HHS Public Access

Author manuscript

Biochim Biophys Acta Gene Regul Mech. Author manuscript; available in PMC 2024 March 01.

Published in final edited form as:

Biochim Biophys Acta Gene Regul Mech. 2023 March ; 1866(1): 194906. doi:10.1016/j.bbagrm.2023.194906.

Disease-associated Non-coding Variants Alter NKX2-5 DNA-Binding Affinity

Edwin G. Peña-Martínez[1], **Alejandro Rivera-Madera**[2], **Diego A. Pomales-Matos**[1], **Leandro Sanabria-Alberto**[1], **Brittany M. Rosario-Cañuelas**[1], **Jessica M. Rodríguez-Ríos**[1], **Emanuel A. Carrasquillo-Dones**[1], **José A. Rodríguez-Martínez**[1,*]

[1]University of Puerto Rico Río Piedras, San Juan, Puerto Rico

[2]University of Puerto Rico Cayey, Cayey, Puerto Rico

Abstract

Genome-wide association studies (GWAS) have mapped over 90% of disease- or trait-associated variants within the non-coding genome, like *cis*-regulatory elements (CREs). Non-coding single nucleotide polymorphisms (SNPs) are genomic variants that can change how DNA-binding regulatory proteins, like transcription factors (TFs), interact with the genome and regulate gene expression. NKX2-5 is a TF essential for proper heart development, and mutations affecting its function have been associated with congenital heart diseases (CHDs). However, establishing a causal mechanism between non-coding genomic variants and human disease remains challenging. To address this challenge, we identified 8,475 SNPs predicted to alter NKX2-5 DNA-binding using a position weight matrix (PWM)-based predictive model. Five variants were prioritized for in vitro validation; four of them are associated with traits and diseases that impact cardiovascular health. The impact of these variants on NKX2-5 binding was evaluated with electrophoretic mobility shift assay (EMSA) using purified recombinant NKX2-5 homeodomain. Binding curves were constructed to determine changes in binding between variant and reference alleles. Variants rs7350789, rs7719885, rs747334, and rs3892630 increased binding affinity, whereas rs61216514 decreased binding by NKX2-5 when compared to the reference genome. Our findings suggest that differential TF-DNA binding affinity can be key in establishing a causal mechanism of pathogenic variants.

[*]Corresponding author.

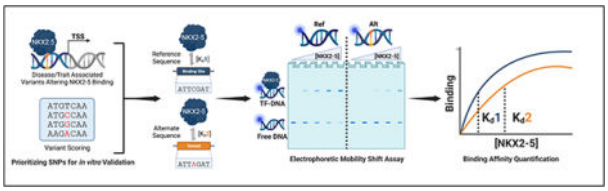
Author Contribution

Edwin G. Peña-Martínez: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Project administration, Writing - Original Draft. **Alejandro Rivera-Madera:** Validation, Formal Analysis, Investigation, Writing - Original Draft. **Diego A. Pomales-Matos:** Investigation, Writing - Review & Editing. **Leandro Sanabria-Alberto:** Investigation, Writing - Review & Editing. **Brittany M. Rosario-Cañuelas:** Investigation, Writing - Review & Editing. **Jessica M. Rodríguez-Ríos:** Validation, Writing - Review & Editing. **Emanuel A. Carrasquillo-Dones:** Investigation, Writing - Review & Editing. **José A. Rodríguez-Martínez:** Conceptualization, Methodology, Resources, Writing - Original Draft, Supervision, Project administration, Funding acquisition.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Graphical Abstract



Keywords

transcription factors; non-coding variants; binding affinity; gene regulation

2. Introduction

Genome-wide association studies (GWAS) have revealed that over 90% of disease/trait-associated variants occur within the non-coding genome.[1–5] Non-coding DNA comprises 98% of the human genome and includes regions like *cis*-regulatory elements (CREs), such as promoters and enhancers, that are essential for regulating gene expression.[6–11] Non-coding single nucleotide polymorphisms (SNPs) within CREs can alter the function of regulatory DNA-binding proteins, like transcription factors (TFs).[12–14] TFs bind to specific DNA sequences within CREs and recruit chromatin remodelers and the transcriptional machinery to regulate gene expression.[15–19] Non-coding SNPs can alter the binding affinity between a TF and its binding site, potentially leading to dysregulation of gene expression and crucial biological processes.[20,21]

NKX2-5 is a conserved homeodomain-containing TF from the NK-2 family that regulates heart development.[22–28] NKX2-5 binds to its cognate site (5'- `TTAAGTG` -3') to regulate cardiac genes needed for cardiomyocyte differentiation and cardiac morphogenesis, like the *atrial natriuretic factor (ANF)*.[29–33] Previous work has proven mutations within the NKX2-5 DNA-binding domain (DBD), a homeodomain, can alter its regulatory function, leading to cardiovascular diseases like congenital heart diseases (CHDs).[34–37] However, most variants associated with cardiovascular disease are non-coding and occur in CREs that can function as cardiac TF binding sites (TFBS).[38–42] Non-coding SNPs can affect TF-DNA binding affinity resulting in the loss of TFBS or the creation of a new TFBS in the genome[40,43,44]. Previous work has shown that coronary artery disease (CAD)-associated SNPs have disrupted the binding of tissue-specific TFs like STAT1, MEF2, and KLF2.[45–47]

In this work, we evaluated the potential of disease-associated variants to alter the binding affinity of the cardiac TF NKX2-5 homeodomain. Using SNP2TFBS[48], a position weight matrix (PWM)-based predictive model, we identified five variants associated with cardiovascular-related traits predicted to impact NKX2-5 binding. Through electrophoretic mobility shift assays (EMSA), we observed differential NKX2-5 binding activity between the reference and alternate genomic sequences. Binding curves were constructed to derive apparent dissociation constants (K_d^{app}) and determine changes in NKX2-5 binding. We found that three out of five predictions were consistent with in vitro observations, whereas the other two variants in vitro experiments contradicted the in silico prediction. Our results show

that non-coding cardiovascular trait-associated SNPs change NKX2-5 DNA binding affinity. This suggests that disruption or creation of TFBS throughout the genome can be a plausible causal mechanism behind cardiovascular diseases like CADs and CHDs.

3. Materials and Methods

3.1 Identification of Non-coding Variants

Non-coding SNPs predicted to alter NKX2-5-DNA binding affinity were identified using SNP2TFBS, a position weight matrix (PWM)-based predictive model.[48] SNP2TFBS uses variants from the 1000 Genomes Project[49] to create an alternate allele genome. It then uses PWM models from the JASPAR Core 2014 vertebrates[50] to perform whole genome scanning and extract overlapping SNPs that create, disrupt, or change TFBS scores. Using SNP2TFBS, we used an NKX2-5 PWM (MA0063.2) to score variants predicted to alter NKX2-5 binding that and intersected with SNPs in the GWAS catalog[3]. Filtered variants were scored using the SNP2TFBS, and the top five variants with the largest predicted change in binding were selected for *in vitro* validation.

3.2 NKX2-5 cloning and expression

The NKX2-5 homeodomain (HD) gene (Asp16 to Leu96) on the pDONR221 vector (DNASU plasmid repository, HsCD00039790) was cloned in pET-51(+) expression vector (Millipore Sigma) containing an N-terminal Strep•Tag II® and a C-terminal 10x His•Tag® through Gibson Cloning and used to transform BL21 DE3 *E. coli* strain (Millipore Sigma). Glycerol stock of transformed bacteria was cultivated in 50 mL Luria Broth (Sigma-Aldrich) for 16 h at 37°C. Following this period, 10 mL of initial bacterial culture was transferred to 500 mL of Terrific Broth (Sigma-Aldrich) in a 2 L Erlenmeyer flask and grown at 37°C, 130 rpm until the optical density at 600 nm (OD600) reached 0.5. Once the OD600 reached 0.5, the bacterial culture was induced with 1 mM IPTG for 20 h at 20°C and shaking at 130 rpm. The bacterial pellet was collected by centrifugation (2,800 × g, 5 min, 4 °C), the supernatant was discarded, and the pellet was stored at −80°C overnight.

3.3 Protein purification

NKX2-5 HD was purified from lysed bacterial cells through Ni-NTA affinity chromatography. Cell pellets were resuspended in 40 mL column buffer (500 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.2% Tween-20, 30 mM imidazole, and EDTA-free protease inhibitor). In addition, 4 mL 5 M NaCl was added to resuspended cells and sonicated in four 30-second cycles at 40% amplitude (QSONICA, Part No. Q125). The column was prepared with 2 mL Ni-NTA Agarose Resin (Qiagen) and equilibrated with 10 column volumes of column buffer. The lysate was centrifuged (2,800 × g, 30 min, 4°C), and the supernatant was loaded into the Ni-NTA affinity chromatography column. Column resin resuspended with the supernatant cell extract for 1 h at 4°C with orbital shaking. The supernatant was passed through the column three times. The column was washed with 20 mL of column buffer thrice, with increasing imidazole concentration (30, 50, and 100 mM) in each wash. The protein was eluted with 1.8 mL of elution buffer (500 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.2% Tween-20, 500 mM imidazole) six times.

NKX2-5 HD purity was evaluated with SDS-PAGE using 15 well Mini-PROTEAN TGX Precast Protein Gels® (Bio-Rad). Samples were prepared with 4X loading buffer containing β-mercaptoethanol (BME) for a total volume of 20 μL (5μL 4x loading buffer: 15 μL sample) and heated at 95°C for 5 min. 15 μL were loaded onto the gel for a 1.5 h run at 100 V at room temperature. The gel was stained using ProtoStain™ Blue Colloidal Coomassie G-250 stain (National Diagnostics).

For the Western Blot analysis, contents from the SDS-PAGE gel were transferred to a PVDF membrane using the Bio-Rad Turbo Transfer System protocol in a Trans-Blot® Turbo™ for 3 min at 25 V. Membranes were blocked using 5% milk in 1X TBST buffer for 1 h in orbital shaking and incubated overnight with 1:10,000 dilution of Anti-His mouse monoclonal antibody (Novus Biologicals, AD1.1.10). The SDS-PAGE and Western Blot Analysis results were imaged using Azure Sapphire Biomolecular Imager (Azure Biosystems).

3.4 Electrophoretic Mobility Shift Assay

NKX2-5 HD binding was evaluated using 40 bp sequences centered on the SNP and an additional 20 bp sequence for IRDye® 700 fluorophore conjugation(Integrated DNA Technologies). All sequences were ordered in IDT and are available in Supplementary Table 2. Reference and alternate oligonucleotides were labeled with the IR-700 fluorophore through a primer extension reaction. Binding reactions were performed in binding buffer (50 mM NaCl, 10 mM Tris-HCl (pH 8.0), and 10% glycerol) and 5 nM fluorescently labeled dsDNA. Binding reactions were incubated for 30 min at 30°C, then 30 min at room temperature before. The 6% polyacrylamide gel in 0.5x TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.4) was pre-runned at 85 V for 15 min, loaded at 30 V, and resolved at 75 V for 1.5 h at 4°C. Gels were imaged with Azure® Sapphire Bio-molecular Imager at 658 nm/710 nm excitation and emission.

3.5 Binding Affinity

Apparent K_d was determined by first quantifying the fluorescence signal in each DNA band using ImageJ.[51] Background intensities obtained from blank regions of the gel were subtracted from the band intensities. The fraction of bound DNA was determined using Equation 1. The fraction of bound DNA was plotted versus the concentration of the NKX2-5 homeodomain. Binding curves, K_d^{app}, and B_{max} were obtained by “one-site specific binding” non-linear regression using Prism software. Equation 1.
Binding Affinity from the integrated density of bound and unbound bands.

$$Fraction\,bound = \frac{bound}{(bound + unbound)}$$

4. Results and Discussion

4.1 Identification of Non-Coding Disease-Associated SNPs

Using SNP2TFBS, a PWM-based predictive model, we identified 8,475 SNPs predicted to change NKX2-5 binding out of the >84 million SNPs cataloged from the 1000 Genomes project (Supplementary File 1).[49] The genomic coordinates of the predicted variants were

intersected with disease or quantitative trait-associated SNPs from the GWAS catalog, resulting in 30 variants (Figure 1, Supplementary File 2). The output of the SNP2TFBS includes a ΔPWM score that predicts if the TF-DNA binding will increase or decrease based on a positive or negative score, respectively. ΔPWM scores were sorted by magnitude, and the five variants with the largest predicted impact on NKX2-5 DNA binding were chosen for *in vitro* validation, including two variants with a predicted increase in binding and three variants with decreased binding (Table 1). The selected SNPs are associated with traits and diseases that impact cardiovascular phenotypes (e.g., hemoglobin and cholesterol levels, red blood cell traits, and systolic blood pressure).

4.2 Expression and Purification of NKX2-5 Homeodomain

Recombinant NKX2-5 homeodomain with an N-terminal Strep•Tag® and a C-terminal 10x His•Tag® was cloned in an expression vector and produced using an IPTG-inducible bacterial system. After overexpression of NKX2-5 homeodomain, bacteria were lysed and centrifuged to obtain soluble fractions and purified using Ni-NTA affinity chromatography and eluted with an imidazole gradient. The NKX2-5 homeodomain was successfully purified as determined by SDS-PAGE and Western Blot (Figure 2A; Supplementary Figure 1). The DNA-binding activity of the purified NKX2-5 homeodomain was determined through EMSA using a known binding site within the ANF promoter (Figure 2B–C; Supplementary Figure 2).

4.3 Non-coding mutations alter NKX2-5 binding

We tested the five variants with the largest predicted change in binding using 40 bp genomic sequences centered at the SNP. Oligonucleotides were synthesized with an additional 20 bp constant region that served to add the IR700 fluorophore via primer extension. Changes in DNA binding affinity between the reference and the alternate allele were determined through EMSA. Purified NKX2-5 homeodomain was equilibrated with reference and variant sequences at seven different protein concentrations. Differential TF-DNA binding was observed for the five predicted variants (Supplementary Figure 3). The impact of the five variants on binding affinity was quantified by generating binding curves and calculating the K_d^{app} values (Figure 3). Three out of the five tested variants agreed with our computational prediction (Supplementary Table 1). Variants rs7350789 and rs3892630 were precited by SNP2TFBS to increase binding affinity and were successfully validated by EMSA. Changes in K_d^{app} for rs3892630 resulted in a 2.3-fold decrease, while rs7350789 could not be quantified due to low binding affinity in the reference sequence. SNP2TFBS predicted variant rs61216514 to decrease in binding affinity and was successfully validated by EMSA, resulting in a 1.3-fold change increase of its K_d^{app} However, the two variants that did not follow the prediction (rs7719885 and rs747334) still demonstrated differential DNA binding. SNP2TFBS predicted variants rs7719885 and rs747334 would decrease binding affinity. Both increased, resulting in a K_d^{app} 1.3- and 1.8-fold change decrease, respectively. The predictions were made using a PWM model generated from chromatin immunoprecipitation sequencing (ChIP-seq) data which has cellular factors (e.g., other TFs or co-factors) not present in our in vitro biochemical assay like an EMSA. Additional DNA-binding specificity models generated by *in vitro* experiments or alternate computational approaches will be

tested in the future. However, all five variants significantly altered NKX2-5 DNA binding affinities as determined by changes in their K_d^{app} values.

5. Conclusion

We conclude that non-coding GWAS variants can alter NKX2-5 affinity for its genomic binding sites. Non-coding SNPs associated with cardiovascular traits altered the NKX2-5-DNA complex formation, which is essential for proper heart development. Changes in the biophysical properties of gene regulation, like TF-DNA binding, are key factors to consider when determining the causal mechanism of genetic variants behind human diseases. Pathogenic SNPs within the NKX2-5 binding sites are potential regulatory targets for healthy cardiovascular development and function. Future work should address some of the limitations of in vitro biochemical assays, which do not consider cellular factors such as chromatin accessibility, TF cellular localization, and TF binding partners that are present in vivo.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Esther Peterson for helping with the statistical analysis. This project was supported by NIH-SC1GM127231. EGPM and JMRR were funded by the NSF BioXFEL Fellowship (STC-1231306). EGPM and DAPM were funded by the NIH RISE Fellowship (5R25GM061151-20). ARM was funded by NSF REU: PR-CLIMB Program (2050493). LSA was funded by NIH ID-GENE Fellowship (1R25HG012702). JMRR was funded NSF Graduate Research Fellowship (1744619). BMRC was funded by the ACS SEED Program and the UPRRP Department of Chemistry. Graphical abstract and Figure 1 were created in Biorender®.

References

1. Oldoni F et al. Post-GWAS methodologies for localisation of functional non-coding variants: ANGPTL3. Atherosclerosis 246, 193–201 (2016). [PubMed: 26800306]
2. Alsheikh AJ et al. The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases. BMC Med Genomics 15, 74 (2022). [PubMed: 35365203]
3. Buniello A et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 47, D1005–D1012 (2019). [PubMed: 30445434]
4. Lee PH et al. Principles and methods of in-silico prioritization of non-coding regulatory variants. Human Genetics 137, 15–30 (2018). [PubMed: 29288389]
5. Zhang F & Lupski JR Non-coding genetic variants in human disease. Hum Mol Genet 24, R102–R110 (2015). [PubMed: 26152199]
6. Saenko VA & Rogounovitch TI Genetic polymorphism predisposing to differentiated thyroid cancer: A review of major findings of the genome-wide association studies. Endocrinology and Metabolism 33, 164–174 (2018). [PubMed: 29947173]
7. Taft RJ, Pheasant M & Mattick JS The relationship between non-protein-coding DNA and eukaryotic complexity. BioEssays 29, 288–299 (2007). [PubMed: 17295292]
8. Elkon R & Agami R Characterization of noncoding regulatory DNA in the human genome. Nat Biotechnol 35, 732–746 (2017). [PubMed: 28787426]
9. Cremer M & Cremer T Nuclear compartmentalization, dynamics, and function of regulatory DNA sequences. Genes Chromosomes Cancer 58, 427–436 (2019). [PubMed: 30520215]

10. Meddens CA, van der List ACJ, Nieuwenhuis EES & Mokry M Non-coding DNA in IBD: From sequence variation in DNA regulatory elements to novel therapeutic potential. Gut 68, 928–941 (2019). [PubMed: 30692146]
11. Shabalina SA & Spiridonov NA The mammalian transcriptome and the function of non-coding DNA sequences. 4, 105 (2004).
12. Shrestha S et al. Discovering human transcription factor physical interactions with genetic variants, novel DNA motifs, and repetitive elements using enhanced yeast one-hybrid assays. Genome Res 29, 1533–1544 (2019). [PubMed: 31481462]
13. Holley G, Wittler R, Stoye J & Hach F Quantifying the Impact of Non-coding Variantson Transcription Factor-DNA Binding. Springer International 2, 50–65 (2017).
14. Weirauch MT et al. Determination and inference of eukaryotic transcription factor sequence specificity. Cell 158, 1431–1443 (2014). [PubMed: 25215497]
15. Rodríguez-Martínez JA, Reinke AW, Bhimsaria D, Keating AE & Ansari AZ Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. Elife 6, 1–29 (2017).
16. Stormo GD Modeling the specificity of protein-DNA interactions. Quantitative Biology 1, 115–130 (2013). [PubMed: 25045190]
17. Inukai S, Kock KH & Bulyk ML Transcription factor–DNA binding: beyond binding site motifs. Curr Opin Genet Dev 43, 110–119 (2017). [PubMed: 28359978]
18. Lambert SA et al. The Human Transcription Factors. Cell 172, 650–665 (2018). [PubMed: 29425488]
19. Song W et al. Tumor-Derived Ligands Trigger Tumor Growth and Host Wasting via Differential MEK Activation. Dev Cell 48, 277–286.e6 (2019). [PubMed: 30639055]
20. Jiang X et al. Variants in a cis-regulatory element of TBX1 in conotruncal heart defect patients impair GATA6-mediated transactivation. Orphanet J Rare Dis 16, (2021).
21. Lu X et al. Global discovery of lupus genetic risk variant allelic enhancer activity. Nat Commun 12, (2021).
22. Elliott DA, Kirk EP, Schaft D & Harvey RP NK-2 Class Homeodomain Proteins. Conserved Regulators of Cardiogenesis in Heart Development and Regeneration 569–597 (Elsevier Inc., 2010).
23. Pradhan L et al. Crystal structure of the human NKX2.5 Homeodomain in complex with DNA target. Biochemistry 51, 6312–6319 (2012). [PubMed: 22849347]
24. Davidson EH & Erwin DH Gene regulatory networks and the evolution of animal body plans. Science 311, 796–797 (2006). [PubMed: 16469913]
25. Tao Y & Schulz RA Heart development in Drosophila. Seminars in Cell and Developmental Biology 18, 3–15 (2007). [PubMed: 17208472]
26. McCulley DJ & Black BL Transcription Factor Pathways and Congenital Heart Disease. in Current Topics in Developmental Biology 100, 253–277 (2012). [PubMed: 22449847]
27. Medicine, M. & Divisions, C. Csx: A murine homeobox-containing gene specifically expressed in the developing heart (cardiac development/transcription factor/tissue-specific gene expression/ embryonic stem cell/evolutionary conservation) ISSEI KOMURO AND SEIGO IZUMO*. Proc. Natl. Acad. Sci. USA vol. 90 (1993).
28. Lyons I et al. Myogenic and morpho.genetic defects in the heart tubes of murlne embryos lacking the homeo box gene Nkx2-5. Genes Dev 9, 1654–1666 (1995). [PubMed: 7628699]
29. Owens GK, Kumar MS & Wamhoff BR Molecular regulation of vascular smooth muscle cell differentiation in development and disease. Physiol Rev 84, 767–801 (2004). [PubMed: 15269336]
30. Bruneau BG Signaling and transcriptional networks in heart development and regeneration. Cold Spring Harb Perspect Biol 5, (2013).
31. Olson EN Gene regulatory networks in the evolution and development of the heart. Science (1979) 313, 1922–1927 (2006).
32. Carlson CD et al. Specificity landscapes of DNA binding molecules elucidate biological function. Proc Natl Acad Sci U S A 107, 4544–4549 (2010). [PubMed: 20176964]

33. Small EM & Krieg PA Transgenic analysis of the atrialnatriuretic factor (ANF) promoter: Nkx2-5 and GATA-4 binding sites are required for atrial specific expression of ANF. Dev Biol 261, 116–131 (2003). [PubMed: 12941624]
34. Chung IM & Rajakumar G Genetics of congenital heart defects: The NKX2-5 gene, a key player. Genes (Basel) 7, (2016).
35. Bouveret R et al. NKX2-5 mutations causative for congenital heart disease retain functionality and are directed to hundreds of targets. Elife 4, 1–30 (2015).
36. Kasahara H & Benson DW Biochemical analyses of eight NKX2.5 homeodomain missense mutations causing atrioventricular block and cardiac anomalies. Cardiovasc Res 64, 40–51 (2004). [PubMed: 15364612]
37. Bruneau BG The developmental genetics of congenital heart disease. Nature 451, 943–948 (2008). [PubMed: 18288184]
38. Yang B et al. Protein-altering and regulatory genetic variants near GATA4 implicated in bicuspid aortic valve. Nat Commun 8, (2017).
39. Smemo S et al. Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. Hum Mol Genet 21, 3255–3263 (2012). [PubMed: 22543974]
40. Villar D, Frost S, Deloukas P & Tinker A The contribution of non-coding regulatory elements to cardiovascular disease. Open Biol 10, 200088 (2020). [PubMed: 32603637]
41. Benaglio P et al. Allele-specific NKX2-5 binding underlies multiple genetic associations with human electrocardiographic traits. Nat Genet 51, 1506–1517 (2019). [PubMed: 31570892]
42. Richter F et al. Genomic analyses implicate noncoding de novo variants in congenital heart disease. Nat Genet 52, 769–777 (2020). [PubMed: 32601476]
43. Khurana E et al. Role of non-coding sequence variants in cancer. Nat Rev Genet 17, 93–108 (2016). [PubMed: 26781813]
44. Deplancke B, Alpern D & Gardeux V The Genetics of Transcription Factor DNA Binding Variation. Cell 166, 538–554 (2016). [PubMed: 27471964]
45. Harismendy O et al. 9p21 DNA variants associated with coronary artery disease impair interferon-γ 3 signalling response. Nature 470, 264–270 (2011). [PubMed: 21307941]
46. Krause MD et al. Genetic variant at coronary artery disease and ischemic stroke locus 1p32.2 regulates endothelial responses to hemodynamics. Proc Natl Acad Sci U S A 115, E11349–E11358 (2018). [PubMed: 30429326]
47. Beaudoin M et al. Myocardial Infarction-Associated SNP at 6p24 Interferes with MEF2 Binding and Associates with PHACTR1 Expression Levels in Human Coronary Arteries. Arterioscler Thromb Vasc Biol 35, 1472–1479 (2015). [PubMed: 25838425]
48. Kumar S, Ambrosini G & Bucher P SNP2TFBS-a database of regulatory SNPs affecting predicted transcription factor binding site affinity. Nucleic Acids Res 45, D139–D144 (2017). [PubMed: 27899579]
49. Devuyst O The 1000 genomes project: Welcome to a new world. Peritoneal Dialysis International 35, 676–677 (2015). [PubMed: 26703842]
50. Fornes O et al. JASPAR 2020: Update of the open-Access database of transcription factor binding profiles. Nucleic Acids Res 48, D87–D92 (2020). [PubMed: 31701148]
51. Schneider CA, Rasband WS & Eliceiri KW NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9, 671–675 (2012). [PubMed: 22930834]

Highlights:

- Identified 30 disease-associated non-coding SNPs impacting NKX2-5 DNA-binding
- Human NKX2-5 homeodomain was successfully cloned, expressed, and purified
- In silico NKX2-5 binding predictions of GWAS SNPs were evaluated through EMSA
- Four cardiovascular GWAS SNPs showed differential NKX2-5 binding affinity

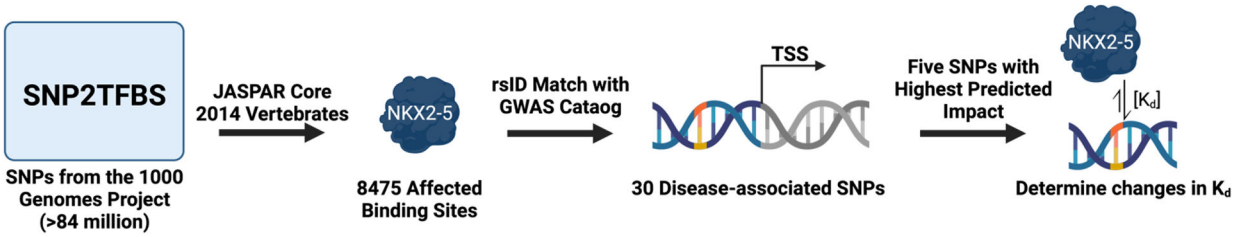


Figure 1:

Identification of disease/trait-associated non-coding SNPs affecting NKX2-5 binding. Predicted SNPs from the 1000 Genomes Project were intersected with disease-associated variants from the GWAS catalog.

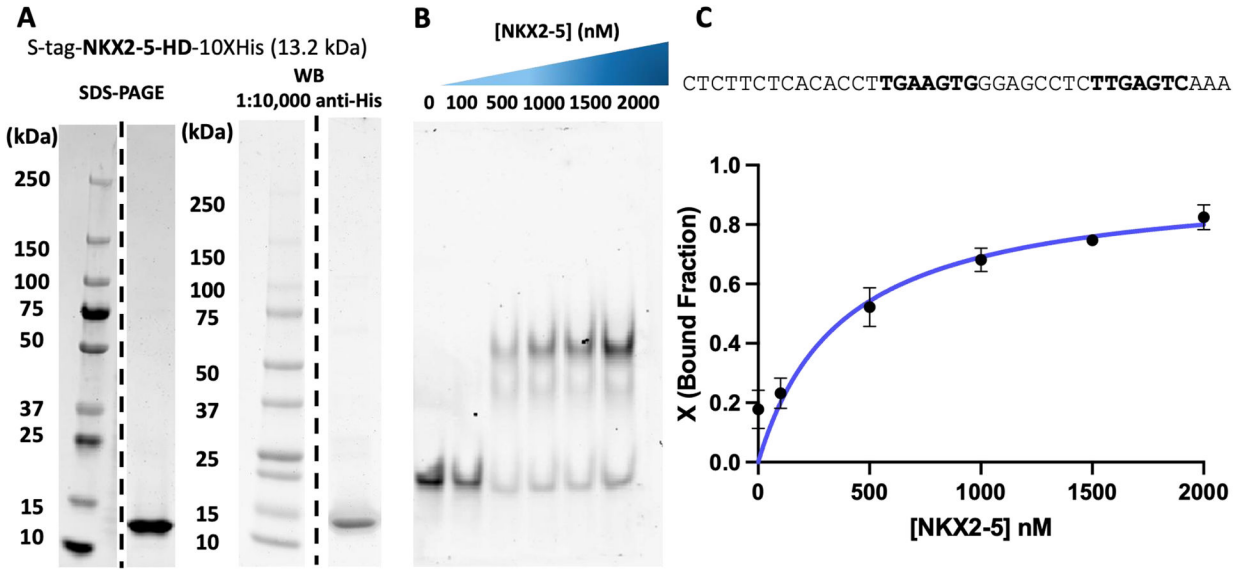


Figure 2:

Expression and purification of functional NKX2-5 homeodomain. A) SDS-PAGE (left) and Western blot (right) of purified NKX2-5 homeodomain after Ni-NTA affinity chromatography. B) Electrophoretic mobility shift assay (EMSA) of known binding site within the ANF promoter. C) Binding curve analysis of NKX2-5 homeodomain. Data points represent the average value of triplicate measurements and error bars the standard error.

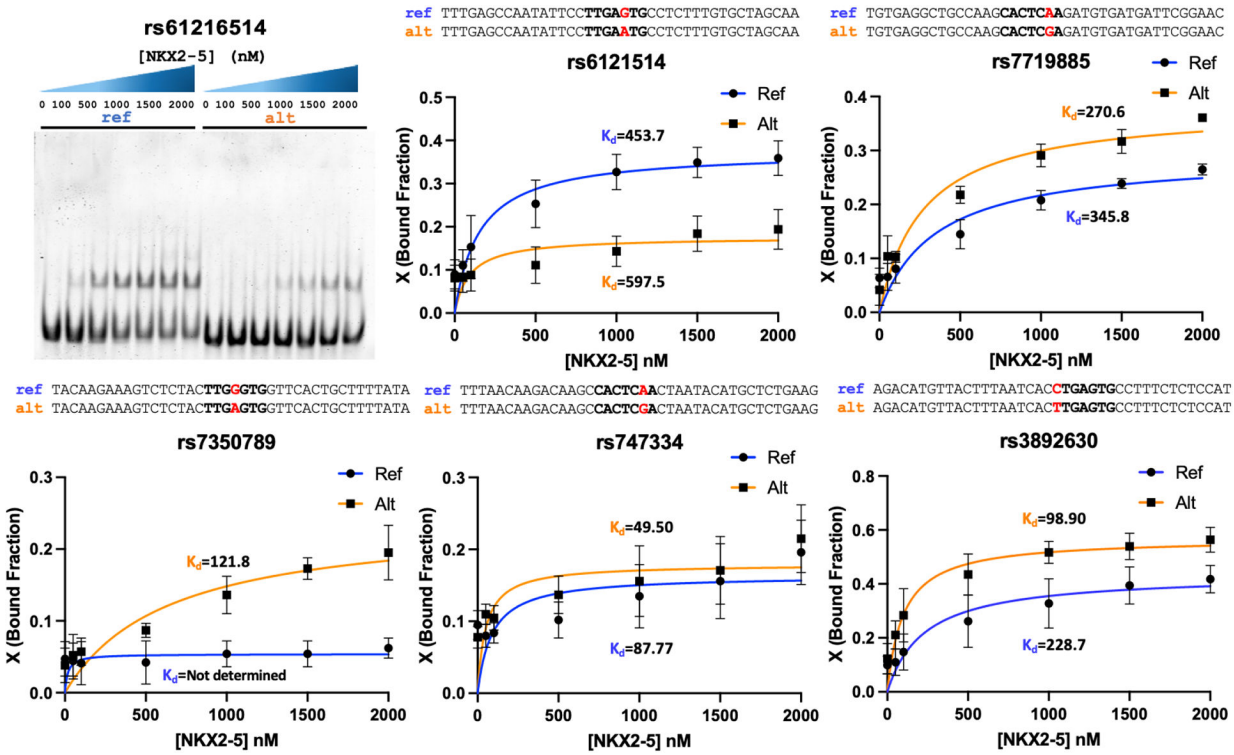


Figure 3:

NKX2-5 binding to reference (ref) and variant (alt) sequences determined through EMSA. (top left) Representative EMSA gel used for binding curve analysis for rs6121514. Binding experiments were performed in triplicates. Binding curves show average X (bound fraction) and error bars are standard error. Parameters of the non-linear regressions are reported in Supplementary Table 1.

Table 1:

Non-coding SNPs prioritized for in vitro validation.

rsID	Mutation	ΔPWM Score	Predicted Binding Impact	Associated Disease or Trait
rs7350789	G → A	258	Increase	Serum metabolite levels, High density lipoprotein cholesterol levels, Postprandial triglyceride response, Total cholesterol levels, Phosphatidylethanolamine levels
rs61216514	G → A	−232	Decrease	Mean corpuscular hemoglobin
rs7719885	A → G	−212	Decrease	Systolic blood pressure
rs747334	A → G	−187	Decrease	Fibroblast growth factor basic levels
rs3892630	C → T	146	Increase	Red blood cell traits