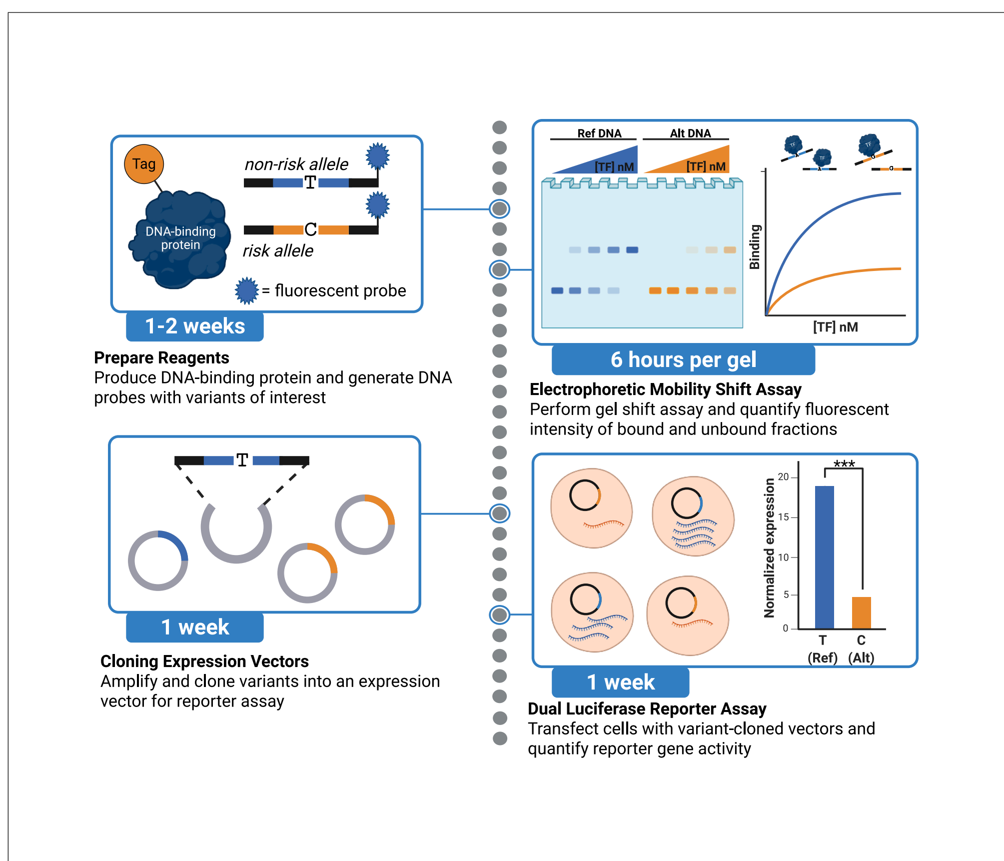
STAR Protocols

CellPress OPEN ACCESS

Protocol

Protocol for evaluating the impact of non-coding variants on transcription factor binding and gene expression



Non-coding variants can alter transcription factor (TF)-DNA binding and dysregulate gene expression. Here, we present a protocol for evaluating the impact of non-coding variants on TF binding and gene expression. We describe steps for expressing and purifying DNA-binding proteins, preparing DNA probes and reporter constructs, and measuring changes in TF-DNA binding affinity. Finally, we provide instructions for quantification and statistical analysis.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Edwin G. Peña-Martínez, Jessica M. Rodríguez-Ríos, Jean L. Messon-Bird, ..., Alejandro Rivera-Madera, Esther A. Peterson-Peguero, José A. Rodríguez-Martínez

edwin.pena1@upr.edu (E.G.P.-M.)
jessica.rodriguez8@upr.edu (J.M.R.-R.)
jose.rodriguez233@upr.edu (J.A.R.-M.)

Highlights

Guidance on variant selection and recombinant protein expression

Steps for *in vitro* evaluation of protein-DNA binding and gene expression assays

Instructions for result quantification and statistical analysis

Peña-Martínez et al., STAR Protocols 6, 103874
June 20, 2025 © 2025 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.xpro.2025.103874

STAR Protocols

CellPress OPEN ACCESS

Protocol

Protocol for evaluating the impact of non-coding variants on transcription factor binding and gene expression

Edwin G. Peña-Martínez,[1,3,4,*] Jessica M. Rodríguez-Ríos,[1,3,4,*] Jean L. Messon-Bird,[1] Adriana C. Barreiro-Rosario,[1] Rosalba Velázquez-Roig,[1] Alejandro Rivera-Madera,[2] Esther A. Peterson-Peguero,[1] and José A. Rodríguez-Martínez[1,5,*]

[1]Department of Biology, University of Puerto Rico-Río Piedras, San Juan, PR 00931, USA

[2]Department of Biology, University of Puerto Rico-Cayey, Cayey, PR 00736, USA

[3]These authors contributed equally

[4]Technical contact

[5]Lead contact

*Correspondence: edwin.pena1@upr.edu (E.G.P.-M.), jessica.rodriguez8@upr.edu (J.M.R.-R.), jose.rodriguez233@upr.edu (J.A.R.-M.)
https://doi.org/10.1016/j.xpro.2025.103874

SUMMARY

Non-coding variants can alter transcription factor (TF)-DNA binding and dysregulate gene expression. Here, we present a protocol for evaluating the impact of non-coding variants on TF binding and gene expression. We describe steps for expressing and purifying DNA-binding proteins, preparing DNA probes and reporter constructs, and measuring changes in TF-DNA binding affinity. Finally, we provide instructions for quantification and statistical analysis.
For complete details on the use and execution of this protocol, please refer to Peña-Martínez and Messon-Bird et al.[1]

BEFORE YOU BEGIN

Non-coding genetic variants encompass over 90% of all disease-associated mutations identified to date, which can alter the regulatory functions of transcription factors (TFs).[2–4] In vitro approaches, such as electrophoretic mobility shift assays (EMSA) and luciferase gene reporter assays, have been proven effective at quantifying the impact of genetic variants on TF-DNA binding and gene expression, respectively.[5–7]

Here, we describe a stepwise approach to quantify the impact of cardiovascular disease (CVD)-associated single nucleotide polymorphisms (SNPs) on GATA4 DNA-binding and regulatory function. However, this protocol can be applied to various genetic variants and regulatory proteins of interest (see limitations). Two critical reagents are required to initiate this protocol: synthesized oligonucleotides containing genetic variants and purified DNA-binding protein of interest. We recommend the following work as guidelines to identify and prioritize biologically relevant genetic variants to assay.[5,8–10]

Purify DNA-binding proteins by affinity chromatography

Timing: 1–2 weeks

Here, we provide an overview of the expression and affinity chromatography purification of recombinant GATA4 DNA-binding domain with a hexahistidine tag.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

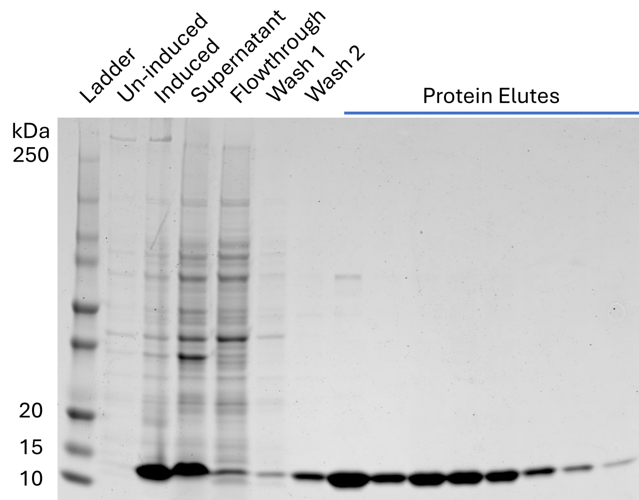


Figure 1. Purification of recombinant GATA DNA-binding domain (DBD) with hexahistidine tag

1. Perform protein expression in competent BL21 DE3 *E. coli* that are Isopropyl β- D-1-thiogalactopyranoside (IPTG) inducible.
   a. Grow BL21 DE3 *E. coli* transformed with the expression vector of interest in 500 mL of Terrific Broth with corresponding antibiotics in a 2,000 mL flask at 37°C 180–220 rpm until optical density (O.D.) is between 0.6–0.8.
   b. Add IPTG to the cell culture at a final concentration of 1 mM and leave shaking at 18°C 180–220 rpm for 18–20 h.
   c. Collect pellets by centrifuging cell culture at 15,000 × *g* at 4°C for 5 min.

   Note: BL21 DE3 *E. coli* cell lines use the T7 RNA polymerase system that is IPTG-inducible. To implement this expression system, the protein-coding gene of interest should be cloned downstream of a T7 promoter in a pET expression vector.

   Note: Protein induction and expression may need to be optimized for different proteins (*see* troubleshooting 1).

2. Perform protein purification using Ni-NTA affinity chromatography with imidazole elution. See buffer recipes at the end of this step.
   a. Resuspend the bacterial pellet in 40 mL of Column Buffer and add 4 mL of 5 M NaCl.
   b. Sonicate resuspended pellets three times at 40% amplitude in 30 sec intervals.
   c. Centrifuge sample at 15,000 × *g* for 30 min at 4°C.
   d. While centrifugation is in process, prepare a column with 2 mL of Ni-NTA resin and equilibrate with 20 mL of Column buffer.
   e. After the sample is centrifuged, incubate the supernatant with the equilibrated resin for 1 h at 4°C with orbital shaking.
   f. After incubation, add resin and supernatant mix back to the column and drain into a collection falcon tube. Collected supernatant will be referred to as flowthrough.
   g. Add flowthrough and re-collect in a falcon tube two additional times.
   h. Perform a 20 mL wash using Column Buffer.
   i. Perform a 20 mL wash using Wash Buffer 1.
   j. Perform a 20 mL wash using Wash Buffer 2.
   k. Perform a 1.8 mL elution with Elution Buffer six times and collect in separate 2 mL tubes.
   l. Evaluate purification through SDS-PAGE (Figure 1).

Note: We recommend collecting 1 mL fractions at each step after 2f to evaluate in the SDS-PAGE.

Note: After purification, columns can be re-used up to five times without reconstituting the resin. For this, wash the resin with 1 M NaOH and store with 20% ethanol at 4°C.

⚠ CRITICAL: Sonication step should be done while sample is on ice to avoid heating sample.

Column buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.2% Tween-20, 30 mM imidazole, and EDTA-free protease inhibitor)

Reagent	Final concentration	Amount
1 M Tris-HCl pH 8.0	20 mM	20 mL
5 M NaCl	500 mM	100 mL
Tween-20	0.2%	2 mL
1 M Imidazole	30 mM	30 mL
EDTA-free Protease Inhibitor	N/A	1 tablet
H_2O	N/A	Up to 1 L
Total	N/A	1 L

Store at 4 °C for up to 6 months

Wash buffer 1 (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.2% Tween-20, 50 mM imidazole, and EDTA-free protease inhibitor)

Reagent	Final concentration	Amount
1 M Tris-HCl pH 8.0	20 mM	20 mL
5 M NaCl	500 mM	100 mL
Tween-20	0.2%	2 mL
1 M Imidazole	50 mM	50 mL
EDTA-free Protease Inhibitor	N/A	1 tablet
H_2O	N/A	Up to 1 L
Total	N/A	1 L

Store at 4 °C for up to 6 months

Wash buffer 2 (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.2% Tween-20, 100 mM imidazole, and EDTA-free protease inhibitor)

Reagent	Final concentration	Amount
1 M Tris-HCl pH 8.0	20 mM	20 mL
5 M NaCl	500 mM	100 mL
Tween-20	0.2%	2 mL
1 M Imidazole	100 mM	100 mL
EDTA-free Protease Inhibitor	N/A	1 tablet
H_20	N/A	Up to 1 L
Total	N/A	1 L

Store at 4 °C for up to 6 months

Elution buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.2% Tween-20, 500 mM imidazole, and EDTA-free protease inhibitor)

Reagent	Final concentration	Amount
1 M Tris-HCl pH 8.0	20 mM	20 mL
5 M NaCl	500 mM	100 mL
Tween-20	0.2%	2 mL
1 M Imidazole	500 mM	500 mL
EDTA-free Protease Inhibitor	N/A	1 tablet
H_2O	N/A	Up to 1 L
Total	N/A	1 L

Store at 4 °C for up to 6 months

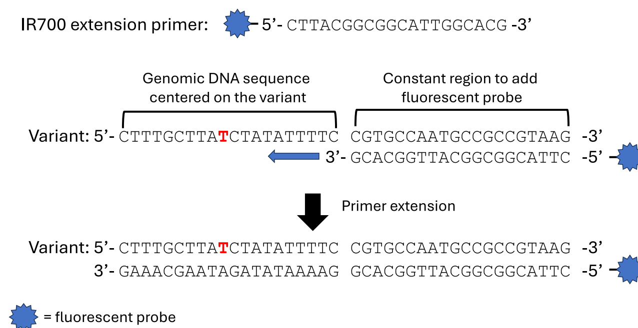


Figure 2. Generating fluorescently labeled double-stranded DNA through primer extension reaction

Note: The expression and purification should be optimized for each protein.

3. After protein purification, perform buffer exchange to the binding buffer needed for binding assays and adjust protein concentration to 10 μM.
4. Store as 50 μL aliquots at −80°C. See the recipe for the binding buffer of this protocol below:

Binding buffer 5× (100 mM NaCl, 10 mM HEPES pH 8.0, 0.5 μM zinc acetate, 200 mM NH4 acetate, 20% glycerol)		
Reagent	Final concentration	Amount
1 M HEPES pH 8.0	10 mM	1 mL
5 M NaCl	100 mM	0.2 mL
1 mM Zinc acetate	0.5 μM	50 μL
1 M NH_4	200 mM	20 mL
Glycerol	20 %	20 mL
H_2O	N/A	Up to 100 mL
Total	N/A	100 mL
Store at 4 °C for up to 6 months		

Note: TF stored in binding buffer 5× should be diluted to 1× in the binding reaction.

Note: We prepare 50 μL aliquots for triplicate experiments for reference and alternate allele (∼24 μL each) to avoid constant freezing and thawing of proteins.

Alternative: Binding experiments can also be done using cell/nuclear lysates overexpressing the TF of interest, but they will require optimization.

Preparation of fluorescent probe for binding assay

Timing: 2–3 h

Here, we describe how to generate fluorescently labeled DNA probes to visualize protein-DNA interactions. These sequences will be used for the electrophoretic mobility shift assay (EMSA).

5. Prepare the oligonucleotides by resuspending lyophilized DNA at the desired volume or concentration, typically 100 μM.

6. Once the DNA is in solution, prepare the following primer extension reaction mix for the fluorophore incorporation into double-stranded DNA (Figure 2).

Primer extension reaction mix	
Reagent	**Amount**
dsDNA (100 μM stock)	2.0 μL
EconoTaq 2× Master Mix	25 μL
Fluorescent probe FW primer (100 μM)	3.0 μL
Nuclease-free water	20 μL
Total	50 μL (per reaction)

Oligonucleotides needed	
Name	**Sequence**
IR700 dsDNA FW Primer	IR700/CTTACGGCGGCATTGGCACG
Variant with reference (non-risk) allele	CTTTGCTTA**C**CTATATTTTC CGTGCCAATGCCGCCGTAAG
Variant with alternate (risk) allele	CTTTGCTTA**T**CTATATTTTC CGTGCCAATGCCGCCGTAAG

Note: During this protocol we will be using the variant rs56992000 (C > T) from *Peña-Martínez and Messon-Bird et al.* (2025) as an example for subsequent experiments.

Extension reaction (1× cycle)			
Step	**Temperature**	**Time**	**Cycles**
Denaturing	95°C	2:00 min	1
Annealing	68°C	1: 00 min	
Extension	72°C	5: 00 min	
Hold	4°C	forever	

7. To remove free primer and dNTPs, purify fluorescently labeled dsDNA using the Qiagen PCR purification kit. For higher concentrations, elute with 30 μL of EB Buffer (included with the kit, 10 mM Tris-HCl pH 8.5) or nuclease-free water.
8. Measure DNA concentration and purity by absorbance at 260 nm using a Nanodrop UV-Vis spectrophotometer.

Note: Fluorescent DNA probe concentration usually ranges from 20–150 ng/μL.

9. Prepare 1 μM aliquots of fluorescent DNA probes by diluting in EB buffer and store at −20°C.

⚠ CRITICAL: Maintain DNA covered throughout procedure to protect from light.

Cloning expression vectors

Timing: 1 week

The candidate regulatory elements are 60 bp synthetic oligonucleotides centered on the non-coding variant to be tested. They are cloned upstream of a minimal promoter driving luciferase reporter gene expression. Cloning is performed by Gibson assembly of the PCR linearized gene reporter plasmid, and the PCR amplified non-coding element.

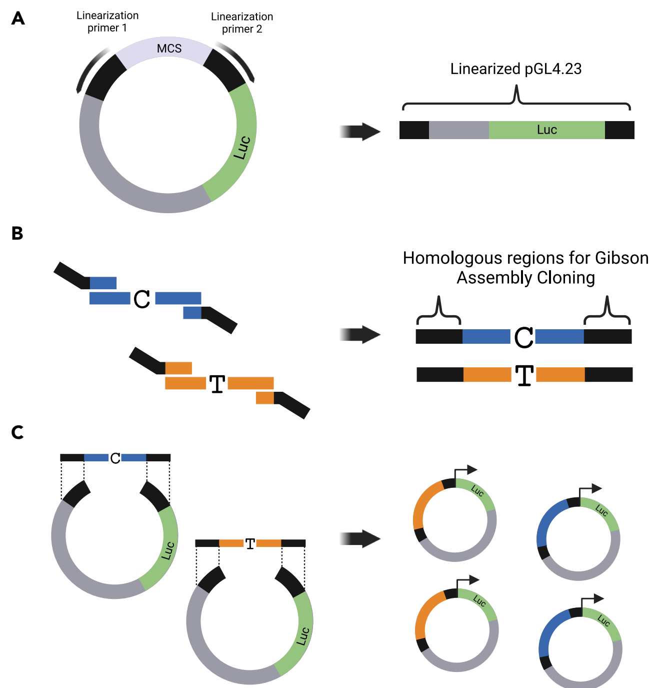


Figure 3. Cloning genomic sequences centered on variant into the reporter construct
Steps for (A) Linearization of pGL4.23, (B) amplification of 60 bp variant sequences and addition of homologous regions, and (C) Gibson assembly cloning are detailed in the Step-by-step methods section.

10. Linearize the pGL4.23[Luc2/minP–firefly luciferase] plasmid using Phusion High-Fidelity DNA Polymerase with GC Buffer (Figure 3A).

Plasmid linearization PCR reaction master mix

Reagent	Amount
pGL4.23	10 ng
2× Phusion GC Master Mix	25 μL
FW primer (10 μM)	2.5 μL
RV primer (10 μM)	2.5 μL
DMSO	1.5 μL
Nuclease-free water	Up to 50 μL
Total	50 μL (per reaction)

Primers to linearize pGL4.23 plasmid

Name	Sequence
luc2/NlucP GA Fw Primer	AGACACTAGAGGGTATATAATGGAAG
luc2/NlucP GA Rv Primer	AGAGAAATGTTCTGGCACCTGCAC

11. Set up the PCR reaction:

PCR cycling conditions (35× cycle)

Step	Temperature	Time	Cycles
Initial Denaturing	98 °C	30 s	1
Denaturing	98 °C	10 s	35 cycles
Annealing	68 °C	30 s	
Extension	72 °C	2:15 min	

(Continued on next page)

Continued			
Step	Temperature	Time	Cycles
Final Extension	72 °C	7:00 min	1
Hold	4 °C	forever	

12. Run the PCR products on a 1% agarose gel.
13. Excise the linearized plasmid band from the gel and extract DNA using the QIAquick Gel Extraction Kit.
14. Measure the DNA concentration using absorbance at 260 nm in Nanodrop UV-Vis spectrophotometer.

Note: The expected DNA concentration ranges from 30–200 ng/µL. About 50 ng of linearized plasmid is needed for cloning in subsequent steps.

15. Amplify the 60 bp sequences centered on the selected SNPs using Phusion High-Fidelity DNA Polymerase with GC Buffer (Figure 3B).

Note: This step also adds constant regions for Gibson assembly cloning to linearized pGL4.23. The genomic sequences can also be amplified using human or mouse genomic DNA.

PCR reaction master mix	
Reagent	Amount
60 bp oligonucleotide	0.5 µL or required ng of DNA
2× Phusion GC Master Mix	25 µL
FW primer (10 µM)	2.5 µL
RV primer (10 µM)	2.5 µL
DMSO	1.5 µL
Nuclease-free water	Up to 50 µL
Total	50 µL (per reaction)

PCR cycling conditions (35× cycle)			
Step	Temperature	Time	Cycles
Initial Denaturing	98 °C	30 s	1
Denaturing	98 °C	10 s	35 cycles
Annealing	68 °C	30 s	
Extension	72 °C	15 s	
Final Extension	72 °C	7:00 min	1
Hold	4 °C	forever	

16. Run the PCR products on a 3% agarose gel, purify the amplified DNA using the QIAquick Gel Extraction Kit, and measure the DNA concentration using absorbance at 260 nm in Nanodrop UV-Vis spectrophotometer.
17. Clone the sequences upstream of the minimal promoter into the pGL4.23 plasmid using Gibson Assembly and transform into competent NEB 5-alpha *E. coli* according to the manufacturer's protocol (NEB High Efficiency Transformation Protocol).

Note: For our 90 bp insert fragment (60 bp + 15 bp homologous regions at each end), we use the manufacturer's recommended 5:1 molar insert: vector ratio and 15-min incubation at 50°C (Figure 3C).

Gibson cloning mix	
Reagent	Amount
Linearized plasmid	50 ng
Insert sequence with ref or alt allele	3.58 ng
2× Gibson Assembly Master Mix	10 μL
Nuclease-free water	Up to 20 μL
Total	20 μL (per reaction)

18. Perform Colony PCR using EconoTaq PLUS 2× Master Mix to confirm the presence and approximate size of the expected insert.
    a. Prepare colony PCR Reaction Master Mix and add a single colony to each PCR tube and gently mix.
    b. Before performing PCR, plate 1 μL of the master mix with the bacterial colony to a Luria Broth (LB) agar plate with ampicillin for downstream bacterial culture after positive colony identification.
    c. Perform colony PCR and evaluate the amplicon size of each colony in a 1% agarose gel.

Note: We recommend evaluating 5–10 colonies per desired construct.

Colony PCR reaction master mix	
Reagent	Amount
EconoTaq PLUS 2× Master Mix	12.5 μL
FW primer (10 μM)	2.5 μL
RV primer (10 μM)	2.5 μL
Nuclease-free water	7.5 μL
Total	25 μL (per reaction)

PCR cycling conditions (35× cycle)			
Step	Temperature	Time	Cycles
Initial Denaturing	95 °C	2:00 min	1
Denaturing	95 °C	30 s	35 cycles
Annealing	47 °C	1 min	
Extension	72 °C	30 s	
Final Extension	72 °C	5:00 min	1
Hold	4 °C	forever	

Oligonucleotides for colony PCR	
Name	Sequence
Luc2/Nluc Fw Primer	CATCAAAACAAAACGAAACAAAACAAA
Luc2 Rv Primer	CGTCTTCGAGTGGGTAGAATG

19. Grow positive bacterial colonies in 50 mL of LB medium with Ampicillin.
20. Extract plasmid DNA using the QIAGEN Plasmid Plus Midi Kit and elute with 200 μL of Elution Buffer.
21. Measure DNA concentration using Qubit dsDNA BR Assay Kit.
22. We recommend verifying the whole plasmid sequence before proceeding with the reporter assay. We use Whole Plasmid Sequencing with Oxford Nanopore (Plasmidsaurus).

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-His mouse monoclonal antibody (1:10,000)	Novus Biologicals	AD1.1.10
Bacterial and virus strains		
BL21 DE3 *E. coli*	MilliporeSigma	70956–4
5-alpha competent *E. coli*	NEB	C2987
Chemicals, peptides, and recombinant proteins		
6xHis-GATA4 recombinant protein	Peña-Martínez et al.[1]	N/A
Luria Broth	Sigma-Aldrich	L3022
Terrific bbroth	Sigma-Aldrich	T0918
β-d-1-thiogalactopyranoside (IPTG)	Sigma-Aldrich	I6758
Ni-NTA resin	QIAGEN	30210
Tween-20	Sigma-Aldrich	P9416
Glycerol	Sigma-Aldrich	G5516
Sodium chloride (NaCl)	Fisher Scientific	S671-3
UltraPure 1 M Tris-HCl, pH 8.0	Invitrogen	15568025
Imidazole	Sigma-Aldrich	I2399
Pierce EDTA-free protease inhibitor	Thermo Scientific	A32965
Amicon ultra centrifugal filter, 3 kDa MWCO	MilliporeSigma	UFC500396
HEPES	VWR	7365-45-9
Zinc acetate	Sigma-Aldrich	383058
Mini-PROTEAN TGX Precast Protein Gels	Bio-Rad	N/A
4× Laemmli Sample Buffer	Bio-Rad	1610747
β-mercaptoethanol (BME)	Sigma-Aldrich	M3148
ProtoStain Blue Colloidal Coomassie G-250 stain	National Diagnostics	EC-727
10× TBE Buffer (Tris/Boric Acid/EDTA)	Bio-Rad	1610770
ProtoGel (30%)	National Diagnostics	EC890
Ammonium persulfate	Sigma-Aldrich	A3678
N,N,N′,N′-tetramethylethane-1,2-diamine (TEMED)	Thermo Scientific	17919
Poly(dI-dC)	Sigma-Aldrich	P4929
BSA	Fisher Scientific	23210
EconoTaq DNA polymerase	Lucigen	30035–2
Phusion High-Fidelity PCR Master Mix with GC Buffer	NEB	M0532S
Eagle's minimum essential medium	ATCC	30–2003
Fetal bovine serum	Sigma-Aldrich	12106C
1× RIPA lysis buffer	Rockland	MB-030-0050
Penicillin-Streptomycin solution stabilized	Sigma-Aldrich	P4333
Hank's balanced salt solution	Sigma-Aldrich	H6648
TrypLE Express Enzyme (1×), phenol red	Thermo Scientific	12605010
FuGENE 6	Promega	E2691
Critical commercial assays		
QIAquick PCR purification kit	QIAGEN	28104
Gibson Assembly Master Mix	NEB	E2611L
QIAquick Gel Extraction Kit	QIAGEN	28704
QIAGEN Plasmid Plus Midi Kit	QIAGEN	12943
Pierce BCA Protein Assay Kit	Thermo Scientific	A55865
Nano-Glo Dual-Luciferase Reporter Assay System	Promega	N1610
Experimental models: Cell lines		
HeLa cervical adenocarcinoma human cells	ATCC	CCL-2
Oligonucleotides		
IR700 dsDNA FW Primer	IDT	N/A
40 bp genomic sequence centered on variant for binding assay (rs56992000-ref)	IDT	N/A
40 bp genomic sequence centered on variant for binding assay (rs56992000-alt)	IDT	N/A
luc2/NlucP GA Fw Primer	IDT	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
luc2/NlucP GA Rv	IDT	N/A
60 bp genomic sequence centered on variant for reporter assay (rs56992000-ref)	IDT	N/A
60 bp genomic sequence centered on variant for reporter assay (rs56992000-alt)	IDT	N/A
Gibson assembly Fw primer for 60 bp genomic sequence	IDT	N/A
Gibson assembly Rv primer for 60 bp genomic sequence	IDT	N/A
Recombinant DNA		
pET-28a(+)-GATA4	Twist Bioscience	N/A
pGL4.23 [Luc2/minP - firefly luciferase]	Promega	E8411
pNL1.1.CMV [Nluc/CMV] Vector	Promega	N1091
GATA4 pTwist CMV BG WPRE Neo	Twist Bioscience	N/A
Software and algorithms		
Prism 10.4.1 (https://www.graphpad.com/features)	GraphPad	N/A
ImageJ 1.54p (https://imagej.net/ij/)	NIH	N/A
Other		
Econo-Pac Chromatography columns	Bio-Rad	7321010
1.5 mL DNA LoBind tubes	Eppendorf	0030108051
PCR tube strips	USA Scientific	1402–2400
Greiner culture flasks, tissue culture treated	Sigma-Aldrich	C6481
Corning cell scrapers	Sigma-Aldrich	CLS3010
Corning 96 well white polystyrene microplate flat-bottom clear white polystyrene (TC-treated)	Sigma-Aldrich	CLS3610
Sonicator	QSonica	Q125
NanoDrop spectrophotometer	Thermo Scientific	ND-1000
T100 Thermo cycler	Bio-Rad	1861096
Electrophoresis Power Supply	VWR	G-58517
Mini-PROTEAN Tetra Vertical Electrophoresis Cell	Bio-Rad	1658004
Sapphire Biomolecular Imager	Azure Biosystem	76318–254
Tecan Infinite 200 Pro microplate reader	Tecan Group Ltd.	1055090

MATERIALS AND EQUIPMENT

- Alternative applications of the protocol.

The binding assays for this protocol were described using the DNA-binding domain of a TF. Alternatively, this protocol can be optimized for full-length proteins or cell or nuclear extracts expressing the protein of interest.

The reporter assays described were tested using HeLa and HEK293T cells. However, this protocol can be optimized for other adherent cell lines. Depending on cell type and experimental requirements, alternative reagents, such as different transfection reagents (FuGENE HD) or culture media (DMEM), may be used. Adjustments may be necessary to achieve optimal assay performance.

- Oligonucleotides used in this work.

Name	Sequence
IR700 dsDNA FW Primer	IR700/CTTACGGCGGCATTGGCACG
40 bp genomic sequence centered on variant for binding assay (rs56992000-ref)	CTTTGCTTA**C**CTATATTTTC CGTGCCAATGCCGCCGTAAG
40 bp genomic sequence centered on variant for binding assay (rs56992000-alt)	CTTTGCTTA**T**CTATATTTTC CGTGCCAATGCCGCCGTAAG
luc2/NlucP GA Fw Primer	AGACACTAGAGGGTATATAATGGAAG

(*Continued on next page*)

Continued	
Name	Sequence
luc2/NlucP GA Rv	AGAGAAATGTTCTGGCACCTGCAC
60 bp genomic sequence centered on variant for reporter assay (rs56992000-ref)	GGTTTATGGATTTTT TTACTCTTTGCTTACCTATATTTTCTACAA TAAACATTATCATTT
60 bp genomic sequence centered on variant for reporter assay (rs56992000-alt)	GGTTTATGGATTTTT TTACTCTTTGCTTATCTATATTTTCTACAA TAAACATTATCATTT
Gibson assembly Fw primer for 60 bp genomic sequence	CCAGAACATTTCTCT GGTTTATGGATTTTT
Gibson assembly Rv primer for 60 bp genomic sequence	TACCCTCTAGTGTCT AAATGATAATGTTTA

Note: Constant regions used for dsDNA generation and reporter construct cloning are underlined. These sequences should not be modified for direct application of this protocol.

- Preparation of buffers for protein purification

Column buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.2% Tween-20, 30 mM imidazole, and EDTA-free protease inhibitor)

Reagent	Final concentration	Amount
1 M Tris-HCl pH 8.0	20 mM	20 mL
5 M NaCl	500 mM	100 mL
Tween-20	0.2 %	2 mL
1 M Imidazole	30 mM	30 mL
EDTA-free Protease Inhibitor	N/A	1 tablet
H_2O	N/A	Up to 1 L
Total	N/A	1 L
Store at 4 °C for up to 6 months		

Wash buffer 1 (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.2% Tween-20, 50 mM imidazole, and EDTA-free protease inhibitor)

Reagent	Final concentration	Amount
1 M Tris-HCl pH 8.0	20 mM	20 mL
5 M NaCl	500 mM	100 mL
Tween-20	0.2 %	2 mL
1 M Imidazole	50 mM	50 mL
EDTA-free Protease Inhibitor	N/A	1 tablet
H_2O	N/A	Up to 1 L
Total	N/A	1 L
Store at 4 °C for up to 6 months		

Wash buffer 2 (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.2% Tween-20, 100 mM imidazole, and EDTA-free protease inhibitor)

Reagent	Final concentration	Amount
1 M Tris-HCl pH 8.0	20 mM	20 mL
5 M NaCl	500 mM	100 mL
Tween-20	0.2 %	2 mL
1 M Imidazole	100 mM	100 mL
EDTA-free Protease Inhibitor	N/A	1 tablet
H_2O	N/A	Up to 1 L
Total	N/A	1 L
Store at 4 °C for up to 6 months		

Elution buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.2% Tween-20, 500 mM imidazole, and EDTA-free protease inhibitor)		
Reagent	Final concentration	Amount
1 M Tris-HCl pH 8.0	20 mM	20 mL
5 M NaCl	500 mM	100 mL
Tween-20	0.2 %	2 mL
1 M Imidazole	500 mM	500 mL
EDTA-free Protease Inhibitor	N/A	1 tablet
H_2O	N/A	Up to 1 L
Total	N/A	1 L
Store at 4 °C for up to 6 months		

- Preparation of buffer for EMSA.

Binding buffer 5× (100 mM NaCl, 10 mM HEPES pH 8.0, 0.5 μM zinc acetate, 200 mM NH4 acetate, 20% glycerol)		
Reagent	Final concentration	Amount
1 M HEPES pH 8.0	10 mM	1 mL
5 M NaCl	100 mM	0.2 mL
1 mM Zinc acetate	0.5 μM	50 μL
1 M NH_4	200 mM	20 mL
Glycerol	20 %	20 mL
H_20	N/A	Up to 100 mL
Total	N/A	100 mL
Store at 4 °C for up to 6 months		

- Preparation of 6% polyacrylamide gel (solution for two 1.5 mm gels)

Reagent	Final concentration	Amount
30% Acrylamide/Bis Solution	6%	6.0 mL
5× TBE Buffer	0.5×	1.8 mL
Glycerol 80 %	4%	1.14 mL
H_20	N/A	20.76 mL
10 % Ammonium persulfate (APS)	1 mg/mL	240 μL
TEMED	0.2%	60 μL
Total		30 mL

Warning: APS, TEMED, and acrylamide are toxic. Read each reagent's safety data sheet and handle it with caution.

> ⚠ CRITICAL: Add ammonium persulfate and tetramethylethylenediamine (TEMED) last to induce polymerization. Quickly cast gels following the addition of these reagents.

STEP-BY-STEP METHOD DETAILS

Electrophoretic mobility shift assay

⏱ Timing: 6 h

In this section we describe the steps and binding conditions to measure TF-DNA interaction changes due to non-coding variants. 40 bp DNA sequences centered on the variant are used during binding assay. At the end of this step, measurements of bound and unbound DNA will be obtained to quantify binding fractions and changes in affinity.

CellPress
OPEN ACCESS

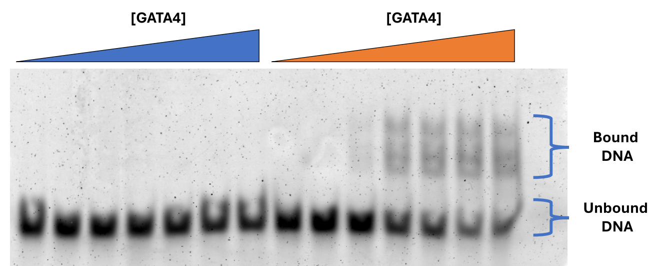


Figure 4. Example gel shift assay to evaluate TF-DNA binding

Lanes used for reference DNA are represented in blue, and alternate DNA in orange.

1. Prepare a 6% polyacrylamide gel (8.3 × 7.3 cm) of 1.5 mm thickness for electrophoresis.

   Note: We prepare 15-well gels to load seven lanes of the reference allele, seven lanes of the alternate allele, and a final lane to load a dye for gel monitoring.

6% polyacrylamide solution for two 1.5 mm gels

Reagent	Final concentration	Amount
30% Acrylamide/Bis Solution	6%	6.0 mL
5× TBE Buffer	0.5 ×	1.8 mL
Glycerol 80%	4%	1.14 mL
H_20	N/A	20.76 mL
10 % Ammonium persulfate (APS)	1 mg/mL	240 μL
TEMED	0.2 %	60 μL
Total		30 mL

Note: Gel percentage might have to be optimized if protein-DNA complex is too large for gel migration.

⚠ CRITICAL: Add Ammonium persulfate (APS) and Tetramethyl ethylenediamine (TEMED) last to induce polymerization. Quickly cast gels following the addition of these reagents.

Warning: APS, TEMED, and acrylamide are toxic. Read each reagent's safety data sheet and handle them with caution.

2. Prepare TF-DNA binding reactions with a final concentration of 5 nM fluorescently labeled dsDNA.

TF binding reactions

Reagent	Final concentration	Amount
5× Binding Buffer	1×	4.0 μL
pdI-dC (500 ng/μL)	1000 ng	2.0 μL
BSA (1 mg/mL)	0.025 mg/mL	0.5 μL
Tween-20 (stock 1%)	0.05 %	1.0 μL
DTT (stock 1 M)	10 mM	0.2 μL
Fluorescent dsDNA (1 μM stock)	5 nM /per reaction	0.1 μL
TF	0-2,000 nM	Up to 20 μL
Nuclease-free water	N/A	Up to 20 μL
Total (Single reaction)		20 μL

Note: We suggest preparing a binding reaction master mix (reagents x number of TF concentration points tested). We usually perform binding assays for seven concentration points.

⚠ CRITICAL: Proteins should be thawed on ice ~15 min prior to beginning experiment and should always remain on ice.

Example for 7 binding reactions		
Reagent	Final concentration	Amount
5× Binding Buffer of TF being assayed	1×	28.0 μL
pdI-dC (500 ng/μL)	50 ng/μL	14.0 μL
BSA (1 mg/mL)	0.025 mg/mL	3.5 μL
Tween-20 (stock 1%)	0.05 %	7.0 μL
DTT (stock 1 M)	1 mM	1.4 μL
dsDNA (1 μM stock)	5 nM	0.7 μL
TF	0-2,000 nM	Add separately
Nuclease-free water	N/A	Add separately
Total		140 μL

Example for titrating TF concentration (nM)							
Reagent	0 nM	50 nM	100 nM	500 nM	1,000 nM	1,500 nM	2,000 nM
Binding Reaction Master Mix (μL)	9.4	9.4	9.4	9.4	9.4	9.4	9.4
1× Binding Buffer of TF being assayed (μL)	5	–	–	–	–	–	–
Nuclease-free water (μL)	5.6	10.5	10.4	9.6	8.6	7.6	6.6
TF (10 μM stock, μL)	–	0.1	0.2	1.0	2.0	3.0	4.0
Total (μL)	20	20	20	20	20	20	20

Note: For this step, you should already have the TF of interest purified and the binding reaction conditions optimized (see troubleshooting 2). For additional troubleshooting and optimization, we recommend the following article on gel shift assay.[11]

⚠ CRITICAL: Return proteins to −80°C after adding to binding reaction.

3. Incubate binding reactions for one h. We perform incubations at 30°C for 30 min and then 30 min at 20–25°C, but this can vary when optimizing for different TFs.
4. Pre-run the polyacrylamide gel for 15 min at 75 V at 4°C during binding reaction incubation.
5. Load 18 μL of binding reaction into a polyacrylamide gel and run it at 30 V. After loading all samples, run the gel for 1 h at 120 V at 4°C.

   Alternative: If a cold room set-up is not available, steps four and five could also be performed at 20–25°C. This might require voltage and running time optimization.

   Note: We recommend leaving an empty well at the end to load a visible dye and monitor gel migration. We use 4× Laemmli sample buffer (Bio-Rad #1610747).

6. To evaluate TF-DNA complex formation, image gel at 658 nm excitation and 710 nm emission. This will appear as a slower migration band (shift) on the gel (Figure 4). To quantify changes in TF binding to reference and alternate alleles, see the quantification and statistical analysis section.

Dual-luciferase reporter assay

⏲ Timing: 1 week

In this section, we describe the steps for cell culture and transfection to measure changes in gene expression due to non-coding variants. Expression vectors containing 60 bp DNA sequences centered on the variant are used for a Luciferase reporter assay. At the end of this step, measurements of luciferase activity will be obtained to quantify changes in gene expression.

7. Grow HeLa Cervical Adenocarcinoma Human cells in Eagle's Minimum Essential Medium (EMEM) supplemented with 10% Fetal Bovine Serum (FBS) and 10 mL/L Penicillin-Streptomycin at 37°C and 5% CO_2 in a T25 flask.

   Note: Confluency may vary by up to two days or more, depending on the cell type. Ensure the medium is changed regularly. Maintain cells healthy at least 90% confluency.

8. Prepare the DNA for reverse transfection. In a 1.5 ml tube, mix the following DNA components:
   a. 7.5 ng of NanoLuc-luciferase-CMV vector.
   b. 75 ng of Ref or Alt Variant-minP-luciferase vector.
   c. 25 ng of the transcription factor vector (pTwist-CMV-Beta-Glob-WPRE-Neo).

   Note: The total amount of the reporter DNA is 82.5 ng (NanoLuc-luciferase + luciferase). For co-transfection of transcription factors (TF), the total DNA amount added should be 50 ng. If transfecting only one TF, add 25 ng of a Green Fluorescent Protein (GFP) vector to maintain a consistent total DNA quantity.

9. Prepare a Reverse Transfection Reaction mix using a 3:1 of FuGENE 6:DNA volume to mass ratio (e.g., 0.6 μL of FuGENE 6 to 0.2 μg of DNA), ensuring a final volume of 10 μL, including DNA. In a clean 1.5 ml tube, add the following components:
   a. Medium EMEM without FBS. (Adjust to ensure a total volume of 10 μL, after adding all components.).
   b. Add 0.6 μL of FuGENE 6 transfection reagent.
   c. Vortex the mix for 5–10 sec, spin and incubate for 5 min at 20–25ºC. If a vortex is not available mix by pipetting up and down.

   Note: We suggest preparing a Reverse Transfection Reaction master mix (reagents x number replicates tested). We perform triplicates for each sequence tested.

Reverse transfection reaction

Reagent	96-Well plate	6-Well plate
NanoLuc-luciferase-CMV vector	7.5 ng	225 ng
Ref/Alt Variant-minP-luciferase vector	75 ng	2,250 ng
Transcription factor	25 ng	750 ng
FuGENE 6	0.6 μL	18 μL
Medium only	Up to 10 μL	Up to 300 μL
Total	10 μL (per reaction)	300 μL (per reaction)

10. Add the Reverse Transfection Reaction mix to the prepared DNA mix (from step 8).
11. Incubate at 20–25ºC for 15–45 min.
12. Add 10 μL of the transfection + DNA mix to each well of a Corning 96 Well White Polystyrene Microplate.
13. Prepare cells for splitting.
    a. Remove the medium through vacuum suction.
    b. Wash cells with 1 mL of Hanks' Balanced Salt Solution (HBSS).
    c. Discard HBSS, and add 1 mL of TrypLE and incubate for 5 min at 37°C and 5% CO_2.
    d. Add 1 mL of medium and transfer the suspension to a 15 mL tube.

e. Centrifuge for 5–7 min at 125 × *g*.
f. Carefully remove the supernatant and resuspend the cells in 1 mL of fresh medium.
g. Count the cells using a cell counter.
h. Seed 20,000 cells per well in the 96-well plate.

Note: The volume for cell splitting may be adjusted based on the flask used. According to Promega's recommendation, 80 μL of cells were seeded per well. Be mindful of the risk of overflow, as two equivalent volumes will be added to each well. For problems during transfection *see* troubleshooting 3.

14. Incubate the cells for 24 h at 37°C and 5% CO_2.

Note: If necessary, incubation can be extended up to 48 h if necessary.

15. After 24 h, remove the plate from the incubator and allow it to equilibrate to 20–25ºC for a few min.
16. Lyse cells according to the Nano-Glo Dual-Luciferase Reporter (NanoDLR) Assay System's protocol from Promega (Nano-Glo Dual-Luciferase Reporter Assay System).
    a. Add one volume of ONE-Glo EX Luciferase Reagent equivalent to the culture medium to each well.
    b. Incubate for 3 min while shaking at 300 rpm.
    c. Measure firefly luciferase activity using a Tecan plate reader Infinite 200 Pro (Tecan) in the Luminescence program with no attenuation and 0.5 sec integration time.
    d. Add one volume of NanoDLR Stop & Glo Reagent equivalent to the culture medium to each well.
    e. Incubate for 10 min, shaking at 300 rpm.
    f. Measure NanoLuc luciferase activity.

Note: To quantify changes in gene expression between reference and alternate allele see the quantification and statistical analysis section.

Note: In this case, 80 μL of each reagent was used. The parameters for both luminescence measurements are the same. For problems during luminescence measurement, see troubleshooting 4.

EXPECTED OUTCOMES

The outcome of this protocol is measurements of binding affinity and TF-dependent gene expression of non-coding variants. The sections below describe the software and equipment used to quantify the binding and reporter assay and the statistical analysis used to determine significance.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantifying changes in fluorescent signal in the gel between reference and alternate allele

Timing: 1 h

For this step, we use the ImageJ software[12] to quantify pixel intensity in bound and unbound DNA fractions. These values will be used to calculate the binding fractions and determine changes in binding between reference and alternate alleles.

1. Open picture of gel in the preferred image analysis software.

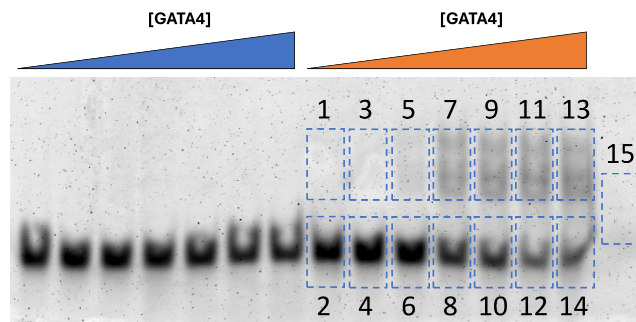


Figure 5. Quantifying TF-DNA binding from gel image
Areas used for pixel quantification are represented with blue dashed squares. Bound fractions are indicated with odd numbers and unbound with even numbers. The area used for subtracting background is #15. This figure was generated from Figure 4.

Note: We use ImageJ software, but other alternatives such as CellProfiler and QuPath should also work. Subsequent steps are described specifically for ImageJ software.

Note: Convert image into black and white. You may need to invert image so that DNA bands appear as white.

2. Quantify bound and unbound fractions across all concentrations using the same surface area (Figure 5).

   Note: This is done using the "Analyze" tab and selecting "Measure". A table will be displayed indicating your pixel intensity in the Integrated Density (IntDen) column.

   Note: We recommend quantifying background intensity from an empty lane and subtracting it from all measurements.

3. Calculate bound fraction for each concentration using Equation 1. Perform this for triplicate gels of both the reference and alternate allele.

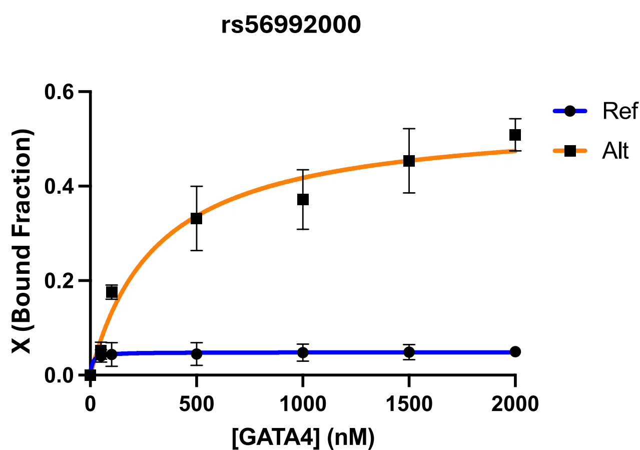


Figure 6. Binding curves of reference and alternate sequences
This binding curve was generated using gel from Figure 4. Error bars represent variation from experimental triplicates. Figure re-used from: Peña-Martínez, E.G. et al, Cardiovascular disease-associated non-coding variants disrupt GATA4-DNA binding and regulatory functions, Copyright Elsevier (2025). Permission license number: 6015491253169.

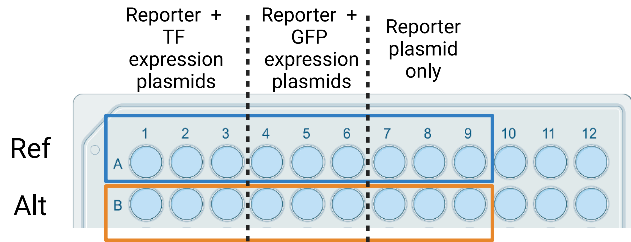


Figure 7. Quantifying luciferase activity on 96-well plates
Luminescent measurements are performed in triplicates for reference (blue rectangle) and alternate (orange rectangle) alleles. Experimental design should also include triplicate controls that only contain reporter plasmids for luciferase activity normalization.

4. Once the binding fraction is calculated for all concentration points, binding curves are constructed using GraphPad Prism (Version 10.4.1) using the "One site–Specific binding" fit from the average of triplicate bound fractions (Figure 6).

$$Bound\ Fraction = \frac{(bound - background)}{(bound - backround)+(unbound - background)} \qquad \text{(Equation 1)}$$

Quantifying changes in gene expression between reference and alternate allele

Timing: 1 h

For this step, we use a Tecan Infinite 200 Pro microplate reader to quantify relative luciferase activity between the reference and alternate allele (Figure 7).

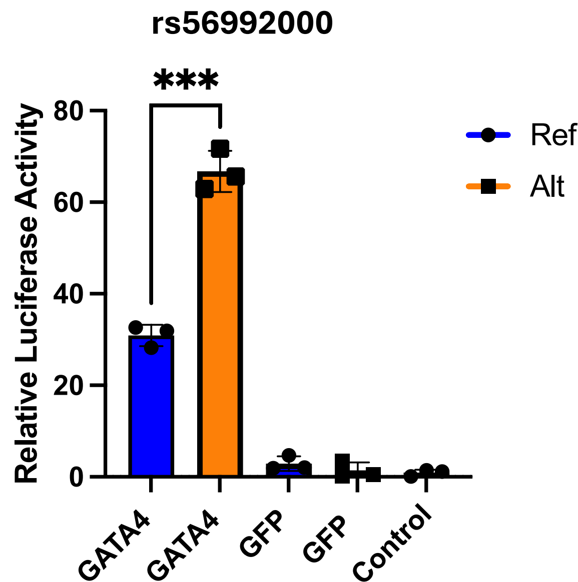


Figure 8. Quantifying changes in gene expression between reference and alternate sequences
Error bars represent variation from experimental triplicates. Significance was determined by un-paired t-test, *** $p<0.001$. Figure re-used from: Peña-Martínez, E.G. et al, Cardiovascular disease-associated non-coding variants disrupt GATA4-DNA binding and regulatory functions, Copyright Elsevier (2025). Permission license number: 6015491253169.

Note: Firefly and NanoLuc luciferase activity measurements should be performed in triplicates for both alleles.

5. Calculate relative luciferase activity by normalizing luminescent measurements with the control wells (reporter plasmids + GFP) using Equation 2.
6. Construct barplots of relative luciferase activity is calculated for both alleles and perform an unpaired t-test to determine significance (Figure 8).

Note: We construct bar plots using GraphPad Prism (Version 10.4.1).

$$\textit{Relative Luciferase Activity} = \frac{\textit{Average}\left(\frac{Luc2}{NanoLuc}\right)\textit{of reporter+TF}}{\textit{Average}\left(\frac{Luc2}{NanoLuc}\right)\textit{of reporter+GFP}} \qquad \text{(Equation 2)}$$

Alternative: Normalization can also be performed using control cells that are only transfected with the reporter plasmid.

LIMITATIONS

The binding assay in this protocol measures protein-DNA interactions in vitro using purified components. Hence, results are limited to binding events outside the cellular context, such as binding partners, co-factors, and chromatin remodeling of topologically associated domains (TADs). This method is not suitable for proteins that do not directly bind to DNA, such as co-factors. Additionally, different proteins should be optimized for their ideal temperature, pH, and buffer conditions. The reporter assays were performed in HeLa cells transfected to overexpress the TF of interest. Alternatively, the reporter assay can be optimized in a cell line that naturally expresses the regulatory protein of interest.

TROUBLESHOOTING

Problem 1

Step 1.

How to optimize conditions for protein expression and purification.

Sub-optimal protein induction conditions and purification can produce low amount of soluble protein.

Potential solution

Three parameters—IPTG concentration, temperature, and induction time—can be optimized for each individual protein to optimize protein induction conditions. For IPTG concentration, inductions are usually performed from 0.4 to 2.0 mM. Perform small-scale induction trials can be performed at multiple IPTG concentrations to evaluate protein expression yield and solubility. Induction time and temperature can be evaluated together by comparing expression at lower temperatures at longer times (e.g., 18°C for 20 h) versus higher temperature for a shorter period (37°C for 4–6 h). During purification, column washing and elution conditions can impact final protein purity and yield. In the associated manuscript, washes are performed with increasing imidazole concentration (30–100 mM) and eluted at 500 mM. Imidazole concentration during the column washing steps should be evaluated by collecting fractions and running them on an SDS-PAGE. Washes should be performed up to an imidazole concentration where most of the unwanted proteins are removed while avoiding eluting the protein of interest. For guidelines on recombinant protein expression and purification using bacterial systems, we recommend starting with the following manuscripts.[13,14]

Problem 2

Steps 2–6.

How to determine optimal conditions for binding reactions and EMSA.

Among possible problems while performing an EMSA are the following: DNA is not visible in the gel, there is no/low protein-DNA binding, the protein-DNA complex is trapped in the wells, and the band shift is smearing across the gel.

Potential solution

If fluorescent DNA probe is not visible in the gel, re-measure the concentration of the 1 μM DNA aliquots. If the aliquot is at the desired concentration, you can add a small amount to the gel imager (∼1 μL) and evaluate if the oligonucleotide is still fluorescent. If not fluorescent, re-examine primers with probes and re-prepare DNA probe. If fluorescent, reduce electrophoresis time or voltage to examine if DNA is running too fast through the gel.

If there is no or low protein-DNA binding, re-measure protein-DNA concentration to assure correct amount of protein is being added. If protein concentration is as desired, binding reaction conditions (buffer, ionic strength, additives, temperature, or incubation time) should be optimized for each protein. If available, a positive control sequence (known bound DNA sequence with high affinity) should also be added to the assay to determine if protein is functional after purification.

If the DNA sample is stuck in the well, the protein-DNA complex may be too large, or the protein could be aggregated. A possible solution is to reduce the acrylamide concentration to increase the gel's pore size or adjust the protein:DNA ratio. If protein aggregation is suspected, optimize binding reaction to improve protein solubility. Remember to maintain protein aliquots in an imidazole-free storage buffer with glycerol. Avoid constant freezing and thawing while keeping protein aliquot on ice prior to binding assay.

If the band shift is smearing across the gel, the protein-DNA complex might be unstable during the binding reaction or electrophoresis. To address this, electrophoresis conditions could be optimized: gel polymerization, voltage, temperature, and time. Check whether gel polymerization is even or at the desired percentage. Gel could also be overheating during electrophoresis. Try reducing voltage or performing EMSA in a cooling system at ∼4ºC. For further information on EMSA optimization and troubleshooting, we recommend Hellman L. and Fried M (2007).[11]

Problem 3

Steps 8-12.

How to determine transfection efficiency and optimal DNA concentration for reporter assay.

After initiating transfection, it's possible to have a low transfection efficiency or high cell mortality rate.

Potential solution

To assess transfection efficiency, it is recommended to include a reporter plasmid such as GFP. Seed 600,000 cells per well in a 6-well plate and test three different concentrations: 1,200 ng, 3,000 ng, and 6,000 ng of GFP and TF plasmid. Adjust FuGENE 6 volumes accordingly based on the total DNA added (4.5 μL, 9 μL, and 18 μL). After 24–48 h, evaluate GFP fluorescence using a microscope or Sapphire Biomolecular Imager to confirm successful transfection. Additionally, prepare cell lysates and perform Western blot analysis to evaluate protein expression.

If cells show high mortality post-transfection, the amount of DNA transfected may be excessive. Compare transfected cells with a negative control without a transfection mix. To mitigate toxicity, decrease the DNA concentration and/or shorten the incubation time to 24 h. If transfection efficiency remains low, consider using lower cell passage numbers and checking for contamination to ensure cell health.

Problem 4

Step 16.

Luminescence measurements were too low or high.

Measurement accuracy can be affected if the signal is below the detection threshold or too intense with a saturated signal.

Potential solution

If luminescence is low or absent, assess GFP fluorescence to confirm that the transfection was successful. Additionally, plasmid concentrations can be verified using Qubit, as a Nanodrop may overestimate the DNA concentration and not be enough for transfection. If necessary, repeat plasmid purification using a midiprep kit. If the problem persists, verify the plasmid sequences for any mutation with the luciferase or NanoLuc gene.

If excessive NanoLuc luminescence is detected, decrease the DNA concentration of the NanoLuc-luciferase-CMV vector and adjust the instrument settings for automatic attenuation.

RESOURCE AVAILABILITY

Lead contact

José A. Rodríguez-Martínez, PhD (jose.rodriguez233@upr.edu).

Technical contact

Edwin G. Peña-Martínez (Edwin.pena1@upr.edu) and Jessica M. Rodríguez-Ríos (Jessica.rodriguez8@upr.edu).

Materials availability

Oligonucleotides, cloned vectors, recombinant proteins, and cell lines are available upon request.

Data and code availability

There is no data or code related to this work.

ACKNOWLEDGMENTS

This project was supported by NIH-SC1GM127231, NSF (1736026), University of Puerto Rico Rio Piedras Institutional Funds (FIPI), Puerto Rico Science, Technology, and Research Trust, and NIH Institutional Development Award (IDeA) INBRE (P20GM103475). E.G.P.-M., J.M.R.-R., R.V.-R., and A.C.B.-R. were funded by NIH RISE Fellowship (5R25GM061151–20). E.G.M.-P. and J.M.R.-R. were funded by NSF BioXFEL Fellowship (STC-1231306). J.M.R.-R. was funded by NIH G-RISE 1T32GM152384-01. J.L.M.-B. and A.R.-M. were funded by NSF PR-LSAMP fellowship (HRD-2008186). E.A. P.-P. was supported by R21CA253609. A.R.-M. was funded by the NSF REU: PR-CLIMB Program (2050493) and NIH 1T34GM145404. J.M.R.-R. was funded by an NSF graduate research fellowship (1744619). The graphical abstract and Figures 3 and 7 were created using BioRender.

AUTHOR CONTRIBUTIONS

E.G.P.-M. and J.A.R.-M. developed and optimized the protein binding assays. J.M.R.-R. and E.A.P.-P. developed and optimized the reporter assays. R.V.-R. performed recombinant protein expression and purification. J.L.M.-B., A.C. B.-R., and A.R.-M. performed the experiments. E.G.P.-M. and J.M.R.-R. performed the data quantification and analysis. E.G.P.-M. and J.M.R.-R. wrote the manuscript. All authors read and approved the manuscript. J.A.R.-M. and E.A.P.-P. supervised the work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Peña-Martínez, E.G., Messon-Bird, J.L., Rodríguez-Ríos, J.M., Velázquez-Roig, R., Pomales-Matos, D.A., Rivera-Madera, A., Sanabria-Alberto, L., Barreiro-Rosario, A.C., Figueroa-Rosado, J.A., Rivera-Del Valle, J., et al. (2025). Cardiovascular disease-associated non-coding variants disrupt GATA4-DNA binding and regulatory functions. HGG Adv. *6*, 100415. https://doi.org/10.1016/j.xhgg.2025.100415.

2. Peña-Martínez, E.G., and Rodríguez-Martínez, J.A. (2024). Decoding Non-coding Variants: Recent Approaches to Studying Their Role in Gene Regulation and Human Diseases. Front. Biosci. (Schol Ed) *16*, 4. https://doi.org/10.31083/j.fbs1601004.

3. French, J.D., and Edwards, S.L. (2020). The Role of Noncoding Variants in Heritable Disease. Trends Genet. *36*, 880–891. https://doi.org/10.1016/j.tig.2020.07.004.

4. Maurano, M.T., Haugen, E., Sandstrom, R., Vierstra, J., Shafer, A., Kaul, R., and Stamatoyannopoulos, J.A. (2015). Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. Nat. Genet. *47*, 1393–1401. https://doi.org/10.1038/ng.3432.

5. Peña-Martínez, E.G., Pomales-Matos, D.A., Rivera-Madera, A., Messon-Bird, J.L., Medina-Feliciano, J.G., Sanabria-Alberto, L., Barreiro-Rosario, A.C., Rivera-Del Valle, J., Rodríguez-Ríos, J.M., and Rodríguez-Martínez, J.A. (2023). Prioritizing Cardiovascular Disease-Associated Variants Altering NKX2-5 and TBX5 Binding through an Integrative Computational Approach. J. Biol. Chem. *299*, 105423. https://doi.org/10.1016/j.jbc.2023.105423.

6. Jiang, X., Li, T., Liu, S., Fu, Q., Li, F., Chen, S., Sun, K., Xu, R., and Xu, Y. (2021). Variants in a cis-regulatory element of TBX1 in conotruncal heart defect patients impair GATA6-mediated transactivation. Orphanet J. Rare Dis. *16*, 334. https://doi.org/10.1186/s13023-021-01981-4.

7. Peña-Martínez, E.G., Rivera-Madera, A., Pomales-Matos, D.A., Sanabria-Alberto, L., Rosario-Cañuelas, B.M., Rodríguez-Ríos, J.M., Carrasquillo-Dones, E.A., and Rodríguez-Martínez, J.A. (2023). Disease-associated non-coding variants alter NKX2-5 DNA-binding affinity. Biochim. Biophys. Acta. Gene Regul. Mech. *1866*, 194906. https://doi.org/10.1016/j.bbagrm.2023.194906.

8. Lu, X., Chen, X., Forney, C., Donmez, O., Miller, D., Parameswaran, S., Hong, T., Huang, Y., Pujato, M., Cazares, T., et al. (2021). Global discovery of lupus genetic risk variant allelic enhancer activity. Nat. Commun. *12*, 1611. https://doi.org/10.1038/s41467-021-21854-5.

9. Dong, S., and Boyle, A.P. (2022). Prioritization of regulatory variants with tissue-specific function in the non-coding regions of human genome. Nucleic Acids Res. *50*, e6. https://doi.org/10.1093/nar/gkab924.

10. Lee, P.H., Lee, C., Li, X., Wee, B., Dwivedi, T., and Daly, M. (2018). Principles and methods of in-silico prioritization of non-coding regulatory variants. Hum. Genet. *137*, 15–30. https://doi.org/10.1007/s00439-017-1861-0.

11. Hellman, L.M., and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat. Protoc. *2*, 1849–1861. https://doi.org/10.1038/nprot.2007.249.

12. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

13. Kimple, M.E., Brill, A.L., and Pasker, R.L. (2013). Overview of affinity tags for protein purification. Curr. Protoc. Protein Sci. *73*, 9.9.1–9.9.23. https://doi.org/10.1002/0471140864.ps0909s73.

14. Young, C.L., Britton, Z.T., and Robinson, A.S. (2012). Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. Biotechnol. J. *7*, 620–634. https://doi.org/10.1002/biot.201100155.