The American Journal of Human Genetics, Volume *108*

Supplemental information

Non-coding region variants upstream of *MEF2C* cause severe developmental disorder through three distinct loss-of-function mechanisms

Caroline F. Wright, Nicholas M. Quaife, Laura Ramos-Hernández, Petr Danecek, Matteo P. Ferla, Kaitlin E. Samocha, Joanna Kaplanis, Eugene J. Gardner, Ruth Y. Eberhardt, Katherine R. Chao, Konrad J. Karczewski, Joannella Morales, Giuseppe Gallone, Meena Balasubramanian, Siddharth Banka, Lianne Gompertz, Bronwyn Kerr, Amelia Kirby, Sally A. Lynch, Jenny E.V. Morton, Hailey Pinz, Francis H. Sansbury, Helen Stewart, Britton D. Zuccarelli, Genomics England Research Consortium, Stuart A. Cook, Jenny C. Taylor, Jane Juusola, Kyle Retterer, Helen V. Firth, Matthew E. Hurles, Enrique Lara-Pezzi, Paul J.R. Barton, and Nicola Whiffin

Supplementary Data: Non-coding variants upstream of *MEF2C* cause severe developmental disorder through three distinct loss-of-function mechanisms

Figure S1: Two non-coding deletions remove the distal 5'UTR exon of *MEF2C* and the entire promoter sequence. Coding exons are shown in black with UTRs in red. The dotted line indicates the start of the coding sequence. The five deletions identified in DDD in the *MEF2C* region are shown as orange bars, the top two of which are entirely non-coding. A representative H3K4me3 dataset from ENCODE is plotted in blue across the top (GN12878) to show active promoter regions.

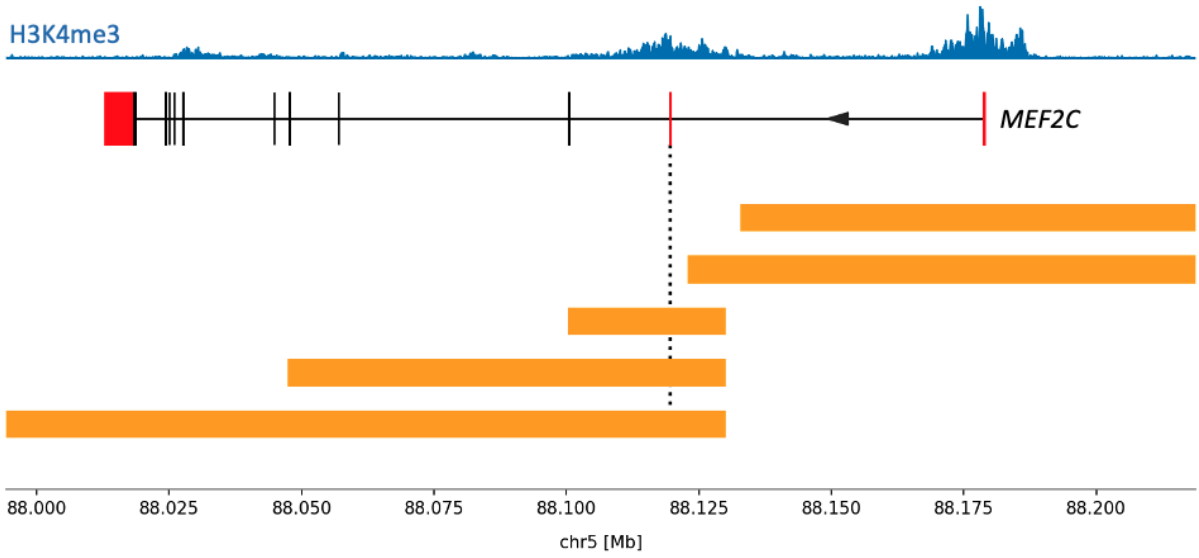


Figure S2: uAUG-creating variants do not alter RNA or protein levels. (A) Relative Gaussia luciferase (GLuc) RNA levels remain unchanged with each out-of-frame oORF-creating variant when normalised to RNA of secreted alkaline phosphatase (SEAP) transfection control. (B and C) The decreases in transactivation seen for the CDS-elongating variants c.-8C>T and c.-26C>T are not accompanied by a significant change in protein levels. For (A) and (C) bars are coloured by Kozak consensus: yellow = weak; orange = moderate; red = strong. ns = not significant.

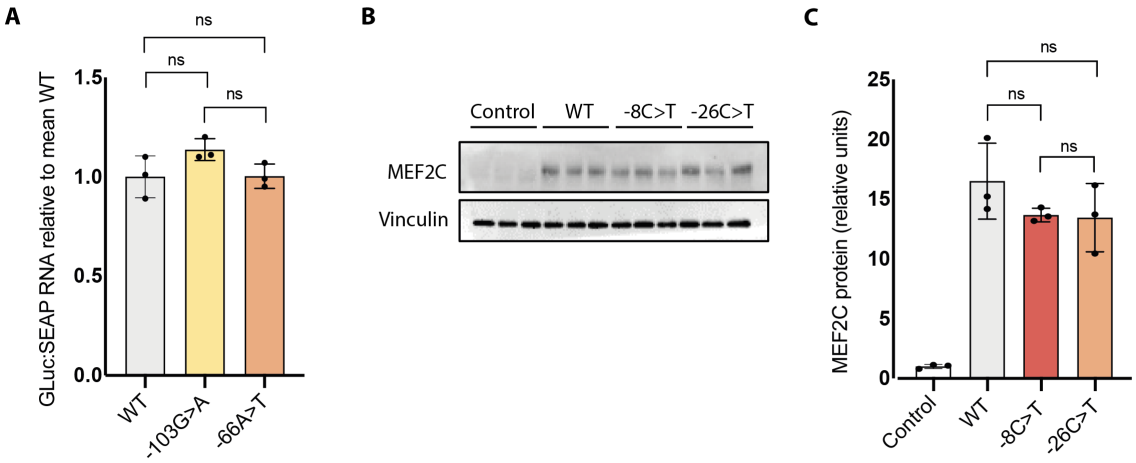


Figure S3: A single base mutation in the context surrounding the c.-103G>A variant which changes a weak Kozak consensus into a moderate consensus significantly reduces translational efficiency. (A) oORF-creating variants c.-103G>A and c.-66A>T reduce downstream luciferase expression relative to wild-type (WT) 5' UTR in a translation reporter assay. Reduction is stronger for c.-66A>T (moderate Kozak context) than for c.-103G>A (weak Kozak context). Modifying the context surrounding the c.-103G>A variant into a moderate Kozak context (as shown in B) reduces downstream luciferase expression compared to the unmodified vector. The translational efficiency of the modified vector is equivalent to the c.-66A>T variant which also has a moderate Kozak consensus. ns = not significant.

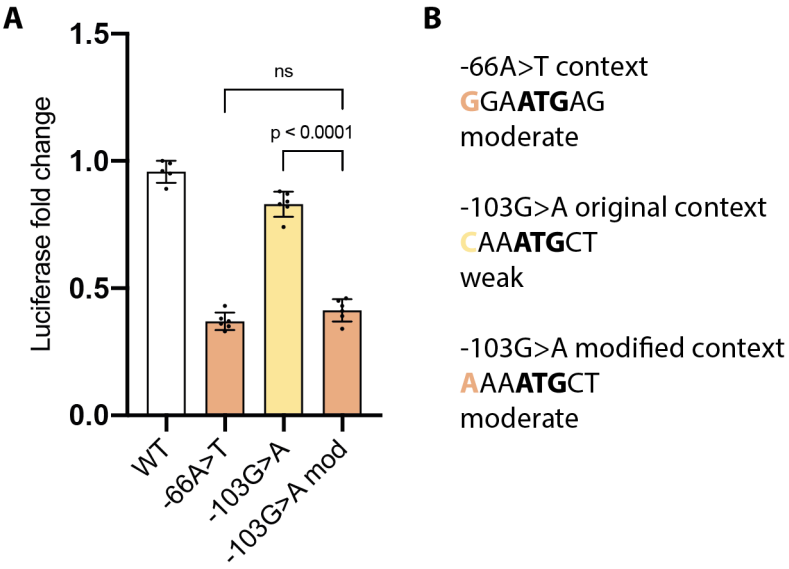


Figure S4: Protein sequence alignment of the four human myocyte enhancer factor 2 proteins proteins (MEF2A-D). The Clustal-Omega default alignment function in UniProt for the first 92 N-terminal residues was used. Coloured by similarity; * = identical amino acids in all 4 proteins; : = similar amino acids in all 4 proteins.

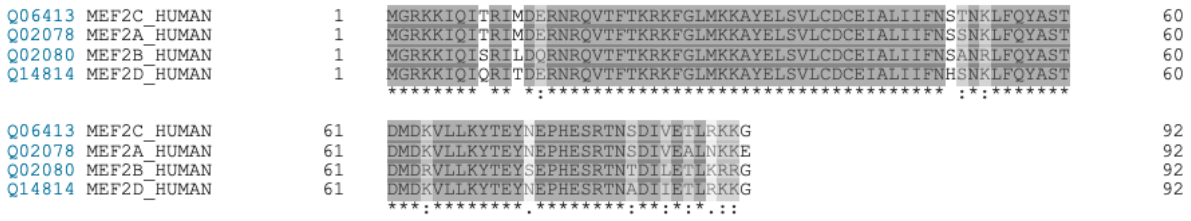
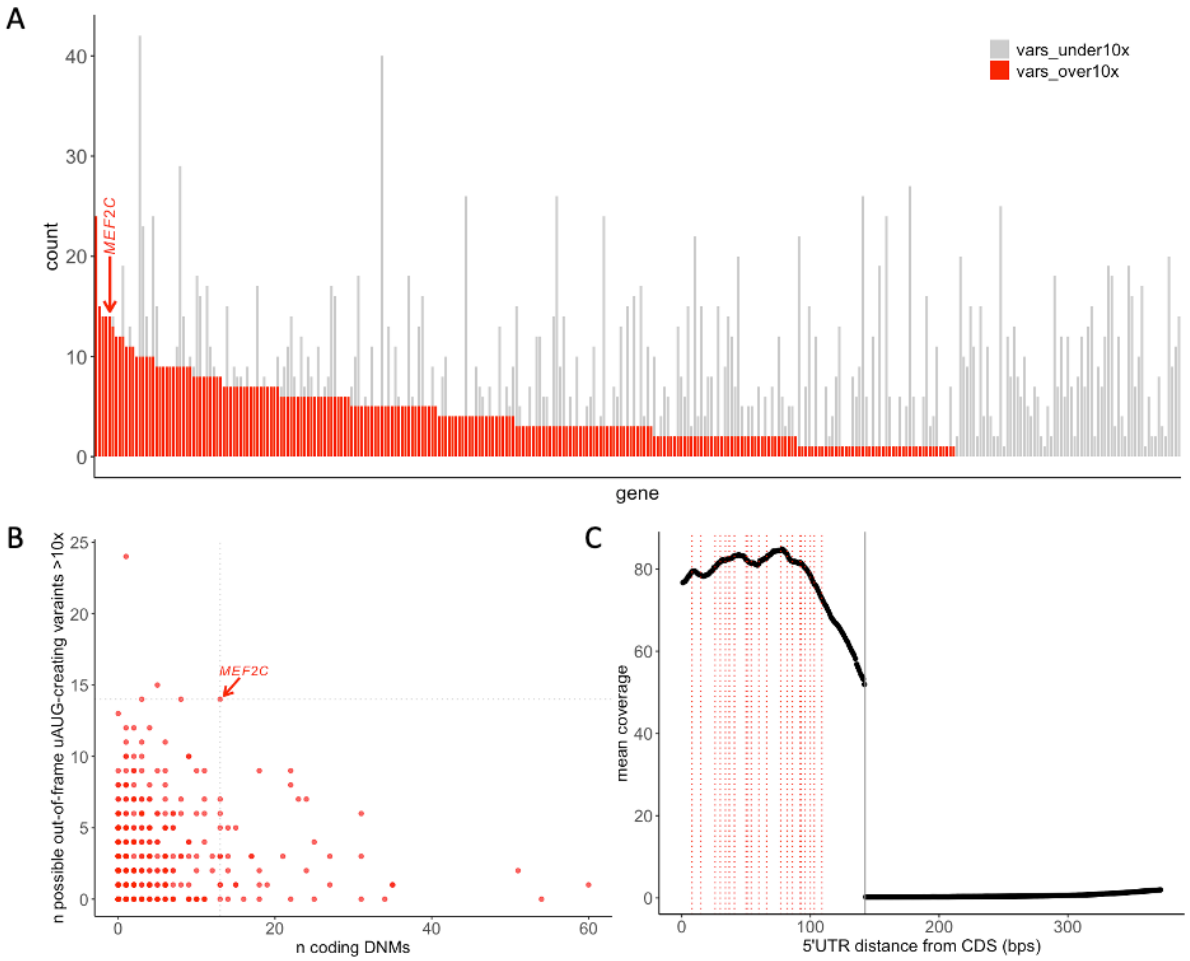


Figure S5: Coverage of uAUG-creating sites of DD haploinsufficient genes and the *MEF2C* 5'UTR. (A) Stacked bar chart showing the count of all possible uAUG-creating variants that would create out-of-frame overlapping ORFs that are covered at mean >10x (red), or ≤10x (grey) per gene. *MEF2C* has a high number of possible variants (n=14), all of which are well covered. (B) The number of well-covered uAUG-creating variants that would create out-of-frame overlapping ORFs plotted against the number of coding missense and protein-truncating *de novo* mutations (DNMs) per gene. *MEF2C* has both a high number of well-covered sites and a high diagnostic yield. (C) The mean coverage across the *MEF2C* 5'UTR. All possible uAUG-creating variants that would create either out-of-frame overlapping ORFs or CDS-elongations are plotted as dotted lines. The 5'UTR exon that is adjacent to the CDS is very well covered (mean >50x).

NB: (A) and (B) do not include CDS-elongating variants as these would not be predicted to cause loss-of-function unless there is an important N-terminal structure or functional domain.


A
vars_under10x
vars_over10x
count
MEF2C
gene
B
n possible out-of-frame uAUG-creating varaints >10x
MEF2C
n coding DNMs
C
mean coverage
5'UTR distance from CDS (bps)

Supplementary Tables

Table S1: List of haploinsufficient developmental disorder genes and their MANEv0.91 transcripts used for analysis.

Table S2: Clinical details for patients with non-coding MEF2C variants.

Table S3: List of missense variants identified in DD cases. ClinVar variants are filtered to only those identified as de novo or with experimental evidence. Protein changes are with respect to the Ensembl canonical transcript ENST00000340208.5.

Table S4: List of gnomAD v2.1.1 missense variants in MEF2 genes used from protein modelling. Protein changes are with respect to the Ensembl canonical transcript ENST00000340208.5.

Table S5: Residues in the structure of MEF2A and their direction with respect to the bound DNA.

Table S6: Comparing the proportion of DD and gnomAD variants that are in contact/pointing towards DNA to those that are distal or pointing away from the DNA-binding interface.

Table S7: Change in Gibbs free energy (ΔΔG) of protein-DNA interaction and complex stability associated with missense variants in MEF2C.

Genomics England Research Consortium

John C. Ambrose[1]; Prabhu Arumugam[1]; Emma L. Baple[1]; Marta Bleda[1]; Freya Boardman-Pretty[1,2]; Jeanne M. Boissiere[1]; Christopher R. Boustred[1]; Helen Brittain[1]; Mark J. Caulfield[1,2]; Georgia C. Chan[1]; Clare E. H. Craig[1]; Louise C. Daugherty[1]; Anna de Burca[1]; Andrew Devereau[1]; Greg Elgar[1,2]; Rebecca E. Foulger[1]; Tom Fowler[1]; Pedro Furió-Tarí[1]; Adam Giess[1]; Joanne M. Hackett[1]; Dina Halai[1]; Angela Hamblin[1]; Shirley Henderson[1,2]; James E. Holman[1]; Tim J. P. Hubbard[1]; Kristina ibáñez[1,2]; Rob Jackson[1]; Louise J. Jones[1,2]; Dalia Kasperaviciute[1]; Melis Kayikci[1]; Athanasios Kousathanas[1]; Lea Lahnstein[1]; Kay Lawson[1]; Sarah E. A. Leigh[1]; Ivonne U. S. Leong[1]; Javier F. Lopez[1]; Fiona Maleady-Crowe[1]; Joanne Mason[1]; Ellen M. McDonagh[1,2]; Loukas Moutsianas[1,2]; Michael Mueller[1,2]; Nirupa Murugaesu[1]; Anna C. Need[1,2]; Peter O'Donovan[1]; Chris A. Odhams[1]; Andrea Orioli[1]; Christine Patch[1,2]; Mariana Buongermino Pereira[1]; Daniel Perez-Gil[1]; Dimitris Polychronopoulos[1]; John Pullinger[1]; Tahrima Rahim[1]; Augusto Rendon[1]; Pablo Riesgo-Ferreiro[1]; Tim Rogers[1]; Mina Ryten[1]; Kevin Savage[1]; Kushmita Sawant[1]; Richard H. Scott[1]; Afshan Siddiq[1]; Alexander Sieghart[1]; Damian Smedley[1,2]; Katherine R. Smith[1,2]; Samuel C. Smith[1]; Alona Sosinsky[1,2]; William Spooner[1]; Helen E. Stevens[1]; Alexander Stuckey[1]; Razvan Sultana[1]; Mélanie Tanguy[1]; Ellen R. A. Thomas[1,2]; Simon R. Thompson[1]; Carolyn Tregidgo[1]; Arianna Tucci[1,2]; Emma Walsh[1]; Sarah A. Watters[1]; Matthew J. Welland[1]; Eleanor Williams[1]; Katarzyna Witkowska[1,2]; Suzanne M. Wood[1,2]; Magdalena Zarowiecki[1].

[1]Genomics England, London, UK
[2]William Harvey Research Institute, Queen Mary University of London, London, EC1M 6BQ, UK.