Supplementary Information

Disruption of an AP-2α binding site in an *IRF6* enhancer is strongly associated with cleft lip

Fedik Rahimov[1], Mary L. Marazita[2], Axel Visel[3], Margaret E. Cooper[2], Michael J. Hitchler[4], Michele Rubini[5], Frederick E. Domann[4], Manika Govil[2], Kaare Christensen[6], Camille Bille[6], Mads Melbye[7], Astanand Jugessur[8], Rolv T. Lie[8], Allen J. Wilcox[9], David R. Fitzpatrick[10], Eric D. Green[11], NISC Comparative Sequencing Program[11], Peter A. Mossey[12], Julian Little[13], Regine P. Steegers-Theunissen[14], Len A. Pennacchio[3], Brian C. Schutte[1] and Jeffrey C. Murray[1]

[1]Department of Pediatrics, University of Iowa, 2182 ML, S Grand Ave, Iowa City, IA 52242 [2]Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Suite 500, Bridgeside Point Building, Pittsburgh, PA 15219 [3]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 [4]Department of Radiation Oncology, University of Iowa, 4202 MERF, Iowa City, IA 52242 [5]Department of Experimental Diagnostic Medicine, Medical Genetics Unit, University of Ferrara, Ferrara, Italy [6]Center for the Prevention of Congenital Malformations, Institute of Public Health, University of Southern Denmark, J.B. Winsløws Vej 9B, 5000 Odense C, Denmark [7] Department of Epidemiology Research, Danish Epidemiology Science Center, Statens Serum Institute, Copenhagen, Denmark [8] Section for Epidemiology and Medical Statistics, Department of Public Health and Primary Health Care, University of Bergen, Kalfarveien 31, N-5018, Bergen, Norway [9] Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, 111 T.W. Alexander Drive, Durham, NC 27709 [10]MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, UK [11]Genome Technology Branch and NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892 [12]Dental Hospital & School, University of Dundee, Dundee, Scotland, UK [13]Department of Epidemiology and Community Medicine, University of Ottawa, 451 Smyth Road, Room 3227, Ottawa, Ontario, Canada K1H 8M5 [14]University Medical Center, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands

Corresponding author: Jeffrey C. Murray, M.D.
Address: University of Iowa, Department of Pediatrics, 2182 ML
S Grand Ave, Iowa City, IA 52242, USA
Phone: (319) 335-6897
Fax: (319) 335-6970
E-mail: jeff-murray@uiowa.edu

Supplementary methods

Comparative genomic sequence analysis

MultiPipMaker, http://pipmaker.bx.psu.edu/pipmaker

WebMCS, http://zoo.nhgri.nih.gov/mcs

Sequencing

Primer 3, http://fokker.wi.mit.edu/primer3/input.htm

Electrophoretic mobility shift assay (EMSA)

Infrared dye (IRDye-700) end-labeled and unlabeled oligonucleotide probes were purchased from LI-COR Biosciences (Lincoln, NE) and IDT (Coralville, IA), respectively. We performed EMSAs by using the LI-COR EMSA Kit. Briefly, 6μg of human recombinant AP-2α protein (Promega) was incubated with 2.5nM of labeled probe in 20μl binding reaction for 20 min at 4^{o}C. Each binding reaction contained 2μl of LI-COR 10X binding buffer, 2.5mM DTT, 2.5% Tween-20 and 1%NP-40. To prevent nonspecific binding, 1μg of poly(dI-dC) and 0.1μg of shared salmon sperm DNA were added to each reaction. The samples were then electrophoresed for 2.5 hours at 280V in 0.5XTBE buffer on an 8% polyacrylamide gel at 4^{o}C following pre-run of the gels for 1hr at 200V. The gel was then visualized by Odyssey® Infrared Imaging System (LI-COR). For competition experiments, a 50-, 100-, and 200-fold molar excess of unlabeled competitor oligonucleotide probe was incubated for 20min before addition of the labeled probe. For supershift assay, 3μl of anti-AP-2α antibody (Upstate Biotechnology) was incubated for 20 min at 4^{o}C prior to the addition of the labeled probe.

Chromatin immunoprecipitation (ChIP) assay

Protein cross-linked chromatin from ~$2x10^7$ Ad-AP-2α-infected and uninfected HaCaT keratinocyte cells was isolated as detailed previously[1]. A small amount of the chromatin sample was removed as input control, while the remainder was split into two equal parts and immunoprecipitated with 10μg of anti-AP-2α antibody or control mouse IgG overnight at 4°C with agitation.

Chromatin/antibody complexes were then collected using Protein G agarose followed by washing and elution according to the manufacturer's instructions. DNA was then purified from input chromatin and immunoprecipitation elutions by reversing crosslinks using 200mM NaCl at 65°C for 4 hr followed by the Qiagen DNeasy Kit according to manufacturer's protocol. The amount of immunoprecipitated target region was determined by SYBR Green (Applied Biosystems) quantitative real-time PCR with primers for the target sequence in MCS-9.7 and control region (**Supplementary Table 4**) at a final concentration of 100nM. Specific amplification of the target sequences were tested on agarose gel. Real-time PCR was carried out in triplicate on 5ng of DNA under the following conditions: 10 min denaturation at 95^{o}C, followed by 40 cycles of 95^{o}C for 10 sec and 60^{o}C for 30 sec. Amplification of the target amplicon was monitored as a function of increased SYBR Green fluorescence. An analysis threshold was set and the cycle threshold (C_t) computed for each sample. Fold enrichment of target sequence was calculated using the following formula (Fold enrichment = $2^{(\mathrm{Ct\ AP\text{-}2\alpha\text{-}Ab\ IPed})-(\mathrm{Ct\ IgG\ IPed})}$).

Luciferase reporter assay

For luciferase reporter assay we generated reporter constructs by inserting 540bp genomic segment (chr1:208055787-208056326, UCSC hg18) containing the entire MCS-9.7Kb region upstream of firefly luciferase open reading frame (ORF) driven by the SV40 promoter. DNA samples from individuals homozygous for -14474A>G, -14523G>A, and rs642961 variants were PCR amplified and cloned into the pGL3-Basic and pGL3-Promoter vectors (Promega) in both orientations. HFK cells were seeded in 12-well plates at ~ $4x10^7$ cells/well density in 1mL of K-SFM media supplemented with rEGF (0.16ng/mL) and BPE (25μg/mL) 24 hours prior to transfection. These cells were then cotransfected with 1μg of reporter construct and 20ng hRL-TK renilla luciferase plasmid using 2μg of Lipofectamine™ LTX reagent in 500μl of reduced-serum Opti-MEM® I media (Invitrogen). Dual Luciferase® Assay (Promega) was performed 24 hours after transfection according to manufacturer's protocol. The pGL3-Basic and pGL3-Promoter vectors were used as controls. Renilla luciferase was used as an internal control to normalize the firefly luciferase activity. Luciferase activity was measured with Wallac Victor^2 luminometer (PerkinElmer).

References

1. Provenzano, M.J. *et al.* AP-2 participates in the transcriptional control of the amyloid precursor protein (APP) gene in oral squamous cell carcinoma. *Exp Mol Pathol* **83**, 277-282 (2007).

Supplementary Table 1: Multispecies conserved sequences (MCSs)

Cromosome	Start*	End*	Length (bp)	Gene	Location	Sequenced MCSs**
chr1	207788205	207788237	33		Intergenic	
chr1	207788244	207788275	32		Intergenic	
chr1	207789550	207789648	99		Intergenic	
chr1	207796706	207796768	63		Intergenic	
chr1	207796846	207796878	33		Intergenic	
chr1	207796931	207796984	54		Intergenic	
chr1	207797241	207797265	25		Intergenic	
chr1	207797307	207797344	38		Intergenic	
chr1	207797431	207797455	25		Intergenic	
chr1	207797537	207797562	26		Intergenic	
chr1	207801420	207801522	103		Intergenic	
chr1	207801526	207801551	26		Intergenic	
chr1	207801554	207801598	45		Intergenic	
chr1	207801625	207801669	45		Intergenic	
chr1	207801704	207801756	53		Intergenic	
chr1	207802642	207802673	32		Intergenic	
chr1	207804077	207804109	33		Intergenic	
chr1	207804937	207805044	108		Intergenic	
chr1	207809952	207809985	34		Intergenic	
chr1	207815270	207815294	25		Intergenic	
chr1	207816864	207816889	26		Intergenic	
chr1	207818596	207818624	29		Intergenic	
chr1	207823399	207823430	32		Intergenic	
chr1	207823554	207823714	161	CAMK1G	**5`UTR**	
chr1	207824257	207824283	27	CAMK1G	Intron 1	
chr1	207824412	207824438	27	CAMK1G	Intron 1	
chr1	207824532	207824589	58	CAMK1G	Intron 1	
chr1	207824685	207824711	27	CAMK1G	Intron 1	
chr1	207824841	207824874	34	CAMK1G	Intron 1	
chr1	207825701	207825729	29	CAMK1G	Intron 1	
chr1	207825732	207825757	26	CAMK1G	Intron 1	
chr1	207825769	207825835	67	CAMK1G	Intron 1	
chr1	207825862	207825889	28	CAMK1G	Intron 1	
chr1	207827641	207827665	25	CAMK1G	Intron 1	
chr1	207827849	207827886	38	CAMK1G	Intron 1	
chr1	207827888	207827916	29	CAMK1G	Intron 1	
chr1	207828777	207828805	29	CAMK1G	Intron 1	
chr1	207830410	207830556	147	CAMK1G	Intron 1	
chr1	207834792	207834817	26	CAMK1G	Intron 1	
chr1	207834926	207835061	136	CAMK1G	**Exon 2**	
chr1	207835888	207835913	26	CAMK1G	Intron 2	
chr1	207835939	207835966	28	CAMK1G	Intron 2	
chr1	207836637	207836663	27	CAMK1G	Intron 2	
chr1	207839947	207840095	149	CAMK1G	**Exon 3**	
chr1	207840330	207840365	36	CAMK1G	Intron 3	
chr1	207843172	207843273	102	CAMK1G	**Exon 4**	
chr1	207845494	207845653	160	CAMK1G	**Exon 5**	
chr1	207846278	207846423	146	CAMK1G	**Exon 6**	
chr1	207847819	207847915	97	CAMK1G	**Exon 7**	
chr1	207848968	207849072	105	CAMK1G	**Exon 8**	
chr1	207849174	207849209	36	CAMK1G	Intron 8	

chr1	207849687	207849792	106	CAMK1G	Intron 8
chr1	207849810	207849892	83	CAMK1G	**Exon 9**
chr1	207851428	207851533	106	CAMK1G	**Exon 10**
chr1	207851746	207851858	113	CAMK1G	**Exon 11**
chr1	207851884	207851923	40	CAMK1G	**Exon 11**
chr1	207851998	207852072	75	CAMK1G	**Exon 11**
chr1	207852079	207852121	43	CAMK1G	**Exon 11**
chr1	207852124	207852196	73	CAMK1G	**Exon 11**
chr1	207852752	207852777	26	CAMK1G	**Exon 12**
chr1	207852794	207852851	58	CAMK1G	**Exon 12**
chr1	207853856	207853928	73	CAMK1G	**3`UTR**
chr1	207854856	207854892	37	LAMB3	**3`UTR**
chr1	207855236	207855326	91	LAMB3	**Exon 22**
chr1	207855352	207855396	45	LAMB3	**Exon 22**
chr1	207856431	207856497	67	LAMB3	**Exon 21**
chr1	207856535	207856602	68	LAMB3	**Exon 21**
chr1	207857541	207857570	30	LAMB3	**Exon 20**
chr1	207857869	207857953	85	LAMB3	**Exon 19**
chr1	207857958	207857983	26	LAMB3	**Exon 19**
chr1	207858419	207858443	25	LAMB3	**Exon 18**
chr1	207858454	207858545	92	LAMB3	**Exon 18**
chr1	207858547	207858638	92	LAMB3	**Exon 18**
chr1	207858872	207858896	25	LAMB3	Intron 17
chr1	207858922	207858960	39	LAMB3	Intron 17
chr1	207859016	207859043	28	LAMB3	Intron 17
chr1	207861480	207861507	28	LAMB3	Intron 17
chr1	207861566	207861592	27	LAMB3	Intron 17
chr1	207861599	207861627	29	LAMB3	Intron 17
chr1	207862500	207862541	42	LAMB3	**Exon 17**
chr1	207862543	207862607	65	LAMB3	**Exon 17**
chr1	207862943	207863017	75	LAMB3	**Exon 16**
chr1	207863058	207863083	26	LAMB3	**Exon 16**
chr1	207863224	207863250	27	LAMB3	Intron 15
chr1	207863467	207863503	37	LAMB3	**Exon 15**
chr1	207863575	207863612	38	LAMB3	**Exon 15**
chr1	207863627	207863654	28	LAMB3	**Exon 15**
chr1	207863675	207863701	27	LAMB3	**Exon 15**
chr1	207863803	207863860	58	LAMB3	**Exon 14**
chr1	207863879	207863907	29	LAMB3	**Exon 14**
chr1	207865140	207865167	28	LAMB3	Intron 13
chr1	207865273	207865313	41	LAMB3	Intron 13
chr1	207865636	207865664	29	LAMB3	**Exon 13**
chr1	207865825	207865954	130	LAMB3	**Exon 13**
chr1	207865971	207866003	33	LAMB3	**Exon 13**
chr1	207866834	207866902	69	LAMB3	**Exon 12**
chr1	207866905	207866954	50	LAMB3	**Exon 12**
chr1	207867347	207867396	50	LAMB3	**Exon 11**
chr1	207867401	207867518	118	LAMB3	**Exon 11**
chr1	207867526	207867554	29	LAMB3	**Exon 11**
chr1	207867790	207867819	30	LAMB3	Intron 10
chr1	207867991	207868170	180	LAMB3	**Exon 10**
chr1	207869506	207869533	28	LAMB3	Intron 8
chr1	207869698	207869726	29	LAMB3	**Exon 9**
chr1	207869732	207869901	170	LAMB3	**Exon 9**

chr1	207870581	207870606	26	LAMB3	**Exon 8**
chr1	207870610	207870711	102	LAMB3	**Exon 8**
chr1	207871136	207871173	38	LAMB3	Intron 7
chr1	207871175	207871206	32	LAMB3	Intron 7
chr1	207872604	207872749	146	LAMB3	**Exon 7**
chr1	207873029	207873072	44	LAMB3	**Exon 6**
chr1	207873077	207873106	30	LAMB3	**Exon 6**
chr1	207874441	207874488	48	LAMB3	**Exon 5**
chr1	207874515	207874598	84	LAMB3	**Exon 5**
chr1	207877073	207877101	29	LAMB3	Intron 4
chr1	207877107	207877136	30	LAMB3	Intron 4
chr1	207877287	207877314	28	LAMB3	Intron 4
chr1	207877605	207877641	37	LAMB3	Intron 4
chr1	207877897	207877969	73	LAMB3	**Exon 4**
chr1	207878496	207878539	44	LAMB3	**Exon 3**
chr1	207878541	207878633	93	LAMB3	**Exon 3**
chr1	207881708	207881756	49	LAMB3	Intron 2
chr1	207881781	207881823	43	LAMB3	Intron 2
chr1	207881868	207881927	60	LAMB3	Intron 2
chr1	207883043	207883076	34	LAMB3	Intron 2
chr1	207883312	207883336	25	LAMB3	Intron 2
chr1	207883382	207883443	62	LAMB3	Intron 2
chr1	207883470	207883504	35	LAMB3	Intron 2
chr1	207885060	207885091	32	LAMB3	Intron 2
chr1	207885778	207885802	25	LAMB3	Intron 2
chr1	207885957	207886049	93	LAMB3	Intron 2
chr1	207888446	207888499	54	LAMB3	Intron 2
chr1	207888559	207888593	35	LAMB3	Intron 2
chr1	207888959	207888983	25	LAMB3	Intron 2
chr1	207889925	207890068	144	LAMB3	**Exon 2**
chr1	207892416	207892535	120	LAMB3	**5`UTR**
chr1	207892538	207892562	25		Intergenic
chr1	207892880	207892907	28		Intergenic
chr1	207893095	207893121	27		Intergenic
chr1	207893167	207893216	50		Intergenic
chr1	207894019	207894055	37		Intergenic
chr1	207894069	207894093	25		Intergenic
chr1	207895926	207895955	30		Intergenic
chr1	207902976	207903001	26		Intergenic
chr1	207909098	207909128	31		Intergenic
chr1	207913313	207913348	36		Intergenic
chr1	207913388	207913412	25		Intergenic
chr1	207913607	207913642	36		Intergenic
chr1	207913674	207913698	25		Intergenic
chr1	207914025	207914085	61		Intergenic
chr1	207914678	207914712	35		Intergenic
chr1	207915652	207915708	57	G0S2	**Exon 2**
chr1	207915714	207915817	104	G0S2	**Exon 2**
chr1	207915931	207915966	36	G0S2	**Exon 2**
chr1	207916172	207916206	35	G0S2	**3`UTR**
chr1	207916278	207916302	25	G0S2	**3`UTR**
chr1	207916313	207916345	33	G0S2	**3`UTR**
chr1	207923818	207923843	26		Intergenic
chr1	207924718	207924743	26		Intergenic

chr1	207925789	207925848	60		Intergenic
chr1	207925992	207926182	191	HSD11B1	**5`UTR**
chr1	207926395	207926421	27	HSD11B1	Intron 1
chr1	207926515	207926574	60	HSD11B1	Intron 1
chr1	207927152	207927190	39	HSD11B1	Intron 1
chr1	207927313	207927367	55	HSD11B1	Intron 1
chr1	207927440	207927464	25	HSD11B1	Intron 1
chr1	207927490	207927517	28	HSD11B1	Intron 1
chr1	207927732	207927758	27	HSD11B1	Intron 1
chr1	207927809	207927833	25	HSD11B1	Intron 1
chr1	207933050	207933086	37	HSD11B1	Intron 1
chr1	207933442	207933466	25	HSD11B1	Intron 1
chr1	207938125	207938187	63	HSD11B1	Intron 1
chr1	207943482	207943507	26	HSD11B1	Intron 1
chr1	207943514	207943585	72	HSD11B1	Intron 1
chr1	207944194	207944223	30	HSD11B1	Intron 1
chr1	207944233	207944264	32	HSD11B1	Intron 1
chr1	207944271	207944319	49	HSD11B1	Intron 1
chr1	207944369	207944421	53	HSD11B1	Intron 1
chr1	207944445	207944469	25	HSD11B1	Intron 1
chr1	207944597	207944886	290	HSD11B1	Intron 1
chr1	207944911	207944948	38	HSD11B1	**Exon 2**
chr1	207944952	207944982	31	HSD11B1	**Exon 2**
chr1	207944984	207945008	25	HSD11B1	**Exon 2**
chr1	207945771	207945892	122	HSD11B1	**Exon 3**
chr1	207946675	207946798	124	HSD11B1	**Exon 4**
chr1	207946898	207946951	54	HSD11B1	**Exon 5**
chr1	207946986	207947089	104	HSD11B1	**Exon 5**
chr1	207948394	207948472	79	HSD11B1	Intron 5
chr1	207948515	207948578	64	HSD11B1	Intron 5
chr1	207963094	207963119	26	HSD11B1	Intron 5
chr1	207969333	207969362	30	HSD11B1	Intron 5
chr1	207972418	207972488	71	HSD11B1	**Exon 6**
chr1	207972498	207972562	65	HSD11B1	**Exon 6**
chr1	207974271	207974301	31	HSD11B1	**Exon 7**
chr1	207974311	207974389	79	HSD11B1	**Exon 7**
chr1	207974419	207974443	25	HSD11B1	**Exon 7**
chr1	207974781	207974838	58	HSD11B1	**3`UTR**
chr1	207987979	207988013	35		Intergenic
chr1	207995964	207996020	57		Intergenic
chr1	207996028	207996063	36	TRAF3IP3	**5`UTR**
chr1	207996065	207996120	56	TRAF3IP3	**5`UTR**
chr1	207996199	207996245	47	TRAF3IP3	**5`UTR**
chr1	207996275	207996305	31	TRAF3IP3	**5`UTR**
chr1	207997859	207997883	25	TRAF3IP3	Intron 1
chr1	207998521	207998553	33	TRAF3IP3	Intron 1
chr1	207999880	208000159	280	TRAF3IP3	**Exon 3**
chr1	208000165	208000192	28	TRAF3IP3	**Exon 3**
chr1	208000194	208000235	42	TRAF3IP3	**Exon 3**
chr1	208001204	208001241	38	TRAF3IP3	Intron 3
chr1	208001252	208001280	29	TRAF3IP3	Intron 3
chr1	208002601	208002645	45	TRAF3IP3	**Exon 4**
chr1	208002763	208002858	96	TRAF3IP3	**Exon 5**
chr1	208003022	208003071	50	TRAF3IP3	**Exon 6**

chr1	208003311	208003374	64	TRAF3IP3	**Exon 7**	
chr1	208003453	208003572	120	TRAF3IP3	**Exon 8**	
chr1	208006254	208006296	43	TRAF3IP3	Intron 8	
chr1	208008512	208008652	141	TRAF3IP3	Intron 8	
chr1	208008818	208008843	26	TRAF3IP3	Intron 8	
chr1	208008901	208008934	34	TRAF3IP3	Intron 8	
chr1	208009376	208009413	38	TRAF3IP3	Intron 8	
chr1	208009423	208009622	200	TRAF3IP3	Intron 8	
chr1	208012884	208013006	123	TRAF3IP3	**Exon 9**	
chr1	208013991	208014030	40	TRAF3IP3	Intron 9	
chr1	208015277	208015468	192	TRAF3IP3	**Exon 10**	
chr1	208015551	208015729	179	TRAF3IP3	**Exon 11**	
chr1	208017307	208017472	166	TRAF3IP3	**Exon 12**	
chr1	208017755	208017794	40	TRAF3IP3	Intron 12	**MCS+28.3**
chr1	208017915	208017943	29	TRAF3IP3	Intron 12	**MCS+28.3**
chr1	208018058	208018177	120	TRAF3IP3	**Exon 13**	
chr1	208019202	208019227	26	TRAF3IP3	Intron 13	
chr1	208019237	208019384	148	TRAF3IP3	**Exon 14**	
chr1	208019395	208019424	30	TRAF3IP3	**Exon 14**	
chr1	208019535	208019565	31	TRAF3IP3	**Exon 14**	
chr1	208019580	208019652	73	TRAF3IP3	**Exon 14**	
chr1	208019654	208019678	25	TRAF3IP3	Intron 14	
chr1	208020412	208020466	55	TRAF3IP3	**Exon 15**	
chr1	208020472	208020594	123	TRAF3IP3	**Exon 15**	
chr1	208020808	208020833	26	TRAF3IP3	Intron 15	
chr1	208021304	208021416	113	TRAF3IP3	**Exon 16**	
chr1	208021420	208021444	25	TRAF3IP3	**Exon 16**	
chr1	208022022	208022110	89	TRAF3IP3	**Exon 17**	
chr1	208022787	208022906	120	C1orf74	**Exon 2**	
chr1	208022908	208022958	51	C1orf74	**Exon 2**	
chr1	208022973	208023127	155	C1orf74	**Exon 2**	
chr1	208023155	208023238	84	C1orf74	**Exon 2**	
chr1	208023240	208023362	123	C1orf74	**Exon 2**	
chr1	208023377	208023501	125	C1orf74	**Exon 2**	
chr1	208023517	208023586	70	C1orf74	**Exon 2**	
chr1	208025534	208025558	25		Intergenic	
chr1	208025730	208025754	25		Intergenic	
chr1	208025763	208025788	26		Intergenic	
chr1	208027706	208027778	73	IRF6	**3`UTR**	**3` of ex9**
chr1	208027805	208027905	101	IRF6	**3`UTR**	**3` of ex9**
chr1	208027929	208027983	55	IRF6	**3`UTR**	**3` of ex9**
chr1	208028208	208028245	38	IRF6	**3`UTR**	
chr1	208028299	208028366	68	IRF6	**3`UTR**	
chr1	208028388	208028635	248	IRF6	**Exon 9**	
chr1	208029618	208029772	155	IRF6	**Exon 8**	
chr1	208030448	208030881	434	IRF6	**Exon 7**	
chr1	208030908	208031027	120	IRF6	Intron 6	**Intron 6**
chr1	208031399	208031430	32	IRF6	Intron 6	**Intron 6**
chr1	208032062	208032091	30	IRF6	Intron 6	**Intron 6**
chr1	208032223	208032411	189	IRF6	**Exon 6**	
chr1	208033080	208033104	25	IRF6	Intron 5	
chr1	208035243	208035405	163	IRF6	**Exon 5**	
chr1	208036291	208036537	247	IRF6	**Exon 4**	
chr1	208041187	208041395	209	IRF6	**Exon 3**	

chr1	208041417	208041445	29	IRF6	Intron 2	**Intron 2**
chr1	208041470	208041533	64	IRF6	Intron 2	**Intron 2**
chr1	208041535	208041702	168	IRF6	Intron 2	**Intron 2**
chr1	208041738	208041792	55	IRF6	Intron 2	**Intron 2**
chr1	208042136	208042160	25	IRF6	Intron 1	**Intron 1e**
chr1	208042285	208042329	45	IRF6	Intron 1	**Intron 1d**
chr1	208042460	208042492	33	IRF6	Intron 1	**Intron 1c**
chr1	208042523	208042552	30	IRF6	Intron 1	**Intron 1b**
chr1	208042896	208043008	113	IRF6	Intron 1	**Intron 1b**
chr1	208045290	208045348	59	IRF6	Intron 1	**Intron 1a**
chr1	208045353	208045380	28	IRF6	Intron 1	**Intron 1a**
chr1	208045404	208045428	25	IRF6	Intron 1	**Intron 1a**
chr1	208046058	208046115	58	IRF6	**5`UTR**	**Promoter**
chr1	208046147	208046180	34		Intergenic	**Promoter**
chr1	208048220	208048249	30		Intergenic	**MCS-2**
chr1	208048321	208048346	26		Intergenic	**MCS-2**
chr1	208049711	208049764	54		Intergenic	**MCS-3.6**
chr1	208054851	208054881	31		Intergenic	**MCS-8.7**
chr1	208055903	208055934	32		Intergenic	**MCS-9.7**
chr1	208055954	208055982	29		Intergenic	**MCS-9.7**
chr1	208055984	208056015	32		Intergenic	**MCS-9.7**
chr1	208056021	208056086	66		Intergenic	**MCS-9.7**
chr1	208056089	208056117	29		Intergenic	**MCS-9.7**
chr1	208056146	208056264	119		Intergenic	**MCS-9.7**
chr1	208056545	208056572	28		Intergenic	
chr1	208057206	208057241	36		Intergenic	**MCS-11**
chr1	208057248	208057297	50		Intergenic	**MCS-11**
chr1	208057300	208057345	46		Intergenic	**MCS-11**
chr1	208063651	208063675	25		Intergenic	
chr1	208065985	208066019	35		Intergenic	**MCS-20**
chr1	208066036	208066069	34		Intergenic	**MCS-20**
chr1	208066184	208066229	46		Intergenic	**MCS-20**
chr1	208067893	208067921	29		Intergenic	**MCS-20**
chr1	208068024	208068144	121	C10orf107	**Exon 1**	
chr1	208070062	208070126	65	C10orf107	**Exon 2**	
chr1	208070758	208070836	79	C10orf107	**Exon 3**	
chr1	208070838	208070887	50	C10orf107	**Exon 3**	
chr1	208070892	208070918	27	C10orf107	**Exon 3**	
chr1	208073228	208073300	73	C10orf107	**Exon 4**	
chr1	208075028	208075123	96	C10orf107	**Exon 5**	
chr1	208076792	208076933	142	C10orf107	**Exon 6**	
chr1	208076936	208077014	79	C10orf107	**Exon 6**	
chr1	208077016	208077199	184	C10orf107	**Exon 6**	
chr1	208078876	208079106	231	C10orf107	**Exon 7**	
chr1	208080818	208081038	221	C10orf107	**Exon 8**	
chr1	208081250	208081276	27	C10orf107	Intron 7	
chr1	208081911	208081936	26	C10orf107	Intron 7	
chr1	208082226	208082405	180	C10orf107	**Exon 9**	
chr1	208082541	208082672	132	C10orf107	**Exon 10**	
chr1	208083410	208083675	266	C10orf107	**Exon 11**	
chr1	208085367	208085422	56	C10orf107	Intron 11	
chr1	208090660	208090684	25	C10orf107	Intron 11	
chr1	208090759	208090783	25	C10orf107	Intron 11	
chr1	208091167	208091226	60	C10orf107	Intron 11	

chr1	208091234	208091378	145	C10orf107	**Exon 12**	
chr1	208091381	208091417	37	C10orf107	**3`UTR**	
chr1	208093906	208093930	25	C10orf107	**3`UTR**	
chr1	208094401	208094439	39	C10orf107	**3`UTR**	
chr1	208095674	208095710	37	C10orf107	**3`UTR**	
chr1	208095781	208095822	42	C10orf107	**3`UTR**	
chr1	208100450	208100474	25		Intergenic	
chr1	208110108	208110224	117		Intergenic	
chr1	208110613	208110639	27		Intergenic	
chr1	208114828	208114853	26		Intergenic	
chr1	208115780	208115839	60		Intergenic	**MCS-70**
chr1	208115852	208115925	74		Intergenic	**MCS-70**
chr1	208117024	208117068	45		Intergenic	
chr1	208117089	208117161	73		Intergenic	
chr1	208117274	208117333	60		Intergenic	
chr1	208118904	208118930	27		Intergenic	
chr1	208119560	208119587	28		Intergenic	
chr1	208119594	208119620	27		Intergenic	
chr1	208119791	208119841	51		Intergenic	
chr1	208119843	208119935	93		Intergenic	
chr1	208119973	208120068	96		Intergenic	
chr1	208120134	208120166	33		Intergenic	
chr1	208123752	208123815	64		Intergenic	
chr1	208123820	208123961	142		Intergenic	
chr1	208124932	208124977	46		Intergenic	
chr1	208125027	208125176	150		Intergenic	
chr1	208130888	208130985	98		Intergenic	
chr1	208131005	208131099	95		Intergenic	
chr1	208131116	208131141	26		Intergenic	
chr1	208131167	208131201	35		Intergenic	
chr1	208131255	208131279	25		Intergenic	
chr1	208140659	208140683	25		Intergenic	
chr1	208141344	208141382	39		Intergenic	
chr1	208141562	208141592	31		Intergenic	
chr1	208141632	208141662	31		Intergenic	
chr1	208141701	208141748	48		Intergenic	
chr1	208142717	208142748	32		Intergenic	
chr1	208149761	208149807	47		Intergenic	
chr1	208154475	208154507	33		Intergenic	
chr1	208155316	208155341	26		Intergenic	
chr1	208162548	208162578	31		Intergenic	
chr1	208163109	208163137	29		Intergenic	
chr1	208166239	208166277	39		Intergenic	
chr1	208166310	208166397	88		Intergenic	
chr1	208166402	208166435	34		Intergenic	
chr1	208166536	208166567	32		Intergenic	
chr1	208167018	208167042	25		Intergenic	
chr1	208167376	208167402	27		Intergenic	
chr1	208167444	208167480	37		Intergenic	
chr1	208167553	208167584	32		Intergenic	
chr1	208174860	208174886	27		Intergenic	
chr1	208174975	208174999	25		Intergenic	
chr1	208175046	208175083	38		Intergenic	
chr1	208175184	208175234	51		Intergenic	

chr1	208176970	208176999	30		Intergenic
chr1	208177001	208177078	78		Intergenic
chr1	208178207	208178255	49	sytXIV	**Exon 1**
chr1	208182285	208182310	26	sytXIV	Intron 1
chr1	208183980	208184004	25	sytXIV	Intron 1
chr1	208188414	208188445	32	sytXIV	Intron 1
chr1	208188480	208188510	31	sytXIV	Intron 1
chr1	208192660	208192740	81	sytXIV	**Exon 2**
chr1	208192928	208192968	41	sytXIV	Intron 2
chr1	208192972	208193117	146	sytXIV	intron 2
chr1	208193129	208193212	84	sytXIV	Intron 2
chr1	208193214	208193311	98	sytXIV	Intron 2
chr1	208193799	208193823	25	sytXIV	Intron 2
chr1	208202374	208202412	39	sytXIV	Intron 2
chr1	208202516	208202567	52	sytXIV	Intron 2
chr1	208202579	208202604	26	sytXIV	Intron 2
chr1	208229701	208229746	46	sytXIV	Intron 3
chr1	208230500	208230543	44	sytXIV	Intron 3
chr1	208246492	208246553	62	sytXIV	Intron 3
chr1	208251644	208251668	25	sytXIV	Intron 3
chr1	208253595	208253737	143	sytXIV	**Exon 4**
chr1	208253740	208253775	36	sytXIV	Intron 4
chr1	208255607	208255647	41	sytXIV	Intron 4
chr1	208255666	208255707	42	sytXIV	Intron 4
chr1	208255744	208255771	28	sytXIV	Intron 4
chr1	208255868	208255900	33	sytXIV	Intron 4
chr1	208255916	208255956	41	sytXIV	Intron 4
chr1	208255958	208255985	28	sytXIV	Intron 4
chr1	208257054	208257086	33	sytXIV	Intron 4
chr1	208260994	208261065	72	sytXIV	**Exon 5**
chr1	208261082	208261107	26	sytXIV	**Exon 5**
chr1	208266579	208266616	38	sytXIV	Intron 5

*Coordinates are given according to UCSC Human Genome Browser Bld 36.1 (March 2006)

Multiple and closely arranged MCSs were amplified in single amplicons (see **Supplementary Table 4 for sequenced regions and primers used)

IRF6 coding sequences are highlighted in yellow

MCS-9.7 is highlighted in red

Supplementary Table 2: Worldwide distribution of rs642961 allele frequencies determined from CEHP Human Genome Diversity Panel, http://www.cephb.fr/HGDP-CEPH-Panel/

Region	Population	rs642961-G	rs642961-A	Total	G-freq	A-freq
Central African Republic	Biaka Pygmies	67	3	70	0.96	0.04
Democratic Republic of Congo	Mbuti Pygmies	18	12	30	0.60	0.40
Senegal	Mandenka	40	6	46	0.87	0.13
Nigeria	Yoruba	48	2	50	0.96	0.04
Kenya	Bantu N.E.	24	0	24	1.00	0.00
Namidia	San	14	0	14	1.00	0.00
South Africa	Bantu	15	1	16	0.94	0.06
Algeria (Mzab)	Mozabite	50	10	60	0.83	0.17
AFRICA		**276**	**34**	**310**	**0.89**	**0.11**
Israel	Bedouin	79	19	98	0.81	0.19
Israel	Druze	79	15	94	0.84	0.16
Israel	Palestinian	84	18	102	0.82	0.18
Pakistan	Brahui	34	16	50	0.68	0.32
Pakistan	Balochi	40	10	50	0.80	0.20
Pakistan	Hazara	42	8	50	0.84	0.16
Pakistan	Makrani	40	10	50	0.80	0.20
Pakistan	Sindhi	38	12	50	0.76	0.24
Pakistan	Pathan	39	11	50	0.78	0.22
Pakistan	Kalash	35	15	50	0.70	0.30
Pakistan	Barusho	39	11	50	0.78	0.22
PAKISTAN		**307**	**93**	**400**	**0.77**	**0.23**
China	Han	70	18	88	0.80	0.20
China	Tujia	13	7	20	0.65	0.35
China	Yizu	19	1	20	0.95	0.05
China	Miaozu	16	4	20	0.80	0.20
China	Oroqen	17	3	20	0.85	0.15
China	Daur	17	3	20	0.85	0.15
China	Mongola	17	3	20	0.85	0.15
China	Hezhen	16	4	20	0.80	0.20
China	Xibo	13	6	19	0.68	0.32
China	Uygur	17	3	20	0.85	0.15
China	Lahu	16	4	20	0.80	0.20
China	She	17	3	20	0.85	0.15
China	Naxi	16	4	20	0.80	0.20
China	Tu	14	6	20	0.70	0.30
CHINA		**278**	**69**	**347**	**0.80**	**0.20**
Siberia	Yakut	35	13	48	0.73	0.27
Japan	Japanese	54	8	62	0.87	0.13
Cambodia	Cambodian	18	4	22	0.82	0.18
New Guinea	Papuan	26	8	34	0.76	0.24
Bougainville	NAN Melanesian	31	13	44	0.70	0.30
France	Frence	49	9	58	0.84	0.16
France	French Basque	30	10	40	0.75	0.25
Italy	Sardinian	44	12	56	0.79	0.21
Italy	North Italian	27	1	28	0.96	0.04
Italy	Tuscan	11	5	16	0.69	0.31
Orkney Islands	Orcadian	28	4	32	0.88	0.13
Russia Caucasus	Adygei	24	6	30	0.80	0.20
Russia	Russian	42	6	48	0.88	0.13
Colombia	Colombian	23	3	26	0.88	0.12
Mexico	Pima	37	13	50	0.74	0.26
Mexico	Maya	35	15	50	0.70	0.30
Brazil	Karitiana	34	14	48	0.71	0.29
Brazil	Surui	32	10	42	0.76	0.24
NATIVE AMERICAN		**138**	**52**	**190**	**0.73**	**0.27**

Supplementary Table 3: *IRF6* locus coordinates and genome assemblies

Organism	UCSC Genome Browser coordinates	Build
Human	chr1:207786994-208286993	March 2006
Chimp	chr1:189982478-190486018	March 2006
Rhesus	chr1:160355582-160859543	January 2006
Rat	chr13:108826445-109255967	November 2004
Mouse	chr1:194795892-195233688	July 2006
Dog	chr7:11156498-11636370	May 2005
Cow	chr16:67677502-67941385	August 2006
Cat	scaffold_129274:24555-91072	March 2006
Tenrec	scaffold_124:63192-75668 (*Ensemble*)	October 2007
Opossum	chr2:116773085-117442209	January 2006
Chicken	chr3:23923223-24003837	May 2006
	chr26:2926576-2995482	May 2006
Frog	scaffold_375:789749-797232	August 2005
	scaffold_61:1446322-1455782	August 2005
Tetraodon	chr11:5978227-5980600	February 2004

UCSC Genome Browser: http://www.genome.ucsc.edu/
Ensemble: http://www.ensembl.org/index.html

BAC clones from organisms sequenced by the NIH Intramural Sequencing Center (NISC)

Organism	Clone name	GenBank ID
Rabbit	LB1-121D24	AC173439
Rabbit	LB1-183A3	AC164495
Rabbit	LB1-183H5	AC163952
Rabbit	LB1-289D12	AC166022
Pig	RP44-357O11	AC164944
Pig	RP44-274O19	AC166031
Pig	RP44-55L3	AC167303
Pig	RP44-327E16	AC164503
Bat (greater horseshoe)	VMRC7-303M11	AC163982
Bat (greater horseshoe)	VMRC7-280A9	AC164951
Bat (greater horseshoe)	VMRC7-285O13	AC165214
Bat (greater horseshoe)	VMRC7-212E18	AC167528
Armadillo (nine-banded)	VMRC5-471F9	AC169168
Armadillo (nine-banded)	VMRC5-148P11	AC171392
Armadillo (nine-banded)	VMRC5-151G4	AC164515
Elephant (African)	VMRC15-220O23	AC166032
Elephant (African)	VMRC15-370J11	AC164528
Elephant (African)	VMRC15-554C21	AC164513
Elephant (African)	VMRC15-515M5	AC164512

BAC sequence data can also be accessed at:
http://www.nisc.nih.gov/open_page.cgi?path=/projects/comp_seq.html

Supplementary Table 4: Primers used to amplify *IRF6* Multispecies Conserved Sequences (MCSs)

MCS	Forward Primer	Reverse Primer	UCSC hg18 coordinates
-70	CATCAGAAAGGAGGGAGCAG	AAGGGCATTGCCAAAATATG	chr1:208115704-208115964
-20	GGGTGGGGCTAGGGTAAAT	ATTCCTAATTAGATTCGGAGAGG	chr1:208065913-208066382
-11	TATACCCCTGCCTTGACCTG	AGAGTGTGGTGGGGATTTCA	chr1:208057098-208057400
-9.7	**CTCCTCTGTTTGCTTAGCTTACCTC**	**CTGGTAAATGGTGAGTAGGAAGTTG**	**chr1:208055755-208056361**
-8.7	ATGTCCTGTGCATGACTCCA	GTCACCCCAGGGAATTAAGC	chr1:208054738-208055007
-3.6	CCACAAGCTCTTGAGGAATCA	GAGAGTGGGATTTGGGTTCA	chr1:208049592-208049937
-2	GTTGAAGAGCCCCAGATTGA	TCCAGGGCACTTGACTCTTT	chr1:208048091-208048702
Promoter	CGTACCCACCTTTGCTCAAT	CCAACCCAGAATCTCAAAACC	chr1:208045906-208046469
Intron 1a	TGAGTGCCACTGTCATACATC	CCAGGAGCTTTGGAGCATT	chr1:208045041-208045487
Intron 1b	GGGCCAATAAAGTGATTGAGG	GCGTGTGTTTTAACCTGAGC	chr1:208044398-208044837
Intron 1c	TAAGGCAATCAGGCCAAGAC	CAGTCAAGTTGTATGCTGTTTCC	chr1:208043409-208043868
Intron 1d	AGGTCACCTTCATACCCAGAAA	TAGTGATAGGACTGGGGAGCA	chr1:208042669-208043188
Intron 1e	GAGAAGGCTTGGGAACACTG	CATTGTGGGGCTTGAGAAGT	chr1:208042070-208042643
Intron 2	AGGGCTTTAGCCGGACTCT	CCCATTCACCTTTAGTTGTCTTG	chr1:208041345-208041856
Intron 6	AGAGTCGGCAGCCCTGAG	TTTAGCTCCTTCCAATGACCA	chr1:208030769-208031193
3` of ex9	AGGCTAAACTGCCTGGAACA	CTGTCACCTACCCCCTCTCC	chr1:208027559-208027969
28.3	TTTTAGCCCATCCCTCTGG	TCTTCGTGTCCCAGTTCTGA	chr1:208017687-208017948

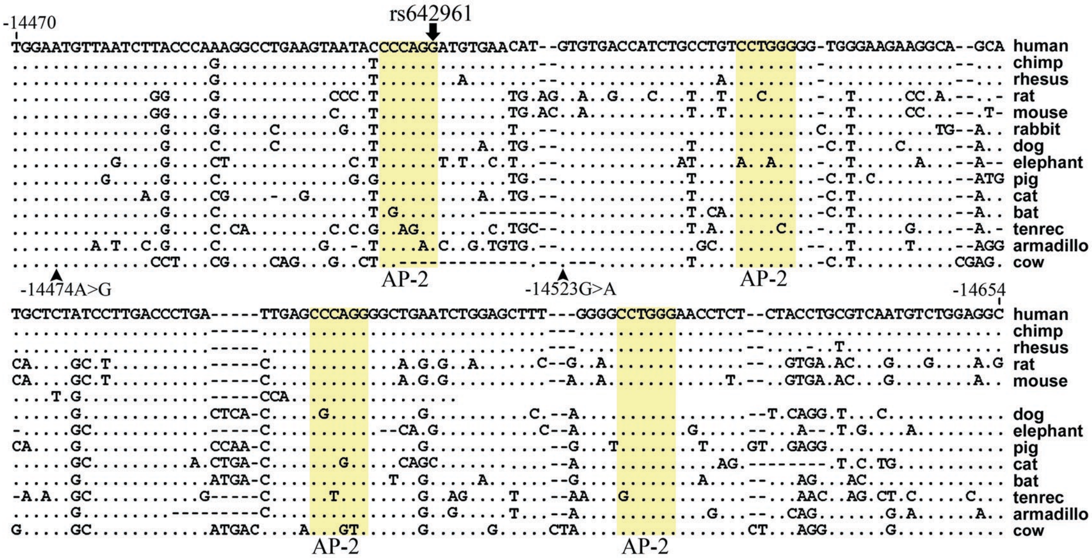
UCSC Genome Browser: http://www.genome.ucsc.edu/

MCSs are identified by their distance in kilobases from *IRF6* transcription start site (chr1:208046102 UCSC hg18)

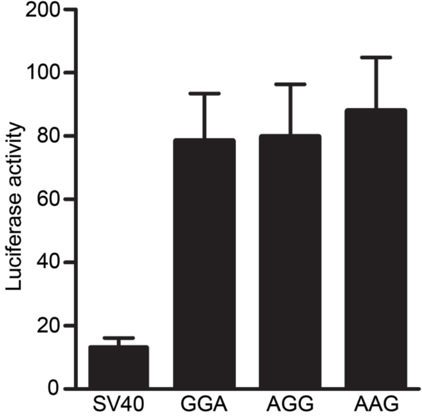
Primers used in Chromatin Immunoprecipitation (ChIP) assay

Amplicon	Forward Primer	Reverse Primer	UCSC hg18 coordinates
Target in MCS-9.7	GATTGGAGCTTTGGAATGTTAATC	ACAGAATAGTAACTGGGAGGTTGC	chr1:208055840-208056093
Control region	GAATAATTTGTGACAACTGCCTTG	AATTTGAGGAGAGTAGACCACCAG	chr1:208022115-208022373

Supplementary Figure 1 Legend: Multispecies sequence alignment of the MCS-9.7 segment corresponding to nucleotides from –14470 to –14654 with respect to *IRF6* translation initiation site, that contains four putative AP-2α binding sites (highlighted in yellow) predicted by the TESS software http://www.cbil.upenn.edu/tess. Variants -14474A>G, rs642961 (G>A) and -14523G>A, are indicated by vertical arrows.


rs642961
-14470
-14474A>G
-14523G>A
-14654
AP-2
AP-2
AP-2
AP-2
human
chimp
rhesus
rat
mouse
rabbit
dog
elephant
pig
cat
bat
tenrec
armadillo
cow

Supplementary Figure 2. Analysis of MCS-9.7 regulatory activity in HFK cell line using luciferase reporter assay. Bars represent mean ± SEM of the luciferase activity data from seven independent experiments performed in triplicates (21 data points total). GGA and AGG are the non-associated haplotype constructs; AAG is the risk haplotype construct. SV40, luciferase expression driven by the SV40 promoter alone.


200
100
80
60
40
20
0
Luciferase activity
SV40
GGA
AGG
AAG