Determining how MAFA and MAFB transcription factors activity is influenced by structural differences predicted by AlphaFold2

Xin Tong[1,5], Jeeyeon Cha[2,5], Min Guo[1], Jin-hua Liu[1], Garrett Reynolds[1], Zach Loyd[2], Emily M. Walker[3], Scott Soleimanpour[3], Richard A. Stein[1,4], Hassane Mchaourab[1,4] and Roland Stein[1*]

[1]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN, USA

[2]Division of Diabetes, Endocrinology, and Metabolism, Vanderbilt University Medical Center, Nashville, TN, USA

[3]Departments of Molecular & Integrative Physiology and Internal Medicine, Division of Metabolism, Endocrinology & Diabetes, University of Michigan, Ann Arbor, MI

[4]Center of Applied Artificial Intelligence in Protein Dynamics, Vanderbilt University

[5]Co-first authors

*Corresponding author: roland.stein@vanderbilt.edu

Running title: Analysis of MAFA and MAFB differences predicted by AlphaFold2

Key words: transcription factor, MAF, basic-leucine-zipper domain, AlphaFold2, structure-function, cellular senescence, gene transcription, protein chimera

ABSTRACT (193 words)

MAFA and MAFB are related basic-leucine-zipper domain containing transcription factors which have important regulatory roles in a variety of cellular contexts, including pancreatic islet hormone producing α and β cells. These proteins have similar as well as distinct functional properties, and here we first used AlphaFold2, an artificial intelligence-based structural prediction program, to obtain insight into the three-dimensional organization of their non-DNA binding/dimerization sequences. This analysis was conducted on the wildtype (WT) proteins as well the pathogenic $MAFA^{Ser64Phe}$ ($MAFA^{S64F}$) and $MAFB^{Ser70Ala}$ ($MAFB^{S70A}$) mutants, with structural differences revealed between $MAFA^{WT}$ and $MAFB^{WT}$ in addition to $MAFA^{S64F}$ and $MAFA^{WT}$, but not $MAFB^{S70A}$ and $MAFB^{WT}$. Functional analysis disclosed that the inability to properly phosphorylate at S70 in $MAFB^{S70A}$, like S65 in $MAFA^{S64F}$, greatly increased protein stability and enabled $MAFB^{S70A}$ to accelerate cellular senescence in cultured cells. Significant differences were also observed in the ability of MAFA, $MAFA^{S64F}$, MAFB, and $MAFB^{S70A}$ to cooperatively stimulate *Insulin* enhancer-driven activity in the presence of other islet-enriched transcription factors. Experiments performed on protein chimeras disclosed that these properties were greatly influenced by structural differences found between the WT and mutant proteins. In general, these results revealed that AlphaFold2 predicts features essential to protein activity.

INTRODUCTION

The musculoaponeurotic fibrosarcoma (MAF) family consists of proto-oncogenes with important cellular regulatory properties that are categorized into two distinct subgroups: small MAFs that lack an N-terminal activation domain (~150-160 amino acids: MAFF, MAFG, and MAFK) and large MAFs (~240-340 amino acids: MAFA, MAFB, c-MAF, and NRL) (1, 2). These MAF proteins act as transcription factors (TFs) and play active roles in many tissue types, such as the pancreas (3, 4), lens (4), and cartilage (5) by contributing to their development, differentiation and establishment of mature adult tissue-specific activities.

The large MAFA and MAFB proteins have been found to be vital in the development and/or maturation of pancreatic islet insulin hormone-producing β- and glucagon hormone-producing α-cells (1, 2). Expression of these proteins is affected in type 2 diabetes, a disease arising from a relative insulin deficiency and concomitant glucose under-utilization, which is compounded by glucagon excess driving glucose over-production (6). Moreover, mutations near or on a priming kinase site in both MAFA and MAFB can cause clinical disease. Thus, the S64F substitution prevents S65 phosphorylation of MAFA and clinically produces either monogenic diabetes or insulinomatosis (7), while S70A in MAFB blocks S70 phosphorylation and causes multicentric carpotarsal osteolysis (i.e., MCTO), a disorder characterized by osteolysis of the carpal and tarsal bones, subtle craniofacial deformities, and nephropathy (8). While MAFB appears to be critical to human insulin (and likely glucagon) production (9), it is unclear if $MAFB^{S70A}$ causes diabetes in MCTO patients, likely due to the early demise of affected patients.

The crystal structure of only the internal basic (b) and leucine-zipper (LZ) regions of the MAFA (10) and MAFB (11) proteins, and not the N- or C-terminal regions has been determined. The bLZ region is responsible for DNA binding and homo- or hetero-dimerization between the large MAF proteins (12). Here we have used the AlphaFold2 algorithm, which employs deep iterative learning to provide predictions of the three-dimensional architecture of a protein solely

based on multiple sequence alignments (13), to obtain structural information on the MAFA and MAFB proteins. AlphaFold2 functional and molecular assays were performed to evaluate the relevance of the predicted differences to the activity of each protein. Our results strongly suggest that the structural dissimilarities between MAFA and MAFB impart unique functional properties.

RESULTS

1) Alphafold2 predicts structurally unique regions within the MAFA, MAFB, and MAFA^{S64F} proteins.

MAFA and MAFB act as homo- or heterodimers under physiological conditions and have shared and distinct functional properties (14). The level of sequence identity between these proteins are variable between domains: the N-terminal transactivation domain (51%), histidine-rich region (29%), basic (b) DNA binding domain (76%), LZ dimerization region (63%), and C-terminal tail (8%) (1) (Figure 1A). Presently, structural information from crystallography is limited to the bLZ portion of the MAFA (i.e., residues 226–318) and MAFB (211–305) (10, 11) with a high content of intrinsically disordered regions predicted within much of the remainder of these proteins.

Here we employed AlphaFold2 for *de novo* modeling of homodimers of full-length MAFA and MAFB. The analysis was performed on unphosphorylated MAFA (MAFA^{un}) and MAFB (MAFB^{un}), fully phosphorylated MAFA (MAFA^{WT}) and MAFB (MAFB^{WT}), as well as partially phosphorylated mutant MAFA (MAFA^{S64F}) and MAFB (MAFB^{S70A}). For the purpose of modeling, aspartic (D) and glutamic (E) acid were used as phosphomimetics at serine and threonine residues, respectively (Figure 1B). Notably, MAFA^{un} (and presumably MAFB^{un}) is unable to bind DNA, migrates as a much larger molecular weight species than detected *in vivo*, and is not believed to be a biologically relevant entity (15). However, they were modeled as a reference in order to examine the effects of the phosphomimetic residues on the modeled structures. MAFA^{S64F} and MAFB^{S70A} have a reduced phosphorylation state relative to the wildtype protein due to the inability of a yet unknown priming kinase to act at S65 in MAFA^{S64F} or S70 in MAFB^{S70A} (Supp Figure 1A), which consequently prevents GSK3 from phosphorylating neighboring serine and threonine residues (16).

Using a locally installed version of AlphaFold2, it was possible to obtain five intrinsic structural models of the $MAFA^{WT}$ and $MAFB^{WT}$ via multiple runs of the algorithm. This allows for a more robust exploration of conformational ensembles. The models for each of the sequences yield dimers with the expected folds within the bLZ-domains, while the transactivation domain had no specific conformation (Figure 1D). The intrinsic disorder of the transactivation domain coincides with the available structure of the MAFA or MAFB bLZ domains. Of note are differences in the non-conserved carboxy-terminal tail between MAFA and MAFB as well as the highly variable loops comprising the histidine-rich regions and the center of the transactivation domain. The amino- and carboxy-termini of the transactivation domain in the models display some secondary structure. Although there is no consensus on their placement, they invariably lie near the DNA binding region of the bLZ domain.

Comparison of $MAFA^{WT}$ to $MAFA^{un}$ or $MAFB^{WT}$ to $MAFB^{un}$ indicates that the phosphomimetics result in a change of the AlphaFold2 predicted models (Supp Figure 1B, C). Moreover, reducing the extent of phosphorylation in $MAFA^{S64F}$ in comparison to $MAFA^{WT}$ leads to further alterations in the placement of the transactivation domain relative to the leucine zipper (Supp Figure 1B, D), which is depicted upon measuring the distance of each amino acid residue to a fixed point in space in the proteins (Supp Figure 1D). The pattern is slightly different in MAFB as the reduction in $MAFB^{S70A}$ phosphorylation does not lead to a change in the location of the transactivation domain (Supp Figure 1C, D). Comparison of $MAFA^{WT}$ to $MAFB^{WT}$ indicates small differences in the regions proximal to the LZ domain and more substantial differences in the unstructured transactivation domains. However, the most strikingly change between $MAFA^{WT}$ and $MAFA^{S64F}$ or $MAFB^{WT}$ and $MAFB^{S70A}$ is in their C-terminal region (Figure 1D and Supp Figure 1B,C), which is much more disordered between $MAFA^{WT}$ and $MAFA^{S64F}$ and, in some models, is found in close proximity to the transactivation domain. The experiments performed below were

focused on determining how these possible structural differences relate to MAFA and MAFB activity.

2) Both MAFA^{S64F} and MAFB^{S70A} cause increased protein stability and cellular senescence.

Mutation in the priming kinase site that prevents phosphorylation at S65 in MAFA (i.e., MAFA^{S64F}(7, 16) and S70 in MAFB (i.e., MAFB^{S70A} (8) of the transactivation domain are pathogenic. Mechanistic studies have demonstrated that MAFA^{S64F} significantly increases protein stability (7), which in male mice results in islet β cell inactivity due the premature induction of senescence and aging (17). These same properties with tested in MAFB^{S70A} presuming that modeled structural differences between MAFA^{WT} and MAFA^{S64F} contributed to these functional changes.

MAFB^{S70A} was significantly more stable than MAFB^{WT} (Supp Figure 2), with changes comparable to MAFA^{S64F} (7). In addition, MAFB^{S70A} induced cellular senescence in the human EndoC-βH3 islet β cell line. Hence, key markers such as endogenous senescence-associated β-galactosidase (SA-β-gal) staining was principally increased in lentiviral-driven MAFA^{S64F} and MAFB^{S70A} in relation to their WT counterparts (Figure 2A). Moreover, expression of other senescence-associated markers like *CDKN1A* (encoding p21), *BCL2*, *ICAM3*, and *IGFBP4*, which were identified in MAFA^{S64F}-producing EndoC β cells (17), were co-induced by these mutants (Figure 2B).

Another property of MAFA^{S64F}-driven senescence includes the development of senescence-associated secretory phenotype (SASP) factors, which upon secretion bring about senescence by reprogramming neighboring cells in a cell non-autonomous manner (17). To determine whether SASP factors were released by human cells expressing MAFB^{S70A}, control medium or conditioned media collected from EndoC-βH3 cells expressing either the WT or priming kinase mutant MAF proteins was applied to untransduced cells (Figure 2C). Significantly,

most but not all of the SASP genes elevated in MAFA^{S64F} conditioned media were also produced by MAFB^{S70A}, and not upon treatment with wildtype conditioned medium (Figure 2D). Collectively these results suggest that only the AlphaFold2 predicted structure for the transactivation domain of MAFA^{S64F}, and not MAFB^{S70A}, is pertinent to activity.

3) Wildtype and mutant MAFA and MAFB differ in their ability to functionally interact with other islet-enriched TFs

MAFA and MAFB, like many other TFs important in cell development and adult function, act in concert by binding and regulating *cis*-acting gene enhancer activity (18). These proteins bind to closely related *cis*-acting element sequences (Figure 1C) and are important in *Insulin* gene transcription in mice and humans (2), with, for example, MAFB absolutely required for *Insulin* expression in human embryonic stem cell derived β-like cells (9). To determine if MAFA^{WT} and MAFB^{WT} differentially impact concerted activation by other islet-enriched TFs of the *Insulin* gene, these proteins were co-transfected with PDX1, NEUROD1, ISL1, and GLIS3 and analyzed for their ability to increase *Insulin*-driven -238 WT:luciferase reporter in HeLa cells.

Notably, only MAFA^{WT} and MAFB^{WT} were able to independently stimulate *Insulin*-reporter activity (Figure 3). MAFB, but not MAFA, provided additional enhanced activation in the presence of PDX1 and ISL1, while both MAFA and MAFB function cooperatively with GLIS3. In contrast, NEUROD1 did not elevate MAFA^{WT} or MAFB^{WT} activity, although expressed effectively (Supp Figure 3). Differences were also observed in the ability of MAFA^{S64F} and MAFB^{S70A} to functionally interact with GLIS3, as MAFB^{S70A} provided additional stimulation over MAFB^{WT}, whereas this enhancement was lost in MAFA^{S64F} (Figure 4). Consequently, these results suggest that the structural differences between wildtype and mutant MAFA and MAFB impact their ability to function in conjunction with other islet TFs.

4) Analyzing how the structurally unique regions of MAFA and MAFB affect ISL1 and PDX1 activation.

To further investigate differences in $MAFA^{WT}$ and $MAFB^{WT}$ activation, chimeric proteins between MAFA and MAFB were generated to evaluate how their N- and C-terminal regions influenced ISL1 and PDX1 activation of the -238 WT:luciferase reporter. Importantly, the chimeric proteins retained the unique three-dimensional properties predicted in the wildtype protein by AlphaFold2 (Figure 5A and Supp Figure 4). Interestingly, ISL1 stimulation was lost in $MAFB^{WT}$ upon exchanging either $MAFA^{WT}$ or $MAFA^{S64F}$ N- (i.e., termed MAFA/B wherein amino acids 1-208 of MAFA were linked to residues 189-223 of $MAFB^{WT}$) or C-terminal sequences (i.e., MAFB/A) (Figure 5B). These results indicated that the structural integrity of both the N- and C-terminal regions of MAFB were necessary for ISL1 stimulation.

In contrast to ISL1, $MAFB^{WT}$ plus PDX1 activity was retained in N-terminal fusions with $MAFA^{WT}$ (Figure 6). Astoundingly, the $MAFA/B^{T}$ chimera that merely retained 35 C-terminal amino acids of $MAFB^{WT}$ also worked effectively with PDX1. Conversely, PDX1 plus MAFA/B activity was lost in $MAFA^{S64F}$/B. Noticeably, stimulation with PDX1 was not observed in MAFB/A and even reduced by $MAFB^{S70A}$/A. Collectively, these results imply that the distinct structural features predicted of the novel C-terminal sequences of MAFB mediate cooperativity between PDX1 and MAFA or MAFB N-terminal transactivation domain activity.

DISCUSSION

MAFA and MAFB play crucial roles in a variety of cell types (14). In this study, we used the deep learning AlphaFold2 algorithm to predict the three-dimensional architecture of the disordered regions within the non-DNA binding/dimerization regions of these proteins, and then analyzed their potential functional impact experimentally. While a structural change in $MAFB^{S70A}$ was not predicted in relation to $MAFB^{WT}$, like $MAFA^{S64F}$, it manifested increased protein stability and induced senescence in a cultured human β cell line. This phenotypic difference is likely associated with MCTO pathogenesis (8). AlphaFold2 did predict differences in the N-terminal and C-terminal disordered regions of $MAFA^{WT}$, $MAFA^{S64F}$ and $MAFB^{WT}$ that were found to regulate activation by the *Insulin* gene PDX1, ISL1, and GLIS3 TFs. Overall, we believe that the newly defined structural regions identified by AlphaFold2 are essential to MAFA and MAFB activity.

The C-terminal tail of MAFB had the most striking impact experimentally, as addition of these sequences to $MAFA^{WT}$ now imparted cooperative activation with PDX1 on the *Insulin* -238 WT:luciferase reporter, while compromising basal $MAFA^{WT}$ transactivation activity (i.e., MAFA/B and $MAFA/B^{T}$ in Figure 6). Earlier studies had also revealed novel regulatory properties for the MAFA C-terminal tail (19). Hence, ectopic expression of MAFA, but not MAFB, promoted $Insulin^{+}$ cell formation within embryonic endoderm in chick *in ovo* electroporation assays. Moreover, only the chimeric MAFB/A protein induced chick *Insulin* activation and not a MAFA/B fusion (19). The inability of MAFB to stimulate *Insulin* expression in the *in ovo* electroporation assay implies that cooperating factors mediating activation during mouse (19) and human (9) pancreatic development are lacking in chick endoderm.

The analysis of PDX1, E47, NEUROD1, ISL1, and GLIS3 on $MAFA^{WT}$ and $MAFB^{WT}$ activity was designed to glean insight into whether their structurally distinct disordered regions predicted by Alphafold2 were regulatory. Indeed, differences in functional interactions with PDX1, ISL1 and GLIS3 were revealed. However, it is unclear if this regulation occurs *in vivo* since the experiments

conducted here were in a cell line that does not express *Insulin* or the analyzed TFs (i.e., HeLa cells). Yet we do propose that collaboration between TFs like those analyzed here with MAFA or MAFB are important in regulating controlled enhancers *in vivo*, many of which will be in common and some distinct. Our results demonstrated that such regulation entails recruitment of shared (e.g., GLIS3) and dissimilar (e.g., ISL1 by MAFB) TFs.

It is noteworthy that MAFA and MAFB are expressed in a distinct manner in mouse and human islet β cells, with only MAFA produced in mouse β cells and MAFA and MAFB in human β cells (20). In fact, mis-expression of MAFB is unable to rescue the many deficiencies associated with mice lacking MAFA in β cells (21). These TFs are not often co-expressed in human tissues, the exception being testes, islets, and skeletal muscle (Supp Figure 5). In these unusual circumstances, MAFA and MAFB presumably act in a heterodimer complex, which has been linked with the most active human islet β cell population (22). Accordingly, the subpopulations of islet β cells expressing both MAFA and MAFB exhibit greater expression of key genes related to glucose sensing and hormone secretion as well as enhanced electrophysiological activity relative to subpopulations expressing only one or neither TF. Future efforts should be aimed at directly determining the structural and functional impact of the MAFA/MAFB heterodimer in this context.

EXPERIMENTAL PROCEDURES

Modeling by AlphaFold2

Multiple sequence alignments were carried out with MMSeqs2 (23) and AlphaFold2 models were generated with ColabFold (24, 25) using model parameters version 2.3. To allow for a more robust exploration of conformational space each alignment was run with 5 seeds leading to 25 models. To examine the effect of phosphorylation on the generated models, we used phosphomimetics based on previously identified phosphorylation sites in MAFA (15). Specifically, serine residues were replaced with aspartic acid, and threonine residues were replaced with glutamic acid (Figure 1B). These substitutions were made across the multiple sequence alignments excluding gaps. All the phosphorylation sites in $MAFA^{WT}$ and their homologous sites in $MAFB^{WT}$ were replaced, whereas in conditions that mimic partial phosphorylation, all the phosphorylation sites were replaced except for the residues 49, 53, 57, 61, and 65 in $MAFA^{S64F}$ or residues 54, 58, 62, and 66 in $MAFB^{S70A}$. The highest predicted interface for each of the modelers is shown. To quantitate the variability in the AlphaFold2 models, distances from Cα for each residue to a fixed point in space were measured. The fixed point was the centroid between R281 in MafA, MafA/B, and $MafA/B^{T}$, R259 in MafB and R264 in MafB/A. The distances were averaged for each set of 25 models.

DNA constructs

All of the cytomegalovirus enhancer-driven human MAFA and MAFB expression constructs used in the transfections and the lentiviral vectors were generated by Vector Builder Inc. (Chicago, IL) as were the lentiviral vectors. DNA sequencing analysis through Genewiz (South Plainfield, NJ) confirmed the fidelity of each construct.

Cell culture and assays

HeLa monolayer cells were maintained in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) supplemented with 10% heat deactivated fetal bovine serum. Cytomegalovirus enhancer-driven MAFA, MAFB, PDX1, GLIS3, ISL1, and NEUROD1 were co-transfected with the rat II *insulin* enhancer -238 WT:firefly luciferase plasmid (2), and the phRL-TK Renilla luciferase internal control plasmid (Promega). Reporter activity was evaluated 48 hours after transfection using a dual-luciferase assay according to the manufacturer's protocol (Promega). Firefly luciferase measurements were normalized to the phRL-TK activity. EndoC-βH3 were grown in DMEM containing 5.6mM glucose, 2% BSA, 50 mM 2-mercaptoethanol, 10mM nicotinamide, 5.5 mg/mL transferrin, 6.7 ng/mL selenite, 100 units/mL penicillin, and 100 units/mL streptomycin (26). The cells were incubated with lentiviral particles (50ng per million cells) producing $MAFA^{WT}$, $MAFA^{S64F}$, $MAFB^{WT}$ or $MAFB^{S70A}$ for one day after plating and then analyzed one week following infection for SA-β-gal staining and expression of various $MAFA^{S64F}$ senescence markers by qPCR as described previously (17). Each experiment was repeated at least three times using independently isolated plasmid preparations.

ACKNOWLEDGEMENTS

We thank Drs. Raphaël Scharfmann and Phillippe Ravassard for generating and providing the original EndoC-βH3 cells.

DATA AVAILABILITY

All data contained within the manuscript are available upon reasonable request to R.S. (roland.stein@vanderbilt.edu)

AUTHOR CONTRIBUTIONS

X.T., J.C., R.S. and H.M. contribute to the conceptualization of this study. G.R., M.G., J.L., Z.L., R.A.S. and J.C. executed data curation and formal analysis with input from X.T. and R.S. The AlphaFold2 data visualization was accomplished by R.A.S. The original draft and figures were prepared by X.T. The review and editing of the manuscript were done by H.M. and R.S. The resources (EndoC-βH3) were maintained and provided by E.W. and S.S.

FUNDING AND ADDITIONAL INFORMATION

This research was performed using resources and/or funding provided by NIH grants to R.S. (R01 DK090570, R01 DK126482), J.C. (K08 DK132507), E.M.W. (F32 DK109577), S.S. (R01 DK108921, R01 DK135268, R01 DK135032, R01 DK136671, U01 DK127747), and to the Vanderbilt Diabetes Research and Training Center (DK20593). J.C. was also supported by a Doris Duke Charitable Foundation Physician Scientist Fellowship (2020063) and Burroughs Wellcome Fund Career Award for Medical Scientists, and X.T. was supported by a JDRF Fellowship (3-PDF-2019-738-A-N) and by the Diabetes Research Connection (2022). Imaging was performed with NIH support from the Vanderbilt University Medical Center Cell Imaging Shared Resource (NCI grant CA68485; NIDDK grants DK20593, DK58404, and DK59637; NICHD grant HD15052; and NEI grant EY08126).

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest with the contents of this article.

REFERENCES

1. Conrad,E., Dai,C., Spaeth,J., Guo,M., Cyphert,H.A., Scoville,D., Carroll,J., Yu,W.-M., Goodrich,L. V, Harlan,D.M., *et al.* (2016) The MAFB transcription factor impacts islet α-cell function in rodents and represents a unique signature of primate islet β-cells. *Am J Physiol Endocrinol Metab*, **310**, E91–E102.

2. Artner,I., Hang,Y., Mazur,M., Yamamoto,T., Guo,M., Lindner,J., Magnuson,M.A. and Stein,R. (2010) MafA and MafB regulate genes critical to beta-cells in a unique temporal manner. *Diabetes*, **59**, 2530–2539.

3. Matsuoka,T., Zhao,L., Artner,I., Jarrett,H.W., Friedman,D., Means,A. and Stein,R. (2003) Members of the large Maf transcription family regulate insulin gene transcription in islet beta cells. *Mol Cell Biol*, **23**, 6049–6062.

4. Sakai,M., Imaki,J., Yoshida,K., Ogata,A., Matsushima-Hibiya,Y., Kuboki,Y., Nishizawa,M. and Nishi,S. (1997) Rat maf related genes: specific expression in chondrocytes, lens and spinal cord. *Oncogene*, **14**, 745–750.

5. MacLean,H.E., Kim,J.I., Glimcher,M.J., Wang,J., Kronenberg,H.M. and Glimcher,L.H. (2003) Absence of transcription factor c-maf causes abnormal terminal differentiation of hypertrophic chondrocytes during endochondral bone development. *Dev Biol*, **262**, 51–63.

6. Guo,S., Dai,C., Guo,M., Taylor,B., Harmon,J.S., Sander,M., Robertson,R.P., Powers,A.C. and Stein,R. (2013) Inactivation of specific β cell transcription factors in type 2 diabetes. *J Clin Invest*, **123**, 3305–3316.

7. Iacovazzo,D., Flanagan,S.E., Walker,E., Quezado,R., De Sousa Barros,F.A., Caswell,R., Johnson,M.B., Wakeling,M., Brändle,M., Guo,M., *et al.* (2018) MAFA missense mutation

causes familial insulinomatosis and diabetes mellitus. *Proc Natl Acad Sci U S A*, **115**, 1027–1032.

8. Ma,N.S., Mumm,S., Takahashi,S. and Levine,M.A. (2023) Multicentric Carpotarsal Osteolysis: a Contemporary Perspective on the Unique Skeletal Phenotype. *Curr Osteoporos Rep*, **21**, 85–94.

9. Russell,R., Carnese,P.P., Hennings,T.G., Walker,E.M., Russ,H.A., Liu,J.S., Giacometti,S., Stein,R. and Hebrok,M. (2020) Loss of the transcription factor MAFB limits β-cell derivation from human PSCs. *Nature Communications 2020 11:1*, **11**, 1–15.

10. Lu,X., Guanga,G.P., Wan,C. and Rose,R.B. (2012) A novel DNA binding mechanism for maf basic region-leucine zipper factors inferred from a MafA-DNA complex structure and binding specificities. *Biochemistry*, **51**, 9706–9717.

11. Pogenberg,V., Consani Textor,L., Vanhille,L., Holton,S.J., Sieweke,M.H. and Wilmanns,M. (2014) Design of a bZip transcription factor with homo/heterodimer-induced DNA-binding preference. *Structure*, **22**, 466–477.

12. Blank,V. and Andrews,N.C. (1997) The Maf transcription factors: regulators of differentiation. *Trends Biochem Sci*, **22**, 437–441.

13. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

14. Tsuchiya,M., Misaka,R., Nitta,K. and Tsuchiya,K. (2015) Transcriptional factors, Mafs and their biological roles. *World J Diabetes*, **6**, 175.

15. Guo,S., Burnette,R., Zhao,L., Vanderford,N.L., Poitout,V., Hagman,D.K., Henderson,E., Ozcan,S., Wadzinski,B.E. and Stein,R. (2009) The stability and transactivation potential of

the mammalian MafA transcription factor are regulated by serine 65 phosphorylation. *J Biol Chem*, **284**, 759–765.

16. Rocques,N., Abou Zeid,N., Sii-Felice,K., Lecoin,L., Felder-Schmittbuhl,M.-P., Eychène,A. and Pouponnot,C. (2007) GSK-3-mediated phosphorylation enhances Maf-transforming activity. *Mol Cell*, **28**, 584–597.

17. Walker,E.M., Cha,J., Tong,X., Guo,M., Liu,J.-H., Yu,S., Iacovazzo,D., Mauvais-Jarvis,F., Flanagan,S.E. and Korbonits,M. (2021) Sex-biased islet β cell dysfunction is caused by the MODY MAFA S64F variant by inducing premature aging and senescence in males. *Cell Rep*, **37**, 109813.

18. Scoville,D.W., Cyphert,H.A., Liao,L., Xu,J., Reynolds,A., Guo,S. and Stein,R. (2015) MLL3 and MLL4 Methyltransferases Bind to the MAFA and MAFB Transcription Factors to Regulate Islet β-Cell Function. *Diabetes*, **64**, 3772–3783.

19. Artner,I., Hang,Y., Guo,M., Gu,G. and Stein,R. (2008) MafA is a dedicated activator of the insulin gene in vivo. *J Endocrinol*, **198**, 271.

20. Hang,Y. and Stein,R. (2011) MafA and MafB activity in pancreatic β cells. *Trends in Endocrinology & Metabolism*, **22**, 364–373.

21. Cyphert,H.A., Walker,E.M., Hang,Y., Dhawan,S., Haliyur,R., Bonatakis,L., Avrahami,D., Brissova,M., Kaestner,K.H., Bhushan,A., *et al.* (2019) Examining How the MAFB Transcription Factor Affects Islet β-Cell Function Postnatally. *Diabetes*, **68**, 337–348.

22. Shrestha,S., Saunders,D.C., Walker,J.T., Camunas-Soler,J., Dai,X.Q., Haliyur,R., Aramandla,R., Poffenberger,G., Prasad,N., Bottino,R., *et al.* (2021) Combinatorial transcription factor profiles predict mature and functional human islet α and β cells. *JCI Insight*, **6**.

23. Mirdita,M., Steinegger,M. and Söding,J. (2019) MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics*, **35**, 2856–2858.

24. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873*, **596**, 583–589.

25. Mirdita,M., Schütze,K., Moriwaki,Y., Heo,L., Ovchinnikov,S. and Steinegger,M. (2022) ColabFold: making protein folding accessible to all. *Nature Methods 2022 19:6*, **19**, 679–682.

26. Scharfmann,R., Pechberty,S., Hazhouz,Y., von Bülow,M., Bricout-Neveu,E., Grenier-Godard,M., Guez,F., Rachdi,L., Lohmann,M., Czernichow,P., *et al.* (2014) Development of a conditionally immortalized human pancreatic β cell line. *J Clin Invest*, **124**, 2087–2098.

27. Xie,X., Rigor,P. and Baldi,P. (2009) MotifMap: a human genome-wide map of candidate regulatory motif sites. *Bioinformatics*, **25**, 167–174.

28. Lonsdale,J., Thomas,J., Salvatore,M., Phillips,R., Lo,E., Shad,S., Hasz,R., Walters,G., Garcia,F., Young,N., *et al.* (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet*, **45**, 580.

29. Alonso,L., Piron,A., Morán,I., Guindo-Martínez,M., Bonàs-Guarch,S., Atla,G., Miguel-Escalada,I., Royo,R., Puiggròs,M., Garcia-Hurtado,X., *et al.* (2021) TIGER: The gene expression regulatory variation landscape of human pancreatic islets. *Cell Rep*, **37**.

ABBREVIATIONS

bLZ: basic leucine-zipper

DMEM: Dulbecco's modified Eagle's medium

E47: E2A immunoglobulin enhancer-binding factors, also known as TCF3

GLIS3: Transcription factor Gli-similar 3

GSK: Glycogen-synthase kinase-3

ISL1: Islet1

MAF: musculoaponeurotic fibrosarcoma

MCTO: multicentric carpotarsal osteolysis

NEUROD1: Neurogenic differentiation 1

PDX1: pancreatic and duodenal homeobox 1

SASP: senescence-associated secretory phenotype

SA-β-gal: senescence-associated β-galactosidase

TF: transcription factor

WT: wildtype

FIGURE LEGENDS

Figure 1: AlphaFold2 modeling of MAFA^{WT} and MAFB^{WT} protein. A) Protein sequence identity (%) between various domains of MAFA^{WT} and MAFB^{WT}, specifically the transactivation (grey), histidine-rich (brown), basic (blue), LZ (green), and non-conserved C-terminal tail (yellow). B) The aligned amino acid sequences of MAFA^{WT} and MAFB^{WT}, with aspartic acid (D) and glutamic acid (E) used as phosphomimetics at *bona fide* sites of phosphorylation (15). The purple labelled E and D represent the phosphosites absent in MAFA^{S64F} and MAFB^{S70A}. The two regions marked in blue (i.e., 83%) and magenta (i.e., 79%) have higher sequence conservation than the overall transactivation domain and have structural elements in the AlphaFold2 models. The intervening sequence is colored red. The region in green is the bLZ domain that is structured in the AlphaFold2 models and determined by crystallography (i.e., MAFA (i.e., residues 226–318) and MAFB (211–305) (10, 11). C) The DNA binding sequence motif of MAFA and MAFB (27) as determined by ChIP-seq databases. D) Comparing the five independently derived models for MAFA^{WT} (left) and MAFB^{WT} (right) illustrates potential structural differences. For example, the longer C-terminal tail of MAFA leads to more divergent conformations that can interact with the N-terminus. The domain coloring refers to the sequences described in A): N- to C-terminal regions of the transactivation domain are blue, red (i.e., less conserved) and magenta while the bLZ domain is green and the poorly conserved histidine-rich and C-terminal sequences in grey.

Figure 2. MAFB^{S70A}, like MAFA^{S64F}, elicits a senescence response in EndoC-βH3 cells. A) SA-β-gal staining on EndoC-βH3 cells transduced to express MAFB^{WT}, MAFB^{S70A}, and MAFA^{S64F}. Scale bar, 100 μm. B) Human expression of β cell-enriched proteins and senescence associated proteins identified in MAFA^{S64F} expressing β cells (17). Two-way ANOVA, *$p<0.05$. C) Conditioned media (CM) from MAFB^{WT}-, MAFB^{S70A}- or MAFA^{S64F}-expressing EndoC-βH3 cells was collected, purified and added to naive EndoC-βH3 cells cultured for 72 hr. Both unique and

overlapping senescence-associated secretory factors were identified in $MAFB^{S70A}$ CM-treated cells. Two-way ANOVA; *p<0.05.

Figure 3. Comparing the ability of $MAFA^{WT}$ and $MAFB^{WT}$ to act with other with islet-enriched TFs to stimulate *insulin*-driven reporter activity. Dual luciferase reporter assays were performed in HeLa cells transfected with vectors producing $MAFA^{WT}$ (Red bars), $MAFB^{WT}$ (Blue bars), ISL1, NEUROD1, PDX1, and/or GLIS3. The fold-change in activity of the -238 WT:luciferase reporter over the no TF transfected controls was shown. Two-way ANOVA with Dunnett's multiple comparisons test against MAFA or MAFB was performed. All error bars indicate SEM. ****P*<0.001; ***P*<0.01. n=3-4.

Figure 4. GLIS3 stimulation with $MAFB^{S70A}$ is greater than with $MAFB^{WT}$, while $MAFA^{S64F}$ is reduced in relation to $MAFA^{WT}$. Dual luciferase reporter assays were performed in HeLa cells transfected with vectors that express $MAFA^{WT}$, $MAFB^{WT}$, $MAFA^{S64F}$, $MAFB^{S70A}$ with or without GLIS3. The fold-change in activity of the -238 WT:luciferase reporter over the no TF transfected control was shown. Two-way ANOVA with Turkey's multiple comparisons test was performed. All error bars indicate SEM. ***P<0.001, **P<0.01; *P<0.05. n=4.

Figure 5. ISL1 does not activate the MAFA/B and MAFB/A chimeras. A) Top: The regions of the MAFA and MAFB chimeric proteins. Bottom: The models for the MAFA/B (left), $MAFA/B^{T}$ (center), and MAFB/A (right). The coloring was explained in the legend to Figure 1: transactivation domain (N-to C-terminal blue, red (i.e., less conserved) and magenta), bLZ domain (green), and the poorly conserved histidine-rich and C-terminal regions (grey). B) Dual luciferase reporter assays were performed in HeLa cells transfected with vectors that express $MAFA^{WT}$, $MAFB^{WT}$,

MAFA/B, or MAFB/A, and/or ISL1. The fold-change in activity of the -238 WT:luciferase reporter over the no TF transfected control was shown. Two-way ANOVA with Turkey's multiple comparisons test was performed. All error bars indicate SEM. ***P<0.001, **P<0.01; *P<0.05. n=4.

Figure 6. MAFB+PDX1 activation was mediated by MAFB C-terminal sequences. Dual luciferase reporter assays were performed in HeLa cells transfected with vectors that express $MAFA^{WT}$, $MAFB^{WT}$, MAFA/B, MAFB/A, $MAFA/B^{T}$, $MAFA^{S64F}$/B, and $MAFB^{S70A}$/A with or without PDX1. The fold-change in activity of the -238 WT:luciferase reporter over the no TF transfected control was shown. The relative fold changes of +PDX1 over -PDX1 were also shown. Two-way ANOVA with Turkey's multiple comparisons test was performed. All error bars indicate SEM. ***P<0.001, *P<0.05. n=4.

Supplemental Figure 1: AlphaFold2 modeling of MAFA and MAFB proteins. A) Priming kinase mutants at $MAFA^{S64F}$ or $MAFB^{S70A}$ prevent GSK3 mediated phosphorylation (Rocques et al., 2007). Schematic adapted from Iacovazzo et al., 2018. B) The structure of the five models developed by AlphaFold2 for $MAFA^{un}$ (left), $MAFA^{WT}$ (center), and $MAFA^{S64F}$ (right). The coloring was as described in Figure 1: transactivation domain (N-to C-terminal blue, red (i.e. less conserved) and magenta), bLZ domain (green), and the poorly conserved histidine-rich and C-terminal regions (grey). C) Plots of the mean distance to a fixed point in space for $MAFA^{un}$, $MAFA^{WT}$, and $MAFA^{S64F}$. The fixed point is the centroid between arginine (R) 281 in MAFA and R259 in MAFB.

Supplementary Figure 2. MAFB^{S70A} increases protein stability. HeLa cells transfected with MAFB^{WT} and MAFB^{S70A} and then incubated with 25 µg/mL cycloheximide for the indicated times. Transfected MAFB and endogenous β-actin protein levels were determined by immunoblotting using anti-MAFB and anti-β-actin antibodies, respectively. A representative blot was shown. n=3.

Supplementary Figure 3. NeuroD1 production in HeLa transfected cells. Transfected NEUROD1 protein levels were determined by immunoblotting using anti-NEUROD1 antibodies. A representative blot was shown. n=3.

Supplemental Figure 4: AlphaFold2 modeling of the WT and mutant MAFA and MAFB chimeras. A) MAFA^{WT}, MAFA^{WT}/B, MAFA^{S64F}/B, and MAFA^{WT}/B^{T}. B) MAFB^{WT}, MAFB^{WT}/A, and MAFB^{S70F}/A. The coloring of the models in A) and B) was described in the legend to Figure 1: transactivation domain (N-to C-terminal blue, red (i.e. less conserved) and magenta), bLZ domain (green), and the poorly conserved histidine-rich and C-terminal regions (grey). C) Plots of the mean distance to a fixed point in space for each set of models. The fixed point is the centroid between R281 in MAFA and MAFA/B chimeras, R259 for MAFB, and R264 for the MAFB/A chimeras.

Supplementary Figure 5. MAFA and MAFB are only co-expressed in testis, skeletal muscle and pancreatic islets in humans. The mRNA expression levels of MAFA and MAFB across different human tissues were queried in A) the Genotype-Tissue Expression (GTEx) project. The testis and skeletal muscle data are highlighted in red circle (Lonsdale et al., 2013).B) Translational human pancreatic islet genotype tissue-expression resource (TIGER) for MAFA and MAFB (29). TPM: transcript per million.

AUTHOR CONTRIBUTIONS

X.T., J.C., R.S. and H.M. contribute to the conceptualization of this study. G.R., M.G., J.L., Z.L., R.A.S. and J.C. executed data curation and formal analysis with input from X.T. and R.S. The AlphaFold2 data visualization was accomplished by R.A.S. The original draft and figures were prepared by X.T. The review and editing of the manuscript were done by H.M. and R.S. The resources (EndoC-βH3) were maintained and provided by E.W. and S.S.

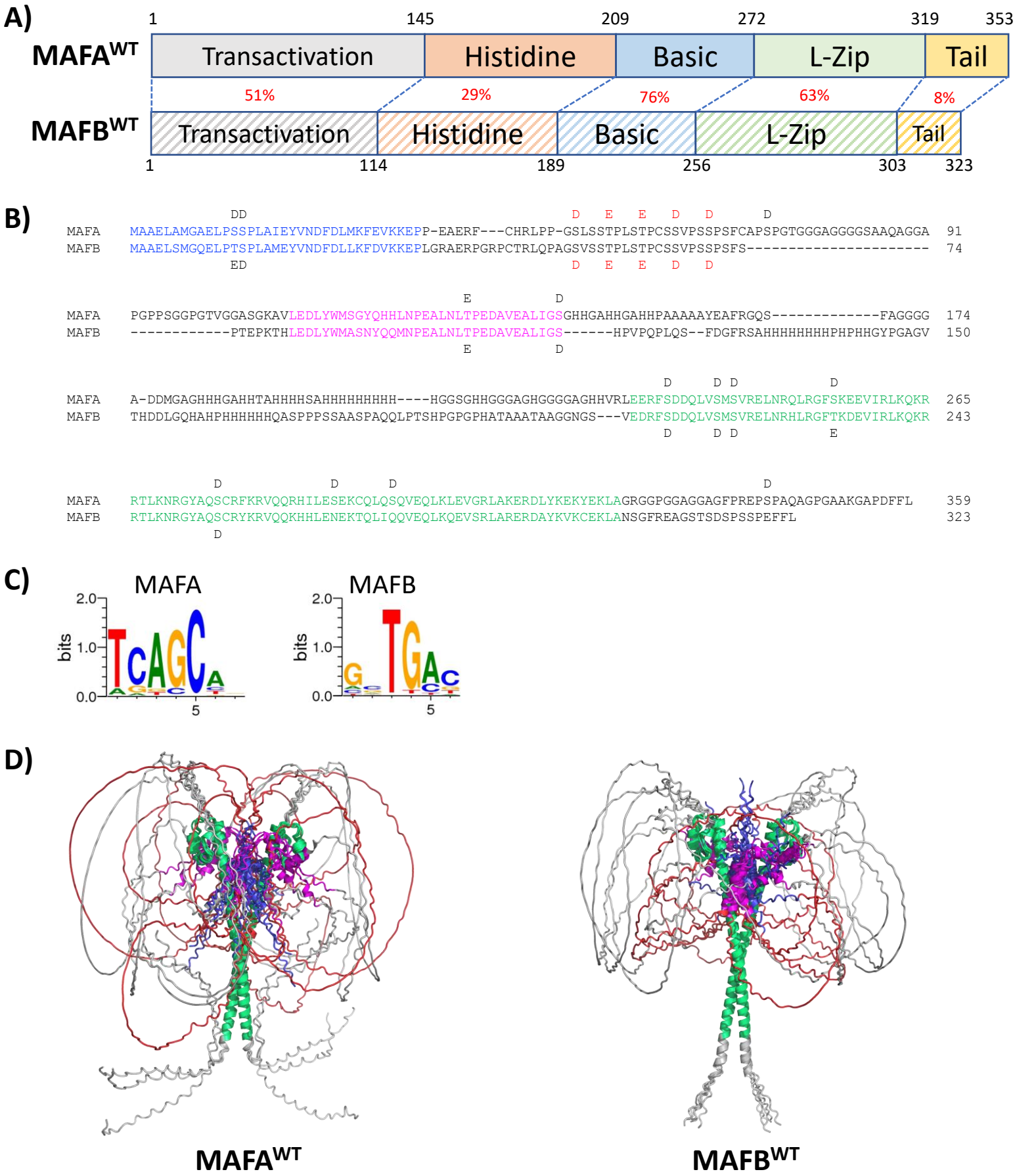

A)
MAFA^WT
1
145
209
272
319
353
Transactivation
Histidine
Basic
L-Zip
Tail
51%
29%
76%
63%
8%
MAFB^WT
Transactivation
Histidine
Basic
L-Zip
Tail
1
114
189
256
303
323
B)
MAFA MAAELAMGAELPSSPLAIEYVNDFDLMKFEVKKEPP-EAERF---CHRLPP-GSLSSTPLSTPCSSVPSSPSFCAPSPGTGGGAGGGGSAAQAGGA 91
MAFB MAAELSMGQELPTSPLAMEYVNDFDLLKFDVKKEPLGRAERPGRPCTRLQPAGSVSSTPLSTPCSSVPSSPSFS---------- 74
MAFA PGPPSGGPGTVGGASGKAVLEDLYWMSGYQHHLNPEALNLTPEDAVEALIGSGHHGAHHGAHHPAAAAAYEAFRGQS----------FAGGGG 174
MAFB ----------PTEPKTHLEDLYWMASNYQQMNPEALNLTPEDAVEALIGS------HPVPQPLQS--FDGFRSAHHHHHHHHPHPHHGYPGAGV 150
MAFA A-DDMGAGHHHGAHHTAHHHHSAHHHHHHHHH----HGGSGHHGGGAGHGGGGAGHHVRLEERFSDDQLVSMSVRELNRQLRGFSKEEVIRLKQKR 265
MAFB THDDLGQHAHPHHHHHHQASPPPSSAASPAQQLPTSHPGPGPHATAAATAAGGNGS---VEDRFSDDQLVSMSVRELNRHLRGFTKDEVIRLKQKR 243
MAFA RTLKNRGYAQSCRFKRVQQRHILESEKCQLQSQVEQLKLEVGRLAKERDLYKEKYEKLAGRGGPGGAGGAGFPREPSPAQAGPGAAKGAPDFFL 359
MAFB RTLKNRGYAQSCRYKRVQQKHHLENEKTQLIQQVEQLKQEVSRLARERDAYKVKCEKLANSGFREAGSTSDSPSSPEFFL 323
C)
MAFA
MAFB
bits
2.0
1.0
0.0
5
D)
MAFA^WT
MAFB^WT

Figure 1

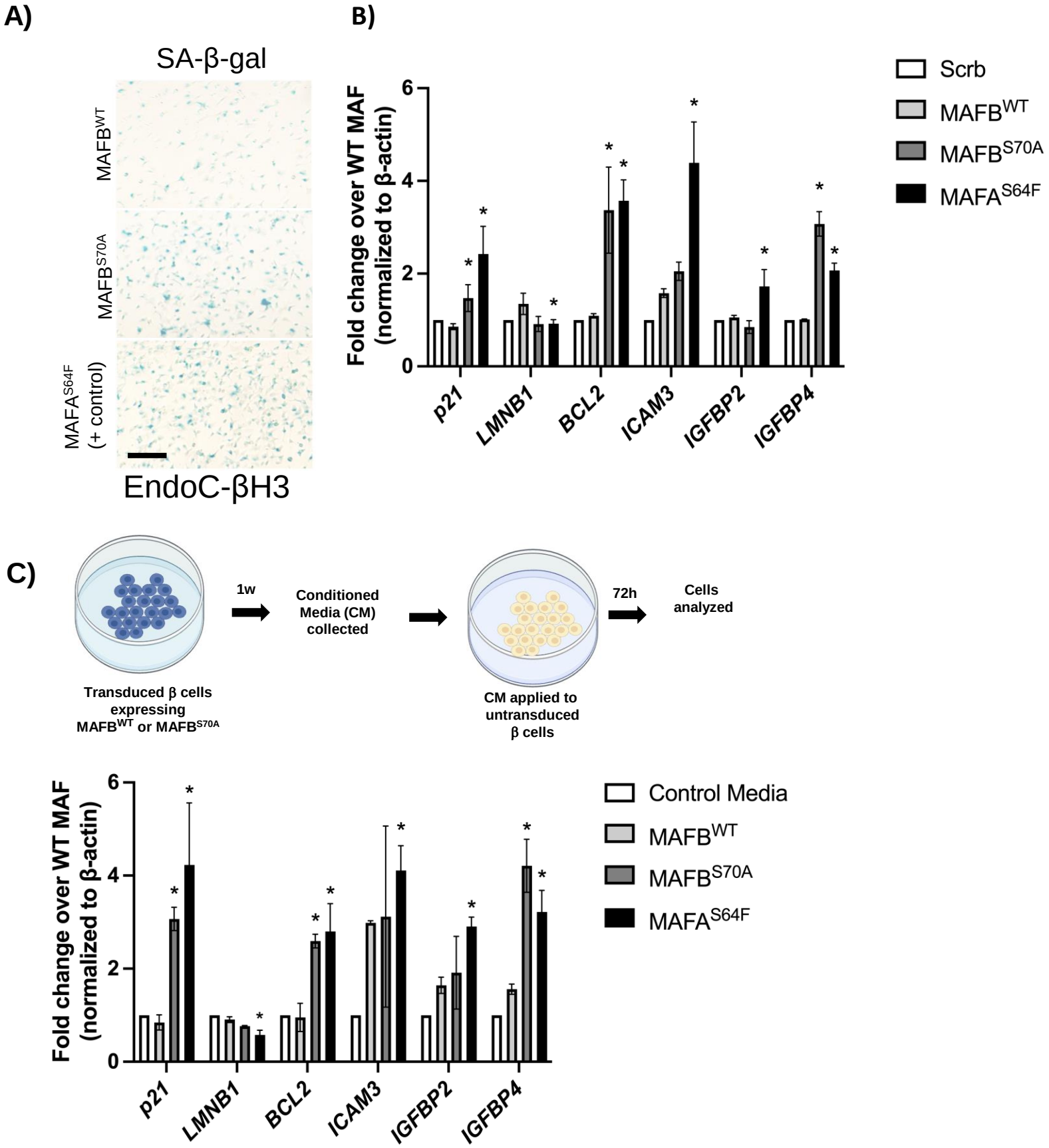

A)
SA-β-gal
MAFB^WT
MAFB^S70A
MAFA^S64F (+ control)
EndoC-βH3
B)
Fold change over WT MAF (normalized to β-actin)
Scrb
MAFB^WT
MAFB^S70A
MAFA^S64F
p21
LMNB1
BCL2
ICAM3
IGFBP2
IGFBP4
C)
1w
Conditioned Media (CM) collected
72h
Cells analyzed
Transduced β cells expressing MAFB^WT or MAFB^S70A
CM applied to untransduced β cells
Control Media

Figure 2

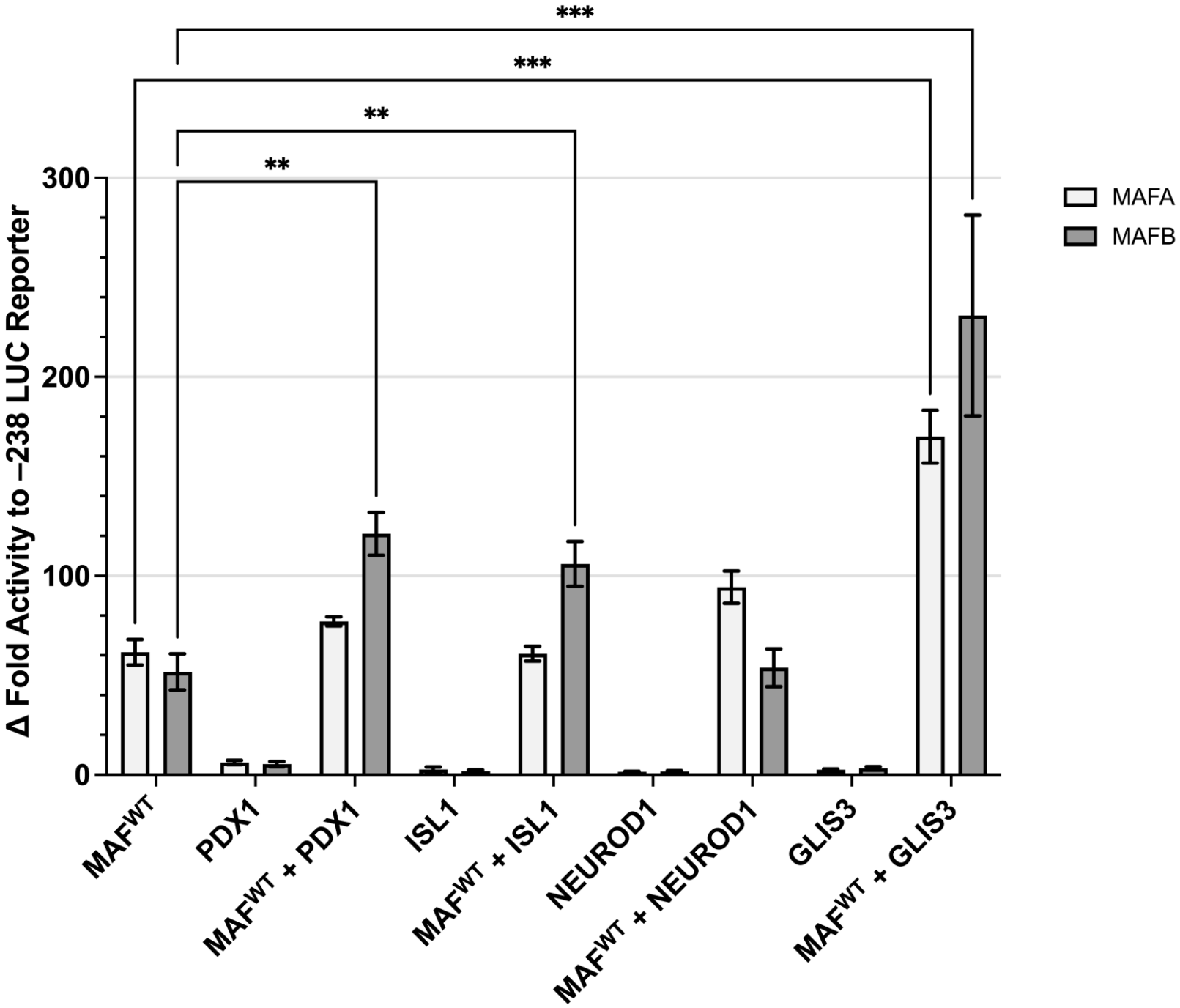

***
***
**
**
300
200
100
0
Δ Fold Activity to −238 LUC Reporter
MAFA
MAFB
MAF^WT
PDX1
MAF^WT + PDX1
ISL1
MAF^WT + ISL1
NEUROD1
MAF^WT + NEUROD1
GLIS3
MAF^WT + GLIS3

Figure 3

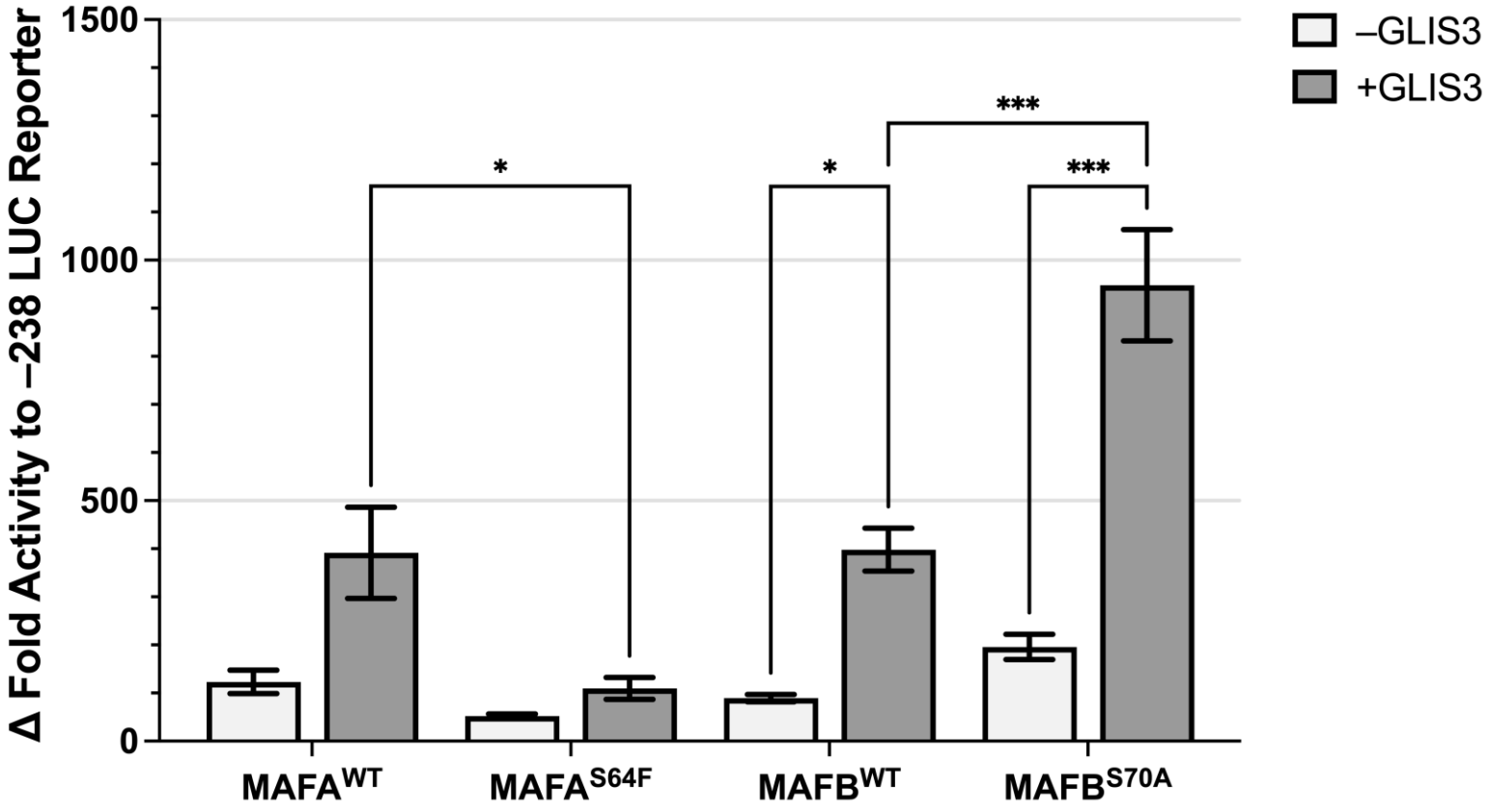
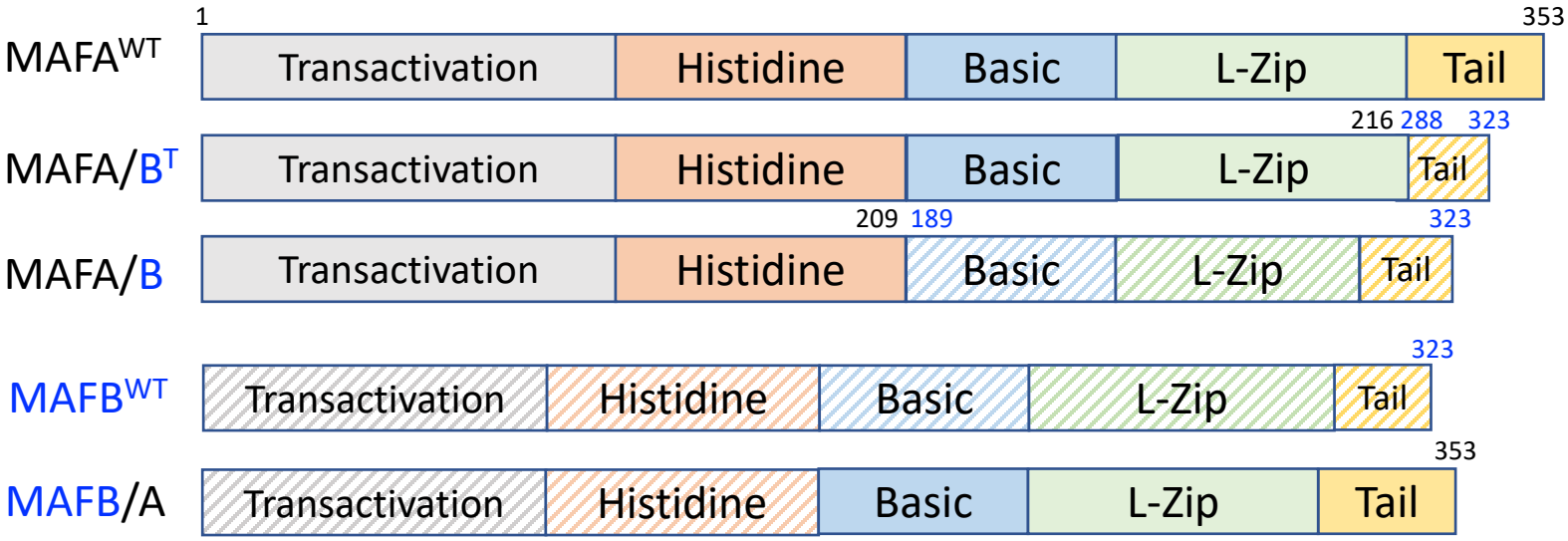
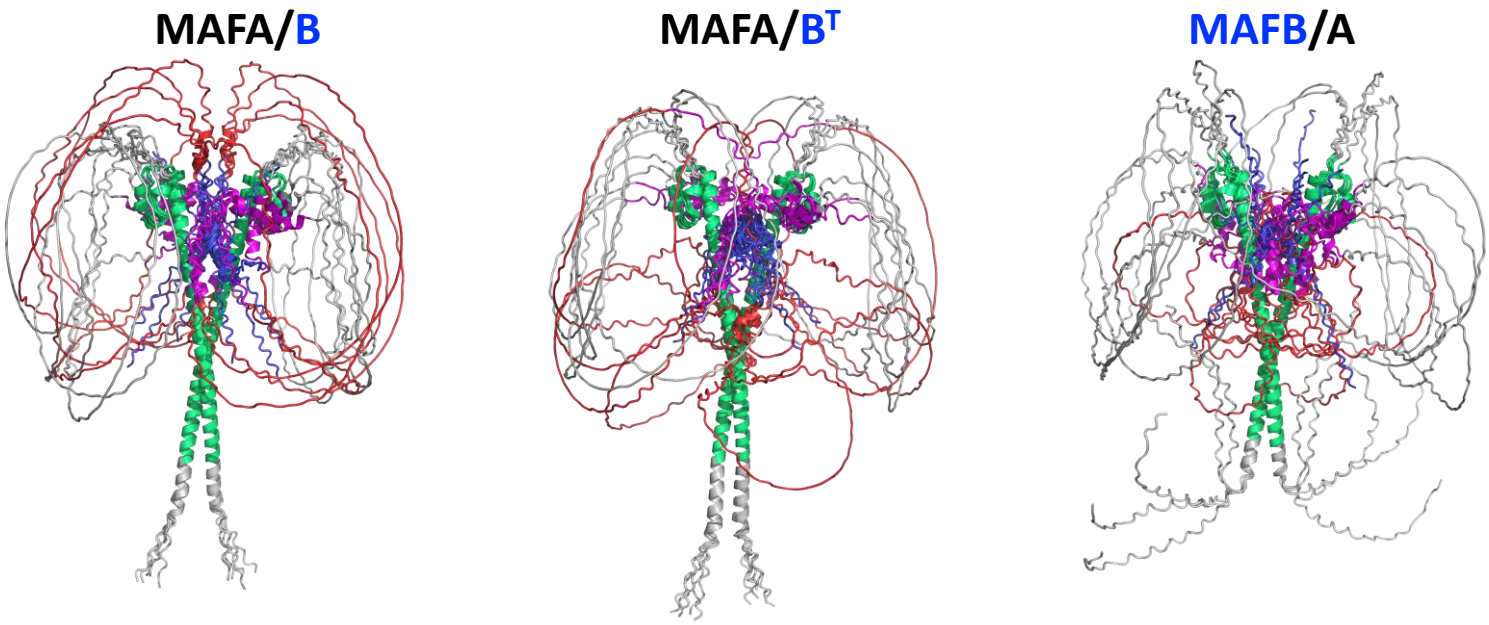

Δ Fold Activity to –238 LUC Reporter
1500
1000
500
0
–GLIS3
+GLIS3
*
*
***
***
MAFA^WT
MAFA^S64F
MAFB^WT
MAFB^S70A

Figure 4

A)


1
353
MAFA^WT
Transactivation
Histidine
Basic
L-Zip
Tail
216 288 323
MAFA/B^T
Transactivation
Histidine
Basic
L-Zip
Tail
209 189
323
MAFA/B
Transactivation
Histidine
Basic
L-Zip
Tail
323
MAFB^WT
Transactivation
Histidine
Basic
L-Zip
Tail
353
MAFB/A
Transactivation
Histidine
Basic
L-Zip
Tail


MAFA/B
MAFA/B^T
MAFB/A

B)

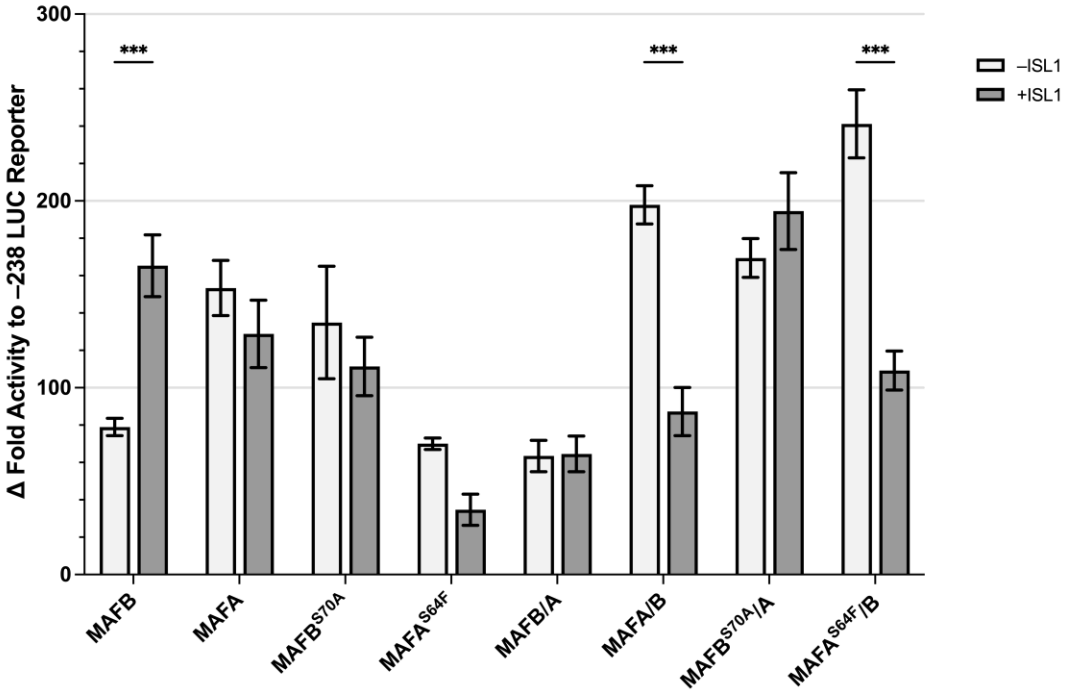

Δ Fold Activity to −238 LUC Reporter
300
200
100
0
***
***
***
−ISL1
+ISL1
MAFB
MAFA
MAFB^S70A
MAFA^S64F
MAFB/A
MAFA/B
MAFB^S70A/A
MAFA^S64F/B

Figure 5

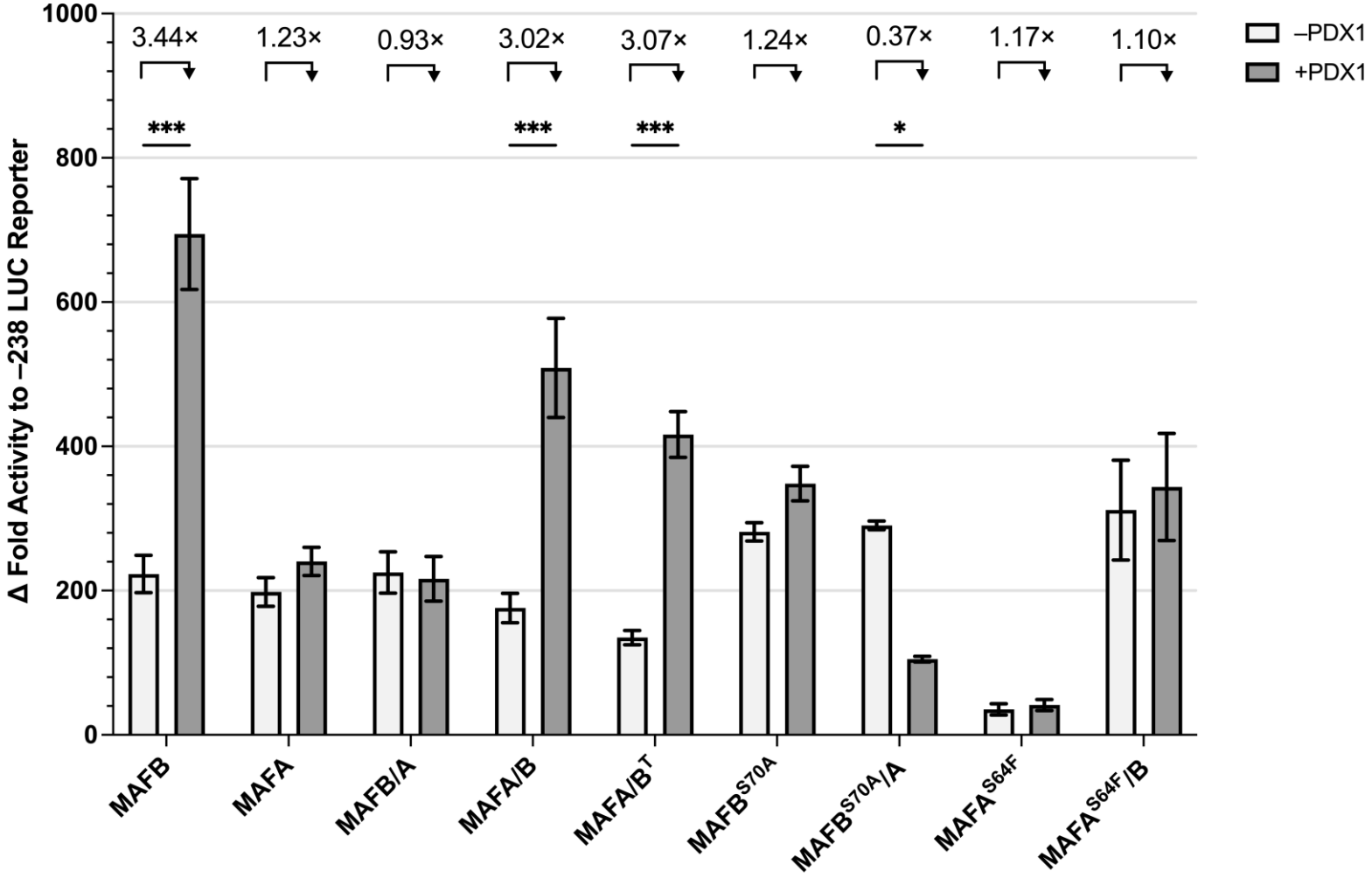

1000
800
600
400
200
0
Δ Fold Activity to −238 LUC Reporter
3.44×
1.23×
0.93×
3.02×
3.07×
1.24×
0.37×
1.17×
1.10×
***
***
***
*
–PDX1
+PDX1
MAFB
MAFA
MAFB/A
MAFA/B
MAFA/B^T
MAFB^S70A
MAFB^S70A/A
MAFA^S64F
MAFA^S64F/B

Figure 6