



# Systems genetics of nonsyndromic orofacial clefting provides insights into its complex aetiology

Zahra Razaghi-Moghadam<sup>1,2</sup> · Atefeh Namipashaki<sup>3</sup> · Saman Farahmand<sup>4</sup> · Naser Ansari-Pour<sup>1,5</sup>

Received: 27 August 2017 / Revised: 22 July 2018 / Accepted: 9 August 2018 / Published online: 25 September 2018  
© European Society of Human Genetics 2018

## Abstract

Nonsyndromic oral clefting (NSOC) is although one of the most common congenital disorders worldwide, its underlying molecular basis remains elusive. This process has been hindered by the overwhelmingly high level of heterogeneity observed. Given that hitherto multiple loci and genes have been associated with NSOC, and that complex diseases are usually polygenic and show a considerable level of missing heritability, we used a systems genetics approach to reconstruct the NSOC network by integrating human-based physical and regulatory interactome with whole-transcriptome microarray data. We show that the network component contains 53% (23/43) of the curated NSOC-implicated gene set and displays a highly significant propinquity ( $P < 0.0001$ ) between genes implicated at the genomic level and those differentially expressed at the transcriptome level. In addition, we identified bona fide candidate genes based on topological features and dysregulation (e.g. *ANGPTL4*), and similarly prioritised genes at GWA loci (e.g. *MYC* and *CREBBP*), thus providing further insight into the underlying heterogeneity of NSOC. Gene ontology analysis results were consistent with the NSOC network being associated with embryonic organ morphogenesis and also hinted at an aetiological overlap between NSOC and cancer. We therefore recommend this approach to be applied to other heterogeneous complex diseases to not only provide a molecular framework to unify genes which may seem as disparate entities linked to the same disease, but to also predict and prioritise candidate genes for further validation, thus addressing the missing heritability.

## Introduction

Nonsyndromic oral clefting (NSOC) is a spectrum of isolated prenatal developmental disorders with complex aetiologies, comprising genetic and environment factors along with possible interactions within and between, consistent with it being a complex genetic disease [1, 2].

Among congenital disorders, NSOC is one of the most prevalent worldwide with a prevalence of ~1.7/1000 births [3]. This relatively-high prevalence along with its clinical, economical and psychological effects on affected, has led to a plethora of studies trying to decipher the underlying aetiology of these disorders given that the genetic and environmental factors are largely undetermined. Many investigations, which started by the pioneer work of Ardinger et al [4]. roughly three decades ago, have been focused on identifying the genetic factors involved in NSOC. Hitherto, multiple genes have been associated with different types of NSOC, particularly with non-syndromic cleft lip with or without palate (NSCL/P) through candidate gene, GWA and whole exome-based studies [5]. Nevertheless, only a few genes that have been confirmed with the

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1038/s41431-018-0263-7>) contains supplementary material, which is available to authorized users.

---

✉ Naser Ansari-Pour  
n.ansaripour@ut.ac.ir

<sup>1</sup> Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran

<sup>2</sup> Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany

<sup>3</sup> Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences, Tehran, Iran

<sup>4</sup> College of Science and Mathematics, University of Massachusetts Boston, Boston, Massachusetts, USA

<sup>5</sup> Present address: Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7LF, UK

most prominent being *IRF6*, a gene confirmed to be strongly associated with NSOC [6, 7]. Nevertheless, even this finding has not been replicated in some populations [8, 9]. More astonishing is non-replication of studies within a single population. For example, we recently showed that *IRF6* is associated with NSCL/P in Iran [10], however, a previous study on the same population suggested otherwise [11]. These observations suggest that substantial heterogeneity exists in the genetics of this complex disease and seems to be unearthed at all levels including aetiological, locus and allelic heterogeneity [10].

In addition, gene–gene interactions have been reported for genes implicated in NSOC. These interactions, which have been shown either statistically [12, 13] or based on functional assays [6, 14, 15], suggest that the inter-relationship of NSOC genes is a major contributory factor of NSOC phenotypic variation. This level of complexity in the genetics of NSOC suggests that genetic analyses based on single genes is no longer effective in understanding the aetiology of these disorders since gene–gene interactions which are central to complex traits [16] are not addressed and considering the significant missing heritability [17], a move must be made towards identifying the underlying ‘network’ as proposed by Kousa and Schutte [18]. This systematic approach has been previously applied to a number of complex phenotypes from autism spectrum disorders [19] to severe spermatogenic failure [20]. For NSOC, however, studies have been limited to either single gene-centric analysis [21] of *IRF6* or gene ontology (GO) enrichment analysis based on NSOC-associated genes [22] or those common between orofacial clefting and neural tube defects [23]. No attempt has therefore been made to expand the gene regulatory network based on multi-omics integrated systems analysis to identify causative genes explaining the missing heritability and in turn deliver more effective diagnosis in the form of targeted gene sequencing panels. We therefore took a systems genetics approach similar to that of Ansari-Pour et al. [20] by integrating physical protein–protein and regulatory interactome data (with NSOC-implicated genes as network ‘seeds’) with whole-transcriptome data to not only generate a tissue-specific network explicating the underlying genetic heterogeneity of NSOC, but to also identify genes, modules and cellular functions likely to be key drivers in biological processes perturbed in this complex disease.

## Methods

### NSOC gene set extraction

Genes implicated in nonsyndromic oral clefts (NSOC) at the genomic level were manually curated by conducting a

systematic literature review by using the keyword “non-syndromic” with “oral cleft”, “cleft lip with or without palate”, “cleft lip”, “cleft palate” and “cleft lip/palate” and “cleft lip and palate” plus “gene”, “genetics”, “locus”, “mutation”, “polymorphism”, “variant”, “haplotype” and “GWAS” in PubMed and Google Scholar. To select seed genes at GWA loci, the gene closest to the top signal was reported. Manually curated GWA-based genes were then compared with entries in the GWAS Catalog (<https://www.ebi.ac.uk/gwas>, accessed on 18/12/2017) [24] to ensure completeness and consistency. Interestingly, of the twenty GWA-based genes, thirteen had been previously associated with NSOC (see Supplementary Table S1). For gene-wide association studies, three quality control criteria were applied, namely a sample size larger than 200, significance remaining after correction for multiple testing and, under a case-control setting, the control group not significantly deviating from Hardy-Weinberg equilibrium expectations at the associated SNP [25]. For genes identified by SNP association (both gene-wide and genome-wide), only those with at least one independent replication were retained. The final gene list fulfilling the above criteria tallied 43 (see Supplementary Table S1).

### Microarray whole-transcriptome data

The Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) database was searched for whole-transcriptome microarray datasets on NSOC. Surprisingly, only one dataset based on NSCL/P samples was available (GEO accession number GSE42589) [26] which comprised seven NSCL/P and six control dental pulp stem cell samples. The dataset was normalised with RMA [27] and expression values were obtained in Log<sub>2</sub> format. Genes were considered not expressed if their median expression level in the control group was below the minimum of median expression levels observed among the seed genes and were removed from the dataset. Given that only one dataset was available, we used a combination of four different well-known statistical methods to identify differentially expressed genes (DEGs) with a higher confidence. The independent *t*-test and the Mann–Whitney *U*-test (both with multiple testing correction by permutation) were undertaken independently. Furthermore, we used the Significance Analysis of Microarrays (SAM) method [28] and a geometrical multivariate approach, namely the Characteristic Direction [29] to also select DEGs. Finally, based on the four independent lists of DEGs obtained, we employed the Robust Rank Aggregation (RRA) method [30] to acquire a single ranked DEG list. The RRA method uses a probabilistic model for aggregation that is robust to noise and also facilitates the calculation of significance probabilities for all the genes in the final ranking. Given that

based on conventional statistical analyses in DEG identification, a large number of genes (sometimes in thousands) show significant expression change [31], we limited the DEGs to the top 500 as a conservative measure. This set was then filtered based on expression fold-change, calculated as  $2^{(\text{mean expression value in disease state}/\text{mean expression value in control state})}$ , and only those that satisfied a fold-change cut-off of 1.7 were retained. All of the above analyses were implemented within the R environment for statistical computing (<http://www.r-project.org/>).

### Global interactome data

The global protein–protein interaction (PPI) network used was constructed based only on physical interactions reported in BioGRID (limited to human interactions only) [32] and HPRD [33] to form the global interactome in this study. This merged interactome consisted of 16,629 nodes and 22,1831 edges, all of which are based on experimental evidence.

### TREEDUST regulatory network generation

We converted all transcription factor–target interactions reported in the TRED database [34] into text format and then merged with those interactions reported in the TRRUST database [35] to generate the comprehensive TREEDUST human regulatory interaction network (meta-dataset available upon request). The generated network was used to identify downstream targets of TFs identified in the PPI network. In addition, wherever a target was also a TF and differentially expressed, its DEG targets were also added to the network. Next, we conducted a literature search to specifically mine functionally validated regulatory interactions among the curated seeds potentially missing from TREEDUST. These pairwise interactions were combined with the physical PPI and TREEDUST interactions for network visualisation, however, given the bias in reported interactions in the literature, these interactions were excluded from any topological analysis.

### Network reconstruction and topological analysis

We used the approach of Ansari-Pour et al. [20] to identify the local network of the seed nodes and their interactions with DEGs, which were identified from the whole-transcriptome microarray dataset. This approach in network reconstruction is based on the ‘local hypothesis’ of disease aetiology [36]. Networks were visualised in Gephi 0.9.1 (<http://gephi.github.io/>) [37]. In a given network, nodes and edges are representative of proteins and pairwise interactions, respectively. Directed and undirected edges represent

regulatory and physical interactions respectively. The NSOC network component was clustered using the fast unfolding clustering algorithm [38] to identify modules within the network. For each node, two critical topological parameters were calculated in Cytoscape [39]. First, degree, the number of interactions, was calculated. Next, betweenness centrality, a measure expressing the importance of a node in intermodular crosstalk, was computed. Modules were named based on the representative node, which displayed the highest betweenness centrality. Shortest path (SP), an algorithm in identifying the shortest distance between any two nodes in a given network [40], was also implemented to compute the distance of all one-to-one interactions among all seeds and DEGs, thus examining the level of functional association between the DEG and seed gene sets.

### Functional enrichment analysis

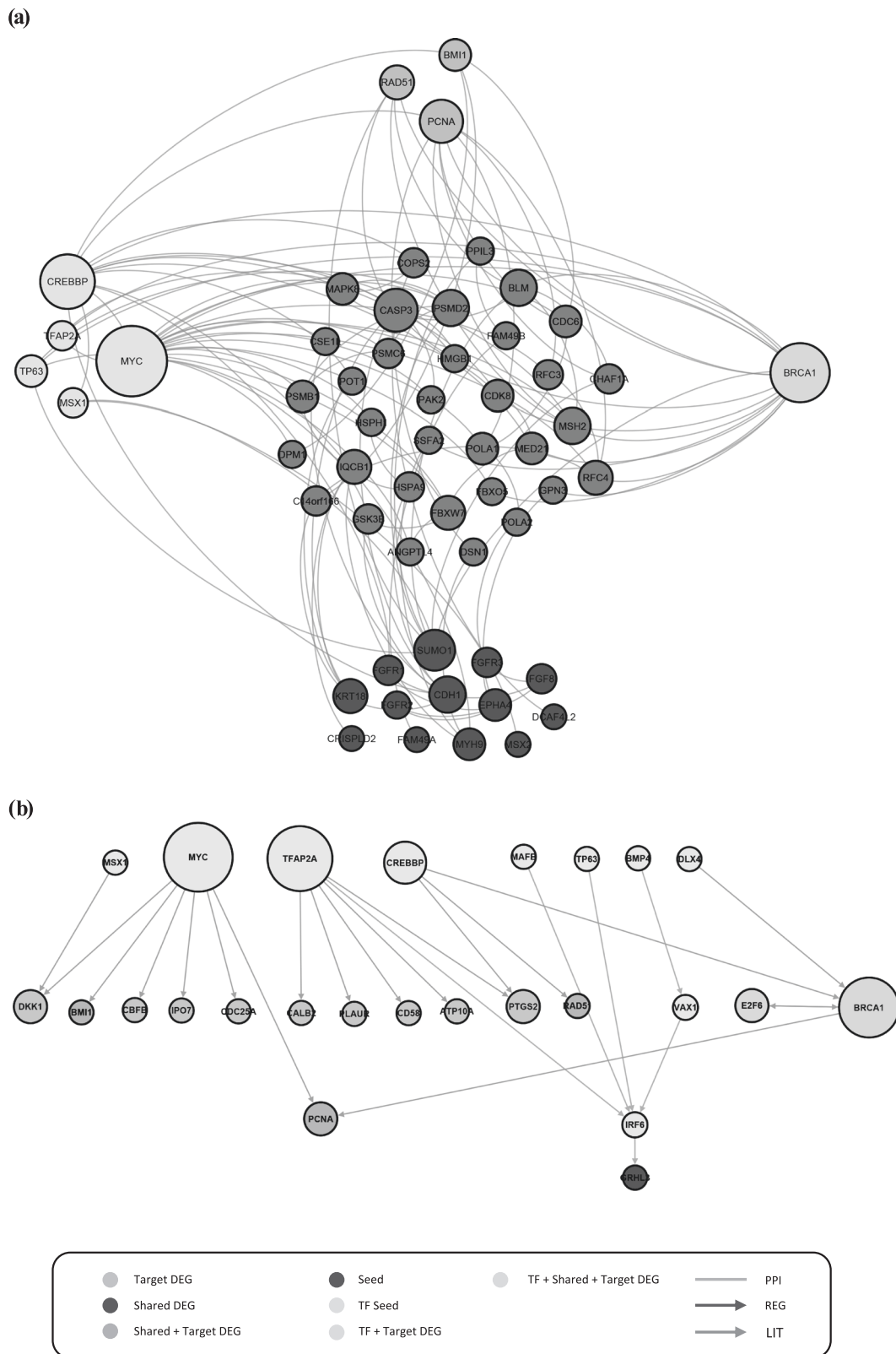
The NSOC network and its constituent modules were analysed for functional enrichment of gene ontology and pathway terms by using PANTHER [41] and based on a custom reference gene set comprising the unique set of genes with reported interactions in either the merged interactome or TREEDUST. Enrichment significance values were corrected for multiple testing by the FDR method and terms covered with over 10% of NSOC genes were retained to exclude terms overlapping scantily with the NSOC network. To potentially identify more unique terms associated with the functional units of the reconstructed network, enrichment analysis was also undertaken for each module of the network independently and terms enriched solely in each module (i.e. not enriched in the overall analysis) were sought.

## Results

### NSOC network reconstruction

Initially, we scoured the merged PPI interactome to identify interactions among the genes implicated in NSOC (i.e. seeds). Of the total 43 seeds, 40 were present in the merged PPI interactome, of which 19 interacted physically with at least one other seed. Interestingly, 15 (79%) of these formed a single network component.

To expand the NSOC network based on dys-regulated genes, we identified DEGs from the whole-transcriptome microarray dataset, tallying to a total of 445 genes. Of these 43 were not mapped to any protein in the PPI interactome and none were among the seeds. To construct the global NSOC network, we sought to identify the interactions between seeds and DEGs using SP. Interestingly, SP ranged



**Fig. 1** Graphical representation of **a** physical protein-protein and **b** regulatory interactions in the local NSOC network. Node colours represent the combination of the source and regulatory attributes of a given gene included in the network. Node size is proportionate to

degree. TF transcription factor, PPI physical protein–protein interaction, REG TREEDUST regulatory interaction, LIT literature-based regulatory interaction. All pairwise interactions in this network are given in Supplementary Table S2

**Table 1** Details of the three hub cut-vertices in the NSOC network and their dependent private neighbours

Cut-vertex	Cut-vertex source	Disconnected gene(s)	Type of interaction
BRCA1	DEG	E2F6, DLX4	REG,REG
MYC	Seed	CBFB, CDC25A, DCAF4L2, IPO7	REG,REG,PPI,REG
TFAP2A	Seed	ATP10A, CALB2, CD58, PLAUR	REG,REG,REG,REG

*PPI* physical protein–protein interaction, *REG* TREEDUST regulatory interaction

from one to six for all pairwise interactions with minimum SP ranging one to three between any DEG and the seed set. More specifically, of the 402 DEGs in the PPI interactome, 57 directly interacted with the seed gene set, 318 were two-step neighbours, and the remaining 27 were three-step neighbours (mean SP = 1.86). We then assessed whether this level of proximity is a non-random observation in the human PPI, we generated a null distribution of ‘mean SP’ based on 10,000 random 402-node gene sets sampled from the global set of expressed genes. The mean of ‘mean SP’ observed between the seeds and the random sets was 21.87 and the propinquity of seed-DEG interactivity was highly significant ( $P < 0.0001$ ). To focus on interactions with more significant functional relevance, we further examined the interactions of directly interacting DEGs. For this, all three-node (V shape) and four-node (U shape) sub-graphs (see Ansari-Pour et al. [20]) of seeds were systematically identified in the global interactome. Of the 57 one-step DEG neighbours, 38 were among those forming V and U sub-graphs with the seeds, suggesting their strong bilateral functional associations. We then merged the seeds with the 38 shared DEGs, as genes with the highest interactivity with seeds, to undertake network analysis. After merging, an extra three seeds were connected to the initial seed component through the DEGs alone, resulting in the presence of 22 interconnected seeds (see Fig. 1a for all physical interactions). We then calculated the number of interactions of each shared DEG with the seed gene set not to only identify the level of connection heterogeneity among the DEGs, but to also identify highly interacting DEGs. The mean number of connections was  $1.45 \pm 0.86$  however, the range of these interactions was 1–5, displaying considerable connection heterogeneity. Subsequently, one hub-like DEG was found with 5 interactions, namely *BRCA1*. Next, to expand the NSOC network by integrating regulatory interactions resulting in expression dysregulation, the TREEDUST dataset was used to identify TFs in the NSOC network. Of the 11 TF seeds, five had at least one regulatory interaction with a DEG, tallying 15 interactions. These five TF seeds directly targeted 13 unique DEGs, of which one (*BRCA1*) was also a TF and regulated two downstream DEGs (see Fig. 1b). Interestingly, of these target DEGs, four overlapped with the 38 shared DEGs of which the only one with downstream regulatory interactions was *BRCA1*. *DLX4*, one of the five TF seeds, which regulates *BRCA1* but has no

reported physical interaction was also added to the network as a seed (see Fig. 1b). The NSOC network thus comprised 23 seeds, 38 shared DEGs and 10 DEGs unique to regulatory interactions, and a total of 139 pairwise interactions. Moreover, we identified a total of six well-defined interactions among the seeds which were absent in the merged PPI interactome and TREEDUST (i.e. green edges in Fig. 1b; Supplementary Table S2). Although these interactions represented only 4% of the network edges, they were removed when calculating topological parameters due to the inherent overrepresentation bias of reported interactions in the literature.

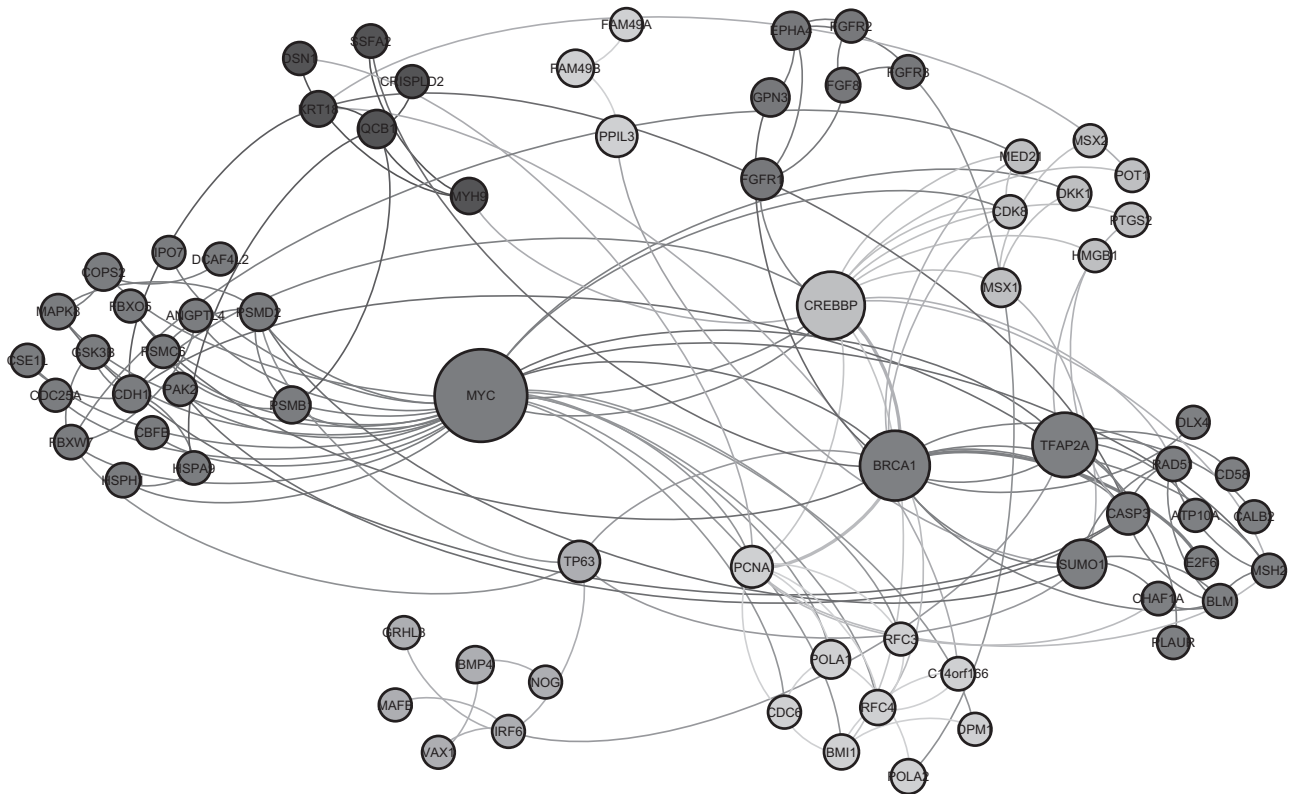
## Network analysis

All nodes in the NSOC network were analysed topologically based on their degree and betweenness centrality. Among all nodes, MYC, CREBBP, BRCA1, and TFAP2A not only had the highest betweenness centrality in the network ( $>0.20$ ), but they were also had high degrees (ranges of 8–27), thus being crucial hubs in the network (see Supplementary Table S3). We then analysed which of these were cut-vertices to identify which nodes in the network solely interact with these highly central nodes. Interestingly, BRCA1, MYC, and TFAP2A were cut-vertices, thus further emphasising the importance of the topological position of these central genes. A total of 10 genes were disconnected by removing the cut-vertices with each cut-vertex disconnecting 2–4 genes from the network (Table 1; Supplementary Figures S1–S3).

Next, to identify potential functional units in the network, we clustered the network based on all edges and identified eight clusters within the NSOC network (see Fig. 2). The number of nodes per module ranged from 3 to 19.

## Gene enrichment analysis

Based on the NSOC network gene set (Supplementary Table S4), we identified 319 enriched terms (see Supplementary Table S5). The top terms with the highest fold enrichment were all biological process terms associated with either embryonic limb/appendage morphogenesis or with DNA replication. Consistently, the top enriched pathways were Hedgehog signalling and DNA replication.



**Fig. 2** Graphical representation of the nine modules within the NSOC network. Modularity class and betweenness centrality can be distinguished by node colour and node size respectively. The node with

the highest betweenness centrality in each module is brought to the centre of the network. Edges (interactions) are coloured based on the combination of the colours of any two nodes

Enrichment analysis based on individual network modules did not result in the enrichment of any further terms.

## Discussion

Isolated NSOC is not only the most common nonsyndromic congenital deformity in humans, but it is also a highly heterogeneous disorder with a complex aetiology. To address this complexity, we undertook a systems genetics approach by conducting network analysis of integrated multi-omics data. Given that a gene is generally either functionally disrupted at the protein level or dysregulated at the transcript level, NSOC-implicated genes identified by variant screening and/or genetic association studies were combined with differentially expressed genes to reconstruct the local network of NSOC (comprising 101 nodes and 9 modules). The NSOC network can explain the likely functional relationships of 23 of the 43 curated NSOC-implicated genes by being unified in a highly connected dense network, many of which may seem as disparate entities associated with the same disorder. Interestingly, all but one connected via a physical interaction, thus

emphasising the importance of integrating physical interactome data in multi-omics studies.

A key issue in whole-transcriptome studies is that the most significantly differentially expressed genes identified either do not overlap or have no plausible association with confirmed genes in a particular disease. Also, in this specific case, the microarray transcriptome dataset used was based on dental pulp stem cells [26], which do not accurately represent the NSOC-affected tissue. Given the difficulty in accurate sampling of the developing palatal shelves, this remains the only published dataset associated with NSOC. However, to test the relevance of this dataset for this analysis, we examined the level of interaction between DEGs and the seed gene set using SP. Although any given DEG could have had an SP value of infinity (i.e. impossible interaction given that the global interactome comprises 15 disconnected super-components) to the seed gene set, we observed a highly significant non-random co-locality of DEGs and seeds in the network (SP ranging from 1 to 3,  $P < 0.0001$ ). According to the 'local' hypothesis in network medicine [36], this robust functional association of DEGs make them functionally relevant to NSOC and therefore the dataset that they emanate

from may be deemed functionally relevant. Moreover, this suggests that a computational systems approach is superior in identifying the most functionally relevant DEGs when compared with the most common status-quo approach of conducting GO analysis on the overall DEG list, since one would expect a ‘molecular’ association among perturbed genes. It should be noted that, based on the FC range, down-regulation of implicated DEGs is relatively low in magnitude (max of fivefold down-regulation). This may be due to the low sample size in the whole-transcriptome experiment. However, a recent report showed that low expression changes of key genes significantly affect facial morphogenesis [14]. Thus, it is likely that the observed FC range may still have a significant impact on NSOC development.

### Identification of bona fide candidate genes

We used the NSOC network as a framework to rank DEGs based on their topological features. *ANGTPL4*, *BRCA1* and *PCNA* were the DEGs with the most favourable topological features in the network. For instance, *ANGTPL4* which has been implicated in disrupting endothelial cell–cell adhesion [42] is not only the second most dysregulated gene in the whole-transcriptome, but it also interacts directly with *MYC*, thus making it a bona fide candidate gene in NSOC development. More interestingly, another top DEG, *BRCA1* (the highly penetrant breast cancer predisposition gene), not only interacts physically with five seeds, but it is also regulated directly by one seed (see Supplementary Figure S1), thus displaying a strong functional association with NSOC. This, however, does not necessarily imply that *BRCA1* is an upstream aetiological factor. Rather, it is more plausible that variants in its regulatory factors such as that observed in *DLX4* may have dysregulated its expression. Although, due to the paucity of exome studies, there is no report of *BRCA1* variants in NSOC patients, no study has also shown *BRCA1*-positive breast cancer patients to have had NSOC in their childhood. Nonetheless, the potential effects it may have on its physical seed interactors can not be ruled out.

### The aetiological overlap of NSOC with cancer

A number of epidemiological studies have suggested an association between different types of cancer and subtypes of nonsyndromic NSOC with the most reliable and significant association reported with breast cancer [43]. This suggestion has been further strengthened by recent genetic studies showing an overlap of predisposition genes [44, 45] and GWA signals [46] between cancer and NSOC. The presence of key cancer genes in the network as hub cut-vertices (e.g. *BRCA1*) and enriched DNA

replication and repair GO and KEGG terms, support the aetiological overlap of these concurrent diseases.

### Prioritisation of GWAS genes

Among all genes identified through GWAS only (see Supplementary Table S1), we prioritised *MYC* as the associated gene at 8q24.21. This is due to the highly favourable topological features of this gene in the network than any other flanking gene. Also, given the suggested aetiological overlap of NSOC with cancer and the genome-wide significant association of multiple cancers with this locus [47], this prioritisation seems more palpable. Interestingly, this finding is further supported with a recent functional study showing that *MYC* regulation by enhancers in the 8q24.21 locus is essential for normal facial morphogenesis [48]. Variant detection studies in large cohorts of patients may validate this finding and explain, to some extent, the missing heritability of syndromic and nonsyndromic forms of orofacial cleft, especially in syndromic patients – known for their strong genetic predisposition – that are *IRF6*-negative.

### Limitations

The global physical interactome and the TREEDUST regulatory interactome used here are inherently incomplete and missing interactions are likely. Consistently, of the six interactions identified in the literature among seeds, none were reported in the combined physical interactome even though all nodes were present in both the PPI and TREE-DUST datasets. Furthermore, the five interactions that we identified for *IRF6* are not reported in either of the two datasets for this well-studied NSOC gene. For instance, one of the best established interactions has been demonstrated between *TFAP2A* and *IRF6* (with rs642961 as the only aetiological SNP at the *IRF6* locus present in the upstream AP-2 $\alpha$  binding site) [6], however, this interaction was among those missing. This shows that further manual curation of the literature is required to update the current physical/regulatory interactome. Another limitation of multi-omics integrated systems analysis for NSOC is the inaccessibility of tissue samples representing the developing palatal shelves. This may become scarcely available with sporadic research-donated foetal tissue samples. However, the human developmental biology resource (HDBR) [49], which is an ongoing effort in human embryonic and foetal tissue collection for scientific research, is likely to resolve this issue in a more systematic way.

There is some evidence for subtype-specific genes in NSOC [12], thus warranting independent systems analysis of each subtype. However, the number of such genes is currently not sufficient to allow appropriate network analysis. Nevertheless, increase in the identification of subtype-specific

genes in the future will create the potential to reconstruct gene networks of specific NSOC subtypes, which are known to be aetiologically distinct [22].

## Conclusion

In conclusion, we show that the reconstructed local network of disease aetiology is a solid molecular framework to address missing heritability of complex diseases by identifying potential bona fide candidate genes and prioritise genes in GWA loci based on topological parameters and the level of dysregulation. It has been recently proposed that mouse models, due to their high similarity in gene signatures, may be a useful tool to unravel the underlying heterogeneity of NSOC given that human studies are cumbersome [23]. However, we show that this approach is able to tackle this issue based on human data and importantly, does not require the translational step, which is often not straightforward. A targeted gene sequencing panel custom-designed to include the genes comprising the reconstructed NSOC network would be able to validate the extent of genetic heterogeneity explicated here.

**Acknowledgements** This study was supported by the Iran National Science Foundation (INSF) with reference 92041708.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Dixon MJ, Marazita ML, Beaty TH, Murray JC. Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet.* 2011;12:167–78.
- Wyszynski DF, Duffy DL, Beaty TH. Maternal cigarette smoking and oral clefts: a meta-analysis. *Cleft Palate Craniofac J.* 1997;34:206–10.
- Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. Cleft lip and palate. *Lancet.* 2009;374:1773–85.
- Ardinger HH, Buetow KH, Bell GI, Bardach J, VanDemark DR, Murray JC. Association of genetic variation of the transforming growth factor- $\alpha$  gene with cleft lip and palate. *Am J Hum Genet.* 1989;45:348–53.
- Leslie EJ, Marazita ML. Genetics of orofacial cleft birth defects. *Curr Genet Med Rep.* 2015;3:118–26.
- Rahimov F, Marazita ML, Visel A, et al. Disruption of an AP-2 $\alpha$  binding site in an IRF6 enhancer is associated with cleft lip. *Nat Genet.* 2008;40:1341–7.
- Zucchero TM, Cooper ME, Maher BS, et al. Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. *N Engl J Med.* 2004;351:769–80.
- Paranaiba LM, Bufalino A, Martelli-Junior H, de Barros LM, Graner E, Coletta RD. Lack of association between IRF6 polymorphisms (rs2235371 and rs642961) and non-syndromic cleft lip and/or palate in a Brazilian population. *Oral Dis.* 2010;16:193–7.
- Pegelow M, Koillinen H, Magnusson M, et al. Association and mutation analyses of the IRF6 gene in families with nonsyndromic and syndromic cleft lip and/or cleft palate. *Cleft Palate Craniofac J.* 2014;51:49–55.
- Kerameddin S, Namipashaki A, Ebrahimi S, Ansari-Pour N. IRF6 is a marker of severity in nonsyndromic cleft lip/palate. *J Dent Res.* 2015;94:226S–232S.
- Nouri N, Memarzadeh M, Carinci F, et al. Family-based association analysis between nonsyndromic cleft lip with or without cleft palate and IRF6 polymorphism in an Iranian population. *Clin Oral Investig.* 2015;19:891–4.
- Carlson JC, Standley J, Petrin A, et al. Identification of 16q21 as a modifier of nonsyndromic orofacial cleft phenotypes. *Genet Epidemiol.* 2017;41:887–97.
- Velazquez-Aragon JA, Alcantara-Ortigoza MA, Estandia-Ortega B, et al. Association of interactions among the IRF6 gene, the 8q24 region, and maternal folic acid intake with non-syndromic cleft lip/palate in Mexican Mestizos. *Am J Med Genet A.* 2012;158A:3207–10.
- Green RM, Feng W, Phang T, et al. Tfp2a-dependent changes in mouse facial morphology result in clefting that can be ameliorated by a reduction in Fgf8 gene dosage. *Dis Model Mech.* 2015;8:31–43.
- McDade SS, Henry AE, Pivato GP, et al. Genome-wide analysis of p63 binding sites identifies AP-2 factors as co-regulators of epidermal differentiation. *Nucleic Acids Res.* 2012;40:7190–206.
- Civelek M, Lusk AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet.* 2014;15:34–48.
- Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461:747–53.
- Kousa YA, Schutte BC. Toward an orofacial gene regulatory network. *Dev Dyn.* 2016;245:220–32.
- Li J, Shi M, Ma Z, et al. Integrated systems analysis reveals a molecular network underlying autism spectrum disorders. *Mol Syst Biol.* 2014;10:774.
- Ansari-Pour N, Razaghi-Moghadam Z, Barneh F, Jafari M. Testis-specific Y-centric protein-protein interaction network provides clues to the etiology of severe spermatogenic failure. *J Proteome Res.* 2016;15:1011–22.
- Dai J, Yu H, Si J, Fang B, Shen SG. Irf6-related gene regulatory network involved in palate and lip development. *J Craniofac Surg.* 2015;26:1600–5.
- Funato N, Nakamura M. Identification of shared and unique gene families associated with oral clefts. *Int J Oral Sci.* 2017;9:104–9.
- Kousa YA, Mansour TA, Seada H, Matoo S, Schutte BC. Shared molecular networks in orofacial and neural tube development. *Birth Defects Res.* 2017;109:169–79.
- MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* 2017;45: D896–D901.
- Namipashaki A, Razaghi-Moghadam Z, Ansari-Pour N. The essentiality of reporting Hardy-Weinberg equilibrium calculations in population-based genetic association studies. *Cell J.* 2015;17:187.
- Kobayashi GS, Alvizi L, Sunaga DY, Francis-West P, Kuta A, Almada BV, et al. Susceptibility to DNA damage as a molecular mechanism for non-syndromic cleft lip and palate. *PLoS One* 2013;8:e65677.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249–264.
- Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA.* 2001;98:5116–21.



29. Clark NR, Hu KS, Feldmann AS, Kou Y, Chen EY, Duan Q, et al. The characteristic direction: a geometrical approach to identify differentially expressed genes. *BMC Bioinformatics* 2014;15:79.
30. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 2012;28:573–80.
31. Wang H, Qiu T, Shi J, Liang J, Wang Y, Quan L, et al. Gene expression profiling analysis contributes to understanding the association between non-syndromic cleft lip and palate, and cancer. *Mol Med Rep* 2016;13:2110–2116.
32. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, et al. The BioGRID interaction database: 2015 update. *Nucleic Acids Res* 2015;43:D470–478.
33. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human Protein Reference Database–2009 update. *Nucleic Acids Res* 2009;37:D767–772.
34. Jiang C, Xuan Z, Zhao F, Zhang MQ. TRED: a transcriptional regulatory element database, new entries and other development. *Nucleic Acids Res*. 2007;35:D137–D140.
35. Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, et al. TRRUST: a reference database of human transcriptional regulatory interactions. *Sci Rep* 2015;5:11432.
36. Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011;12:56.
37. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Web and Social Media*. 2009;8:361–2.
38. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp*. 2008;2008:P10008.
39. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–2504.
40. Dijkstra EW. A note on two problems in connexion with graphs. *Numer Math*. 1959;1:269–71.
41. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* 2017;45:D183–D189.
42. Padua D, Zhang XH, Wang Q, Nadal C, Gerald WL, Gomis RR, et al. TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* 2008;133:66–77.
43. Bille C, Winther JF, Bautz A, Murray JC, Olsen J, Christensen K. Cancer risk in persons with oral cleft—a population-based study of 8093 cases. *Am J Epidemiol*. 2005;161:1047–55.
44. Andrade Filho PA, Letra A, Cramer A, Prasad JL, Garlet GP, Vieira AR, et al. Insights from studies with oral cleft genes suggest associations between WNT-pathway genes and risk of oral cancer. *J Dent Res* 2011;90:740–746.
45. Frebourg T, Oliveira C, Hochain P, Karam R, Manouvrier S, Graziadio C, et al. Cleft lip/palate and CDH1/E-cadherin mutations in families with hereditary diffuse gastric cancer. *J Med Genet* 2006;43:138–142.
46. Dunkhase E, Ludwig KU, Knapp M, Skibola CF, Figueiredo JC, Hosking FJ, et al. Nonsyndromic cleft lip with or without cleft palate and cancer: Evaluation of a possible common genetic background through the analysis of GWAS data. *Genom Data* 2016;10:22–29.
47. Grisanzio C, Freedman ML. Chromosome 8q24-associated cancers and MYC. *Genes Cancer*. 2010;1:555–9.
48. Uslu VV, Petretich M, Ruf S, Langenfeld K, Fonseca NA, Marioni JC, et al. Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nat Genet* 2014;46:753–758.
49. Lindsay S, Copp AJ. MRC–Wellcome Trust Human Developmental Biology Resource: enabling studies of human developmental gene expression. *Trends Genet*. 2005;21:586–90.