorhinolaryngol 71(10):1509-1519.

Carreno H, Suazo J, Paredes M, Sola J, Valenzuela J, Blanco R. 2002. [Association between cleft lip/palate phenotype and non syndrome microsatellite markers located in 6p, 17q and 19q]. Rev Med Chil 130(1):35-44.

Davies SJ, Wise C, Venkatesh B, Mirza G, Jefferson A, Volpi EV, Ragoussis J. 2004. Mapping of three translocation breakpoints associated with orofacial clefting within 6p24 and identification of new transcripts within the region. Cytogenet Genome Res 105(1):47-53.

Eiberg H, Bixler D, Nielsen LS, Conneally PM, Mohr J. 1987. Suggestion of linkage of a major locus for nonsyndromic orofacial cleft with F13A and tentative assignment to chromosome 6. Clin Genet 32(2):129-132.

Franke D, Philippi A, Tores F, Hager J, Ziegler A, Konig IR. 2005. On confidence intervals for genotype relative risks and attributable risks from case parent trio designs for candidate-gene studies. Hum Hered 60(2):81-88.

Grant SF, Wang K, Zhang H, Glaberson W, Annaiah K, Kim CE, Bradfield JP, Glessner JT, Thomas KA, Garris M, Frackelton EC, Otieno FG, Chiavacci RM, Nah HD, Kirschner RE, Hakonarson H. 2009. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. J Pediatr 155(6):909-913.

Horvath S, Xu X, Laird NM. 2001. The family based association test method: strategies for studying general genotype--phenotype associations. Eur J Hum Genet 9(4):301-306.

Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM. 2004. Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. Genet Epidemiol 26(1):61-69.

Jung J, Mysliwiec MR, Lee Y. 2005. Roles of JUMONJI in mouse embryonic development. Dev Dyn 232(1):21-32.

Kim TG, Kraus JC, Chen J, Lee Y. 2003. JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor. J Biol Chem 278(43):42247-42255.

Mangold E, Reutter H, Birnbaum S, Walier M, Mattheisen M, Henschke H, Lauster C, Schmidt G, Schiefke F, Reich RH, Scheer M, Hemprich A, Martini M, Braumann B, Krimmel M, Opitz C, Lenz JH, Kramer FJ, Wienker TF, Nothen MM, Diaz Lacava A. 2009. Genome-wide linkage scan of nonsyndromic orofacial clefting in 91 families of central European origin. Am J Med Genet A 149A(12):2680-2694.

Marigo V, Nigro A, Pecci A, Montanaro D, Di Stazio M, Balduini CL, Savoia A. 2004. Correlation between the clinical phenotype of MYH9-related disease and tissue distribution of class II nonmuscle myosin heavy chains. Genomics 83(6):1125-1133.

Martinelli M, Di Stazio M, Scapoli L, Marchesini J, Di Bari F, Pezzetti F, Carinci F, Palmieri A, Carinci P, Savoia A. 2007. Cleft lip with or without cleft palate: implication of the heavy chain of non-muscle myosin IIA. J Med Genet 44(6):387-392.

Moreno LM, Arcos-Burgos M, Marazita ML, Krahn K, Maher BS, Cooper ME, Valencia-Ramirez CR, Lidral AC. 2004. Genetic analysis of candidate loci in non-syndromic cleft lip families from Antioquia-Colombia and Ohio. Am J Med Genet A 125(2):135-144.

Olson EN. 2004. A decade of discoveries in cardiac biology. Nat Med 10(5):467-474.

Pagani F, Buratti E, Stuani C, Romano M, Zuccato E, Niksic M, Giglio L, Faraguna D, Baralle FE. 2000. Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. J Biol Chem 275(28):21041-21047.

Pagani F, Stuani C, Tzetis M, Kanavakis E, Efthymiadou A, Doudounakis S, Casals T, Baralle FE. 2003. New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. Hum Mol Genet 12(10):1111-1120.

Park JW, Cai J, McIntosh I, Jabs EW, Fallin MD, Ingersoll R, Hetmanski JB, Vekemans M, Attie-Bitach T, Lovett M, Scott AF, Beaty TH. 2006. High throughput SNP and expression analyses of candidate genes for non-syndromic oral clefts. J Med Genet 43(7):598-608.

Pasini D, Cloos PA, Walfridsson J, Olsson L, Bukowski JP, Johansen JV, Bak M, Tommerup N, Rappsilber J, Helin K. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature.

Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, Wysocka J. 2009. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell 139(7):1290-1302.

Prescott NJ, Lees MM, Winter RM, Malcolm S. 2000. Identification of susceptibility loci for nonsyndromic cleft lip with or without cleft palate in a two stage genome scan of affected sib-pairs. Hum Genet 106(3):345-350.

Scapoli C, Collins A, Martinelli M, Pezzetti F, Scapoli L, Tognon M. 1999. Combined segregation and linkage analysis of nonsyndromic orofacial cleft in two candidate regions. Ann Hum Genet 63(Pt 1):17-25.

Scapoli L, Pezzetti F, Carinci F, Martinelli M, Carinci P, Tognon M. 1997. Evidence of linkage to 6p23 and genetic heterogeneity in nonsyndromic cleft lip with or without cleft palate. Genomics 43(2):216-220.

Scherag A, Dempfle A, Hinney A, Hebebrand J, Schafer H. 2002. Confidence intervals for genotype relative risks and allele frequencies from the case parent trio design for candidate-gene studies. Hum Hered 54(4):210-217.

Schliekelman P, Slatkin M. 2002. Multiplex relative risk and estimation of the number of loci underlying an inherited disease. Am J Hum Genet 71(6):1369-1385.

Schultz RE, Cooper ME, Daack-Hirsch S, Shi M, Nepomucena B, Graf KA, O'Brien EK, O'Brien SE, Marazita ML, Murray JC. 2004. Targeted scan of fifteen regions for nonsyndromic cleft lip and palate in Filipino families. Am J Med Genet A 125(1):17-22.

Topping A, Harris P, Moss AL. 2002. The 6p deletion syndrome: a new orofacial clefting syndrome and its implications for antenatal screening. Br J Plast Surg 55(1):68-72.

Figure legends

Figure 1 - In situ RNA hybridization of *Jarid2* during fusion of palatal shelves in mouse between E14.5 and E15.5. The palatal epithelium shows a strong *Jarid2* hybridization signal along the anterior–posterior axis when the shelves adhere to each other at their medial edge epithelia (arrow). At E15.5 the fusion of palatal shelves terminates and the expression is limited in spots (arrow) before disappearing completely at the end of the process. Ps. Palatal shelves; t, tongue; nc, nasal cavity.

Figure 2 - Schematic view of linkage disequilibrium results. From the top to the bottom: chromosome 6 coordinates, genes located within the region, hapmap consortium hotspot recombination data, positions of SNPs genotyped in this study, and identified haplotype blocks. The D' values indicating the linkage disequilibrium relationships between each SNP pair are reported in each box.

Figure 3 – Effect of SNP rs2076056 of the human *JARID2* gene in a minigene transient transfection assay. A) Schematic representation of the genomic PCR product inserted within NdeI restriction site of pTBNde(min). Alpha-globin, fibronectin and human *JARID2* exons are indicated in black, shaded and white boxes, respectively. The sequence surrounding the underlined C to G polymorphism (rs2076056), which is localized at nucleotide +9 of intron 6, is also shown. Sequences from exon 6 and intron 6 are in uppercase and lowercase, respectively. B) RT-PCR analysis of transient transfection of *JARID2* exon 6 minigene in HeLa cells using primers indicated by arrows in the schematic representation of the minigene. Lane 1, molecular marker (100 bp DNA ladder; New England, Biolobs); lane 2, pTBNde(min) alone showing a product of 239 bp; lanes 3 and 4, pTBNde(min) including *JARID2* exon 6 with variant C and G, respectively both generating a product of 475 bp; Lane 5, negative control.

Table 1. Markers information and results of allelic association based on 100,000x MonteCarlo permutations.

Marker	dbSNP ID[a]	genomic position[b]	alleles[c]	MAF[d]	Z score[e]	P value
1	rs6915344	chr6:15348423	C/T	0.403	-1.781	0.08058
2	rs9464779	chr6:15362784	T/C	0.490	0.420	0.75238
3	rs2072820	chr6:15482523	T/C	0.497	2.434	0.01720
4	rs2237149	chr6:15520044	C/A	0.382	-2.523	0.01370
5	rs2299043	chr6:15565149	G/A	0.316	3.260	0.00089
6	rs2237138	chr6:15571374	T/C	0.330	3.677	0.00012
7	rs71932765	chr6:15576710	in/del	0.340	-1.755	0.08893
8	rs2076056	chr6:15595761	C/G	0.321	3.681	0.00023
9	rs2282819	chr6:15611488	G/A	0.289	-2.286	0.02297
10	rs34554301	chr6:15621379	in/del	0.214	1.229	0.20040
11	rs2072821	chr6:15626015	G/C	0.253	-1.249	0.22333
12	rs2056942	chr6:15650277	C/A	0.241	0.474	0.56145
13	rs35861734	chr6:15716490	A/C	0.242	0.000	0.97672

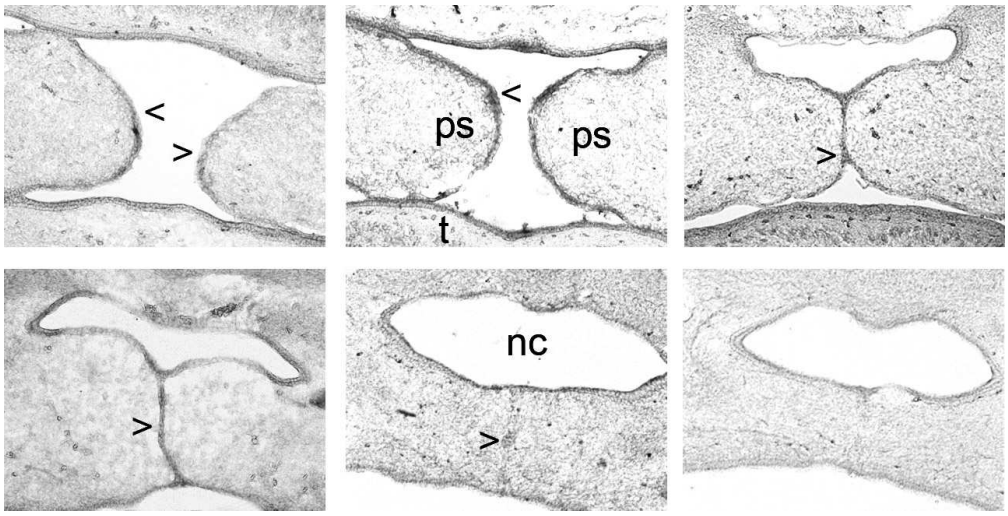
[a]NCBI-SNPs data base accession numbers
[b]UCSC Genome Browser on Human March 2006 assembly
[c]Alleles in JARID2 coding frame, major allele first
[d]Minor Allele Frequence
[e]FBAT Z score for the major allele, positive scores means overtransmission, while negative undertransmission

Table 2. Association analysis for haplotypes of block 2 with frequency >1%. Haplotipe 3 showed positive association, while haplotype 9 negative association.

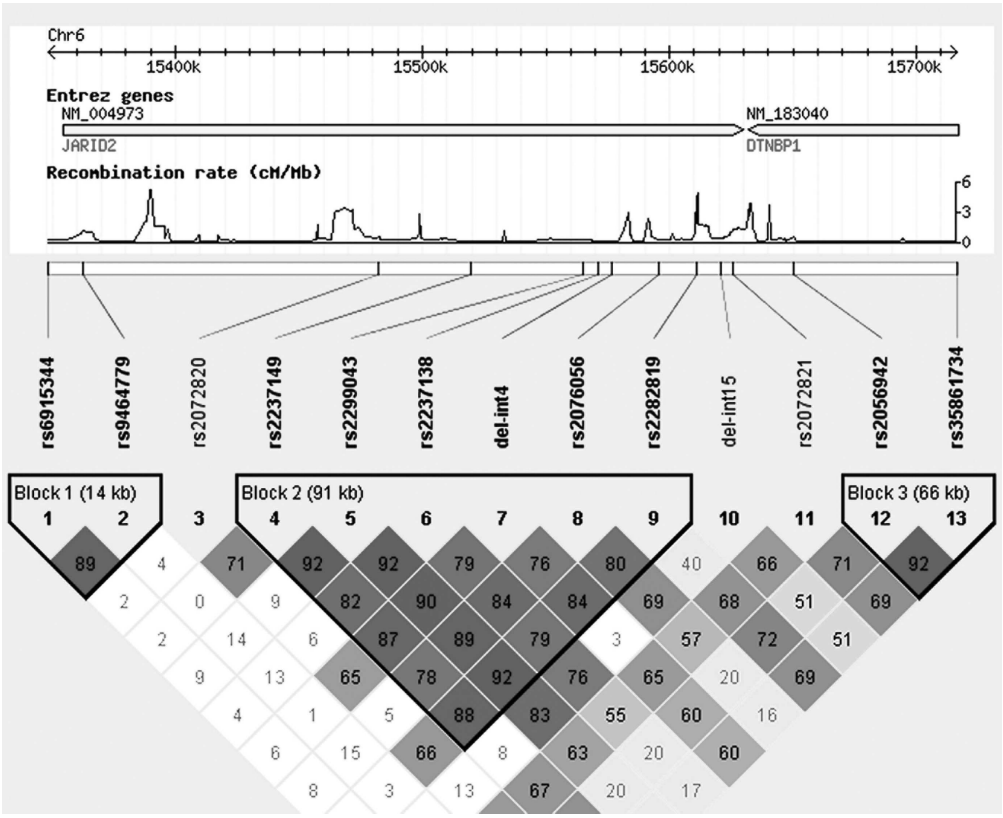
haplotype	Marker						freq[a]	P value[b]
	4	5	6	7	8	9		
h1	c	g	t	in	c	g	0.27	1.0
h2	c	a	c	in	g	g	0.26	1.0
h3	a	g	t	del	c	a	0.21	0.000108
h4	a	g	t	del	c	g	0.06	1.0
h5	a	g	t	in	c	g	0.04	1.0
h6	a	g	t	in	c	a	0.02	1.0
h7	c	a	c	in	c	g	0.02	1.0
h8	c	g	t	del	c	g	0.01	1.0
h9	a	g	t	del	g	a	0.01	0.035226

[a]Estimated haplotype frequency

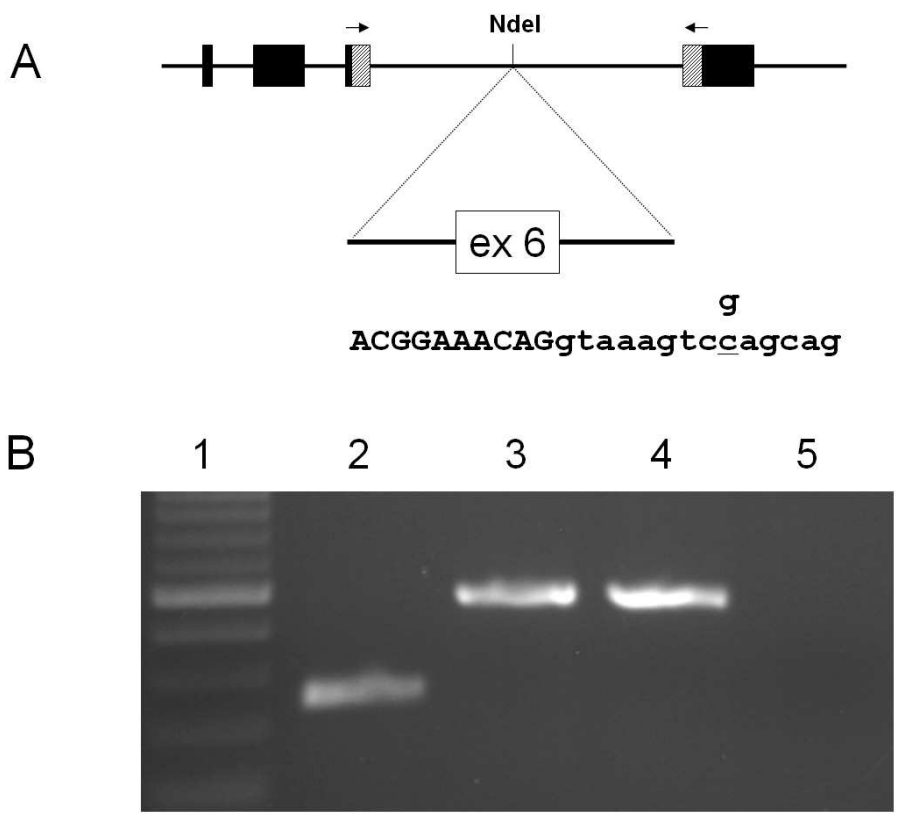
[b]Empirical P values corrected for multiple testing



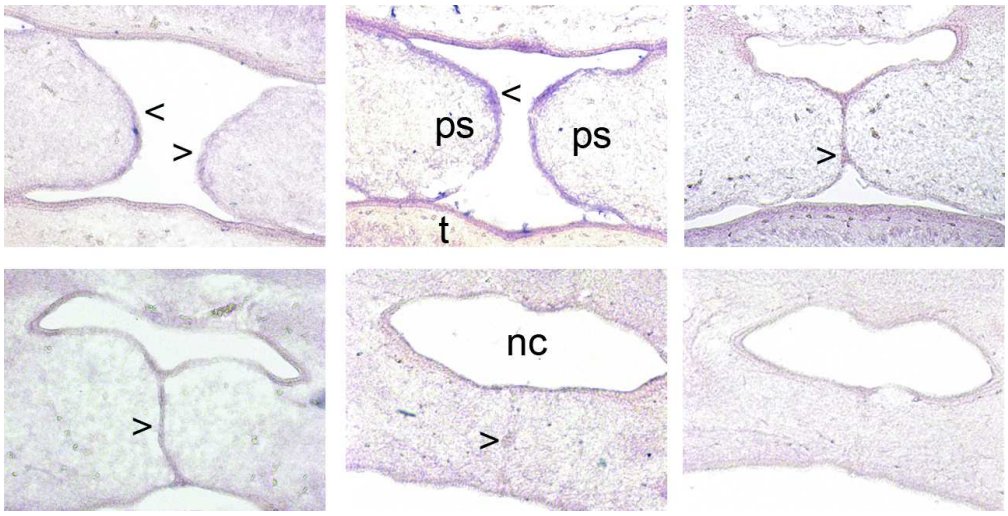
108x55mm (300 x 300 DPI)



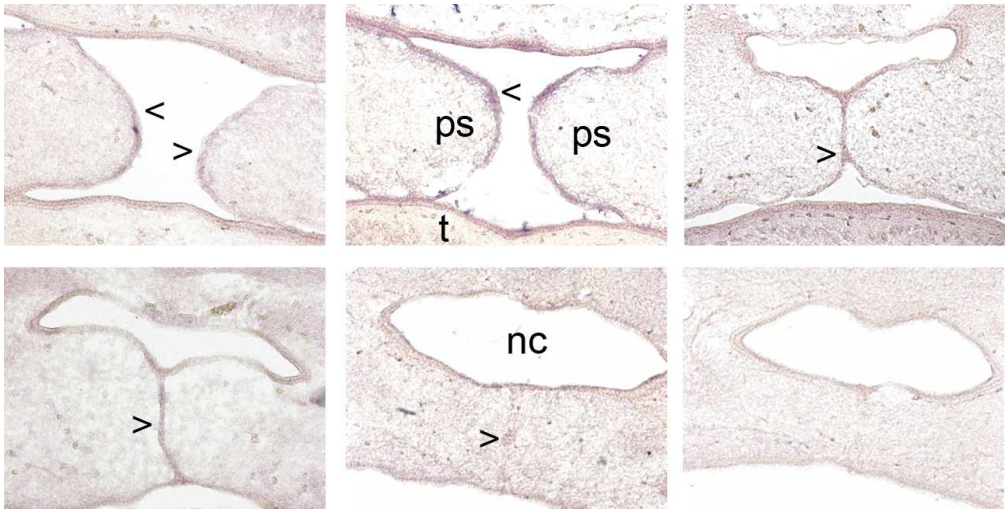
232x188mm (300 x 300 DPI)



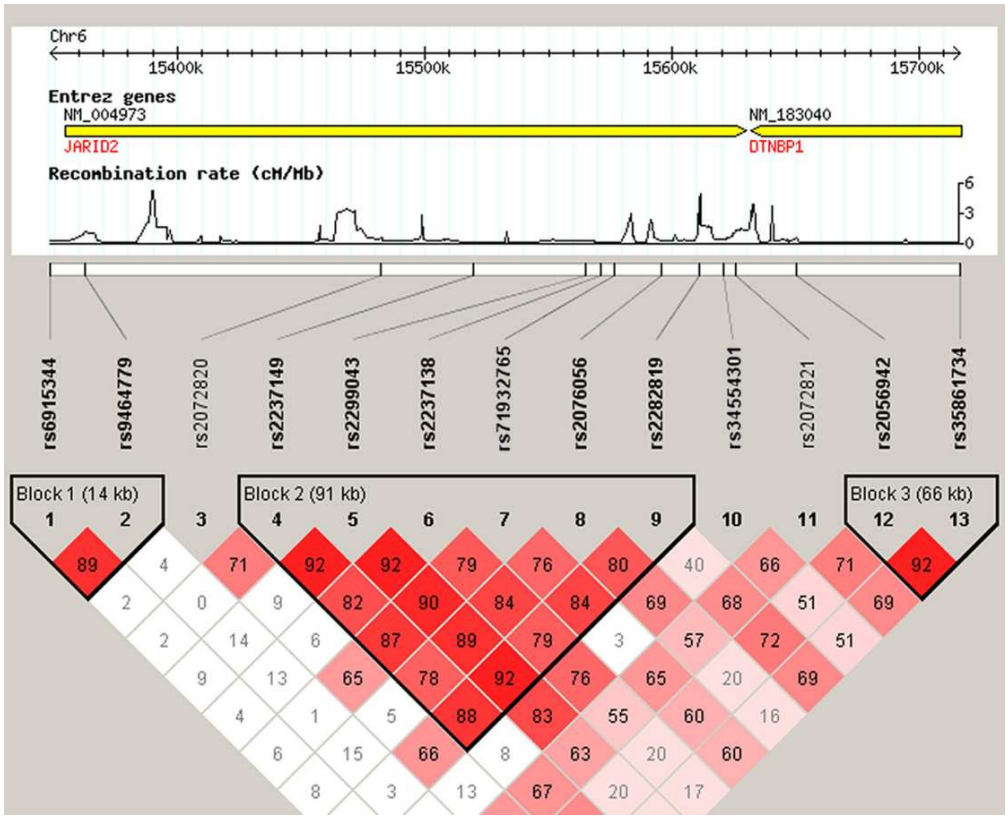
48x40mm (600 x 600 DPI)



108x55mm (300 x 300 DPI)



108x55mm (300 x 300 DPI)



47x38mm (600 x 600 DPI)

Primers for RT-PCR (mouse)

MRNF182F aagaaatgcctgcctgagaa
MRNF182R ggacgtctgaaagagcaagg
CCDC90AF GAGAGCTCAGCCTGTCTGCT
CCDC90AR ATTTCCTGCTGCATCTTGCT
RANBP9f gatggctacatgggaattgg
RANBP9r tccattcttcgtgtaaaagca
HMjarid2F GATTGCACAAGCAGAAGCAA
HMjarid2R AGCCACTTTCGAGGCTTTTT
HMnol71F GAGGCGATGGTGGACGAG
HMnol71R CTGGCGAAAGTCAGCTCCT
HMsirt5F TGGTCATCACCCAGAACATC
HMsirt5R ACGTGAGGTCGCAGCAAG
Mcd83F Tccagctcctgtttctaggc
Mcd83R agctgttttgcttgctctcc
CD83musF AAACCTGGTACGGAACAAGC
CD83musR TGCTGGGGAATAGTTCACTG

Primers for RNA in situ hibridization (mouse)

RNF182-F-T3 attaaccctcactaaagggaTCCTTTAATCCAAGTCA
RNF182-R-T7 taatacgactcactatagggACACGGTGGTGACAGAG
CCDC90A-F-T3 attaaccctcactaaagggaTTTTATCGCCTGTGGA
CCDC90A-R-T7 taatacgactcactatagggaacccagagacaaagg
RANBP9-F-T3 attaaccctcactaaagggaCAGGCCACACAGTGTCTA
RANBP9-F-T7 taatacgactcactatagggtcggaacagctcccaat
NOL7-F-T7 taatacgactcactatagggTTCATTTCTCAAAATTTAGTC
NOL7-F-T3 attaaccctcactaaagggaCACACATGGTAGGGAAGA
SIRT5-F-T7 taatacgactcactatagggCACTAACGGGAAAAAT
SIRT5-F-T3 attaaccctcactaaagggaagtaagcactgaaaaga
CD83musT7 gtgtaatacgactcactatagggAGTCACCTCCCCAAGCAAAC
CD83musT3 agaattaaccctcactaaagggACCTTCGCACTGGGAAATTA
JARID2-F-T3 taatacgactcactatAGGGCTTCGAGACTGCCAA
JARID2-R-T7 attaaccctcactaaAGGGACCTCTTTTTGGTGTGG

Primers for exon sequencing (human)

PrJARID2-1bisf AGAGGGGTGACCTCGGACTAC
PrJARID2-1bisr AGAATGGTCCCCTTGATCTTCT
PrJARID2-2bisf GACACGTCCAAGCGTTTGTT
PrJARID2-2bisr ACAAGTTAGGGCTCCTCGGATA
PrJARID2-3f GAATTATCCGAGGAGCCCTAAC
PrJARID2-3r CTGATTGCAAAAGGGGACAAT
JARID2-1F ACAACAATAAAAACCACCAGGA
JARID2-1R aacttgacattcacagccattg
JARID2-2F ttgactctgaatattgccttgc
JARID2-2R ccaattcaggaacaagtcca

JARID2-3F cgttgtgtagggttaactgtgg
JARID2-3R ggtgaaaactggcacctgag
JARID2-4F attttattgccgcacgtagg
JARID2-4R ggttggggacagaaaaatga
JARID2-5F tcggttttgtttcattgtgg
JARID2-5R cagcagaggctctcttccag
JARID2-6F ccatatcagtagtttgcgtggt
JARID2-6R tggtgaagtgaagtgggaaa
JARID2-7AF gtgggccggtcattttagt
JARID2-7AR GCCGATTCCTCTCCAGACT
JARID2-7BF AGGCACACCAGGCGGAGAAGC
JARID2-7BR AGGGAACTGCAGGGATCTGGGG
JARID2-8F caagaagtacaggttgaatgagga
JARID2-8R gtgcatagcacgctccact
JARID2-9F ctggcctcatttgcagtagg
JARID2-9R cagggcaggacaggatgt
JARID2-10/11F ttgggttgaactgtgctctg
JARID2-10/11R agccctggttctcagcaag
JARID2-12F acttactgttcctttgagaccttg
JARID2-12R gccccacttcatgtggtaat
JARID2-13F ggcccagtacatgcaggag
JARID2-13R gggaaaccatgttcaagtgc
JARID2-14F gtgcgcatacgtcacctg
JARID2-14R gcagggatgcttctgtgtg
JARID2-15/16F ctgccccctccaccaagaa
JARID2-15/16R agagagcgccctccctcttc
JARID2-17F ttccagcatttccagtcctc
JARID2-17R ccggagtgcctacatccag
JARID2-18F cctaaacttgcccctgcat
JARID2-18R GTTCAACAGTTTAATGCTAAAACAAA

Primers for insertion/deletion polymorphism typing

rs71932765F GCAGGAGAACCTGGTGGAAT
rs71932765R TCGGTTTTGTTTCATTGTGG
rs34554301f cctgacaggagggtgtgtct
rs34554301r AGCTCTGTATCCctgccaga

Primers for Transient expression of hybrid minigene

6F gtagtttgcgtggtagtgg
6R ctcatgcaaaggtgcgctc
2-3alpha caacttcaagctcctaagccactgc
B2 taggatccggtcaccaggaagttggttaaatca

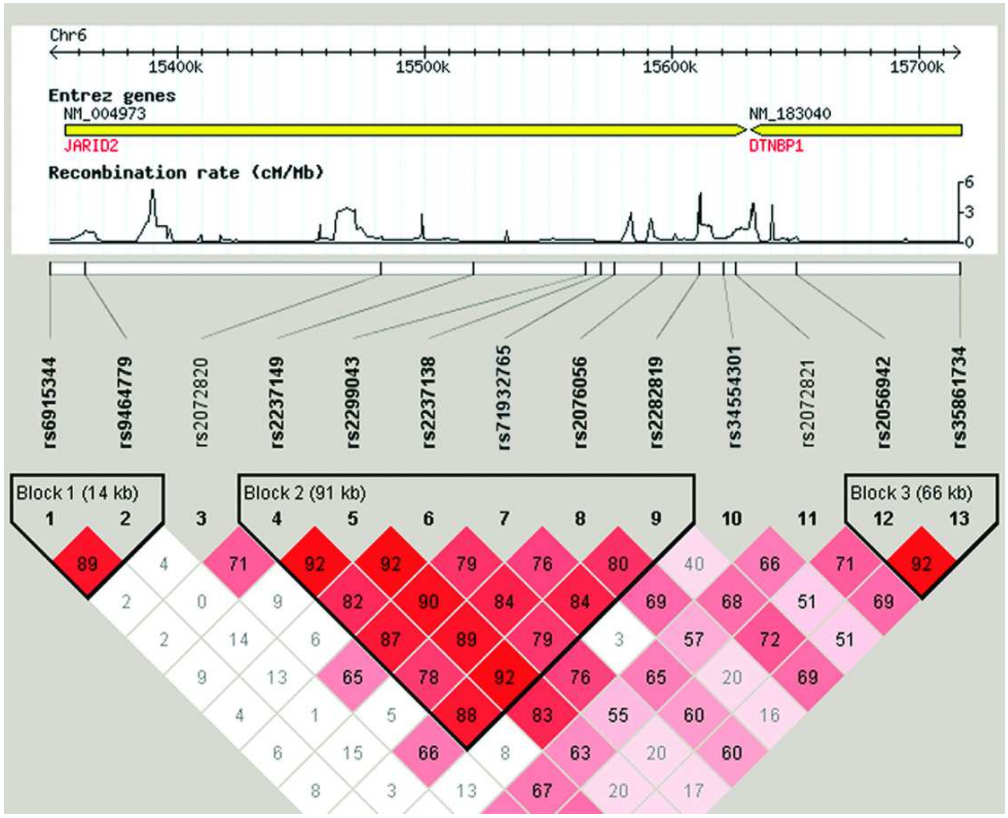


Fig 2
47x38mm (600 x 600 DPI)