TWIST1 homo- and heterodimers orchestrate specificity control in embryonic stem cell lineage differentiation and craniofacial development

Xiaochen Fan[1,2, a+], Ashley J. Waardenberg[3, b], Madeleine Demuth[1, c], Pierre Osteil[1], Jane Sun[1], David A.F. Loebel[1,2], Mark Graham[4], Patrick P.L. Tam[1,2] and Nicolas Fossat[1,2, d]

[1] *Embryology Unit, Children's Medical Research Institute, The University of Sydney,* [2] *The University of Sydney, School of Medical Sciences, Faculty of Medicine and Health,* [3] *Bioinformatics Group, and* [4] *Synapse Proteomics Group, Children's Medical Research Institute, The University of Sydney.*

[+]Corresponding author: Xiaochen Fan email: x6fan@ucsd.edu

Present address:

[a] Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA

[b] Centre for Tropical Bioinformatics and Molecular Biology, James Cook University, QLD 4870 Australia.

[c] Developmental Dynamics Laboratory, The Francis Crick Institute, London NW1 1AT, UK.

[d] Copenhagen Hepatitis C Program (CO-HEP), Department of Immunology and Microbiology, University of Copenhagen, and Department of Infectious Diseases, Hvidovre Hospital, Denmark.

Abstract

Combinations of bHLH factors dimers generate great functional diversity required for cell type specification in development. The bHLH factor TWIST1 plays pleiotropic roles in craniofacial development. However, which combination of TWIST1 dimers are involved in craniofacial development, and what impact each dimer impose on gene regulation network remained elusive. Proteome profiling of human-TWIST1 expressing cell lines and transcriptome analysis of mouse cranial mesenchyme revealed TWIST1 homodimer and heterodimers with TCF3, TCF4 and TCF12 E-proteins as preferred combinations. We found that balance of homo- and heterodimers were impaired by human disease mutations in TWIST1 Helix domains, which may underpin the developmental defects in haploinsufficiency. Further, loss or gain of function analysis of TWIST1-E-protein dimers in differentiating embryonic stem cells revealed their previously unappreciated roles in lineage specification. TWIST1-E-protein heterodimers activate mesoderm and neural crest differentiation through epithelial-to-mesenchymal transition, whilst TWIST1 homodimers maintained a progenitor-like state and blocked entry to endoderm lineages.

Short title: TWIST1 dimers in lineage differentiation

Keywords: bHLH factor, Twist1, E-protein, embryonic stem cell, lineage differentiation, craniofacial development

Introduction

The basic helix-loop-helix (bHLH) superfamily is an evolutionarily conserved group of transcription factors required for lineage differentiation and early organogenesis. Through dimerization with different partner proteins, bHLH factors drive the specification of a plethora of cell types during neurogenesis, haematopoiesis and myogenesis (Schlaeger *et al.*, 2004; Belle and Zhuang, 2014; Comai and Tajbakhsh, 2014; Dennis *et al.*, 2019). Eukaryotic helix-loop-helix (HLH) factors can be classified based on their structural and functional attributes (Massari and Murre, 2000). Class I, II and V factors are known to engage in homo- and heterodimerization. Class I proteins, comprised of all E-proteins (e.g. E12/TCF3, HEB/TCF12 and E2-2/TCF4), are broadly expressed in various tissues (Wang and Baker, 2015). Class II proteins generally have restricted expression in specific cell lineages, for example, MyoD in myogenic cells, NeuroD in neurogenic cells and TWIST1 in the tissue mesenchyme (e.g. craniofacial and limb mesenchyme) (Thisse *et al.*, 1987; Lee *et al.*, 1995; Spicer *et al.*, 1996; Ghouzzi *et al.*, 2000; Bildsoe *et al.*, 2009; Bildsoe *et al.*, 2013; Bildsoe *et al.*, 2016). Class V represents a group of non-DNA binding HLH proteins, such as IDs, that competes for E-protein binding (Benezra *et al.*, 1990). It was hypothesised that the bHLH dimer composition is balanced by the expression level of bHLH factors, competition between them, and phospho-regulation (Centonze *et al.*, 2004; Firulli *et al.*, 2005; Firulli *et al.*, 2007). However, the identity of dimer combinations and their specificity during development remained enigmatic.

The bHLH factor TWIST1, highly expressed in cranial mesoderm- and neural crest- derived mesenchyme (Brault *et al.*, 2001; Bildsoe *et al.*, 2009; Bildsoe *et al.*, 2013), is critical for craniofacial development. *Twist1*$^{+/-}$ mice display craniosynostosis (Carver *et al.*, 2002; Connerney *et al.*, 2006) that partly phenocopies skeletal defects associated with TWIST1

haploinsufficiency in Saethre-Chotzen Syndrome (SCS, AHC [MIM: 101400]). Conditional ablation of *Twist1* from the cranial mesoderm or the cranial neural crest leads to malformations of the cranium, brain, cranial nerves, facial skeleton and muscles (Bildsoe *et al.*, 2013) (Chen and Behringer, 1995 Soo et al; Ota *et al.*, 2004). At cellular level, *Twist1* is required for maintaining the mesenchymal potency of progenitor cells for osteo-, chondro-, and adipogenic tissues (Brault *et al.*, 2001; Bildsoe *et al.*, 2009; Miraoui *et al.*, 2010; Bildsoe *et al.*, 2013). *Twist1* deficient craniofacial tissues surrounding midbrain and hindbrain acquire atypical epithelial features whilst losing mesenchymal features (Brault *et al.*, 2001; Bildsoe *et al.*, 2009; Miraoui *et al.*, 2010; Bildsoe *et al.*, 2013). Previous studies highlight the differential functions of TWIST1 dimers in the osteogenic differentiation of the cranial sutural mesenchyme (Connerney *et al.*, 2006; Connerney *et al.*, 2008), which is mediated by their targeted action on FGF signalling (Guenou *et al.*, 2005; Rice, 2005; Miraoui *et al.*, 2010). For example, the TWIST1-TCF3 heterodimer promotes mesenchymal expansion, while TWIST1-homodimer activates FGFR, OCN and BSP expression for ossification. Identifying TWIST1 dimerization partners and their transcriptional targets in the cranial mesenchyme, is therefore likely to allow for a better understanding of TWIST1 function during early development.

In this study, dimerization partners of TWIST1 were identified by mass-spectrometry analysis of TWIST1-containing complexes immunoprecipitated from human-TWIST1 (hTWIST1) expressing mesenchymal cells, in conjugation with *Twist1* co-expression analysis in mouse embryonic head tissues. A bi-fluorescent complementation (BiFC) assay was used to elucidate the balance between hetero- and homo-dimerization and the impact of pathogenic mutations. Finally, to delineate the immediate downstream targets and the function of TWIST1-dimers, we analysed differentiating ESC with *Twist1* knockout mutations or TWIST1-E-protein

inducible cassettes. Our results revealed a concert of TWIST1-dimers activities in mesoderm and neural crest determination and cellular behaviour: TWIST1-E-proteins heterodimers promote mesoderm and neural crest differentiation through epithelia-mesenchymal transition, whilst homodimer activity maintained a progenitor-like state and blocked entry to endoderm lineages. Taken together, our findings illuminated a previously unappreciated mechanism of specificity control in craniofacial development and disease, encoded by TWIST1 molecular interactions.

Results

Identification of dimerization partners of TWIST1 in embryonic head

Candidate dimerization factors were identified by their co-expression with *Twist1* in the mouse embryonic head tissues. Microarray analysis of cranial neural crest cells (CNC, Wnt1-Cre active) and cranial mesoderm cells (CM, Mesp1-Cre active) revealed that 58 out of 158 known bHLH factors (Skinner *et al.*, 2010) were expressed in the head mesenchyme (Bildsoe *et al.*, 2016; Fan *et al.*, 2016; Supplemental Table S1). Twelve bHLH factors were significantly enriched in CNC or CM (Fig. 1A) and forty-six were expressed in both tissues (Supplemental Table S1). Based on their known roles in craniofacial development, seven candidates were selected for validation, including SIM2, TCF4, EBF1, EBF3, TAL1, TWIST2, and TCF3 (isoform of E2A and a known TWIST1 partner as a positive control). HA-tagged protein constructs were transfected into Madin-Darby Canine Kidney (MDCK) cells that stably over-expressed hTWIST1 (designated as MDCK/hTWIST1-OE cells), which displayed TWIST1 induced mesenchymal phenotypes (Xue *et al.*, 2012; Fig. 1B). These factors were detected in complex with TWIST1 by co-immunoprecipitation using α-TWIST1 antibody. Reciprocally, TWIST1 was detected in

together with HA-tagged EBF1, EBF3 and TCF4 in α-HA immunoprecipitated fraction (Fig. 1B), demonstrating interaction with the three bHLH partners. Genes encoding EBF1, EBF3 and TCF4 were co-expressed with *Twist1* in the frontonasal tissues, the cranial mesoderm and the first branchial arch of the mouse embryo (Fig. 1C).

Proteomic screening of MDCK/TWIST1-OE cells revealed TCF factors as the prevalent TWIST1 partners

Further delineation of TWIST1 dimer partners was undertaken by unbiased proteomic screening by mass spectrometry following affinity purification of TWIST1-complexes from MDCK/hTWIST1-OE cells (Fig. 2A). Protein complexes were purified from lysate using α-TWIST1 antibody or mouse IgG (as the control) in triplicate. 846 proteins were identified and divided into three sets using the ranked-product method for differential analysis (Breitling *et al.*, 2004) (Fig. 2B and Table S1). "Set 1" represented 103 identified proteins that specifically bound to TWIST1, "Set 2" were those proteins that bound specifically to IgG beads (377), while "Set 3" contained proteins bound both IgG beads and TWIST1 (362). As expected, TWIST1 was one of the top enriched proteins in "Set 1" (Fig. 2B). Only two other bHLH factors, TCF12 and TCF4, were highly enriched, and the rest of “Set 1” putative partners included RNA-binding proteins and transcription factors such as Cyclin T2, TRIP6, GATA1, FHL2 and CRTC2 (Table S1). *Tcf4 and Tcf12*, like *Tcf3,* were expressed in the facial primordia and overlapped with *Twist1* expression domain in the frontonasal, the branchial arches and the mesoderm in embryonic head (Fig. 2C). We also demonstrated the interaction of TWIST1 with endogenous TCF4 and TCF12 in mouse embryoid bodies (EBs) overexpressing TWIST1 (Fig. 2D).

SCS mutations impair TWIST1-TCF12 dimer formation and function

The TWIST1 bHLH dimerisation domain shows 84–100% sequence identity among vertebrate orthologues (including human, mouse, frog and zebrafish). It is also highly conserved (~70%) among the TWIST1 family proteins: TWIST1, TWIST2, HAND and HAND2, reflecting functional conservation and significance (Castanon and Baylies, 2002; Qin *et al.*, 2012). There is a high incidence of 63 base clinically significant missense mutations in the bHLH domain, many of which are identified from SCS patients (Fig. 3A; Human genomic mutation database; Stenson *et al.*, 2003; Centonze *et al.*, 2004; Firulli *et al.*, 2005; Firulli *et al.*, 2007; Barnes and Firulli, 2009) Some of these mutations are located directly at or around the threonine and serine phosphosites (Fig.3A), and were found to inhibit TWIST1 phosphorylation (Firulli *et al.*, 2005; Firulli and Conway, 2008).

To validate the identified interactions and test the hypothesis that TWIST1 pathogenic mutations at bHLH region may shift the balance of homo- versus hetero-dimerization, we employed a multicolour BiFC assay (Hu and Kerppola, 2003). It was previously shown that TWIST1 SCS mutations in the basic-helix I region, such as R118H, may impair phosphorylation at T121 and S123 residues, and alter the preference of dimerization with TCF3 and HAND2 (Firulli *et al.*, 2005). A subset of mutations in the basic-helix I region were selected (Fig.3A, red): TWIST1(S123X) truncated at the bHLH, phosphorylation-inhibitory mutations (S123A and R118H) and phospho-mimetic mutations (T121E and S123E). We also selected previously uncharacterised residues on helix II that were conserved among TWIST-family members (T148A, A152P and I156N). When examined on a 3D reconstruction of TWIST1/TCF12 dimer (bHLH regions), the phospho-regulatory residues appeared to be in closer proximity to the DNA interface, rather than the site of E-protein interaction (Fig. 3Bi). On other hand, clinical variants on helix II, affected residues at the protein interaction interface (Fig. 3 Bii).

Since TCF12 is also associated with SCS and other craniofacial anomalies (Sharma *et al.*, 2013), we tested the impact of these mutations on TWIST1 dimerization with TCF12. TWIST1 and TCF12 were conjugated with different split fluorescent domains (N- or C-terminal) and co-expressed in C3H10T1/2 cells (Fig. 3C). Using BiFC, we monitored dimerization of SCS mutant TWIST1 proteins with TCF12. The interaction of TCF12-CrN173 versus TWIST1-VN173 with TWIST1* mutant- or WT-CC155 was assessed by imaging the emission intensity of the two fluorophores (Fig. 3C). When expressed at similar levels, TWIST1 mutants displayed significant changes in their preference to form homo- (GFP) or heterodimers (CFP; Fig. 3D, E). Amongst all mutations tested, truncation at the bHLH domain (S123X) most severely compromised dimerization (Fig. 3D). S123A and R118H, however, significantly increased the preference for heterodimer formation, compared to WT. In contrast, the phospho-mimetic form (TSE or T121E & S123E) of TWIST1 favoured homo-dimerization. All three mutations at helix II (in the C-Term region) were associated with reduced TWIST1/TCF12 and TWIST1/TWIST1 ratios and (Fig. 3D), and overall lower dimerization activity (Supplemental Fig. S1B). Meanwhile, TWIST1 with a deletion of the N-term region did not affect dimer ratio (Supplemental Fig. S1A).

Mutations associated with a significant shift in dimer balance were further tested for the impact on cell migration, a TWIST1-driven biological process. Scratch assay was performed on C3H10T1/2 cells co-transfected with *Twist1**mutant-*CC155* and *Tcf12-CrN173* constructs expressing the recombinant proteins (Fig. 3E). Cells expressing TWIST1 with helix II mutations (A152P, I156N), showed reduced cell migration compared to TWIST1 wildtype/TCF12 expressing cells (Fig. 3F). These results demonstrate that mutations in Helix II that severely disrupted TWIST1/TCF12 dimerization also impaired downstream function in mesenchymal cell migration.

Loss of *Twist1* impairs ESC pluripotency exit and cell fate specification

We detected TWIST1 expression during the differentiation of ESCs, initially from Day 3 and peaking at Day 6-7 of differentiation (Fig. 4A). To examine TWIST1 functionality in differentiation, we generated mono- and bi-allelic *Twist1* knockout mouse ESC cell lines using CRISPR-Cas9 technique (Fig. 4B, C), and compared expression of pluripotency and lineage markers with the parental cell line on Day 3 of differentiation (Fig. 4D-F). Pairwise statistical tests revealed that both full and partial loss of *Twist1* disrupted lineage specification in a gene dosage-dependent manner. Pluripotency genes such as *Dppa2, Nanog*, *Sox2* and *Pou5f1* retained expression at the pre-differentiated levels in the mutant cells, significantly higher than the Day-3 wildtype counterpart (Fig. 4D). Concurrently, ectodermal and mesodermal genes failed to be activated in the mutant cells (Figure 4E, F). Expression of EMT genes, such as *Flk1, Pdgfra, Snail2* and *Vegfa* were reduced, whilst the epithelial marker, *Cdh1*, was up-regulated (Fig. 4F). In contrast, endodermal markers (such as *Gsc, Rbm47, Foxa2*) and WNT signalling components (such as WNT ligand *Wnt3*, receptors *Fzd10* and *Prickle1,* target gene *Axin2*) were up-regulated in the mutant cells (Fig. 4G, H). These findings implicated various roles of TWIST1, potentially attributable to different dimer-patterns, in the maintenance of the mesenchymal progenitor and selectively directing lineage differentiation of the embryonic stem cells.

Comparison of the target specificity of different TWIST1 dimers

To further elucidate the scope of TWIST1-dependent function in lineage differentiation, a gain of function approach was taken to analyse the downstream genes of the homodimer and heterodimers. We tethered TWIST1 and TCF proteins into dimers via flexible poly-glycine

linkers and expressed these dimers in ESCs (Fig. 5A, B). In *Drosophila*, tethered Twist-Twist and Twist-Dorsal dimers could recapitulate the function of dimers formed naturally by Twist1 and Dorsal; Twist homodimer induced *Mef2* expression and muscle differentiation *in vivo*, whereas the Twist-Dorsal dimer antagonised this activity (Castanon and Baylies, 2002). Following the same principle, we constructed dimers for TWIST1 and E-proteins partners (Fig. 3A): TWIST1-TWIST1 (TT), TWIST1-TCF3 (T3), TWIST1-TCF4 (T4), TWIST1-TCF12 (T12) and TWIST1 monomers (T0). Coding sequences were stably integrated into the genome of A2LoxCre ESCs, downstream of a doxycycline inducible promotor (Iacovino *et al.*, 2014). ESCs were then differentiated for three days as EBs, and assayed for the expression of developmental/lineage genes (Fig. 5B). Before doxycycline induction, no expression of the transgene was detected (Supplemental Fig. S2A). Dimer expression was induced on Day 2 in synchrony with the onset of upregulation of endogenous *Twist1* (Figure 4A, 5B). To ensure that the effects of the expression of each dimer was not confounded by the level of expression, we first determined and used optimised level of doxycycline concentration for each construct to achieve comparable levels of protein expression (Methods, Supplemental Fig. S2A).

To decipher the gene regulatory networks imposed by different TWIST1 dimer combinations, we conducted RNA-seq analysis of EBs expressing comparable levels of T0 monomer and TT, T3, T4 and T12 dimers, with the A2loxcre parental cell line differentiated under the same conditions as control (Fig. 5C, Supplemental Fig. S3, Materials and Methods). Unsupervised principle component analysis (PCA) and hierarchical clustering confirmed groupings of replicates (Fig. 5D, F). T3, T4 and T12 heterodimers separated from TT, T0 and A2LC dimer combinations. Among the three heterodimers, cells expressing T4 and T12 dimer exhibited more similar gene expression profile (Fig. 5F). Examination of genes driving the differences

(Fig. 5G) revealed that heterodimers activated the expression of genes associated with specification of mesenchymal tissues (cardiovascular system, skeletal system) and neurons; molecular pathways of neural crest cell differentiation and TGFβ signalling (Fig. 5H). Pairwise DEG analysis was performed for T0, TT, T3, T4 and T12 against the parental A2LoxCre cells. For each group between 100-600 genes were significantly upregulated, and 100-200 down-regulated (adjusted pvalue < 0.05, fold change >= 2; Fig. 6A).

It was previously demonstrated by *in vitro* mobility shift assay that dimer formation alters DNA sequence specificity (Firulli *et al.*, 2007; Chang *et al.*, 2015). Sequence motif analysis was performed on the promotors of dimer-specific DEGs. Top ranking regulatory motifs for the heterodimers comprised of canonical E-boxes “CATCTG” or “CAGCTG" (Fig. 6B). TT targets were uniquely enriched in double E-boxes with a five nucleotide spacer, as reported in a ChIP-seq study in human mammary epithelial cells over-expressing *TWIST1* (Chang *et al.*, 2015). Binding of the double E-box motif, found to this date only for *Drosophila* Twist and human hTWIST1, may suggest higher order tetramer formation between TWIST1 and E47 (a spliceoform of TCF3; Chang *et al.*, 2015). Our results showed that recognition of a double E-box might be enhanced by the tethered TWIST1 homodimer. TWIST1 monomer recapitulated partially the activity of TT and T3 dimers, but also regulated a unique group of genes (177/336 upregulated and 87/218 down regulated). In consistence, the monomer showed reduced recognition of canonical E-boxes compared with the dimer counterparts, and were redirected to GATA motifs, suggesting additional co-factor binding.

To dissect the functionality of each dimer individually, we performed additional gene-set enrichment analysis based on gene expression values (see Methods). EBs expressing monomer or homodimers displayed repression of developmental gene-sets relative to the

parental A2LoxCre cells (Supplemental Fig. S4A; Supplemental Table S2). Differentiation of neural-ectoderm derivatives were repressed by the monomer (T0), and mesoderm and endoderm cell fates were repressed by TT. On the other hand, heterodimers activated overlapping sets of genes associated with mesoderm and neural ectoderm development (Supplemental Fig. S4A; Supplemental Table S2). T3 dimer activated heart developmental processes, including heart vasculature, muscle, and blood cells, while the most frequent processes induced by T4 were related to neuron differentiation (Supplemental Fig. S4A; Supplemental Table S2). T12 activity was intermediate between T3 and T4. A comparison down to specific cell lineage and pathway genes also reflected these changes (Fig. 6C, D). Evident in the heterodimer groups, was the downregulation of pluripotency genes and a burst in cell lineage markers, especially those for mesoderm and neural crest development (Fig. 6C). T3 dimer preferentially activated mesoderm markers, whilst T4 dimer favoured ectoderm and neural crest genes (Fig. 6C). Heterodimers also showed a stronger capacity to activate genes associated with EMT pathway (Supplemental Table S3, Fig. 6D)

Taken together, our data demonstrated that TWIST1 dimerization partners drive different gene expression programs that may be influenced by the preference for specific DNA motifs. Heterodimers appeared to drive neural and mesoderm differentiation, whereas the monomer or homodimer of TWIST1 repressed the expression of genes associated with endoderm differentiation.

TCF12-TWIST1 dimer activates EMT more robustly and antagonise TWIST1 homodimer activity for critical craniofacial pathway genes

Given that the synergy between *Twist1* and *Tcf12* in normal cranial tissue development

(Sharma *et al.*, 2013), we compared the ability of T12 and TT dimer to induce transcription of developmental pathways (Supplemental Table S3, Supplemental Fig. S4B, C). T12-expressing cells significantly activated more genes associated with TGFβ signaling (FDR 0.007), WNT signaling (FDR 0.017) and EMT compared to TT (FDR 0.002; Fig. 7A). Although both T12 and TT could promote the expression of EMT genes, around 50% of all EMT genes were enriched in T12 cells compared to TT (Fig. 7A). T12 expressing cells were also enriched for genes associated with developmental processes of neuronal, skeletal, face and cardiovascular system development (Fig. 7B, Supplemental Fig. S4B). We then examined transcriptional regulators that has been previously associated with craniofacial development (Fig. 7B, Face development). As mentioned earlier, studies in the skull suture suggested that antagonistic regulation by the homo- and heterodimers (Connerney *et al.*, 2006; Connerney *et al.*, 2008),. Indeed, we noted that a subset of critical developmental regulators, such as those associated with TGFβ signaling (e.g. *Tgfbr1, Pdgfra*) and WNT (*Prickle1, Dkk1*) were antagonistically regulated by the homodimer and heterodimers (Fig. 7C). Most other genes, represented by *Fgfr2, Jun, Chd7* and *Mmp2* were activated to a greater extent by T12 than TT (Fig. 7C). The differential expression of downstream genes between T12 and TT was likely to be the outcome of the balancing activity of these two forms of dimers in transcriptional regulation.

Discussion

TCF factors are the major dimerization partner of TWIST1

Our study identified TCF3, TCF4 and TCF12 E-proteins, which were expressed in cranial mesenchyme, as the predominant interacting partners of TWIST1. E-proteins partner with

several other Class II bHLH factors (Forrest *et al.*, 2014). The three E-proteins also interacted with HAND2 (preferentially for TCF3) and shown to greatly enhance transcriptional activity in GAL4 reporter assays in yeast (Murakami *et al.*, 2004). Consistently, we find that TWIST1 activated ESC differentiation programs more robustly when coupled with E-proteins. Both HAND2 and TWIST1, when coupled to different E-proteins, also leads to altered E-box binding sequence specificity (Murakami *et al.*, 2004; Firulli *et al.*, 2007). This altered sequence specificity is likely to underpin the regulation of the expression of different genes, leading to a mechanism whereby dimer combinations may target to developmental gene regulatory networks in specific cellular and timing context.

Molecular pathology of SCS mutations in context of bHLH factor dimerization

Dimerization is modulated by the availability of partner proteins, competing factors and the phosphorylation of the bHLH domain (Centonze *et al.*, 2004; Firulli *et al.*, 2005; Firulli *et al.*, 2007). Hypo-phosphorylation mutations (T121A, S123A) near or at the Helix I phosphorylation sites was previously shown to favour homodimerization (Firulli *et al.*, 2005). Yet a triple phospho-mutant model of TWIST1 (S123A, T148A and A184A) in oncogenic cell lines showed preference for dimerization with TCF3 and HAND2 compared to wildtype TWIST1 (Wang *et al.*, 2017). In line with this, we find that two individual hypo-phosphorylation mutants (R118H, S123A) both increased preference for TCF12 heterodimer formation, whilst phospho-mimetic forms at T121 and S123 reduced heterodimerization, compared to wildtype TWIST1. We note that incorporation of the negatively charged phosphate group to these residues may reduce binding to DNA, rather than directly interfering with the dimerization surface (Figure 3B). Therefore, T121/S123 dephosphorylated TWIST1 may favour heterodimerization through stabilising DNA-binding or adapting a helix conformation that enhances E-protein binding.

Mutations at TWIST1 Helix II compromised both types of dimerization. This impaired dimerization functionally impacts on the migration of mesenchymal cells, which was not encountered for variants near Helix I. Together, these conserved residues, in the context of their structural role in helix conformation, act as an important dimerization interface.

Our finding that TWIST1 SCS mutations disrupted the relative ratio of T12 and TT dimers may underpin the dysregulation of developmental genes in disease models. RNA-seq analysis in differentiating ESCs revealed that the TWIST1-TCF12 dimer was a potent driver of EMT and mesenchymal differentiation, compared to TWIST1 homodimer. T12 also induced signalling pathways and key transcription regulators associated with craniofacial development. These genes were either inactive or repressed in the homodimer expressing cells. The *Twist1* heterozygous (*Twist1*$^{+/-}$) mouse model that recapitulates premature fusion of the suture displays excessive FGF and TGFβ signalling at the osteogenic front to mid-suture mesenchyme, as revealed by the enhanced expression of FGFR2, pSmad 1/5/8, and ID proteins (Connerney *et al.*, 2006; Connerney *et al.*, 2008). Our transcriptome analysis suggests that these molecular responses may be related to the enhanced activity of TWIST1-E-protein heterodimer. The Basic-Helix I mutations (R118H, S123A) could directly lead to increased heterodimer ratio. In contrast, the helix II variants generated dimerization-defective TWIST1 protein, leading to situation similar as haploinsufficiency. These findings support a defect-causing mechanism driven by bHLH haploinsufficiency.

Dimerization-dependent activities of TWIST1 in progenitor cell lineage specification

TWIST1 regulates multiple pathways in linage specification *in vitro*. Loss of *Twist1* function led to retention of pluripotency, defective mesoderm and ectoderm differentiation, as well as

disrupted EMT process. EMT is tightly associated with ESC differentiation, especially the specification of mesoderm (Nakaya and Sheng, 2008; Evseenko *et al.*, 2010; Iacovino *et al.*, 2014). Canonical WNT signaling directs ESC germ layer differentiatoin (Ten Berge *et al.*, 2008), and induces major EMT drivers including *Twist1, Zeb2, Snail1* and *Snail2*. However, constitutive activation of WNT primes the differentiation of precursor cells towards primitive endoderm (Price *et al.*, 2013). Both *Twist1* and *Snail1* are required for WNT induced mesoderm differentiation (this study; Sharma *et al.*, 2013), but only *Twist1* appeared to negatively feedback on WNT signaling regulation and repress endoderm cell fates. Over-expression of the TWIST1 homodimer blocked WNT activity and endoderm differentiation, whilst loss of TWIST1 shifted EBs towards endoderm. On the other hand, ablation of *Snail1* promoted ectodermal fate instead of mesendoderm, without affecting WNT pathway genes (Sharma *et al.*, 2013). These observations suggested two arms of control by TWIST1; an EMT-dependent role in promoting mesoderm and neural crest commitment, and an EMT-independent role in endodermal repression, potentially by negatively feedback on WNT activity.

These pleiotropic activities TWIST1 during lineage commitment were directed by E-proteins. E-proteins display overlapping expression patterns and developmental functions in embryonic tissues (Murakami *et al.*, 2004; Wang and Baker, 2015). They are expressed in the cranial mesenchyme and proliferating neuroepithelium in E9.5 - E11 mouse embryos (Murakami *et al.*, 2004). *Tcf3* and *Tcf12* expression are restricted in the proliferative ventricular and sub-ventricular zones and are down-regulated during neurogenesis (Uittenbogaard and Chiaramello, 2002; Pfurr *et al.*, 2017). *Tcf4* expression persists in the adult brain regions that overlaps with neuronal bHLH factors such as ATHO1, ASCl1 and NEUROG2, and initiates

neuronal specification (Forrest *et al.*, 2014). Consistent with these findings, tethered TWIST1-TCF4 was found to modulate the activity of neural-ectoderm genes, whereas tethered TWIST1-TCF3 acted on the mesoderm genes. TWIST1-E-protein heterodimers activated the TGF-β signalling pathway and neural crest cell differentiation, and up-regulated ID proteins, which may provide a negative feedback mechanism to modulate the activity of the heterodimer (Connerney *et al.*, 2006).

In conclusion, our study uncovered patterns of TWIST1 dimerisation that offered additional controls in lineage differentiation. Our finding suggest that TWIST1 cooperates with TCF partners to promote the differentiation of neural crest and mesoderm, and represses endoderm as homodimers. Different heterodimers may fine tune mesenchyme precursors for myogenesis, skeletogenesis and neurogenesis fates. Our study also provided resource for the interpretation of results and controversies from TWIST1 over-expression systems, in which the phenotypic effect could be a net result of the action of homodimers and heterodimers. In the context of craniofacial morphogenesis, we showed that loss of function mutations of *Twist1* may disrupt dimer balance, which may underpin the mesenchymal tissue defects and the aberrant neural crest cell development. Ultimately, using advanced molecular biology tools, this and future studies would shed light on precision control in complex developmental processes, through molecular information encoded and transmitted through protein interactions.

Materials and Methods

Cell culture

The Madin-Darby Canine Kidney (MDCK) cell line over expressing human *TWIST1* was received from Xue and colleagues (Xue *et al.*, 2012). Cells were maintained in high glucose Dulbecco's modified Eagle's medium (DMEM, Gibco), 10% Fetal calf serum (FCS) and 10 mM β-mercaptoethanol at 37°C and 5% CO2. *TWIST1* expression was constant selection using 4 µg/mL of puromycin (Xue *et al.*, 2012).

C3H10T1/2 cell line was received from ATCC. Cells were maintained in high glucose DMEM with 10% FCS (Life Tech, Australia) and 10 mM β-mercaptoethanol at 37°C in 5% CO2. Both cell lines were typically split at ratio 1:6- 1:10 to maintain desired density.

Mouse ESCs (A2loxCre) was a gift from Kyba Lab (Lillehei Heart Institute, Minnesota, USA), originally described in Mazzoni et al., (2011). The ESCs were cultured in complete mESC medium (high glucose DMEM (Gibco), 12.5 % (v/v) heat-inactivated FBS (Fisher Biotec Australia), 10 mM β-mercaptoethanol, 0.09 % (v/v), 1X non-essential amino acids (100X Thermo Fisher Scientific), 1 % (v/v) nucleosides (100X, Merck) and 10 mil U/mL ESGRO® mouse leukaemia inhibitory factor (Merck)). ESCs were fed with fresh media daily and passaged every 2–3 days. All cell lines were tested free of mycoplasma

Protein Affinity Purification and Mass spectrometry

2x10^7 MDCK/*hTWIST1* over-expressing cells were collected for each experiment. Cell pellet were thawed in 500 µl hypotonic lysis buffer (HEPES 20 mM, MgCl2 1 mM, Glycerol 10%,

Triton 0.5 %, DTT 1 mM, 1x Complete protease inhibitor [Roche], Benzo nuclease) and incubated at room temperature for 15 mins (for nuclease activity). Equal volume of hypertonic lysis buffer (HEPES 20 mM, NaCl2 500 mM, MgCl2 1 mM, Glycerol 10%, DTT 1 mM, 1x Complete protease inhibitor) was applied to the lysate. Cells are further broken down by passage through gauge 25 needle for 10 strokes and rotated at 4°C for 30 min. After centrifugation at 14,000 x g, 15 min, 450 µl lysate were incubated with 5 ug a-TWIST1/ mouse IgG antibody at 4°C overnight with rotation. Meanwhile, Dynabead® protein G slurry (Invitrogen™, cat. 10003D) was prepared by washing and blocking for 1 hr at RT in 1% BSA 0.1% triton in PBS before use. 50 µl Dynabead was used for each treatment to capture antibody-protein complexes by rotating at RT for 30 mins. Beads were washed in 500 µl ice-cold wash buffer (1:1 mixture of the lysis buffer, without nuclease) for 6 times, and transferred to new 1.5 ml tube before the last wash. Beads were then washed quickly with cold 1 ml 50 mM Tris·HCl, pH 7.4, and then 500 µl TEAB (75 mM). Lastly, beads were collected after spinning down for 5 min at 2,000 × *g,* before processing for Mass-spectrometry analysis.

Tryptic digestion of bead-bound protein was performed by adding trypsin 1:20 (ug) of the sample directly onto the washed beads in 50mM TEAB buffer. After a brief vortex to resuspend the beads, they were incubated at 37 °C overnight. A second digest was performed by adding trypsin 1:40 to the sample the next morning, incubated for 4 more hours. The tubes were placed on a magnet to allow the beads to collect at the tube wall; the supernatant was removed and directly placed into TFA to give a final concentration of 0.5%.

Proteolytic peptides were desalted using Oligo R3 reversed phase resin (Thermo Fisher Scientific) in stage tips (Rappsilber *et al.*, 2007) made in-house. Mass spectrometry (MS) was performed using an LTQ Velos-Orbitrap MS (Thermo Fisher Scientific) coupled to an UltiMate

RSLCnano-LC system (Thermo Fisher Scientific). Raw MS data files were processed using Proteome Discoverer v.1.3 (Thermo Fisher Scientific). Processed files were searched against the UniProt mammalian database (downloaded Nov 2016) using the Mascot search engine version 2.3.0. Searches were done with tryptic specificity allowing up to two missed cleavages and a tolerance on mass measurement of 10 ppm in MS mode and 0.3 Da for MS/MS ions. Using a reversed decoy database, a false discovery rate (FDR) threshold of 1% was used. The lists of protein groups were filtered for first hits.

Rank Product analysis for MS data

We generated three sets using the non-parametric rank product method (Breitling *et al.*, 2004) based on how many times a protein was reported across independent experiments (Supplemental Table S4). If proteins were present in at least two Twist and two control samples, we deemed this set suitable for a two-class comparison, i.e. direct comparison of protein levels (set 1). To test for cases where proteins were preferentially detected (here "preferentially" is defined as present in at least 2 samples for Twist, i.e. Set 3, and 2 samples for Control, i.e. Set 2, and not detected in at least 2 samples of their respective contrast), we employed a single class test. The overall schema is represented in Fig. 2B and summarised below:

Set 1:present preferentially in IPs ($n \geq 2$)
Set 2: present preferentially in control IPs ($n \geq 2$)
Set 3: present in Twist and control cell IPs ($n \geq 2$)

P-values and FDR were then determined using the rank product method implemented in the RankProd R package (version 2.34) using 1000 permutations (Hong *et al.*, 2006). We used an FDR threshold of 0.20 in each set. If a protein for Set 1 was present in Sets 2 or 3 it was removed after FDR calculation.

Protein Immunoprecipitation

For the analysis of exogenous protein localisation, 3T3 cells were transfected 24 h before immunostaining using Lipofectamine® 3000 (Life Tech) according to manufacturer instructions with the following combinations of plasmids: pCMV-*Twist1-FLAG* plus one of (*pCMV-gfp-HA*, *pCMV-Tcf3-HA, pCMV-Prrx1-HA, pCMV-Prrx2-HA, pCMV-Chd7-HA, pCMV-Chd8-HA, pCMV-Dvl1-HA, pCMV-Smarce1-HA, pCMV-Tfe3-HA, pCMV-Whsc1-HA, pCMV-Hmg20a-HA)*

Cell pellet were thawed in 300 µl hypotonic lysis buffer (HEPES 20 mM, MgCl2 1 mM, Glycerol 10%, Triton 0.5 %, DTT 1 mM, 1x Complete protease inhibitor [Roche], Benzo nuclease) and incubated at room temperature for 15 mins nucleic acid digestion. Equal volume of hypertonic lysis buffer (HEPES 20 mM, NaCl2 500 mM, MgCl2 1 mM, Glycerol 10%, DTT 1 mM, 1x Complete protease inhibitor) was applied to the lysate. Cells are further broken down by passage through gauge 25 needle for 10 strokes and rotated at 4°C for 30 min, and centrifuged at 14,000 x g for 15 min. The supernatant was incubated with a-TWIST1/ a-FLAG antibody (2 ug/ ml) at 4°C for 2 hrs. 1/10 volume of Protein-G agarose beads (Roche) was then added and the sample rotated for 30 min at RT °C. Beads were washed with ice-cold wash buffer (1:1 mixture of the two lysis buffers) for 6 times, and transferred to new 1.5 ml tube before eluted in 2x LDS loading buffer (Life Technologies) at 70°C for 10 mins. Half the eluate was loaded on SDS-PAGE with the “input” controls for western blot analysis.

Western Blotting

Protein was extracted using RIPA buffer lysis (1× PBS, 1.5% Triton X-100, 1% IGEPAL, 0.5% Sodium Deoxycholate, 0.1% SDS, 1 mM DTT, 1x Complete protease inhibitor [Roche]) for 30

minutes at 4°C under rotation. Lysed samples were then centrifuged at 15000 g and supernatant collected. Protein concentration was determined using Direct Detect spectrometer (Millipore). 20 µg of protein per sample was denatured at 70°C for 10 mins in 1× LDS loading buffer. Protein electrophoresis and transfer was performed using NuPage system (Life Technologies, Cat. NP0322BOX), following manufacturer instructions. Membranes were revealed using SuperSignal West Pico PLUS (Thermo Scientific, Cat. #34580) as per manufacturer's protocol. Blots were imaged using ImageReader LAS 4000 (GE Healthcare Life Sciences).

Primary antibodies used were mouse monoclonal anti-TWIST1 (1:1000, abcam, Cat. #ab50887), mouse anti-α-tubulin (1:1000, Sigma, Cat. #T6199), rabbit anti-HA (1:1000, abcam, Cat. #ab9110) and mouse anti-FLAG M2 (Sigma, Cat. #F1804). Secondary antibodies used were HRP-conjugated donkey anti-Rabbit IgG (1:8000, Jackson Immunoresearch, Cat. #711-035-152) and HRP-conjugated donkey anti-Mouse IgG (1:8000, Jackson Immunoresearch, Cat. #711-035-150).

Generation of *Twist1* mutant ESC by CRISPR-Cas9

ESC editing and selection by GFP was performed as described in detail in (Sibbritt *et al.*, 2019). Briefly, 3µg of pSpCas9(BB)-2A-GFP (addgene plasmid #48138, a gift from Feng Zhang) containing the gRNA were electroporated into 1x106 A2loxCre ESCs using the Neon® Transfection System (Thermo Fisher Scientific). mESC colonies were picked by GFP expression on day 4 after electroporation.

For mESC genotyping, One-third of each dissociated clone was used to grow on gelatine coated plate for three passages, to remove residue MEFs contamination. Region surrounding

the mutation site (+/- 200 bp) was amplified from cell lysate and purified for sequencing. Clones with potential bi- or mono-allelic frameshift mutations were expanded for further validation. Targeted genomic regions were again amplified and cloned into bacteria vector to identify mutations on each allele. Isogenic ESC clones with desired mutations were used for subsequent experiments.

Generation of dimer inducible ESC line

The inducible TWIST1-dimer ESC lines was generated using the inducible cassette exchange method described previously (Mazzoni *et al.*, 2011). The protein coding sequences were cloned from the mouse embryo cDNA library into the p2lox plasmid downstream of Flag-tag (Mazzoni *et al.*, 2011). Plasmid was transfected into A2loxCre treated with 1 µg/mL doxycycline for 24 hrs. Selection was performed in 300 µg/mL of G418 (Gibco) antibiotic for one week. Colonies were then picked and tested for TWIST1 expression following doxycycline treatment.

Embryoid body formation

Embryoid bodies (EB) were maintained in EB-medium consisting of high glucose DMEM supplemented with 15% heat-inactivated FCS, 1% Non-essential amino-acid supplement and 10 mM β-mercaptoethanol. Single ES cells were seeded onto aggrewells in a 24 well plate (Stem Cell Technologies, Cat. No. 34411) at 1.5 × 10^6 cells per well in EB-medium. Cell aggregates were transferred to non-tissue culture treated 10 cm plates and EB-medium 24 hrs later, and placed on an orbital-shaker under slight rotation at 37 °C, 5% CO2, to prevent single EBs from attaching to each other or surfaces. Drug treatment conditions was optimised as shown in Supplemental Fig. S2. Doxycycline was supplied at 50 ng/ml for T3 and T12, 100

ng/ml for TT, and 1000 ng/ml for other cell lines 2 days in differentiation media. Cells were collected 16 hours later, as expression of known *TWIST1* targets peaked 12–16 hrs after induction (Supplemental Fig. S2B).

RNA-seq and data analysis

The raw RNA-seq sequence data were deposited into NCBI GEO database and can be accessed with the accession number GSE130252. Token for data access: elajseaoxnmbtwv

RNA library prep (Truseq mRNA kit) and sequencing service (Hiseq2500, 100bpPE) was provided by Macrogen (Seoul, Korea). Raw Fastq sequence files (20 Million reads per sample) were imported into Galaxy instance (https://galaxy-mel.genome.edu.au/) from Macrogen Server and analysed using the following standard pipeline outlined in Fig. 4.8 (Anders *et al.*, 2013). FastQC package was used to assess data quality and adapter sequences and low-quality reads were trimmed from the original reads (Slidingwindow method in Trimmomatic package). A minimum of 50 bp read length filter was applied. Alignment was performed using Tophat2 package, against the built-in mm10 genome. The resulting bam file was sorted and assessed visually, before read was counted by htseq-count package, using Genecode mouse basic genome assembly (GRCm38, vM10).

Read count values were imported into R software for DEG analysis (Afgan *et al.*, 2016; Supplemental Fig. S3). Genes that are not expressed at a biologically meaningful level under any conditions were discarded. A standard count per million (CPM) reads value threshold of 1 (or log-CPM value of 0) was used for defining a gene to be expressed. To be kept for downstream analysis, a gene had to be expressed in at least three samples across the entire

experiment. After filtering using this criterion, the number of genes was reduced to 14,165, approximately half the number that was started with. One replicate of T12 sample was removed from further analysis due to low library size. For EdgeR and limma-voom analysis, the data were normalised by library size before further processing.

Pre-processed data were analysed with three different R/Bioconductor packages: DESeq2, EdgeR and limma- voom (Robinson *et al.*, 2010; Love *et al.*, 2014; Ritchie *et al.*, 2015). For T12 and TT, more than 60% of total DEG were shared for all three methods (Supplemental Fig. S3A). However, when using the common adjusted P-value (adjp) threshold of 0.05, limma-voom generated the most stringent list while EdgeR gave the most permissive list (Supplemental Fig. S3). When -LOGadjp from different calculations are visualised on the scatter plot, DESeq2 and EdgeR had good correlation (R^2 = 0.79), while limma-voom tends to assign lower -LOGadjp values in pairwise comparisons (Supplemental Fig. S3B). Based on these results, DESeq2 was used for further DEG analysis, which gave intermedium-range stringency.

Two types of gene-set analysis were performed: competitive gene-set analysis using Toppgene suite in Figure 5H, and self-contained gene-set analysis using ROAST package in Table S2 and summarised in Supplemental Fig. S4 (Chen *et al.*, 2009; Wu *et al.*, 2010). In contrast to commonly used competitive gene-set analysis, the self-contained test takes gene expression into account and examines whether the genes in the set/pathway are differentially expressed as a whole for different dimers. Motif analysis on DEGs was performed using Homer package (Heinz *et al.*, 2010).

Plasmid construction

Plasmids used in the BiFC experiments were generous gifts from Hu Lab, imported via Addgene depository. The plasmid backbone and the describing publication is listed in Table S4 (Shyu *et al.*, 2006). *Twist1* and *Tcf12* were cloned from cDNA library generated from mouse head (Embryonic day 9.5), and inserted into relevant plasmids. Mouse homologous of S123A, R118H, T121E; S123E, T148A, A152P and I156N variants were generated by site-directed mutagenesis, using primers listed in Supplemental Table S5.

Bi-Fluorescent Complementation (BiFC) Experiments

BiFC experiments and quantitative analysis were performed as described previously (Hu and Kerppola, 2003). The day before transfection, C3H10T1/2 cells were seeded on to 0.1% gelatine coated coverslips, at a density of 1.66x 10^6 per 9.6 cm^2 (6- well). 24 hours later, transient transfection was performed using Lipofectamine 3000 (Life Technologies) according to manufacturer's instructions. Plasmid encoding TWIST1-VN173, TCF12-CrN173 were co-transfected with wildtype or mutant TWIST1-CC155 (Fig. S1A) at 1 μg each.

Cells were imaged 16-24 hours following the transfection. Before imaging, they were brought to room temperature, washed twice with PBS and fixed in 4% PFA for 10 mins. Coverslips were then washed three times in PBT (1% Tween 20 in PBS), once in 1% Trition X-100 in PBS for 5 minutes, and three times in PBT. This was followed by incubation with 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich; 1:10000) for 10 mins at room temperature. Coverslips were washed three times in PBT for 5-minute incubation in dark conditions, mounted using Fluoromount-G (Invitrogen) and stored at 4°C.

All immunofluorescence slides were imaged using Zeiss Axio Imager A1 (Carl Zeiss, Australia). TWIST1*-CC155/ TCF12-CrN173 heterodimers were visualized using a CFP filter set. A GFP

filter set was used for detection of TWIST1*-CC155/ TWIST1-VN173 homodimers and DAPI was imaged at 405 nm. For BiFC quantification, images were taken across the slides with randomised pattern, at 20 x magnification. At least 1000 cells from each treatment (based on DAPI counting) were sampled across three biological replicates. Fluorescence intensity from the nucleus (DAPI stained) was quantified using ImageJ software and analysed in R (codes are available upon request).

After filtering for background fluorescent, using the same intensity threshold for all treatment groups, the ratio between CFP/ GFP was taken for each cell. And mean value of this ratio was used to compare the CC155/ CrN173 and CC155/ VN173 dimer formation efficiency between wildtype or mutant TWIST1. Non-parametric Mann-Whitney U Test was used for statistical analysis. TWIST1* represents mouse homologous of wildtype, S123A, R118H, T121E; S123E, T148A, A152P or I156N variants.

Scratch Assay

TWIST1*-CC155 (1 ug) and TCF12-CrN173 (1 ug) was co-transfected into C3H10T1/2 cells in 6-well pates. 24 hours after transfection, when cells had reached confluency, a p200 pipet tip was used to make a scratch on the cell monolayer. Plate was washed once with cell culture medium to remove floating cells after the scratch. Meanwhile Cell Observer Widefield microscope (ZEISS international) was equilibrated to normal cell culture conditions (37°C, 5% CO2). After plate was installed, bright-field images were taken at pre-set tile regions covering the scratch every 15 mins over a 12 hours' period. ImageJ was used to process the image data. Essentially, the original image was enhanced by subtracting background and binary images were created to assist quantification of the gap closure. Total migration area from start of

imaging until the time point when the first cell line closed gap was calculated. One-way ANOVA followed by wo-tailed t-test was used to determine the significance of difference between groups.

Web Resources

OMIM http://www.omim.org/

Galaxy Australia https://usegalaxy.org.au/

ToppGene Suite https://toppgene.cchmc.org/

The Human Gene Mutation Database (HGMD®) http://www.hgmd.cf.ac.uk/

AQUARIA: http://aquaria.ws

Acknowledgement

Imaging analysis was performed in the ACRF Telomere Analysis Centre and proteomics analysis was performed in the Biomedical Proteomics Facility, both supported by the Australian Cancer Research Foundation. Our work was supported by the Australian Research Council (DP 1094008, DP 160100933) and Mr James Fairfax (Bridgestar Pty Ltd). XF was supported by the University of Sydney International Postgraduate Research Scholarship, Australian Postgraduate Award and CMRI Scholarship. PPLT is a NHMRC Senior Principal Research Fellow (Grant ID 1110751).

Author contribution

P.P.L.T., N.F. and X.F. designed the project; X.F., M.D. and J.S. conducted the experiments; M.G. and D.L. provided technical assistance with proteomics and transcriptome analysis, X.F., A.W. and P.O. performed the bioinformatics analysis, X.F., A.W. and P.P.L.T. wrote the manuscript.

Declaration of interest

The authors declare no competing interests

References

Afgan, E., Baker, D., Van Den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Gruning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A. & Goecks, J. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res,* 44**,** W3-W10.

Anders, S., Mccarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W. & Robinson, M. D. 2013. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature protocols,* 8**,** 1765-1786.

Barnes, R. M. & Firulli, A. B. 2009. A twist of insight - the role of Twist-family bHLH factors in development. *The International journal of developmental biology,* 53**,** 909-924.

Belle, I. & Zhuang, Y. 2014. E proteins in lymphocyte development and lymphoid diseases. *Current topics in developmental biology,* 110**,** 153-187.

Benezra, R., Davis, R. L., Lockshon, D. & Cell, T.-D. L. 1990. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell*.

Bildsoe, H., Loebel, D. A., Jones, V. J., Chen, Y.-T. T., Behringer, R. R. & Tam, P. P. 2009. Requirement for Twist1 in frontonasal and skull vault development in the mouse embryo. *Developmental biology,* 331**,** 176-188.

Bildsoe, H., Loebel, D. A., Jones, V. J., Hor, A. C. C., Braithwaite, A. W., Chen, Y.-T. T., Behringer, R. R. & Tam, P. P. 2013. The mesenchymal architecture of the cranial mesoderm of mouse embryos is disrupted by the loss of Twist1 function. *Developmental biology,* 374**,** 295-307.

Bildsoe, H., Fan, X., Wilkie, E. E., Ashoti, A., Jones, V. J., Power, M., Qin, J., Wang, J., Tam, P. P. L. & Loebel, D. a. F. 2016. Transcriptional targets of TWIST1 in the cranial mesoderm regulate cell-matrix interactions and mesenchyme maintenance. *Developmental biology,* 418**,** 189-203.

Brault, V., Moore, R., Kutsch, S., Ishibashi, M., Rowitch, D. H., Mcmahon, A. P., Sommer, L., Boussadia, O. & Kemler, R. 2001. Inactivation of the beta-catenin gene by Wnt1-Cre-mediated deletion results in dramatic brain malformation and failure of craniofacial development. *Development (Cambridge, England),* 128**,** 1253-1264.

Breitling, R., Armengaud, P., Amtmann, A. & Letters, H.-P. 2004. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS letters*.

Carver, E. A., Oram, K. F. & Gridley, T. 2002. Craniosynostosis in Twist heterozygous mice: a model for Saethre-Chotzen syndrome. *The Anatomical record,* 268**,** 90-92.

Castanon, I. & Baylies, M. K. 2002. A Twist in fate: evolutionary comparison of Twist structure and function. *Gene,* 287**,** 11-22.

Centonze, V. E., Firulli, B. A. & Firulli, A. B. 2004. Fluorescence resonance energy transfer (FRET) as a method to calculate the dimerization strength of basic Helix-Loop-Helix (bHLH) proteins. *Biological Procedures Online,* 6**,** 78-82.

Chang, A. T., Liu, Y., Ayyanathan, K., Benner, C., Jiang, Y., Prokop, J. W., Paz, H., Wang, D., Li, H.-R. R., Fu, X.-D. D., Rauscher, F. J. & Yang, J. 2015. An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors. *Genes & development,* 29**,** 603-616.

Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res,* 37**,** W305-11.

Chen, Z. F. & Behringer, R. R. 1995. twist is required in head mesenchyme for cranial neural tube morphogenesis. *Genes & development,* 9**,** 686-699.

Comai, G. & Tajbakhsh, S. 2014. Molecular and cellular regulation of skeletal myogenesis. *Current topics in developmental biology,* 110**,** 1-73.

Connerney, J., Andreeva, V., Leshem, Y., Muentener, C., Mercado, M. A. & Spicer, D. B. 2006. Twist1 dimer selection regulates cranial suture patterning and fusion. *Developmental dynamics : an official publication of the American Association of Anatomists,* 235**,** 1345-1357.

Connerney, J., Andreeva, V., Leshem, Y., Mercado, M. A., Dowell, K., Yang, X., Lindner, V., Friesel, R. E. & Spicer, D. B. 2008. Twist1 homodimers enhance FGF responsiveness of the cranial sutures and promote suture closure. *Developmental biology,* 318**,** 323-334.

Dennis, D. J., Han, S. & Schuurmans, C. 2019. bHLH transcription factors in neural development, disease, and reprogramming. *Brain research,* 1705**,** 48-65.

Evseenko, D., Zhu, Y., Schenke-Layland, K., Kuo, J., Latour, B., Ge, S., Scholes, J., Dravid, G., Li, X., Maclellan, W. R. & Crooks, G. M. 2010. Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of*

America, 107**,** 13742-13747.

Fan, X., Loebel, D. a. F., Bildsoe, H., Wilkie, E. E., Qin, J., Wang, J. & Tam, P. P. L. 2016. Tissue interactions, cell signaling and transcriptional control in the cranial mesoderm during craniofacial development. *AIMS Genetics,,* 3**,** 74.

Firulli, A. B. & Conway, S. J. 2008. Phosphoregulation of Twist1 provides a mechanism of cell fate control. *Current medicinal chemistry,* 15**,** 2641-2647.

Firulli, B. A., Krawchuk, D., Centonze, V. E., Vargesson, N., Virshup, D. M., Conway, S. J., Cserjesi, P., Laufer, E. & Firulli, A. B. 2005. Altered Twist1 and Hand2 dimerization is associated with Saethre-Chotzen syndrome and limb abnormalities. *Nature genetics,* 37**,** 373-381.

Firulli, B. A., Redick, B. A., Conway, S. J. & Firulli, A. B. 2007. Mutations within Helix I of Twist1 Result in Distinct Limb Defects and Variation of DNA Binding Affinities. *Journal of Biological Chemistry,* 282**,** 27536-27546.

Forrest, M. P., Hill, M. J., Quantock, A. J., Martin-Rendon, E. & Blake, D. J. 2014. The emerging roles of TCF4 in disease and development. *Trends in molecular medicine,* 20**,** 322-331.

Ghouzzi, E. V., Legeai-Mallet, L. & Molecular …, A.-S. 2000. Saethre–Chotzen mutations cause TWIST protein degradation or impaired nuclear location. *Human molecular ….*

Guenou, H., Kaabeche, K., Mée, S. L. & Marie, P. J. 2005. A role for fibroblast growth factor receptor-2 in the altered osteoblast phenotype induced by Twist haploinsufficiency in the Saethre-Chotzen syndrome. *Human molecular genetics,* 14**,** 1429-1439.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. & Glass, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell,* 38**,** 576-589.

Hong, F., Breitling, R., Mcentee, C. W., Wittner, B. S., Nemhauser, J. L. & Chory, J. 2006. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics,* 22**,** 2825-7.

Hu, C.-D. & Kerppola, T. K. 2003. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nature Biotechnology,* 21**,** 539-545.

Iacovino, M., Roth, M. E. & Kyba, M. 2014. Methods in Molecular Biology. *Methods in molecular biology (Clifton, N.J.),* 1101**,** 339-351.

Lee, J. E., Hollenberg, S. M., Snider, L., Turner, D. L., Lipnick, N. & Weintraub, H. 1995. Conversion of Xenopus ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. *Science (New York, N.Y.),* 268**,** 836-844.

Love, M. I., Huber, W. & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

Massari, M. E. & Murre, C. 2000. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Molecular and cellular biology,* 20**,** 429-440.

Mazzoni, E. O., Mahony, S., Iacovino, M., Morrison, C. A., Mountoufaris, G., Closser, M., Whyte, W. A., Young, R. A., Kyba, M., Gifford, D. K. & Wichterle, H. 2011. Embryonic stem cell-based mapping of developmental transcriptional programs. *Nat Methods,* 8**,** 1056-8.

Miraoui, H., Severe, N., Vaudin, P., Pagès, J.-C. C. & Marie, P. J. 2010. Molecular silencing of Twist1 enhances osteogenic differentiation of murine mesenchymal stem cells: implication of FGFR2 signaling. *Journal of cellular biochemistry,* 110**,** 1147-1154.

Murakami, M., Kataoka, K., Tominaga, J., Nakagawa, O. & Kurihara, H. 2004. Differential cooperation between dHAND and three different E-proteins. *Biochem Biophys Res Commun,* 323**,** 168-74.

Nakaya, Y. & Sheng, G. 2008. Epithelial to mesenchymal transition during gastrulation: an embryological view. *Development, growth & differentiation,* 50**,** 755-766.

Ota, M. S., Loebel, D., O'rourke, M. P., Wong, N., Tsoi, B. & Tam, P. 2004. Twist is required for patterning the cranial nerves and maintaining the viability of mesodermal cells. *Developmental Dynamics,* 230**,** 216-228.

Pfurr, S., Chu, Y.-H. H., Bohrer, C., Greulich, F., Beattie, R., Mammadzada, K., Hils, M., Arnold, S. J., Taylor, V., Schachtrup, K., Uhlenhaut, N. H. & Schachtrup, C. 2017. The E2A splice variant E47 regulates the differentiation of projection neurons via p57(KIP2) during cortical development. *Development (Cambridge, England),* 144**,** 3917-3931.

Price, F. D., Yin, H., Jones, A., Van Ijcken, W., Grosveld, F. & Rudnicki, M. A. 2013. Canonical Wnt Signaling Induces a Primitive Endoderm Metastable State in Mouse Embryonic Stem Cells. *STEM CELLS,* 31**,** 752-764.

Qin, Q., Xu, Y., He, T., Qin, C. & Xu, J. 2012. Normal and disease-related biological functions of Twist1 and underlying molecular mechanisms. *Cell research,* 22**,** 90-106.

Rappsilber, J., Mann, M. & Ishihama, Y. 2007. Protocol for micro-purification, enrichment, pre-fractionation and

storage of peptides for proteomics using StageTips. *Nature protocols,* 2**,** 1896-1906.

Rice, D. P. 2005. Craniofacial anomalies: from development to molecular pathogenesis. *Current molecular medicine,* 5**,** 699-722.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. & Smyth, G. K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res,* 43**,** e47.

Robinson, M. D., Mccarthy, D. J. & Smyth, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England),* 26**,** 139-140.

Schlaeger, T. M., Schuh, A., Flitter, S., Fisher, A., Mikkola, H., Orkin, S. H., Vyas, P. & Porcher, C. 2004. Decoding hematopoietic specificity in the helix-loop-helix domain of the transcription factor SCL/Tal-1. *Molecular and cellular biology,* 24**,** 7491-7502.

Sharma, V. P., Fenwick, A. L. L., Brockop, M. S., Mcgowan, S. J., Goos, J. A., Hoogeboom, A. J., Brady, A. F., Jeelani, N. O., Lynch, S. A., Mulliken, J. B., Murray, D. J., Phipps, J. M., Sweeney, E., Tomkins, S. E., Wilson, L. C., Bennett, S., Cornall, R. J., Broxholme, J., Kanapin, A., Consortium, Johnson, D., Wall, S. A., Van Der Spek, P. J., Mathijssen, I. M., Maxson, R. E., Twigg, S. R. & Wilkie, A. O. 2013. Mutations in TCF12, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis. *Nature genetics,* 45**,** 304-307.

Shyu, Y. J., Liu, H., Deng, X. & Hu, C.-D. D. 2006. Identification of new fluorescent protein fragments for bimolecular fluorescence complementation analysis under physiological conditions. *BioTechniques,* 40**,** 61-66.

Sibbritt, T., Osteil, P., Fan, X., Sun, J., Salehin, N., Knowles, H., Shen, J. & Tam, P. P. L. P. L. 2019. Gene Editing of Mouse Embryonic and Epiblast Stem Cells. *Methods in molecular biology (Clifton, N.J.),* 1940**,** 77-95.

Skinner, M. K., Rawls, A., Wilson-Rawls, J. & Roalson, E. H. 2010. Basic helix-loop-helix transcription factor gene family phylogenetics and nomenclature. *Differentiation,* 80**,** 1-8.

Spicer, D. B., Rhee, J., Cheung, W. L. & Lassar, A. B. 1996. Inhibition of myogenic bHLH and MEF2 transcription factors by the bHLH protein Twist. *Science (New York, N.Y.),* 272**,** 1476-1480.

Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shiel, J. A., Thomas, N. S., Abeysinghe, S., Krawczak, M. & Cooper, D. N. 2003. Human Gene Mutation Database (HGMD): 2003 update. *Human mutation,* 21**,** 577-581.

Ten Berge, D., Koole, W., Fuerer, C., Fish, M., Eroglu, E. & Nusse, R. 2008. Wnt signaling mediates self-organization and axis formation in embryoid bodies. *Cell stem cell,* 3**,** 508-518.

Thisse, B., El Messal, M. & Perrin-Schmitt, F. 1987. The twist gene: isolation of a Drosophila zygotic gene necessary for the establishment of dorsoventral pattern. *Nucleic acids research,* 15**,** 3439-3453.

Uittenbogaard, M. & Chiaramello, A. 2002. Expression of the bHLH transcription factor Tcf12 (ME1) gene is linked to the expansion of precursor cell populations during neurogenesis. *Brain research. Gene expression patterns,* 1**,** 115-121.

Wang, J., Nikhil, K., Viccaro, K., Chang, L., Jacobsen, M., Sandusky, G. & Shah, K. 2017. Aurora A-Twist1 axis promotes highly aggressive phenotypes in pancreatic carcinoma. *J Cell Sci,* 130.

Wang, L.-H. H. & Baker, N. E. 2015. E Proteins and ID Proteins: Helix-Loop-Helix Partners in Development and Disease. *Developmental cell,* 35**,** 269-280.

Wu, D., Lim, E., Vaillant, F., Asselin-Labat, M.-L. L., Visvader, J. E. & Smyth, G. K. 2010. ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics (Oxford, England),* 26**,** 2176-2182.

Xue, G., Restuccia, D. F., Lan, Q., Hynx, D., Dirnhofer, S., Hess, D., Ruegg, C. & Hemmings, B. A. 2012. Akt/PKB-mediated phosphorylation of Twist1 promotes tumor metastasis via mediating cross-talk between PI3K/Akt and TGF-beta signaling axes. *Cancer Discov,* 2**,** 248-59.

Fan_Fig1

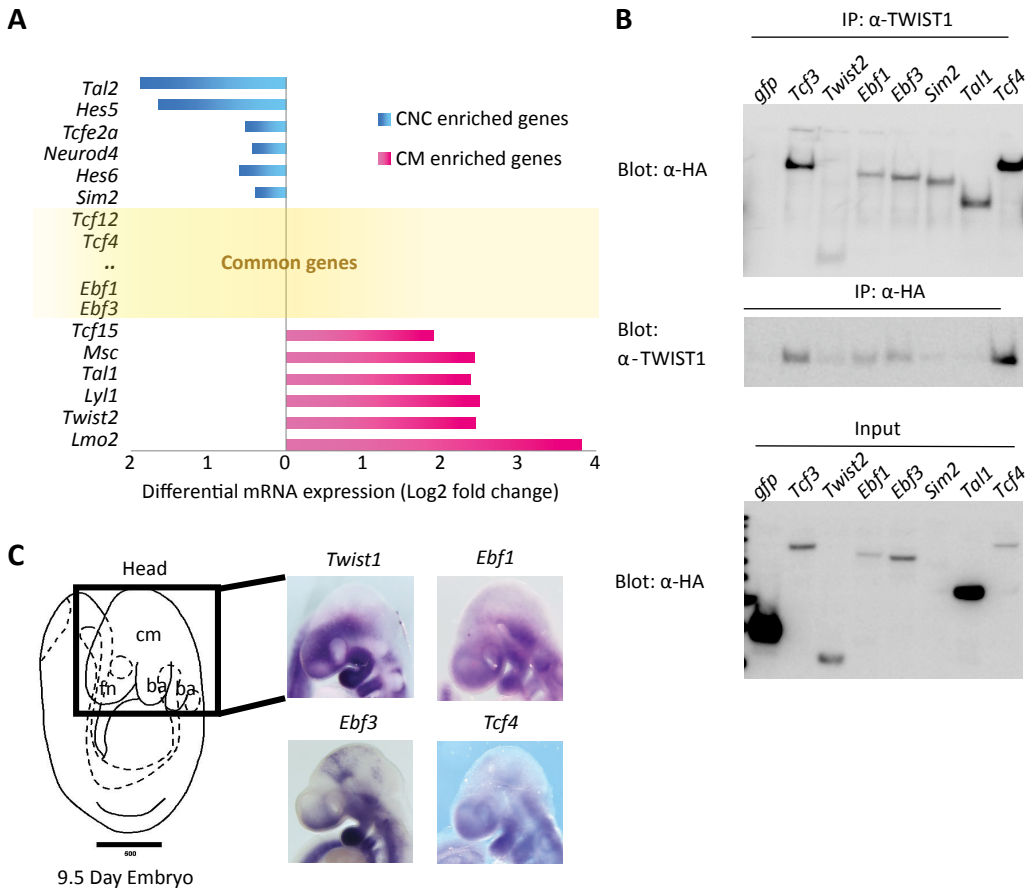

Figure 1. Identification of TWIST1 bHLH interaction partners in mouse embryonic head tissue
A. Differentially expressed bHLH factors in cranial neural cells (CNC) and cranial mesoderm (CM) of E9.5 embryos. Data from microarray analysis (Fan et al. 2016). **B.** Reciprocal co-IP in hTWIST1-expressing MDCK cell line that was transfected with constructs of individual HA-tagged candidate partner. Immunoprecipitation was performed using α-TWIST1 or α-HA. Construct for gfp used as control. **C.** Overlapping expression regions of three positive candidates and Twist1 in E9.5 mouse head tissues, revealed by whole mount in situ hybridization. cm = cranial mesoderm, fn = frontonasal tissue, ba = brachial arch. Bar = 500 μm.

Fan_Fig2

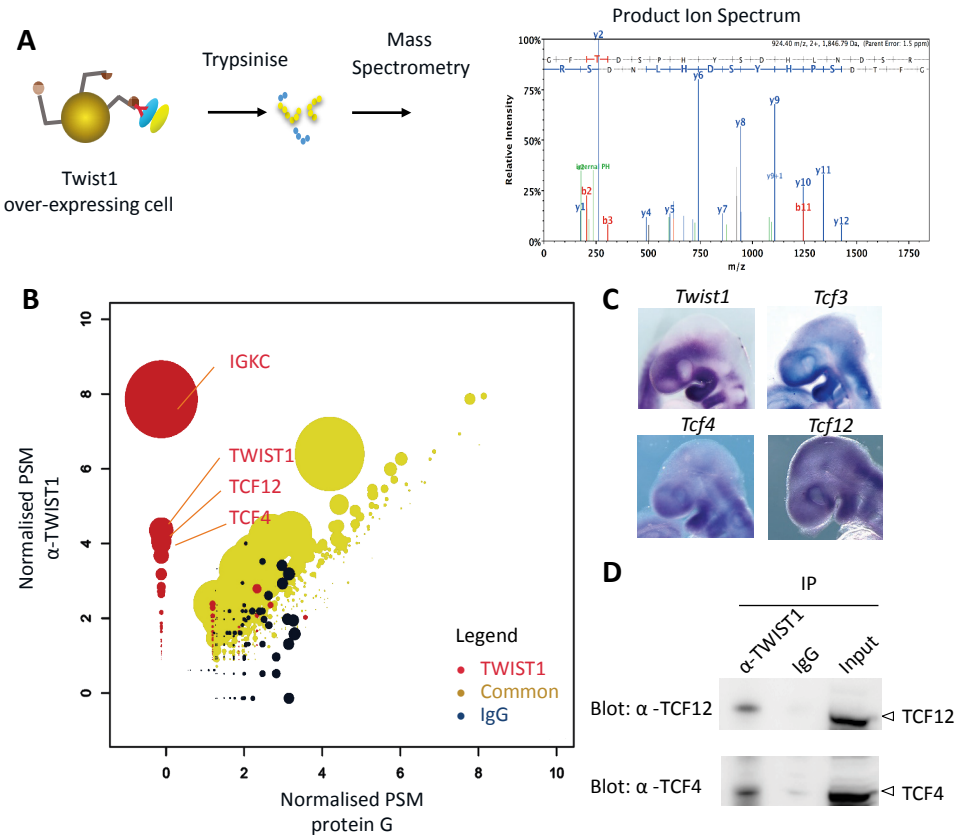

Figure 2. Identification of proteins co-precipitated with TWIST1

A. Experimental strategy for identifying TWIST1 interacting partners by mass-spectrometry. **B.** Groups of proteins identified using the non-parametric rank product method, based on cumulative count of reporting a protein across independent experiments: Sets: red = TWIST1 IP; blue = IgG IP control; yellow = common to α-TWIST1 IP and control. Normalised unique peptide spectrum match (PSM) for proteins are shown on the axis. The size of the dots corresponds to –log (10) P-value. C. Overlapping expression of the transcripts of *Tcf3, Tcf4* (from Fig. 1C) and *Tcf12* with *Twist1* (from Fig. 1C) in the craniofacial tissues of E9.5 mouse embryo. **D.** Immunoprecipitation (IP) of TWIST1 using α-TWIST1 from embryoid bodies over-expressing TWIST1 and co-IP candidate factors detected using α-TCF4 and α-TCF12.

Fan_Fig3

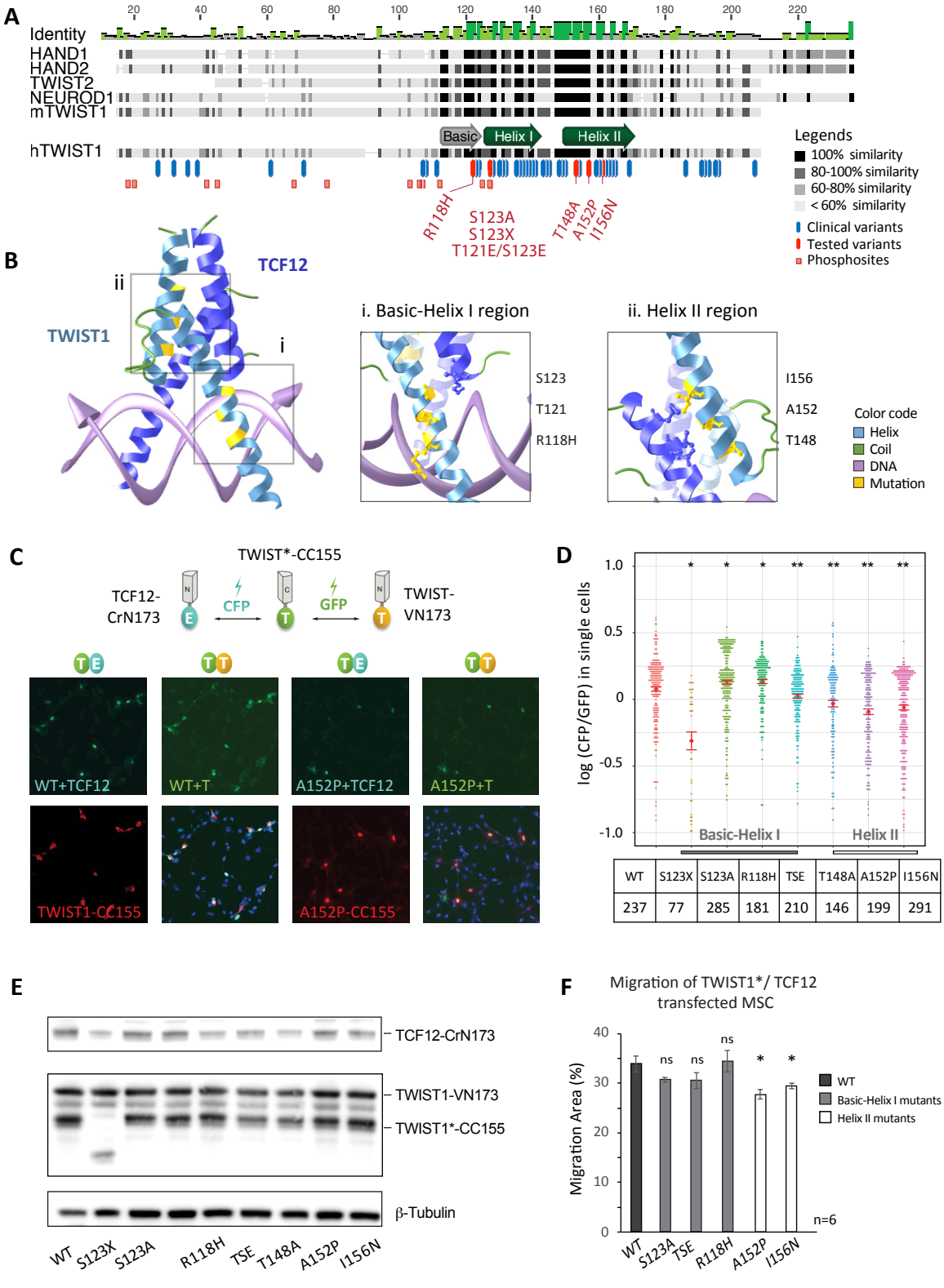

Figure 3. Human SCS mutations disrupt TWIST1-TCF12 dimer function

A. Alignment of sequences between human bHLH proteins and mouse mTWIST1 protein. Homologous sequences are indicated in dark shades or green regions. Blue = 63 missense pathogenic mutations in human gene mutation database (Stenson et al. 2003); red = mutations for this study. square dots = Predicted phosphorylation sites (bottom row). **B.** TWIST1-TCF12 dimer mapped onto NeuroD-TCF3 crystography, over the bHLH region (http://aquaria.ws) (O'Donoghue et al. 2015). Chain colours represent TWIST1 or E-protein molecules as labelled. Zoomed representation of i. basic-Helix I and ii. Helix II showing molecular structure of TWIST1 residues mutated in disease (yellow) and their atomic interaction with E-protein partner. **C.** Experimental setting and imaging result showing TWIST1*-CC155 interaction with TWIST1-VN173 or TCF12-CrN173 partners that compete each other. TWIST1*=wildtype (WT) or mouse equivalent of SCS variants. **D.** Quantified imaging data of heterodimer/homodimer CFP/GFP ratio for TWIST1 WT and pathogenic variants. TSE= T121E;S123E. Count = number of cells for quantification. Bar = standard errors. Data analysed by non-parametric Mann-Whitney U test; p-values. * = $P < 0.05$, ** = $P < 0.001$. E. Western blot analysis of recombinant protein expression in BiFC experiment. F. Quantified result of images measuring the migration (% total area) of TWIST1*-CC155 and TCF12-CrN173 transfected C3H10T1/2 cells over 12 hours in scratch assay in 6 replicate experiments. Bar = standard errors. Data analysed by one-way ANOVA followed by two-tailed t-test; p-values * = $P < 0.05$.

Fan_Fig4

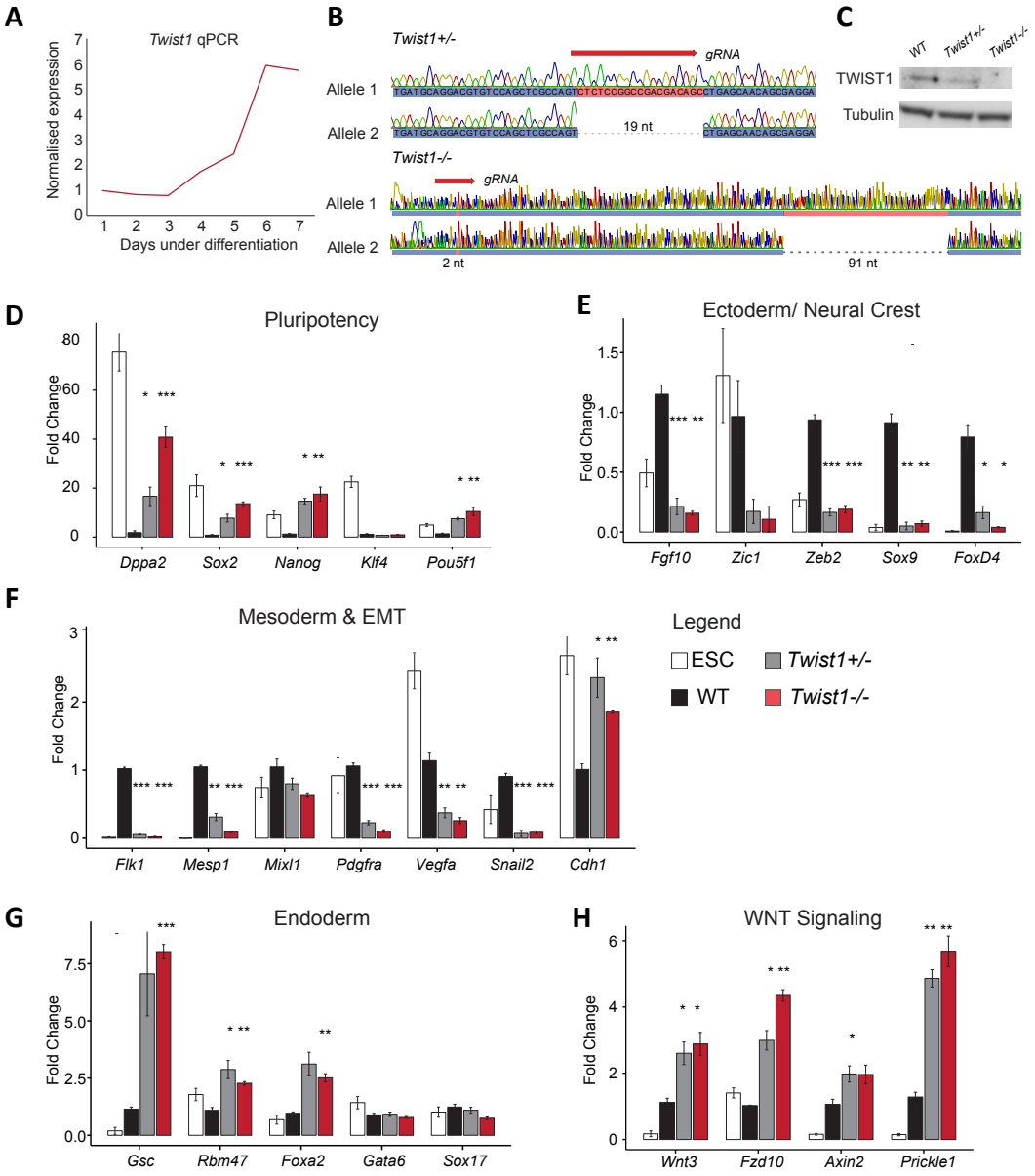

Figure 4. Twist1 is required for mesoderm and ectoderm differentiation. **A.** qPCR result of Twist1 expression during ESC differentiation. Mean expression value normalized against *Gapdh* from n=3 were shown. Twist1 CRISPR KO clones (*Twist1+/-* and *Twist1+/-*) were validated by **B.** genotyping and **C.** western blot. Frame-shift mutations were found around gRNA targeted nucleotides 29-47 in exon 1. Mutant and wildtype clones were subjected to serum-induced differentiation for 3 days, before RNA was collected and quantified by RT-qPCR. Transcript for stem cell and germ layer markers were amplified using specific primers (**D-H**). Target gene expressions were normalized against the mean value of β*Actin* and *Gapdh*, and then the wildtype sample (WT). Student t-test were performed between *Twist1* mutants and the wildtype cell line. n=3; *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001; bar= standard error.

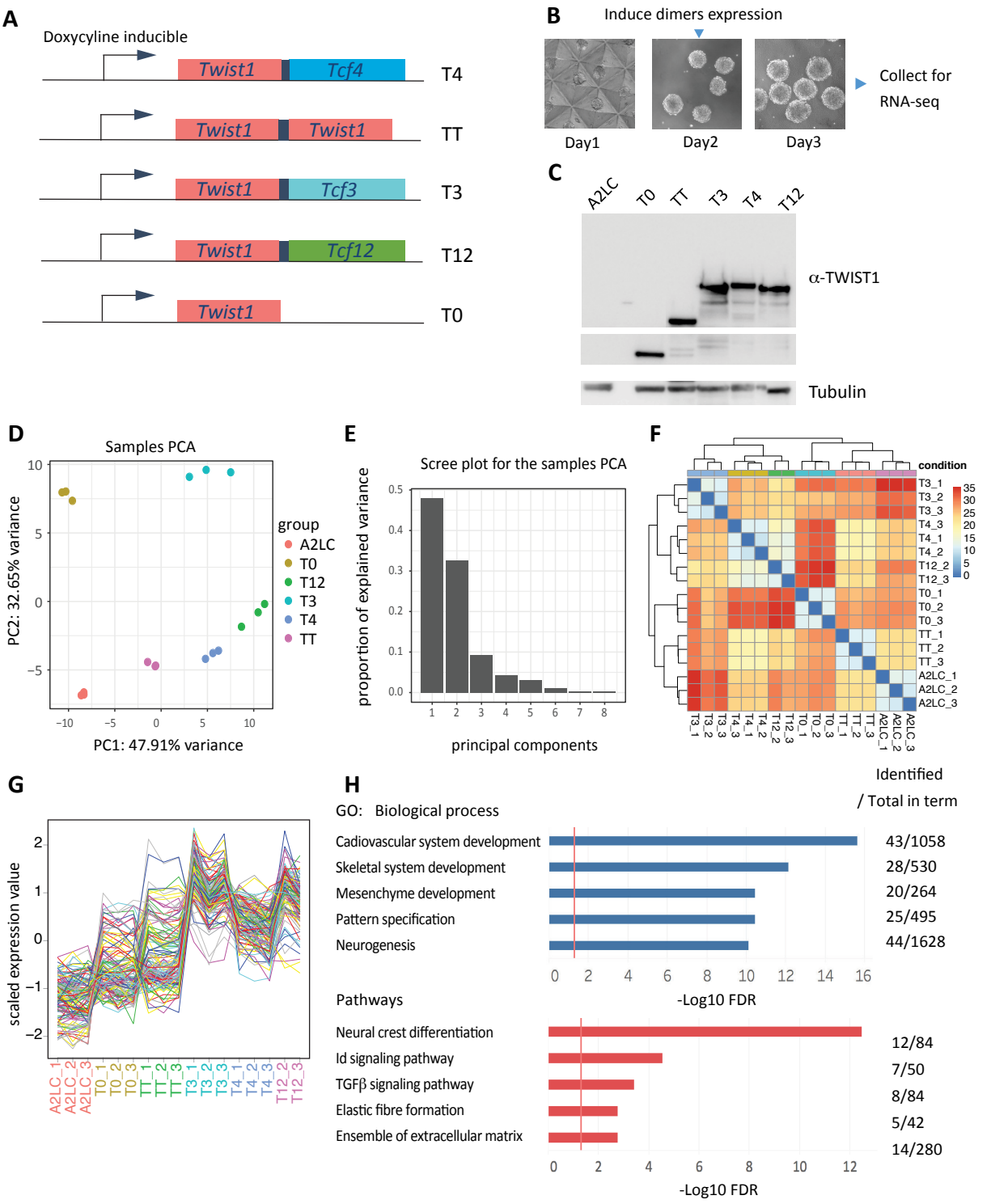

Figure 5. Comparison of transcriptional profiles of differentiating embryonic stem cells expressing TWIST1 dimers. **A.** Tethered dimers generated by joining TWIST1 with itself or E-Proteins by a flexible poly-glycine linker. Doxycycline-inducible expression constructs: Twist1-Tcf3 (T3), Twist1-Tcf4 (T4), Twist1-Tcf12 (T12), Twist1-Twist1 (TT) and Twist1 only (T0 monomer). Control: unmodified parental A2loxCre cells. **B.** Differentiation of embryonic stem cells into embryoid bodies (EBs), collected on Day 3 for RNA sequencing analysis. **C.** Western blot analysis of dimer expression in EBs harvested for RNA-seq analysis, showing comparable levels of protein expression across the experiments. **D.** Principle component analysis of the dataset showing first two primary components. **E.** Scree plot showing the proportion of explained variance for each component. **F.** Correlation map of six groups of cells based on RNA-seq data. **G.** Scaled gene expression trends of top heterodimer enriched genes. **H.** Ontology of heterodimer enriched genes.

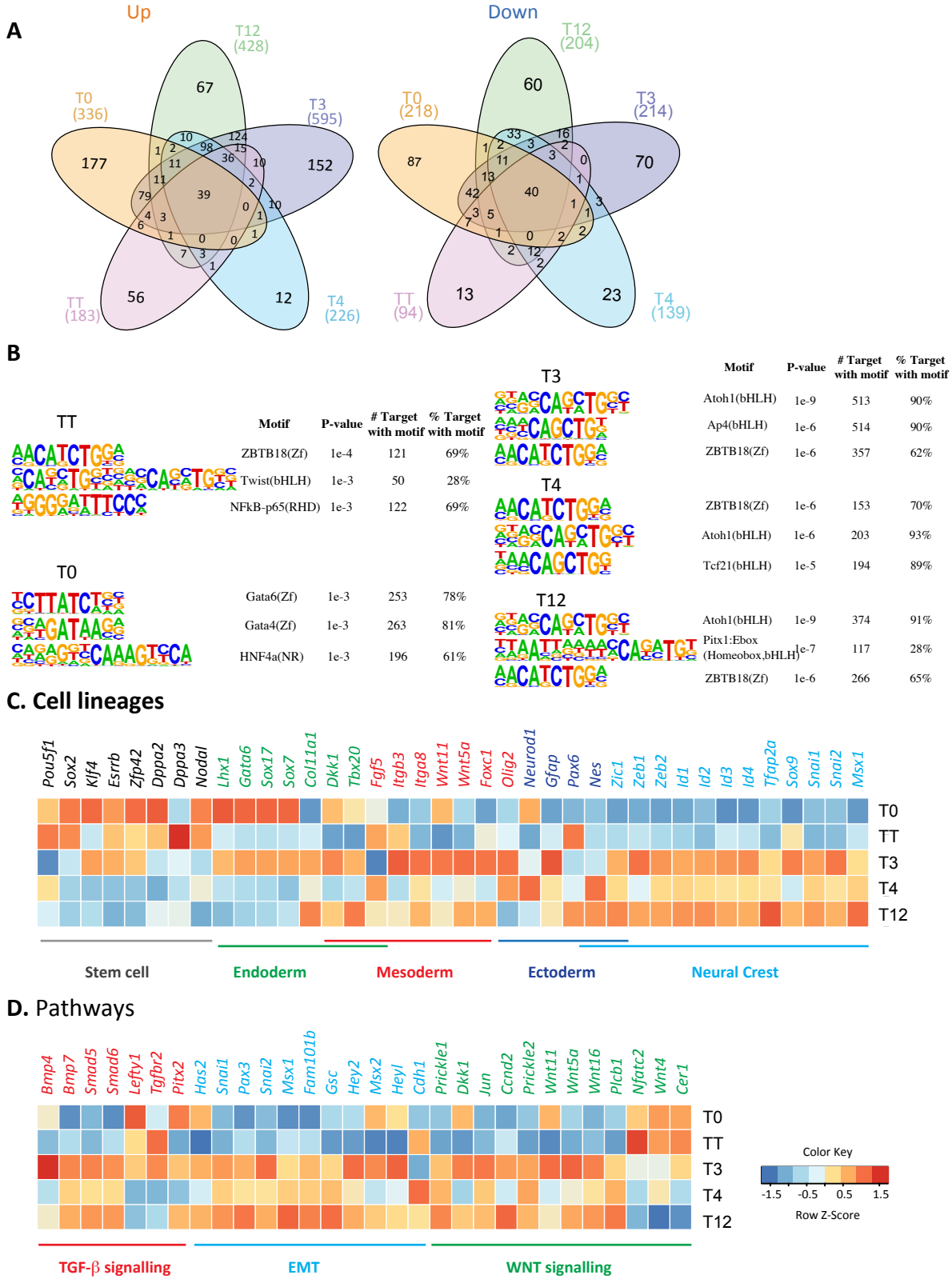

Figure 6. Differential specificity of TWIST1-dimers. **A.** Venn diagrams of pairwise DEG analysis against A2LoxCre generated in DESeq2 (adjp< 0.05, fold change >= 2). Up-regulated (Up) and down-regulated (Down) genes are shown separately. **B.** Top three binding motifs of TWIST1 dimers and monomer, based on Homer analysis of differentially regulated target sequences. **C.** Heatmap showing relative expression of lineage genes among dimer expressing EBs, based on Log2-fold change against A2loxCre EBs. The colour scale for Z-score is shown below. **D.** Heatmap comparing expression of TGF signaling, EMT and WNT signaling pathway genes among the EBs.

Fan_Fig7

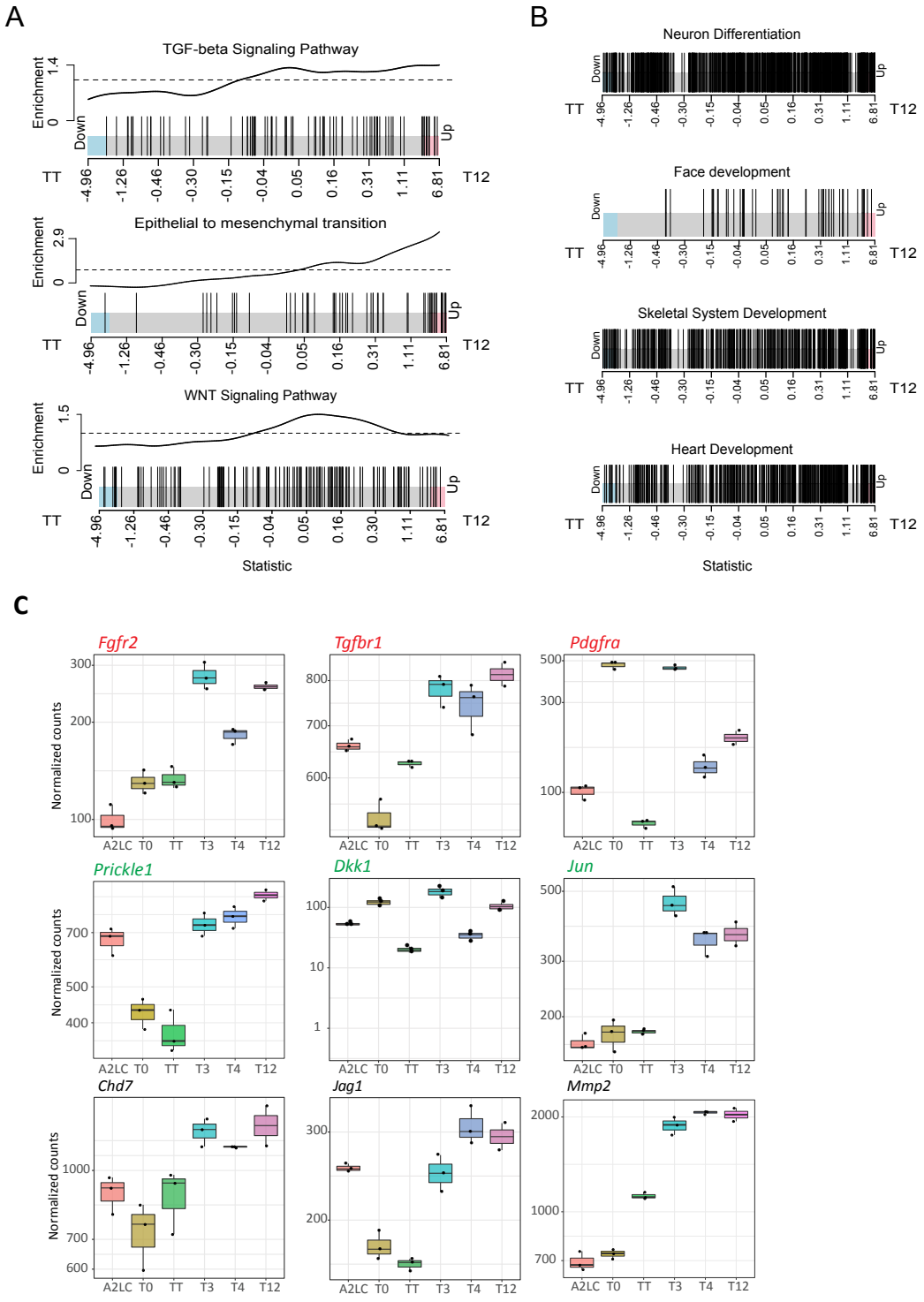

Figure 7. Comparison of gene-set enrichment between TWIST1-TCF12 and TWIST1 homodimer dimers
A. Barcode plots of the enrichment of TGF-β, EMT ,and WNT related genes in TWIST1-TCF12 versus TWIST1-TWIST1. In the plot, genes are ranked from left to right by increasing wald statistics (log-fold change/ Standard error), and individual genes in the gene-set are represented by vertical bars. The enrichment line shows the relative local enrichment of the bars in the plot. The dotted horizontal line indicates neutral enrichment; values above the dotted line shows enrichment while values below the dotted line shows depletion. **B.** Developmental process-related genes in T12-expressing EBs compared to TT group. **C.** Examples of genes in craniofacial development and pathways, enriched in T12 dimer expressing cells relative TT. Red= TGFβ signalling; Green= WNT signalling