RESEARCH ARTICLE

In silico analysis of single nucleotide polymorphisms (SNPs) in human FOXC2 gene [version 1; peer review: 2 approved with reservations]

Mohammed Nimir ¹, Mohanad Abdelrahim ², Mohamed Abdelrahim ³, Mahil Abdalla¹, Wala eldin Ahmed¹, Muhanned Abdullah³, Muzamil Mahdi Abdel Hamid ⁴

¹Soba Center for Audit and Research, Soba University Hospital, University of Khartoum, Khartoum, 11111, Sudan

²Department of Human Anatomy, Ahfad University for Women, Khartoum, 11111, Sudan

³Department of Internal Medicine, Faculty of Medicine, University of Khartoum, Khartoum, 11111, Sudan

⁴Institute of Endemic Diseases, University of Khartoum, Khartoum, 11111, Sudan

v1 First published: 09 Mar 2017, 6:243
<https://doi.org/10.12688/f1000research.10937.1>
 Latest published: 09 Oct 2017, 6:243
<https://doi.org/10.12688/f1000research.10937.2>

Abstract

Introduction: Lymphedema is abnormal accumulation of interstitial fluid, due to inefficient uptake and reduced flow, leading to swelling and disability, mostly in the extremities. Hereditary lymphedema usually occurs as an autosomal dominant trait with allelic heterogeneity. **Methods:** We identified single nucleotide polymorphisms (SNPs) in the FOXC2 gene using dbSNP, analyzed their effect on the resulting protein using VEP and Biomart, modelled the resulting protein using Project HOPE, identified gene – gene interactions using GeneMANIA and predicted miRNAs affected and the resulting effects of SNPs in the 5' and 3' regions using PolymiRTS. **Results:** We identified 448 SNPs - 429 were nsSNPs and 44 SNPs were in the 5' and 3' UTRs. In total, 2 SNPs have deleterious effects on the resulting protein, and a 3D model confirmed those effects. The gene – gene interaction network showed the involvement of FOXC2 protein in the development of the lymphatic system. hsa-miR-6886-5p, hsa-miR-6886-5p, hsa-miR-6720-3p, which were affected by the SNPs rs201118690, rs6413505, rs201914560, respectively, were the most important miRNAs affected, due to their high conservation score. **Conclusions:** rs121909106 and rs121909107 were predicted to have the most harmful effects, while hsa-miR-6886-5p, hsa-miR-6886-5p and hsa-miR-6720-3p were predicted to be the most important miRNAs affected. Computational biology tools have advantages and disadvantages, and the results they provide are predictions that require confirmation.

Keywords

Primary lymphedema, FOXC2, SNP, In silico, Bioinformatics, miRNA

Open Peer Review

Approval Status

	1	2
version 2		
(revision)		
09 Oct 2017		
version 1		
09 Mar 2017		

1. **Pascal Brouillard** , Université catholique de Louvain, Brussels, Belgium

2. **Fengkai Zhang**, National Institutes of Health (NIH), Bethesda, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Mohammed Nimir (mohammednimir7@gmail.com)

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2017 Nimir M *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Nimir M, Abdelrahim M, Abdelrahim M *et al.* *In silico analysis of single nucleotide polymorphisms (SNPs) in human FOXC2 gene [version 1; peer review: 2 approved with reservations]* F1000Research 2017, 6:243
<https://doi.org/10.12688/f1000research.10937.1>

First published: 09 Mar 2017, 6:243 <https://doi.org/10.12688/f1000research.10937.1>

Introduction

Lymphedema is abnormal accumulation of interstitial fluid, due to inefficient uptake and reduced flow, leading to swelling and disability, which mostly affects the extremities. It can be divided into primary and secondary lymphedema according to the underlying cause. Primary or hereditary lymphedema results from genetic damage, while secondary or acquired lymphedema is caused by lymphatic system malfunction, resulting from trauma, including surgery, radiotherapy, tumors, or infections (for example, parasitic infections)¹.

Hereditary lymphedema usually occurs as an autosomal dominant trait with allelic heterogeneity. The most common type of primary hereditary lymphedema, Milroy disease, can develop due to mutations in the vascular endothelial growth-factor receptor-3 gene (VEGFR-3; FLT4)².

Forkhead box (Fox) proteins are a family of transcription factors (TFs) that play a key role in cell development, cell cycle regulation, and other important biological processes³. FOXC2, was first identified as a TF that plays a key role in the morphogenesis of the cardiovascular system⁴. Further studies revealed that FOXC2 was involved in lymphatic vascular development. In both humans and mice, FOXC2 is expressed in large amounts in the developing lymphatic vessels and in adult lymphatic valves⁵. FOXC2-deficient mice were demonstrated to have abnormal lymphatic patterning and failure to form lymphatic valves, which reveals the critical role of FOXC2 in lymphatic vascular development⁶. Truncating and missense mutations of FOXC2 have been discovered in patients with late-onset lymphedema (hereditary lymphedema II; OMIM 153200), often associated with distichiasis (double row of eyelashes), and sometimes ptosis (Lymphedema Distichiasis Syndrome [LDS]; OMIM 153400), and/or yellow nails (OMIM 153300)^{7–9}. LDS patients develop defects characterized by lymphatic and venous blood reflux, which means failure and/or absence of lymphatic and venous valves^{10,11}. On a molecular level, FOXC2 DNA binding sites are enriched in NFATC1 consensus sequences, and the two TFs work in tandem during lymphatic vascular morphogenesis¹². FOXC2 controls the expression of proteins that are essential for lymphatic valve development, such as connexins. Thus, adequate control of their activity is extremely important for proper lymphatic vessel development and function¹³.

The aetiology governing the phenotypic variability remains unclear¹⁴. Lymphatic malformations could also result from slightly mutated germline alleles, which are difficult to access (and so to identify), mutations in regulatory regions of the DNA, epigenetic changes, or a combination of the three. Epigenome sequencing and whole exome sequencing may be needed for in-depth study of these mutations. The part genetics play in the development of lymphatic anomalies is highly complex, shown by the discovery of 23 mutated human genes¹⁵.

Few studies addressing FOXC2 mutations from a bioinformatics point of view have been published. Out of those, none have specifically focused on single nucleotide polymorphisms (SNPs).

SNPs are located in vital regions of the gene, which may code for proteins located at the forkhead active domain of the protein or other sites, may severely affect the function of the TF. Furthermore, SNPs are feasible and cost effective to study using *in silico* analysis via available bioinformatics tools.

This study aimed to analyze all SNPs in the human FOXC2 gene and predict their effect on the structure, function, stability and regulation of its respective protein. The results of our study can be used in population studies to screen patients with hereditary lymphedema, and more importantly in phenotypic variations in lymphatic malformations of affected individuals.

Methods

Mining the database for SNPs

We selected the National Center for Biotechnology Information (NCBI) database, dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP>) for the retrieval of SNPs and their related protein sequence of FOXC2 gene. We used “FOXC2” as our search term and used filters to narrow down our search results into two categories: 3' + 5' UTR SNPs only and all SNPs in FOXC2, except for those in the 3' + 5' UTR. This gene was chosen because it's the one that is known to be associated with LDS, used for our computational analysis.

Evaluation of coding SNPs

We chose three complementary algorithms for functional impact prediction of nsSNPs: Sorting Intolerant From Tolerant (SIFT; <http://sift.bii.a-star.edu.sg/>), Polymorphism Phenotyping (PolyPhen; <http://genetics.bwh.harvard.edu/pph/>) and CONsensus DELeteriousness (Condel; <http://bg.upf.edu/fannsdbs/>)^{15–17}.

SIFT version 2.0 was used to distinguish between tolerant and intolerant coding mutations, and is used to predict whether an amino acid substitution in a protein will have a phenotypic effect. SIFT is based on the premise that protein evolution is correlated with protein function. Variants that occur at conserved alignment positions are expected to be tolerated less than those that occur at diverse positions. PolyPhen is a computational tool for identification of potentially functional nsSNPs. Predictions are based on a combination of phylogenetic, structural and sequence annotation information characterizing a substitution and its position in the protein.

We uploaded the SNP accession numbers of all the SNPs into VEP (<http://www.ensembl.org/Tools/VEP>)¹⁸ and enabled “SIFT”, “PolyPhen” and “Condel” to retrieve the corresponding predictions of the functional significance of each SNP from all of the three algorithms. We repeated this step using Biomart (<http://www.ensembl.org/biomart/martview/3beeb39b00d4f063ebfe8016f3a3b4ba>)¹⁹, which is similar to VEP and part of Ensembl. It provides a SIFT and PolyPhen prediction and score, like VEP, but it does not provide a Condel prediction. We took the SNPs predicted by both databases to be deleterious and pathogenic (2 SNPs; rs121909106 and rs121909107) to perform the next steps in the analysis.

Predicting the molecular phenotypic effects of deleterious SNPs

Project HOPE (<http://www.cmbi.ru.nl/hope/>) is an online web-server used to search protein 3D structures by collecting structural information from a series of sources. We input the two SNPs (rs121909106 and rs121909107) that we obtained from the last step together with the primary structure of the FOXC2 protein (obtained from the Protein Data Bank <http://www.rcsb.org/pdb/explore/explore.do?structureId=1D5V>) into HOPE²⁰. HOPE lets you choose the amino acid that was affected but the mutation and allows you to change that specific amino acid and then analyse the resulting change in the 3D (tertiary/quaternary) structure and outputs the change predicted, plus the explanation of such a change (in both structure and function of the protein).

Outlining gene – gene interactions

GeneMANIA ver. 3.1.2.8 (<http://genemania.org/>) finds other genes that are related to a set of input genes, using a very large set of functional association data²¹. Association data include protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity. We input FOXC2 as our query gene and the website generated a network of genes along with their gene-gene interactions, according to gene ontology terms.

Characterization of SNPs in 3' and 5' untranslated regions

PolymiRTS v3.0 (<http://compbio.uthsc.edu/miRSNP/>) is an integrated platform for analyzing the functional impact of genetic polymorphisms in miRNA seed regions and miRNA target sites²². We input a second set of SNPs (containing only 5' and 3' SNPs) and acquired a list of the miRNAs affected by these mutations. The affected miRNAs might lead to a decrease/increase of the expression of FOXC2.

The analysis steps performed are summarized in **Figure 1**.

Results

Predictions of deleterious and damaging coding nsSNPs

SNP information for FOXC2 was retrieved from dbSNP. For our investigations, we selected SNPs in coding and UTR (5' and 3') regions. Among the 448 SNPs, 429 were nsSNPs and 44 SNPs were in the 5' and 3' UTRs of FOXC2.

VEP and Biomart

We found two SNPs, rs121909106 and rs121909107, which correspond to S125L and R121H, to have significant deleterious effect on the structure and function of FOXC2 protein. We summarized the information of the two SNPs predicted by the two databases in **Table 1**.

Project HOPE

Figure 2 shows a 3D model of rs121909107 and rs121909106 and the spatial effects of each on the respective domains they are a part of, in addition to their effects on neighbouring domains.

GeneMANIA

Figure 2 shows the gene-gene interactions of FOXC2. The most important interactions are with FOXB1, FOXD1, FOXD2 and FOXD3 (**Figure 3**) (additional data in **Supplementary material**).

PolymiRTS

Table 3–Table 5 show the effect that SNPs in the 3' and 5' region of FOXC2 might have, along with the conservation score (CS) of the site in question. The higher the CS, the more profound the effect of the SNP is predicted to be.

Figure 1. Flow chart of the analysis process.

Table 1. Results of analysis by VEP and BioMart.

Accession number		rs121909106	rs121909107
Analysis by VEP	Location	16:86567709	16:86567697
	Allele	T	A
	Position	125	121
	Amino Acids	S/L	R/H
	Codons	tCg/tTg	cGc/cAc
	SIFT	Deleterious (0)	Deleterious (0)
	PolyPhen	Probably damaging (0.994)	Probably damaging (0.998)
	Condel	Deleterious (0.997)	Deleterious (0.999)
	Clinical Significance	Pathogenic	Pathogenic
Analysis by BioMart	Chromosome	16	16
	Position (bp)	86567709	86567697
	Alleles	C/T	G/A
	Clinical significance	Pathogenic	Pathogenic
	PolyPhen prediction	Probably damaging (0.994)	Probably damaging (0.998)
	SIFT prediction	Deleterious (0)	Deleterious (0)
	Condel	Deleterious (0.997)	Deleterious (0.999)

Figure 3. Gene-gene interaction network of FOXC2 with paths colored according to their functions.

Dataset 1. SNPs used for PolymiRTS analysis

<http://dx.doi.org/10.5256/f1000research.10937.d153064>

Dataset 2. SNPs used for other analyses

<http://dx.doi.org/10.5256/f1000research.10937.d153065>

Discussion

This study aimed to analyze the SNPs identified in the FOXC2 gene. We found a total of 472 SNPs, 2 of which were predicted to adversely affect the function of the resulting protein. The mutation leads to transcription of a histidine instead of an arginine at position 121, and the mutant residue is located near a highly conserved region²⁰. The mutation was reported previously in a patient suffering from lymphedema-distichiasis syndrome by Berry *et al.* (2005), and the effects that the R121H mutation would have on the DNA-binding part of the forkhead domain (FHD) of FOXC2 was predicted. It was predicted that R121H would impair FOXC2 protein's ability to bind DNA and act as a transcription activator, and this was confirmed by biochemical studies. Also, the mutation resulted in the mislocalization of FOXC2. Berry *et al.* (2005) determined that the mutation results in a non-functional protein and that this leads to hereditary LDS²³. The residue affected by rs121909107 is part of an inter pro-domain named Fork Head Domain Conserved Site 2 (Interpro, IPR030456), and it is annotated with the following Gene-Ontology (GO) terms to indicate its function: sequence-specific DNA binding (GO:0043565) and sequence-specific DNA binding transcription factor activity (GO:0003700).

The mutation of a serine into a leucine at position 125 means that each amino acid has its own specific size, charge, and hydrophobicity-value. The original wild-type residue and newly introduced mutant residue often differ in these properties. The mutation results in incorporating an amino acid with a different level of hydrophobicity, this will affect hydrogen bond formation, and the mutant residue is located near a highly-conserved region²⁰. This mutation matches a previously described variant in affected members of families with lymphedema-distichiasis syndrome, previously reported by Mangion *et al.*²⁴ and Bell *et al.*⁷. The mutation was not identified in 100 normal chromosomes. This mutation results in a wide range of phenotypes from minimal distichiasis to severe lymphoedema and congenital heart disease⁷. The residue affected by rs121909106 is part of an inter prodomain named Winged Helix-Turn-Helix Dna-Binding Domain (IPR011991) and it is annotated with the following GO terms to indicate its function: nucleic acid binding (GO:0003676) and nucleic acid binding transcription factor activity (GO:0001071).

Our results from GeneMANIA showed that FOXC2 interacts with a lot of genes (Figure 2), mainly functioning to control connective tissue, tube and epithelial tissue development. These results were proved important by Mangion *et al.*²⁴ and Bell *et al.*⁷, which showed that mutations in FOXC2 lead to developing LDS.

PolymiRTS results showed SNPs and INDELs in miRNA target sites: target sites disrupted by SNPs and INDELs in miRNA seeds and target sites created by SNPs and INDELs in miRNA seeds (Table 3–Table 5). Three miRNAs that are worth noting are hsa-miR-6886-5p (with a CS of 7), hsa-miR-6886-5p (with a CS

Table 2. SNPs and INDELs in miRNA target sites.

Location	dbSNP ID	Wobble base pair	Ancestral allele	Allele	miR ID	Conservation	miRSite	Function Class	Exp Support	context+ score change
86602454	rs200353178	N	C	C	hsa-miR-1247-3p hsa-miR-4449 hsa-miR-4532	2 3 3	cgtgTCCCGGGAc cgtgcCCGGGAC cgtgTCCCGGGAc	D D D	N N N	No Change No Change No Change
86602489	rs3751795	N	C	T	hsa-miR-4740-3p	2	cgtgTCTCGGGAc	C	N	No Change
				C	hsa-miR-4747-5p hsa-miR-5196-5p	2 2	gctcgCTTCCCA gctcgCTTCCCA	D D	N N	-0.054 -0.057
				T	hsa-miR-3153 hsa-miR-4668-5p hsa-miR-6733-5p hsa-miR-6739-5p	2 2 2 2	gctcgTTTCCCA gctcgTTTCCCA gctcgTTTCCCA gctcgTTTCCCA	C C C C	N N N N	-0.024 -0.022 -0.021 -0.027
86602513	rs142773766	N	C	C	hsa-miR-4288 hsa-miR-632	2 2	acCAGACAAttaa acCAGACAAttaa	D D	N N	-0.117 -0.117
				T	hsa-miR-1298-3p hsa-miR-3126-3p	3 3	aCCAGATAattaa aCCAGATAattaa	C C	N N	-0.114 -0.132
Location	SNP location in the mRNA transcript. It is a zero-based number.									
SNPID	Link to dbSNP.									
Wobble Pair	Whether the SNP can form a G:U wobble basepair with the miRNA. Y: Yes; N: No.									
Ancestral Allele	If applicable, the ancestral allele is denoted.									
Allele	Two alleles of the SNP in the mRNA transcript.									
miRID	Link to miRBase.									
Conservation	Occurrence of the miRNA site in other vertebrate genomes in addition to the query genome. By clicking the hyperlink, the users can examine the genomes in which this miRNA target site occurs.									
FuncClass	D: The derived allele disrupts a conserved miRNA site (ancestral allele with support >= 2). N: The derived allele disrupts a nonconserved miRNA site (ancestral allele with support < 2). C: The derived allele creates a new miRNA site. O: The ancestral allele can not be determined.									
miRSite	Sequence context of the miRNA site. Bases complementary to the seed region are in capital letters and SNPs are highlighted in red.									
ExpSupport	LT: The miRNA-mRNA interaction is supported by a low-throughput experiment (e.g., luciferase reporter assay or Western blot). HT: The miRNA-mRNA interaction is supported by a high-throughput experiment (e.g., microarray or pSILAC). LTL: The miRNA targeting the specific location is supported by a low-throughput experiment (e.g., allelic imbalance sequencing). HTL: The miRNA targeting the specific location is supported by a high-throughput experiment (e.g., HTS-CLIP).									
Context+ score change	Context+ scores predict the binding of a miRNA to the entire 3'-UTR by summing over contributions made by individual sites within the 3'-UTR that have perfect sequence complementarity to the miRNA seed region. The change the differences in context+ scores between the reference and derived alleles for each SNP or INDEL in putative miRNA target sites. A more negative value of the context+ score difference indicates an increased likelihood that the miRNA targeting is disrupted or newly created by the mutation in the target sites									

Table 3. Target sites disrupted by single nucleotide polymorphisms (SNPs) and INDELs in miRNA seeds.

Location	miR ID	dbSNP ID	miR seed	Allele	Wobble base pair	miRSite	Conservation	context+score change
86602466	hsa-miR-6889-5p	rs146254801	C[G/A]GGGAG	G/A	1	CUCCCCG	2	-0.182
86602496	hsa-miR-5090	rs3823658	C[G/A]GGGCA	G/A	1	GCCCCGA	2	-0.227
86602496	hsa-miR-6727-5p	rs202175375	U[C/T]GGGGC	C/T	0	GCCCCGA	2	-0.291
86602466	hsa-miR-6777-5p	rs56155608	[C/T]GGGGAG	C/T	0	CUCCCCG	2	-0.144

Table 4. Target sites created by single nucleotide polymorphisms (SNPs) and INDELs in miRNA seeds.

Location	miR ID	dbSNP ID	miR Seed	Allele	Wobble base pair	miRSite	Conservation	context+score change
86602524	hsa-miR-6886-5p	rs201118690	[C/T]CGCAGG	C/T	0	CUGCAGA	7	-0.074
86602524	hsa-miR-6886-5p	rs6413505	C[C/T]GCAGG	C/T	0	CUGCAGA	7	-0.074
86602530	hsa-miR-6720-3p	rs201914560	GCG[C/T]CUG	C/T	0	AGACGCA	6	-0.108
86602497	hsa-miR-4472	rs28655823	GU[G/C]GGGG	G/C	0	CCCCGAC	2	-0.244
86602497	hsa-miR-4472	rs202127912	GU[TG/-]GGGG	TG/-	0	CCCCGAC	2	-0.244

of 7) and hsa-miR-6720-3p (with a CS of 6), which are affected by the SNPs rs201118690, rs6413505, rs201914560, respectively. This points to possibility that the areas affected by those SNPs have an evolutionary important function.

Conclusions

In conclusion, there are many SNPs that affect FOXC2 gene; some are predicted to be harmful, such as rs121909106 and rs121909107, but most are not. miRNAs were affected by SNPs in the 3' and 5' untranslated regions of FOXC2 gene and three are noteworthy, hsa-miR-6886-5p, hsa-miR-6886-5p and hsa-miR-6720-3p, due their high conservation score.

Computational biology tools are very powerful, especially when provided with good data and used by experts. However, bioinformatics tools have their limitations; most importantly the fact that their results are but predictions, meaning that the information they provide us with requires confirmation by probing.

Data availability

Dataset 1: SNPs used for PolymiRTS analysis. doi, [10.5256/f1000research.10937.d153064](https://doi.org/10.5256/f1000research.10937.d153064)²⁵

Dataset 2: SNPs used for other analyses. doi, [10.5256/f1000research.10937.d153065](https://doi.org/10.5256/f1000research.10937.d153065)²⁶

Author contributions

MN conceived the study, contributed to data collection and analysis, and preparation of the manuscript. MA, MA, MA, WA and MA contributed to data collection. MH contributed to preparation of the manuscript and supervised the whole process. All authors were involved in the revision of the manuscript and have agreed to the final content.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgements

Many thanks to Prof. Muntasir Ibrahim, Sarmad Haydar and Mutaz Amin for their encouragement and help.

Supplementary material

Supplementary File 1: GeneMANIA network as an Excel file.

[Click here to access the data.](#)

Supplementary File 2: Print-out of the full report obtained from the GeneMANIA website for FOXC2 gene.

[Click here to access the data.](#)

References

1. Kurland LT, Molgaard CA: **The patient record in epidemiology.** *Sci Am.* 1981; **245**(4): 54–63.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Karkkainen MJ, Ferrell RE, Lawrence EC, *et al.*: **Missense mutations interfere with VEGFR-3 signalling in primary lymphoedema.** *Nat Genet.* 2000; **25**(2): 153–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Wijchers PJ, Burbach JP, Smidt MP: **In control of biology: of mice, men and Foxes.** *Biochem J.* 2006; **397**(2): 233–46.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Iida K, Koseki H, Kakinuma H, *et al.*: **Essential roles of the winged helix transcription factor MFH-1 in aortic arch patterning and skeletogenesis.** *Development.* 1997; **124**(22): 4627–38.
[PubMed Abstract](#)
5. Petrova TV, Karpanen T, Norrmen C, *et al.*: **Defective valves and abnormal mural cell recruitment underlie lymphatic vascular failure in lymphedema distichiasis.** *Nat Med.* 2004; **10**(9): 974–81.
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Kriederman BM, Myloyde TL, Witte MH, *et al.*: **FOXC2 haploinsufficient mice are a model for human autosomal dominant lymphedema-distichiasis syndrome.** *Hum Mol Genet.* 2003; **12**(10): 1179–85.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Bell R, Brice G, Child AH, *et al.*: **Analysis of lymphoedema-distichiasis families for FOXC2 mutations reveals small insertions and deletions throughout the gene.** *Hum Genet.* 2001; **108**(6): 546–51.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Fang J, Dagenais SL, Erickson RP, *et al.*: **Mutations in FOXC2 (MFH-1), a forkhead family transcription factor, are responsible for the hereditary lymphedema-distichiasis syndrome.** *Am J Hum Genet.* 2000; **67**(6): 1382–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Finegold DN, Kimak MA, Lawrence EC, *et al.*: **Truncating mutations in FOXC2 cause multiple lymphedema syndromes.** *Hum Mol Genet.* 2001; **10**(11): 1185–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
10. Brice G, Mansour S, Bell R, *et al.*: **Analysis of the phenotypic abnormalities in lymphoedema-distichiasis syndrome in 74 patients with FOXC2 mutations or linkage to 16q24.** *J Med Genet.* 2002; **39**(7): 478–83.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Mellor RH, Brice G, Stanton AW, *et al.*: **Mutations in FOXC2 are strongly associated with primary valve failure in veins of the lower limb.** *Circulation.* 2007; **115**(14): 1912–20.
[PubMed Abstract](#) | [Publisher Full Text](#)
12. Ivanov KI, Agalarov Y, Valmu L, *et al.*: **Phosphorylation regulates FOXC2-mediated transcription in lymphatic endothelial cells.** *Mol Cell Biol.* 2013; **33**(19): 3749–61.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Brouillard P, Boon L, Vikkula M: **Genetics of lymphatic anomalies.** *J Clin Invest.* 2014; **124**(3): 898–904.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Yu Z, Wang J, Peng S, *et al.*: **Identification of a novel VEGFR-3 missense mutation in a Chinese family with hereditary lymphedema type I.** *J Genet Genomics.* 2007; **34**(10): 861–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. González-Pérez A, López-Bigas N: **Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel.** *Am J Hum Genet.* 2011; **88**(4): 440–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
16. Ramensky V, Bork P, Sunyaev S: **Human non-synonymous SNPs: server and survey.** *Nucleic Acids Res.* 2002; **30**(17): 3894–900.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. Ng PC, Henikoff S: **Predicting deleterious amino acid substitutions.** *Genome Res.* 2001; **11**(5): 863–74.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. McLaren W, Pritchard B, Rios D, *et al.*: **Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor.** *Bioinformatics.* 2010; **26**(16): 2069–70.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Kinsella RJ, Kähäri A, Haider S, *et al.*: **Ensembl BioMart: a hub for data retrieval across taxonomic space.** *Database (Oxford).* 2011; **2011**: bar030.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Venselaar H, Te Beek TA, Kuipers RK, *et al.*: **Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces.** *BMC Bioinformatics.* 2010; **11**: 548.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Warde-Farley D, Donaldson SL, Comes O, *et al.*: **The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function.** *Nucleic Acids Res.* 2010; **38**(Web Server issue): W214–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Bhattacharya A, Ziebarth JD, Cui Y: **PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways.** *Nucleic Acids Res.* 2014; **42**(Database issue): D86–91.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Berry FB, Tamimi Y, Carle MV, *et al.*: **The establishment of a predictive mutational model of the forkhead domain through the analyses of FOXC2 missense mutations identified in patients with hereditary lymphedema with distichiasis.** *Hum Mol Genet.* 2005; **14**(18): 2619–27.
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Mangion J, Rahman N, Mansour S, *et al.*: **A gene for lymphedema-distichiasis maps to 16q24.3.** *Am J Hum Genet.* 1999; **65**(2): 427–32.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Nimir M, Abdelrahman M, Abdelrahman M, *et al.*: **Dataset 1 in: In silico analysis of single nucleotide polymorphisms (SNPs) in human FOXC2 gene.** *F1000Research.* 2017.
[Data Source](#)
26. Nimir M, Abdelrahman M, Abdelrahman M, *et al.*: **Dataset 2 in: In silico analysis of single nucleotide polymorphisms (SNPs) in human FOXC2 gene.** *F1000Research.* 2017.
[Data Source](#)

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 19 May 2017

<https://doi.org/10.5256/f1000research.11789.r22567>

© 2017 Zhang F. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fengkai Zhang

Laboratory of Systems Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), Bethesda, MD, USA

The manuscript presents nice work in analyzing the SNPs in the human FOXC2 gene with a set of publicly available bioinformatics tools and identified 2 nsSNPs and several other UTR region SNPs with potential significance in Lymphedema. The pipeline of tools has been well described and it could benefit research and analysis of other important genes associated with diseases. As the authors pointed out, the findings need further confirmation. I would suggest the authors to address the following aspects: (1) Is there any relationship or coordination among the identified SNPs? Do they belong to the same haplotype(s)? (2) What is the consideration or motivation to select the tools used for this study? For example, why did the authors choose GeneMANIA? All other tools in this study are used to analyze the SNPs in FOXC2 gene. GeneMANIA gives the result of the network involving FOXC2 with other genes. However, FOXC2 has been well introduced with multiple references at the beginning of this manuscript about its importance in Lymphedema. GeneMANIA does not show obvious help in SNP analysis focused in this study. (3) What is the criteria to identify the positive and negative SNPs for FOXC2, and why? How can false negative be avoided?

Besides the above major reservations, I would like to ask the authors to clarify/correct the following:

- As pointed by the first referee, what is the total number of SNPs? The abstract and result sections indicated 448 SNPs and the discussion indicated 472 SNPs. How can the two numbers be calculated from 429 nsSNPs and 44 UTR region SNPs ($429+44=473$)?
- In both abstract and discussion sections, the authors indicate the results require further confirmation. Can the authors describe some possible methods, like experimental, clinical or statistical?
- Page 3, spell out NFATC1 when it appears at the first time
- Page 3, rephrase

“SNPs are located in vital regions of the gene, which may code for proteins located at the forkhead active domain of the protein or other sites, may severely affect the function of the TF.”

- Page 6, figure 3, the color codes are hard to distinguish in the network visualization. Use alternative method, or give a more detailed explanation.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Systems biology, bioinformatics, genomics, immunology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 18 April 2017

<https://doi.org/10.5256/f1000research.11789.r21906>

© 2017 Brouillard P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pascal Brouillard

Laboratory of Human Molecular Genetics, de Duve Institute, Université catholique de Louvain, Brussels, Belgium

This paper describes an elegant bioinformatic analysis of variations in a known gene (FOXC2) for

primary lymphedema. It uses numerous freely available tools to identify not only variants that can possibly alter the function of the protein, but also affect miRNA (binding) sites. However, one weakness is that it only focuses on changes present in dbSNP, which lacks known amino acid-changing mutations such as e.g. those identified by Van Steenstel et al. or Sholto-Douglas-Vernon et al.^{1,2}. The results of the PolymiRTS should also be explained a bit more, especially on how to interpret the scores and what to consider as important. Indeed, contrary to the two known amino acids that have been previously characterized, these have not been proven. Focusing on the conservation score is probably not enough because some scores of 2 have a stronger "context+score change" effect than the scores of 7.

There are several mistakes that need to be corrected:

- The number of SNPs is unclear. The abstract and results mention a total of 448 and the discussion 472. Yet, neither of these corresponds to 429 nsSNPs + 44 in UTRs (= altogether 474)
- Page 3, right column, second line, SNPs may "code for amino acids", not proteins.
- Page 3, last paragraph, the Biomart link can stop after /martview/. The alphanumerical term after is a changing each time you open a session.
- Page 4, line 10: ... affected BY the mutation ..., not "but"
- Page 4, GeneMANIA is Figure 3, not 2, and PolymiRTS refers to table 2 to 4, not 3-5.
- The bolded title of Figure 2 cannot be solely rs121909107.
- Page 6, Discussion, 4th line "the mutation leads to TRANSDUCTION of a histidine ..., not transcription.
- Page 6, right column, the first sentence is strange. It is not the mutation into a leucine at position 125 that means that "each amino acid has its own specific size, charge and hydrophobicity-value". This is a general fact!
- Page 6, right col, line 5: AN amino acid, not as. Line 14: interpro domain. Line 15, DNA in capital letters. Second paragraph, GeneMania links to Figure 3, not 2. Second-last line, refer to Tables 2-4.

References

1. van Steensel MA, Damstra RJ, Heitink MV, Bladergroen RS, et al.: Novel missense mutations in the FOXC2 gene alter transcriptional activity. *Hum Mutat.* 2009; **30** (12): E1002-9 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Sholto-Douglas-Vernon C, Bell R, Brice G, Mansour S, et al.: Lymphoedema-distichiasis and FOXC2: unreported mutations, de novo mutation estimate, families without coding mutations. *Hum Genet.* 2005; **117** (2-3): 238-42 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research