Restless Legs Syndrome-associated intronic common-variant in *Meis1* alters enhancer function in the developing telencephalon

Supplemental Material

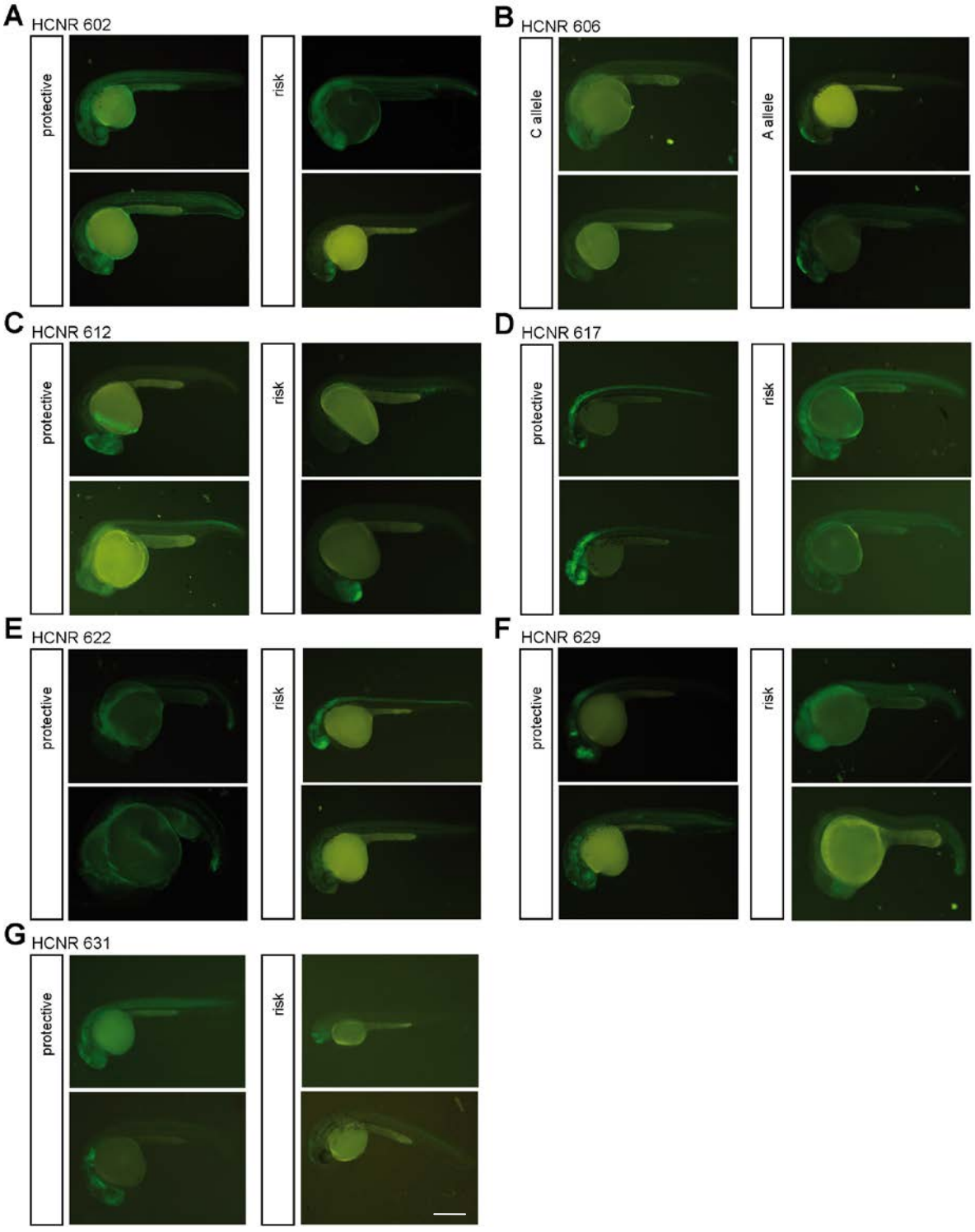
Derek Spieler[1, 25], Maria Kaffe[1,2, 25], Franziska Knauf[1, 25], José Bessa[3], Juan J. Tena[3], Florian Giesert[4], Barbara Schormair[1,2], Erik Tilch[1,2], Heekyoung Lee[5,6,7,8], Marion Horsch[9], Darina Czamara[10], Nazanin Karbalai[10], Christine von Toerne[11], Melanie Waldenberger[12], Christian Gieger[12], Peter Lichtner[1], Melina Claussnitzer[5,6,7,8,13], Ronald Naumann[14], Bertram Müller-Myhsok[10,15,16], Miguel Torres[17], Lillian Garrett[18], Jan Rozman[9,7,19], Martin Klingenspor[5,6,9,19], Valérie Gailus-Durner[9], Helmut Fuchs[9], Martin Hrabě de Angelis[9,7,20], Johannes Beckers[7,9,20], Sabine M. Hölter[18], Thomas Meitinger[1,2, 15], Stefanie M. Hauck[7,11], Helmut Laumen[5,6,7,8,9], Wolfgang Wurst[4,10,15,18,21,22], Fernando Casares[3], Jose Luis Gómez-Skarmeta[3], Juliane Winkelmann[1,15,23,24,26]

Supplemental Figures

Supplemental Tables

Supplemental Experimental Methods

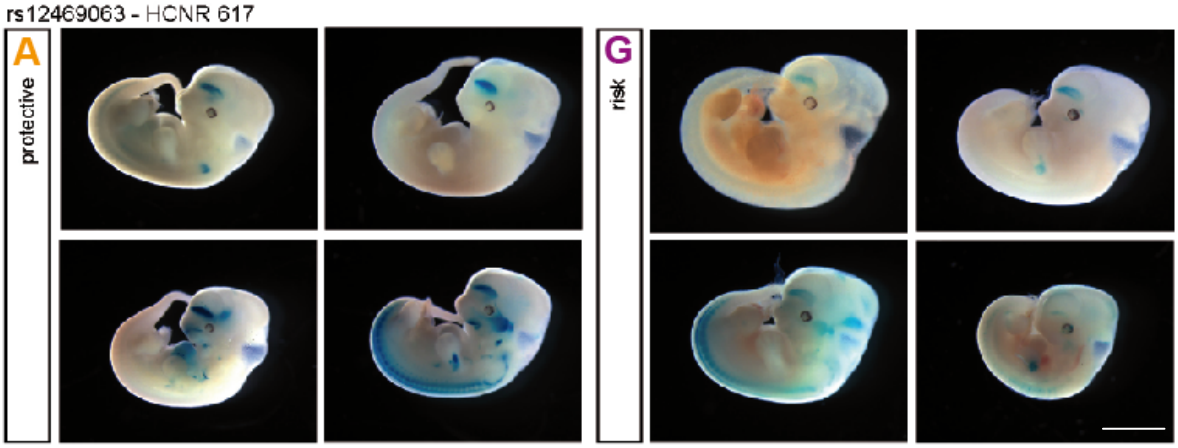
Figure S1: Results of the enhancer screen in zebrafish.



Representative transgenic fish ~24 hpf. (A) Booster activity of HCNR 602 in zebrafish with the protective alleles (left) and the risk alleles (right), respectively. Green EGFP signal refers to variable regions in the fish. (B) Robust enhancer function of HCNR 606 in the midbrain. Representative transgenic fish with the A allele (left) and the C allele (right) in HCNR 606,

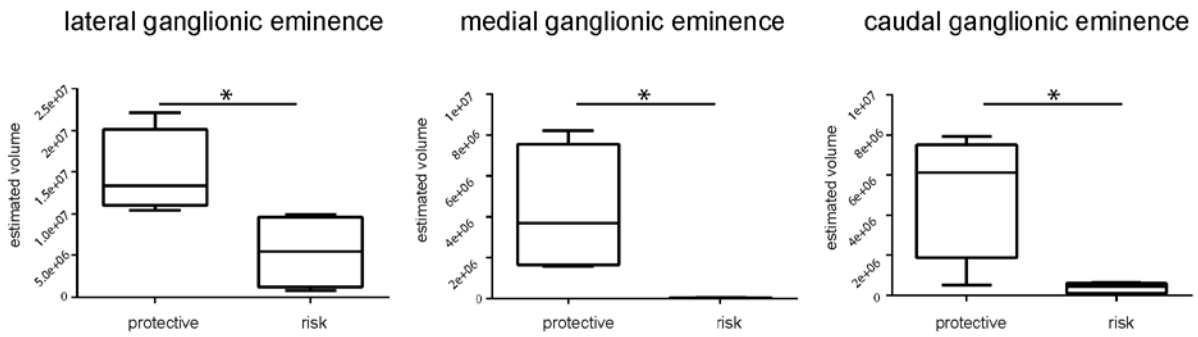
respectively. Green EGFP signal refers to the tectum. (C) Likely lack of enhancer function of HCNR 612. No reproducible pattern could be identified, although the results remained inconclusive due to the low number of identified founders. (D) Allele-specific enhancer HCNR 617 in retina, forebrain, hindbrain, and spinal cord. EGFP signal represents the regions of enhancer function of the protective allele construct – retina, forebrain, hindbrain and spinal cord. Fish in the right column displayed EGFP signal at background level. (E) Inconclusive enhancer function of HCNR 622 in hematopoietic stem cells. EGFP signal in the protective allele carriers was located in the hematopoietic stem cells. (F) Neural booster activity of HCNR 629. Green EGFP signal refers to variable regions in the central nervous system (CNS), risk allele carriers displayed EGFP only at background level. (G) Neural booster activity of HCNR 631. EGFP signal refers to variable regions in the CNS, risk allele carriers displayed EGFP only at background level (bar in G represents 250µm).

Figure S2: Theiler-staged transgenic mouse embryos for intensity analysis and volume estimation of the reporter signal.



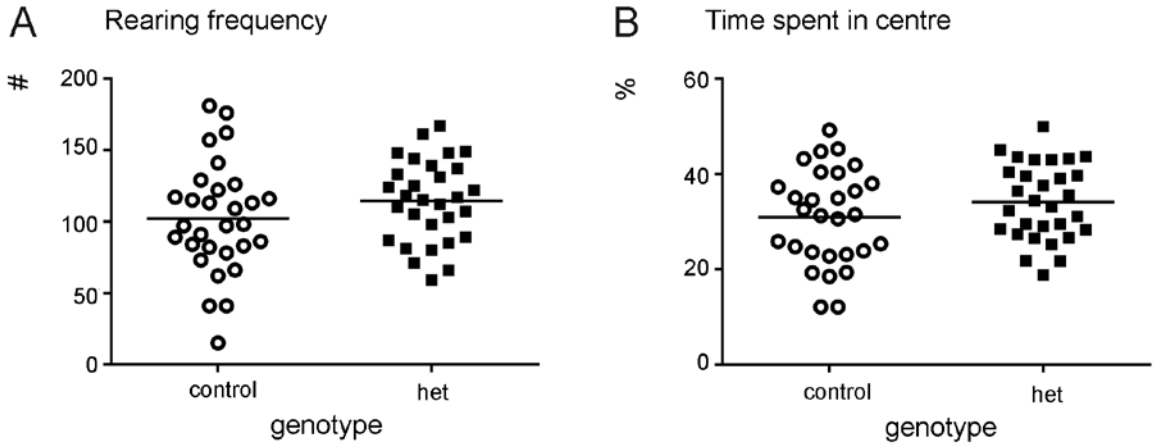
Transgenic embryos were staged according to Theiler (Theiler stages 19 and 20) (Theiler 1989). Blue color refers to regions expressing beta-galactosidase. Note the difference in signal intensity (compare left column, protective allele carriers, with right column, risk allele carriers) (bar represents 2,5 mm).

Figure S3: Volume estimations of the beta-galactosidase signal according to Cavalieri.



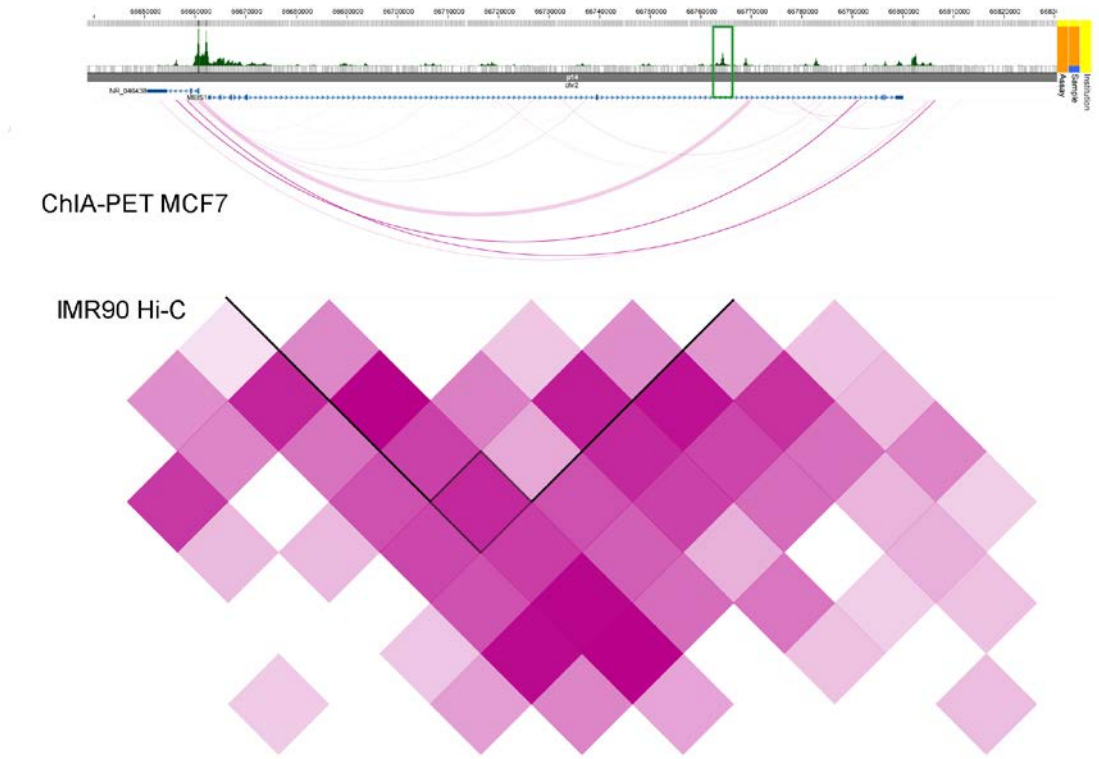
Analysis of transient transgenic mouse embryos via stereomicroscopy. The ganglionic eminences are split in lateral, medial and caudal. Asterisks refer to a significant difference in volume (Wilcoxon rank sum test, $P \leq 0.05$).

Figure S4: Missing difference in hyperactivity due to fear related behavior.



Rearing frequency (A) and the percentage of time spend in the aversive centre of the test arena (B), indicating anxiety related behavior, in a 20 min Open Field test for spontaneous locomotor activity in a novel enrichment. No differences could be detected between the genotypes (n.s.; n = 29-30).

Figure S5: Poll II ChIA-PET and Hi-C experiments in two human cell lines focusing on the RLS-associated intronic region.



Analysis of two human cell lines (human fetal lung cells, IMR90; breast cancer cells, MCF7) illustrates interaction (cf. three curved lines) between the 32 kb RLS-associated block and the promoter of *MEIS1*. Chromatin Interaction Analysis by Paired-End Tag Sequencing (Pol II ChIA-PET) and chromosome conformation capture (3C) coupled with next-generation sequencing (Hi-C) (Dixon et al. 2012; Li et al. 2012) identified interaction of the intronic RLS-associated 32 kb LD block with the regulatory landscape of *MEIS1*. Publicly available data illustrated with the human epigenomic browser of the Washington University School of Medicine, St.Louis, USA, (http://epigenomegateway.wustl.edu/info/) (Zhou et al. 2013). The triangle shapes in the Hi-C track depict chromatin domains in IMR90 cells and the arcs in the ChIA-PET tracks indicate similar domain structures in MCF7 cells.

Table S1: Results of the high resolution melting curve analysis ("SNP discovery panel") of the RLS-associated 32 kb LD block (NCBI36/hg18).

Variant	HCNR element	Genome position	Minor allele	Major allele	MAF	Selection for follow-up genotyping
rs147475118	602	66602604	A	C	0.005	
rs114033747	602	66602608	A	C	0.005	
rs4316926	602	66602612	G	A	0.125	tagged by rs4430933 and rs4433986 with r2 = 1.0
rs62144055	602	66602814	C	A	0.11	tagged by rs11688599 and rs6742861 with r2 = 1.0
rs74559538	602	66603066	C	G	0.01	
rs4430933	602	66603114	A	G	0.13	genotyped
rs6724747	602	66603332	A	G	0.11	genotyped
ss947429183	602	66603477	A	G	0.003	
rs4544423	602	66603521	T	G	0.025	genotyped
rs6742861	602	66603656	C	T	0.315	genotyped
rs6728018	602	66603769	T	C	0.11	genotyped
rs73937943	602	66603788	A	C	0.003	
rs73937944	602	66603806	T	C	0.003	
rs113851554	**602**	**66604068**	**T**	**G**	**0.42**	**genotyped**
rs59088239	602	66604108	G	A	0.03	
rs34545468		66604989	A	G	0.003	
ss947429184		66605464	T	A	0.003	
ss947429185		66605609	G	C	0.003	
rs4233937		**66605755**	**A**	**G**	**0.138**	**genotyped**
rs145264250		**66605850**	**DEL**	**AAAC**	**0.138**	**genotyped**
rs62144057		66606060	G	C	0.109	tagged by rs11688599 and rs11678354 with r2 = 1.0
rs114947399		66606301	C	G	0.008	
ss947429186	606	66606746	A	G	0.016	
rs72822902	606	66607381	A	G	0.005	
rs4480973	606	66607557	A	C	0.088	genotyped
ss947429187	606	66607732	G	A	0.003	
ss947429188	606	66607735	T	C	0.005	
ss947429189	606	66607870	T	C	0.005	
ss947429190		66608378	C	T	0.003	
rs72824804		66608457	G	A	0.064	tagged by rs6710341 with r2 = 0.941
rs147083355		66608510	C	T	0.003	
rs56383906		66608878	C	T	0.011	
rs79464024		66609106	T	C	0.027	
ss947429191		66609324	A	G	0.003	
rs11687267		66609453	A	G	0.109	tagged by rs11688599 and rs11678354 with r2 = 1.0
rs12465087		66609493	T	A	0.019	
rs13003040		66609774	G	A	0.13	tagged by rs430933 and rs4316926 with r2 = 1.0
rs11677371		66609917	A	G	0.13	tagged by rs430933 and rs4316926 with r2 = 1.0
rs11688599		66610054	T	C	0.13	genotyped
rs11694675		66610140	A	T	0.109	tagged by rs11688599 and rs11678354 with r2 = 0.927
ss947429192		66610215	C	G	0.003	
rs3891585		66610480	A	G	0.13	tagged by rs430933 and rs4316926 with r2 = 1.0
rs55742206		66610571	A	G	0.109	tagged by rs11688599 and

						rs11678354 with r2 = 1.0
ss947429193		66610676	T	G	0.003	
rs62145760		66610954	T	C	0.109	tagged by rs11688599 and rs11678354 with r2 = 1.0
rs182588061		66611213	T	G	0.005	
rs6710341		66611926	A	G	0.197	genotyped
rs12621948	612	66612416	G	C	0.1	genotyped
ss947429194	612	66612421	G	C	0.003	
rs115566126	612	66612680	C	T	0.01	
ss947429195	612	66612744	C	A	0.005	
rs11693646	612	66613347	A	G	0.11	genotyped
rs11678354	612	66613583	A	T	0.136	genotyped
rs4433986		66613815	A	G	0.136	genotyped
ss947429196		66613849	A	G	0.005	
rs55855986		**66613965**	**DEL**	**T**	**0.13**	**assay design not possible, no tagging proxies available**
ss947429197		66614166	A	T	0.003	
ss947429198		66614210	A	G	0.005	
rs4316931		66614409	T	G	0.088	genotyped
rs142456824		**66614805**	**AT**	**DEL**	**0.106**	**genotyped**
rs191999528		66614962	A	G	0.003	
ss947429199		66615029	C	T	0.003	
ss947429200		66615134	T	C	0.005	
rs7600007		66615159	G	C	0.003	
rs9789607		66615652	T	A	0.106	genotyped
ss947429201		66615877	G	C	0.003	
rs7603415		66615880	G	C	0.019	tagged by rs2216125 and rs2216120 with r2 = 0.915
rs7603236		66615915	C	T	0.125	genotyped
rs150221482		66615978	TATA	DEL	0.021	
rs5006732		66615984	T	C	0.021	
rs74514787		66616019	G	A	0.003	
rs150277491		66616114	T	C	0.035	
ss947429203		66616178	T	C	0.003	
rs189634091	616	66616545	T	C	0.02	
rs147099389	616	66616717	T	G	0.003	
ss947429205		66617103	A	G	0.005	
ss947429206		66617173	A	G	0.003	
rs12469063	617	66617812	G	A	0.25	genotyped
ss947429207	617	66617888	T	C	0.003	
rs141399173	617	66617969	T	C	0.003	
ss947429208	617	66618106	G	A	0.003	
rs11681729	617	66618204	G	A	0.11	genotyped
rs139989120		66618445	A	G	0.011	
rs10865355		66618501	A	G	0.128	genotyped
rs9789535		66619082	A	G	0.144	genotyped
ss947429209		66619204	T	G	0.003	
ss947429210		66619219	G	C	0.003	
rs186720691		66619330	G	A	0.003	
ss947429221		66619520	DEL	T	0.021	
rs11678321		66619522	T	G	0.021	
ss947429211		66620016	G	A	0.003	
rs2216125		66620287	A	G	0.016	genotyped
rs2192960		66620468	C	T	0.271	tagged by rs12469063 with r2 = 0.823
rs189306611		66620521	A	G	0.011	
rs7563429		66621051	A	T	0.016	
ss947429202		66621636	C	G	0.003	

rs13000004		66621734	T	C	0.258	tagged by rs12472245 with r2 = 0.810
rs17625694		66622087	G	A	0.237	genotyped
ss947429213	622	66622320	G	A	0.003	
ss947429214	622	66622328	A	G	0.003	
rs17625724	622	66622479	A	G	0.11	genotyped
ss947429215	622	66622598	C	T	0.003	
rs17625742	622	66622658	T	C	0.125	genotyped
ss947429222	622	66623108	DEL	T	0.019	
rs11678796	622	66623539	A	G	0.11	genotyped
rs12472245	622	66624057	A	T	0.095	genotyped
rs997153	622	66624551	T	C	0.125	genotyped
rs10194077	622	66625367	C	T	0.135	genotyped
rs11897119	622	66625504	C	T	0.135	tagged by rs10194077 and rs17625694 with r2 = 1.0
ss947429216		66626044	A	C	0.003	
rs13029520		66626466	T	C	0.098	genotyped
rs139413629		66626528	C	T	0.003	
rs75313256		66626823	A	G	0.016	
rs10188003		66626973	T	C	n.d.	genotyped
rs367929915	628	66628521	G	A	0.005	
ss947429217	629	66629795	G	A	0.01	
rs2216120	629	66630010	C	T	0.02	genotyped
rs13019989	629	66630028	A	G	0.08	genotyped
rs141179647	629	66630214	T	C	0.003	
rs2300477	629	66630233	G	A	0.195	genotyped
rs2192954		66631025	G	A	0.41	genotyped
rs62145814	631	66631567	T	C	0.308	genotyped
ss947429259	631	66631639	T	C	n.d.	genotyped
rs189356944	631	66631764	G	A	0.01	
rs192448151	631	66631875	C	T	0.003	
rs149534896		66632176	C	T	0.005	
rs76977770		66632785	C	T	0.003	
rs3772026		66633370	T	C	0.13	tagged by rs10194077 and rs17625694 with r2 = 0.933
rs3755517		66633878	C	T	0.021	genotyped
ss947429260		66634004	T	C	0.005	
rs75226702		66634198	T	C	0.032	
ss947429218		66634481	A	G	0.003	
ss947429219		66634815	C	T	0.003	
rs76205499		66634930	T	A	0.003	
rs2300478		66634957	T	G	0.25	genotyped
rs2300480		66635212	C	T	0.093	genotyped

Results of the high resolution melting curve analysis ("SNP discovery panel") of the RLS-associated 32 kb LD block in 188 unrelated RLS-affected individuals (94 patients homozygote for the risk allele of rs12469063 (G/G), 94 patients heterozygote (G/A)). MAF was calculated in the respective sample, n.d. – not determined, refers to common variants for which exact MAF calculation was not possible due to overlapping melting curve profiles. Variants are annotated based on dbSNP 138 with novel variants identified in this study with their new ss-id. Positions refer to genome build NCBI36/hg18. Selection for genotyping indicates common variants selected for genotyping in additional samples and with additional

information if SNP was genotyped directly or tagged. Bold highlighting indicates SNPs, which did not pass quality control and were not included in the final analysis.

Table S2: Results of the case/control association study with the identified variants covering the RLS-associated 32 kb LD block.

Variant	HCNR element	Genome position	Risk allele	MAF (minor allele)	P_{nom}	P_{corr}
rs4430933	602	66603114	G	0.37 (A)	2.88E-05	0.00010
rs6724747	602	66603332	G	0.32 (A)	6.68E-04	0.00922
rs4544423	602	66603521	G	0.37 (T)	1.30E-06	0.00010
rs6742861	602	66603656	T	0.31 (C)	3.08E-04	0.00431
rs6728018	602	66603769	C	0.31 (T)	4.71E-04	0.00772
rs4480973	606	66607557	C	0.26 (A)	1.03E-01	0.63502
rs11688599		66610054		0.31 (T)	1.47E-04	0.00140
rs6710341		66611926		0.14 (G)	4.00E-01	0.99008
rs12621948	612	66612416	C	0.28 (G)	6.89E-02	0.50241
rs11693646	612	66613347	G	0.3 (A)	1.08E-05	0.00010
rs11678354	612	66613583	T	0.3 (A)	2.01E-05	0.00010
rs4433986		66613815		0.37 (A)	8.65E-06	0.00010
rs4316931		66614409		0.27 (T)	1.10E-01	0.65457
rs9789607		66615652		0.31 (T)	7.25E-05	0.00010
rs7603236		66615915		0.36 (C)	6.93E-08	0.00010
rs12469063	617	66617812	G	0.29 (G)	7.68E-18	0.00010
rs11681729	617	66618204	A	0.31 (G)	5.38E-05	0.00010
rs10865355		66618501		0.37 (A)	1.52E-05	0.00010
rs9789535		66619082		0.41 (A)	2.83E-04	0.00391
rs2216125		66620287		0.06 (A)	6.35E-02	0.47604
rs17625694		66622087		0.37 (G)	9.35E-06	0.00010
rs17625724	622	66622479	G	0.3 (A)	6.53E-05	0.00010
rs17625742	622	66622658	C	0.31 (T)	1.48E-05	0.00010
rs11678796	622	66623539	G	0.07 (A)	4.84E-01	0.99810
rs12472245	622	66624057	T	0.28 (A)	1.13E-01	0.66510
rs997153	622	66624551	C	0.37 (T)	3.38E-06	0.00010
rs10194077	622	66625367	T	0.37 (C)	1.53E-06	0.00010
rs13029520		66626466		0.37 (T)	2.30E-05	0.00010
rs10188003		66626973		0.37 (T)	1.99E-06	0.00010
rs2216120	629	66630010	T	0.06 (C)	2.54E-01	0.91680
rs13019989	629	66630028	G	0.27 (A)	8.05E-02	0.55443
rs2300477	629	66630233	A	0.44 (A)	1.21E-09	0.00010
rs2192954		66631025		0.48 (A)	1.61E-08	0.00010
rs62145814	631	66631567	C	0.31 (T)	5.11E-05	0.00010
ss947429259	631	66631639	C	0.31 (T)	4.96E-04	0.00792
rs3755517		66633878		0.07 (C)	2.64E-01	0.92562
rs2300478		66634957		0.3 (G)	4.24E-18	0.00010
rs2300480		66635212		0.28 (C)	9.68E-03	0.12390

Results of the case/control association study performed in 1,302 unrelated RLS-affected individuals from Germany and 1,259 age-, sex-, and ethnicity-matched population-based controls of the KORA cohort. Variants within the RLS-associated 32 kb LD block with a MAF $\geq$ 0.05, identified in the "SNP discovery panel" of 188 unrelated RLS patients, were included in statistical analysis. P-values were calculated using logistic regression (P_{nom}) with gender as a covariate and corrected for multiple testing by permutation with n = 10.000 permutations (P_{corr}). Positions refer to genome build NCBI36/hg18.

Table S3: List of the 51 proteins identified by affinity chromatography/mass spectrometry as differentially bound by rs12469063.

The acquired spectra were loaded into the Progenesis LC-MS software (version 4.0, Nonlinear) for label free quantification based on peak intensities. Spectra were searched against the Ensembl mouse database (Release 62; 54576 sequences). Normalized abundances of all unique peptides were summed up and allocated to the respective protein. Fold changes were calculated separately for every independent experiments based on normalized intensities, comparing the protective (NR) with the risk (R) condition. Mean fold change was calculated and proteins with fold changes ≥ 1.5 were considered differentially binding.

[a] number of peptides assigned to the respective protein, [b] the number of unique peptides used for quantification of the assigned protein, [c] Mascot score: number reflecting the combined mascot scores of all observed spectra matching to amino acid sequences within the respective protein. A higher score indicates a more confident match, [d] DNA binding: "yes" indicates, if the identified protein binds to the DNA in a sequence-specific manner based on GeneRanker software (Genomatix software suite) (Berriz et al. 2003), which uses annotation data from various sources, like Gene Ontology or Genomatix proprietary annotation,[e] Consistent allele-specific binding in three biological replicates: "yes" indicates, if the protein was identified consistently to bind preferentially the protective or the risk allele, in all three independent biological experiments, [f-h] Fold change (R/NR) 1st/2nd/3rd: the ratio R/NR of normalized abundances in the 1st/2nd/3rd biological replicate, respectively, [i] the mean of the ratios R/NR of normalized abundances for all three biological replicates, [j-k] the respective allele (NR or R) with the highest mean normalized abundance, [l]detailed protein description including abbreviation according to Mouse genome Informatics (http://www.informatics.jax.org/), [m-o] raw normalized abundances for the risk allele in all three independent experiments, [p-r] raw normalized abundances for the protective (non-risk) allele in all three independent experiments

[Please refer to separate excel file in the online Supplemental Material.]

Table S4: Heat map from unpaired two class analysis (SAM) of genes regulated between *Meis1*$^{tm1Mtor/wt}$ and wildtype forebrain tissue at E12.5 expanded to the level of individual genes.

Fold changes were calculated as ratios of signal intensities of individual mutant embryos and the arithmetic mean signal of the respective wildtype control group. Fold changes are given as log2 values and are color encoded according to the scale bar at the bottom of the heat map. Blue is down- and yellow is up-regulated in mutant embryos. The corresponding Illumnina IDs, official gene symbols and gene names are given to the right of the heat map. Genes (rows in heat map) are clustered according to similarities in expression profiles between individual mutant embryos (columns in heat map). Letters on the right of the figure refer to groups of genes with similar expression patterns and are identical to those in Fig. 5.

[Please refer to separate excel file in the online Supplemental Material.]

Table S5: Results of calorimetry trial: oxygen consumption and mean distance travelled.

	female		**male**		**linear model**	**linear model**	**linear model**	**linear model**
	control	**mutant**	**control**	**mutant**	**sex**	**genotype**	**body mass**	**sex:genotype**
	n = 11	**n =9**	**n = 11**	**n = 9**				
	mean ± sd	**mean ± sd**	**mean ± sd**	**mean ± sd**	**P-value**	**P-value**	**P-value**	**P-value**
avg.mass	22.6±1.5	20.9±1.7	29.8±3.1	27.7±2.5	< 0.001	0.002	NA	0.719
food intake [g]	1.9±0.6	2.2±0.5	2.2±0.9	2.5±1	0.75	< 0.001	< 0.001	0,215
avg. VO2 [ml/(h animal)]	82.24±5.41	82.08±7.48	92.98±9.25	94.83±8.71	0.003	< 0.001	< 0.001	0.406
min. VO2 [ml/(h animal)]	57.8±5.17	57.13±6.64	66.53±10.95	66.8±6.84	< 0.001	0.001	< 0.001	0.534
max. VO2 [ml/(h animal)]	111.27±7.32	116.4±11.39	128.8±13.59	129.47±15.11	0.023	< 0.001	< 0.001	0.534
avg. RER[a] [VCO2/VO2]	0.85±0.003	0.85±0.03	0.850±0.04	0.86±0.03	0.73	0.426	NA	0.975
avg. distance [cm]	1101±271	1192±296	815±154	945±123	< 0.001	0.06	NA	0.734
avg. rearing [counts]	143±52	148±53	98±30	114±20	0.001	0.319	NA	0.476

[a] RER = respiratory exchange ratio

Table S6: PreSTIGE-analysis focusing on the RLS-associated LD block suggests solely *MEIS1* as regulated gene.

Genome	Chr	Start (position)	Midpoint (position)	End (position)	HCNR	Predicted enhancer	Predicted Genes (Cell lines)[a]
hg18	2	66601963	66603150	66604337	602	chr2:66602948-66605355	***MEIS1*** (NHLF)
hg18	2	66606673	66607387	66608102	606	chr2:66606229-66607067	***MEIS1*** (NPC)
hg18	2	66612005	66612805	66613605	612	NA	NA
hg18	2	66617411	66617882	66618354	617	chr2:66617222-66618938	***MEIS1*** (NHLF, NPC)
hg18	2	66622163	66623916	66625669	622	NA	NA
hg18	2	66628857	66629739	66630622	629	NA	NA
hg18	2	66631456	66631695	66631935	631	NA	NA

Only *MEIS1* is predicted to be regulated by the RLS-associated 32 kB LD block (indicated in **BOLD**).

[a] NHLF, NPC = cell lines in which *MEIS1* was predicted to be regulated by the RLS-associated LD block

Table S7: Oligonucleotide sequences used for "high resolution melting curve analysis"amplicons

Amplicons are listed for highly conserved non-coding regions (HCNR) and the remaining areas (non-highly conserved non-coding regions (NHCNR) separately.

[Please refer to separate excel file in the online Supplemental Material.]

Table S8: Oligonucleotide sequences used in the "Sequenom iPLEX Gold" genotyping

For each genotyped SNP oligonucleotide sequences of the respective forward-, reverse- and extension-primer are given.

[Please refer to separate excel file in the online Supplemental Material.]

Supplemental Experimental Methods

Selection of variants for follow-up genotyping

42 variants were genotyped directly. For 16 variants, tagging SNPs with r^2-values of 0.85 and above were contained in the 42 genotyped variants. One variant (rs55855986) could not be converted to a genotyping assay and had no proxy SNPs in the set of genotyped SNPs. Quality control criteria for inclusion into statistical analysis were a call rate ≥ 90% for individuals, and call rate ≥ 95% and P-value for Hardy-Weinberg-Equilibrium > 0.001 for SNPs.

HCNR selection criteria and generation of HCNR reporter vectors

For determination of HCNRs we used the VISTA Genome Browser v2.0 *(http://pipeline.lbl.gov/cgi-bin/gateway2)* (Frazer et al.2004). First, we looked for ≥ 70% sequence identity over 100 bp between human *(March 2006; hg18)* and the respective species: mouse *(Jul. 2007; mm9)*, chicken *(May 2006; galGaL3)*, frog *(Xenopus tropicalis v4.1)*, and pufferfish *(Takifugu rubripes v4.0*). In case a region met this criterion in at least two species a HCNR was appointed. Distinct peaks were defined as one HCNR block, when the conservation in mouse was continuous and the corresponding peaks in the chicken conservation were closer than 700 bp. Thus, eight HCNRs were determined, which were named according to their genomic location (for example: HCNR 602 refers to the block beginning at genomic position ~66,*602*,000 on chromosome 2 etc.).

HCNR were cloned using Phusion polymerase (Finnzymes, Vantaa, Finland) and Qiagen Taq Polymerase to adapt the blunt-ended PCR product for the TOPO-TA-Cloning (pCR8GW/TOPO. Based on the allele-specific pCR8GW/TOPO vectors, all HCNRs were transferred via the Gateway-technology (Invitrogen) into the respective reporter vector and Sanger sequenced. For zebrafish, we used the zebrafish reporter detection vector (ZED; EGFP-gata2-promoter reporter vector flanked by Tol2 transposase recognition sites) (Bessa et al. 2009). This vector includes an intrinsic control with red fluorescent protein (dsRed; (Baird et al. 2000; Wall et al. 2000)). For mice, we used the beta-galactosidase reporter vector "Hsp68-*lacZ*" (Pennacchio et al. 2006). The following primers were used (in brackets length of PCR product): 602_F: 5'-AAACCTTCTAACACAGAATTTAGCTC and 602_R: 5'-TGCCACATTTGAATGCTACTTTAC (2375 bp), 606_F: 5'-

TGTATTCCCACTGCCTTGTG and 606_R: 5'-AAAGGCATGACTCTGATGAGG (1429 bp), 612_F: 5'-TGTGAAGTCTCTGTTTAAATAGGAAGG and 612_R: 5'-ATTTGATGGCAGGATTTTGG (1601 bp), 617_F: 5'-AATGCATAAAAAGTGGGCATT and 617_R: 5'-ACGCCATTTTGGAATGAGTC (944 bp), 622_F: 5'-ACTGGCAACTTCTTTTAACTGC and 622_R: 5'-TTGCATGCCTGTTTATGAGC (3507 bp), 629_F: 5'-TCCTTTATAAGTTGACAATTTTATGC and 629_R: 5'-GCTCTCCGGCAGAGACTGT (1766 bp), 631_F: 5'-CCAGGCTGGTCTCTAACTCC and 631_R: 5'-TCTCCTCTTTTGCCTTTCTCC (652bp).

Genotyping of transient transgenic mice

Genotyping was performed using extra-embryonic tissue and the transgene-specific PCR was performed using two primer pairs:
5'-GGTGGCGCTGGATGGTAA, 5'-CGCCATTTGACCACTACC
and 5'-GATGTTCCTGGAGCTCGGTA, 5'-GTGGATTGAAGCCCAGCTAA.

Histological methods

Templates for in vitro transcription of riboprobes: for Skor1 a full-length cDNA clone from Imagenes: (IMAGp998I071138Q), for *Meis1* cDNA as described previously (Mercader et al. 2009) and cDNA fragments amplified by following primers: 5'-gacgttgggggtcattgaag-3'and 5'-tgaggagagcgtgcagacag-3' for *Ptprd*; 5'-tggtgccttgttgattggtg-3' and 5'-cggtcccacaggagtagacg-3' for *Btbd*9; 5'-ccaccgagacactcaggac-3' and 5'-aggccatgtatgttccgttc-3' for *Map2K5*; 5'-caaatccagtcgcagactca-3', 5'-gggctaacagtggccataga-3' for *Tox3* and 5'-tgtagtttgacgcggtgtgt-3' and 5'-cttgagggcagaagtggaag-3' for *Creb1*.
Immunohistochemistry on paraffin sections was performed according to standard procedures using polyclonal mouse anti-MEIS1 antibody (Abnova, 1:200 dilution), monoclonal rabbit anti-CREB antibody (Ab32515, 1:200 dilution) and donkey anti-mouse/anti-rabbit IgG Alexa Fluor 488/ Fluor 594, respectively (Invitrogen, 1:1,000 dilution). The specificity was verified by omission of the primary antibody and pre-absorption of the antibody with an excess of MEIS1 peptide (Abnova, 1:100 dilution). Histological images were taken with an Axioplan2 microscope or StemiSV6 using a AxioCam MRc camera and Axiovision 4.6 software (Zeiss, Germany) and processed with Adobe Photoshop and Illustrator CS3 software (Adobe Systems Inc., USA).

Electrophoretic mobility shift assay (EMSA) and overexpression of CREB1 in 293T cells

5'-terminus Cy5-labeled oligonucleotides were annealed with the respective unlabeled complementary strand and subsequently purified using a native 12% polyacrylamide gel: 5'-GCTTCCAGCTGTGGCAGGCATGATGCAGTGAATTGCTTTT-3' (protective allele, underlined) and 5'-GCTTCCAGCTGTGGCAGGCGTGATGCAGTGAATTGCTTTT-3' (risk allele, underlined) ,
5' GATTCCAGACACAAACCCCAAGATGCAGTGAATTGCTTTT-3' (oligonucleotide with predicted CREB1 binding site mutated, underlined), CREB consensus (AGAGATTGCCTGACGTCAGAGAGCTAG) and mutant (AGAGATTGCCTGTGGTCAGAGAGCTAG) oligonucleotide (Santa Cruz, sc-2504 and sc-2517, respectively).

5 µg of nuclear protein extract were incubated with 1x EMSA binding buffer (3% glycerol, 0.75 mM MgCl2, 0.375 mM EDTA, 0.375 mM DTT, 37.5 mM NaCl, 7.5 mM TrisHCl pH 7.5) and 90 ng/µl poly-(dIdC) for 10 min on ice, prior to the addition of the labeled oligonucleotide probe. For competition the unlabeled probe was added in increasing amounts (33x, 66x, 100x nM) together with the labeled probe. After adding the oligonucleotide probe the reaction was incubated for additional 20 min on ice and in the dark. DNA-protein complexes were resolved by electrophoresis for 4 hours on a 5.3% native polyacrylamide gel in 0.5% TBE buffer, at 4 C and in the dark. The fluorescence of the Cy5-labeled oligonucleotides was detected with the Typhoon Trio+ imager (GE Healthcare).

For the supershift assays, 1 µg of rabbit anti-CREB (Abcam, ab32515) or 1 µg control antibody (rabbit IgG, same isotype, Santa Cruz) was added to the reaction before adding the oligonucleotide probe and incubated for 30 min on ice.

293T cells were transiently transfected with the CMV-*Creb1* overexpression vector (pcDNA3.1^{+}, cDNA of *Creb1* cloned in NheI and NotI sites), using standard methods (Lipofectamine 2000 Invitrogen). Nuclear extract was obtained using the NE-PER Extraction Reagents (Thermo Fisher Scientific).

RNA isolation and transcriptome analysis

Each embryo was genotyped for *Meis1* alleles (Azcoitia et al. 2005) and gender, and stored in liquid nitrogen. Forebrains of E12.5 mouse embryos were carefully dissected in order to isolate *Meis1* expressing areas of the MGE/LGE. Total RNA of eleven samples (males: *Meis1*$^{tm1Mtor/-}$ n = 3, wt n = 3; females: *Meis1*$^{tm1Mtor/-}$ n = 3, wt n = 2) was extracted according

to a standard protocol (RNeasy, QiagenTM) and RNA quality was examined using an Agilent Bioanalyzer 2100. Using the Illumina TotalPrep RNA Amplification kit (Ambion), 500 ng of high quality total RNA was amplified and hybridized to MouseRef-8v2 Expression BeadChips (Illumina, San Diego, CA, USA). After 16 hours hybridization, staining and scanning, data were normalized applying cubic spline normalization and background subtraction (GenomeStudioV2011.1, Illumina). Statistical analyses were performed using SAM (Significant Analysis of Microarrays) included in the TM4 software package (Saeed et al. 2003; Horsch et al. 2008). To estimate the false discovery rate, nonsense genes were identified by calculating 1000 random permutations of measurements. The selection of the top differentially expressed genes between heterozygous mutant and wildtype embryos with reproducible up- or down-regulation for the one class and two-class SAM analysis was based on an FDR < 10% in combination with a fold change above 1.5 fold. Hierarchical cluster analysis (HCL) (Eisen et al. 1998) was employed for identification of similar expression patterns applying the average-linkage-method as distance and the Euclidean Distance as distance-metric. Overrepresented functional annotations within the data sets were provided as GO terms of the category biological processes (Ingenuity Pathway Analysis, IPA). The complete array data set is available from the GEO database under accession number GSE44592.

Behavioral and metabolic phenotyping of *Meis1*$^{tm1Mtor/-}$ mice

For the evaluation of energy balance, single mice (9-11 per sex and genotype) were kept in respirometry cages (Phenomaster System, TSE Systems, Germany) (Gailus-Durner et al. 2009). The indirect calorimetry measurement started at 1 pm (CET) after about two hours adaptation to the cages and continued until 10 am the next morning. The set-up allowed the analysis of gas samples of individual mice every 20 minutes resulting in 63 readings per individual and trial. Mean oxygen consumption (ml/h) and respiratory exchange ratio (VCO_2/VO_2) were calculated from gas exchange data. Minimum and maximum rates were estimated from the lowest and the highest VO_2 detected during the trial. Two infrared light beam frames allowed the monitoring of physical activity (lower frame: distance travelled per 20 minutes, upper frame: number of rearings per 20 minutes). Food hoppers were attached to electronic scales to determine food consumption. Two way ANOVA was calculated for genotype, sex and genotype:sex effects on respiratory exchange ratio, distance travelled and number of rearings. Oxygen consumption and food intake were analyzed by linear regression

models including the factors genotype, sex, genotype:sex effects, and body mass as covariate (TIBCO Spotfire S+ 8.1 for Windows).

The Open Field test was assessed according to the standardized phenotyping screens developed by the Eumorphia partners (Mandillo et al. 2008), available as EMPReSSslim protocols (see www.eumodic.org). The Open Field apparatus consisted of a transparent and infrared light permeable acrylic test arena with a smooth floor (internal measurements: 45.5 x 45.5 x 39.5 cm). Illumination levels were set at approximately 150 lux in the corners and 200 lux in the middle of the test arena. Data were recorded and analyzed using the ActiMot system (TSE, Bad Homburg, Germany).

Poll II ChIA-PET and Hi-C data and "PreSTIGE"-software analysis

For all seven enhancers of the 32 kb RLS-associated region, interactions were analyzed based on predictions from "PreSTIGE" software using thirteen human cell lines, as available at http://genetics.case.edu/prestige/ (Corradin et al. 2014). Poll II ChIA-PET (Fullwood and Ruan 2009; Li et al. 2012) and Hi-C (Belton et al. 2012; Dixon et al. 2012) data from the human epigenomic browser of the Washington University School of Medicine, St.Louis, USA, (http://epigenomegateway.wustl.edu/info/) (Zhou et al. 2013) were analyzed with respect to the 32 kb RLS-associated LD block.

References

Azcoitia V, Aracil M, Martinez AC, Torres M. 2005. The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. *Dev Biol* **280**(2): 307-320.

Baird GS, Zacharias DA, Tsien RY. 2000. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci U S A* **97**(22): 11984-11989.

Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**(3): 268-276.

Berriz GF, King OD, Bryant B, Sander C, Roth FP. 2003. Characterizing gene sets with FuncAssociate. *Bioinformatics* **19**(18): 2502-2504.

Bessa J, Tena JJ, de la Calle-Mustienes E, Fernandez-Minan A, Naranjo S, Fernandez A, Montoliu L, Akalin A, Lenhard B, Casares F et al. 2009. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev Dyn* **38**(9): 2409-2417.

Corradin O, Saiakhova A, Akhtar-Zaidi B, Myeroff L, Willis J, Cowper-Sal Lari R, Lupien M, Markowitz S, Scacheri PC. 2014. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res* **24**(1): 1-13.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**(7398): 376-380.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**(25): 14863-14868.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* **32**: W273-279.

Fullwood MJ, Ruan Y. 2009. ChIP-based methods for the identification of long-range chromatin interactions. *J Cell Biochem* **107**(1): 30-39.

Gailus-Durner V, Fuchs H, Adler T, Aguilar Pimentel A, Becker L, Bolle I, Calzada-Wack J, Dalke C, Ehrhardt N, Ferwagner B et al. 2009. Systemic first-line phenotyping. *Methods Mol Biol* **530**: 463-509.

Horsch M, Schadler S, Gailus-Durner V, Fuchs H, Meyer H, de Angelis MH, Beckers J. 2008. Systematic gene expression profiling of mouse model series reveals coexpressed genes. *Proteomics* **8**(6): 1248-1256.

Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J et al. 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**(1-2): 84-98.

Mandillo S, Tucci V, Holter SM, Meziane H, Banchaabouchi MA, Kallnik M, Lad HV, Nolan PM, Ouagazzal AM, Coghill EL et al. 2008. Reliability, robustness, and reproducibility in mouse behavioral phenotyping: a cross-laboratory study. *Physiol Genomics* **34**(3): 243-255.

Mercader N, Selleri L, Criado LM, Pallares P, Parras C, Cleary ML, Torres M. 2009. Ectopic Meis1 expression in the mouse limb bud alters P-D patterning in a Pbx1-independent manner. *Int J Dev Biol* **53**(8-10): 1483-1494.

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M et al. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**(2): 374-378.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD et al. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**(7118): 499-502.

Wall MA, Socolich M, Ranganathan R. 2000. The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nat Struct Mol Biol* **7**(12): 1133-1138.

Zhou X, Lowdon RF, Li D, Lawson HA, Madden PA, Costello JF, Wang T. 2013. Exploring long-range genome interactions using the WashU Epigenome Browser. *Nat Methods* **10**(5): 375-376.