Trends in Developmental Biology
Vol. 8, 2014

Original Communication

The contribution of developmental biology to human genetics in the era of next-gen genome sequencing: what are we learning from studying the Kallmann syndrome?

Ivan Molineris, Roberto Santoro, Giulia Garaffo, Paolo Provero and Giorgio R. Merlo*
Department of Molecular Biotechnology and Health Sciences, University of Torino, Italy.

ABSTRACT

In the era of Next-Generation Sequencing (NGS) applied to human congenital diseases, developmental biology meets with transcriptome-wide analyses and predictive bioinformatics to address the following issues: a) identify novel disease-genes, b) clarify functions of detected mutations and/or polymorphisms, c) identify relevant networks and core regulations from 'omic' data. With the appropriate animal models, sets of high-throughput data can be intersected with relevant datasets, and used to predict/prioritize disease-genes. The same data can also be used to draw the blueprint of the core cellular and molecular processes altered in specific conditions. Finally, simple models such as *Danio rerio* or *Caenorhabditis elegans* can be used to functionally examine the effect of identified mutated genes in developmental processes. Kallmann syndrome (KS) is a paradigm of a genetically heterogeneous and complex set of conditions, in which many genes have been found mutated in patients' DNA, but in the majority of cases the disease-gene is unknown. We illustrate a general workflow that combines coding and non-coding RNA data from animal models to infer relevant pathways and predict novel disease-genes. The choice of the appropriate animal model, the generation of coherent sets of high-throughput data and ad hoc bioinformatic (meta)analyses are effective in revealing novel relationships between groups of genes and pathways, and work towards innovative hypotheses such as the involvement of microRNA (miR) misregulation. Predictions can rapidly be verified via resequencing of patients' genomes, in progress in several laboratories. Conversely, NGS data need parallel developmental models and profiling data to acquire functional relevance. We anticipate that similar approaches will become routine along with the evaluation of high-throughput genome-sequencing data from patients' DNAs.

KEYWORDS: mouse model, Kallmann syndrome, microRNA, disease gene prediction, semaphorin, FGF receptor

INTRODUCTION

When addressing polygenic diseases, having multiple inheritance modes and variable phenotypes and being mostly sporadic, the discovery of new disease-associated allelic variants by standard approaches is unrealistic. Such is the case with Cleft-Lip/Palate, Kallmann Syndrome (KS), and Autism Spectrum Disorders (ASD) or intellectual disability, to name a few. More global approaches that include developmental biology, bioinformatics and mutation screens are required. With the advent of Next-Generation Sequencing (NGS) in human genetics, genome-wide data can rapidly be collected from large populations (normal or patients); however the wealth of allelic variants identified by NGS approaches has little information content, if not accompanied by functional analyses. Alternatively,

*Corresponding author: giorgioroberto.merlo@unito.it

resequencing can be conducted only on selected genes, but in such cases the genome coverage is minimal [1]. Numerous bioinformatic tools exist for the computational assessment of the function of gene variants [2]. Animal models of congenital diseases and developmental high-throughput studies may be of considerable help to identify novel disease-genes more rapidly. Here we aim to illustrate the contribution of high-throughput studies on developmental models, focusing on the KS disease.

During embryonic development, immature olfactory receptor neurons (ORNs) differentiate and extend their axons to reach the anterior forebrain and make contact with the projection neurons. Axonal extension is accompanied by pools of migratory cells including the GnRH neurons, which reach the forebrain around E12-E13 and home in the hypothalamus and various preoptic areas of the brain [3-9]. Defects in olfactory development and/or GnRH neuron migration are the primary cause of a congenital disorder known as KS (OMIM 308700), characterized by central hypogonadism combined with anosmia.

Several protein-coding genes are known to be mutated in patients with KS and/or normosmic CHH (nCHH), and these include *KAL1*, *FGFR1*, *FGF8*, *PROK-2*, *PROKR-2*, *Kiss1R/GPR54*, *NELF*, *CHD7*, *GnRH-R*, *GnRH-R*, *HS6ST1*, *TAC3*, *TACR3*, *SOX10*, *SEMA3A* and 5 members of the 'FGF8-synexpressome' [10-21]. However, mutations in these genes account for less than 40% of the cases. It is expected, therefore, that more disease-genes remain to be identified.

Several mutant mouse strains display a KS-like phenotype [10, 15, 22-35]. The targeted inactivation of the homeodomain transcription factor *Dlx5* leads to a fully penetrant KS-life phenotype, consisting of delayed ORN differentiation, impaired axonal connectivity and failure of GnRH neurons to reach the forebrain [28, 29, 31]. We recently showed that a KS-like phenotype is also induced in the *Danio rerio* (zebrafish) model upon downmodulation of the endogenous *z-dlx5a* with morpholino injection [36]. Thus the depletion of *Dlx5* seems to provide an ideal model system on which to conduct 'omic' analyses and attempt to predict/prioritize KS disease-genes.

Here we combine the knowledge we have gathered from profiling of coding-genes and microRNAs (miRs) from the developing olfactory system of the mouse, with genomic data and with coexpression networks closely linked with the KS-causing disease-genes. Then, we apply tools to prioritize candidate genes [37, 38] likely to be relevant for KS. We also show that two relevant classes of genes, the FGF8 and the semaphorin signalling, are under control of olfactory-relevant miRs and Dlx5 transcription regulation. In the upcoming era of NGS applied to patients' genome, these approaches will become essential parts of disease-gene discovery. Future efforts in identifying mutated genes in KS patients could be directed towards these emerging classes.

MATERIALS AND METHODS

RNA profiling from *Dlx5* mutant embryos

Mice with targeted disruption of *Dlx5* have been previously reported [39], and the olfactory phenotype has been previously characterized [28, 31]. Transcriptome-wide gene profiling has been done on samples from E12.5 normal and *Dlx5*$^{-/-}$ olfactory tissues by hybridization on Affymetrix microarray, as described in [36]. miR expression profiling was carried out on the same samples, labelled using the one-colour method, with the miR Complete Labelling and Hybridization kit (Agilent), on the mouse 8x15K arrays (Agilent), with the mouse V2 microarray platform. These arrays comprise probes for 627 mature mouse miRs and 39 viral miRs. Data were extracted using conventional spot-recognition and significantly-above-background tools; the signal intensity was normalized across samples using the Lowess cyclic normalization algorithm. Differentially expressed miRs were detected using the SAM two-class unpaired statistical tool, using a FDR = 5%. Of the 627 mouse miR probes present on the arrays, 118 miRs were found to be expressed significantly above the background.

Softwares and databases

For preliminary Gene Ontology (G.O.) analyses we used DAVID (http://david.abcc.ncifcrf.gov/) and KEGG (http://www.genome.jp/kegg/pathway.html). For improved categorization and visualization, we used ClueGO [40]. To examine the embryonic

expression of individual RefSeq we used the publicly available transcriptome-wide *in situ* expression databases Genepaint (www.genepaint.org) and Eurexpress (www.eurexpress.org). For the prediction of miR targets we used TargetScan6.2 (www.targetscan.org).

Additional softwares and databases

Ensembl Genome Browser, http://www.ensembl.org/index.html
UCSC Genome Browser, http://genome.ucsc.edu
RefSeq, http://www.ncbi.nlm.nih.gov/refseq/
Mouse Genome Informatix, http://www.informatics.jax.org/
On-line Mendelian Inheritance in Man (OMIM), http://www.omim.org/
TS-CoExp Browser, http://www.mbcunito.it/cbu/ts-coexp

RESULTS

KS disease-genes examined via networks of conserved coexpression

With the exception of some genes evidently linked (*PROK2* with *PROKR2*; *FGF8* with *FGFR1*; *GNRH* with *GNRHR*) the remaining ones appear to be unrelated, or distantly related on a functional basis. We reasoned that relationships might exist between the KS-disease-genes that are not obvious, or not easily detected, or that genes may have pleiotropic functions, not known as yet. Tools have been developed that identify such relationships using either available databases or newly generated data, or both, and use the extracted information to propose putative disease-genes [41].

We compiled a list of genes known to either cause KS, or to cause KS and normosmic idiopathic central hypogonadism (nICH), but excluding those causing only nICH; the list included *FGFR1*, *FGF8*, *KAL1*, *PROK2*, *PROKR2*, *CHD7*, *GNRH*, *GNRHR*, *HS6ST1*, *TAC3*, *TACR3*, *SOX10*, *SEMA3A*, *FEZF1* and *TSHZ1*. We also included *FLRT3*, *IL17RD*, *FGF17*, *SPRY4* and *DUSP6*, members of the 'FGF8 synexpression' group [18] and named these 'human reference genes'. We positioned the 'human reference genes' within the global conserved coexpression network, using TS-CoExp (with the exception of *KAL1/ANOSMIN1* that lacks a mouse ortholog and for which the conservation criterion cannot be applied, and *FEZF1* for which no probe is present in profiling databases).

Using these data we generated a network of most connected genes (Fig. 1). Genes connected with at least six (N = 2), at least five (N = 5), at least four (N = 13), at least three (N = 70), at least two (N = 393) and at least one (N = 2265) reference disease-genes are listed in Suppl. Table 1. The most connected ones include: *SPTBN1*, *HIPK2* (connected with 6), *SNRPN*, *GATAD2A*, *NCAM*, *TRIO*, *ZFHX3* (connected with 5) and *CELF2*, *CCNY*, *MAST4*, *NFIB*, *SRPK2*, *CALD1*, *FOXP1*, *TNS1*, *NPTXR*, *SYNJ2*, *MRC2*, *CEACAM3*, *TRIM2* (connected with 4). Among these, *HIPK2*, *GATAD2A*, *NCAM*, *SNRPN*, *TRIM2* and *SRPK2* are expressed in the murine embryonic OE (www.genepaint.org); thus they represent interesting genes. Among the genes connected with 3 or more KS disease-genes, we do not find any gene causing KS-like phenotypes in mice (listed in Suppl. Table 2). For this reason we did not proceed to generate a coexpression network using the mouse disease-genes as input.

We categorized the most connected genes by Gene Ontology (GO) and detected an enrichment of the following G.O. categories: phosphoproteins, kinase/transmembrane receptors, cell adhesion, cell junctions, regulation of cytoskeleton, cell migration/motility and neuronal projection.

Kallmann disease-genes used to prioritize profiling data from the *Dlx5*$^{-/-}$ model

Gene prioritization establishes the ranking of lists of candidate genes on the basis of their relevance in a biological process, or on the basis of other known disease-genes. We used the 'human input genes' known to cause KS in humans (indicated above) as reference genes to compute a KS-signature and with this prioritize lists of genes derived from other sources, such as profiling experiments conducted on the *Dlx5* mouse model of KS-like phenotype [28, 31]. This procedure was carried out using the disease-gene prediction tool of TS-CoExp, previously shown to be robust and effective [41, 42]. Specifically, the following lists of differentially expressed genes were examined: *Dlx5*$^{-/-}$ OE vs. WT at E12 (list A);

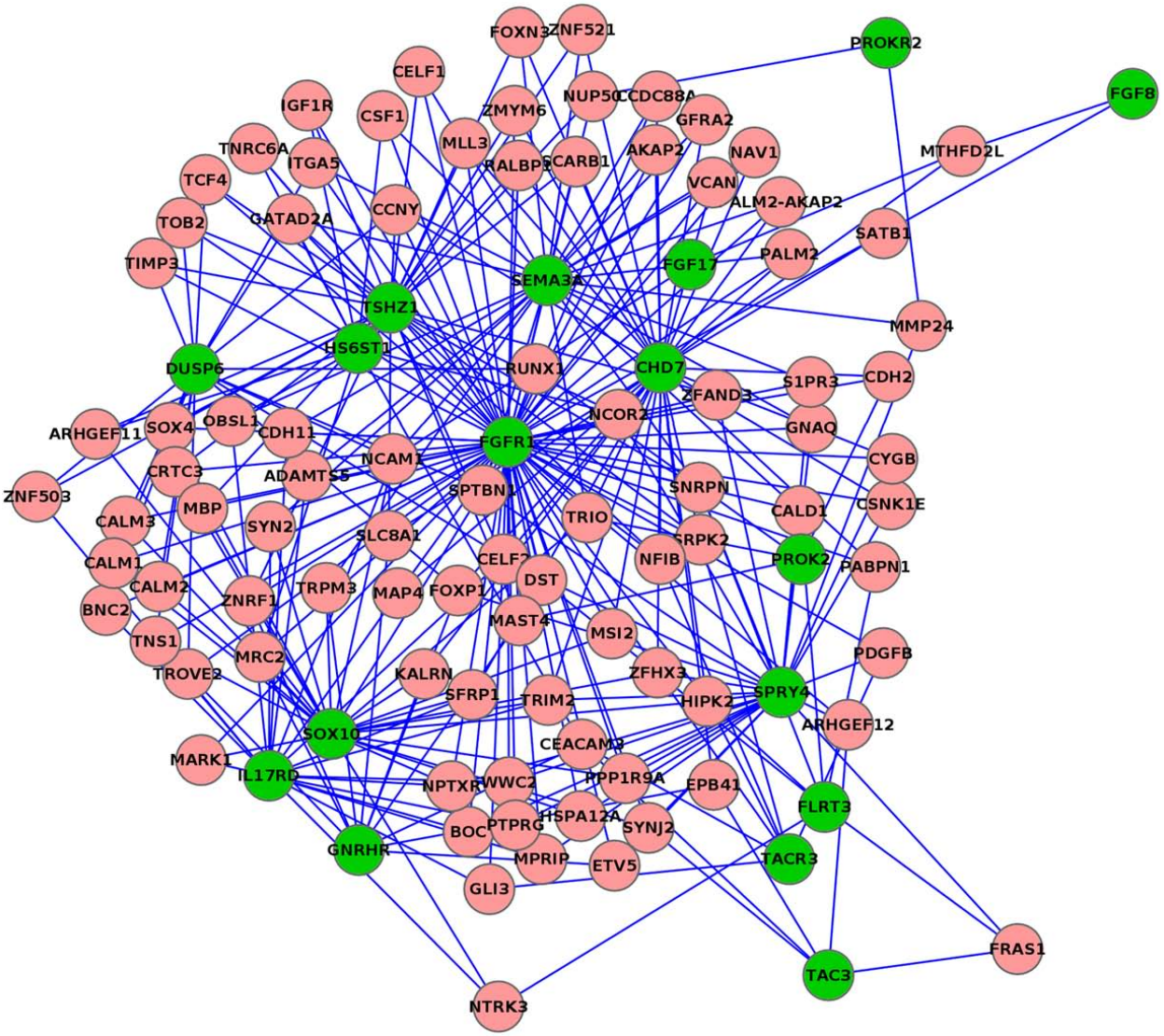


Figure 1. Network diagram illustrating the position of human KS-causing genes within the global conserved coexpression network, computed with the TS-CoExp algorithm. The list of the input human disease-genes used for this analysis is provided in the text. For simplicity, only the genes connected with at least three input genes are shown; the genes connected with 'at least 1' or 'at least 2' input genes are available upon request. Green circles represent the input genes, pink circles represent the connected genes, lines represent statistically significant coregulations.

EPI OE 14 vs. OPL E11 (WT) (list B); EPI VNO 14 vs. OPL E11 (WT) (list C); MES OE14 vs. OPL E11 (WT) (list D); MES VNO 14 vs. OPL E11 (WT) (list E). From the list A we found 17 genes significantly related to the KS-signature, four of which (*RGS5*, *F2RL1*, *DPF3* and *DSCAM*) are expressed in the murine embryonic OE, while two (*GATA3* and *ADAMTS5*) are expressed in the murine olfactory mesenchyme (Suppl. Table 3).

From the lists B and C we predict 21 and 78 genes respectively, related to the KS-signature, 90% of which are present in both lists, and the majority of these are expressed in the nasal mesenchyme (Suppl. Table 4). Notably, two genes known to cause KS in the mouse when disrupted, namely *EBF2* and *NRP1*, are present in these lists, confirming that our analysis is in principle correct. From the lists D and E of differentially

expressed genes (DEG) we predict 23 and 49 genes significantly linked to the KS phenotype, 50% of which are present in both lists (Suppl. Table 5). Among these, 30% show expression in the murine embryonic OE (*ATF5*, *NDRG1* and *TPD52*) while 70% are expressed in the murine olfactory-associated mesenchyme (*ANXA1*, *DCN*, *PAPS2*, *PTRF*, *RUNX1* and *TGM2*). Notably, the search predicted two genes known to cause a KS-like phenotype in mice: *EBF2* and *NRP1*. A number of interesting genes appear from these analyses, including *SEMA3C* (known to play a role in axon guidance), *RGS5*, *FGF7*, *ADAMTS5*, *NTRK2* and *TSHZ2* (related to the *TSHZ1* KS disease-gene), discussed in a previous work [36].

miRNAs involved in olfactory/GnRH development and the KS disease

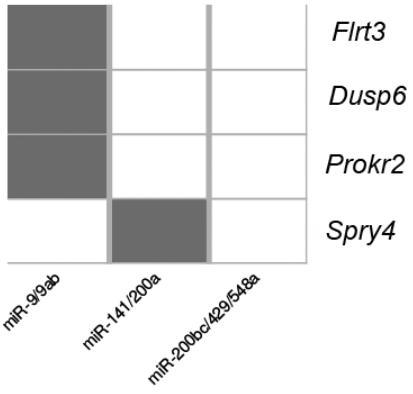
microRNAs (miR) are increasingly recognized as important molecular players in most biological processes, including embryonic development, neuronal differentiation and neuronal migration [43-50]. However, miR predictive biology is seldom used in attempts to identify core regulatory modules, or to prioritize/predict novel disease-genes, relevant for human genetics.

We carried out miR profiling comparing the *Dlx5*$^{-/-}$ embryonic OE with the wild-type counterpart. Notably, *Dlx5*$^{-/-}$ embryos are considered a model of KS-like phenotype in the mouse [28, 31]. The screen identified three miR classes that are downregulated in the *Dlx5*$^{-/-}$ OE, namely *miR-9/9ab*, *miR-141/200a* and *miR-200bc/429/548a* (Garaffo *et al.*, under revision). We reasoned that gene prioritization based on known KS disease-gene and coexpression networks could be applied to the best-predictable targets of the miR indicated above. We used TargetScan [51] and examined these gene lists for the presence of genes causing KS in humans (including the 'FGF8 synexpression' genes [18]. Three KS disease-genes (out of 20) are among the predicted targets for *miR-9/9a*, and two KS disease-genes (out of 20) are among the targets predicted for the *miR-200*-class (comprising both *miR-141/200a* and *miR-200b/c/429/548a* subclasses) (Table 1A). Initially the *FGFR1* gene was not found, however we repeated the analysis with miRecords [52], which integrates data from several miR target prediction tools, and we found that human (not mouse) *FGFR1* is potentially regulated by the *miR-200bc/429/548a*.

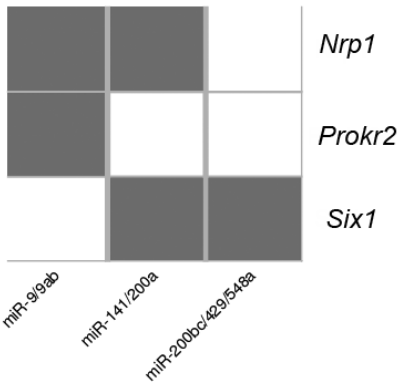
We then intersected the lists of best-predicted targets with the genes known to cause a KS-like phenotype in mice, upon targeted knockout (Suppl. Table 2); three mouse genes causing a KS-like phenotype (out of 12) are among these targets (Table 1B). Thus, human and mouse KS genes are clearly over-represented among the predicted *miR-9* and *miR-200* targets, implying that the networks regulated by these two miRs during olfactory system might contribute, when altered, to the KS phenotype.

Table 1. KS-causing disease-genes among the predicted targets for miRs relevant for olfactory/GnRH development. *FGFR1* is a predicted target of *miR-200bc/429/548* only in humans and not in mouse. Thus, this gene is also included. Grey boxes indicate the presence of a target sequence; open boxes indicate absence.

A. Human KS disease-genes.



B. Mouse genes causing KS-like phenotype.



Next, we used the lists of best-predicted *miR-9/9ab*, *miR-141/200a* and *miR-200bc/429/548a* targets to prioritize/predict novel putative disease-genes, using the human KS disease-signature. The procedure is based on previous bioinformatic meta-analyses conducted by our team, uses the disease-gene-prediction tool of TS-CoExp [41, 42], and has recently been reported [36]. The prioritization yielded 6 genes from the list of *miR-9/9ab* targets, and 16 genes from the *miR-141/200a* and *miR-200bc/429/548a* targets, taken together (Suppl. Table 6). Finally, we examined the genes 'most connected' with KS disease-genes in the conserved coexpression network (see Fig. 1 & Suppl. Table 1) for enrichments in miR target sequences in their 3' UTR. Several miRs were over-represented (Table 2); notably *miR-9* and *miR-200*-class are over-represented in the human network (while only the *miR-200* class is enriched in the mouse network) and several of the over-represented miRs are expressed in the embryonic olfactory system [46]. These results suggest that the coexpression of KS disease-genes and of the most connected ones is strongly based on regulation by miRs.

The 'FGF8 synexpression' genes and FGFR1 signalling network

Molecules implicated in the FGF8 signalling pathways are heavily implicated in the onset of the KS disease. So far, the genes in the signalling network implicated in KS are: *FLRT3*, *IL17RD*, *FGF17*, *SPRY4* and *DUSP6* [18], with the addition of *FGFR1* and *FGF8*, previously identified as KS disease-genes. The involvement of the FGF8 signalling pathway also emerges strongly from developmental studies in model organisms [53, 54].

We examined the 'FGF8 synexpression group' genes as a whole, asking the question whether there are indications they are transcriptionally regulated by transcription factors (TFs) implicated in the KS phenotype, or whether they could be regulated by the miRs specifically implicated in the olfactory and GnRH development. Specifically we examined Dlx5, a TF causing a KS-like phenotype in mice [28, 31]. Recent observations suggest that *FGF8* is a target of Dlx5, at least in the developing limb bud [55]. We have previously generated a genome-wide prediction of Dlx5 DNA-binding sites, based on its position-weight matrix (JASPAR PH0024.1) [56, 57], and experimentally validated some of them. We examined the promoter region (up to 10 kb upstream of the TSS) and the first intron of the mammalian 'FGF8 synexpression' genes. The results show that two genes, namely *IL17RD* and *FGFR1*, contain *bona-fide* conserved Dlx5 sites (Table 3A). Thus, some 'FGF8 synexpression' genes are expected to be controlled by a TF relevant for the onset of KS-like phenotypes in rodents.

Second, we intersected the predicted targets of *miR-9/9ab*, *miR-141/200a* and *miR-200bc/429/548a*, with the 'FGF8 synexpression' group genes, and found that four of these are expected to be regulated by either *miR-9*or *miR-200*-class (Table 1A). Thus, the FGF8 signalling pathway as a whole is strongly regulated by miRs, specifically by those linked to GnRH and olfactory development.

The semaphorin class

The signalling molecules semaphorins and their receptors (plexins, neuropilins and others) are emerging to be strongly implicated in the development of the olfactory/GnRH system, as well as mutated in KS and ICH patients, and causing KS-like phenotypes in the mouse model when mutated [10, 25, 58].

First we examined whether semaphorins and/or their receptors appear in the lists of predicted/ prioritized genes. *SEMA3C* and *NRP1* appear among the genes prioritized from the profiling data (lists D, E); *SEMA6A*, *SEMA6D*, *SEMA3F*, *NRP1* and *NRP2* appear among those prioritized from *miR-200* class targets, and *NRP1* appears among those prioritized from the *miR-9* targets. Second, we examined the possibility that this gene category might be transcriptionally controlled by Dlx5. We considered the promoter region (up to 10 kb upstream of the TSS) and the first intron of all mammalian semaphorins and their receptors/ coreceptors (plexins, neuropilins, etc.) (Suppl. Table 7); we identified 8 semaphorin genes (out of 21) and 7 receptor/coreceptor genes (out of 12) with conserved Dlx5 sites (Tables 3 B & C). These results suggest that the semaphorin function as a whole is globally controlled by the transcriptional activity of Dlx5, although the known disease-causing semaphorins are not among these ones.

Table 2. miR target sequences enriched in the 3' UTR of genes connected with ‘at least three’ human KS disease-genes in coexpression networks. This table is partial, full table is available on request.

A. Human network.

miR	p-value	Common	Expected	Common list
miR-128/ 128ab	8.7 E-08	22	6.28	*ADAMTS5; CCDC88A; CCNY; CDH11; CSF1; EPB41; GATAD2A; HIPK2; ITGA5; MAST4; NCAM1; NFIB; NPTXR; NTRK3; PALM2; PPP1R9A; RUNX1; SFRP1; SPTBN1; SRPK2; TROVE2; ZFHX3*
miR-144	5.6 E-07	19	5.27	*ADAMTS5; AKAP2; BNC2; CCDC88A; CDH11; CELF2; DST; EPB41; FOXP1; HIPK2; NFIB; PABPN1; PALM2; PALM2-AKAP2; RUNX1; SFRP1; TRIO; TROVE2; ZFHX3*
miR-27abc/ 27a-3p	4.5 E-06	21	7.27	*ADAMTS5; CALD1; CALM3; CCNY; CDH11; CELF2; CSF1; EPB41; GATAD2A; HIPK2; ITGA5; NAV1; NCAM1; NFIB; NPTXR; PALM2; RUNX1; SFRP1; TRIM2; TROVE2; ZFHX3*
miR-141/ 200a	5.8 E-06	16	4.46	*AKAP2; BNC2; CSNK1E; FOXP1; HIPK2; KALRN; MARK1; MMP24; MPRIP; NCAM1; PALM2-AKAP2; PTPRG; RUNX1; TNS1; TRIM2; VCAN*
miR-543	2.3 E-05	15	4.43	*AKAP2; BNC2; CDH2; CELF2; EPB41; FOXP1; HIPK2; HSPA12A; MARK1; NFIB; NPTXR; PALM2-AKAP2; SPTBN1; TRIM2; VCAN*
miR-495/ 1192	6.5 E-05	16	5.42	*BNC2; CCDC88A; CELF2; CSNK1E; FRAS1; GLI3; MAST4; MPRIP; MTHFD2L; NFIB; NPTXR; PPP1R9A; S1PR3; SATB1; SPTBN1; ZFHX3*
miR-129-5p/ 129ab-5p	8.3 E-05	12	3.27	*CALM1; CCDC88A; CDH2; CELF2; ETV5; GATAD2A; GFRA2; MAST4; NFIB; NUP50; RUNX1; TRIO*
miR-200bc/ 429/548a	0.0001	17	6.35	*AKAP2; BNC2; CDH11; CELF2; ETV5; FOXP1; GLI3; HIPK2; MPRIP; NCAM1; NFIB; PALM2; PALM2-AKAP2; PPP1R9A; RUNX1; TRIO; WWC2*
miR-300/ 381/539-3p	0.0001	15	5.29	*BNC2; CALM2; CCNY; CELF2; EPB41; GATAD2A; HIPK2; KALRN; MAP4; MARK1; NFIB; SATB1; TRIO; WWC2; ZFHX3*
miR-25/32/ 92abc/363/ 363-3p/367	0.0022	13	5.36	*CALM3; CELF2; ITGA5; KALRN; MARK1; MAST4; NFIB; PALM2; PPP1R9A; SRPK2; SYN2; TRIO; ZFHX3*
miR-1ab/ 206/613	0.0023	12	4.74	*CALM1; CALM2; FOXP1; FRAS1; GATAD2A; HIPK2; NUP50; PTPRG; RUNX1; SFRP1; SPTBN1; TRIM2*
miR-374ab	0.006	10	4.08	*ADAMTS5; CALD1; CELF2; FOXP1; HIPK2; NFIB; RUNX1; SATB1; SRPK2; WWC2*
miR-758	0.013	5	1.41	*ADAMTS5; AKAP2; PALM2-AKAP2; WWC2; ZFHX3*
miR-9/9ab	0.013	14	7.42	*ADAMTS5; BNC2; CCDC88A; FOXP1; HIPK2; NFIB; OBSL1; PALM2; RUNX1; S1PR3; TNS1; TRIM2; VCAN; ZFHX3*
miR-142-3p	0.014	6	1.98	*ARHGEF12; BNC2; CCDC88A; GNAQ; NFIB; S1PR3*
miR-182	0.014	13	6.73	*ARHGEF12; BNC2; CCDC88A; CCNY; CELF2; CSNK1E; GNAQ; MAST4; NUP50; PPP1R9A; PTPRG; TNS1; WWC2*

Table 2 continued..

B. Mouse network.

miR	p-value	Common	Expected	Common list
miR-128/128ab	4.79 E-06	17	5.03	*Ccdc88a; Ccny; Cdh11; Csf1; Epb4.1; Gatad2a; Itga5; Msi2; Ncam1; Nfib; Ntrk3; Palm2; Ppp1r9a; Runx1; Sfrp1; Trove2; Zfhx3*
miR-27abc/27a-3p	9.7 E-05	16	5.69	*Ccny; Cdh11; Csf1; Epb4.1; Gatad2a; Itga5; Kalrn; Msi2; Nav1; Ncam1; Nfib; Palm2; Runx1; Sfrp1; Trove2; Zfhx3*
miR-141/200a	0.0012	10	3.24	*Akap2; Bnc2; Csnk1e; Kalrn; Ncam1; Ptprg; Runx1; Tns1; Trim2; Vcan*
miR-218/218a	0.0012	12	4.45	*Arhgef12; Ccdc88a; Cdh2; Csnk1e; Kalrn; Mast4; Mprip; Msi2; Nup50; Palm2; Ptprg; Trio*
miR-182	0.0013	13	5.13	*Arhgef12; Bnc2; Ccdc88a; Ccny; Celf2; Csnk1e; Mast4; Ncam1; Nup50; Ppp1r9a; Ptprg; Tns1; Wwc2*
miR-223	0.0014	6	1.23	*Celf2; Map4; Msi2; Nfib; Srpk2; Zfhx3*
miR-30abcdef/ 30abe-5p/384-5p	0.0024	15	6.86	*Bnc2; Epb4.1; Itga5; Kalrn; Mast4; Msi2; Nav1; Ncam1; Nfib; Palm2; Ppp1r9a; Runx1; Satb1; Syn2; Trio*
miR-137/137ab	0.0024	13	5.47	*Akap2; Ccny; Celf2; Itga5; Kalrn; Mark1; Mast4; Msi2; Nfib; Ppp1r9a; Srpk2; Syn2; Trove2*
miR-200bc/ 429/548a	0.0025	12	4.83	*Akap2; Ccny; Cdh11; Celf2; Gli3; Kalrn; Mprip; Nfib; Palm2; Ppp1r9a; Srpk2; Trio*
miR-219-5p/508/ 508-3p/4782-3p	0.0028	7	1.921	*Arhgef12; Celf2; Etv5; Nav1; Palm2; Satb1; Sptbn1*
miR-494	0.003	7	2.00	*Arhgef12; Celf2; Gli3; Map4; Nfib; Palm2; Zfhx3*

Table 3. Dlx5 binding sites near the 'FGF8 synexpression group' (A), the semaphorin group (B) and the semaphorin receptor/coreceptor group (C) of genes.

A. The 'FGF8 synexpression' group.

Gene ID	Gene symbol	Score
54756	*IL17RD*	6
2260	*FGFR1*	1
8822	*FGF17*	0
1848	*DUSP6*	0
81848	*SPRY4*	0
23767	*FLRT3*	0
2253	*FGF8*	0

Table 3 continued..

B. The semaphorin group.

Gene ID	Gene symbol	Gene name	Score
20358	*Sema6a*	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6A	14
108151	*Sema3d*	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3D	10
20357	*Sema5b*	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5B	10
214968	*Sema6d*	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D	6
20349	*Sema3e*	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3E	2
20348	*Sema3c*	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	1
20356	*Sema5a*	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	1

C. The semaphorin receptor/coreceptor group.

Gene ID	Gene symbol	Gene name	Score
243743	*Plxna4*	plexin A4	9
67448	*Plxdc2*	plexin domain containing 2	6
18845	*Plxna2*	plexin A2	6
18187	*Nrp2*	neuropilin 2	4
18844	*Plxna1*	plexin A1	4
140570	*Plxnb2*	plexin B2	1
54712	*Plxnc1*	plexin C1	1

Third, we examined the possibility that semaphorin and semaphorin receptor/coreceptor genes are enriched among the predicted *miR-9* or *miR-200* targets. To do this, we carried out a prediction of all miR seed sequences present in the 3' UTR of the genes indicated in Suppl. Table 7. Semaphorin mRNAs are enriched in the miR target sequences indicated in Table 4A, which do not include *miR-9* and *miR-200* class. On the contrary, the semaphorin receptor/coreceptor mRNAs are enriched in miR target sequences reported in Table 4B, among which *miR-200* seed sequence for 4 of these (out of 12), namely *NRP2, PLXNA2, PLXNA4* and *VEGFA*. These observations suggest that the *miR-200* class, relevant for olfactory development, globally impacts on the semaphorin receptors/coreceptors gene category.

DISCUSSION

To address complex, genetically heterogeneous, mostly sporadic developmental disorders, two general workflows are proposed, depicted in Figure 2: A) developmental biologists help human geneticists identify and prioritize candidate

Table 4. miR target sequences in the 3' UTR of semaphorin and their receptor/coreceptor mRNAs.

A. Semaphorin genes.

miR	p-value	Found	Expected	Gene symbol
miR-24/24ab/24-3p	0.0017	5	0.91	*Sema4a; Sema4b; Sema4g; Sema5a; Sema6a*
miR-125a-5p/125b-5p/ 351/670/4319	0.043	4	1.35	*Sema4b; Sema4c; Sema4d; Sema4f*
miR-17/17-5p/20ab/20b-5p/ 93/106ab/427/518a-3p/519d	0.047	5	2.05	*Sema3c; Sema4b; Sema4g; Sema5a; Sema7a*

B. Semaphorin receptor/coreceptor genes.

miR	p-value	Found	Expected	Gene symbol
miR-361-5p	0.0011	3	0.22	*Nrp1; Plxna2; Vegfa*
miR-299/299-3p/3563-3p	0.0041	2	0.097	*Itgav; Vegfa*
miR-1ab/206/613	0.0089	4	0.87	*Met; Nrp1; Plxna4; Vegfa*
miR-183	0.0097	3	0.459	*Itgb1; Nrp2; Plxna4*
miR-199ab-5p	0.0137	3	0.521	*Plxna2; Plxnd1; Vegfa*
miR-200bc/429/548a	0.0258	4	1.187	*Nrp2; Plxna2; Plxna4; Vegfa*
miR-320abcd/4429	0.0366	3	0.757	*Itgb1; Nrp1; Plxnc1*
miR-15abc/16/16abc/195/322/424/497/1907	0.04547	4	1.414	*Nrp2; Plxna2; Plxna4; Vegfa*

disease-genes, to be screened for mutations in patients' genomes, B) human geneticists provide the developmental biologists with lists of non-synonymous allelic variants, of unknown function and with uncertain pathogenic roles. The proposed interaction appears both necessary and highly informative, especially in developmental disorders of complex etiology and variable phenotype.

A. Developmental biology provides human geneticists prioritized putative disease genes

The choice of an appropriate (mutant) animal model that recapitulates the main features of the disease of interest is the first critical issue. The model should be used to identify the tissue that primarily is defective in the earliest step of the disease onset, and in parallel identify the most appropriate time-window for the molecular analyses. Expression profiling data can be generated according to the following schemes: a) adopting a time-series approach, i.e. comparing embryonic tissues at the relevant age windows, b) comparing the normal tissue with that from a mutant model of the disease, c) screening for coding and non-coding RNAs, and searching for RNA-based regulatory pathways. On the other hand, protein-protein interactions (PPI) are difficult to examine in embryonic cells and tissues; nevertheless PPI networks can be used in bioinformatic meta-analyses to infer relevant information (see [18]), provided that data are generated from a cell type relevant for the developmental process of interest.

Time series and mutant vs. wt expression profiles from embryonic mouse tissues or sorted cells are used to identify differentially-expressed genes, and define biologically-relevant pathways or temporal expression trends. This has been done, for instance, to identify new candidate genes for congenital diaphragmatic hernia, combining various bioinformatic tools and existing relevant datasets [59].

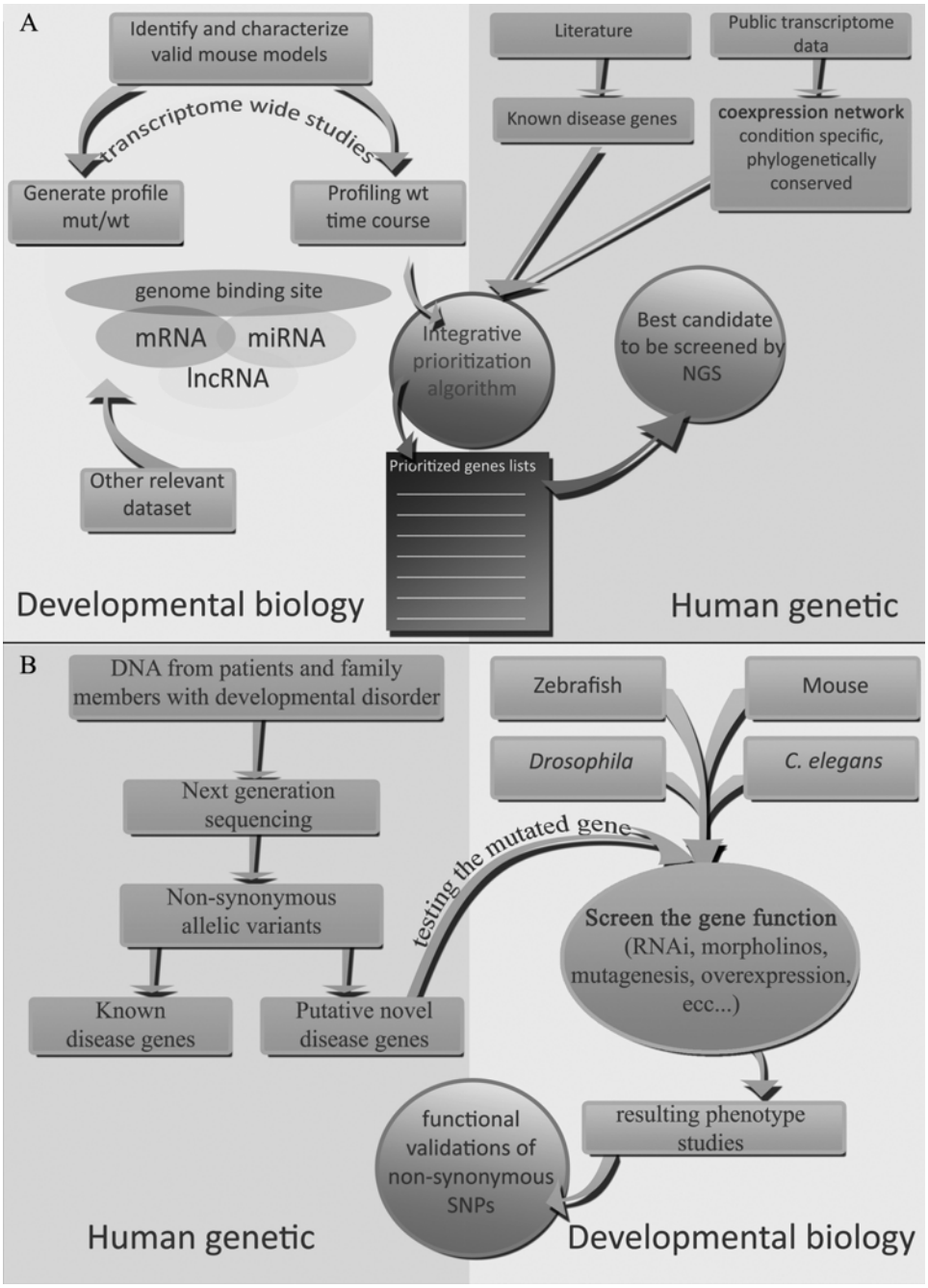


Figure 2. Workflow describing the contribution of developmental biology to human genetics, in search for novel disease-causing mutations in genetically heterogeneous diseases (**A**), and conversely the need for developmental biology studies to assess functional significance of non-synonymous allele variants identified by NGS approaches applied to patients' genomes (**B**).

A popular approach is to prioritize/predict candidate disease-genes based on the identification of specific biological networks enriched in cases as compared to controls [37, 38]. Although this approach cannot be used unequivocally to define cause-effect relationships, membership of a specific gene in a particular PPI or coexpression network may increase the likelihood of its association

with disease. Another widely used approach uses known disease-genes, or *de novo* copy number variations (CNV) and single-nucleotide variant (SNV) mutations in specific set of diseases [60-63]. Indeed, a role for a given gene in disease is often provided by related pathologies, either in humans or in model organisms. In many cases, vertebrate models (mouse, zebrafish, etc.) already exist and relevant phenotypes have been documented. Invertebrate models might be useful as well, mainly for pathway analyses and complementation tests [64].

Lists of prioritized genes are then used in further studies, mainly NGS in human genetics. There are now hundreds of new candidate genes and targeted resequencing technologies that allow screening of dozens of genes in large numbers of individuals with high specificity and sensitivity.

The decision of which genes to pursue depends on many factors, including recurrence, previous evidence of overlap with pathogenic CNV, the position of the mutation in the protein, the mutational burden among healthy individuals and membership of the candidate gene in disease-implicated protein network modules [65].

B. Human genetics provides developmental biologists allele variants for functional tests

Recent NGS technologies allow the full spectrum of genetic variation to be monitored in genomic DNA of patients. Whole-exome and whole-genome sequencing applied to large numbers of genomes from patients with ASD, intellectual disability, epilepsy and schizophrenia has provided clear evidence of the importance of *de novo* and gene-disruptive events [65]. The increasing deep reads allow *de novo* mutation discovery which is frequently used for 'disease-gene discovery' in complex genetic conditions, such as ASD or intellectual disability. The hope is to identify the most likely causal mutation in diseased individuals. For this approach to work efficiently, first a detailed phenotypic characterization of patients is needed, to select candidate genomes for resequencing. Next, known and most frequent mutations should be searched first. Third, in the case of rare inherited diseases, familial clusters should be the first choice, and resequencing should be applied to several affected and unaffected family members.

One example of combining NGS data with biological network from related developmental processes is reported on ASD [66]. Using NGS they identified nine high-confidence ASD genes, and used previously published spatially- and temporally-rich transcriptome data set to computed coexpression dynamics and functionally evaluate the proposed disease-genes. Interestingly, when testing statistically significant enrichment of high-confidence ASD genes within a set of spatiotemporal expression network modules, previously identified by hierarchical clustering, they identified a key point of convergence in a specific time and brain region. This approach informs when, where, and in what cell types, mutations in these specific genes may be most relevant for ASD. In general, this supports the hypothesis that the genetic heterogeneity underlying complex diseases can be referred to a much smaller set of underlying pathological mechanisms.

Since a typical NGS screen generates several non-synonymous allelic variants, of which only a few are likely to turn out as disease-genes, developmental biology comes in handy to screen for gene functions. Mouse embryo studies would be ideal, however they are very demanding, time-consuming and increasingly less well accepted by the public opinion and legislators. As an alternative, other vertebrate (*D. rerio* and *O. latipes*) or invertebrate (*C. elegans* and *Drosophila*) models can be used.

miR in development and disease

Most research, so far, has focused on the identification of protein-coding genes, while little information is available on the role of miRs (and non-coding RNAs in general) in the molecular pathogenesis of congenital disorders. Instead, it is increasingly recognized that biological processes are governed by complex regulatory modules and networks of molecular interaction, rather then simplistically by individual genes. miRs and other non-coding RNAs are essential players in these networks, being negative post-transcriptional regular of mRNA stability and translation, and recently have been implicated in more complex regulations such as RNA::RNA competition [67, 68]. Few studies have documented a role of miRs in developmental pathologies. A frame-shift mutation was found in the binding site for *miR-189* in Tourette's syndrome patients [69], and

miR-153 has been found misregulated in a murine model of Alzheimer's disease [70]. In general altered expression, rather than mutation, is thought to be a likely mechanism of disease. Or, alternatively, mutations in proteins that interact with mRNAs and miRs and that selectively control gene expression at translational level can modify miRs level [71].

Taking advantage of miR profiling data from a developmental model of KS, here we have attempted to incorporate miRs in our disease-prediction workflow, and show that two classes of KS-relevant genes (semaphorin receptors and FGF8) are controlled by miRs relevant for the development of the olfactory/GnRH system in the mouse embryo. Thus, the participation of these miR and some of their OE/GnRH-selective targets is a most likely possibility.

Novel information on Kallmann syndrome and future advancements

Using the workflow in Figure 2, starting from a mouse model of KS we inferred sets of gene functions and predicted/prioritized candidate disease-genes and miRs. This effort is justified by the fact that in < 40% KS patients the causative mutation(s) has been found. A number of gens emerging from these analyses have been previously discussed as putative KS disease-genes [36]. They should rapidly be included in NGS-based screening approaches on DNA from KS patients.

The role of FGFR1 signalling

FGFR1 mutations cause KS and related conditions, alone or in combination with other mutations. In 2013, Miraoui and colleagues identified 5 additional genes of the *FGF8-FGFR1* functional group that are mutated in KS patients, which were initially predicted by computational methods [18]. Here we identified *FGF7* as a novel likely candidate, belonging to the FGF family. We also show that the 'FGF8 synexpression' genes are globally regulated by those miRs found to be down-modulated in the *Dlx5*$^{-/-}$ model of KS disease. Indeed, we detect an enrichment of these genes among the predicted targets of *miR-9* and *miR-200* class. In KS patients having no mutations in these genes, *miR-9* and *miR-200* levels could be altered and this would lead to changes in expression and/or stability of target miRs.

The role of semaphorin signalling

Semaphorins and their receptors/coreceptors are emerging to be strongly implicated in the development of the olfactory/GnRH system [58]. Our analyses show that the semaphorin and the semaphorin receptor/coreceptor genes are collectively under the transcriptional control of Dlx5 and under the control by the *miR-9* and *miR-200* class of ncRNAs. Interestingly, *SEMA3A*, a gene found to be mutated in a subset of KS patients, and mouse models strongly support its role in GnRH development [10, 25]. However, *SEMA3A* is not under control by Dlx5 or *miR-9* and *miR-200*. Thus, it is tempting to speculate that, in the absence of *SEMA3A* mutations, or mutations in other genes of this family, other events could affect semaphorin signalling via miR-based mechanisms, and thus contribute to the onset of the KS phenotype, a possibility to be explored experimentally.

miRs in olfactory/GnRH development and KS

During neural development miRs participate in all processes, from differentiation to physiology of mature neurons, survival, homeostasis, activity and plasticity [43-45, 47-50, 72-74]. Little information is available on the role of miRs (or other non-coding RNAs) in olfactory/GnRH neuron development. In mice, miRs are required quite early for ORN differentiation, as shown by conditional knockout of *Dicer* [46]. The *SEMA3A* gene, which causes KS in humans and a KS-like phenotype in mice [10, 25], is directly regulated by *miR-124* in retinal neurons, and this control modulates the response of these cells to the SEMA3A ligand [75]. We have recently detected specific alterations of *miR-9* and *miR-200*-class in olfactory/GnRH development, and confirm their developmental functions in zebrafish embryos; indeed their depletion leads to defective axonal organization and connectivity and to altered GnRH neuron migration (Garaffo *et al*., under revision). A similar study in zebrafish embryos showed a role of *miR-200*-class for olfactory neuron differentiation [46].

Based on these observations, miRs are likely to play a role in the pathogenesis of KS and related diseases. To begin to address this hypothesis we intersected the best-predicted targets of *miR-9*, *miR-200a/141* and *miR-200bc/429/548a* with the

known KS genes and found an enrichment of the 'FGF8 synexpression' genes, and of the semaphorin and semaphorin receptor/coreceptor genes. These findings suggest the existence of an intricate miR::mRNA network involved in olfactory and GnRH development. A misregulation of this network could contribute to the onset of the KS phenotype. It is temping to speculate that in KS patients with no mutations in *SEMA3a* or in genes of the 'FGF8 synexpression', altered *miR-9* or *miR-200* could have taken place during embryonic development. How could miR be altered? Most likely not by mutation, rather by altered transcription, a hypothesis that cannot be easily tested. Considering that we observe a downregulation of *miR-9* and *miR-200*-class, linked to the KS phenotype of *Dlx5*$^{-/-}$ mice, and that KS-causing mutations are by large loss-of-function, a direct link is difficult to envision. One possible scenario could be that miR targets might include a repressor TF, such as Foxp1. However this gene is expressed in the mesenchyme and thus cannot repress transcription in the OE or GnRH cells. Alternatively, the loss or reduced expression of a miR could result in mRNA depletion in case the miR normally prevents the annealing of another miR, more potent in destabilizing the target mRNA.

ACKNOWLEDGEMENTS

We thank Drs. Paolo Giacobini (INSERM, Lille France), Anna Cariboni and Roberto Maggi (Univ. of Milano, Italy), Ferdinando DiCunto (Univ. of Torino, Italy), Giovanni Levi (CNRS Paris. France), Nelly Pitteloud and Andrew Dwyer (CHUV, Endocrinology Dept., Lausanne, Switzerland) for helpful criticism on the manuscript. This work was supported by grants from the Italian Fondazione Telethon (GGP11097) and Fondazione Cassa di Risparmio di Torino.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could constitute a potential conflict of interest.

Supplementary Table 1. Genes most connected with (human) KS disease-genes, in the conserved coexpression network.

Entrez	Gene symbol	Gene name	N° of connected KS genes in the coexpression network
6711	*SPTBN1*	spectrin, beta, non-erythrocytic 1	6
28996	*HIPK2*	homeodomain interacting protein kinase 2	6
463	*ZFHX3*	zinc finger homeobox 3	5
4684	*NCAM1*	neural cell adhesion molecule 1	5
6638	*SNRPN*	small nuclear ribonucleoprotein polypeptide N	5
7204	*TRIO*	trio Rho guanine nucleotide exchange factor	5
54815	*GATAD2A*	GATA zinc finger domain containing 2°	5
800	*CALD1*	caldesmon 1	4
1084	*CEACAM3*	carcinoembryonic antigen-related cell adhesion molecule 3	4
4781	*NFIB*	nuclear factor I/B	4
6733	*SRPK2*	SRSF protein kinase 2	4
7145	*TNS1*	tensin 1	4
8871	*SYNJ2*	synaptojanin 2	4

Supplementary Table 1 continued..

9902	*MRC2*	mannose receptor, C type 2	4
10659	*CELF2*	CUGBP, Elav-like family member 2	4
23321	*TRIM2*	tripartite motif containing 2	4
23467	*NPTXR*	neuronal pentraxin receptor	4
27086	*FOXP1*	forkhead box P1	4
219771	*CCNY*	cyclin Y	4
375449	*MAST4*	microtubule associated serine/threonine kinase family member 4	4
667	*DST*	Dystonin	3
801	*CALM1*	calmodulin 1 (phosphorylase kinase, delta)	3
805	*CALM2*	calmodulin 2 (phosphorylase kinase, delta)	3
808	*CALM3*	calmodulin 3 (phosphorylase kinase, delta)	3
861	*RUNX1*	runt-related transcription factor 1	3
949	*SCARB1*	scavenger receptor class B, member 1	3
1000	*CDH2*	cadherin 2, type 1, N-cadherin (neuronal)	3
1009	*CDH11*	cadherin 11, type 2, OB-cadherin (osteoblast)	3
1112	*FOXN3*	forkhead box N3	3
1435	*CSF1*	colony stimulating factor 1 (macrophage)	3
1454	*CSNK1E*	casein kinase 1, epsilon	3
1462	*VCAN*	Versican	3
1903	*S1PR3*	sphingosine-1-phosphate receptor 3	3
2035	*EPB41*	erythrocyte membrane protein band 4.1 (elliptocytosis 1, RH-linked)	3
2119	*ETV5*	ets variant 5	3
2675	*GFRA2*	GDNF family receptor alpha 2	3
2737	*GLI3*	GLI family zinc finger 3	3
2776	*GNAQ*	guanine nucleotide binding protein (G protein), q polypeptide	3
3480	*IGF1R*	insulin-like growth factor 1 receptor	3
3678	*ITGA5*	integrin, alpha 5 (fibronectin receptor, alpha polypeptide)	3
4134	*MAP4*	microtubule-associated protein 4	3
4139	*MARK1*	MAP/microtubule affinity-regulating kinase 1	3
4155	*MBP*	myelin basic protein	3
4916	*NTRK3*	neurotrophic tyrosine kinase, receptor, type 3	3
5155	*PDGFB*	platelet-derived growth factor beta polypeptide	3
5793	*PTPRG*	protein tyrosine phosphatase, receptor type, G	3
6304	*SATB1*	SATB homeobox 1	3

Supplementary Table 1 continued..

6422	*SFRP1*	secreted frizzled-related protein 1	3
6546	*SLC8A1*	solute carrier family 8 (sodium/calcium exchanger), member 1	3
6659	*SOX4*	SRY (sex determining region Y)-box 4	3
6738	*TROVE2*	TROVE domain family, member 2	3
6854	*SYN2*	synapsin II	3
6925	*TCF4*	transcription factor 4	3
7078	*TIMP3*	TIMP metallopeptidase inhibitor 3	3
8106	*PABPN1*	poly(A) binding protein, nuclear 1	3
8997	*KALRN*	kalirin, RhoGEF kinase	3
9204	*ZMYM6*	zinc finger, MYM-type 6	3
9612	*NCOR2*	nuclear receptor corepressor 2	3
9826	*ARHGEF11*	Rho guanine nucleotide exchange factor (GEF) 11	3
10658	*CELF1*	CUGBP, Elav-like family member 1	3
10762	*NUP50*	nucleoporin 50kDa	3
10766	*TOB2*	transducer of ERBB2, 2	3
10893	*MMP24*	matrix metallopeptidase 24 (membrane-inserted)	3
10928	*RALBP1*	ralA binding protein 1	3
11096	*ADAMTS5*	ADAM metallopeptidase with thrombospondin type 1 motif, 5	3
11217	*AKAP2*	A kinase (PRKA) anchor protein 2	3
23164	*MPRIP*	myosin phosphatase Rho interacting protein	3
23363	*OBSL1*	obscurin-like 1	3
23365	*ARHGEF12*	Rho guanine nucleotide exchange factor (GEF) 12	3
25925	*ZNF521*	zinc finger protein 521	3
27327	*TNRC6A*	trinucleotide repeat containing 6°	3
54796	*BNC2*	basonuclin 2	3
55607	*PPP1R9A*	protein phosphatase 1, regulatory subunit 9A	3
55704	*CCDC88A*	coiled-coil domain containing 88A	3
58508	*MLL3*	myeloid/lymphoid or mixed-lineage leukemia 3	3
60685	*ZFAND3*	zinc finger, AN1-type domain 3	3
64784	*CRTC3*	CREB regulated transcription coactivator 3	3
80014	*WWC2*	WW and C2 domain containing 2	3
80036	*TRPM3*	transient receptor potential cation channel, subfamily M, member 3	3
80144	*FRAS1*	Fraser syndrome 1	3
84858	*ZNF503*	zinc finger protein 503	3
84937	*ZNRF1*	zinc and ring finger 1, E3 ubiquitin protein ligase	3

Supplementary Table 1 continued..

89796	*NAV1*	neuron navigator 1	3
91653	*BOC*	BOC cell adhesion associated, oncogene regulated	3
114299	*PALM2*	paralemmin 2	3
114757	*CYGB*	cytoglobin	3
124540	*MSI2*	musashi RNA-binding protein 2	3
259217	*HSPA12A*	heat shock 70kDa protein 12A	3
441024	*MTHFD2L*	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2-like	3
445815	*PALM2-AKAP2*	PALM2-AKAP2 readthrough	3

Supplementary Table 2. Genes causing a KS-like phenotype when mutated in mice.

Symbol	**Notes**
Prok2	the only true model of KS disease: lack of connection in the presence of normal olfactory neurogenesis/axonogenesis
Prok-R2	the only true model of KS disease: lack of connection in the presence of normal olfactory neurogenesis/axonogenesis
Dlx5	causes severe KS-like phenotype, defect of neurogenesis/axonogenesis
Emx2	causes KS-like phenotype
Lhx2	causes KS-like phenotype
FezF1	causes KS-like, expressed in the peripheral olfact. neurons.
Shep1	causes KS-like phenotype
Klf7	causes mild KS-like phenotype
Six1	causes KS-like phenotype
Ebf2	causes severe KS-like phenotype
Nrp1	causes KS-like phenotype
Sema3a	causes KS-like phenotype

Supplementary Table 3. Genes differentially expressed in the normal and *Dlx5*$^{-/-}$ developing murine olfactory system, prioritized using human KS-causing genes and the coexpression network.

Entrez	**Gene symbol**	**Gene name**	**p-value**
8490	*RGS5*	regulator of G-protein signaling 5	2.11 E+00
84870	*RSPO3*	R-spondin 3 homolog (Xenopus laevis)	2.11 E+00
4211	*MEIS1*	Meis homeobox 1	2.20 E+00
5727	*PTCH1*	patched homolog 1 (Drosophila)	2.20 E+00
9444	*QKI*	quaking homolog, KH domain RNA binding (mouse)	0.00012
2625	*GATA3*	GATA binding protein 3	0.00012

Supplementary Table 3 continued..

5015	*OTX2*	orthodenticle homeobox 2	0.00013
2150	*F2RL1*	coagulation factor II (thrombin) receptor-like 1	0.00019
7163	*TPD52*	tumor protein D52	0.00037
7022	*TFAP2C*	transcription factor AP-2 gamma (activating enhancer binding protein 2 gamma)	0.00039
8110	*DPF3*	D4, zinc and double PHD fingers, family 3	0.00058
57708	*MIER1*	mesoderm induction early response 1 homolog (Xenopus laevis)	0.00064
57685	*CACHD1*	cache domain containing 1	0.00111
2562	*GABRB3*	gamma-aminobutyric acid (GABA) A receptor, beta 3	0.00111
27071	*DAPP1*	dual adaptor of phosphotyrosine and 3-phosphoinositides	0.00213
1826	*DSCAM*	Down syndrome cell adhesion molecule	0.00258
11096	*ADAMTS5*	ADAM metallopeptidase with thrombospondin type 1 motif, 5	0.00676
92949	*ADAMTSL1*	ADAMTS-like 1	0.01008

Supplementary Table 4. Genes differentially expressed in lists B and C**,** prioritized using human KS-causing genes and the coexpression network. Notably, the search predicted two genes known to cause KS in the mouse, namely *Ebf2* and *Nrp1*.

List B

Entrez	**Gene symbol**	**Gene name**	**p-value**
1277	*COL1A1*	collagen, type I, alpha 1	1.60 E-05
1289	*COL5A1*	collagen, type V, alpha 1	1.60 E-05
1634	*DCN*	decorin	1.60 E-05
2252	*FGF7*	fibroblast growth factor 7 (keratinocyte growth factor)	1.60 E-05
284119	*PTRF*	polymerase I and transcript release factor	1.60 E-05
4147	*MATN2*	matrilin 2	1.60 E-05
1291	*COL6A1*	collagen, type VI, alpha 1	1.60 E-05
1293	*COL6A3*	collagen, type VI, alpha 3	1.60 E-05
7052	*TGM2*	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	8.12 E-03
1292	*COL6A2*	collagen, type VI, alpha 2	8.12 E-03
301	*ANXA1*	annexin A1	4.96 E-02
1805	*DPT*	dermatopontin	4.96 E-02
4015	*LOX*	lysyl oxidase	5.28 E-02
64641	*EBF2*	early B-cell factor 2	2.41 E-01
309	*ANXA6*	annexin A6	5.80 E-01

Supplementary Table 4 continued..

30846	*EHD2*	EH-domain containing 2	2.30 E+00
1803	*DPP4*	dipeptidyl-peptidase 4	2.30 E+00
387758	*FIBIN*	fin bud initiation factor homolog (zebrafish)	2.38 E+00
4915	*NTRK2*	neurotrophic tyrosine kinase, receptor, type 2	3.84 E+00
3371	*TNC*	tenascin C	3.95 E+00
347	*APOD*	apolipoprotein D	0.00092

List C

Entrez	**Gene symbol**	**Gene name**	**p-value**
1278	*COL1A2*	collagen, type I, alpha 2	2.46 E-09
92689	*FAM114A1*	family with sequence similarity 114, member A1	2.46 E-09
1277	*COL1A1*	collagen, type I, alpha 1	2.46 E-09
4147	*MATN2*	matrilin 2	2.46 E-09
2252	*FGF7*	fibroblast growth factor 7 (keratinocyte growth factor)	2.46 E-09
1291	*COL6A1*	collagen, type VI, alpha 1	2.46 E-09
1290	*COL5A2*	collagen, type V, alpha 2	2.46 E-09
1634	*DCN*	decorin	2.46 E-09
1289	*COL5A1*	collagen, type V, alpha 1	2.46 E-09
56944	*OLFML3*	olfactomedin-like 3	2.46 E-09
81792	*ADAMTS12*	ADAM metallopeptidase with thrombospondin type 1 motif, 12	2.46 E-09
284119	*PTRF*	polymerase I and transcript release factor	2.46 E-09
1293	*COL6A3*	collagen, type VI, alpha 3	2.46 E-09
4071	*TM4SF1*	transmembrane 4 L six family member 1	2.46 E-09
51313	*C4orf18*	chromosome 4 open reading frame 18	2.46 E-09
1303	*COL12A1*	collagen, type XII, alpha 1	2.46 E-09
2013	*EMP2*	epithelial membrane protein 2	2.46 E-09
112464	*PRKCDBP*	protein kinase C, delta binding protein	2.46 E-09
1307	*COL16A1*	collagen, type XVI, alpha 1	2.46 E-09
857	*CAV1*	caveolin 1, caveolae protein, 22kDa	2.46 E-09
151887	*CCDC80*	coiled-coil domain containing 80	6.64 E-05
7052	*TGM2*	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	6.64 E-05
1292	*COL6A2*	collagen, type VI, alpha 2	6.64 E-05
57333	*RCN3*	reticulocalbin 3, EF-hand calcium binding domain	6.64 E-05
2014	*EMP3*	epithelial membrane protein 3	9.80 E-05

Supplementary Table 4 continued..

309	*ANXA6*	annexin A6	9.80 E-05
64175	*LEPRE1*	leucine proline-enriched proteoglycan (leprecan) 1	9.80 E-05
8829	*NRP1*	neuropilin 1	9.80 E-05
59	*ACTA2*	actin, alpha 2, smooth muscle, aorta	9.80 E-05
26010	*SPATS2L*	spermatogenesis associated, serine-rich 2-like	1.71 E-04
2202	*EFEMP1*	EGF-containing fibulin-like extracellular matrix protein 1	1.71 E-04
302	*ANXA2*	annexin A2	1.71 E-04
4015	*LOX*	lysyl oxidase	1.71 E-04
8840	*WISP1*	WNT1 inducible signaling pathway protein 1	1.71 E-04
8974	*P4HA2*	prolyl 4-hydroxylase, alpha polypeptide II	1.71 E-04
3371	*TNC*	tenascin C	4.29 E-02
30846	*EHD2*	EH-domain containing 2	1.44 E+00
948	*CD36*	CD36 molecule (thrombospondin receptor)	1.44 E+00
22885	*ABLIM3*	actin binding LIM protein family, member 3	1.44 E+00
9358	*ITGBL1*	integrin, beta-like 1 (with EGF-like repeat domains)	1.44 E+00
1803	*DPP4*	dipeptidyl-peptidase 4	1.44 E+00
7099	*TLR4*	toll-like receptor 4	1.44 E+00
8436	*SDPR*	serum deprivation response (phosphatidylserine binding protein)	1.44 E+00
7010	*TEK*	TEK tyrosine kinase, endothelial	1.44 E+00
5268	*SERPINB5*	serpin peptidase inhibitor, clade B (ovalbumin), member 5	2.18 E+00
8490	*RGS5*	regulator of G-protein signaling 5	9.13 E+00
167681	*PRSS35*	protease, serine, 35	9.13 E+00
2246	*FGF1*	fibroblast growth factor 1 (acidic)	9.34 E+00
2934	*GSN*	gelsolin (amyloidosis, Finnish type)	9.34 E+00
1805	*DPT*	dermatopontin	9.34 E+00
301	*ANXA1*	annexin A1	9.34 E+00
85364	*ZCCHC3*	zinc finger, CCHC domain containing 3	9.78 E+00
4060	*LUM*	lumican	0.0001133
6423	*SFRP2*	secreted frizzled-related protein 2	0.0001199
3880	*KRT19*	keratin 19	0.0001199
4784	*NFIX*	nuclear factor I/X (CCAAT-binding transcription factor)	0.0001360
4915	*NTRK2*	neurotrophic tyrosine kinase, receptor, type 2	0.0002110
960	*CD44*	CD44 molecule (Indian blood group)	0.0002395
5272	*SERPINB9*	serpin peptidase inhibitor, clade B (ovalbumin), member 9	0.0002395
360	*AQP3*	aquaporin 3 (Gill blood group)	0.0002395

Supplementary Table 4 continued..

116039	*OSR2*	odd-skipped related 2 (Drosophila)	0.000315
3486	*IGFBP3*	insulin-like growth factor binding protein 3	0.000315
387758	*FIBIN*	fin bud initiation factor homolog (zebrafish)	0.000629
1131	*CHRM3*	cholinergic receptor, muscarinic 3	0.000682
2078	*ERG*	v-ets erythroblastosis virus E26 oncogene homolog (avian)	0.001003
5744	*PTHLH*	parathyroid hormone-like hormone	0.001003
23704	*KCNE4*	potassium voltage-gated channel, Isk-related family, member 4	0.001003
131	*ADH7*	alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide	0.001119
2052	*EPHX1*	epoxide hydrolase 1, microsomal (xenobiotic)	0.002516
1641	*DCX*	doublecortin	0.002661
10581	*IFITM2*	interferon induced transmembrane protein 2 (1-8D)	0.00469
1674	*DES*	desmin	0.005341
1280	*COL2A1*	collagen, type II, alpha 1	0.005642
30061	*SLC40A1*	solute carrier family 40 (iron-regulated transporter), member 1	0.009319
54873	*PALMD*	palmdelphin	0.012907
9509	*ADAMTS2*	ADAM metallopeptidase with thrombospondin type 1 motif, 2	0.012907
8862	*APLN*	apelin	0.012907
255488	*RNF144B*	ring finger protein 144B	0.012907

Supplementary Table 5. Genes differentially expressed in lists D and E**,** prioritized using human KS-causing genes and the coexpression network.

List D

Entrez	**Gene symbol**	**Gene name**	**p-value**
1277	*COL1A1*	collagen, type I, alpha 1	6.83 E-05
7052	*TGM2*	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	8.09 E-03
960	*CD44*	CD44 molecule (Indian blood group)	4.73 E+00
57211	*GPR126*	G protein-coupled receptor 126	6.83 E-05
1291	*COL6A1*	collagen, type VI, alpha 1	6.83 E-05
4147	*MATN2*	matrilin 2	6.83 E-05
1634	*DCN*	decorin	6.83 E-05
284119	*PTRF*	polymerase I and transcript release factor	6.83 E-05
2252	*FGF7*	fibroblast growth factor 7 (keratinocyte growth factor)	6.83 E-05
1292	*COL6A2*	collagen, type VI, alpha 2	8.09 E-03
2675	*GFRA2*	GDNF family receptor alpha 2	8.09 E-03

Supplementary Table 5 continued..

1293	*COL6A3*	collagen, type VI, alpha 3	6.83 E-05
301	*ANXA1*	annexin A1	8.09 E-03
1805	*DPT*	dermatopontin	8.09 E-03
3371	*TNC*	tenascin C	5.46 E+00
2078	*ERG*	v-ets erythroblastosis virus E26 oncogene homolog (avian)	0.000147
57118	*CAMK1D*	calcium/calmodulin-dependent protein kinase ID	0.000147
7163	*TPD52*	tumor protein D52	0.000175
9060	*PAPSS2*	3'-phosphoadenosine 5'-phosphosulfate synthase 2	0.000339
861	*RUNX1*	runt-related transcription factor 1	0.001244
10397	*NDRG1*	N-myc downstream regulated 1	0.001330
57471	*ERMN*	ermin, ERM-like protein	0.001330
22809	*ATF5*	activating transcription factor 5	0.005079

List E

Entrez	**Gene symbol**	**Gene name**	**p-value**
8490	*RGS5*	regulator of G-protein signaling 5	0.000273
7052	*TGM2*	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	0.001246
57211	*GPR126*	G protein-coupled receptor 126	1.95 E+00
284119	*PTRF*	polymerase I and transcript release factor	1.95 E+00
2252	*FGF7*	fibroblast growth factor 7 (keratinocyte growth factor)	1.95 E+00
1634	*DCN*	decorin	1.95 E+00
4915	*NTRK2*	neurotrophic tyrosine kinase, receptor, type 2	3.77 E+00
55450	*CAMK2N1*	calcium/calmodulin-dependent protein kinase II inhibitor 1	0.000405
1490	*CTGF*	connective tissue growth factor	0.000565
57670	*KIAA1549*	KIAA1549	0.001033
948	*CD36*	CD36 molecule (thrombospondin receptor)	0.001096
64641	*EBF2*	early B-cell factor 2	1.30 E+00
6423	*SFRP2*	secreted frizzled-related protein 2	1.66 E+00
3880	*KRT19*	keratin 19	1.66 E+00
2150	*F2RL1*	coagulation factor II (thrombin) receptor-like 1	1.66 E+00
10631	*POSTN*	periostin, osteoblast specific factor	1.95 E+00
857	*CAV1*	caveolin 1, caveolae protein, 22kDa	1.95 E+00
10512	*SEMA3C*	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	1.95 E+00

Supplementary Table 5 continued..

1303	*COL12A1*	collagen, type XII, alpha 1	1.95 E+00
8829	*NRP1*	neuropilin 1	2.18 E+00
59	*ACTA2*	actin, alpha 2, smooth muscle, aorta	2.18 E+00
2014	*EMP3*	epithelial membrane protein 3	2.18 E+00
25830	*SULT4A1*	sulfotransferase family 4A, member 1	3.77 E+00
4784	*NFIX*	nuclear factor I/X (CCAAT-binding transcription factor)	0.000118
5015	*OTX2*	orthodenticle homeobox 2	0.000239
7852	*CXCR4*	chemokine (C-X-C motif) receptor 4	0.000405
93664	*CADPS2*	Ca++-dependent secretion activator 2	0.000537
128553	*TSHZ2*	teashirt zinc finger homeobox 2	0.00053
1301	*COL11A1*	collagen, type XI, alpha 1	0.000565
7163	*TPD52*	tumor protein D52	0.000629
2052	*EPHX1*	epoxide hydrolase 1, microsomal (xenobiotic)	0.000921
1910	*EDNRB*	endothelin receptor type B	0.000921
80036	*TRPM3*	transient receptor potential cation channel, subfamily M, member 3	0.001011
54502	*RBM47*	RNA binding motif protein 47	0.001011
360	*AQP3*	aquaporin 3 (Gill blood group)	0.001419
116039	*OSR2*	odd-skipped related 2 (Drosophila)	0.001567
80731	*THSD7B*	thrombospondin, type I, domain containing 7B	0.001728
23704	*KCNE4*	potassium voltage-gated channel, Isk-related family, member 4	0.001728
2686	*GGT7*	gamma-glutamyltransferase 7	0.002173
861	*RUNX1*	runt-related transcription factor 1	0.002173
1809	*DPYSL3*	dihydropyrimidinase-like 3	0.002663
57471	*ERMN*	ermin, ERM-like protein	0.004343
10397	*NDRG1*	N-myc downstream regulated 1	0.004343
301	*ANXA1*	annexin A1	0.004916
3778	*KCNMA1*	potassium large conductance calcium-activated channel, subfamily M, alpha member 1	0.005003
9060	*PAPSS2*	3'-phosphoadenosine 5'-phosphosulfate synthase 2	0.005778
9358	*ITGBL1*	integrin, beta-like 1 (with EGF-like repeat domains)	0.005778
928	*CD9*	CD9 molecule	0.010787
22809	*ATF5*	activating transcription factor 5	0.016386

Supplementary Table 6. Putative KS disease-genes obtained by prioritizing lists of best-predicted *miR-9/9a* targets (**A**) or *miR200*-class targets (including both *miR-200a/141* and *miR-200bc/429/548a*) (**B**), using human KS-causing genes as reference. In both cases, only the top 50 ones are shown.

A.

Gene ID	Gene name	p-value
8503	PIK3R3; phosphoinositide-3-kinase, regulatory subunit 3 (gamma)	1.02 E-04
5286	PIK3C2A; phosphoinositide-3-kinase, class 2, alpha polypeptide	1.02 E-04
246175	CNOT6L; CCR4-NOT transcription complex, subunit 6-like	1.02 E-04
23095	KIF1B; kinesin family member 1B	1.02 E-04
9043	SPAG9; sperm associated antigen 9	1.02 E-04
23387	SIK3; SIK family kinase 3	1.20 E-04
10121	ACTR1A; ARP1 actin-related protein 1 homolog A, centractin alpha (yeast)	1.20 E-04
51341	ZBTB7A; zinc finger and BTB domain containing 7A	1.24 E-04
4781	NFIB; nuclear factor I/B	1.24 E-04
84445	LZTS2; leucine zipper, putative tumor suppressor 2	1.24 E-04
27303	RBMS3; RNA binding motif, single stranded interacting protein	1.27 E-04
91252	SLC39A13; solute carrier family 39 (zinc transporter), member 13	1.27 E-04
285590	SH3PXD2B; SH3 and PX domains 2B	1.27 E-04
9823	ARMCX2; armadillo repeat containing, X-linked 2	1.27 E-04
7837	PXDN; peroxidasin homolog (Drosophila)	1.27 E-04
5159	PDGFRB; platelet-derived growth factor receptor, beta polypeptide	1.27 E-04
25942	SIN3A; SIN3 homolog A, transcription regulator (yeast)	1.28 E-04
23551	RASD2; RASD family, member 2	1.41 E-04
8396	PIP4K2B; phosphatidylinositol-5-phosphate 4-kinase, type II, beta	1.41 E-04
1729	DIAPH1; diaphanous homolog 1 (Drosophila)	1.41 E-04
132864	CPEB2; cytoplasmic polyadenylation element binding protein 2	1.43 E-04
2926	GRSF1; G-rich RNA sequence binding factor 1	1.95 E-04
83892	KCTD10; potassium channel tetramerisation domain containing 10	2.03 E-04
27303	RBMS3; RNA binding motif, single stranded interacting protein	2.03 E-04
64375	IKZF4; IKAROS family zinc finger 4 (Eos)	2.03 E-04
2744	GLS; glutaminase	2.03 E-04
4082	MARCKS; myristoylated alanine-rich protein kinase C substrate	2.03 E-04
23095	KIF1B; kinesin family member 1B	2.03 E-04
23767	FLRT3; fibronectin leucine rich transmembrane protein 3	2.03 E-04
8087	FXR1; fragile X mental retardation, autosomal homolog 1	2.18 E-04
11052	CPSF6; cleavage and polyadenylation specific factor 6, 68kDa	2.18 E-04

Supplementary Table 6 continued..

257397	MAP3K7IP3; mitogen-activated protein kinase kinase kinase 7 interacting protein 3	2.18 E-04
23767	FLRT3; fibronectin leucine rich transmembrane protein 3	2.26 E-04
136319	MTPN; myotrophin	2.26 E-04
1848	DUSP6; dual specificity phosphatase 6	2.26 E-04
51232	CRIM1; cysteine rich transmembrane BMP regulator 1 (chordin-like)	2.29 E-04
55161	TMEM33; transmembrane protein 33	2.33 E-04
9341	VAMP3; vesicle-associated membrane protein 3 (cellubrevin)	2.33 E-04
154810	AMOTL1; angiomotin like 1	2.33 E-04
26100	WIPI2; WD repeat domain, phosphoinositide interacting 2	2.33 E-04
23168	RTF1; Rtf1, Paf1/RNA polymerase II complex component, homolog (S. cerevisiae)	2.33 E-04
55333	SYNJ2BP; synaptojanin 2 binding protein	2.33 E-04
5634	PRPS2; phosphoribosyl pyrophosphate synthetase 2	2.33 E-04
56061	UBFD1; ubiquitin family domain containing 1	2.33 E-04
5725	PTBP1; polypyrimidine tract binding protein 1	2.33 E-04
84444	DOT1L; DOT1-like, histone H3 methyltransferase (S. cerevisiae)	2.58 E-04
23186	RCOR1; REST corepressor 1	2.58 E-04
84445	LZTS2; leucine zipper, putative tumor suppressor 2	2.58 E-04
23187	PHLDB1; pleckstrin homology-like domain, family B, member 1	2.58 E-04
64319	FBRS; fibrosin	2.58 E-04

B.

Gene ID	**Gene name**	**p-value**
114885	OSBPL11; oxysterol binding protein-like 11	9.76347 E-06
79832	QSER1; glutamine and serine rich 1	9.76347 E-06
23168	RTF1; Rtf1, Paf1/RNA polymerase II complex component, homolog (S. cerevisiae)	9.76347 E-06
254048	UBN2; ubinuclein 2	9.76347 E-06
51592	TRIM33; tripartite motif-containing 33	9.76347 E-06
27125	AFF4; AF4/FMR2 family, member 4	9.76347 E-06
23169	SLC35D1; solute carrier family 35 (UDP-glucuronic acid/ UDP-N-acetylgalactosamine dual transporter), member D1	9.76347 E-06
6198	RPS6KB1; ribosomal protein S6 kinase, 70kDa, polypeptide 1	9.76347 E-06
54469	ZFAND6; zinc finger, AN1-type domain 6	9.76347 E-06
1456	CSNK1G3; casein kinase 1, gamma 3	9.76347 E-06
2744	GLS; glutaminase	9.76347 E-06

Supplementary Table 6 continued..

8880	FUBP1; far upstream element (FUSE) binding protein 1	9.76347 E-06
9637	FEZ2; fasciculation and elongation protein zeta 2 (zygin II)	9.76347 E-06
55074	OXR1; oxidation resistance 1	9.76347 E-06
23332	CLASP1; cytoplasmic linker associated protein 1	9.76347 E-06
25843	MOBKL3; MOB1, Mps One Binder kinase activator-like 3 (yeast)	9.76347 E-06
5937	RBMS1; RNA binding motif, single stranded interacting protein 1	9.76347 E-06
523	ATP6V1A; ATPase, H+ transporting, lysosomal 70kDa, V1 subunit A	9.76347 E-06
25842	ASF1A; ASF1 anti-silencing function 1 homolog A (S. cerevisiae)	9.76347 E-06
143684	FAM76B; family with sequence similarity 76, member B	9.76347 E-06
95681	TSGA14; testis specific, 14	9.76347 E-06
23435	TARDBP; TAR DNA binding protein	9.76347 E-06
6830	SUPT6H; suppressor of Ty 6 homolog (S. cerevisiae)	9.76347 E-06
58508	MLL3; myeloid/lymphoid or mixed-lineage leukemia 3	9.76347 E-06
220972	MARCH8; membrane-associated ring finger (C3HC4) 8	9.76347 E-06
9857	CEP350; centrosomal protein 350kDa	9.76347 E-06
9972	NUP153; nucleoporin 153kDa	9.76347 E-06
8833	GMPS; guanine monphosphate synthetase	9.76347 E-06
57122	NUP107; nucleoporin 107kDa	9.76347 E-06
1385	CREB1; cAMP responsive element binding protein 1	9.76347 E-06
9706	ULK2; unc-51-like kinase 2 (C. elegans)	9.76347 E-06
10915	TCERG1; transcription elongation regulator 1	9.76347 E-06
5529	PPP2R5E; protein phosphatase 2, regulatory subunit B', epsilon isoform	9.76347 E-06
91408	BTF3L4; basic transcription factor 3-like 4	9.76347 E-06
22823	MTF2; metal response element binding transcription factor 2	9.76347 E-06
5728	PTEN; phosphatase and tensin homolog	9.76347 E-06
5783	PTPN13; protein tyrosine phosphatase, non-receptor type 13 (APO-1/CD95 (Fas)-associated phosphatase)	9.76347 E-06
64864	RFX7; regulatory factor X, 7	9.76347 E-06
57478	USP31; ubiquitin specific peptidase 31	9.76347 E-06
1024	CDK8; cyclin-dependent kinase 8	9.76347 E-06
5326	PLAGL2; pleiomorphic adenoma gene-like 2	9.76347 E-06
11127	KIF3A; kinesin family member 3A	9.76347 E-06
29035	C16orf72; chromosome 16 open reading frame 72	9.76347 E-06
8816	DCAF5; DDB1 and CUL4 associated factor 5	9.76347 E-06
8936	WASF1; WAS protein family, member 1	9.76347 E-06

Supplementary Table 6 continued..

134957	STXBP5; syntaxin binding protein 5 (tomosyn)	9.76347 E-06
4684	NCAM1; neural cell adhesion molecule 1	9.76347 E-06
55288	RHOT1; ras homolog gene family, member T1	9.76347 E-06
6885	MAP3K7; mitogen-activated protein kinase kinase kinase 7	9.76347 E-06
153222	C5orf41; chromosome 5 open reading frame 41	9.76347 E-06

Supplementary Table 7. Semaphorin and semaphorin receptor/coreceptor genes, considered in our analysis for conserved Dlx5 binding sites and for miR seed sequences.

Ligands	
Semaphorin 3a	ID:20346
Semaphorin 3b	ID:20347
Semaphorin 3c	ID:20348
Semaphorin 3d	ID:108151
Semaphorin 3e	ID:20349
Semaphorin 3f	ID:20350
Semaphorin 3g	ID:218877
Semaphorin 4a	ID:20351
Semaphorin 4b	ID:20352
Semaphorin 4c	ID:20353
Semaphorin 4d	ID:20354
Semaphorin 4f	ID:20355
Semaphorin 4g	ID:26456
Semaphorin 5a	ID:20356
Semaphorin 5b	ID:20357
Semaphorin 6a	ID:20358
Semaphorin 6b	ID:20359
Semaphorin 6c	ID:20360
Semaphorin 6d	ID:214968
Semaphorin 7a	ID:20361
Collapsin response mediator protein	ID:12933
Receptors	
Plexin A1	ID:18844
Plexin A2	ID:18845
Plexin A3	ID:18846

Supplementary Table 7 continued..

Plexin A4	ID:243743
Plexin B1	ID:235611
Plexin B2	ID:140570
Plexin B3	ID:140571
Plexin C1	ID:54712
Plexin C2	ID:67448
Plexin D1	ID:67784
Plexin domain containing 1	ID:72324
Plexin-like	ID:104008
Neuropilin 1	ID:18186
Neuropilin 2	ID:18187
Coreceptors	
Met	ID:17295
ErbB2	ID:13866
VEGF receptor a	ID:22339
Integrin β1	ID:16412
Integrin αV	ID:16410

REFERENCES

1. Wang, Z., Liu, X., Yang, B. Z. and Gelernter, J. 2013, Front Genet., 4, 160.
2. Bromberg, Y. 2013, J. Mol. Biol., 425, 3993-4005.
3. Astic, L., Pellier-Monnin, V. and Godinot, F. 1998, Neuroscience, 84, 295-307.
4. Cariboni, A., Maggi, R. and Parnavelas, J. G. 2007, Trends Neurosci., 30, 638-644.
5. Forni, P. E., Taylor-Burds, C., Melvin, V. S., Williams, T. and Wray, S. 2011. J. Neurosci., 31, 6915-6927.
6. Tarozzo, G., Peretto, P. and Fasolo, A. 1995, Zoolog Sci., 12, 367-383.
7. Valdes-Socin, H., Rubio Almanza, M., Tome Fernandez-Ladreda, M., Debray, F. G., Bours, V. and Beckers, A. 2014, Front Endocrinol. (Lausanne), 5, 109.
8. Wray, S., Grant, P. and Gainer, H. 1989. Proc. Natl. Acad. Sci. USA, 86, 8132-8136.
9. Whitlock, K. E., Illing, N., Brideau, N. J., Smith, K. M. and Twomey, S. 2006, Mol. Cell Endocrinol., 254-255, 39-50.
10. Hanchate, N. K., Giacobini, P., Lhuillier, P., Parkash, J., Espy, C., Fouveaut, C., Leroy, C., Baron, S., Campagne, C., Vanacker, C., Collier, F., Cruaud, C., Meyer, V., Garcia-Pinero, A., Dewailly, D., Cortet-Rudelli, C., Gersak, K., Metz, C., Chabrier, G., Pugeat, M., Young, J., Hardelin, J. P., Prevot, V. and Dode, C. 2012. PLoS Genet., 8(8), e1002896.
11. Bonomi, M., Libri, D. V., Guizzardi, F., Guarducci, E., Maiolo, E., Pignatti, E., Asci, R. and Persani, L. 2012, Asian J. Androl., 14, 49-56.
12. Cadman, S. M., Kim, S. H., Hu, Y., Gonzalez-Martinez, D. and Bouloux, P. M. 2007, Horm. Res., 67, 231-242.
13. Cariboni, A. and Maggi, R. 2006, Cell Mol. Life Sci., 63, 2512-2526.
14. Dode, C. and Hardelin, J. P. 2009, Eur. J. Hum. Genet., 17, 139-146.
15. Hardelin, J. P. and Dode, C. 2008, Sex Dev., 2, 181-193.
16. Hu, Y., Tanriverdi, F., MacColl, G. S. and Bouloux, P. M. 2003, Int. J. Biochem. Cell Biol., 35, 1157-1162.

17. Kotan, L. D., Hutchins, B. I., Ozkan, Y., Demirel, F., Stoner, H., Cheng, P. J., Esen, I., Gurbuz, F., Bicakci, Y. K., Mengen, E., Yuksel, B., Wray, S. and Topaloglu, A. K. 2014, Am. J. Hum. Genet., 95, 326-331.
18. Miraoui, H., Dwyer, A. A., Sykiotis, G. P., Plummer, L., Chung, W., Feng, B., Beenken, A, Clarke, J., Pers, T. H., Dworzynski, P., Keefe, K., Niedziela, M., Raivio, T., Crowley, W. F. Jr., Seminara, S. B., Quinton, R., Hughes, V. A., Kumanov, P., Young, J., Yialamas, M. A., Hall, J. E., van Vliet, G., Chanoine, J. P., Rubenstein, J., Mohammadi, M., Tsai, P. S., Sidis, Y., Lage, K. and Pitteloud, N. 2013, Am. J. Hum. Genet., 92, 725-743.
19. Ragancokova, D., Rocca, E., Oonk, A. M., Schulz, H., Rohde, E., Bednarsch, J., Feenstra, I., Pennings, R. J., Wende, H. and Garratt, A. N. 2014, J. Clin. Invest., 124, 1214-1227.
20. Semple, R. K. and Topaloglu, A. K. 2010, Clin. Endocrinol. (Oxf), 72, 427-435.
21. Topaloglu, A. K. and Kotan, L. D. 2010, Curr. Opin. Obstet. Gynecol., 22, 264-270.
22. Laub, F., Dragomir, C. and Ramirez, F. 2006, Brain Res., 1103, 108-113.
23. Corradi, A., Croci, L., Broccoli, V., Zecchini, S., Previtali, S., Wurst, W., Amadio, S., Maggi, R., Quattrini, A. and Consalez, G. G. 2003, Development, 130, 401-410.
24. Berghard, A., Hagglund, A. C., Bohm, S. and Carlsson, L. 2012, FASEB J., 26, 3464-3472.
25. Cariboni, A., Davidson, K., Rakic, S., Maggi, R., Parnavelas, J. G. and Ruhrberg, C. 2011, Hum. Mol. Genet., 20, 336-344.
26. Hirata, T., Nakazawa, M., Yoshihara, S., Miyachi, H., Kitamura, K., Yoshihara, Y. and Hibi, M. 2006, Development, 133, 1433-1443.
27. Ikeda, K., Ookawara, S., Sato, S., Ando, Z., Kageyama, R. and Kawakami, K. 2007, Dev. Biol., 311, 53-68.
28. Levi, G., Puche, A. C., Mantero, S., Barbieri, O., Trombino, S., Paleari, L., Egeo, A. and Merlo, G. R. 2003, Mol. Cell Neurosci., 22, 530-543.
29. Long, J. E., Garel, S., Depew, M. J., Tobet, S. and Rubenstein, J. L. 2003, J. Neurosci., 23, 568-578.
30. Matsumoto, S., Yamazaki, C., Masumoto, K. H., Nagano, M., Naito, M., Soga, T., Hiyama, H., Matsumoto, M., Takasaki, J., Kamohara, M., Matsuo, A., Ishii, H., Kobori, M., Katoh, M., Matsushime, H., Furuichi, K. and Shigeyoshi, Y. 2006, Proc. Natl. Acad. Sci. USA, 103, 4140-4145.
31. Merlo, G. R., Mantero, S., Zaghetto, A. A., Peretto, P., Paina, S. and Gozzo, M. 2007, J. Mol. Histol., 38, 612-623.
32. Ng, K. L., Li, J. D., Cheng, M. Y., Leslie, F. M., Lee, A. G. and Zhou, Q. Y. 2005, Science, 308, 1923-1927.
33. Shimizu, T. and Hibi, M. 2009, Dev. Growth Differ, 51, 221-231.
34. Watanabe, Y., Inoue, K., Okuyama-Yamamoto, A., Nakai, N., Nakatani, J., Nibu, K., Sato, N., Iiboshi, Y., Yusa, K., Kondoh, G., Takeda, J., Terashima, T. and Takumi, T. 2009, J. Comp. Neurol., 515, 565-584.
35. Yoshida, M., Suda, Y., Matsuo, I., Miyamoto, N., Takeda, N., Kuratani, S. and Aizawa, S. 1997, Development, 124, 101-111.
36. Garaffo, G., Provero, P., Molineris, I., Pinciroli, P., Peano, C., Battaglia, C., Tomaiuolo, D., Etzion, T., Gothilf, Y., Santoro, M. and Merlo, G. R. 2013, Front Endocrinol. (Lausanne), 4, 203.
37. Doncheva, N. T., Kacprowski, T. and Albrecht, M. 2013, Wiley Interdiscip Rev. Syst. Biol. Med., 4, 429-442.
38. Masoudi-Nejad, A., Meshkin, A., Haji-Eghrari, B. and Bidkhori, G. 2012, Mol. Genet. Genomics, 87, 679-698.
39. Acampora, D., Merlo, G. R., Paleari, L., Zerega, B., Postiglione, M. P., Mantero, S., Bober, E., Barbieri, O., Simeone, A. and Levi, G. 1999, Development, 126, 3795-3809.
40. Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W. H., Pages, F., Trajanoski, Z. and Galon, J. 2009, Bioinformatics, 25, 1091-1093.

41. Piro, R. M. and Di Cunto, F. 2012, FEBS J., 279, 678-696.
42. Piro, R. M., Ala, U., Molineris, I., Grassi, E., Bracco, C., Perego, G. P., Provero, P. and Di Cunto, F. 2011, Eur. J. Hum. Genet., 19, 1173-1180.
43. Agostini, M., Tucci, P., Steinert, J. R., Shalom-Feuerstein, R., Rouleau, M., Aberdam, D., Forsythe, I. D., Young, K. W., Ventura, A., Concepcion, C. P., Han, Y. C., Candi, E., Knight, R. A., Mak, T. W. and Melino, G. 2011, Proc. Natl. Acad. Sci. USA, 108, 21099-21104.
44. Bian, S. and Sun, T. 2011, Mol. Neurobiol., 44, 359-373.
45. Brett, J. O., Renault, V. M., Rafalski, V. A., Webb, A. E. and Brunet, A. 2011, Aging (Albany NY), 3, 108-124.
46. Choi, P. S., Zakhary, L., Choi, W. Y., Caron, S., Alvarez-Saavedra, E., Miska, E. A., McManus, M., Harfe, B., Giraldez, A. J., Horvitz, H. R., Schier, A. F. and Dulac, C. 2008, Neuron, 57, 41-55.
47. Cui, Y., Xiao, Z., Han, J., Sun, J., Ding, W., Zhao, Y., Chen, B., Li, X. and Dai, J. 2012, BMC Neurosci., 13, 116.
48. Fiore, R., Khudayberdiev, S., Saba, R. and Schratt, G. 2011, Prog. Mol. Biol. Transl. Sci., 102, 47-100.
49. Li, Q., Bian, S., Hong, J., Kawase-Koga, Y., Zhu, E., Zheng, Y., Yang, L. and Sun, T. 2011, PLoS One, 6, e26000.
50. Olde Loohuis, N. F., Kos, A., Martens, G. J., van Bokhoven, H., Nadif Kasri, N. and Aschrafi, A. 2012, Cell Mol. Life Sci., 69, 89-102.
51. Farh, K. K., Grimson, A., Jan, C., Lewis, B. P., Johnston, W. K, Lim, L. P., Burge, C. B. and Bartel, D. P. 2005, Science, 310, 1817-1821.
52. Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T. 2009, Nucleic Acids Res., 37(Database issue), D105-110.
53. Chung, W. C. and Tsai, P. S. 2010, Front Horm. Res., 39, 37-50.
54. Forni, P. E., Bharti, K., Flannery, E. M., Shimogori, T. and Wray, S. 2013, J. Neurosci., 33, 19620-19634.
55. Restelli, M., Lopardo, T., Lo Iacono, N., Garaffo, G., Conte, D., Rustighi, A., Napoli, M., Del Sal, G., Perez-Morga, D., Costanzo, A., Merlo, G. R. and Guerrini, L. 2014, Hum. Mol. Genet., 23, 3830-3842.
56. Portales-Casamar, E., Thongjuea, S., Kwon, A. T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W. W. and Sandelin, A. 2010, Nucleic Acids Res., 38(Database issue), D105-110.
57. Vieux-Rochas, M., Bouhali, K., Mantero, S., Garaffo, G., Provero, P., Astigiano, S., Barbieri, O., Caratozzolo, M. F., Tullo, A., Guerrini, L., Lallemand, Y., Robert, B., Levi, G. and Merlo, G. R. 2013, PLoS One, 8, e51700.
58. Messina, A. and Giacobini, P. 2013, Front Endocrinol. (Lausanne), 4, 133.
59. Russell, M. K., Longoni, M., Wells, J., Maalouf, F. I., Tracy, A. A., Loscertales, M., Ackerman, K. G., Pober, B. R., Lage, K., Bult, C. J. and Donahoe, P. K. 2012, Proc. Natl. Acad. Sci. USA, 109, 2978-2983.
60. Gilman, S. R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M. and Vitkup, D. 2011, Neuron, 70, 898-907.
61. O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., Levy, R., Ko, A., Lee, C., Smith, J. D., Turner, E. H., Stanaway, I. B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J. M., Borenstein, E., Rieder, M. J., Nickerson, D. A., Bernier, R., Shendure, J. and Eichler, E. E. 2012, Nature, 485, 246-250.
62. Parikshak, N. N., Luo, R., Zhang, A., Won, H., Lowe, J. K., Chandran, V., Horvath, S. and Geschwind, D. H. 2013, Cell, 155, 1008-1021.
63. Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., Ercan-Sencicek, A. G., DiLullo, N. M., Parikshak, N. N., Stein, J. L., Walker, M. F., Ober, G. T., Teran, N. A., Song, Y., El-Fishawy, P., Murtha, R. C., Choi, M., Overton, J. D., Bjornson, R. D., Carriero, N. J., Meyer, K. A., Bilguvar, K., Mane, S. M., Sestan, N., Lifton, R. P., Gunel, M., Roeder, K., Geschwind, D. H., Devlin, B. and State, M. W. 2012, Nature, 485, 237-241.

64. Di Schiavi, E. and Andrenacci, D. 2013, Curr. Genomics, 14, 2-10.
65. Hoischen, A., Krumm, N. and Eichler, E. E. 2014, Nat. Neurosci., 17, 764-772.
66. Willsey, A. J., Sanders, S. J., Li, M., Dong, S., Tebbenkamp, A. T., Muhle, R. A., Reilly, S. K., Lin, L., Fertuzinhos, S., Miller, J. A., Murtha, M. T., Bichsel, C., Niu, W., Cotney, J., Ercan-Sencicek, A. G., Gockley, J., Gupta, A. R., Han, W., He, X., Hoffman, E. J., Klei, L., Lei, J., Liu, W., Liu, L., Lu, C., Xu, X., Zhu, Y., Mane, S. M., Lein, E. S., Wie, L., Noonan, J. P., Roeder, K., Devlin, B., Sestan, N. and State, M. W. 2013, Cell, 155, 997-1007.
67. Karreth, F. A. and Pandolfi, P. P. 2013, Cancer Discov., 3, 1113-1121.
68. Salmena, L., Poliseno, L., Tay, Y., Kats, L. and Pandolfi, P. P. 2011, Cell, 146, 353-358.
69. Abelson, J. F., Kwan, K. Y., O'Roak, B. J., Baek, D. Y., Stillman, A. A., Morgan, T. M., Mathews, C. A., Pauls, D. L., Rasin, M. R., Gunel, M., Davis, N. R., Ercan-Sencicek, A. G., Guez, D. H., Spertus, J. A., Leckman, J. F., Dure, L. St., Kurlan, R., Singer, H. S., Gilbert, D. L., Farhi, A., Louvi, A., Lifton, R. P., Sestan, N. and State, M. W. 2005, Science, 310, 317-320.
70. Liang, C., Zhu, H., Xu, Y., Huang, L., Ma, C., Deng, W., Liu, Y. and Qin, C. 2012, Brain Res., 1455, 103-113.
71. Liu, T., Wan, R. P., Tang, L. J., Liu, S. J., Li, H. J., Zhao, Q. H., Liao, W. P., Sun, X. F., Yi, Y. H. and Long, Y. S. 2014, Mol. Neurobiol., PMID 24906954
72. Gao, F. B. 2010, Neural Dev., 5, 25.
73. Schratt, G. M., Tuebing, F., Nigh, E. A., Kane, C. G., Sabatini, M. E., Kiebler, M. and Greenberg, M. E. 2006, Nature, 439, 283-289.
74. Shi, Y., Zhao, X., Hsieh, J., Wichterle, H., Impey, S., Banerjee, S., Neveu, P. and Kosik, K. S. 2010, J. Neurosci., 30, 14931-14936.
75. Baudet, M. L., Zivraj, K. H., Abreu-Goodger, C., Muldal, A., Armisen, J., Blenkiron, C., Goldstein, L. D., Miska, E. A. and Holt, C. E. 2011, Nat. Neurosci., 15, 29-38.